

# Multidimensional scaling visualization of earthquake phenomena

António M. Lopes · J. A. Tenreiro Machado ·  
C. M. A. Pinto · A. M. S. F. Galhano

**Abstract** Earthquakes are associated with negative events, such as large number of casualties, destruction of buildings and infrastructures, or emergence of tsunamis. In this paper, we apply the Multidimensional Scaling (MDS) analysis to earthquake data. MDS is a set of techniques that produce spatial or geometric representations of complex objects, such that, objects perceived to be similar/distinct in some sense are placed nearby/distant on the MDS maps. The interpretation of the charts is based on the resulting clusters since MDS produces a different locus for each similarity measure.

In this study, over three million seismic occurrences, covering the period from January 1, 1904 up to March 14, 2012 are analyzed. The events, characterized by their magnitude and spatiotemporal distributions, are divided into groups, either according to the Flinn–Engdahl seismic regions of Earth or using a rectangular grid based in latitude and longitude coordinates. Space-time and Space-frequency correlation indices are proposed to quantify the similarities among events. MDS has the advantage of avoiding sensitivity to the non-uniform spatial distribution of seismic data, resulting from poorly instrumented areas, and is well suited for accessing dynamics of complex systems. MDS maps are proven as an intuitive and useful visual representation of the complex relationships that are present among seismic events, which may not be perceived on traditional geographic maps. Therefore, MDS constitutes a valid alternative to classic visualization tools, for understanding the global behavior of earthquakes.

**Keywords** Multidimensional scaling · Seismic events · Correlation indices · Visualization

## 1 Introduction

Big earthquakes are characterized by a huge severity, usually defined by a large number of casualties, destruction of buildings, and high probability of tsunamis. Predicting the next big earthquake would be of extreme importance in developing pre-disaster strategies.

However, it should be highlighted that reliable short-term prediction of earthquakes is not possible at present. Economic issues are also of great significance, since big earthquakes translate in millions of dollars of expenses, to rebuild, assemble emergency teams, and gather medical and food supplies, amongst others.

Statistical processing of real data and clustering analysis have been used to recognize precursory events for earthquake prediction (Geller 1997; Kagan and Jackson 2000; Jafari 2010). In this paper, we apply multidimensional scaling (MDS) to earthquake data. MDS is a statistical and visualization technique used to tackle high dimensional data, by reducing it into a low-dimensional map, where objects are interpreted by their clustering rather than by the geometric shape of the chart itself. MDS has relevant properties for the analysis of complex phenomena since it does not require a priori assumptions on the system nature. Furthermore, since MDS maps do not need data in a uniform grid in time and space, it may constitute a valuable tool for studying the non-homogeneous data registered in global earthquake catalogues.

Earthquakes are due to the movement of the Earth's crust. Two empirical laws, the Gutenberg–Richter (GR) and the Omori laws, are valuable tools for a better understanding of the earthquake cycle. The GR law is given by:

$$\log N = a - bM \quad (1)$$

where  $N \in \mathbb{N}$  is the number of earthquakes with amplitude greater or equal than  $M$  and  $a, b \in \mathbb{R}^+$  are parameters. Namely,  $a$  is the so-called productivity and  $b$  is a tectonic parameter, that is typically in the interval of  $0.8 \leq b \leq 1.1$ . The GR distribution gives the estimated number of earthquakes with a magnitude greater or equal than  $M$ , which may occur in a given region during a given period. This distribution is a power law (Gutenberg and Richter 1944). The (modified) Omori's law describes the rate of decay of an aftershock sequence, following a main event, and is formulated as:

$$N(t) = C_1(t + C_2)^{-\alpha} \quad (2)$$

where  $t$  is the time after the main shock,  $C_1, C_2 \in \mathbb{R}_0^+$  are constants, and  $\alpha > 0$  is the rate of decay. For  $\alpha = 1$  and  $C_2 = 0$ , we obtain the classical Omori's law (Omori 1894). It is shown that  $\alpha$  values for aftershock sequences in southern California vary between 0.5 and 1.5 (Reasenber and Jones 1989), with a mean value

close to 1 (Utsu et al. 1995). The variability of  $\alpha$  may be due to the tectonic condition of the region. Stress, structural heterogeneity, and temperature, are among those conditions; nevertheless, a significant factor controlling the  $\alpha$  value is still to be found (Utsu et al. 1995). The parameter  $C_1$  is a measure of the aftershock productivity and is a function of the magnitude of the main event. The physical meaning of  $C_2$  is involved in controversy. Some advocate that it relates to the physics of aftershock generation right after the main event, whereas others say it models the incomplete detection of aftershocks at short times after the main earthquake. Lindman et al. (2006) use a pore pressure diffusion model and conclude that  $C_2$  is dependant, at short times after the main earthquake, of the reducing process of high pore pressure gradients existing across a fault zone.

Aftershocks have lower frequencies and magnitudes than those of the main earthquake. The stronger the earthquake, the larger and long lasting are the aftershocks. Moreover, the probability of small aftershocks is higher than for greater amplitude aftershocks. One example of a big earthquake is the Indonesia earthquake, in 2004, with magnitude 9.1. One of its aftershocks had magnitude 8.7 and was felt in March 2005. An anomalous decrease in aftershock activity below the level predicted by the modified Omori formula sometimes precedes a large aftershock (Utsu et al. 1995).

The epicenters of the earthquakes occur on a fractal-like distribution of faults. (Kagan 1992) computed rotations of focal mechanisms and observed it followed a Cauchy distribution. Cauchy distribution has a power-law tail, thus, its control of stress implies that fractal geometry is behind earthquake rupture.

The goal of using MDS is to add a different, but helpful, description of the distribution of earthquakes embedding space, time, and magnitude. The seismic events are divided into groups, either according to the Flinn–Engdahl (F-E) seismic regions of Earth, or using a rectangular grid based in latitude and longitude coordinates. Space-time and space-frequency correlation indices are proposed to quantify the similarities among events. The indices used for MDS are not unique and are essentially determined by the user experience and intuition. Moreover, MDS has the advantage of avoiding sensitivity to the non-uniform spatial distribution of seismic events that result from poorly instrumented areas.

MDS tools have been widely used in finance, social sciences, biology, and other disciplines (Preoteasa et al.

2013), but studies in seismology remain scarce. Previous works may be found in Dzwiniel et al. (2005). There, the authors used both multi-resolutional clustering and nonlinear MDS of earthquake patterns to analyze synthetic and observed earthquake data. Observed seismic activity was related to Japanese islands between 1997 and 2003. They found that the combination of clustering results, in low- and high-resolution spaces, helped the recognition of precursory events more precisely than at a single resolution.

Bearing these ideas in mind, this paper is organized as follows. In Section 2, we give a brief review of the MDS techniques. Section 3 analyzes earthquakes using MDS tools and discusses results, adopting F-E seismic regions. Section 4 extends the analysis to an alternative seismic regionalization of Earth. Finally, in Section 5, we infer the main conclusions of this paper.

## 2 Multidimensional scaling

MDS is a set of techniques that produce spatial or geometric representations of complex objects. It helps understand people's judgments (preference, relatedness), concerning elements in sets of objects (Torgerson 1958). Given  $s$  objects in an  $m$ -dimensional space and an  $s \times s$  matrix of proximity measures among the objects, MDS produces a  $q \leq m$  dimensional configuration of points,  $X$  representing the items. This configuration preserves the proximities of the higher dimensional space, and eases the understanding of data underlying structure. MDS is, consequently, different from other similar techniques, such as factor analysis and cluster analysis because there are no assumptions concerning which factors might drive each dimension. Additionally, MDS algorithms have better convergence rates and are less complex than other algorithms (Bronstein et al. 2006). MDS treats every type of data, negative, non-negative, correlations, amongst others (Borg and Groenen 2005).

The proximity measures the (dis)similarities among the items and is, usually, a distance measure. Smaller (larger) distances between two objects translate into more (less) similarities between them. Let the dissimilarity for objects  $i$  and  $j$  be given by  $\delta_{ij}$ . The coordinates are gathered in an  $s \times q$  matrix  $X$ , where  $q$  is specified in advance by the user. Thus, row  $i$  from  $X$  gives the coordinates of object  $i$ . For example, the Minkowski distance metric provides a general way to

specify distance for quantitative data in a multidimensional space:

$$d_{ij} = \left( \sum_{k=1}^q w_k |x_{ik} - x_{jk}|^r \right)^{1/r} \quad (3)$$

where  $x_{ik}$  is the value of dimension  $k$  for object  $i$  and  $w_k$  is a weight factor. Setting  $w_k=1$  the Euclidean and the city-block distances are obtained for  $r=2$  and  $r=1$ , respectively.

Nevertheless, the MDS technique allows users to choose any metric for comparing objects, which leads to a rich pool of possibilities.

The main purpose of MDS is to find a matrix  $X$  such that  $d_{ij}$  approximates  $\delta_{ij}$  as closely as possible. Mathematically, this problem is equivalent to minimize the raw Stress function,  $\sigma^2$ , given by (Kruskal 1964):

$$\sigma^2 = \sum_{i=2}^s \sum_{j=1}^{i-1} z_{ij} (\delta_{ij} - d_{ij})^2 \quad (4)$$

where  $z_{ij}$  is a user nonnegative chosen weight. Usually, a value of  $z_{ij}=0$  means that dissimilarities are absent. MDS uses numerical algorithms to find matrices  $X$  for which  $\sigma^2$  is a minimum. There are other stress measures, such as the normalized raw Stress, that is  $\sigma^2$  divided by the sum of squared dissimilarities.

Other measures are Kruskal's stress-1 and stress-2, which divide  $\sigma$  by the sum of squared distances or by a function of the variances of distances, respectively. Another example is the S-stress measure given by the sum of squared errors between squared distances and squared dissimilarities.

The Shepard diagram can be used to infer the quality of the MDS solution. Let  $p_{ij}$  denote the similarities between objects  $i$  and  $j$ . A Shepard diagram consists of pairs  $(p_{ij}, d_{ij})$ ,  $(p_{ij}, \delta_{ij})$ . A line is drawn connecting the pairs  $(p_{ij}, \delta_{ij})$ . The approximation error, concerning dissimilarities of each object, is given by  $d_{ij} - \delta_{ij}$ . The Shepard diagram is thus useful to visualize the residuals and outliers resulting from the MDS application to the data.

MDS can be divided according to the classification of data similarities, number of similarity matrices, and the nature of the MDS model. We thus have non metric MDS, if similarity data are qualitative, and a metric MDS for quantitative similarities. In what concerns the number of similarity matrices and nature of the model, there are several MDS types. Classical MDS has one matrix and unweighted models, replicated MDS deals with several

matrices and unweighted models, and finally, weighted MDS, with several matrices and weighted models.

Usually, MDS geometric configurations are in two or three dimensions, nevertheless, any dimension  $q$ , with  $q < m$  is possible.

MDS has been used in a large variety of real data, from finances, marketing, sociology, physics, political science, biology, and biomedics (Ramsay 1980; Matheus et al. 2006; Tzagarakis et al. 2009; Tenreiro Machado et al. 2011a, b, c; Ionescu et al. 2011). Other relevant research areas, such as the one in wireless network sensors, are also becoming interested in the MDS analysis (Ji and Zha 2004; Chan and So 2009).

To conclude this brief review in MDS techniques, we emphasize that some caution must be exerted when analyzing MDS maps. One must take into account that the axes of the maps are meaningless, the orientation of the geometrical configuration is arbitrary, and the substantive dimensions or attributes under analysis do not need to correspond in number or direction, to the mathematical dimensions (axes) which define the vector space. The latter is explained by the fact that mathematical and human dimensions are different. Mathematical dimensions are orthogonal, by definition, whereas human dimensions may have a highly degree of correlation.

### 3 Multidimensional scaling analysis of real data

This section starts by characterizing the dataset and preparing the catalogue to be used in the sequel. Afterwards, the seismic data is analyzed by means of MDS technique. Moreover, the events are divided into groups, according to their geographic location. Firstly, it adopted the Flinn–Engdahl regionalization of Earth (Young et al. 1996; Flinn et al. 1974; Flinn and Engdahl 1965), which corresponds to seismic zones commonly used by seismologists for localizing earthquakes (Table 1). Secondly, an alternative seismic regionalization, based on a rectangular grid, is also analyzed. Space-time and space-frequency correlation indices are proposed to quantify the similarities among regions, feeding MDS for generating maps to be interpreted.

#### 3.1 Brief description of the dataset

In this study, the Bulletin of the International Seismological Centre (ISC), available online at [http://](http://www.isc.ac.uk/)

[www.isc.ac.uk/](http://www.isc.ac.uk/), is used (International Seismological 2010). The ISC Bulletin contains seismic events since 1904, relying on data contributed by seismological agencies from around the world. To date, a total of 487 agencies have contributed to the ISC Bulletin, reporting data collected by more than 17,000 seismic stations located worldwide (Fig. 1).

The data was retrieved in April, 2012. Each data record contains information about date and time, geographic location, and magnitude of the event. As illustrated in Fig. 2, the seismic activity is far from being uniform across the Earth. In fact, the data reveals that most seismic activity occurs in three large zones (<http://earthquake.usgs.gov/>): (1) the Circum-Pacific belt (“Ring of Fire”) which extends from Chile, northward along the South American coast through Central America, Mexico, the West Coast of the USA, and the southern part of Alaska, through the Aleutian Islands to Japan, the Philippine Islands, New Guinea, the island groups of the Southwest Pacific, and to New Zealand (about 90 % of the world's earthquakes occur in this zone); (2) the second most seismic region (representing approximately 6 % of all earthquakes) is the Alpide belt, that extends from the Mediterranean region, eastward through Turkey, Iran, and northern India; (3) the firth major region is the submerged mid-Atlantic Ridge.

#### 3.2 Preparation of the earthquake catalogue

The ISC Bulletin comprises of different magnitude scales and events contributed by various seismological networks. Before using the seismic data, a homogeneous earthquake catalogue must be obtained, in terms of magnitude of the events (Wiener 2001; Hussein et al. 2008). In this study, we adopt the moment magnitude,  $M_W$ , as the scale of reference. Magnitudes expressed in different scales, namely local magnitude,  $M_L$ , surface-wave magnitude,  $M_S$ , body-wave magnitude,  $M_B$ , and duration magnitude  $M_D$  are converted into  $M_W$ . This procedure is accomplished by taking samples of events with magnitudes expressed in various scales and computing the relations between scales using an orthogonal regression algorithm. Contrary to the least-squares linear regression procedure, orthogonal regression gives accurate results even when both compared magnitudes are affected by measurement errors (Das et al. 2012). Figure 3 just illustrates the relation between  $M_W$  and  $M_B$  scales, as all other cases are similar.

**Table 1** Flinn–Engdahl regions and characterization of the data

Region number	Region name	Number of seismic events	Minimum Magnitude	Maximum Magnitude
1	Alaska–Aleutan arc	39,196	1.2	8.3
2	Southeastern Alaska to Washington	19,650	0	8.1
3	Oregon, California and Nevada	26,801	0	8.2
4	Baja California and Gulf of California	7,711	0.5	7.5
5	Mexico - Guatemala area	30,294	0	8.1
6	Central America	20,671	0.9	7.8
7	Caribbean loop	48,932	0.3	8.1
8	Andean South America	81,766	0	8.6
9	Extreme South America	2,557	1.1	7.8
10	Southern Antilles	6,227	1.9	7.8
11	New Zealand region	58,759	0	8.1
12	Kermadec–Tonga–Samoa Basin area	50,764	0.7	8.3
13	Fiji Islands area	23,957	1.2	7.9
14	Vanuatu Islands	29,477	0	8.2
15	Bismarck and Solomon Islands	29,895	0.3	8.1
16	New Guinea	25,221	−0.1	7.9
17	Caroline Islands area	5,097	1.7	7.9
18	Guam to Japan	34,335	1	8.1
19	Japan–Kuril Islands–Kamchatka Peninsula	867,008	−1.4	8.5
20	Southwestern Japan and Ryukyu Islands	584,156	−1.4	8.2
21	Taiwan area	286,413	−0.2	8
22	Philippine Islands	31,723	0	8.4
23	Borneo–Sulawesi	34,807	1.2	8.2
24	Sunda arc	47,076	0	8.4
25	Myanmar and Southeast Asia	8,327	0.1	8
26	India–Xizang–Sichuan–Yunnan	32,256	−0.8	8.6
27	Southern Xinjiang to Gansu	17,563	0	8.5
28	Lake Issyk-Kul to Lake Baykal	34,596	0	8.4
29	Western Asia	22,092	0	8.2
30	Middle East–Crimea–Eastern Balkans	223,922	0	8.4
31	Western Mediterranean area	195,228	−0.6	7.5
32	Atlantic Ocean	37,849	0	8.3
33	Indian Ocean	13,175	1.3	7.9
34	Eastern North America	15,258	0	7.8
35	Eastern South America	68	3.1	5.7
36	Northwestern Europe	92,045	−0.5	6.3
37	Africa	49,498	−0.3	7.4
38	Australia	7,813	0	7.8
39	Pacific Basin	3,040	−2.1	7.2
40	Arctic zone	18,993	0	7.1
41	Eastern Asia	15,423	0	7.8
42	Northeast. Asia, North. Alaska to Greenland	6,941	0	7.6
43	Southeastern and Antarctic Pacific Ocean	7,076	2.2	7.1
44	Galápagos Islands area	2,402	2.3	6.8

**Table 1** (continued)

Region number	Region name	Number of seismic events	Minimum Magnitude	Maximum Magnitude
45	Macquarie loop	1,792	2.1	7.8
46	Andaman Islands to Sumatera	21,038	0.6	9.2
47	Baluchistan	4,155	1.8	7.6
48	Hindu Kush and Pamir area	41,023	0	8
49	Northern Eurasia	60,156	−0.6	6.8
50	Antarctica	68	2.2	5.5

After homogenization, the magnitude completeness of the catalogue was studied in order to find the magnitude threshold above which the catalogue is complete. Magnitude completeness is defined as the lowest magnitude at which all earthquakes in a given space-time volume are detected. A correct estimate of the magnitude completeness is important since a value too high may lead to usable data being discarded, while a value too low might lead to the inclusion of incomplete data and to a biased analysis (Stucchi et al. 2004).

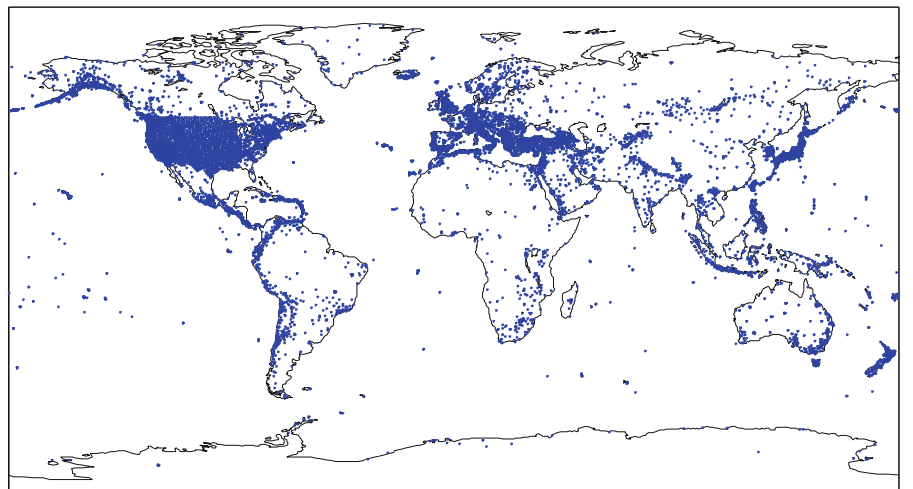
In this study, the magnitude completeness is estimated by fitting a GR model to the observed frequency–magnitude distributions of every F-E region. The magnitudes at which the lower ends of the frequency–magnitude distributions diverge from the GR lines are calculated (Mignan and Woessner 2012). The value  $M_W=4.5$  was found as the threshold above which magnitude completeness is achieved for all regions. Therefore, this value is taken as an estimate of catalogue completeness and events of magnitude below the threshold value are discarded.

For example, Fig. 4 depicts the frequency–magnitude distribution corresponding to F-E region Alaska–Aleutan arc, as well as the GR line that best fits to the data. In this case, completeness would be achieved for a slightly inferior threshold, namely for magnitudes above  $M_W=4.3$ .

### 3.3 MDS analysis based on space-time correlation index

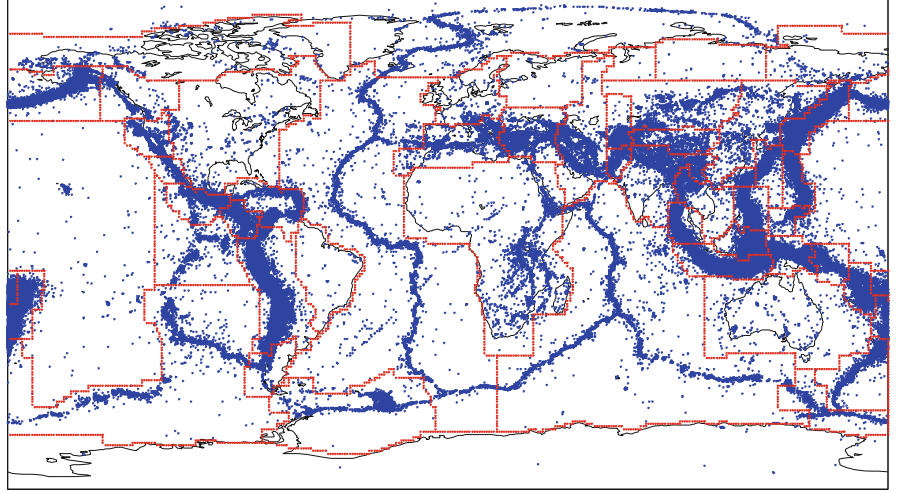
In this subsection, we consider the construction of matrix  $X$  using a measure based on the signal time correlation. We apply the MDS method to the  $s=50$  F-E regions. The data consists on the magnitudes of the seismic events and the period of analysis goes from January 1, 1904 up to March 14, 2012. We firstly adopt a correlation index, representing a measurement of the similarity between regions  $i$  and  $j$  ( $i, j=1, \dots, s$ ). Then, based on that index, an  $s \times s$  symmetric matrix,  $C$ , is computed. Finally, in order to reveal the possible relationships between the F-E zones under analysis, the MDS technique is used. In this perspective, while

**Fig. 1** Geographical location of the seismic stations that contributed to the ISC Bulletin





**Fig. 2** Geographical location of seismic events, according to the ISC Bulletin. Only the events with magnitude greater than four are depicted. The red lines represent the borders of the Flinn–Engdahl regions



several MDS criteria were tested, the Sammon criterion revealed good results and was adopted in all simulations. It should be noted that this criterion tries to optimize a cost function that describes how well the pairwise distances in a data set are preserved (Ahrens 2006; Duarte et al. 2010).

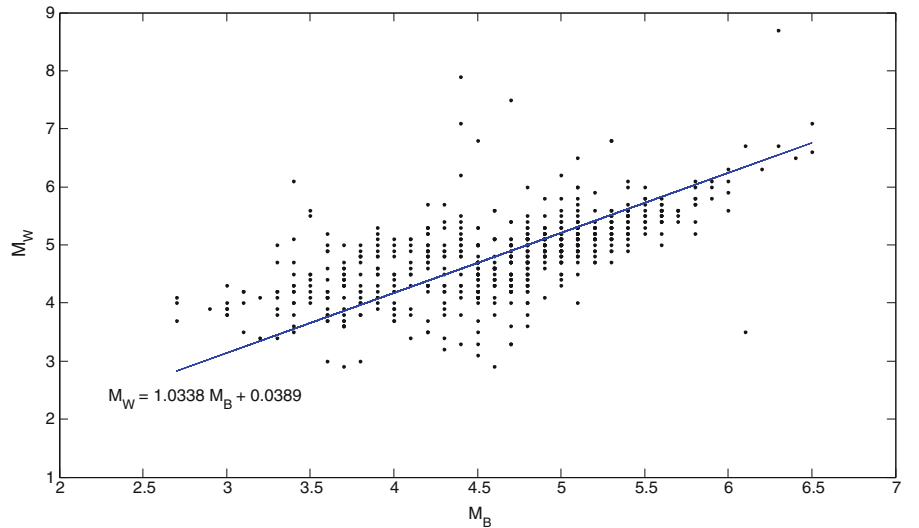
The space-time correlation index,  $c'_{ij}$ , adopted in this subsection is given by Eq. (5).

$$c'_{ij} = 1 - \left( p \frac{a_{ij}}{a_{\max}} + (1-p) \frac{d_{ij}}{d_{\max}} \right) \quad (5)$$

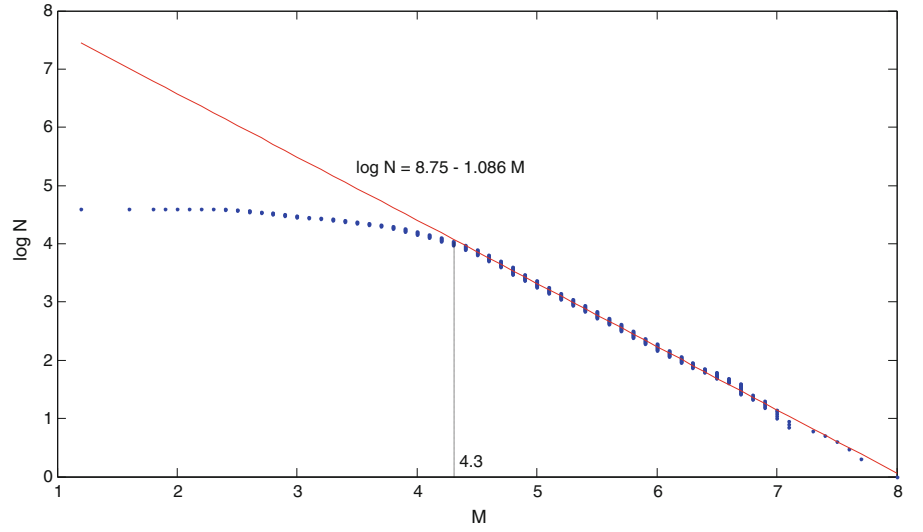
where  $a_{\max} = \max_{i,j} \{a_{ij}\}$  and  $d_{\max} = \max_{i,j} \{d_{ij}\}$ . The index  $c'_{ij}$  relies on two functions,  $a_{ij}$  and  $d_{ij}$ , as defined by Eqs. (6) and (7), respectively.

$$a_{ij} = r_{\text{Earth}} \cdot \arctan_2 \left( \frac{\sqrt{(\cos\phi_j \sin(\lambda_i - \lambda_j))^2 + (\cos\phi_i \sin\phi_j - \sin\phi_i \cos\phi_j \cos(\lambda_i - \lambda_j))^2}}{\cos\phi_i \sin\phi_j + \cos\phi_i \cos\phi_j \cos(\lambda_i - \lambda_j)} \right) \quad (6)$$

**Fig. 3** Relation between  $M_W$  and  $M_B$  scales using orthogonal regression



**Fig. 4** Frequency–magnitude distribution, corresponding to F-E region Alaska–Aleutan arc, as well as the GR line that best fits to the data



$$d_{ij} = \sqrt{1 - \left( \frac{\sum_{t=1}^T x_{FE_i}(t) \cdot x_{FE_j}(t)}{\sqrt{\sum_{t=1}^T x_{FE_i}^2(t) \cdot \sum_{t=1}^T x_{FE_j}^2(t)}} \right)^2} \quad (7)$$

Equation (6) gives the great-circle distance between two  $(i, j)$  regions. We assume that Earth is a sphere and the latitude/longitude coordinates  $(\phi_i, \lambda_i)$  and  $(\phi_j, \lambda_j)$  of regions, correspond to their “centroids.” “Centroids” meant values that are calculated as geographic mid-points, using the coordinates of all events in each region, instead of the coordinates of the regions' borders. The parameter  $r_{Earth}$  represents the Earth's radius.

Equation (7) represents the cosine correlation between the vectors  $x_{FE_i}(t)$  and  $x_{FE_j}(t)$ , corresponding to every  $(i, j)$  F-E regions that are defined in Eq. (8). Parameter  $t$  represents time and  $T$  is the total time period of study, both expressed in seconds.

For obtaining  $x_{FE}(t)$  we consider:

$$x_{FE}(t) = \sum_{k=1}^T M_k \delta(t - k) \quad (8)$$

meaning that the seismic events are modeled as Dirac impulses,  $M_k \delta(t - t_0)$ , where  $M_k$  represents the magnitude and  $t_0$  is the time of the occurrence (Tenreiro Machado et al. 2011b). Hence, each  $x_{FE}(t)$  is a time-domain signal that corresponds to the sequence of all seismic events, registered in every F-E region, over the time of study.

The constant  $a_{ij\max}$ , in Eq. (5), denotes the maximum distance between F-E regions and serves as a normalization parameter. The constant  $p$  is a weighting parameter used to establish a compromise between geographic and seismic similarities among the F-E regions. This parameter was adjusted experimentally to  $p=0.2$ . Smaller values tend to overlook geographic similarities between regions, while larger values tend to generate MDS maps where the F-E regions form single spherical-shaped clouds of points, hiding the relationships among the data. In practical terms, this result occurs when the MDS is not capable of producing a “good” clustering and, therefore, is of no usefulness.

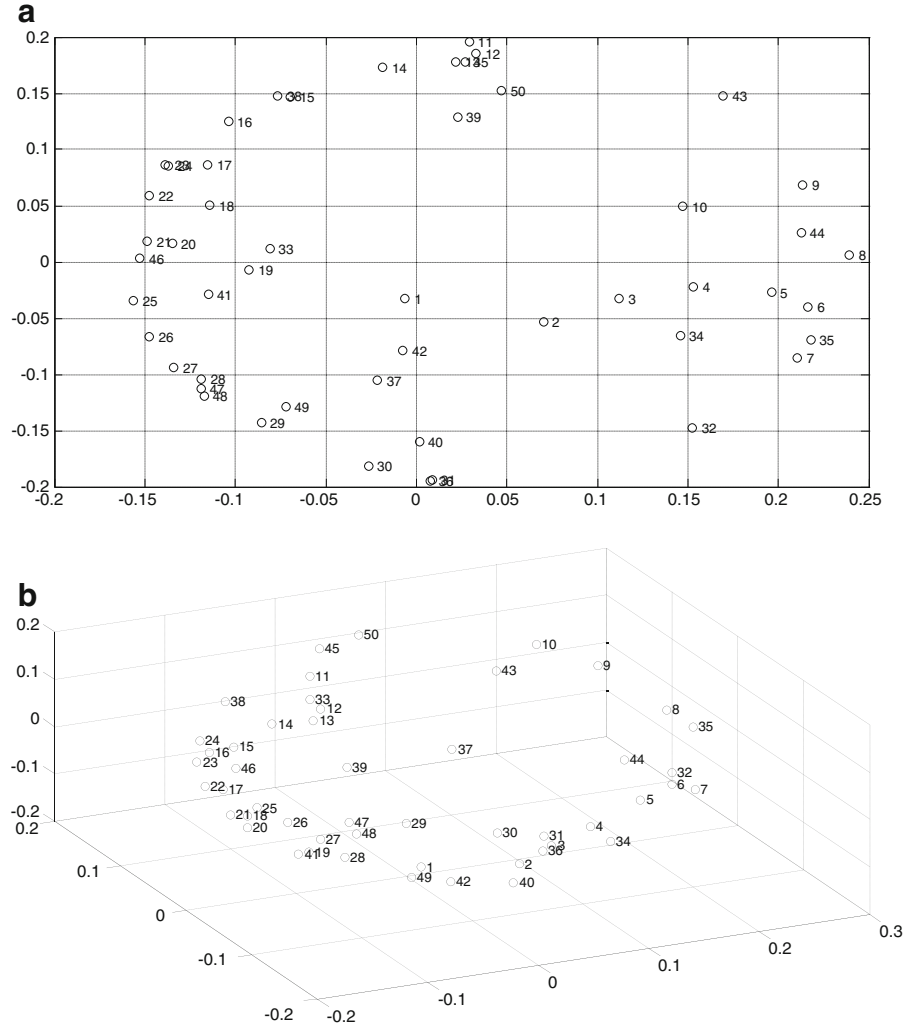
Figure 5 depicts the 2D and 3D locus of the F-E regions produced by the MDS. Each point on the map represents a region, denoted by the corresponding label shown in Table 1. We can notice clusters such as (1) regions  $FE8 \cup FE35$ ; (2)  $FE5 \cup FE6 \cup FE7 \cup FE44$ ; (3)  $FE4 \cup FE34$ ; (4)  $FE2 \cup FE3$ ; (5)  $FE1 \cup FE40 \cup FE42$ ; and (6)  $FE30 \cup FE31 \cup FE36$ . Therefore, regions belonging to the same cluster can be interpreted as being similar to each other, in the perspective of the space-time index (5).

We should note that MDS is merely a mathematical visualization tool and that a physical perspective of the reported results must be found in the light of the comparison index. Therefore, a further explanation about physical mechanisms associated with the results must be envisaged by standard complementary procedures.

Figures 6 and 7 depict the Shepard and stress plots, respectively. The Shepard diagram shows a reasonable distribution of points around the  $45^\circ$  line, particularly when the dimensionality is three, which means a good



**Fig. 5** MDS maps: **a** 2D; **b** 3D. The space-time correlation  $c_{ij}^t$  is used. Each point represents an F-H region



fit of the distances to the dissimilarities. The stress plot reveals that a three-dimensional space describes well the locus of the 50 F-E regions. In fact, the stress diminishes strongly until the dimensionality is two, moderately towards dimensionality three and weakly from then on. The maximum curvature point of the stress plot is often adopted as the criterion for deciding the dimensionality of the MDS maps. This means that, although four or five dimensions would represent the data more accurately, 3D maps are a good compromise between accuracy and easiness of visualization.

### 3.4 MDS analysis based on space-frequency correlation index

This subsection is considered a second index for construction matrix  $X$ .

Each signal  $x_{FE}(t)$  is processed using the Fourier Transform (FT):

$$F\{x_{FE}(t)\} = X_{FE}(j\omega) = \int_{-\infty}^{+\infty} x_{FE}(t) \cdot e^{-j\omega t} \cdot dt \quad (9)$$

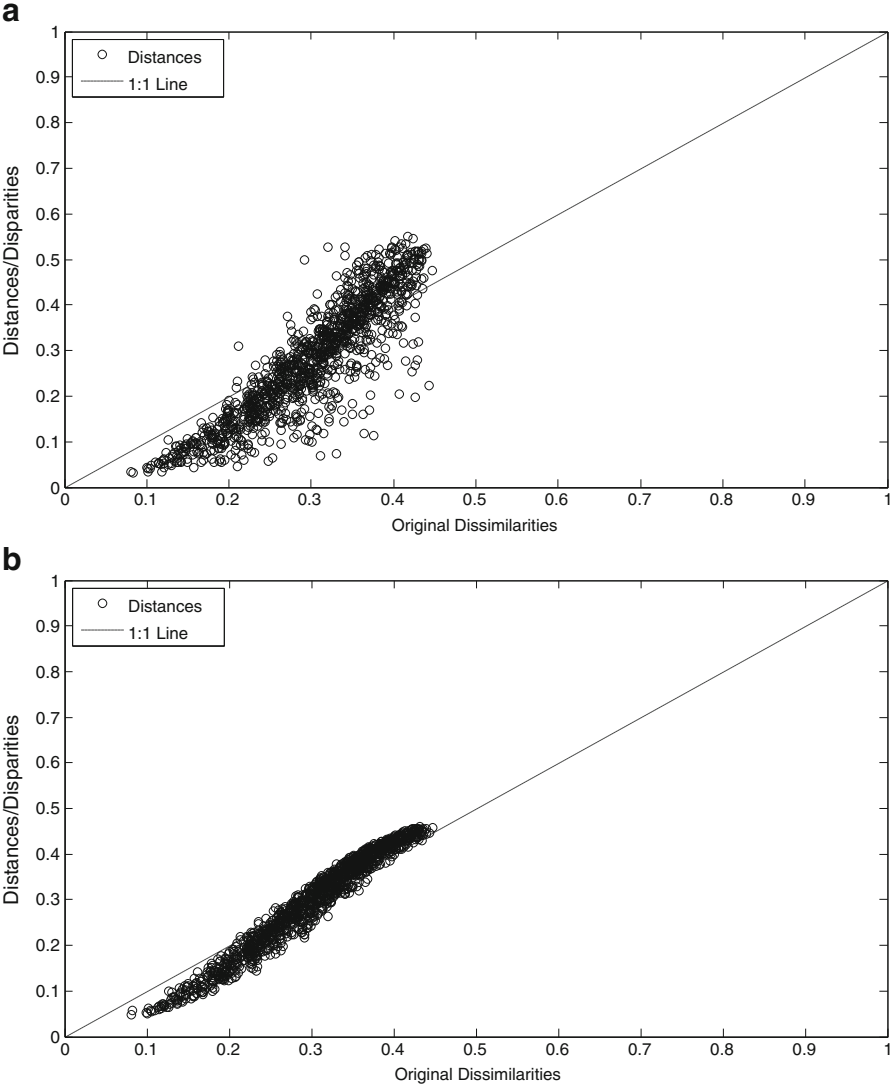
where  $F\{\cdot\}$  represents the Fourier operator,  $\omega$  is the angular frequency and  $j = \sqrt{-1}$ .

Figure 8 illustrates a typical spectrum of one signal  $x_{FE}(t)$ , corresponding to the F-E region number 19 (Southwestern Japan and Ryukyu Islands).

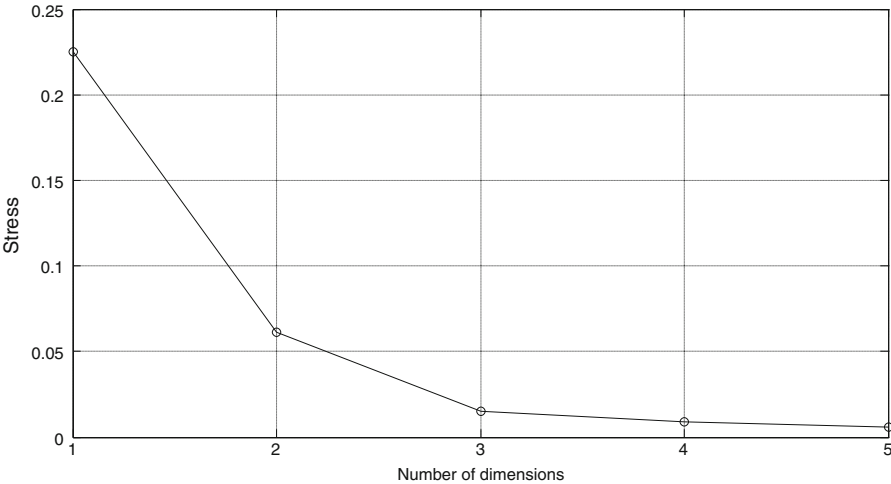
Adopting now the methodology established in subsection 3.3, the space-frequency correlation index can be defined as:

$$c_{ij}^{\omega} = 1 - \left( p \frac{a_{ij}}{a_{\max}} + (1-p) \frac{h_{ij}}{h_{\max}} \right) \quad (10)$$

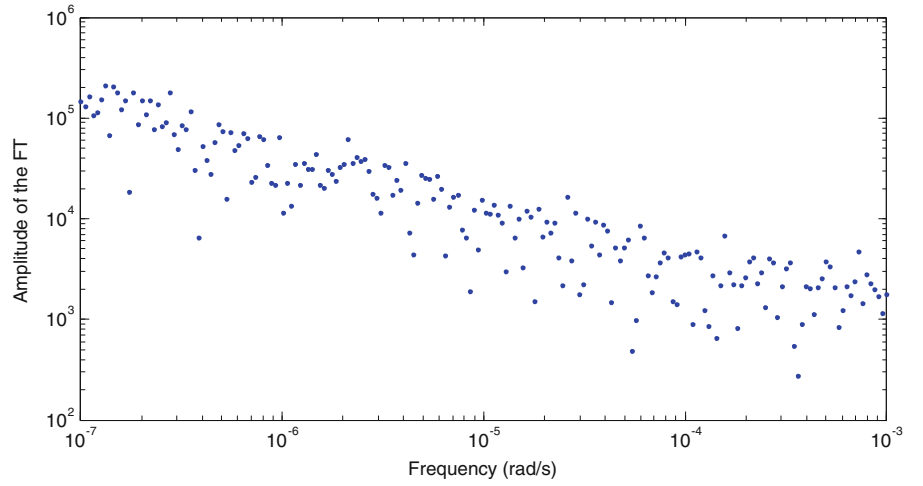
**Fig. 6** Shepard plots: **a** 2D; **b** 3D. The space-time correlation  $c_{ij}^t$  is used



**Fig. 7** Stress plot. The space-time correlation  $c_{ij}^t$  is used



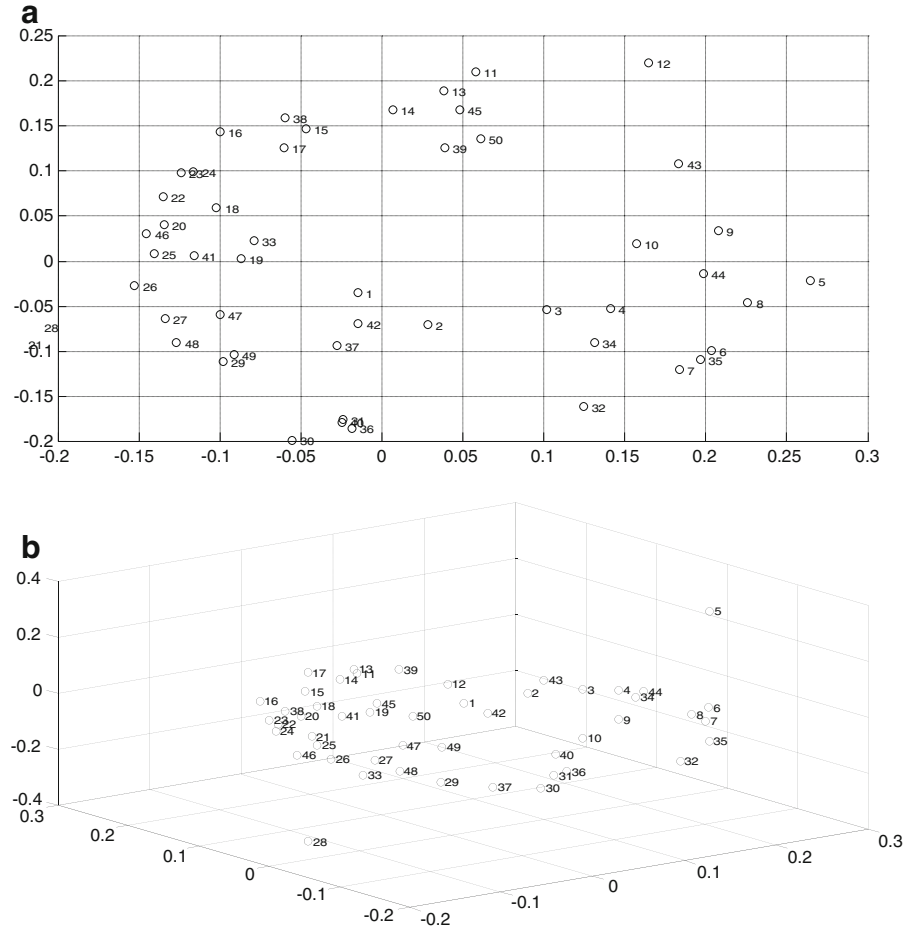
**Fig. 8** Magnitude of the FT of the signal  $x_{FE}(t)$ , which represents the seismic activity in the F-E region number 19. Events with magnitude above 4.5 were considered



which relies on  $a_{ij}$ , defined previously in Eq. (6), and the quantity  $h_{ij}$  (Ionescu et al. 2011) formulated as:

$$h_{ij} = \sum_{\omega=1}^{\Omega} \frac{[\text{Re}_i(\omega) - \text{Re}_j(\omega)]^2 + [\text{Im}_i(\omega) - \text{Im}_j(\omega)]^2}{[\text{Re}_i(\omega) + \text{Re}_j(\omega)]^2 + [\text{Im}_i(\omega) + \text{Im}_j(\omega)]^2} \quad (11)$$

**Fig. 9** MDS maps: **a** 2D; **b** 3D. The space-frequency correlation  $\tilde{c}_{ij}^{\omega}$  is used. Each point represents a F-H region



The index  $h_{ij}$  represents the distance between the real, Re, and the imaginary, Im, parts of the FT of two  $(i, j)$  time signals  $x_{FE}(t)$ . The parameter  $\Omega$  is the total number of points in the frequency domain used to evaluate the FT, which was fixed to  $\Omega=200$ , distributed logarithmically. The constant  $h_{\max} = \max_{i,j} \{h_{ij}\}$ , represents the maximum value of  $h_{ij}$  and the weigh  $p$  was adjusted as previously to  $p=0.2$ .

The 2D and 3D locus of the F-E regions resulting from the MDS analysis are represented in Fig. 9. We observe now a different pattern or, in other words, a distinct “shape,” but the main idea of clustering remains. This observation is usual in MDS plots, where alternative indices, capturing different characteristics of the

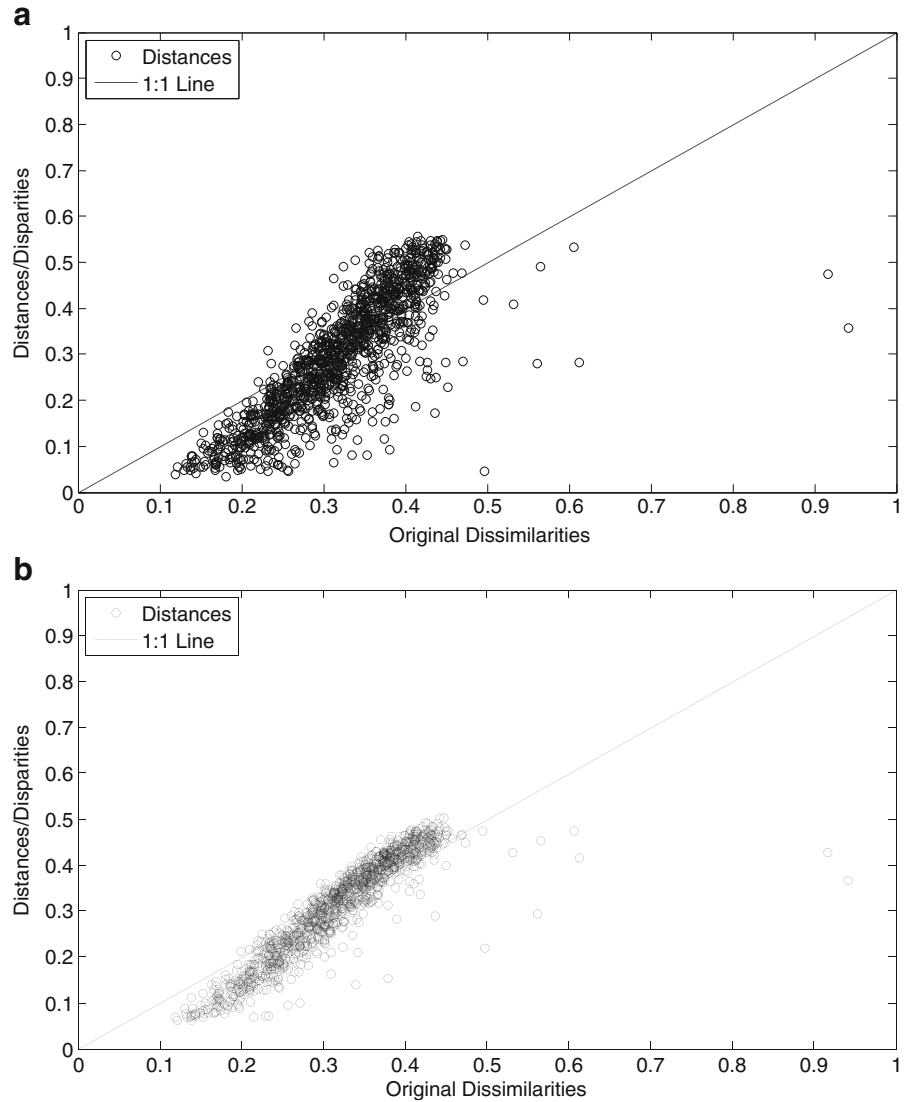
phenomena, lead to unidentical plots, but revealing the same type of conclusions.

The Shepard plots depicted in Fig. 10 reveal a good distribution of points around the 45 ° line. This distribution is even better than the one obtained for the space-time correlation index, especially for the 3D maps. The stress plot (Fig. 11) shows that a three dimensional space describes well the locus of the regions and, again, that  $c_{ij}^{\omega}$  gives results more accurate than  $c_{ij}^t$ .

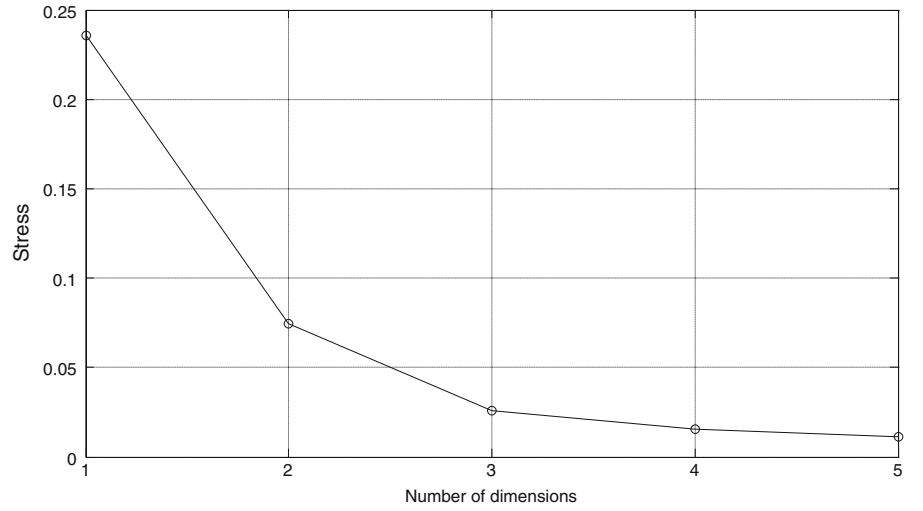
### 3.5 MDS sensitivity analysis

In this subsection, we analyze the MDS sensitivity to errors in the magnitude of the earthquake events. A

**Fig. 10** Shepard plots: **a** 2D; **b** 3D. The space-frequency correlation  $c_{ij}^{\omega}$  is used



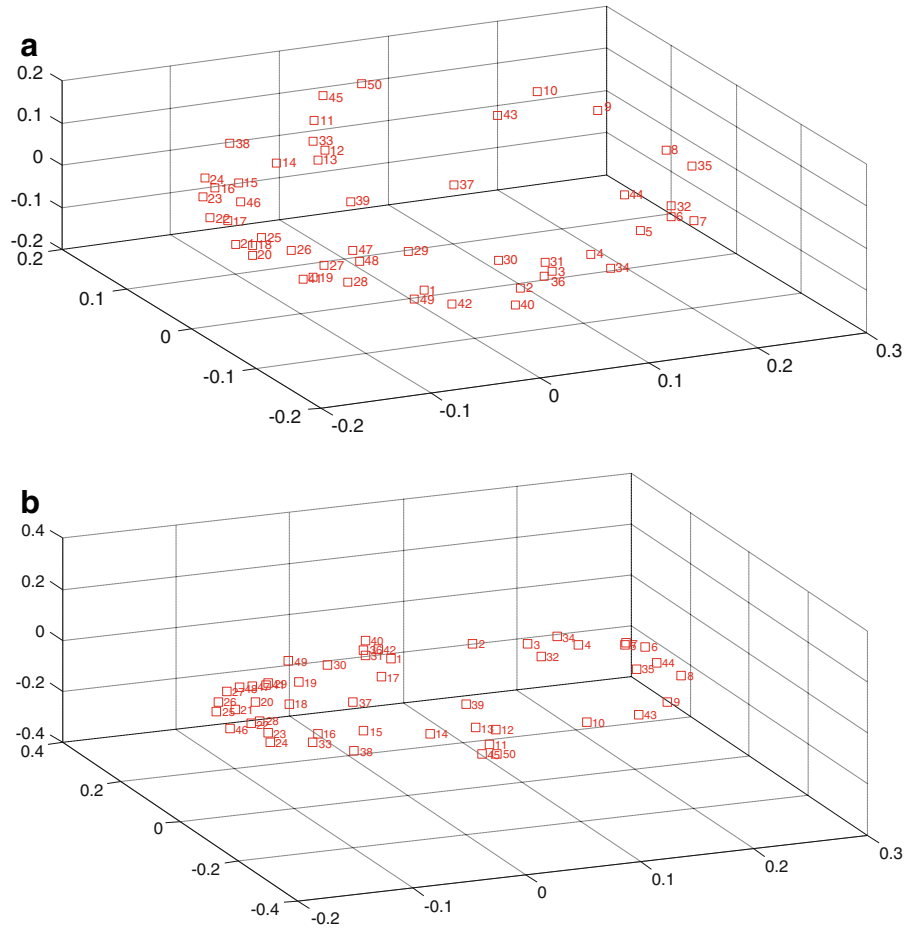
**Fig. 11** Stress plot. The space-frequency correlation  $c_{ij}^{sf}$  is used



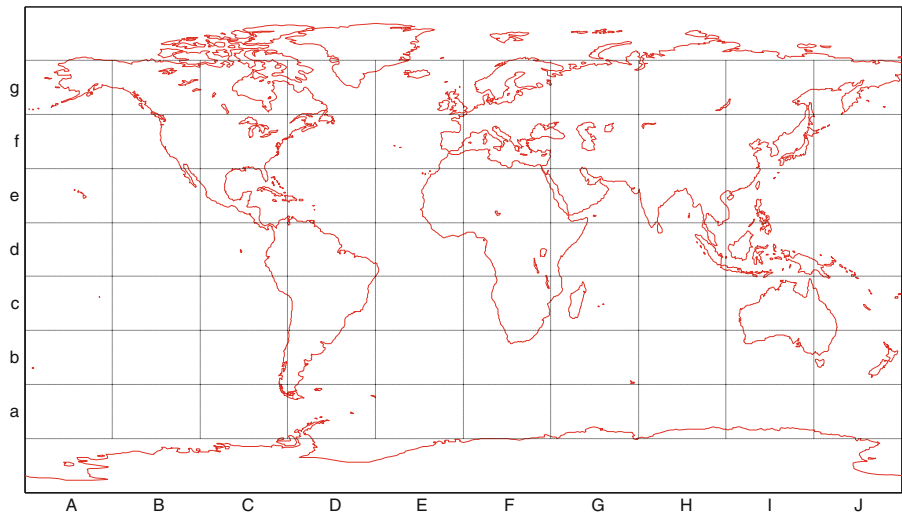
random error,  $\Delta_k$ , is added to the seismic time-series, Eq. (8), resulting in:

$$x_{FE}(t) = \sum_{k=1}^T M_k(1 + \Delta_k)\delta(t-k) \quad (12)$$

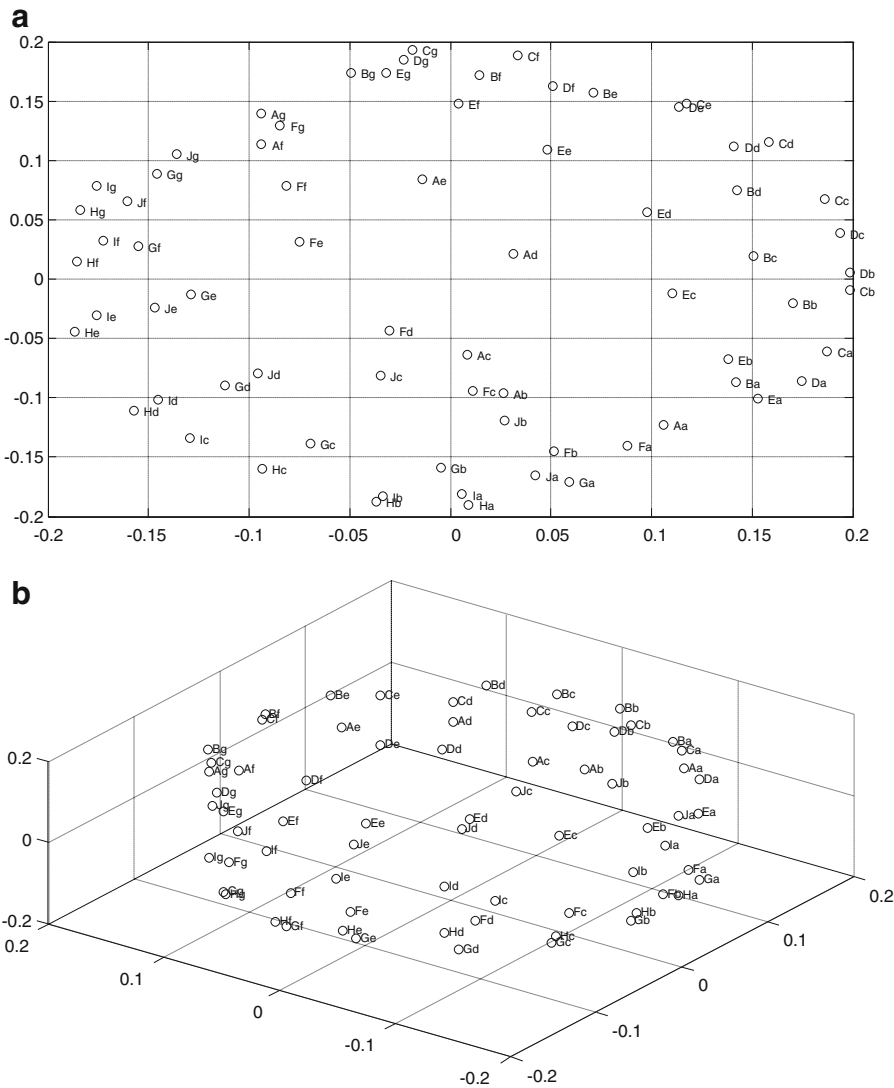
**Fig. 12** MDS map for the **a** time-space correlation and **b** space-frequency correlation ( $\Delta_k = \pm 5\%$ )



**Fig. 13** Earth's seismic regionalization based in a  $10 \times 7$  rectangular grid



**Fig. 14** MDS maps based on a  $10 \times 7$  rectangular grid. The space-time correlation  $c_{ij}^t$  is used. Each point represents a rectangular region





The space-time and space-frequency correlations ( $c_{ij}^t$ ,  $c_{ij}^\omega$ ) are calculated and the maps generated by the MDS are analyzed. Figure 12 depicts the MDS charts obtained for  $c_{ij}^t$  and  $c_{ij}^\omega$ , respectively, adopting  $\Delta_k = \pm 5\%$ . The MDS map based on the space-time correlation (Fig. 12a) is particularly robust and remains almost unchanged, when compared to Fig. 5b. The map obtained for the space-frequency correlation index is more sensitive to errors in magnitude (Fig. 12b). Nevertheless, while the geometric shape of the chart changes, when compared to Fig. 9b, for the adopted  $\Delta_k$  the clustering remains almost identical.

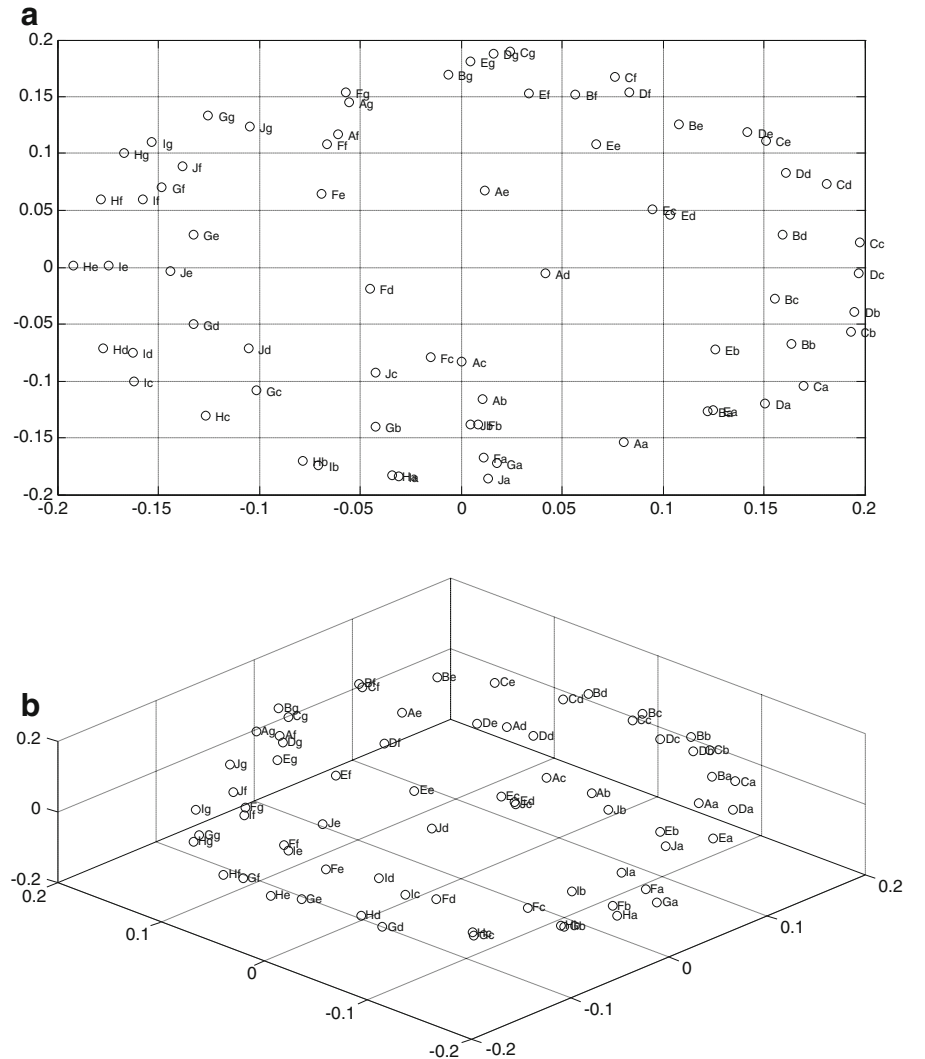
In conclusion, both space-time and space-frequency correlation indices proved to be adequate for representing the similarities between seismic regions.

The choice of one of the alternative indices as the “best” or even the definition of another index is not a matter of being correct or wrong. Therefore, the choice of the “adequate” index for MDS construction is strictly based on the user experience and intuition, remaining this issue to be further explored.

#### 4 MDS analysis of rectangular grid-based regions

Flinn–Engdahl regions are very general seismotectonic areas and/or seismic source zones which may not adequately capture all geological and geophysical conditions and regional factors. In this section, a different seismic regionalization is considered in alternative to

**Fig. 15** MDS map based on a  $10 \times 7$  rectangular grid. The space-frequency correlation  $c_{ij}^\omega$  is used. Each point represents a rectangular region



the F-E regions and the MDS tools, presented in section 3, are adopted. We now propose dividing Earth into  $10 \times 7$  rectangular cells,  $36^\circ$  wide in longitude per  $20^\circ$  height in latitude. Latitudes outside the interval  $\pm 70^\circ$  are not considered due to the reduced number of occurrences (Fig. 13). As shown previously, data is analyzed in a per region basis.

Figures 14 and 15 depicted the MDS graphs corresponding to the analysis based on space-time and space-frequency correlation indices, respectively. Shepard and stress plots are not included as they are identical to those shown in section 3. We observe that the analysis can lead to a more comprehensive visualization of the information than the F-E regions. In fact, as the rectangular regionalization is based in longitude and latitude information, without any other assumption, it can better capture local phenomena not revealed when F-E regions are used.

## 5 Conclusion

The Multidimensional Scaling analysis was proposed to visualize the similarities among Earth's seismic regions. The Bulletin of seismic events of the International Seismological Centre, available online at <http://www.isc.ac.uk/>, was used. The dataset covers the period from 1904 up to the present date, relying on more than three million events from around the world. Seismic occurrences were divided into groups, either according to the Flinn–Engdahl seismic regions of Earth or using a rectangular grid based in latitude and longitude coordinates. Two correlation indices were used to quantify the similarities between regions: a space-time correlation index and a space-frequency correlation index. MDS maps were proven as an intuitive and useful visual representation of the complex relationships that are present among seismic events, which are not perceived on traditional maps. The dynamics of earthquakes is complex and difficult to model. Furthermore, real data has an unstructured spatial distribution caused by the existence of geographic areas with a reduced instrumentation. However, MDS constitutes a valid alternative to analyze the available data, while avoiding modeling exercises that can lead to imprecise conclusions. This paper demonstrates the feasibility of the proposed method and encourages further research taking into account present day visualization techniques.

## References

- Ahrens B (2006) Distance in spatial interpolation of daily rain gauge data. *Hydrol Earth Syst Sci* 10:197–208
- Borg I, Groenen P (2005) *Modern multidimensional scaling theory and applications*, 2nd edn. Springer, New York
- Bronstein M, Bronstein A, Kimmel R, Yavneh I (2006) Multigrid multidimensional scaling. *Numer Linear Algebra Appl* 13: 149–171
- Chan FKW, So HC (2009) Efficient weighted multidimensional scaling for wireless sensor network localization. *IEEE Trans Signal Process* 57:4548–4553
- Das R, Wason HR, Sharma ML (2012) Magnitude conversion to unified moment magnitude using orthogonal regression relation. *J Asian Earth Sci* 50:44–51
- Duarte FB, Tenreiro Machado JA, Duarte GM (2010) Dynamics of the Dow Jones and the Nasdaq stock indexes. *Nonlinear Dynam* 61:691–705
- Dzwiniel W, Yuen DA, Boryczko K, Ben-Zion Y, Yoshioka S, Ito T (2005) Nonlinear multidimensional scaling and visualization of earthquake clusters over space, time and feature space. *Nonlinear Process Geophys* 12:117–128
- Flinn EA, Engdahl ER (1965) A proposed basis for geographical and seismic regionalization. *Rev Geophys* 3:123–149
- Flinn EA, Engdahl ER, Hill AR (1974) Seismic and geographical regionalization. *Bull Seismol Soc Am* 64:771–993
- Geller RJ (1997) Earthquake prediction: a critical review. *Geophys J Int* 131:425–450
- Gutenberg B, Richter CF (1944) Frequency of earthquakes in California. *Bull Seismol Soc Am* 34:185–188
- Hussein HM, Abou Elenean KM, Marzouk IA, Peresan A, Korrat IM, Abu El-Nader E, Panza GF, El-Gabry MN (2008) Integration and magnitude homogenization of the Egyptian earthquake catalogue. *Nat Hazards* 47:525–546
- International Seismological Centre (2010) On-line Bulletin, Internatl Seis Cent, Thatcham, United Kingdom. <http://www.isc.ac.uk>. Accessed 2 Apr 2012.
- Ionescu CM, Tenreiro Machado JA, Keyser RD (2011) Is multidimensional scaling suitable for mapping the input respiratory impedance in subjects and patients? *Comput Methods Prog Biomed* 104:189–200
- Jafari MA (2010) Statistical prediction of the next great earthquake around Tehran, Iran. *J Geodyn* 49:14–18
- Ji X, Zha H (2004) Sensor positioning in wireless ad-hoc sensor networks using multidimensional scaling. In: *Proc. IEEE INFOCOM*.
- Kagan YY (1992) Correlations of earthquake focal mechanism. *Geophys J Int* 110:305–320
- Kagan YY, Jackson DD (2000) Probabilistic forecasting of earthquakes. *Geophys J Int* 143:438–453
- Kruskal J (1964) Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika* 29:1–27
- Lindman M, Lund B, Roberts R, Jonsdottir K (2006) Physics of the Omori law: inferences from interevent time distributions and pore pressure diffusion modeling. *Tectonophysics* 424:209–222
- Matheus J, Dourado A, Henriques J, Antonio M, Nogueira D (2006) Iterative multidimensional scaling for industrial process monitoring. In: *IEEE International Conference on Systems, Man, and Cybernetics*, Taipei, Taiwan.

- Mignan A, Woessner J (2012) Estimating the magnitude of completeness for earthquake catalogs, Community Online Resource for Statistical Seismicity Analysis, doi:[10.5078/corssa-00180805](https://doi.org/10.5078/corssa-00180805). Available at <http://www.corssa.org>.
- Omori F (1894) On the aftershocks of earthquakes. *J Coll Sci Imp Univ Tokyo* 7:111–200
- Preoteasa CT, Ionescu E, Preoteasa E, Tenreiro Machado JA, Baleanu MC, Baleanu D (2013) Multidimensional scaling for orthodontic root resorption. *Mathematical Problems in Engineering*. 2013:383698 doi:[10.1155/2013/383698](https://doi.org/10.1155/2013/383698)
- Ramsay JO (1980) Some small sample results for maximum likelihood estimation in multidimensional scaling. *Psychometrika* 45:139–144
- Reasenberga PA, Jones LM (1989) Earthquake hazard after a mainshock in California. *Science* 243:1173–1176
- Stucchi M, Albin P, Mirto C, Rebez A (2004) Assessing the completeness of Italian historical earthquake data. *Ann Geophys* 47:659–673
- Tenreiro Machado JA, Costa AC, Lima MFM (2011a) Dynamical analysis of compositions. *Nonlinear Dynam* 65:399–412
- Tenreiro Machado JA, Duarte FB, Duarte GM (2011b) Analysis of financial data series using fractional Fourier transform and multidimensional scaling. *Nonlinear Dynam* 65:235–245
- Tenreiro Machado JA, Duarte GM, Duarte FB (2011c) Identifying economic periods and crisis using the multidimensional scaling. *Nonlinear Dynam* 63:611–622
- Torgerson W (1958) *Theory and methods of scaling*. Wiley, New York
- Tzagarakis C, Jerde TA, Lewis SM, Ugurbil K, Georgopoulos AP (2009) Cerebral cortical mechanisms of copying geometrical shapes: a multidimensional scaling analysis of fMRI patterns of activation. *Exp Brain Res* 194:369–380
- Utsu T, Ogata Y, Matsuura RS (1995) The centenary of the Omori formula for a decay law of aftershock activity. *J Phys Earth* 43:1–33
- Wiemer S (2001) A software package to analyze seismicity: ZMAP. *Seismol Res Lett* 72:373–382
- Young JB, Presgrave BW, Aichele H, Wiens DA, Flinn EA (1996) The Flinn-Engdahl regionalisation scheme: the 1995 revision. *Phys Earth Planet Inter* 96:223–297