Probabilistic modeling and identifying fluctuations in annual extreme heatwave regimes of Karachi

Shahnillah Haider Rizvi1 · Muhammad Jawed Iqbal2 · Muhammad Ali3

Received: 11 July 2021 / Accepted: 21 August 2022 / Published online: 17 September 2022
© The Author(s), under exclusive licence to Springer-Verlag GmbH Austria, part of Springer Nature 2022

Abstract
Climatic warming in the global mean has significantly increased the probability of occurrence of heat extremes on time scales ranging from months to seasons. As extreme heat events are most likely to become intense and frequent over the next decades, it is important to examine these events to mitigate its negative impacts on public health and society. This study focuses on Karachi heat extremes over the last 23 years. The power spectral analyses of Karachi heat extremes records have been carried out by two indices: heat index (HI) and effective temperature index (TEE), which are also found to be significantly correlated. The result indicates a regular cyclic pattern of 4.5 years which is estimated to face a heat index of more than 73.63 °C, associated with the El Niño–Southern Oscillation (ENSO). Other peaks are observed at 2.8 and 2.2 years with the expected value of the Karachi heat index of 70.53 and 68.71 °C, respectively. The probabilistic approach is also used to predict the future heatwave events of Karachi. Generalized extreme value (GEV) distribution is found to be the best-fitted probability distribution for the extreme heatwave events on the basis of goodness-of-fit test. Furthermore, the estimation of the return period of the heatwave event reveals that Karachi will be facing a maximum heat index of 84.37 °C or more in the coming 33 years, which suggests an urgent need for mitigation strategies in Karachi to overcome the effects of extreme heatwave events.

1 Introduction
Extreme heat events have gained increasing attention for a few years because of the related effects on the increased mortality risk (IPCC 2012). Densely populated areas are most susceptible to synergistic effects of extreme heat and humidity. These conditions may produce lethal effects and exert significant influence on the human-habitable atmosphere (Pal and Eltahir 2015). The average temperature of a human body is 37 °C. Although, the body temperature can sustain at less than 35 °C for releasing heat (Sherwood and Huber 2010), the human skin cannot regulate the temperature below 35 and may contribute to develop troubles like cardiovascular disorders, hyperthermia and heat strokes.

The characterization and prediction of heat intensity, its varying frequency and duration of heat waves can be estimated in a proficient way through the combination of both statistical and physical approaches with large atmospheric driving forces in extreme value analyses (EVA). Extreme value theory (EVT) offers a statistical framework that is useful for the modeling of extreme events. It is one of the well-established techniques in the climate literature and has been extensively used for modeling of extreme meteorological and hydrological events (Goto-Maeda et al. 2008; Waylen et al. 2012; Keellings and Waylen 2014). Furthermore, most of the naturally occurring phenomena, specifically extreme climatic events are described through the frequency–functions which rely on fluctuations of the event. Variations in precipitation and temperature for different areas have also been extensively studied in this regard (Hasanean 2001; Morala et al. 2003; Tošić and Unkašević 2005). The intra-seasonal estimation of extreme events is beneficial in predicting future energy production and consumption (Roulston et al. 2003; Taylor and Buizza 2003), water resources plans
The metropolitan areas have become true hot spots of climate change because of the multiple environmental stresses, including perplexing phenomena of urban heat island, growing population, and air pollution. Karachi is the largest city in Pakistan which is more susceptible to extreme events. Previous study indicates that Karachi heat events are linked with the phenomenon of UHI (Rizvi et al. 2019) and intensified due to rapid rate of urbanization, deforestation, industrialization, and change in LULC pattern (Arshad et al. 2020). Moreover, the built-up areas of Karachi were increased by 40% from 2009 to 2017; consequently, the intensity of UHI was also increased (Rizvi et al. 2020a). The regions of the maximum intensity of surface UHI nearly correspond with dense urban areas in Karachi, which covers the area of 201.28 km² (Rizvi et al. 2020b). The heatwave event occurred in 2015 was one of the most severe climate extremes to hit Sindh province which caused more than 800 deaths in Karachi between 17 and 24 June 2015. An extreme event occurring across the province can increase the mortality rate and potential economic damages. Therefore, for enacting the mitigation strategies, a precise estimation of the risk in the region is needed.

Previous studies shows that heat related natural hazards and their associated socioeconomic impact in Pakistan have received considerable attention of researchers in past few years but they are mostly focused on projection models for future predictions (Arshad et al. 2020; Zahid and Rasul 2012). The complexities occur in the study of urban environment due to which the narrative of climate vulnerability becomes more complicated to understand, and are still to be addressed. The facts and causes of numerous climate migrants in Punjab has also been investigated in recent years (Mueller et al. 2014; Umer and Saeed 2018). The study suggests that in addition to the other factors such as economic opportunity, workers migrate to different cities due to heat stress impact of the climate on labor productivity. However, the profound literature on the occurrence of such phenomenon in the areas of Sindh is not available. This study focuses on the spectral analysis to assess the regular cyclic patterns of heat waves and examines the extreme heat events with its future variability in Karachi using generalized extreme value and Gumbel maximum distribution to estimate the long return periods and levels via the best-fitted model.

2 Materials and methods

2.1 Data

In this study, different climate indices are utilized for the multivariate analysis of heat extremes. All the indices are accessible for the entire study period, 1997–2019. The Nino 3.4 (hereafter, N3.4) index is generally used to describe ENSO state (Kenyon and Hegerl 2008). It represents sea surface temperature (SST) anomalies in the N3.4 region (5°N − 5°S; 170°W − 120°W), computed from the SST data taken from the Hadley Centre Sea Ice and Sea Surface Temperature (HadISST) data set, available at (https://www.esrl.noaa.gov/psd/data/correlation/nina34.data). The index has positive value during El Niño (warming phase of ENSO) but it shows negative value during La Niña (cooling phase of ENSO). Further details about N3.4 index can be obtained from Trenberth and Stepaniak (2001). The Quasi-Biennial Oscillation (QBO) index is also used in this study, which is identified as Singapore winds having the spatial extent of 60°Nto 60°S; 73°Eto 90°W, available at NOAA Climate Prediction Center (https://www.cpc.ncep.noaa.gov/data/indices/). The daily observations of maximum temperature, relative humidity, and wind speed are taken from Weather Underground Organization (WUO) to calculate the heat index (HI) and effective temperature index (TEE) of Karachi over a period of 23 years, during the summer season (AMJJ).

2.2 Heat index and effective temperature index

To estimate the heatwave intensity, various indices are used in environmental health research to analyze the combined effects of different climatic factors. In present study, two indices of extremely hot climate are investigated: (i) heat index (HI) which is also known as apparent temperature (ii) effective temperature index (TEE). Thus, the annual extreme values of HI are calculated through the succeeding regression equation which was achieved using multiple regression analysis (Rothfusz 1990):

$$HI = -42.379 + 2.049T + 10.143(RH) - 0.2247(T)(RH) - 6.83 \times 10^{-3}T^2 - 5.48 \times 10^{-2}(RH)^2 + 1.22 \times 10^{-3}(T)^2(RH) + 8.5 \times 10^{-4}(T)(RH)^2 - 1.99 \times 10^{-6}T^2(RH)^2,$$

where $T$ represents the temperature in degree Fahrenheit ($^\circ F$) and RH shows relative humidity in percent. HI represents the heat index indicated as an apparent temperature in degree Fahrenheit ($^\circ F$). To examine the influence of wind speed on Karachi heat extreme, the annual extreme values of effective temperature index (TEE) are also computed as (Gregorczuk 1968):

$$TEE = 37 - \frac{37 - t}{0.68 - 0.0014(RH)} - 0.29\nu(1 - 0.01RH),$$

where $\nu$ is the wind speed in m/s, $t$ represents temperature in $^\circ C$, and RH shows Relative Humidity. The rating scales of TEE varies with the latitude of the region, such as for UK, 14.4°–20.6°; for Yakovenko region of Russian Federation,
16.7°–21.8°; for USA, 18°–22°, and for tropical regions it varies from 23.3° to 29.4° (Teodoreanu and Bunescu 2007) (Fig. 1).

2.3 Spectral analysis

Statistical spectral analysis has been traditionally used to detect regular cyclical patterns of the time series. In this study, Fourier-based spectral analysis has been chosen for the assessment to recover the periodicities in the Karachi heat extreme which has been used in previous climatological studies (Negrón-Juárez and Liu 2001; Aldriana and Djamila 2008; Iqbal et al. 2011; Hasanean et al. 2013). The annual HI records are transformed through a finite Fourier transform that decomposes it in different frequency waves. The power spectrum of the considered time series of heat extreme is computed via

Fig. 1 Timeseries of an annual extreme heat index (HI) and effective temperature index (TEE) of Karachi for the last 23 years (1997–2019)
autocorrelation spectral analysis (ASA), which is the Fourier transform of the autocorrelation function of HI time series (Blackman and Tukey 1959). As the autocorrelation function possesses even symmetry, the terms having sine function in Fourier series have become zero. As a result, the power spectrum ($B_i$) of the natural phenomenon becomes simplified, containing only real cosine terms:

$$B_i = \frac{r_i}{n} + \frac{2}{n} \sum_{l=1}^{n-1} r_l \cos \left( \frac{360}{2n} il \right) + (-1)^i \frac{r_n}{n}, \quad i \in \{1, 2, 3, \ldots, n\}, \quad (3)$$

where $r_i$ and $n$ represents the autocorrelation function of lag $L$ and sample size, respectively (Sprott 2003). The above spectrum function does not provide a reliable estimate until it will be smoothened by using lag features in a frequency domain. The latter estimates ($F_i$) can be computed using the smoothing technique called Hamming method with a five-term weighted average (Chatfield 1989)

$$F_i = 0.0357B_{i-2} + 0.2411B_{i-1} + 0.4464B_i + 0.2411B_{i+1} + 0.0357B_{i+2}, \quad (4)$$

### 2.4 Extreme value distribution

The time series of extreme events are usually analyzed through extreme value theory (EVT), particularly, in the fields of meteorology, hydrology and in the prediction of financial time series (Rusticucci and Tencer 2008; Iqbal and Ali 2012; Mandal and Choudury 2015; Du et al. 2019). This theory provides the statistical framework that can estimate the occurrence probability of an extreme event with its corresponding return period in risk analysis (Wilks 2006). In this study, the daily heat index is denoted by $X_1, X_2, \ldots, X_n$ the conventional model for extremes is achieved by observing the behavior of $M = \max \{X_1, X_2, \ldots, X_n\}$ where the annual maxima is represented by $n = 23$, and $M$. The generalized extreme value distribution (GEV) is formulated in the EVT framework whose probability density function (PDF) is given by:

$$f(x) = \frac{1}{\beta} (1 + Z)^{-1-\frac{1}{\beta}} e^{-[1 + (Z)^{-\frac{1}{\beta}}]}, \quad 1 + Z > 0, \quad (5)$$

where $Z = \frac{k(x - \xi)}{\beta}$,

where $k$, $\beta$ and $\xi$ are the shape, scale and location parameters, respectively. By integrating the above equation, the following CDF function $F(x)$ can be obtained (Coles 2001).

For the purpose of the explicit formulation of the quantile function, Eq. (6) can be inverted which gives

$$F(x) = e^{-[(1 + Z)^{\frac{1}{k}}} \quad (6)$$

$$X = F^{-1}(p) = \xi + \frac{\beta}{k} \{[-\ln(P)]^{-k} - 1\}, \quad (7)$$

where $P = F(x)$, represents cumulative probability. The L-moment or maximum likelihood method (MLM) is applied to estimate the parameters of GEV distribution (Hosking 1990; Stedinger et al. 1993). The method of L-moment is ideal for a small sample size (Hosking 1990), but the MLM can be simply modified to include covariate effects, or other factors, such as the possibility of some distribution estimates to show a trend caused by climate change (Katz et al. 2005; Zhang et al. 2005). However, both methods show similar results for large and moderate sample data. The GEV distribution consists of three cases based on the shape parameter. The type-I extreme value distribution is a special case having ($k = 0$) which is also called Gumbel distribution. Gumbel primarily focused on EVT application to engineering challenges, particularly in the modeling of climatological phenomena like annual extreme flood flows. The distribution function of Gumbel maximum is given as:

$$f(x) = \frac{1}{\beta} e^{[-(x - \xi) / \beta]}, \quad (8)$$

where $\xi$ and $\beta$ are location and scale parameters, respectively. The shape of this distribution is independent of its parameters. The Gumbel distribution probabilities can be found by its cumulative distribution function (CDF)

$$F(x) = e^{-(x - \xi) / \beta}. \quad (9)$$

To obtain the quantile function, the inverse of CDF of Gumbel distribution is given by

$$F^{-1}(X) = \xi - \beta \ln \{\ln \left( \frac{1}{1 - P} \right)\}. \quad (10)$$

### 2.5 Testing of the probability distribution model

The Goodness of Fit tests measures the compatibility between a theoretical PDF and a random sample. Thus, the results of these tests reveal how well a certain distribution fits the data set. The goodness-of-fit test is employed to test the null hypothesis, $H_0$: the annual maximum daily heat index follow the particular distribution and the corresponding alternative hypothesis is, $H_A$: the annual maximum daily heat index does not follow the particular distribution. It is also highlighted that the test statistics of different goodness-of-fit test can produce a varied ranking position among the pdfs (Zhou et al. 2010; Chang 2011). Though the Kolmogorov–Smirnov (K–S) test is extensively applied (Lo brano...
et al. 2011; Vivekanandan et al. 2012; Whan et al. 2015), so the distribution is selected on the basis of the (K–S) test at 5% level of significance. The methodology of the K–S test is as follows:

### 2.5.1 Kolmogorov–Smirnov (K–S) test

The supreme class of empirical distribution function (EDF) statistics is a K–S test statistic, originated on the maximum vertical difference among the empirical and theoretical distributions (Conover 1999). Comparison of the CDF of an assumed theoretical distribution \( F_X(x) \) and the empirical cumulative frequency \( S_n(x) \) is a fundamental objective of this test. The K–S test statistic \( D_n \) is the maximum difference between \( F_X(x) \) and \( S_n(x) \). For ‘n’ samples, the data set is arranged in ascending order, \( X_1 < X_2 < \ldots X_n \), and the statistic of (K–S) test is evaluated for each ordered value

\[
S_n(x) = \begin{cases} 0; & \text{if } X < X_1, \\ \frac{k}{n}; & \text{if } X_k \leq X < X_{k+1}, \\ 1; & \text{if } X > X_n, \end{cases}
\]

\[D_n = \max[F_X(x) - S_n(x)].\]  

(11)

\[ P(D_n \leq D^*_n) = 1 - \alpha, \]

where \( D^*_n, \alpha, \) and \( k \) represent critical value, significance level and the ranking order of the data, respectively.

### 3 Results and discussion

#### 3.1 Power spectrum and coherency

The power spectral analysis has been applied by the Fourier transform method to the annual time series of the Karachi heat index for the period of 1997–2019. The HI series is converted from real space (time-domain) to Fourier space (frequency domain). The Fourier transform provides a 1–1 correspondence into the signals at particular times and examines how specific frequency contributes to the signal. Hence, it would be possible to determine the signal statistics in Fourier space instead of real space. First, the time series of HI is linearly detrended, and then the frequency spectrum is estimated by smoothing a periodogram. As the smooth periodogram can identify the specific frequencies with high spectral densities that have a significant contribution in the cyclic pattern of series. The outcomes of the FFT spectral analysis are depicted in Fig. 2 which shows that the annual heat index series of Karachi for 23 years have the dominant peaks around 4.5, 2.8, and 2.2 years, and annual TEE series have the dominant peaks at 4.5 and 2.8 years, which reveals the regular cyclic behavior of heat events. This result is compatible with the results obtained in previous studies. It can be concluded that the number of heat wave events and its duration has been increased in Pakistan (Ali et al. 2018; Khan et al. 2020).

The time series of HI and TEE has a significant correlation of 0.86 at 0.01 level and both series show an almost similar return period. The 2.8-year middle frequency wave and 2.2-year short wave show some association with the quasi-biennial oscillations (Lamb 1972; Iqbal et al. 2011). There would be multiple time scales in the trend and interannual variability of the monsoon–atmosphere–ocean systems such as a quasi (2-year) cycle linked with tropical biennial oscillation (TBO), along with (4–6) year cycle related with ENSO. Thus, the other peak at 4.5 years seems to be related to ENSO. The El Niño southern oscillation (ENSO) is recognized as a nexus of climate change throughout the world (Glantz et al. 1991; Allan 2000; Oldenborgh et al. 2005; Stevenson et al. 2017). Studies also reveal that oceanic forces had dominant effects on the extreme heatwave events of the twentieth century (De and Mukhopadhyay 1998; Della-Marta et al. 2007; Jenamani 2012). To determine the association between the climate and ENSO phenomena, different circulation indices have been developed. El Niño phase is linked with an increased chance of extreme heat events while the La Niña phase is associated with a reduced chance of extreme heat events (White et al. 2014).

To assess the association between the spectral peak frequencies of Karachi heat extreme and the large-scale atmospheric system, spectral coherence analysis has been performed. The coherence function reveals whether or not variation in a particular series responds to a similar kind of variability in the other series. It is the correlation among different variables in the frequency domain and its value is normalized between 0 and 1. For a small number of samples, i.e., \( n = 23 \), the significant coherence and coherence-squared at 95% confidence interval are 0.35 and 0.12, respectively (Thompson 1979). Figure 3 shows a lack of correlation among HI and QBO, for the frequencies, lie between 0.26 and 0.37. On the other hand, high coherence is observed at 0.46. The significant squared coherence is observed among the frequencies of heat index series and N3.4 frequencies except 0.26–0.29 (Fig. 4). The obtained results related to the impact of ENSO on Karachi heatwaves are in line with the previous studies focused on the occurrence of heatwaves in any specific region (De and Mukhopadhyay 1998; Parker et al. 2014). Perkins et al. (2015) also claims that ENSO-strength provides a good prediction for the expected heatwave days when the N3.4 index increases positively;
consequently, the number of heatwave days increases. Another study of Pearson’s correlation analysis revealed that ENSO driven changes cause extremes in the environment of the region. In particular, the arid and semi-arid climate zones of Pakistan are observed to have extremely high temperature events. The severity of such events is greatly influenced by the potential stability of La Niña occurrence (Saleem et al. 2021).
3.2 Probability analysis of annual maximum daily heat index

The annual HI data of Karachi is utilized for extreme value analysis, as both indexes have a significant correlation. A threshold value of the heat index is 40.5 °C as mentioned by NOAA (National Oceanic and Atmospheric Administration) corresponds to danger and extreme danger conditions. During the study period, the highest annual value of HI was 84.37 °C, observed in 2010. However, there exist some other relative maxima of HI such as 81.63, 78.70, 76.39 °C. This is due to the reason that the summer season in Karachi starts earlier than the usual and stays for longer time period now. It is observed that the daytime ambient temperature average exceeded 35 °C in
April (Ullah et al. 2019) and $T_{\text{max}}$ recorded in June also averaged above 35 °C (Anwar et al. 2022).

To predict the future heatwave events using extreme value analysis, a univariate extreme value model is required that focuses on heat index observations. Statistical goodness-of-fit techniques and graphical representation of PDFs are capable of revealing the consistency of the fitted distributions with a certain data set (Stedinger et al. 1993). This section comprises the comparison of results for the performance of each probability distribution to obtain a best-fitted probability distribution that could efficiently provide the estimate of HW in Karachi. The GEV is suggested as an appropriate distribution for the representation of extreme events because of its theoretical basis for maxima analysis and the procedure of estimating the model’s parameter is quite simple. However, the Gumbel model has also been used to fit the annual maximum values of different indices including temperature (Kharin and Zwiers 2000; Hong et al. 2013).

Figure 5 presents the distribution function of the GEV and Gumbel maximum distributions, fitted to the extreme HW events to visually assess the level of Goodness-of-fitting for both distributions. It is observed from the PDF plot that both distribution models fit on the HI data quite well. Furthermore, the K–S test has been used to examine the goodness-of-fit. The test statistics ($D$) from the K–S goodness-of-fit test using GEV and Gumbel for extreme HW events are 0.086 and 0.087, respectively, which are smaller than the 95th percentile value of 0.274 (critical value) with 5% significance level. As the test statistic value in the goodness-of-fit techniques specifies the difference among the set of observations and the fitted distributions, it is evident that the model having the lowest test statistic is a good fit. Hence, it is concluded on the basis of the K–S goodness-of-fit test that the GEV is the best-fitted probability distribution for the extreme HW events.

After the selection of best-fitted distribution, the probability–probability (P–P) plot has been plotted for the extreme annual heatwave of Karachi in Fig. 6 to validate the results of the above goodness-of-fit test. It shows the relation between empirical and theoretical CDF values to identify how much a specific distribution agrees with HI data. If theoretical distribution fits precisely then the P–P plot will be roughly linear. The P–P plots are more advantageous than quantile–quantile (Q–Q) plots since the respective plotting positions are unbiased for the examined distributions (Soukissian 2013). The result also reveals that the theoretical CDF values of GEV distribution have less deviation from the observed data points than Gumbel maximum distribution. Hence, it is concluded that the GEV distribution provides the best fit for the probabilistic modeling of heatwave events.

### 3.3 Estimates of the return period of heatwave events

The outcomes of extreme value analysis (EVA) can be considered to be those quantiles which correspond to large cumulative probabilities, like an event that has an annual exceedance probability of 0.01. The direct assessment of these extreme quantiles is not possible unless $n$ is a large number; however, a best-fitted extreme value distribution presents a practical approach for extrapolation of the
probabilities that may be significantly larger than $1 - \frac{1}{n}$. The average return periods of such extreme probabilities are often expressed as (Wilks 2011):

\[
\text{Return}(R) = \frac{1}{\omega [1 - F(x)]},
\]  

Return periods $R(x)$ related to a quantile is considered to be the average period concerning the occurrences of extreme
events of that extent or more. It is a function of the average sampling frequency $w$ and the CDF of a specific distribution. For annual maximum observations $w = 1$/year, thus the event $x$ having the cumulative probability $F(x) = 0.99$ will be the probability of exceedance $1 - F(x)$ at any certain time. The return period of this event would be 100 years and could be named as a 100-year event. The results reveal that the highest heat index of 84.37 °C is observed on June 19, 2010, during the whole study period from 1997 to 2019. The return period is estimated to determine the number of years after which there will be a possibility of occurrence for the maximum heatwave in the future. Therefore, the fitting of the GEV distribution to the annual maximum daily heat index data of Karachi yielded the parameter $k = -0.11$, $\beta = 6.19$, and $\xi = 65.69$. To estimate the cumulative probability Eq. (6) is used

$$F(x) = P(x \leq 84.37) = e^{-(1+\xi)\frac{x}{\beta}}.$$

$$F(x) = 0.97.$$

Equation (12) yields;

$$R(x) = \frac{1}{\Gamma[1-0.97]} = 33 \text{ (years)}.$$

The above calculation reveals that there is a chance for occurring heat index equals to 84.37 °C or more in the coming 33 years. However, return levels are estimated against the different return periods by using Eq. (7). The regular cyclic pattern of 4.5 years for future heatwaves is estimated in Sect. 3.1 through spectral analysis i.e.; at 4.5 years, $p = 1/1.45 = 0.22$ and $F(x) = P = 1 - 0.22 = 0.78$, substituting these values in the inverse CDF of GEV Distribution yields HI $= 73.63^\circ$C, which shows that in the next 4.5 years, the expected value of Karachi heat index will be 73.63°C, with a 95% confidence interval [69.82, 77.43]. Similarly, a small peak of 2.8 years, observed from the spectral analysis is projected to have a heat index of 70.53°C, with a 95% confidence interval [67.09, 73.96]. These results are similar to the results obtained in the previous studies. This reveals that during the years 2060–2099, the recurrence of HW’s is estimated to increase up to 12 events per annum, whereas the duration of each event is estimated to increase up to 100 days in a year in case of highest emission scenario (Khan et al. 2020). Further calculations are summarized for different return levels against the different return periods with a 95% confidence interval in Table 1. Here, the sample size of data is quite small compared to the return levels. Anticipating up to 100-year return periods with only 23 years of data available unavoidably yields much lower confidence in the results as the return period increases.

### 4 Summary and conclusion

Large-scale atmospheric systems can greatly affect the local temperature and extreme heatwave events. The analysis in this study reveals that besides anthropogenic heat, ENSO is also a significant aspect of varying HW activity. To estimate the periodicity of heatwave events, the HI data is subjected to power spectrum analysis and the results reveal the regular cyclic pattern of 4.5, 2.8, and 2.2 years seem to be affected by the well-known phenomenon of ENSO and QBO. As it has been observed that the periodicities of heat index series are linked with El Niño, this hypothesis is greatly strengthened by computing coherence between the annual HI series with QBO and an El Niño index which shows significant and strong coherence among them. Results indicate a prolonged association of these indices on Karachi heat extremes. The frequency of occurrence of extreme heatwave events in Karachi is also examined by comparing the performance of GEV and Gumbel maximum probability distributions to identify an appropriate model that could indicate extreme HW estimates, accurately. The performance of fitted distribution is evaluated by the statistical goodness-of-fit K–S test in which the selected distributions are ranked to identify the best-fitted distribution to the HI data of Karachi city. Among both probabilistic models, GEV is spotted as the best probability distribution to be fitted on the basis of (K–S) test.

Furthermore, the estimation of the return period of the heatwave event reveals that Karachi will be facing a heat index of 78.03 °C or more in the coming 10 years, while a 50-years return period is estimated to face more than 85.33 °C heat index. This paper provides the future prediction of extreme heatwave events in Karachi to establish early warning alerts for residents and city planners to endorse.
smart sustainable strategies. The results that are drawn in this study contribute to open the doors of awareness towards the severity of climate extremes. They also highlight the causes, impact and some safety measures for the case of one of the vulnerable cities of Pakistan–Karachi. It is suggested to investigate the heat wave regime in more detail including urban and rural regions. Moreover, the research can be extended further on the precautionary measures and the policy making regarding heat wave events in the country. In addition, sector and locality-wide global development programs are required to modify to overcome the impact of climate change together with heat waves.

**Funding** No funding was received for conducting this study.

**Data availability** The data and material used in this research are partly available in the body of the article; other data/material is also available to public and will be provided through writing to the corresponding author.

**Declarations**

**Conflict of interest** The authors declare that they have no competing interests.

**References**

Aldriana E, Djamila YS (2008) Spatio-temporal climatic change of rainfall in East Java Indonesia. Int J Climatol 28:435–448. https://doi.org/10.1002/joc.1543

Ali J, Syed KH, Gabriel HF, Saeed F, Ahmad B, Bukhari SAA (2018) Centennial heat wave projections over Pakistan using ensemble NEX GDDP data set. Earth Syst Environ (2):437–454. https://doi.org/10.1007/s41748-018-0064-8

Allan R (2000) ENSO and climatic variability in the past 150 years. In: Diaz H, Markgraf V (eds) El Niño and the Southern Oscillation: multiscale variability and global and regional impacts. Cambridge University Press, Cambridge, pp 3–35

Anwar NH, Khan HF, Abdullah A, Macktoom S, Fatima A (2022) Designed to fail? Heat governance in urban South Asia: the case of Karachi a scoping study. Edinb Res Arch. https://doi.org/10.1007/s12517-021-0548-z

Ashraf M, Arshad A, Sundari RS, Qamar H, Wajid M, Hasan M (2020) Anwar NH, Khan HF, Abdullah A, Macktoom S, Fatima A (2022) Designed to fail? Heat governance in urban South Asia: the case of Karachi a scoping study. Edinb Res Arch. https://doi.org/10.1007/s12517-021-0548-z

Chang TP (2011) Estimation of wind energy potential using different probability density functions. Appl Energy 88(5):1848–1856. https://doi.org/10.1016/j.apenergy.2010.11.010

Chatfield C (1989) The analysis of time series: an introduction, 4/e. Chapman and Hall, New York

Coles S (2001) An introduction to statistical modeling of extreme values. Springer, London. https://doi.org/10.1007/978-1-4471-3675-0

Conover WJ (1999) Practical nonparametric statistics, 3rd edn. Wiley, New York, pp 428–433

De US, Mukhopadhyay RK (1998) Severe heat wave over Indian sub-continent in 1998 in a perspective of global climate. Curr Sci 75(12):1308–1311

Della-Marta PM, Luterbacher J, Weissenhuh HV, Xoplaki E, Brunet M, Wanne H (2007) Summer heat waves over western Europe 1880–2003, their relationship to large-scale forcing and predictability. Clim Dyn 29:251–275. https://doi.org/10.1007/s00382-007-0233-1

Du S, Tang G, Li S (2019) Risk measurement of chinese stock market based on GARCH model and extreme value theory. Open J Bus Manag 7(2):963–997. https://doi.org/10.4236/ojbm.2019.72065

Glantz MH, Katz RW, Nicholls N (eds) (1991) Teleconnections linking worldwide climate anomalies: scientific basis and societal impact. Cambridge University Press, Cambridge

Goto-Maeda Y, Shin DW, O’Brien JJ (2008) Freeze probability of Florida in a regional climate model and climate indices. Geophys Res Lett 35:L11703. https://doi.org/10.1029/2008GL033720

Gregorczuk M (1968) Bioclimates of the world related to air enthalpy. Int J Biometeorol 12(1):35–39. https://doi.org/10.1007/BF01525976

Hasanean HM (2001) Fluctuations of surface air temperature in the Eastern Mediterranean. Theor Appl Climatol 68:75–87. https://doi.org/10.1007/s007040170055

Hasanean HM, Almazroui M, Jones PD, Alamoudi AA (2013) Siberian high variability and its teleconnections with tropical circulations and surface air temperature over Saudi Arabia. Clim Dyn 41:2003–2018. https://doi.org/10.1007/s00382-012-1657-9

Hong HP, Li SH, Mara TG (2013) Performance of the generalized least-squares method for the Gumbel distribution and its application to annual maximum wind speeds. J Wind Eng Ind Aerodyn 119:121–132

Hosking JRM (1990) L-moments: analysis and estimation of distributions using linear combinations of order statistics. J R Stat Soc B 52(1):105–124

IPCC (2012) Managing the risks of extreme events and disasters to advance climate change adaptation, a special report of working groups I and II of the intergovernmental panel on climate change. Cambridge University Press, Cambridge. p 582

Iqbal MJ, Ali M (2012) A probabilistic approach for estimating the return period of extreme annual rainfall in different cities of Punjab. Arab J Geosci 6:2599–2606. https://doi.org/10.1007/s12517-012-0548-z

Iqbal MJ, Quamar J, Youusufzai MAK (2011) Spectral analysis of local climatic fluctuations. Arab J Geosci 4:291–298. https://doi.org/10.1007/s12517-010-0154-x

Jenamani RK (2012) Analysis of Ocean-Atmospheric features associated with extreme temperature variation over east coast of India—a special emphasis to Orissa heat waves of 1998 and 2005. Mavasum 63(3):401–422

Jewson S, Caballero R (2003) The use of weather forecasts in the pricing of weather derivatives. Meteorol Appl 10:377–389. https://doi.org/10.1017/S1350482703001099

Katz R, Brush G, Parlane M (2005) Statistics of extremes: modeling ecological disturbances. Ecology 86(5):1124–1134

Keelings D, Waylen P (2014) Increased risk of heat waves in Florida: characterizing changes in bivariate heat wave risk using extreme value analysis. Appl Geogr 46:90–97. https://doi.org/10.1016/j.apgeog.2013.11.008

Kenyon J, Hegerl GC (2008) Influence of modes of climate variability on global temperature extremes. J Clim 21:3872–3889. https://doi.org/10.1175/2008JCLI2125.1

Khan N, Shahid S, Ahmed K, Wang X, Ali R, Ismail T, Nawaz N (2020) Selection of GCMs for the projection of spatial distribution of heat waves in Pakistan. Atmos Res 233:104688. https://doi.org/10.1016/j.atmosres.2019.104688

Kharin VV, Zwiers FW (2000) Changes in the extremes in an ensemble of transient climate simulations with a coupled atmosphere-ocean GCM. J Clim 13:3760–3788. https://doi.org/10.1175/1520-0442(2000)1013%3C3760:CTIE32.0.CO;2
Lamb HH (1972) Climate: present, past, and future, I: fundamentals and climate now. Methuen, London

Lo Brano V, Orioli A, Ciulla G, Culotta S (2011) Quality of wind speed fitting distributions for the urban area of Palermo Italy. Renew Energy 36(3):1026–1039. https://doi.org/10.1016/j.renene.2010.09.00

Mandal S, Choudhury BU (2015) Estimation and prediction of maximum daily rainfall at Sagar Island using best-fit probability models. Theor Appl Climatol 121:87–97. https://doi.org/10.1007/s00704-014-1212-1

Negrón-Juárez RI, Liu WT (2001) FFT analysis on NDVI annual cycle. J Atmos Sci 58:219–232.

Mueller V, Gray C, Kosec K (2014) Heat stress increases long-term human migration in rural Pakistan. Nat Clim Change 4(3):182–185

Mueller V, Gray C, Kosec K (2014) Heat stress increases long-term human migration in rural Pakistan. Nat Clim Change 4(3):182–185.

Mørk-Andreassen I, Lien AG (2013) Distribution of precipitation and climate feedbacks. Theor Appl Climatol 112:575–593.

Negrón-Juárez RI, Liu WT (2001) FFT analysis on NDVI annual cycle. J Atmos Sci 58:219–232.

Perkins SE, Argüeso D, White CJ (2015) Relationships between climate variability, soil moisture, and Australian heatwaves. J Geophys Res Atmos 120:8144–8164. https://doi.org/10.1002/2015JD023592

Rizvi SH, Alam K, Iqbal MJ (2019) Spatial-temporal variations in urban heat island and its interaction with heat wave. J Atmos Sol Terr Phys 185(2019):50–57

Rizvi SH, Fatima H, Iqbal MJ, Alam K (2020a) The effect of urbanization on the intensification of SUHIs: analysis by LULC on Karachi. J Atmos Sol Terr Phys 207:105374. https://doi.org/10.1016/j.jastp.2020.105374

Rizvi SH, Fatima H, Alam K, Iqbal MJ (2020b) The surface urban heat island intensity and urban expansion: a comparative analysis for the coastal areas of Pakistan. Environ Dev Sustain. https://doi.org/10.1007/s10668-020-00828-5

Rothfusz LP (1990) The heat index “equation” NWS tech. Attatch SR 2:90–23

Roulston MS, Kaplan DT, Hardenberg J, Smith LA (2003) Using a range of weather forecasts to improve the value of wind energy production. Renew Energy 28:585–602. https://doi.org/10.1016/S0960-1481(02)00054-X

Rusticiucci M, Tencer B (2008) Observed changes in return values of precipitation products against in-situ observations over different climate regions of Pakistan. Remote Sens 11(6). https://doi.org/10.3390/rs11060628

Rizvi SH, Fatima H, Iqbal MJ, Alam K (2020a) The effect of urbanization on the intensification of SUHIs: analysis by LULC on Karachi. J Atmos Sol Terr Phys 207:105374. https://doi.org/10.1016/j.jastp.2020.105374

Sankarasubramanian A, Lall U, Espinueva S (2009) The role of monthly updated climate forecasts in improving intraseasonal water allocation. J Appl Meteorol Clim 48:1464–1482. https://doi.org/10.1175/2009JAMC2122.1

Sherwood SC, Huber M (2010) An adaptability limit to climate change due to heat stress. Proce Natl Acad Sci USA 107(21):9552–9555. https://doi.org/10.1073/pnas.0913352107

Soukissian T (2013) Use of multi-parameter distributions for offshore wind speed modeling: the Johnson SB distribution. Appl Energy 111:982–1000. https://doi.org/10.1016/j.apenergy.2013.06.050

Sprott JC (2003) Chaos and time series analysis. Oxford University Press, Oxford, pp 226–291

Stedinger JR, Vogel RM, Foufoula-Georgiou E (1993) Frequency analysis of extreme events. Handbook of hydrology. McGraw-Hill, New York

Stovall BD, Thompson RORY (1979) Notes and correspondence: coherence significance levels. J Atmos Sci 36:2020–2021

Taylor JW, Buizza R (2003) Using weather ensemble predictions in electricity demand modeling. Int J Forecast 19:57–70. https://doi.org/10.1016/S0169-2070(01)00123-6

Teodoreanu E, Bunescu I (2007) Thermal comfort. Present Environ Sustain Dev 1:135–142

Thompson RORY (1979) Notes and correspondence: coherence significance levels. J Atmos Sci 36:2020–2021

Tošic I, Uništečević M (2005) Analysis of precipitation series for Belgrade. Theor Appl Climatol 8:67–77. https://doi.org/10.1007/s00704-004-0076-1

Trenberth KE, Stepaniak DP (2001) Indices of El Niño evolution. J Clim 14:1697–1701. https://doi.org/10.1175/1520-0442(2001)014%3c1697:LIENOS%3e2.0.CO;2

Ullah W, Wang G, Ali G, Hagan DFT (2019) Comparing multiple precipitation products against in-situ observations over different climate regions of Pakistan. Remote Sens 11(6). https://doi.org/10.3390/rs11060628

Umar MA, Saeed F (2018) The role of heat stress in migration decisions: a case study of Faisalabad Islamabad. Sustainable Development Policy Institute, Islamabad

Van Oldenborgh GJ, Philip S, Collins M (2005) El Niño in a changing climate: a multimodel study. Eur Geosci Union 1(2):81–95 (hal-00298272)

Vivekanandan N, Roy SK, Jagtap RS (2012) Assessment of rainfall and temperature using osa estimators of extreme value distributions. Bonfiring Int J Softw Eng Soft Comput 2(3):16–21

Waylen PR, Keellings D, Qin Y (2012) Climate and health in Florida: changes in risks of annual maximum temperatures in the second half of the twentieth century. Appl Geogr 33:73–81. https://doi.org/10.1016/j.apgeog.2011.06.007

Whan K, Zscheischler J, Orth R, Shongwe M, Rahimi M, Asae EO, Seneviratne SI (2015) Impact of soil moisture on extreme maximum temperatures in Europe. Weather Clim Extreme 9:57–67. https://doi.org/10.1016/j.wace.2015.05.001

White CJ, Hudson D, Alves O (2014) ENSO, the IOD and the intraseasonal prediction of heat extremes across Australia using POAMA-2. Clim Dyn 43:1791–1810

Wilks DS (2006) Statistical methods in the atmospheric sciences. Cornell University, 2nd edn. Elsevier, Amsterdam

Wilks DS (2011) Parametric probability distributions. Statistical methods in the atmospheric sciences. International geophysics series, vol 100, 3rd edn. Elsevier, Amsterdam

Zahid M, Rasul G (2012) Changing trends of thermal extremes in Pakistan. Clim Change 113(3):883–896

Zeng L (2000) Weather derivatives and weather insurance: concept, application, and analysis. Bull Am Meteorol Soc 81:2075–2082. https://doi.org/10.1175/1520-0442(2000)081%3c2075:WDWAIC%3e2.0.CO;2

Zhang X et al (2005) Trends in Middle East climate extreme indices from 1950 to 2003. J Geophys Res 110:D22104. https://doi.org/10.1029/2005JD006181

Zhou J, Erdem E, Li G, Shi J (2010) Comprehensive evaluation of wind speed distribution models: a case study for North Dakota sites. Energy Convers Manag 51(7):1449–1458. https://doi.org/10.1016/j.enconman.2010.01.020

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.