Integer and fractional-order entropy analysis of earthquake data series

António M. Lopes · J. A. Tenreiro Machado

Abstract This paper studies the statistical distributions of worldwide earthquakes from year 1963 up to year 2012. A Cartesian grid, dividing Earth into geographic regions, is considered. Entropy and the Jensen–Shannon divergence are used to analyze and com-pare real-world data. Hierarchical clustering and multi-dimensional scaling techniques are adopted for data visualization. Entropy-based indices have the advan-tage of leading to a single parameter expressing the relationships between the seismic data. Classical and generalized (fractional) entropy and Jensen–Shannon divergence are tested. The generalized measures lead to a clear identification of patterns embedded in the data and contribute to better understand earthquake distrib-utions.

Keywords Complex systems · Dynamical systems · Data series · Entropy · Visualization

1 Introduction

An earthquake is caused by a sudden release of energy in the Earth's crust. Such energy propagates as seismic waves that manifest by shaking and ground displacement. Depending on the location, magnitude or depth, earthquakes can cause serious damage.

Several models use the dynamics of tectonic plates to explain earthquakes [46]. In fact, tectonic plates on the Earth's surface move with respect to each other, due to the convection currents that exist within the mantle [48]. The asperities between the plates cause friction and stick-slip motion [5,6,11]. Such behavior increases stress, while strain energy accumulates around the fault surfaces. Whenever the stress is high enough to break through the asperities, the accumulated energy is released and the plates move suddenly causing an earthquake [31].

Sotolongo-Costa and Posadas [47] present a model for earthquake dynamics consisting of two rough profiles interacting via fragments filling the gaps between adjacent plates. The irregularities of the fault surfaces can interact with the fragments and develop a mechanism for triggering earthquakes. The authors propose an energy distribution function that yields the Gutenberg–Richter (GR) law as a particular case. Hallgass et al. [17] explain the fractal scaling behavior observed in earthquake phenomena through the asperities on the fault surfaces, which act as fractals sliding over each other. Peng and Gomberg [38] found that slow-slip phenomena is not exclusive of subduc-

tion zone plate interfaces. They concluded that slow-slip events are not a distinct mode of fault slip when compared with what occurs in earthquakes. Ekström et al. [13] detected moderate earthquakes beneath large glaciers and observed that the seismic source is well represented by stick-slip, downhill sliding of glacial ice masses.

In certain models, the tectonic plates are viewed as a complex system (CS) owing to interactions among seismic faults [26,41,52]. The loading rates are not uniform in time [49], and the tectonic plates motion and strain accumulation processes interact in a wide range of scales.

Statistical tools applied to real data and clustering analysis have been used to detect precursory seismic events as well as for earthquake forecasting [2,14,25]. However, reliable short-term prediction of earthquakes is not possible at present time [30].

A common property to CS is the absence of a characteristic length-scale, meaning that CS reveals frequency-size power law (PL) behavior [7,31,34,39]. PL distributions have been associated with systems with memory, as is the case of fractional-order systems [4,22]. The space–time distributions of earthquakes have been studied by several authors. The short-time and short-space scaling behavior is well established and is described by the GR and Omori laws, for example. However, providing clear evidence for supporting long-range correlations has been more difficult to reach [34].

The GR law describes the overall frequency distribution of earthquakes [16] in a given region:

$$\log N(m) = q - r \cdot m \tag{1}$$

where $N(m)$ is the frequency of earthquakes with magnitude greater than m, parameter r describes the relation between small and large earthquake numbers and q is a measure of the seismic activity, or earthquake productivity, of a region [24].

The GR law can also be stated in terms of energy released by earthquakes. In this case the number of earthquakes with energy greater than ε, $N(\varepsilon)$, behaves as a PL, with critical exponent $\tau \in [0.8; 1.05]$ [31,47], while:

$$N(\varepsilon) \sim \varepsilon^{-\tau} \tag{2}$$

The (modified) Omori law describes the short-time rate of decay of aftershock sequences following a main event [53]:

$$N(t) = C_1(t + C_2)^{-\beta} \tag{3}$$

where t represents the time after the main shock, C_1 and C_2 are constants, and β is the rate of decay. The classical Omori law is obtained for $\beta = 1$ and $C_2 = 0$ [37].

Recent studies focus on long-range correlations, motivated by the critical point concept for earthquakes. This viewpoint assumes that the Earth's crust is a CS that produces large non-random earthquakes whenever it is in a critical state. On the contrary, the occurrence of strong events is rare while in a subcritical state [27,45]. The tendency to remain near a critical point corresponds to the behavior of self-organized criticality, in which a system, by itself, converges to an ordered state characterized by the emergence of a coherent global pattern.

Mega et al. [34] use the diffusion entropy to study the temporal distribution of seismicity in California. They found that the distribution of time intervals between main shocks is a PL. The authors propose a long-range model, reproducing the main properties of the diffusion entropy and describing the seismic triggering mechanisms induced by large earthquakes. Nicholson et al. [36] use entropy to measure the degree of clustering within earthquake distributions. They propose a method for calculating the entropy of earthquake foci that uses Voronoi cells to measure point density in three dimensions. Goltz and Böse [15] present an approach to describe the evolution of distributed seismicity by means of configurational entropy. Their findings support the assumption of intermittent criticality in the Earth's crust. De Santis et al. [12] introduced the Shannon entropy of earthquakes and showed how such entropy is related to the GR law parameters. They applied the proposed concepts to earthquakes empirical data.

This paper studies the statistical distributions of worldwide earthquakes over the time period from year 1963 up to 2012. The data are from the Bulletin of the International Seismological Centre (ISC), freely available at http://www.isc.ac.uk/iscbulletin/. First, we prepare the seismic catalog for data analysis. We consider a Cartesian grid to divide Earth in geographic regions, and we analyze earthquake distributions by means of entropy. Second, we use the Jensen–Shannon divergence to compare the empirical data, and we adopt hierarchical clustering algorithms and multi-dimensional scaling (MDS) techniques for data analysis and visualization. Entropy allows relationships and patterns in

the data to be expressed using a single parameter. Classical and generalized (fractional) entropy and Jensen–Shannon divergence are used. The generalized measures lead to a clear identification of patterns embedded in the data and contribute to a better understanding of earthquake distributions.

Bearing these ideas in mind, this paper is organized as follows. Sections 2 and 3 present the experimental dataset and the main mathematical tools used, respectively. Section 4 characterizes the data using entropy. Section 5 compares the events using Jensen–Shannon divergence and different visualization methods. Finally, Sect. 6 outlines the main conclusions.

2 Dataset

In this study we use data from the Bulletin of the International Seismological Centre (ISC), freely available at http://www.isc.ac.uk/iscbulletin/. Each data record contains information about the earthquake magnitude, epicenter location (latitude, longitude and depth), and time (with one-second resolution), among other [21]. The data were retrieved in October, 2014. As the magnitudes of the events have been reported by different entities and are expressed in different scales, the original earthquake catalog is preprocessed in order to homogenize the data [20,42,55]. We adopt the moment magnitude, M_W, as the scale of reference and convert different scales (e.g., local magnitude, M_L, surface-wave magnitude, M_S, body-wave magnitude, M_B, and duration magnitude M_D) to M_W. This is accomplished by taking samples of events with magnitudes expressed in various scales and computing the relations between scales using an orthogonal regression algorithm [10,30]. For example, Fig. 1 illustrates the relation between M_W and M_S scales, for events in the period from year 1963 up to 2012.

After homogenization, the catalog was analyzed for magnitude completeness. This procedure is defined as the lowest magnitude at which all earthquakes in a given space-time volume are detected [50]. Magnitude completeness is estimated by fitting a GR model to the distributions of empirical data and determining the magnitude at which the lower ends of the modeled and experimental distributions diverge [35]. The value $M_C = 4.5$ was found as the threshold that gives a good compromise between inclusion of usable data and rejection of data that could lead to biased analysis. All events of magnitude below M_C are discarded. Figure 2 depicts the frequency–magnitude distribution and the GR model approximation for the time period under study.

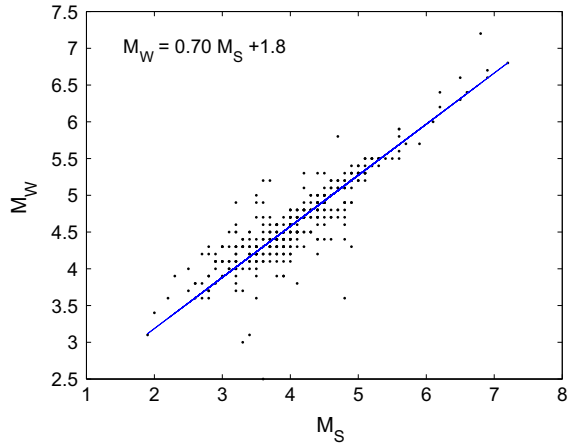


Fig. 1 Relation between M_W and M_S scales, for events in the period from year 1963 up to 2012

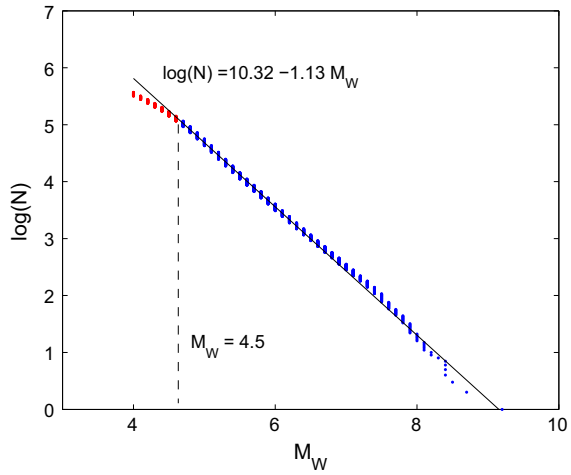


Fig. 2 Frequency–magnitude distribution of earthquakes for the time period 1963–2012 and the GR model approximation

The geographic location of all seismic occurrences with $M_W \geq 4.5$ is depicted in Fig. 3. We can observe that most events are located along the tectonic plate boundaries, mainly in three large zones (http://earthquake.usgs.gov/): (1) the circum-Pacific belt or 'Ring of Fire' (about 90 % of the world's earthquakes occur in this zone); (2) the Alpide belt (the second most seismic region, representing approximately 6 % of all earthquakes); (3) the Mid-Atlantic Ridge (the firth major region).

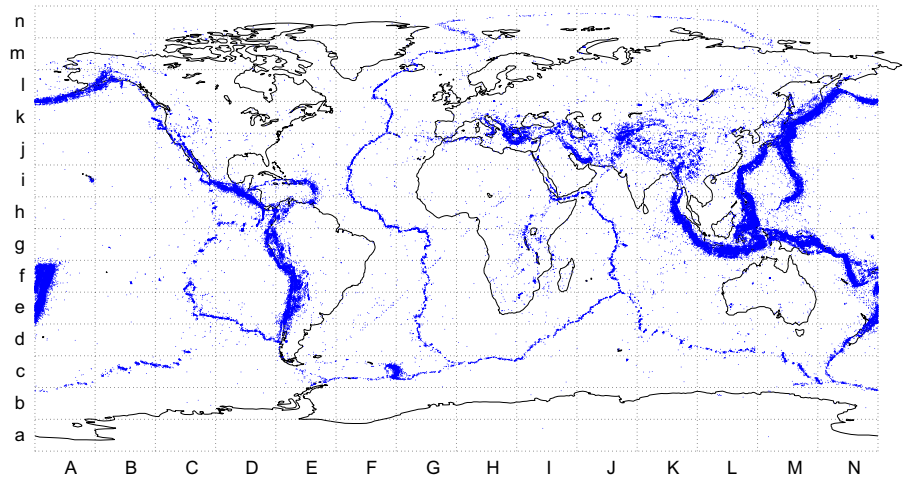


Fig. 3 Geographic location of the occurrences with $M_W \geq 4.5$ for the time period 1963–2012

3 Mathematical tools

This section presents the mathematical tools used to process the data, namely entropy, hierarchical clustering and MDS.

3.1 Entropy

Entropy has been used not only in the context of CS, but also in many other scientific areas [3,29,43]. In information theory, entropy was introduced by Shannon for studying the amount of information contained in a message. Shannon entropy satisfies the so-called Khinchin axioms [28]. It is given by:

$$S = \sum_i (-\ln p_i) p_i \tag{4}$$

and represents the expected value of the information content, $I(p_i) = -\ln p_i$, of an event with probability of occurrence p_i, where $\sum_i p_i = 1$.

Recently, the concepts of information content and entropy of order $\alpha \in \mathbb{R}$ were proposed [32,54], inspired on fractional calculus. These generalizations of the classical definitions are given by:

$$\begin{aligned} I_\alpha(p_i) &= D^\alpha I(p_i) \\ &= -\frac{p_i^{-\alpha}}{\Gamma(\alpha+1)} [\ln p_i + \psi(1) - \psi(1-\alpha)] \end{aligned} \tag{5}$$

$$S_\alpha = \sum_i \left\{ -\frac{p_i^{-\alpha}}{\Gamma(\alpha+1)} [\ln p_i + \psi(1) - \psi(1-\alpha)] \right\} p_i \tag{6}$$

where $D^\alpha(\cdot)$ is the derivative of order α and $\Gamma(\cdot)$ and $\psi(\cdot)$ represent the gamma and digamma functions, respectively.

The generalized entropy (6) does not obey some of the Khinchin axioms except for $\alpha = 0$ [32]. In this case it leads to the classical Shannon entropy (4).

3.2 Hierarchichal clustering

Clustering is a technique for data analysis that has been used in many scientific areas [18]. This technique groups objects that are similar to each other. In hierarchical clustering, two alternative algorithms can generate a hierarchy of clusters: (1) agglomerative and (2) divisive clustering. In (1) each object starts in its own cluster. The algorithm iterates, merging the two most similar clusters, until there is a single cluster containing all objects. In (2) all objects start in one cluster. The algorithm iterates, removing the 'outsiders' from the least cohesive cluster, until each object is in its own singleton cluster. For both algorithms, clustering requires the definition of a linkage criterion, for quantifying the dissimilarity between clusters, which is a function of the pairwise distances between objects. Mathematically, for the clusters R and S, the distance between objects $x_R \in R$ and $x_S \in S$ is given by $d(x_R, x_S)$ [1]. Based on this metric, the maximum, minimum, and average linkages are used, being, respectively:

$$d_{\max}(R, S) = \max_{x_R \in R, x_S \in S} d(x_R, x_S) \tag{7}$$

$$d_{\min}(R, S) = \min_{x_R \in R, x_S \in S} d(x_R, x_S) \tag{8}$$

$$d_{\text{ave}}(R, S) = \frac{1}{\| R \| \| S \|} \sum_{x_R \in R, x_S \in S} d(x_R, x_S) \qquad (9)$$

The results of hierarchical clustering are presented in a dendrogram or a visualization tree.

To assess the clustering, we use the cophenetic correlation (CC) coefficient [44]. The CC gives a measure of how well the cluster tree generated by the linkage function preserves the pairwise distances between the original unmodeled data points. If the clustering is meaningful, the linking of objects in the cluster tree has a strong correlation with the distances between objects in the original dataset. The closer the value of the cophenetic correlation coefficient is to 1, the more accurately the clustering reflects the original data. The results of this analysis are often plotted in a Shepard graph comparing the original and the cophenetic distances. The better the clustering the closer to the 45 degree line the points will lie.

3.3 Multi-dimensional scaling

MDS is a technique for clustering and visualizing data [9]. Given s distinct objects and a measure of similarity, a $s \times s$ symmetric matrix, C, of item to item similarities is calculated and passed to the MDS. The algorithm assigns points to the objects in a multi-dimensional space in order to reproduce the observed similarities. MDS evaluates different configurations for maximizing a goodness-of-fit function, arriving at a configuration that best approximates the observed similarities. A common measure to evaluate how a particular configuration reproduces the observed similarities is the raw stress:

$$\mathcal{S} = \left[d_{ij} - f(\delta_{ij})\right]^2 \qquad (10)$$

where d_{ij} represents the reproduced similarities, given the respective number of dimensions, δ_{ij} corresponds to the observed similarities and $f(\cdot)$ indicates some type of transformation. The smaller the stress, $\mathcal{S}$, the better is the fit between d_{ij} and δ_{ij}.

The MDS interpretation is based on the emerging clusters and distances between points in the map, rather than on their absolute coordinates, or the geometrical form of the locus. Thus, we can rotate or translate the MDS map since the distances between points remain identical.

The quality of the MDS approach can be evaluated by means of the stress and Shepard plots. The stress plot represents $\mathcal{S}$ versus the number of dimensions m of the MDS map. We get a monotonic decreasing chart and we choose m as a compromise between reducing $\mathcal{S}$ and having a low dimension for the MDS map. The Shepard diagram compares the d_{ij} distances, for a particular value m, versus the δ_{ij} distances. Therefore, a narrow scatter around the 45° line indicates a good fit between d_{ij} and δ_{ij}.

4 Entropy analysis of the dataset

In this section we adopt a 14 × 14 rectangular grid, dividing Earth into 196 geographic regions (Fig. 3). For each region we calculate the Shannon, S, and generalized (fractional), S_α, entropies of the distributions of seismic events, where the probabilities are estimated from the histograms of relative frequencies. The results are visualized on geographic maps.

Using a rectangular grid has the advantage of imposing no a priori restrictions regarding the regional seismic activity. Alternative regionalization schemes, for example, along the seismic faults, or based on the Flinn–Engdahl seismic zones, assume that earthquakes are mostly clustered along plate boundaries. Furthermore, the grid resolution adopted establishes a good compromise between having a representative number of samples per cell (i.e., the statistical significance of the sample) and the ability of capturing regional seismicity.

For the generalized entropy, we choose the parameter $\alpha = 0.6$, for which the entropy of all regions, S_α, is maximum. Figure 4 illustrates the evolution of S_α

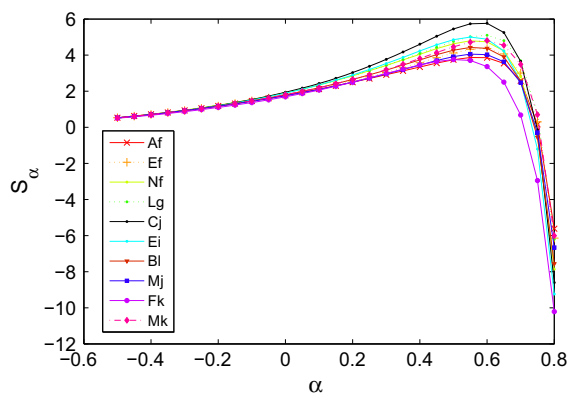


Fig. 4 Variation of S_α versus α for the regions {Af, Ef, Nf, Lg, Cj, Ei, Bl, Mj, Fk, Mk}. The time period of analysis is 1963–2012

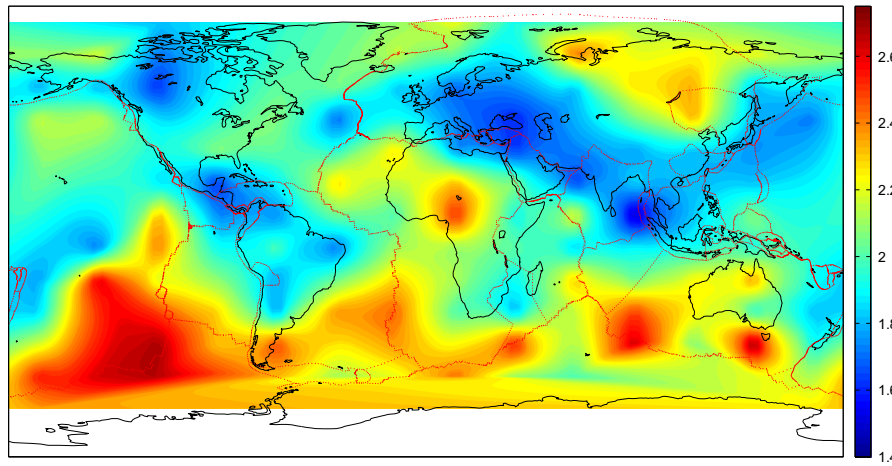


Fig. 5 Geographic map of S for the time period 1963–2012

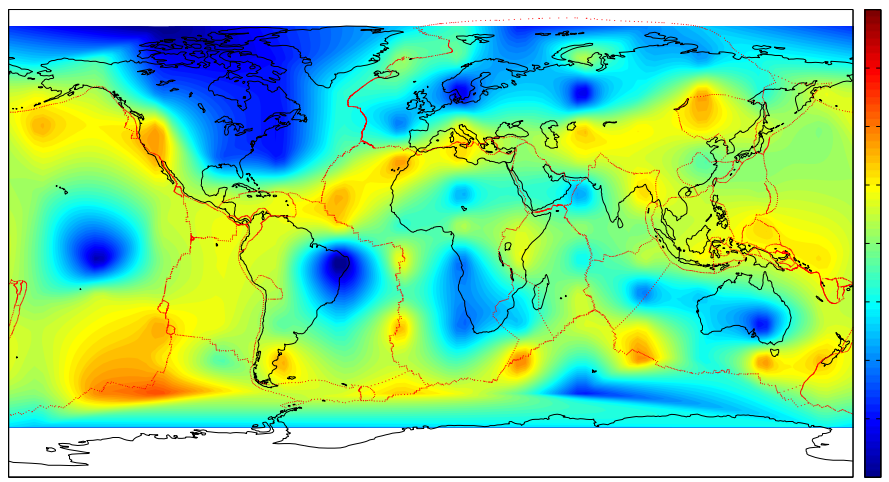


Fig. 6 Geographic map of $S_{0.6}$ for the time period 1963–2012

versus α for a subset of regions (to facilitate the visualization), namely {Af, Ef, Nf, Lg, Cj, Ei, Bl, Mj, Fk, Mk}. As can be seen S_α has a maximum for $\alpha \approx 0.6$.

Figures 5 and 6 represent the geographic maps of S and $S_{0.6}$ for the time period 1963–2012, respectively.

As can be seen, the map of S reveals an unbalanced geographic distribution of entropy between northern and southern hemispheres. Furthermore, most regions located in Eurasia, Far East Asia and Indian ocean, mid-north-Atlantic and Central America and South America seismic faults exhibit unjustified smaller entropy [51]. On the contrary, $S_{0.6}$ has the advantage of leading to a good distribution of entropy along the main fault zones, being more representative/informative of the seismicity distribution.

5 Jensen–Shannon comparison and visualization

We adopt the Jensen–Shannon divergence to compare the data in an inter-regional basis. For that purpose, in Sects. 5.1 and 5.2, we use hierarchical clustering and MDS algorithms for data analysis and visualization, respectively.

The Jensen–Shannon divergence, JSD, measures the 'distance' between two probability distributions, P and Q [8]. The JSD can be viewed as a symmetrical and smoothed version of the Kullback–Leibler divergence, KLD, given by:

$$\mathrm{KLD}\,(P \parallel Q) = \sum_i p_i \ln \frac{p_i}{q_i} \tag{11}$$

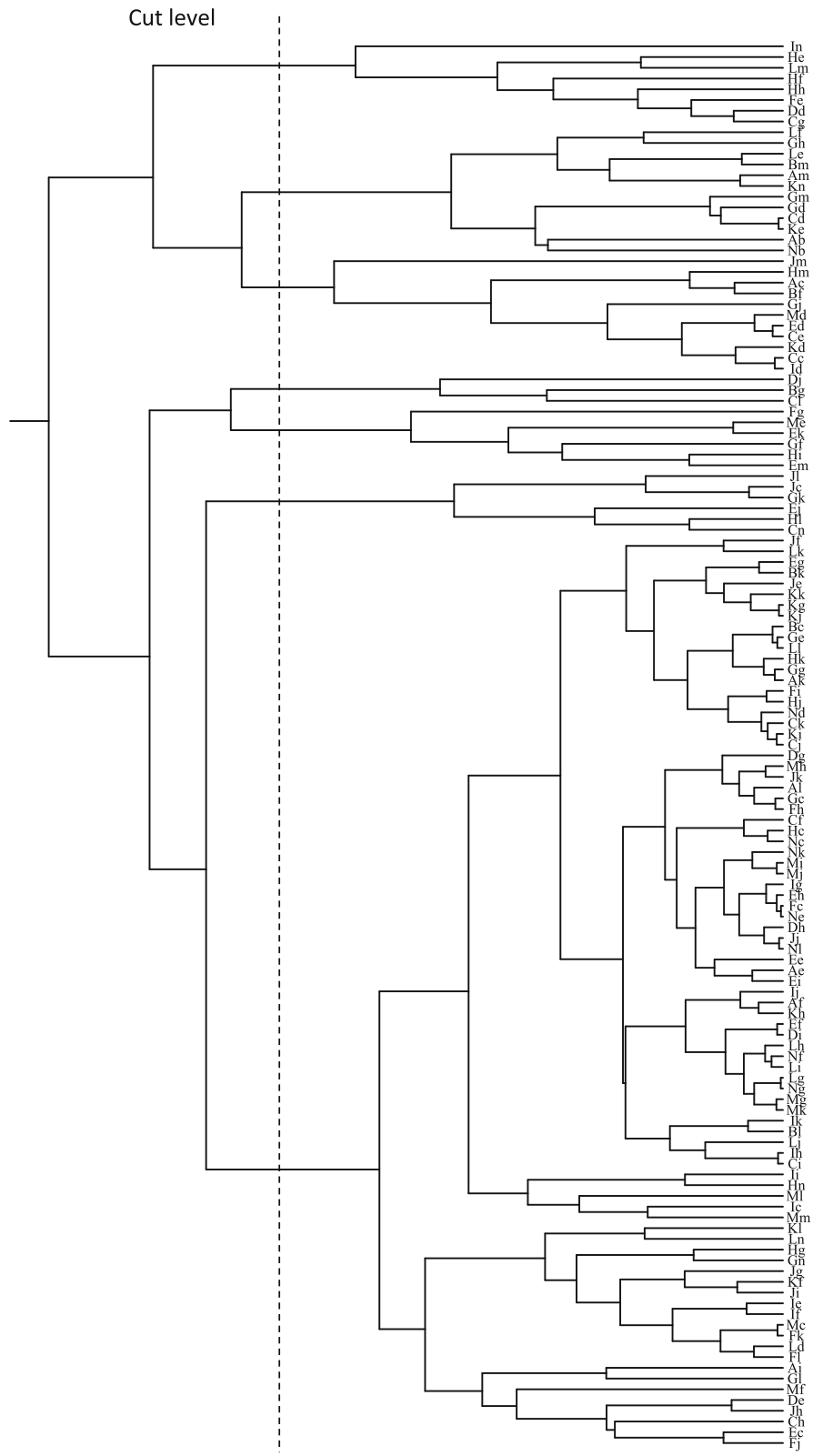


Fig. 7 Dendrogram representing similarities between probability distributions of regions (R_i, R_j), $i = \{A, \ldots, N\}$, $j = \{a, \ldots, n\}$, based on fractional Jensen–Shannon divergence, JSD_α

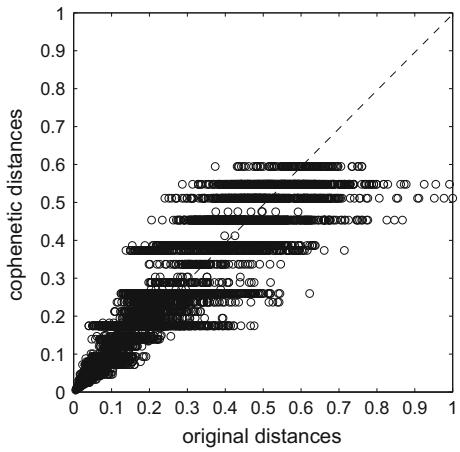


Fig. 8 Shepard plot for the hierarchical clustering corresponding to JSD_α

Resulting in:

$$\mathrm{JSD}\,(P \parallel Q) = \frac{1}{2}\left[\mathrm{KLD}\,(P \parallel M) + \mathrm{KLD}\,(Q \parallel M)\right] \tag{12}$$

where $M = \frac{P+Q}{2}$.

Alternatively, we can write $\mathrm{JSD}\,(P \parallel Q)$ as:

$$\mathrm{JSD}\,(P \parallel Q) = \frac{1}{2}\left[\sum_i p_i \ln p_i + \sum_i q_i \ln q_i\right] - \sum_i m_i \ln m_i \tag{13}$$

which, from Eqs. (5) and (13), leads to the generalized (fractional) Jensen–Shannon divergence:

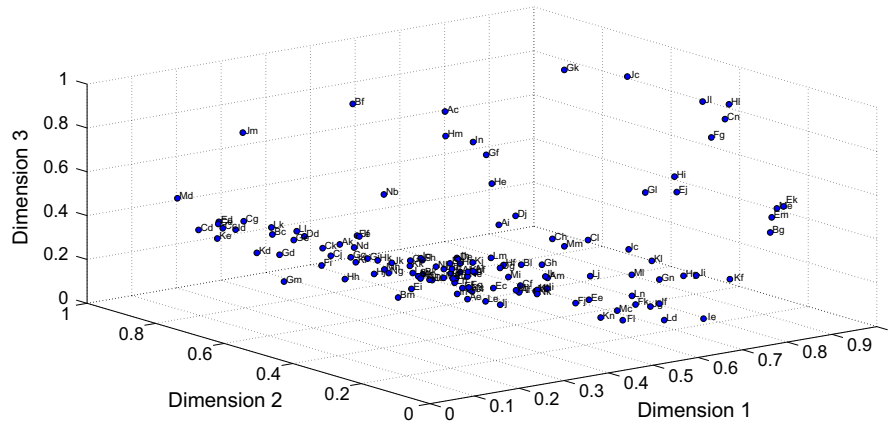


Fig. 9 MDS three-dimensional map representing similarities between probability distributions of regions (R_i, R_j), $i = \{A, \ldots, N\}$, $j = \{a, \ldots, n\}$, based on fractional Jensen–Shannon divergence, JSD_α

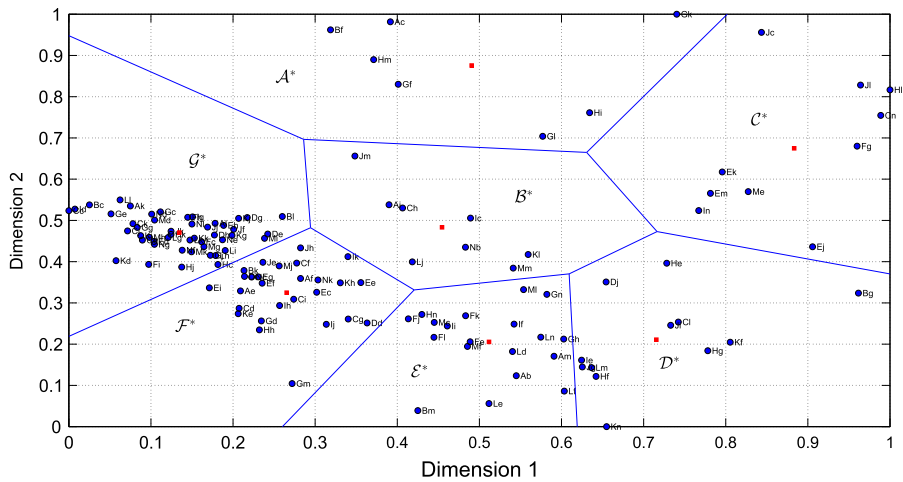


Fig. 10 MDS two-dimensional map and Voronoi cells, corresponding to $c = 7$ clusters and fractional Jensen–Shannon divergence, JSD_α

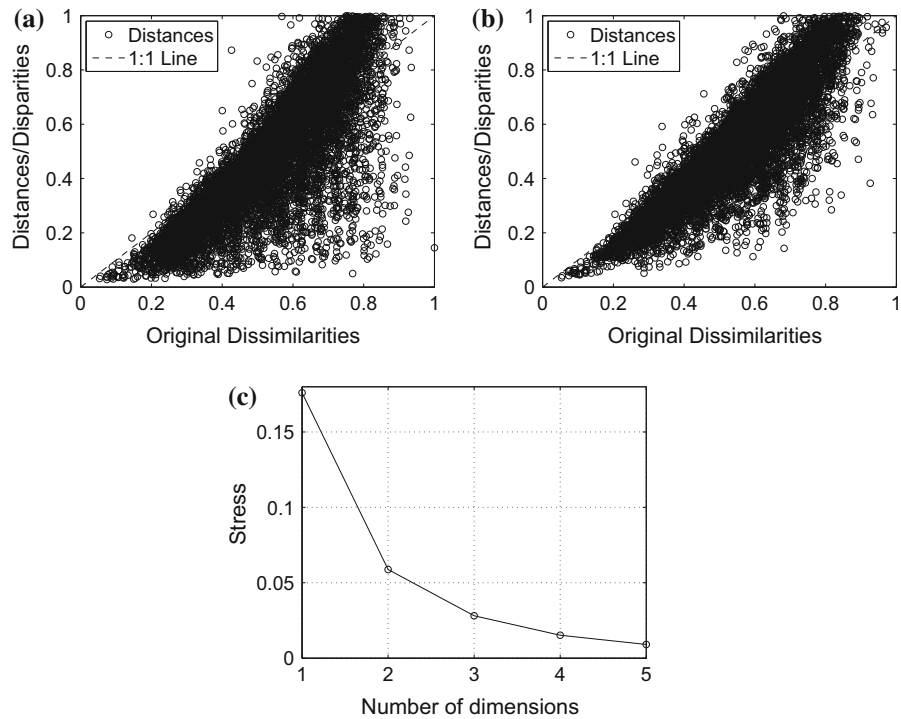


Fig. 11 Shepard and stress plots corresponding to the MDS maps: **a** Shepard, two-dimensional map; **b** Shepard, three-dimensional map; **c** stress

$$
\begin{aligned}
&\mathrm{JSD}_{\alpha}\left(P \parallel Q\right) \\
&= \frac{1}{2}\sum_{i} p_i \left\{ \frac{p_i^{-\alpha}}{\Gamma\left(\alpha+1\right)} \left[\ln p_i + \psi\left(1\right) - \psi\left(1-\alpha\right)\right] \right\} \\
&\quad + \frac{1}{2}\sum_{i} q_i \left\{ \frac{q_i^{-\alpha}}{\Gamma\left(\alpha+1\right)} \left[\ln q_i + \psi\left(1\right) - \psi\left(1-\alpha\right)\right] \right\} \\
&\quad - \sum_{i} m_i \left\{ \frac{m_i^{-\alpha}}{\Gamma\left(\alpha+1\right)} \left[\ln m_i + \psi\left(1\right) - \psi\left(1-\alpha\right)\right] \right\}
\end{aligned}
\tag{14}
$$

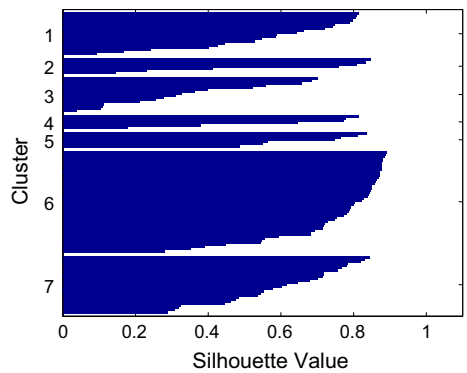


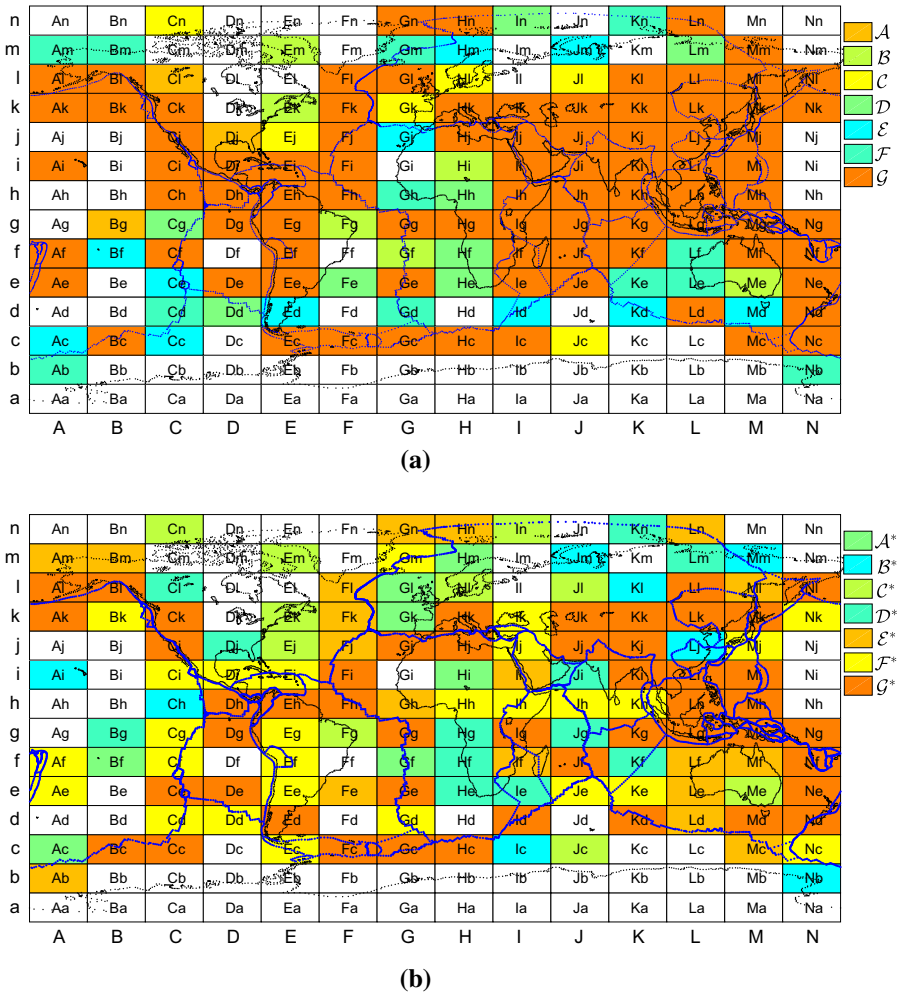
Fig. 12 Silhouette diagram for assessing the K-means clustering

5.1 Hierarchical clustering analysis and visualization

We calculate the 196×196 matrix $C_{\alpha} = [\mathrm{JSD}_{\alpha}^{ij}]$, $i = \{A, \ldots, N\}$, $j = \{a, \ldots, n\}$. The distance $\mathrm{JSD}_{\alpha}^{ij}$ ($\alpha = 0.6$) represents the fractional Jensen–Shannon divergence between the probability distributions (P, Q) of regions (R_i, R_j). The probabilities are estimated by histograms of relative frequencies with $N = 20$ bins.

A hierarchical clustering algorithm based on the successive (agglomerative) clustering and average-linkage method for data comparison is adopted.

Figure 7 depicts the resulting dendrogram. We can observe that the number of clusters c depends on the level adopted for 'cutting' the tree. We adopt $c = 7$ clusters as a compromise between good discrimination of differences and good identification of patterns embedded in the data. We obtain the sets $\mathcal{A}$ ={Bg, Dj, Cl}, $\mathcal{B}$ ={Me, Gf, Fg, Hi, Ek, Em}, $\mathcal{C}$ ={Jc, Ej, Gk, Hl, Jl, Cn}, $\mathcal{D}$ ={Dd, Fe, He, Hf, Cg, Hh, Lm, In}, $\mathcal{E}$ ={Ac, Cc, Ed, Id, Kd, Md, Ce, Bf, Gj, Hm, Jm}, $\mathcal{F}$ ={Ab, Nb, Cd, Gd, Ke, Le, Lf, Gh, Am, Bm, Gm, Kn}, as well as a larger and denser cluster, $\mathcal{G}$, that includes all the remaining regions.



(a)

(b)

Fig. 13 Geographical representation of the clusters obtained for the fractional Jensen–Shannon divergence, JSD_α: **a** hierarchical clustering; **b** MDS plus K-means

Figure 8 represents the Shepard plot for assessing the hierarchical clustering. The chart reflects an accurate clustering of the original data, yielding a cophenetic correlation coefficient of 0.87.

5.2 MDS analysis and visualization

For an alternative and powerful clustering and visualization, we adopt MDS considering the same matrix C_α. Furthermore, we compare the advantage of using two- and three-dimensional MDS maps. For example, Fig. 9 depicts the three-dimensional map, where the points represent regions.

Clusters can be determined by the user, based on the observation of the relative distances between points in the charts, or using an 'automatic' clustering method [33]. In this work we use the classical K-means algorithm [19,23], based on the euclidean distance, with the two-dimensional MDS map. In fact, 2D is slight less accurate then the 3D counterpart, but it is much easier to visualize.

Figure 10 depicts the chart, as well as the Voronoi cells, corresponding to $c = 7$ clusters. We see that the regions are included in four smaller sets, namely $\mathcal{A}^* = \{\text{Ac, Bf, Gf, Hi, Gk, Gl, Hm}\}$, $\mathcal{B}^* = \{\text{Nb, Ic, Ch, Ai, Lj, Kl, Jm, Mm}\}$, $\mathcal{C}^* = \{\text{Jc, Me, Fg, Ej, Ek, Hl, Jl, Em, Cn, In}\}$ and $\mathcal{D}^* = \{\text{He, Ie, Hf, Kf, Bg, Hg, Jg, Ji, Dj, Cl, Lm, Kn}\}$, together with three large clusters, $\mathcal{E}^*$, $\mathcal{F}^*$ and $\mathcal{G}^*$.

Figure 11 represents the Shepard charts for the two- and three-dimensional maps, used to assess the quality of the MDS results. The points around the 45° line reveal a good correspondence between the original and

the reproduced distances. The curve elbow in the stress chart means that both the two- and three-dimensional maps are a good compromise between facilities of representation and good visualization. As expected, the three-dimensional map yields a slight better results at the expense of a more complex visualization procedure.

Figure 12 shows the silhouette diagram that assesses the K-means results. The silhouette value, S, for each object, is a measure of how well each object lies within its cluster [40]. Silhouette values close to $+1$ correspond to objects that are very distant from neighboring clusters, and therefore, they are assigned to the correct cluster. For values close to 0 it means that there is doubt and that the objects could have been assigned to another cluster. When the silhouette values tend to -1 the objects are assigned to the wrong cluster.

Dendrograms and MDS maps are alternative techniques for data analysis and visualization. We verify, that both representations yield good representation of the correlations between the data. However, the MDS maps are more intuitive, specially when dealing with large number of objects. The MDS three-dimensional maps are sightly more precise, but need to be rotated for allowing users to perceive all the details embedded in the graphs.

Figure 13 depicts the geographical location of the clusters identified by hierarchical clustering and the MDS plus K-means. Regions belonging to the same cluster are represented with the same color. Regions with insufficient data are kept in blank. These maps complement Figs. 7 and 10 by embedding geographical information. We can see that both methods capture well the geographic distribution of earthquakes along the boundaries of the tectonic plates. Furthermore, the MDS and K-means yield better identification of the emerging patterns.

6 Conclusions

We studied the statistical distributions of worldwide earthquakes over the time period from year 1963 up to 2012. First, we adopted a Cartesian grid to divide Earth in 14×14 geographic regions, and we analyzed earthquake probability distributions by means of entropy. Second, we used the Jensen–Shannon divergence to compare the empirical data. We adopted hierarchical clustering algorithms and MDS techniques for data analysis and visualization. Entropy allows using a single parameter to represent relationships between data. The results reveal patterns in the data and show that the adopted methodology can contribute to a comprehensive explanation of the phenomena.

Acknowledgments The authors acknowledge the International Seismological Centre (ISC) for the data (http://www.isc.ac.uk/).

Compliance with Ethical Standards

Conflict of interest The authors declare that they have no conflict of interest.

References

1. Aggarwal, C.C., Hinneburg, A., Keim, D.A.: On the Surprising Behavior of Distance Metrics in High Dimensional Space. Springer, New York (2001)
2. Ashtari Jafari, M.: Statistical prediction of the next great earthquake around Tehran, Iran. J. Geodyn. **49**(1), 14–18 (2010)
3. Balasis, G., Daglis, I.A., Papadimitriou, C., Anastasiadis, A., Sandberg, I., Eftaxias, K.: Quantifying dynamical complexity of magnetic storms and solar flares via nonextensive Tsallis entropy. Entropy **13**(10), 1865–1881 (2011)
4. Baleanu, D., Diethelm, K., Scalas, E., Trujillo, J.J.: Fractional Calculus. World Scientific, Singapore (2012)
5. Bhattacharya, P., Chakrabarti, B.K., et al.: A fractal model of earthquake occurrence: theory, simulations and comparisons with the aftershock data. In: Journal of Physics: conference series, vol. 319, pp. 012004. IOP Publishing (2011)
6. Carlson, J.M., Langer, J.S., Shaw, B.E.: Dynamics of earthquake faults. Rev. Mod. Phys. **66**(2), 657 (1994)
7. Clauset, A., Young, M., Gleditsch, K.S.: On the frequency of severe terrorist events. J. Confl. Resolut. **51**(1), 58–87 (2007)
8. Cover, T.M., Thomas, J.A.: Elements of Information Theory. Wiley, New York (2012)
9. Cox, T.F., Cox, M.A.: Multidimensional Scaling. CRC Press, Boca Raton (2000)
10. Das, R., Wason, H., Sharma, M.: Magnitude conversion to unified moment magnitude using orthogonal regression relation. J. Asian Earth Sci. **50**, 44–51 (2012)
11. De Rubeis, V., Hallgass, R., Loreto, V., Paladin, G., Pietronero, L., Tosi, P.: Self-affine asperity model for earthquakes. Phys. Rev. Lett. **76**(14), 2599 (1996)
12. De Santis, A., Cianchini, G., Favali, P., Beranzoli, L., Boschi, E.: The Gutenberg–Richter law and entropy of earthquakes: two case studies in Central Italy. Bull. Seismol. Soc. Am. **101**(3), 1386–1395 (2011)
13. Ekström, G., Nettles, M., Abers, G.A.: Glacial earthquakes. Science **302**(5645), 622–624 (2003)
14. Geller, R.J.: Earthquake prediction: a critical review. Geophys. J. Int. **131**(3), 425–450 (1997)
15. Goltz, C., Böse, M.: Configurational entropy of critical earthquake populations. Geophys. Res. Lett. **29**(20), 51 (2002)
16. Gutenberg, B., Richter, C.F.: Frequency of earthquakes in California. Bull. Seismol. Soc. Am. **34**(4), 185–188 (1944)

17. Hallgass, R., Loreto, V., Mazzella, O., Paladin, G., Pietronero, L.: Earthquake statistics and fractal faults. Phys. Rev. E **56**(2), 1346 (1997)
18. Hartigan, J.A.: Clustering Algorithms. Wiley, New York (1975)
19. Hartigan, J.A., Wong, M.A.: Algorithm as 136: A K-means clustering algorithm. Appl. Stat. **28**(1), 100–108 (1979)
20. Hussein, H., Elenean, K.A., Marzouk, I., Peresan, A., Korrat, I., El-Nader, E.A., Panza, G., El-Gabry, M.: Integration and magnitude homogenization of the egyptian earthquake catalogue. Nat. Hazards **47**(3), 525–546 (2008)
21. International Seismological Centre: On-line Bulletin. International Seismological Centre, Thatcham, United Kingdom (2012). http://www.isc.ac.uk
22. Ionescu, C.M.: The Human Respiratory System; An Analysis of the Interplay between Anatomy, Structure, Breathing and Fractal Dynamics. Springer, New York (2013)
23. Jain, A.K.: Data clustering: 50 years beyond k-means. Pattern Recognit. Lett. **31**(8), 651–666 (2010)
24. Kagan, Y.Y.: Seismic moment distribution revisited: I. Statistical results. Geophys. J. Int. **148**(3), 520–541 (2002)
25. Kagan, Y.Y., Jackson, D.D.: Probabilistic forecasting of earthquakes. Geophys. J. Int. **143**(2), 438–453 (2000)
26. Kanamori, H., Brodsky, E.E.: The physics of earthquakes. Rep. Prog. Phys. **67**(8), 1429 (2004)
27. Keilis-Borok, V.: The lithosphere of the earth as a nonlinear system with implications for earthquake prediction. Rev. Geophys. **28**(1), 19–34 (1990)
28. Khinchin, A.I.: Mathematical Foundations of Information Theory, vol. 434. Courier Dover Publications, New York (1957)
29. Levada, A.: Learning from complex systems: on the roles of entropy and Fisher information in pairwise isotropic gaussian Markov random fields. Entropy **16**(2), 1002–1036 (2014)
30. Lopes, A.M., Tenreiro Machado, J.A., Pinto, C., Galhano, A.: Multidimensional scaling visualization of earthquake phenomena. J. Seismol. **18**(1), 163–179 (2014)
31. Lopes, A.M., Tenreiro Machado, J.A., Pinto, C., Galhano, A.: Fractional dynamics and MDS visualization of earthquake phenomena. Comput. Math. Appl. **66**(5), 647–658 (2013)
32. Machado, J.T.: Fractional order generalized information. Entropy **16**(4), 2350–2361 (2014)
33. Martínez-Torres, M.R., García, F.B., Marín, S.T., Vázquez, S.G.: A digital signal processing teaching methodology using concept-mapping techniques. IEEE Trans. Educ. **48**(3), 422–429 (2005)
34. Mega, M.S., Allegrini, P., Grigolini, P., Latora, V., Palatella, L., Rapisarda, A., Vinciguerra, S.: Power-law time distribution of large earthquakes. Phys. Rev. Lett. **90**(18), 188,501 (2003)
35. Mignan, A., Woessner, J.: Estimating the magnitude of completeness for earthquake catalogs. Community Online Resour. Stat. Seism. Anal. (2012). doi:10.5078/corssa-00180805
36. Nicholson, T., Sambridge, M., Gudmundsson, O.: On entropy and clustering in earthquake hypocentre distributions. Geophys. J. Int. **142**(1), 37–51 (2000)
37. Ōmori, F.: On the After-Shocks of Earthquakes, vol. 7. The University, Tokyo (1894)
38. Peng, Z., Gomberg, J.: An integrated perspective of the continuum between earthquakes and slow-slip phenomena. Nature Geosci **3**(9), 599–607 (2010)
39. Pinto, C., Mendes Lopes, A., Tenreiro Machado, J.A.: A review of power laws in real life phenomena. Commun. Nonlinear Sci. Numer. Simul. **17**(9), 3558–3578 (2012)
40. Rousseeuw, P.J.: Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. J. Comput. Appl. Math. **20**, 53–65 (1987)
41. Sarlis, N., Christopoulos, S.R.: Natural time analysis of the centennial earthquake catalog. Chaos Interdiscip. J. Nonlinear Sci. **22**(2), 023,123 (2012)
42. Scordilis, E.: Empirical global relations converting m_s and m_b to moment magnitude. J. Seismol. **10**(2), 225–236 (2006)
43. Seely, A.J., Newman, K.D., Herry, C.L.: Fractal structure and entropy production within the central nervous system. Entropy **16**(8), 4497–4520 (2014)
44. Sokal, R.R., Rohlf, F.J.: The comparison of dendrograms by objective methods. Taxon. **11**(2), 33–40 (1962)
45. Sornette, A., Sornette, D.: Earthquake rupture as a critical point: Consequences for telluric precursors. Tectonophysics **179**(3), 327–334 (1990)
46. Sornette, D.: Earthquakes: from chemical alteration to mechanical rupture. Phys. Rep. **313**(5), 237–292 (1999)
47. Sotolongo-Costa, O., Posadas, A.: Fragment-asperity interaction model for earthquakes. Physical Rev. Lett. **92**(4), 048,501 (2004)
48. Stadler, G., Gurnis, M., Burstedde, C., Wilcox, L.C., Alisic, L., Ghattas, O.: The dynamics of plate tectonics and mantle flow: from local to global scales. Science **329**(5995), 1033–1038 (2010)
49. Stein, S., Liu, M., Calais, E., Li, Q.: Mid-continent earthquakes as a complex system. Seismol. Res. Lett. **80**(4), 551–553 (2009)
50. Stucchi, M., Albini, P., Mirto, M., Rebez, A.: Assessing the completeness of italian historical earthquake data. Ann. Geophys. **47**(2–3), 659–673 (2004)
51. Tenreiro Machado, J.A., Lopes, A.M.: Analysis and visualization of seismic data using mutual information. Entropy **15**(9), 3892–3909 (2013)
52. Turcotte, D.L., Malamud, B.D.: Earthquakes as a complex system. Int. Geophys. **81**, 209-IV (2002)
53. Utsu, T., Ogata, Y., Matsu'ura, R.S.: The centenary of the Omori formula for a decay law of aftershock activity. J. Phys. Earth **43**(1), 1–33 (1995)
54. Valério, D., Trujillo, J.J., Rivero, M., Tenreiro Machado, J.A., Baleanu, D.: Fractional calculus: a survey of useful formulas. Eur. Phys. J. Spec. Top. **222**, 1827–1846 (2013)
55. Wiemer, S.: A software package to analyze seismicity: Zmap. Seismol. Res. Lett. **72**(3), 373–382 (2001)