



Quantifying the Rarity of Extreme Wildfires: Translating Fire Size into Return Period Using Extreme Value Theory Methods

Nikos Koutsias¹ · Frank A. Coutelieris¹

Received: 13 February 2025 / Accepted: 24 June 2025
© The Author(s) 2025

Abstract

This study presents a comparative study of Extreme Value Theory (EVT) concepts, applied to wildland fire data collected in Greece over a 25-year period (1983 – 2007). The dataset comprises 28,658 fire records from the Hellenic Forest Service, including details such as coordinates and size of the burned area. The primary objective is to develop a method for quantifying extreme values and estimating return periods and thus translating fire size into a more interpretable measure of fire significance. The study applies and evaluates both the block maxima generalized extreme value (GEV) approach and the peaks over threshold (POT) approach, using frequentist and Bayesian frameworks. The results from the GEV distribution indicate an asymptotic leveling off at a constant value, suggesting that the burned area size remains unchanged over time. This outcome is clearly unrealistic, implying that the GEV model provides reliable predictions only for time frames closely aligned with the period during which the empirical data were collected. Although conceptually different, the Bayesian POT approach also demonstrates unrealistic behavior over longer periods, with the burned area size continuously increasing along a power-law curve. In contrast, the frequentist POT model performs well with the available data, providing a more realistic and accurate representation of the increasing burned area size over time. According to the frequentist POT approach, the largest forest fire ever recorded in Greece (Dadia Forest in 2023)—covering approximately 95,000 hectares—corresponds to a return period of 2000 years. Over the past few years, many wildland fires in Greece have resulted in exceptionally large size of burned areas.

Keywords Extreme values analysis · GEV · POT · Bayesian · Wildfires · Return period

1 Introduction

When a wildfire occurs and the event is announced, especially to the general public, the size of the fire becomes one of the most important characteristics communicated. This statistical value provides people with an understanding of the fire's magnitude and significance. To achieve this, individuals already have an idea of what a particular fire size means in terms of scale, based on their empirical knowledge derived from the fire history of the area. In essence, the interpretation of fire size varies based on the local distribution of fire sizes to which the area is exposed. For example, a 5000

Ha fire would have a different meaning if occurred in Southern Europe, Northern Europe, Africa, or North America and also depending on whether it occurred in the Mediterranean, Northern, or Savannah ecosystems. To overcome the limitations associated with characterizing fire events solely based on burned area, an alternative approach involves integrating the concept of the return period (also referred to as recurrence or repeat interval), which accounts for the statistical rarity and temporal frequency of such events. This statistical measure, calculated using the Extreme Value Theory (EVT), corresponds to the average time between extreme events [1]. If a fire size of 5000 Ha corresponds to a return period of 50 years, then it can be described as a “50-year event” rather than a “5,000-Ha event.” Depending on the local distribution of fire sizes, the 5000-Ha fire would have a different return period if occurred in Southern Europe, Northern Europe, Africa, or North America, or in the Mediterranean, Nordic, or Savannah ecosystems. This approach better reflects the expected fire magnitude based on the local fire history [2].

✉ Frank A. Coutelieris
fcoutelieris@upatras.gr

Nikos Koutsias
nkoutsia@upatras.gr

¹ Department of Sustainable Agriculture, University of Patras,
2 Seferi Str, 30132 Agrinio, Greece

From a statistical perspective, Extreme Value Theory (EVT) represents the most rigorous and specialized methodology for characterizing extreme deviations from the central tendency of a probability distribution. Traditional statistical approaches are primarily concerned with the typical or average behavior of data series, offering limited insight into the tail behavior where rare events reside. In contrast, EVT is expressly developed to model the stochastic behavior of extreme observations, providing a mathematically robust framework for quantifying the likelihood and recurrence intervals of infrequent but potentially high-impact events [3]. Its applicability extends across diverse scientific and engineering disciplines, particularly in scenarios involving events that substantially exceed historical observations in magnitude or rarity [4]. Crucially, the accurate estimation of return periods for such extremes is of paramount importance [5], often superseding the relevance of their mere probability of occurrence, as it informs long-term risk assessments and strategic planning under uncertainty. Extreme Value Theory (EVT) admits both frequentist and Bayesian formulations, allowing for flexible inference under varying epistemological frameworks. In the context of univariate modeling of rare events—where the stochastic process is characterized by extreme deviations—three principal methodologies are prevalently employed: the GEV distribution for block maxima, the generalized Pareto distribution (GPD) for threshold exceedances, and the Poisson point process (PP) framework, which often incorporates GEV characteristics to model the occurrence and magnitude of extremes in continuous time or space [6]. Each of these approaches is grounded in asymptotic theory and is distinguished by its suitability for different sampling schemes and inferential objectives. The mathematical foundations, model assumptions, and conditions for applicability of these formulations will be rigorously examined in the forthcoming sections. Considerable efforts have been directed towards the application of extreme value analysis to wildfire events. Beverly and Martell [2] utilized EVT to characterize dry spell and fire size extremes in both the east and west divisions of the Boreal Shield ecozone in Ontario. They estimated the return time of a 30-day dry spell event and a 10,000 Ha fire, highlighting substantial differences between the compared event locations [2]. In the context of characterizing large fire behaviors, Jiang and Zhuang [7] employed the GEV distribution, the GP distribution, and the GEV distribution combined with a Poisson point process representation. They applied these methods to fit Canadian historical fire data spanning from 1959 to 2010. Their findings indicated notable disparities between anthropogenic and natural extreme fires in terms of extreme statistics. However, fire return levels aligned well with observations in relation to the magnitude and frequency of extreme events [7]. de Zea Bermudez et al. [8] modeled the spatial and temporal patterns of large fire events in Portugal from 1984 to 2004

using the POT approach. Their analysis revealed distinct regional variations in return periods [8]. Moreover, drawing from 64,474 fire records collected between 1991 and 2007 in Mississippi, Sun and Tolver [1] employed extreme value statistics, specifically the POT approach. Their study concluded that wildfires in Mississippi followed a GP distribution [1]. Lastly, Evin et al. [9] employed a Bayesian approach to EVT, utilizing it to evaluate fire policy in France. They incorporated the return period of extreme events to assess policy effectiveness in terms of spatial and temporal discretization [9]. Finally, the POT method, combined with classification techniques, has been applied to cluster time series data of daily area burned in 18 Portuguese districts from 1980 to 2010. Their results showed that the distributions of area burned are heavy-tailed with considerable density in the tail, which indicates a non-negligible probability of occurrence of days with very large area burned [10].

Given that wildfires with extensive burned areas are important events in the context of the planet's evolving climate and landscapes, their accurate assessment of rarity is of substantial scientific and societal importance. This information could play an important role in policy making, aiding in the prevention of such events and facilitating the implementation of necessary actions, including training fire crews, maintaining tactical aircraft and trucks, and educating support forces [11]. This study aims to apply and explore Extreme Value Theory approaches to wildland fire observations collected over a period of 25 years from 1983 to 2007 in Greece. The dataset includes 28,658 fire observations obtained from the Hellenic Forest Service and reports several parameters, including the x and y coordinates, the size of the burned area, and others. Our primary objective is to develop a methodological approach for quantifying extreme values and assessing the return period. To this end, the size of a fire is translated into a return period, providing people with a better understanding of the fire's magnitude and significance. We apply and evaluate the block maxima GEV approach and the POT approach, both frequentist and Bayesian.

2 Theoretical Background

2.1 Extreme Value Theory

Extreme Value Theory (EVT) constitutes a specialized domain within probability theory and statistical inference that systematically characterizes the stochastic behavior of observations located in the distributional tails. Rather than addressing the central tendency or variance of a dataset, EVT is explicitly formulated to model rare, high-impact events that manifest as extreme deviations from the expected value of an underlying distribution. The framework enables

rigorous quantification of the likelihood and severity of such outliers, which may represent maxima or minima within a finite sample. In operational terms, EVT facilitates the estimation of tail probabilities associated with observations that surpass all previously recorded values in a univariate time series [12]. More precisely, it allows for the derivation of the asymptotic distribution of exceedances or block maxima, thereby enabling probabilistic assessments of future extreme events in relation to historically observed data [13]. In general, we focus on the behavior of a random variable.

$$M_n = \max\{X_1, X_2, \dots, X_n\}, \quad (1)$$

where $X_1, X_2, \dots, X_n$ is an ordered sequence of values from an independent random variable X and n is the number of the different values, assumed to follow a distribution function.

Obviously, the above Eq. (1) can be easily used for the case of minimum rather than maximum. By denoting the distribution function of M_n values as F , it is rather straightforward that the probability P for M_n to be lower than a value z is given as

$$P(M_n < z) = [F(z)]^n \quad (2)$$

As far as F is generally an unknown function, the approximate families of P should be identified, based only on the extreme data, and by considering that $n \rightarrow \infty$. From a frequentist point of view, two widely used methods study this statistical behavior. The first method has been developed in the context of deriving singularities (maxima or minima) by applying the GEV distribution [14]. The second method identifies values that exceed or fall below a certain threshold, referred as POT method [15]. Both approaches are valid if and only if the random variable is independently and identically distributed [16]. The assumption of independence might be violated when extreme values are grouped in clusters, while the assumption of identically distributed observations can be relaxed when non-stationary processes such as trends render the assumption invalid [17]. Both methods are presented in detail in the next section.

An alternative to the frequentist approach is the Bayesian method, where prior distributions are chosen for the parameters of the GP distribution (GPD), and posterior distributions are calculated, providing a measure of expected uncertainty [18].

2.2 Extreme Value Theory Approaches

2.2.1 GEV Approach

The conventional EVT considers maxima (or minima) over specific blocks of time, known as block maxima/minima. By incorporating the External Types Theorem, also referred to as the Fisher-Tippett-Gnedenko Theorem [19], along

with the aforementioned concept for defining the distribution function, the maxima in Eq. (1) follow the generalized extreme value distribution, given as

$$G(z) = \exp\left\{-\left[1 + \xi\left(\frac{z - \mu}{\sigma}\right)\right]^{-\frac{1}{\xi}}\right\} \quad (3)$$

where μ is the location parameter, σ is the scale parameter, and ξ is the shape parameter. In accordance with the above-mentioned theorem, the value of ξ outlines the family of non-degenerate distributions that $G(z)$ belongs to. Precisely, if $\xi = 0$ then $G(z)$ follows the Gumbel distribution, if $\xi > 0$ then $G(z)$ follows the Fréchet distribution and, if $\xi < 0$ then $G(z)$ follows the Weibull distribution [13].

To apply the GEV distribution to a dataset, obtaining various conventional yet necessary statistical quantities is also required. Specifically, one must estimate the log-likelihood, along with the variance-covariance matrix, to calculate the standard errors for the parameters μ , σ , and ξ , as well as the 95% confidence intervals for them. Additionally, it is important to identify the profile log-likelihood for ξ and, finally, to approximate confidence intervals for pre-defined return levels [20]. The GEV distribution is typically fitted to block maxima (or minima), and it is worth noticing that the main drawback of the block maxima approach is that it does not utilize all the available information, especially about the upper tail of the distribution, since only each maximum value is able to be used. As a result, it inaccurately represents the needed statistical information [21].

2.2.2 POT Frequentist Approach

This approach focuses on the number of extreme events that have values higher than a certain threshold. More specifically, considering the random variable defined in Eq. (1) and assuming that M_i satisfies the GEV distribution (Eq. (3)), then the distribution function of the transformed variable $X - u$ for sufficiently large u must adhere to the constraint

$$P(X_i - u > y \mid X_i > u) \quad (4)$$

where y follows the GP distribution of the form:

$$H(y) = 1 - \left(1 + \frac{\xi y}{\bar{\sigma}}\right)^{-\frac{1}{\xi}} \quad (5)$$

where

$$\bar{\sigma} = \sigma + \xi(u - \mu) \quad (6)$$

Note that the parameters μ , σ , and ξ have been defined previously (refer to Eq. (3) and the related discussion there). It is interesting to provide a simple description of how ξ influences the distribution: if $\xi < 0$, then the distribution is bounded; if $\xi = 0$, the distribution is light-tailed

(exponential); and if $\xi > 0$, the distribution is heavy-tailed (Pareto).

To perform a POT estimation of the GEV distribution (Eq. (3)), it is essential to determine a reasonable threshold u [22]. To avoid empirical estimation based on the raw data and the physical problem under consideration, a maximum likelihood estimation should be conducted by fitting a GPD to the data series. This implies that, above a specific threshold at which the GPD assures a valid approximation, the mean residuals should be roughly linear with u (Eq. (4)). Consequently, a plot of mean residuals can be quite useful for defining the appropriate threshold. Following this, a 95% confidence interval must be recalculated based on the maximum likelihood data. Finally, the return levels can be derived using the same methodology as in the GEV approach [23].

The accuracy of the POT approach strongly depends on the threshold value, whose identification is usually under question due to the empiricism involved here [24]. To overcome this, a balance between bias and variance [25] should be assured, either by applying the mean residual life plot and identifying the point at which this plot becomes linear [26], or by comparing the fitting of the GP distribution across a range of different thresholds while adjusting the corresponding parameters [1].

2.2.3 POT Bayesian Approach

As far as frequentist statistical methods present some drawbacks for the treatment of extreme values, some Bayesian approaches are also available [9]. A typical application of Bayesian approach initialized by putting all the parameters affecting the problem to a vector θ . Then, each element of the vector is treated as random variable that is assumed to follow a specified probability density distribution (named “prior distribution”). This prior distribution is the distribution that the vector elements follow, before related to any data. Then, the likelihood of $f(\theta)$ is expressed as $f(x|\theta)$ and, by applying Bayes’ Theorem, is

$$f(\theta|x) = \frac{f(\theta)f(x|\theta)}{\sum_{\theta} f(\theta)f(x|\theta)} \quad (7)$$

to the extent that x is a discrete random variable. This $f(\theta|x)$ is the posterior distribution of θ .

Precisely, the direct computation of the denominator in Eq. (7) can be tricky. As a result, Markov-chain Monte Carlo (MCMC) techniques are frequently employed to investigate the quality of fitting that the simulated values present regarding the posterior distribution. A common practice to produce an evaluated fitting is the utilization of the Metropolis–Hastings algorithm within the MCMC procedure. This algorithm generates and accepts new samples for $f(\theta|x)$ in a

randomized manner [27]. Ultimately, the Bayesian approach also facilitates the estimation of return levels, based on the maximum likelihood analysis, as below discussed.

2.2.4 Return Levels

The return period represents the expected average interval between occurrences of extreme events. It is mathematically derived as the inverse of the exceedance probability, which is itself estimated within a defined confidence interval. In the framework of EVT, specific emphasis is placed on deriving return levels, which correspond to quantiles associated with a given return period [18]. For calculating return levels over a specific period of years, we employ the approach presented by Sun and Tolver [1]. When considering an event as extreme (regardless of the method used for characterization), let us denote the desired return period as N years and the corresponding return level as x^N . This implies that the probability of x^N being exceeded in any given year is $P = 1/N$.

The specific return level x^N is anticipated to be surpassed on average once every N years. In this context and for a given extreme values distribution, characterized by parameters σ and ξ , the N -year return level associated with m observations (x_m^N) can be expressed as:

$$x_m^N = u + \frac{\sigma}{\xi} \left[N n_y P(X_i > u)^{\xi} - 1 \right] \quad (8)$$

where n_y is the number of observations per year. Note here that the above theoretical approach stands for both GEV and POT frequentist methods, where μ , σ , and ξ are the parameters of either GEV or GP distributions. For the case of Bayesian approach, a prior distribution can be described by parameters μ , σ , and ξ , where Monte-Carlo algorithms usually are applied to obtain the corresponding posterior distribution, and therefore, the return level in analogy to Eq. 8 [28]. To summarize the return period concept, it serves as a metric for the extremeness of an event: the greater the return period, the rarer the event.

3 Materials and Methods

3.1 Study Area and Data

3.1.1 Study Area

Greece is situated at the end of the Balkan Peninsula and stretches between latitudes 34° and 43° north, and longitudes 19° and 28° east (Fig. 1). The country covers an approximate area of 132,000 km², with a total population of nearly 10.5 million people, based on the 2021 census data provided by the Hellenic Statistical Authority. The Greek climate is

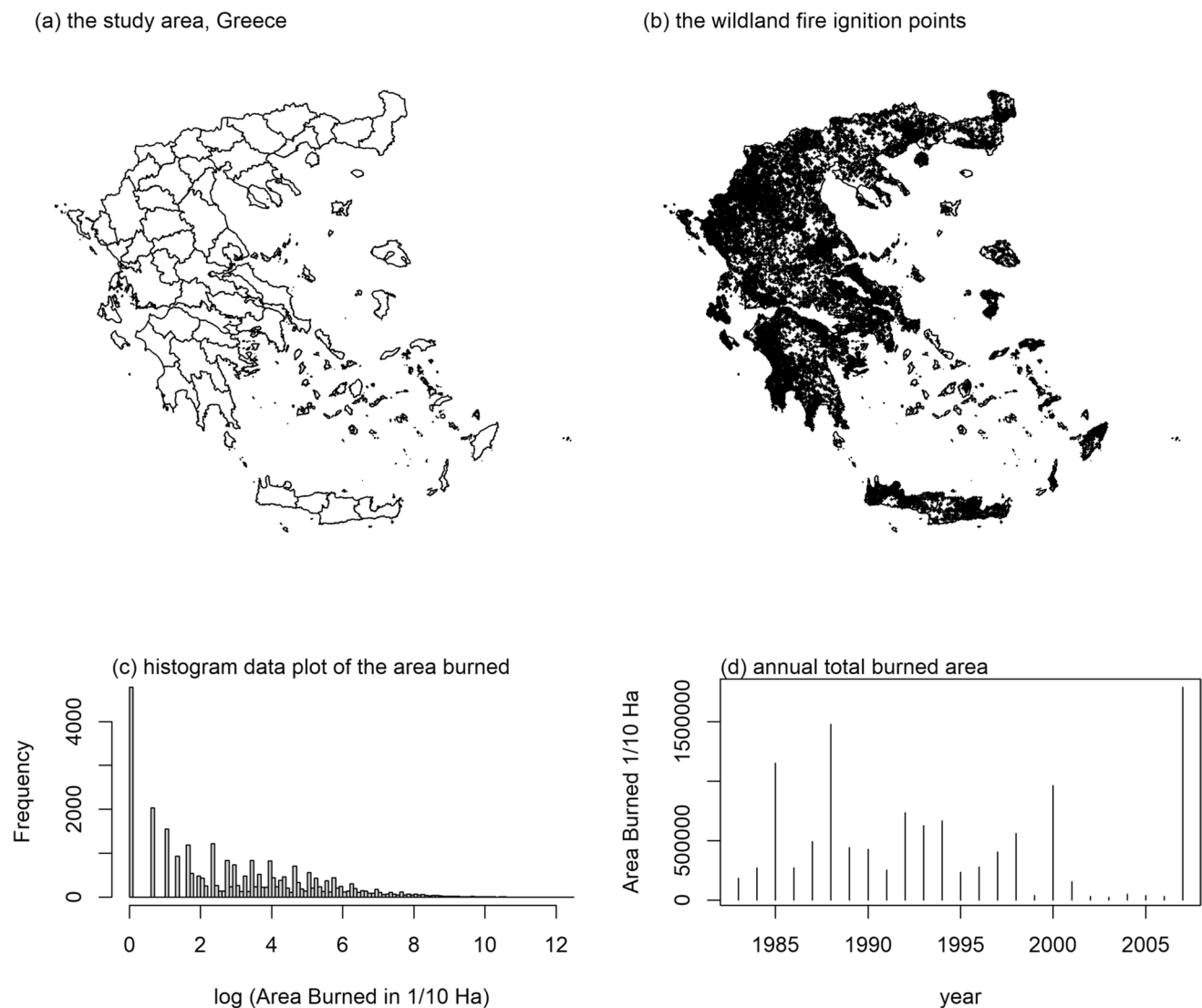


Fig. 1 **a** the study area, Greece, **b** the wildland fire ignition points, **c** histogram data plot of the area burned, and **d** annual total burned area

mainly Mediterranean, with a large range of variations due to complex terrain. The dominant vegetation types affected by fires, according to official fire records between 1985 and 2004 (Hellenic Forest Service), are phryganic ecosystems (32.34% of the total burned area), *Pinus halepensis* (15.43%) and *Pinus brutia* (6.17%) forests, *Quercus coccifera* shrublands (12.22%), and grasslands (7.36%). Additionally, 10.34% of the total burned area was recorded in forests under regeneration, where no information on the dominant affected vegetation types was provided [29].

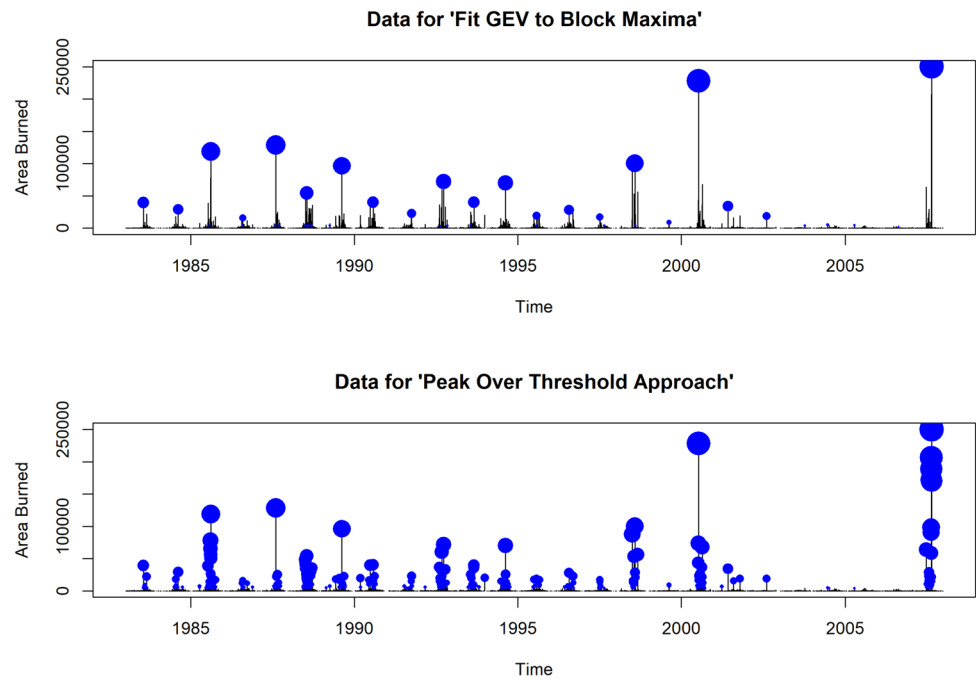
3.1.2 Wildland Fire Observations

The fire database used in this study consists of fire events occurred in Greece during the period 1983–2007. The 25-year database contains in total 28,658 fire events with

many parameters that have been recorded during each fire event and attached like for example the area burned, the date, and others. Additionally, the x and y coordinates recorded in latitude and longitude using degrees and first minutes have been registered and helped to position the fire events into a map (Fig. 1). This positional information helped us to register the fire events geographically. Histogram data plots and aggregated statistics to create the total annual burned area of the considered time period are shown in Fig. 1.

To apply the EVA methods, data pre-processing is necessary. Starting with the GEV approach, the block maxima has been identified and recorded (Fig. 2). The block maxima approach in Extreme Value Theory (EVT) consists of dividing the observation period into non-overlapping periods of equal size and restricts attention to the maximum observation in each period [30]. The block maxima in our study

Fig. 2 Data preparation to run the Extreme Value Theory approaches (i) “Fit GEV to Block Maxima” (upper figure) and (ii) “Peaks Over Threshold Approach” using the frequentist and the Bayesian approach (lower figure)



were considered on a yearly basis due to the seasonality of the fire events within the year. There are several months during one single year when no incidents occur, as for example mainly during the winter. On the other hand, block sizes higher than 1 year correspond to available data less than 25, which is a very limited amount to obtain accurate statistical estimations. As mentioned before, the POT method is based on the use of a threshold to isolate values considered extreme from the rest of the data and create a model for the extreme values by considering the tail of all the values that exceed this threshold. In general, data exceeding the threshold approximate to a general Pareto distribution [31]. The thresholded data used in this study are presented in Fig. 2. Regarding the Bayesian approach, it involves the derivation of posterior distributions. This requires defining prior distributions and utilizing Markov Chain Monte Carlo (MCMC) methods for sampling. Finally, the comparison of the prior with the posterior distribution is achieved empirically through the relative Q-Q plots.

It is also worth noticing that the results should be checked against the geographical (location) dependence, as far as the non-stationarity assumption is crucial for both GEV and GP distributions that are employed in the analysis presented here. More precisely, the block maxima method on the collated fire size data will choose the yearly maximum from the location with the largest fire size only, so the resulting GEV distribution is only appropriate for a single location where the maximum-maximums occur. Similarly, for the GPD, a high threshold choice will ignore locations with lower fire size, even if the lower fire size is extreme for those areas. Figure 3 shows the spatial distribution of the size of the

burned area for each fire ignition point and the longitudinal and latitudinal trends of the maximum values of the size of the fires. It seems that there is not any significant trend, therefore, the stationarity in our data seems that it is not a real problem. The above assures that the resulting distributions estimate return periods and return levels that are not locationally dependent.

Finally, for the POT approach, the assumption of absolutely no clustering of the above-threshold data values has been also checked through an estimated Extremal Index as Sun and Tolver (2012). Precisely, we estimated the External Index both annually and for the whole data set, finding values high enough to ensure independence. These values vary between 0.46 the lowest and 1.0 the highest, while the overall average value for the whole data in our work set is around 0.39.

3.2 Methodology

3.2.1 Applying the Extreme Value Theory

From a technical standpoint, the statistical analyses and graphical representations were conducted using the R programming language [32], which is universally acknowledged as one of the optimal environments for statistical computation and visualization due to its vast repository of well-established and rigorously assessed subroutines that are freely distributed. The preparation of raw data for extreme value analysis in R is a nuanced process that demands careful consideration of data quality, distributional assumptions, and the specific requirements of the chosen statistical methods.

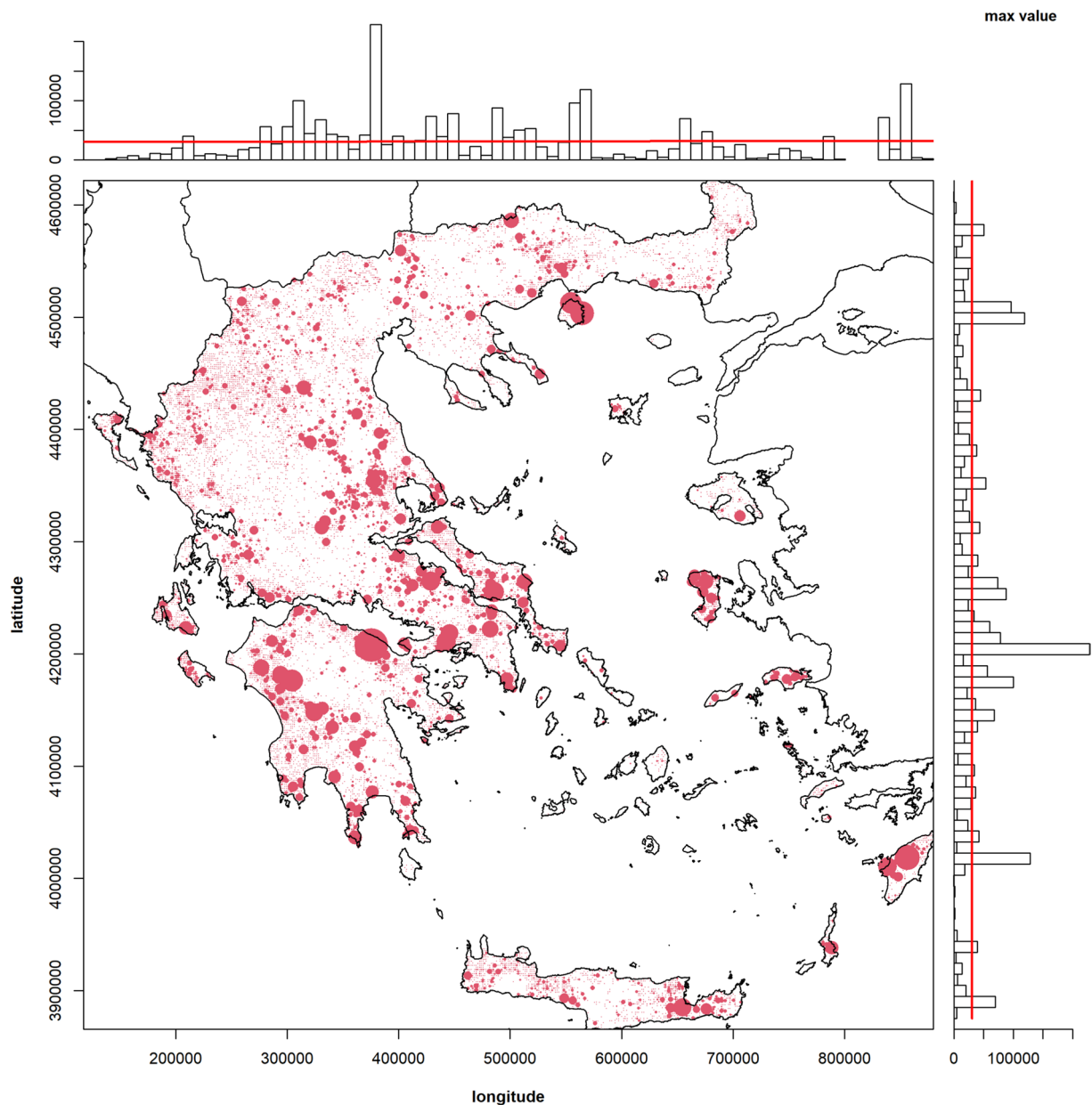


Fig. 3 The spatial distribution of the size of the burned area for each fire ignition point and the longitudinal and latitudinal trends of the maximum values of the size of the fires

Therefore, it involves a systematic and meticulous process to ensure the reliability and accuracy of the subsequent statistical models, particularly those related to GEV, POT, and Bayesian approaches [33]. The initial step includes data cleaning, addressing any missing or erroneous values, and ensuring a consistent time series or distribution for analysis. Once cleaned, data is often transformed (i.e., by log function), if needed, to satisfy distributional assumptions or stabilize variance.

Conceptually, our emphasis lies in interpreting the probability to observe specific magnitudes of burned areas into

a quantifiable measure of the rarity of such occurrences. We achieve this by utilizing the concept of the return period, derived from extreme value analysis, as above discussed. Essentially, the return period estimates the time expected for the re-appearance of a specific event, recorded through the value of the burned area.

To validate the accuracy of our model against real data, we initiate the process with a comprehensive four-plot visualization. The initial component involves the presentation of two quantile–quantile (Q–Q) plots, which serve as effective tools for assessing the congruence between two distinct data

sets. In essence, a quantile represents the proportion of data points within a dataset that are below a specified threshold. When comparing two sets of data originating from the same distribution, the points on the Q-Q plot should closely follow a diagonal line at a 45° angle. Deviations from this reference line indicate potential disparities between the distributions of the two datasets. Notably, substantial deviations from this line of equality may suggest that the two datasets have been generated from dissimilar distributions. The third plot visually illustrates the level of alignment between the data model and the empirical distribution function (i.e., how closely the data model or satisfies the empirical distribution function). From a qualitative point of view, a higher degree of proximity between the two depicted curves signifies enhanced accuracy of the model. It is important to acknowledge that the bandwidth showcased on the same plot serves as an indicator of the desired level of conformity between the density and the distribution. Finally, the fourth plot depicts the return level plot, a graphical representation of the level that is expected to be exceeded by the process on average once in T -years against (the logarithm of) return period T . Maximum likelihood estimation is shown by the solid black line on the plots while the gray dashed lines are approximate pointwise 95% confidence intervals. If there are n points in each data set, the largest point will correspond to the empirical n -year quantile, the second largest to the empirical $(n - 1)$ -year empirical quantile, and so forth [34].

3.2.2 Applying the Extreme Value Theory to Real-World Cases

In this study, we apply the Extreme Value Theory (EVT) as a method to analyze and quantify the magnitude of unprecedented wildfire events occurred in Greece, one in Euboea (2021) and one in Dadia (2023). As already underlined, the study and especially the prediction of when a catastrophic fire incident is expected, is of great scientific and societal importance, especially regarding the extreme values they might present.

4 Results

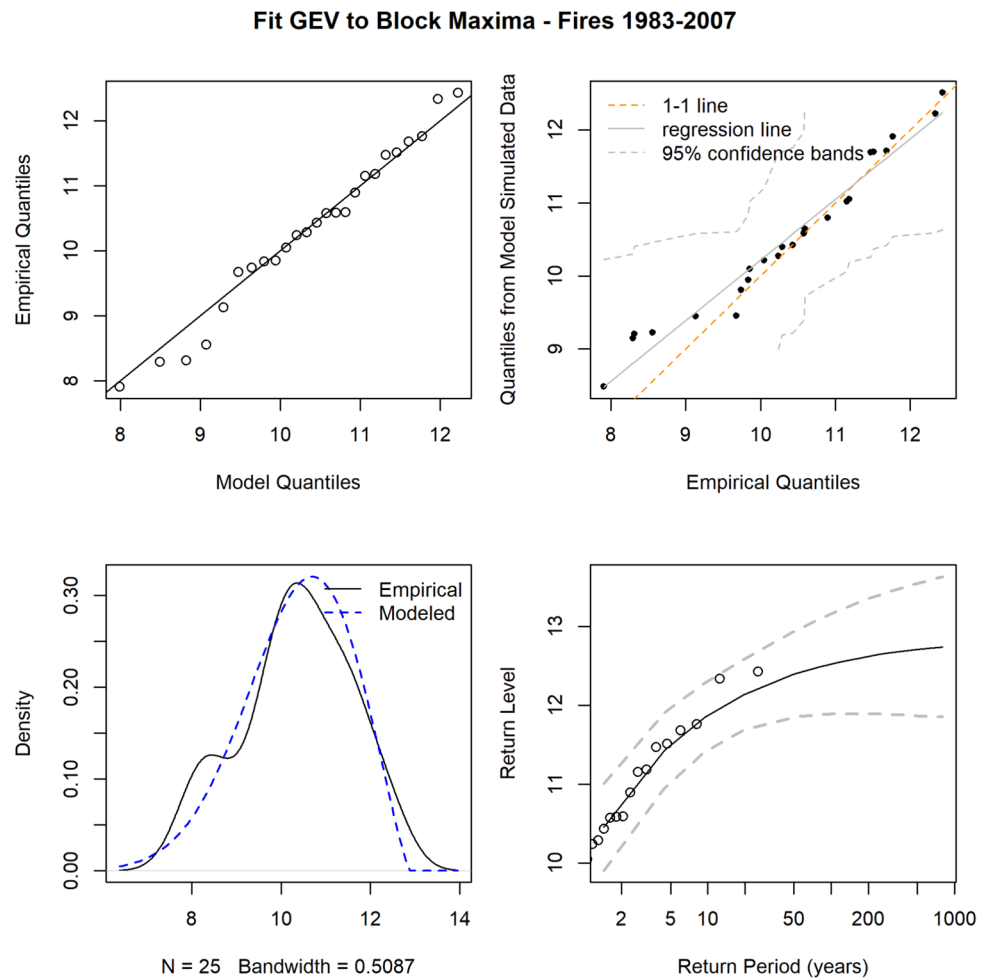
4.1 The Extreme Value Theory Models

The results of the GEV approach that has been applied using the function `fevd()` from the `extRemes` package [35] are graphically presented in Fig. 4. The q-q plots indicate that data following GEV distribution is in an acceptable agreement with deviations observed at low and high values of burned area size along the regression line. Furthermore, the comparison between modeled data and empirical distribution shows specific insufficiencies, especially for not being

able to capture the second peak. Finally, the estimations of return period presents similar divergences for both low and high return period values. This is expected due to the limited amount of the length of the available time series data; in the GEV, each time interval within the time series is represented by one single value (the maximum).

The results for the POT are graphically presented in Fig. 5. In this case, the q-q plots indicate that data follow the model distribution in a very good agreement except for very high sizes of burned area where certain deviations are observed. Furthermore, the comparison between modeled data and empirical distribution seems to be good for values higher than the threshold. Finally, the accuracy of the return period is good, even for high return periods (> 50 years) since all the values, higher than the threshold, are involved in the calculations. The above results suffer from the empiricism that is used to select the appropriate threshold, a process that introduces a level of arbitrariness in the method itself. To overcome this weakness and to obtain an appropriate threshold value, based on a robust scientific approach rather than empirical estimations, some work is necessary, as described below [36]. Figure 6 shows the excess of the dataset's mean value over the selected threshold in terms of likelihood confidence interval. In our case, it is depicted that values of the threshold higher than 8 indicate a wider interval of confidence, thus producing less reliable results. The latter becomes clear through the representation of the effect of the threshold's choice over the parameters of distribution, i.e., over both the shape and obviously the modified shape, a parameter introduced to describe the GP distribution that the excess should follow [24]. Again, the variation of parameters, corresponding to the uncertainty of the distribution and therefore the difficulty in matching the available data, becomes significant for threshold values > 8 .

The results of the POT approach when the Bayesian method is used rather than the frequentist are presented in Fig. 7. The functions and libraries applied in R are the same as above, while the selected threshold is also the same. Again, the q-q plots indicate that data follow the model distribution in a good agreement, except for very high sizes of burned area where certain deviations are observed, similarly to POT frequentist. Furthermore, the comparison between modeled data and empirical distribution seems to be sufficient for the values higher than the threshold. Finally, the estimation of return period is good for low return periods, while the predictions become uncertain for higher periods. The 95% confidence bands, as shown in Fig. 7, increase with the size of the burned area, becoming very large for high values, indicating a high uncertainty in the estimations. Therefore, the intrinsic drawback of the method is also presented here, in accordance with what has been observed for the frequentist POT approach, although at much lower levels. Finally, the weakness of the empiricism that was previously

Fig. 4 Output plots of the Fit GEV to block maxima

underlined is also observed here, as far as the selection of the appropriate threshold is again arbitrary. Compared with the frequentist POT approach, the Bayesian one shows matching between empirical and modeled data of the same accuracy as for the frequentist POT approach, while the predictions for the return period become of weak confidence for long time periods comparing with the size of the available data [37].

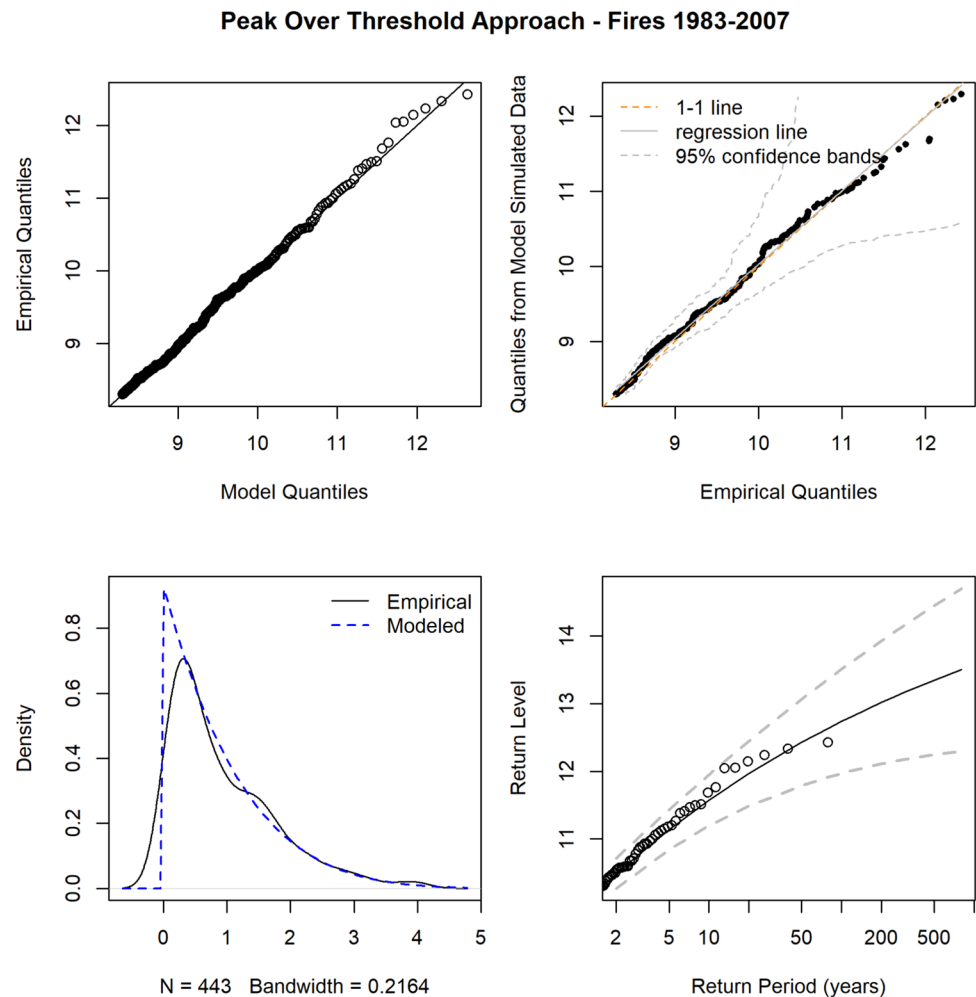
The estimated return periods, which predict the expected time until fire events of specific sizes re-occur, are detailed for the three methods in Table 1 and visually represented in Fig. 8. The analysis includes ten return periods (2, 5, 10, 25, 50, 100, 200, 500, 1000, and 2000 years), with the corresponding burned areas for each method reported in Table 1. Up to the 25-year return period, the GEV is less accurate compared to both the frequentist and Bayesian POT methods, as it associates larger burned areas with each return period, while both POT methods are of equivalent efficiency. However, beyond the 50-year return period, this trend reverses, with the GEV method underestimating the burned area following an asymptotic-type behavior, while both frequentist and Bayesian POT approaches present predictions increasing with the return period. When comparing

the two POT approaches, the Bayesian method diverges with a rate much higher than the frequentist one (as seen in Table 1 and Fig. 8).

4.2 Specific Large Fires in Greece

Return periods have been estimated using the three methods for few real-world fire events that have occurred in Greece in recent years namely in Euboea (2021) and Dadia (2023) (see Fig. 9).

The North Euboea fire in 2021 marks a significant shift in defining the scale of exceptionally large fires in Greece, accounting for 45,000 ha of burned area. As stated by Gemitzis and Koutsias [38], the fire in Euboea in the summer of 2021 destroyed more than 400 km² of forest in the northern part of the island. The areas affected by the fire were economically active, serving as sites for honey and resin production, as well as hosting cattle breeding and mild tourist activities. Consequently, the fire had profoundly negative effects on the local economy, in addition to the destruction of an environment of exceptional natural beauty. Regarding the reasons behind it, many sources indicate a combination of human negligence, fuel load, and persisting favorable

Fig. 5 Output plots of the POT approach (frequentist)

meteorological conditions possibly related to climate change as the main driving factors [38]. The size of about 45,000 ha corresponds to a return period of 200 years based on the POT frequentist, between 100 and 200 years based on the POT Bayesian, and much more than 2000 years based on the GEV. In this case, the GEV approach overestimates significantly and probably enormously wrongly the return period for this size of fire.

In the summer of 2023, we experienced one more significant event concerning the fire activity in Greece, in Dadia, which unfortunately burned more than 95,000 ha—an unreal number according to the fire history of the last century in Greece. The Dadia case is considered an exceptional fire that burned almost consecutively for 16 days. Although the number of 95,000 ha has not been recorded before in the recent history of the country, its real magnitude remains uncertain. The second largest fire in terms of size is that of Euboea in 2021, which is almost half the size of Dadia's fire. Therefore, translating the size of the fire into a return interval helps to define its real magnitude, as the fire in this case is reported as an event expected once within a very large

period of time. The size of about 95,000 ha corresponds to a return period of 2000 years based on the POT frequentist, between 500 and 1000 years based on the POT Bayesian and enormously wrongly higher than 2000 based on the GEV. In this case, the GEV approach enormously overestimates the return period for this size of fire while POT Bayesian underestimates it.

5 Discussion

5.1 Limitations Relating to Extreme Value Theory Implementation

Starting with the GEV method, the results seem to be not accurate enough, particularly for long return periods (> 100 years). This concern can be attributed to the limited size of the available data (only 25 years) as well as to the block maxima approach itself, because it introduces a degree of arbitrariness in the calculations, as far as the number of incidents that build a block is one of the parameters of the

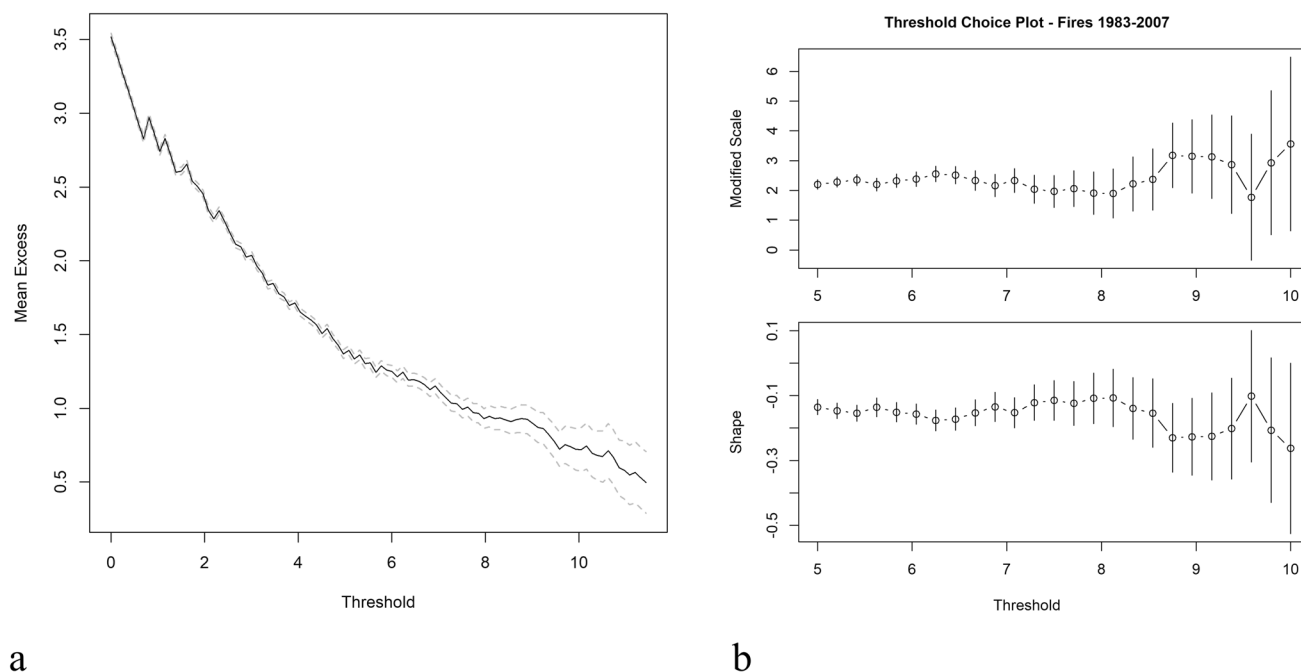


Fig. 6 **a** Mean residual life plot for POT frequentist approach, and **b** effect of threshold on GP distribution parameters

Fig. 7 Output plots of the POT approach (Bayesian)

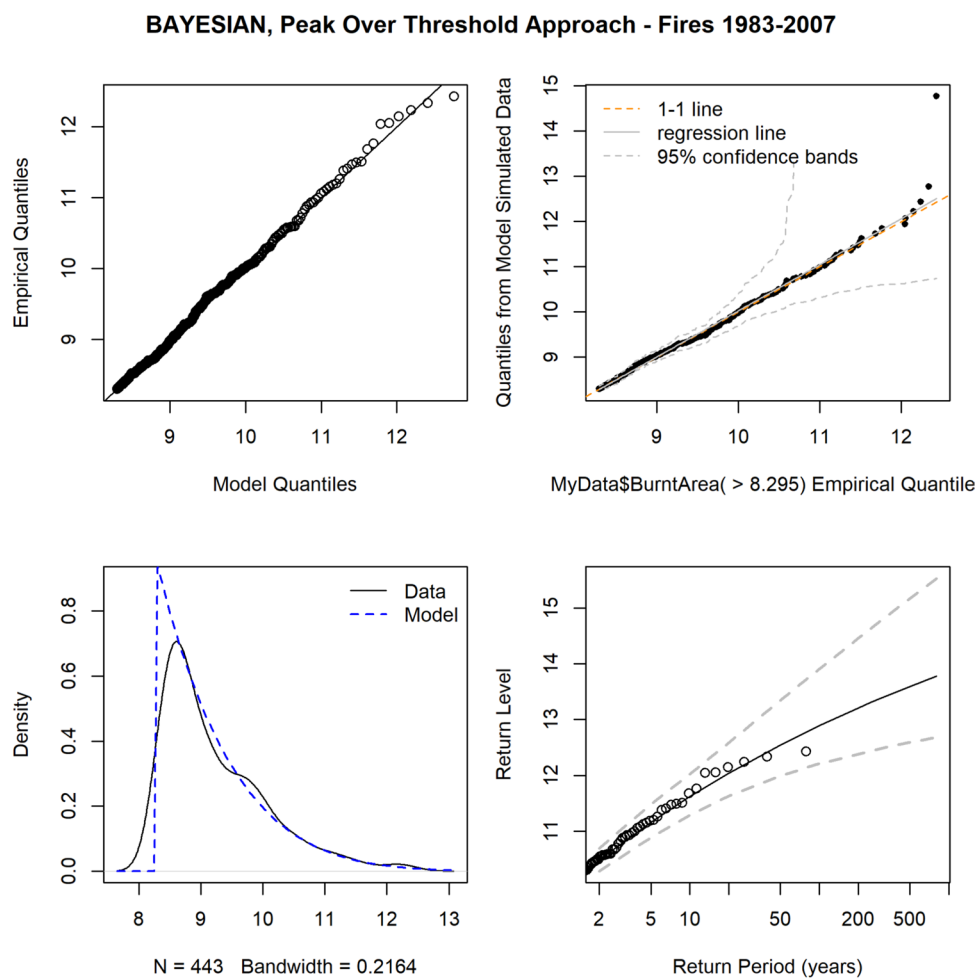


Table 1 Return periods and the corresponding area burned for the three EVT approaches

Return period	Burned area (Ha)		
	GEV	POT frequentist	POT Bayesian
2	3462	3597	3642
5	9173	6871	7191
10	13,942	10,614	11,485
25	20,003	17,698	20,196
50	24,035	24,957	29,826
100	27,445	34,046	42,797
200	30,223	45,074	59,824
500	33,020	62,685	89,832
1000	34,585	78,244	119,197
2000	35,775	95,599	155,153

model [39]. In our case, the definition of block size to 1 year is the only reasonable option, and it could strongly affect the accuracy of the procedure, because the total length of the data series (25 maxima used) is only an order of magnitude larger than the length of the blocks. Indeed, the size of available data has already been recognized as crucial for the effectiveness and accuracy of the GEV approach, where it is found that the return period calculated by GEV is successful only when it is of comparable magnitude with the size of the available data (see, for instance, [36]). The core of this weakness is identified in the fact that the parameters of the distribution are estimated through mathematical limits, which hold only when the amount of data tends to be infinite. However, in practice, the quantity of data that is available is finite, and this is a serious reason for biased results. Furthermore, the selected method to obtain explicit expressions for the GEV parameter estimates is important. Here, we made use of the maximum likelihood method, which is a widely acceptable and preferred estimation method, since it is quite general and more flexible than others. However, in ordinary extreme value analyses like the ones we are cited in this work, the flexibility provided by the maximum likelihood method is not necessary because we have a small sample size and, at the same time, no too heavy-tailed distributions [40].

To continue with the POT approach, it is well known that it is not as sensitive to the magnitude of the available data as GEV, but, unfortunately, it suffers from the arbitrariness that the selection of the threshold introduces to the procedure. This weakness has been pointed out several times in the past [26, 41]. As mentioned before, the selection of the threshold is an empirical practice that affects the results for high-level return periods. In particular, the determination of the threshold involves the mean excess function as also defined by Kotz and Nadarajah [12] where graphical determination of the function's linearity is used. With an appropriately

selected threshold for a given dataset, POT seems to be a reliable alternative for the return period estimations, even for very rare extreme events [22, 37]. In this context, the method can be proposed as an appropriate option, especially when the available dataset is of limited size once specific assumptions are valid, such as those referring to stationarity, independence, and clustering. Analogous conclusions have been encountered several times in the relative literature when both methods were applied to hydrological [37], oceanic [22], and temperature [3] data. Finally, the use of the log function on the dataset supports a more accurate selection of the threshold, as the narrow range of the available values allows for less sensitive choices of threshold value [42]. It is also worth noticing that the raw data sampling rate is another parameter affecting the return period values predicted by the POT approach. This is more significant when data is recorded on a regular basis, a case where the period should be defined and considered for the mathematical estimations. In our case, the incidents are recorded whenever they happen; thus, the sampling rate is not constant. Specific issues can also arise due to multi-sampling, since more than one incident can occur during the same date. Although the return period is always estimated in years, the sampling rate time unit should be considered through the calculations.

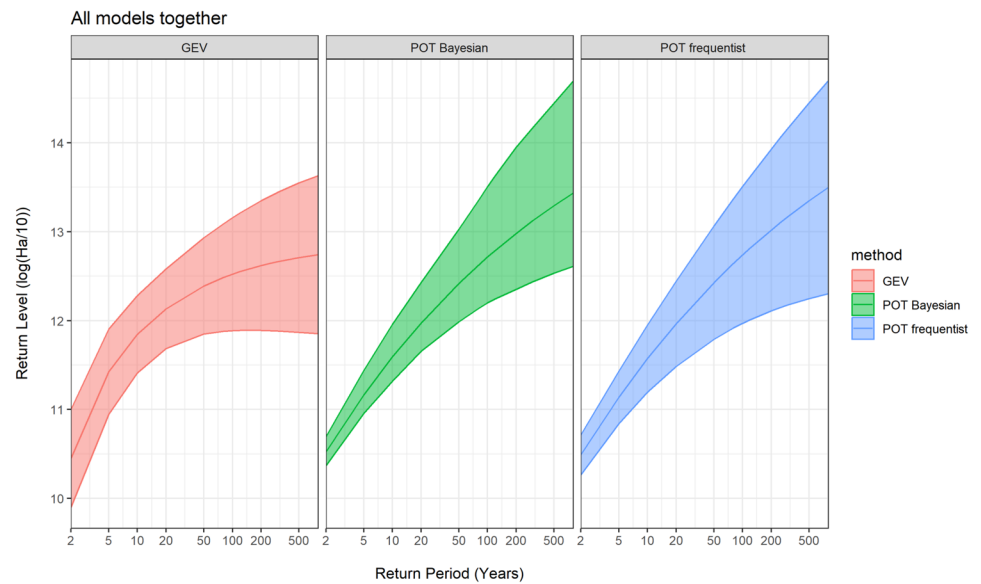
Theoretically speaking, the Bayesian approach is expected to perform particularly well when working with short data sets. As the available thresholded data are quite limited here, the information incorporated as a prior distribution is more than valuable. On the other hand, an analysis using an uninformative prior closely mirrors results obtained via profile likelihood methods. However, this outcome contrasts sharply with analyses based on maximum likelihood and asymptotic standard errors, due to the pronounced skewness in profile likelihood surfaces for return levels, particularly when very few years of data are available. The key point is that Bayesian inference, when using flat priors, essentially recovers a likelihood-based analysis. In contrast, the empirical prior distribution we adopted is genuinely informative, leading to significantly different inferences.

To conclude with, the frequentist POT approach shows a better performance as compared to GEV and POT-Bayesian methods. This can be observed through the range of the confidence intervals, where the wider the intervals, the higher the uncertainties of the estimations. There are several works in literature with similar findings, using various datasets and conditions [43–45], among others, where this higher efficiency is attributed to lower uncertainties at the distribution tails [46].

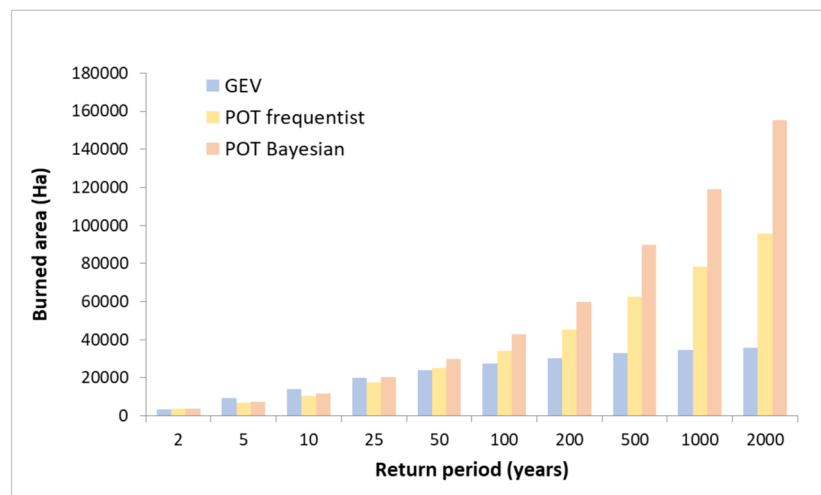
5.2 Limitations Using Fire Magnitudes as Extremal Predictors

As mentioned in the introduction, when a wildfire occurs and the event is communicated—especially to the general

Fig. 8 **a** Comparison of the three methods in terms of return-period prediction, and **b** graph of the return periods and the corresponding area burned for the three EVT approaches



a



b

public—the size of the fire becomes one of the most important characteristics to report. This statistic provides people with an understanding of the fire's size and thus people get an idea about the magnitude and significance of the fire. However, this assumes that people have an idea of what a particular fire means in terms of size and how it compares to their empirical knowledge derived from the fire history of the area. In essence, the interpretation of fire size varies depending on the local distribution of fire sizes to which the area is exposed. Therefore, a fire of a specific size can have a completely different significance depending on the area in which it occurs and the type of ecosystem affected. To address the limitations of relying solely on fire

size to characterize a wildfire event, an alternative approach involves incorporating the concept of return period (also known as recurrence interval or repeat interval). This statistical measure, calculated by the Extreme Value Theory (EVT), represents the average time between extreme events [1]. It offers a more context-sensitive reflection of fire magnitude, grounded in the local fire history [2].

When modeling large-scale phenomena—especially across extensive geographic areas such as at the global level—using fire size as the dependent variable can introduce bias into the analysis. This is because fires of the same size may have vastly different meaning depending on their location and ecological context. Relying solely on

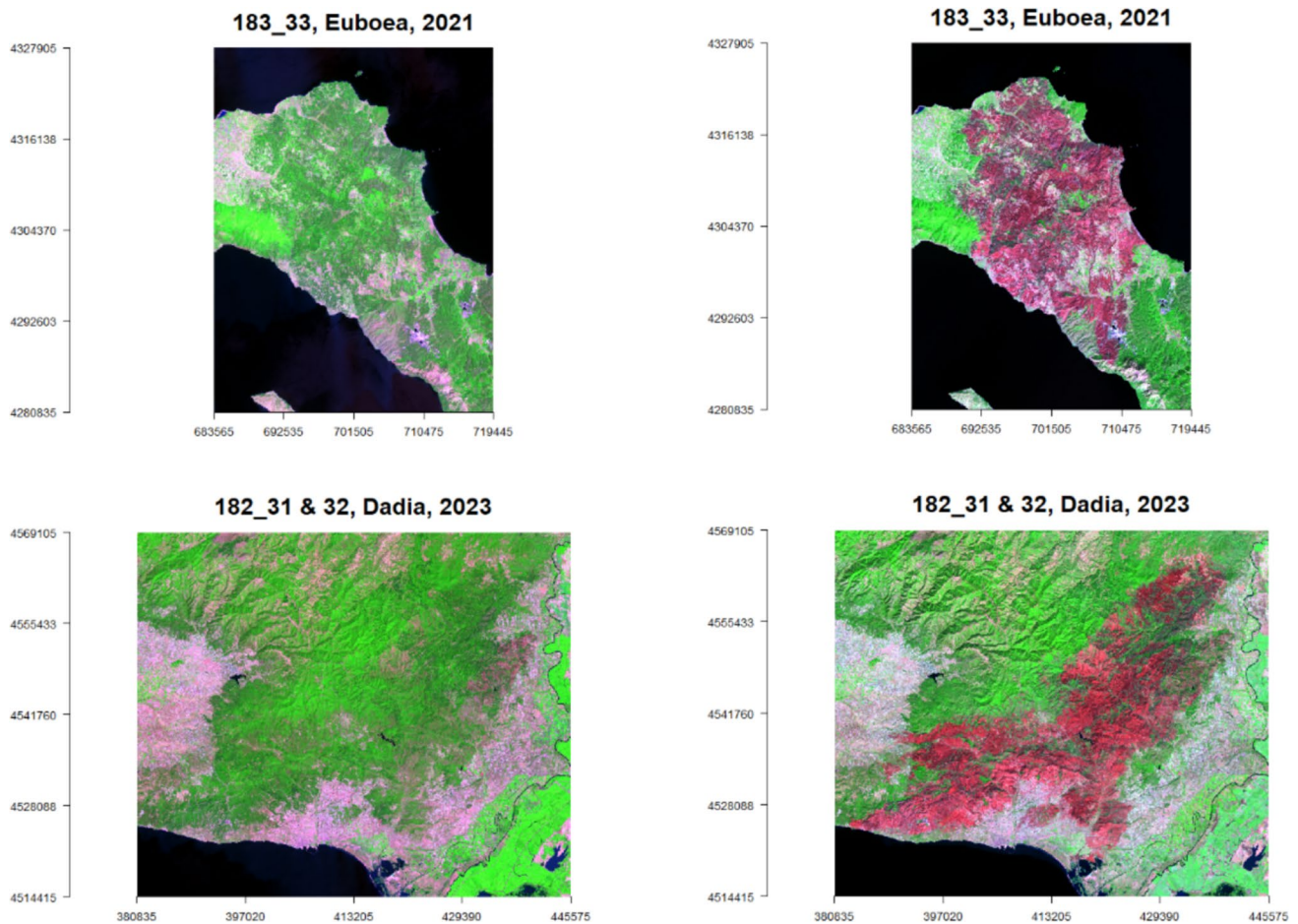


Fig. 9 Pre-fire Landsat images (left two images) and post-fire Landsat images (right two images) of the two exceptional large fires occurred in Euboea (2021) and Dadia (2023), Greece

fire size may lead to inconsistent interpretations and potentially misleading conclusions. An alternative approach involves incorporating the return period as an additional or even primary response variable. As previously discussed, two fires of equal size can correspond to very different return periods, indicating differing levels of rarity and recurrence. Using a return period provides a more context-sensitive measure that reflects local fire history and variability. This shift in focus—from fire size alone to return period—not only refines the analytical approach but also reframes the research question. Instead of identifying global factors that influence fire size, the analysis shifts towards identifying local drivers of fire magnitude. This perspective offers a more nuanced understanding on fire behavior and risk at regional scales.

To illustrate this concept more concretely, consider a hypothetical research question aimed at explaining burned areas, with fire size as the dependent variable. In such a case, two fires of the same size—one in a fire-prone region where large fires are common, and another

in a non-fire-prone region where large fires are rare—would be treated equivalently based solely on size. This approach, however, overlooks important contextual differences. If we instead use the return period as the dependent variable, these fires would be differentiated. The fire in the fire-prone area would have a shorter return period, indicating frequent occurrence, while the fire in the non-fire-prone area would have a much longer return period, reflecting its rarity. This distinction offers a more meaningful interpretation of fire dynamics and help assess fire risk more accurately across different regions.

Finally, for local authorities, understanding the expected frequency of events of a particular magnitude is essential for effective planning and resource allocation—including infrastructure, personnel, and emergency services. For example, it is critical for communities to assess the likelihood of experiencing a fire of a specific size. Knowing whether a 5000- or 10,000-ha fire is expected once every 10 years versus once every 50 years can significantly influence their preparedness efforts and plans. This information

shapes fire prevention strategies, emergency response capacity, and long-term planning. The return period is also highly relevant for the insurance sector. Insurance companies use it to assess risk and determine premiums more accurately. By understanding how frequently large fires are likely to occur, both public and private sectors can better manage and mitigate fire-related risks.

6 Conclusions

In this study, EVT has been applied to wildland fire data recorded for 25 years from 1983 to 2007 in Greece. Data include details about the size of the burned area as well as other parameters related to fires, such as the x and y coordinates, dates, etc. The main objective of this study is to explore the performance of each method regarding the prediction of the return period for extremely rare events. For the sake of comparison and evaluation, three methods (GEV, frequentist POT and Bayesian POT) have been applied to the same raw data after the appropriate elaboration. As far as the return period is introduced here as the most appropriate indicator for the approximation of extremeness, all the methods were compared in terms of accuracy on the estimation of the return period for events of specific magnitude.

The results from the GEV approach tend to level off asymptotically at a constant value, suggesting that the size of the burned area remains approximately constant after a certain initial time, regardless of the period considered for the return level. This behavior is clearly unrealistic, highlighting that GEV provides reliable forecasts only for time periods that are of comparable size with the available dataset. Although conceptually different, the Bayesian POT approach also exhibits unrealistic behavior over extended time periods, as the size of the burned area constantly increases with time, following a power-law curve. Lastly, the frequentist POT model appears to perform well with the available data, predicting a more realistic and accurate increase in burned area size over time, as has also been evaluated here.

Acknowledgements We would like to thank the two anonymous reviewers and the editor for their comments that helped to improve the original manuscript. The publication fees of this manuscript have been financed by the Research Council of the University of Patras.

Author Contributions N.K. and FC wrote the main manuscript text. All authors reviewed the manuscript.

Funding Open access funding provided by HEAL-Link Greece.

Data Availability Data sets generated during the current study are available from the corresponding author on reasonable request.

Declarations

Competing interests The authors declare no competing interests.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

1. Sun, C., & Tolver, B. (2012). Assessing the distribution patterns of wildfire sizes in Mississippi, USA. *International Journal of Wildland Fire*, 21, 510–520.
2. Beverly, J. L., & Martell, D. L. (2005). Characterizing extreme fire and weather events in the Boreal Shield ecozone of Ontario. *Agricultural and Forest Meteorology*, 133, 5–16.
3. Chavez-Demoulin, V., & Davison, A. C. (2005). Generalized additive modelling of sample extremes. *Journal of the Royal Statistical Society Series C, Applied Statistics*, 54, 207–222.
4. Santhanam, M.S. (2013). Extreme event recurrence time distributions and long memory, in A.S. Sharma, A. Bunde, V.P. Dimri & D.N. Baker (eds.), *Extreme events and natural hazards: The complexity perspective*, Wiley Blackwell, pp. 335–344.
5. Read, L. K., & Vogel, R. M. (2015). Reliability, return periods, and risk under nonstationarity. *Water Resources Research*, 51, 6381–6398.
6. Hernandez, C., Keribin, C., Drobinski, P., & Turquety, S. (2015). Statistical modelling of wildfire size and intensity: A step toward meteorological forecasting of summer extreme fire risk. *Annales Geophysicae*, 33, 1495–1506.
7. Jiang, Y., & Zhuang, Q. (2011). Extreme value analysis of wildfires in Canadian boreal forest ecosystems. *Canadian Journal of Forest Research*, 41, 1836–1851.
8. de Zea Bermudez, P., Mendes, J., Pereira, J. M. C., Turkman, K. F., & Vasconcelos, M. J. P. (2009). Spatial and temporal extremes of wildfire sizes in Portugal (1984–2004). *International Journal of Wildland Fire*, 18, 983–991.
9. Evain, G., Curt, T., & Eckert, N. (2018). Has fire policy decreased the return period of the largest wildfire events in France? A bayesian assessment based on extreme value theory. *Natural Hazards and Earth System Sciences*, 18, 2641–2651.
10. Scotto, M. G., Gouveia, S., Carvalho, A., Monteiro, A., Martins, V., Flannigan, M. D., San-Miguel-Ayanz, J., Miranda, A. I., & Borrego, C. (2014). Area burned in Portugal over recent decades: An extreme value analysis. *International Journal of Wildland Fire*, 23, 812–824.
11. Malamud, B. D., & Turcotte, D. L. (1999). Self-organized criticality applied to natural hazards. *Natural Hazards*, 20, 93–116.
12. Kotz, S., & Nadarajah, S. (2000). *Extreme value distributions: Theory and applications*. Imperial College Press.
13. Garrido, M., & Lezaud, P. (2013). Extreme value analysis: An introduction. *Journal de la Societe Française de Statistique*, 154, 66–97.
14. Makkonen, L. (2008). Problems in the extreme value analysis. *Structural Safety*, 30, 405–419.

15. Castillo, E., & Hadi, A. S. (1997). Fitting the generalized pareto distribution to data. *Journal of the American Statistical Association*, 92, 1609–1620.
16. Katz, R. W., Parlange, M. B., & Naveau, P. (2002). Statistics of extremes in hydrology. *Advances in Water Resources*, 25, 1287–1304.
17. Schoenberg, F. P., Peng, R., & Woods, J. (2003). On the distribution of wildfire sizes. *Environmetrics*, 14, 583–592.
18. Coles, S. G., & Powell, E. A. (1996). Bayesian methods in extreme value modelling: A review and new developments. *International Statistical Review/Revue Internationale De Statistique*, 64, 119–136.
19. Embrechts, P., Klüppelberg, C., & Mikosch, T. (1997). *Modelling extremal events: For insurance and finance, stochastic modelling and applied probability*. Springer-Verlag.
20. Smith, R. L., & Weissman, I. (1985). Maximum likelihood estimation of the lower tail of a probability distribution. *Journal of the Royal Statistical Society. Series B, Statistical Methodology*, 47, 285–298.
21. Gründemann, G. J., Zorzetto, E., Beck, H. E., Schleiss, M., van de Giesen, N., Marani, M., & van der Ent, R. J. (2023). Extreme precipitation return levels for multiple durations on a global scale. *Journal of Hydrology*, 621, Article 129558.
22. Bakali, H., Serhir, N., & Aouiche, I. (2021). Comparison between maximum annual and peak over threshold methods for the determination of extreme waves in Moroccan Atlantic Coast. *IOP Conference Series: Earth and Environmental Science*, 849, Article 012002.
23. Önöz, B., & Bayazit, M. (2001). Effect of the occurrence process of the peaks over threshold on the flood estimates. *Journal of Hydrology*, 244, 86–96.
24. Scarrott, C., & MacDonald, A. (2012). A review of extreme value threshold estimation and uncertainty quantification. *Revstat Statistical Journal*, 10, 33–60.
25. Beirlant, J., Goegebeur, Y., Segers, J., & Teugels, J. (2004). *Statistics of extremes: Theory and applications*. Wiley.
26. Davison, A. C., & Smith, R. L. (1990). Models for exceedances over high thresholds. *Journal of the Royal Statistical Society. Series B (Methodological)*, 52, 393–442.
27. Mendes, J. M., de Zea Bermudez, P. C., Pereira, J., Turkman, K. F., & Vasconcelos, M. J. P. (2010). Spatial extremes of wildfire sizes: Bayesian hierarchical models for extremes. *Environmental and Ecological Statistics*, 17, 1–28.
28. Coles, S. G., & Tawn, J. A. (1996). A bayesian analysis of extreme rainfall data. *Applied Statistics*, 45, 463–478.
29. Koutsias, N., Xanthopoulos, G., Founda, D., Xystrakis, F., Nioti, F., Pleniou, M., Mallinis, G., & Arianoutsou, M. (2013). On the relationships between forest fires and weather conditions in Greece from long-term national observations (1894–2010). *International Journal of Wildland Fire*, 22, 493–507.
30. Ferreira, A., & de Haan, L. (2015). On the block maxima method in extreme value theory: PWM estimators. *The Annals of Statistics*, 43, 276–298.
31. Franke, J., Härdle, W.K. & Hafner, C.M. (2010). *Statistics of financial markets: An introduction*. Springer Cham, pp.585
32. R Core Team. (2023). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing.
33. Gilleland, E., Ribatet, M., & Stephenson, A. G. (2013). A software review for extreme value analysis. *Extremes*, 16, 103–119.
34. Jones, D. A. (1997). Plotting positions via maximum-likelihood for a non-standard situation. *Hydrology and Earth System Sciences*, 1, 357–366.
35. Gilleland, E., & Katz, R. W. (2016). ExtRemes 2.0: An extreme value analysis package in R. *Journal of Statistical Software*, 72, 1–39.
36. Smith, R. L. (1989). Extreme value analysis of environmental time series: An application to trend detection in ground-level ozone. *Statistical Science*, 4, 367–377.
37. Bezak, N., Brilly, M., & Šraj, M. (2014). Comparison between the peaks-over-threshold method and the annual maximum method for flood frequency analysis. *Hydrological Sciences Journal*, 59, 959–977.
38. Gemitzi, A., & Koutsias, N. (2022). A Google Earth Engine code to estimate properties of vegetation phenology in fire affected areas – A case study in North Evia wildfire event on August 2021. *Remote Sensing Applications: Society and Environment*, 26, Article 100720.
39. Hosking, J. R. M., Wallis, J. R., & Wood, E. F. (1985). Estimation of the generalized extreme-value distribution by the method of probability-weighted moments. *Technometrics*, 27, 251–261.
40. Caires, S., & Sterl, A. (2005). 100-year return value estimates for ocean wind speed and significant wave height from the ERA-40 data. *Journal of Climate*, 18, 1032–1048.
41. Lang, M., Ouarda, T. B. M. J., & Bobée, B. (1999). Towards operational guidelines for over-threshold modeling. *Journal of Hydrology*, 225, 103–117.
42. Joe, H., Smith, R. L., & Weissman, I. (1992). Bivariate threshold methods for extremes. *Journal of the Royal Statistical Society. Series B (Methodological)*, 54, 171–183.
43. Niu, D., Sayed, T., Fu, C., & Mannering, F. (2024). A cross-comparison of different extreme value modeling techniques for traffic conflict-based crash risk estimation. *Analytic Methods in Accident Research*, 44, Article 100352.
44. Ali, Y., Haque, M. M., & Zheng, Z. (2022). An extreme value theory approach to estimate crash risk during mandatory lane-changing in a connected environment. *Analytic Methods in Accident Research*, 33, Article 100193.
45. Farah, H., & Azevedo, C. L. (2017). Safety analysis of passing maneuvers using extreme value theory. *Iatss Research*, 41, 12–21.
46. MacPherson, L. R., Arns, A., Fischer, S., Méndez, F. J., & Jensen, J. (2023). Bayesian extreme value analysis of extreme sea levels along the German Baltic coast using historical information. *Natural Hazards and Earth System Sciences*, 23, 3685–3701.

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