



Robust regional clustering and modeling of nonstationary summer temperature extremes across Germany

Meagan Carney and Holger Kantz

Max Planck Institute for the Physics of Complex Systems, Dresden, Germany

Correspondence: Meagan Carney (meagan@pks.mpg.de)

Received: 28 November 2019 – Revised: 24 March 2020 – Accepted: 5 May 2020 – Published: 8 June 2020

Abstract. We use sophisticated machine-learning techniques on a network of summer temperature and precipitation time series taken from stations throughout Germany for the years from 1960 to 2018. In particular, we consider (normalized) maximized mutual information as the measure of similarity and expand on recent clustering methods for climate modeling by applying a weighted kernel-based k -means algorithm. We find robust regional clusters that are both time invariant and shared by networks defined separately by precipitation and temperature time series. Finally, we use the resulting clusters to create a nonstationary model of regional summer temperature extremes throughout Germany and are thereby able to quantify the increase in the probability of observing high extreme summer temperature values ($> 35^{\circ}\text{C}$) compared with the last 30 years.

1 Introduction

Extremes in temperature have lasting affects on human health (Gasparrini et al., 2017), power consumption (Thornton et al., 2016; Bartos et al., 2016; Burillo, 2019), biodiversity, and ecosystem services (Bellard et al., 2012; Balvanera et al., 2013). As extreme events occur in the tails of probability distributions, studying them allows us to make predictions of rare, and more catastrophic, weather events. For this reason, recent climate literature has become more focused on extreme event statistics (Brown et al., 2008; Cheng et al., 2014; Christidis et al., 2011; Finkel and Katz, 2018).

Modeling extreme events typically involves fitting a generalized extreme value (GEV) probability distribution to these data under the assumption that the parameters of this distribution do not change in time or are “stationary” (Coles, 2001). Climate variability has resulted in challenges to these assumptions. Steady increasing trends in the mean air temperature have been observed worldwide (Hansen et al., 2010; Wergen and Krug, 2010) and have been attributed to anthropogenic green house gas emissions (Erkwurzel et al., 2017; Hansen et al., 2012; Christidis et al., 2011). With the continued warming trend and recurring heat waves sweeping across Europe (Kysely, 2002; Schär et al., 2004), most re-

cently occurring in the summers of 2015, 2018, and 2019, the need for nonstationary modeling for accurate extreme temperature prediction has become more apparent (Christidis et al., 2015; Hasan et al., 2013; Hamdi et al., 2018; Rahmstorf and Coumou, 2011). We cite Caroni and Panagoulia (2016), Gao and Zheng (2018), and Hamdi et al. (2018) for examples of nonstationary extreme value modeling; however, there are numerous other examples in the literature. Nonstationary adaptations are particularly important for countries that are less equipped to handle these high temperature events, such as Germany (Tomczyk and Sulikowska, 2018; Kosanic et al., 2019).

Along with nonstationary modeling, there is the need to create regional models of extremes as policy making and governmental response are most influenced when extreme weather events are predicted on a larger scale. With the introduction of machine-learning techniques such as maximized modularity algorithms (Newman, 2006; Duch and Arenas, 2005) and spectral clustering (Hastie et al., 2008; Luxburg, 2007) in many scientific areas including genetics, physics, computer science, and neuroscience, among others (Guimera and Amaral, 2005; Treviño et al., 2012; Zscheischler et al., 2012; Azencott et al., 2018), we are better equipped to deal with large data which provide a more accurate, global un-

derstanding of networks. These techniques are particularly useful for dense clusters, such as those observed in climate data where meteorological towers are gridded; networks that are affected by many external factors; and where obvious divisions are unknown.

Recently, spectral clustering methods have been shown to provide reasonable clustering results for climate networks defined by a set of time series (Carney et al., 2019); however, these methods may fail when networks with high intra-cluster similarity are considered. We expand on the methods of Carney et al. (2019) by introducing a weighted kernel approach based on a spectral view of the network that allows for more reliable clustering of dense climate networks where simple *k*-means methods cannot be reliably implemented. This method provides a robust regional clustering of summer temperature time series from meteorological stations throughout Germany. We show that these regions are both time invariant and equivalent over summer temperature and precipitation time series. We use these clusters to create nonstationary regional models of temperature extremes. Our results show a significant increase in the probability of returns for high summer temperature values ($> 35^{\circ}\text{C}$) for each region in Germany.

This paper is outlined as follows. Section 2 is broken into three parts: Sect.2.1 contains details on the chosen subset of data; Sect. 2.2 provides relevant background on clustering methods and a detailed description of the clustering algorithm; and Sect. 2.3 provides relevant background on extreme value modeling and details on the methods used for regional modeling of summer temperature extremes. Section 3 is broken into two parts: Sect. 3.1 contains results on the clustering algorithm applied to summer temperature and precipitation time series taken from the stations across Germany; and Sect. 3.2 outlines the extreme value results on the station level and their incorporation into the regional models formed from the clustering outcome. Section 4 is a brief summary of the paper containing future work motivated by the results in this paper.

2 Data and methods

2.1 Data

The data in this analysis were taken from the Deutscher Wetterdienst Climate Data Center (DWD CDC, 2018a), which contains climate records from stations across Germany. We use three data sets from the DWD in total: hourly temperature values from 1960 to 2018 (measured in Celsius), hourly precipitation values from 1995 to 2018 (measured in millimeters), and daily precipitation amounts from 1960 to 2018 (measured in millimeters). All three data sets are used in clustering analysis, although only the hourly temperature data set is used for extremes analysis. We choose two precipitation data sets because hourly precipitation values are not available across Germany before 1995 and daily precipita-

Table 1. DWD data sets used in clustering.

DWD data set	Time interval	No. of stations
Hourly temperature	1960–2018	56
Hourly precipitation	1995–2018	115
Daily precipitation	1960–2018	65

Geographic locations of stations in Germany

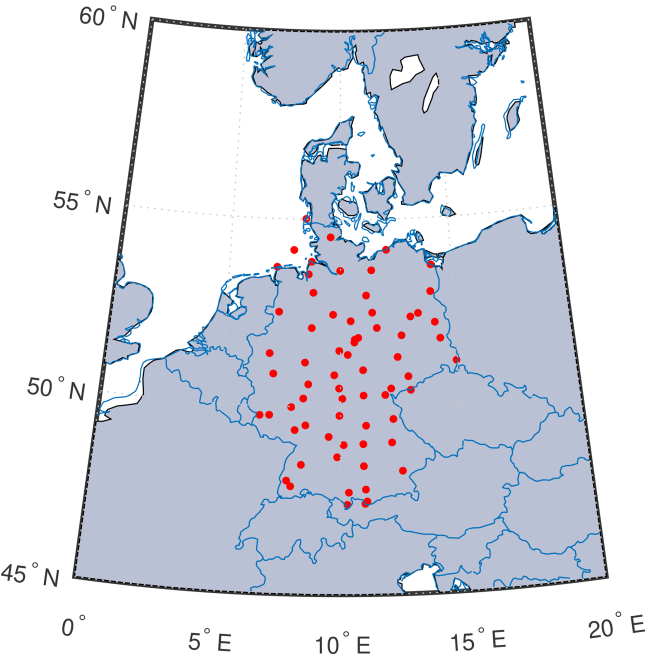


Figure 1. Geographic locations of 68 stations used in regional clustering and modeling of temperature extremes.

tion values do not provide as much resolution. A summary of these data sets can be found in Table 1.

The time interval from 1960 to 2018 provides a general representation of all climatology in Germany and a reasonable quantity of extreme values for modeling. For information on station locations used in this analysis, see Fig. 1; detailed station location and identification information can be found in Table A1.

Quality control was performed by the DWD CDC with details provided in Kaspar et al. (2013). We remark that all values used in this analysis passed this preprocessing procedure. These data are not detrended to account for the urban heat island; however, we do not expect this to affect the quality of this analysis because most stations are located several kilometers outside of major cities. With the transfer from manual to automatic weather recording stations, additional considerations are made while investigating the statistical parameters in these data. Some stations in our subset contain missing values; while there are methods that allow us to replace

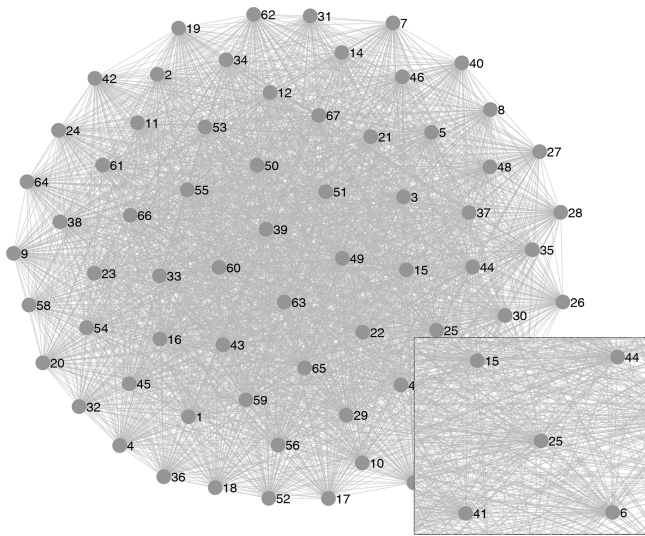


Figure 2. Graphical form of the network defined by temperature time series emphasizing the density of the network. Each node represents a station in the network. Each edge represents a positive similarity between two stations. Note that this graph is not complete because low values of similarity (< 0.1) are taken to be zero in this analysis.

missing data in the time series, such as linear interpolation and the like, these methods require an a priori assumption on the trend in the time series. Stations containing missing values are removed from extremal analysis; however, they are still used for regional clustering because missing data will not notably affect the algorithm.

2.2 Regional clustering

This section discusses some machine-learning tools used to perform regional clustering by viewing the set of time series, corresponding to our subset of weather stations, as a network. The set of time series can be viewed as a weighted network (in graph form: a set of nodes and edges) where each station is a node. The weighted edge between two nodes is calculated by a similarity measure that relies on the corresponding time series (Fig. 2).

Given a network, the goal of clustering analysis is to find a partition with high intra-cluster similarity and low inter-cluster similarity; however, solving this system is an np-hard (non-polynomial) problem. Relaxations of this problem can be split into two approaches: (1) maximizing modularity through partitioning algorithms and redefined modularity metrics (Newman, 2006; Duch and Arenas, 2005; Sun et al., 2009; Li et al., 2008; Chen et al., 2018) or (2) minimizing a distance function through spectral clustering, k -means, and the like (Hastie et al., 2008; Luxburg, 2007; Dhillon et al., 2004). Under standard measures of modularity, Li et al. (2008) showed that these two approaches have an equiv-

alence relation: modularity is maximized if and only if the kernel k -means distance is minimized.

The authors in Carney et al. (2019) approach regional clustering of temperature time series using the spectral clustering method equipped with k -means (Luxburg, 2007). Given n time series, each time series is represented by a point in \mathbb{R}^n where the coordinate is given by the corresponding value of similarity to every time series in the network. The goal of this algorithm is then to find hyperplane separations between the clusters in \mathbb{R}^n by minimizing the objective function defined by the sum of the intra-cluster distances. The spectral portion of this algorithm allows for a reduction in the dimension of the system which results in a set of more easily separable clusters in the lower-dimensional space. The advantage of this approach is that it is computationally efficient and reliable provided that the clusters can be linearly separated. Linear separability is rarely the case with larger, dense clusters (such as in this study) where all similarity measures are close in value. Instead, we employ a weighted Gaussian kernel k -means approach. This method provides a more reliable regional clustering of the time series by allowing the separation of the clusters using a nonlinear function at some computational expense. The weights w_h are given by eigenvalues λ_h of a normalized symmetric, positive definite matrix so that all $\lambda_h \in [0, 1]$. These weights serve to compress the clusters along coordinates of less importance, in terms of the corresponding eigenvalue, $w_h = \lambda_h \ll 1$ and maintain space along coordinates of most importance $w_h = \lambda_h \approx 1$. We outline the main ideas of the algorithm here; for a more rigorous explanation we refer the reader to the Appendix A1. We also refer to Wang et al. (2019) for a nice introduction on kernel k -means and Hastie et al. (2008, Sect. 14.5.3) and Luxburg (2007) for more information on the spectral portion of this algorithm.

- *Step 1.* Given any matrix of similarities $S_{i,j} \geq 0$, calculate the normalized Laplacian

$$\mathbf{L} = \mathbf{D}^{-1/2} (\mathbf{D} - \mathbf{S}) \mathbf{D}^{-1/2}, \quad (1)$$

where \mathbf{D} is the diagonal matrix with entries $D_{i,i} = \sum_{j=1}^n S_{i,j}$.

- *Step 2.* Calculate the spectrum of \mathbf{L} and choose a cutoff ℓ used to create a subspace of eigenvectors \mathbf{V}_h for $h = 1, \dots, \ell$ of \mathbf{L} corresponding to the first ℓ eigenvalues.
- *Step 3.* Calculate the weighted Gaussian kernel

$$K(\hat{\mathbf{W}}_i, \hat{\mathbf{W}}_j) = \exp\left(\frac{[\|\hat{\mathbf{W}}_i\|^2 + \hat{\mathbf{W}}_i^T \hat{\mathbf{W}}_j - \|\hat{\mathbf{W}}_j\|^2]/s^2}{s^2}\right), \quad (2)$$

where $\hat{\mathbf{W}}_i$ and $\hat{\mathbf{W}}_j$ are the respective weighted i th and j th row vectors \mathbf{W}_i and \mathbf{W}_j of the matrix of eigenvectors \mathbf{V}_h and $s \in \mathbb{R}$. $\hat{\mathbf{W}}_{i,h} = w_h \times \mathbf{W}_{i,h}$ where $w_h = 1 - \lambda_h$.

– Step 4. Run kernel k -means.

- Step 4.1. Start with k random partitions P_k .
- Step 4.2. Compute the centers $C_k = \sum_{m \in P_k} K(\hat{W}_m, \cdot) / \text{card}(P_k)$, where $K(\hat{W}_m, \cdot)$ is the m th row vector of the kernel matrix \mathbf{K} and $\text{card}(\cdot)$ is the cardinality function.
- Step 4.3. Assign \hat{W}_h to the partition P_k with centroid C_k such that

$$\sum_k \sum_{\hat{W}_h \in P_k} \|K(\hat{W}_h, \cdot) - C_k\|^2 \quad (3)$$

is minimized.

- Step 4.4. Recalculate the centroids based on new assignments of \hat{W}_h .
- Step 4.5. Repeat assignment until a stable minimum is reached.

It is possible that the stable minimum reached is not the global minimum. We run the algorithm for 1000 different initial sets of partitions and choose the cluster associated with the lowest minimized distance. In addition, the number of partitions (clusters) in kernel k -means needs to be assumed a priori. A standard practice is to run the algorithm for different values of k and choose the optimal value as the smallest k corresponding to the greatest decrease in variance. This technique is typically referred to as the “elbow” method Yuan and Yang (2019).

Mutual information has become a more common measurement of similarity in recent literature because it is capable of measuring nonlinear relationships between two time series. For two random variables (or time series) X_1 and X_2 taking discrete values, mutual information is defined as

$$\mathcal{I}(X_1, X_2) = \sum_{i=1}^q \sum_{j=1}^q p(x_{1,i}, x_{2,j}) \log \left(\frac{p(x_{1,i}, x_{2,j})}{p(x_{1,i}) \cdot p(x_{2,j})} \right), \quad (4)$$

where $p(x_{1,i}) = \mathbb{P}(X_1 = x_{1,i})$ and $p(x_{1,i}, x_{2,j}) = \mathbb{P}(X_1 = x_{1,i}, X_2 = x_{2,j})$. Intuitively, mutual information compares the probability distributions of X_1 and X_2 by measuring their dependence in terms of the joint distribution relative to an assumption of independence. A positive value for Eq. (4) indicates that X_1 and X_2 are dependent in some sense. On the other hand, it is straightforward to see that Eq. (4) equals zero if X_1 and X_2 are independent. We may rewrite mutual information in terms of the entropy of X_1 and X_2 as

$$\mathcal{I}(X_1, X_2) = \mathcal{H}(X_1) + \mathcal{H}(X_2) - \mathcal{H}(X_1, X_2), \quad (5)$$

with entropy $\mathcal{H}(X_1)$ defined by

$$\mathcal{H}(X_1) = - \sum_{i=1}^q p(x_{1,i}) \log p(x_{1,i}), \quad (6)$$

and joint entropy $\mathcal{H}(X_1, X_2)$ given by

$$\mathcal{H}(X_1, X_2) = - \sum_{i=1}^q \sum_{j=1}^q p(x_{1,i}, x_{2,j}) \log p(x_{1,i}, x_{2,j}). \quad (7)$$

This reconstruction provides a natural expression of mutual information in terms of information and allows for the calculation of error under the compression of the continuous time series. As the true probability distributions of the data are unknown, a standard approach to calculating the entropy of the time series is to compress the data into discrete bins which form an approximate uniform distribution. From here, the error on the mutual information can be calculated by comparing the resulting compressed distribution with proportion probabilities $\hat{p}(x_{1,i}, x_{1,j})$ to the uniform distribution with equal proportion $p(x_{1,i}, x_{1,j})$ (Roulston, 1999).

Compressing the time series can result in an obvious loss of information from the system. Following the work of Carney et al. (2019) and Zhang et al. (2018), we perform gradient ascent by varying the end points of the compression interval to maximize the mutual information between each pair of time series. We will refer to this measurement as the maximized mutual information (MMI). This approach should not change the order of estimated error on the mutual information as entropy is maximized on uniform distributions.

We use a normalized version of the maximized mutual information (NMMI),

$$\mathcal{U}(X_1, X_2) = \mathcal{I}(X_1, X_2) / \min\{\mathcal{H}(X_1), \mathcal{H}(X_2)\}, \quad (8)$$

for the network defined by the temperature time series. Normalized versions of the mutual information can produce better clustering results by allowing the mutual information to be viewed relative to the entropy of the time series. In the case of the precipitation network, we find that normalizing the maximized mutual information causes a large error in the spectral decomposition of the normalized Laplacian. This error is a product of naturally low entropy in all time series causing a large value of NMMI and, hence, a large value of D , so that the values of L become too low for optimal spectral decomposition. For this reason, we use the unnormalized version of the maximized mutual information for the precipitation network. Higher values in the entropy for the temperature time series can be attributed to hourly and annual cycles, resulting in a distribution with less dispersion than that of the precipitation time series.

Mutual information may return a large positive value for negatively correlated variables. In this analysis, we cross reference our clustering outcomes from mutual information and correlation. Summer hourly temperature time series (with the daily cycle removed) for all stations in the network are positively correlated. This is expected because Germany is smaller than the average warm air mass. If we attempt to cluster the network by correlation, we find the minimized Euclidean distance plot loses its structure and no longer provides an obvious value for the number of clusters k . We make

two observations: first, the correlation network does not allow for reliable clustering of regional temperature values; and, second, mutual information adds substance to the network.

We remark that mutual information provides a reasonable estimate of similarity in the context of clustering extremes because it compares the probability distributions of the time series to determine their shared information. Recent literature has also introduced measures of similarity defined by the extremal index (EI; Bador et al., 2015). This measure incorporates a comparison of the location and scale parameters from the time series' corresponding GEVs in the sense that two GEVs with equal shape parameters and differing EIs are equivalent up to a shift of their location and scale parameters (Coles, 2001, Theorem 5.2).

2.3 Local and regional extreme value modeling

The block maximum method is used in this analysis to model the maximum values of the time series (Coles, 2001). In this method, the time series is broken into blocks of a chosen length l so that asymptotic results hold: where M is the length of the time series, the number of maximum values is M/l . Provided the block maxima are independent and stationary, classical extreme value results ensure that they will converge to a generalized extreme value (GEV) distribution described by a shape k , location μ , and scale σ parameter.

$$F(x) = \exp\left[-\left(1 + k\left(\frac{x - \mu}{\sigma}\right)\right)^{-1/k}\right] \quad (9)$$

under the requirement $1 + k\left(\frac{x_i - \mu}{\sigma}\right) > 0 \forall i$. The shape parameter describes the tail behavior of the maximum values and defines the exact probability distribution to which they will converge, Frechét (Type II) for $k > 0$, Weibull (Type III) for $k < 0$, or Gumbel (Type I) for $k = 0$.

When modeling real-world data, it is natural to use maximum likelihood estimation (MLE; Coles, 2001) or the L -moments method (Hosking and Wallis, 1997) to estimate the three parameters of the GEV model. L -moments is a common technique in climate modeling, and its advantage comes from its stability for low quantities of data (Hosking, 1990); however, the limiting distributions of the estimates are unknown, and bootstrapping methods are required to obtain confidence intervals. Moreover, adaptations to the nonstationary model are not well studied. As we have a relatively large pool of data, maximum likelihood estimation is chosen to estimate the parameters and confidence intervals of the GEV distribution. This is done by maximizing the following log-likelihood function for a set of m block maxima:

$$\mathcal{L}_{k,\mu,\sigma} = -m \log \sigma - (1 + 1/k) \sum_{i=1}^m \log \left[1 + k \left(\frac{x_i - \mu}{\sigma} \right) \right] - \sum_{i=1}^m \left[1 + k \left(\frac{x_i - \mu}{\sigma} \right) \right]^{1/k}, \quad (10)$$

Clustering result
hourly temperature values from 1960 to 2018

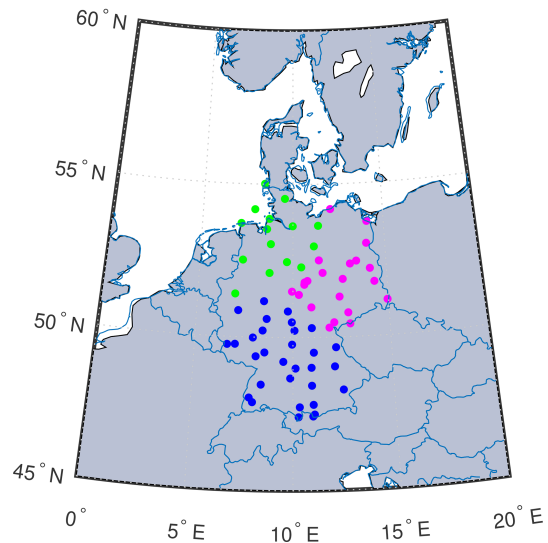


Figure 3. Clustering result for the hourly temperature network with similarity defined by NMMI over the time interval from 1960 to 2018.

provided that

$$1 + k \left(\frac{x_i - \mu}{\sigma} \right) > 0 \forall i \quad (11)$$

Strict independence is not always necessary for convergence of the maxima to the GEV (Lucarini et al., 2016); however, in practice it is common to choose blocks that are “long enough” for independence to be observed. As temperature time series follow a daily cycle, this daily cycle is removed during independence analysis by subtracting the average cycle over all days in the data. Seasonal cycles are not considered as the data were taken only from summer months. Compression is performed on the centralized data vector by assigning +1 to values that increase in the next step, −1 for values that decrease, and 0 for values that stay the same. These quantities are tabulated, and the chi-square test of independence is performed.

Stationary assumptions require the location and scale parameters of the maxima to not change in time. We test this assumption using the Mann–Kendall trend test and a mean difference comparison of the time intervals from 1960 to 1990 and from 1991 to 2018 for the temperature maximum at each station where we find evidence of a linear trended location parameter.

Finally, we perform a likelihood-ratio test on the stationary and nonstationary model equipped with a linear time-dependent location parameter $\mu(t) = \beta_0 + \beta_1 \times t$. Estimates of the parameters of the nonstationary GEV model were obtained by maximum likelihood estimation of the following

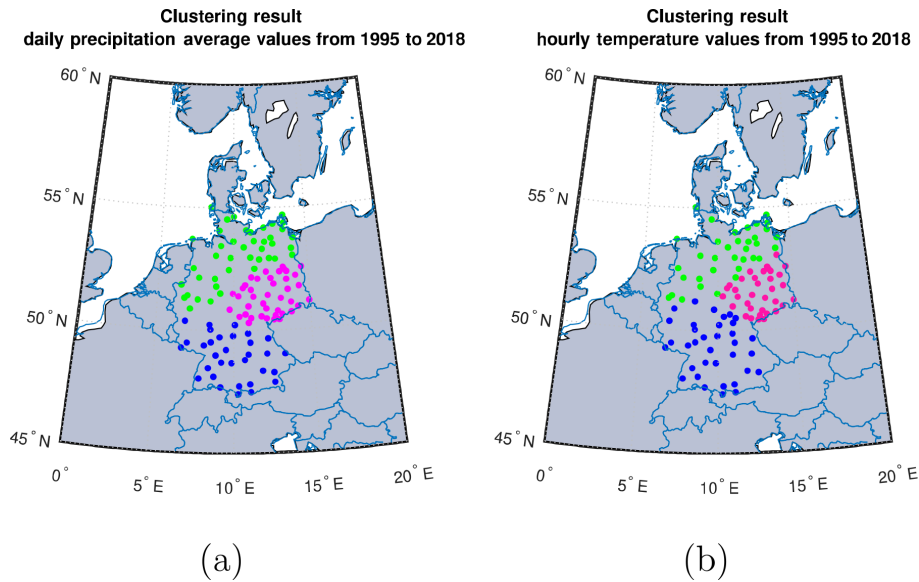


Figure 4. Clustering result for the (a) daily precipitation amounts network taken from the hourly DWD data set with similarity defined by MMI, and (b) the hourly temperature network with similarity defined by NMMI over the time interval from 1995 to 2018.

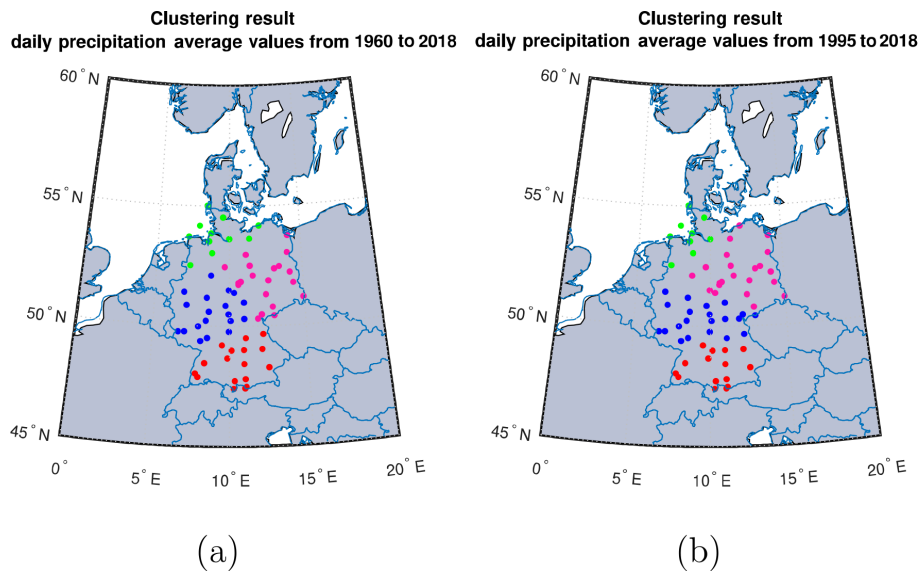


Figure 5. Clustering result for the daily precipitation amounts network taken from the daily DWD data set over (a) the time interval from 1960 to 2018 and (b) the time interval from 1995 to 2018 with similarity defined by MMI.

log-likelihood function:

$$\begin{aligned} \mathcal{L}_{k,\beta_0,\beta_1,\sigma} = & -m \log \sigma \\ & - (1 + 1/k) \sum_{i=1}^m \log \left[1 + k \left(\frac{x_i - \mu(t_i)}{\sigma} \right) \right] \\ & - \sum_{i=1}^m \left[1 + k \left(\frac{x_i - \mu(t_i)}{\sigma} \right) \right]^{1/k}, \end{aligned} \quad (12)$$

where t_i is the index of a block over which the maxima is taken.

The likelihood-ratio test compares the final log-likelihood values \mathcal{L}_1 and \mathcal{L}_0 of the nonstationary model \mathcal{M}_1 and the stationary model \mathcal{M}_0 via the “deviance statistic”:

$$\mathcal{D} = 2\{\mathcal{L}_1(\mathcal{M}_1) - \mathcal{L}_0(\mathcal{M}_0)\}, \quad (13)$$

where $\mathcal{D} \sim \chi^2(1)$. To test whether the resulting nonstationary model fits the data, we transform the original time series of maximum values, Z_t by

$$\tilde{Z}_t = \frac{1}{k} \log \left\{ 1 + k \left(\frac{Z_t - \mu(t)}{\sigma} \right) \right\}. \quad (14)$$

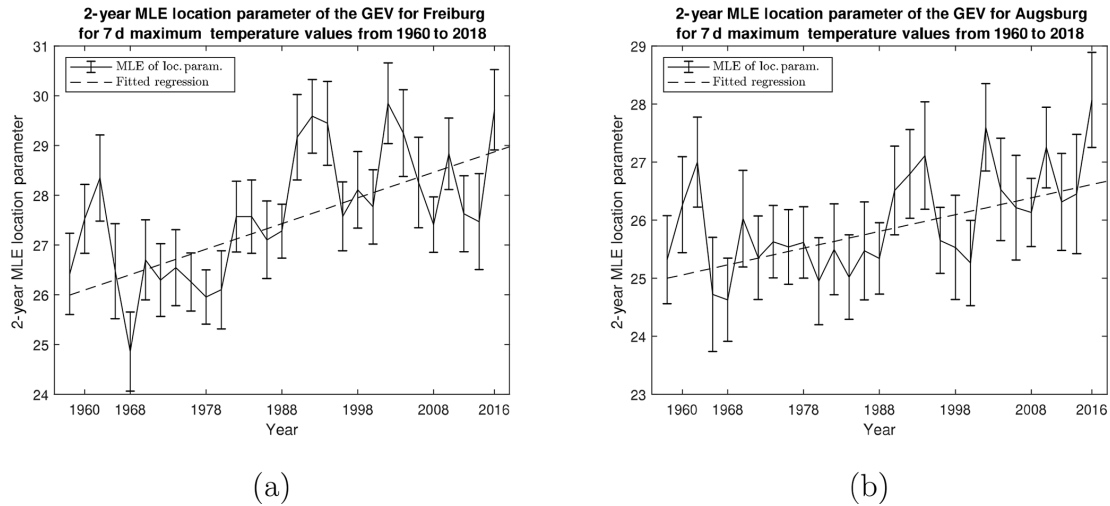


Figure 6. Example time plots for the 2-year maximum likelihood estimate of the location parameter with standard error bars for weekly maximum temperature values taken from (a) Freiburg with fitted regression $y = 25.89 + 0.10 \times t$ and (b) Augsburg with fitted regression $y = 24.94 + 0.06 \times t$ (where t is 2-years). Time plots from other stations are similar and can be attributed to the previously mentioned correlations between stations throughout Germany.

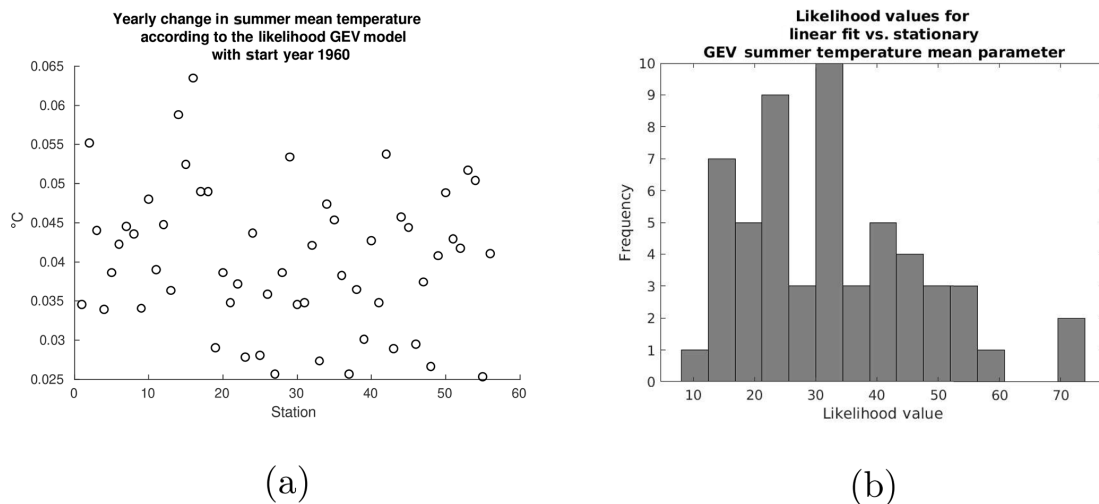


Figure 7. (a) Scatter plot of the yearly change in the temperature mean according to the likelihood GEV model for all stations. (b) Likelihood values comparing the linear fit of the summer temperature mean to that of a stationary mean. All likelihood values surpass the chi-square value of $\chi^2(1, 0.05) = 3.841$.

It is known that \tilde{Z}_t follows a stationary Gumbel distribution with $\mu = 0$ and $\sigma = 1$ (Coles, 2001). We perform the Anderson–Darling goodness of fit test on \tilde{Z}_t under these assumptions for every station in this analysis.

Regional models are created by formulating a mixed GEV distribution from all of the stations in a regional cluster. The mixed GEV distribution is given by

$$F(z_1, \dots, z_N) = \frac{1}{N} \sum_{i=1}^N \exp \left[- \left(1 + k_i \left(\frac{z_i - \mu_i(t)}{\sigma_i} \right) \right)^{-1/k_i} \right], \quad (15)$$

where N is the number of stations in the regional cluster. This distribution describes the probability of an observed temperature extreme occurring in the regional population defined by the cluster of subpopulations (stations) given an equal probability of sampling from any subpopulation in the regional population. We refer to Yoon et al. (2013) for an example of mixed GEV distributions; however, we remark that fitting results are not required in our context because subpopulations are known.

3 Results and conclusions

3.1 Regional clustering results

As discussed in Sect. 2.2, we compute the maximized mutual information between each pair of stations represented by their corresponding hourly temperature time series from 1960 to 2018. Throughout this analysis, all error for maximized mutual information is on the order of 10^{-5} . Outcomes of our clustering algorithm applied to the maximized mutual information matrix suggest that Germany has three ($k = 3$) distinct climate regions located in (roughly) the northwest, the northeast, and the south (Fig. 3).

To test for time invariance of the clusters, we apply the algorithm to the intervals from 1960 to 2000, from 1970 to 2010, and from 1980 to 2018. Resulting clusters are equivalent for all three time intervals.

For further investigation, we consider clusters obtained by maximizing the mutual information between pairs of stations represented by their corresponding daily precipitation amounts. Daily precipitation amounts for this analysis are taken from the same hourly recording data set as the temperature data. Precipitation data in this set are only available for the time period from 1995 to 2018. All available data over the period from 1995 to 2018 are used to show the robustness of the regional clusters; however, these data are not used in the final regional cluster analysis of temperature extremes. Hourly precipitation values are summed over each 24 h period and the resulting time series is used to compute the MMI matrix. Clustering results suggest the network of daily precipitation amounts is regionally equivalent to that of hourly temperature (Fig. 4).

For added measure, we consider the daily precipitation amounts taken from the daily DWD CDC set which have values over the same time period (1960–2018) as our temperature data set.

Results from our clustering algorithm show preference for four ($k = 4$) regional clusters over Germany. This fourth cluster is created from a splitting of the southern cluster (Fig. 5). The period from 1995 to 2018 is checked for this data set with similar results. An investigation into the raw data shows slight discrepancies when compared with that of the summed hourly precipitation data which may be the result of differences in the time of recording.

3.2 Local and regional extreme value modeling results

For the remaining analysis, we define an extreme (or maximum) value of the temperature time series as the maximum taken over data blocks of 7 d (168 data points). We find independence over almost all blocks at the $\alpha = 0.05$ significance level. The quantity of blocks that do not return independence are less than 5 % and can be attributed to the type I error of the test.

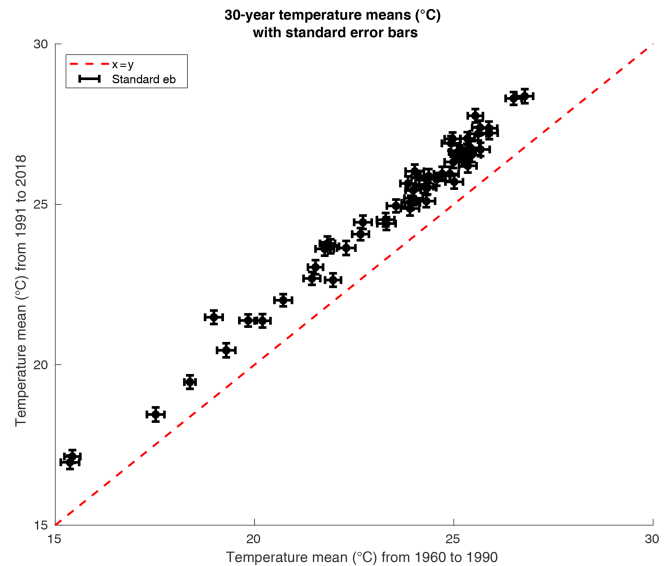


Figure 8. Scatter plot illustrating the difference in mean summer temperature taken over the period from 1960 to 1990 (x axis) and from 1991 to 2018 (y axis) with standard error bars. All values and standard error bars are located above the identity.

Recall from Sect. 2.3 that fitting a GEV model to the maximum values in a time series requires stationarity of the maximum values, e.g., their location and scale parameters do not change in time. We first test this assumption using the Mann–Kendall trend test on the 2-year ($m = 24$ block maxima) maximum likelihood estimate of the location and scale parameters for weekly summer temperature extremes with the p values indicating that a significant trend ($\alpha = 0.05$) exists for the location parameter and no trend exists for the scale parameter for all stations. We choose 2 years ($m = 24$) of maximum values because this exceeds the minimum sample size requirement for estimating the mean while providing enough data to investigate the trend. Low quantities of block maxima produce larger standard errors on the location parameter. As a way of obtaining additional information to make conclusions regarding the trend, we perform a mean-difference test on the maximum likelihood estimate of the location parameter over the time intervals from 1960 to 1990 and from 1991 to 2018 ($m \approx 373$ block maxima per interval) with the alternative hypothesis $\mu_{1991-2018} > \mu_{1960-1990}$. As the MLE location parameter is asymptotically normal, we use the 95 % confidence intervals to calculate the standard error used in the mean-difference test. We find that this difference is significant (at the $\alpha = 0.05$ significance level) for every station; see Fig. 7 for these results. We also refer the reader to Table A2 for more detailed information. We note that, due to the variability in the location parameter described earlier, low p values from the Mann–Kendall test imply low p values for the mean-difference test, whereas the converse is not true. Time plots are also evaluated as a superficial method of determining whether this trend is the result of a single change-point

Table 2. Confidence intervals for transformed regional models.

Regional cluster	k	95 %CI	μ	95 %CI	σ	95 %CI
Northwest	−0.0157	[−0.0240, −0.0075]	0.0087	[−0.0061, 0.0234]	1.0045	[0.9940, 1.0152]
Northeast	0.0154	[0, 0.0308]	−0.0083	[−0.0303, 0.0138]	0.9946	[0.9785, 1.0109]
South	0.0009	[−0.0135, 0.0154]	−0.0006	[−0.0233, 0.0222]	0.9997	[0.9833, 1.0164]

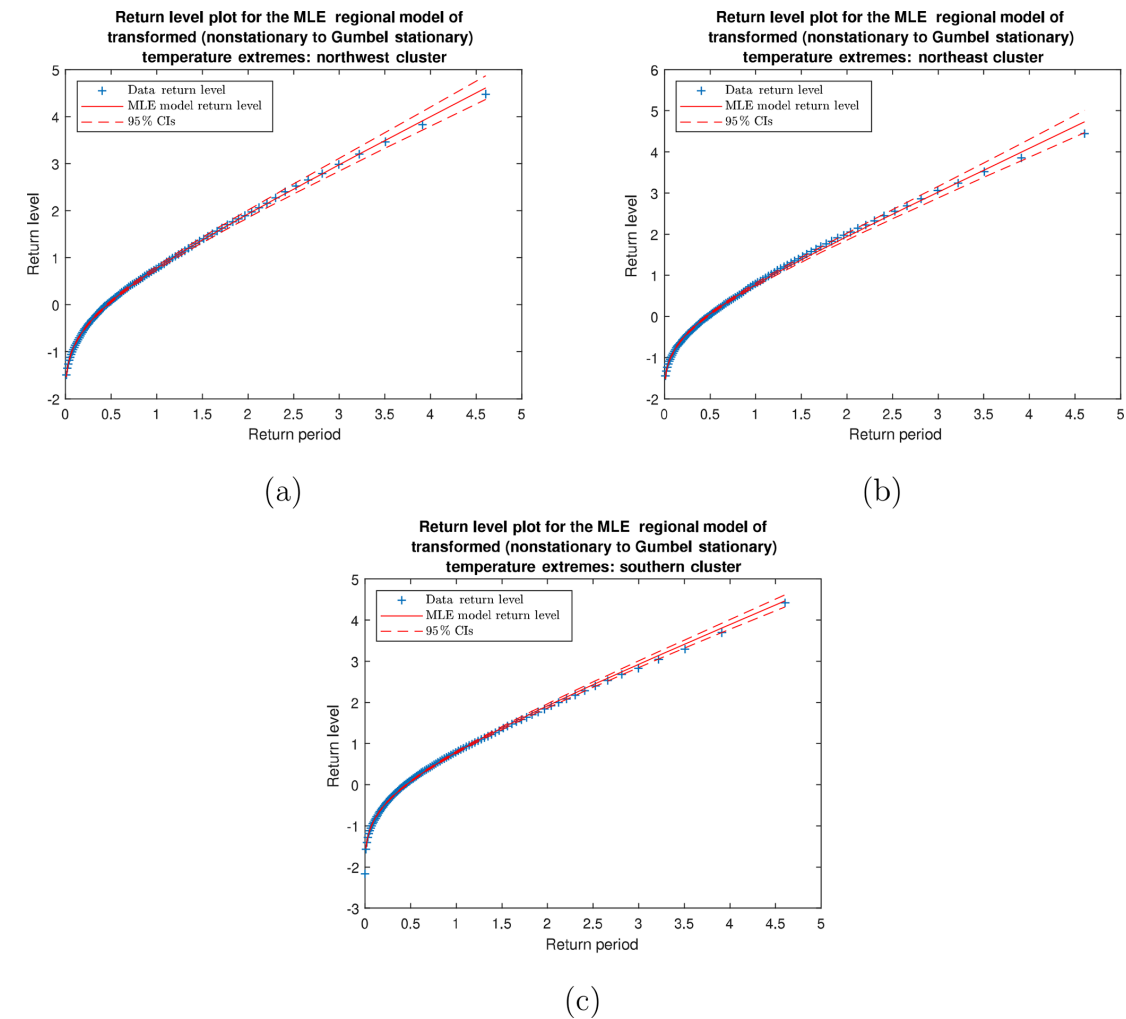


Figure 9. Return level plots comparing the parametric mixed GEV distribution with confidence intervals and actual mixed temperature maxima for the (a) northwest, (b) northeast, and (c) south clusters. These plots emphasize the regional model’s fit to the pooled data.

in the 1990s (Fig. 6). These plots show a continued upward trend, suggesting that this phenomenon is not the result of manual/automatic instrumental change.

Results from these tests provide justification for considering nonstationary modeling of the weekly summer temperature extremes. We use maximum likelihood estimation of the log-likelihood function equipped with a linear trended location parameter to generate the nonstationary GEV model for every station in the data set. We compare these results with the stationary model using the likelihood ratio test, and

we find that the nonstationary model provides a significant ($\alpha = 0.05$) improvement over the stationary model for every station; see Fig. 8 for these results. We also refer the reader to Table A3 for more detailed information.

We form regional GEV models using the mixed GEV distribution described by Eq. (14). Stations located at high altitude, including Zugspitze and (the mountains near) G rlitz, were removed from the regional model. These stations have location parameter values that are further than 2 standard deviations away from the regional location parameter.

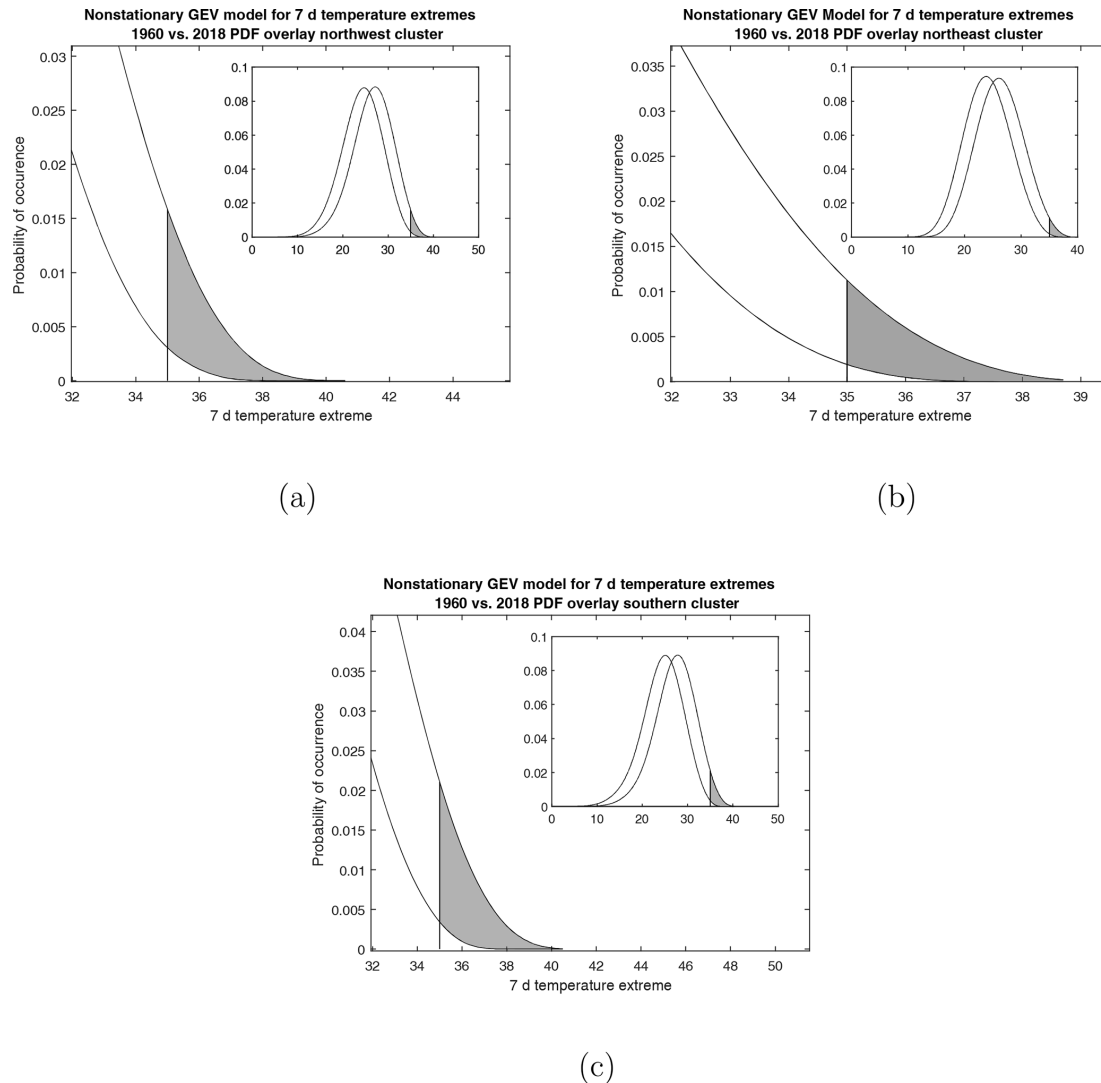


Figure 10. Nonstationary generalized extreme value model overlay of 1960 and 2018 weekly temperature extremes for the (a) northwest (b) northeast, and (c) south regional clusters. The shaded region indicates the probability of observing a weekly extreme higher than 35 °C.

This deviation causes issues with the modality and accuracy of regional temperature extreme probability estimation. When creating regional models of extremes, care needs to be taken regarding the homogeneity of the region in terms of the shape parameter. This is especially important for regional models of extremes where local phenomena are observed. We remark that the mixed GEV distribution has been used to describe regional dynamics with unknown local phenomena through likelihood estimation (Yoon et al., 2013); however, this method is not relevant for our purposes because the subpopulation GEVs (station GEVs) are known and the shape parameters for temperature maxima from each station are nearly equivalent.

To determine how well the regional models represent the pooled time series, we first transform the time series of maxima for each station into a stationary Gumbel distribution us-

ing Eq. (13) and perform the Anderson–Darling goodness of fit test on the transformed maxima under the null hypothesis that these data follow a stationary Gumbel distribution with $\mu = 0$ and $\sigma = 1$. We find that all stations follow the null hypothesis at the $\alpha = 0.05$ significance level. Using Eq. (14), we create a transformed (stationary) regional model from the transformed time series that is also expected to follow a Gumbel distribution of $\mu = 0$ and $\sigma = 1$. We perform maximum likelihood estimation on the transformed regional models to test this expectation. All 95 % confidence intervals contain the values $\mu = 0$, $\sigma = 1$, and $k = 0$ (see Table 2). This transformation allows us to create return level plots for the regional model (Fig. 9). Our results suggest that the return levels obtained for the transformed regional model fit the transformed time series. We remark that these return level plots can only be used in model diagnostics to ensure that the re-

Table 3. Regional tail probability estimates (in units of occurrences per year) for weekly temperature extremes above 35 °C.

Regional cluster	1960	2018	2018/1960
Northeast	1.90×10^{-3}	1.13×10^{-2}	5.95
Northwest	3.10×10^{-3}	1.58×10^{-2}	5.10
South	3.40×10^{-3}	0.02	6.21

gional model is a good representation of the pooled data. Return times for probabilities of temperature maxima exceeding the threshold must be derived from the mixed GEV with a fixed time value.

We use the MLE parameters k , $\mu(t)$, and σ of the nonstationary GEV models for the mixed GEV regional distribution with fixed t related to the beginning of 1960 and the beginning of 2018. Comparing these models shows an increase in the probability of observing maximum weekly summer temperatures greater than 35 °C of up to 6 times for 2018 compared with 1960 (Fig. 10, Table 3). We remark that the largest difference in the probability of observing an extreme temperature event occurs in the southern cluster, which is likely the result of ocean regulation on the warming trend in the northern clusters (Pörtner et al., 2020).

4 Summary

In this analysis, we expand on the work of Carney et al. (2019) by considering a more dense set of networks and adding a weighted kernel k -means approach to the algorithm. This addition allows for more reliable clustering where the k -means method was shown to fail. We find regional clusters for weather stations throughout Germany by defining a network based on their corresponding summer, hourly temperature values. The resulting clusters are not only time invariant but are shared by the network defined by the corresponding daily precipitation amounts. The addition of more stations to the network does not radically change the clustering outcome.

Shared clusters could reflect an underlying relationship between temperature and precipitation dynamics. This result is nontrivial because similarity at the station level does not generalize in an obvious way due to the numerous external factors that affect regional weather. We remark that the clustering equivalence does not (necessarily) indicate that temperature and precipitation move together, rather that their individual networks contain relationships between stations within equivalent spacial regions.

We use the clusters we obtain from our algorithm to create regional models of the weekly summer temperature extremes. We find significant increasing linear trends in the location parameter of these models and a preference for nonstationary modeling for all stations in the network. Regional nonstationary models are created by defining a mixed nonstationary GEV distribution from likelihood results. Return time plots are used to validate the results. Regional distribution models reveal an increase in the probability of observing a weekly summer temperature maximum above 35 °C of up to 6 times for 2018 compared with 1960.

In future work, we will consider more complicated networks defined by both temperature and precipitation values, simultaneously. We plan to use these results to form joint probability densities of temperature and precipitation extremes. Viewing the data in this way will allow us to make conclusions about the conditional extreme probabilities. We are also interested in determining whether these clusters are seasonally dependent or consistent throughout the year.

Appendix A: Complementary material

A1 A spectral weighted kernel k -means algorithm

The similarity matrix $\mathbf{S}_{i,j}$ in this study is transformed into the normalized Laplacian given by

$$\mathbf{L} = \mathbf{D}^{-1/2}(\mathbf{D} - \mathbf{S})\mathbf{D}^{-1/2}. \quad (\text{A1})$$

Spectral decomposition is performed on \mathbf{L} , and a matrix is formed with columns given by the first h eigenvectors $[\mathbf{V}_1, \dots, \mathbf{V}_h]$ of \mathbf{L} . As the Laplacian is normalized, the first h eigenvalues represent the eigenvectors carrying the most information from the network (represented by the Laplacian matrix). The corresponding subspace made of the span of $[\mathbf{V}_1, \dots, \mathbf{V}_h]$ is the reduced dimensional network that approximates the true network well. The rows of this matrix correspond to the h -dimensional projections of the input vectors (e.g., the h -dimensional projections of the rows of the similarity matrix). Each row gives a point representation $x \in \mathbb{R}^h$ to the stations in the network. The goal of the next step in the algorithm is to find divisions that create clusters of these points in \mathbb{R}^h .

Kernel k -means allows for separation of clusters by a nonlinear function. This is done by taking a nonlinear mapping $\phi(x): \mathbb{R}^h \rightarrow \mathcal{F}$ of each input vector $x \in \mathbb{R}^h$ into a higher-dimensional (possibly infinite) feature space. Once in the feature space, the goal of clustering is to find partitions that minimize the objective function now given by

$$\min \frac{1}{N} \sum_{k=1}^{\ell} \sum_{j \in P_k} \|\phi(x_j) - \frac{1}{\text{card}(P_k)} \sum_{i \in P_k} \phi(x_i)\|_{\mathcal{F}}^2, \quad (\text{A2})$$

where P_k is the k th partition, and $i \in P_k$ indicates the index assigned to P_k . Schölkopf et al. (1998) defined the “kernel trick” by recognizing that ϕ can be viewed as a function on the Hilbert space \mathcal{F} equipped with the inner product norm $\langle \cdot, \cdot \rangle$. This allows us to redefine the objective function

$$\min \frac{1}{N} \sum_{k=1}^{\ell} \sum_{j \in P_k} \|K(x_j, \cdot) - \frac{1}{\text{card}(P_k)} \sum_{i \in P_k} K(x_i, \cdot)\|_2^2 \quad (\text{A3})$$

in terms of a kernel $K(x_i, x_j) = \langle \phi(x_i), \phi(x_j) \rangle$. Now, we may minimize the objective function without having to explicitly compute the feature vectors $\phi(x)$. Choosing $K(x_i, x_j)$ as the Gaussian kernel used in this analysis, given by

$$K(x_i, x_j) = \exp\left(-\|x_i - x_j\|^2/s^2\right), \quad (\text{A4})$$

ensures some universality in the sense that any function required to separate the cluster can be approximated arbitrarily well by a function of the kernel form. The weighted Gaussian kernel is calculated by

$$K(\hat{x}_i, \hat{x}_j) = \exp\left([|\hat{x}_i|^2 + \hat{x}_i^T \hat{x}_j - |\hat{x}_j|^2]/s^2\right), \quad (\text{A5})$$

where $w = \mathbf{1} - \bar{\lambda}$; here, $\bar{\lambda}$ is the vector of h first eigenvalues of \mathbf{L} and $\hat{x}_{i,h} = w_h \times x_{i,h}$. The values of w represent the amount of information (given by the associated eigenvalue of \mathbf{L}) carried by a component of $x \in \mathbb{R}^h$. By using w to weight the input vectors of the kernel, we are able to tighten clusters along coordinates with the least amount of information, or $w_h = \lambda_h \ll 1$, while maintaining distance along coordinates with the most amount of information, or $w_h = \lambda_h \approx 1$.

Table A1. DWD station IDs and locations.

ID	Latitude	Longitude	Location
3987	52.3813	13.0622	Potsdam
5100	49.7479	6.6582	Trier – Petrisberg
1580	49.9859	7.9549	Geisenheim
867	50.3066	10.9679	Lautertal – Oberlauter
2290	47.8009	11.0108	Hohenpeißenberg
4271	54.1803	12.0808	Rostock – Warnemünde
656	51.7234	10.6021	Braunlage
1197	48.9895	10.1312	Ellwangen – Rindelbach
1550	47.4830	11.0621	Garmisch-Partenkirchen
1691	51.5002	9.9507	Göttingen
2261	50.3122	11.8761	Hof
2597	50.2241	10.0792	Bad Kissingen
2601	50.2218	8.4469	Kleiner Feldberg
3730	47.3986	10.2760	Oberstdorf
4104	49.0425	12.1019	Regensburg
4887	48.6657	9.8646	Stötten
5371	50.4973	9.9428	Wasserkuppe
5705	49.7703	9.9577	Würzburg
5906	49.5062	8.5585	Mannheim
691	53.0450	8.7979	Bremen
701	53.5332	8.5761	Bremerhaven
1975	53.6332	9.9881	Hamburg – Fuhlsbüttel
2014	52.4644	9.6779	Hannover
3032	55.0110	8.4125	List on Sylt
3631	53.7123	7.1519	Norderney
1639	50.6017	8.6439	Gießen – Wetttenberg
4371	52.1042	8.7521	Bad Salzungen
4625	53.6425	11.3872	Schwerin
5792	47.4209	10.9847	Zugspitze
433	52.4675	13.4021	Berlin – Tempelhof
662	52.2915	10.4464	Braunschweig
722	51.7986	10.6183	Brocken
853	50.7913	12.8720	Chemnitz
891	53.8713	8.7058	Cuxhaven
1270	50.9829	10.9608	Erfurt–Weimar
1303	51.4041	6.9677	Essen – Bredeney
1358	50.4283	12.9535	Fichtelberg
1443	48.0233	7.8344	Freiburg
1468	48.4537	8.4091	Freudenstadt
1544	52.5129	11.3941	Gardelegen
1684	51.1622	14.9506	Görlitz
2171	50.8519	9.7378	Bad Hersfeld
2483	51.1803	8.4891	Kahler Asten
2928	51.3151	12.4462	Leipzig – Holzhausen
3015	52.2085	14.1180	Lindenberg
3023	52.5181	7.3081	Lingen
3668	49.5030	11.0549	Nürnberg
3946	50.4819	12.1300	Plauen
4466	54.5275	9.5487	Schleswig
5142	53.7444	14.0697	Ueckermünde
5397	49.6663	12.1845	Weiden
5490	51.8454	10.7686	Wernigerode
5629	51.8892	12.6445	Wittenberg
1346	47.8749	8.0038	Feldberg – Schwarzwald
2115	54.1750	7.8920	Helgoland

Table A1. Continued.

ID	Latitude	Longitude	Location
5426	49.3758	8.1213	Weinbiet
3093	52.9724	11.1374	Lüchow
232	48.4254	10.9420	Augsburg
953	49.7619	7.0541	Deuselbach
2559	47.7233	10.3348	Kempten
3126	52.1029	11.5827	Magdeburg
3366	48.2791	12.5024	Mühlendorf
3761	49.2070	9.5175	Öhringen
5440	49.0116	10.9308	Weißenburg – Emetzheim
164	53.0316	13.9908	Angermünde
880	51.7760	14.3168	Cottbus
2925	51.3933	10.3123	Leinefelde
2667	50.8646	7.1575	Köln-Bonn

Table A2. Location trend results for the hourly temperature time series. $p_{\text{MEAN DIFF}} H_A: \mu_{1960-1990} < \mu_{1991-2018}$ and $p_{\text{MK}} H_A: \text{trend}$.

Station	$\mu_{1960-1990}$	STE	$\mu_{1991-2018}$	STE	$p_{\text{MEAN DIFF}}$	p_{MK}
3987	25.67	0.23	26.71	0.21	3.43×10^{-6}	7.50×10^{-3}
5100	24.93	0.23	26.90	0.23	0	1.73×10^{-5}
1580	25.88	0.21	27.37	0.21	8.29×10^{-13}	7.89×10^{-4}
867	24.99	0.22	26.32	0.19	1.46×10^{-9}	1.7×10^{-3}
2290	21.53	0.19	23.04	0.22	8.88×10^{-16}	3.08×10^{-5}
4271	21.76	0.23	23.61	0.21	1.11×10^{-16}	3.25×10^{-7}
656	20.72	0.22	22.01	0.19	2.64×10^{-9}	4.08×10^{-5}
1550	24.13	0.19	25.83	0.21	0	2.00×10^{-5}
1691	24.90	0.22	25.95	0.22	7.72×10^{-7}	7.01×10^{-4}
2261	22.72	0.22	24.44	0.21	1.11×10^{-15}	3.55×10^{-5}
2597	25.42	0.22	26.76	0.21	4.89×10^{-10}	2.7×10^{-3}
2601	19.84	0.22	21.38	0.19	7.30×10^{-13}	1.73×10^{-5}
3730	23.98	0.19	25.44	0.20	1.22×10^{-15}	2.00×10^{-5}
4104	25.54	0.19	27.76	0.21	0	3.55×10^{-5}
4887	21.83	0.19	23.78	0.22	0	3.25×10^{-7}
5371	18.98	0.22	21.48	0.21	0	1.46×10^{-6}
5705	25.61	0.21	27.20	0.21	4.04×10^{-14}	2.64×10^{-4}
5906	26.78	0.21	28.37	0.22	3.78×10^{-15}	0.01
691	23.90	0.23	24.87	0.22	8.30×10^{-6}	2.20×10^{-3}
701	22.30	0.23	23.64	0.22	3.35×10^{-9}	2.67×10^{-5}
1975	23.31	0.23	24.40	0.20	9.96×10^{-7}	9.30×10^{-5}
2014	24.34	0.22	25.52	0.21	3.20×10^{-8}	4.70×10^{-5}
3631	20.20	0.20	21.37	0.21	2.97×10^{-9}	6.06×10^{-6}
1639	25.10	0.22	26.63	0.22	1.64×10^{-12}	2.64×10^{-4}
4371	24.31	0.22	25.10	0.19	1.93×10^{-4}	2.64×10^{-4}
4625	23.29	0.22	24.52	0.21	1.42×10^{-8}	9.30×10^{-5}
5792	6.72	0.20	8.18	0.22	1.81×10^{-13}	2.31×10^{-5}
662	24.27	0.22	25.55	0.20	1.90×10^{-9}	1.58×10^{-4}
722	15.37	0.23	16.96	0.21	1.54×10^{-12}	1.11×10^{-5}
853	23.92	0.23	25.05	0.22	2.34×10^{-7}	5.00×10^{-3}
891	21.44	0.21	22.69	0.20	2.11×10^{-9}	4.08×10^{-5}
1303	23.55	0.23	24.95	0.20	2.88×10^{-10}	2.31×10^{-5}
1358	17.52	0.22	18.45	0.22	7.37×10^{-6}	0.02
1443	26.50	0.20	28.30	0.20	0	0.02
1468	21.91	0.21	23.69	0.22	0	5.19×10^{-6}
1684	24.71	0.21	25.95	0.22	1.15×10^{-9}	1.10×10^{-3}
2171	25.35	0.22	26.20	0.21	6.07×10^{-5}	0.02
2483	19.29	0.23	20.45	0.22	1.99×10^{-7}	1.80×10^{-4}
3015	25.45	0.23	26.54	0.22	1.27×10^{-6}	0.03
3023	23.99	0.23	25.19	0.22	6.21×10^{-8}	5.93×10^{-5}
3668	25.89	0.21	27.22	0.19	4.49×10^{-11}	3.30×10^{-3}
3946	24.02	0.23	26.03	0.21	0	1.49×10^{-5}
4466	21.97	0.20	22.64	0.21	4.20×10^{-4}	9.95×10^{-4}
5397	24.37	0.21	25.88	0.22	9.09×10^{-14}	9.30×10^{-5}
1346	15.43	0.20	17.14	0.20	0	4.45×10^{-6}
2115	18.38	0.14	19.46	0.21	3.89×10^{-15}	2.00×10^{-5}
3093	24.57	0.22	25.78	0.20	1.02×10^{-8}	2.64×10^{-4}
232	25.24	0.20	26.39	0.21	2.18×10^{-9}	4.89×10^{-4}
953	22.66	0.21	24.07	0.19	1.78×10^{-11}	2.31×10^{-5}
2559	23.85	0.19	25.65	0.22	0	2.74×10^{-7}
3126	25.33	0.22	26.68	0.21	3.27×10^{-10}	8.86×10^{-4}
3366	25.34	0.18	27.05	0.20	0	8.13×10^{-5}
3761	25.66	0.20	27.39	0.22	0	3.83×10^{-4}
5440	24.97	0.20	27.04	0.20	0	1.06×10^{-4}
164	25.01	0.22	25.70	0.21	7.75×10^{-4}	0.02
2667	25.15	0.23	26.56	0.22	1.73×10^{-10}	9.30×10^{-5}

Table A3. MLE nonstationary models for temperature maxima. scale σ , shape k , location $\mu(t) = \beta_0 + \beta_1 \times t$, and likelihood D parameters. Here, t is 1 week.

Station	σ	k	β_0	β_1	D
3987	4.01	−0.30	25.07	2.88×10^{-3}	20.50
5100	4.08	−0.31	24.18	4.60×10^{-3}	51.26
1580	3.76	−0.33	25.23	3.67×10^{-3}	39.48
867	3.98	−0.35	24.58	2.83×10^{-3}	22.39
2290	3.48	−0.31	21.05	3.22×10^{-3}	34.39
4271	3.88	−0.21	21.29	3.52×10^{-3}	32.38
656	3.89	−0.33	20.01	3.71×10^{-3}	33.66
1550	3.49	−0.29	23.60	3.63×10^{-3}	42.84
1691	3.86	−0.30	24.29	2.84×10^{-3}	20.83
2261	3.92	−0.34	22.08	4.0×10^{-3}	44.08
2597	3.95	−0.33	24.87	3.25×10^{-3}	28.22
2601	3.85	−0.33	19.19	3.73×10^{-3}	38.37
3730	3.43	−0.32	23.53	3.03×10^{-3}	32.29
4104	3.53	−0.27	24.78	4.90×10^{-3}	71.25
4887	3.62	−0.31	21.13	4.37×10^{-3}	58.74
5371	3.98	−0.32	18.22	5.29×10^{-3}	73.61
5705	3.86	−0.31	24.87	4.08×10^{-3}	45.83
5906	3.74	−0.33	26.06	4.08×10^{-3}	47.58
691	3.92	−0.28	23.46	2.42×10^{-3}	14.81
701	3.94	−0.26	21.75	3.22×10^{-3}	25.06
1975	3.97	−0.26	22.73	2.90×10^{-3}	19.94
2014	3.84	−0.29	23.72	3.10×10^{-3}	25.33
3631	3.53	−0.14	19.92	2.32×10^{-3}	15.24
1639	3.86	−0.33	24.48	3.64×10^{-3}	36.06
4371	3.85	−0.29	23.80	2.34×10^{-3}	14.36
4625	3.92	−0.26	22.74	2.99×10^{-3}	21.96
5792	3.61	−0.36	6.62	2.14×10^{-3}	17.54
662	3.82	−0.28	23.66	3.22×10^{-3}	27.30
722	3.98	−0.32	14.46	4.45×10^{-3}	49.22
853	3.89	−0.33	23.39	2.88×10^{-3}	23.40
891	3.70	−0.22	20.91	2.90×10^{-3}	21.74
1303	3.92	−0.29	22.92	3.51×10^{-3}	31.13
1358	3.85	−0.30	17.13	2.28×10^{-3}	14.43
1443	3.65	−0.30	25.89	3.95×10^{-3}	45.18
1468	3.85	−0.33	21.33	3.78×10^{-3}	39.97
1684	3.69	−0.32	24.13	3.19×10^{-3}	30.79
2171	3.94	−0.33	24.96	2.14×10^{-3}	13.08
2483	3.99	−0.35	18.70	3.04×10^{-3}	24.80
3015	3.98	−0.31	25.03	2.51×10^{-3}	16.22
3023	3.90	−0.29	23.25	3.56×10^{-3}	31.53
3668	3.85	−0.31	25.44	2.90×10^{-3}	23.66
3946	4.04	−0.35	23.36	4.48×10^{-3}	55.51
4466	3.48	−0.24	21.38	2.41×10^{-3}	17.58
5397	3.77	−0.23	23.65	3.81×10^{-3}	35.75
1346	3.74	−0.33	14.84	3.70×10^{-3}	41.42
2115	2.58	−0.23	17.99	2.46×10^{-3}	32.70
3093	3.89	−0.28	23.98	3.12×10^{-3}	24.72
232	3.61	−0.30	24.94	2.22×10^{-3}	15.29
953	3.84	−0.31	22.07	3.4×10^{-3}	32.02
2559	3.56	−0.33	23.17	4.07×10^{-3}	53.63
3126	3.87	−0.31	24.62	3.58×10^{-3}	34.25
3366	3.39	−0.29	24.83	3.48×10^{-3}	40.97
3761	3.83	−0.34	24.92	4.31×10^{-3}	53.83
5440	3.72	−0.29	24.39	4.20×10^{-3}	48.65
164	3.82	−0.29	24.54	2.11×10^{-3}	11.69
2667	3.94	−0.29	24.55	3.42×10^{-3}	29.72

Code and data availability. The hourly temperature and precipitation data sets and the daily precipitation data set are freely available from the DWD CDC website: https://www.dwd.de/EN/climate_environment/cdc/cdc_node.html (DWD CDC, 2018a, b, 2019). The code used to perform the methods outlined in the paper was written in MATLAB and R using a combination of pre-built and newly created functions. For access to this code, please contact the corresponding author.

Author contributions. The majority of the work for this study was performed by MC with heavy guidance from HK. Both authors worked closely together on this analysis and held regular discussions on all aspects of this paper.

Competing interests. The authors declare that they have no conflict of interest.

Acknowledgements. The authors thank Kevin Bassler for helpful discussions and the Deutscher Wetterdienst Climate Data Center for the hourly and daily data sets used in this analysis. The authors are also grateful to Jakob Zscheischler and the anonymous reviewer for their helpful comments and suggestions.

Review statement. This paper was edited by William Hsieh and reviewed by Jakob Zscheischler and one anonymous referee.

References

- Azencott, R., Muravina, V., Hekmati, R., Zhang, W., and Paldino, M.: Automatic clustering in large sets of time series, *Contributions to Partial Differential Equations and Applications*, Comput. Meth. Appl. Sci., 47 65–75, 2018.
- Bador, M., Naveau, P., Gilleland, E., Castellá, M., and Arivelo, T.: Spacial clustering of summer temperature maxima from the CNRM-CM5 climate model ensembles and E-OBS over Europe, *Weather and Climate Extremes*, 9, 17–24, <https://doi.org/10.1016/j.wace.2015.05.003>, 2015.
- Balvanera, P., Siddique, I., Dee, L., Paquette, A., Isbell, F., Gonzalez, A., Byrnes, J., O'Connor, M., Hungate, B., and Griffin, J.: Linking biodiversity and ecosystem services: current uncertainties and the necessary next steps, *BioScience*, 64, 49–57, 2013.
- Bartos, M., Chester, M., Johnson, N., Gorman, B., Eisenberg, D., Linkov, I., and Bates, M.: Impacts of rising air temperatures on electric transmission ampacity and peak electricity load in the United States Environ. Res. Lett., 11, 114008, <https://doi.org/10.1088/1748-9326/11/11/114008>, 2016.
- Bellard, C., Bertelsmeier, C., Leadley, P., Thuiller, W., and Courchamp, F.: Impacts of climate change on the future of biodiversity, *Ecol. Lett.*, 15 367–377, 2012.
- Brown, S. J., Caesar, J., and Ferro, C. A. T.: Global changes in extreme daily temperature since 1950, *J. Geophys. Res.*, 113, D05115, <https://doi.org/10.1029/2006JD008091>, 2008.
- Burillo, D.: Effects of Climate Change in Electric Power Infrastructures, *Power System Stability*, IntechOpen, 5, <https://doi.org/10.5772/intechopen.82146>, 2019.
- Carney, M., Azencott, R., and Nicol, M.: Nonstationarity of summer temperature extremes in Texas, *Int. J. Climatol.*, 40, 620–640, <https://doi.org/10.1002/joc.6212>, 2019.
- Caroni, C. and Panagoulia, D.: Non-stationary modelling of extreme temperatures in a mountainous area of Greece, *REVSTAT*, 14, 217–228, 2016.
- Chen, T., Singh, P., Bassler, K.: Network community detection using modularity density measures, *J. Stat. Mech.*, 2018, 053406, <https://doi.org/10.1088/1742-5468/aabfc8>, 2018.
- Cheng, L., AghaKouchak, A., Gilleland, E., and Katz, R. W.: Non-stationary Extreme Value Analysis in a Changing Climate, *Climatic Change*, 2, 353–369, <https://doi.org/10.1007/s10584-014-1254-5>, 2014.
- Christidis, N., Stott, P., and Brown, S.: The role of human activity in the recent warming of extreme warm daytime temperatures, *J. Climate*, 24, 1922–1930, 2011.
- Christidis, N., Jones, G. S., and Stott, P. A.: Dramatically increasing chance of extremely hot summers since the 2003 European heatwave, *Nat. Clim. Change*, 5, 46–50, 2015.
- Coles, S.: *An Introduction to Statistical Modeling of Extreme Values*, Springer, 2001.
- Cover, T. and Thomas, J.: *Elements of Information Theory*, John Wiley and Sons, 1991.
- Dhillon, I., Guan, Y., and Kulis, B.: Kernel k-means, spectral clustering and normalized cuts, *Conference on Knowledge Discovery and Data Mining (KDD)* 22–25 August, Seattle, Washington, USA, 2004.
- Duch, J. and Arenas, A.: Community detection in complex networks using extremal optimization, *Phys. Rev. E*, 72, 027104, <https://doi.org/10.1103/PhysRevE.72.027104>, 2005.
- DWD Climate Data Center (CDC): Historical hourly station observations of 2m air temperature and humidity for Germany, version v006, available at: https://opendata.dwd.de/climate_environment/CDC/observations_germany/climate/hourly/air_temperature/historical/ (last access: 13 June 2019), 2018a.
- DWD Climate Data Center (CDC): Historical hourly station observations of precipitation for Germany, version v006, available at: https://opendata.dwd.de/climate_environment/CDC/observations_germany/climate/hourly/precipitation/historical/ (last access: 26 June 2019), 2018b.
- DWD Climate Data Center (CDC): Historical daily precipitation observations for Germany, version v007, available at: https://opendata.dwd.de/climate_environment/CDC/observations_germany/climate/daily/more_precip/historical/, last access: 5 August 2019.
- Erkwrzel, B., Boneham, J., Dalton, M. W., Heede, R., Mera, R. J., Allen, M. R., and Frumhoff, P. C.: The rise in global atmospheric CO₂, surface temperature, and sea level from emissions traced to major carbon producers, *Climatic Change*, 144, 579–590, 2017.
- Finkel, J. and Katz, J. I.: Changing world extreme temperature statistics, *Int. J. Climatol.*, 38, 2613–2617, 2018.
- Gao, M. and Zheng, H.: Nonstationary extreme value analysis of temperature extremes in China, *Stoch. Env. Res. Risk A.*, 32, 1299–1315, 2018.

- Gasparri, A., Guo, Y., Sera, F., Vicedo-Cabrera, A. M., Huber, V., Tong, S., Coelho, M. S. Z. S., Saldiva, P. H. N., Lavigne, E., Correa, P. M., Ortega, N. V., Kan, H., Osorio, S., Kyselý, J., Urban, A., Jaakkola, J. J. K., Rytí, N. R. I., Pascal, M., Goodman, P. G., Zeka, A., Michelozzi, P., Scortichini, M., Hashizume, M., Honda, Y., Hurtado-Díaz, M., Cruz, J. C., Seposo, X., Kim, H., Tobias, A., Iñiguez, C., Forsberg, B., Åström, D. O., Ragettli, M. S., Guo, Y. L., Wu, C., Zanobetti, A., Schwartz, J., Bell, M. L., Dang, T. N., Van, D. D., Heaviside, C., Vardoulakis, S., Hajat, S., Haines, A., and Armstrong, B.: Projections of temperature-related excess mortality under climate change scenarios, *The Lancet Planetary Health*, 1, E360–E367, 2017.
- Guimera, R. and Amaral, L. N.: Functional cartography of complex metabolic networks, *Nature*, 433, 895–900, 2005.
- Hamdi, Y., Duluc, C. M., and Rebour, V.: Temperature extremes: estimation of non-stationary return levels and associated uncertainties, *Atmosphere*, 9, 129, <https://doi.org/10.3390/atmos9040129>, 2018.
- Hansen, J., Ruedy, M., Sato, M., and Lo, K.: Global surface temperature change, *Rev. Geophys.*, 48, RG4004, <https://doi.org/10.1029/2010RG000345>, 2010.
- Hansen, J., Sato, M., and Ruedy, R.: Perception of climate change, *P. Natl. Acad. Sci. USA*, 109, E2415–E2423, 2012.
- Hasan, H., Salam, N., and Adam, M.: Modelling extreme temperature in Malaysia using generalized extreme value distribution, *Int. J. Math. Comput. Sc.*, 7, <https://doi.org/10.5281/zenodo.1080273>, 2013.
- Hastie, T., Tibshirani, R., Friedman, J.: *The Elements of Statistical Learning: Data Mining, Inference and Prediction*, 2nd Edn., Springer, 2008.
- Hosking, J. R. M.: L-moments: Analysis and Estimation of Distributions using Linear Combinations of Order Statistics, *J. R. Stat. Soc. B.*, 52, 105–124, 1990.
- Hosking, J. R. M. and Wallis, J. R.: *Regional Frequency Analysis: An approach based on L-moments*, Cambridge University Press, 1997.
- Kaspar, F., Müller-Westermeier, G., Penda, E., Mächel, H., Zimmermann, K., Kaiser-Weiss, A., and Deutschländer, T.: Monitoring of climate change in Germany – data, products and services of Germany's National Climate Data Centre, *Adv. Sci. Res.*, 10, 99–106, <https://doi.org/10.5194/asr-10-99-2013>, 2013.
- Kosanic, A., Kavcic, I., van Kleunen, M., and Harrison, S.: Climate change and climate change velocity analysis across Germany, *Sci. Rep.*, 9, <https://doi.org/10.1038/s41598-019-38720-6>, 2019.
- Kyselý, J.: Temporal fluctuations in heat waves at Prague–Klementinum, the Czech Republic, from 1901–97, and their relationships to atmospheric circulation, *Int. J. Climatol.*, 22, 33–50, 2002.
- Li, Z., Zhang, S., Wang, R.-S., Zhang, X.-S., and Chen, L.: Quantitative function for community detection, *Phys. Rev. E*, 77, 036109, <https://doi.org/10.1103/PhysRevE.77.036109>, 2008.
- Lucarini, V., Faranda, D., Freitas, A. C. M., Freitas, J., Tobias, K., Holland, M., Nicol, M., Todd, M., and Vienti, S.: *Extremes and Recurrence in Dynamical Systems*, Wiley, 2016.
- Luxburg, U.: A Tutorial on Spectral Clustering, *Stat. Comput.*, 17, 395–416, <https://doi.org/10.1007/s11222-007-9033-z>, 2007.
- Newman, M. E. J.: Modularity and community structure in networks, *P. Natl. Acad. Sci. USA*, 103, 8577–8582, 2006.
- Pörtner, H.-O., Roberts, D. C., Masson-Delmotte, V., Zhai, P., Tignor, M., Poloczanska, E., Mintenbeck, K., Nicolai, M., Okem, A., Petzold, J., Rama, B., and Weyer, N.: IPCC, 2019: Summary for Policymakers, IPCC Special Report on the Ocean and Cryosphere in a Changing Climate, in press, 2020.
- Rahmstorf, S. and Coumou, D.: Increase of extreme events in a warming world, *P. Natl. Acad. Sci. USA*, 108, 17905–17909, 2011.
- Roulston, M.: Estimating the errors on measured entropy and mutual information, *Physica D*, 125, 285–294, 1999.
- Schär, C., Vidale, P., Lüthi, D., Frei, C., Häberli, C., Liniger, M., and Appenzeller, C.: The role of increasing temperature variability in European summer heatwaves, *Nature*, 427, 332–336, 2004.
- Schölkopf, B., Smola, A., and Müller, K.-R.: Nonlinear component analysis as a kernel eigenvalue problem, *Neural Comput.*, 10, 1299–1319, 1998.
- Sun, Y., Danila, B., Josić, K., and Bassler, K.: Improved community structure detection using modified fine-tuning strategy, *Europhysics Letters Association*, 86, 2, <https://doi.org/10.1209/0295-5075/86/28004>, 2009.
- Thornton, H. E., Hoskins, B. J., and Scaife, A. A.: The role of temperature in the variability and extremes of electricity and gas demand in Great Britain, *Environ. Res. Lett.*, 11, 11, <https://doi.org/10.1088/1748-9326/11/11/114015>, 2016.
- Tomczyk, A. and Sulikowska, A.: Heat waves in lowland Germany and their circulation-related conditions, *Meteorology and Atmospheric Physics*, 130, 499–515, 2018.
- Treviño III, S., Sun, Y., Cooper, T., and Bassler, K.: Robust detection of hierarchical communities from *Escherichia coli* gene expression data, *PLOS Comput. Biol.*, 8, E1002391, <https://doi.org/10.1371/journal.pcbi.1002391>, 2012.
- Wang, S., Gittens, A., and Mahoney, M.: Scalable Kernel K-Means Clustering with Nyström Approximation: Relative-Error Bounds, *J. Mach. Learn. Res.*, 20, 1–49, 2019.
- Wergen, G. and Krug, J.: Record-breaking temperatures reveal a warming climate, *Europhys. Lett.*, 92, 3, <https://doi.org/10.1209/0295-5075/92/30008>, 2010.
- Yoon, P., Kim, T.-W., and Yoo, C.: Rainfall frequency analysis using a mixed GEV distribution: a case study for annual maximum rainfalls in South Korea, *Stoch. Env. Res. Risk A.*, 27, 1143–1153, 2013.
- Yuan, C. and Yang, H.: Research on K-Value Selection Method of K-Means Clustering Algorithm, *Multidisciplinary Scientific Journal*, 2, 226–235, <https://doi.org/10.3390/j2020016>, 2019.
- Zhang, W., Muravina, V., Azencott, R., Chu, Z., and Paldino, M.: Mutual information better quantifies brain network architecture in children with epilepsy, *Comput. Math. Methods M.*, 2018, 6142898, <https://doi.org/10.1155/2018/6142898>, 2018.
- Zscheischler, J., Mahecha, M., and Harmeling, S.: Climate Classifications: the Value of Unsupervised Clustering, *Procedia Comput. Sci.*, 9, 897–906, 2012.