

Nat. Hazards Earth Syst. Sci. Discuss., author comment AC1  
<https://doi.org/10.5194/nhess-2021-66-AC1>, 2021  
© Author(s) 2021. This work is distributed under  
the Creative Commons Attribution 4.0 License.



## Reply on RC1

Benjamin Poschlod

---

Author comment on "Using high-resolution regional climate models to estimate return levels of daily extreme precipitation over Bavaria" by Benjamin Poschlod, Nat. Hazards Earth Syst. Sci. Discuss., <https://doi.org/10.5194/nhess-2021-66-AC1>, 2021

---

Dear Referee,

Thank you for your comments and helpful suggestions. I agree with most of your thoughts and I am sure that addressing your suggestions will further improve the article. I will answer the comments point by point below (your comments marked in *italic*), but first I would like to provide an explanatory introduction about the motivation for this study, which will hopefully allow a better understanding of the model and data selection.

This investigation is motivated by the findings about the representation of extreme precipitation in the CRCM5 large ensemble (CRCM5-LE) about historical (<https://doi.org/10.5194/essd-13-983-2021>) and future conditions (<https://doi.org/10.1088/1748-9326/ac0849>). The CRCM5-LE is a single model initial-condition large ensemble featuring 50 members and driven by a 50-member global circulation model large ensemble (CanESM2-LE). The first study has shown that the model is able to reproduce observed extreme rainfall return levels with good skill over Europe and also over Germany and Bavaria (similar to the CRCM5 in this study). The second study has shown the large projected changes of extreme precipitation under the RCP8.5 scenario.

The results of these studies with a focus on Germany and Bavaria were presented to a large selection of institutional experts and users of extreme rainfall return level data from the Bavarian Environmental Agency. For them, the KOSTRA dataset (as legal guideline) is the basis for applications and therefore the "benchmark". However, KOSTRA is based on historical observation data between 1951 and 2010.

As has already been the case for floods (HF100) since 2004 in Bavaria (see L495-497 in the article), a climate change surcharge for heavy precipitation was discussed. From my perspective as a climate scientist, the results from the CRCM5-LE would be sufficient to recommend adaptive measures such as a climate change surcharge. However, local biases of the CRCM5-LE rainfall return levels have greatly reduced the acceptance of the results by the institutional users (which I have hinted at in L63-66).

After this presentation and discussion, I asked myself whether currently available RCM simulations with higher resolution could lower the local biases in order to increase the acceptance of extreme rainfall return levels projections.

Therefore, the CRCM5 was chosen as the "climate model reference" and KOSTRA as the target to be simulated.

I hope that the description of this starting point provides a good understanding of the choice of data (KOSTRA) and model (CRCM5). I will mention this motivation more clearly and in more detail in the article.

*Major comments:*

*1) In the conclusions the author clearly stated the uncertainties arising from different model setups regarding internal climate variability, parametrizations, and further assumptions. Saying so, why did you then choose different RCMs and not only a single one with similar setups, e.g., a COSMO-CLM version in the given (slightly different) resolutions? Furthermore, why did you use ERA-Interim and ERA5 as forcing data and not only the higher resolved and newer ERA5 data for all simulations?*

The CRCM5 simulation was chosen as the "reference" RCM simulation due to the previous studies based on this model (see explanation above). However, the higher-resolved and convection-permitting successor CRCM6 is still being developed. Therefore, I chose higher resolution simulations from freely available data sources covering the study area with a time period of 30 years driven by reanalysis data. The 5km WRF as representative for high-resolution simulations with parametrization of convection, and the 1.5 km WRF as the highest-resolution simulation known to me without any parametrization of deep and shallow convection.

*2) The author put lots of effort into the homogenization of pointwise observational data sets. There are several high-res gridded precipitation data sets on the market like REGNIE/HYRAS for Germany (1km, Rauthe et al., 2013), RADOLAN (DWD, 1km), or SPARTACUS (Austria, 1km, Hiebl and Frej, 2017). I agree that even at this high resolution these data sets have limitations when it comes to convection. Nevertheless, DWD and ZAMG put a lot of effort into calibrating these data sets not only with ground measurements but also with radar data and vice versa in the case of RADOLAN. So, I assume these data sets have a higher quality than the homogenized point observations by the author and they have a higher resolution which made the validation of the 1.5km WRF model more robust.*

I agree with you on your assessment of the gridded data sets. However, the public authorities rely on KOSTRA as legal guideline. RADOLAN, for example shows substantial differences compared to KOSTRA (see Fig. 26 in [https://www.dwd.de/DE/leistungen/pbfb\\_verlag\\_berichte/pdf\\_einzelbaende/251\\_pdf.pdf?\\_\\_blob=publicationFile&v=2](https://www.dwd.de/DE/leistungen/pbfb_verlag_berichte/pdf_einzelbaende/251_pdf.pdf?__blob=publicationFile&v=2) sorry, only available in German). In a personal communication with a representative of the DWD, the technical suitability and high quality of the KOSTRA data set for daily rainfall return levels was also confirmed to me.

The REGNIE data set is based on the same observational data as KOSTRA, but the interpolation is carried out on a finer grid. Therefore, multiple regression using the elevation and exposition are used as co-variables for interpolation. However, the scope of REGNIE is to provide daily high-resolution precipitation fields, whereas the scope of KOSTRA is to provide rainfall return levels. Hence, the workflow also differs:

For KOSTRA, the order is the following: Rain gauge observations -> extreme value analysis -> spatial interpolation.

For REGNIE: Rain gauge observations -> spatial interpolation. As next step extreme value analysis could be carried out.

As you suggest, a comparison of REGNIE (and similar products in Austria and Switzerland) to the 1.5km WRF return levels would be very interesting as well. However, this comparison would be more of an evaluation of the REGNIE interpolation method versus the 1.5 km WRF return levels, and therefore I would refrain from doing so in this article.

*3) When it comes to different extreme value techniques, a proper validation would use every method with every data set and not only a couple of possible combinations like currently presented.*

Thanks for this suggestion. I will calculate the remaining combinations and add the results (probably) in the Supplementary Materials. If unexpected or noteworthy findings from the remaining combinations evolve, I will of course show and discuss them in the main article.

*4) The authors conclude that RCMs are better in terms of spatial representativeness of return levels. Saying so I expect cross-validation with existing products like KOSTRA for Germany to clearly point out the benefit of RCMs compared to raw or existing gridded observations.*

I guess, you are referring to L475ff. I do not conclude that RCMs are generally better than any KOSTRA/ÖKOSTRA/Swiss gauges in terms of spatial representativeness of return levels.

The spatially homogeneous return levels based on RCMs can mostly support areas, where observations are scarce. The study area shows a high rain gauge density, which is necessary to validate the RCM return levels. Therefore, the comparison of RCM return levels to the broad database of KOSTRA/ÖKOSTRA/Swiss gauges implies that RCMs can be used in areas with less observational coverage in order to enhance the spatial representativeness.

Second, the spatial representativeness of each rain gauge differs depending on the topography. Especially in complex terrain, a rain gauge based in the valley may be only representative for a very small area in this valley, but not for the surrounding slopes. That's why the ÖKOSTRA design rainfall is based on a combination using observed data with limited spatial representativity and simulations of a convection-permitting weather model. The conclusion of this study supports this ÖKOSTRA-approach.

I think that this conclusion can be sufficiently justified on the basis of the existing results (spatial correlation based on the Spearman coefficient). I think a separate cross-validation would be out of the scope of this article.

*5) The author concentrated on the return level of 10 years and stated that this is the most important value for the targeted applications. At least for the insurance industry, minimum the 100-year return level better the 200-year values (PML200) are the relevant levels. As all results are specifically related to the 10-year level, I am wondering if the methodology can be adapted/used for higher return levels or if further validation/calibration is necessary in that case. I miss some statements on that in the discussion and conclusions*

sections.

Thanks for this comment. I agree that the even more extreme return levels are very important and very interesting. I will calculate and add the 100-year return levels in the revised version of the article.

*Additionally to the major comments above, I have some minor comments [page-line/paragraph]:*

*[Sect. 1] I recommend clearly state the key research questions you are focusing on in this study. For me, it is not clear what the main aims are.*

The main questions are:

Are there high-resolution RCM setups available which lead to lower local biases of rainfall return levels compared to observations?

How much do the return levels based on different state-of-the-art extreme value theoretical methods differ?

I'll add them at the end of the introduction.

*[P3 L81ff] Schröter et al. (2015) analyzed three major flood events in Germany during the past 70 years (1954,2002,2013), which also partly affected your investigation area, concluding that it is not daily/multi-day precipitation amount that triggers major flood events.*

Thanks for this hint. I did not want to state that daily extreme precipitation events are the only relevant trigger for floods in the whole study area. I will add the Schröter et al. (2015) reference and state that daily extreme precipitation is not alone responsible for all floods in the area. Of course, the antecedent wetness state plays a major role as well. Further, the character and size of the respective river catchment determines, which precipitation duration is more probable to cause floods.

*[P3 L88ff] "RCM can bridge the gaps" – what about stochastic weather generator or other approaches? Ehmele and Kunz (2019), for example, introduced a semi-physical, 2D, and high-resolved precipitation model mainly based on orographic precipitation which in a statistical sense, gives good results in terms of return levels even for higher return periods.*

Thanks for this reference: I will include and discuss stochastic weather generators, such as Ehmele & Kunz, and other statistical methods e.g. the French SHYPRE as representatives of different methods. Of course, there are many more methods to investigate extreme precipitation.

However, RCMs incorporate the advantage of including climate scenarios in their boundary conditions.

*[Sect. 2] I recommend a reordering of the paragraphs in this section. As your investigation area is restricted to the given data sets, I suggest first describe the data sets and the investigation area afterward.*

Thanks for this suggestion, I rearrange the paragraphs accordingly for better readability.

*[Fig.1] Is the study area equal to the model domain? If so, how do you deal with boundary effects?*

No, the whole model domain of the 1.5 km model setup covers 351x351 grid cells, whereas the study area is reduced to 271x271 grid cells by Collier & Mölg (2020). 40 grid cells at each edge of the domain are discarded for the analysis. Hence, boundary effects can be assumed to be excluded.

I'll add an according statement in the text.

*[P4 L97f] In Fig.2 you give the reference for the data set, I suggest giving it in the text, too.*

Will be added in L98.

*[Fig.2] Do you have an explanation for the strong "drying" signal in the main Alpine valleys? Please use discrete color separations. See also <https://www.nature.com/articles/s41467-020-19160-7>*

Thanks for this suggestion. The figure will be re-drawn with appropriate colors.

The underestimation of annual precipitation in the Alpine valleys is also stated by Warscher et al. (2019), but without any explanation. Maybe it would be better to show an observational product - however the EURO4-APGD (Isotta et al., 2014 <https://doi.org/10.1002/joc.3794>) does not cover the northern part of the study area.

Even though E-OBS with a spatial resolution of 0.11° (~12 km) cannot fully resolve the spatial heterogeneity, I will re-draw Figure 2 based on E-OBS.

*[Sect. 2.2] So I understand that you estimate daily precipitation or at least 24h sums in the moving window by hourly station data, right? If so, please clarify in the text.*

The KOSTRA and ÖKOSTRA data are partly based on (sub-)hourly station data using a moving window and partly based on daily observations. However, the 24h return level values are provided representing the value of extreme precipitation for a 24-hourly moving window.

I transferred these values to "daily estimations" (fixed window) reducing them by 14%, as this relationship between "daily" and "24-hourly moving window" has been found stable (Boughton & Jakob 2008, Barbero et al. 2019 ref. in the article, own calculations).

I'll enhance the description in the text to clarify this. The Swiss return levels are provided as daily estimates, and therefore no reduction is applied.

*[P7 L133] "24h RLs are adjusted to daily values using a reduction". I do not understand what this reduction is about. Please clarify this in the text.*

see comment above; will be clarified.

*[Sect. 2.3] Why did you choose exactly these models and not others? There is a huge variety of RCM in 0.11° resolution within the CORDEX project and also high-resolution simulations mainly Germany and Alpine region in the CORDEX FPS convection project. Furthermore, you used WRF v3.6.1 for the 5km and v4.1 for the 1.5km simulations. Are there major differences between the versions? For consistency, the same model version would be better.*

The motivation to use the CRCM5 as "12-km RCM reference" is explained in the introduction of this answer. Further, the CRCM5 has shown a relatively good performance at the reproduction of 10-year return levels (<https://doi.org/10.5194/essd-13-983-2021>) compared to EURO-CORDEX models (see Berg et al., 2019; ref in article).

I agree that the same WRF model versions would be better to investigate the added value of higher resolution. The two WRF setups have been chosen as they are publicly available and cover 30 years driven by reanalysis data. The 5km WRF represents a setup with a very high resolution and parametrization of convection, whereas the 1.5 km setup is the highest-resolution setup known to me. Further it explicitly resolves deep and shallow convection.

The main differences between the WRF versions are summarized here:

<https://github.com/wrf-model/WRF/releases/tag/v4.1>

In the newer version, additional schemes are available (microphysics, radiation, cumulus). However, these have not been applied in the 1.5 km setup. Further, some minor improvements and bug fixes have been implemented. Hence, I conclude that the differences of the model version do not play a major role.

*[P8 L161ff] For WRF 1.5km, you have 30 simulations with a 1-year length each. Does this have an impact on the comparability with the continuous simulations at coarser resolution?*

Of course, the model initialization impacts the simulations due to the model representation of internal climate variability. In that sense, transient simulations would yield slightly different results than the sliced simulations. As the WRF domain is forced by the lateral boundary conditions of the ERA5 reanalysis data at 3-hourly resolution, I would not assume that slicing the simulation period does have a systematic impact on the magnitude of rainfall return levels.

For other variables in the WRF with longer lag times such as deeper soil moisture, transient simulations would be more appropriate. Hence, I conclude that the slicing does not have an impact on the comparability of daily rainfall return levels with the transient simulations.

*[Sect.3] I suggest a reordering here, too. Instead of first describing strategies and distributions and then how they are applied in this study, I recommend a structure like 3.1 BM; 3.2 POT, 3.3 MEV each with a short introduction to the method and then directly saying how you will apply it in this study.*

I will revise the text according to your suggestion.

*[P9 L180ff] It would be helpful for the reader if you can give typical values or magnitude orders of  $t_{wet}$  and  $t_{decluster}$ .*

Values are given in L248 and L261. The reordering of the section (see comment above) will provide these values directly in the respective subsection.

*[P9 L192]  $G$  is also a CDF, right? Please indicate it.*

Will be corrected.

*[P12 L242] Can you explain why the low-res simulations have higher return values than the high-res?*

There is no simple relationship between spatial resolution and extreme precipitation intensity. Of course, GCMs show smaller rainfall intensities than RCMs, and there is a general tendency that higher spatial resolution leads to higher precipitation intensity. However, the chosen model, the model setup, and the chosen parametrization schemes can overlay this tendency.

*[P13 L277] You mean Fig.5d instead of 5b?*

yes, will be corrected.

*[P15 L289] The 5km WRF seems to have a much stronger orographic signal than the 1.5km, especially the "drying" in the main valleys. Is there any explanation for that?*

Warscher et al. (2019) also note this strong orographic signal in their setup compared to observational data. I don't have an explanation for this behavior, but I will describe this behavior in the revised text.

*[P15 293] Fig.5b and later 5e?*

Yes, will be corrected.

*[P15 L301] Sure you mean Fig 3d here?*

It should read 4d.

*[P15 L305] Fig.5c and 5f, I guess*

Yes, will be corrected.

*[P19 L394-398] Maybe I miss something, but I do not get the message from these two paragraphs*

These two paragraphs discuss the differences in the driving data (75km ERA-Interim vs. 30km ERA5) and the temporal coverage (1980–2009 and 1988–2017) as sources of uncertainty and discrepancy between the three RCM setups.

The message is: Even though driving data, model (version) and time period differ, the resulting return levels are quite similar in terms of intensity and spatial patterns.

L399ff: However, for the evaluation of single extreme events, the setup can result in large differences.

*[Fig. S5+S6] There is data missing for Switzerland and Austria. Why? I thought you have the data for that regions and time periods.*

I only have the return level data as described in section 2.2.

For Fig. S5&6 REGNIE data is used for Germany, as it is publicly available. For Switzerland and Austria, similar products are not publicly available as far as I know. Only for the event in August 2005, Meteo Swiss provided daily precipitation here: <https://www.meteoswiss.admin.ch/home/climate/swiss-climate-in-detail/extreme-value-analyses/high-impact-precipitation-events/19-23-august-2005/precipitation-and-temperature.html>

I hope that my answers can satisfyingly address your comments and suggestions.

Kind regards,

Benjamin Poschlod