Reply on RC2

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-AC2, 2021

Dear Referee,

Thank you for your comments and helpful suggestions. I really think that your thoughts and additional references help to improve the quality of this study and encouraged me to dive deeper into EVT and the MEV framework.

I will address 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) The use of the high-resolution products (REGNIE, RADOLAN, SPARTACUS) would avoid to homogenize the gauge precipitation values and would make possible a more accurate validation of the RCMs with the finest resolution. Why not considering them?

Yes, these data sets provide values at higher spatial resolution. However, for rainfall return levels, the public authorities rely on KOSTRA as legal guideline (as described in the introduction of this answer). Furthermore, 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 confirmed to me.

The radar product RADOLAN, for example shows substantial differences compared to the observation-based KOSTRA (see Fig. 26 in https://www.dwd.de/DE/leistungen/pbfb_verla g_berichte/pdf_einzelbaende/251_pdf.pdf?__blob=publicationFile&v=2 sorry, only available in German).

The REGNIE data set is based on the same observational data as KOSTRA, but just the interpolation is carried out on a finer grid. Therefore, multiple regression using the elevation and exposition are used as co-variates 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 (~ 8km).

For REGNIE: Rain gauge observations -> spatial interpolation (1km). 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 in general be very interesting. However, in my opinion 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.

2) Why only return level of 10 years? I understand the concern of the author that 30 years of data are few for estimating higher quantiles, but return periods higher than 10 (e.g., 100) years are more relevant for engineering applications/(re)insurance purposes and the challenge is indeed to estimate them with the availability of short time series. How would the estimation of higher return levels compare e.g. with the official ones from KOSTRA? As the manuscript is presented now, the conclusion stated in the abstract "it follows that high-resolution regional climate models are suitable for generating spatially homogenous

rainfall return level products" is not fully supported by the analysis, since only the 10-years return levels have been evaluated.

I totally agree. I will add the 100-year return levels in the revised version of the manuscript. It will also be interesting if the MEV can outperform the GEV and GPD for the longer return period.

3) The study area is characterized by some high-elevated regions affected by orographic precipitation. I'm wondering if using all the values as "ordinary events" in the MEV might not respect the independence hypothesis required by the MEV framework. See for example Marra et al. (2018) and Miniussi et al. (2020) for some discussion on temporal correlation.

Thank you for this comment on the methodology and these two very interesting additional references. I will check the temporal autocorrelation to ensure that all "ordinary events" can be assumed to be independent events (similarly to Fig S3 of Zorzetto et al., 2016). For grid cells, where the independence hypothesis is not respected, I will implement an "event separation method" as in Marra et al. (2018). However, in this study only daily (not hourly) rainfall amounts are available, so the "running parameter" (minimal time between two rainfall events) will amount to [0, 1, 2, 3, …] days.

4) Why using a GEV distribution with a constant shape parameter and not, for example, a Gumbel? Previous studies (e.g., Grieser et al. (2007)) have shown that the Gumbel distribution is a good model for precipitation in the Bavarian area, and its location parameter has a strong correlation with altitude, while its scale parameter has a noisy pattern (except for the Bavarian Alps). Moreover, you say that the shape parameter based on all the three RCM setups is centered around a value close to 0.114, in line with the one recommended by Papalexiou and Koutsoyiannis (2013): is this really a fair comparison, as these shape parameter values are already affected by estimation uncertainty?

Grieser et al. 2007 stated that the Gumbel distribution is an adequate fit for the German rain gauge data – however they also write: "The fitting of this GEV to a data sample allows the identification of which of the three extreme value distributions may adequately describe the data. However, this approach of model identification is not the topic of this paper. It is already known that daily precipitation in Germany can be adequately described by a Gamma distribution (ZOLINA et al., 2005). Since the Gamma distribution is in the basin of the Gumbel distribution we a priori expect the highest value of a sample of daily precipitation to be describable by the Gumbel distribution."

They also checked the goodness of fit via the value of R^2 for the empirical annual maxima and the theoretical fitted Gumbel distribution.

The cited underlying paper of Zolina et al. 2005 (https://doi.org/10.1029/2005GL023231), which used the Gamma distribution, only investigated 96 rain gauges over Europe, from which 20 – 30 gauges are in the study area of this paper.

Papalexiou and Koutsoyiannis (2013) found that in the study area the value of the shape parameter amounts to 0.1 – 0.14 based on rain gauge observations. Hence, also their regional investigation in the study area supports their finding that a value of 0.114 is an appropriate choice for most areas in the world.

The finding in this study that the distribution of the values of the shape parameter fitted

via L-moments is centered around roughly 0.114 for all three RCMs is of course affected by estimation uncertainty. But it supports the argumentation by Papalexiou and Koutsoyiannis (2013).

Encouraged by your comment, I carried out a similar investigation to Papalexiou and Koutsoyiannis (2013) for over 1100 rain gauges in the study area with more than 30 years of daily precipitation data. I fitted the GEV via L-moments, which results in a distribution of the shape parameter centered around 0.09. Of course, again this value is affected by estimation uncertainty.

Further, I carried out the Anderson-Darling test at 0.05 significance for the GEV with shape=0.114 and the Gumbel distribution for the 1.5 km WRF data. For the GEV with shape=0.114, less grid cells have a p-value < 0.05 than for the Gumbel distribution (GEV114: 4097 versus Gumbel: 9180 grid cells from 73441 grid cells in total). When applying the adjusted p-values following Wilks (2016) the GEV with shape=0.114 is not rejected at any grid cell, whereas the Gumbel fits are rejected at 870 grid cells. Therefore, I think that the GEV with shape=0.114 is more appropriate for the study area.

Minor comments.

Section 3.

L225: Another title for section 3.3 would be more appropriate

Section 3 will be reordered according to the suggestion of the other referee. Hence, this title will be discarded.

L226-227: please add a couple of words about the adjustment, so that the reader understands it directly from here without the need to go looking at the reference.

I will add an explanation of the general problem of false positives and the principle of the approach by Wilks (2016).

L239: you state that "the location and scale parameter are governed by the topography". From Figure 3 one can notice that the spatial pattern of the location parameter is somehow coherent with topography, but the noise for the scale parameter does not make its pattern straightforward to understand. Maybe also the colors scale is not helping.

For the scale parameter, some topographical features are visible (e.g. the Alpine valleys in Fig. 3e,h; the Swabian Jura, the Odenwald, the Ore Mountains, and the Bavarian Forest in Fig. 3e,h), but as you mentioned, the pattern is noisier.

I will clarify in the text that the patterns of the scale parameter are noisier than for the location parameter and discuss the topographical features, which are visible.

L240: why a chaotic pattern for the shape parameter? Is it related to the uncertainty that one can get due to the limited series available to estimate it?

The explanation is given in L438-441 as I did not want to anticipate the content of the discussion in the results section. Indeed, this explanation refers to the uncertainty that one can get due to the limited series available to estimate it.

L259: you mention you made a "goodness of fit" (despite its limitation in prediction) for the GEV and the GP distributions. Have you made a similar analysis also for the Weibull distribution?

No, I haven't. I will add such analysis for the Weibull distribution as well.

L264: in L253-255 you mention that for sample sizes > 50 estimation via ML is recommended. Why then using PWM for the Weibull distribution in the MEV framework?

For the MEV Framework, I followed the approach of Zorzetto et al. (2016), who applied the PWM to fit the Weibull distribution to non-zero rainfall events within one year. They claimed that the PWM attributes a greater weight to the tail of the distribution than ML and is also more robust to outliers. As within the MEV framework, the Weibull distribution is fitted to "ordinary" events, the tails are especially important to investigate the extremes.

Section 4.

L287 and 310 (captions of Figures 4 and 6): "difference calculated as climate model return level minus observational return level" -> difference between the return level from the climate model and the observational one. Why using of the absolute error instead of the relative error?

The relative error is shown. I wanted to express that the relative error is calculated as $(RL\_RCM - RL\_OBS)/RL\_OBS$. I will clarify the captions accordingly.

L454-457: in Zorzetto et al. (2016) the analysis has been made by means of a cross-validation approach, so that the sample used for parameter calibration is independent from the one used for testing the performance of GEV and MEV distributions. When GEV is fitted and tested on the same sample (unless the sample is shorter, i.e. 10-20 years, when issues in the parameter estimation –especially for the shape parameter- might arise), it usually outperforms MEV, but it is not flexible in prediction.

I will correct these sentences accordingly and distinguish between the performance when tested on the same sample and the performance at prediction.

A curiosity: are you considering or discarding snow events?

All daily values of precipitation are considered.

Supplementary material.

FigS2: how are the 95% confidence intervals computed?

The confidence intervals are computed based on the delta method using the covariance matrix. I will add the method to the caption of FigS2.

You have the example for the Munich grid cell, and only for GEV-LMOM and GP models, why not for GEV-ML and MEV? Moreover, a comprehensive validation of all the extreme value models for the whole area would add value to the analysis.

I will add plots for the GEV-FIX and MEV.

FigS5-S6: now the REGNIE product is shown; why not showing the observation-based product used in the analysis? It would be also useful to evaluate differences among the products (even if for some events only).

REGNIE is also observation-based and provides continous daily precipitation values. The observation-based KOSTRA (as used to evaluate the return levels in this study) only provides return levels but no continuous daily precipitation values. Hence, it can't be used to analyze specific events.

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

Kind regards,

Benjamin Poschlod