Review of **An analysis on temporal scaling behavior of extreme rainfall of Germany based on radar precipitation QPE data** by Pöschmann et al.

October 27, 2020

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 km^2 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.

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.

(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.

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).

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.

-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.

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

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

-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.

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

-Eq. 7: Something has gone wrong here. If $\log(M) = B + b \cdot \tau$, then $M = 10^{(B+b \cdot \tau)}$. Also, there's an open bracket in Eq. 6.

-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.

-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 final 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.

-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.

-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?

-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?

-L163: You've repeated "0.9999" here.

-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?

-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.

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

-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.

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

-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.

-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".

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 fine scales

and short durations. *Environmental Research Letters*, 15(8):085003, 2020. doi: 10.1088/1748-9326/ab98b4.