nhess-2020-1982: An analysis on temporal scaling behaviour of extreme rainfall of Germany based on radar precipitation QPE data (Pöschmann et al.)

Reply to the comments from Referee #2:

We appreciate the unbiased review and are thankful for all comments and suggestions from anonymous referee #2 that helped to improve our manuscript. We have provided answers to all comments in blue and updated the original manuscript accordingly. All line numbers given in our responses refer to the original version of the manuscript.

Comments:

General Comments

The authors present an analysis of precipitation intensity-duration relationships over Germany based on the RADKLIM radar dataset, with a spatial resolution of 1 km² and a temporal resolution of 5 minutes. They find a non-smooth scaling relationship, with indications of regime transitions between different temporal aggregation lengths. I think the study is interesting and I'm not aware of a similar study using radar data over Germany. I didn't see any fundamental flaws in the work and therefore think it could be publishable.

Before it can be published though, I think there is room for improvement. This mostly relates to the text and presentation, rather than the science. In some parts the manuscript can be a bit confusing and hard to follow, with aspects explained in a sub-optimal manner. I've outlined my suggestions below.

I should note that, in order to perform an independent and unbiased review, I refrained from reading the already published reviewer comment and therefore apologise for any repetition of what may have already been said!

Main Comments

1. The "three-regime" scaling curve. It wasn't clear to me if the "three-regime" scaling curve you report is a new finding or not, i.e. is there any other literature which report a multiple-regime scaling curve? If the three-regime scaling curve is a novel result, then you should emphasize this. If it is not, then you should cite other studies where multiple-regime scaling curves were reported.

Response: We replied together with your main comment 3.

2. The RADKLIM data set.

(a) As the whole study and its results hinge on the RADKLIM dataset, I think the reader needs to be given more information about this dataset and its limitations, particularly how they might affect the results of an extreme precipitation study. This is particularly important because there is no documentation for the RADKLIM dataset available in English. As far as I know, the only available source is DWD Report No. 251, which is in German (https://www.dwd.de/DE/leistungen/pbfb_verlag_berichte/pdf_einzelbaende/251_pdf).

For example, RADAR data are known to often contain artefacts due to interference (wind turbines, WLAN networks, etc.), which can be particularly problematic when looking at intense events. What

steps were taken in the production of RADKLIM to eliminate or reduce such artefacts? You'll find all the necessary information in Section 4 of DWD Report No. 251 (see link). Of course we can't expect the data and results to be perfect, so we need to transparently present these issues to the reader to help them form their own opinions.

Response: The publication of Kreklow et al. (2019) shows that the remaining weaknesses of RADKLIM are a higher number of missing values as well as an overall negative bias causing an underestimation of high intensity rainfall due to spatial averaging and rainfall-induced attenuation of the radar beam. The authors assume that almost half a million pixels will still provide an adequate representation of the characteristics and feel safe that there is no danger of getting unreasonably high values for the maxima.

Lines 81 – 84 were edited as follows:

"Since the quality enhancement of RADOLAN is ongoing without post-correcting previous data, the so–called radar climatology project of the DWD, RADolanKLIMatologie (RADKLIM,Winterrath et al., 2017) has consistently reanalysed the complete radar data archive set since 2001 for improved homogeneity despite the originally different processing algorithms. Compared to RADOLAN, RADKLIM has implemented additional algorithms leading to consistently fewer radar artefacts, improved representation of orography as well as efficient correction of range-dependent path-integrated attenuation at longer time scales (Kreklow et al., 2019). Whereas RADOLAN is not well suited for climatological applications with aggregated precipitation statistics, RADKLIM is a promising data set for these climatological applications. The RADKLIM data is available.."

Lines 91 – 94 were edited as follows:

"The YW product covers the area composed of 1100 x 900 pixels with the spatial resolution of 1 km (improved compared to former version of RADOLAN). Remaining weaknesses of RADKLIM (as outlined in Kreklow et al. (2019)) are the greater number of missing values (compared below) compared to RADOLAN as well as negative bias causing an underestimation of high intensity rainfall due to spatial averaging and rainfall-induced attenuation of the radar beam."

The following sentence was added at line 151 (centre):

"As mentioned in the data quality description, it is possible that these sub-hourly values do not represent the true extreme across Germany for 2001-2016 since radar-based measurements at fine timescale (e.g. xx minutes) are highly sensitive to the averaging effects."

Following sentence was added at the end of the conclusion:

"Also, the known issue of rainfall extreme underestimation by RADKLIM-YW and the potential impact on the results need further investigation."

(b) The scaling regime transitions at 1 hour and 1 day got me thinking. As we know, RADKLIM uses station data to adjust radar-measured precipitation. According to DWD Report No. 251 (Sections 4c and 4d), this is done with hourly station data where available. If no hourly station data are available, then daily station data are used. The number of hourly and daily stations used can be seen in Fig. 5 of the aforementioned report. The radar data are summed to the temporal resolution of the station data and adjusted, before a "disaggregation" procedure is applied to return the radar data to their original temporal frequency.

Could it be that the different regimes result from this adjustment process? Maybe those pixels adjusted with hourly gauge data tend towards a scaling curve with one characteristic slope, while those pixels adjusted with daily gauge data tend towards a scaling curve with a different characteristic slope? What do the authors think? I think this question underlines the importance of my first main point about whether your multiple-regime scaling curve is a unique finding or not. If a multiple-regime scaling curve has never been reported before (even in radar-based studies), then it would arouse concern that your three-regime scaling curve may be a data artefact. If multiple-regime scaling curves are common, then my comment can likely be ignored.

Response: Thank you for your interesting comment, this has previously been a topic of discussions between the authors. However, we did not agree after all, mainly because Figure 9 would in our opinion show different curve characteristics than it does.

We furthermore hesitate to emphasize on the multiple-regime scaling curve as a novel finding, since it most likely simply results from the lengths of the radar time series. The lack of time length cannot be compensated by more information in space to the extent we hoped. Obviously, some maxima are captured that would not be captured without the detail of spatial information, however for a single scaling regime additional big contributing events and maxima would be needed, which seems only possible with more years of data.

Additionally, for example the study of Galmarini et al. 2004 proposes a simple rainfall model, however, their curves of real data (they evaluated station-wise) in the beginning of the paper in fact show some similar deviations from a straight line. This is similar to what you see when looking at just one pixel in our data set.

One finding of our paper is, however, as mentioned in the discussion, that it might be possible that the regional or world record curves which are often based on fixed interval sampling and "limited" to rain gauge locations COULD potentially look different (= without the clear scaling regime) if more spatial information is considered. Especially when looking at Figure 10, we could show that even a sample size of 1000 pixels can give a very wrong representation of the "real" information. This is similarly true for rain gauge data, since the gauges themselves already represent an exceedingly small unit.

We thus hope that our study opens more discussion on that topic, but we would refrain from making statements that we after all are not able to fully prove (for now).

3. Figure Captions. I think it would help the readers if the figure captions were a bit more descriptive. Generally, they are just one sentence. For example, in the captions you could add some more text highlighting the interesting aspects of the figures, so that it is clear to the reader what exactly the motivation for showing the plot is and why the presented result is interesting. This saves the reader from having to flip back and forth between the text and image (which may be several pages apart).

Response: We edited the captions of the following figures as follows:

Figure 2: Results of NaN analyses of the QPE RADKLIM-YW from 2001–2016: (a) Spatial distribution of the proportion of NaNs (in %) for each pixel, (b) Maximum intensity per time step of 5 min that need to be interpolated (= maximum intensity difference within one time step to overcome)

(Figure 3: removed from manuscript)

Figure 4: Overview of maximum rainfall records in Germany. Chart: Depth-duration relationship of rainfall records based on QPE RADKLIM-YW (data of this study) (blue dots), and as reference the relationships based on the German ground network Rudolf and Rapp, 2003; DWA, 2015; DWD, 2020) (non-filled triangles) and the "world records" (World Meteorological Organization, 1994; NWS, 2014). Map: Locations of rainfall maxima (based on QPE RADKLIM-YW) for the considered duration.

(Note: some studies had been updated due to a suggestion of referee#1)

Figure 5: Depth-duration relationships of rainfall values of whole Germany based on QPE RADKLIM-YW for 2001 – 2016 from maximum values down to the 3921st greatest per duration

Figure 6: Locations of the 0.99999, 0.9999, 0.999, and 0.99 quantile rainfall with varying durations from 5 min to 3 d. Point colours represent the corresponding rainfall duration, similar for each quantile. Different numbers of data points in panels a-d result from several data points being at the same location.

Figure 7: Spatial distribution of the maximum rainfall values retrieved from QPE RADKLIM-YW (2001 – 2016) for different durations (5 min to 3 days).

Figure 8: Depth-duration relationships of rain records for single pixels at rain gauge locations within state capitals of German federal states.

Figure 9: Resulting 6 groups after clustering the maximum depth- duration relationships of rainfall for all pixels. The left panel shows the spatial distribution of the groups, distinguishable by colour. The corresponding curve shapes of 100 randomly selected radar pixels from each group are displayed on the right side with the same colours as the map.

Figure 10: Dependency of maximum depth-duration relationship characteristic on underlying pixel sample size. The maximum rainfall values are derived from (a) 10, (b) 100, (c) 1000, and (d) 10000 random pixels from all considered pixels (n=392 128) within Germany. For each sample size, 30 ensembles are displayed and compared to the overall maximum curve from Fig. 4 and 5 (yellow top line in (a) – (d)).

Minor Comments and Technical Corrections

-Language: As far as I know, NHESS publishes using British English. The manuscript currently uses American English spellings. If you change the language of your spell-checker to British English you should easily be able to find all of these misspellings. For example, "behavior → behaviour", "modeled → modelled", "color → colour", "neighboring → neighbouring", etc.

Response: Thank you for this comment! We have change it accordingly.

-Superfluous text: I think there's a fair bit of superfluous text which could be eliminated. As just a small example, I don't think anyone is interested that (L78) the DWD "is providing different free and purchasable rainfall data derived from it". Maybe I'm being a bit picky here, so you can ignore my comment if you want! Less redundant text is, in general, always appreciated by the reader.

Response: Thank you for this remark. We will try to eliminate it.

-Title: "An analysis *of* temporal scaling behaviour of extreme rainfall *in* Germany based on radar precipitation QPE data" (not `on' or `of')

Response: Thank you! We will request a change of the title!

-L38: Here you've switched from “d" to “D".

Response: We have changed d to D for consistency reasons and also edited lines 33 ff as follows (based on the first referee’s comments):

“Jennings discovered that this unique scaling behavior holds at the rainfall duration between 1 min through 24 months. Paulhus (1965) showed that the same power law relationship holds after addition of new world rainfall record observed at the island of La Réunion at the duration between 9 h and 8 d. The envelope for this extreme values can be expressed as: $P = \alpha D^{\beta}$ (1) where P is the maximum precipitation (in mm) occurring in duration D (in h), the coefficient α (425 in Paulhus (1965)) represents the value at one hour of the depth–duration relationship plotted on the log–log plane, and the exponent β (0.47 in Paulhus (1965)) is the parameter characterizing the scaling behavior of the depth–duration relationship. The Spanish study of Gonzales and Bech (2017) updated the world envelope’s slope to 0.51, showing a remarkable stability. Multiple exponents describing the scaling property of .....”

-L55-: For your discussion of the impact of sparse rain gauge networks (also elsewhere in the manuscript), the very new publication of Lengfeld et al. (2020) based on the RADKLIM network may be particularly interesting for you.

Response: Thank you for suggesting this interesting publication. We have included it in the introduction.

-L88: What are RADOLAN-RY and RADOLAN-RH?

Response: We removed RADOLAN-RH in the text, since it is just the 1-hour sum of RADOLAN-RY. We added a short information in brackets: “... 5 min product RADOLAN-RY (rainfall estimate after basic quality correction and refined z-R-relationship) with the help of RADKLIM-RW ....”

-Eq. 7: Something has gone wrong here. If log(M) = B + b*τ, then M = 10(B+b*τ). Also, there's an open bracket in Eq. 6.

Response: We merged sections 2.2.1 and 2.2.2 into one paragraph 2.2. “Depth-Duration relationships” and removed most of the equations for a better reading.

The methodology section is now shortened as follows:

“2.2 Depth-Duration relationships
Maximum rainfall values for each duration τ between 2001-2016 were calculated with rolling sums applied over moving windows using the R package RcppRoll (Ushey, 2018). Durations of up to 3 d were chosen for the analysis, with multiple steps for minutes and hours out of our interest for sub–hourly and sub–daily pattern. The records may include non-rainfall data and thus do not imply continuous precipitation for the period considered. Values were not aggregated spatially, since this usually reduces the maximum intensity values (Cristiano et al., 2018). First, the extreme values for each pixel and duration $M_{max}^{\tau,pixel}$ are calculated. Afterwards, the overall maxima for whole Germany for each τ (M_{max}^{τ}) is extracted from these calculated extreme values. Based on these results, the depth–duration relationships can be built for each pixel as well as for the whole of Germany.”

Section 2.2.3 is thus changed to 2.3.

-L121: Please state how many pixels exist for the whole of Germany, i.e. N = ? This is also useful to know when we look at the subsamples in Fig. 10. I therefore also suggest repeating the value of N in the caption of Fig. 10 (like you show in Fig. 5) and also somewhere around the line 235-238.

Response: The whole of Germany has around 392875 pixels (uncertainty involved by extracting the data with the polygon of Germany). We evaluated those N = 392128 pixels (compare Figure 2) covered by the German radar data. We left all pixels with data in the analysis, knowingly that some (grey colour in Figure 2) have shorter time series than 16 years.

We added the total number of analysed pixels in line 85: "The RADKLIM data is available in the following two formats, with around 392 128 filled pixels within the German border" and in line 238. Additionally, we repeated the value in caption of Fig. 10 (compare related comment with new figure captions).

-L138: There's an open bracket here too.

-Fig. 3: Personally I think that Fig. 3 is superfluous. You could just state the result in one sentence without showing the plot, and take up less space. k-Means and elbow plots are pretty standard and are unlikely to confuse readers. Do you know that the journal publication charge will be based on the number of pages? Alternatively, if you really like the plot you could put it in a supplementary info file or an appendix.

Response: Referee #1 asked us to give a more detailed explanation of the classification. We still followed your suggestion and removed the Figure 3. The text edit of section 2.2.3 (new: section 2.3) is below:

"The depth-duration relationships ($M_{max}^{\tau,pixel}$ vs τ) for each pixel derived from Sect. 2.2 are individually clustered with the K-Mean clustering algorithm (Scott and Knott, 1974). "Erroneous" pixels (=having NaNs as resulting maxima) were excluded from the cluster process in order to avoid disturbances. The data was rescaled to make the characteristics more comparable with each other. If the number of clusters is not predefined, it can be identified by drawing an elbow chart. For different numbers of clusters K the measure of the variability of the observations within each cluster (Total within-cluster sum of squares, y-axis) is calculated and the curve should bend like an elbow at the optimal value. Since the algorithm did not suggest a number of clusters, we chose six clusters for a sufficiently detailed analysis since it gave consistent results when repeating the automatic algorithm for several times (each time the algorithm clusters slightly differently)." [end of section]

-L143 and Fig. 4: I can't see any "blue solid lines" or "red dotted lines" in Fig. 4! I only see blue dots, black triangles and empty triangles. Also, shouldn't the unit in Fig. 4 y-axis just be mm? The caption for Fig. 4 is also confusing, because it talks about "Spanish ground gauge records" but these aren't visible in the plot.

Response: We changed it to the characteristics shown in Figure 4: Triangles (filled/empty) and dots. The Spanish study indeed is not included in the Figure, so we removed it in the caption.

-Fig. 6: I'm a bit confused by Fig. 6. Why are there a different number of data points in panels a-d? Maybe this could be cleared up with a comment in the text or the caption of Fig. 6. Are some data points "invisible" due to several being at the same location?

Response: Indeed, some data points are "invisible" due to several being at the same location! We added a comment in the caption (see new figure captions above)

-L163: You've repeated “0.9999" here.

Response: We did not notice, thanks a lot! We changed the first one to 0.99999.

-L162: “The lower the chosen quantile, the clearer the scaling regime appears." Is this supposed to mean that lower quantiles show a smoother curve rather than the 3-regime form?

Response: Yes! We changed the sentence to “Lower quantiles thus show a smoother curve rather than the 3-regime form.”

-L164: “The lower the quantile, the sparser the location of the quantile rainfall occurrence ..." I'm confused by this sentence. “Sparse" means “not dense". How can a location be sparse? Are you trying to say that for lower quantile events, the location of the maxima (Fig. 6) are more spread out across Germany? **(s. below)**

-L169: “... the curve shows a very smooth ...". This is the curve in Fig. 5? I suggest writing the sentence as ”... the curve (Fig. X) shows a very smooth ..." This makes it easier to follow for people who are reading the manuscript for the first time.

Response: find the text edition from line 164 (including referee #1’s suggestions/comments):

“It shows that the number of locations increases the lower the quantile of maximum rainfall is. This suggests the reduction of the influence of one single rainfall event on the depth-duration relationship causing inflection in the curve. Additionally, from a certain degree of quantile (Fig. 6 d) the locations of maximum rainfall contributing to the development of the rainfall-duration relationship seem to happen mainly in the wider Alpine region in South Germany. This suggests that rather natural rainfall mechanisms are dominating the scaling relationship, such as regional characteristics and meteorological conditions (e.g. orographic lifting or leeward effects). Naturally, one would assume that this heterogeneity of the meteorological conditions and rainfall generating mechanisms will reflect rather regional characteristics and will exhibit some irregular scaling behavior. Contrary to this conjecture, the curves in Fig. 5 (99.9% and 99%) show a rather smooth scaling behavior” [end of subsection 3.2]

-L177: “The influence of . . . persists until *the* hourly timescale ..."

Response: Thank you! We changed it!

-L188: ”This result also implies the real rainfall process significantly deviates from the assumptions of the simple rainfall models suggested by Galmarini et al. (2004) and Zhang et al. (2013)." Would another possibility be that the 16-year time series used here is too short to see smooth behaviour at the point scale? I don't know how long the time series in the cited literature are.

Response: Galmarini et al. (2004) uses time series of different lengths and resolutions, from 1 year of 1 min resolution up to 160 years with 1day resolution. Zhang 2013 does actually not provide any information on the time length and focuses on 1 time series. Figure 3 in Galmarini et al. (2004) shows that longer time series do not necessarily show smoother scaling behaviour. The 6 min/99 years curve for example shows multiple scaling (also with bends at 1 hour/1 day), and there seems to be no clear dependency on time length or resolution.

We changed the wording of the passage after reading your comment, to: [replacing l.188 ff]

“Furthermore, Galmarini et al. (2004) and Zhang et al. (2013) both showed that the maximum rainfall-duration relationship at a given point location follows a smooth and simple power law if

the rainfall process can be modelled with a set of simple stochastic processes. Our results imply that natural rainfall processes might significantly deviate from this rather simple assumption, also their model frameworks are based on very few time series of very different lengths and resolutions.

-L182: I think here you mean “unprecedented" instead of “precedent".

Response: We mean that the 2002 flooding in Saxony had long heavy rainfall before the event. But we agree, that “unprecedented” would also be a good term to use here thus we changed it accordingly.

-L193: ”The maximum dept-duration relationships in Fig. 8 were clustered since some show a similar shape with each other." Did you really perform a k-Means clustering based on the 15 data points of Fig. 8, as suggested by the text here? This would be highly non-robust. I presume you really did the clustering based on all data points over DE, right? If so, please make this clearer in the text.

Response: We added a half-sentence that Figure 9 clustering is for all pixels as described in the corresponding section 2.2.3
Edition of first sentence in line 193:
“The maximum depth–duration relationships for all pixels within Germany were clustered since Fig. 8 indicated that they might show similar shapes.”

Additionally, we mentioned in Figure 9 also (compare new figure caption) that the clustering is based on all pixels in Germany.

-L194: ”The k-mean clustering successfully classified the depth-duration relationship into six categories ..." The word ”successfully" here is a bit problematic without an objective method for deciding what “successful" is. K-Means will always categorize the data into the chosen number of classes, even if the data are completely unrelated to each other. If “successful" is based on the appearance of the right-hand panel of Fig. 9, a critic could say that to the naked eye there's little discernible difference between categories 3 and 4, or 2 and 6. You could just delete the word “successfully".

Response: We deleted it!

References

K. Lengfeld, P.-E. Kirstetter, H. J. Fowler, J. Yu, A. Becker, Z. Flamig, and J. J. Gourley. Use of radar data for characterizing extreme precipitation at _ne scales and short durations. Environmental Research Letters, 15(8):085003, 2020. doi: 10.1088/1748-9326/ab98b4.

Reference

J. Kreklow, B. Tetzlaff, G. Kuhnt, and B. Burkhard. A Rainfall Data Intercomparison Dataset of RADKLIM, RADOLAN, and Rain Gauge Data for Germany, Data, 4, 118, https://doi.org/10.3390/data4030118, 2019.