Estimation of extreme daily precipitation return levels at-site and in ungauged locations using the simplified MEV approach

Arianna Miniussi[1] and Francesco Marra[2]

[1] Department Catchment Hydrology, Helmholtz Centre for Environmental Research – UFZ, Halle (Saale), Germany

[2] Institute of Atmospheric Sciences and Climate, National Research Council of Italy (CNR-ISAC), Bologna, Italy

Abstract

Estimating extreme precipitation return levels at ungauged locations is key for hydrological applications and risk management, and demands improved techniques to decrease the large uncertainty of traditional methods. Here, we leverage the perks of the simplified Metastatistical extreme value (SMEV) approach with a twofold aim: we show how it can be effectively used in situations in which the ordinary daily precipitation events cannot be fully described using a two-parameter distribution, and we examine the performance of different interpolation techniques for the estimation of return levels in ungauged locations. SMEV proved adequate at representing at-site extremes for a set of 4000+ stations in Germany, with a general tendency to underestimate the probability of the largest annual maxima. At the same time SMEV tends to overestimate with respect to the design return levels currently adopted in the country, suggesting that these might actually underestimate the distribution tail. Among the investigated methods, the inverse distance weighted interpolation of SMEV parameters provides the most accurate estimates of extreme return levels for ungauged locations, with typical standard errors of 0.79 (0.83) for rain gauge densities of 1/500 km^{-2} (1/1000 km^{-2}). Albeit only less than 10% of the variance in estimation errors is explained by elevation, the correlation between SMEV parameters and orography (up to 43% explained variance) suggests that future applications should test the inclusion of such information in spatial estimates.

Highlights

- At-site and spatial Simplified Metastatistical Extreme Value in Germany

- Light-tails for precipitation in Germany may lead to dangerous underestimation

- We identify a correlation between orography and parameters of daily rainfall

- Inverse distance weighting of parameters minimizes error in ungauged locations

1. Introduction

Accurate estimates of extreme daily precipitation amounts associated to rare yearly exceedance probabilities, generally termed *return levels*, constitute the basis for hydraulic design, risk assessment and mitigation, environmental policies, and insurance/reinsurance business (Katz et al., 2002; Groenemeijer et al., 2015; Hailegeorgis and Alfredsen, 2017). Deriving such quantities requires observational records and statistical models able to reproduce the upper tail of the intensity distributions. The extreme value theorem shows that, under some general hypotheses, *extremes* (i.e., the maximum values recorded in each year or all the values exceeding a high threshold) may only converge to known classes of three-parameter distributions, independent of the distribution class describing the underlying *ordinary* events (Fisher and Tippett, 1928; Gnedenko, 1943). Three-parameter distributions are thus traditionally fitted to the observed extremes and used to extrapolate the information for low yearly exceedance probabilities (Jenkinson, 1955; Coles, 2001). Doing so, the large stochastic uncertainties characterizing the sampled extremes are inherited by the parameters, in particular by the shape parameter (Lu and Stedinger, 1992; Morrison and Smith, 2002); this is highly undesirable because the estimated return levels are highly sensitive to it. Specific techniques are used to reduce this uncertainty and generally entail the a-priori choice of a shape parameter (of which the use of two-parameter distributions, such as the Gumbel distribution, is a particular case; Papalexiou and Koutsoyiannis, 2013), the use of a regionally-derived shape parameter (Hosking and Wallis, 1997), or of prior distributions to restrict its value (Martins and Stedinger, 2000). As they are based on strong homogeneity assumptions, these methods do not permit to adequately capture small-scale variations, a problem which could become important in presence of local climatological gradients (e.g., orographic effects; Avanzi et al., 2015).

The global need for local information collides with the sparseness of rain gauge networks worldwide (Kidd et al., 2017), a problem which is exacerbated by the need for long data records. Considering the direct hydrological implications of extreme precipitation, improving our ability to estimate extreme return levels at ungauged locations could contribute to addressing important open questions in the hydrological community, especially for what concerns our understanding of the variability of extremes (Blöschl et al., 2019). However, the procedures currently available to estimate extreme return levels at ungauged locations are based on regionalization approaches or on the geospatial modeling of the distribution parameters (Hosking and Wallis, 1997; Wallis et al., 2007; Blanchet and Lehning, 2010; Ceresetti et al., 2012; Das,

2019); being based on the spatial representation of traditional estimates, they suffer from the uncertainties characterizing these quantities. Remotely sensed precipitation products are getting more and more attention as alternative to observations from rain gauges. On the one hand they can provide valuable information in ungauged regions with a quasi-global coverage, but some important issues remain still open, especially for what concerns the propagation of estimation errors to the estimated return levels, the availability of short time series, and the lack of suitable validation methods (Marra, Nikolopoulos, et al., 2019).

Here, we address the problem of estimating extreme return levels in ungauged locations by evaluating the spatial representativeness of a novel statistical approach based on the distribution of the *ordinary*, as opposed to *extreme*, events (Marani and Ignaccolo, 2015). Differently from traditional methods, this metastatistical extreme value (MEV) approach assumes that the distribution class describing the ordinary events tail is known, and derives an extreme value distribution explicitly considering the occurrence frequency of ordinary events. Thanks to the use of a much larger amount of information (ordinary events versus extremes only) this approach proved to be highly accurate in representing rare return levels. This is the case of probabilities corresponding to return periods much longer than the record length (Zorzetto et al., 2016; Marra et al., 2018). Clearly, the reasoning is valid whenever the basic assumption about the ordinary events distribution is met; when this does not hold, the method leads to biased estimates. In particular, it is worth recalling that the theoretical results by Wilson and Toumi (2005) only hold for the tail of the ordinary events distribution. In a preliminary study over Germany, Wang et al. (2020) showed that the two-parameter Weibull distribution, typically used to describe precipitation ordinary events in MEV approaches (e.g., Zorzetto et al., 2016), is not adequate to represent the full distribution of the ordinary events, and thus extremes, in the country. Conversely, they showed that thanks to the different handling of the available information, the Simplified MEV framework (SMEV, Marra et al., 2019) could better isolate independent events and, through left-censoring, reproduce extremes. We hence use here SMEV, which allows to (i) represent the tail of the ordinary events distribution under more general conditions (see details in Marra et al., 2020), thus extending the validity of the assumption and limiting the impact of the above issue, and to (ii) examine fine-scale variations of the distribution parameters by fully leveraging the at-site information (Marra et al., 2021). We refer to Marra et al. (2020) and Serinaldi et al. (2020) for a discussion of the differences between MEV, SMEV, and traditional approaches based on extremes.

The method was previously tested in the context of spatial modeling (Schellander et al., 2019). Here, we extend upon Schellander et al. (2019) and Marra et al. (2019) by (a) demonstrating the use of SMEV in regions where the ordinary daily precipitation events are not well described by a two-parameter Weibull distribution, and (b) examining the performance of different interpolation techniques for the estimation of

SMEV parameters in ungauged locations under varying densities of available observations in a controlled Monte Carlo experiment.

After a brief introduction about the data used and the theoretical framework, we explain the procedures developed for the evaluation of the suitability of SMEV for German daily precipitation extremes and for the spatial estimation of extreme return levels in ungauged locations; results for both the at-site and spatial estimations are then presented and discussed. A final section provides a short summary and the main conclusions.

2. Data and methodology

2.1 Study area and data

We focus on the whole of Germany, a study area chosen for its dense daily rainfall station network (average density 1/60 km^{-2}), along with the presence of diverse precipitation patterns and generating mechanisms. Precipitation in Germany mostly occurs due to westerly circulation patterns, whose influence decreases moving towards the east; topography significantly affects precipitation spatial patterns (Figure 1a). Extratropical cyclones of type "Vb" (first defined by Van Babber (1891); Nissen et al. (2013); Grams et al. (2014)) are a peculiar atmospheric phenomenon linked to extreme precipitation and flash floods in Central-Eastern Europe. The characteristic pathway of the "Vb" cyclones originates in the Bay of Biscay, the Balearic Sea or the Ligurian Sea and then moves north-eastward around the Alps into eastern Europe and the Baltic Sea. These cyclones sometimes (around 2.3 events/year, Messmer et al., 2015) trigger heavy precipitation over the northern Alpine region and central and eastern Europe, affecting the distribution of extremes. In fact, about 40% of the "Vb" cyclones are associated with precipitation exceeding the 95^{th} percentile of daily precipitation over Central Europe (Nissen et al., 2013). Considering the reference period 1961-1990, the lowest annual precipitation amounts occur in north-east Germany (577 mm/year) and the highest in the Alps (1935 mm/year) (Deutscher Wetterdienst, 2017).

We collected daily precipitation data from the Deutsche Wetter Dienst (DWD), which provides a network of about 6000 daily stations covering the whole Germany, from the year 1781. A first screening consists in the control of missing values: we retain only series with more than 10 complete years (i.e., including at least 330 days of observations), resulting in 5019 available stations (density of approximately 1/71 km^{-2}). These are then further skimmed on the basis of the performance of the SMEV model (see Section 2.3).

2.2 The Metastatistical Extreme Value (MEV) distribution and its simplified form (SMEV)

The MEV distribution, originally introduced by Marani and Ignaccolo (2015), builds the distribution of the extremes considering all the so-called *ordinary* events, i.e., the all independent realizations of the variable

of interest. The cumulative distribution of maxima emerges from the full distribution of the ordinary events which is sampled a variable number of times, namely the number of occurrences of the events themselves. The discrete expression of the MEV cumulative distribution function $\zeta(x)$ reads as follows:

$$\zeta(x) = \frac{1}{M}\sum_{j=1}^{M}\left[F(x;\boldsymbol{\theta}_{\mathbf{j}})\right]^{\mathrm{n_j}} \qquad (1)$$

where M is the length of the available time series, $F(x;\boldsymbol{\theta}_{\mathbf{j}})$ the yearly distributions of the ordinary events and n_j their yearly number (in this formulation, the inter-annual variability is taken into account, as indicated by the subscript j).

It is thus straightforward to notice that one of the greatest perks of the approach is the fact that it effectively uses most of the available data, and does not limit the estimation to a small subset of it (such as the annual maxima or the exceedances of a high threshold).

Marra et al. (2019) proposed a simplified version of the MEV (SMEV) which neglects inter-annual variability (dependence of the parameters and events occurrences on j in Eq. (1)) in favor of parameter estimation accuracy, an approach that was also adopted by Schellander et al. (2019) and discussed in Miniussi and Marani (2020). Given the preliminary results obtained over Germany using MEV and SMEV (Wang et al., 2020), this is deemed as a crucial advantage as it allows to accurately estimate parameters describing only the portion of ordinary events which are well approximated by the chosen model (details in Marra et al., 2020). Here we will use SMEV considering one rainfall type; in these conditions, the cumulative distribution function describing yearly exceedance probabilities can be written as:

$$\zeta(x) = F(x;\boldsymbol{\theta})^{\bar{\mathrm{n}}} \qquad (2)$$

where $F(x;\boldsymbol{\theta})$ is the distribution describing the portion of ordinary events of interest (see definition in Sec. 2.3), $\boldsymbol{\theta}$ is a set of two parameters describing the distribution, and $\bar{n}$ is their average yearly number.

2.3 Identification of wet periods and parameter estimation

Many precipitation records in Germany are characterized by snow events, but to avoid dealing with two potentially different populations of events, a problem which could affect the quality of the estimates (Marra et al., 2018), we discard snow days (i.e., values flagged with 7 in the DWD archive, while mixed precipitation –flag 8– is here considered as liquid precipitation) as they do not significantly contribute to annual maxima. Following the unified approach presented in Marra et al. (2020), we define as storms the sequences of consecutive wet days (i.e., any day with a precipitation value greater than or equal to 0.1 mm, Marra et al., 2019) separated by at least one dry day. Ordinary events are then defined as the maximum

values observed within each storm. The same methodology was used by Miniussi et al. (2020) to identify cyclone-generated rainfall.

Following theoretical reasoning (Wilson and Toumi, 2005) and previous MEV works on daily precipitation (e.g., Marani and Ignaccolo (2015)), we assume the high-intensity portion of the ordinary events is described by a two-parameter Weibull distribution, in the form:

$$F(x; C, w) = 1 - \exp(-(\frac{x}{C}))^w \qquad (3)$$

where C is the scale parameter and w the shape parameter. Distributions with larger shape parameter are characterized by lighter tails, and vice versa. We estimate the parameters by left-censoring the portion of ordinary events which is not well described by our two-parameter model and using a least-squares regression in Weibull transformed coordinates on the remaining data points (Marani and Ignaccolo, 2015); this means that the weight in the probability of the left-censored events is retained. The need of censoring the upper part of the distributions was also highlighted by Wang et al. (2020) when analyzing sub-daily precipitation in Germany. Following the procedure described in Marra et al. (2019) and Marra et al. (2020), which permits to maximize the ability of ordinary events to represent extremes in the area, we optimized the threshold for left-censoring to 90%. Although locally different thresholds may define the tail of the ordinary events distribution, we expect this to be a climatological characteristic of a region. In particular, we find that left-censoring 90% of the ordinary events allows for an accurate identification of the tail throughout the study region. To avoid stochastic uncertainties that would affect our results, we use this threshold for our parameter estimation procedure meaning that, at each station, the largest 10% of the ordinary events are used to estimate the parameters. Given the typical number of ordinary events in the region, the two parameters of the Weibull distributions are estimated using a number of points between 44 and 1086 (median value 211), which is 3-6 times larger than the number of points used by traditional approaches to estimate three parameters.

In order to evaluate the suitability of SMEV on the German dataset, we proceed by using the test presented in Marra et al. (2020): 1) we estimate the scale and shape parameters of the Weibull distribution on the ordinary events identified as explained above, while $\bar{n}$ is directly computed as the mean of the wet days in the series; 2) given the parameters of SMEV (Weibull scale and shape, and $\bar{n}$), we generate 1000 series of Weibull-distributed synthetic precipitation values M years long (with M indicating the original length of the time series) and with a number of rainy days equal to $\bar{n}$; 3) if 90% of the original maxima is within the 90% confidence interval of the maxima from the synthetic-generated time series, we consider SMEV a robust model for the station, otherwise we discard it. This procedure is performed by both retaining and explicitly censoring annual maxima for the estimation of the Weibull parameters, in order to provide also

a complete independent validation of the hypothesis. It is worth here to point out that synthetic experiments conducted using typical ordinary events tails (stretched-exponential such as Weibull, and power-type) show that the test is rather strict and applied to the generalized extreme value distribution as a model for annual maxima, rejects about 20–40% of the samples.

2.4 Parameter estimation at the ungauged locations

The algorithm for estimating parameters and, consequently, quantiles associated to specific return periods at ungauged locations consists in the following steps: 1) we select a target station (i.e., the station that is considered as ungauged); 2) we randomly select a subset of gauges modifying the station density δ between 1/200 and 1/100000 km^{-2}. To avoid the creation of clusters, we impose a minimum inter-distance between stations, set equal to $d_{\min} = (0.5\delta^{0.5})$ (following Nikolopoulos et al., 2015, which used a network characterized by inter-gauge distances similar to ours) with δ = 200, 500, 1000, 2000, 5000, 10000, 20000 and 100000 km^2; 3) we estimate the SMEV parameters of the "ungauged" station using three widely used interpolation methods, namely two deterministic interpolation methods (nearest neighbor, NN and inverse distance weighting, IDW) and one geostatistical method (ordinary kriging, OK, with a spherical variogram) (details in Webster and Oliver, 2007). Finally, we compute the SMEV quantiles for the "ungauged" station. We iterate the procedure outlined above 100 times for each "ungauged" station, in order to have a full statistical description of the errors at each location.

We chose these interpolation methods as they have been extensively applied in interpolation studies, and have been shown to provide different performances according to the objectives of the investigations. The NN method, despite being extremely simple, has been proved to be affected by smaller biases than more sophisticated methods in some applications (e.g., Nikolopoulos et al., 2015); IDW is an efficient methodology that allows to include information from many stations without a significant increase in the computation time; OK is a more sophisticated geostatistical method but presents drawbacks like the higher computational time along with the higher sensitivity to the station network density (for example, Webster and Oliver, 2007, indicate the need of availability of at least 100 measurement locations –ideally 150- to adequately estimate the variogram).

We focus on target return periods (where the return period T_r is defined as the inverse of the yearly exceedance probability, which we estimate using the Weibull plotting position: $p(i) = 1-[i/(M+1)]$, with $i = 1,2\ldots M$ being the rank of the value, and M the length of the times series) to compare the new estimations with the ones estimated in the at-site approach.

2.5 Evaluation metrics

We evaluate the accuracy of the estimations for each interpolation method (m) and station density (d), computing the relative error on the parameters (Weibull shape and scale parameters and average number of rainy days) as:

$$\varepsilon_{\theta}(m,d) = \frac{\theta_{\mathrm{int(m,d)}} - \theta_{\mathrm{est}}}{\theta_{\mathrm{est}}} \qquad (4)$$

where $\theta_{\mathrm{int(m,d)}}$ is the parameter estimated in the location considered as ungauged using the interpolation method m and with a station density d, and θ_{est} is the at-site estimated parameter, here used as a reference.

To summarize with one single number the relative errors on the 100 iterations with respect to the reference return periods, we finally computed the fraction standard error (FSE) as:

$$\varepsilon_{FS}(m,d) = \frac{1}{x_{\mathrm{T_r}}}\left(\sum_{j=1}^{N_r}\left(x_{\mathrm{int(m,d)_{Tr}}}{}^{(\mathrm{j})} - x_{\mathrm{T_r}}\right)^2\right)^{0.5} \qquad (5)$$

where $x_{\mathrm{int(m,d)_{Tr}}}{}^{(\mathrm{j})}$ is the *j-th* quantile corresponding to the return period T_r computed for the ungauged location with the interpolation method m and a density d, and x_{Tr} is the at-site estimated quantile for the same T_r.

2.6 Assessment of tail heaviness

We use here the obesity index (introduced by Cooke and Nieboer, 2011) to evaluate the heaviness of the tail of the distributions. It is a non-parametric measure based on order statistics and is defined as the probability that the sum of the smallest and largest value of a four values random sample is higher than the sum of the other two. Being a probability, it ranges between 0 and 1 and its formulation reads as follows:

$$\mathrm{Ob(X)} = \mathrm{P}(\mathrm{X}_{(4)} + \mathrm{X}_{(1)} > \mathrm{X}_{(2)} + \mathrm{X}_{(3)} | \mathrm{X}_{(1)} \geq \mathrm{X}_{(2)} \geq \mathrm{X}_{(3)} \geq \mathrm{X}_{(4)}) \qquad (6)$$

where $\mathrm{X}_{(\mathrm{k})}$ are independent and identically distributed values randomly extracted 1000 times from the original sample (here, the series of ordinary events). Cooke and Nieboer (2014) showed that the obesity index for an exponential distribution is equal to 0.75, and that obesity index greater than 0.75 represent tails decreasing slower than the exponential ones.

2.7 Benchmarking with the state of the art

Finally, in order to provide a benchmark with previous research, we compare our results with the ones by Schellander et al. (2019), who implemented the smooth modeling method by Blanchet and Lehning (2010) originally devised for the generalized extreme value distribution parameters in the SMEV framework, but could not show significant improvements over the original approach. Mirroring the analysis presented in

Figure 9 in Schellander et al. (2019), we focus on the relative error $\varepsilon = (x_{\mathrm{int}} - x_{\mathrm{obs}})/x_{\mathrm{obs}}$, between the interpolated and the observed 50-year quantile.

3. Results and Discussion

3.1 Evaluation of SMEV model and comparison with official design precipitation

First, to evaluate the suitability and robustness of SMEV as an extreme value approach for precipitation in Germany, we perform the experiment described in Section 2.2, for which the left-censoring threshold was optimized in order to maximize the number of long stations with time series longer than 60 years to be retained in the analysis. The final number of stations (irrespective of their length) for which at least 90% of the annual maxima is within the 90% confidence intervals is 4015 out of 5019 (the median length of the time series is 37 years) resulting in an average density of approximately 1/90 km^{-2} (Figure 1b). This result is corroborated by a similar analysis in which the annual maxima were explicitly censored from the estimation of the Weibull parameters (red and yellow dots in Figure 1b represent stations in which less than 90% of these maxima lie within the 90% confidence interval). Despite the fact that some stations could fail the test due to stochastic uncertainty, this choice allows us to be on the safe side. Especially when evaluating the spatial approaches, it would be extremely complicated to disentangle potential uncertainties related to sub-optimal choices of the left-censoring from errors due to the interpolation methods, thus decreasing the robustness of our results.

Among the 1004 stations that we discarded, in the majority of the cases SMEV was underestimating the large annual maxima. When considering the upper tail of the distribution of the maxima –from 90^{th} to 99^{th} percentile- in about 50% of the stations at least one annual maxima was above the 95^{th} percentile of the samples, while the opposite (maxima below the 5^{th} percentile) occurred in a very small portion (1.8%) of the cases.

Figure S1 in the Supporting Information shows results of the test run either using all the here-defined ordinary events for estimating the distribution parameters (Figure S1a), or using the traditional MEV approach (i.e., defining ordinary events as independent daily amounts greater or equal to 1 mm, and allowing inter-annual variability of the distribution parameters, as in Zorzetto and Marani, 2019 – Figure S1b). Results of the test show that (a) when all ordinary events are used (i.e., SMEV without left-censoring), the test is successful in only 27% of the stations (green filled circles in Figure S1); (b) this percentage drops to 0.5% when inter-annual variability of the parameters is allowed as in the standard MEV. This confirms what found by Wang et al. (2020) when analyzing sub-daily rainfall in the same region.

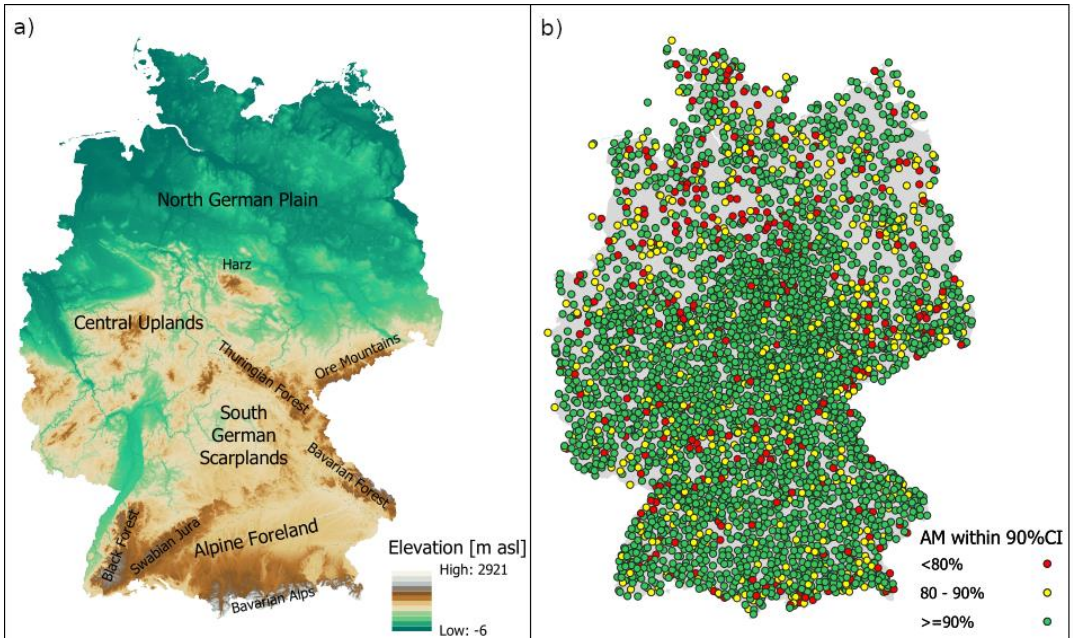

Figure 1 - Panel a): Topography of Germany with five main landscape units (the North German Plain, Central Uplands, South German Scarplands, and Alpine Foreland regions). Other minor natural areas in Germany are indicated, to ease the discussion. Panel b): The stations selected for the analysis are those for which the percentage of observed annual maxima that lie in the area where the 90% of the annual maxima are expected if they were sampled from SMEV (1000 times sampling procedure) is greater than 90%, when parameters are estimated considering annual maxima. The color scale represents the percentage of annual maxima within the 90% confidence interval of the 1000 times resampling procedure when parameters are estimated explicitly censoring the annual maxima (yellow and red colors show decreasing performance, 80-90% and <90% respectively, while green colors represents a percentage ≥90%). Note that the color of the dots in panel b) would be green for all the stations, when annual maxima are included in the parameter estimation.

Figure 2 shows the distribution of the ordinary events in a Weibull plot for four example stations characterized by contrasting topographic and precipitation features (in eastern Germany, Northern German Plains, Central Uplands and in the Bavarian Alps). Note that in these coordinates Weibull distributions become linear. We can clearly see how the lower portion of the ordinary events detaches more or less markedly from the linear behavior which seems to characterize the high-intensity portion (the dashed lines, whose color matches the stations location in the map in the inset, show the Weibull distributions describing the largest 10% of the events, as done in this study), supporting the findings by Wang et al. (2020) in sub-hourly rainfall in Germany. In all but the station in the East (purple color), highest values tend to align with the regression line, while in the latter case the less frequent rainfall depths tends to stay over it. In such situations, the slope of the regression line, whose inverse is the estimated Weibull shape parameter, should be higher (or, in other words, the distribution should be characterized by a heavier tail).

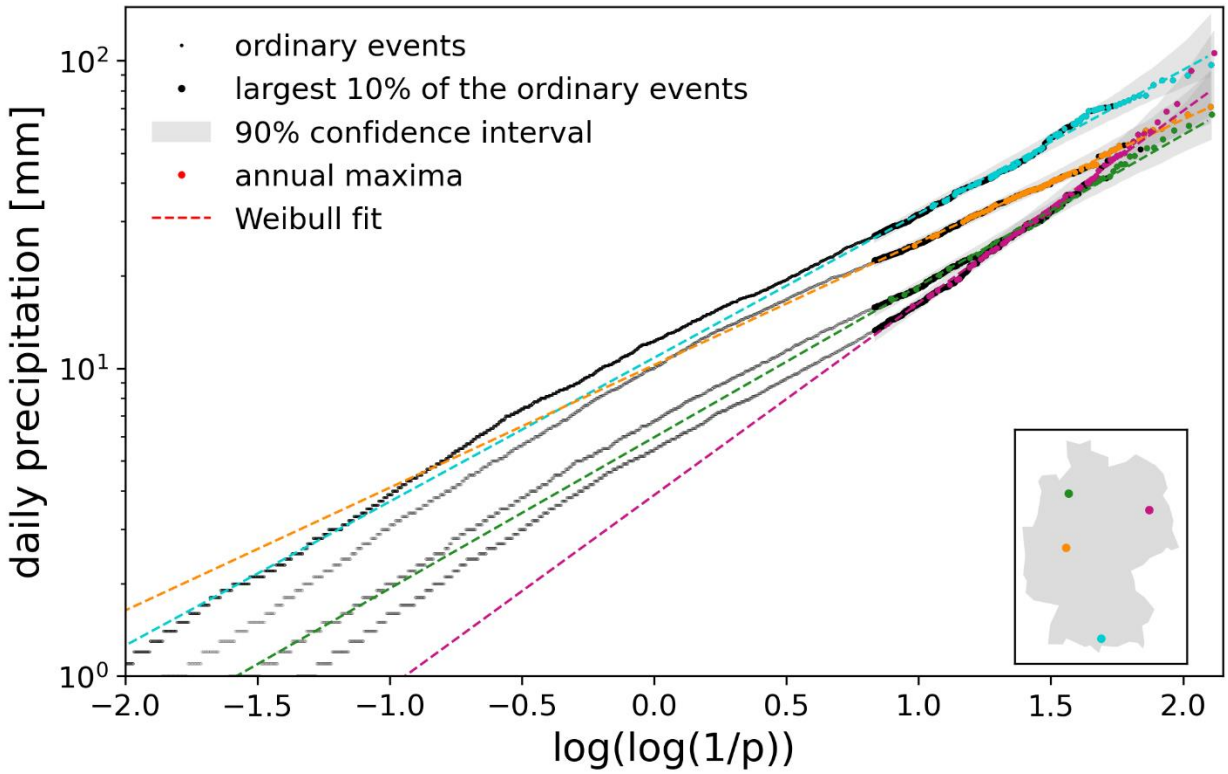

Figure 2 - Weibull plot (on the x axis p is the exceedance probability) of the ordinary events (small black dots) for four example stations characterized by contrasting topographic and precipitation features: the Berlin-Dahlem (Berlin, ID: 00403) station is located in Eastern Germany (purple), the Delmenhorst station (Lower Saxony, ID: 00936) is located in the Northern German Plains (green), the Berleburg, Bad-Wunderthausen station (North Rhine Westphalia, ID: 00392) is located in the Central Uplands (orange), and the Aitrang station (Bavaria, ID: 00063) is located in the Bavarian Alps (turquoise). Thicker black dots represent the largest 10% of the data used for parameter estimation, colored dots indicate annual maxima, dashed lines represent the Weibull distributions describing the 10% right tail, and gray shaded areas display the 90% sampling uncertainty (1000 random samples) from the Weibull distribution. Note that in the legend the colored thick dot and dashed lines are red to indicate a general color.

With the aim of quantifying the error emerging from a wrong estimate of the Weibull shape parameter, in Figure S2 we compare the rainfall frequency curves (i.e., daily rainfall maxima as a function of the logarithm of the return period, black filled circles) for the four exemplary cases in Figure 2, obtained by SMEV without left censoring (red shaded area), the standard MEV (green shaded area) and the SMEV with left censoring that we found to be the best choice for German daily precipitation (blue shaded area). The most relevant case is surely the one presented in panel a of Figure S2, namely the daily precipitation maxima for the Berlin-Dahlem (Berlin, ID: 00403) station in eastern Germany, which we noted as an area where heavy tails emerge and the risk of underestimating extremes is therefore particularly critical. In general, Figure S2 allows for a straightforward evaluation of the error that can be caused if wrong assumptions (in this case, the whole distribution of ordinary events following the Weibull model) are taken.

We now extend the analysis of the Weibull parameters over the whole country. The spatial patterns that arise when estimating the parameters of the Weibull distribution visibly resemble the German landscape characteristics (Figure 3, compared to Figure 1a). Higher values of the scale parameter *C* along with higher values of the shape parameter *w* can be noticed in correspondence of higher elevated areas: in the central-west part of the country (Central Uplands), in the centre (Harz), centre-east (Thuringian Forest), and in the south-west (low elevation mountains as the Black Forest are impacted by frequent precipitation). The southwestern Swabian Jura -lee side- is instead characterized by low precipitation rates (Sasse et al., 2013), and does not indeed display higher values of *C* or *w*. The same behavior is highly marked in the south (Bavarian Alps) and south-east (in correspondence of the Bavarian Forest). This implies that these areas are generally characterized by larger typical wet days and lighter-than-exponential tails. Lower values of *C* and values of *w* significantly lower than 1, indicating the emergence of heavier-tails, characterize eastern Germany, coherently with studies focusing on the evaluation of heavy-tails behaviors, e.g. the recent global analysis by means of the mean excess function by Nerantzaki and Papalexiou (2019). As mentioned in Section 2.6, we employ the obesity index (Cooke and Nieboer, 2011) for assessing tail heaviness, which confirms the pattern shown in Nerantzaki and Papalexiou (2019) (Figure 4). Red dots in Figure 4 indicate stations for which a heavy tail behavior is expected (an obesity index equal to 0.75, white dots in Figure 4, indicates the tail of an exponential, while an obesity index greater than 0.75 represents tails that are decreasing slower than the exponential ones (Cooke and Nieboer, 2014)), which are mainly concentrated in eastern Germany. This area is characterized by lower precipitation rates and magnitudes in comparison to other German regions, but precipitation triggered by "Vb" cyclones contributes to events with large magnitudes, which tend to make the tail of the distribution of daily rainfall become heavier. Figure 3b and Figure 4 both highlight an extensive presence of heavy tails in Germany (red dots in the two figures indicate shape parameter values lower than 1 and obesity index values greater than 0.75, respectively), therefore distributions with exponential (or lighter) tails, such as the Gumbel distribution (Papalexiou and Koutsoyiannis, 2013), should be used carefully. Also SMEV, despite being able to better describe the tail of the distribution when heavy tails emerge, generally underestimates the largest extremes. The behavior highlighted in the example station for eastern Germany in Figure 2 (purple) is common in stations whose distribution of daily precipitation values has a heavy tail. Among the stations with an obesity index greater than 0.75 (2642, red points in Figure 4), in 544 the highest maxima are underestimated, while in only 45 cases the tail of the distributions was overestimated. Regarding the parameter $\bar{n}$ (Figure 3c), it is important to remind that it represents the average yearly number of storms (i.e., consecutive wet days, as defined above). In areas characterized by longer storms, for example where the elevation is higher with an average storm length of about 4 days, this number is thus considerably lower than the typical number of wet days which can exceed 150 wet days per year. On the other hand, regions characterized by shorter storms are

likely to display higher values of $\bar{n}$ (notably, eastern Germany). Interestingly, elevation explains about 43% of the variance in the scale parameter (positive relation), 11% for the shape (positive), and 8% for the average number of yearly ordinary events (negative; insets in Fig. 3).

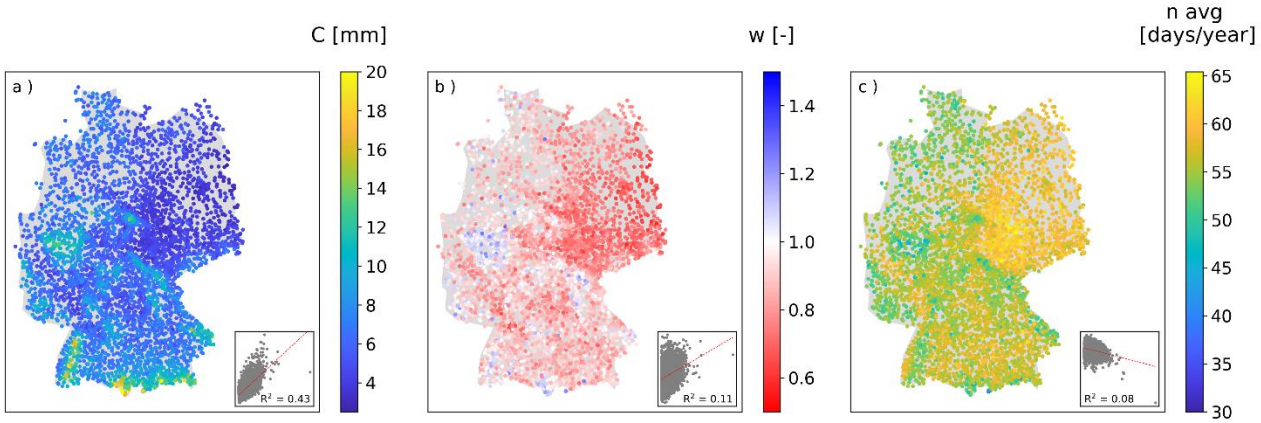

Figure 3 - Maps showing the estimated SMEV parameters: scale (panel a) and shape parameter (panel b) of the Weibull distribution and average yearly number of ordinary events (panel c). The insets represent the value of the respective parameter plotted against elevation; red dashed lines show the regression line fitted on these points (R^2 values for the three parameters are 0.43, 0.11 and 0.08 for C, w and $\bar{n}$ respectively).

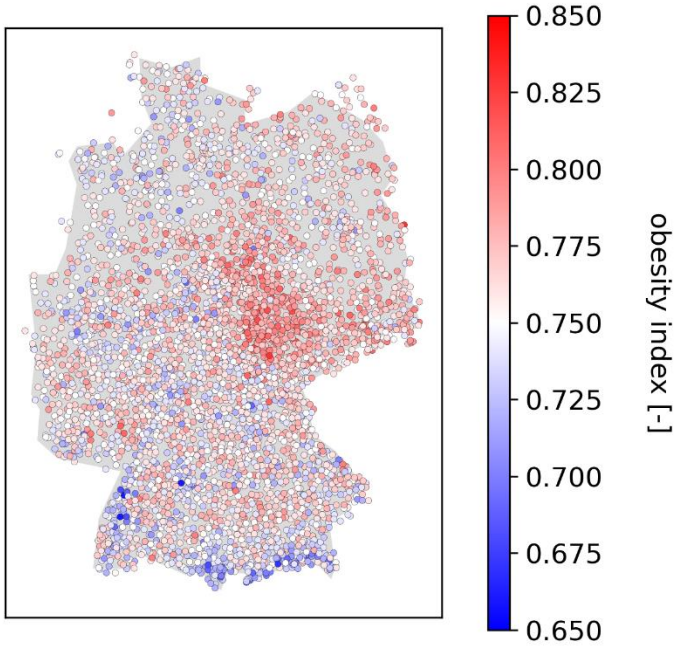

Figure 4 - Obesity index (Cooke and Nieboer, 2011) computed for the series of ordinary values of the 4015 stations analyzed. Blue (<0.75), white (=0.75) and red (>0.75) indicate light, exponential and heavy tails respectively.

To quantitatively evaluate SMEV estimates, we compared them with the design precipitation gridded values provided by the DWD Climate Data Center (CDC), which is in charge of providing official design values

for Germany. The grid cell size is approximately 8km, the temporal coverage from 1951 and 2010 (see details in Malitz and Ertel, 2015) and the tolerance range is $\pm$15% for return periods between 5 and 50 years and $\pm$20% for return periods greater than 50 years (see panels a and d in Figure 5 for the 50 and 100 years return values of design precipitation estimates). The guidelines given by the German Association for Water, Wastewater and Waste (DWA, 2012) for the estimation of return levels of daily precipitation consist in fitting an exponential distribution on the partial duration series including the heaviest rainfall events, whose length is at least twice the length of the available time series. Thus, the derived extreme value distribution is characterized by an exponential tail.

We estimated precipitation from SMEV for a return period of 50 and 100 years at each station for which at least 10 complete years in the 1951-2010 range (1046 stations) were available, and linearly interpolated them to match the grid provided by DWD-CDC (Figure 5b and e). Panels c and f in Figure 5 show the relative error between SMEV gridded values and the grid from DWD-CDC. Error values between $\pm$0.15 and $\pm$0.20 for 50 and 100 years return periods (Figure 5c and f) indicate that SMEV estimates are within the uncertainty range provided. SMEV tends to overestimate the 50 years design precipitation in some high-elevated areas (mainly in the Alps, slightly in the Black Forest and in the Bavarian Forest, more saturated red values in Figure 5c), but otherwise its estimations are between the lower and upper limits from DWD-CDC. Things change when focusing on the 100 return period quantiles. In addition to the Alpine region, SMEV overestimates systematically (see the marked red pattern in eastern Germany, Figure 5f) with respect to DWD-CDC, especially where the emergence of heavy tails was indicated by the heavy-tail indicator and by the Weibull shape parameter (the Spearman correlation coefficient computed between the relative error and the shape parameter is -0.4, hinting to the fact that the higher differences between DWD-CDC and SMEV-based estimated are at least partially explained by the presence of heavier tails). This relative over-estimation from SMEV is likely due to the approach used in the official German product that, being characterized by an exponential tail, might be locally too light.

Remarkably, SMEV estimates accurately resemble the orographic patterns, highlighting the fact that the model is be able to capture small-scale variations, (e.g., orographic effects; Marra et al., 2021), a feature of primary importance when estimating precipitation extremes.

So far, we found that the main drawback with the use of SMEV in Germany is a tendency towards underestimating the distribution tail heaviness. At the same time, we show that the design values currently in use in Germany seem to be even more affected by this issue as they rely on lighter-tailed distributions. The analyses presented up to now thus indicate SMEV as a more than adequate model for extreme daily precipitation in Germany. In the following, we evaluate the effect of the reduction of station density on high return period quantiles in ungauged locations.

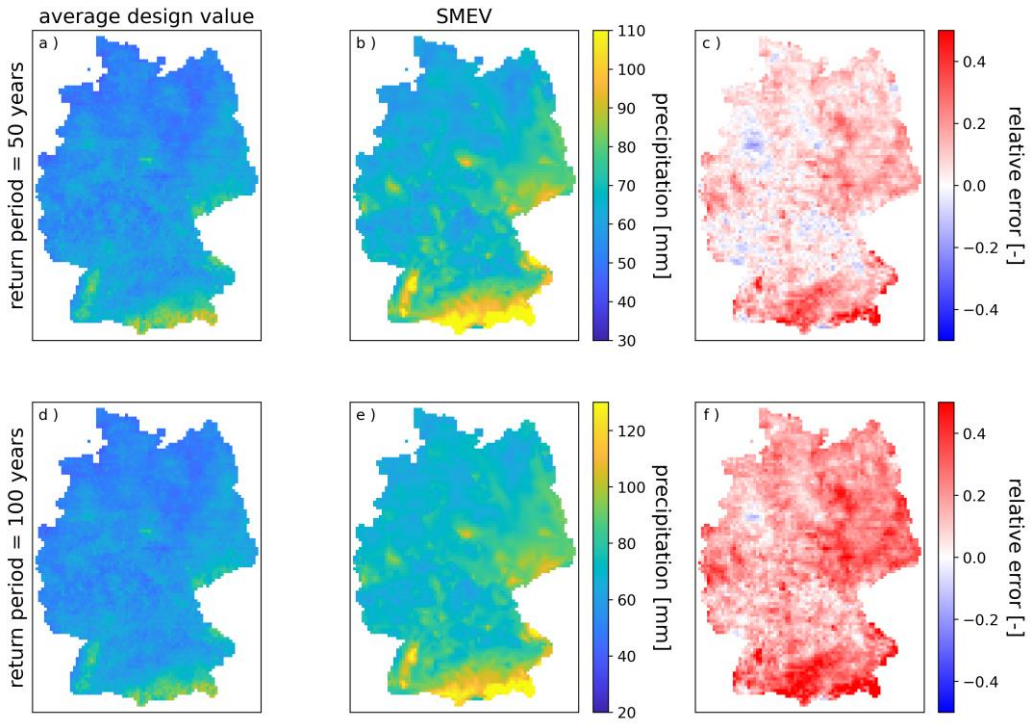

Figure 5 - Comparison between 50 (a) and 100 (d) years return period design values over Germany provided by the official source (DWD Climate Data Center (CDC): Grids of return periods of heavy precipitation (design precipitation) over Germany (KOSTRA-DWD), version 2010) and obtained with a linear interpolation of SMEV quantiles on the same grid (panels b and e). Panel c (f) shows the relative error between the SMEV and DWD-CDC gridded estimates for a return period of 50 (100) years. The spatial resolution is 8.15km×8.20km.

3.2 SMEV for ungauged locations

We now proceed with the SMEV estimates for ungauged locations, implemented as described in Section 2.4. We will focus on errors on parameters and quantiles associated to target return periods when stations are considered as ungauged. We present results for three representative densities, 1/500 km^{-2}, 1/1000 km^{-2} and 1/10000 km^{-2}, to evaluate to what extent the performance is degrading when the network becomes significantly sparser, which is an information of primary importance for real-world applications. In these cases, the number of stations used for estimating values in the "ungauged" station is respectively 714, 357 and 35.

Errors are computed as explained in Eq.(4) for what concerns parameters and following Eq.(5) to quantify the error on the estimation of high return period quantiles. We present here results for T_r = 100 years, but errors on other return periods show the same overall patterns and can be found in the supporting information (see Figures S3 and S4 for T_r = 50 and T_r = 500 years respectively).

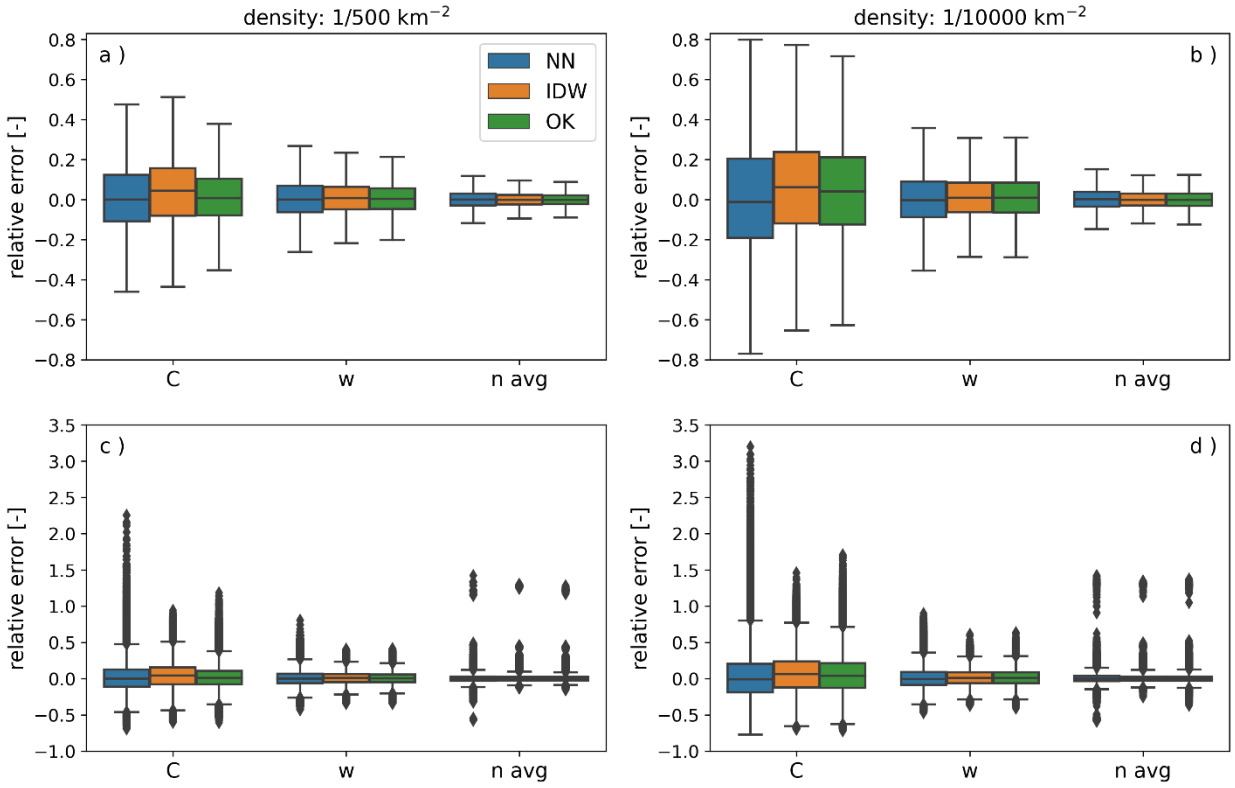

Figure 6 - Relative error on the parameters of SMEV (C, w and average n) for the three interpolation methods evaluated (nearest neighbor in navy blue, inverse distance weighting in orange and ordinary kriging in green). The two rows differ for the presence of outliers and the columns for the station densities. Namely, panels a (c) and b (d) show the boxplots of the relative errors without (with) outliers and for a density of 500 1/km^{-2} and 10000 1/km^{-2} respectively.

Boxplots in Figure 6 indicate the number of ordinary events as the parameter more accurately estimated by all three interpolation methods; its errors tend to remain fairly limited also when the density is significantly reduced (most of the density of the relative error distribution on the number of rainy days lies in the range -0.2, 0.2, as it can be seen in the third group of box plots in the first row of Figure 6). This indicates that, as expected, local conditions tend to affect precipitation intensity distributions more than storms occurrence. The distribution of the relative error on the shape parameter *w* shows fewer outliers than the one on $\bar{n}$, but it is more likely to have higher errors (range -0.4, 0.4, second group of boxplots in the first row of Figure 6). Moreover, differences between the three methods can be highlighted: NN displays a larger range of outliers with respect to the other two, which are instead comparable. The largest errors occur for the scale parameter *C* (see first group of boxplots in Figure 6), with their uncertainty being minimized by the inverse distance weighting method.

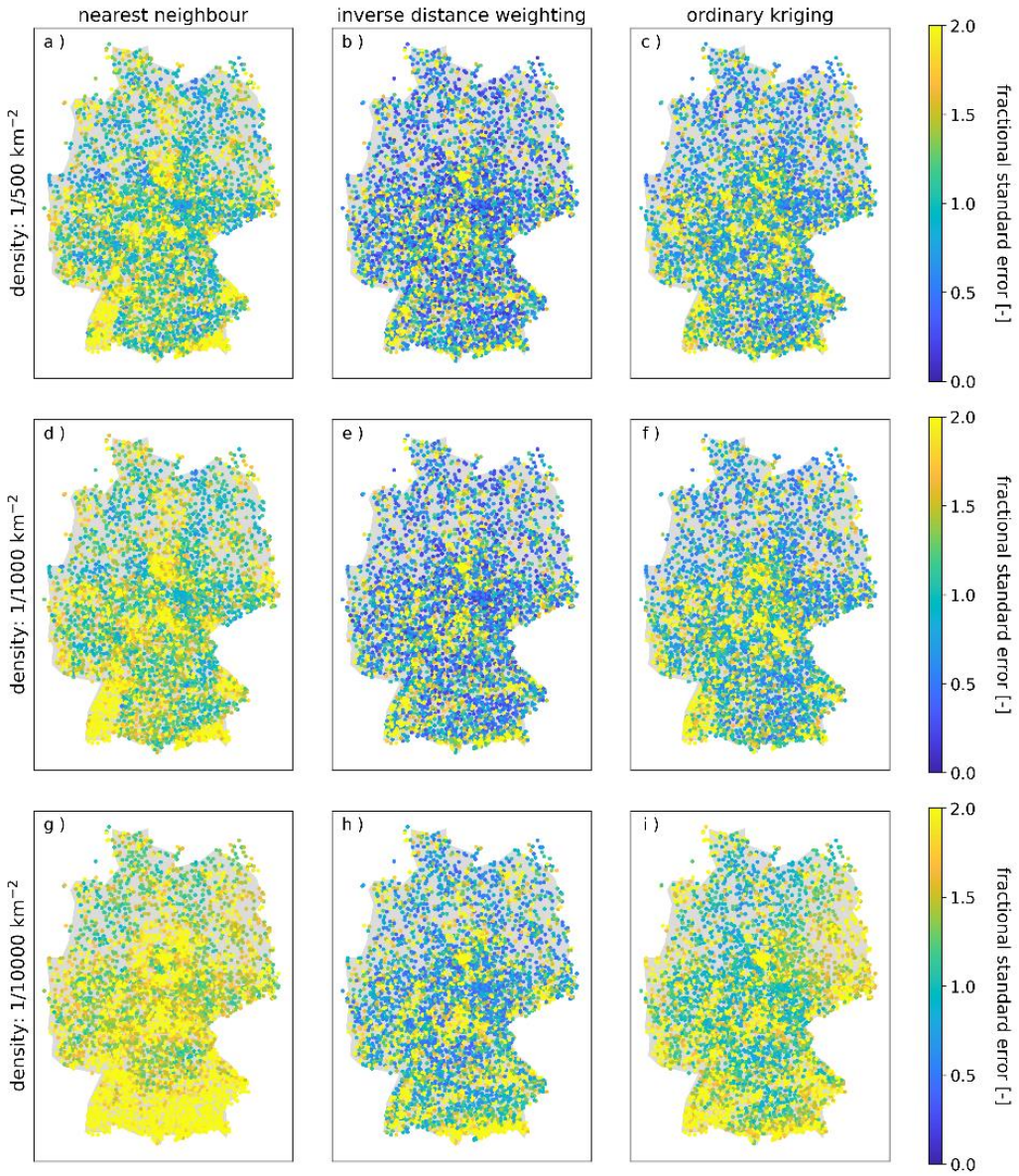

Figure 7 - Map of the fractional standard error (FSE) computed for the return period 100 years from three different stations density and different interpolation methods. Panels a)-b)-c) show the FSE for a station density of 1/500 km^{-2} and nearest neighbour, inverse distance weighting and ordinary kriging interpolation methods; panels d)-e)-f) show the FSE for a station density of 1/1000 km^{-2} and nearest neighbour, inverse distance weighting and ordinary kriging interpolation methods; panels g)-h)-i) show the FSE for a station density of 1/10000 km^{-2} and nearest neighbour, inverse distance weighting and ordinary kriging interpolation methods.

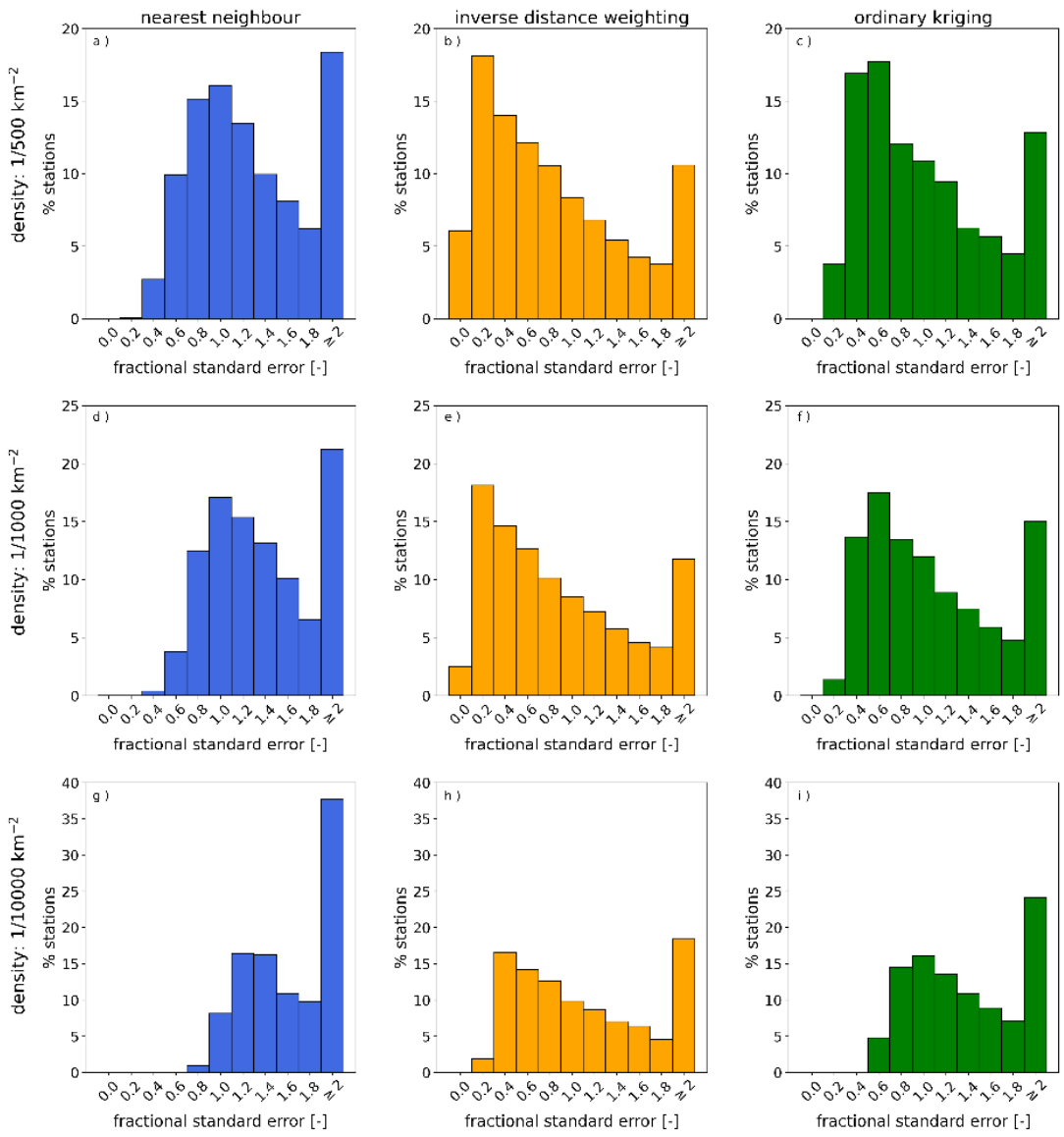

Figure 8 – Distributions of the fractional standard error (FSE) for the return period 100 years computed for three different station densities for different interpolation methods (nearest neighbour in blue, inverse distance weighting in orange and ordinary kriging in green). Panels a)-b)-c) show the distribution of the FSE for a station density of 1/500 km^{-2} and nearest neighbour, inverse distance weighting and ordinary kriging interpolation methods; panels d)-e)-f) show the distribution of the FSE for a station density of 1/1000 km^{-2} and nearest neighbour, inverse distance weighting and ordinary kriging interpolation methods; panels g)-h)-i) show the distribution of the FSE for a station density of 1/10000 km^{-2} and nearest neighbour, inverse distance weighting and ordinary kriging interpolation methods. Tick labels on the x-axis indicate the lower limit (included) of the interval considered.

Errors in the parameters affect the estimation accuracy: in Figure 7 we present the map of the fractional standard error (FSE) computed for each station on the 100 random extraction of the gauges used for parameter interpolation, for all the interpolation methods, a return period of 100 years and three different densities (1/500, 1/1000 and 1/10000 km^{-2}). Figure 7 is complemented by Figure 8, which shows the

447 histograms of the frequency (in %) of stations displaying a FSE within specific ranges. Remarkably, the
448 interpolation method displaying the lowest FSE is IDW (panels b-e-h in Figure 7 and Figure 8), also when
449 the station density is reduced (almost 50% of the stations shows FSE lower than 1 when the density is
450 1/10000 km^{-2}, panel h in Figure 8). The FSE computed on 100 years return values estimated via NN
451 manifests the highest values, also for the higher station densities (panels a and d in Figures 7 and 8, where
452 about 80% of the stations have FSE greater than 1, and panel g, with FSE always greater than 1). Finally,
453 the FSE from the OK is in between those from the other two interpolation methods (c-f-i and corresponding
454 panels in Figure 8). The determination coefficients of the regression line between FSE and elevation are
455 low (R^2 less than 0.1) for all the cases, implying that little of the variability of the FSE can be explained by
456 the variability of the elevation.

457 The results described above indicate the IDW as the most stable interpolation method, at least for our study
458 case. Even if geostatistical methods were often been proved to be superior to IDW, depending on the
459 application of interest, other studies found similar performances between IDW and OK, mainly in
460 dependence to the network density and the geographic region. For example, Wagner et al. (2012)
461 highlighted comparable performances between IDW and OK methods, and in Ly et al. (2011) neither OK
462 nor kriging with external drift were able to improve the IDW interpolation accuracy.

463 As indicated in Section 2.7, as a last step we benchmark our approach with previous research (Schellander
464 et al., 2019), namely by comparing the relative error between the interpolated and the observed 50-year
465 quantile as in figure 9 in Schellander et al. (2019). We consider as target stations those with a time series at
466 least 50 (100) years long, whose spatial distribution is shown in the first (second) row in Figure 9. The
467 stations selected for this analysis are 1515 and 102 respectively, and we consider only our higher station
468 densities (1/200 km^{-2} and 1/500 km^{-2}), as they are the closest, albeit much sparser, to the 1/67 km^{-2} density
469 of the calibration stations in Schellander et al. (2019). Notably, the most significant underestimation issues
470 appear in eastern Germany, where tails tend to be heavier than those of our model. This is true despite the
471 fact that our estimates are higher than the operational design values for Germany. In terms of comparison
472 with Schellander et al. (2019), the range of the median relative error that we get is (-0.6,0.4) when
473 considering the stations with the shorter records and (-0.4, 0.2) in the case of the longer, in contrast with
474 their range which reaches 0.95; the fraction of stations with small relative error (range -0.05, 0.05) is 24%
475 and 33% in our cases, as opposed to their 21%. These represent solid results, also in consideration of the
476 denser calibration network used in Schellander et al. (2019).

477

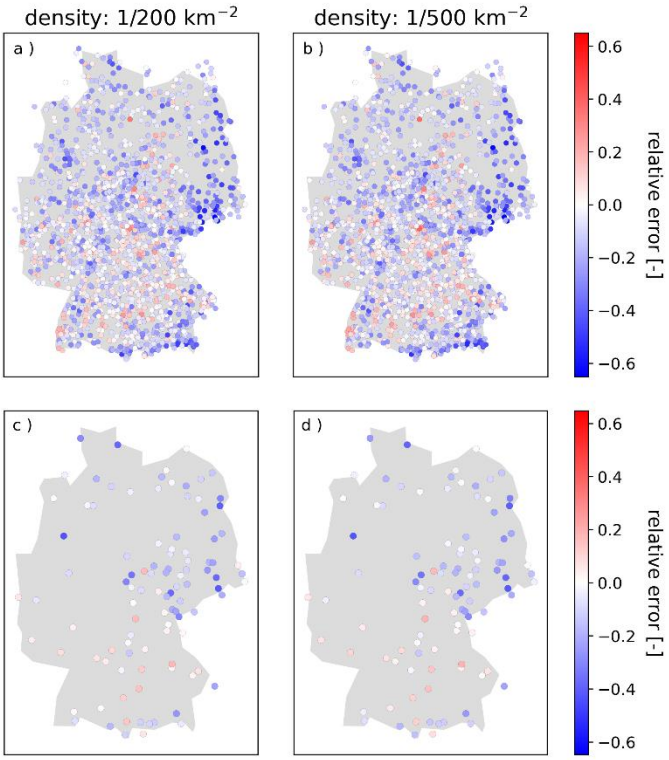

Figure 8 - Relative error obtained by the IDW method computed on the empirical 50 years return value (for the series in which the exact empirical 50 years return period is not available, the closest one is considered) on the stations for which the time series available is at least 50 (100) years long in the first (second) row. Panels a (b) show the relative error from the IDW method for a time series at least 50 years long and a station density of 1/200 km^{-2} (1/500 km^{-2}), while panels c (d) show the relative error from the IDW method for a time series at least 100 years long and a station density of 1/200 km^{-2} (1/500 km^{-2}). Blue (red) dots indicate under- (over-) estimation of the empirical quantile.

4. Summary and Conclusions

We presented an at-site and spatial analysis of extreme daily precipitation return levels over Germany by means of the simplified Metatastatistical Extreme Value (SMEV) approach, an emergent methodology for the estimation of extremes from ordinary events. Here we leverage the enhanced flexibility of SMEV in the definition of the ordinary events to (i) avoid time-dependence issues and (ii) properly model ordinary events by left-censoring values which are not described by the same two-parameter distribution describing extremes. We found that SMEV with left-censoring 90% of the values is a robust extreme value method for Germany. Strict tests evaluating whether observed annual maxima are likely sampled from this model showed that this is the case for more than 80% of the examined stations, even when annual maxima are explicitly censored from the parameter estimation procedure. In general, our approach tends to underestimate the tail heaviness in some areas, but remarkably showed higher estimates of 50-year and 100-year quantiles than the German official design precipitation values (DWD Climate Data Center (CDC)),

meaning that these values could be subject to even larger underestimation. The emergence of heavy tails in several regions of the country, mainly (but not limited to) the eastern part, hints to the fact that the use of the approach recommended by the official German guidelines, whose tail is equivalent to exponential, should be revised in some areas. Notably, the regions in which the SMEV (and DWD) underestimation is more marked are also characterized by a peculiar climatology, in which specific relatively rare events (the "Vb" cyclones) contribute to an important portion of the extremes. The presence of multiple types of ordinary events characterized by different distributions could easily explain the underestimation for these models based on one type of ordinary events only (e.g., Berg et al (2013); Marra et al. (2018); Marra et al. (2019)).

Interestingly, at-site Weibull parameters show spatial patterns which closely resemble the German orography. Notwithstanding the orographic enhancement of precipitation observed in many mountainous regions being a widely studied process, the orographic impact on the statistical properties of the distributions used to describe extremes has been studied less (e.g., Grieser et al. (2007); Allamano et al. (2009); Blanchet et al. (2009); Avanzi et al. (2015); Ragulina and Reitan (2017)). Here, we highlight a positive correlation between the scale and, to a lesser extent, the shape parameters of the Weibull distribution with the elevation, which aligns with recent results in 6-hour precipitation intensities on different climates (Marra et al., 2021).

In the spatial approach, the inverse distance weighting method is the most robust, showing lower FSE values than those obtained with the other two methods. Only when the density decreases significantly (1/10000 km^{-2}, i.e., the parameter at each station in the country are estimated using only 35 other stations), high-elevated stations are affected by the highest FSE, for example in correspondence of the Bavarian Alps and the Harz mountains. Comparison of our results with the pioneer investigations on the spatial use of SMEV by Schellander et al. (2019) showed that, despite the availability of a sparser calibration network, our methodology showed a smaller range of the errors as well as a larger fraction of stations with very-small error (around 0).

These promising results indicate that the approach here proposed is both flexible and robust. It might hence be used for estimation of extremes in ungauged locations. Albeit a small fraction of the variance in estimation errors is explained by elevation, the correlation between SMEV parameters and orography (up to 43% explained variance) suggests that future applications should test the inclusion of such information in spatial estimates. Similarly, future applications should consider the use of multiple types of ordinary events to explain the underestimated extreme quantiles in regions where rare-type events dominate the distribution of extremes.

Data statement and acknowledgements

Daily rainfall time series are available at: https://opendata.dwd.de/climate_environment/CDC/observations_germany/climate/daily/more_precip/historical/

A.M. acknowledges support from the German Research Foundation (Deutsche Forschungsgemeinschaft, DFG) through the project 421396820 Propensity of rivers to extreme floods: climate-landscape controls and early detection (PREDICTED). F.M. thanks the Institute of Atmospheric Sciences and Climate of the National Research Council of Italy, and Vincenzo Levizzani in particular, for the support. We would like to thank Larisa Tarasova for the valuable suggestions. The scientific results have been in part computed at the High-Performance Computing (HPC) Cluster EVE (Schnicke et al., 2019), a joint effort of both the Helmholtz Centre for Environmental Research - UFZ (http://www.ufz.de/) and the German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig (http://www.idiv-biodiversity.de/).

References

Allamano, P., Claps, P., Laio, F., & Thea, C. (2009). A data-based assessment of the dependence of short-duration precipitation on elevation. *Physics and Chemistry of the Earth*, *34*(10–12), 635–641. https://doi.org/10.1016/j.pce.2009.01.001

Avanzi, F., De Michele, C., Gabriele, S., Ghezzi, A., & Rosso, R. (2015). Orographic signature on extreme precipitation of short durations. *Journal of Hydrometeorology*, *16*(1), 278–294. https://doi.org/10.1175/JHM-D-14-0063.1

Berg, P., Moseley, C. & Haerter, J. O. (2013). Strong increase in convective precipitation in response to higher temperatures. *Nature geoscience, 6*, 181-185. https://doi.org/10.1038/NGEO1731

Blanchet, J., & Lehning, M. (2010). Mapping snow depth return levels: Smooth spatial modeling versus station interpolation. *Hydrology and Earth System Sciences*, *14*(12), 2527–2544. https://doi.org/10.5194/hess-14-2527-2010

Blanchet, J., Marty, C., & Lehning, M. (2009). Extreme value statistics of snowfall in the Swiss Alpine region. *Water Resources Research*, *45*(5), 1–12. https://doi.org/10.1029/2009WR007916

Blöschl, G., Bierkens, M. F. P., Chambel, A., Cudennec, C., Destouni, G., Fiori, A., Kirchner, J. W., McDonnell, J. J., Savenije, H. H. G., Sivapalan, M., Stumpp, C., Toth, E., Volpi, E., Carr, G.,

559 Lupton, C., Salinas, J., Széles, B., Viglione, A., Aksoy, H., … Zhang, Y. (2019). Twenty-three
560 unsolved problems in hydrology (UPH)–a community perspective. *Hydrological Sciences Journal*,
561 *64*(10), 1141–1158. https://doi.org/10.1080/02626667.2019.1620507

562 CDC, D. C. D. C. (2010). *Grids of return periods of heavy precipitation (design precipitation)*
563 *overGermany (KOSTRA-DWD), version 2010R.*

564 Ceresetti, D., Ursu, E., Carreau, J., Anquetin, S., Creutin, J. D., Gardes, L., Girard, S., & Molinié, G.
565 (2012). Evaluation of classical spatial-analysis schemes of extreme rainfall. *Natural Hazards and*
566 *Earth System Science*, *12*(11), 3229–3240. https://doi.org/10.5194/nhess-12-3229-2012

567 Coles, S. (2001). *An Introduction to Statistical Modeling of Extreme Values* (S. Series in Statistics (ed.)).

568 Cooke, R. M., & Nieboer, D. (2011). Heavy-tailed distributions: data, diagnostics, and new
569 developments. *SSRN Electronic Journal*, *March*. https://doi.org/10.2139/ssrn.1811043

570 Cooke, R. M. & Nieboer, D. (2014). Fat-tailed distributions: data, diagnostics and dependence. Vol. 1.
571 Hoboken, NJ: John Wiley & Sons.

572 Das, S. (2019). Extreme rainfall estimation at ungauged sites: Comparison between region-of-influence
573 approach of regional analysis and spatial interpolation technique. *International Journal of*
574 *Climatology*, *39*(1), 407–423. https://doi.org/10.1002/joc.5819

575 Deutscher Wetterdienst. (2017). *National Climate Report. Climate - yesterday, today and in future*.
576 https://www.dwd.de/EN/ourservices/nationalclimatereport/download_report_edition-
577 3.pdf?__blob=publicationFile&v=4

578 DWA, 2012: Starkregen in Abhängigkeit von Wiederkehrzeit und Dauer (engl. title: Extreme
579 precipitation in dependence of return period and duration). – Deutsche Vereinigung für
580 Wasserwirtschaft, Abwasser und Abfall e.V., Hennef, Germany

581 Fisher, R. A., & Tippett, L. H. C. (1928). Limiting forms of the frequency distribution of the largest or
582 smallest member of a sample. *Mathematical Proceedings of the Cambridge Philosophical Society*,
583 *24*(2), 180–190. https://doi.org/10.1017/S0305004100015681

584 Gnedenko, B. (1943). Sur La Distribution Limite Du Terme Maximum D'Une Serie Aleatoire. *The*
585 *Annals of Mathematics*, *44*(3), 423. https://doi.org/10.2307/1968974

586 Grams, C. M., Binder, H., Pfahl, S., Piaget, N., & Wernli, H. (2014). Atmospheric processes triggering
587 the central European floods in June 2013. *Natural Hazards and Earth System Sciences*, *14*(7), 1691–

588 1702. https://doi.org/10.5194/nhess-14-1691-2014

589 Grieser, J., Staeger, T., & Schonwiese, C. D. (2007). Estimates and uncertainties of return periods of
590 extreme daily precipitation in Germany. *Meteorologische Zeitschrift*, *16*(5), 553–564.
591 https://doi.org/10.1127/0941-2948/2007/0235

592 Groenemeijer, P., Becker, N., Djidara, M., Gavin, K., Hellenberg, T., Holzer, A. M., Juga, I., Jokinen, P.,
593 Jylhä, K., Lehtonen, I., Mäkelä, H., Napoles, O. M., Nissen, K., Paprotny, D., Prak, P., Púčik, T.,
594 Tijssen, L., & Vajda, A. (2015). *Past cases of extreme weather impact on critical infrastructure in*
595 *Europe*. http://rain-project.eu/wp-content/uploads/2015/11/D2.2-Past-Cases-final.compressed.pdf
596 (accessed 16 March, 2020)

597 Hailegeorgis, T. T., & Alfredsen, K. (2017). Analyses of extreme precipitation and runoff events
598 including uncertainties and reliability in design and management of urban water infrastructure.
599 *Journal of Hydrology*, *544*, 290–305. https://doi.org/10.1016/j.jhydrol.2016.11.037

600 Hosking, J. R. M., & Wallis, J. R. (1997). *Regional frequency analysis: an approach based on L-*
601 *moments*. Cambridge University Press.

602 Jenkinson, A. F. (1955). The frequency distribution of the annual maximum (or minimum) values of
603 meteorological elements. *Q. J. R. Meteorol. Soc*, *81*. https://doi.org/10.1002/qj.49708134804

604 Katz, R. W., Parlange, M. B., & Naveau, P. (2002). Statistics of extremes in climatology and hydrology.
605 *Advances in Water Resources*, *25*, 1287–1304.

606 Kidd, C., Becker, A., Huffman, G. J., Muller, C. L., Joe, P., Skofronick-Jackson, G., & Kirschbaum, D.
607 B. (2017). So, how much of the Earth's surface is covered by rain gauges? *Bulletin of the American*
608 *Meteorological Society*, *98*(1), 69–78. https://doi.org/10.1175/BAMS-D-14-00283.1

609 Lu, L. H., & Stedinger, J. R. (1992). Variance of two- and three parameter GEV/PWM quantile
610 estimators: formulas, confidence intervals, and a comparison. *Journal of Hydrology*, *138*, 247–267.

611 Ly, S., Charles, C., & Degré, A. (2011). Geostatistical interpolation of daily rainfall at catchment scale:
612 The use of several variogram models in the Ourthe and Ambleve catchments, Belgium. *Hydrology*
613 *and Earth System Sciences*, *15*(7), 2259–2274. https://doi.org/10.5194/hess-15-2259-2011

614 Malitz, G., & Ertel, H. (2015). KOSTRA-DWD-2010 - Starkniederschlagshöhen für Deutschland
615 (Bezugszeitraum 1951 bis 2010). Abschlussbericht. *Deutscher Wetterdienst*.
616 https://www.dwd.de/DE/leistungen/kostra_dwd_rasterwerte/download/bericht_kostra_dwd_2010_p
617 df.pdf?__blob=publicationFile&v=9

Marani, M., & Ignaccolo, M. (2015). A metastatistical approach to rainfall extremes. *Adv. Wat. Res.*, *79*, 121–126. https://doi.org/10.1016/j.advwatres.2015.03.001

Marra, F., Armon, M., Borga, M., & Morin, E. (2021). Orographic effect on extreme precipitation statistics peaks at hourly time scales. *Geophysical Research Letters*. https://doi.org/10.1029/2020gl091498

Marra, F., Borga, M., & Morin, E. (2020). A Unified Framework for Extreme Subdaily Precipitation Frequency Analyses Based on Ordinary Events. *Geophysical Research Letters*, *47*(18). https://doi.org/10.1029/2020GL090209

Marra, F., Nikolopoulos, E. I., Anagnostou, E. N., Bárdossy, A., & Morin, E. (2019). Precipitation frequency analysis from remotely sensed datasets: A focused review. *Journal of Hydrology*, *574*, 699–705. https://doi.org/10.1016/j.jhydrol.2019.04.081

Marra, F., Nikolopoulos, E. I., Anagnostou, E. N., & Morin, E. (2018). Metastatistical Extreme Value analysis of hourly rainfall from short records: Estimation of high quantiles and impact of measurement errors. *Adv. Wat. Res.*, *117*, 27–39. https://doi.org/10.1016/j.advwatres.2018.05.001

Marra, F., Zoccatelli, D., Armon, M., & Morin, E. (2019). A simplified MEV formulation to model extremes emerging from multiple nonstationary underlying processes. *Adv. Wat. Res.*, *127*(March), 280–290. https://doi.org/10.1016/j.advwatres.2019.04.002

Martins, E. S., & Stedinger, J. R. (2000). Generalized maximum-likelihood generalized extreme-value quantile estimators for hydrologic data. *Water Resources Research*, *36*(3), 737–744. https://doi.org/10.1029/1999WR900330

Messmer, M., Gómez-Navarro, J. J., & Raible, C. C. (2015). Climatology of Vb cyclones, physical mechanisms and their impact on extreme precipitation over Central Europe. *Earth System Dynamics*, *6*(2), 541–553. https://doi.org/10.5194/esd-6-541-2015

Miniussi, A., & Marani, M. (2020). Estimation of Daily Rainfall Extremes Through the Metastatistical Extreme Value Distribution: Uncertainty Minimization and Implications for Trend Detection. *Water Resources Research*, *56*(7). https://doi.org/10.1029/2019WR026535

Miniussi, A., Villarini, G., & Marani, M. (2020). Analyses Through the Metastatistical Extreme Value Distribution Identify Contributions of Tropical Cyclones to Rainfall Extremes in the Eastern United States. *Geophysical Research Letters*, *47*(7). https://doi.org/10.1029/2020GL087238

Morrison, J. E., & Smith, J. A. (2002). Stochastic modeling of flood peaks using the generalized extreme

value distribution. *Water Resources Research*, *38*(12), 41-1-41–12. https://doi.org/10.1029/2001wr000502

Nerantzaki, S. D., & Papalexiou, S. M. (2019). Tails of extremes: Advancing a graphical method and harnessing big data to assess precipitation extremes. *Advances in Water Resources*, *134*. https://doi.org/10.1016/j.advwatres.2019.103448

Nikolopoulos, E. I., Borga, M., Creutin, J. D., & Marra, F. (2015). Estimation of debris flow triggering rainfall: Influence of rain gauge density and interpolation methods. *Geomorphology*, *243*, 40–50. https://doi.org/10.1016/j.geomorph.2015.04.028

Nissen, K. M., Ulbrich, U., & Leckebusch, G. C. (2013). Vb cyclones and associated rainfall extremes over central Europe under present day and climate change conditions. *Meteorologische Zeitschrift*, *22*(6), 649–660. https://doi.org/10.1127/0941-2948/2013/0514

Papalexiou, S. M., & Koutsoyiannis, D. (2013). Battle of extreme value distributions : A global survey on extreme daily rainfall. *Water Resources Research*, *49*, 187–201. https://doi.org/10.1029/2012WR012557

Ragulina, G., Reitan, T. (2017). Generalized extreme value shape parameter and its nature for extreme precipitation using long time series and the Bayesian approach. *Hydrological Sciences Journal*, *62*, 863-879. https://doi.org/10.1080/02626667.2016.1260134

Sasse, R., Schädler, G., & Kottmeier, C. (2013). The regional atmospheric water budget over Southwestern Germany under different synoptic conditions. *Journal of Hydrometeorology*, *14*(1), 69–84. https://doi.org/10.1175/JHM-D-11-0110.1

Schellander, H., Lieb, A., & Hell, T. (2019). Error structure of Metastatistical and Generalized Extreme Value Distributions for modeling extreme rainfall in Austria. *Earth and Space Science*, 2019EA000557. https://doi.org/10.1029/2019EA000557

Schnicke, T., Langenberg, B., Schramm, G., Krause, C., & Strempel, T. (2019). *EVE - High-Performance Computing Cluster*. https://wiki.ufz.de/eve/

Serinaldi, F., Lombardo, F., & Kilsby, C. G. (2020). All in order: Distribution of serially correlated order statistics with applications to hydrological extremes. *Adv. Wat. Res.*, *144*. https://doi.org/10.1016/j.advwatres.2020.103686

Van Bebber, W.: Die Zugstrassen der barometrischen Minima nach den Bahnenkarten der deutschen Seewarte für den Zeitraum 1875–1890, *Meteorol. Z.*, 8, 361–366, 1891.

Wagner, P. D., Fiener, P., Wilken, F., Kumar, S., & Schneider, K. (2012). Comparison and evaluation of spatial interpolation schemes for daily rainfall in data scarce regions. *Journal of Hydrology*, *464–465*, 388–400. https://doi.org/10.1016/j.jhydrol.2012.07.026

Wallis, J. R., Schaefer, M. G., Barker, B. L., & Taylor, G. H. (2007). Regional precipitation-frequency analysis and spatial mapping for 24-hour and 2-hour durations for Washington State. *Hydrology and Earth System Sciences*, *11*(1). https://doi.org/10.5194/hess-11-415-2007

Wang, L., Marra, F., & Onof, C. (2020). Modelling sub-hourly rainfall extremes with short records – a comparison of MEV, Simplified MEV and point process methods. *European Geosci. Union (EGU) General Assembly 2020 (Online)*. https://presentations.copernicus.org/EGU2020/EGU2020-6061_presentation.pdf

Webster, R., & Oliver, M. A. (2007). *Geostatistics for Environmental Scientists*. John Wiley & Sons. https://doi.org/0.1002/9780470517277

Wilson, P. S., & Toumi, R. (2005). A fundamental probability distribution for heavy rainfall. *Geophysical Research Letters*, *32*(14), 1–4. https://doi.org/10.1029/2005GL022465

Zorzetto, E., Botter, G., & Marani, M. (2016). On the emergence of rainfall extremes from ordinary events. *Geophys. Res. Lett.*, *43*, 8076–8082. https://doi.org/10.1002/2016GL069445

Zorzetto and Marani (2019). Downscaling of rainfall extremes from satellite observations. *Water Resources Research*, *55*, 156–174. https://doi.org/10.1029/2018WR022950