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 #4:

We are thankful for anonymous referee #4's comments and suggestions in order to improve our original draft. We have provided responses to all comments in blue and updated the manuscript according to the suggestions. All line numbers given in our responses refer to the original version of the manuscript.

Comments:

In this paper radar derived Quantitative Precipitation Estimates with high temporal and spatial resolution are used to derive depth-duration relation for Germany. The result indicates that the scaling behaviour between the maximum rainfall depth and duration curves don't follow a power law function as previously derived by historical records. Instead, three distinct scaling regimes are identified which boundaries are 1h and 1d. The results are shown for different quantile levels and cities in Germany. Moreover the maximum rainfall depth-relation curves are derived for all radar pixels and clustered according to their shapes. This gave a presentation of the spatial relations and the different rainfall event type occurring over each pixel. The topic is very interesting and relevant and justifies a publication, however the manuscript suffers from several issues that need to be addressed/discussed.

Major comments

- In the introduction, a lot of focus is put on PMP estimation. This is part of the story, however, other aspects related to this topic should be considered and discussed here as well, e.g. rainfall extremes, the problems associated with radar QPE, rainfall extremes as not being a point event but rather a space-time phenomena, scaling properties of extremes, trading space for time, etc.

Response: Thank for this remark, we have added some changes to the introduction:

We have added some text in the first paragraph:

"Extreme rainfall poses significant threats to natural and anthropogenic systems (Papelxiou et al., 2016). The frequency and magnitude of extreme rainfall is expected to increase in the future (Blanchet et al., 2016; Gado et al., 2017; Garcia-Marin et al., 2012; Ghanmi et al., 2016; Lee et al., 2016; Madsen et al., 2009; Marra and Morin, 2015; Marra et al., 2017; Overeem et al., 2009; Yang et al., 2016} especially at sub- daily timescales (Barbero et al., 2017; Fadhel et al., 2017; Guerreiro et al., 2018; Westra et al., 2013, 2014} leading potentially to more urban and non-urban flash floods (Dao et al., 2020), riverine floods, and landslides. A thorough understanding on magnitude, duration, and frequency of extreme rainfall is thus necessary for efficient design, planning, and management of these systems, with many needing (sub-)hourly information especially.

Obstacles to identifying and investigating extreme and record rainfall events are their rare occurrence as well as the spatiotemporal resolution and coverage of rainfall information in general. Lengfeld et al. (2020) analysed the problems of rain gauge observations, missing more than 50% of the extreme rainfall events observed, with even more missed at higher temporal resolutions. Remotely sensed precipitation products with high spatiotemporal resolution such as the ones provided by radar, satellite or microwave link networks may solve this issue. For rainfall extremes, weather radar systems are seen to be appropriate to capture the spatial variability and extreme events with limited spatial extent (Borga et al., 2008).

However, most of the currently available radar QPE (quantitative precipitation estimates) data sets do not cover very long periods (Lengfeld et al., 2020), while their high spatiotemporal resolution is superior to many other rainfall products. Radar products also have well-known uncertainties, like variation of reflectivity with height, relating radar reflectivity to precipitation rates, clutter and beam blocking). Therefore, their processing is subject to improvements, and reprocessing these data sets is necessary in order to achieve homogeneous and consistent products that can be evaluated for rainfall characteristics over space and time."

The next paragraphs on the PMP were shortened and changed (including comments by other referees):

"Probable maximum precipitation (PMP) is one way to define extreme rainfall. It is defined as the theoretically greatest depth of precipitation for a given duration that is physically possible over a particular drainage basin at a particular time of year (Ralph E. Huschke, 1959). One of the methods to estimate the PMP is the maximum rainfall envelope curve method, that plots the depth (y)–duration (x) relationship of the record rainfall events observed across a large geographical boundary (e.g. entire country or globe) on the log–log plane. The PMP is then derived as a straight line on the plot representing the upper boundary of the envelope containing all depth–duration relationships. This maximum rainfall envelope curve method was first proposed by Jennings (1950), who showed that the depth of the extreme rainfall events observed across the globe is a power function of their duration. Jennings discovered that this unique scaling behaviour holds at rainfall durations between 1 min through 24 months. Paulhus (1965) showed that the same power law relationship holds after the addition of a new world rainfall record observed at the island of La Réunion at the duration between 9 h and 8 d. The envelope for these extreme values can be expressed as:

:

$$P = \alpha D^{\beta}$$

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 Gonzalez and Bech (2017) updated the world envelope's slope to 0.51, showing a remarkable stability. Multiple exponents .."

Continued change from P3L64:

In this study, we analyse the rainfall depth-duration relationship for the whole of Germany based on 16 years of RADKLIM-YW, a carefully reprocessed QPE radar product with 1 km - 5 min space-time resolution. We want to answer the following questions regarding the scaling behaviour of the maximum rainfall values: (1) Does ….."

- Furthermore, I would expect the extremes detected by radar to look differently depending on the distance from the radar and the height above ground since the size of the radar bins increase with increasing distance and so does the elevation above ground. Thus, I would expect less severe extremes towards the outer areas of the radar circles. For example, many of the 5 min extremes in Figure 7 seem to be located near the sites of the radars. Last but not least, even though data correction was applied by the DWD, there is still uncertainty in the observed data, especially for extreme events. This should be mentioned since the results are derived from this product.

Response: Thank you for this comment. As shown in Figure 1 below (different colour scheme than Figure 7 in the original manuscript), the radius-dependency could not be identified by our analysis. When starting from the finest resolution of 5 minutes (compare Figure below with different colour scheme than Figure 7), we do not see any accumulation of extremes located near the sites of the radar and it is also not the case the more we aggregate. One possible reason is that a principle of RADOLAN is to take the maximum value where radar circles overlap (=outer areas of the radar circles) which would lead to a rather accumulation of higher values in the outer areas (this is not the case in our results). Additionally, because of orography we don't expect a general dependency on distance from radar and height above ground within the data. As Kreklow et al., 2019 explained, remaining weaknesses of RADKLIM are a higher number of missing values as well as an overall negative bias causing a rather "underestimation" of high intensity rainfall due to spatial averaging and rainfall-induced attenuation of the radar beam. We generally think that with the huge number of pixels evaluated, the analysis still provides an adequate representation of the characteristics.

We agree that a little more text on the data quality would add to a better understanding and thus edited several passages in the documents as follows:

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."

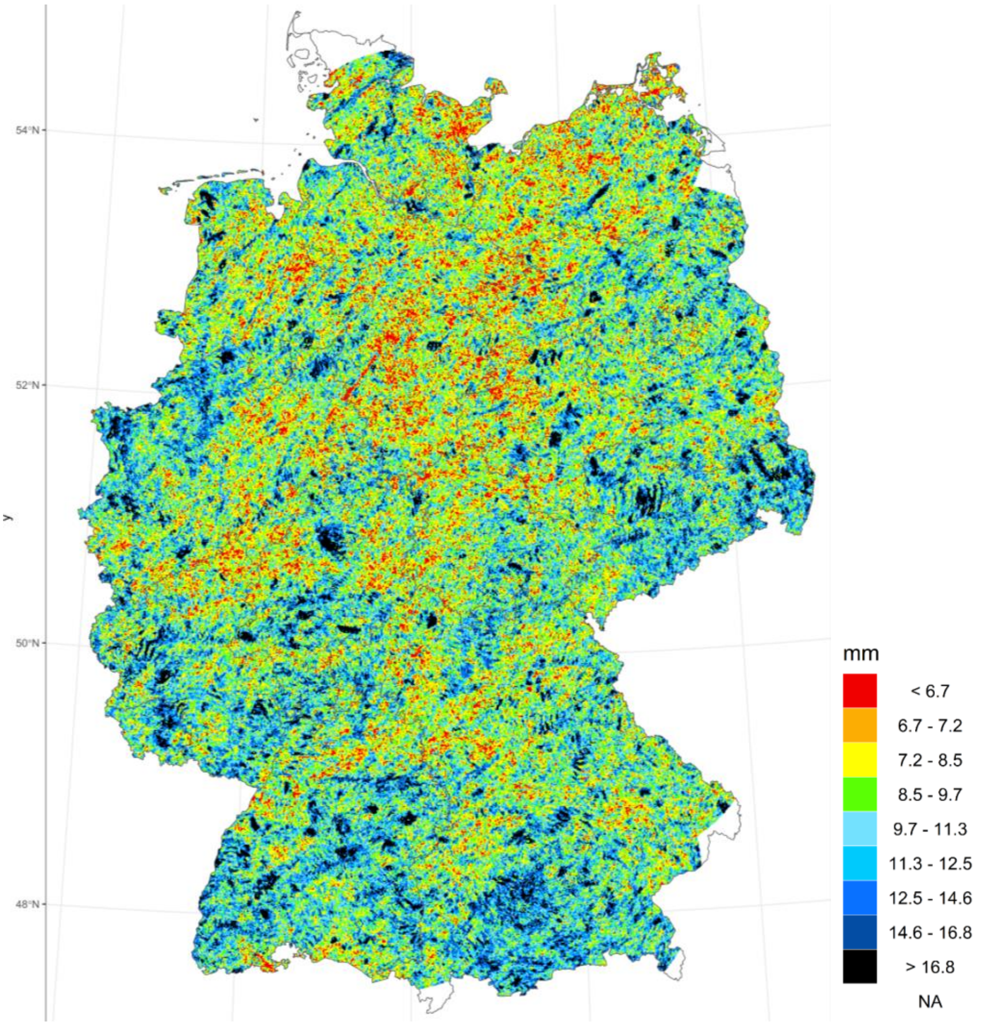


FIG. 1: 5 Minutes Extremes for 2001 - 2016 derived from RADKLIM-YW

- I'm not convinced by the clustering applied for the scaling behaviour. The number of 6 clusters seems arbitrary and section 2.2.3 is poorly written, also with respect to the missing values. The k-means does not provide any measure for the quality of the classification. This all has implications on the results discussed in Section 3.5. How would the results look like if you chose 4 or 7 clusters? Did you perform a sensitivity analysis on how the results change of the number of clusters is changed? Could e.g. a fuzzy-logic based algorithm maybe yield better results?

Response: Thank you for this comment. We agree with the reviewer in that our K-means clustering analysis may be subjective even though the number of the clusters chosen in this study (6) was the result of our careful visual analysis based on 2 through 8 clusters. However, we also would like to argue that the primary focus of this cluster analysis should not be the validity of the cluster numbers or the analysis methodology, but the fact that the scaling behaviour of the rainfall extremes can be classified based on the inflection points in the scaling relationship and this inflection points are primarily governed by the history of the extreme events that each of the radar pixels experienced. We believe that this analysis is particularly meaningful in that it specifically showed, from a spatial perspective, why the rainfall extreme value scaling behaviour deviates from a simple power law.

Since Referee#1 wanted a more detailed explanation, we have edited section 2.2.3 as follows (including an adaptation of the NaN part, but removed Fig. 3, since Ref#2 found it too superfluous). Please note that sections 2.2.1 and 2.2.2 were also summarized to 2.2, thus 2.2.3 will become section 2.3. in the revised paper:

"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]

- Furthermore, I do not understand the concept behind Figure 8b. It shows the maximum difference before and after a gap, but what does this mean if >50% data are missing as e.g. in the northernmost part in Germany? Could you explain this more clearly?

Response: Thank you for your comment. Are you referring to Figure 2b instead of Figure 8b? The reason for >50% of missing data is because radar coverage started relatively late for the concerning pixels (sometimes only from 2014 onwards or so), thus the time series are a lot shorter. The Figure should support that NaN imputation might be problematic for our data. Other reasons than the large amount of data is that firstly, too many pixels are with quite a lot of missing values (4% of data means already almost 70,000 timesteps of 5 minutes). Furthermore, the imputation bridge may yield very high values adding large uncertainty. As mentioned in the text, imputing unrealistic high values in these gaps is likely to add bias to our results more severely than if we just keep everything as is. Maybe the NaN section might be unnecessary for you, we however feel that it is more transparent to display our thoughts behind what we did.

- Why are the values generally higher the closer they are to the radar site? Could this also point to the different behaviour of extremes depending on the distance from the radar?

Response: The values are lower in the outer areas of the radar circles potentially due to the overlap of radar information in these areas. Fewer and shorter NaNs sequences in the time series will also reduce the number of high imputation values.

Minor comments

- In general, the manuscript should be proof-read again, there a many awkward formulations,spelling mistakes, etc.

Response: Sorry for that! We have sent the revised document to proof-read again.

Specific Comments:

- P1L17ff.: This whole sentence sounds weird, fatal disasters don't react to anything

Response: see above (first paragraph in new introduction)

- P1L20: Introduce the acronym PMP here

Response: We added it.

- P2L24: PMP can be estimated

Response: We changed it.

- P2L52: AR(1) -> first-order autoregressive process?

Response: We changed the sentence to "Zhang et al. (2013) showed that the scaling exponent varies around 0.5, if the vertical moisture flux and rainfall can be modelled as a censored (or truncated) first-order autoregressive process AR(1)."

- P3L58: Breña-Naranjo (this needs to be corrected in the references as well)

Response: Sorry! We changed it.

- P3L63: 16 years

Response: We changed it.

- P4L77: Aren't there currently 17 C-Band Radars?

Response: Sorry! We change it.

- P4L78: delete "free and purchasable"

Response: We deleted it.

- P4L81: ground information

Response: We changed it.

- P4L90f.: I find the justification that "Due to comparison reasons with another study at our institute only years 2001 to 2016 had been used for this study" rather weak. It would have been worthwhile to use the data until 2018, since you also mention that "With longer available time series of radar in the future, the deviation can be further investigated and tested" in the discussion.

Response: Thank you for your comment, we are sorry that we cannot find a new reason to make the justification sound better. We agree, that longer time series are always better, however, we would need to find a better data handling approach in order to analyse years 2017 + 2018. The preparation of the manuscript together with the final analyses took longer than expected. With the current data handling approach it would take too long to 1) process the data, 2) redo the complete analysis and 3) interpret the results, that is why we kept the analysis to the years 2001 – 2016. We will see in further work if our argumentation was correct or not. Thank you for your understanding!

- P4L95-100: there is no need to mention the data size or how the data was saved.

Response: The data size seems very relevant for us. Without high-performance computing and sufficient storage capacity, it will not possible to do complex analysis of RADKLIM-YW in a reasonable amount of time. Due to these reasons, it took a while to convert to time series (obviously better programmers will have no problems with this issue).

We removed the following passage: "Analysis was conducted in R ……… was chosen to store the data." (2 sentences)

- P4L99: Why don't you use "NaN" for missing values?

Response: We changed NA to NaN in all 6 occurrences.

- P5L103-105: Are data of overlapping radar coverage areas similar? Since the data was measured by two different radars, the values can differ significantly (e.g. Yan and Bárdossy, 2019)

Response: We totally agree that radar data within overlapping radar cones can differ depending on the radar. In lines 103-105 we are however not evaluating the quality of the radar data, but focus on the data coverage, which increases when having more available (even different) information. The work of the DWD has been to merge the different radar information in a consistent way.

- P5L107ff.: This whole paragraph is difficult to read, please rewrite this in a clearer way.

Response: We edited the text as follows:

(P5L107ff):

"It is hard to handle NaNs in highly episodic geophysical events such as rainfall. Based on Fig. 2, we chose to not do any data interpolation, since the consequence of imputing potentially too high extreme values is more severe and uncertain for our study than the missing of any extreme values."

- P6L115: How was the aggregation done considering the missing values? And how where events separated? Did you use a threshold? If yes, which?
  P6L116: Durations of up to 3 h or 3 d? This whole sentence is difficult to read.
  P6L130: Is the scaling relationship formulas are not correct

Response: The aggregation was done with rolling sums applied over moving windows (compare text edit below), ignoring the missing value (=treating them as zero when "rolling over" them). The evaluation is not event-based but time based, thus events are not separated. We also have done the analysis event-based out of interest, but obviously most events stop after a few days, thus this approach is not useful if looking at maxima across different scales. The event-based analysis will also not necessarily give the maxima for a certain time period, since some maxima are the sum of several short heavy events.

The authors agree that the methodology part might create unnecessary confusion (following Referee #1's comment). Thus, 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."

- P7L138: Please reformulate sentence (also see major comments above)

Response: Please see our text edits that are included at the response to the corresponding major comment above.

- P7L143: The temporal resolution of the ground truth reference should be mentioned. Furthermore, "world record" should be used (also in caption of Fig. 4)

Response: We used "world record". We have tried to find the temporal resolution on which the world record curve is built. However, even when checking papers that individually treat certain historical rainfall events/maxima, the temporal resolution was not found for most cases.

- P8L149: what are "very distant places of Germany"? Distant from what?

Response: This means that the places are not close to each other, rather distant from each other. We edited the sentence as such: "..even though the maxima are observed rather randomly across the whole of Germany."

- P8L151f.: Which temporal resolutions did you use for your analysis? 5 Min increments up to 16h? 3 days? Please specify!

Response: All analysis is based on the 5 min temporal resolution of RAKLIM-YW. P6L116: Maximum rainfall intensities were retrieved from aggregating the 5 min values for each duration of interest .....

- P8L156: Data uncertainty is mentioned here but it's effect is not discussed!

Response: Data uncertainty is also mentioned in the data description; however, the effect cannot be estimated well, without doing a separate study on the radar QPE data processing scheme/algorithm itself. This is partly already done by the DWD, for example in Kreklow et al. (2019), however, we expect the DWD to do further publications about it. We could find many references discussing the effect of radar rainfall uncertainty on hydrologic responses, but not on scaling behaviour. We guess this is because this is one of the first studies in this regard. If the reviewer suggests references, we will carefully address them in the manuscript.

- P8L163: the first two quantiles are identical

Response: We changed the first one to 0.99999

- P8L167: development the rainfall-duration ! development of the rainfall-duration

Response: Thank you! We changed it.

- P8L170f.: The statement that extreme rainfall events share common characteristics such as peak rainfall depth and correlation structure regardless of time-scale is a 'strong' statement that somehow contradicts the fact the rainfall extreme are spatially and temporally variant and their correlation structure differs.

Response: We agree to the referee and have removed the statement.

The final sentence of Chapter 3.2 is changed to: "Contrary to this conjecture, the curves in Fig. 5 (99.9% and 99 %) show a rather smooth scaling behaviour."

- Figure 6: A discrete colour bar should be used, moreover the spacing between the durations does not reflect the real spacing. An additional suggestion would be to add a second colour bar showing the associated rainfall values. This can help relate the quantiles to the rainfall values.

Response: We changed the colour bar to a discrete one (see Figure below), but think that a second colour bar will add more confusion than that it helps. We further believe that a reader can relate the quantiles to the rainfall values with the help of Figure 5 if necessary.

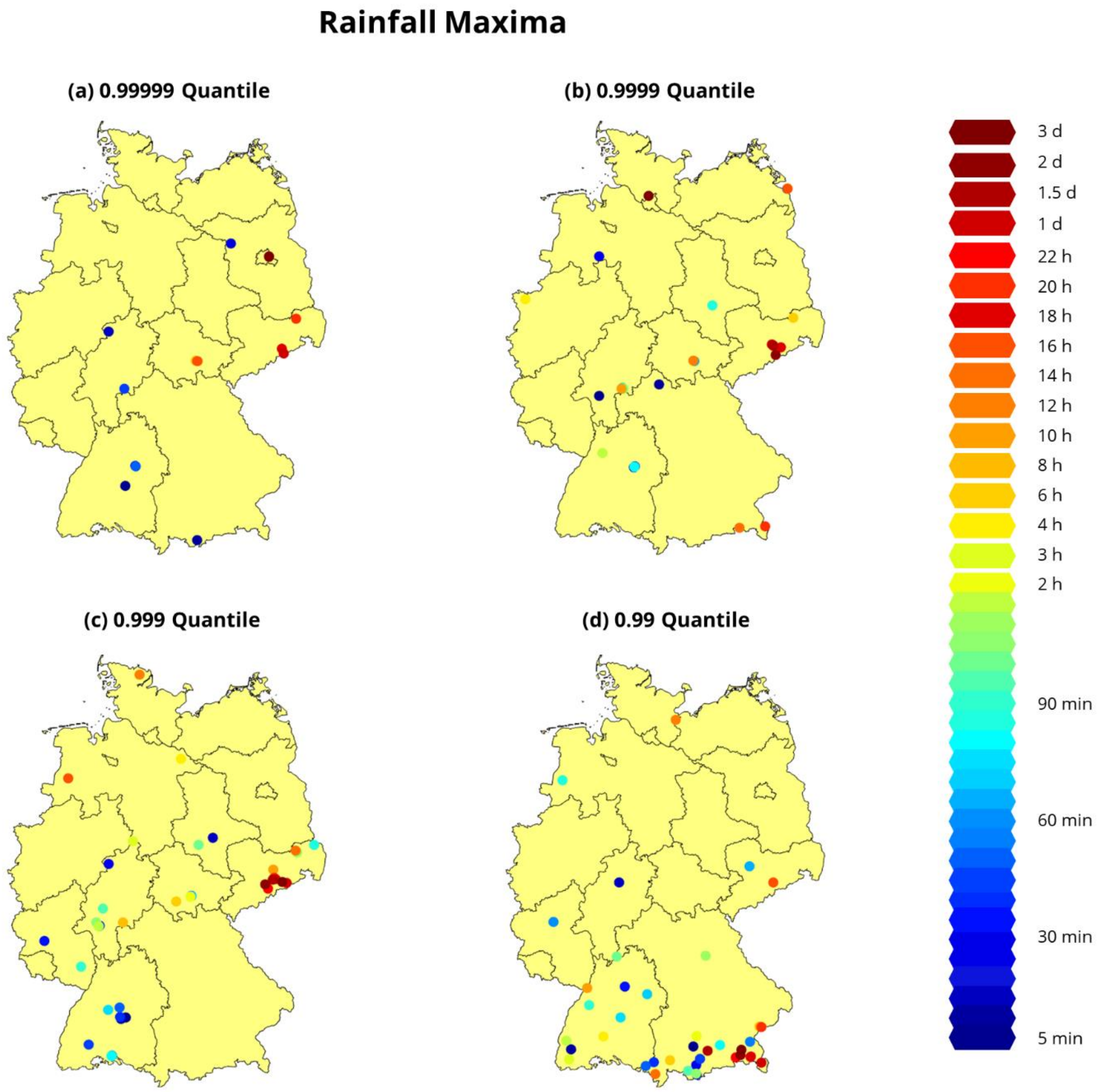


FIG. 2: Figure 6 with new color bar. Caption: Locations of the 0.99999, 0.9999, 0.999, and 0.99 quantile rainfall with varying durations from 5 min to 3 d. Point colors represent the corresponding rainfall duration, similar for each quantile. Different numbers of data points in panels a-d result from several data points being at the same location.

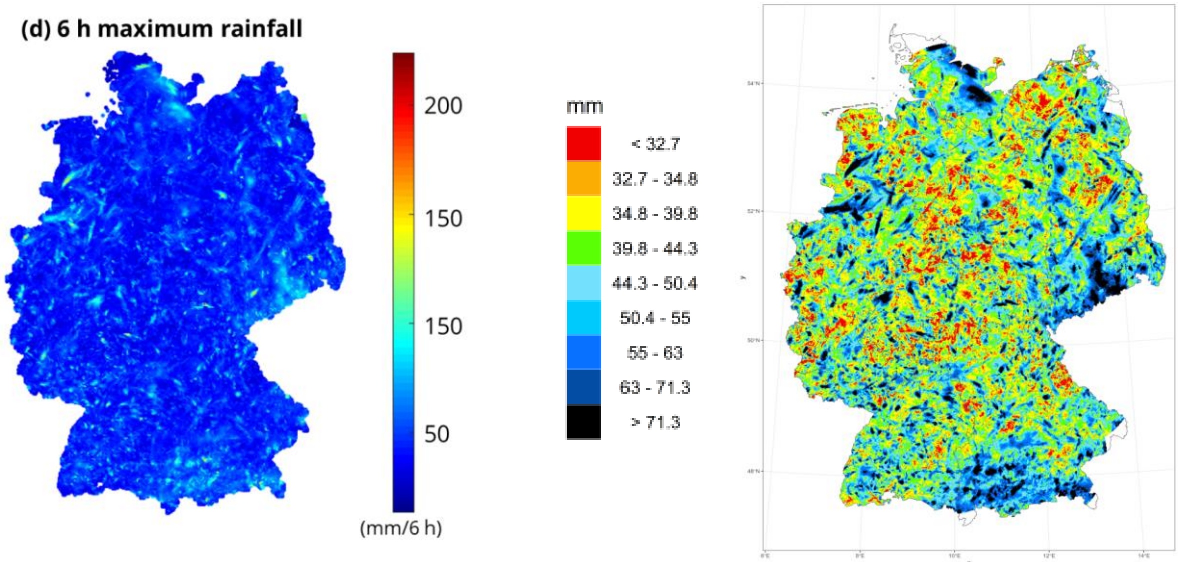
- Figure 6: Why are the 0.99 quantiles are mostly located in the south of Bayern?

Response: We have edited the text corresponding to Figure 6 (P8L163 ff) and hope that it will answer your question:

"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 leewards effects)."

- Figure 7: please redo with a discrete colour bar and maybe scale it to the values so the details in the map are more visible.

Response: If the referee would find it more suitable, we could change the colour scheme as given in the following example (colours represent quantiles):



- P12L180: avoid formulations such as "really high"

Response: We changed it to "These maxima seem to be dominated by single events or single heavy rainfall occurrence."

- P12L185: Stuttgart

Response: We changed it.

- P12L187: who

Response: Sorry! We changed it.

- P12L188 what do you mean with "real rainfall process"?

Response: "Real" corresponds to "observed" in comparison to "modelled/simulated" process. We edited the sentence to: "This result also implies that natural rainfall processes significantly deviate from the assumption of the simple rainfall models suggested by ...... "

- Figure 8: How are curves for the cities calculated, mean of all cells in the city or maximum cell? This might be relevant to explain why neighbouring cities show very distinct behaviour.

Response: We changed the caption of Figure 8 as follows: "Depth-duration relationships of rain records for single pixels at rain gauge locations within state capitals of German federal states."

The gauge locations as x-y coordinates (provided by the DWD via an online repository) are transformed into RADKLIM projection (with potential slight geographic shift due to assignment of point data to raster data). We chose these cities since they are relatively well spread over Germany. We could have also selected random pixels, but preferred to choose known locations.

We don't understand the second part of the comment. The minimum distance between the cities is approx. 80 km, a distance which usually guarantees different characteristics. Even in the small structure of Germany, already the two different oceans influence the meteorological characteristics significantly besides multiple other influences (continentality, luv/lee-ward effect, ..). Only two cities, Mainz and Wiesbaden are rather close to each other. What distinct behaviour of the two needs explanation?

- P13L193ff. Rewrite this paragraph and be more specific, what is "successfully classified", what is a "certain colour", etc.
- P13L193: dept-duration ! depth-duration

Response: We have edited the paragraph as follows:

"The maximum depth-duration relationships for all pixels within Germany were clustered since Fig. 8 indicated that they might show similar shapes. The k-mean clustering algorithm classified the depth-duration relationship into six categories revealing different curve characteristics regarding the curve shapes. Figure 9 shows a categorical map of Germany representing each category with an individual colour. Additionally, depth- duration relationships at 100 randomly chosen grid elements from each category are shown with the regression line from Category 5 as reference."

- P13L199: "that pour for around 1 h and move on or weaken" -> rewrite this

Response: We have edited it as follows:

"The behaviour of the curve between 5 min and 1 h is associated with strong convective rainfall events of around 1 h within the corresponding pixel."

- Figure 9: Does this relate to topography?

Response: We would partly relate it to topography as orographic rainfall does play an important role for certain rainfall durations. As written on P15L209 ff, we attribute some large clusters to orographic rainfall, whereas some clusters can be identified as other large scale events without a direct relation to topography.

- Figure 9: Legend of plot 'mm/uration' ! mm/duration

Response: Following the first referee's recommendation, we have changed all mm/duration to simply mm.

- Figure 9: How would the clustering and this map look, if the data was divided between summer and winter period? Did you look into this?

Response: We unfortunately did not look into a division between summer and winter. It is a good idea for a follow up analysis!

- P15L204f.: If the look similar and occur together why do you distinguish these categories? (c.f. major comments above)

Response: Please compare our answer to the corresponding major comment above. We preferred to first to the automatic grouping, afterwards reduced the groups as best as possible with our expert knowledge.

- P15L205: a slope is steeper instead if higher

Response: Thank you! We have changed it!

- P15L213: the term 'super-daily' is confusing, please consider changing it.

Response: We believe that the term is an appropriate way of addressing everything beyond "day".

- P15L219: saying that areas with category 5 have never been hit by any 'extreme' extreme event needs more evidence. It could be that the occurred events were not well captured due to data uncertainty.

Response: Thank you for your comment! We made this statement based on what we get from the data. Every pixel's "extreme" extreme value for different duration was extracted, thus it is not wrong to say that the pixels of category 5 show certain characteristics. In our opinion, we generally have no "hard" evidence" that products from remotely sensed data deliver "true" information, especially for remote areas with lack of supporting information, we mention the data uncertainty in the beginning and should then work with what we got.

We added in the beginning of the corresponding sentence in L219: "Based on the data set, these regions/locations … " and hope that it makes it clear enough.

- P15L232: areas don't "experience" a rainstorm…

Response: We changed it to:

"… at a given point varies location by location based on the occurred rainstorms." [removing the "areas"]

- P16L232: … the same goes for pixels!

Response: We changed the sentence in L262:

"The shape of the curve was governed by the temporal structure of the extreme rainfall events at the pixel location."

- P16L264f.: Reformulate this sentence

Response: We edited it as follows:

"The scaling behaviour thus can be significantly different for each pixel, because the rainfall characteristics for each pixel are very different as well."

- Figure 10: Please add legend and increase the grid visibility.

Response: We added a text in the caption and increased the grid visibility

Caption 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)).

- P17L268f. If you have the data until 2018, why didn't you use them? (c.f. P4L90f.)

Response: compare answer to P4L90f.

References: Several issues with capitalization of titles and author's names.

Response: Thank you for your helpful comments. The data was mainly extracted via the doi automatically in JabRef/Mendeley. Afterwards, we have revised the references to the best of our knowledge. We don't know exactly which titles you are referring to, but we think that we took the capitalization as given in the journals, e.g., Jennings 1950 and Paulhus 1965: all capitalized (Monthly Weather Review had it that way at the time). "Intensity-Duration-Frequency" is sometimes capitalized, sometimes not. We kept it in all cases as in the journal.

We have corrected the authors' names in:

- Breña-Naranjo et al. 2015

We have changed the following references after revisions:

- We have removed Marra et al. 2016 since it is the preprint of Marra et al. 2017
- We have replaced Cristiano et al. 2018 (preprint) by the final revised paper
- We have updated Lengfeld et al. 2019 (preprint) to the final revised paper

We also have removed some double urls.

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.