Contents lists available at ScienceDirect

Weather and Climate Extremes

journal homepage: www.elsevier.com/locate/wace

Extreme heat events in the Iberia Peninsula from extreme value mixture modeling of ERA5-Land air temperature

Susana Barbosa [a,*], Manuel G. Scotto [b]

[a] INESC TEC - INESC Technology and Science, Porto, Portugal
[b] CEMAT and Department of Mathematics, IST University of Lisbon, Lisbon, Portugal

A R T I C L E I N F O

Keywords:
Air temperature
Iberia
ERA5-Land
Extreme value mixture models

A B S T R A C T

Extreme summer temperatures in the Iberia Peninsula are analyzed from ERA5-Land reanalysis data based on an extreme value mixture model combining a Normal distribution for the bulk distribution (i.e. for the non-extreme values) and a Generalized Pareto Distribution for the extremes in the upper tail. This approach allows to treat the threshold of temperature exceedances as being one of model parameters rather than fixed a priori, enabling to take into account its corresponding uncertainty. Extreme value mixture models are estimated individually for each location, and the analysis is performed separately for two distinct periods, namely from 1981 to 2000 and from 2000 to 2019, respectively. The results show significant differences in the extreme value mixture models for the two periods, and in their corresponding 20-year return levels. The mean of the bulk distribution of summer maximum temperature increases significantly, particularly in Eastern Iberia. The largest differences in the tails of the data distribution between the two periods occur in the eastern Mediterranean area, and are characterized by a significant increase in the threshold for temperature exceedances and in their corresponding return levels.

1. Introduction

Excessive heat events are one of the largest direct cause of increased weather-related mortality (Lee, 2014), with extreme high temperatures being associated with a near universal increase in human mortality and morbidity (Sheridan and Allen, 2015). Despite the decreased sensitivity to heat in the developed world, associated with increased awareness and improvement in life conditions (e.g. Sheridan and Allen, 2018 and DeCastro et al., 2011), climate change is expected to exacerbate extreme high temperatures. Furthermore, demographic changes lead to increasingly elderly populations, the most vulnerable to extreme heat events. In addition to health concerns, extreme temperatures are also associated with important economic impacts, for example in terms of agriculture yields (Sun et al., 2019) or electricity demand (Hyndman and Fan, 2009).

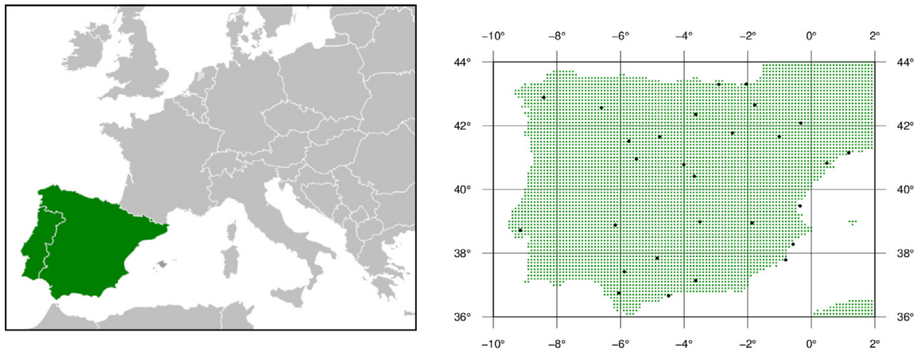
The Iberian Peninsula is traditionally exposed to extreme heat events, but climate change has the potential to further intensify already typically high temperatures. Climate projections indicate an increase in the frequency and intensity of extreme heat events associated with the increase in global mean temperature (e.g. Cardoso et al., 2019; Molina et al., 2020; Suarez-Gutierrez et al., 2020; Carvalho et al., 2020 and Lorenzo et al., 2021). Lisboa and Madrid are among the most vulnerable European capitals to extreme heat in the future (Smid et al., 2019). This study aims to describe extreme temperature events in the Iberia Peninsula.

Non-stationarity is a fairly common feature of climate records (e.g. Milly et al., 2008 and Slater et al., 2020), and the analysis of extreme temperatures in a non-stationary framework often considers changes in time of indices or percentiles of the data distribution within a given period, such as the warmest/coolest day (e.g. Ramos et al., 2011; Fernández-Montes and Rodrigo, 2012 and Donat et al., 2013). Traditional methods for the assessment of extreme events are based on the assumption of long-term stationarity, enabling the use of past information to infer the risk of future extreme events. An alternative enabling the extrapolation to extremes outside the observation record is to parameterize the tail of the data distribution by considering a GEV (Generalized Extreme Value) distribution for block extremes (e.g. Kharin et al., 2018) or GPD (Generalized Pareto) distribution for excesses above a high threshold (e.g. Katz et al., 2002; Lucio et al., 2010). Non-stationary is then taken into consideration by estimating trends in the parameters of the extreme value distributions (e.g. Parey et al., 2007; Brown et al., 2008) or by standardizing the variable in order to produce extremes which are approximately stationary (Parey et al., 2019).

Although the modeling of excesses above a threshold has the advantage of making a more efficient use of the available data, its main

* Corresponding author.
E-mail addresses: susana.a.barbosa@inesctec.pt (S. Barbosa), manuel.scotto@tecnico.ulisboa.pt (M.G. Scotto).

https://doi.org/10.1016/j.wace.2022.100448
Received 11 June 2021; Received in revised form 27 January 2022; Accepted 22 April 2022



(a) (b)

Fig. 1. Study area: (a) geographic location of the Iberia Peninsula (license CC0 1.0); (b) gridpoints of air temperature time series over the Iberia peninsula and location of individual stations (•).

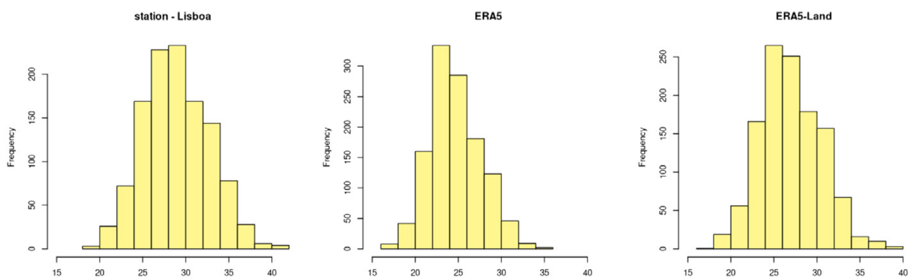


Fig. 2. Empirical distribution of summer temperature (declustered) maxima for the station of Lisboa, and the corresponding distribution for the nearest ERA5 and ERA5-Land gridpoints.

challenge is the selection of such threshold, which needs to be set high enough to describe a relevant extreme event while low enough to ensure an adequate sample size of extreme values. It is important to stress that most approaches devised for threshold selection do not account for the uncertainty associated with the threshold choice. This limitation is overcome by extreme value mixture models, which treat the threshold as being an unknown parameter which has to be estimated, enabling the uncertainty on the threshold to be taken into account in subsequent inferences (e.g. Behrens et al., 2004 and Scarrott and MacDonald, 2012; Scarrott et al., 2016). Furthermore, in contrast with extreme value models, mixture models do not ignore the non-extremal data, but rather combine in the same modeling framework a GPD threshold model for the extreme values with a model for the bulk distribution, thus allowing to characterize the full data distribution.

In this paper, the complete distribution of extreme summer daily temperatures in the Iberia Peninsula is analyzed by means of an extreme value mixture model applied to homogeneous air temperature time series on a regular spatial grid. The rest of the paper is organized as follows: data is described in Section 2 and the modeling approach is detailed in Section 3. The results are presented in Section 4 and the conclusions of the study are given in Section 5.

2. Data

Temperature extremes in Iberia are assessed from Iberia reanalysis data, resulting from the combination of models and observations, as it allows to overcome the limitations in terms of length and spatial coverage of available station records. Over North America, the comparison of atmospheric reanalysis and surface observations in terms of extreme temperatures indicates that reanalysis data are broadly able to replicate the spatial pattern of extreme temperature events in terms of their climatological frequency, with the ERA5-Land reanalysis having in general a lower match with station observations than the ERA5 reanalysis (Sheridan et al., 2020). For the Iberia, ERA5 (Hersbach et al., 2019) and ERA5-Land reanalysis (Muñoz Sabater, 2019) are here compared with station data from the European Climate Assessment (ECA&D) dataset. (Klein Tank et al., 2002).

2.1. Comparison of reanalysis and station data

The ERA5 reanalysis is available on a $0.25 \times 0.25°$ grid, while the ERA5-Land is available on a higher resolution $0.1 \times 0.1°$ grid. Fig. 1 shows the ERA5-land grid and the location of individual stations. Hourly reanalysis time series of air temperature at 2 m above the surface from January 1981 to December 2019 are obtained for the gridpoint closest to each station. As the focus of the analysis is on extreme heat events, and to avoid non-stationarity due to the seasonal cycle, only summer values from June to August (JJA) are considered. Furthermore, in order to reduce the serial dependence in daily temperature values, the reanalysis data are declustered by computing the maximum over clusters of 72 hourly values (3-days).

A total of 30 stations from the non blended ECA&D dataset with data available for the 1981–2019 period are considered. For consistency, the same approach is adopted as for the reanalysis data. Daily temperature time series of maximum temperature are considered only for the summer period and maximum values are taken over 3-days clusters.

Fig. 2 shows the data distribution for an individual station (Lisboa) and for the nearest ERA5 and ERA5-Land reanalysis gridpoints. In terms of range the higher resolution ERA5-Land distribution is closer to the station data distribution than the ERA5 reanalysis. The range and median values for all the individual stations and the corresponding reanalysis data are displayed in the Appendix (Table A.1).

Fig. 3 summarizes for all the stations the differences in terms of median and range between station and reanalysis summer temperature data. The differences are dominantly positive, indicating that the reanalysis temperature values are lower than the station ones. The largest difference between station and reanalysis summer maxima is found for a mountain station, Navacerrada, at 1894 m altitude. Differences are in general smaller for ERA5-Land than ERA5 data.

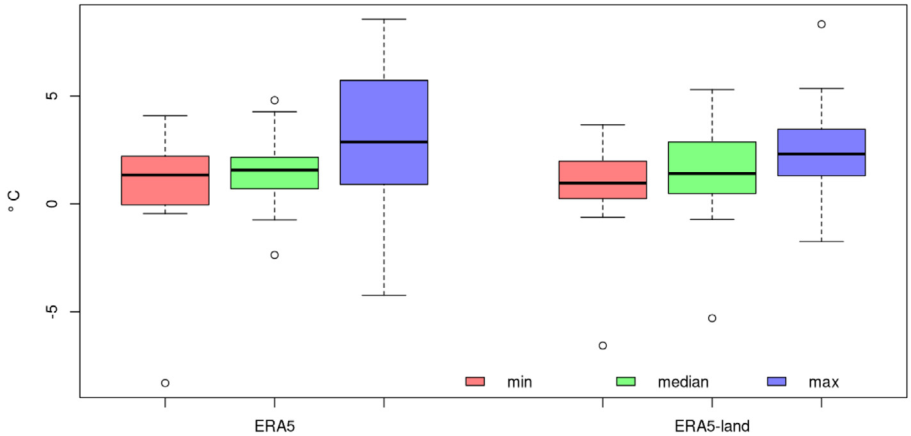


Fig. 3. Boxplots of differences in minimum, median and maximum values of summer temperatures between stations and nearest reanalysis time series. The ∘ indicates outlying values which correspond to the mountain station of Navacerrada.

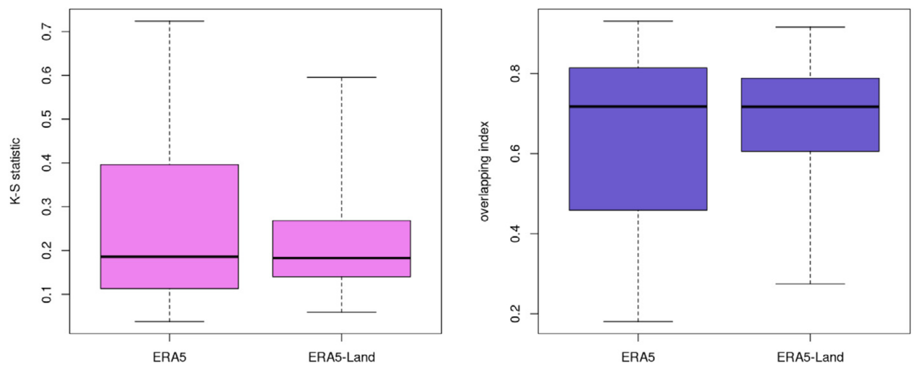


Fig. 4. Kolmogorov–Smirnov statistic (left) and overlapping index (right) for station and reanalysis empirical data distributions.

Further comparison of the empirical distributions of station and reanalysis data is carried out by computing the Kolmogorov–Smirnov distance (Arnold and Emerson, 2011) and the overlapping index (Pastore, 2018; Pastore and Calcagni, 2019). The Kolmogorov–Smirnov statistic reflects the distance between empirical cumulative distribution functions, while the overlapping index provides an estimate of the percentage of overlapping of the empirical distributions. The results are displayed in Table A.2 and summarized for all the stations in Fig. 4. On average ERA5 and ERA5-Land results are similar, but the distance to the station data is more frequently higher for ERA5 than ERA5-Land data. The overlapping index is also lower in the case of ERA5 data.

2.2. ERA5-Land reanalysis data

Taken into consideration the results of the previous section, the ERA5-Land reanalysis data is used for the study of summer temperature maxima in Iberia. The main advantage of ERA5-Land is its higher spatial resolution, corresponding to a total number of ~ 6000 gridpoints over the Iberian Peninsula (compared to ~ 1000 for ERA5).

Each one of these ~ 6000 hourly time series of declustered summer maxima is analyzed individually. As an illustration, Fig. 5a shows one of the time series (8.4°W, 41.6°N). The analysis focuses on the data distribution (displayed in Fig. 5(b)), thus ignoring the temporal dependence between observations (reflected in the autocorrelation and partial autocorrelation functions shown in Fig. 5c–d).

3. Models and methods

An extreme value mixture model combining a normal distribution for the bulk and a GPD distribution for the upper tail is fitted to the temperature data. The model takes the form (MacDonald et al., 2011):

$$F(x|\mu,\sigma,u,\sigma_u,\xi,\phi_u) = \begin{cases} \frac{(1-\phi_u)}{\mathcal{N}(u|\mu,\sigma)}\mathcal{N}(x|\mu,\sigma) & x \le u \\ (1-\phi_u) + \phi_u G(x|u,\sigma_u,\xi) & x > u, \end{cases} \tag{1}$$

where $\mathcal{N}(\mu,\sigma)$ is the bulk normal distribution with mean $\mu \in \mathbb{R}$ and standard deviation $\sigma > 0$. Furthermore, $G(u,\sigma_u,\xi)$ denotes the GPD distribution with threshold $u > 0$, scale parameter $\sigma_u > 0$ and shape parameter $\xi \in \mathbb{R}$. The parameter ϕ_u represents the tail fraction (threshold exceedance probability), $\phi_u = Pr(X > u)$. The tail fraction ϕ_u can be either treated as a model's parameter as indicated in Eq. (1), or directly obtained from the bulk model. In the more general parameterized tail fraction case ϕ_u is an explicit parameter affecting both bulk and tail estimates. In the bulk-based tail fraction case ϕ_u in Eq. (1) is replaced by the survival probability of the bulk model assuming it continues above the threshold, $\phi_u = 1 - \mathcal{N}(u|\mu,\sigma)$. This form has the advantage of using the information from the larger sample of bulk data to estimate the tail fraction, but it can make the tail fraction estimation susceptible to misspecification of the bulk model. Either specification provides a proper density, with the parameterized tail fraction approach including the bulk model based tail fraction approach as a particular case (Hu and Scarrott, 2018).

The return level q_t for a return period T, corresponding to the temperature exceeded on average once every T summer days is computed from the model parameters as

$$q_t = u + \frac{\sigma}{\xi}[(\phi_u T)^{\xi} - 1]. \tag{2}$$

The analysis is illustrated for the data displayed in Fig. 5. The parameters of an extreme value mixture model with normal distribution for the bulk and GPD distribution for the upper tail are estimated

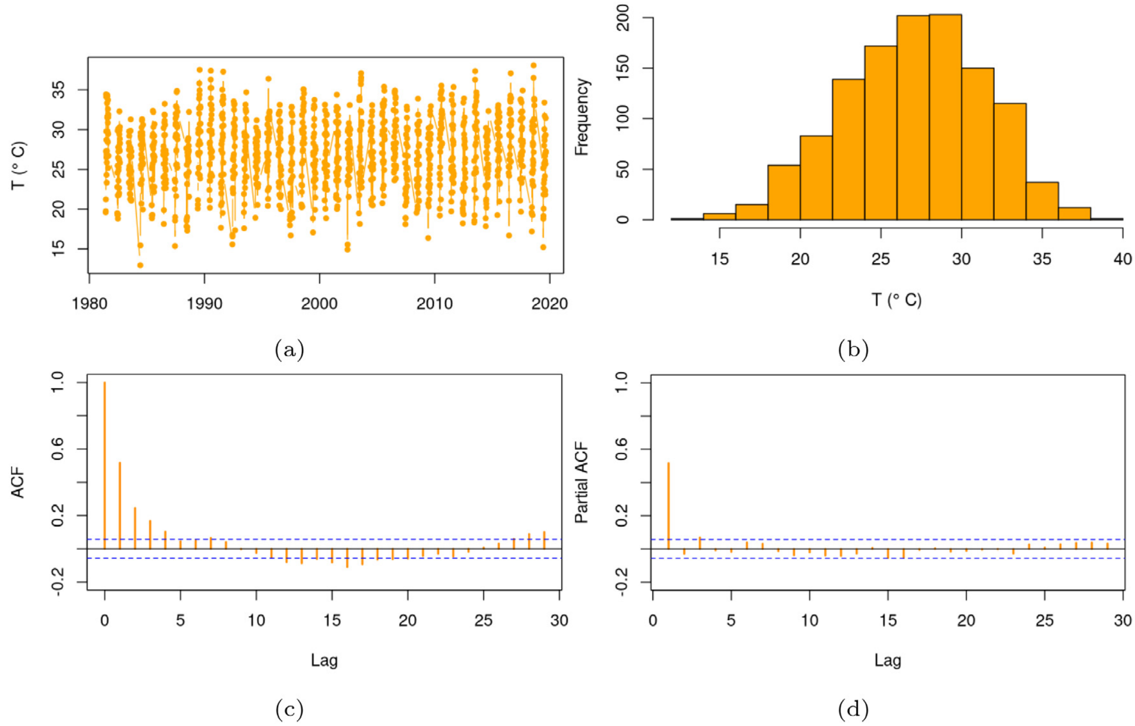


Fig. 5. (a) Observed time series (8.4°W, 41.6°N) of summer daily temperatures from the period 1980–2020. (b) Empirical marginal distribution. (c) sample autocorrelation function (ACF) and (d) sample partial ACF.

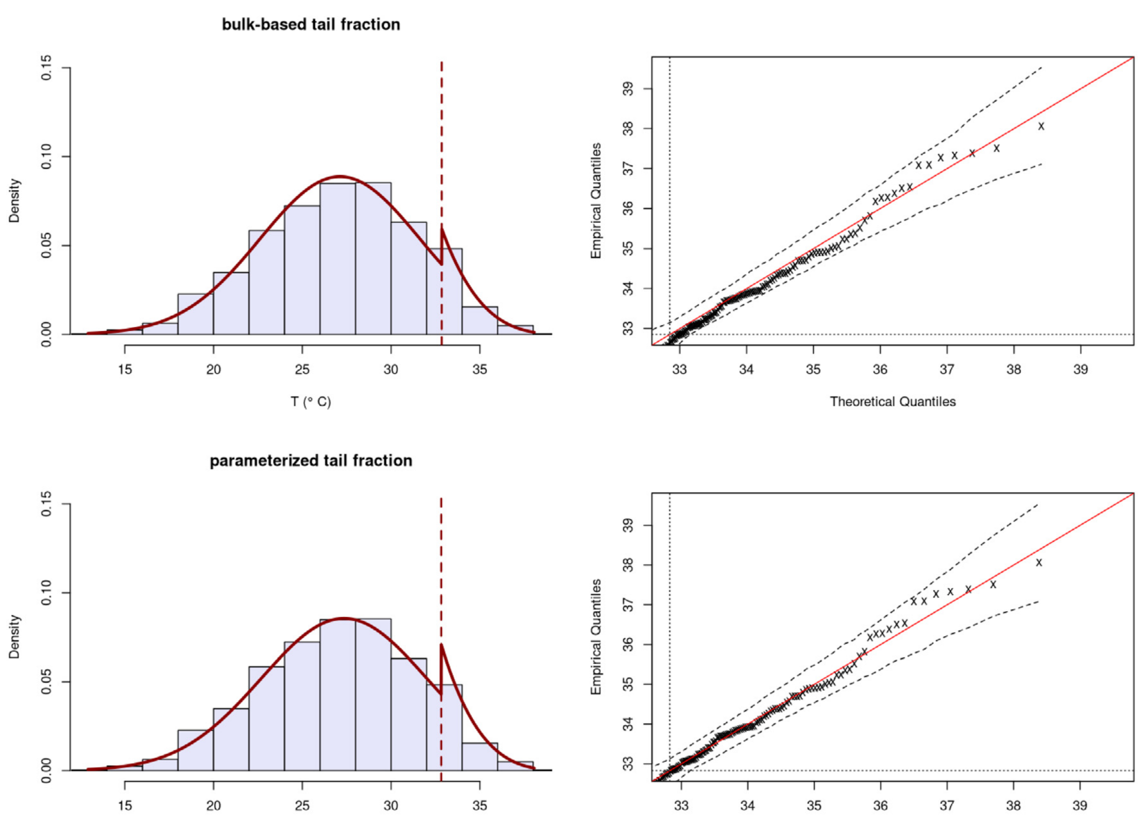


Fig. 6. Extreme value mixture model. Left: histogram of daily summer temperatures and fitted mixture model (solid curve); the vertical dashed line represents the estimated threshold u. Right: quantile–quantile plot of observed and modeled values; dashed lines represent 95% confidence intervals, and dotted lines the threshold u. The solid (red) line corresponds to the fitted extreme value mixture model..

by maximum likelihood using the R-package *evmix* (Hu and Scarrott, 2018). Two different possibilities are considered for the tail fraction ϕ_u, namely: it is either obtained directly from the bulk distribution (bulk-based tail fraction) or estimated as an additional parameter (parameterized tail fraction). The results are displayed in Fig. 6. The quantile–quantile plots (right) show that both estimated mixture models are an adequate representation of the upper tail data distribution. The mixture model fits (left) are represented by the curve superimposed on the data distribution. The vertical line indicates the estimated threshold u above which the GPD distribution is considered. The discontinuity at the threshold indicates that the GPD tail is heavier than the one that would be represented by a normal distribution. Although a constraint of continuity could have been included, the benefit is typically limited and can even be damaging in case of bulk model misspecification (Hu and Scarrott, 2018).

Furthermore, comparison of the upper and lower plots in Fig. 6 also shows that very similar results are obtained for the different tail fraction options. The parameterized tail fraction gives slightly better estimates, but the results are very similar for the highest quantiles. Although a parameterized tail fraction could provide a tail fraction estimate more robust to bulk model misspecification, the model estimation is based on maximum likelihood and the optimization can be sub-optimal due to local modes, leading to a poor model fit, even in the case of parameterization of the tail fraction. The assessment of model results for the data under consideration shows that the bulk-based tail fraction provides more robust results for intermediate quantiles, and very similar results for highest quantiles. Thus the strategy adopted in this study is to consider the bulk-based tail fraction option, obtaining the tail fraction from the normal survivor function. This enables the tail estimation to benefit from the larger sample of bulk data compared to the limited tail data. The model parameters obtained for the bulk-based tail fraction estimation are presented in Table 1.

In order to assess potential non-stationary behavior in time, the analysis is performed separately for two distinct periods, from 1981 to 2000, and from 2000 to 2019. The modeling results for the two different time periods indicate a reasonable good fit (Fig. 7), although confidence intervals can be wider, which is expected due to the smaller number of observations than for the complete time period. Difference in the extreme value mixture models for the two periods are described by the difference between estimated model parameters, with the uncertainty of the difference computed as the square root of the sum of squared uncertainties (Table 1). The estimated parameters of the bulk data distribution are very similar for the two periods, but the upper tail estimates are significantly different. The threshold for temperature exceedances is significantly higher in the 2000–2019 than in the 1981–2000 period and the GPD distribution is steeper.

Furthermore, return levels obtained from the extreme mixture models for each period (Fig. 8) reflect the significant difference in the upper tail parameters from each period. Return levels for the later period are higher and with a smaller return period than for the earlier period. A conservative approach to take into account the uncertainty in the threshold as well as in the remaining model parameters is to plug-in the lower and upper bound values of each parameter in Eq. (2), yielding corresponding lower and upper bounds for return levels. Table 2 shows the values corresponding to 20-, 50- and 100-year return periods, including uncertainty derived from the difference between upper and lower bounds. This approach yields much larger uncertainties than the simulated confidence intervals displayed in Fig. 8 as these are obtained from the estimated parameters by Monte Carlo simulation and ignore the parameter estimation uncertainty. The uncertainty increases with the return period, as expected, and indicates that in this case the differences in return levels between the two periods are not significant, despite their differences in the tails of the mixture model estimated for each period.

Table 1
Parameters of the extreme value mixture model with normal bulk distribution and GPD distribution for the upper tail with bulk-based tail fraction. The corresponding uncertainties are displayed in parentheses.

	μ	σ	u	σ_u	ξ	ϕ_u
1981–2019	27.12 (0.13)	4.50 (0.10)	32.85 (0.0032)	1.70 (0.24)	−0.21 (0.10)	0.10
1981–2000 (P1)	27.06 (0.19)	4.55 (0.15)	31.27 (0.0023)	2.08 (0.28)	−0.22 0.092	0.18
2000–2019 (P2)	27.24 (0.19)	4.55 (0.14)	33.06 (0.0021)	1.54 (0.31)	−0.16 (0.15)	0.10
P2-P1	0.18 (0.26)	0.0032 (0.21)	1.79 (0.0031)	−0.55 (0.42)	0.062 (0.17)	−0.08

Table 2
Return levels for 20-, 50- and 100-year return periods. The corresponding uncertainties are displayed in parentheses.

	$T = 20$	$T = 50$	$T = 100$
1981–2009 (P1)	33.6 (0.9)	34.9 (1.6)	35.7 (2.2)
2009–2019 (P2)	34.1 (0.5)	35.2 (1.4)	36.0 (2.2)
P2-P1	0.5 (1.0)	0.4 (2.1)	0.3 (3.1)

4. Results

Extreme value mixture models are estimated at every gridpoint shown in Fig. 1b. Individual models are fitted for the two periods considered before, 1981–2000 and 2000–2019. In general a bulk-based tail fraction model is implemented based on the rational described in Section 3. However in some cases the bulk distribution is not well represented by a normal distribution, as indicated by bulk-mean point estimates higher than the threshold. This occurs only for a few gridpoints (<1%) typically exhibiting a long right tail. In these few cases the parameterized tail fraction model is adopted as it provides more robust estimates. Furthermore, for five gridpoints a mixture model with normal bulk and GPD for the upper tail (either with bulk-based or parameterized tail fraction) is not an adequate representation of the data distribution, therefore results for those gridpoints are not presented. The modeling results are displayed as maps of the fitted model parameters in Section 4.1 and in terms of return levels in Section 4.2.

4.1. Extreme value mixture modeling

An extreme value mixture model with normal distribution for the bulk and GPD distribution for the upper tail is estimated for each gridpoint, and their corresponding parameter's estimates are presented as maps of the parameters at each point. The difference between the two periods is represented in the map if significant taking into account their corresponding uncertainties (as illustrated in Section 3), i.e. non-significant differences (within the computed uncertainty) are displayed as zero in the maps.

The mean and standard deviation of the normal bulk distribution estimated for each gridpoint are displayed in Figs. 9 and 10, respectively. The mean reflects the climatic regions of Iberia, with lower mean values of the bulk distribution in the Pyrenees and in the north of the Peninsula, and higher values in the southeast, particularly the Andalusian plain, with highest values in the Guadalquivir valley. The spatial pattern is similar for the two periods considered, but the map of differences (Fig. 9(c)) shows in general an increase in the mean of the distribution of maximum summer temperatures in Iberia. The largest increase occurs in southeastern Mediterranean region and the smallest increase in the north and western Iberia.

The estimated standard deviation of the normal bulk distribution is lower in coastal areas and higher at in-land locations. Differences

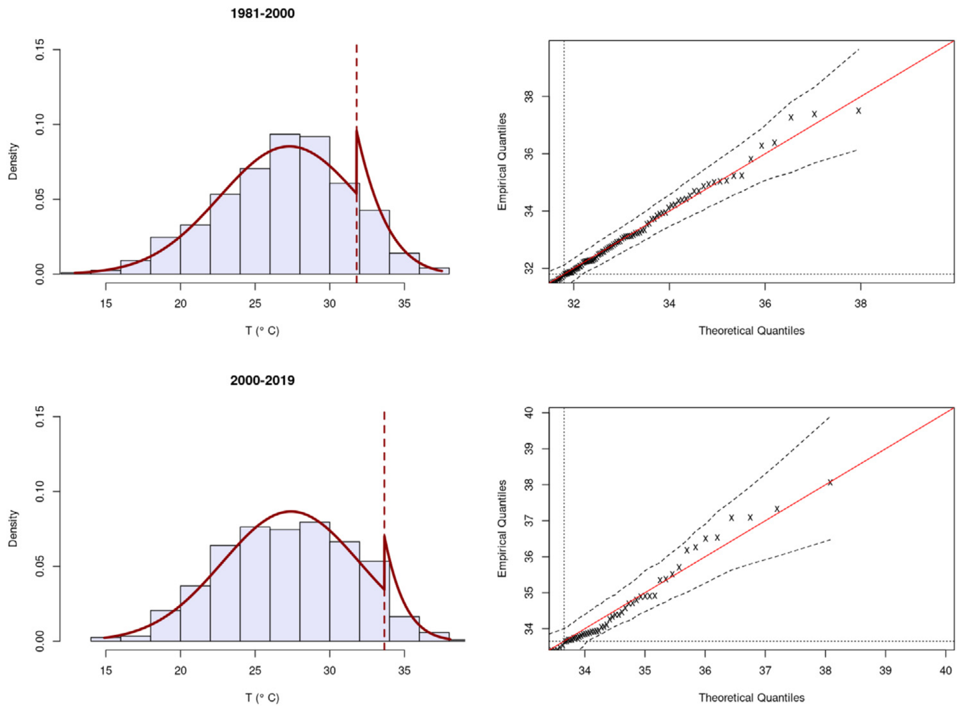


Fig. 7. Same as in Fig. 6 but for two distinct time periods: 1981–2000 (top) and 2000–2019 (bottom).

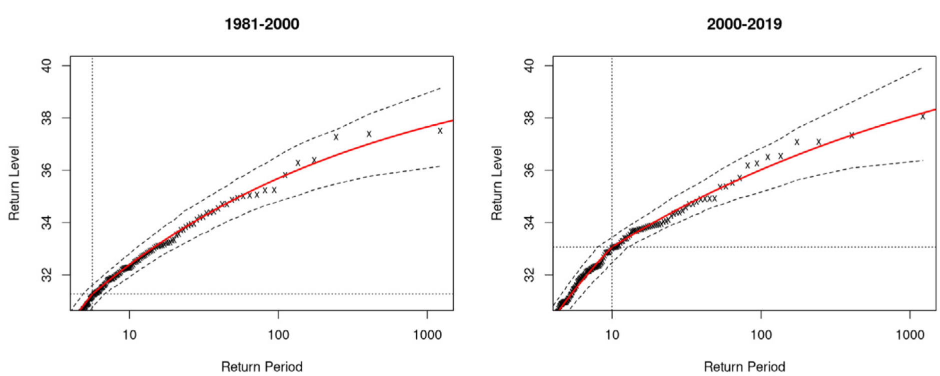
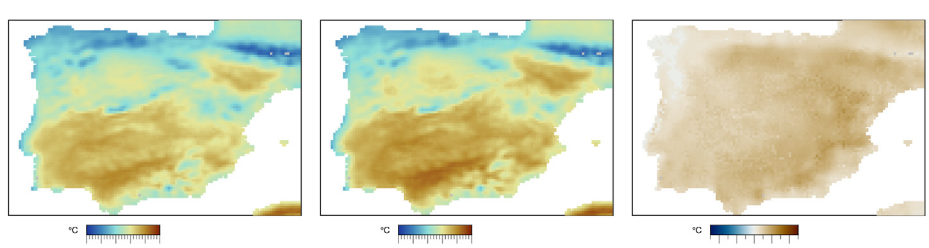


Fig. 8. Return levels from the estimated extreme value mixture model (solid curve) and simulated pointwise 95% confidence intervals (dashed curve); the dotted lines represent the threshold level u and the corresponding return period $1/\phi_u$.



(a) (b) (c)

Fig. 9. Estimated mean of the normal distribution model: (a) 1981–2000, (b) 2000–2019, (c) difference.

between the two analyzed periods are not significant except in the western coast and northeast Iberia regions in which the standard deviation increases slightly.

In terms of the parameters of the GPD distribution, the shape parameter ξ is predominantly negative (Fig. 11) in the two analyzed periods. Thus the estimated extreme value distribution is in general of the Weibull type, corresponding to a light tail with finite upper bound. The difference in the shape parameter between the two periods is small, typically representing a slight increase in localized areas of the northeast and southern Iberia.

Fig. 12 shows the estimated threshold value for temperature exceedances. The spatial pattern tends to be similar to the mean, with lowest values in the coldest regions of northern Spain and the Pyrenees. However the spatial pattern of the differences in the threshold values

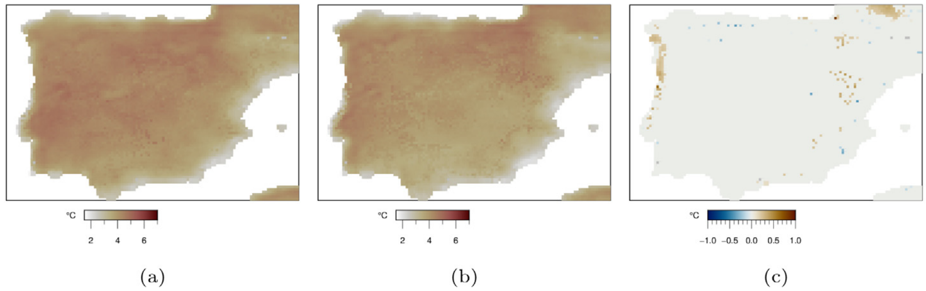


Fig. 10. Estimated standard deviation of the normal distribution model: (a) 1981–2000, (b) 2000–2019, (c) difference.

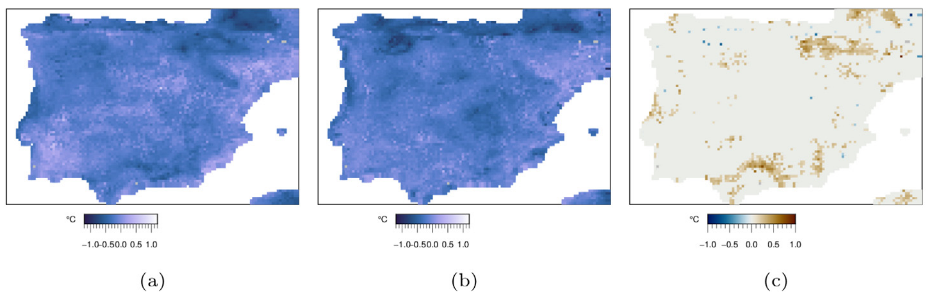


Fig. 11. Estimated shape parameter of the GPD distribution (a) 1981–2000, (b) 2000–2019, (c) difference.

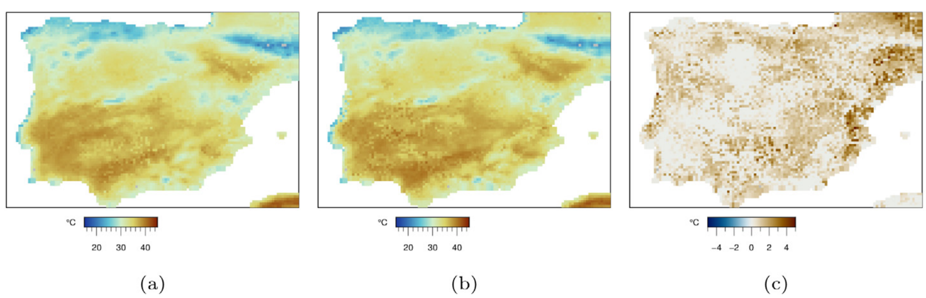


Fig. 12. Estimated threshold (u) of the Generalized Pareto distribution: (a) 1981–2000, (b) 2000–2019, (c) difference.

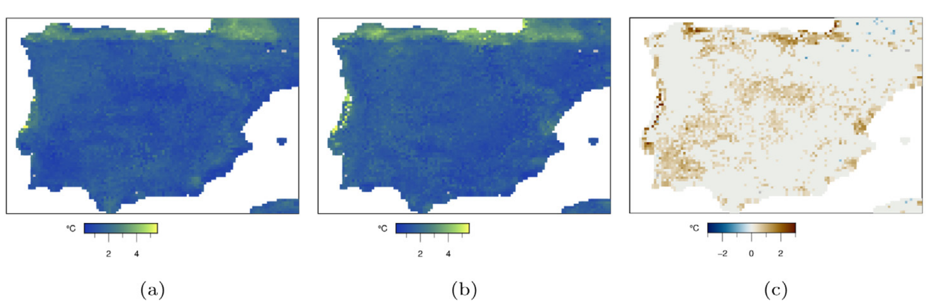


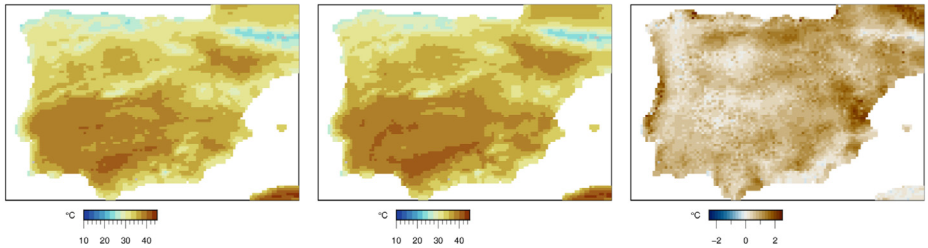
Fig. 13. Estimated standard deviation, σ_u, of the Generalized Pareto distribution: (a) 1981–2000, (b) 2000–2019, (c) difference.

is different than the one from the bulk mean, although it is also predominantly positive. Highest increases in the exceedance threshold occur in the southeastern Iberia area, particularly the upper Spanish Mediterranean coast.

The scale parameter of the GPD distribution, σ_u, displayed in Fig. 13, is highest in the northern and western coast of the Iberia peninsula, and increases there from the initial to the later period. A slight but significant increase is also found in the flat areas of south Portugal and central Iberia.

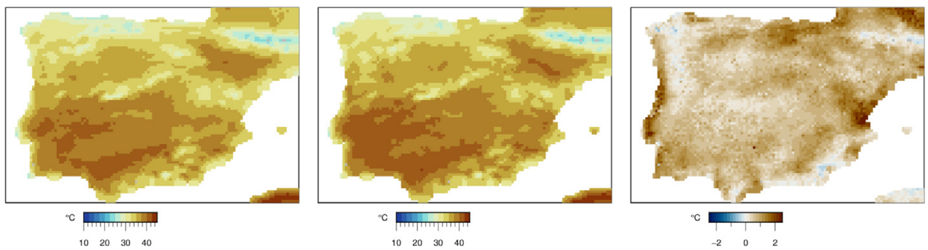
4.2. Return levels

Return levels corresponding to return periods of 20-, 50- and 100-year are show in Figs. 14, 15 and 16, respectively. The return levels computed for the later period are consistently higher than the corresponding return levels computed from the estimates based on the earlier period. The southern Iberia is the most affected area in terms of extreme temperatures. The spatial pattern is closely associated with the relief of the peninsula, with highest return levels in the low-lying plains,



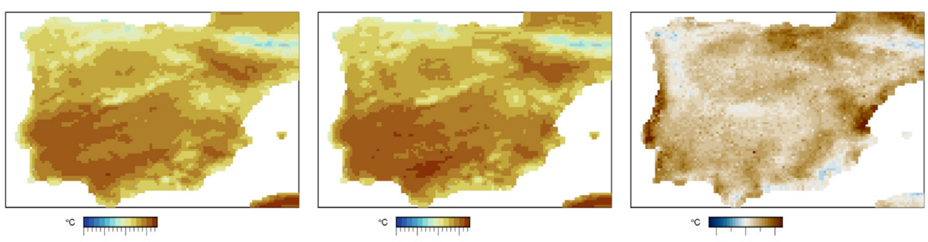
(a) (b) (c)

Fig. 14. Return levels corresponding to a return period of 20 years: (a) 1981–2000, (b) 2000–2019, (c) difference.



(a) (b) (c)

Fig. 15. Same as Fig. 14 but for a return period of 50 years.



(a) (b) (c)

Fig. 16. Same as Fig. 14 but for a return period of 100 years.

notably in the Guadalquivir and in the Ebro depressions, and lowest maximum temperatures in the mountainous areas of the Pyrenees and of the Cantabrian mountains in the North. Although the spatial pattern is consistent with the average of summer maximum temperatures, the pattern of return values shows very high summer temperatures in the southwestern Iberia (central and south regions of Portugal). The difference in return levels computed from the data distributions for the two periods is positive, showing largest increases in return levels computed from the two periods in the French Pyrenees area, Valencia Mediterranean coast and western Iberia coast.

Fig. 17 displays the differences in the return levels for the two different periods when taking into account the uncertainty in the threshold estimate as well as in the remaining model parameters, as illustrated in Section 3. With this conservative assessment the uncertainty on the return values is large enough to prevent the detection of any significant differences except for the lower 20-year return period.

5. Discussion and conclusions

In the present study extreme temperature events are analyzed focusing on declustered summer maxima from ERA5-Land reanalysis data. Temperature time series are considered individually, and an extreme value mixture model with normal bulk and GPD upper tail is estimated at each point. Although this approach has the disadvantage of ignoring the spatial dependency (e.g. Coelho et al., 2008) it allows the straightforward assessment of the complete data distribution based on existing tools for univariate data.

The comparison of reanalysis and station data in Section 2.1 indicates that the ERA5-Land reanalysis is as close as or closest to station data than ERA5 reanalysis. However, both reanalysis show differences to the station data, most obvious in the highest altitude station (Navacerrada). Trends in upper quantiles tend to be higher in mountain stations (Barbosa et al., 2011) emphasizing the importance of performing detailed study of extremes based on station rather than reanalysis data for high-altitude stations.

The short length of the considered time series is another limitation of this study. Despite the inevitable restricted size of the sample of extreme events in ∼40 years time series, the use of the ERA-5 Land reanalysis data allows to analyze homogeneous data on a regular grid and obtain a spatial perspective on extreme summer temperature over the Iberia domain. The sample size is further impacted by performing the analysis separately for two different periods, from 1981 to 2000 and from 2000 to 2019. This arbitrary separation in time attempts to somewhat handle eventual nonstationarities, by assuming a stationary

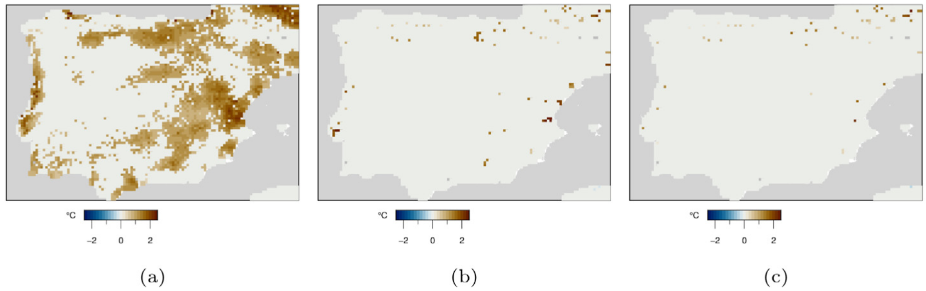


Fig. 17. Difference in return levels for the two periods (as in (a) Figs. 14c, (b) 15c and (c) 16c) but taking into account uncertainty from threshold and model parameters estimates.

behavior within each of the time blocks and evaluating the differences between the two periods. Although the modeling results are still reasonable for the smaller sample in each period, as illustrated in Section 3, ≲20 years is a too short period to encompass the dominant variability modes of the climate system. Here the extreme models derived from each period are compared from a statistical point of view. The physical explanation of eventual differences is not addressed in the current study.

The fit of extreme value mixture models can be hindered by local modes impacting the optimization results. Furthermore, the strategy for handling the tail fraction depends on the specific data distribution. Although graphical diagnostics are very useful and allow in general to select the optimal strategy, its use in a large study such as this one involving ~6000 time series is not feasible. The approach adopted here was to consider a bulk-based tail fraction, since the normal distribution is in general a good approximation for the bulk distribution of maximum summer temperatures. The assessment of the fits obtained showed however that this was not the case for every data distribution and in some cases (≲1%) the parameterized tail fraction option was considered. The selection of these cases was made by identifying outlying or obvious misfits in the bulk-based tail estimates, which is not an entirely satisfactory strategy. Still the impact of the specific tail fraction option is typically very small for the highest quantiles, and tail-based results such as return levels are quite robust.

A crucial aspect contributing to the robustness of the results is the deliberate assessment of uncertainties. One of the main advantages of extreme values mixture models is the possibility to avoid the issue of threshold selection in the analysis of exceedances. By considering the threshold as a parameter that is estimated rather than fixed a priori allows to take into account in the extreme value analysis the uncertainty associated with the threshold specification. Here a conservative approach is taken by considering the widest possible range of parameter values in the computation of return values from the model estimates.

In terms of the bulk data distribution of maximum summer temperatures, the mean reflects mainly the climatic regions of Iberia determined by the Peninsula relief, and the standard deviation reflects the contrast between coastal and in-land locations. In terms of the tails of the distribution, it is represented by a GPD distribution with negative shape parameter. The Weibull distribution is a reasonable representation for temperature exceedances as it corresponds to a tail with a finite upper end point and indeed air temperature is not expected to attain indefinitely higher values. The scale parameter of the GPD distribution typically exhibits an homogeneous spatial pattern, except in the coastal regions of northern and western Iberia where it is slightly higher. The threshold estimates exhibit a similar pattern to the average of the bulk distribution.

The comparison of the results for the two different periods allows to evaluate whether a similar description, in terms of extreme values, is obtained from the two data distributions. The large uncertainties due to the small size of the sample preclude the detection of any significant differences between the two periods for high return periods. However, significant differences are obtained in the extreme value mixture models for the two periods, and in the corresponding 20-year return levels. The mean increases significantly, particularly in the Eastern Iberia. This indicates a shift of the data distribution to the right, towards higher values of maximum temperature, with no major changes in the dispersion of the bulk data distribution. Changes in the tails of the data distribution do not coincide with the areas with highest mean and also highest threshold for temperature exceedances (south of Portugal, Andalucia, Extremadura). Instead the largest differences between the two period occur in the eastern Mediterranean region, corresponding to significant changes in the threshold estimate and an increase in return levels. An increase in the 20-year return levels is also identified in western Iberia (driven by an increase in the threshold and scale of the GPD distribution) and in the French Haute Garonne region (driven by an increase in the threshold). Although these differences between the two ~20-year periods merely reflect the distinct descriptions of extreme summer temperatures that are obtained from the observations available for the two different periods, with no predictive value nor aiming to support climate policies, the results indicate a detectable change in extreme temperatures in Iberia.

CRediT authorship contribution statement

Susana Barbosa: Conceptualization, Investigation, Software, Writing – original draft. **Manuel G. Scotto:** Methodology, Software, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work is financed by National Funds through the Portuguese funding agency, FCT - Fundação para a Ciência e a Tecnologia, within projects LA/P/0063/2020 and UIDB/04621/2020 (CEMAT/UL). Maps were made using Generic Mapping Tools (GMT; Wessel et al., 2013). Perceptually uniform color maps are used in order to prevent visual distortion of the data (Crameri, 2018).

Appendix

See Tables A.1 and A.2

Table A.1
Summary statistics for station's maximum declustered summer time series and for nearest reanalysis gridpoints.

		Station	ERA5	ERA5-Land
	min	18.7	16.6	17.7
Lisboa	median	28.7	24.3	26.5
(−9.15 38.72)	max	42.0	34.7	39.5
	min	20.2	16.6	17.4
Badajoz	median	36.0	34.5	35.0
(−6.17 38.88)	max	45.4	43.7	43.9
	min	17.5	16.5	16.6
Madrid	median	33.0	33.5	32.8
(−3.68 40.41)	max	40.7	41.0	40.3
	min	23.2	20.3	19.6
Malaga	median	31.6	28.8	28.8
(−4.49 36.67)	max	44.0	38.1	38.8
	min	4	12.3	10.6
Navacerrada	median	23.8	29.1	26.2
(−4.01 40.78)	max	32.0	36.2	33.7
	min	12.6	12.5	12.4
Salamanca	median	31.4	30.3	29.8
(−5.50 40.96)	max	41.0	37.8	38.0
	min	12.5	13.0	13.1
San Sebastian	median	22.7	23.0	22.9
(−2.04 43.31)	max	39.0	33.3	34.7
	min	23.1	19.0	19.4
Tortosa	median	33.3	28.1	29.0
(0.49 40.82)	max	43.0	34.9	37.7
	min	22.1	20.1	21.2
Valencia	median	30.2	28.5	30.9
(−0.35 39.48)	max	43.0	35.8	42.4
	min	20	20.3	19.3
Zaragoza	median	34.2	34.9	33.7
(−1.01 41.66),	max	44.5	44.1	43.4
	min	23.2	21.0	21.1
Alicante	median	31.0	28.2	30.1
(−0.57 38.28)	max	41.4	35.5	38.5
	min	19.0	19.2	17.8
Barcelona	median	29.2	28.1	27.0
(2.12 41.42)	max	39.8	36.8	34.9
	min	20.3	18.8	18.5
Albacete	median	34.0	33.1	32.9
(−1.86 38.95)	max	42.0	40.9	40.0
	min	24.3	21.0	20.9
Cordoba	median	37.7	36.1	36.0
(−4.85 37.84)	max	46.9	44.9	45.2
	min	14.2	12.8	12.3
Burgos	median	29.5	27.9	26.4
(−3.63 42.36)	max	38.8	38.4	36.2
	min	19	18.4	18.4
Ciudad Real	median	35.4	34.2	33.6
(−3.50 38.99)	max	43.7	42.3	41.0
	min	19.6	18.3	18.1
Granada	median	34.8	31.8	31.1
(−3.63 37.14)	max	43.5	40.1	39.0
	min	18.8	19.0	17.9
Huesca	median	32.6	32.8	31.4
(−0.33 42.08)	max	42.6	41.7	40.6
	min	15.7	14.1	13.7
Pamplona	median	31.2	28.0	28.0
(−1.78 42.65)	max	41.4	37.7	39.3
	min	22.2	21.9	21.5
Sevilla	median	37.0	35.7	35.4
(−5.88 37.42)	max	46.6	43.9	43.8
	min	13.8	13.4	12.7
Soria	median	30.2	29.7	28.8
(−2.48 41.78)	max	37.9	37.9	37.1
	min	15.8	13.9	13.7
Valladolid	median	32.0	31.0	30.2
(−4.77 41.65)	max	40.2	40.2	39.4
	min	16.6	13.9	13.9
Bilbao	median	27.2	22.8	22.4
(−2.90 43.30)	max	41.9	33.4	33.6
	min	13.4	13.6	13.4
Santiago	median	26.0	23.1	24.3
(−8.41 42.89)	max	39.4	34.3	35.9
	min	14.0	14.1	14.1
Vigo	median	25.6	25.1	24.9
(−8.62 42.24)	max	40.8	36.3	35.5
	min	12.8	9.5	11.2
Ponferrada	median	31.2	25.9	27.8
(−6.6 42.56)	max	39.6	34.0	36.3
	min	13.6	13.6	13.9
Zamora	median	32.0	30.3	30.4
(−5.73 41.52)	max	41.0	38.4	38.9
	min	21.6	20.0	20.9
Reus	median	30.2	30.0	29.5
(1.18 41.15)	max	39.8	38.9	38.0
	min	22.2	19.7	22.1
Murcia	median	29.5	26.1	29.2
(−0.80 37.79)	max	39.6	33.3	38.3
	min	21.0	20.8	20.9
Jerez	median	35.1	34.4	34.1
(−6.05 36.75)	max	45.1	43.9	44.2

Table A.2
Kolmogorov–Smirnov statistic (left) and overlapping index (right) for station and reanalysis empirical data distributions.

	Kolmogorov–Smirnov statistic		Overlapping index	
	ERA5	ERA5-Land	ERA5	ERA5-Land
Lisboa	0.49	0.24	0.37	0.64
Badajoz	0.19	0.14	0.72	0.80
Madrid	0.07	0.06	0.89	0.92
Malaga	0.40	0.38	0.47	0.48
Navacerrada	0.55	0.27	0.31	0.61
Salamanca	0.14	0.18	0.78	0.72
San Sebastian	0.10	0.08	0.78	0.82
Tortosa	0.72	0.60	0.18	0.27
Valencia	0.37	0.15	0.50	0.77
Zaragoza	0.11	0.06	0.84	0.89
Alicante	0.50	0.18	0.36	0.73
Barcelona	0.18	0.34	0.71	0.52
Albacete	0.14	0.17	0.80	0.75
Cordoba	0.21	0.22	0.68	0.66
Burgos	0.18	0.28	0.73	0.59
Ciudad Real	0.16	0.25	0.76	0.64
Granada	0.40	0.47	0.46	0.39
Huesca	0.04	0.17	0.93	0.74
Pamplona	0.30	0.26	0.57	0.61
Sevilla	0.17	0.20	0.74	0.71
Soria	0.07	0.18	0.91	0.75
Valladolid	0.12	0.21	0.81	0.69
Bilbao	0.49	0.53	0.37	0.33
Santiago	0.31	0.18	0.55	0.72
Vigo	0.09	0.12	0.86	0.80
Ponferrada	0.52	0.32	0.35	0.54
Zamora	0.19	0.18	0.71	0.74
Reus	0.05	0.12	0.92	0.81
Murcia	0.64	0.12	0.24	0.81
Jerez	0.11	0.14	0.82	0.79

References

Arnold, T.B., Emerson, J.W., 2011. Nonparametric goodness-of-fit tests for discrete null distributions. R J. 3 (2).

Barbosa, S.M., Scotto, M.G., Alonso, A.M., 2011. Summarising changes in air temperature over central europe by quantile regression and clustering. Nat. Hazards Earth

Syst. Sci. 11 (12), 3227–3233.
Behrens, C.N., Lopes, H.F., Gamerman, D., 2004. Bayesian Analysis of extreme events with threshold estimation. Stat. Model. 4 (3), 227–244.
Brown, S.J., Caesar, J., Ferro, C.A.T., 2008. Global changes in extreme daily temperature since 1950. J. Geophys. Res.: Atmos. 113 (D5).
Cardoso, R.M., Soares, P.M., Lima, D.C., Miranda, P.M., 2019. Mean and extreme temperatures in a warming climate: EURO CORDEX and WRF regional climate high-resolution projections for Portugal. Clim. Dynam. 52 (1–2), 129–157.
Carvalho, D., Pereira, S.C., Rocha, A., 2020. Future surface temperature changes for the Iberian Peninsula according to EURO-CORDEX climate projections. Clim. Dynam. 1–16.
Coelho, C.A.S., Ferro, C.A.T., Stephenson, D.B., Steinskog, D.J., 2008. Methods for exploring spatial and temporal variability of extreme events in climate data. J. Clim. 21 (10), 2072–2092.
Crameri, F., 2018. Scientific colour-maps. Zenodo. http://dx.doi.org/10.5281/Zenodo.1243862.
DeCastro, M., Gomez-Gesteira, M., Ramos, A.M., Alvarez, I., DeCastro, P., 2011. Effects of heat waves on human mortality, galicia, Spain. Clim. Res. 48 (2–3), 333–341.
Donat, M.G., Alexander, L.V., Yang, H., Durre, I., Vose, R., Dunn, R.J.H., Willett, K.M., Aguilar, E., Brunet, M., Caesar, J., et al., 2013. Updated analyses of temperature and precipitation extreme indices since the beginning of the twentieth century: The HadEX2 dataset. J. Geophys. Res.: Atmos. 118 (5), 2098–2118.
Fernández-Montes, S., Rodrigo, F., 2012. Trends in seasonal indices of daily temperature extremes in the Iberian Peninsula, 1929–2005. Int. J. Climatol. 32 (15), 2320–2332.
Hersbach, H., Bell, B., Berrisford, P., Horányi, A., Sabater, J.M.n., Nicolas, J., Radu, R., Schepers, D., Simmons, A., Soci, C., et al., 2019. Global reanalysis: goodbye ERA-interim, hello ERA5. ECMWF Newsletter 159, 17–24.
Hu, Y., Scarrott, C., 2018. Evmix: An r package for extreme value mixture nodeling, threshold estimation and boundary corrected kernel density estimation. J. Stat. Softw. 84 (5), 1–27.
Hyndman, R.J., Fan, S., 2009. Density forecasting for long-term peak electricity demand. IEEE Trans. Power Syst. 25 (2), 1142–1153.
Katz, R.W., Parlange, M.B., Naveau, P., 2002. Statistics of extremes in hydrology. Adv. Water Resour. 25, 1287–1304.
Kharin, V.V., Flato, G.M., Zhang, X., Gillett, N.P., Zwiers, F., Anderson, K.J., 2018. Risks from climate extremes change differently from 1.5c to 2.0c depending on rarity. Earth's Future 6 (5), 704–715.
Klein Tank, A.M.G., Wijngaard, J., Können, G.P., Böhm, R., Demarée, G., Gocheva, A., Mileta, M., Pashiardis, S., Hejkrlik, L., Kern-Hansen, C., et al., 2002. Daily dataset of 20th-century surface air temperature and precipitation series for the European climate assessment. Int. J. Climatol.: J. Royal Meteorol. Soc. 22 (12), 1441–1453.
Lee, W.V., 2014. Historical global analysis of occurrences and human casualty of extreme temperature events (ETEs). Nat. Hazards 70 (2), 1453–1505.
Lorenzo, N., Diaz-Poso, A., Roye, D., 2021. Heatwave intensity on the iberian peninsula: Future climate projections. Atmos. Res. 258, 105655.
Lucio, P.S., Silva, A.M., Serrano, A.I., 2010. Changes in occurrences of temperature extremes in continental Portugal: a stochastic approach. Meteorol. Appl. 17 (4), 404–418.
MacDonald, A., Scarrott, C.J., Lee, D., Darlow, B., Reale, M., Russell, G., 2011. A flexible extreme value mixture model. Comput. Statist. Data Anal. 55 (6), 2137–2157.
Milly, P., Betancourt, J., Falkenmark, M., Hirsch, R.M., Kundzewicz, Z.W., Lettenmaier, D.P., Stouffer, R.J., 2008. Stationarity is dead: Whither water management? Earth 4, 20.
Molina, M.O., Sánchez, E., Gutiérrez, C., 2020. Future heat waves over the mediterranean from an euro-cordex regional climate model ensemble. Sci. Rep. 10 (1), 1–10.
Muñoz Sabater, J., 2019. ERA5-L And hourly data from 1981 to present (access on 2020-06-08), copernicus climate change service (c3s) climate data store (cds). http://dx.doi.org/10.24381/cds.e2161bac.
Parey, S., Hoang, T.T.H., Dacunha-Castelle, D., 2019. Future high-temperature extremes and stationarity. Nat. Hazards 98 (3), 1115–1134.
Parey, S., Malek, F., Laurent, C., Dacunha-Castelle, D., 2007. Trends and climate evolution: Statistical approach for very high temperatures in France. Clim. Change 81, 331–352.
Pastore, M., 2018. Overlapping: a r package for estimating overlapping in empirical distributions. J. Open Source Softw. 3 (32), 1023.
Pastore, M., Calcagni, A., 2019. Measuring distribution similarities between samples: a distribution-free overlapping index. Front. Psychol. 10, 1089.
Ramos, A.M., Trigo, R.M., Santo, F.E., 2011. Evolution of extreme temperatures over Portugal: recent changes and future scenarios. Clim. Res. 48 (2–3), 177–192.
Scarrott, C., MacDonald, A., 2012. A review of extreme value threshold estimation and uncertainty quantification. REVSTAT-Stat. J. 10 (1), 33–60.
Scarrott, C., Yan, J., Dey, D.K., 2016. Univariate extreme value mixture modeling. Extrem. Value Model. Risk Anal. Methods Appl. 41–67.
Sheridan, S.C., Allen, M.J., 2015. Changes in the frequency and intensity of extreme temperature events and human health concerns. Curr. Clim. Chang. Rep. 1 (3), 155–162.
Sheridan, S.C., Allen, M.J., 2018. Temporal trends in human vulnerability to excessive heat. Environ. Res. Lett. 13 (4), 043001.
Sheridan, S.C., Lee, C.C., Smith, E.T., 2020. A comparison between station observations and reanalysis data in the identification of extreme temperature events. Geophys. Res. Lett. 47 (15), e2020GL088120.
Slater, L.J., Anderson, B., Buechel, M., Dadson, S., Han, S., Harrigan, S., Kelder, T., Kowal, K., Lees, T., Matthews, T., Murphy, C., Wilby, R.L., 2020. Nonstationary weather and water extremes: a review of methods for their detection, attribution, and management. Hydrol. Earth Syst. Sci. Discuss. 1–54.
Smid, M., Russo, S., Costa, A.C., Granell, C., Pebesma, E., 2019. Ranking European capitals by exposure to heat waves and cold waves. Urban Clim. 27, 388–402.
Suarez-Gutierrez, L., Müller, W.A., Li, C., Marotzke, J., 2020. Dynamical and thermodynamical drivers of variability in European summer heat extremes. Clim. Dynam. 1–16.
Sun, Q., Miao, C., Hanel, M., Borthwick, A.G.L., Duan, Q., Ji, D., Li, H., 2019. Global heat stress on health, wildfires, and agricultural crops under different levels of climate warming. Environ. Int. 128, 125–136.
Wessel, P., Smith, W.H.F., Scharroo, R., Luis, J., Wobbe, F., 2013. Generic mapping tools: improved version released. EOS Trans. Am. Geophys. Union 94 (45), 409–410.