Dear Editor and dear Reviewers,

Thank you for your time and efforts to review the revised version of the manuscript.
I will briefly address the remaining comments point-by-point. Please also find the tracked-changes version of the manuscript.

Kind regards,

Benjamin Poschlod

Point-by-point answers to anonymous referee 1:

(comments in blue, answers in black)

Line 140: So what is finally meant by daily values? Is it calender day, 00-00UTC/Local Time or anything else?

This sentence is added: “The daily measurement window spans from 05:50 to 05:50 UTC.” The Austrian and Swiss daily measurement windows are now defined in the article as well. They differ by 10 minutes each, so there is no big discrepancy for rainfall events at the country borders.

Lines 170ff: You state that uncertainties or biases prevent the data of a previous study forced with GCMs to be included in further decision making etc at local authorities. Is this the main reason why you now change to reanalysis forcing?

It is an “indirect” reason. The study (Poschlod et al., 2021; ref. in article) using the CRCM5 forced by a 50-member single model initial-condition large ensemble has shown that internal climate variability has major impacts on the estimations of return levels. For this study, where I wanted to test higher spatial resolution setups, no large ensembles are available. Hence, driving the RCM by reanalysis data minimizes internal variability as reason for biases.

This is added to the article:

"The CRCM5 driven by a global climate model ensemble has proven to reproduce rainfall return levels over Europe with good skill (Poschlod et al., 2021). However, the study has shown that internal climate variability has major impacts on the estimation of return levels. Using reanalysis data as boundary conditions strongly reduces this source of uncertainty when comparing with observation-based return levels. As described in Section 1, the resulting return levels of this RCM driven by a global climate model ensemble were presented to local authorities, but local biases prevented further implementation of the results. Therefore, the CRCM5 setup serves as a benchmark."

Sect. 2.2.1/2.2.2: WRF-ERA-I is nested from 75x75km² to 45x45km² to 15x15km² down to th 5x5km² resolution. I wonder why you then perform a direct nest from 75x75km² to 0.11° in case of the CRCM-ERA-I run and not again some kind of nesting.

The "nesting strategy" is chosen by the executing modeller group / institute based on their experience. The CRCM-ERA-I run has been set up by my Canadian colleagues Leduc et al. (2019; ref. in article). This step from 75x75km² ERA-I to 0.11° without nesting is common also within EURO-CORDEX (see Kotlarski et al. 2014, table 1: https://gmd.copernicus.org/articles/7/1297/2014/gmd-7-1297-2014.pdf).
Hence, I add for clarification in the article: "No nesting was applied, as with the RCM setups presented in Kotlarski et al. (2014), which are also driven by ERA-Interim and have a spatial resolution of 0.11°."

Line 280ff and related figures: In some captions you state the CI95 is calculated with 1000 times bootstrapping and in case of Fig.S3 it says "via delta method". Why did you use a different method in this particular case and what can you say about the "accuracy" of both methods?

For this particular case (GEV with fixed shape parameter), the bootstrapping method is not implemented in any R or Python package known to me. Hence, in order to provide confidence intervals, the delta method is applied. The disadvantage of the delta method is the symmetry of the CIs, which is an unrealistic assumption, especially for long return periods. The bootstrapping method shows a slight tendency for the bootstrap sample to generate shorter tails than the true sample distribution (Coles and Simiu, 2003) resulting in slightly more narrow CIs for long return periods (longer than 100 years; Caires, 2007:

https://repository.tudelft.nl/islandora/object/uuid:8d38ef9c-ead4-4b9d-850c-d4dd2e71a34f/datastream/OBJ/download). So in the case of this study (return periods up to 100 years) I would prefer the bootstrapping over the delta method.

Coles, S., and E. Simiu, 2003: Estimating uncertainty in the extreme value analysis of data generated by a hurricane simulation model. J. Engrg. Mech., 129 (11), 1288-1294.

Fig. 3: Is there any serious explanation why the WRF-ERA5 simulates opposite sign in shape parameters e.g. over the Franconia region?

Generally, the chaotic pattern of the shape parameter for all three setups is governed by estimation uncertainty due to the small sample size. If you compare all three maps, there are areas for all three setups where one setup differs from the other two. The pattern you describe for WRF-ERA5 in Franconia is indeed the most prominent. There is no “physical explanation” (RCM, reanalysis data set) for this behaviour. As “high estimation variance of the shape parameter based on the limited available sample size” is already mentioned in the article, no further comment is added to the article.

Line 335ff: are the given thresholds from the observations or the simulations?

Thanks a lot for pointing at this. These thresholds were given based on the simulation of WRF-ERA5. In the old version of the paper the POT approach was only carried out for this model setup. I add the statistical properties for the threshold values for all three RCM setups by including a small table.

Figure S4: What do you mean with empirical estimated return periods and how are they calculated?

The “empirical return periods” are based on the empirical annual maxima (for GEV and MEV) and the 90 events over the respective threshold (GPD), which are plotted via plotting position formula. With n as sample size, k as order rank and P_k as empirical distribution function (EDF), the EDF can be expressed as $P_k = k/n$ (de Haan & Ferreira, 2006). However, inspired by your comment, I now follow Makkonen (2006), who strongly recommends the Weibull plotting position formula, where $P_k = k/(n+1)$. The figures in the supplement are adapted (very slightly) and an according comment is added to the captions.

De Haan, L. and Ferreira, A.: Extreme Value Theory: An Introduction, Springer, 436 pp., New York, 2006.

Makkonen, L.: Plotting positions in extreme value analysis, J. Appl. Meteorol. Clim., 45, 334–340, https://doi.org/10.1175/JAM2349.1, 2006.

Table 1+2: I suggest some kind of sorting/ordering in both tables to make it more readable, e.g. by resolution, bias, etc. If you decide to do so, it has to be stated in the caption as well.

The Tables were sorted by the EVT approaches (GEV-LMOM, GP-MLE, GEV-MLE, MEV-PWM). I admit that this order is not intuitive, and therefore I change the order as you suggest (sorted by resolution and then bias).

Thank you for the useful comments and hints. I hope that your questions and suggestions are answered and implemented sufficiently.

Kind regards,

Benjamin Poschlod