

Evaluation of daily precipitation extremes in reanalysis and gridded observation - based datasets over Germany

Article

Published Version

Creative Commons: Attribution 4.0 (CC-BY)

Open Access

Hu, G. ORCID: <https://orcid.org/0000-0003-4305-3658> and Franzke, C. L. E. ORCID: <https://orcid.org/0000-0003-4111-1228> (2020) Evaluation of daily precipitation extremes in reanalysis and gridded observation based datasets over Germany. *Geophysical Research Letters*, 47 (18). e2020GL089624. ISSN 0094-8276 doi: <https://doi.org/10.1029/2020GL089624> Available at <http://centaur.reading.ac.uk/92583/>

It is advisable to refer to the publisher's version if you intend to cite from the work. See [Guidance on citing](#).

To link to this article DOI: <http://dx.doi.org/10.1029/2020GL089624>

Publisher: AGU

including copyright law. Copyright and IPR is retained by the creators or other copyright holders. Terms and conditions for use of this material are defined in the [End User Agreement](#).

www.reading.ac.uk/centaur

CentAUR

Central Archive at the University of Reading

Reading's research outputs online

Evaluation of daily precipitation extremes in reanalysis and gridded observation-based datasets over Germany

Guannan Hu ^{*1,2,3} and Christian L. E. Franzke^{2,3,4}

¹Department of Meteorology, School of Mathematical, Physical and Computational Sciences, University of Reading, UK

²Meteorological Institute, University of Hamburg, Germany

³Center for Earth System Research and Sustainability, University of Hamburg, Germany

⁴School of Engineering and Science, Jacobs University, Germany

Key Points

- Most examined gridded data products underestimate the magnitude of observed daily precipitation extremes.
- Gridded observation-based datasets generally perform better than reanalyses in representing daily precipitation extremes.
- Reanalyses often fail to capture the timing of observed daily precipitation extremes.

Abstract

Accurate and reliable gridded datasets are important for analyzing extreme weather and climate events. Specifically, these datasets should produce extreme value statistics that are close to reality. Here we use various statistical methods to evaluate the quality of four gridded data products in representing daily precipitation extremes. The data products are the COSMO-REA6 regional reanalysis, the ERA5 global reanalysis, and the E-OBS and HYRAS gridded observation-based datasets. The statistical methods we use offer a thorough insight into the quality of the different datasets by providing temporal and spatial extreme value statistics of daily precipitation. Our results show that all datasets except HYRAS underestimate the magnitude of daily precipitation extremes when compared with weather station data. Moreover, the reanalysis datasets give generally worse extreme value statistics of daily precipitation than the gridded observation-based datasets. In particular, the reanalysis datasets often fail in reproducing the accurate timing of observed daily precipitation extremes.

Plain Language Summary

Gridded data products provide long-term estimates of climate variables such as temperature and precipitation at regularly spaced grids on the Earth. They are an important source for the research of extreme weather. For example, for investigating the change in frequency and intensity of heavy precipitation over time. To achieve reliable results, we need such data products to be able to accurately represent extreme weather events. To verify if this is the case, we evaluate the quality of several gridded data products in representing heavy daily precipitation events and find that there is room for improvement.

*Correspondence email address: guannan.hu@reading.ac.uk

1 Introduction

Reanalysis and gridded observation-based datasets provide long-term estimates of climate variables on a grid covering the globe or a region. Reanalysis data are produced by blending a large number of observations and forecasts from a numerical weather prediction system (Bollmeyer et al., 2015; Hersbach et al., 2020), while the gridded observation-based datasets are created by interpolating weather station observations onto a regularly spaced grid (Cornes et al., 2018; Rauthe et al., 2013). These gridded data products have found wide application in atmospheric and climate sciences. For instance, they are used to identify climate variability and change, to evaluate climate model simulations, and to investigate extreme weather and climate events.

Extreme climate and weather events are of utmost importance to many aspects of ecosystems, the economy and society. They are not only of interest to the climate and weather communities, but are also a big concern for the financial sector and the insurance industry, since they can have severe impacts on economic damages (Houser et al., 2015; Franzke, 2017; Franzke and Czupryna, 2020). Because reanalysis and gridded observation-based datasets typically cover several decades or longer, they are able to provide a decent amount of data for us to analyze extreme weather and climate events, which are rare by definition. Considering the wide use of gridded data products for analyzing extreme weather and climate events (Blender et al., 2017; Garfinkel and Harnik, 2017; Sienz et al., 2010; Bach et al., 2016; Donat et al., 2016; Dulière et al., 2011), there is a need to carefully assess how well these products actually perform in representing these events. Temperature and precipitation are two meteorological variables that are highly relevant to our lives. Previous studies have shown that precipitation extremes in gridded data products are often less reliable, compared to temperature extremes (Kharin et al., 2005; Donat et al., 2014; Lockhoff et al., 2019; Mannshardt-Shamseldin et al., 2010; Zolina et al., 2004). Therefore, in this paper we evaluate four gridded data products in terms of their performance on representing daily precipitation extremes over Germany. The investigation of daily precipitation extremes is relevant for flood risk assessment. We take advantage of the high-density precipitation observations provided by the German Weather Service (Deutscher Wetterdienst; DWD) to validate the daily precipitation extremes represented in the gridded data products.

In many previous studies climate extreme indices are computed to investigate temperature and precipitation extremes (Donat et al., 2014, 2016; Dulière et al., 2011). These indices are recommended by the Expert Team on Climate Change Detection and Indices (ETCCDI; Zhang et al., 2011), such as annual maximum value of daily maximum temperature, and number of days with daily precipitation greater than a threshold. Moreover, approaches from extreme value theory (EVT; Coles, 2001) provide a more in-depth investigation of extreme weather and climate events through modeling the extreme values by a statistical model, such as a Gamma distribution (Zolina et al., 2004), a generalized extreme value distribution (Kharin et al., 2005; Mannshardt-Shamseldin et al., 2010), and a generalized Pareto (GP) distribution (Sienez et al., 2010; Blender et al., 2017; Zahid et al., 2017). These models are able to model data which are highly skewed and have heavy-tailed distributions and, thus, are highly non-Gaussian. There are also some other methods that were used in previous studies, including using the fractions skill score to assess the skill of a data set to represent the occurrence of extreme events in a reference data set (Lockhoff et al., 2019), and using the Taylor diagram to characterize the spatial structure of extreme events (Kharin et al., 2005).

In this paper we use several statistical methods to evaluate daily precipitation extremes. We first use the GP distribution to model daily precipitation extremes at different locations. We then use the extremal coefficient (Cooley et al., 2006) to analyze the spatial characteristics of daily precipitation extremes. The spatial characteristics were often not investigated in previous extreme precipitation studies. Then, we use the conditional probability to examine the concurrent event occurrence properties of daily precipitation extremes. Finally, we use the extremal index (Ferro and Segers, 2003) to investigate the temporal clustering characteristics of daily precipitation extremes. The use of these statistical methods provide us with new insights into the quality and accuracy of gridded precipitation data products.

This paper is organized as follows: in Sec. 2 we introduce the used statistical methods, in Sec. 3 we present the four gridded data products and station observations, in Sec. 4 we compare daily precipitation extremes in gridded data products with the observations, and in Sec. 5 we conclude and discuss our results.

2 Methods

2.1 Statistical Modeling of Daily Precipitation Extremes

We select extreme values as the daily rainfall amount that exceeds the 90th percentile. The GP distribution has theoretical justification for fitting such threshold exceedances, and the probability density function (PDF) is given by (see e.g., Coles, 2001):

$$f(x; u, \sigma, \xi) = \begin{cases} \frac{1}{\sigma} \left(1 + \xi \frac{(x-u)}{\sigma}\right)^{-1/\xi-1}, & \xi \neq 0, \\ \frac{1}{\sigma} \exp\left(-\frac{(x-u)}{\sigma}\right), & \xi = 0, \end{cases} \quad (1)$$

where u , σ and ξ are referred to as the threshold, scale and shape parameters, respectively. As we use a relative *threshold*, it becomes a good indicator for the magnitude of the selected precipitation extremes. The *scale parameter* reflects the statistical dispersion of the probability distribution, namely, to which extend it is squeezed or stretched. The *shape parameter* determines the decay rate of the distribution, i.e., how fast the relative likelihood of the occurrence of precipitation extremes decreases as their magnitudes increase. A positive shape parameter characterizes a power-law decay, a zero-valued one means an exponential decay, and a negative one indicates a fast decay and thus a bounded distribution. As seen from Eq. (1) an increase of shape or scale parameter results both in a higher probability of the rainfall of the same amount. To make the scale parameter independent from the threshold we introduce a modified scale parameter: $\sigma^* = \sigma - \xi u$. For determining whether the GP distribution is a valid model for the values above the chosen threshold, we adopt two methods: (i) the mean excess function (MEF) (Coles, 2001) and (ii) a chi-square goodness-of-fit test (Bódai, 2017). Both measures indicate that our chosen threshold is high enough for a GPD to fit daily precipitation extremes well (not shown). For estimating the scale and shape parameters we use a Bayesian estimation approach from the R package `extRemes` (Gilleland and Katz, 2016). The uncertainty of the estimates is given by the credible interval, which is the Bayesian equivalent of the confidence interval.

2.2 The Extremal Coefficient

Accurately modeling the spatial dependence of daily precipitation extremes is a key factor for the generation of gridded precipitation datasets since this will ensure that the spatial structures of extreme events are captured. However, standard correlation analysis is not able to reliably measure the dependence of extremes at two locations, because correlation analysis focuses more on normal events and not extremes (Porcu et al., 2012). Hence, we use the extremal coefficient (Cooley et al., 2006):

$$\hat{\theta}(h) = \frac{1 + 2\nu(h)}{1 - 2\nu(h)}, \quad (2)$$

with

$$\nu(h) = \frac{1}{2} \mathbb{E}[|F(z(x+h)) - F(z(x))|], \quad (3)$$

where h is the distance between two locations and $F(z) = \exp(-1/z)$. The extremal coefficient varies from 1 to 2, with 1 denoting perfect dependence and 2 independence between two locations. For more details and applications of the extremal coefficients see Schlather and Tawn (2003), Davison et al. (2012), Ribatet (2017), and Yang et al. (2020a). We used the R package `SpatialExtremes` (Ribatet, 2019) to compute the extremal coefficient.

2.3 The Conditional Probability

We use the conditional probability to measure the temporal correspondence between the daily precipitation extremes in gridded data products and observed daily precipitation extremes, which is calculated by using the approach of Blender et al. (2017):

$$P(p \geq u | p^* \geq u) = \frac{P(p \geq u \cap p^* \geq u)}{P(p^* \geq u)}, \quad (4)$$

where p denotes the precipitation values in gridded data products, p^* denotes observed precipitation values, and u is the threshold used to select precipitation extremes. The joint probability

$P(p \geq u \cap p^* \geq u)$ is estimated by dividing the frequencies of the precipitation extremes occurring simultaneously in two data sets by the size of precipitation data. The marginal probability $P(p^* \geq u)$ is estimated in a similar way. Eq. (4) gives the probability of an extreme event to occur in gridded data products given that the event has been observed. A value of 1 means that all observed extreme events are reproduced by gridded data products at the same time. Therefore, it measures the skill of gridded data products to represent the exact occurrence time of observed extreme events.

2.4 The Extremal Index

We use the extremal index to measure the degree of clustering, or temporal dependence, of daily precipitation extremes. One interpretation of the extremal index is that its inverse is approximately the mean cluster length of extremes in time. Typically flooding events are triggered by extreme precipitation events occurring over consecutive days. These consecutive occurring extremes constitute a cluster of extremes. Taking account of this clustering is important for risk assessment (Moloney et al., 2019). For estimating the extremal index, we use the interval method described by Ferro and Segers (2003)

$$\tilde{\theta} = \frac{2\{\sum_{i=1}^{N-1}(T_i - 1)\}^2}{(N - 1)\sum_{i=1}^{N-1}(T_i - 1)(T_i - 2)}, \quad (5)$$

where N is the number of extremes and T_i is the time between the $(i + 1)^{th}$ and i^{th} extremes. The extremal index has a value range between 0 and 1, where a value of 1 indicates that the extremes occur independently from each other, while a value smaller than 1 means that they appear in clusters and there is temporal dependency, with long-range dependence (Franzke et al., 2020) for a value of 0. See Ferro and Segers (2003) for more details. The computation of the extremal index is available with the R package `extRemes` (Gilleland and Katz, 2016).

3 Data

We evaluate daily precipitation extremes given by a regional reanalysis, a global reanalysis and two gridded observation-based datasets over the time period 1995-2018 over Germany. The daily precipitation extremes are defined as the daily rainfall amounts that exceed the 90th percentile of the rainfall amounts within the investigated period. As reference data, we use weather station observations provided by the DWD Climate Data Center: Historical daily precipitation observations for Germany, version v007, 2019 (Freydank, 2014; Kaspar et al., 2013; Spengler, 2002). The daily precipitation observations are measured at 06:30 UTC. We choose 701 stations which have data over the used time period and are homogeneously located throughout Germany.

3.1 The COSMO-REA6 Regional Reanalysis

COSMO-REA6 is provided by the Hans Ertel Centre for Weather Research of the DWD, which covers the European domain with a spatial resolution of 0.055° , corresponding to about 6 km grid point distance (Bollmeyer et al., 2015). COSMO-REA6 is created based on the DWD’s COSMO operational numerical weather prediction model and uses a nudging technique for data assimilation. The COSMO model uses the scheme proposed by Tiedtke (1989) for the parametrization of convection. Precipitation observations are not assimilated in COSMO-REA6, however, the assimilation of other observations—the observations of prognostic variables for precipitation such as updraft, temperature, and pressure—still contributes to constraining of precipitation. The regional reanalyses can show large systematic errors for precipitation due to the lack of assimilation of observations into the reanalysis dataset (Bach et al., 2016).

3.2 The ERA5 Global Reanalysis

ERA5 is produced by the European Centre for Medium-Range Weather Forecasts (ECMWF), covering the Earth with a horizontal resolution of 31 km (Hersbach et al., 2020). ERA5 is created based on the ECMWF Integrated Forecasting System (IFS) and a hybrid increment 4D-Var data assimilation system (Bonavita et al., 2016). The precipitation in ERA5 is generated by a large-scale

cloud and precipitation scheme (Tiedtke, 1993; Forbes and Ahlgrimm, 2014; Forbes and Tompkins, 2011) and a convection scheme (Tiedtke, 1989; Bechtold et al., 2008; Hirons et al., 2013). In addition to the conventional observations assimilated in COSMO-REA6, ERA5 assimilates the NCEP stage IV quantitative precipitation estimates produced over the USA (Hersbach et al., 2020).

3.3 The E-OBS Gridded Observation-based Dataset

E-OBS (version v20.0e) is provided by the European Climate Assessment and Dataset consortium (Cornes et al., 2018). E-OBS provides gridded fields with a spatial resolution of 0.1° which corresponds to approximately 10 km grid point distance. This gridded observation-based data set is derived from meteorological station observations. In the gridding procedure, a generalized additive model (Wood, 2017) is fitted to the station values of daily precipitation to represent the long-range spatial correlations in the data. It models the square of the nonzero daily precipitation as a smoothed function of longitude and latitude, plus a smoothed function of squared monthly precipitation totals. The square root transformation reduces the skewness in the precipitation data. The model is fitted using penalized likelihood maximization (Wood, 2017). The monthly totals are gridded using a trivariate thin-plate spline (Cornes et al., 2018).

3.4 The HYRAS (REGNIE) Gridded Precipitation Dataset

HYRAS is provided by the DWD Climate Data Center, covering Germany with a high spatial resolution of 1 km (Rauthe et al., 2013). This gridded daily precipitation dataset is formed from station observations using the REGNIE method, in which a multiple linear regression is applied to create background fields of precipitation. The response variables in the regression are monthly precipitation totals, and the explanatory variables are geographical longitude and latitude, height above sea level, exposition and mountain slope. The regression coefficients for one area are calculated using all available stations in that area and using the least squares method. The residuals of the regression are interpolated from the closest stations using inverse distance weighting if a grid box contains no observations. The ratios of daily precipitation to the background fields at the stations are also interpolated using inverse distance weighting. The interpolated values of daily precipitation at grid points are finally determined by multiplying the ratios with the background fields (Rauthe et al., 2013). A big advantage of the REGNIE method is that it does not smooth the observed precipitation extremes in the gridded field (Rauthe et al., 2013).

For consistency, we linearly interpolate the daily precipitation data of COSMO-REA6, ERA5 and E-OBS to station locations, and choose the HYRAS grid point that is at the nearest to the respective station. Our results are robust against the choice of the interpolation method. The daily precipitation value in reanalysis datasets is obtained by accumulating hourly values from 0700 to 0600 UTC, in order to make it consistent with the observed daily value that is valid for the time period from 0630 to 0630 UTC.

4 Results

4.1 The Probability Density Distribution of Daily Precipitation Extremes

Fig. 1 shows the geographical distribution of the threshold values, shape and scale parameter estimates for station observations. The highest threshold values are shown along the southern boundary of Germany and also in a small region in western Germany. One common feature of these areas is that they are at relatively higher elevation ($\geq 300\text{m}$). In contrast, smaller threshold values are found in northeastern Germany where the altitude is lower. Moreover, the threshold values in summer are generally larger than in other seasons. The geographical structure of the scale parameter estimates is similar to that of the threshold values. The smallest scale parameter estimates are found in winter and the largest estimates are in summer. In comparison with the threshold and scale parameter, a geographical structure of the shape parameter is unclear. However, at most stations the shape parameter is positive, indicating that the probability of the occurrence of extreme daily precipitation events decreases following a power-law as their magnitude increases. This implies that there is a non-negligible probability of getting a daily rainfall amount that is much larger than the threshold value.

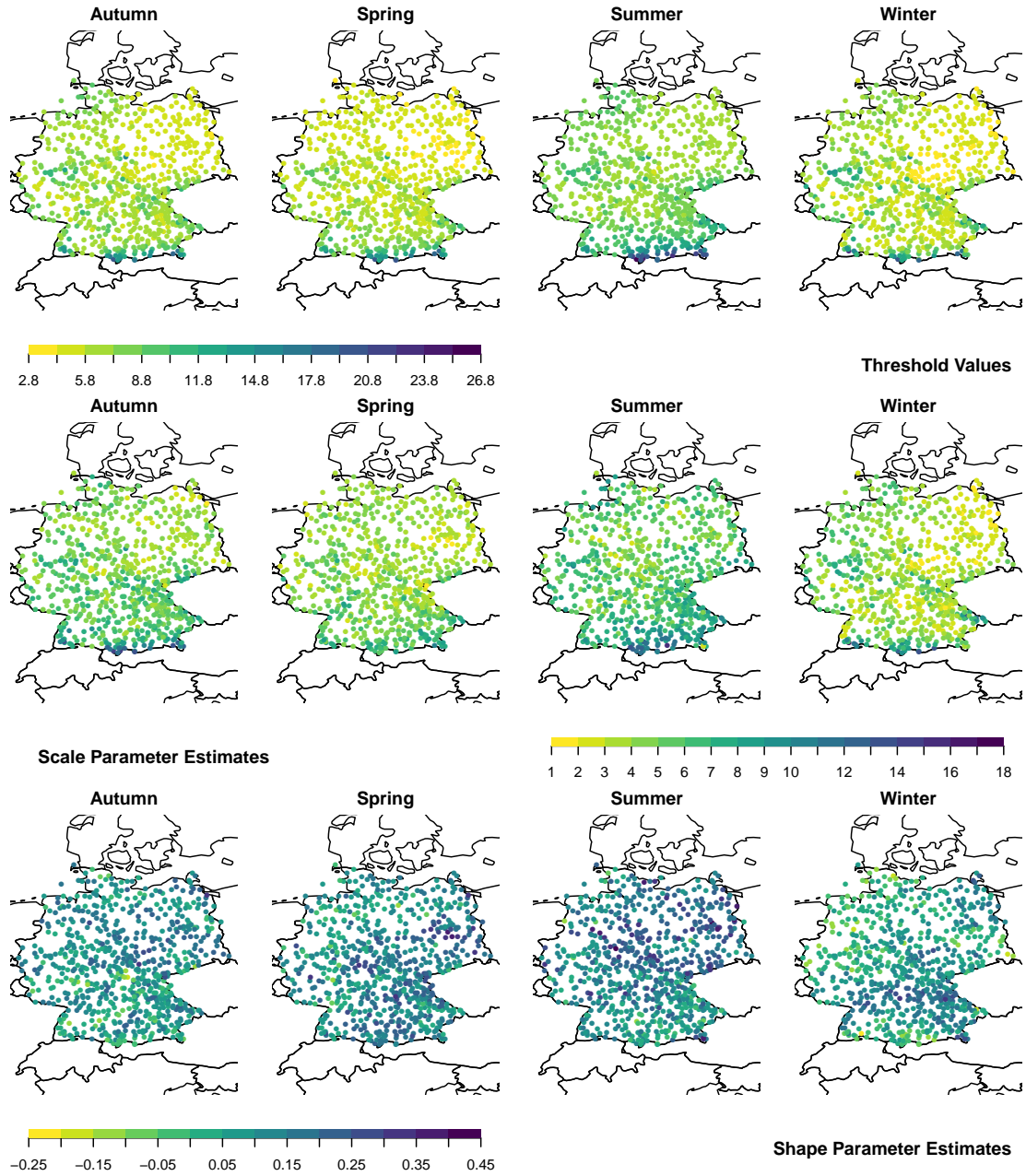


Figure 1: The threshold values and shape and scale parameter estimates of the fitted GP distributions for observed daily precipitation extremes at 701 weather stations throughout Germany. The threshold values are the 90th percentile of daily precipitation values at each location. The scale and shape parameters describe the statistical dispersion and tail behaviour of the distribution of daily precipitation extremes, respectively.

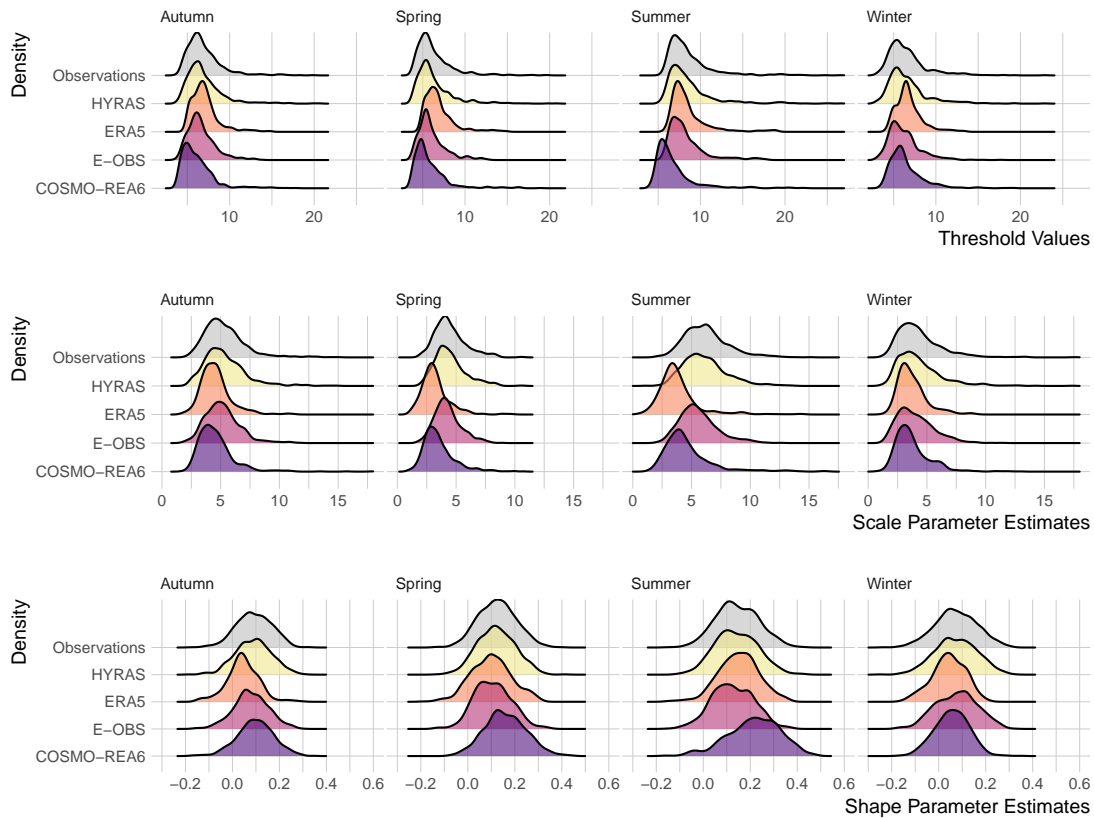


Figure 2: Density ridgeline plots of threshold values, and shape and scale parameter estimates of the fitted GP distributions for daily precipitation extremes in station observations and four gridded data products.

The threshold value and the scale and shape parameters characterize different aspects of the distribution of daily precipitation extremes. The threshold value gives the minimal magnitude of the top 10% of the largest daily precipitation values, whilst the shape and scale parameters describe how these precipitation extremes are distributed. The difference between the shape and scale parameters is that the former emphasises how the distribution decays, while the scale parameter shows how wide is the range of the magnitudes for most of daily precipitation extremes. In our computations, the shape parameter estimates vary from -0.24 to 0.43 with a credible interval of $[-0.15, +0.18]$ for each station, and the scale parameter estimate changes between 1.2 and 17.6 with a credible interval of about ± 1.0 . Therefore, the geographical variation of the scale parameter is more pronounced.

Fig. 2 compares the distribution of threshold values and shape and scale parameter estimates between different gridded data products and station observations. The two gridded observation-based datasets (E-OBS and HYRAS) give closer threshold values and parameter estimates to the station observations than the two reanalysis datasets (COSMO-REA6 and ERA5). The largest disagreement between the reanalysis datasets and station observations is found in the scale parameter; the reanalyses have generally smaller scale parameter estimates than the observations, especially in summer and spring. Additionally, COSMO-REA6 gives lower threshold values and slightly larger shape parameter estimates than station observations in general in summer. Compared to the scale parameter and threshold, the difference in the shape parameter is small, so that the major difference between daily precipitation extremes in gridded data products and station observations is mainly revealed by the threshold and scale parameter. A smaller scale parameter indicates that the distribution of daily precipitation extremes is more squeezed to its lower bound rather than stretched out. In other words, it means that more extremes have values close to the respective threshold. Therefore, a smaller scale parameter reflects an underestimation of the magnitude of the observed daily precipitation extremes in gridded data products. For instance, the observed maximum daily precipitation in the investigated period at 701 weather stations is

251.1mm, while the maximum values in HYRAS, E-OBS, COSMO-REA6 and ERA5 are 243.8mm, 192.9mm, 164.4mm and 131.6mm, respectively. However, it should be noted that precipitation observations are measured at a particular location, whereas the precipitation data in gridded data products represent averages over a grid box. As a result the magnitude of daily precipitation extremes in a gridded dataset with a lower resolution tends to be smaller than that in a gridded dataset with a higher resolution.

4.2 Temporal and Spatial Dependence of Daily Precipitation Extremes

Fig. 3 presents the spatial and temporal dependence of daily precipitation extremes, and how often the gridded data products represent the occurrence of the observed extreme events. The two reanalysis datasets have lower conditional probabilities than the two gridded observation-based datasets (see Fig. 3a). In summer, the conditional probabilities of COSMO-REA6 and ERA5 are only about 0.5, indicating that almost half of all extreme events in these two datasets do not occur at the same time as in the observations. Also for the gridded observation-based datasets, the timing of daily precipitation extremes is captured worse in summer than in the other seasons. Precipitation is often caused by convection in summer and this type of precipitation is generally intense and may suddenly start and stop with the formation and dissipation of active cumulus and cumulonimbus clouds. Therefore, its timing is hard to be captured by a forecast model. Assimilating precipitation observations can potentially improve the timing of daily precipitation extremes in reanalysis datasets.

The extremal index reveals a seasonal change of the strength of the clustering of daily precipitation extremes (see Fig. 3b): the extremes tend to appear in clusters in winter, whereas they occur more independently in summer. This is again a result of different types of precipitation. In summer, the convective precipitation such as rain showers are dominant, which usually occur independently from each other. Whilst in winter large-scale precipitation such as precipitation caused by frontal uplift occurs more often and can last from several hours to days. This seasonal change of the extremal index is not clearly shown by COSMO-REA6. Furthermore, we observe some spatial features of the degree of clustering of daily precipitation extremes over Germany: the daily precipitation extremes at the stations in Northeast Germany have a lower degree of clustering, while the precipitation extremes at the southern stations, particularly the stations along the southern boundary, are more likely to occur in clusters (not shown). The reason of this could be that orographic precipitation (caused by rising terrain, such as a mountain) often takes place in the southern boundary of Germany.

We finally compare the extremal coefficients between gridded data products and station observations. As shown in Fig. 3c, all gridded data products except HYRAS disagree to a relatively large extent with the station observations. The daily precipitation extremes in E-OBS have systematically a higher spatial dependence compared to station observations regardless of their distance. The daily precipitation extremes in COSMO-REA6 and ERA5 show a higher dependence when the distances are small but a lower dependence as the distance increases. COSMO-REA6 gives closer extremal coefficients to station observations than ERA5 when the distance is small.

5 Conclusion and Discussion

We evaluate the quality of four gridded data products in representing daily precipitation extremes over Germany. We find that daily precipitation extremes in gridded observation-based datasets are generally closer to observed daily precipitation extremes than in reanalysis datasets, in terms of their density distribution, spatial and temporal dependence, and timing. The performance of these datasets is not only influenced by their resolution, but also, most importantly, influenced by the different methods used to produce precipitation data.

The precipitation data in the two reanalysis datasets are generated by precipitation parameterization schemes in the respective forecast model. Because COSMO-REA6 does not assimilate precipitation observations, the contribution of data assimilation to the precipitation output is limited to an indirect constraint of precipitation due to the assimilation of observations of prognostic variables that influence precipitation such as vertical wind. ERA5 assimilates precipitation observation equivalents produced over the USA but not over our investigated region. Therefore, the contribution of data assimilation to precipitation data in ERA5 is similar to that to COSMO-REA6. An improvement of precipitation data in reanalyses can be potentially achieved by assimilating

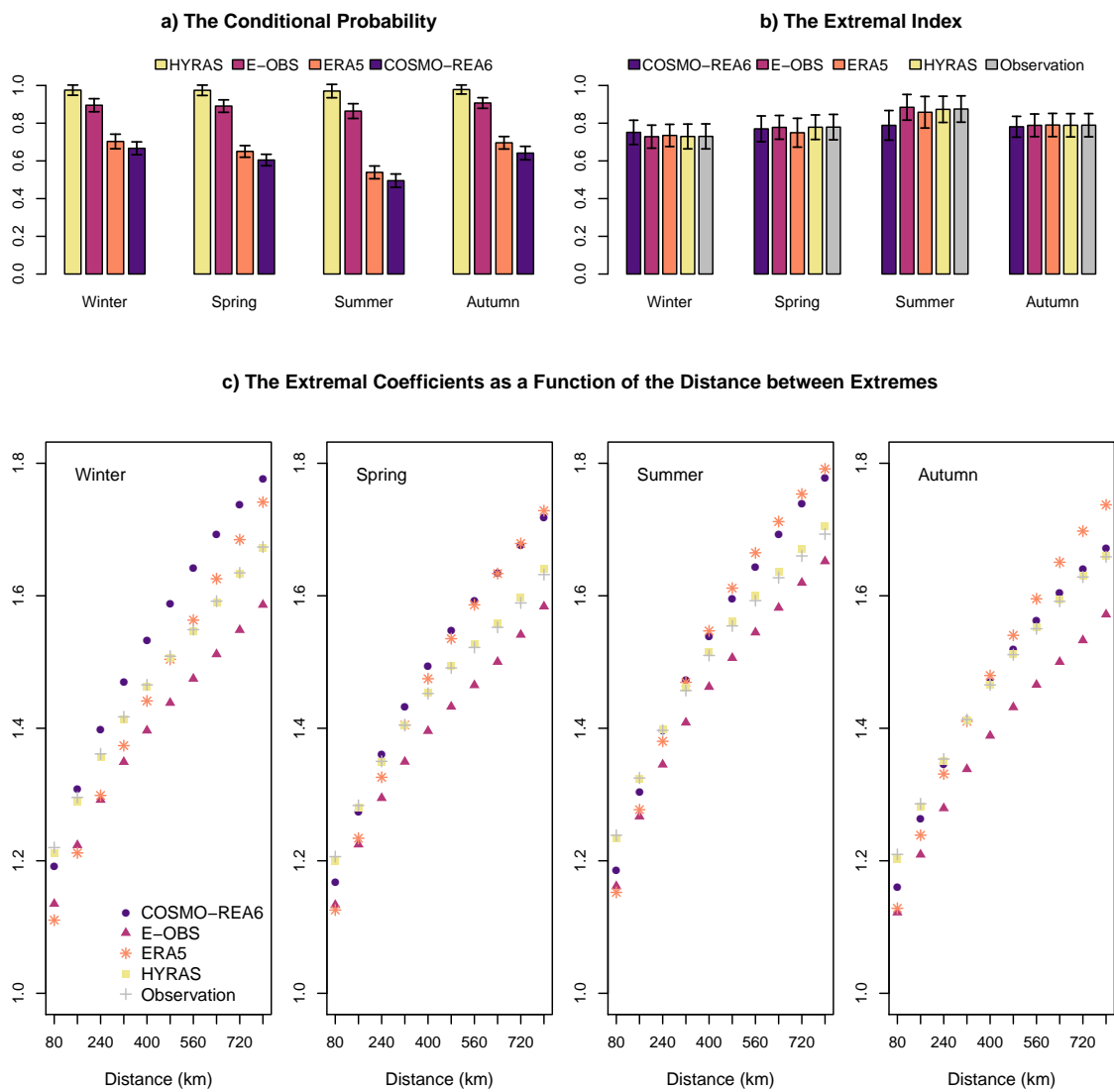


Figure 3: The comparison of daily precipitation extremes between gridded data products and station observations in terms of a) the conditional probability of concurrent occurrence of daily precipitation extremes; b) the extremal index, which reveals the temporal dependency of daily precipitation extremes, with lower dependence as the value increases; c) the extremal coefficients, which indicates the spatial dependence of daily precipitation extremes, with lower dependence as the value increases.

precipitation observations or equivalents. The precipitation observations should not be limited to station gauge data, but also include the data measured by other instruments such as rain radar systems. In addition to COSMO-REA6, DWD provides a convective-scale regional reanalysis for Central Europe (COSMO-REA2), which has a spatial resolution of 2 km, and indeed assimilates radar precipitation estimates using latent-heat nudging (Wahl et al., 2017). However, COSMO-REA2 is only available over a short period from 2007 to 2013 and for this reason we have not include it in our study. The impact of a higher model resolution and assimilation of radar precipitation estimates on precipitation can be found in Wahl et al. (2017). The ECMWF model (IFS) and DWD model (COSMO) both use convection schemes based on Tiedtke (1989), but in IFS a large upgrade has been made (Bechtold et al., 2008; Hirons et al., 2013). Moreover, IFS uses a large-scale cloud scheme to generate large-scale precipitation (Tiedtke, 1993; Forbes and Ahlgrimm, 2014; Forbes and Tompkins, 2011). The upgrades of the convection and cloud schemes in IFS are probably the reason for the better representation of daily precipitation extremes in ERA5 than COSMO-REA6. Currently, a non-hydrostatic extension of IFS is in use at ECMWF for research purposes. Since 2003, DWD has been developing a convective-scale model called COSMO-DE, which is a version of COSMO and partially resolves organized convection (Stephan et al., 2008). The use of such models might improve extreme precipitation in reanalysis data.

The daily precipitation extremes in E-OBS improve over the reanalyses. However, they still show some discrepancy in representing the statistics of the observed daily precipitation extremes. Particularly, the daily precipitation extremes in E-OBS have a systematically larger spatial dependence than observed. E-OBS is available on a 0.1 and 0.25 degree grid, and the high-resolution one performs better in our evaluation of daily precipitation extremes (not shown). HYRAS performs even better than E-OBS in representing daily precipitation extremes. The extreme value statistics of daily precipitation in HYRAS are almost the same as that in station observations. This is a benefit of using a different gridding method and a denser observations network. In E-OBS daily precipitation data are modeled by a smoothed function of geographical factors, while in HYRAS monthly precipitation totals are modeled, so that the daily precipitation proportions are not smoothed. This results in that HYRAS does not smear out daily precipitation extremes in contrast to other interpolation methods with smoothing (Rauthe et al., 2013). Moreover, HYRAS uses a higher number of station observations over Germany and has a higher resolution than E-OBS. The high density of observations also makes the preservation of the station values for the respective grids possible (Rauthe et al., 2013). As the density of observations is a key factor in determining the quality of gridded precipitation data, we suggest to consider the additional use of satellite and radar data for the production of gridded datasets. Another area of potential improvement would be to explicitly consider extreme value statistics in the gridding procedure so that the tail behaviour will be better matched. For example, the coefficients of the statistical model in the gridding procedure are determined by the bulk of precipitation data not extremes. Therefore, they are optimized for the normal precipitation events.

The better performance of the gridded observation-based datasets in representing daily precipitation extremes does not necessarily mean that the gridded observation-based datasets are better than reanalyses. Reanalysis is a three-dimensional dataset, consisting of a large number of atmospheric, land and oceanic climate variables, while gridded observation-based datasets are two-dimensional and contains only a few climate variables. Moreover, gridded observation-based datasets require dense station observations, whilst reanalysis uses a numerical model and can assimilate various types of observations that are measured from different instruments, e.g., satellite observations. This allows reanalysis to cover the region where in situ observations are sparse, such as over the ocean.

In addition to the validation of daily precipitation extremes in gridded data products using station observations, we demonstrate the use of several statistical methods to evaluate extreme values in gridded data products. These methods go beyond standard correlation analysis and provide new insights into the behaviour of extreme events such as daily precipitation extremes. These methods can also be used to evaluate climate models (Yang et al., 2020a,b) and can be applied to other variables as well. The choice of using daily data is limited by data availability. Subdaily or hourly extremes are also relevant to flood forecasting, especially for urban flooding. The evaluation of these precipitation extremes is as important as for daily extremes. The gridding method REGNIE can also be applied to create a hourly gridded precipitation dataset (Van Osnabrugge et al., 2017). This dataset can be updated to near real-time, making it suitable for operational flood forecasting and drought monitoring (Van Osnabrugge et al., 2017).

Acknowledgements

We acknowledge the historical daily precipitation observations for Germany, version v007, 2019, COSMO-REA6 reanalysis, and REGNIE grids of daily precipitation (HYRAS) from DWD; the datasets are accessible via <https://opendata.dwd.de>. We acknowledge the E-OBS dataset from the EU-FP6 project UERRA (<https://www.uerra.eu>) and the Copernicus Climate Change Service, and the data providers in the ECA&D project (<https://www.ecad.eu>); the dataset is accessible via <https://www.ecad.eu/download/ensembles/download.php>. We acknowledge the ERA5 dataset, citable as: Copernicus Climate Change Service (C3S) (2017): ERA5: Fifth generation of ECMWF atmospheric reanalyses of the global climate. Copernicus Climate Change Service Climate Data Store (CDS), date of access. <https://cds.climate.copernicus.eu/cdsapp#!/home>; ERA5 hourly total precipitation data is accessible via <https://cds.climate.copernicus.eu/cdsapp#!/dataset/reanalysis-era5-single-levels?tab=overview>.

We thank two anonymous reviewers for providing valuable comments that helped us to improve our manuscript. GH is funded in part by the UK EPSRC DARE project (EP/P002331/1). CF is supported by the German Research Foundation (DFG) through the collaborative research center TRR 181 "Energy Transfers in Atmosphere and Ocean" at the University of Hamburg, DFG Grant FR3515/3-1 and by the Federal Ministry of Education and Research (BMBF) project ClimXtreme.

References

- Liselotte Bach, Christoph Schraff, Jan D Keller, and Andreas Hense. Towards a probabilistic regional reanalysis system for europe: evaluation of precipitation from experiments. *Tellus A: Dynamic Meteorology and Oceanography*, 68(1):32209, 2016.
- Peter Bechtold, Martin Koehler, Thomas Jung, Francisco Doblas-Reyes, Martin Leutbecher, Mark J Rodwell, Frederic Vitart, and Gianpaolo Balsamo. Advances in simulating atmospheric variability with the ecmwf model: From synoptic to decadal time-scales. *Quarterly Journal of the Royal Meteorological Society: A journal of the atmospheric sciences, applied meteorology and physical oceanography*, 134(634):1337–1351, 2008.
- Richard Blender, CC Raible, and CLE Franzke. Vorticity and geopotential height extreme values in era-interim data during boreal winters. *Quarterly Journal of the Royal Meteorological Society*, 143(703):634–640, 2017.
- Tamás Bódi. Extreme Value Analysis in Dynamical Systems: Two Case Studies. In Christian L. E. Franzke and T. J. O’Kane, editors, *Nonlinear and Stochastic Climate Dynamics*, chapter 14, pages 392–429. Cambridge University Press, Cambridge, 2017.
- C Bollmeyer, JD Keller, C Ohlwein, S Wahl, S Crewell, P Friederichs, A Hense, J Keune, S Kneifel, I Pscheidt, et al. Towards a high-resolution regional reanalysis for the european cordex domain. *Quarterly Journal of the Royal Meteorological Society*, 141(686):1–15, 2015.
- Massimo Bonavita, Elias Hólm, Lars Isaksen, and Mike Fisher. The evolution of the ecmwf hybrid data assimilation system. *Quarterly Journal of the Royal Meteorological Society*, 142(694):287–303, 2016.
- Stuart Coles. *An Introduction to Statistical Modeling of Extreme Values*. Springer, 2001.
- Dan Cooley, Philippe Naveau, and Paul Poncet. Variograms for spatial max-stable random fields. In Patrice Bertail, Philippe Soulier, and Paul Doukhan, editors, *Dependence in Probability and Statistics*, pages 373–390. Springer New York, New York, NY, 2006.
- Richard C. Cornes, Gerard van der Schrier, Else J. M. van den Besselaar, and Philip D. Jones. An Ensemble Version of the E-OBS Temperature and Precipitation Data Sets. *Journal of Geophysical Research: Atmospheres*, 123(17):9391–9409, 2018.
- A. C. Davison, S. A. Padoan, and M. Ribatet. Statistical modeling of spatial extremes. *Statistical Science*, 27(2):161–186, 2012.

- Markus G Donat, Jana Sillmann, Simon Wild, Lisa V Alexander, Tanya Lippmann, and Francis W Zwiers. Consistency of temperature and precipitation extremes across various global gridded in situ and reanalysis datasets. *Journal of Climate*, 27(13):5019–5035, 2014.
- Markus G Donat, Lisa V Alexander, Nicholas Herold, and Andrea J Dittus. Temperature and precipitation extremes in century-long gridded observations, reanalyses, and atmospheric model simulations. *Journal of Geophysical Research: Atmospheres*, 121(19):11–174, 2016.
- Valérie Dulière, Yongxin Zhang, and Eric P Salathé Jr. Extreme precipitation and temperature over the us pacific northwest: A comparison between observations, reanalysis data, and regional models. *Journal of Climate*, 24(7):1950–1964, 2011.
- Christopher AT Ferro and Johan Segers. Inference for clusters of extreme values. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 65(2):545–556, 2003.
- R Forbes and A Tompkins. An improved representation of cloud and precipitation. *ECMWF Newsletter*, 129:13–18, 2011.
- Richard Forbes and Maike Ahlgrimm. On the representation of high-latitude boundary layer mixed-phase cloud in the ecmwf global model. *Monthly Weather Review*, 142(9):3425–3445, 2014.
- Christian L. E. Franzke. Impacts of a changing climate on economic damages and insurance. *Economics of Disasters and Climate Change*, 1(1):95–110, 2017.
- Christian L. E. Franzke and Marcin Czumryna. Probabilistic assessment and projections of US weather and climate risks and economic damages. *Climatic Change*, 158:503–515, 2020.
- Christian L. E. Franzke, Susana Barbosa, Richard Blender, Hege-Beate Fredriksen, Thomas Laepple, Fabrice Lambert, Tine Nilsen, Kristoffer Rypdal, Martin Rypdal, Manuel G Scotto, Stephane Vannitsem, Nicholas W. Watkins, Lichao Yang, and Naiming Yuan. The structure of climate variability across scales. *Rev. Geophys.*, page e2019RG000657, 2020.
- Eberhard Freydanck. *150 Jahre staatliche Wetter-und Klimabeobachtungen in Sachsen: Ergänzungs- und Sondernetze, Messungen in der freien Atmosphäre*. Institut für Hydrologie und Meteorologie, 2014.
- Chaim I Garfinkel and Nili Harnik. The non-Gaussianity and spatial asymmetry of temperature extremes relative to the storm track: The role of horizontal advection. *J. Climate*, 30(2):445–464, 2017.
- Eric Gilleland and Richard W. Katz. extRemes 2.0: An extreme value analysis package in R. *Journal of Statistical Software*, 72(8):1–39, 2016.
- Hans Hersbach, Bill Bell, Paul Berrisford, Shoji Hirahara, András Horányi, Joaquín Muñoz-Sabater, Julien Nicolas, Carole Peubey, Raluca Radu, Dinand Schepers, et al. The era5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, 2020.
- LC Hirons, P Inness, F Vitart, and P Bechtold. Understanding advances in the simulation of intraseasonal variability in the ecmwf model. part ii: The application of process-based diagnostics. *Quarterly Journal of the Royal Meteorological Society*, 139(675):1427–1444, 2013.
- Trevor Houser, Solomon Hsiang, Robert Kopp, and Kate Larsen. *Economic risks of climate change: an American prospectus*. Columbia University Press, 2015.
- F Kaspar, G Müller-Westermeier, E Penda, H Mächel, K Zimmermann, A Kaiser-Weiss, and T Deutschländer. Monitoring of climate change in germany—data, products and services of germany’s national climate data centre. *Advances in Science and Research*, 10(1):99–106, 2013.
- Viatcheslav V Kharin, Francis W Zwiers, and Xuebin Zhang. Intercomparison of near-surface temperature and precipitation extremes in amip-2 simulations, reanalyses, and observations. *Journal of Climate*, 18(24):5201–5223, 2005.
- M Lockhoff, O Zolina, C Simmer, and J Schulz. Representation of precipitation characteristics and extremes in regional reanalyses, satellite-and gauge-based estimates over western and central Europe. *Journal of Hydrometeorology*, 20(6):1123–1145, 2019.

- Elizabeth C Mannshardt-Shamseldin, Richard L Smith, Stephan R Sain, Linda O Mearns, Daniel Cooley, et al. Downscaling extremes: A comparison of extreme value distributions in point-source and gridded precipitation data. *The Annals of Applied Statistics*, 4(1):484–502, 2010.
- Nicholas R Moloney, Davide Faranda, and Yuzuru Sato. An overview of the extremal index. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 29(2):022101, 2019.
- Emilio Porcu, José-María Montero, and Martin Schlather. *Advances and challenges in space-time modelling of natural events*, volume 207. Springer Science & Business Media, 2012.
- Monika Rauthe, Heiko Steiner, Ulf Riediger, Alex Mazurkiewicz, and Annegret Gratzki. A central european precipitation climatology—part i: Generation and validation of a high-resolution gridded daily data set (hyras). *Meteorologische Zeitschrift*, 22(3):235–256, 2013.
- M. Ribatet. Modelling spatial extremes using max-stable processes. In C. L. E. Franzke and T. O’Kane, editors, *Nonlinear and Stochastic Climate Dynamics*, pages 369–391. Cambridge University Press, Cambridge, 2017.
- Mathieu Ribatet. *SpatialExtremes: Modelling Spatial Extremes*, 2019. URL <https://CRAN.R-project.org/package=SpatialExtremes>. R package version 2.0-7.2.
- Martin Schlather and Jonathan A. Tawn. A dependence measure for multivariate and spatial extreme values: Properties and inference. *Biometrika*, 90(1):139–156, 03 2003.
- Frank Sienz, Andrea Schneidereit, Richard Blender, Klaus Fraedrich, and Frank Lunkeit. Extreme value statistics for North Atlantic cyclones. *Tellus*, 62(4):347–360, 2010.
- Reinhard Spengler. The new quality control and monitoring system of the deutscher wetterdienst. In *Proceedings of the WMO Technical Conference on Meteorological and Environmental Instruments and Methods of Observation, Bratislava, 2002*.
- Klaus Stephan, S Klink, and C Schraff. Assimilation of radar-derived rain rates into the convective-scale model cosmo-de at dwd. *Quarterly Journal of the Royal Meteorological Society: A journal of the atmospheric sciences, applied meteorology and physical oceanography*, 134(634):1315–1326, 2008.
- M Tiedtke. Representation of clouds in large-scale models. *Monthly Weather Review*, 121(11):3040–3061, 1993.
- MICHAEL Tiedtke. A comprehensive mass flux scheme for cumulus parameterization in large-scale models. *Monthly weather review*, 117(8):1779–1800, 1989.
- B Van Osnabrugge, AH Weerts, and R Uijlenhoet. genre: A method to extend gridded precipitation climatology data sets in near real-time for hydrological forecasting purposes. *Water Resources Research*, 53(11):9284–9303, 2017.
- Sabrina Wahl, Christoph Bollmeyer, Susanne Crewell, Clarissa Figura, Petra Friederichs, Andreas Hense, Jan D Keller, and Christian Ohlwein. A novel convective-scale regional reanalysis cosmo-rea2: Improving the representation of precipitation. *Meteorol. Z*, 26(4):345–361, 2017.
- Simon N Wood. *Generalized additive models: an introduction with R*. CRC press, 2017.
- Lichao Yang, Christian L. E. Franzke, and Zuntao Fu. Evaluation of the ability of regional climate models and a statistical model to represent the spatial characteristics of extreme precipitation. *International Journal of Climatology*, pages 1–17, 2020a. doi: 10.1002/joc.6602.
- Lichao Yang, Christian L. E. Franzke, and Zuntao Fu. Power-law behaviour of hourly precipitation intensity and dry spell duration over the united states. *International Journal of Climatology*, 40(4):2429–2444, 2020b. doi: 10.1002/joc.6343.
- Maida Zahid, Richard Blender, Valerio Lucarini, and Maria Caterina Bramati. Return levels of temperature extremes in southern pakistan. *Earth System Dynamics*, 8(4):1263–1278, 2017.

Xuebin Zhang, Lisa Alexander, Gabriele C Hegerl, Philip Jones, Albert Klein Tank, Thomas C Peterson, Blair Trewin, and Francis W Zwiers. Indices for monitoring changes in extremes based on daily temperature and precipitation data. *Wiley Interdisciplinary Reviews: Climate Change*, 2(6):851–870, 2011.

Olga Zolina, Alice Kapala, Clemens Simmer, and Sergey K Gulev. Analysis of extreme precipitation over europe from different reanalyses: a comparative assessment. *Global and Planetary Change*, 44(1-4):129–161, 2004.