



Stationary and Non-stationary Extreme Value Approaches for Modelling Extreme Temperature: the Case of Riyadh City, Saudi Arabia

Bechir Raggad^{1,2,3} 

Received: 9 June 2016 / Accepted: 4 December 2017 / Published online: 15 December 2017
© Springer International Publishing AG, part of Springer Nature 2017

Abstract

Climate change is one of the most fiercely debated scientific issues in recent decades, and the changes in climate extremes are estimated to have greater negative impacts on human society and the natural environment than the changes in mean climate. Extreme value theory is a well-known tool that attempts to best estimate the probability of adversarial risk events. In this paper, the focus is on the statistical behaviour of extreme maximum values of temperature. Under the framework of this theory, the methods of block maxima and threshold exceedances are employed. Due to the non-stationary characteristic of the series of temperature values, the generalized extreme value distribution and the generalized Pareto distribution were extended to the non-stationary processes by including covariates in the parameters of the models. For the purpose of obtaining an approximately independent threshold excesses, a declustering method was performed and then the de-clustered peaks were fitted to the generalized Pareto distribution. The stationary Gumbel distribution was found a reasonable model for the annual block maxima; however, a non-stationary generalized extreme value distribution with quadratic trend in the location was recommended for the half-yearly period. The findings also show that there is an improvement in modelling daily maxima temperature when it is applied to the declustered series and the given model outperforms the non-stationary generalized Pareto distribution models. Furthermore, the retained generalized Pareto distribution model proved better than the generalized extreme value distribution. Estimates of the return levels obtained from both extreme value models show that new records on maximum temperature event could appear within the next 20, 50 and 100 years.

Keywords Climate change · Extreme value theory · Generalized extreme value distribution · Generalized Pareto distribution · Non-stationary · Return levels · Temperature

1 Introduction

Climate change is one of the most fiercely debated scientific issues in recent decades, and the changes in climate extremes are estimated to have greater negative impacts on human society and the natural environment than the changes in mean climate. Extremely high temperatures are among the most frequently investigated extreme events. Numerous studies

were conducted to investigate possible effects of high temperature on the main human activities. Most analysts believe that extreme temperatures have considerable consequences: they can damage agricultural production (e.g. [1]), increase energy demand (e.g. [2]) and water consumption (e.g. [3]) and also badly affect human well-being, human health and even cause loss of human lives (e.g. [4–7]).

According to the synthesis of the Inter-governmental Panel on Climate Change (IPCC) [8], global mean temperature has shown a 0.85 °C (0.65–1.06 °C) increase over the period of 1800–2012 and a 0.74 °C increase during the last hundred years (1906–2005). The IPCC [9] report concludes that warming of the climate system is now “unequivocal,” based on observations of increases in global average air temperatures and the surface temperature is projected to rise over the 21st and it is very likely that heat waves will occur more often and last longer.

This work aims to model and analyse maximum temperature records in Riyadh city, the capital of Saudi Arabia, during

✉ Bechir Raggad
b.raggad@mu.edu.sa

¹ Department of Business Administration, College of Science and Humanities in Rumah, Majmaah University, Majmaah 11952, Saudi Arabia

² Faculty of Economics and Management of Nabeul, University of Carthage, Tunis, Tunisia

³ BESTMOD Laboratory, Higher Institute of Management of Tunis, University of Tunis, Tunis, Tunisia

the time period 1985–2014, using stationary and non-stationary extreme value theory (EVT) tools.

Within the climate literature, there are two main approaches for the diagnostic analysis of extreme events in climate data: the parametric and the non-parametric. The non-parametric approach is recognized by the definition of indices for extreme events, initially introduced by the Expert Team on Climate Change Detection Monitoring and Indices (ETCCDI). During the last years, there has been wide interest in analyzing climate with climate extreme indices (e.g. [10–20]). The main drawback of this approach is that it lacks the ability to extrapolate the results towards damage-relevant extremes with much larger return periods than those observed [21].

The second approach, initially developed by Fisher and Tippett [22], is founded on EVT techniques. EVT focuses directly on the tails of the sample distribution and, therefore, could potentially perform better than other approaches that model the whole distribution, in terms of predicting unexpected extreme changes [23]. EVT has been successfully applied in various fields, such as hydrology (e.g. [24–26]), engineering (e.g. [27]), finance and insurance (e.g. [28, 29]), oil price risk management (e.g. [23, 30]) and environment and climate fields (e.g. [31–38]). Within climate context, EVT has gained fast acceptance and popularity among both researchers and climate experts.

In the context of climate processes, it is common to observe non-stationarity due to seasonal variations and long-term trends owing to climate change. Consequently, it is essential to take into account the non-stationarity when modelling extremes. There are broadly two common strategies for dealing with the non-stationarity issues [39]: first, to use the full data set to detect and estimate non-stationarities, and then to apply methods for stationary extreme-value modelling to the resulting residuals, and, second, to fit a non-stationary extremal model to the original data. The first strategy¹ consists of using the full data set to detect and estimate non-stationarities induced by the volatility clustering using a GARCH type model and then to apply methods for stationary extreme value modelling to the resulting residuals (e.g. [23, 41]). The main advantage of this approach is that it reflects two stylized facts exhibited by most financial return series, namely stochastic volatility and the fat-tailedness of conditional return distributions over short time horizons.

The second strategy for modelling the extremes of a non-stationary process was initially introduced by Davison and Smith [24], who explicitly model the dependence structure using time series or covariate in the parameters of the model. Following this pioneering work, numerous studies describing methodologies have been proposed for the estimation of time-

varying extreme value distributions on non-stationary time series. For example, Chavez-Demoulin and Davison [42] describe smooth non-stationary generalized additive modelling for sample extremes, in which spline smoothers are incorporated into models for exceedances over high thresholds. Scotto and Guedes-Soares [43] illustrate a modelling with non-linear thresholds. Besides that, the existing literature reveals that various studies have been conducted, under the non-stationarity conditions, by using different ideas and techniques (e.g. [44–50]).

As climate changes increase due to global warming, it is expected that more extreme weather events will occur. According to a recent study of Pal and Eltahir [51], the Arab gulf region and parts of southwest Asia could be uninhabitable before the turn of the century as temperatures are expected to rise to intolerable levels. These alarming results justify the need for further analysis of temperature records in the region. For the case of Saudi Arabia, changes in temperature have already begun to be detected, using climate indices (e.g. [11, 12, 52]), although a comprehensive analysis on the subject, based on rigorous statistical tools, has yet to be conducted. EVT may provide a rigorous statistical method for the investigation of extreme temperature events in response to current and future climate change and for the accurately assess the changes in extreme temperature.

The objective of this study is to provide a more comprehensive analysis of observed changes in temperature extremes in Riyadh city, during the time period 1985–2014, by applying stationary and non-stationary EVT tools. In the first step, stationary and non-stationary generalized extreme value (GEV) models were conducted for the yearly and half-yearly maxima temperature data. The results reveal that the stationary Gumbel distribution was found a reasonable model for the annual block maxima; however, a non-stationary GEV with a quadratic trend in the location was recommended for the half-yearly period. In the second step, generalized Pareto distribution (GPD) models, under stationary and non-stationary conditions, were fitted for the temperature data. The findings show that there is an improvement in modelling daily maxima temperature when it is applied to the declustered series and the given model outperforms the non-stationary GPD models. Furthermore, the retained GPD model was found better than the GEV. The estimates of the return levels obtained from both EVT models show that new records on maximum temperature event could appear within the next 20, 50 and 100 years.

To the best of our knowledge, no studies focused on modelling extreme weather conditions relative to Saudi Meteorological data, using stationary and non-stationary EVT models, despite the significant interest to further analyse the extreme temperatures data. Such analysis is essential to gain an understanding of their possible occurrence and dimensions in present and future climate.

¹ Similar approaches for dealing with non-stationarity can be found in the literature. One of them is based on the removal of trends, obtaining a net data series suitable for the classical extreme value analysis (e.g. [40]).

The paper is organized as follows: Sect. 2 presents a review of extreme value theory. Section 3 provides the methods of estimation and model selection. Data and preliminary analysis are presented in Sect. 4. Results and discussions are presented in Sect. 5, and Sect. 6 concludes the paper.

2 Extreme Value Theory

In this section, we revisit the extreme value theory to provide a basis to our modelling of extreme temperature events in climate model data. Readers interested in a more detailed background may consult various texts on EVT such as Embrechts et al. [53]. EVT has two important results. First, the asymptotic distribution of a series of maxima (minima) is modelled and under certain conditions, the distribution of the standardized maximum of the series is shown to converge to the GEV distribution. The second significant result concerns the distribution of excess over a given threshold. EVT shows that the limiting distribution is a GPD.

2.1 GEV Distribution

2.1.1 GEV Distribution for Stationary Processes

EVT is concerned with the asymptotic distribution of standardized maxima (or minima) from a series of independent and identically (i.i.d) random variables with unknown common cumulative distribution function (cdf), $F(x) = P(X \leq x_i)$.

By the Fisher–Tippett theorem, the normalized maximum converges in the GEV distribution whose cumulative distribution is as follows:

$$G(x, \mu, \sigma, \xi) = \begin{cases} \exp\left\{-\left(1 + \xi\left(\frac{x-\mu}{\sigma}\right)\right)^{-\frac{1}{\xi}}\right\}, \xi \neq 0, & 1 + \xi\left(\frac{x-\mu}{\sigma}\right) > 0 \\ \exp\left\{-\left(\exp\left(-\frac{x-\mu}{\sigma}\right)\right)\right\}, & \xi = 0 \end{cases} \tag{1}$$

where $\mu \in \mathbb{R}$, $\sigma > 0$ and $\xi \in \mathbb{R}$ are the location, scale and shape

$$G(x, \mu(t), \sigma(t), \xi(t)) = \begin{cases} \exp\left\{-\left(1 + \xi(t)\left(\frac{x-\mu(t)}{\sigma(t)}\right)\right)^{-\frac{1}{\xi(t)}}\right\}, \xi \neq 0, & 1 + \xi(t)\left(\frac{x-\mu(t)}{\sigma(t)}\right) > 0 \\ \exp\left\{-\left(\exp\left(-\frac{x-\mu(t)}{\sigma(t)}\right)\right)\right\}, & \xi(t) = 0 \end{cases} \tag{3}$$

Following El Adlouni et al. [46], Cannon [54] and Hasan et al. [55], we consider possible non-stationary behaviour of the location μ and scale σ parameters but keep the shape

parameters, respectively. The Fisher–Tippett theorem suggests that the asymptotic distribution of the maxima belongs to a Frechet, Weibull or Gumbel distribution, regardless of the original distribution of the observed data. The shape parameter ξ describes the tail behaviour of the maximum distribution. For $\xi = 0$, the GEV is the Gumbel distribution. For $\xi > 0$, the tail of the GEV is “heavier” than the tail of the Gumbel distribution, and for $\xi < 0$, it is “lighter” than that of the Gumbel distribution. The GEV is said to have a type II tail for $\xi > 0$ and a type III tail for $\xi < 0$.

Having modelled the upper tail of a distribution by fitting a GEV distribution, it remains to use such a model for inference. One of the main applications of extreme value analysis is the estimation of the once per N year return levels. An event exceeding such a level is expected to occur once every N years. The $1/N$ year return value based on GEV distribution, z_N , is given by

$$z_N = \begin{cases} \hat{\mu} - \frac{\hat{\sigma}}{\hat{\xi}} \left\{ 1 - \left(-\log\left(1 - \frac{1}{N}\right) \right)^{-\hat{\xi}} \right\}, & \text{for } \hat{\xi} \neq 0 \\ \hat{\mu} - \hat{\sigma} \log\left\{ -\log\left(1 - \frac{1}{N}\right) \right\}, & \text{for } \hat{\xi} = 0 \end{cases} \tag{2}$$

The classical $G(x, \mu, \sigma, \xi)$ model assumes that the three parameters of location, scale and shape are time independent [32]. This model is frequently called “the stationary approach”. However, if trends are detected in the data sample, the non-stationary case, where parameters are no longer constants but expressed as covariates (e.g. time), should be considered.

2.1.2 GEV Distribution for Non-stationary Processes

The stationary GEV in Sect. 2.1.1 can be extended to non-stationary processes by including covariates in the parameters of the model (e.g. Coles [32]). In this paper, we limit our investigation to the case of dependence only on time. The non-stationary GEV distribution can be denoted as $G(\mu(t), \sigma(t), \xi(t))$ with distribution function

parameter ξ constant. More specifically, nine models of varying complexity that may be defined in this way (three choices for each of j and k), allowing up linear and non-linear

dependence on time of both the location μ and scale σ parameters, were developed with parameters as described as follows: where

$$\begin{aligned} \mu(t) &= \mu_0 + \mu_1 t + \mu_2 t^2 \\ \sigma(t) &= \exp(\sigma_0 + \sigma_1 t + \sigma_2 t^2) \end{aligned} \tag{4}$$

We denote by GEV_{jk} the model with time dependence of order j in the location parameter and order k in the scale parameter. For example, the stationary GEV distribution is GEV_{00} , obtained when the location and scale parameters are both independent of time ($\mu_1 = \mu_2 = 0$ and $\sigma_1 = \sigma_2 = 0$), whilst the GEV_{21} non-stationary model assumes a quadratic trend in location and a log-linear trend in scale ($\sigma_2 = 0$). The choice of the appropriate specification was performed using the selection criteria such as Akaike information criterion (AIC) and Bayesian information criterion (BIC) and statistical tests.

2.2 Generalized Pareto Distribution

2.2.1 Stationary GPD Models

In practice, modelling the maximum of a collection of random variables is wasteful if other data on extreme values are available. Therefore, a more efficient approach to modelling extreme events is to attempt to focus not only the largest events, but on all events greater than some large preset threshold. This is referred to as peaks over threshold (POT) modelling.

Let us define the excess distribution above the threshold u as the conditional probability:

$$F_u(y) = \Pr(X - u \leq y | X > u) = \frac{F(u + y) - F(u)}{1 - F(u)}, \quad y \geq 0 \tag{5}$$

Balkema and de Haan [56] and Pickands [57] theorem showed that the generalized Pareto distribution (GPD) is the limiting distribution for $F_u(y)$ as the threshold tends to the right endpoint. They stated that if the distribution function belongs to the maximum domain of attraction of the extreme value distribution of G , then it is possible to find a positive measurable function $\sigma(u)$ such that

$$\lim_{u \rightarrow x_F} \sup_{0 \leq y \leq u - x_F} |F_u(y) - G_{\xi, \sigma(u)}(y)| = 0 \tag{6}$$

where $G_{\xi, \sigma(u)}(y)$, the generalized Pareto distribution (GPD), is

$$G_{\xi, \sigma(u)}(y) = \begin{cases} 1 - \left\{ 1 + \xi \left(\frac{y}{\sigma(u)} \right) \right\}^{-\frac{1}{\xi}}, & \xi \neq 0, \quad 1 + \xi \left(\frac{y}{\sigma(u)} \right) > 0 \\ 1 - \exp\left(-\frac{y}{\sigma(u)}\right), & \xi = 0 \end{cases} \tag{7}$$

where $y \geq 0$ for $\xi \geq 0$ and $0 \leq y \leq -\frac{\sigma(u)}{\xi}$ for $\xi < 0$.

The GPD embeds a number of other distributions. If $\xi > 0$, then G is a reparametrized version of the ordinary Pareto distribution, $\xi = 0$ corresponds to the exponential distribution and $\xi < 0$ is known as a Pareto type II distribution.

The estimated return level x_m that is exceeded on average once every m observations is

$$x_m = \begin{cases} \hat{\mu} + \frac{\hat{\sigma}}{\hat{\xi}} \left((m\zeta_u)^{\hat{\xi}} - 1 \right), & \text{for } \hat{\xi} \neq 0 \\ \hat{\mu} + \hat{\sigma} \log(m\zeta_u), & \text{for } \hat{\xi} = 0 \end{cases} \tag{8}$$

where u is the selected threshold value, $\zeta_u = P(X > u) = k/n$, k is the number exceedances and n is the number of observations. By construction, x_m is the t -observation return level; however, it is often more convenient to give return levels on an annual scale, so that the N year return level is the level expected to be exceeded once every N years. If there are m_x observations per year, this corresponds to the m -observation return level with $m = m_x \times N$. Hence, an estimate of the N year return level x_N is defined by

$$x_N = \begin{cases} \hat{\mu} + \frac{\hat{\sigma}}{\hat{\xi}} \left((Nm_x\zeta_u)^{\hat{\xi}} - 1 \right), & \text{for } \hat{\xi} \neq 0 \\ \hat{\mu} + \hat{\sigma} \log(Nm_x\zeta_u), & \text{for } \hat{\xi} = 0 \end{cases} \tag{9}$$

The selection of the threshold u , which is crucial for the success of the GPD modelling, involves a delicate trade-off between bias and variance. If the threshold is chosen too high, then there are not enough exceedances over the threshold to obtain good estimators of the extreme value parameters, and consequently, the variances of the estimators are high. Contrariwise, if the threshold is too low, the GPD may not be good fit to the excesses over the threshold and there will be bias in the estimations. Several diagnostic techniques exist for this purpose, including graphical and bootstrap methods (e.g. [53]).

In the present study, we have chosen our threshold u using standard exploratory techniques, based on parameter stability plot (e.g. [32]), to aid the choice.

The POT model requires that the exceedances be mutually independent. However, especially in the case of climate data, this assumption may be violated because of serial correlation. In fact, at high levels, this appears as clustering of values of the series. To overcome this issue, a declustering approach that involves filtering the data series to remove the clustering and gain a set of independent threshold exceedances is used.

Most declustering methods are based on the estimation of a statistic called the extremal index θ . The extremal index is defined as the reciprocal of the limiting mean cluster size [32]. In the presence of no autocorrelation (clustering) in the series, then $\theta = 1$. Else if $\theta < 1$, then there is clustering in the data. The most widely adopted method for dealing with this

problem is declustering, which filters the dependent observations to obtain a set of threshold exceedances that are approximately independent.

Mainly, there are two approaches for estimating the extremal index based on run length r : In the first case, clusters are formed by arbitrarily specifying run length r , such that a cluster is considered active until r consecutive values fall below the threshold u . The extremal index is then estimated as the quotient of the number of clusters over the number of exceedances of the threshold. This approach is called “runs estimator”.

In the second case, optimal run length is obtained through the estimation of the extremal index assuming the exceedance times to be observed values of a point process whose limit is the Poisson process distribution. This latter method “intervals estimator” was formulated by Ferro and Segers [58], extending on the work of Hsing et al. [59] and Smith and Weissman [60], amongst others. The choice of r affects the bias-variance trade-off. A value of r that is too small raises concerns over the validity of the assumption of independent of cluster maxima. Conversely, large values of r could result in too few cluster maxima, hence raising concern over the precision of the GP distribution’s parameter estimates. This procedure does not involve any arbitrary choice such as that mentioned above, but it uses the extremal index, a parameter which measures the degree of clustering of extremes in a stationary process and takes values in the interval $[0, 1]$.

In the present work, we have used an automatic declustering scheme developed by Ferro and Segers [58] to obtain independent clusters of exceedances. For a detailed review on different declustering methods, the interested reader is referred to Smith et al. [61], Ferro and Segers [58] and Fawcett and Walshaw [62].

2.2.2 Non-Stationary GPD Models

The POT method explained previously is only valid for extremes for which we can assume stationarity. The idea is now to describe non-stationarity, as it is the case for temperature series that generally shows a strong seasonal pattern, by allowing the GPD parameters to depend on time or other covariates. In the literature, various approaches have been developed to deal with non-stationarity arising from seasonality or trend.

The most widely adopted technique to deal with data that vary seasonally is to partition the data into seasons and perform a separate extremal analysis on each season (e.g. [20, 63]). The second approach is based on fitting continuous parametric functions to capture the seasonality (e.g. [64]).

Thus, attempting to model any seasonal variation, we opt for the second approach and we include in our analysis a non-stationary GPD model with cycle variation in the scale

parameter. From Eq. (7), the shape parameter is assumed to be constant, that is $\xi(t) = \xi$, whilst $\sigma(t)$ is given by

$$\log(\sigma(t)) = \sigma_0 + \sigma_1 \sin\left(\frac{2\pi}{365.25} t\right) + \sigma_2 \cos\left(\frac{2\pi}{365.25} t\right) \quad (10)$$

For comparison, we introduce a second non-stationary GPD model with linear time covariate in the scale parameter given by

$$\sigma(t) = \exp(\sigma_0 + \sigma_1 t) \quad (11)$$

It should be noted that the exponential function has been adopted to introduce time dependency into the scale parameter to ensure its positivity. Note that it is also possible to incorporate the non-stationarity into the shape parameter. However, it is very difficult to estimate the shape parameter of the extreme values distribution with precision when it is time dependent, and therefore, it is not realistic to attempt to estimate the shape parameter as a smooth function of time [32]. Superiority of the non-stationary GPD models over the stationary GPD models was investigated through the likelihood-ratio test.

3 Estimation and Model Selection

The model parameters for GEV and GPD can be estimated in a variety of ways. Possible methods include maximum likelihood techniques (ML), the L-moment approach of Hosking [65] and Bayesian methods (e.g. [66]). Most common methods for parameter estimation in climate research are the ML estimation (e.g. Katz et al. [26, 36]) and the method of moments. The ML method, although problematic when applied to very small samples, is the preferable method due to its universal applicability and its nice asymptotic properties. Moreover, the method allows for the introduction of covariates such as time into the model [36]. The ML approach seems to be most common within the literature, and so we concentrate on this method here.

The analysis of extremes with EVT has been performed using the free software R and the extRemes package, which is designed for problems of extreme weather events and climate. For a brief introduction to the capabilities of extremes, we refer to the paper of Gilleland and Katz [67].

In the literature, there exist various methods to identify the best model out of a set of cautiously selected candidate models. One approach involves the information criteria, e.g. the Bayesian information criterion (BIC, e.g. [36]), the Akaike information criterion (AIC, e.g. [68]) and the Hannan-Quinn information criterion (HIC, e.g. [69]).

Table 1 Descriptive statistics of maximum temperature

	Number	Min	Mean	Max	S. Dev	Coefficient of variation	Skewness	Kurtosis
Daily	10,957	2.5	33.38	48.2	9.1	0.269	-0.4	-1.05
Half yearly	60	43.6	46.08	48.2	1.03	0.022	-0.05	-0.35
Yearly	30	45.40	46.65	48.2	0.811	0.0174	-0.36	-1.05

In this study, the AIC and the BIC statistics were used to determine which of the candidate models is most applicable. These two statistics identify the best model, which is supposed to fulfil the individual Student's t test on the parameters, when minimized. For a discussion on the use of these criteria for model selection, we refer to Burnham and Anderson [70].

4 Data and Preliminary Analysis

4.1 Data

The data used in this study consists of daily maximum temperatures for 30 years spanning from 1 January 1985 to 31 December 2014. The data was collected from the Saudi Arabia Presidency of Meteorology and Environment (PME). The city of Riyadh, the capital of Saudi Arabia, was selected as the case study area. Riyadh is located at 24.7° N and 46.71° E at an elevation of 635 m above sea level and is a typical station experiencing the hot and dry climate of central Saudi Arabia. The city has experienced significant population growth and urban expansion during the last decades. Such development may have been accompanied by changes in the local extreme temperature patterns [11]. In fact, new industrial zones have come up in hitherto green and open areas, leading to increased air pollution within the urban environment. The ever-increasing number of motor vehicles also

points to more pollution. Rising greenhouse gases and changes in the reflective properties of the Earth's surface are predicted to raise global temperatures [16].

This data set is challenging to model given the variety of extreme weather and climate that make it vulnerable to a changing climate.

4.2 Preliminary Analysis of Temperature Data

Table 1 shows some descriptive statistics for the daily temperatures together with the two selection periods: the yearly and the half yearly. The daily maximum temperature in Riyadh fluctuates between 2.5 and 48.2 °C over the period 1985–2014. The 10,957 daily maximum temperatures has a standard deviation 9.1 and a coefficient of variation 0.269.

After partitioning the data into yearly and half-yearly selection periods, it is observed that as the selection period increases, the difference between the minimum and maximum gets smaller, and the coefficient of variation decreases. This indicates that the maximum temperature data is less dispersed from the mean as the selection period increases. The skewness (SK) is negative for both periods, which indicates that the left tail of the distribution is relatively longer than the right side, implying that most of the data is concentrated on the right of the mean; the reverse is true for positive skewness. Figure 1 shows the time series plot of the annual maximum temperature

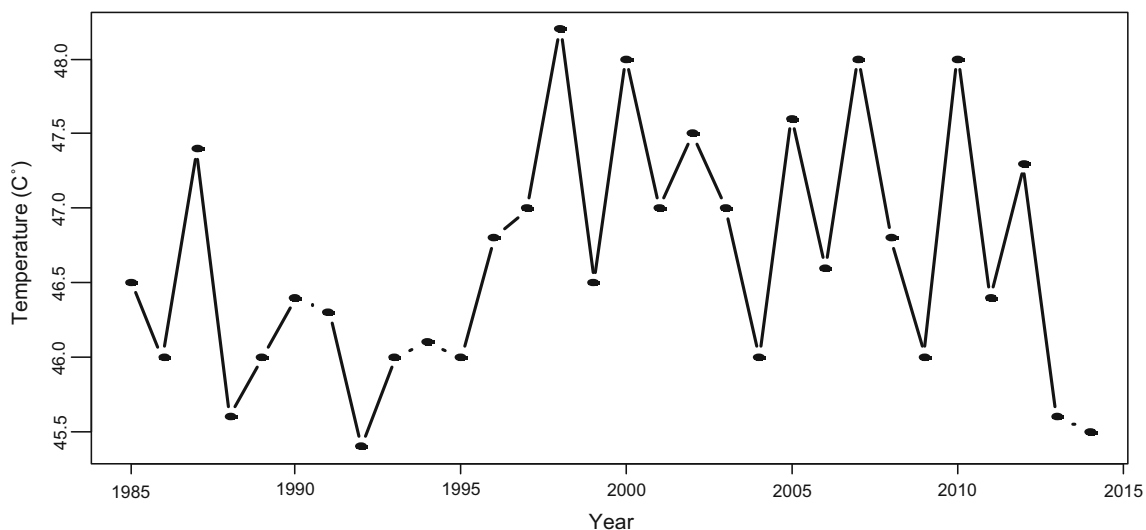
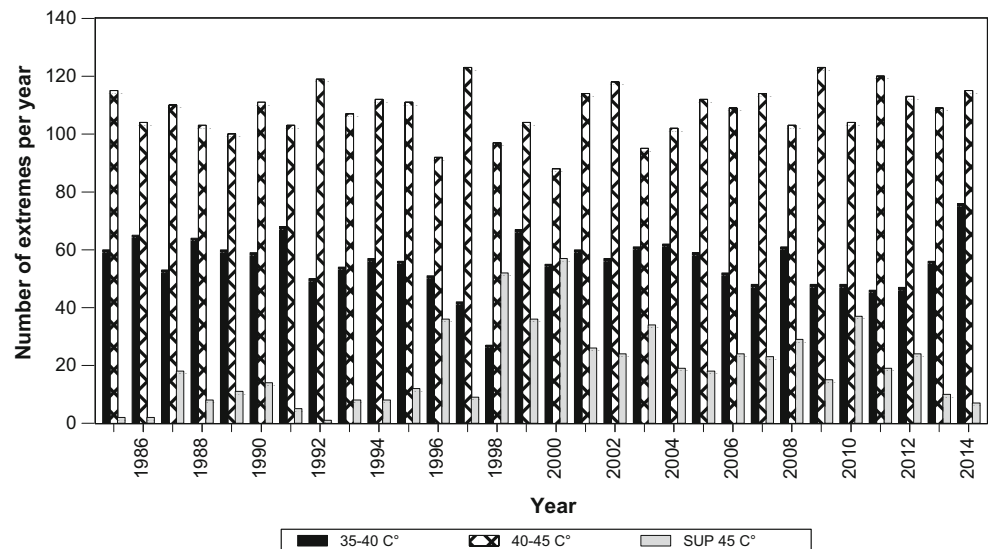


Fig. 1 Annual maximum temperature (°C) observed at Riyadh, 1984–2014

Fig. 2 Absolute frequencies of daily extreme temperatures (1 January 1985–31 December 2014)



(°C) observed at Riyadh over the period 1984–2014. The graph shows no clear trend in the annual maximum data. This visual impression will be confirmed by statistical tests.

In order to get an idea about the extremes of our data, we look at the absolute frequency² histograms of daily extreme temperature presented in Fig. 2. The graph displays the frequencies of extreme daily temperature per year for the time period of 1 January 1985 through 31 December 2014. We can see that as we increase the cut-off level as to what we define as extreme temperature, the number of observations declines substantially. In addition, there were some years that had many extreme temperature superior to $T \geq 45$ °C (e.g. 2000, 1998, 2010, 1999 and 1996). Broadly, we can see that the frequency of extremes has been higher in late 1990 until it reaches a peak in 2000 and it seems that the very extremely observations are more frequently after 1996. Overall, we can see that about the third of the year where the daily maximum temperatures belong to the interval; $40 \leq T < 45$ °C and about the sixth of them are between $35 \leq T < 40$ °C.

Prior to the application of the EVT models, some tests are required to assess the proprieties of our data. Three types of tests were performed for testing the assumptions of independence, stationarity and existence of trend. In particular, the GEV distribution needs that the random variables must be independent with common distribution. Violation of the assumption of independence will occur since seasonality will cause temperature data to vary accordingly. Failure to consider non-homogeneity into account will affect the analysis of the data and then produce inaccurate results. Therefore, larger

blocks are considered to assure that the assumption of having common distribution is plausible, even though it will generate only few block maxima. In addition, the choice of this block length can often avoid the need to account for yearly seasonal variation in environmental data, which may be seen as an advantage over other block lengths for this type of application.

One way for checking the assumption of independence is to perform the Ljung–Box test. The results of the test, presented in Table 2, show that the null hypothesis of independence could not be rejected for both selection periods³ (yearly and half-yearly). Such results are crucial for modelling with GEV distribution, which is based on the assumption of independence of random variables.

In the same table, we observe the results of the Augmented Dickey Fuller (ADF) and the Kwiatkowski–Phillips–Schmidt–Shin (KPSS) tests that have been performed to check the stationarity assumption of the data relative to different selection periods. The ADF tests suggest that the null hypothesis of unit root cannot be rejected here for both yearly and the half-yearly periods; however, the KPSS test concludes to the stationarity of the yearly maxima data but not for the half-yearly.

The Mann–Kendall (MK) trend test is performed in order to detect the presence of monotonic trend in the data under the null hypothesis of absence of trends. The obtained result, in Table 2, reveals that the null hypothesis was accepted for both selection periods at the 5% level whilst we fail to reject the null hypothesis for the half-year period at the 10% level.

² The absolute frequencies were calculated by counting the number of days per year where the daily maximum temperatures (T) belong to the three following arbitrary intervals: $35 \leq T < 40$ °C, $40 \leq T < 45$ °C and $T \geq 45$ °C.

³ For the other selection periods (Quarterly, Monthly), we fail to accept the null hypothesis of independence. Consequently, we limit our GEV yearly and modelling to half-yearly block lengths.

Table 2 Statistical tests for maximum temperature data

Tests	Selection period	
	Yearly	Half yearly
Ljung–Box	4.79 (0.19)	4.58 (0.20)
Mann–Kendall	60 (0.29)	268 (0.09)
Stationarity	ADF	− 1.12 (0.90)
	KPSS	0.33 (> 0.1)

Note: *P* values are in parentheses

From the MK test result, it seems that there is some concordance with the previous results of the stationarity test, essentially if we consider the KPSS test and we relax the level of significance to 10%. These results suggest that we ought to model for both stationarity and non-stationarity of the data set.

Based on the above analysis, we are justified in trying to fit extreme value theory to the data over the studied time period.

5 Results and Discussions

5.1 Application of GEV Distribution to Annual Temperature Extremes

In this study, data of maximum temperature are firstly blocked into annual lengths. An analysis of annual maximum data is

likely to be more robust compared to shorter blocks. Consequently, larger blocks are considered in our study to assure that the assumption of independence is plausible, even though it will generate only few block maxima. The choice of an annual block length can often avoid the need to account for yearly seasonal variation in environmental data, which may be seen as an advantage over other block lengths for this type of application.

Since the time series available for this study typically cover only a few decades, the sample sizes of annual maxima data are relatively small. In fact, if a 1-year block is used, this study will only have 30 annual maximum temperatures for the purpose of modelling, which may affect the accuracy of the estimated parameters of the GEV distribution. We tried to overcome this shortcoming by considering shorter blocks length (monthly, quarterly, and half-yearly block lengths) for the GEV model. However, we finally retained only the yearly and half-yearly periods, as they are the only block lengths that lead to sample data that meet the assumption of independence, required for the stationary GEV.

The model selection is based on nine models of varying complexity, as defined on Sect. 2.2, that incorporate both stationary and non-stationary GEV. The extreme value analysis is first performed by fitting the GEV distribution, assuming constant location and scale parameters to the sequence of annual maxima. Then, the likelihood ratio

Table 3 Model selection of GEV candidate models (for annual maxima) and return levels for the retained model

Parameters	GUM 00	GUM10	GUM20	GUM01	GUM02	GUM11	GUM12	GUM21	GUM22
μ	46.27 (0.13)	NA	NA	46.24 (0.13)	46.21 (0.15)	NA	NA	NA	NA
σ	0.67 (0.09)	0.67 (0.09)	0.60 (0.08)	NA	NA	NA	NA	NA	NA
μ_0	NA	46.270 (0.13)	46.64 (0.18)	NA	NA	46.30 (0.13)	46.35 (0.16)	46.63 (0.21)	46.63 (0.20)
μ_1	NA	0.070 (0.18)	0.242 (0.200)	NA	NA	0.304 (0.231)	0.40 (0.30)	0.24 (0.20)	0.26 (0.25)
μ_2	NA	NA	− 0.950 (0.391)	NA	NA	NA	NA	− 0.92 (0.48)	− 0.922 (0.48)
σ_0	NA	NA	NA	− 0.42 (0.14)	− 0.36 (0.23)	− 0.44 (0.14)	− 0.54 (0.24)	− 0.51 (0.14)	− 0.59 (0.24)
σ_1	NA	NA	NA	0.221 (0.238)	0.25 (0.24)	0.41 (0.24)	0.41 (0.25)	0.03 (0.30)	0.03 (0.30)
σ_2	NA	NA	NA	NA	− 0.18 (0.51)	NA	0.27 (0.54)	NA	0.066 (0.57)
AIC	<i>74.10</i>	75.94	72.10	75.16	77.035	75.40	77.16	77.10	76.08
BIC	<i>76.90</i>	80.15	77.71	79.36	82.640	81	84.16	81.10	84.49
Estimated return levels (95% confidence interval)									
20 years	49.02 (48.57;49.93)								
50 years	49.25 °C (49.87; 50.72)								
100 years	50.34 °C (49.36;51.32)								

Notes:

- Numbers in parentheses are standard errors
- The lowest AIC and BIC values are printed in italics
- Whilst the Gum20 model has the lowest AIC, it was excluded from the candidate models since it has the parameter μ_1 statistically insignificant
- NA (Not available) indicates that the parameters are not estimated to these models

Table 4 Model selection of Gumbel candidate models (for half-yearly maxima) and return levels for the retained model

Parameters	GEV00	GEV10	GEV20	GEV01	GEV02	GEV11	GEV12	GEV21	GEV22
μ	45.73 (0.15)	NA	NA	45.74 (0.15)	45.73 (0.16)	NA	NA	NA	NA
σ	1.04 (0.11)	0.99 (0.10)	0.95 (0.09)	NA	NA	NA	NA	NA	NA
ξ	-0.32 (0.09)	-0.31 (0.09)	-0.31 (0.09)	-0.31 (0.11)	-0.32 (0.12)	-0.31 (0.09)	-0.33 (0.10)	-0.313 (0.089)	-0.295 (0.100)
β_0	NA	45.74 (0.14)	46.11 (0.19)	NA	NA	45.74 (0.14)	45.72 (0.15)	46.120 (0.192)	46.101 (0.197)
β_1	NA	0.44 (0.22)	0.38 (0.21)	NA	NA	0.44 (0.22)	0.42 (0.21)	0.414 (0.217)	0.411 (0.205)
β_2	NA	NA	-1.03 (0.39)	NA	NA	NA	NA	-1.064 (0.401)	-1.031 (0.394)
σ_0	NA	NA	NA	0.03 (0.11)	0.08 (0.17)	-0.002 (0.11)	0.07 (0.16)	-0.054 (0.101)	-0.020 (0.128)
σ_1	NA	NA	NA	-0.06 (0.19)	-0.05 (0.20)	-0.011 (0.16)	-0.04 (0.15)	-0.063 (0.148)	-0.091 (0.156)
σ_2	NA	NA	NA	NA	-0.13 (0.29)	NA	-0.17 (0.28)	NA	-0.138 (0.316)
AIC	178.41	176.56	<i>172.03</i>	180.31	182.12	178.56	180.20	173.84	175.64
BIC	184.69	184.94	<i>182.50</i>	188.68	192.59	189.03	192.77	186.40	190.30

Notes:

- Numbers in parentheses are standard errors
 - The lowest AIC and BIC values are printed in italics
 - NA indicates that the parameters are not estimated to these models
- NA not available

test⁴ is then used to compare the goodness-of-fit of the model to the Gumbel distribution (GUM). It depends on the result of the test; if the null-hypothesis $H_0: \xi = 0$ against the two-sided alternative ($\xi \neq 0$) is not rejected at the significance level of 0.05, we retain in a first step the Gumbel distribution and we try to improve our modelling approach by allowing for time dependence on the location μ and/or on the scale σ parameters. However, if we fail to accept the null-hypothesis, the analysis will be based on the GEV_{jk} models defined in Sect. 2.1.2. Model selection and parameter estimates of different candidate models are given in Tables 3 and 4.

Table 3 presents the estimated results for annual maxima together with their resulting AIC and BIC values. The estimated shape parameters for the annual period were found to be very close to 0. The null hypothesis $H: \xi = 0$ was tested against the two-sided alternative ($\xi \neq 0$) using the maximum likelihood test. The null hypothesis is not rejected at the significance level of 0.05, which justifies the application of the Gumbel distribution to the annual maxima of temperature. Since a stationary behaviour can be unrealistic for maxima, particularly in a climate change context, the present model is compared with eight candidate models that are considered as defined above. The results show that the best specification is the simple

stationary Gumbel distribution (GUM00). It is important to note that this model has the smallest AIC and BIC criteria, after removing the models with insignificant parameters (e.g. GUM20) from the competing models. This finding may be confirmed by the Mann–Kendall (MK) trend test and the stationary tests results. In fact, in the former, we have failed to detect the presence of monotonic trend in the annual data and in the latter, we have showed that our data are level stationary.

The various diagnostic plots for assessing the accuracy of the Gumbel model are shown in Fig. 3. Both the probability plot and the quantile plot show the reasonability of the Gumbel fit: each set of points follows a quasi-linear behaviour. The return level plot shows approximate linearity, since it corresponds to $\xi = 0$ of the Gumbel distribution. Finally, the corresponding density estimate seems consistent with the empirical density of the data. Consequently, all four diagnostics plots support the fitted Gumbel model.

Once the best model for the data has been selected, the interest is in deriving the return levels of extreme maximum temperature. Estimates and confidence intervals for return levels for 20, 50 and 100 years are obtained by Table 3 based on the GEV00 model. It can be seen from the table that the return levels for maximum temperature gradually increase for higher and higher return periods. From the above results, one would expect that the maximum temperature (°C) at Riyadh will exceed about 49.02 on average every 20 years, will exceed about 49.25 °C on average every 50 years and will exceed about 50.34 °C every 100 years. The 95% confidence intervals, in degrees

⁴ The likelihood ratio test is used to compare the goodness-of-fit of two hierarchically nested GEV_{jk} models. If \hat{l}_0 denotes the maximized log-likelihood of the simpler model M_0 and \hat{l}_1 the maximized log-likelihood of a more complex model M_1 that contains M_0 , then, under standard conditions, $-2(\hat{l}_0 - \hat{l}_1)$ asymptotically follows the chi-squared distribution with q df, where q is the number of additional parameters in M_1 compared with M_0 .

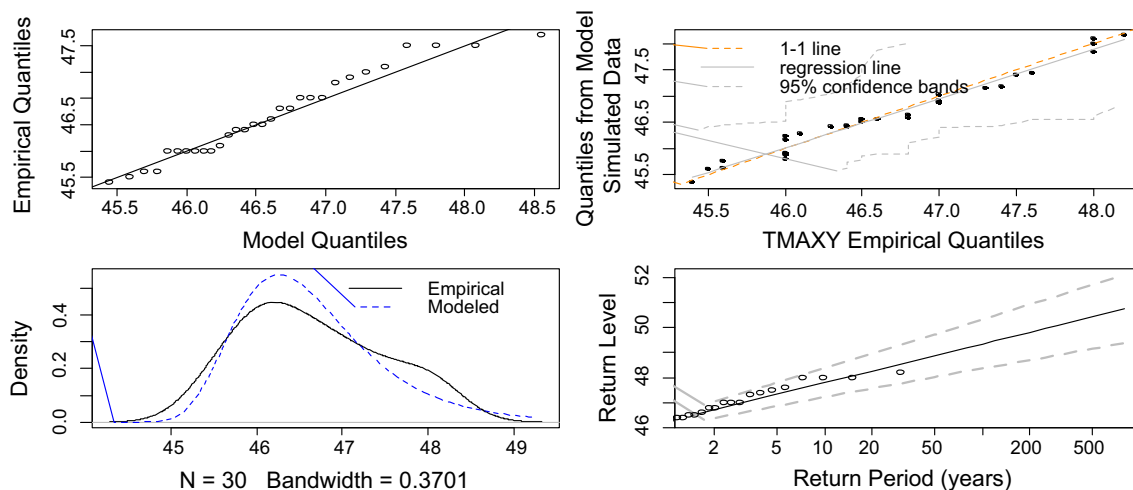


Fig. 3 Model diagnostics for the GUM00 fit to the Riyadh maximum temperature data shown. Quantile–quantile plot (top left), quantiles from a sample drawn from the fitted model df against the empirical data quantiles with 95% confidence bands (top right), density plots of

empirical data and fitted model df (bottom left) and return level plot with 95% point-wise normal approximation confidence intervals (bottom right)

Celsius, were (48.57; 49.93), (49.87; 50.72) and (49.36; 51.32), respectively. It is almost certain that the yearly maximum will exceed the current maximum, which is 48.2 °C.

The estimated results for annual maxima together with their resulting AIC and BIC values for the half-yearly period are shown in Table 4. The null hypothesis $H: \xi = 0$ was tested against the two-sided alternative ($\xi \neq 0$) using the maximum likelihood test. The null hypothesis is rejected at the significance level of 0.05, which justifies the application of the GEV distribution to the half-yearly maxima. We follow our analysis by examining the existence of any trend in GEV parameters.

Our results show that the best specification is non-stationary GEV20 is the best model. This model satisfies the individual Student’s t test and minimizes both criteria, the AIC and BIC. The shape parameter of temperature was found negative $\xi = -0.31$, indicating that the distribution of the extremes values is of Weibull form. Hence, the variable exhibits tail behaviour such that the upper tail is bounded at a finite upper point.

Diagnostic plots of Fig. 4 show that the fitted model is satisfactory. The return levels for the GEV20 model vary through the years because the model is non-stationary, which prevents the straightforward presentation of results.

Fig. 4 Model diagnostics for the GEV20 fit to the Riyadh maximum temperature data. Quantile–quantile plot (top left), quantiles from a sample drawn from the fitted GEV df against the empirical data quantiles with 95% confidence bands (top right), density plots of empirical data and fitted GEV df (bottom left) and return level plot with 95% point-wise normal approximation confidence intervals (bottom right)

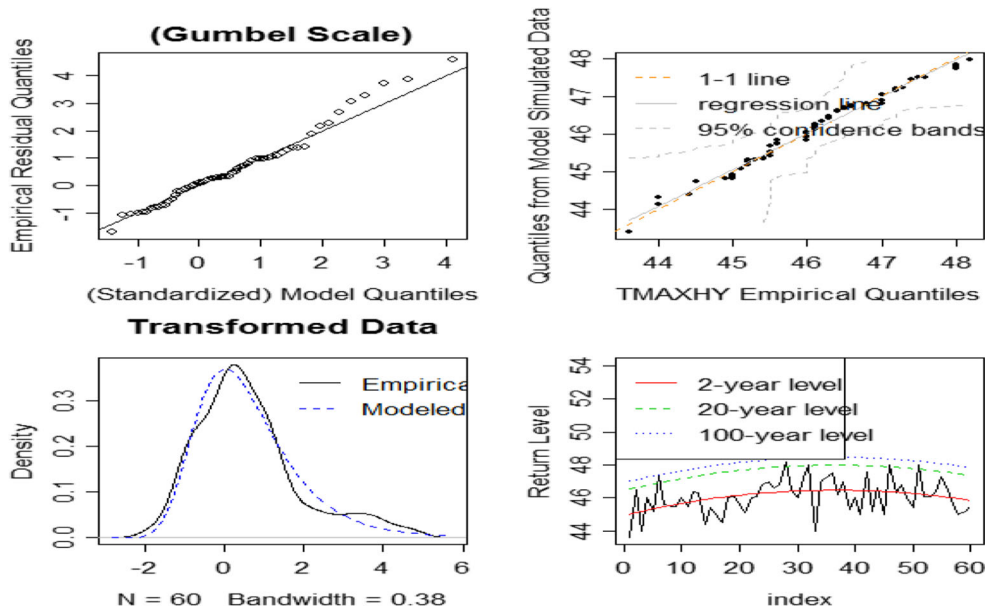
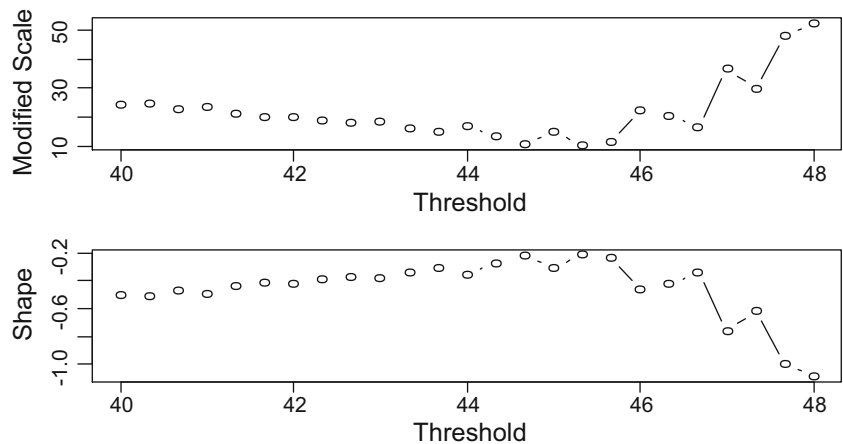


Fig. 5 Adjustment of the GPD for a range of 25 threshold values from 40 to 48 °C. The modified scale (σ^*) and the shape parameter (ξ) versus the threshold for the daily maximum temperature are shown for Riyadh, KSA during the period 1980–2014



5.2 Application of the GPD to the Maximum Daily Temperatures

The POT approach requires that the exceedances be mutually independent. However, in the case of temperature data, this assumption may be violated because of serial correlation. For example, a hot day is likely to be followed by another hot day. To overcome this issue, a declustering approach, which involves filtering the data series to remove the clustering and gain a set of independent threshold exceedances, is used.

Prior to the declustering, we need to specify a threshold for the GPD model. As we have explained in Sect. 2.2, the selection of the threshold is crucial for the GPD. However, there is no systemic approach that allows for the choice of an optimal threshold. In order to solve this problem, we find an appropriate threshold for GP models by fitting them to a sequence of thresholds in order to find the lowest threshold that yields roughly the same parameter estimates as any higher threshold (see [32] for further details).

Therefore, the GPD was adjusted to the maximum daily temperatures. The maximum likelihood estimators of the modified scale (σ) and shape parameter (ξ) plotted versus u for the daily maximum temperatures are shown in Fig. 5. If the GPD is a reasonable model for the exceedances above a certain threshold u , the estimates σ and ξ should remain near-constant [32].

The plot⁵ indicates a range of acceptable values for u . It appears that the results are not very sensitive to the choice of threshold ranging between 41 and 44. Parameter perturbations are small until the chosen threshold of 42 is reached. Such choice was found reasonable. The reasoning underlying this choice is that we need to have a value which is large enough so that the limiting GPD of Eq. (6) is a good approximation for the

⁵ The idea of this plot is that if the exceedances of a high threshold u_0 follow a GPD with parameters ξ and σ_{u_0} , then for any threshold u such that $u > u_0$, the exceedances still follow a GPD with shape parameter $\xi_u = \xi$ and scale parameters are related by $\sigma_u = \sigma_{u_0} + \xi(u - u_0)$. Let $\sigma_u = \sigma_u + \xi_u(u - u_0)$. This new parametrisation does not depend on u any longer, given that u_0 is a reasonably high threshold. The threshold should be chosen at the value where the shape and scale parameters remain constant.

exceedance distribution, whilst not so large as to reduce unnecessarily the number of exceedances available for the analysis. In order to confirm the rightness of our threshold choice, QQ plots were created using data screened at different thresholds. From the plots in Fig. 6, we can see that the chosen threshold of 42 appears to do an adequate job of predicting most of the data. Based on this, one could recommend $u = 42$ as the one to be used.

Before deciding whether there is a need for declustering or not, we begin by looking at the auto-tail dependence function plot, shown in Fig. 7, which allows us to check whether there exists temporal dependence in threshold excess data.

The sample auto tail-dependence function⁶ based on $\hat{\rho}$, an auto-tail dependence estimate suggested by Reiss and Thomas [71], along with a plot against increasing lags is produced. It takes on values between zero and one (inclusive). If the values over a high threshold are stochastically independent, then the values of $\hat{\rho}$ should be close to $(1 - u)$ at each lag, where u is the quantile threshold, which in the example above $u = 0.90$. Inspection of Fig. 7 shows that all lags equal or greater than five are fairly close to 0.4, but that the lag-one, two, three and four terms are between 0.45 and 0.63, so the assumption of independence may not be reasonable for these data.

In order to evaluate the strength of tail dependence, the extremal index was estimated, using the threshold 42, for maximum temperature data using the method of Ferro and Segers [58]. Consequently, the optimal run length was determined. The estimate of the extremal index⁷ was found 0.031,

⁶ The auto-tail-dependence functions can be defined as follows: Let X_1, \dots, X_n be a series of identically distributed random variables with common df F. Assume, in addition, that the series has stationary dependencies in the upper tail in the sense that for $i \leq n - h$, $P(X_{i+h} > F^{-1}(u) | X_i > F^{-1}(u)) = P(X_{i+1} > F^{-1}(u) | X_i > F^{-1}(u)) = \rho(u, h)$. The auto-tail-dependence functions $\rho(u, h)$ at the level u can be estimated by the sample version $\hat{\rho}_n(u, h) = \frac{1}{n(1-u)} \sum_{i < n} I(\min(x_i, x_{i+h}) > x_{[nu]:n})$ based on data x_1, \dots, x_n .

⁷ The extremal index is a useful indicator of how much clustering of exceedances of a threshold occurs in the limit of the distribution. The extremal index is bounded between 0 and 1, with large values indicating complete independence and values near zero an indication of perfect dependence between the exceedances.

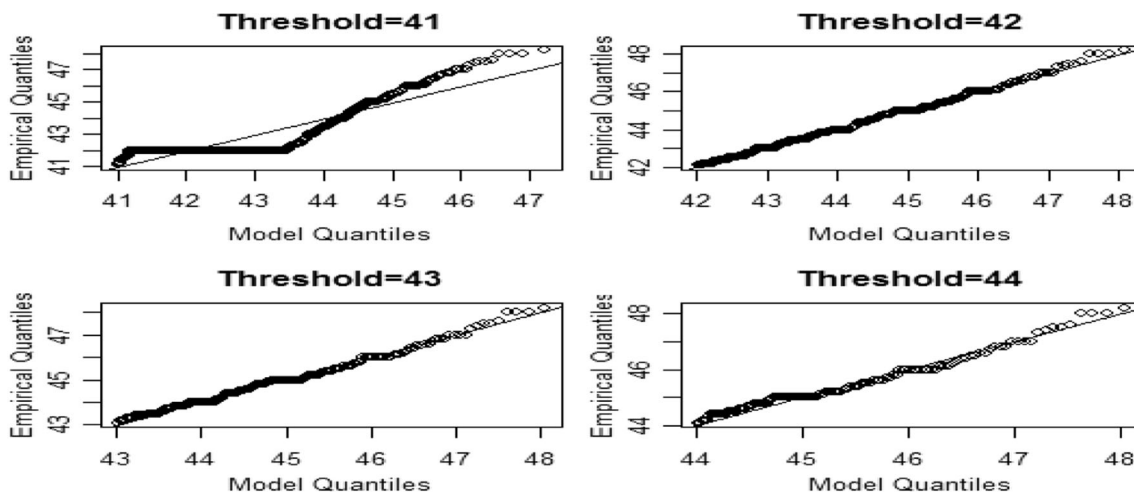


Fig. 6 QQ plots for temperature data for a range of threshold

with a 95% confidence level of (0.023, 0.043), which is suggestive of strong tail dependence (in agreement with our findings in Fig. 7). The estimated run length was found 7. Having evidence of tail dependence, we decide to employ a declustering scheme to filter out a set of approximately independent threshold excesses.

In order to check the existence of any improvement in our modelling based on the declustered data, we present the results of the estimation of two stationary GPD models. The first model, noted (model M_1), is applied to the data assuming the hypothesis of the independence and the second, noted (model M_2), is fitted to the declustered series of daily maximum temperature using the optimal run lengths determined by the method of Ferro and Segers [58].

Table 5 displays the parameter estimates for both estimated models. For the model M_2 , we can see that the shape parameter estimates is negative ($\xi = -0.39$), indicating a short-tailed distribution. Hence, the variable exhibits tail behaviour such that the upper tail is bounded at a finite upper point. The 95% confidence interval for ξ is $(-0.483; -0.311)$ does not contain zero which implies the upper tail here may not decay exponentially. The result shows that there is a strong evidence, at 5%

significance level to support the suitability of an upper bounded distribution $\xi < 0$, for the declustered data.

For the model M_1 , we also obtain strong evidence in support of GPD with $\xi < 0$. The shape parameter estimate is given by $\xi = -0.420$, which is different from the estimate value found in the model M_2 .

From the Table 5, we note that when the GPD is fitted to the declustered series, the scale parameter σ is overestimated, and the shape parameter ξ is underestimated relative to the approach which uses all exceedances (without declustering).

Figure 8 shows the QQ plots for the declustered series having used a run length of 7. Such plots indicate a good model fit as the points follow the path of the line of equality. In the same figure, we observe the QQ plot of the GPD fitted to all exceedances. It seems that GPD fitted to the declustered data (model M_2) is the most appropriate for modelling our data. We note that there is a substantial improvement in GPD fit after clustered observations were filtered from the analysis. In addition, the model M_2 showed the lowest values for both AIC and BIC information criteria.

Of greater practical interest are the estimated return levels. Table 6 shows that the estimates are slightly different for the 20-year return periods but are greater for the declustered data for the 50- and 100-year return periods.

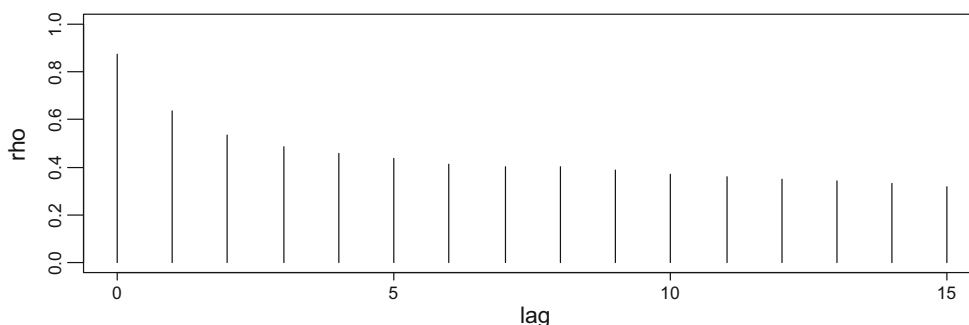


Fig. 7 Auto-tail dependence function for daily temperature data using a quantile threshold of 0.90

Table 5 Summary of results from fitting the GPD models with and without declustering

	σ	ξ	AIC	BIC
Model M_1	2.62 (0.027)	-0.42 (0.003)	7611.49	7623.11
Model M_2	2.68 (0.18)	-0.39 (0.04)	1047.41	1055

Note: Numbers in parentheses are standard errors

We can expect that a maximum temperature event will reappear within the next 20, 50 and 100 years.

Within climate data, it is common to observe non-stationarity due to different seasons having different climate patterns or due to more long-term trends owing to climate change. Consequently, it is essential to take into account of the non-stationarity when modelling extremes. As an example, we have plotted the daily maxima for the years 2013 and 2014 (Fig. 9). It is clear that the temperatures follow a seasonal pattern: the temperature increases from January until it reaches a peak during summer (June, July and August) and then decreases until it reaches a minimum during winter (December, January and February).

To avoid these seasonal variations when dealing with modelling non-stationarity, one possibility to describe the variation due to seasonality in exceedances could be a separate season analysis. An alternative way consists to incorporate sinusoidal functions of time to the parameter estimates of the GPD model.

In this study, two non-stationary GPD models were included in our analyses with the shape parameter is assumed to be constant, that is $\xi(t) = \xi$, whilst $\sigma(t)$ is given by either a cycle variation in the scale parameter

(Eq. 10) or a linear time covariate in the scale parameter (Eq. 11).

The results of fitting the non-stationary GPD models are given in the Table 7. The results show that there is a significant trend in the scale parameter for both models.

By comparing the log-likelihood values of the two non-stationary models, it seems that the non-stationary GPD with cycle variation in the scale parameter is more appropriate than one with linear time covariate in scale parameter.

The likelihood ratio test is used to compare the three models and the test statistic and p values are listed in Table 8. According to the p values of the likelihood ratio comparison between the stationary GPD model (M_1) and the GPD model with cycle variation in the scale parameter (model M_3) and that with model with linear time covariate in the scale parameter (model M_4), the last two models provide a significant improvement over the model M_1 whilst the model M_3 is more appropriate.

However, by comparing the AIC and the BIC criteria in Table 7 to those presented in Table 5, it appears that the GPD fit after clustered observations is more reasonable. There is strong evidence that the GPD model adequately estimates the tail behaviour of the distribution of the data series once the data was declustered.

5.3 Comparison of EVT Models

It is well known that theoretically the parameter ξ is the same in the GEV and the GPD. However, due to different amounts of data used in the block maxima analysis versus

Fig. 8 QQ plots of the GPD models fitted to the data with declustering (left) and without declustering (right)

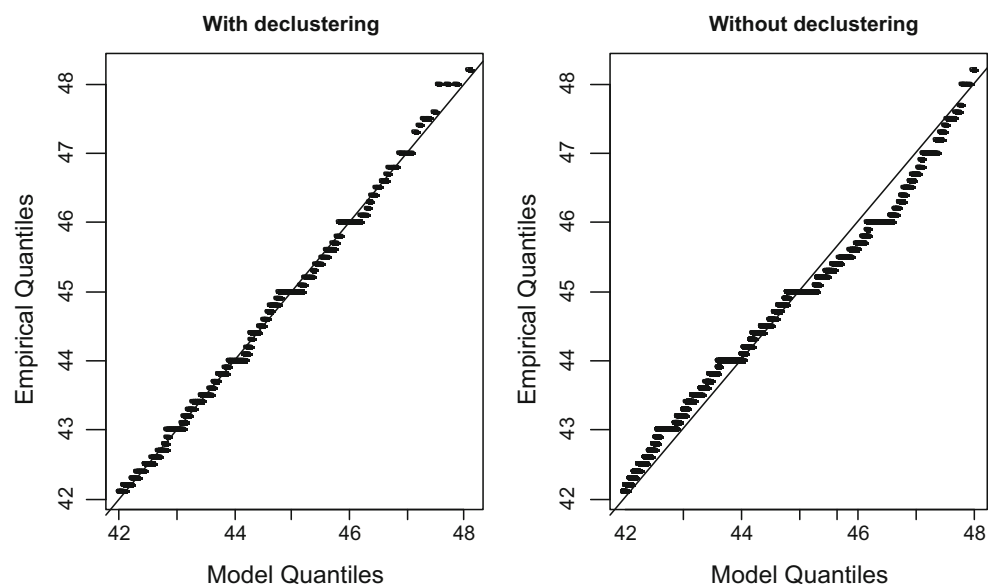


Table 6 Estimates of return levels for the GPD models fitted to the data without and with declustering

	20 years	50 years	100 years
Model M_1 (without declustering)	48.96 (46.91;51)	49.05 (46.04; 52.05)	49.094 (45.07;53.12)
Model M_2 (with declustering)	48.97 (46.51;51.43)	49.21 (45.69; 52.73)	49.343 (44.72;53.97)

Note: Numbers in parentheses are standard errors

exceedances, the parameter estimates are different. In fact, if the block is annual, the GEV only considers annual maxima, whilst the GPD tries to model all values that exceed a certain threshold u .

In this paper, the Gumbel was found to be a reasonable choice for modelling the annual block maxima; however, by considering half-yearly block maxima, we have found that the most applicable model is a non-stationary GEV with a quadratic trend in the location parameter. This result may be due the relatively small sample sizes (30 observations) of annual maxima data used in the first case. In fact, a simple look to the Figs. 2 and 3 confirms that the GEV20 is more suitable than the stationary Gumbel.

Concerning the GPD approach, we have found that the GPD applied to the declustered data is better than the fitting of the GPD to all exceedances, ignoring dependence. There is a substantial improvement in GPD fit after clustered observations were filtered from the analysis.

By taking account of non-stationary in POT analyses, it seems that the non-stationary GPD with sinusoidal functions in the scale parameter is more appropriate for modelling our data. However, it appears that the GPD fit after clustered observations is more reasonable.

Based on the analysis above, we may retain the GEV20 and the GPD fitted to the declustered data as the most suitable models. In order to compare the performance of

them, we put the diagnostic plots for the two models side-by-side (Fig. 10). We see that the three models fit the data fairly well, but the GPD, when it was applied to the declustered data, it appears to do a better job at the first. This result is not surprising since it has been well confirmed in the literature that the GPD is the more efficient approach in modelling extreme events. The discrepancies seem to be because the GPD uses much higher proportion of the original data values.

6 Conclusions

In this paper, we have studied extreme temperatures in Riyadh city, KSA. Different approaches coming from extreme value theory (EVT) are applied and compared to model, analyse and forecast extreme weather conditions under stationary and non-stationary contexts. In particular, we have modelled the daily maximum temperatures recorded in Riyadh, the capital of Saudi Arabia, using the GEV and GPD models.

Major findings and conclusions of this study are as follows:

- The stationary Gumbel model was found capable of fitting extreme temperature series when applied to annual block maxima. The developed non-stationary GEV models did not show any advantage over the stationary model according to the used criteria for

Fig. 9 Maximum daily temperature records from January 2013 to December 2014

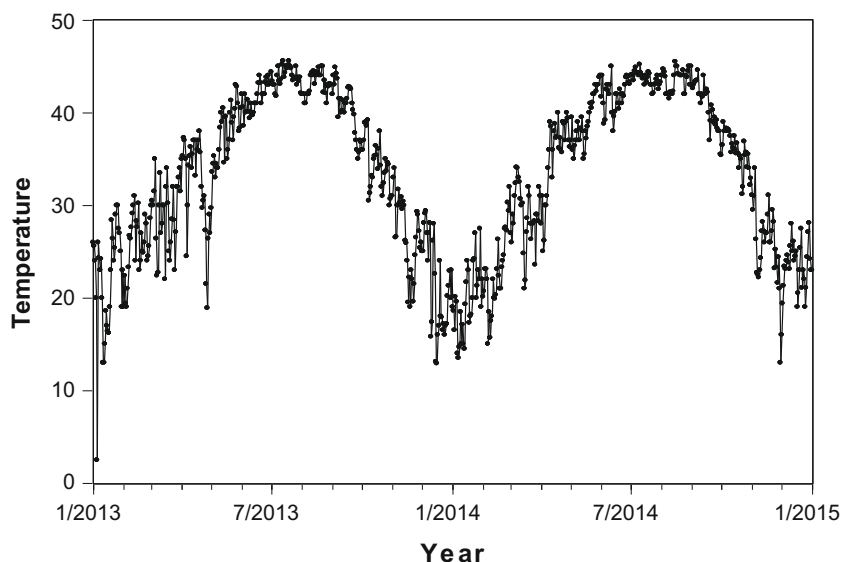


Table 7 Summary of results from fitting the non-stationary GPD models

	σ_0	σ_1	σ_2	ξ	AIC	BIC
Model M_3 (GPD with linear time covariate in scale parameter)	0.88 (0.007)	0.00012 (0.00002)	–	–0.41 (0.002)	7605.74	7623.17
Model M_4 (GPD with cycle variation in scale parameter)	–3.27 (0.63)	1.24 (0.15)	4.10 (0.62)	–0.43 (0.01)	7575.68	7598.92

Note: Numbers in parentheses are standard errors

models selection. The last choice could also be justified by the main time series of the data that shows no obvious evidence of climate change, which may well not yet have an impact at that location. The return level estimates reveal that the temperature that exceeds the current maximum temperature (48.2 °C) will start to appear within the next 20, 50 and 100 years.

- Despite the absence of any obvious trends in yearly extreme data, there was evidence of trend in the half-yearly data. The non-stationary GEV model with a quadratic trend in the location shows advantage over the stationary models. This result is partially supported by Mann–Kendall (MK) and the KPPS tests.
- The GPD approach fitted to the declustered data was found to be superior to the GPD fitted to all exceedances. Similar results were found for the return levels; as for the retained stationary Gumbel model fitted to annual block maxima, new records on maximum temperature event could appear within the next 20, 50 and 100 years.
- The non-stationary GPD with cycle variation in the scale parameter is found to be more appropriate for modelling the temperature data when compared to the non-stationary GPD model with linear time covariate in the scale parameter. However, it appears that the GPD fit, when applied to declustered observations, is more reasonable among all the GPD candidate models.
- By comparing EVT models, our findings seem to suggest that the three retained models (stationary

GEV, non-stationary GEV and the GPD with declustering) to the temperature data are fairly well, but the GPD, when applied to the declustered data, appears to do a better job at the first. These results are not surprising since it is well confirmed in the literature that the GPD is a more efficient approach than the GEV.

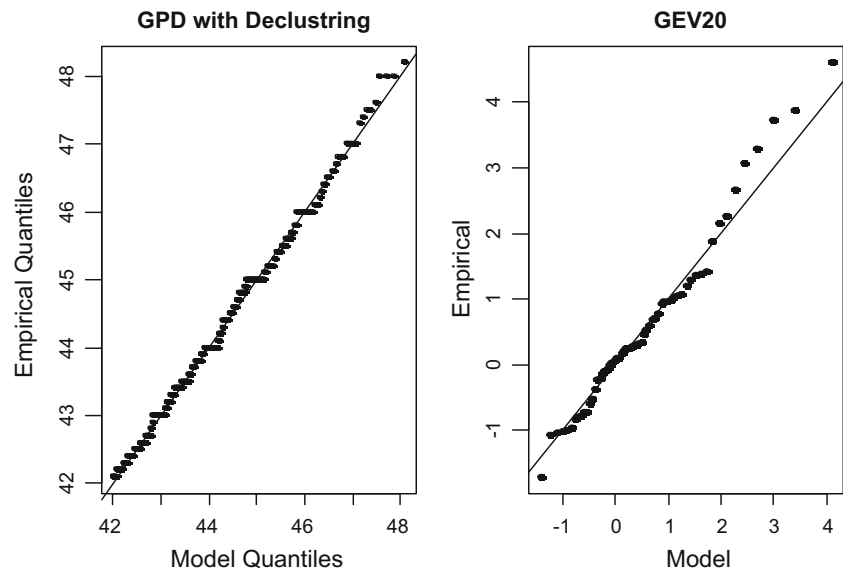
The main outcomes of this study are broadly consistent with the findings of Tanarhte et al. [72], who studied the characteristics of heat waves in the eastern Mediterranean and Middle East region during the period 1973–2010. Their results, based on stationary extreme value theory, reveal that the return levels calculated for the individual hot days and found to be very high in the Arab Gulf region. In particular, they reported that the high temperature of 50 °C is expected to be exceeded every 20 years in Dhahran (Saudi Arabia) and every 10 years in Kuwait, 48 °C every 10 years in Doha (Qatar) and 46 °C every 10 years in Eilat (Israel). Our results are also in line with previous studies that used climate indices for the analysis of extreme temperatures. For example, we can cite the results of Zhang et al. [73] for the Middle East region and Athar [12] for Saudi Arabia, in which significant, increasing trends have been found in the annual maximum of daily maximum temperature.

In this paper, we have shown how extreme value theory serves as a useful analysis tool in describing extreme temperature events. The study improves our understanding of temperature events in Riyadh city, and it may be beneficial for quantifying future heat wave properties as part of forecasts or longer-term projections from climate models. Additionally, it may offer some insights, for policy makers and planners, such as the prediction of long run temperature in future relative. However, our study concludes with the importance of regarding climate change exposure of this region, instead of using a single station that may not be representative for the whole area,

Table 8 Likelihood ratio test results for comparison of models 1 versus 3 and 1 versus 4

	LR statistic	<i>p</i> value
Model 1 versus model 3	39.81	0
Model 1 versus model 4	7.75	0.005

Fig. 10 QQ plots of the GPD fitted to the declustered data and the GEV20 fitted to half-yearly maxima temperatures



as we were not able to confirm the impact of climate change based only in these observations.

The results we have presented show the need for more investigation and therefore can be extended in several ways. In this study, we only considered data from a single station to demonstrate the EVT methodology for climate studies. It is not realistic to extrapolate the findings of this study for larger spatial scales such as the entire Saudi area without further analysis using temperature data from multiple observation stations within the area. In addition, non-stationarity in extremes increases the complexity in modelling. This is mainly due to the subjectivity involved in choosing appropriate parametric functions for seasonality and long-term trends.

Accordingly, the statistical modelling can further be expanded and improved in various ways. For example, multiple observation stations could be used to further test the statistical significance of the results presented in this study and their generalization for the whole region. Two scientific concerns that have recently emerged in the field of climate research motivate the extension of the present work as previously mentioned. First, the Nature Climate Change paper of Pal and Eltahir (2016), “Future temperature in southwest Asia projected to exceed a threshold for human adaptability”, addresses this region specifically and shows that the Persian/Arab gulf region and parts of southwest Asia could be uninhabitable before the turn of the century as temperatures are expected to rise to intolerable levels. Second, the IPCC reports an expected profound climate change that may radically change the picture of our climate.

Finally, researchers could extend our framework to jointly model, via copula functions, the dependence relation between extreme temperatures and air pollution as they are well known to be closely coupled and generally move together.

Acknowledgements The author would like to thank Deanship of Scientific Research at Majmaah University for supporting this work under Project Number No. 37/103.

References

1. Shah, F., Huang, J., Cui, K., Nie, L., Shah, T., Chen, C., & Wang, K. (2011). Impact of high-temperature stress on rice plant and its traits related to tolerance. *Journal of Agricultural Science*, 149(5), 545–556. <https://doi.org/10.1017/S0021859611000360>.
2. Chikobvu, D., & Sigauke, C. (2013). Modelling influence of temperature on daily peak electricity demand in South Africa. *Journal of Energy in Southern Africa*, 24(4), 63–70 Energy Research Centre.
3. Chang, H., Praskievicz, S., & Parandvash, H. (2014). Sensitivity of urban water consumption to weather and climate variability at multiple temporal scales: the case of Portland, Oregon. *International Journal of Geospatial and Environmental Research*, 1(1), Article 7. Available at: <http://dc.uwm.edu/ijger/vol1/iss1/7>.
4. CDC. (2013). Heat-related deaths after an extreme heat event—four states, 2012, and United States, 1999–2009. *Morbidity and Mortality Weekly Report*, 62(22), 433–436.
5. Comrie, A. (2007). Climate change and human health. *Geography Compass*, 1(3), 325–339.
6. Ebi, K. L. (2008). Using health models to prepare for and cope with climate change. *Climatic Change*, 88, 1–3.
7. Karl, T. R., & Easterling, D. R. (1999). Climate extremes: selected review and future research directions. *Climatic Change*, 42, 309–325.
8. IPCC (2007) Climate change 2007: the physical science basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge, United Kingdom and New York, NY, USA.
9. IPCC (2014). Climate Change, Synthesis Report. Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change [Core Writing Team, Pachauri, R.K. & Meyer, L.A. (eds.)]. IPCC, Geneva, Switzerland, 2014.
10. Alexander, L. V., Zhang, X., Peterson, T. C., Caesar, J., Gleason, B., Klein Tank, A. M. G., Haylock, M., Collins, D., Trewin, B.,

- Rahimzadeh, F., Tagipour, A., Kumar Kolli, R., Revadekar, J. V., Griffiths, G., Vincent, L., Stephenson, D. B., Burn, J., Aguilar, E., Brunet, M., Taylor, M., New, M., Zhai, P., Rusticucci, M., & Vazquez Aguirre, J. L. (2006). Global observed changes in daily climate extremes of temperature and precipitation. *Journal of Geophysical Research-Atmospheres*, *111*, D05109. <https://doi.org/10.1029/2005JD006290>.
11. Alghamdi, A. S., & Moore, T. W. (2014). Analysis and comparison of trends in extreme temperature indices in Riyadh city, Kingdom of Saudi Arabia, 1985–2010. *Journal of Climatology*, *2014*, 1–10. <https://doi.org/10.1155/2014/560985>.
 12. Athar, H. (2014). Trends in observed extreme climate indices in Saudi Arabia during 1979–2008. *International Journal of Climatology*, *34*(5), 1561–1574. <https://doi.org/10.1002/joc.3783>.
 13. Frich, P., Alexander, L. V., Della-Marta, P., Gleason, B., Haylock, M., Klein Tank, A. M. G., & Peterson, T. (2002). Observed coherent changes in climatic extremes during the second half of the twentieth century. *Climate Research*, *19*, 193–212.
 14. Klein, T. A., & Konnen, G. P. (2003). Trends in indices of daily temperature and precipitation extremes in Europe, 1946–1999. *Journal of Climate*, *16*, 3665–3680.
 15. Kostopoulou, E., & Jones, P. D. (2005). Assessment of climate extremes in the Eastern Mediterranean. *Meteorology and Atmospheric Physics*, *89*, 69–85.
 16. Krishna, L. V. (2014). Long term temperature trends in four different climatic zones of Saudi Arabia. *International Journal of Applied Science and Technology*, *4*(5), 233–242.
 17. Lorenz, R., Pitman, A. J., Donat, M. G., Hirsch, A. L., Kala, J., Kowalczyk, E. A., et al. (2014). Representation of climate extreme indices in the ACCESS1.3b coupled atmosphere-land surface model. *Geoscientific Model Development*, *7*(2), 545–567. <https://doi.org/10.5194/gmd-7-545-2014>.
 18. Manton, M. J., Della-Marta, P. M., Haylock, M. R., Hennessy, K. J., Nicholls, N., Chambers, L. E., Collins, D. A., Daw, G., Finet, A., Gunawan, D., Inape, K., Isobe, H., Kestin, T. S., Lefale, P., Leyu, C. H., Lwin, T., Maitrepierre, L., Ouprasitwong, N., Page, C. M., Pahalad, J., Plummer, N., Salinger, M. J., Suppiah, R., Tran, V. L., Trewin, B., Tibig, I., & Yee, D. (2001). Trends in extreme daily rainfall and temperature in Southeast Asia and the South Pacific: 1961–1998. *International Journal of Climatology*, *21*, 269–284.
 19. New, M., Hewitson, B., Stephenson, D., Tsiga, A., Kruger, A., Manhique, A., Gomez, B., Coelho, C., Masisi, D., Kululanga, E., Mbambalala, E., Adesina, F., Saleh, H., Kanyanga, J., Adosi, J., Bulane, L., Fortunata, L., Mdoka, M., & Lajoie, R. (2006). Evidence of trends in daily climate extremes over southern and west Africa. *Journal of Geophysical Research*, *111*, 1–11. <https://doi.org/10.1029/2005JD006289>.
 20. Vincent, L., Peterson, T., Barros, V., Marino, M., Rusticucci, M., Carrasco, G., Ramirez, E., Alves, L., Ambrizzi, T., Berlato, M., Grimm, A., Marengo, J., Molion, L., Moncunill, D., Rebello, E., Anunciacao, Y., Quintana, J., Santos, J., Baez, J., Coronel, G., Garcia, J., Trebejo, I., Bidegain, M., Haylock, M., & Karoly, D. (2005). Observed trends in indices of daily temperature extremes in South America 1960–2000. *Journal of Climate*, *18*(23), 5011–5023. <https://doi.org/10.1175/JCLI3589.1>.
 21. Frei, C. (2003). Statistical limitation for diagnosing changes in extremes from climate model simulations. *Proc. 14th Sympos. on global change and climate variations*. AMS Annual Meeting 2003, Long Beach, CA. On CD-ROM, p 6.
 22. Fisher, R., & Tippett, L. (1928). Limiting forms of the frequency distribution of the largest or smallest member of a sample. *Proceedings of the Cambridge Philosophical Society*, *24*, 180–190.
 23. Marimoutou, V., Raggad, B., & Trabelsi, A. (2009). Extreme value theory and value at risk: application to oil market. *Energy Economics*, Elsevier, *31*(4), 519–530.
 24. Davison, A. C., & Smith, R. L. (1990). Models for exceedances over high thresholds. *Journal of Royal Statistic Society Ser B*, *52*(3), 393–442.
 25. Eastoe, E. F., & Tawn, J. (2012). Modelling the distribution of the cluster maxima of exceedances of sub-asymptotic thresholds. *Biometrika*, *99*, 43–55.
 26. Katz, R., Parlange, M., & Naveau, P. (2002). Statistics of extremes in hydrology. *Advances in Water Resources*, *25*, 1287–1304.
 27. Naess, A. (1998). Estimation of long return period design values for wind speeds. *Journal of Engineering Mechanics*, *124*(3), 252–259.
 28. Chavez-Demoulin, V., & Embrechts, P. (2004). Smooth extremal models in finance and insurance. *Journal of Risk and Insurance*, *71*, 183–199. <https://doi.org/10.1111/j.0022-4367.2004.00085.x>.
 29. Gençay, R., & Selçuk, F. (2004). Extreme value theory and value-at-risk: relative performance in emerging markets. *International Journal of Forecasting*, *20*, 287–303.
 30. Raggad, B. (2009). Fundamentals of extreme value theory, its principal applications and its contribution to the risk management of the oil market. *Mathematics and Social Sciences*, *186*, 29–63 Online since 15 October 2009, connection on 04 March 2017. URL: <http://msh.revues.org/11069>.
 31. Bador, M., Naveau, P., Gilleland, E., Sánchez, M., & Arivelo, T. (2015). Spatial clustering of summer temperature maxima from the CNRM-CM5 climate model ensembles and E-OBS over Europe. *Journal of Weather and Climate Extremes*, *9*, 17–24. <https://doi.org/10.1016/j.wace.2015.05.003>.
 32. Coles, S. (2001). An introduction to statistical modelling of extreme values. Springer-Verlag.
 33. Furio, D., & Meneu, V. (2011). Analysis of extreme temperatures for four sites across Peninsular Spain. *Theoretical Applied Climatology*, *104*, 83–99.
 34. Jonathan, P., & Ewans, K. (2013). Statistical modelling of extreme ocean environments for marine design: a review. *Ocean Engineering*, *62*, 91–109.
 35. Katz, R. W., & Brown, B. G. (1992). Extreme events in a changing climate: variability is more important than averages. *Climatic Change*, *21*, 289–302.
 36. Katz, R., Brush, G., & Parlange, M. (2005). Statistic of extremes: modelling ecological disturbances. *Ecology*, *86*, 1124–1134.
 37. Keellings, D., & Waylen, P. (2014). Increased risk of heat waves in Florida: characterizing changes in bivariate heat wave risk using extreme value analysis. *Applied Geography*, *46*, 90–97.
 38. Smith, R. L. (1989). Extreme value analysis of environmental time series: an application to trend detection in ground-level ozone. *Statistical Science*, *4*(4), 389–393.
 39. Chavez-Demoulin, V., & Davison, A. C. (2012). Modelling time series extremes. *REVSTAT—Statistical Journal*, *10*, 109–133.
 40. Eastoe, E. F., & Tawn, J. A. (2009). Modelling non-stationary extremes with application to surface level ozone. *Journal of the Royal Statistical Society: Series C*, *58*(1), 25–45. <https://doi.org/10.1111/j.1467-9876.2008.00638.x>.
 41. McNeil, A. J., & Frey, R. (2000). Estimation of tail-related risk measures for heteroscedastic financial time series: an extreme value approach. *Journal of Empirical Finance*, *7*, 271–300.
 42. Chavez-Demoulin, V., & Davison, A. C. (2005). Generalized additive modelling of sample extremes. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, *54*(1), 207–222.
 43. Scotto, M. G., & Guedes Soares, C. (2000). Modelling the long-term time series of significant wave height with non-linear threshold models. *Coastal Engineering*, *40*(4), 313–327. [https://doi.org/10.1016/S0378-3839\(00\)00016-8](https://doi.org/10.1016/S0378-3839(00)00016-8).
 44. Clarke, R. T. (2002). Estimating trends in data from the Weibull and a generalized extreme value distribution. *Water Resources Research*, *38*(6), 1089. <https://doi.org/10.1029/2001WR000575>.
 45. Cooley, D. (2009). Extreme value analysis and the study of climate change, commentary on Wigley 1988. *Climatic Change*, *97*, 77–83.

46. El Adlouni, S., Ouarda, T. B. M. J., Zhang, X., Roy, R., & Bobée, B. (2007). Generalized maximum likelihood estimators for the non-stationary generalized extreme value model. *Water Resources Research*, *43*, W03410. <https://doi.org/10.1029/2005WR004545>.
47. Gaetan, C., & Grigoletto, M. (2007). A hierarchical model for the analysis of spatial rainfall extremes. *Journal of Agricultural, Biological, and Environmental Statistics*, *12*, 434–449.
48. Hall, P., & Tajvidi, N. (2000). Nonparametric analysis of temporal trend when fitting parametric models to extreme-value data. *Statistical Science*, *15*(2), 153–167.
49. Nogaj, M., Parey, S., & Dacunha-Castelle, D. (2007). Non-stationary extreme models and a climate application. *Nonlinear Processes in Geophysics*, *14*, 305–316.
50. Westra, S., & Sisson, S. A. (2011). Detection of non-stationarity in precipitation extremes using a max-stable process model. *Journal of Hydrology*, *406*(1), 119–128. <https://doi.org/10.1016/j.jhydrol.2011.06.014>.
51. Pal, J. S., & Eltahir, E. A. B. (2016). Future temperature in south-west Asia projected to exceed a threshold for human adaptability. *Nature Climate Change*, *6*, 197–200. <https://doi.org/10.1038/nclimate2833>.
52. Al-Saleh, M. A. (1997). Variability and frequency of daily rainfall in Riyadh, Saudi Arabia. *The Geographical Bulletin*, *39*, 48–57.
53. Embrechts, P., Kluppelberg, C., & Mikosch, T. (1997). *Modelling extremal events for insurance and finance*. Berlin: Springer.
54. Cannon, A. J. (2010). A flexible nonlinear modelling framework for non-stationary generalized extreme value analysis in hydroclimatology. *Hydrological Processes*, *24*, 673–685. <https://doi.org/10.1002/hyp.7506>.
55. Hasan, H., Salam, N., & Adam, M. B. (2014). Modelling extreme temperature in Malaysia using generalized extreme value distribution. World Academy of Science, Engineering and Technology. *International Journal of Mathematical, Computational, Natural and Physical Engineering*, *7*(6), 618–624.
56. Balkema, A., & De Haan, L. (1974). Residual life at great age. *Annals of Probability*, *2*(1974), 792–804.
57. Pickands, J. (1975). Statistical inference using extreme order statistics. *Annals of Statistics*, *3*, 119–131.
58. Ferro, C. A. T., & Segers, J. (2003). Inference for clusters of extreme values. *Journal of the Royal Statistical Society B*, *65*, 545–556.
59. Hsing, T., Husler, J., & Leadbetter, M. R. (1988). On the exceedance point process for a stationary sequence. *Probability Theory and Related Fields*, *78*, 97–112.
60. Walshaw, D. (2000). Modelling extreme wind speeds in regions prone to hurricanes. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, *49*(1), 51–62. <https://doi.org/10.1111/1467-9876.00178>.
61. Smith, R. L., Tawn, J. A., & Coles, S. G. (1997). Markov chain models for threshold exceedances. *Biometrika*, *84*, 249–268.
62. Fawcett, L., & Walshaw, D. (2003). Improved estimation for temporally clustered extremes. *Environmetrics*, *18*, 173–188.
63. Smith, R. L., & Weissman, I. (1994). Estimating the extremal index. *Journal of Royal Statistical Society: Series B*, *56*, 515–528.
64. Sugahara, S., Rocha, R. P., & Silveirac, R. (2009). Non-stationary frequency analysis of extreme daily rainfall in Sao Paulo, Brazil. *International Journal of Climatology*, *29*, 1339–1349.
65. Hosking, J. R. M. (1990). L-moments: analysis and estimation of distributions using linear combinations of order statistics. *Journal of the Royal Statistical Society, Series B*, *52*, 105–124.
66. Coles, S., & Powell, E. A. (1996). Bayesian methods in extreme value modelling: a review and new developments. *International Statistical Review*, *64*, 119–136.
67. Gilleland, E., & Katz, R. W. (2011). New software to analyze how extremes change over time. *Eos*, *92*(2), 13–14.
68. Burnham, K. P., & Anderson, D. R. (2002). *Model selection and multimodel inference: a practical information-theoretic approach* (2nd ed.). New York: Springer.
69. Grasa, A. A. (1989). *Econometric model selection: a new approach*. Netherlands: Springer.
70. Burnham, K. P., & Anderson, D. R. (2004). Multimodel inference: understanding AIC and BIC in model selection. *Sociological Methods Research*, *33*, 261–304.
71. Reiss, R., & Thomas, M. (1997). *Statistical analysis of extreme value with applications to assurance, finance, hydrology and other fields*. Basel: Birkhauser Verlag.
72. Tanarhte, M., Hadjinicolaou, P., & Lelieveld, J. (2015). Heat wave characteristics in the Eastern Mediterranean and Middle East using extreme value theory. *Climate Research*, *63*, 99–113.
73. Zhang, X., et al. (2005). Trends in Middle East climate extreme indices from 1950 to 2003. *Journal of Geophysical Research*, *110*, D22104. <https://doi.org/10.1029/2005JD006181>.