DOI:10.25092/baunfbed. 1695930 *J. BAUN Inst. Sci. Technol., 28(1), 171-190, (2026)*

Anomaly detection of earthquakes with Benford's law and extreme value approach

Ceren ÜNAL[1,*], Gamze ÖZEL[2]

[1,2] *Hacettepe University, Faculty of Science, Department of Statistics, 06800, Ankara, Türkiye.*

Geliş Tarihi (Received Date): 09.05.2025
Kabul Tarihi (Accepted Date):30.09.2025

Abstract

Türkiye is located on active fault lines where significant tectonic movements occur regularly leading to earthquakes. Western Anatolia, being in the North Anatolian Fault Zone, has a higher earthquake risk. This study focuses on detecting anomalies in earthquake magnitudes in Türkiye, spanning the years 1970 to 2021. The analysis utilizes Benford's Law and the Generalized Extreme Value distribution within the Block Maxima framework to examine the conformity of earthquake magnitude data and identify potential anomalies. This study is one of the first to apply these methods specifically to the Aegean Region, providing new insights into the seismic risk in this tectonically active region. The dataset, comprising 15845 seismic events with magnitudes of 3.0 and above, was evaluated using both statistical methods. Benford's Law indicated that the data conforms to natural patterns, with statistical tests confirming the expected distribution of first digits in the magnitude data. The Generalized Extreme Value distribution, applied through the Block Maxima approach, revealed that the expected magnitude of extreme events increases with longer return periods, with return levels estimated at 4.77 Mw for 2 years, 5.81 Mw for 20 years, and 6.36 Mw for 100 years. These findings highlight the importance of using these statistical methods to assess seismic risks and identify potential anomalies in earthquake data.

Keywords: *Benford's Law, block maxima, earthquake hazard, generalized extreme value approach.*

[*]Ceren ÜNAL, cerenunal@hacettepe.edu.tr, http://orcid.org/0000-0002-9357-1771
Gamze ÖZEL KADILAR, gamzeozl@hacettepe.edu.tr, http://orcid.org/0000-0003-3886-3074

Benford yasası ve uç değer yaklaşımı ile depremlerde anomali tespiti

Öz

Türkiye, depremlere yol açan önemli tektonik hareketlerin düzenli olarak meydana geldiği aktif fay hatları üzerinde yer almaktadır. Kuzey Anadolu Fay hattında yer alan Batı Anadolu, daha yüksek bir deprem riskine sahiptir. Bu çalışma, Türkiye'de 1970-2021 yıllarını kapsayan deprem büyüklüklerindeki anomalileri tespit etmeye odaklanmaktadır. Analizde deprem büyüklüğü verilerinin uygunluğunu incelemek ve potansiyel anomalileri belirlemek için Benford Yasası ve Blok Maksima çerçevesinde Genelleştirilmiş Uç Değer dağılımını kullanmaktadır. Bu çalışma, söz konusu yöntemleri özellikle Ege Bölgesi'ne uygulayan ilk çalışmalardan biri olmakla beraber tektonik açıdan aktif bu bölgede deprem riski üzerine yeni bakış açıları sunmaktadır. Büyüklüğü 3.0 ve üzerinde olan 15845 sismik olaydan oluşan veri seti her iki istatistiksel yöntem kullanılarak değerlendirilmiştir. Benford Yasası, verilerin doğal örüntülere uygun olduğunu göstermiş ve istatistiksel testler büyüklük verilerindeki ilk rakamların beklenen dağılımını doğrulamıştır. Blok Maksima yaklaşımıyla uygulanan Genelleştirilmiş Uç Değer dağılımı, ekstrem olayların beklenen büyüklüğünün daha uzun geri dönüş süreleriyle arttığını, geri dönüş seviyelerinin 2 yıl için 4,77 Mw, 20 yıl için 5,81 Mw ve 100 yıl için 6,36 Mw olarak tahmin edildiğini ortaya koymuştur. Bu sonuçlar, sismik riskleri değerlendirmek ve deprem verilerindeki potansiyel anomalileri belirlemek için bu istatistiksel yöntemlerin kullanılmasının önemini vurgulamaktadır.

Anahtar kelimeler: *Benford yasası, blok maksima, deprem tehlikesi, genelleştirilmiş uç değer yaklaşımı.*

1. Introduction

Earthquakes are among the most devastating natural disasters, frequently occurring along seismic belts worldwide. These events often result in significant loss of life and property, profoundly affecting communities [1, 2]. The sudden and destructive nature of earthquakes makes them a major concern for scientists and policymakers alike. Understanding the underlying causes and characteristics of earthquakes is crucial for mitigating their effects. Over the years, research has focused on identifying the triggers of seismic activity, exploring why certain regions are more prone to earthquakes, and developing predictive models for earthquake magnitudes [3].

Türkiye is situated at the convergence of several major tectonic plates, including the Anatolian, European, and Asian plates, as well as the African and Arabian plates. This complex tectonic setting makes the region highly susceptible to seismic activity [4]. Historical records show that Anatolia has experienced numerous significant earthquakes over the centuries. The Aegean region, particularly Western Anatolia, is a notable area of intense seismic activity, even though it lies outside the North Anatolian Fault Zone (NAFZ). This region is one of the fastest-changing areas in terms of geological deformation. The western parts of Türkiye, especially the Marmara Region, are particularly vulnerable to earthquakes, with the NAFZ being a critical fault line capable

of triggering major seismic events [5]. Türkiye's provinces are home to dynamic cities and industrial hubs with large populations and considerable economic significance. Cities such as Istanbul, Izmir, Bursa, and Ankara serve as major trade centre's, supporting diverse commercial, tourism, and industrial sectors. However, it is essential to consider the earthquake risks in these areas, as seismic events can severely disrupt commercial operations [6].

Several researchers have extensively studied the tectonic structure of Western Anatolia. Major cities like Izmir and surrounding provinces in the Aegean region are highly susceptible to earthquake hazards due to active fault lines [7-12]. The earthquake in the Yenice-Gönen area of Balıkesir on March 18, 1953, which had a magnitude of 7.4, resulted in approximately 265 fatalities and widespread structural damage [13]. Similarly, a 7.2 magnitude earthquake affected Izmir, situated near these fault lines, on June 6, 1970, in the Gediz River valley area, causing about 1,086 deaths and extensive damage in Gediz [14]. More recently, on October 30, 2020, a powerful 6.9 magnitude earthquake hit near the coast of Samos close to Izmir. It was one of the largest seismic events in the region and was accompanied by 5,799 aftershocks in the subsequent two months, emphasizing the area's seismic risk [15].

There are various approaches to earthquake prediction, which can generally be divided into two groups. The first is based on the experimental observation of precursor changes. This type includes studies on precursor seismic activity, ground movements, and other similar changes. The second approach is based on the statistical occurrence of seismicity. In the literature, there are numerous studies on the analysis and modelling of earthquake data. Yegulalp and Kuo [16] made statistical predictions of the maximum intensity earthquakes in their study. Knopoff and Kagan [17] focused on extreme value analysis has been used to assess earthquake data. Kagan and Jackson [18] examined the probabilistic forecasts of earthquakes. Pisarenko and Sornette [19] focused on deviations from the Gutenberg-Richter distribution for earthquakes with magnitudes 8.0 and larger and their statistical detection. Apostol [20] proposed a model for seismic predictions using a logarithmic distribution. Lombardi and Marzocchi [21] employed a stationary Poisson process for the temporal distribution of major worldwide earthquakes. Kagan [22] worked on discrete statistical distributions for modelling earthquake counts, including Poisson, geometric, logarithmic, and negative binomial distributions. Pisarenko et al. [23] studied the distributions of maximum intensity earthquakes using the generalized Pareto and extreme value distributions. Pisarenko et al. [24] focused on the statistical estimation of tail distributions of earthquake magnitudes, using extreme value theory for maximum magnitudes in successive time intervals. In addition to the studies mentioned, recent years have also seen significant contributions. Dutfoy [25] proposed an innovative approach that employs extreme value models derived from the Poisson process to estimate the tail distribution of earthquake magnitudes and the maximum annual earthquake magnitude. On the other hand, the Western Anatolian region of Türkiye has not been studied in detail with extreme value theory before.

Benford's Law is a mathematical rule that examines the distribution of the first digits of data and is often used in the analysis of natural phenomena and economic data. Applying Benford's Law to studies on air pollution values is an effective method for ensuring the accuracy of the data and detecting potential anomalies. Benford's Law is widely used to detect data manipulation [26]. In the environmental sciences, Benford's Law has been utilized in the analysis of various data sets, such as climate change data or environmental

pollution levels [27]. In Özkan's [28] study, Benford's Law was used to evaluate the naturalness of ecosystems. This research contributed to detecting human intervention or unnatural changes by examining numerical patterns in ecosystem data using Benford's Law. Benford's Law have also become popular to detect anomaly of the earthquake parameters. De and Sen [29] found that the first digit distribution of Benford's Law altered at the critical phase transition point in a quantum magnetic field system, suggesting its applicability in detecting natural phenomena like quantum phase transitions and earthquakes. Sottili et al. [30] reported that the recurrence intervals of 17,000 seismic events in Italy closely conformed to Benford's Law. Diaz et al. [31] showed that longer observation windows better identified local seismic events, and that richer frequency signals conformed to Benford's Law. However, Benford's Law has not been used for the detection of anomalies of earthquakes in Türkiye.

The extreme value theory has not been studied in the Western Anatolian region of Türkiye previously. Additionally, Benford's Law has not been applied to detect earthquake anomalies in Türkiye. This study utilizes Benford's Law alongside extreme value theory to identify anomalies in seismic activities in Türkiye for the first time. The structure of this article is as follows: Section 2 outlines the study's methodology, while Section 3 provides a description of the study area and data catalogue. The results are presented in Section 4, and the article concludes with Section 5.

2. Methodology

2.1. Benford's law

Benford's Law, also referred to as the first-digit law, is a significant pattern found in data sets that was first identified more than a century ago, then forgotten, and then rediscovered in the 20th century. According to this law, the initial digit's frequency of recurrence in many real-world data sets is not constant; rather, it indicates a preference for smaller digits. Benford's Law states that the probability that a number in a data set would have a 1 as its initial digit is greater than the probability of 2 and that the likelihood of 2 is greater than the probability of 3. The least likely number to appear as the first digit is 9. Specifically, the probability distributions from 1 to 9 are, in accordance with this law, 0.3010, 0.1761, 0.1249, 0.0969, 0.0792, 0.0669, 0.0580, 0.0512, and 0.0458. The probability were calculated using the following formula.

$$\mathrm{P(d)} = \log_{10}\left(1 + (1/\mathrm{d})\right),\ \mathrm{d} = 1, 2, \ldots, 9. \quad (1)$$

In the late nineteenth century, Newcomb [32] first proposed the law. However, Benford's Law remained largely ignored by the scientific community for more than five decades. In 1938, engineer Frank Benford rediscovered the law while analysing data from various sources, including financial information, city populations, and statistics from the American Baseball League. Each of these datasets followed the expected proportionality described by the law [33]. Benford's rediscovery brought significant attention to the principle, leading to its widespread acceptance and eventual attribution of the law to him. He expanded the rule to apply to any logarithmic base and confirmed its relevance beyond just the leading digit. Varian [34] was the first to suggest that the degree of fit to Benford's Law could be used to assess the authenticity of certain datasets. This concept was further developed by Nigrini [35, 36], who proposed applying Benford's Law in forensic accounting to verify the integrity of data. When it is confirmed that a dataset

follows Benford's Law under normal conditions, any deviations can be considered evidence of data manipulation. This method has proven effective in detecting fraud in business accounts, tax filings, and stock markets, and is now legally recognized as evidence in the U.S. legal system [28].

The application of Benford's Law to the natural sciences was examined by Nigrini and Miller [37], who found that hydrological streamflow statistics adhere to the law, while the distribution of lake and wetland sizes does not. Sambridge et al. [38] investigated the law's applicability to datasets from physics, chemistry, geophysics, and astronomy, including pulsar rotation frequencies, atmospheric temperature anomalies, and the spread of infectious diseases. Geyer and Martí [39] assessed the conformity of various volcanology-related datasets, such as the area and age of collapse calderas and volcanic eruption durations. In seismology, Sambridge et al. [38] showed that seismic wave speeds beneath the southwest Pacific and a large database of global earthquake depths follow Benford's Law.

In analyses performed according to Benford's Law, various statistical tests are used to examine the results [40]. In this evaluation process, the differences between the observed and expected ratios are analysed. The tests used to analyse these deviations are Z-Statistics, Kolmogorov-Smirnov, Chi-Square and Mean Absolute Deviation. The Z-Statistic test, the confidence intervals (lower bound and upper bound) at 95% are, respectively, given as

$$Z = \left[|p_0 - p_e| - \frac{1}{2n}\right] / \sqrt{\left(p_e * \frac{1-p_e}{n}\right)}, \tag{2}$$

$$Z_{lower} = p_e - \left[1.96 * \sqrt{\left(p_e * \frac{1-p_e}{n}\right)} - \left(\frac{1}{2n}\right)\right], \tag{3}$$

$$Z_{upper} = p_e + \left[1.96 * \sqrt{\left(p_e * \frac{1-p_e}{n}\right)} + \left(\frac{1}{2n}\right)\right]. \tag{4}$$

Here, p_0 is the observed value according to Benford's Law, p_e is the expected value in the dataset [27]. The number of observations is symbolized as n. For a two-tailed 95% confidence interval, the corresponding critical value is 1.96 [41].

Chi-square test is used to test the suitability of the data to the expected distributions, and the formula is given as

$$\chi^2 = \sum \frac{(p_0 - p_e)^2}{p_e}. \tag{5}$$

The chi-square value obtained in this test is compared with a determined critical value. If this value is below the critical value, it indicates that the data conform to the expected distribution; if it is not below the critical value, it indicates that the data do not conform to the expected distribution [42].

2.2. Extreme value theory

Extreme Value Theory (EVT) was developed as an alternative to traditional statistical methods that do not give effective results in data sets containing a small number of

observations and has emerged as an important statistical method for the quantification of the stochastic process of extraordinary events. Fisher and Tippet [43] introduced the theoretical basis of the extreme value theory.

EVT deals with the frequency and magnitude of events that have a very low probability of occurring. Therefore, it is mainly used to make future predictions from data obtained in the past regarding extraordinary events. Instead of the entire sample, extreme values are dealt with, and the aim is to model these values [44]. EVT is similar to the Central Limit Theorem, but unlike the Central Limit Theorem, it is used in modelling the tail part of the distribution [45]. Today, EVT has a wide range of applications in many fields of science, including economics, earth science, weather forecasting, geological engineering, hydrogeological engineering, etc.

In the EVT, some techniques are needed to make statistical inferences about rare events using a limited amount of data. There are two widely used methods in theory. These methods are the Block Maxima (BM) approach and the Peaks-Over-Threshold (POT) method. In the BM Approach, the data is divided into independent periods of equal length, and the smallest or largest values in each period are determined. With this approach, it is assumed that the selected extreme values comply with one of the extreme distributions specified in the theory. The disadvantage of the method is the need for a large number of observations. The POT Approach was developed after the BM Method. It generally provides more effective use of unusual data, which is limited in number. For this reason, it is considered more practical in practice and is preferred more because it is relatively easier. In this method, the researcher generally models the values above a high threshold value determined by graphical methods by not determining them as extreme values. The selected high threshold value prevents some extreme values from being included in the analysis; a low threshold value may cause data that are not large enough to be included in the analysis. Therefore, determining the appropriate threshold value is very important at this stage. The methods used to estimate the threshold value are mostly visual. These methods include the Hill plot, the mean excess plot, and Q-Q graphs [46, 47].

The BM Approach is based on the Generalized Extreme Values (GEV) distribution. Extreme value theory is the modelling of the asymptotic distribution of the series obtained from the maximum or minimum values of the sample data set and shows that under similar conditions, the standardized maximum or minimum series converge to one of the Gumbel, Frechet or Weilbull distributions. It is possible to collect these three limit distributions under a single standardized distribution and this distribution is called the Generalized Extreme Values (GEV) distribution. In the first step of the method, there is a stage of dividing the observed data series into equal-sized blocks. The minimum or maximum value within that block is selected from each block. This is known as block maxima. Parameter estimates are obtained from the selected maximum values, usually using the GEV distribution. The probability density function of the GEV distribution is given below:

$$f(x) = \frac{1}{\sigma} t(x)^{\xi+1} e^{-t(x)}, \tag{6}$$

where

$$t(x) = \begin{cases} \left[1 + \xi\left(\frac{x-\mu}{\sigma}\right)\right]^{-\frac{1}{\xi}} & if\ \xi \neq 0 \\ exp\left(-\frac{x-\mu}{\sigma}\right) & if\ \xi = 0. \end{cases} \quad (7)$$

There are three important parameters in the GEV distribution. These are μ location parameter, σ scale parameter and ξ shape parameter. The location parameter (μ), indicating the central tendency of the maxima; the scale parameter (σ), reflecting the variability among the annual maxima; and the shape parameter (ξ), defining the tail behaviour of the distribution. The shape parameter was used to determine whether the tail of the distribution is heavy, light, or has an exponential decay, which directly influences the estimation of extreme event probabilities.

3. Study area and earthquake catalogue

Türkiye is primarily located in Western Asia and forms part of the larger Eurasian landmass. It lies between latitudes 36-42°N and longitudes 26-45°E, positioned within a seismically active region at the collision zone of the Eurasian, Arabian, and African Plates. Several active fault lines influence the country, where tectonic movements result in frequent earthquakes. Each region must be analysed independently due to its distinct tectonic structures and fault lines, which lead to variations in earthquake risks and magnitudes. The North Anatolian Fault Zone, which extends from west to east, is known for producing significant earthquakes, making the western part of Türkiye particularly vulnerable to seismic activity. Conversely, the eastern part of Türkiye lies along the East Anatolian Fault Zone, which also has the potential to trigger major earthquakes. Notably, two powerful earthquakes occurred in eastern Türkiye in 2023. Despite this, the western region remains at high risk for destructive earthquakes, which is why our focus is on this area.

The western part of Türkiye is also influenced by the southward movement of the Hellenic arc, contributing to seismic activity [48, 49]. The Aegean Region is particularly important due to its active tectonics, as the relative movement of the African-Arabian plates help maintain the extensional system in western Türkiye. The extension is driven by the rollback of the Hellenic subduction zone [48, 50, 51].

Recent earthquakes, such as the Acıpayam (Denizli) earthquake on March 20, 2019 (Mw 5.5), the Bozkurt (Denizli) earthquake on August 8, 2019 (Mw 6.0), the Bodrum (Muğla) earthquake on July 21, 2017 (Mw 6.6), the Manisa earthquake on June 26, 2020 (Mw 5.5), and the Seferihisar (İzmir) earthquake on October 30, 2020 (Mw 6.6), highlight the ongoing seismic activity in the region [52]. For example, the Izmir-Seferihisar earthquake on October 30, 2020, caused 115 fatalities and significant structural damage in Izmir-Bayraklı, 70 km from the epicentre. This earthquake also affected the coastal areas between Akarcik and Sigacik. Therefore, research on earthquake prediction in the Western Anatolia Region, particularly in İzmir (the third-largest city), is crucial [53].

The study focuses on the Aegean Region and its surrounding 200 km buffer zone. This region, characterized by a complex tectonic structure, encompasses multiple independent fault systems. Şengör [54, 55] identified four distinct neotectonic provinces within Türkiye: (1) the North Anatolian province, (2) the Eastern Anatolian contractional

province, (3) the Central Anatolian 'Ova' province, and (4) the Western Anatolian extensional province, each exhibiting unique tectonic characteristics.

The catalogue created for the study area was compiled by reviewing three earthquake catalogues covering the study area. The first of these earthquake catalogues is the Disaster and Emergency Management Presidency, Earthquake Department catalogue (AFAD-DDA) catalogue, which is considered as the basic catalogue. As the other two earthquake catalogues, Kandilli Observatory Earthquake Catalogue [4] and Tan [56] earthquake catalogues were considered. To control the parameters of earthquakes in the AFAD-DDA catalogue, it is aimed to detect common earthquakes between AFAD-DDA and KOERI and AFAD-DDA and Tan [56] catalogues. The following criteria were taken into consideration for the determination of common earthquakes: Time difference between earthquakes $\leq$ 3 minutes, position difference: (Longitude 1-Longitude 2 $\leq$ 0.5 degrees, Latitude 1-Latitude 2 $\leq$ 0.5 degrees). The results show that there is a good correlation between AFAD-DDA and Tan (2021) catalogues in terms of earthquake magnitudes, especially for earthquakes of magnitude 3.0 and above. The fact that the comparisons between the two earthquake catalogues are based on a total of 153,533 earthquakes indicate that the AFAD-DDA catalogue is reliable. As the next step, the study catalogue was homogenized via earthquake magnitude and all earthquakes were converted to Mw earthquake scale. Various empirical relations have been used to convert earthquake magnitudes. Scordilis [57], Kadirioğlu and Kartal [58], and Tan [56] magnitude transformation relations were used.

The most important factor in the selection of these relations is that they are produced from earthquake catalogues and that they are transformation relations that are frequently used for earthquakes in the region [57]. Finally, Gardner and Knopoff [59] declustering method was chosen as the catalogue declustering approach. It is one of the standard time-space window methods. The project catalogue was cleared of foreshocks and aftershocks using the ZMAP program, using Gardner and Knopoff [59] delustering algorithm. The decoupled catalogue obtained contains 16,503 earthquakes. Totally, 5284 clusters were identified in the decoupling analysis and 30,107 earthquakes were removed from the catalogue.

4. Results

In this study, Benford's Law and the BM approach with the GEV distribution were chosen as they directly aligned with the objectives of the research. Benford's Law has long been used to evaluate the naturalness and reliability of numerical datasets, with the first-digit analysis being the most common and sensitive approach in the literature for anomaly detection. Although higher-order digit analyses can also be performed, the first-digit analysis was considered sufficient for testing the overall structure of the dataset in this study. For extreme value analysis, the BM approach with the GEV distribution was studied. Since the data set covers a long time period (1970-2021), it is well suited for the construction of annual blocks. An alternative method, the POT approach, is highly sensitive to the choice of threshold, which, if determined subjectively or incorrectly, may lead to unstable estimates. Therefore, to avoid such subjectivity and to ensure comparability with related studies in seismology, the BM and GEV distribution framework was preferred.

Earthquake data from 1970 to 2021 were analysed, covering 15845 recorded events, in this study. To provide a clear overview of the dataset, figures and descriptive statistics were presented, categorizing the earthquakes by magnitude. The analysis included graphical representations of magnitude distribution, depth profiles, depth-magnitude scatter plots, and time-dependent magnitude trends. These analyses were conducted for earthquakes with magnitudes of 3.0, 4.0, 5.0, and 6.0 and above, offering a detailed understanding of the seismic activity in the region across different magnitude levels. As seen from Figures 1 and 4, the earthquake data reveals that as earthquake magnitude increases, there is a corresponding increase in the depth at which these events occur, indicating that more significant seismic events are generally associated with deeper tectonic processes.

Specifically, earthquakes with magnitudes of 3.0 and above tend to be more frequent and occur at shallower depths, typically between 0 to 30 km (Figure 1). However, as the magnitude threshold increases to 4.0 (Figure 2), 5.0 (Figure 3), and 6.0 (Figure 4), there is a clear trend toward deeper seismic activity, with larger earthquakes often originating between 20 to 50 km below the Earth's surface.

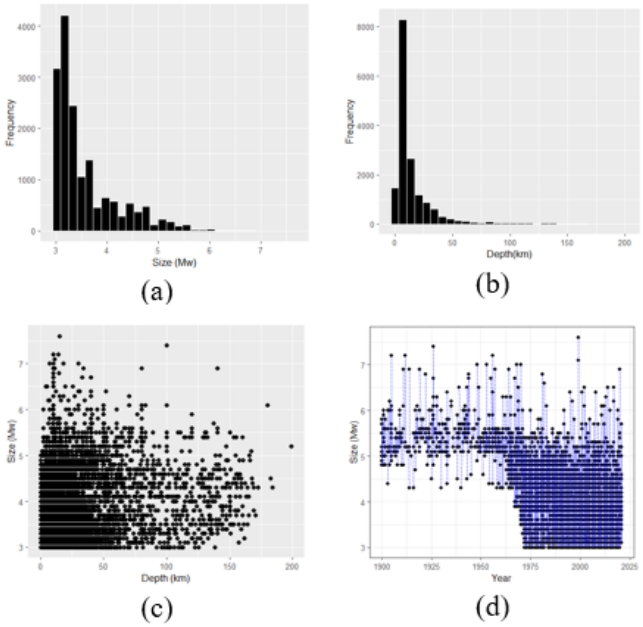

Figure 1. Magnitude (a), depth (b), depth-magnitude scatter (c) and time-dependent magnitude line (d) graphs of earthquakes with magnitude 3.0 and above.

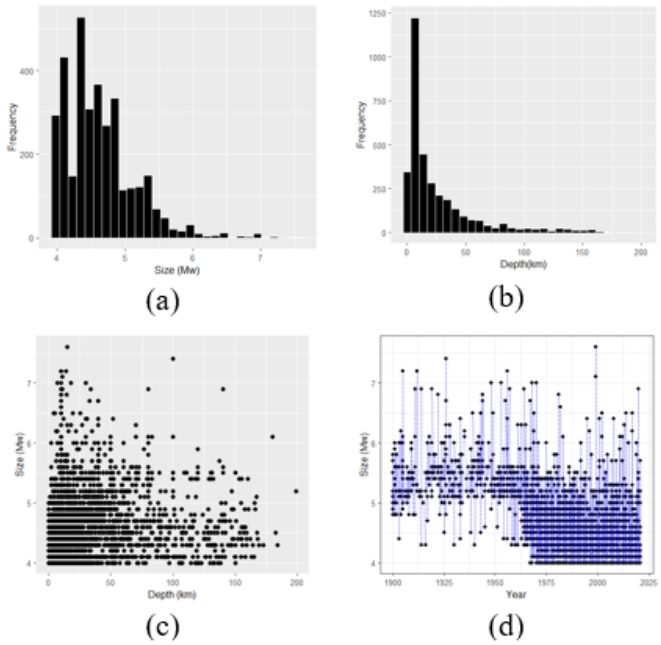

Figure 2. Magnitude (a), depth (b), depth-magnitude scatter (c) and time-dependent magnitude line (d) graphs of earthquakes with magnitude 4.0 and above.

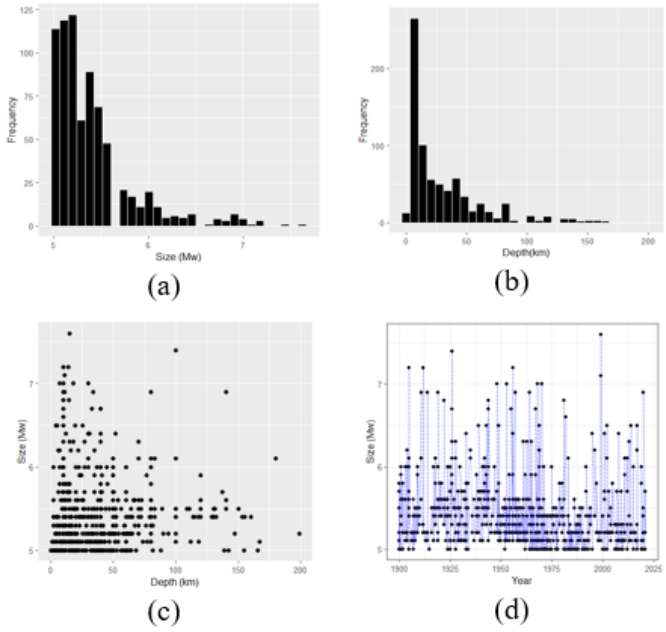

Figure 3. Magnitude (a), depth (b), depth-magnitude scatter (c) and time-dependent magnitude line (d) graphs of earthquakes with magnitude 5.0 and above.

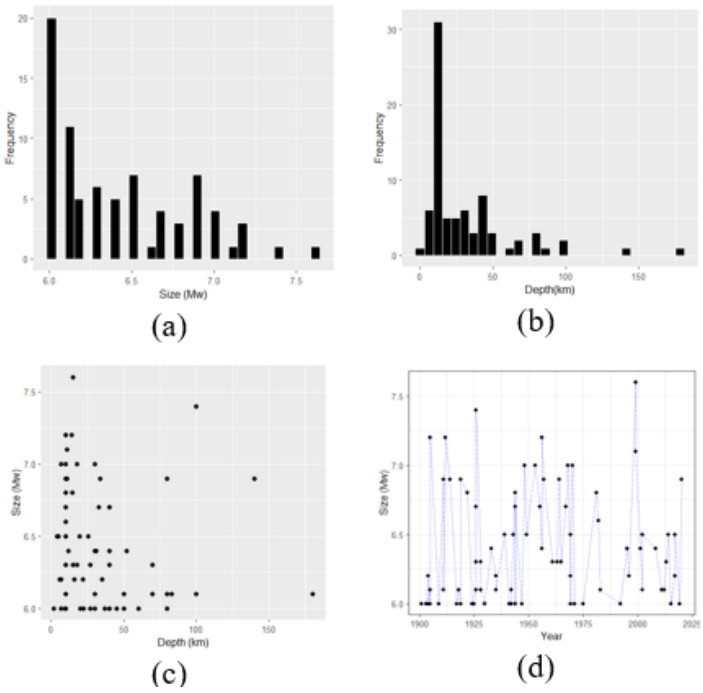

Figure 4. Magnitude (a), depth (b), depth-magnitude scatter (c) and time-dependent magnitude line (d) graphs of earthquakes with magnitude 6.0 and above.

Table 1 presents a detailed statistical analysis of earthquake magnitudes and depths recorded between 1970 and 2021, focusing on key descriptive statistics such as minimum and maximum values, mean, skewness, kurtosis, and quartile values (Q_1, Q_2, and Q_3) across different magnitude thresholds.

Table 1. Statistical analysis of earthquake magnitude and depth by thresholds.

	3.0 and above		4.0 and above		5.0 and above		6.0 and above	
	M	D	M	D	M	D	M	D
Minimum	3.0	0	4.0	0	5.0	0	6.0	2
Q1	3.1	6.9	4.2	9.6	5.1	10	6.05	10
Q2	3.3	10	4.6	2	5.3	16.55	6.3	18
Q3	3.8	17	4.9	33	5.5	40	6.75	40
Mean	3.5	16.62	4.6	26.58	5.4	31.35	6.42	30.32
Maximum	7.6	199	7.6	199	7.6	199	7.6	180
Skewness	1.63	3.68	1.34	2.25	1.96	2.06	0.76	2.31
Kurtosis	5.63	19.28	5.79	8.15	7.47	7.67	2.56	9.57

M: Magnitude (Mw), D: Depth (km)

As seen in Table 1, earthquakes with magnitudes of 3.0 and above, seismic events often occur at surface levels, with depths starting from zero kilometres. As the magnitude threshold increases, the data indicate that earthquakes tend to occur at greater depths, with

those of 6.0 and above typically originating from depths of 2 km or more. The quartile distribution further supports this trend, showing that higher magnitude earthquakes are associated with deeper origins, with the third quartile for 6.0 and above reaching depths of up to 40 km. The average depth for the highest magnitude category (6.0 and above) is approximately 30.32 km, suggesting that larger earthquakes are more likely to occur deeper within the Earth's crust. The maximum recorded magnitude across all thresholds is 7.6, with maximum depths reaching 199 km. The data also exhibit right-skewed distributions, indicating that smaller and shallower earthquakes are more common, while extreme values (higher magnitudes and greater depths) are less frequent. The high kurtosis values observed indicate a leptokurtic distribution, with significant outliers representing exceptionally deep or shallow seismic events.

Benford's Law was used to evaluate the extreme values on earthquake magnitudes for the period 1970-2021. The frequencies of earthquake magnitudes were obtained from 15845 earthquakes with magnitude 3.0 and above. The analysis of compliance with Benford's Law was performed on the R Studio program. Mantissa statistics refer to the descriptive measures derived from the decimal parts of the logarithmic values of earthquake magnitudes. In the context of Benford's Law, the mantissa is defined as the decimal part of a number's logarithm, which is expected to be uniformly distributed in the interval [0,1) if the data conform to natural phenomena. The Mantissa statistics are given in Table 2. The mean and variance show the central tendency and distribution of the data, while the excess kurtosis and skewness show whether the data is symmetrical and in which direction the tail is.

Table 2. Mantissa statistics.

Mantissa Statistics	Value
Mean	0.403
Variance	0.091
Kurtosis	-1.033
Skewness	0.096

The mean of the mantissa values was found to be 0.403. This value shows that the data generally has a positive distribution. The variance value of 0.091 shows that the data is relatively tightly clustered around the mean. Low variance indicates that the data is homogeneous and contains values close to the mean. Kurtosis (-1.033) indicates that the data has a lower peak and thinner tails compared to a normal distribution. This shows that the data is closer to the central values rather than the extreme values. The positive skewness value (0.096) indicates that the data has a long tail to the right. This shows that the data has mostly low values and some high values. In general, the earthquake magnitude data has a distribution that is concentrated in the centre and shows a long tail to the right.

Pearson's Chi-square test results and Figure 5 show how well the data fits the Benford's Law. The Pearson's Chi-square test value is 3.7378 $\left(\chi^2 = 3.7378 < \chi^2_{0.95;8} = 15.51\right)$ and the p-value $(p = 0.88 > 0.05)$ shows that the data fits the Benford Law.

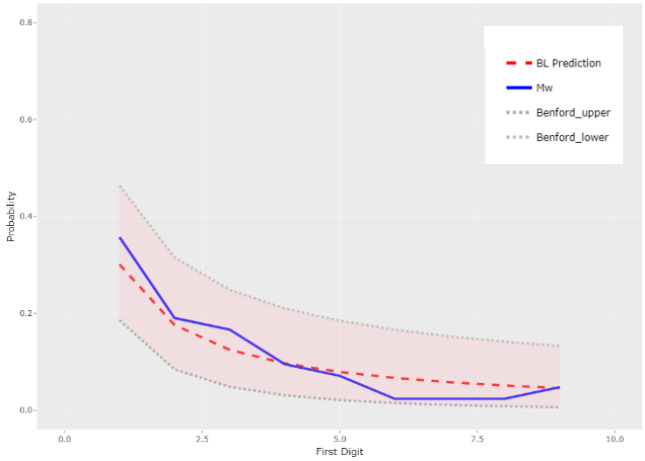

Figure 5. Curves with lower and upper confidence intervals at 95% significance level.

Figure 5 shows that the curve (Mw) is within the 95% confidence interval and Table 3 presents Benford's Law analysis of magnitudes with confidence intervals.

Table 3. Benford's Law analysis of earthquake magnitudes with confidence intervals.

Benford's Law	Mw	Confidence Interval Lower	Confidence Interval Upper
0.30103000	0.35714286	0.186111054	0.4635680
0.17609126	0.19047619	0.084704125	0.3150974
0.12493874	0.16666667	0.048748442	0.2487481
0.09691001	0.09523810	0.031248748	0.2101903
0.07918125	0.07142857	0.021326954	0.1846546
0.06694679	0.02380952	0.015168896	0.1663437
0.05799195	0.02380952	0.011113978	0.1524890
0.05115252	0.02380952	0.008333067	0.1415910
0.04575749	0.04761905	0.006370678	0.1327634

Table 3 presents the results of applying Benford's Law to earthquake magnitude data, highlighting the observed proportions of leading digits (Mw) alongside their corresponding 95% confidence intervals. The analysis reveals that the observed frequency of the leading digit 1 (0.3571) slightly exceeds the expected value (0.3010) but remains within the confidence interval (0.1861 - 0.4636), indicating conformity to Benford's Law. Similarly, the digit 2 shows an observed frequency (0.1905) close to the expected value (0.1761), also falling within its confidence interval (0.0847 - 0.3151). As expected, the frequencies of higher leading digits (3 through 9) progressively decrease, aligning with Benford's Law. Although the observed frequencies for digits 6 through 9 are lower than the expected values, they still fall within the confidence intervals, suggesting no significant anomalies. Overall, the analysis demonstrates that the earthquake magnitude data conforms well to Benford's Law, indicating the dataset's reliability and lack of unusual patterns or anomalies.

The analysis of earthquake magnitude data reveals significant insights into the distribution and behaviour of seismic events. By utilizing visualizations such as boxplots, time series plots, and histograms, a deeper understanding of the characteristics of these events can gain. One key observation from the data is the presence of several outliers and

a highly skewed distribution, which underscores the importance of applying EVT to model the tail behaviour of the distribution. Figure 6 presents three key visualizations to understand the earthquake magnitude data. Figure 6.a shows the time series of earthquake magnitudes over time with a blue line, Figure 6.b provides a horizontal boxplot to visualize the distribution and outliers, and Figure 6.c shows the histogram in blue tones to depict the frequency distribution of magnitudes.

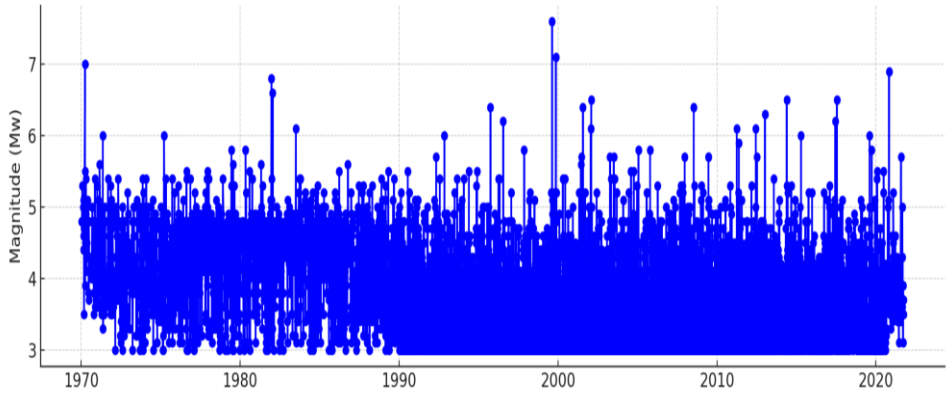

Figure 6.a. Time series of earthquake magnitudes.

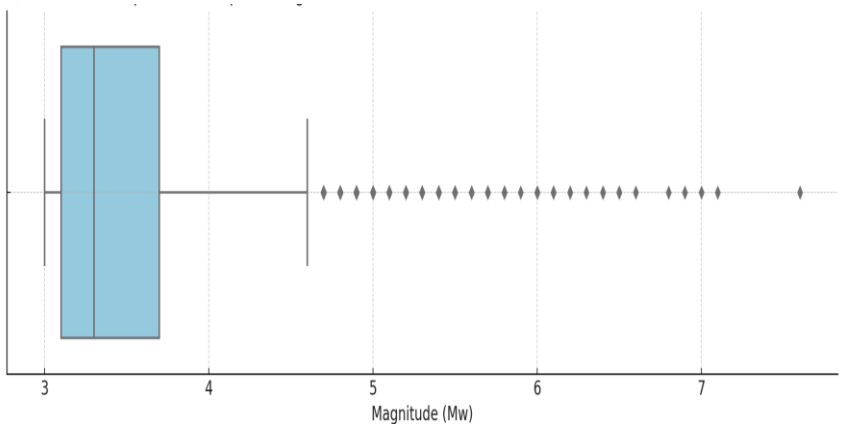

Figure 6.b. Box plot of earthquake magnitudes.

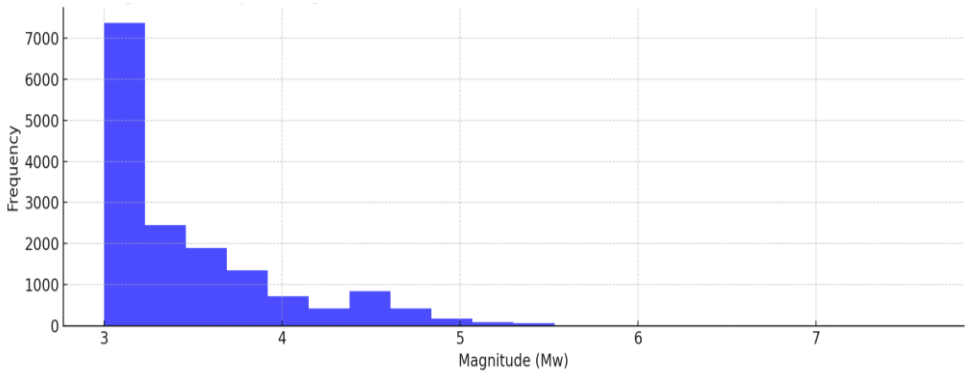

Figure 6.c. Histogram of earthquake magnitudes.

The boxplot in Figure 6.a shows the distribution of earthquake magnitudes, highlighting the median, interquartile range, and outliers. The presence of multiple outliers above 5.0 Mw indicates the occurrence of significant seismic events, although they are less frequent. This distribution suggests that while most earthquakes are of lower magnitude, the potential for extreme events exists. The time series plot of earthquake magnitudes over time in Figure 6.b shows that smaller earthquakes are more frequent, while larger magnitudes occur sporadically throughout the period from 1970 to 2021. The histogram in Figure 6.c illustrates the frequency distribution of earthquake magnitudes, showing a concentration of events around 3.0 Mw and a steep decline in frequency as the magnitude increases. This skewed distribution further supports the application of extreme value theory, as it highlights the tail behaviour that is crucial for understanding the risk of rare but severe earthquakes. The presence of outliers and the highly skewed distribution of earthquake magnitudes make it essential to apply to EVT when modelling this data. The visualizations confirm that the data is dominated by smaller magnitude events, but the potential for extreme, high-magnitude earthquakes necessitates a focus on tail behaviour. For this aim, the BM approach combined with the GEV distribution was employed to analyse the extreme earthquake magnitudes.

Table 4. Maximum Likelihood Estimations (MLEs) and information criteria results.

-log L	457.3836		
Estimated Parameters	Value	Standard Error	
Location (μ)	4.6076	0.0200	
Scale (σ)	0.4590	0.0136	
Shape (ξ)	-0.0840	0.0189	
Estimated Parameter Covariance Matrix	Location	Scale	Shape
Location	0.0004	0.0001	-0.0001
Scale	0.0001	0.0002	-0.0001
Shape	-0.0001	-0.0001	0.0004
Information Criteria		Value	
AIC		920.7672	
BIC		934.0612	

The results of the MLE in Table 4 indicate that the negative log-likelihood value is 457.3836. The estimated parameters for the model are as follows: the location parameter is 4.6076 with a standard error of 0.0200, the scale parameter is 0.4590 with a standard error of 0.0136, and the shape parameter is -0.0840 with a standard error of 0.0189. The covariance matrix of the estimated parameters shows that the location and scale parameters have a covariance of 6.1658e-05, the location and shape parameters have a covariance of -1.0606e-04, and the scale and shape parameters have a covariance of -6.7705e-05. The model's performance is further evaluated by AIC and BIC, which are 920.7672 and 934.0612, respectively. These criteria suggest the relative quality of the model in balancing goodness of fit with complexity.

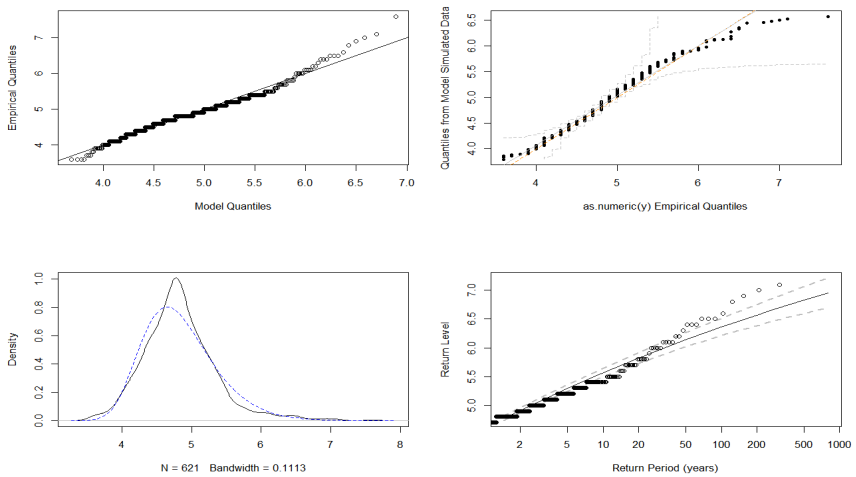

Figure 7. Diagnostic plots and return level analysis for earthquake magnitude data using the generalized extreme value distribution.

Figure 7 presents diagnostic plots for the GEV distribution applied to the dataset. The top left plot is a Q-Q (quantile-quantile) plot comparing the empirical quantiles of the data to the quantiles predicted by the GEV model. The close alignment of points along the 45-degree line suggests that the GEV model provides a good fit to the data. The top right plot is a P-P (probability-probability) plot, which also compares the empirical probabilities with the probabilities predicted by the model. The points closely follow the 45-degree line, further confirming the adequacy of the model. The bottom left plot shows the density estimation of the data, where the GEV distribution seems to capture the peak and spread of the observed data well, indicating a reliable model fit. Finally, the bottom right plot displays the return level plot, which estimates the return level (magnitude) associated with different return periods (in years). Overall, the diagnostic plots indicate that the GEV distribution is a suitable model for the data, effectively capturing the key characteristics of the observed values and providing reliable predictions for extreme events.

Table 5. Estimated return levels for earthquake magnitude.

	95% Lower CI	Estimate	95% Upper CI
2-year return level	4.731194	4.773269	4.815343
20-year return level	5.724967	5.814235	5.903503
100-year return level	6.208971	6.359094	6.509216

Table 5 presents the estimated return levels for earthquake magnitudes corresponding to different return periods (2 years, 20 years, and 100 years). The 2-year return level is estimated to be 4.77, with a 95% confidence interval ranging from 4.73 to 4.82. The 20-year return level is higher at 5.81, with a confidence interval of 5.72 to 5.90. For the 100-year return period, the estimated magnitude is 6.36, with a wider confidence interval from 6.21 to 6.51. These estimates suggest that as the return period increases, the expected magnitude of the earthquake also increases, with a higher degree of uncertainty reflected in the widening confidence intervals for longer return periods.

5. Conclusion

Located at the junction of the Eurasian, Arabian, and African tectonic plates, Türkiye is one of the most active regions in the Mediterranean in terms of seismic activity. The continuous movement and interaction of these major plates contribute significantly to the tectonic dynamics of the region. As a result, the Anatolian Plate, which covers a significant part of Türkiye, carries a significant seismic hazard. This study performed a comprehensive analysis of earthquake magnitudes in the Aegean Region of Türkiye between 1970 and 2021 using Benford's Law and the Generalised Extreme Value distribution. The findings reveal that the earthquake data generally conform to Benford's Law, showing natural patterns in the magnitude distribution, and statistical tests support this conformity. The application of the Generalised Extreme Value distribution allowed the estimation of return levels for extreme seismic events, highlighting the increasing magnitude of such events over longer return periods. In particular, the 100-year return period estimates a magnitude of approximately 6.36 with a wider confidence interval, highlighting the uncertainty associated with predicting rare but severe events. Compared to the existing literature, this study strengthens the applicability of Benford's Law and the extreme value theory in seismic risk assessment. Previous studies such as Sottili et al. [30] and Diaz et al. [31] have demonstrated the effectiveness of Benford's Law in identifying anomalies in seismic data, while studies such as Pisarenko et al. [24] have used extreme value theory to model earthquake magnitudes. However, this study is one of the first to apply these methods specifically to the Aegean Region, providing new insights into the seismic risk in this tectonically active region. The results contribute to the growing body of evidence supporting the use of these statistical methods in understanding and reducing the risks posed by significant seismic events in similar regions. While the primary objective of this study is methodological, the findings also have practical implications. In particular, extreme value estimates and return periods can contribute to the development of regional earthquake hazard scenarios. This could indirectly support more scientifically based decisions in disaster risk management, early warning systems, and urban planning.

References

[1] Stein, S. and Wysession, M., Introduction to Seismology, **Earthquakes, and Earth Structure**, Blackwell, Malden, Mass, (2003).

[2] Kanamori, H., "Real-time seismology and earthquake damage mitigation", **Annual Review of Earth and Planetary Sciences**, 33(1), 195–214, (2005).

[3] Turcotte, D. L. and Schubert, G., **Geodynamics**, 2. Baskı, Cambridge University Press, Cambridge, (2002).

[4] KOERI, "6 Şubat 2023 ML7.5 Ekinözü-Kahramanmaraş Depremi", http://www.koeri.boun.edu.tr/sismo/2/06-subat-2023-ml7-5-ekinozu-kahramanmaras-depremi/, (15.02.2023).

[5] Duman, T. Y., Çan, T., Emre, Ö., Kadirioğlu, F. T., Başarır, B. B., Kılıç, T., Arslan, S., Özalp, S., Kartal, R. F., Kalafat, D., Karakaya, F., Eroğlu, A. T., Özel, N. M., Ergintav, S., Akkar, S., Altınok, Y., Tekin, S., Cingöz, A. and Kurt, A. İ., Türkiye Sismotektonik Haritası Ölçek 1:1.250.000, **Maden Tetkik ve Arama Genel Müdürlüğü, Özel Yayın Serisi – 34**, Ankara, (2017).

[6] Öcal, A., "Natural disasters in Turkey: Social and economic perspective", **International Journal of Disaster Risk Management**, 1(1), 51–61, (2019).

[7] Barka, A. and Reilinger, R., "Active tectonics of the Eastern Mediterranean region: Deduced from GPS, neotectonic and seismicity data", **Annali di Geofisica**, 40, 587–610, (1997).
[8] Seyitoğlu, G. and Scott, B. C., "The age of the Menderes graben (west Turkey) and its tectonic implications", **Geological Magazine**, 29, 239–242, (1992).
[9] Seyitoğlu, G. and Scott, B. C., "Age of Alagehir graben (west Turkey) and its tectonic implications", **Geological Journal**, 31, 1–11, (1996).
[10] Koçyiğit, A., Yusufoğlu, H. ve Bozkurt, E., "Evidence from the Gediz graben for episodic two-stage extension in western Turkey", **Journal of the Geological Society of London**, 156, 605–616, (1999).
[11] Şalk, M., Altıner, Y. and Ergün, M., "Geodynamics of western Turkey and implications", **Proceedings of International Conference on Earthquake Hazard and Risk in the Mediterranean Region**, I, 179–189, (2000).
[12] Kahraman, S., Baran, T. and Şalk, M., "Frequency and Risk Analyses for İzmir and its Surrounding Region Earthquake", **Proceedings of VIth International Conference on Advances in Civil Engineering**, I, 500–509, İstanbul, (2004).
[13] Çağatay, İ. H., "Experimental evaluation of buildings damaged in recent earthquakes in Turkey", **Engineering Failure Analysis**, 12(3), 440–452, (2005).
[14] Bayrak, Y. and Bayrak, E., "An evaluation of earthquake hazard potential for different regions in Western Anatolia using the historical and instrumental earthquake data", **Pure and Applied Geophysics**, 169(10), 1859–1873, (2012).
[15] Karadaş, A. and Öner, E., "The Effects of Alluvial Geomorphology of the Bornova Plain on the Damage Caused by the 30 October 2020 Samos Earthquake in İzmir-Bayraklı", **Journal of Geography**, 42, 139–153, (2021).
[16] Yegulalp, T. M. and Kuo, J. T., "Statistical prediction of the occurrence of maximum magnitude earthquakes", **Bulletin of the Seismological Society of America**, 64(2), 393–414, (1974).
[17] Knopoff, L. and Kagan, Y., "Analysis of the theory of extremes as applied to earthquake problems", **Journal of Geophysical Research**, 82(36), 5647–5657, (1977).
[18] Kagan, Y. Y. and Jackson, D. D., "Probabilistic forecasting of earthquakes", **Geophysical Journal International**, 143(2), 438–453, (2000).
[19] Pisarenko, V. F. and Sornette, D., "Statistical detection and characterization of a deviation from the Gutenberg-Richter distribution above magnitude 8", **Pure and Applied Geophysics**, 161, 839–864, (2004).
[20] Apostol, B. F., "A model of seismic focus and related statistical distributions of earthquakes", **Physics Letters A**, 357(6), 462–466, (2006).
[21] Lombardi, A. M. and Marzocchi, W., "Evidence of clustering and nonstationarity in the time distribution of large worldwide earthquakes", **Journal of Geophysical Research: Solid Earth**, 112, 1–15, (2007).
[22] Kagan, Y. Y., "Statistical distributions of earthquake numbers: consequence of branching process", **Geophysical Journal International**, 180(3), 1313–1328, (2010).
[23] Pisarenko, V. F., Sornette, D. and Rodkin, M. V., "Distribution of maximum earthquake magnitudes in future time intervals: application to the seismicity of Japan (1923–2007)", **Earth, Planets and Space**, 62, 567–578, (2010).
[24] Pisarenko, V. F., Sornette, A., Sornette, D. and Rodkin, M. V., "Characterization of the tail of the distribution of earthquake magnitudes by combining the GEV and GPD descriptions of extreme value theory", **Pure and Applied Geophysics**, 171, 1599–1624, (2014).

[25] Dutfoy, A., "Estimation of tail distribution of the annual maximum earthquake magnitude using extreme value theory", **Pure and Applied Geophysics**, 176(2), 527–540, (2019).
[26] Diekmann, A., "Not the first digit! Using Benford's law to detect fraudulent scientific data", **Journal of Applied Statistics**, 34(3), 321–329, (2007).
[27] Nigrini, M. J. and Mittermaier, L. J., "The use of Benford's Law as an Aid in Analytical Procedures", **Auditing: A Journal of Practice and Theory**, 16(2), 52, (1997).
[28] Özkan, K., "Estimating ecosystem naturalness using Benford's law and generalized Benford's law", **Turkish Journal of Forestry**, 22(2), 73–82, (2021).
[29] De, A. S. and Sen, U., "Benford's law detects quantum phase transitions similarly as earthquakes", **Physics Letters A**, 95(5), 5008, (2011).
[30] Sottili, G., Palladino, D. M. and Giaccio, B., "Benford's law in time series analysis of seismic clusters", **Mathematical Geosciences**, 44, 619–634, (2012).
[31] Diaz, J., Gallart, J. and Ruiz, M., On the ability of the Benford's law to detect earthquakes and discriminate seismic signals, **Seismological Research Letters**, 86, 1, 192–201, (2015).
[32] Newcomb, S., Note on the frequency of use of the different digits in natural numbers, **American Journal of Mathematics**, 4, 1, 39–40, (1881).
[33] Benford, F., The law of anomalous numbers, **Proceedings of the American Philosophical Society**, 551–572, (1938).
[34] Varian, H., Benford's Law, **The American Statistician**, 26, 3, 65, (1972).
[35] Nigrini, M. J., The detection of income tax evasion through an analysis of digital distribution, PhD Thesis, **University of Cincinnati**, Cincinnati, (1992).
[36] Nigrini, M. J., A taxpayer compliance application of Benford's law, **Journal of the American Taxation Association**, 18, 1, 72–91, (1996).
[37] Nigrini, M. J. and Miller, S. J., Benford's law applied to hydrology data—results and relevance to other geophysical data, **Mathematical Geology**, 39, 469–490, (2007).
[38] Sambridge, M., Tkalčić, H. and Jackson, A., Benford's law in the natural sciences, **Geophysical Research Letters**, 37, 22, 1–5, (2010).
[39] Geyer, A. and Martí, J., Applying Benford's law to volcanology, **Geology**, 40, 4, 327–330, (2012).
[40] Cleary, R. and Thibodeau, J. C., Applying digital analysis using Benford's law to detect fraud: The dangers of type I errors, **Auditing: A Journal of Practice and Theory**, 24, 1, 77–81, (2005).
[41] Durtschi, C., Hillison, W. and Pacini, C., The effective use of Benford's law to assist in detecting fraud in accounting data, **Journal of Forensic Sciences**, 5, 1, 17–34, (2004).
[42] Hindls, R. and Hronová, S., Benford's Law and Possibilities for Its Use in Governmental Statistics, **Statistika: Statistics & Economy Journal**, 95, 2, (2015).
[43] Fisher, R. A. and Tippett, L. H. C., Limiting forms of the frequency distribution of the largest or smallest member of a sample, **Proceedings of the Cambridge Philosophical Society**, (1928).
[44] Liu, X., Data analysis in extreme value theory, Master Thesis, University of North Carolina at Chapel Hill, **Department of Statistics and Operations Research**, (2009).
[45] Mallor, F. and Omey, E., An introduction to statistical modelling of extreme values, **Hub Research Paper**, 36, 5–31, (2009).

[46] Hill, B. M., A simple general approach to inference about the tail of a distribution, **Annals of Statistics**, 3, 1163–1174, (1975).

[47] Davison, A. C. and Smith, R. L., Models for exceedances over high thresholds, **Journal of the Royal Statistical Society: Series B (Statistical Methodology)**, 52, 3, 393–425, (1990).

[48] Kalafat, D. and Görgün, E., An example of triggered earthquakes in western Turkey: 2000–2015 Afyon-Akşehir Graben earthquake sequences, **Journal of Asian Earth Sciences**, 146, 103–113, (2017).

[49] Turkan, S. and Özel, G., Modeling destructive earthquake casualties based on a comparative study for Turkey, **Natural Hazards**, 72, 2, 1093–1110, (2014).

[50] Flerit, F., Armijo, R., King, G. and Meyer, B., The mechanical interaction between the propagating North Anatolian Fault and the back-arc extension in the Aegean, **Earth and Planetary Science Letters**, 224, 3–4, 347–362, (2004).

[51] Bohnhoff, M., Harjes, H. P. and Meier, T., Deformation and stress regimes in the Hellenic subduction zone from focal mechanisms, **Journal of Seismology**, 9, 3, 341–366, (2005).

[52] TUBITAK 1002, Earthquake Magnitude Estimation for the Western Anatolia Region with Bidirectional Deep Learning Methods, Proje No: 124F059, (2024).

[53] AFAD Earthquake Department, 30 October 2020 Aegean Sea, Seferihisar (İzmir) Offshore (17.26 km) **Preliminary Evaluation Report on Mw 6.6 Earthquake**, (2020).

[54] Şengör, A. M. C., The North Anatolian transform fault: its age, offset and tectonic significance, **Journal of the Geological Society**, 136, 269–282, (1979a).

[55] Şengör, A. M. C., Mid-Mesozoic closure of Permo-Triassic Tethys and its implications, **Nature**, 279, 590–593, (1979b).

[56] Tan, O., Turkish Homogenized Earthquake Catalogue (TURHEC), **Natural Hazards and Earth System Sciences**, https://doi.org/10.5281/zenodo.5056801, (2021).

[57] Scordilis, E. M., Empirical global relations converting MS and mb to moment magnitude, **Journal of Seismology**, 10, 225–236, (2006).

[58] Kadirioglu, F. T. and Kartal, R. F., The new empirical magnitude conversion relations using an improved earthquake catalogue for Turkey and its near vicinity, **Turkish Journal of Earth Sciences**, 25, 4, 300–310, (2016).

[59] Gardner, J. K. and Knopoff, L., Is the sequence of earthquakes in Southern California with aftershocks removed, Poissonian?, **Bulletin of the Seismological Society of America**, 64, 5, 1363–1367, (1974).