Reviews (Round 2) of

Do convection-permitting regional climate models have added value for hydroclimatic simulations? A test case over small and medium-sized catchments in Germany

By Oakley Wagner, Verena Maleska, and Laurens M. Bouwer

Review 2 (RC2)

Comments

I would like to thank the authors for their careful reading of the different reviews and the amount of work they put into replying to comments and editing the document.

The hydrological section has been positively improved and clarifies reviewers' comments by better understanding the experiment and simulated processes. The significant biases of extreme precipitation in the ICON3km simulation have been clearly described. The explanation of the authors regarding these biases and their transfer on hydrological modeling is convincing with fair analyses and discussion. The manuscript quality has been greatly improved and can be addressed for publication after considering some modifications.

> We thank the reviewer for the positive comments. We appreciate the time taken to assess our paper.

1) In their response, the authors emphasised the focus of the hydrological aspect of this study on summer convective events. This is stated line 185, but I think this should be clearly stated at the end of the introduction when describing the aim of the study.

> We thank the reviewer for pointing this out. In fact, in the referred line 185 we discuss our methodology and justify why snow melt is not investigated in our study. Our methodology is built upon prior findings, which show that CPRCMs are unlikely to offer their main added value in winter, when large synoptic weather systems dominate, as these are already well represented by the driving RCMs (Strandberg & Lind, 2021). It is in summer, when convective processes are at the forefront, that CPRCMs and RCMs are likely to differ most strongly in their results (Strandberg & Lind, 2021), which is of interest in this study. We go on to confirm this, e.g. for precipitation (see Fig. S14) and discharge (see Fig. 13). Our findings thereby support that summer is indeed the period of interest and that our choice to not investigate winter processes more in depth is indeed justified.

2) line 191: We understand after the author's response and Fig 14 that 8hours is the value chosen for computing the antecedent rainfall. Why this value ? In Ross et al. (2021), the minimum antecedent rainfall value to compute their analysis is 1 day. Other studies from neighbouring countries have chosen values from 1 to several days (Froidevaux et al., 2015; Staudinger et al., 2025). 8h seems to be too short a window to capture rainfall phenomena preceding the rainfall-runoff event, given the geographical context of the basins. Could you adapt the analysis of the

non-linearity of the rainfall-runoff relationship with a value of 1 to several days and see if this changes the results?

> Many thanks to the reviewer for catching what turns out to be a typo. A check of the code showed a chosen antecedent rainfall window of 8 days, not 8 hours. Indeed, the latter would be too short, reason why we opted for a longer time window. Our methodology is hence in line with the reasoning stated above and only a correction of the text is needed. We have changed „h“ to „days“ in line 440 of the new version and in the caption of Fig. 14.

3) Section 3.1.5 - Precipitation. I thank the author for better highlighting the strong overestimation of ICON3km precipitation in this section from line 295 compared to the first version of the manuscript. However, I sincerely believe the choice of the percentile for hourly precipitation could better highlight the strong biases on the precipitation extremes through the medians of values exceeding this percentile.

In your analysis, the point of studying heavy rainfall is to highlight the potential biases of climate models and their propagation in the hydrological reaction these "extremes" can cause, such as flood events.

The chosen percentile leads to medians of around 3 to 4mm used to compare model biases on extreme rainfall. This order of magnitude of values is problematic for two reasons:

- It does not show the truly significant positive bias of ICON3km, as can be seen in Figure 6.
- In terms of hydrological consistency, if the objective of this assessment is to compare the biases of ICON3km and ICON11km on extreme precipitation, 2 to 3 mm is not sufficient to produce hydrological responses to short events, and this comparison cannot be informative for drawing conclusions about high flows or floods, knowing that the focus on the paper is the summer convective events.

I recommend reconsidering this percentile threshold in order to compare extreme precipitation values that are meaningful in terms of potential hydrological response representative of summer convective events (99.9th or 99.95th percentile of all hours or 99th or 99.5th percentile of wet hours). You will need to adapt Fig7 accordingly. This will strengthen this part of analysis and give more credit to the evaluation.

> We agree with the reviewer on the need to better highlight the highest precipitation quantiles. The 99.95th-percentile of all hours is at 11.7 mm for ICON3km, 7.3 mm for RADOLAN3km, 4.7 mm for ICON12km and 6.6 mm for RADOLAN12km. For comparison of the frequency distributions, a consideration over a same range of intensities would be preferable. This is challenged by the given large spread of the mentioned quantile thresholds. Furthermore, cropping the quantiles to the bins that they fully cover does not imply that the resulting subsets have an equal number of time steps, which entails a potentially misleading comparison between their frequency density polygons. Considering alternatively data above a common absolute threshold, e.g. for data of the same resolution, would introduce the same issue of a differing number of time steps between the resulting subsets. The same holds true for a consideration of wet-hour frequency, as the number of time steps considered wet differs

between the data groups. We have hence chosen to work with all time steps in the data series and to use histograms showing the absolute frequencies to keep the highest quantiles visible against those with much greater frequencies. Given the strong overestimation of precipitation intensity by ICON3km, its comparison with RADOLAN3km requires a frequency analysis over a wide range of intensities. While bins of high intensity register only a few counts, those grouping lower intensities have considerably high frequencies, making a relative or linear representation unfavourable. This can be solved through a logarithmic scale (see Fig. 1 below).

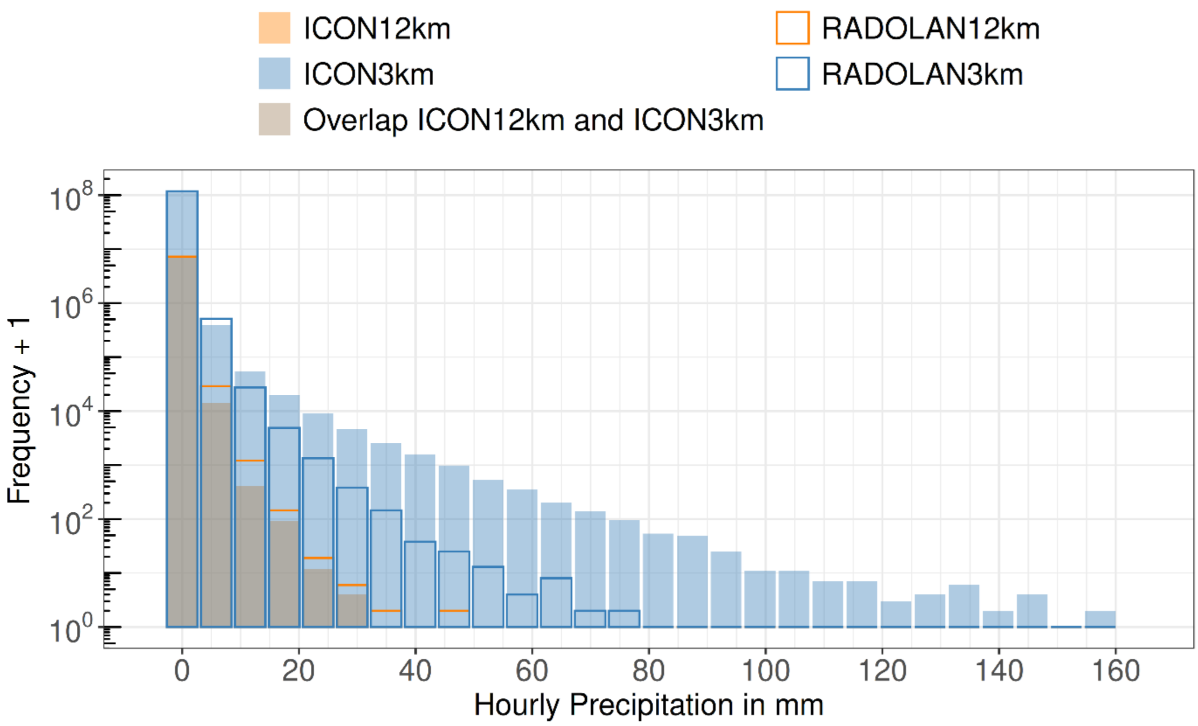

Fig. 1: Histograms for ICON12km, ICON3km, RADOLAN12km and RADOLAN3km for hourly precipitation for the period from 2005 to 2014. Note the logarithmic scale of the vertical axis.

We consequently suggest the following modifications to the manuscript:

- Increase the quantile used for the general performance analysis of the climate models (before: 99.5th percentile, now: 99.9th percentile)

  L. 288 - 291 now state: "With median hourly precipitation of 4.6 mm over the 99.9th percentile (left-closed interval), ICON12km shows an underestimation compared to adjusted radar data upscaled to the same grid (RADOLAN12km) with a respective estimate of 6.6 mm. In turn, ICON3km overestimates the median of the 99.9th percentile of hourly precipitation compared to upscaled adjusted radar estimates (RADOLAN3km) with 11.6 mm to 7.3 mm respectively."
- Substitute the frequency density polygons by histograms showing the absolute frequencies on a logarithmic scale.

  L. 310 - 315 now state: "The histograms for hourly precipitation are shown in Fig. 7. ICON12km overestimates the frequency of time steps of 0 to 2.9 mm of precipitation (with 7,259,900 counts compared to 7,244,504 counts in

RADOLAN12km; bias by ICON12km of 0.21%; not differentiable in the plot). The overestimation seen in ICON12km has been improved through ICON3km, with ICON3km giving similar counts as RADOLAN3km for precipitation of 0 to 2.9 mm (with 118,011,717 counts and 117,954,139 counts respectively; bias by ICON3km of 0.05%). The frequency of heavy precipitation on the other hand is consistently underestimated by ICON12km, but overestimated by ICON3km."

4) l309 : Change "discussed" to "shown". Results are discussed later on in the discussion section.

We support the suggestion and have adapted it in the revised version of the manuscript.

5) l333 : Something is wrong with the statement "2.5-times as high for ICON3km than for ICON11km (1.0 mm to 0.4 mm)". It gives the reader the feeling of a consistent positive bias for ICON3km for all summer. When looking at Fig S14, we can observe that ICON3km systematically simulates higher daily catchment average precipitation from April to October, but not necessarily stronger biases in summer. July is only a particular case of a bias "2.5 times as high for ICON3km than for ICON11km", whereas the absolute bias is lower for ICON3km than ICON11km in June and August.

I can advise to replace the Fig S14 by one by seasons to illustrate your point.

We thank the author for the comment and have adapted the respective paragraph as follows:

„The monthly means of the deviations of daily catchment-average precipitation computed by ICON3km and ICON12km to the adjusted radar data estimates upscaled to their respective resolution (Fig. S14 in the appendix) are similar between the climate models during autumn (SON) and winter (DJF) but differ more strongly during mid to late spring (April by 0.3 mm, May by 0.4 mm) and during summer (June by 0.3 mm, July by 0.6 mm, August by 0.3 mm). In July, the monthly mean bias of daily catchment-average precipitation is 2.5-times as high for ICON3km than for ICON12km (1.0 mm to 0.4 mm), given the overestimation of heavy precipitation by ICON3km. Throughout all of April to August, ICON3km gives consistently higher estimates than ICON12km, thereby improving on the negative bias ICON3km shows, with an exception for July, where ICON12km already showed a positive bias, which is only further amplified by ICON3km. It should be kept in mind, that the consideration of daily and catchment-spatial averages entails a smoothing of local extreme hourly precipitation values, dampening their effect."

Figure 2 below shows the seasonal mean bias of daily catchment-average precipitation of ICON3km and ICON12km and highlights the greater difference in bias in summer. As commented, the picture is strongly shaped by the month of July during which the deviation in performance between the models was exceptionally pronounced. It shall be worth showing the months individually to highlight this point.

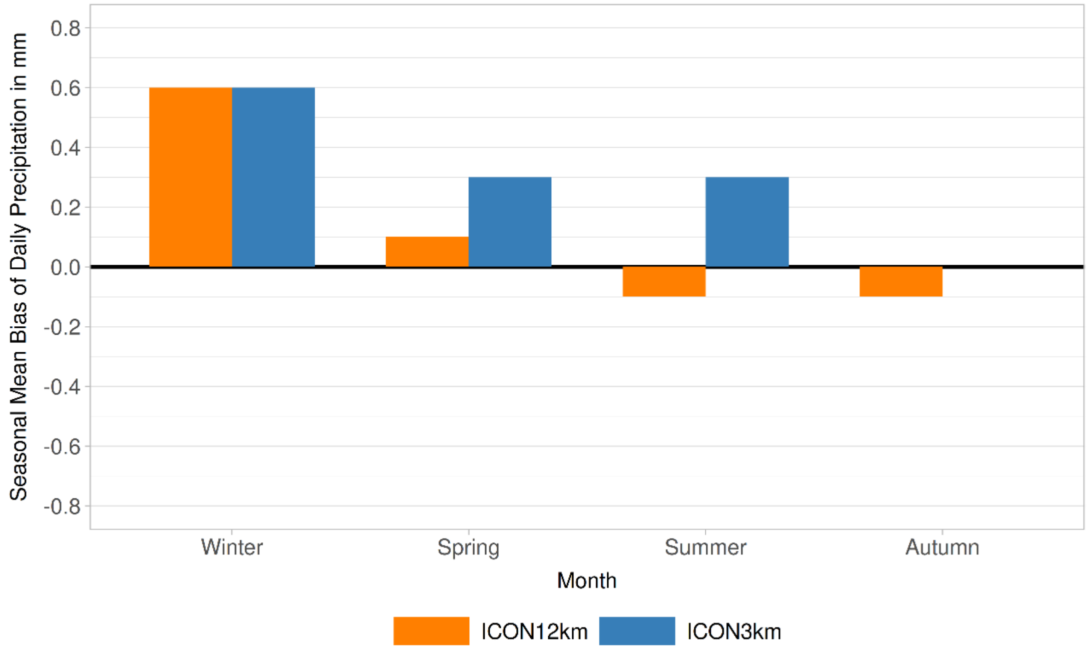

Fig. 2: Seasonal mean bias of daily catchment-average precipitation of ICON3km and ICON12km for the period from 2005 to 2014 for Winter (DJF), Spring (MAM), Summer (JJA) and Autumn (SON)

6) Fig 10 - Why does the maximum relative soil moisture differ for ICON11km and ICON3km ? Intuitively, I would have thought that this parameter would reach the same maximum relative soil moisture when saturation is reached by each of the models. Figure 15 also depicts this. Is there an issue with the reservoir capacity between models ?

In addition, is there a link to the bias differences of hourly ET if the saturation is never reached for ICON11km ?

The considered catchments of the Weiße Elster cover 2,960 km² and are located in East Central Germany. We study catchment-average hourly soil moisture for the period from 2006 to 2014. It is not to be excepted that every grid cell reaches saturation and that the catchment-average comes to maximum relative soil moisture. In the simulation of the flood of 2013, soil moisture indeed leveled out when the hydrological model was driven with RADOLAN data, however this does not mean that catchment-average soil moisture was at maximum capacity, but that the soil grids upon which the rainfall in early June 2013 fell were saturated. Would the rainfall have fallen over different parts of the catchment, which were not yet saturated (as it is assumed to have in ICON3km), the catchment-average relative soil moisture would have risen further (as it did for ICON3km). To summarise, none of the models reached saturation for every grid box, reason why the different simulations come to different catchment-average maximum relative soil moisture values.

7) Fig S18 : In my opinion, this figure is a main result and is more important for the paper than Fig14. I suggest you switch the figures. FigS18 in the paper body and Fig 14 in Supportive information.

We thank the reviewer for the suggestion. However, Fig. 14 explores the representation of non-linear behaviour in flood generation using CPRCM-data, thereby covering a research question which has seldomly been studied in literature to-date. Even though the data shows to be widely scattered and doesn't allow for conclusions on this topic, its representation highlights the open research question. It would be interesting to follow it up with ensembles and longer time series. Furthermore, the inclusion of Fig. 14 supports a topic in the paper, which has not been elaborated graphically in another context in the manuscript. Figure S18 in contrast represents boxplots of the 99.5th percentiles of hourly discharge, showing a positive bias in the CPRCM data, just as could be seen in the boxplots across the entire data series (Fig. 12). We therefore have decided to keep Fig. 14 in the main body of the paper and Fig. S18 in the appendix.

8) l461-462 : "leading to a reduction of the flood peak". This part of the sentence is misleading. The increase of the soil moisture does not lead to a reduction of the flood peak, it shows that the flood could have been worse. There is no causal link between an increase in soil moisture and a reduction in flooding. I suggest rephrasing this sentence as follows: "In fact, the average soil moisture in the hydrological model driven with ICON3km rose, showing the buffering of the large amount of precipitation simulated by ICON3km."

Soils have the capacity to store water through a large network of pores, their size and connectivities being a function of the soil type. Rainfall, which rains down on a soil, partly runs off and partly infiltrates in an amount dependent on the infiltration capacity of the respective soil. As water fills up the pores of the soil, the latter's soil moisture rises. In the event of 2013, a rise in the soil moisture was simulated as a reflection that part of the rainfall had contributed to the water storage in the soil. This amount is not available for runoff, leading to a reduced runoff peak. Though it might contribute in parts to the river discharge as exfiltrating interflow or groundwater, a significant share of it will remain stored in the soil. To conclude, soils have the capacity to buffer rainfall events and flood waves, with an uptake of water meaning higher soil moisture in the soil. The referred lines summarise these processes and we hence suggest keeping them in the paper unchanged.

9) l538-539 : "These results agree with the consensus from literature (Lucas-Picher et al., 2021), e.g. with Ban et al. (2021) and Adinolfi et al. (2021).". I disagree with this statement. The literature you cite shows mainly a reduction in biases in the simulation of extreme hourly precipitation by the CPM compared to the RCM (Fig 7 and Fig 8 of Ban et al. (2021) for a wide range of CPM in Europe). In your study, the CPM changes the sign of the bias from negative to positive for extreme precipitation, with an increase in the absolute bias by the CPM (see Figure 6).

You should rephrase this sentence to take this point into account.

Indeed, Ban et al. (2021) conclude that the climate model ensemble run at 3 km resolution shows lower wet-hour frequency, but more intense precipitation than the driving RCMs. The authors note that higher estimates in precipitation intensity were particularly pronounced for heavy hourly precipitation. While the ensemble mean at 12 km resolution showed a negative bias for all seasons and all study regions (Switzerland, France, Italy) for the 99.9th-percentile of hourly precipitation, indeed the 3 km ensemble produced results that were in the mean in the acceptable bias range considering the measurement uncertainties (Ban et al., 2021). We agree with the reviewer's comment that while this does prove higher intensities by the CPRCMs, it does not necessarily mean a positive bias in the CPRCM-estimates. We suggest to rephrase the corresponding sentences as follows:

„ICON12km was found to compute too frequent light precipitation, while underestimating the intensity of heavy rainfall. ICON3km improves on the frequency estimation of light precipitation but overestimates the intensity of the highest quantiles. A reduction in the frequency of light precipitation and an increase in the intensity of heavy precipitation by the CPRCM compared to its driving RCM is in line with the consensus from literature (Lucas-Picher et al., 2021), e.g. with Ban et al. (2021) and Adinolfi et al. (2021). Ban et al. (2021) studied a 23-member ensemble of CPRCMs run at 3 km resolution and their driving RCMs at 12 km and found the CPRCMs to be able to alleviate the RCM-ensemble negative mean bias seen in the 99.9th-precipitation-percentile. In fact, the CPRCM-ensemble shows a respective mean bias within the acceptable range considering measurement uncertainty for all study areas and seasons, except winter (DJF) in Switzerland and summer in Italy (JJA) where the negative bias is weakened but remaining (Ban et al., 2021). Ban et al. (2021) furthermore found improvement through reduced wet-hour frequency by the studied CPRCM-ensemble. Adolfini et al. (2021) show similar improvement by the CPRCM they analysed over an Alpine domain, stretching from central Italy to northern Germany, reducing precipitation frequency and increasing precipitation intensity compared to the driving RCM. Nevertheless, within the 99.9th percentile, the CPRCM intensity estimates remain slightly too low for precipitation values lower than 6 mm/h, but the CPRCM overestimates heavy precipitation higher than 6 mm/h (Adolfini et al., 2021)."

10) l640-641 - This sentence contradicts line 631 "Xie et al. (2025) attributed no added value to the use of their studied CPRCM for hydrological impact modelling"

We thank the reviewer for the careful reading of the manuscript.

In fact, Xie et al. (2025) state in their paper abstract that: „The study concludes that CPRCMs improve the simulation of extreme precipitation and temperature but not show clear added value for flood simulations." (Xie et al., 2025) In their conclusion however they state that: „HCLIM3 does not consistently outperform HCLIM12 in flood simulations across the two study basins, although it demonstrates added value in simulating severe flood peaks."

We agree with the reviewer on a contradiction. We have chosen to cross the paper from those mentioned showing no added value for the use of CPRCMs for hydrological impact modelling (L. 642 and L. 698 in the new manuscript version).

11) l650-654 - This discussion is interesting and could be enriched by literature on the role of antecedent condition of rainfall-runoff climatic event and their influence on flood intensity (Sharma et al., 2018; Tramblay et al., 2019; Wasko & Nathan, 2019)

We thank the reviewer for this suggestion. We conclude from the referred studies that higher precipitation intensity does not necessarily result in greater flooding per se, with antecedent soil moisture being a key factor able to dampen peak discharge (Tramblay et al., 2019; Wasko & Nathan, 2019). The correlation between peak discharge and soil moisture is particularly high for frequent flooding events and large basins (i.e. above 100 km², c.f. subcatchments feeding Zeitz sum to 2.535 km², Wasko & Nathan, 2019). However, for very high intensity rainfall events, the soils' flood attenuation capacity was found to be limited and the correlation between soil moisture and peak discharge diminishes (Wasko & Nathan, 2019).

While the topic of the importance of antecedent soil moisture on flood generation is highly interesting, we consider it to be outside the scope of our paper, and given the length of the present manuscript, we have decided not to include these additional references.

12) l655 : Why no added value can be found for hydrological modelling ? It is important here to add a statement on the transfer of ICON3km extreme precipitation biases on hydrological modelling, in particular high flows and floods.

We suggest to adapt the respective sentence as follows: „To conclude, no added value could be found in the use of uncorrected climate model data from ICON3km for hydrological impact modelling, given the overestimation of the intensity and frequency of heavy rainfall by ICON3km."

13) l686-701 : I think this paragraph should be displaced in its own section in the discussion. Maraun (2016) is another citation that could complete the discussion on bias-correction.

We thank the reviewer for this suggestion. However, we consider a more in-depth elaboration on the topic of bias-correction to be outside the scope of our paper, which did not apply bias-correction. Still, the referred paper of Maraun (2016) is a valuable contribution to the discussion and has been added to the revised version of the manuscript.

Comment by the authors

The RCM is run on the Euro-CORDEX domain at a resolution of 0.11° (EU-11, approximately 12.5 km). Throughout the paper, we erroneously referred to the resolution as being of 11 km. We have updated the paper to correctly state the resolution as 12 km.

We have furthermore deleted lines 175 f. of the first submitted revised manuscript, as the expressed statements have already been covered in the preceding lines 167 and 170.

References

Strandberg, G. and Lind, P.: The importance of horizontal model resolution on simulated precipitation in Europe – from global to regional models, 2, 181–204, https://doi.org/10.5194/wcd-2-181-2021, 2021.

Review 3 (RC3)

General Comments

The paper has been well extended by the suggestions of the reviewers. By now it provides a comprehensive overview about the CPRCM results and their usage for hydrological modelling. Some important aspects regarding the hydrological simulation results were added, such as soil moisture, which bring light into the model behaviour at the analysed event with the different inputs. The model calibration results as well as the reasons of mismatching in some catchments are now well discussed. The discussion part and the conclusions have been extended by important parts (e.g., L650 ff), from which it can be concluded that it is important to capture both the spatial distribution of the input and the spatial differences in catchment states (snow and soil moisture). The conclusions further point out the challenges using raw or corrected climate model results in climate change impact studies and underline the need for ensemble studies also for convection permitting models. I recommend publication together with the appendix with some minor, rather technical, corrections (see below, page and line numbers refer to the document with the author's tracked changes).

> We thank the reviewer for the positive comments. We appreciate the time taken to assess our paper.

Technical Comments

P. 8 / L. 139: I would leave out "however"

> We thank the reviewer for the suggestion, which we have adopted in the revised version of the manuscript.

P. 9 / L. 171: Insert "focussing": ... calibrated focussing on (1) ...

> We support this comment and have implemented it.

P. 17 / L. 337 – L. 341: I believe, the inserted paragraph could be left out, this should be clear for modellers. The question is, do we want a better resolution or is smoothing ok, when data are highly uncertain?

> We have followed the suggestion and deleted this paragraph.

P. 23 / L. 440: In S18 it seems the results with the climate models have a larger variation in the 99.5th percentiles. Is this connected to the generally higher number of simulated high discharge peaks? Perhaps you can add a sentence to this.

We thank the reviewer for this valuable comment, which draws a connection between the results from the rainfall-runoff-analysis (Fig. 14) and the boxplots showing the 99.5th percentile of hourly discharge (Fig. S18). Indeed, we could see in the scatter plot of total runoff to the sum of total rainfall and 8-day antecedent rainfall (Fig. 14) that the use of climate model data leads to a greater number of simulated rainfall-runoff events of high return period. This is reflected through a more strongly scattered rainfall-runoff-plot in comparison to the use of observational meteorological data, where the events are more strongly clustered. In the boxplots of the 99.5th percentile of hourly discharge (Fig. S18) this translates into a larger interquartile range in the discharge simulations driven with climate model data than when driven with observational meteorological data.

We suggest the following additions to the paper:

- L. 437 ff.: "The interquartile range of the 99.5th percentile of hourly discharge computed by the hydrological model driven with climate model data (ICON12km/ ICON3km) is found to be larger than when driven with meteorological observations (RADOLAN12km/ RADOLAN3km)."
- L. 656 ff.: „In fact, the rainfall-runoff-analysis (Fig. 14) showed a greater number of rainfall-runoff events of high return period when using the climate model data for hydrological modelling than when using meteorological observations. This is reflected through a more strongly scattered rainfall-runoff-plot in comparison to the use of observational meteorological data and translates to a higher interquartile range and a higher median of the 99.5th percentile of hourly discharge for the hydrological model driven with climate model data (Fig. S18)."

P. 24 / Fig. 14: Please distinct the plots above (e.g., by adding a, b, c, d).

We present the new figure below and have updated it in the manuscript.

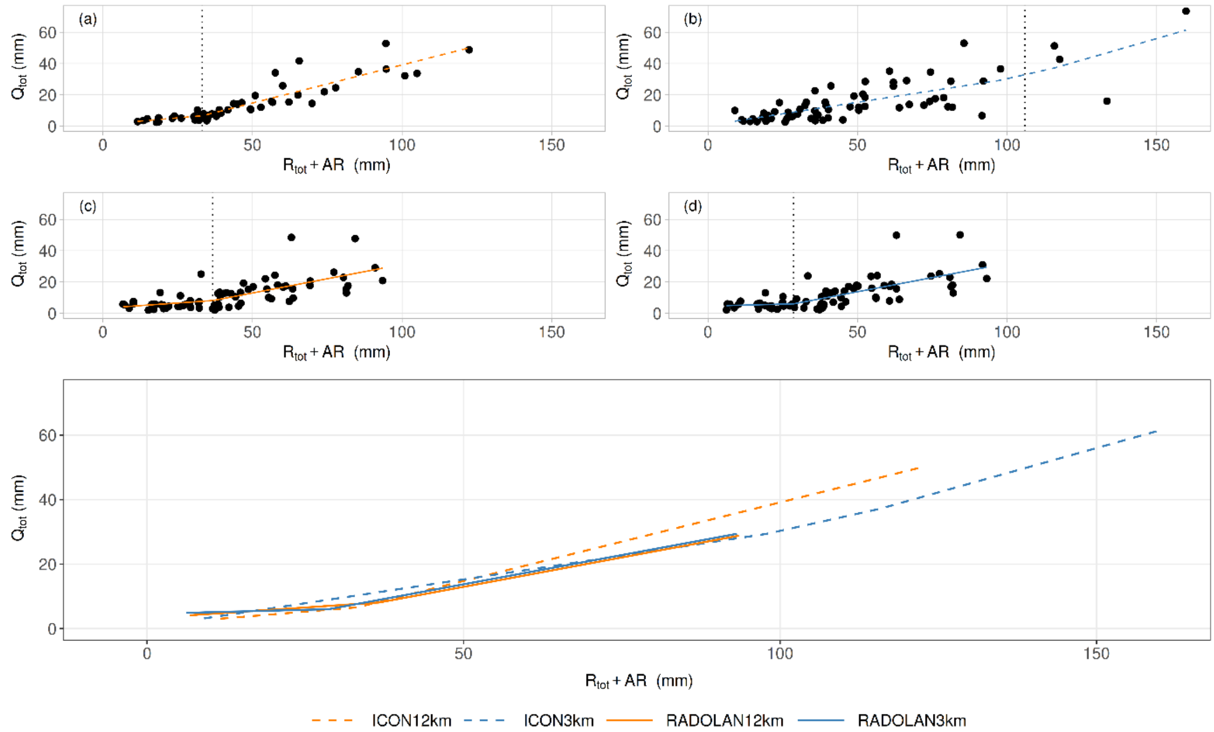

Fig. 1: Upper half: Scatter plots of total runoff (Q_{tot}) to the sum of total rainfall (R_{tot}) and 8-day antecedent rainfall (AR) over the catchment of Kleindalzig and those upstream, fitted with a piecewise linear regression model for (a) ICON12km, (b) ICON3km, (c) RADOLAN12km and (d) RADOLAN3km. The thresholds are indicated with a dotted vertical line. Lower half: Combined piecewise linear regression model results of Q_{tot} to R_{tot} + AR

Comment by the authors

The RCM is run on the Euro-CORDEX domain at a resolution of 0.11° (EU-11, approximately 12.5 km). Throughout the paper, we erroneously referred to the resolution as being of 11 km. We have updated the paper to correctly state the resolution as 12 km.

We have furthermore deleted lines 175 f. of the first submitted revised manuscript, as the expressed statements have already been covered in the preceding lines 167 and 170.