Reply to the Comments from Referee #1 for HESS-2018-517

This is the second time I review this paper. Most of my previous comments were taken into account and the authors have made substantial changes to the original manuscript. As a result, the quality has improved. That being said, there are still a few important issues that remain to be addressed before publication can be envisaged.

We appreciate the time and effort the reviewer put into the review. In the following, we have provided an item-by-item reply to the comments. Please note that lines and pages mentioned here are based on the marked-up manuscript.

Major Comments:

1. Usefulness and novelty:

While the WegenerNet is an amazing and unique source of data, the paper itself is rather dull and empty, providing very few ideas and findings worth publishing. The conclusions can be summarized as follows:

- The spatial variability of rainfall varies with years, seasons and events (no big surprise there). The spatial autocorrelation structure is anisotropic and approximately decreases exponentially with distance (which has been known for a long time already).

- The small-scale variability during the cold season is higher while the decorrelation range is longer, in agreement with previous findings.

- The uncertainty in areal-rainfall estimates depends on the number of gauges available (obviously!). High-resolution gridded data provide more reliable information about extremes (also obvious).

Many of the numbers are specific to the WegenerNet and of little usefulness to others. The only new idea in my opinion is the power-law model for predicting the number of gauges needed to ensure the uncertainty affecting mean areal rainfall estimates is below a given threshold. But this part of the paper is not detailed enough and the model has not been properly validated (see comment 3). Otherwise, the paper remains very descriptive, with only a single equation in it. There's nothing wrong with it but I just don't think that it will be useful. A more useful approach in my opinion would have been to use the WegenerNet data to formulate predictive models for assessing uncertainties and validating findings on other networks or situations.

=> As pointed out by Reviewer, the main limitation of studies using a local network such as WEGN is that it is hard to generalize the studies' findings and results to other regions. However, such local-scale studies can support a comprehensive understanding of rainfall processes, not just by delivering results found from different climate/rainfall regimes, but also by showing the potential use of the network for future (collaborative) research.

In addition, WEGN data have been widely used by science communities and the need for the data is increasing. Therefore, this study is also motivated by the practical needs of on-

1

going/future remote sensing applications and modeling studies.

It might be true that our general findings regarding the spatial variability of rainfall are not totally new as pointed out by the reviewer, however we believe that the quantitative information derived from WEGN is valuable not only for local science communities (who are using WEGN data directly) but also for hydrology communities (who want to compare rainfall characteristics before applying our results, e.g., the power-law model, to their own regions).

The knowledge about the approximately exponential autocorrelation decrease with distance is based on a limited number of studies in different geographic regions and climate regimes; very few of them have been performed with a comparable spatial resolution. We believe that it is also important to confirm previous knowledge.

It may not be surprising that the uncertainty in areal-rainfall estimates depends on the number of gauges, but we provide detailed information on *how* it depends on the number auf gauges.

We agree with the reviewer that the conclusions did not fully reflect the main results of the paper and added some quantitative information (also to the abstract).

Please refer to Comment 3 for more details on the power-law model.

2. Too many inconsistencies:

The paper contains many inconsistencies. Often, the text says A while the figures show B.

- in the autocorrelation function, the text says that the largest considered distance is 15 km. Yet Figure 3 shows distances up to 20 km and more. This was already an issue in the previous version and has not been addressed properly.

> *=> The correlation model **is built** using data of separation distance <=15 km and, for the Fig. 3, the model **is used to present** the correlation function till ~23 km. We clarify this in the caption of Fig. 3.*

- On page 4, the text says that a data transformation is necessary to deal with zeros and make the data more Gaussian distributed. Yet Figure 3 shows the autocorrelation without data transformation (or at least I assume, it's not 100% clear). If indeed the values change after transformation, why don't you show the ones matching the text?

> *=> Data after the log-transformation are used for Figure 3. We added a sentence to make it clear. 6-7 lines, Page5.*

- For the exponential model, the text mentions that another two-parameter model was tried with very similar results. Yet somehow the authors decided to go for the three parameter model

2

anyway. The explanations to justify this choice are not convincing at all and proper validation is required to motivate this choice. Indeed, **one of the parameters seems to be almost constant across scales**, which means its probably not useful.

> *=> The three-parameter model is selected because the model shows the smallest RMSE (fitting errors) among the tested models (lines 34-35, page5). We think "the one of parameters seems to be almost constant" means Figure4-d (RMSE); the model shows almost constant errors across scales and that is another reason why we chose the model.*

- Page 8, line 24: the text says that the study has confirmed that the WEGN provides very accurate areal-precipitation estimates. But actually, there is no evidence for this in the paper. Only comparisons where the average of the 150 stations is assumed to represent the truth.

> *=> Figures 6 and 7 show the convergence of error in areal rainfall estimation with increasing gauge number. This allows to infer that the WEGN 150 gauge provides very accurate areal-precipitation estimates. We rephrased the sentence, now at Lines 22-23, Page 9.*

- Page 8, lines 32-33: "More than 10 gauges guarantee that we can obtain constant results regardless of the number of gauges." This is not true! The only evidence you show is that the average error will be lower than 20%.

> *=> We wanted to say that the magnitude and spread of errors are not significantly changed with >=10 gauges. The sentence is rephrased (list line, Page 9 to first line, Page 10)*

3. Validity of power-law model:

The power law model proposed in Figure 6 is potentially interesting as one could imagine situations in which people need to estimate the required gauge density for achieving a certain accuracy. However, it should be pointed out that (a) it would be better to formulate it in terms of gauge density (#gauges/km2) and (b) that such a model needs to be properly validated/assessed. Right now, it is just given without any further evaluation or critical discussion. One big question is how well does the power-law model generalize to other cases. For example, what if I work with a catchment of 100 km2, that is, 3x times smaller than the area of the WegenerNet? Does this mean that I can divide the number of gauges by 3? I assume not since the autocorrelation varies faster over the first 10 km compared with the 10-20 km range. There is a lot of potential here for formulating a model that can be used by others. But this requires additional work.

> *=> Thanks for the comment and the suggestion. We tested the plot additionally with areas of 50 km^2, 100 km^2 and 150 km^2 (about 1/6, 1/3 and 1/2 of the WEGN area, respectively) and added the results to the Figure 6-b and -c. For any cases, we still observe the power-low relation between the gauge numbers to reach a certain accuracy*

level vs time resolution. But, as the reviewer has assumed, the required gauge numbers do not linearly decrease as the considered network area decreases. This is discussed at Lines 18-28 Page 7. Please note that the gauge density information is added to the Figure 6.

4 Data transformation:

On page 4 line 26 it is said that a transformation $x \rightarrow \log(x+1)$ is applied to the data to avoid issues due to zeros. However, there are crucial details missing. For example, you need to be aware of the fact that due to the non-linearity of the log transform, the results of the correlation analysis will depend on the units in which x is expressed (e.g., using mm/h or mm). This is a common statistical fallacy and needs to be addressed more carefully. Moreover, it is still unclear to me how zeros are actually handled by the authors. Do you take all time steps (including the ones where all gauges record zeros) or only a subset? I asked for more details about this in the previous round of review but there hasn't been much progress on this aspect. Same for the anisotropy maps in Figure 5. You need to specify how zeros were handled and whether the maps show the values obtained after data transformation or before.

> *=> We used mm for the units, not mm/h (i.e., no unit conversion); indeed, log-transformation after the unit conversion will significantly alter the correlation of the raw data. Data pairs, when both are no-rain, are excluded for calculation. We added this information at Line27, Page4 & Line 4, Page 5. Please note that the same data (i.e., no unit conversion, exclusion of zero-rainfall pairs, and log-transformed) are used for Figs 3-5.*

5 Selection of events:

It's not 100% clear from the text how the events in Section 4 were selected. The corresponding sentence on Page 6, lines-6-7 is not well formulated and open to interpretation. Please revise. Also, the selection seems to be done based on total rainfall accumulation, which tends to favor certain types of events (i.e., persistent and widespread), potentially providing biased results. Please comment on this in the paper.

> *=> The sentence is rephrased (Lines 25-26, Page 6). The heavy event is selected based on total rainfall accumulation, without a consideration of rain type. We agree with the reviewer that the way of selecting rainfall events can heavily influence quantitative results, however, we intended to define the heavy events in a general and common way – using a fixed threshold.*
>
> *We should note that the selected rainfall type is rather mixed, not biased to persistent and widespread events. This is inferred from our visual-inspection on rainfall time series (Lines 27-28, Page 6).*

Minor issues:

- Page 1, abstract: "these dense networks are only available at sub-pixel scales and over short

periods of time". Too vague. Please explicitly state what pixels we are talking about. There are large differences in weather radar products and some of the latest X-band products have resolutions as high as as 30 meters.

> *=> The sentence is modified (Lines 2-3, Page 1). We wanted to say that the area of other dense gauge networks is limited, i.e., smaller than remote-sensing pixel sizes in a general sense (Lines 24-25, Page 2), while the WEGN covers an area of 300 km².*

- In Equation 1, I suggest to replace c1 by 1-c1. This would make the interpretation easier, making c1 the drop in autocorrelation (= the nugget) instead of the intercept. Higher c1 would then mean higher small-scale variability which makes more sense in my opinion.

> *=> Given that many studies of rainfall correlation refer to "c1" as "nugget parameter" or "nugget effect", we would prefer to follow the convention for this parameter (see, e.g., Ciach and Krajewski, 1999, 2006; Villarini et al., 2008; Peleg et al., 2013; Svoboda et al, 2014; Tokay et al., 2014;all listed in manuscript). This is somehow in contrast to the studies of rainfall semivariograms where higher "nugget" means higher small-scale variability.*

- Figure 4: the units are missing
  > *=>We added the units. Thanks for the comment.*

- Figure 4, the RMSE values (<0.01) you provide are misleading. They give the impression to the reader that the exponential fit is very good whereas the RMSE in the plot is the one that you calculated by taking the average value of the yearly-averaged autocorrelation values. However, individual fits (for a given year or a season) are not that good and the spread remains large, as shown in Figure 3.

> *=> The spread of RMSE between individual fits for a given year or a season is not that large as shown in the figure below. This is because we calculated correlation average for each distance bin first (Line 28, Page 4) and then fitted the model to the averaged correlation. This might lead to the "too-good" fitting, however, does not affect general behaviors of the parameters described in Fig4. We agree that this can be misunderstood to readers so we clarify this point at lines 3-4, page 6.*
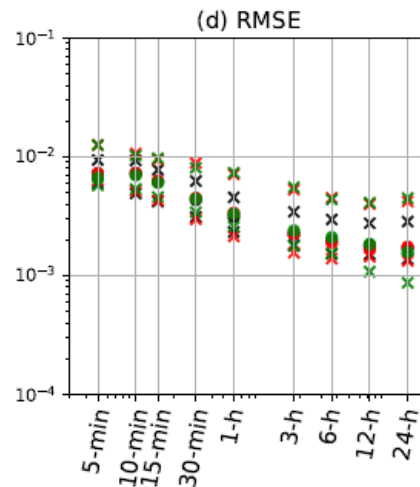
Figure 1the fitting error of the models for yearly-averaged correlation (dots) and for each year (Xs).

- Figure 4: Most decorrelation distances shown here are much larger than the maximum observable size of the network (20 km). So how much do you trust these values? The text provides a small warning and a reference to a paper but I believe this warrants more attention than that. Having worked on spatial structures myself, I know that such large range estimates are often the result of bad fits or the choice of the model rather than physical. In any case, I think it's illusory to think that you can infer a 200 km range from data extending over 20 km. A better approach would be to limit the range of scales.

> => *we agree with the reviewer and added the 'warning sign' in the Fig.4 and modified the sentences at Lines 4-6, Page 6.*

- There seems to be a confusion between the e-folding distance (which is a distance and should be in units of km) and the autocorrelation corresponding to the e-folding distance (which is unitless).

> => *Thanks for pointing this out; we added the units to the former (Figure. 4) and revised the related sentence (Line 14, Page 6)*

- Figure 9: it's almost impossible to see any difference in color here. Please adapt the scales.

> => *the figure is updated; we added another color-scale for the second plot.*

- Section 4. It would be more useful to talk about gauges/km2 to avoid introducing too much dependence on the total area. In fact, the entire discussion should be centered around gauge density rather than the number of gauges. Also, other thresholds than 20% should be considered, as this is rather large in my opinion and is more characteristic of the accuracy one would like to achieve at the point scale.

> => *We added density information to inset-plots of Fig. 6-b and 6-c. Thanks for the*

6

suggestion. The value of 20% is adopted from the well-known study of Villarini et al., 2008; we believe that readers can infer the results for other thresholds from Fig. 6-(a).

- Page 5, lines 10-14: You mainly attribute the higher small-scale variability to solid or mixed precipitation here. Another explanation is that winter-type events are more heterogeneous and spatially disorganized than convective cases. Moreover, since they are lower intensity, the uncertainty affecting the measurements of two neighboring stations plays a bigger role. Bottom line: there are more explanations that can be given here and it's not clear from the evidence that you present that the higher small-scale variability is indeed attributed to snow and ice. Please rephrase.

=> Thanks for the comments. The sentence is rephrased; lines 20-24 Page 5.

Typos:

- Page 2, line 7: "On the contrary, gridded rainfall data are nowadays available..." The expression "On the contrary" does not appear adequate here.

=> fixed.

- Page 2, line 34, "in order to contribute to the effort better and more broadly assessing the uncertainty". Bad English, please rephrase.

=> rephrased.

- Page 7, line 25: "decreases" instead of "deceases"   - Figure 3: there is a typo in the caption (separation) - Figure 5: there is a typo in the caption (north-south)

=> All fixed. We apologize for the typos.

Assessment of spatial uncertainty of heavy rainfall at catchment scale using a dense gauge network

Sungmin O[1,2,*] and Ulrich Foelsche[1,2,3]

[1]Institute for Geophysics, Astrophysics, and Meteorology/Institute of Physics (IGAM/IP), NAWI Graz, University of Graz, Austria
[2]FWF-DK Climate Change, University of Graz, Austria
[3]Wegener Center for Climate and Global Change (WEGC), University of Graz, Austria
[*]Now at Biogeochemical Integration, Max Planck Institute for Biogeochemistry, Jena, Germany

Correspondence: Sungmin O (sungmin.o@uni-graz.at)

Abstract. Hydrology and remote-sensing communities have made use of dense rain-gauge networks for studying rainfall uncertainty and variability. However, in most regions, these dense networks are only available at ~~sub-pixel scales~~ small spatial scales (e.g., within remote-sensing subpixel areas) and over short periods of time. Just a few studies have applied a similar approach, i.e., employing dense gauge networks to catchment-scale areas, which limits the verification of their results in other regions. Using 10-year rainfall measurements from a network of 150 rain gauges, WegenerNet (WEGN), we assess the spatial uncertainty in observed heavy rainfall events. The WEGN network is located in southeastern Austria over an area of 20 km × 15 km with moderate orography. First, the spatial variability of rainfall in the region was characterised using a correlogram at daily and sub-daily scales. Differences in the spatial structure of rainfall events between warm and cold seasons are apparent and we selected heavy rainfall events, the upper 10% of wettest days during the warm season, for further analyses because of their high potential for causing hazards. Secondly, we investigated the uncertainty in estimating mean areal rainfall arising from a limited gauge density. The average number of gauges required to obtain areal rainfall with errors less than a certain threshold (<=20% normalized root-mean-square error is considered here) tends to increase roughly following a power law as the time scale decreases, while the errors can be significantly reduced by establishing regularly distributed gauges. Lastly, the impact of spatial aggregation on extreme rainfall was examined, using gridded rainfall data with various horizontal grid spacings ~~from 0.1° to 0.01°~~. The spatial scale dependence was clearly observed at high intensity thresholds and high temporal resolutions; the 5-min extreme intensity increases by 44% for the 99.9th and by 25% for the 99th percentile, with increasing horizontal resolution from 0.1° to 0.01°. Quantitative uncertainty information from this study can guide both data users and producers to estimate uncertainty in their own observational datasets, consequently leading to the sensible use of the data in relevant applications. Our findings are generalisable to moderately hilly regions at mid-latitudes, however the degree of uncertainty could be affected by regional variations, like rainfall type or topography.

1 Introduction

Rainfall data are one of the most important inputs for hydrological as well as climatological studies and applications. Furthermore, fit-for-purpose information derived from rainfall data is crucial for a wider range of users, such as civil engineers, water resource managers and governments. To meet the needs of diverse user groups, rainfall observational datasets from in-situ measurement and remote sensing have been greatly enhanced in terms of both data quality and resolution (e.g., Berezowski et al., 2016; Hou et al., 2014; Keller et al., 2015; Yatagai et al., 2012). Often, rainfall data are required as areal estimates at the scale of interest, for instance, at grid or catchment scales. Point measurements from in-situ gauge observations are spatially aggregated or interpolated to estimate the areal distribution of rainfall, and hence the accuracy of areal rainfall data is highly dependent on spatiotemporal variability of rainfall events and density of observation points (Girons Lopez et al., 2015; Hofstra et al., 2010; Villarini et al., 2008; Wood et al., 2000). This limits the understanding of fine-scale rainfall processes, particularly of extreme events (Sillmann et al., 2017). ~~On the contrary, gridded~~ Gridded rainfall data are ~~nowadays~~ also available from remotely sensed observations at high spatial resolutions (e.g., 1-5 km^2 for radar data or $0.1° \times 0.1°$ for satellite data). While those data sets are good alternatives to address a number of the issues relating to the scarcity of gauges, rainfall variability at ~~sub-pixel~~ subpixel scales can still not be fully resolved (Peleg et al., 2013; Tokay et al., 2014). Moreover, systematic errors can be large, and the quality of remotely sensed data therefore strongly relies on gauge-based data that are used for their regional validation and correction (Kann et al., 2015; O et al., 2018b; Steiner et al., 1999).

Addressing the issue of spatial variability and uncertainty of rainfall has been tackled over many years with various purposes. For instance, evaluation of satellite or radar rainfall products involves investigation of larger-scale rainfall processes to assess the ability of remote sensing in capturing the inter-pixel rainfall variability (e.g., Chaudhary et al., 2017; Dhib et al., 2017; Lockhoff et al., 2014). On the other hand, small-scale rainfall processes are of interest to identify the effect of intra-pixel variability of rainfall on the performance of remote sensing (e.g., Ciach and Krajewski, 1999, 2006; Gebremichael and Krajewski, 2004; Habib and Krajewski, 2002; Peleg et al., 2013; Tan et al., 2018; Tokay et al., 2014). To quantify the rainfall uncertainty, observational data from highly dense rain-gauge networks have been employed as a ground truth. Peleg et al. (2013) used multiple rain gauges within a radar subpixel area (4 km^2) and examined the contribution of gauge sampling error to the total radar-rainfall estimation error. Using relatively long-term gauge data (5-years), Tokay et al. (2014) analyzed the spatial correlation of rainfall for different seasons and weather systems within the footprint size of microwave satellite sensors.

A similar approach employing dense gauge networks can be adopted to diagnose the spatial variability and uncertainty of rainfall at catchment scales (e.g., 100 - 500 km^2). Such scales are of great interest not only for the evaluation of remotely sensed data, but also for hydrological applications like runoff modelling or gauge network design. Wood et al. (2000) examined the accuracy of areal estimates of rainfall over a 135 km^2 basin according to the HYdrological Radar EXperiment network consisting of 49 rain gauges. The network later provided a six-year rainfall dataset (from 50 gauges) for the study of Villarini et al. (2008), where a comprehensive analysis of temporal and spatial sampling uncertainties was conducted. However, most of the local areas do not have adequately dense gauge networks, which limits the comparison and verification of findings from the aforementioned studies across diverse rainfall regimes. Schroeer et al. (2018) recently employed the WegenerNet Feldbach

region (WEGN) and the surrounding operational rain gauge stations to sample summertime convective extreme events at sub-hourly to hourly scales and found a power law decay of the event maximum area rainfall with increasing interstation distance (1 km to 35 km).

In this paper, in order to contribute to the effort for better and ~~more broadly assessing the uncertainty of rainfall at fine scales~~
5 ~~associated with the spatial variability of rainfall~~broader assessment of the rainfall spatial variability and associated uncertainty, we employed 10-year rainfall data from the WEGN, a high-density network in southeastern Austria (Kirchengast et al., 2014). The network includes 150 rain gauges deployed over an area of $\simeq 300$ km^2, approximately corresponding to one gauge per 2 km^2. First, following previous studies (e.g., Villarini et al., 2008; Peleg et al., 2013; Tokay et al., 2014), we quantified the spatial variability of rainfall utilizing a corrologam between the gauges to understand the spatial characteristics of rainfall in
10 the region.

Second, we investigated the uncertainty in estimating areal rainfall caused by a limited number of point observations. Given that the properties of individual rainfall events can be different from all-event averages (Ciach and Krajewski, 2006; Eggert et al., 2015), we focused on events with a potentially high impact, which we defined as the top 10% wettest days during the warm season (May–September). The accuracy of areal rainfall estimation is a long-standing issue, e.g., in catchment mod-
15 elling, because error and uncertainty in rainfall data can propagate into large variations in simulated runoff, and thus it has been dealt with in diverse manners. For instance, the influence of spatial representations of rainfall input to runoff errors has been demonstrated through modelling studies (e.g., Bárdossy and Das, 2008; Xu et al., 2013)~~or the~~. The error in catchment-scale areal mean rainfall has also been directly quantified by employing high-resolution gauge data (e.g., Villarini et al., 2008; Wood et al., 2000; Ly et al., 2011). We followed the latter approach using the WEGN rainfall data.

20 Finally, we compared extreme rainfall at different spatial and temporal scales using gridded rainfall fields to quantitatively assess the impact of spatial averaging on the definition of extremes. The identification of rainfall extremes based on intensity thresholds is common practice, however, the considered spatial scale of rainfall data defines different sets of extreme events (Eggert et al., 2015), potentially affecting threshold-based early warning systems (Marra et al., 2017). Although gridded datasets have been used in a range of applications like assessments of climate change impacts or evaluation of climate models,
25 a common caveat of using the datasets in the study of extreme rainfall is that the quality of gridded rainfall data is highly con-strained by the location and density of input weather station data (Hofstra et al., 2010; Prein and Gobiet, 2017). By contrast, the quasi-regular configuration of WEGN on an approximately 1.4 km x 1.4 km grid permits robust examination of the frequency and intensity of rainfall extremes at various horizontal resolutions.

Consequently, this study aims to assess spatial uncertainty of rainfall at catchment scale using rain gauge data, with a focus
30 on heavy and extreme rainfall events. This paper is structured as follows. Section 2 describes WEGN rain gauge data and regional rainfall climatology. Section 3, Sect. 4, and Sect. 5 present the results of the data analysis. We close with discussion and conclusions in Sect. 6.

2 WEGN rainfall data and regional rainfall climatology

The 10-year rainfall data (2007-2016) are obtained from the WEGN Feldbach region network in southeastern Austria (Kirchengast et al., 2014). Of 154 weather stations, 150 stations that are equipped with tipping-bucket rain gauges are used in this study (Fig. 1). Raw rain gauge data are aggregated every five minutes. Errors in the rainfall data were comprehensibly analysed and
5 corrected by O et al. (2018a). The gauges are almost uniformly spaced over an area of 20 km × 15 km with moderate topography (about 260 to 520 m altitude). The inter-gauge distances range from approximately 0.7 km to 23.4 km. The gridded fields of rainfall are constructed by an inverse distance weighting (IDW) on a 200 m × 200 m Universal Transverse Mercator grid. WEGN station and gridded data products are available at www.wegenernet.org.

Southeastern Austria including the Feldbach region is influenced by both continental and Mediterranean climates. The re-
10 gion receives high amounts of rainfall during summer months. The occurrence of thunderstorms and hail is higher than in other parts of Austria (Matulla et al., 2003). Figure 2 shows average diurnal variations of rainfall and temperature over the entire network during the study period. The WEGN area is characterized by warm and wet months from May through September (hereafter "warm season") and relatively cold months without much rainfall during the remaining seven months (hereafter "cold season"). The average monthly rainfall is 102.8 mm in the warm season, while 48.9 mm in the cold season. The diurnal
15 signal is more clearly seen in the warm season for both rainfall and temperature. Rainfall maxima occur often in the early afternoon through midnight, shortly after maximum temperature, implying that a major contribution to the warm season rainfall is from short-duration convective events. Because diurnal heating plays an important role in triggering thermal convection, most inland regions show afternoon rainfall maxima (Dai et al., 2007). Extreme daily precipitation, however, can also be caused by Genoa Lows, which transport humidity from the Mediterranean Sea, yielding intense rainfall with long-duration.

3 Spatial variability of rainfall

The spatial structure of rainfall events is studied using the Pearson's correlation coefficient between all pairs of rain gauges. Pearson's r is the most commonly used rainfall correlation estimator (e.g., Ciach and Krajewski, 2006; Jaffrain and Berne, 2012; Peleg et al., 2013; Tokay et al., 2014; Villarini et al., 2008). At sub-daily and daily timescales from 5-min to 24-h (06-06 UTC), the correlation of rainfall among rain gauges is calculated for each year. One year period includes a set of warm season
25 (May to September) and cold season (October to next April). The incomplete years (i.e., first and last years) are excluded from the calculation of all-months (May to next April), whereas the warm and cold seasons have 10 annual curves each. The data pairs when both record zero rainfall are discarded. The correlation values in each period were then sorted according to the separation distance of gauge pairs and averaged into the nearest 1-km distance bins. We fitted a three-parameter exponential function to the average correlations. The distance bins for fitting the model were taken up to and including 15 km
30 given the network dimension, which means that rainfall data pairs were sampled uniformly for any spatial direction. The spatial correlation (r) at separation distance h is:

$$r(h) = c_1 \, exp \left[- \left(\frac{h}{c_2} \right)^{c3} \right] \tag{1}$$

where $c1$ represents the nugget effect, $c2$ is the correlation distance, and $c3$ is the shape factor. The parameters are determined by least-squares curve fitting. Figure 3 shows the spatial correlation of all-months, warm, and cold seasons for four selected accumulation times. A logarithmic transformation ~~(is applied to the data;~~ $log(x+1)$ to keep zero rainfall~~) is applied to the data~~, where x is in rainfall in mm. As the transformation make rainfall data conform more closely to the normal distribution, the effects of extreme values on correlation coefficients is mitigated (Habib et al., 2001; Jaffrain and Berne, 2012). This results in slightly lower correlations (not shown), however, the overall pattern of correlation decay curves remains unaffected. The data after the log-transformation are used in the figure.

Many factors are known to affect the spatial correlation structure in rainfall. For instance, Habib et al. (2001) examined the sensitivity of correlation estimation in rainfall to sample size or extreme rainfall events and Huff and Shipp (1969) demonstrated how the rate of correlation decay varied with different rainfall types. We therefore do not make a direct comparison of correlation values with those from other studies, yet we still observe that the behaviors of the correlation decay found in this study are in broad agreement with spatial rainfall correlation structures reported in the aforementioned studies. First, longer accumulation times show higher $c1$ (i.e., smaller microscale variations) and longer correlation distance values. Second, short-range correlation decreases rapidly with increasing separation distance, particularly at sub-hourly scales.

The warm season shows higher spatial variability of rainfall compared to the cold season, due to a higher proportion of convective events. The correlation curves of all-months show a more similar pattern with the warm season, as expected, given that most of the rainfall events are concentrated during the warm season (see Sect. 2). Tokay et al. (2014) found substantial year-to-year variations especially during autumn and spring. Similarly, WEGN rainfall shows marked interannual variability, but also during the warm season. It should be noted that the correlation functions of the cold season start with lower $c1$ values than of the warm season, meaning larger measurement errors and microscale variability of rainfall. ~~Because~~ This could be related to winter precipitation types in the region. For instance, uncertainty affecting the gauge measurements (e.g., wind-induced bias) may play a bigger role in determining the spatial heterogeneity of neighboring stations during low-intensity precipitation events, than during warm season convective events. Another possible reason is that WEGN does not accurately capture solid precipitation (O et al., 2018a), since only few gauges are heated, and thus systematic errors between neighboring gauges can be greater during the cold season, possibly yielding the low $c1$ values.

Figure 4a-c summarizes the time dependence of the three parameters. Synthesized parameters here are obtained from the fitting function that is constructed by averaging yearly correlation values in each distance bin. Nugget effect values range from 0.71 to 0.98 for the cold season, while from 0.85 to 0.99 for the warm season. The correlation distance of the cold season at the 3-h is nearly corresponding to the correlation distance at the 24-h scale in the warm season. The parameter values of all-months are located between those of warm season and cold season. We found that the dependency of nugget effect and correlation distance on times scale is similar to the results by Villarini et al. (2008). The nugget effect parameter changes sharply at smaller timescales, while the correlation distance appears to be more sensitive for larger timescales. The shape factor of this study, however, does not show a clear increasing or decreasing trend. This is consistent with findings from Peleg et al. (2013) and Tokay et al. (2014). We selected the three-parameter model for the function fitting, because the model shows the minimum root-mean-square error (RMSE) between observed and fitted correlation values across all time scales among the several tested

models (Fig. 4d). ~~However, we also found that a two-parameter function (i.e., we set shape factor =1)is fitted comparably well and furthermore, correlation distance over large time scales decreases significantly when the two-parameter model is used~~Note that we fitted the correlation models to bin-averaged values and thus obtained relatively small fitting errors compared to other studies (e.g., Ciach and Krajewski (2006) or Tokay et al. (2014)). During multiple tests with different fitting models, we found that the fitted correlation distances over 100 km (e.g., values at accumulation times of >6-h in Fig. 4b) are often highly impacted by the selected fitting models. However, this model uncertainty does not affect the ~~characteristics~~ general behaviors of the parameters including their dependence on time scale and their seasonal differences. Nonetheless, when the spatial scale of observed correlations is limited to a distance of a few kilometers ~~, the fitted correlation distances~~ (e.g., ~~accumulation times of >6-for warm season)~~15 km in our study), the correlation distances estimated from the fitting model should be interpreted with caution. Interested readers may obtain a more detailed discussion of the fitting model in Svoboda et al. (2015).

So far we assume that the correlation structure is isotropic. To check the directionality (anisotropy) of the spatial correlations, we remapped them onto the two-dimensional space (Velasco-Forero et al., 2009; Mandapaka and Qin, 2013). Fig. 5 shows the 10-year averaged correlations ~~plotting~~ plotted over 1 km × 1 km grid cells but only till the e-folding distance (~~about~~ when the correlation drops to around 0.37~~which is corresponding to the parameter of $c2$ in eq. 1~~). While the correlation drops rapidly in all directions over short distances, a strong correlation is observed in an approximate southwest–northeast direction as separation distance increases. The directionality is more pronounced during the cold season, which can be interpreted as a consequence of movement of large-scale weather systems (contrary to summertime convective) along the favoured wind direction during the season, rather than as an effect of orographic barriers. Such directional characteristics of the correlations are averaged out in Fig. 4.

4   Accuracy of areal rainfall estimation during heavy rainfall events

In this section we investigate data uncertainty associated with areal rainfall estimation. In particular, the study focuses on high-impact rainfall events. While heavy rainfall is one of the major hydrological hazards, its accurate spatial representation over an area remains a subject worthy of inquiry. ~~Based on daily rainfall (≥ 0.2), those days falling in the 90th–100th percentiles during the warm season are defined as heavy rainfall events~~Heavy rainfall events are defined as days with total rainfall exceeding the 90th percentile of the daily rainfall, without a consideration of rainfall type. Only the warm season is considered. As a result, a total of 71 ~~events~~ days are selected. According to our visual-inspection on rainfall time series, the selected days likely include mixed rainfall types (short- and long-duration rainfall) rather than a specific type. The median of gauge-averaged accumulations is 28.1 mm d^{-1}, with a range of 19.8 mm d^{-1} to 64.1 mm d^{-1} ~~. General information on the selected events can be found in Table A1~~(more information is given in Appendix A).

We assume that the mean areal rainfall of a full density network represents the "truth". The areal rainfall of n-gauge networks (n = number of gauges) is calculated and compared with the true rainfall to quantify the accuracy of areal rainfall estimation with low-density networks (see also Villarini et al., 2008). Each n-gauge network consists of randomly selected 1,000 possible

gauge combinations. The 1-gauge network has 150 cases. As shown in Fig. 6a, the average and spread of normalized RMSEs (NRMSEs) of areal rainfall estimation tend to decrease with rising gauge number. The mean number of gauges required to obtain areal rainfall with NRMSEs lower than 20% is given as a function of time resolution in Fig. 6b. The curve (in black) roughly exhibits power-law behavior; ~~$74.19 \times t^{-0.44}$~~$74.2 \times t^{-0.4}$, where t is the time resolution (minute). At the daily scale, more than one gauge per 300 km^2 would be sufficient to reach a <20% estimation error. Correspondingly, at the temporal scales of 1-h, 30-min, and 5-min, on average more than 12, 18, and 33 gauges, respectively, are needed to achieve the same level of accuracy. Villarini et al. (2008) found that four gauges are necessary at the daily scale for the same accuracy level for an area of 135 km^2. Heavy events are not explicitly considered in their study.

One should note that the use of randomly selected gauge combinations only offers a rule of thumb about the required number of gauges to minimize uncertainty in areal rainfall estimates. ~~Additionally, we wanted to see~~ To demonstrate the role of gauge distribution in determining the estimation error~~. So~~, we selected 'good' and 'bad' distributions, 100 cases, respectively, out of the 1,000 combinations for each n-gauge networks that ranked in the top 10% and bottom 10% based on the area-of-influence (see Appendix B). As seen in Fig 6a (red crosses), the smallest estimation error is obtained with regularly distributed gauges. In other words, a well-designed gauge network allows to meet the desired error limit with a smaller number of gauges (grey curve in Fig 6~~b~~c). For example, at a 1-h scale, the 20% estimation error can be reached using uniformly distributed 8 gauges, however, the same level of accuracy cannot be guaranteed even with 23 rain gauges if their spatial configuration is not properly structured.

We repeated the calculation of the required gauge number to reach the certain accuracy, using sub-areas of 150, 100, and 50 km^2, i.e., 1/2, 1/3, and 1/6 of WEGN area size, respectively (gray lines in Fig 6b and c). For each sub-area, the mean rainfall of all gauges within the area is assumed to be truth and the 1,000 possible gauge combinations are randomly selected, as we did above. For the 50 km^2 area where only 25 gauges are available, all combinations are included when the total possible combinations of n-gauges are less than 1,000 cases. For any case, the dependence of the accuracy of areal rainfall estimates on the gauge number shows the power-law behavior across time scales. However, the required gauge numbers do not linearly decrease as the considered network area decreases. For smaller areas, we need more number of gauges per km^2 (i.e., higher gauge density) to reach the same level of accuracy at the same time scale (see inset plots in Fig 6b and c). Because rainfall variability varies faster within the first few kilometer, more dynamic rainfall variations in the smaller areas cannot be properly captured when the inter-distances of gauges remain the same (i.e., constant gauge density between the sub-areas), particularly for short time scales.

Additionally, the effect of gauge density on event-based rainfall statistics is assessed in Fig. 7. Daily rainfall accumulation and peak hourly rainfall of the 71 heavy daily events are recalculated using predefined sub-networks with gauges ranging from 1 to 16. The gauges are uniformly spread; the definition of the sub-networks can be found in Appendix B. While the sub-network with only one gauge exhibits large overestimation errors for both total and peak rainfall, employing an additional gauge already significantly reduces the degree of errors and yields underestimation error more frequently than overestimation. Note that Austrian weather service (ZAMG) has two operational stations over the actual WEGN area. Given that convective storms occur on scales of a few kilometers, low-density gauges over the region are likely to miss the core of storm. On the

contrary, low-density gauges can also overestimate rainfall intensities by capturing only the core of storm, but the magnitude and frequency of these errors appear slightly less than those of the underestimation errors. There is no significant difference in either average error or spread of errors from more than 10 gauges, as expected from Fig. 6.

5   Impact of spatial aggregation on extreme rainfall

5   We next focus on the uncertainty of area- or grid-averaged rainfall relating to spatial data resolution for the heavy rainfall events. Figure 8 compares rainfall percentiles among the gauges. Grey lines mean a 10-90th percentile range of rainfall intensities at a given percentile bin. For example, at the 30-min scale, the 99.9th percentile (the top 0.1%) rainfall intensity corresponds to roughly 45 $\mathrm{mm\,h^{-1}}$ at most gauges, while it exceeds 52 $\mathrm{mm\,h^{-1}}$ at certain gauges. It is also seen that 10% of WEGN gauges (i.e., 15 gauges) records are found to be lower than 38 $\mathrm{mm\,h^{-1}}$. The upper tail of the rainfall distribution shows
10   strong spatial variation. Such point-scale extreme rainfall features will be completely missed unless there exist dense rainfall observations, or they are inherently smoothed out in gridded data.

In fact, many studies have pointed out that the use of gridded rainfall data can lead to erroneous analyses of small-scale extremes because of the limited number of point observations (Contractor et al., 2015; Hofstra et al., 2010; Prein and Gobiet, 2017; Tozer et al., 2012). In addition to the high-resolution, the regular distribution of WEGN gauges enables generating grid-
15   ded rainfall fields that are homogeneous in space, and, consequently, robustly assessing uncertainty in rare and extreme rainfall represented in the data.

We generated gridded data using all 150 WEGN gauges and rescaled the data into horizontal resolutions from 0.01 to 0.1 degree (hereafter HR01 to HR10). Spatial aggregation begins from the top-left corner towards the bottom right and the remaining southern and/or eastern part of the grid is discarded (see Fig 10). HR01 corresponds to about 1.1 km and 0.8 km in latitudinal
20   and longitudinal directions, respectively. Figure 9 shows the 99.9th and 99th percentiles of heavy rainfall intensities as a function of space-time scales. Although temporal aggregation more significantly alters the definition of extremes, the impact of spatial aggregation is also notable, particularly at the sub-hourly scales. The 5-min extreme intensity ~~deceases~~ decreases from HR01 to HR10 by 30% for the 99.9th percentile while it decreases by 20% for the 99th percentile.

Meanwhile, although the spatial aggregation impact is much less pronounced at a daily scale, the selected spatial scale still
25   affects statistics of extreme areal rainfall, such as daily extreme frequency. This is shown in Fig. 10, which illustrates the occurrence of days above a selected threshold; top 5% of heavy rainfall events at HR01. The concept of the exceedance probability above thresholds is widely used in analyses of rainfall-triggered risk. Some HR01-scale sites appear to experience extreme rainfall more frequently than other part of the region. In other words, high-resolution data well-represent spatial variation and frequency of rainfall extremes, neither of which is seen in lower-resolution data. Many existing gridded datasets are not likely
30   to fully sample such site-level extreme events owing to limited spatial resolution. The exceedance probability of extreme rainfall across spatial resolutions is given in Fig. 11. The impact of different data resolutions on extreme rainfall occurrence is pronounced in both lower and upper tails. The highest daily rainfall during 10-years appears to be 68.4 ~~mm/day~~ $\mathrm{mm\,day^{-1}}$ at

HR10, but 104.4 ~~mm/day~~ mm day^{-1} at HR01; the maximum record over the entire WEGN area is 64.1 ~~mm/day~~ mm day^{-1}, so the ratio of the site-to-areal extreme rainfall ranges from 1.07 to 1.63 depending on the considered spatial scale.

6 Discussion and conclusions

The understanding of spatial uncertainty in heavy rainfall at fine scales has been hampered by the limited availability of suitable and reliable observational datasets. Although high-resolution radar data are often used to study small-scale rainfall variability, the use of ~~the~~ radar data is dubious, as indicated by Svensson and Jones (2010), owing to their indirect measurements of rain and relatively short records. In this study, we used ~~the~~ 10-year rainfall measurement data from the 150 rain gauges, uniformly spaced over the WEGN network in southeastern Austria. First, to quantify rainfall variability, spatial correlation between the gauge records was examined. We found that the degree of spatial rainfall variability can be substantially different not only within years (warm versus cold seasons) but also between years. This implies that long-term data should be considered to obtain comprehensive perspectives on regional rainfall variability. In fact, individual weather systems can exhibit varied spatial characteristics (Habib and Krajewski, 2002; Ciach and Krajewski, 2006; Tokay et al., 2014). In southeastern Austria, including the WEGN area, Schroeer et al. (2018) found much steeper decay in a correlogram function when only extreme summertime convective events are accounted for. We also found that during the cold season, the density of gauges is less of a concern (showing longer correlation distance) compared to the warm season. However, low values of the nugget effect parameter imply that snow measurements during winter time remain a challenge, especially at short time scales. Additionally, we checked the directional features of the spatial correlations. The spatial correlations in the southwest–northeast direction decay faster than in the southeast–northwest direction. This anisotropic pattern appears to be more pronounced over large separation distances during the cold season, which is probably linked to large-scale weather systems. This indicates that the assumption of isotropic correlations can be another source of uncertainty, under certain weather conditions, when modeling spatial correlation of rainfall (especially for estimation of correlation distance).

Secondly, we ~~confirm that the 150 gauges of WEGN offer very highly accurate areal precipitation estimates~~ demonstrate how high density and regular distribution of WEGN gauges contribute to delivering accurate areal-precipitation estimation. The overall uncertainty in mean areal rainfall shows a clear dependence on the number of gauges and the temporal resolution considered for the estimation. To reach the same level of accuracy, the average number of gauge has to be increased roughly following a power law as time scale decreases. Given that only two operational meteorological stations exist over the WEGN area, the insufficient gauge density may hamper the use of the station data to construct spatial rainfall fields in the region, especially at sub-daily scales. Further analysis shows that there is no linear relation between the required number of gauges and the ratio of considered area size. The accuracy of areal rainfall estimation is also significantly dependent on the spatial configuration of the network. Assuming that we have a well-distributed gauge network, it is observed that at least 2-5 gauges are required in the WEGN area (300 km^2) for accurate areal rainfall estimates such that we can obtain reliable rainfall event statistics (e.g., total amount and peak hourly intensity of daily heavy rainfall events) with no significant error. ~~More~~ When more than 10 gauges ~~guarantee that we can obtain constant results, regardless of number of the gauge~~ are available in the area,

9

the impact of gauge number or configuration on the spread and mean of errors in areal rainfall estimation becomes marginal. Our findings have implications concerning the use of sparse observational gauge data, for instance, in hydrologic modeling or rainfall estimates evaluation (e.g., Syed et al., 2003; Tian et al., 2018).

Lastly, using gridded WEGN data, rainfall extremes are reproduced at multiple spatial scales; approximately from the grid resolution of regional to convective-permitting models (about 11.1 km to 1.1 km in latitudinal direction). We show how different rainfall events can be considered extreme depending on the spatial and temporal resolutions. We found that the 5-min extreme intensity increases by 44% for the 99.9th and by 25% for the 99th percentile, when the horizontal resolution increases from $0.1°$ to $0.01°$. The results also demonstrate that high-resolution gridded data provide more reliable information not only in terms of the magnitude and frequency of extremes, but also in terms of the exact location of the extremes. As a result, limited resolution of rainfall data can alter interpretations of rainfall statistics; extreme rainfall events at a location of interest (a $0.01° \times 0.01°$ site in our example) could occur more frequently and more intensely versus the local average. For instance, the highest daily rainfall during 10 years appears to be 68.4 mm day^{-1} at $0.1°$, but 104.4 mm day^{-1} at $0.01°$ resolution. Localized information from high-resolution observation is the key for developing prevention and protection plans to mitigate potential damages of extreme rainfall in an efficient and adequate way. Our results highlight the need to evaluate uncertainty in extreme statistics derived from the existing datasets for supporting data selection among available rainfall data products.

In conclusion, the WEGN network provides a unique opportunity to empirically assess spatial variability and uncertainty of surface rainfall directly based on gauge data. The network provides long-term records, of more than a decade, which permits to exclusively focus on heavy rain events. Nonetheless, as stated in Villarini et al. (2008), there are only a few dense gauge networks on the catchment scale, so the verification of findings from studies in other regions is challenging. Regional variations, such as topography or rain type, can lead to differences in the degree of rainfall variability and uncertainty (e.g., Buytaert et al., 2006; Prein and Gobiet, 2017). Therefore, some of the general conclusions of this study may only be representative for mid-latitude regions with moderate topography. In addition, more robust interpretation of the rainfall spatial structure beyond the network dimension (> 15 km) needs to be complemented by additional larger-scale gauge data. For instance, Schroeer et al. (2018) used three different scales of networks, including the WEGN, to estimate the underestimation of maximum area precipitation of extreme convective over the range of 1 km to 30 km. It should be noted that WEGN has a high flexibility in terms of providing rainfall data within various spatial scales thanks to both high-resolution and quasi-grid configuration of the gauges. In this context, WEGN will continue providing observational evidence to explore local-to-catchment scale rainfall processes over the next years.

Data availability. WegenerNet data products are available at www.wegenernet.org.

Appendix A: Heavy rainfall events

Table A1 ~~shows~~ gives general information about the selected heavy rainfall events studied in Sect. 4 and Sect. 5. The events are corresponding to the days when total rainfall is greater than the 90th percentiles of daily rainfall (06-06 UTC), during the warm season. Peak ratio is given as a ratio of peak hourly rainfall to daily total. Rainfall in the region during the summer months is triggered by the advection of humid air masses from the Adriatic Sea. Heavy rainfall events are closely linked with local thunderstorms (Matulla et al., 2003, see also Sect. 2). The rain type is not explicitly considered for the event selection.

Appendix B: Definition of rain-gauge sub-networks

Figure A1 shows the selection order of WEGN gauges for defining the low-density sub-networks that were used in Fig. 7 of Sect. 4. Priority consideration was given to the actual location of operational weather stations within the WEGN network; the selected gauges 1 and 2 are located nearest to the member stations of the Austrian weather service (ZAMG) and the gauges 3, 4, and 5 are nearest to the rain gauges operated by the Austrian hydrographic services (AHYD). The gauges afterward were arbitrarily selected, ensuring a spatially uniform distribution. Normalized standard deviation of area-of-influence was used as an index for the uniformity of gauge configuration, which fluctuated between 0.37 and 0.23 with a decreasing trend as the number of the selected gauges increases. The area-of-influence is defined as follows: small grid boxes (approx. $0.01° \times 0.01°$, a total of 406 boxes) were defined over the WEGN network and each box is assigned to the nearest gauges of a given sub-network. Then, with an assumption that the most regular gauge configuration would share the same number of boxes, standard deviation of the area-of-influence of n-gauges is calculated. For instance, for the $five$-gauges sub-network, each gauge is expected to share around 80 boxes under an ideal situation. However, in this study, the five gauges share 71 to 113 boxes each, resulting in the uniformity index of 0.35. Note that this simple method does not consider the degree of centralization. The uniformity index defined here is also used for Fig. 6 to select well- and badly-distributed n-gauge networks.

References

Bárdossy, A. and Das, T.: Influence of rainfall observation network on model calibration and application, Hydrol. Earth Syst. Sci., 12, 77–89, https://doi.org/10.5194/hess-12-77-2008, 2008.

Berezowski, T., Szcześniak, M., Kardel, I., Michałowski, R., Okruszko, T., Mezghani, A., and Piniewski, M.: CPLFD-GDPT5: High-resolution gridded daily precipitation and temperature data set for two largest Polish river basins, Earth Syst. Sci. Data, 8, 127–139, https://doi.org/10.5194/essd-8-127-2016, 2016.

Buytaert, W., Celleri, R., Willems, P., Biévre, B. D., and Wyseure, G.: Spatial and temporal rainfall variability in mountainous areas: A case study from the south Ecuadorian Andes, J. Hydrol., 329, 413–421, https://doi.org/10.1016/j.jhydrol.2006.02.031, 2006.

Chaudhary, S., Dhanya, C., and Vinnarasi, R.: Dry and wet spell variability during monsoon in gauge-based gridded daily precipitation datasets over India, J. Hydrol, 546, 204–218, https://doi.org/10.1016/j.jhydrol.2017.01.023, 2017.

Ciach, G. J. and Krajewski, W. F.: On the estimation of radar rainfall error variance, Adv. Water Resour., 22, 585–595, https://doi.org/10.1016/S0309-1708(98)00043-8, 1999.

Ciach, G. J. and Krajewski, W. F.: Analysis and modeling of spatial correlation structure in small-scale rainfall in Central Oklahoma, Adv. Water Resour., 29, 1450–1463, https://doi.org/10.1016/j.advwatres.2005.11.003, 2006.

Contractor, S., Alexander, L. V., Donat, M. G., and Herold, N.: How well do gridded datasets of observed daily precipitation compare over Australia?, Adv. Meteorol., 2015, 1–15, https://doi.org/10.1155/2015/325718, 2015.

Dai, A., Lin, X., and Hsu, K.-L.: The frequency, intensity, and diurnal cycle of precipitation in surface and satellite observations over low- and mid-latitudes, Clim. Dyn., 29, 727–744, https://doi.org/10.1007/s00382-007-0260-y, 2007.

Dhib, S., Mannaerts, C. M., Bargaoui, Z., Retsios, V., and Maathuis, B. H.: Evaluating the MSG satellite Multi-Sensor Precipitation Estimate for extreme rainfall monitoring over northern Tunisia, Weather Clim Extrem, 16, 14–22, https://doi.org/10.1016/j.wace.2017.03.002, 2017.

Eggert, B., Berg, P., Haerter, J. O., Jacob, D., and Moseley, C.: Temporal and spatial scaling impacts on extreme precipitation, Atmos. Chem. Phys., 15, 5957–5971, https://doi.org/10.5194/acp-15-5957-2015, 2015.

Gebremichael, M. and Krajewski, W. F.: Assessment of the statistical characterization of small-scale rainfall variability from radar: analysis of TRMM ground validation datasets, J. Appl. Meteorol. Clim., 43, 1180–1199, https://doi.org/10.1175/1520-0450(2004)043<1180:AOTSCO>2.0.CO;2, 2004.

Girons Lopez, M., Wennerström, H., Nordén, L.-Å., and Seibert, J.: Location and density of rain gauges for the estimation of spatial varying precipitation, Geogr. Ann.: Ser. A, Phys. Geogr., 97, 167–179, https://doi.org/10.1111/geoa.12094, 2015.

Habib, E. and Krajewski, W. F.: Uncertainty analysis of the TRMM ground-validation radar-rainfall products: Application to the TEFLUN-B field campaign, J. Appl. Meteorol. Clim., 41, 558–572, https://doi.org/10.1175/1520-0450(2002)041<0558:UAOTTG>2.0.CO;2, 2002.

Habib, E., Krajewski, W. F., and Ciach, G. J.: Estimation of rainfall interstation correlation, J. Hydrometeorol., 2, 621–629, https://doi.org/10.1175/1525-7541(2001)002<0621:EORIC>2.0.CO;2, 2001.

Hofstra, N., New, M., and McSweeney, C.: The influence of interpolation and station network density on the distributions and trends of climate variables in gridded daily data, Clim. Dyn., 35, 841–858, https://doi.org/10.1007/s00382-009-0698-1, 2010.

Hou, A. Y., Kakar, R. K., Neeck, S., Azarbarzin, A. A., Kummerow, C. D., Kojima, M., Oki, R., Nakamura, K., and Iguchi, T.: The Global Precipitation Measurement mission, Bull Am Meteorol Soc., 95, 701–722, https://doi.org/10.1175/BAMS-D-13-00164.1, 2014.

Huff, F. A. and Shipp, W. L.: Spatial correlations of storm, monthly and seasonal precipitation, J. Appl. Meteorol. Climatol, 8, 542–550, https://doi.org/10.1175/1520-0450(1969)008<0542:SCOSMA>2.0.CO;2, 1969.

Jaffrain, J. and Berne, A.: Quantification of the small-scale spatial structure of the raindrop size distribution from a network of disdrometers, J. Appl. Meteorol. Clim., 51, 941–953, https://doi.org/10.1175/JAMC-D-11-0136.1, 2012.

Kann, A., Meirold-Mautner, I., Schmid, F., Kirchengast, G., Fuchsberger, J., Meyer, V., Tˇuchler, L., and Bica, B.: Evaluation of high-resolution precipitation analyses using a dense station network, Hydrol. Earth Syst. Sci.,, 19, 1547–1559, https://doi.org/10.5194/hess-19-1547-2015, 2015.

Keller, V. D. J., Tanguy, M., Prosdocimi, I., Terry, J. A., Hitt, O., Cole, S. J., Fry, M., Morris, D. G., and Dixon, H.: CEH-GEAR: 1 km resolution daily and monthly areal rainfall estimates for the UK for hydrological and other applications, Earth Syst. Sci. Data, 7, 143–155, https://doi.org/10.5194/essd-7-143-2015, 2015.

Kirchengast, G., Kabas, T., Leuprecht, A., Bichler, C., and Truhetz, H.: WegenerNet: A pioneering high-resolution network for monitoring weather and climate, Bull. Amer. Meteor. Soc., 95, 227–242, https://doi.org/10.1175/BAMS-D-11-00161.1, 2014.

Lockhoff, M., Zolina, O., Simmer, C., and Schulz, J.: Evaluation of satellite-retrieved extreme precipitation over Europe using gauge observations, J. Clim, 27, 607–623, https://doi.org/10.1175/JCLI-D-13-00194.1, 2014.

Ly, S., Charles, C., and Degré, A.: Geostatistical interpolation of daily rainfall at catchment scale: The use of several variogram models in the Ourthe and Ambleve catchments, Belgium, Hydrol. Earth Syst. Sci., 15, 2259–2274, https://doi.org/10.5194/hess-15-2259-2011, 2011.

Mandapaka, P. V. and Qin, X.: Analysis and characterization of probability distribution and small-scale spatial variability of rainfall in Singapore using a dense gauge network, J. Appl. Meteorol. Clim., 52, 2781–2796, https://doi.org/10.1175/JAMC-D-13-0115.1, 2013.

Marra, F., Destro, E., Nikolopoulos, E. I., Zoccatelli, D., Creutin, J. D., Guzzetti, F., and Borga, M.: Impact of rainfall spatial aggregation on the identification of debris flow occurrence thresholds, Hydrol. Earth Syst. Sci., 21, 4525–4532, https://doi.org/10.5194/hess-21-4525-2017, 2017.

Matulla, C., Penlap, E. K., Haas, P., and Formayer, H.: Comparative analysis of spatial and seasonal variability: Austrian precipitation during the 20th century, Int. J. Climatol., 23, 1577–1588, https://doi.org/10.1002/joc.960, 2003.

O, S., Foelsche, U., Kirchengast, G., and Fuchsberger, J.: Validation and correction of rainfall data from the WegenerNet high density network in southeast Austria, J. Hydrol., 556, 1110–1122, https://doi.org/10.1016/j.jhydrol.2016.11.049, 2018a.

O, S., Foelsche, U., Kirchengast, G., Fuchsberger, J., Tan, J., and Petersen, W. A.: Evaluation of GPM IMERG Early, Late, and Final rainfall estimates using WegenerNet gauge data in southeastern Austria, Hydrol. Earth Syst. Sci., 21, 6559–6572, https://doi.org/10.5194/hess-21-6559-2017, 2018b.

Peleg, N., Ben-Asher, M., and Morin, E.: Radar subpixel-scale rainfall variability and uncertainty: Lessons learned from observations of a dense rain-gauge network, Hydrol. Earth Syst. Sci., 17, 2195–2208, https://doi.org/10.5194/hess-17-2195-2013, 2013.

Prein, A. F. and Gobiet, A.: Impacts of uncertainties in European gridded precipitation observations on regional climate analysis: Uncertainty in European Precipitation, Int. J. Climatol., 37, 305–327, https://doi.org/10.1002/joc.4706, 2017.

Schroeer, K., Kirchengast, G., and O, S.: Strong dependence of extreme convective precipitation intensities on gauge network density, Geophys. Res. Lett., https://doi.org/10.1029/2018GL077994, accepted, 2018.

Sillmann, J., Thorarinsdottir, T., Keenlyside, N., Schaller, N., Alexander, L. V., Hegerl, G., Seneviratne, S. I., Vautard, R., Zhang, X., and Zwiers, F. W.: Understanding, modeling and predicting weather and climate extremes: Challenges and opportunities, Weather. Clim. Soc., 18, 65–74, https://doi.org/10.1016/j.wace.2017.10.003, 2017.

Steiner, M., Smith, J. A., Burges, S. J., Alonso, C. V., and Darden, R. W.: Effect of bias adjustment and rain gauge data quality control on radar rainfall estimation, Water Resour. Res., 35, 2487–2503, https://doi.org/10.1029/1999WR900142, 1999.

Svensson, C. and Jones, D.: Review of methods for deriving areal reduction factors: Review of ARF methods, J. Flood Risk Manag., 3, 232–245, https://doi.org/10.1111/j.1753-318X.2010.01075.x, 2010.

Svoboda, V., Máca, P., Hanel, M., and Pech, P.: Spatial correlation structure of monthly rainfall at a mesoscale region of north-eastern Bohemia, Theor. Appl. Climatol, 121, 359–375, https://doi.org/10.1007/s00704-014-1241-9, 2015.

5   Syed, K. H., Goodrich, D. C., Myers, D. E., and Sorooshian, S.: Spatial characteristics of thunderstorm rainfall fields and their relation to runoff, J. Hydrol., 271, 1–21, https://doi.org/10.1016/S0022-1694(02)00311-6, 2003.

Tan, J., Petersen, W. A., Kirchengast, G., Goodrich, D. C., and Wolff, D. B.: Evaluation of Gobal Precipitation Measurement rainfall estimates against three dense gauge networks, J. Hydrometeorol., 19, 517–532, https://doi.org/10.1175/JHM-D-17-0174.1, 2018.

Tian, F., Hou, S., Yang, L., Hu, H., and Hou, A.: How does the evaluation of the GPM IMERG rainfall product depend on gauge density and
10   rainfall intensity?, J. Hydrometeorol., 19, 339–349, https://doi.org/10.1175/JHM-D-17-0161.1, 2018.

Tokay, A., Roche, R. J., and Bashor, P. G.: An experimental study of spatial variability of rainfall, J. Hydrometeorol., 15, 801–812, https://doi.org/10.1175/JHM-D-13-031.1, 2014.

Tozer, C. R., Kiem, A. S., and Verdon-Kidd, D. C.: On the uncertainties associated with using gridded rainfall data as a proxy for observed, Hydrol. Earth Syst. Sci., 16, 1481–1499, https://doi.org/10.5194/hess-16-1481-2012, 2012.

15   Velasco-Forero, C. A., Sempere-Torres, D., Cassiraga, E. F., and Jaime Gómez-Hernández, J.: A non-parametric automatic blending methodology to estimate rainfall fields from rain gauge and radar data, Adv. Water Resour., 32, 986–1002, https://doi.org/10.1016/j.advwatres.2008.10.004, 2009.

Villarini, G., Mandapaka, P. V., Krajewski, W. F., and Moore, R. J.: Rainfall and sampling uncertainties: A rain gauge perspective, J. Geophys. Res., 113, https://doi.org/10.1029/2007JD009214, 2008.

20   Wood, S. J., Jones, D. A., and Moore, R. J.: Accuracy of rainfall measurement for scales of hydrological interest, Hydrol. Earth Syst. Sci., 4, 531–543, https://doi.org/10.5194/hess-4-531-2000, 2000.

Xu, H., Xu, C.-Y., Chen, H., Zhang, Z., and Li, L.: Assessing the influence of rain gauge density and distribution on hydrological model performance in a humid region of China, J. Hydrol., 505, 1–12, https://doi.org/10.1016/j.jhydrol.2013.09.004, 2013.

Yatagai, A., Kamiguchi, K., Arakawa, O., Hamada, A., Yasutomi, N., and Kitoh, A.: APHRODITE: Constructing a long-term
25   daily gridded precipitation dataset for Asia based on a dense network of rain gauges, Bull Am Meteorol Soc., 93, 1401–1415, https://doi.org/10.1175/BAMS-D-11-00122.1, 2012.
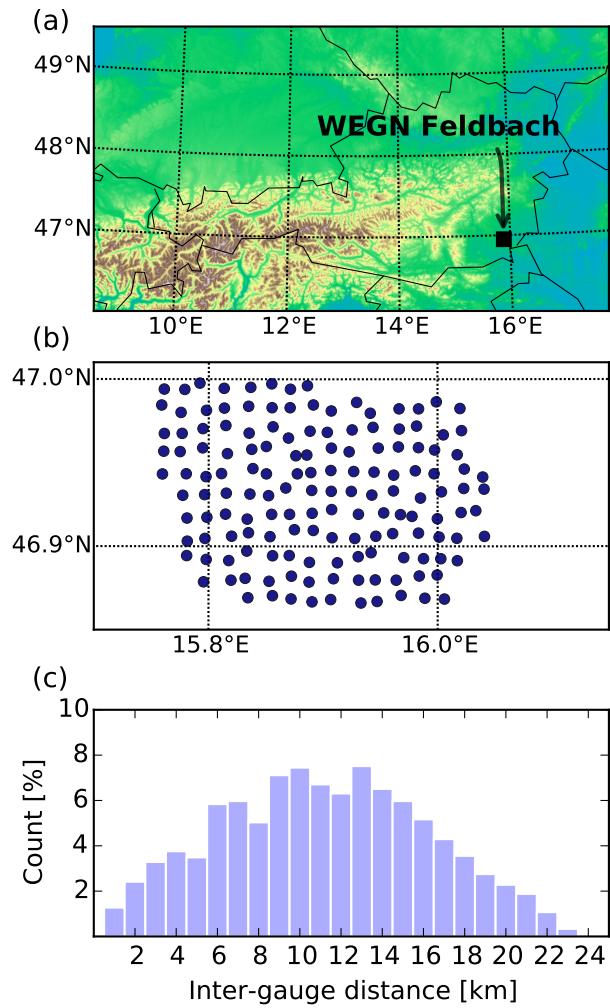
Figure 1. (a) WegenerNet Feldbach region (WEGN) network in southeastern Austria, (b) location of 150 tipping-bucket rain gauges, and (c) inter-gauge distances, rounded to the nearest 1-km bins.
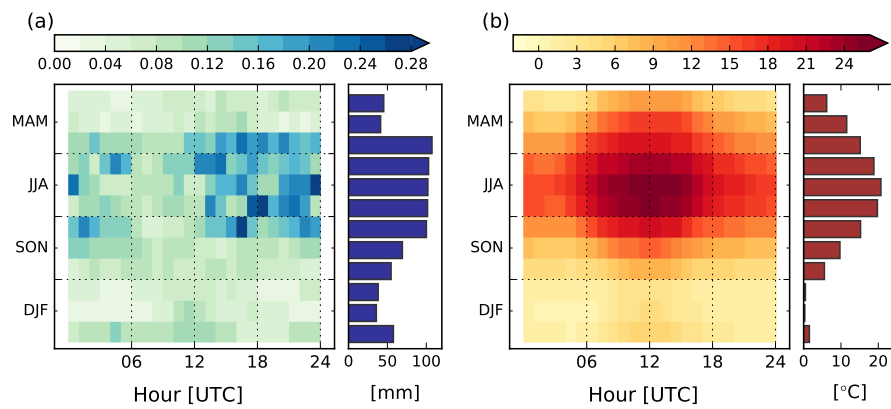
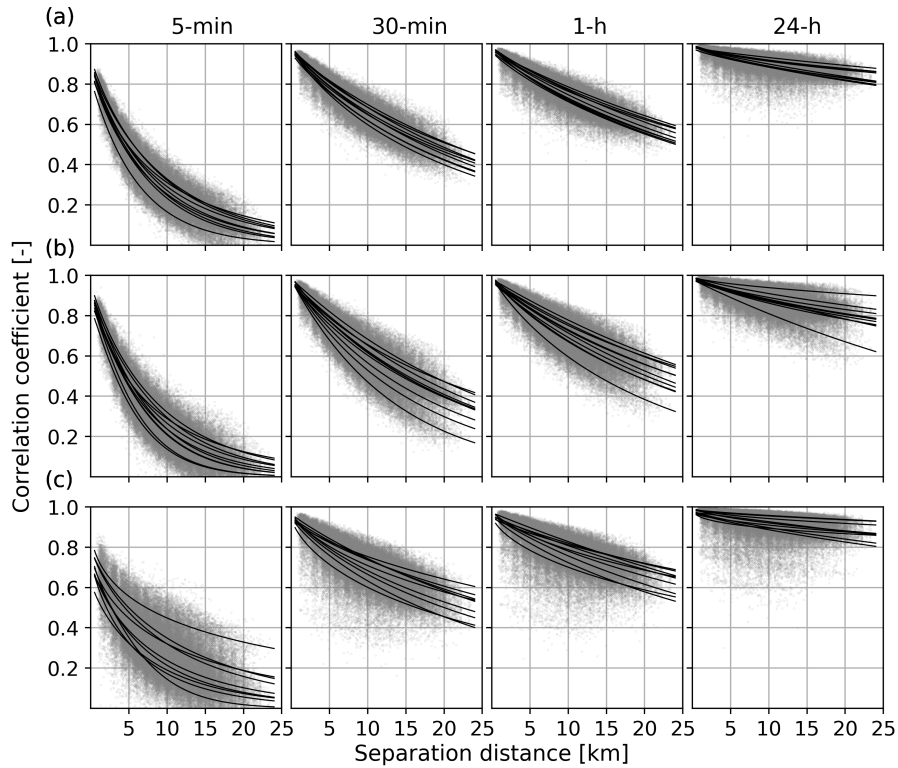**Figure 2.** Diurnal cycles of (a) rainfall and (b) temperature derived from WEGN observational data.

Figure 3. Spatial correlation of rainfall among rain gauges for (a) all-months, (b) warm season, and (c) cold season. Four selected accumulation times are shown. Each solid line represents a fitted exponential function for each year, up to 23 km (the longest inter-gauge distance of WEGN). Note that the function is fitted to correlation for separation distances <= 15 km to sample data uniformly for any spatial direction.
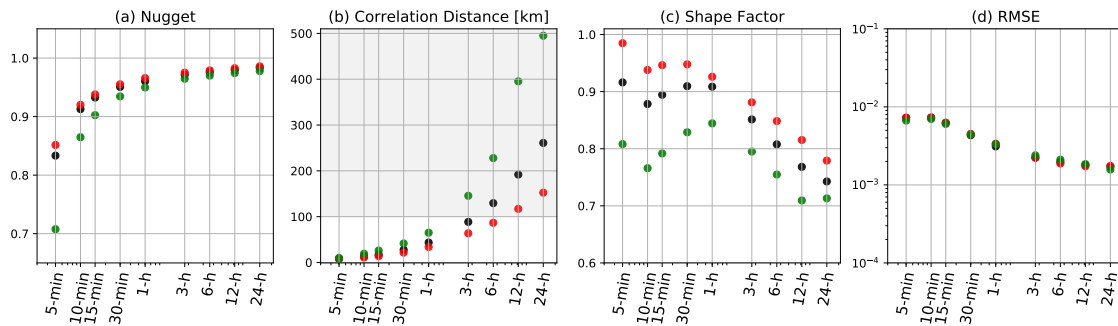
Figure 4. Dependence of (a) nugget effect, (b) correlation distance, and (c) shape factor of the fitted exponential functions on timescale. (d) shows RMSE of fitted correlation values compared to observed values (red: warm season, green: cold season, black: all-months). Note that the correlation distance values over 100 km (shaded in gray) are highly affected by the model selected to estimate the values.

18
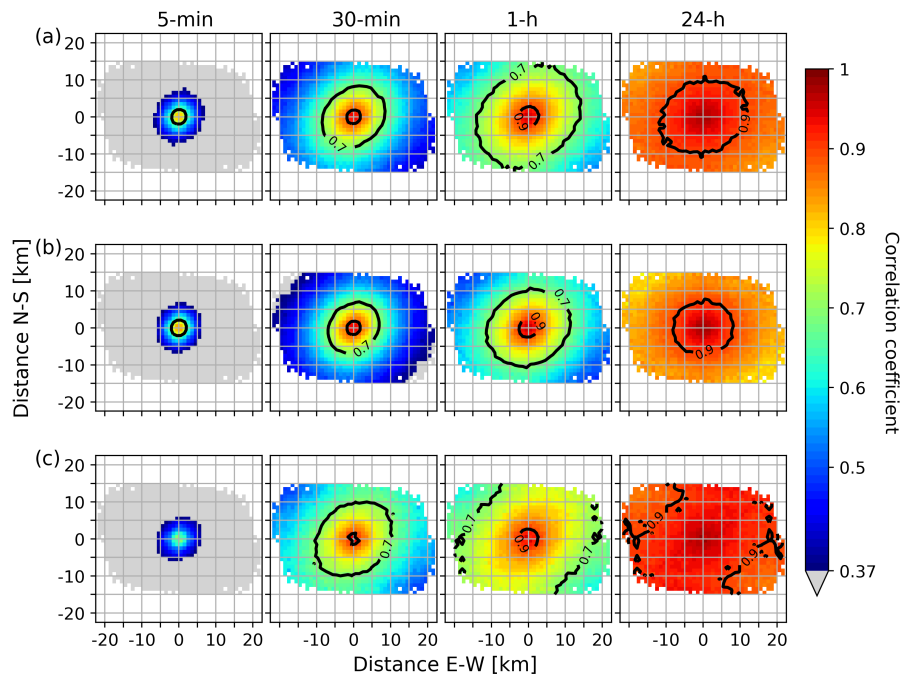
Figure 5. Same as Fig. 3, but for 10-year averaged spatial correlations in a two-dimensional space which is defined according to the distance between the gauges in east–west direction (x-axis) and ~~nort–south~~ north–south direction (y-axis). (a) all-months, (b) warm season, and (c) cold season are shown. The values beyond the e-folding distance ($\approx$0.37) are colored grey.
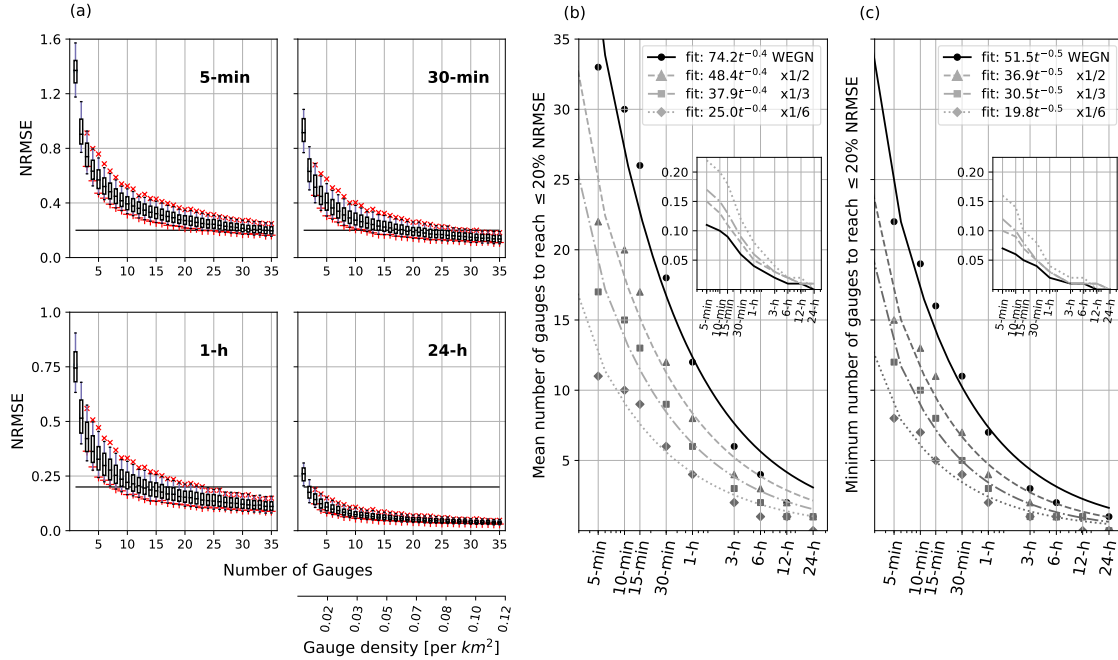
Figure 6. (a) Dependence of the accuracy of WEGN areal rainfall estimates on the number of gauges during heavy rainfall. Normalised RMSEs (NRMSEs) of 1,000 ~~different~~ random gauge combinations are used to assess the accuracy for each n-gauge network. ~~(a)~~ Four selected time accumulation are shown. Black horizontal lines correspond to 20% NRMSE. Box plots display the median, 25th and 75th percentiles of NRMSE distribution, and whiskers extend to the 10th and 90th percentiles. Red crosses and Xs show the median NRMSE for good and bad gauge configurations; 100 cases are selected, respectively, for each of the 1,000 combinations. (b) The average ~~and minimum~~ number of gauges required to obtain areal rainfall estimates with the NRMSE < 20% within the whole WEGN area (black) and ~~gray~~within the WEGN×1/2, ~~respectively~~×1/3, and ×1/6 areas (gray)~~required~~. Inset shows the results in term of gauge density. (c) Same as (b), but for the minimum number of gauges require to obtain areal rainfall estimates with ~~an~~the NRMSE <20%. Note that (a) shows the results with respect to the whole WEGN area.
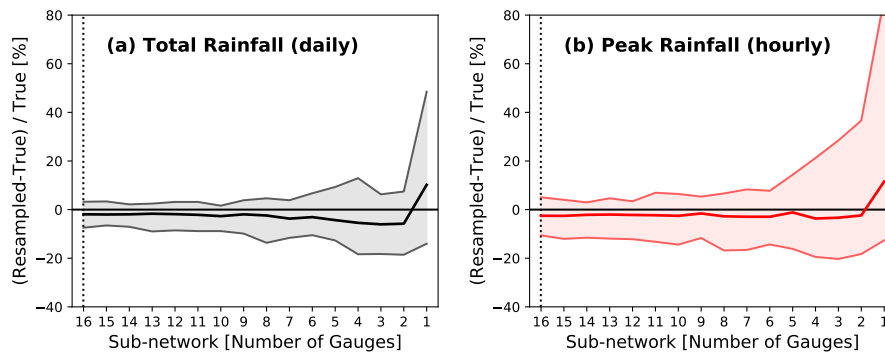
Figure 7. Dependence of the accuracy of (a) daily rainfall and (b) hourly peak intensity on the number of gauges. 71 heavy rain events are considered. The y axis displays the relative difference between resampled and true rainfall. Resampled rainfall is calculated from n-gauge sub-networks, while true rainfall is calculated using the full density WEGN network. The thick lines show the median and the shaded areas show the 10th to 90th percentile spread.
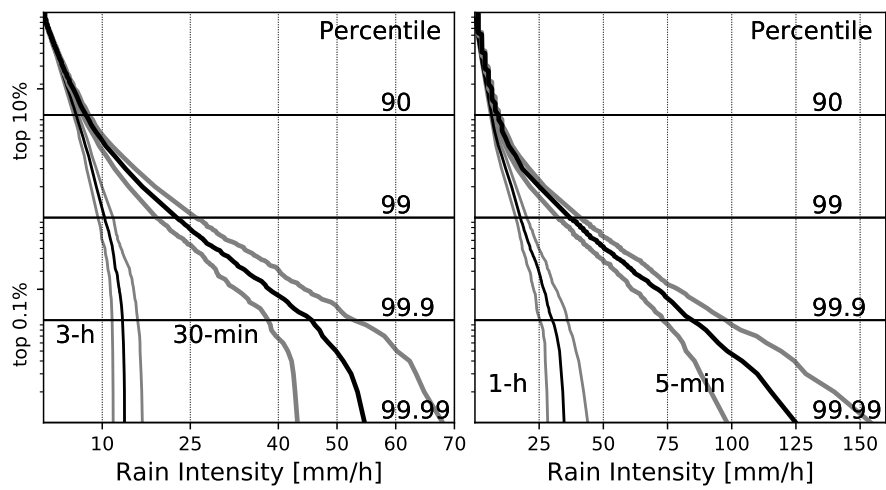
Figure 8. Distribution of gauge-level rainfall intensities corresponding to given percentile thresholds during heavy rainfall events. Four time scales are selected. Black lines show median values, gray lines show a 10th-90th percentile range among the gauges at a given threshold bin.
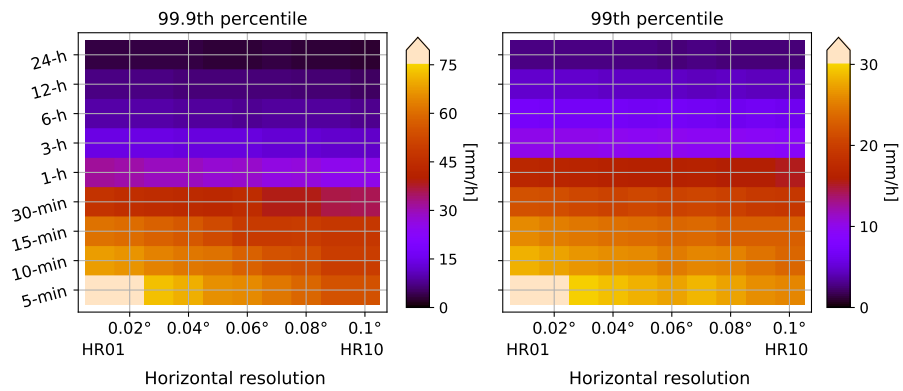
Figure 9. 99.9th and 99th percentiles of rainfall intensities derived from gridded rainfall fields with different spatial and temporal scales. Note that different color scale for the plots.
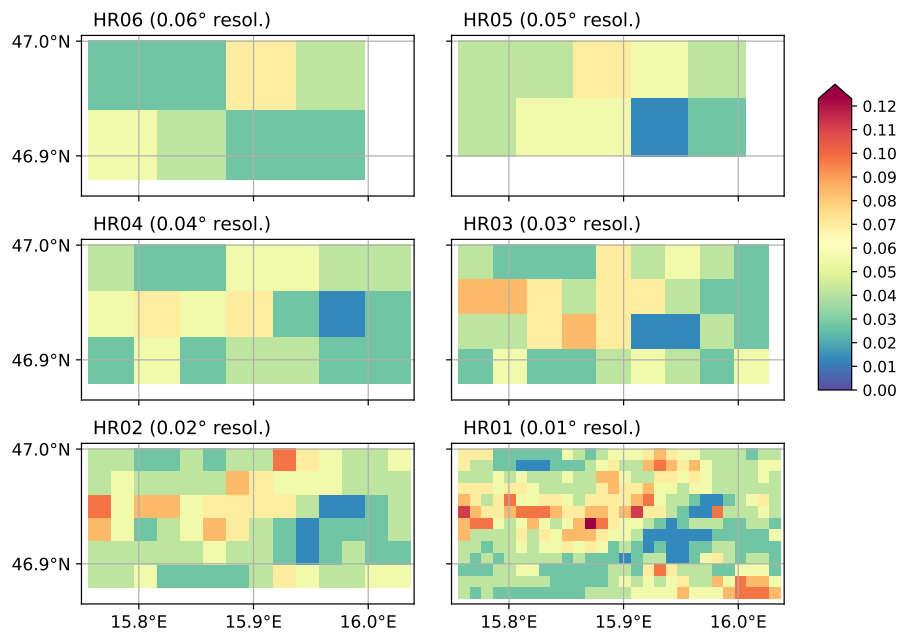
Figure 10. Occurrence of extreme events (defined as days with total rainfall $\geq$ 95th percentile of daily rainfall intensity during heavy rainfall events at HR01) at different horizontal grid spacing.
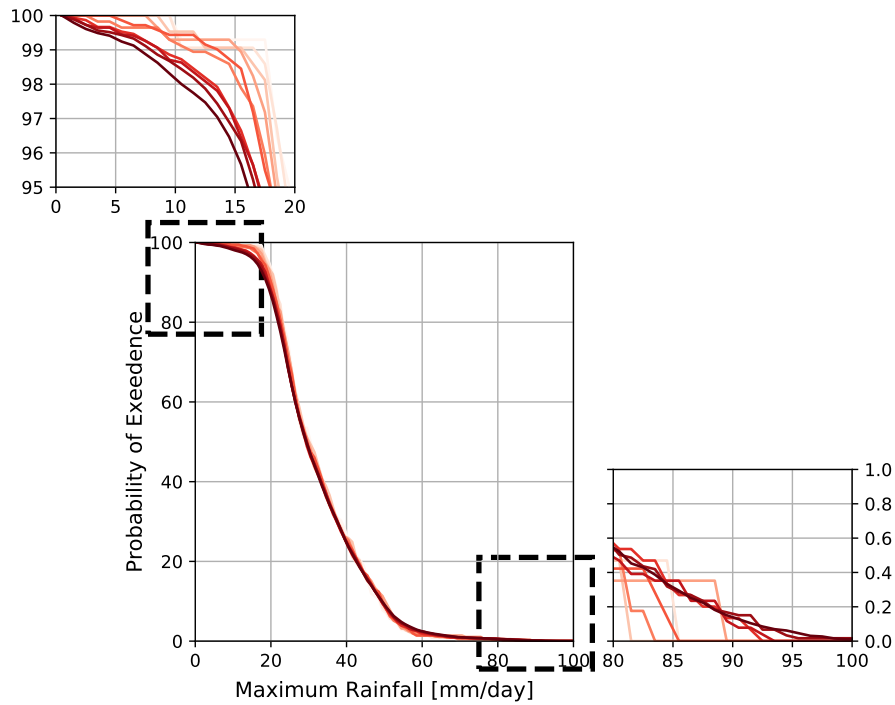
Figure 11. Probability of occurrence of heavy rainfall for different horizontal resolutions. Darker red represents higher horizontal resolution (from $0.1°$ to $0.01°$).

Table A1. Information of selected heavy rainfall events

	Min	Median	Max
Total rainfall ($\mathrm{mm\,d^{-1}}$)	19.8	28.1	64.1
Peak hourly rainfall ($\mathrm{mm\,h^{-1}}$)	2.6	8.6	26.2
Peak ratio (Peak hourly rainfall/Total rainfall)	7.8	25.4	91.0
Duration (h)	2.0	9.5	22.5

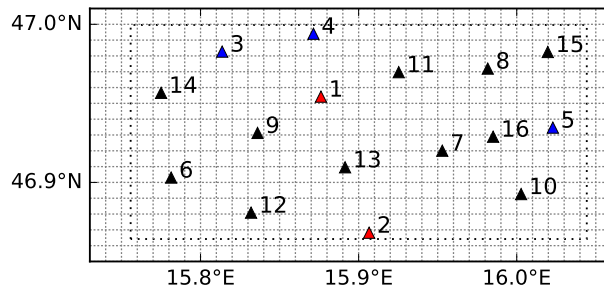**Figure A1.** Selected WEGN gauges for Fig. 7. The gauges nearest to operational weather stations of the ZAMG and AHYD are in red and blue, respectively.