We are grateful for the careful, proficient and helpful review of our manuscript. Our replies are inserted in blue in the following.

Interactive comment on “Rain erosivity map for Germany derived from contiguous radar rain data” by Franziska K. Fischer et al.

Anonymous Referee #2

Received and published: 29 October 2018

The data were noisy so significant data treatment was applied to produce a “typical” erosivity map and an annual cycle of erosivity. The new values show greater erosivity than previously produced maps and the seasonal distribution shows an increase in winter erosivity. The main advantage of the new approach is the use of continuous data over the region.

General comments

The article is very well written, the analyses are clearly described and figures are well chosen. The results will clearly be of use.

However, some of the choices made in data treatment require further justification. My primary concern is about the level of data treatment that has been applied, which is, as the authors state, “extraordinarily large”. Because the amount of smoothing applied is indeed more than normal, it should be carefully justified.

All existing erosivity maps (mostly unintendedly) employ pronounced smoothing (presumably much stronger than we did). We explain this now in more detail in the Introduction.

Due to our high data availability we were able to and had to replace the unintended and uncontrolled smoothing of existing maps by a statistical protocol because of the large small-scale spatial variability (which was not known previously due to a lack in suitable data). In contrast to existing maps where the smoothing steps and their effects are largely unknown, we define a statistical protocol, we quantify the effects of this protocol, and we justify the protocol.

Importantly, we also provide the maps without the different smoothing steps in the Appendix and leave it to the decision of the reader, which one to use.

The aim is to produce a map of “typical” erosivity over Germany, but the erosivity distributions in time are skewed and contain outliers (from rare, extreme events) that make finding one representative value per pixel a challenge. A related problem is possible sampling effects, meaning differences between the sampled and true distribution of values (the authors mention this with respect to measurements from gauge networks that may miss entire events). The authors have applied data transformation techniques to find typical values, smooth them in space, and smooth the evolution of the average erosion index over time.

I’d like to comment on each data transformation undertaken, first to produce the per-pixel values:

1. Winsorizing:

For each pixel, the mean erosivity over 17 yearly values is taken using winsorizing. In this case the authors only replace the lowest and the highest value (with the second-lowest and second-highest respectively). How was the choice made to use winsorizing over, say, the sample median? The choice of the method used (e.g. sample median or winsorizing, and the amount of winsorizing used) should be justified – for example through the use of a density plot of erosivity values, in which the skewness will be clearly visible, to show that the final values produced are representative of “typical” values of erosivity.

From a statistical point of view a median or a geometric mean has the advantage of being more robust compared to the arithmetic mean, while the arithmetic mean has the advantage of being unbiased. From a modelling point of view, a median or a geometric mean is inacceptable because modelling individual events would then lead to a different total soils loss than modelling long-term mean soil loss using a robust estimator (due to its bias). Winsorizing was thus an (accepted) method to reduce the effect of extreme outliers (not to reduce the effect of skewness, which is an important property of erosive rains).

The effect of one-step winsorizing, which considers only the most extreme years, was already small. Using a two-step winsorising (replacing the two highest and the two lowest values) would have had an even smaller additional effect.

An acceptable alternative to our approach would have been not to use any winsorizing and kriging but to calculate arithmetic county means (on average 35 km² in size) because county means are likely used in the final application of the map (this is why we also will provide a table with county means). We decided to use the more laborious approach because we expected that would better preserve landscape features smaller than 35 km².

2. Bias-correction:
The authors state that the winsorized mean is biased for long- tailed variables. But, for skewed distributions the winsorized mean should be closer to a “typical value” of the population than the sample mean because of the removal of outlier values. So is it not the case that winsorizing produces a less biased estimate of a central tendancy than the sample mean, and the bias correction suggested by the authors undoes the benefit of the winsorizing by matching back to the (spatial) sample means which are themselves affected by outliers?

The reviewer is right that a winsorized mean should be closer to the expected long-term average. Winsorizing reduces the overweight of an extreme event (upper outlier) at a certain pixel. This reduced overweight should be balanced by reducing the underweight (lower outlier) at a second pixel in order not to cause a bias of the average. For normal distributed variable this works well while for skewed variables this balancing is not fully achieved. This is why we had to correct for the bias in order to arrive at the arithmetic average. This does, however, not cancel the effect of winsoring, because the mismatch in balancing one pixel was not put back to the same pixel but it was distributed among all 455 309 pixel. This is why the bias correction was so small (2.3%) that it could almost be neglected.

3. Ordinary kriging:
Kriging is used to fill gaps not covered by the radar data (due to beam-blocking, for example), and block kriging is used to smooth the output field. Kriging requires at least roughly symmetrically distributed input data (ideally they would be normally distributed) so that mean values are representative. It should therefore be mentioned in the article whether the distribution of “typical” values after the winsorizing procedure is symmetric, and if not whether steps have been taken to correct for this (possible options are a log transformation and/or use of the Cressie variogram estimator). Block kriging is being used in a non-standard way, as a smoother, so that each 1×1 km2 pixel is estimated as the mean of values across a 10×10 km2 block. How was the block size of 10×10 km2 chosen?

Long-term average erosivity does not have the extreme skewness of individual events. The statistical distribution is not normal but may be multimodal because it reflects the contribution of different landscapes (mountains, plains, coastal regions etc.) to total land. The assumption of normal distribution is even more justified as we use only distances of up to 100 km (i.e., within this distance the data have to be roughly normal distributed). Furthermore, as we do only use kriging as a smoother, the weights of the data would only slightly change with a different semivariogram. A moving window presumably would have done the same job.

The block size was arbitrarily chosen. The entire smoothing steps are rather irrelevant because they do not change the pattern but only provide a slightly smoother map that makes reading easier (please compare the untreated data in Fig. A3 with the final map in in Fig. 2). We provide both maps and leave it entirely to the reader whether he prefers using the unsmoothed Fig. A3 or the smoothed Fig. 2.

After the spatial processing, the annual cycle of erosivity is calculated. Afterwards, smoothing was applied to the daily timeseries of averages. Again commenting on each step:

1. Daily erosion index:
The erosion index is calculated for each pixel and then averaged across space for each day of the year. It was not clear to me whether the pixel values used to make this average were treated in any way (kriged perhaps?) or were raw 1 km^2 values (I assume it was the raw values so that they were daily). If the distribution of daily EI values (across space for each day) is heavily skewed, then the mean of their values may not be representative (the median or a winsorized mean, for example, may be better). Was any testing for this done?

Every day was the sum of 17 yr and 455 309 pixel (i.e., n = 7.7 million). To our surprise there was this large scatter between subsequent days (indicating that 17 yr are still short and that pixel cannot replace time). We did not remove any data because a high average value with n = 7.7 million can only happen, if very many pixels at that day had a high value (i.e., erosive events occurred in many regions at the same day; this is the opposite of an outlier). The lowest value is zero and it can also not be removed because this would be the majority of all days (days without an erosive event).

The main question is whether the daily erosion index varies among landscapes. In order to analyse this, we added in the revision the daily erosion index for the SE, SW, NE and NW quadrant of Germany. These four quadrants differ in climatological properties (e.g., continentality; annual rainfall and altitude) but the resulting pattern was identical for all quadrants (correlation between them yielded $r^2 > 0.998$) indicating a stability of the seasonal pattern. This close correlation also indicates that the influence of outliers is small.

2. 13-day centred median, 3-day skip mean, and 25-day centred hanning mean:
This choice of smoothing routines needs to be better justified. Why was this combination of window sizes (13, 3, and 25 day) and operators chosen, and how was it judged whether the smoothed values represented the true signal? As a suggestion that may provide more information to the reader: the authors could consider displaying maps not only of winsorized mean of annual values, but also per- pixel median, 10th percentile, and 90th percentile, to show not only “typical” values but maps of extreme erosivity values as well, and to show the spread of values for each pixel.

We added a paragraph on smoothing in the Introduction.

In short, the statistical recommendation is to use smoothing until the information can be seen that is intended to be shown. This recommendation may unsatisfactory for the inexperienced and appears arbitrary. However, there cannot be a statistical criterion because the degree of required smoothing depends on the intended application, which is outside statistics.

We could have used a cumulative density function (cdf) instead of a probability density function (pdf) to describe seasonality. This would have required no smoothing at all but with the cdf it would have been more difficult for the user to assess the effect of different crop management options. The cdf of the unsmoothed data correlates with the cdf of the smoothed data with $r^2=0.9998$ (n=365), proofing that smoothing has not caused any relevant change in seasonality. We added to the manuscript:

"Despite the strong smoothing that was necessary for the probability density function, the smoothing did not change the cumulative density function (which is used for calculating C factors). The cumulative density functions of the original data and of the smoothed data correlated with $r^2=0.9998$ (n=365)."

The window sizes are typically used values. A median with n=2 would produce an arithmetic mean. The advantage of the median filter compared to the arithmetic mean increases with increasing window size. At a window size of n=365 identical values for all days would result and the seasonality would be completely destroyed. A window size of n=13, which is also about the time window of individual crop management practices, ascertains that the seasonality is maintained. A skip mean is always applied with n=3. For the hanning mean a similar reasoning can be put forward as for the median filter. Due to the decreasing weights, the window size can be somewhat larger than for the median filter and window size has less influence (even for n=365 some seasonality would be retained). Also the order of the filters is quasi standard. A median filter only makes sense if it is applied in first place. After some averaging has occurred, the median filter loses its advantage compared to the arithmetic mean and both would produce very similar values. The skip mean cannot be the last filter because it distributes a high (or low) value completely to its neighbours and a high (or low) value is replaced by the low (or high) values of the neighbours. The skip mean thus inverts the pattern within the window size of n=3. Hence, it has to be followed by some averaging, which puts the high (or low) value back in place because during averaging some information is inherited from the neighbours on both sides. The hanning mean is especially versatile in this case because the weights (or the window size) allow adjusting to some degree how much of the high (or low) values is put back in place and how much is left with the neighbours.

Again, the statistical treatment is not critical because we also display the original data. The reader can decide whether he likes our smoothing or not. He may draw his own line where he thinks that the line should be (this is how it was done by Wischmeier and Smith, 1978, and still by Rogler and Schwertmann, 1981).

Specific comments

1. Page 2, line 24–25: "Unstable and unreliable transfer functions result that differ pronouncedly" – I do not understand the sentence, could you please rephrase?

Rephrased

2. Page 2, line 30: Please include a general reference for the radar measurement principle. One such reference could be the book by Bringi and Chandrasekar (2001), Polarimetric Doppler Weather Radar, Cambridge Uni. Press.

Thank you for the suggestion. To put this into the European context, we additionally cite:

Meischner, P., Collier, C., Illingworth, A., Joss, J., Randeu W.: Advanced weather radar systems in Europe: The COST 75 action, Bull. Amer. Meteorol. Soc., 78, 1411-1430, 1997.

3. Page 3, line 12: I see your point that the use of continuous radar data avoids missing large and rare events that could be missed completely by gauge networks. But you do a lot of processing to the radar data, including winsorizing and smoothing, which reduces the influence of rare extreme events on the summary statistics. There are two separate problems here: sampling (gauges may miss an event) and then what value to use as a "typical" value for a skewed distribution. It is important that justifications for the data treatment show that the chosen "typical" measure is appropriate.

We fully agree. This is why we examine in detail the effects of the different statistical steps during smoothing. Now we also provide in the Introduction an overview over the previously used, unintended and uncontrolled steps of data manipulation that were necessary to arrive from gauge data at erosivity maps. The origin of the problem, however, rests in the characteristics of erosive events and can thus not be avoided. The use of contiguous radar data allows replacing several poorly controlled steps by clearly defined statistical methods, which is a major step forward but certainly better methods will be found in future the better we understand the origin of variation, the more data are gathered and the better the computing methods will become.

4. Page 3, line 31: Which Z-R relationship is used?

We provide a reference now

5. Page 4, line 6: Is the figure of 1 gauge per 80 km^2 an average value?

Yes; we added "equivalent to" to make clear that this density is calculated from the number of 4000 stations per Germany

6. Page 4, line 20: For clarity, it would be helpful to include the units of Imax30 and Ekin when the variables are introduced here; this is especially important because Ekin [kJ m^{-2}] and $Ekin_{,i}$ [kJ m^{-2} mm^{-1}] have different units.

We included the units when the variables were introduced

7. Page 4, line 24: You should reference Fischer et al 2018 (from your references list) here since your definitions, units, and descriptions are very similar to those used in your previous paper.

We cite Rogler and Schwertmann (1981) now because they were the first who transferred the Wischmeier equations in US units to SI units.

8. Page 4, line 29: "the R_e sum" – do you mean "the sum of R_e"?

Yes; we rephrased the sentence

9. Page 5, lines 14–15: For a given pixel, if too many years were excluded then the sampling may become less representative. How often were pixels affected by this exclusion of years, and were there pixels for which many years were excluded?

The requested information had already been given ("If the effective number of excluded years was larger than one, the respective pixel was excluded. This was the case for 0.6% of all pixels.") but now we moved it to a different place where it may be expected more by the reader.

This loss of years mainly occurred in marginal regions in the very North or very South that had only been captured by radar data before or after the displacement of radar sites (e.g. in the far North due to the shift of the radar Hamburg to Boostedt in 2014).

10. Page 5, lines 16–22: "replaced by the maximum 1-h rain depth" - should this read rain intensity?

Usually rainfall is reported for a certain period of time (per day/month/year) and hence this should always be called intensity instead of depth. However, depth is commonly used if longer periods of time (days, years) are considered during which short-term intensity varies.

We replaced "depth" by "intensity".

11. Page 5, lines 16–22: As I understand it, the scaling factors are being used to adjust the method of calculating erosivity to put a "virtual rain gauge" in each radar pixel, to account for the fact that radar measurements are areal and integrated over time and therefore smooth out rainfall intensity peaks. Since rain intensity depends on temporal resolution, and you require 30 minute maximum rain rates which would be smoothed by the use of 1 hour radar data, I see why a temporal scaling factor could be used. But spatially, the areal measurements at 1 km^2 resolution can be assumed to be representative of each 1 km^2 pixel, and since you are producing erosivity values at the same resolution, I don't understand why a spatial scaling factor (or indeed the method scaling factor) is required. Please could you explain more here why the scaling factors are used and how they are applied (e.g. it is not clear which threshold is lowered to 5.8 mm h^{-1}).

We agree that this is difficult to understand. We had to dedicate an entire publication to this tricky question (Fischer et al., 2018). An additional explanation with one or two sentences in this manuscript would open more questions that it would answer. We prefer to leave it like this.

In the discussion we now explain (and cite a reference) that rain erosivity on average drops to one fourth at only 2 km distance from the centre of a convective rain cell. This pronounced patchiness requires a spatial scaling factor.

We repeated "Imax30" in this sentence to make clear, which threshold had to be adjusted.

12. Page 6, line 14: The use of some independent data to test the spatial representativity of the smoothed data is a good idea, but is this test data independent? It is also based on radar data. Has the test region data been compared to gauges or other ground truth data to ensure it is accurate?

In this case we only wanted to examine the effects of our data treatment but not the comparison between radar-derived and gauge-derived erosivities. We have done this extensively in Fischer et al. (2018), which we cite. In this publication we compared 19 944 events observed at 115 stations with radar-derived erosivities for the same location.

13. Page 7, lines 1–2: "The cumulative distribution curve for the test region calculated from 5-min data will then be a fair estimate of the return periods anywhere in the research area" – I do not think this is proven. Even if the test region and the whole area agree at 1 hour resolution, extreme intensities are smoothed out at this lower resolution, so it does not necessarily follow that the 5 minute cumulative distributions are the same across all regions.

We weakened the statement:

"The cumulative distribution curve for the test region was also calculated using the 5-min rain data. Given that the cumulative distribution curves of the entire study area and the test region agree for the relative erosivities calculated from 1-h data, we expect that the relative erosivities calculated from 5-min rain data of the test region can serve as a first estimate for the entire study region."

14. Page 7, lines 19–20: Please include a reference for these statements about radar accuracy (they are correct but require a citation).

We added a reference

15. Section 3.1: I suggest that to back up your observation that the regional pattern in erosivity is dominated by orography, you should include a topographic map showing ground elevation for comparison with the map in Figure 2.

We added a detailed topographic map in the Appendix

16. Page 8, line 13: "Using the normal distribution" – but are the erosivity values normally distributed?

This is clearly not the case although long-term average erosivity does not have the extreme skewness of individual events or the skewness of individual years. The statistical distribution is not normal but may be multimodal because it reflects the contribution of different landscapes (mountains, plains, coastal regions etc.) to total land. The assumption of a normal distribution is only used here for illustration purposes to "translate" the semivariance, which may be difficult to grasp by some readers, to the ordinary R factor space. This assumption has no further relevance for our analysis. We deleted this phrase.

17. Page 8, line 32: Which variogram was the kriging conditioned by? I would expect kriging to maintain the spatial stucture even in a block kriging case, so it is odd that the kriging changes the variogram.

The semivariogram used for kriging is indicated as "1 h, 17 yr" in Fig. 3. We added a numbering to the semivariograms in Fig. 3 so that we can clearly refer to each of the semivariograms in the text.

The semivariogram after kriging always differs from the semivariogram before kriging because at least the nugget effect is removed by kriging. Block kriging will additionally remove any pattern that is smaller than the block size. It can then not appear anymore in the semivariogram after kriging. This is a well-known phenomenon in geostatistics and usually addressed as change of support or as regularization or as modifiable areal unit problem. Such a change of support may even turn a spherical model into a Gaussian model. See for instance:

Clark I.: Regularization of a semivariogram, Computers & Geosciences, 3, 341-346, 1977.

Gotway, C.A., Young, L.J.: Combining incompatible spatial data, J. Amer. Statistical Assoc., 97, 632-648, 2002.

Emery, X.: On some consistency conditions for geostatistical change-of-support models, Mathematical Geology, 39, 205-223, 2007.

18. Section 3.3: I suggest adding more lines to your Figure 4 to show all the lines you mention in the text.

We do not understand this comment. In the text we mention return periods of 2 yr, 30 yr and 100 yr. In Fig. 4 we show lines indicating return periods of 2 yr, 5 yr, 10 yr, 30 yr and 100 yr.

19. Page 9, line 21: I'm surprised that you would expect less than the mean erosivity for an event with a return period of 2 years. Any comment there?

This is due to the fact that total erosivity is determined by extreme events. To illustrate this: the largest event recorded by Fischer et al (2018) during only two months was 1270 N h^{-1}. This means that if at that location no other erosive rain would fall during the next 10 yr, the average annual rain erosivity would still be the same as that of all other locations.

20. Page 9, line 27: "d" is presumably for days but should be spelled out.

This would not comply with the recommendations of NIST and other institutions of standards, which require that unit symbols and unit names must not be mixed.

Thompson, A., Taylor, B.N., Guide for the Use of the International System of Units (SI), NIST Special Publication 811, 2008

21. Page 9, line 32: To see exactly what is going on here, did you compare the distributions of erosivity values for each of these example days? I suspect that the median values would be more stable.

The value was calculated as the accumulated erosivity of an individual day over 17 yr and all pixels divided by the accumulated erosivity over all days and all pixels during 17 yr. Hence there was only one value for each day. Calculating the erosivity index individually for each year and each pixel would yield a (very stable) median of zero for all days (because there are only about 20 events per year at a location).

What was surprising to us was the fact that despite this huge sample size for every day (almost 8 million pixel days), the values still differed so much for consecutive days. This is because averaging over years cannot be replaced by averaging over locations. During some days many locations (pixels) received erosive rainfall while on the next day the weather system may have changed and no pixel receive erosive rain. It is clear that with 20 events per year and 17 years (=340 events per pixel) some days must exist in each pixel that cannot have received an event even if the events would be evenly distributed.

22. Page 10, line 20: I think Fig. 5 should be Fig. 6.

We corrected the typo

23. Page 11, line 18: No definition of C-factor calculations is given; please add one.

We added a reference here. A short description was already given in the Introduction

24. Page 12, line 22: Please define (R)USLE.

RUSLE is explained in the Introduction now

25. Figures 2, A1, and A2: Units for the plotted variable should be stated either in the key or caption.

Information was added

Technical corrections

- Page 5, line 19: The word "occurred" can be removed.

Sentence was rephrased

- Page 7, line 12: By "in the smooth" do you mean "in the smoothing operation"?

This wording had been taken from the cited statistical reference. Now we reworded the sentence and used common language.

- Figure 2: In the caption the average sizes of the local authority and community areas should be areas (km_2).

Changed as requested

- Page 8, line 8: "very extreme" is redundant when "extreme" will do.

Changed as requested

- Page 9, line 23: "extremer" should be replaced by "more extreme".

Changed as requested