

Analysis of temperature time-series: Embedding dynamics into the MDS method

António M. Lopes, J.A. Tenreiro Machado

a b s t r a c t

Global warming and the associated climate changes are being the subject of intensive research due to their major impact on social, economic and health aspects of the human life. Surface temperature time-series characterise Earth as a slow dynamics spatiotemporal system, evidencing long memory behaviour, typical of fractional order systems. Such phenomena are difficult to model and analyse, demanding for alternative approaches. This paper studies the complex correlations between global temperature time-series using the Multidimensional scaling (MDS) approach. MDS provides a graphical representation of the pattern of climatic similarities between regions around the globe. The similarities are quantified through two mathematical indices that correlate the monthly average temperatures observed in meteorological stations, over a given period of time. Furthermore, time dynamics is analysed by performing the MDS analysis over slices sampling the time series. MDS generates maps describing the stations' locus in the perspective that, if they are perceived to be similar to each other, then they are placed on the map forming clusters. We show that MDS provides an intuitive and useful visual representation of the complex relationships that are present among temperature time-series, which are not perceived on traditional geographic maps. Moreover, MDS avoids sensitivity to the irregular distribution density of the meteorological stations.

Keywords:

Complex systems, Temperature time-series, Multidimensional scaling, Global warming

1. Introduction

Complex systems constitute a challenging frontier in science. Complex systems study the relationships between multiple objects that contribute to a collective behaviour often revealing surprising dynamic phenomena. Observations and descriptions are frequent in natural (e.g., economy, biology, genetics) and man-made (e.g., computer science, chemical and physical apparatus) systems [52–54]. The modelling of these systems can adopt sophisticated mathematical tools, but often we verify that present day human knowledge and scientific tools are still far from capturing the overall richness of the system dynamics. Therefore, a fruitful interplay is possible by experimenting in a given case mathematical tools usual in completely distinct applications. This paper addresses the Earth global warming and in particular the evolution of its dynamics during the last decades, by adopting statistical and computer visualising tools. This strategy leads not only to avoid limitations of classical techniques, but also to evaluate a novel perspective for tackling complexity.

Understanding the climatic variability is extremely important nowadays. The Earth's climate is changing and we are facing an increasing number of extreme weather events like floods, droughts and anomalous temperatures. Since the early of the 20th century the Earth's surface average temperature has increased about $0.8\text{ }^{\circ}\text{C}$ [1] and several projections based on

climate models point out that during this century the average temperature may increase between 1.1 °C and 6.4 °C [2]. Earth is warming and it seems that human activity is one of the main causes for such phenomenon [3–5]. In fact, since the beginning of the industrial revolution, the burning of fossil fuels has significantly increased the non-natural emissions of carbon dioxide to the atmosphere. Estimates from the Intergovernmental Panel on Climate Change (IPCC) revealed an increase of CO₂ emissions of about 30% and, in 2007, the increasing rate was around 0.4% per year [6]. Carbon dioxide is a greenhouse gas that absorbs the infrared radiation produced by the reflexion of the sunlight on the Earth's surface, trapping the heat in the atmosphere. Some impacts of global warming such as the record of high temperatures, the melting glaciers and severe flooding are becoming increasingly common across the countries and around the world [7–13]. Besides the direct effect on temperature, warming leads to the modification of wind patterns, the development of humidity and the alteration of the rates of precipitation. These phenomena are being the subject of intensive research due to major impact on social, economic and health aspects of the human life [14–17].

The analysis of temperature time-series is an important matter that can help us understanding the climate variability. This knowledge may be used in developing pro-active strategies, both for prevention and for minimizing the expected negative impacts of global warming. Although several indicators of climate change can be used, the time evolution of Earth's surface temperature is probably the metric that is easiest to understand [18]. Moreover, vast and reliable temperature data records, comprising information from meteorological stations around the world, are freely available for supporting temperature data analysis [19–21].

Many strategies have been applied in the study of temperature time-series [22]. For example, Grieser et al. [23] studied the monthly mean temperatures of European meteorological stations. They applied statistical tools to decompose the time-series into significant components and found that the phase of the annual cycle is shifted within the year, backward and forward in western and Eastern Europe, respectively. Moreover, they observed that extreme events increased significantly and correspond mostly to cold peaks in winter. They also found significant harmonic components with period 92.3 months in data from northern and western Europe.

Hughes et al. [24] point out that the Antarctic Peninsula region is warming faster than the rest of the world. They analyse the minimum and maximum temperatures time-series at the Faraday/Vernadsky (Lat 65.25 S, Lon 64.25 W) station using a multiple regression model with non-Gaussian correlated errors. They have found that the increase in the minimum monthly temperature was around 6.78 °C over the period 1951–2003.

Nonlinear tools are used to analyse temperature time-series in Viola et al. [25]. The method of delay coordinates is employed for state space reconstruction and delay parameters are evaluated using the method of average mutual information and the method of false nearest neighbours. A nonlinear prediction method is employed to extrapolate temperature values until 2028.

Founda et al. [26] employed statistical methods to analyse the air temperature time-series of Athens from 1897 up to 2001. They observed a tendency towards warmer years, with significantly warmer summers and springs and somewhat warmer winters.

In Ge et al. [27] temperature variation through 2000 years in China was analysed. The results revealed the cold periods 1620–1710 and 1800–1860 and the warming registered during the 20th century.

Deser et al. [28] analyse trends in sea surface and marine air temperature using data from different sources. They found warming everywhere except the in the north-western Atlantic. The largest warming trends were found in the middle latitudes of both northern and southern hemispheres.

Surface temperatures characterise Earth as a slow dynamics spatiotemporal system, evidencing long memory behaviour and complex relationships, typical of fractional order systems [29–31]. Large amounts of data are available for analysis but, in fact, the data is geographically irregular and temporally insufficient to be managed by the analysis tools commonly used in dynamical systems.

In this paper we propose applying the Multidimensional scaling (MDS) technique to study and visualise the complex correlations between global temperature time-series. The MDS method is a powerful tool to expose clusters formed by objects with similarities and is an attractive way for mapping data into a low dimensional space.

MDS has been applied in many areas [32,33]. For example, in [34] this approach was used in the analysis of pilot performance data obtained during simulated air-to-air combat. Costa et al. [35] used MDS to analyse the DNA code in the perspective of identifying structural patterns in the nuclear and mitochondrial genomes. In [36] MDS is used to study fifteen stock markets and to expose possible time-varying correlations between them. Other applications in the area of computational chemistry, machine learning, concept maps, wireless network sensors, data mining and image processing can also be cited [37–41].

In the area of geosciences, MDS was proposed by Oñate and Pou [42] to analyse temperature time-series from eleven meteorological stations over the Iberian Peninsula. Trends are identified by means of the Mann–Kendall test, while MDS produces an automatic grouping. More recently, Stephenson and Doblas-Reyes [43] used MDS as an exploratory tool for describing ensembles of forecasts.

MDS has advantages over other methods, such as standard clustering algorithms and Principal Component Analysis (PCA). In fact, for standard cluster algorithms, central clusters surrounded by rings of points are generally difficult to identify [43]. On the other hand, since MDS can use any similarity/dissimilarity matrix based on distinct metrics, it is a more general projection method than PCA.

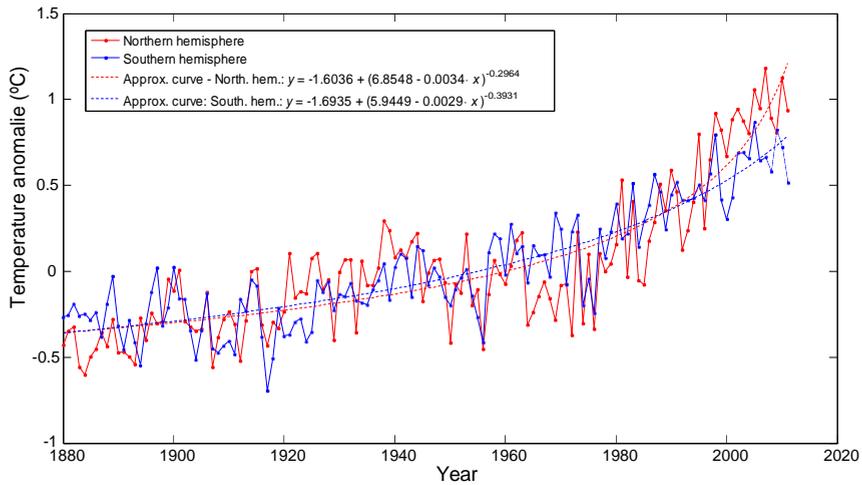


Fig. 1. Annual land temperature anomalies, starting in 1880 up to 2011, for both northern and southern hemispheres. The average of the annual mean temperatures, over the period 1951–1980, on both hemispheres, is taken as reference.

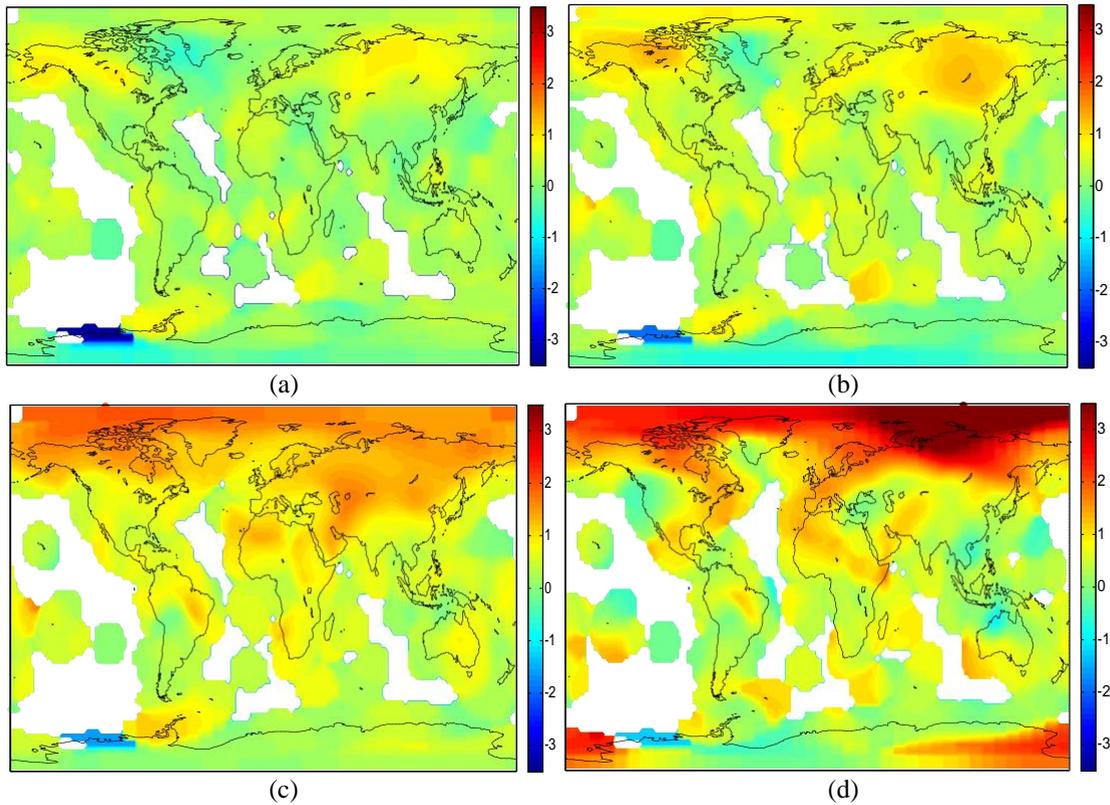


Fig. 2. Worldwide land temperature anomalies. The average of the annual mean temperatures, over the period 1951–1980, on each location, is taken as reference: (a) period 1981–1990; (b) period 1991–2000; (c) period 2001–2010; (d) year 2011. (The white regions on the maps are due to insufficient data).

In our approach, the similarities among geographic regions are quantified through mathematical indices that correlate the monthly average temperatures observed in the meteorological stations over a given period of time. The MDS generates maps describing the stations' locus, such that, those stations, perceived to be similar in some sense, are placed on the map forming clusters. The MDS constructs the object representations in a new space. Usually units are not allocated to that space,

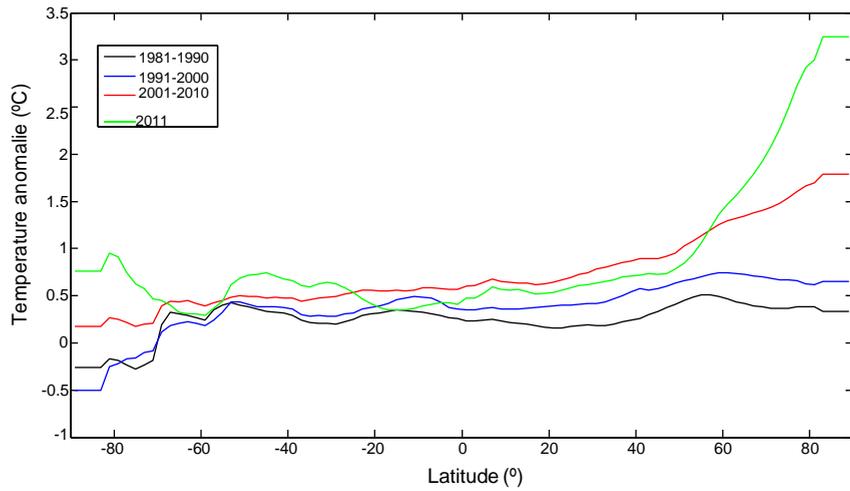


Fig. 3. Mean temperature anomalies as a function of the latitude.

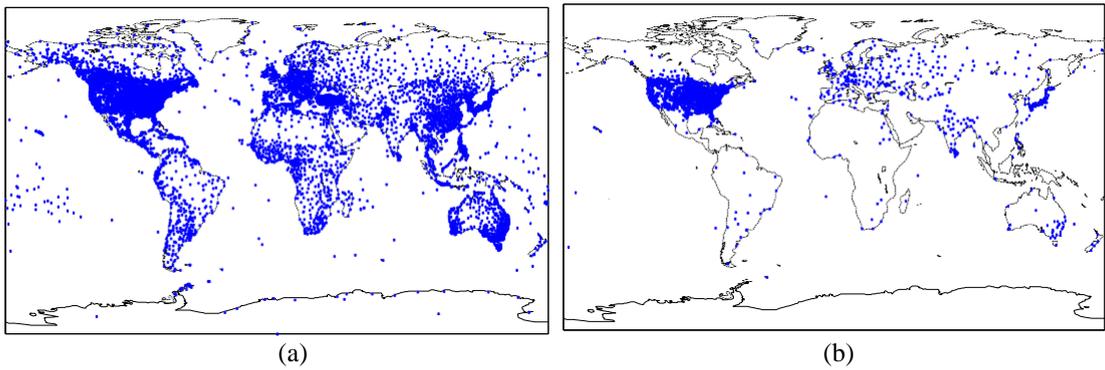


Fig. 4. Geographical location of the meteorological stations in the GHCN-M version 3 dataset: (a) all stations; (b) stations with more than 100-year records.

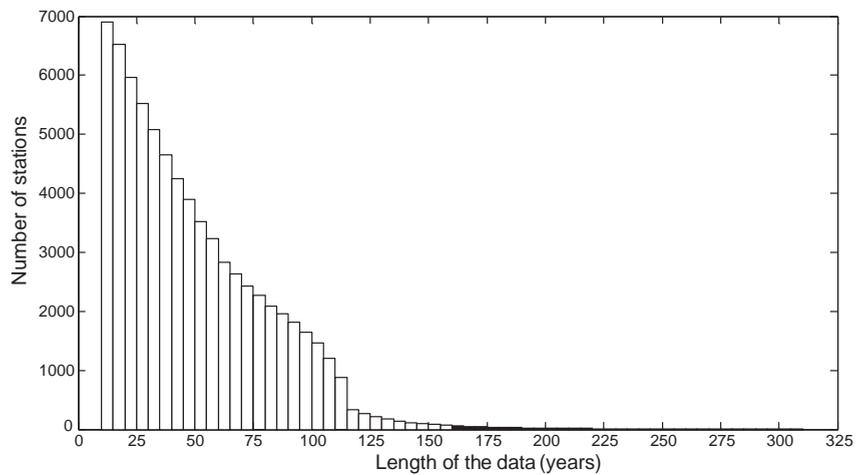


Fig. 5. Cumulative histogram depicting the number of meteorological stations with data records longer than a certain period of years.

but, in fact, the units should be those of the index adopted for comparing items. Therefore, distinct indices can lead to different units. In many cases indices are non-dimensional. Whatever be the case, the objects of our study are a ‘mixture’ of temperature values and time of events. The geographic coordinates of the stations that measure the temperatures are

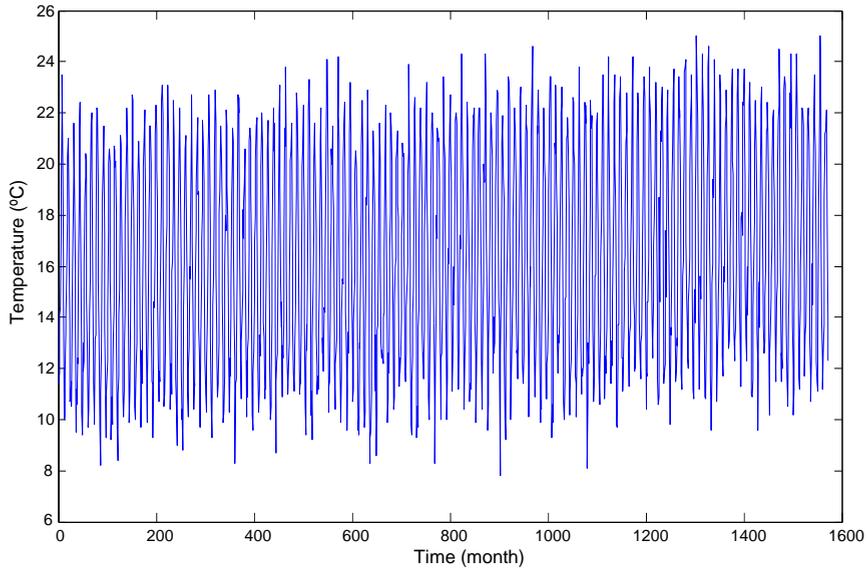


Fig. 6. Monthly average temperature registered in the meteorological station Lisbon, Portugal (January 1881–August 2011).

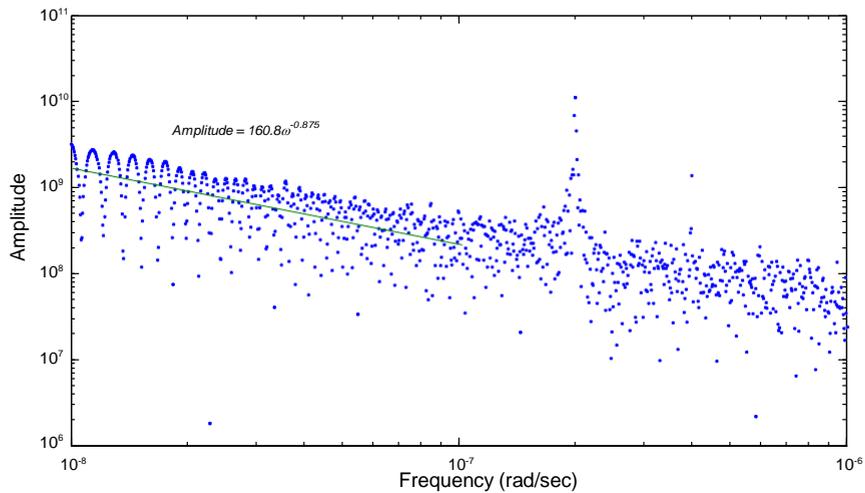


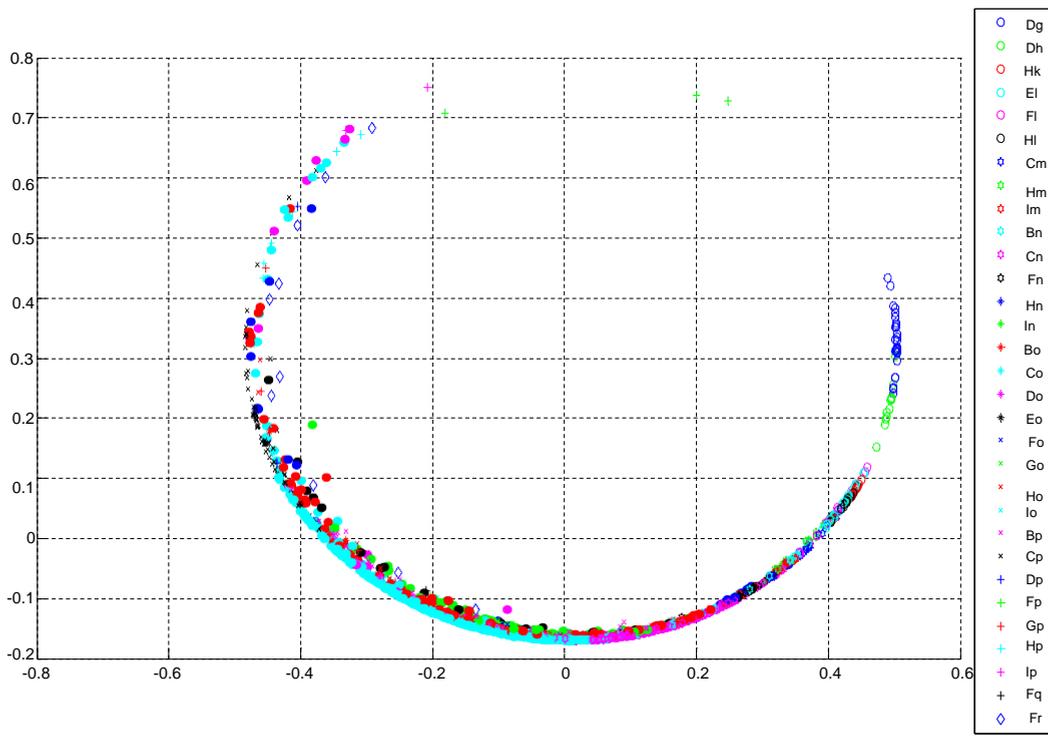
Fig. 7. Amplitude spectrum corresponding to the monthly average temperatures of Lisbon, Portugal (January 1881–August 2011).

not the coordinates of the MDS points, which are those of the ‘mixture’ of variables that were used. Therefore, points in the MDS maps have a one-to-one relationship with the original stations without using explicitly their coordinates.

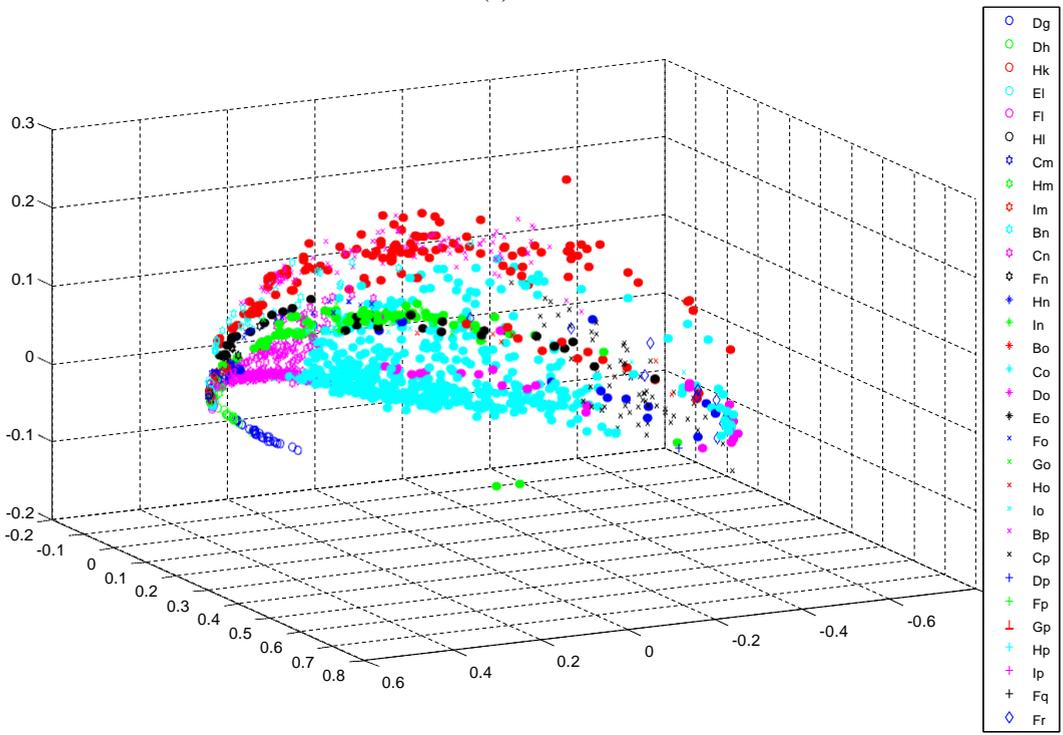
We show that MDS provides an intuitive and useful visual representation of the complex relationships that are present among temperature time-series, which are not perceived on traditional geographic maps. Moreover, MDS surpasses the sensitivity to the irregular geographic distribution of the meteorological stations. The MDS method by a priori modelling assumptions relies only on the data and comparison method. Therefore, it is well suited for accessing phenomena where a plethora of actions with distinct nature is present.

Besides including distinct measures and a considerable volume of data, this paper includes a further generalisation of the classical schemes. In fact, standard MDS representations capture the system dynamics only by means of the comparison index. In this case it is also included a new scheme that results of subdividing each time series into several smaller slices that are shed into the MDS map as separate points depicting, therefore, the time evolution of the phenomena.

Having these ideas in mind, the paper is organised as follows. In Section 2 some facts related to the global warming are illustrated. Section 3 overviews the MDS technique. In Section 4 we formulate the framework of the analysis in the perspective of MDS. The approach is applied to the data and the main results are interpreted and analysed. Finally, Section 5 outlines the main conclusions.



(a)



(b)

Fig. 8. MDS maps: (a) 2D; (b) 3D. The cosine correlation c_{ij}^c is used and the period of analysis is 1951–2010. Each point represents a meteorological station.

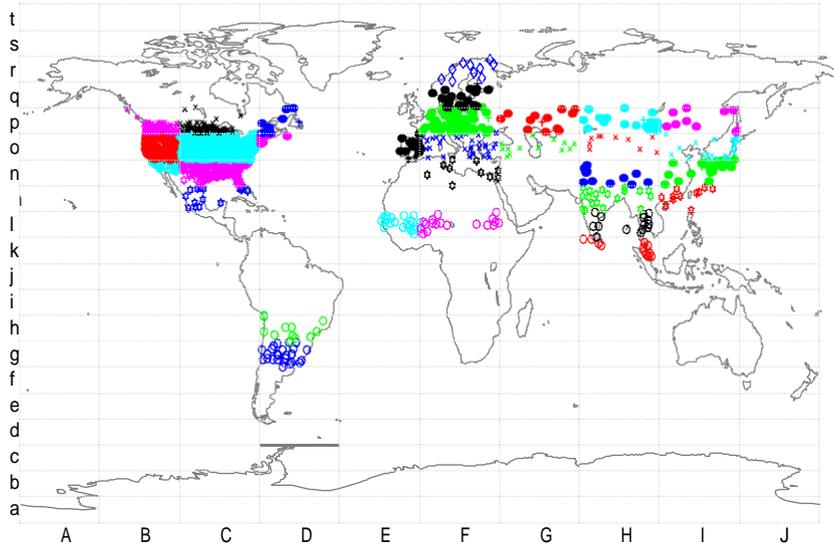


Fig. 9. Location of the stations on a geographic map. The cosine correlation c_{ij}^c is used and the period of analysis is 1951–2010. The same symbols adopted for the MDS maps are used.

2. Global warming

In this section, we use data available at the NASA, Goddard Institute for Space Studies website (<http://www.giss.nasa.gov/>) to illustrate and highlight some facts and evidences of global warming [14]. Fig. 1 depicts the annual land temperature anomalies starting in 1880 up to 2011, for northern and southern hemispheres. The average of the annual mean temperatures, over the period 1951–1980, on both hemispheres, is taken as reference. It must be noted that a temperature anomaly represents the deviation from a reference value or long-term average [14]. Positive values mean that the observed temperatures are higher than the reference value, whereas negative anomalies indicate that the observed temperatures are lower than the reference. As can be seen from Fig. 1, there has been a continuous increase in the annual mean temperature and the rate of increase was higher in the last thirty years. The time evolution of the temperature anomalies has power-law behaviour typically observed in complex and fractional-order systems.

Fig. 2 shows four contour plots corresponding to the geographical distributions of land temperature anomalies over the decades 1981–1990 (Fig. 2a), 1991–2000 (Fig. 2b) and 2001–2010 (Fig. 2c), while Fig. 2d corresponds only to the year 2011. The results show that the Earth’s warming is undoubted. Moreover, the year 2011 was extremely anomalous, with vast regions of the globe registering temperatures more than 2°C above the reference values of these regions.

The annual land temperature anomalies were also analysed as a function of the latitude, which means that, for each latitude coordinate, the average value of temperature anomalies was calculated. Fig. 3 depicts the results corresponding to the four periods mentioned previously. The results show that, on average, the northern hemisphere has been more affected by warming.

3. Multidimensional scaling

Multidimensional scaling (MDS) is a statistical technique for visualising information in the perspective of exploring similarities in data [44–49]. The main idea is to detect meaningful underlying dimensions that allow the researcher to observe similarities, or dissimilarities (distances), between the objects.

The MDS algorithm requires the definition a measure of similarity (or, alternatively, of distance) and the construction an $s \times s$ symmetric matrix C of item to item similarities (or, alternatively, of distances), where s is the total number of objects. In classical MDS matrix C is symmetric and its main diagonal is composed of “1” for similarities (or of “0” for dissimilarities).

MDS assigns a point to each item in a multi-dimensional space and arranges the set in order to reproduce the observed similarities. Often, instead of similarities there are considered dissimilarities, or distances, between the objects. For two or three dimensions the resulting locations may be displayed in a “map” that can be analysed.

By rearranging the object’s positions in the space, MDS tries to arrive at a configuration that best approximates the observed similarities (or distances). For this purpose MDS uses a function minimization algorithm that evaluates different configurations with the goal of maximising the goodness-of-fit. The most common measure that is used to evaluate how well a particular configuration reproduces the observed distance matrix is the raw stress measure defined by

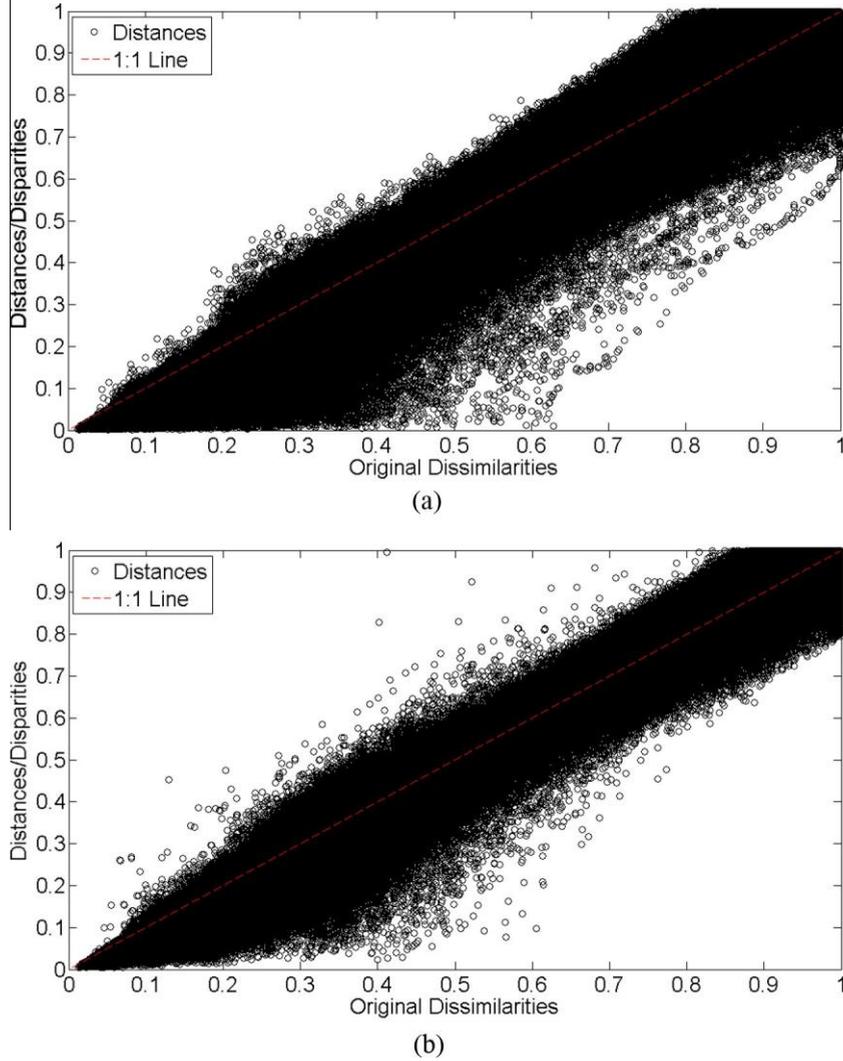


Fig. 10. Shepard plots: (a) 2D; (b) 3D. The cosine correlation c_{ij}^c is used and the period of analysis is 1951–2010.

$$S = [d_{ij} - f(\delta_{ij})]^2 \quad (1)$$

where d_{ij} stands for the reproduced distances, given the respective number of dimensions, and δ_{ij} represents the input data (i.e., the observed distances). The expression $f(\delta_{ij})$ indicates a transformation of the input data. Therefore, the smaller the stress, S , the better is the fit between the observed and the reproduced matrices.

It should be noted that the actual orientation of axes in the final solution is arbitrary. We can rotate the MDS map and the distances between points remain the same. Identically, the MDS map is insensitive to translations. Therefore, the final orientation of axes in the space is to be decided by the researcher that chooses the one that produces a clearer visualisation.

For accessing the quality of the MDS map are adopted the stress and Shepard plots. The stress plot, that is, of S versus the number of dimensions k of the MDS representation, leads to a monotonic decreasing chart. Therefore, we choose the “best” dimension k as a compromise between stress reduction and number of dimensions for the MDS map. The Shepard diagram consists of comparing the reproduced distances, for a particular value k of the MDS dimension, versus the observed input data (distances). Therefore, in the Shepard diagram, a narrow scatter around a 45 degree line indicates a good fit of the distances to the dissimilarities, while a large scatter indicates a lack of fit, for that value k .

The measure for constructing matrix C depends of the researcher’s choice and, therefore, we can have multiple criteria that can prove to be “better” or “worse” depending on the MDS resulting map. Finally, the MDS interpretation is based on the emerging clusters and relative distances, rather than on the absolute coordinates or shapes.

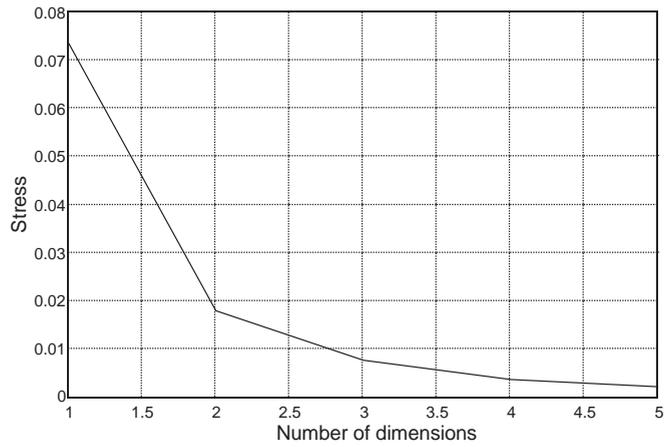


Fig. 11. Stress plot. The cosine correlation c_{ij}^c is used and the period of analysis is 1951–2010.

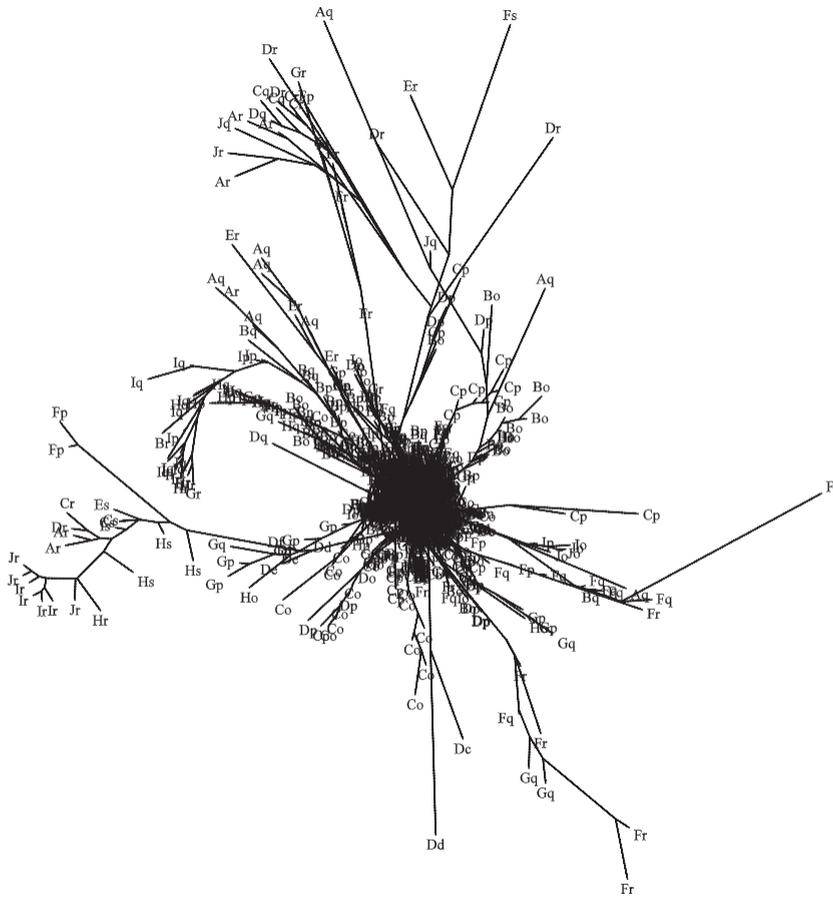
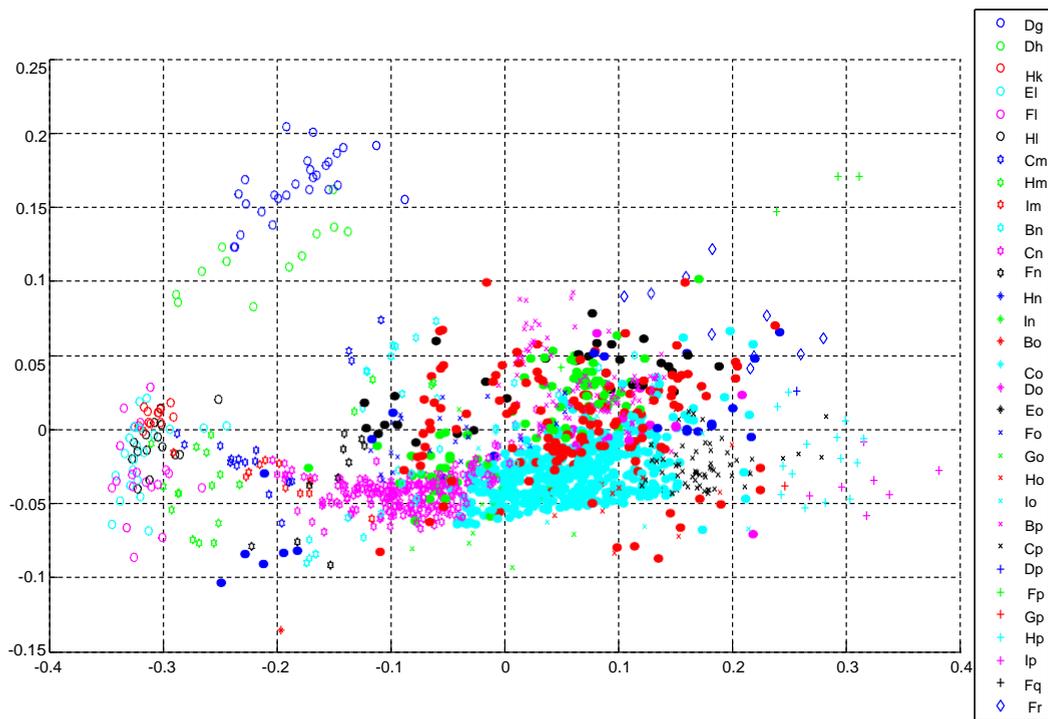


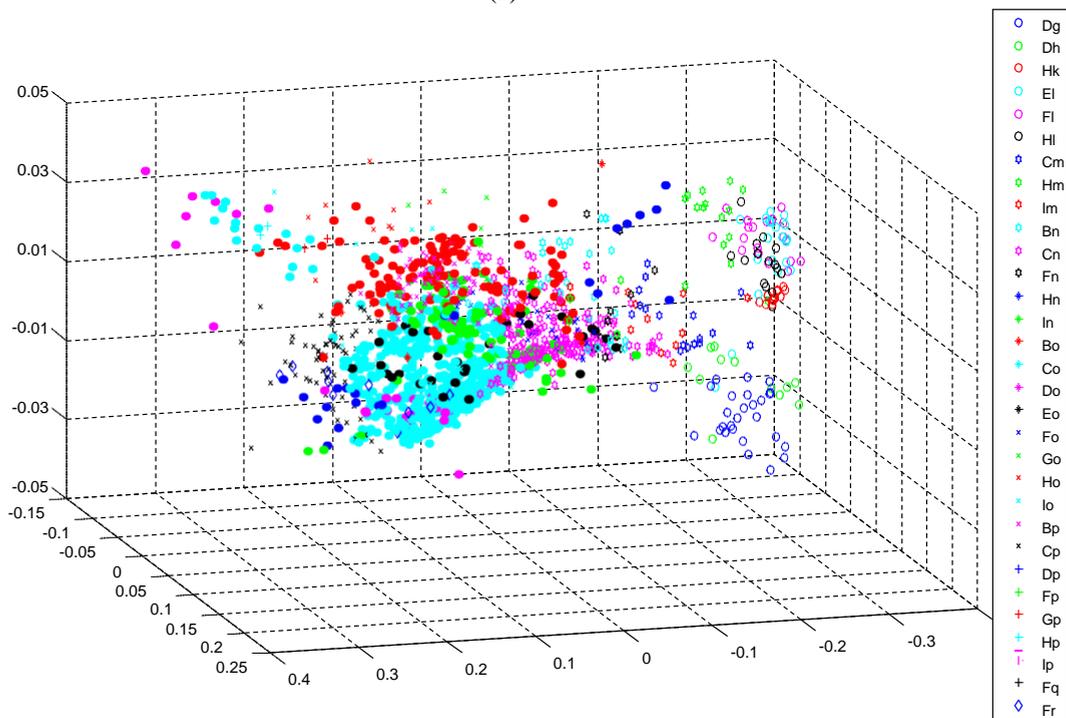
Fig. 12. Clustering tree based on the distance matrix $1 - c_{ij}^c$

4. Multidimensional scaling analysis of real data

This section begins by briefly characterising the dataset used in this study. After that, we apply the MDS technique to the data. This representation cannot be interpreted as better, or worse, than the classical one. Otherwise, the MDS map is meaningful in the sense that provides visualisation based on different concepts. Therefore, it is advantageous to use it as complementary to standard approaches.



(a)



(b)

Fig. 13. MDS maps: (a) 2D; (b) 3D. The Euclidean distance d_{ij}^d is used and the period of analysis is 1951–2010. Each point represents a meteorological station.

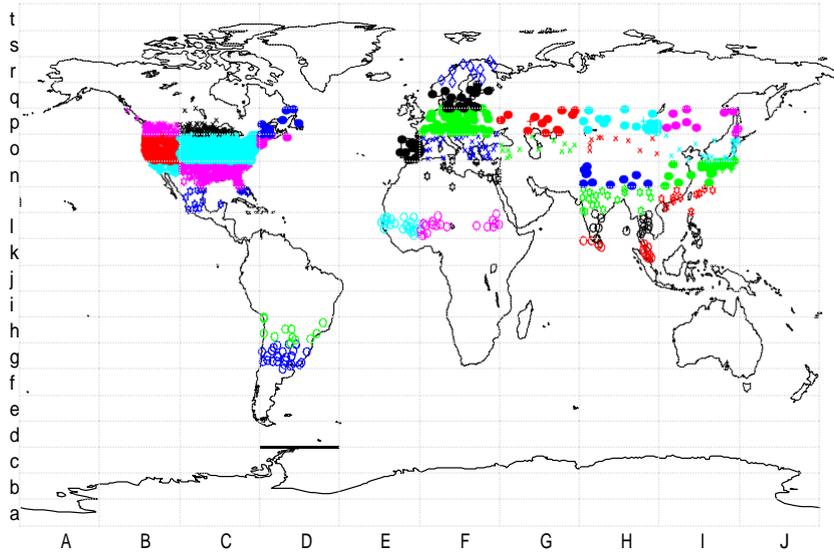


Fig. 14. Location of the stations on a geographic map. The Euclidean distance d_{ij}^d is used and the period of analysis is 1951–2010. The same symbols adopted for the MDS maps are used.

Two indices that correlate the monthly average temperatures observed in meteorological stations, over a certain period of time, are used to quantify the similarities among geographic regions. Therefore, in subsection 4.2, the cosine correlation index is considered over a six-decade period, which corresponds to a compromise between time and spatial data coverage. In subsection 4.3 we use the Euclidean distance correlation index. Firstly, the same 60-year period is considered. Then, to increase the number of stations and geographic coverage, a shorter time period of 30 years is analysed. In subsections 4.4 and 4.5 time dynamics is analysed by performing the MDS analysis over slices sampling the stations' time series.

4.1. Brief description of the dataset

In this study the Global Historical Climatology Network-Monthly (GHCN-M), version 3 dataset of monthly mean temperature [19], available at the National Oceanic and Atmospheric Administration, National Climatic Data centre (NOAA-NCDC) (<http://www.ncdc.noaa.gov/ghcnm/v3.php>) is used. The data was retrieved on the 29th of March, 2012. The archive contains temperature records from 7280 meteorological stations located almost exclusively on land areas (Fig. 4a). The number of stations with long records is quite moderate and those are essentially located on the northern hemisphere (United States, Europe, India and Japan) (Fig. 4b). Fig. 5 shows a cumulative histogram depicting the number of meteorological stations with data records longer than a certain number of years.

Each data record consists of the monthly average temperatures of a meteorological station. Some occasional gaps of one month in the data (represented on the original data by the value -9999) are substituted by a linear interpolation between the two adjacent values.

Fig. 6 depicts the time evolution of the monthly average temperature of one typical station. We have chosen Lisbon, Portugal (Lat 38.7 N, Lon 9.2 W). Analysing the temperature time-series, three processes are visible, namely (i) a continuous, almost linear, temperature increase, (ii) an annual periodic variation and (iii) a 'random' temperature variation that is the symptom of a rich dynamical behaviour. Applying the Fourier transform to the time domain signal, the amplitude spectrum depicted in Fig. 7 is obtained. The peak at the angular frequency $\chi = 1.99 \times 10^{-7}$ rad/s corresponds to the periodicity of one year. Moreover, at low frequencies the spectrum can be approximated by a power law with amplitude $160.8\chi^{-0.875}$. While frequency domain analysis could be used to characterise every station, long time-series are needed in order to have accurate results. As mentioned before, only few stations meet this condition, suggesting that alternative analyses are necessary.

4.2. MDS analysis based on cosine correlation index

In the sequel we apply the MDS method to the time correlation between the station's monthly average temperatures. Depending on the stations, the data is available for different periods of time. Thus, the period from January 1951 up to December 2010 is adopted in the analysis. For this 60-year period the number of stations with valid data is $s = 1751$. This option is a compromise between spatial and temporal data volatility. Choosing longer periods of analysis results in fewer stations, covering just restricted geographic areas, whereas opting for shorter periods of time makes difficult capturing the long-term dynamics of the involved phenomena.

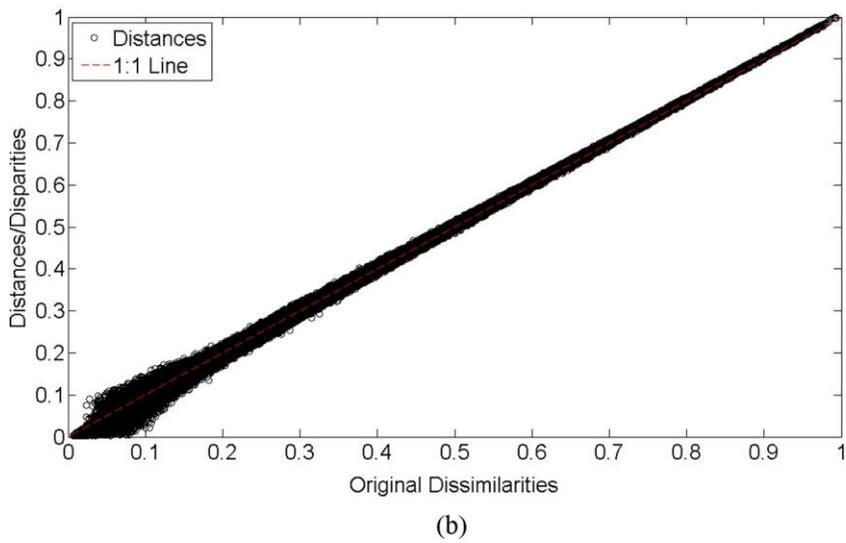
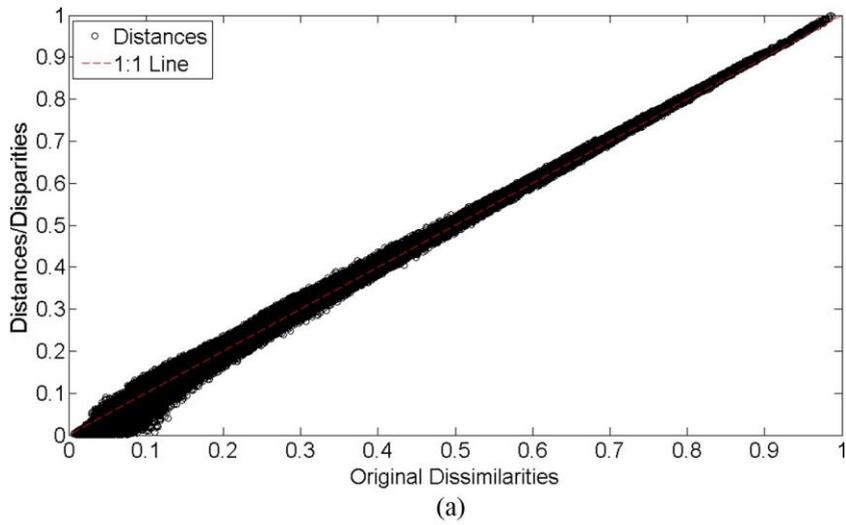


Fig. 15. Shepard plots: (a) 2D; (b) 3D. The Euclidean distance c_{ij}^d is used and the period of analysis is 1951–2010.

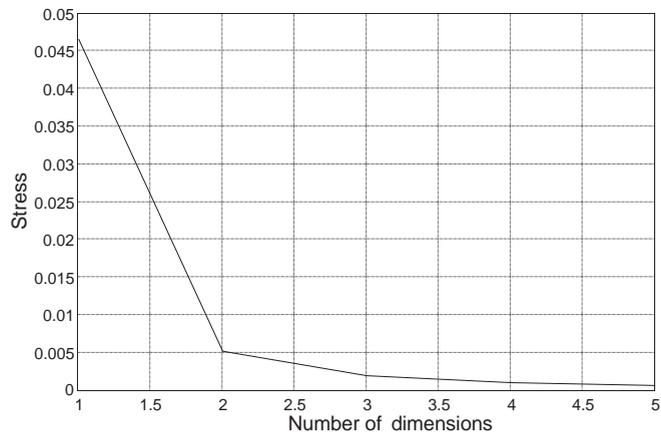
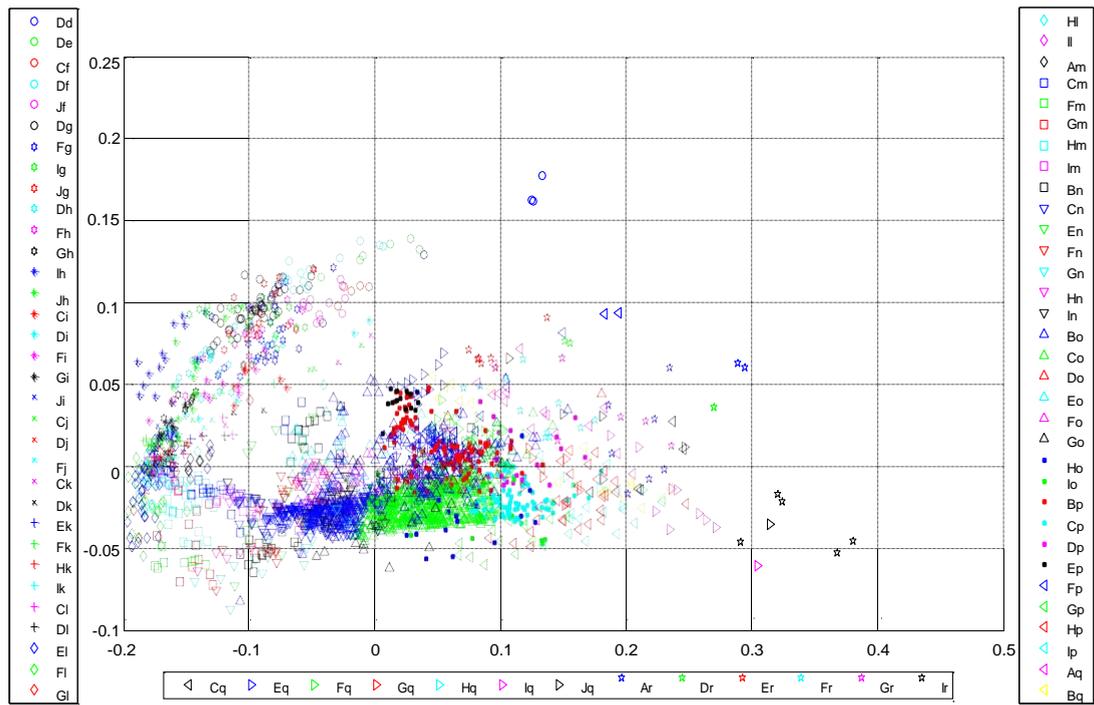
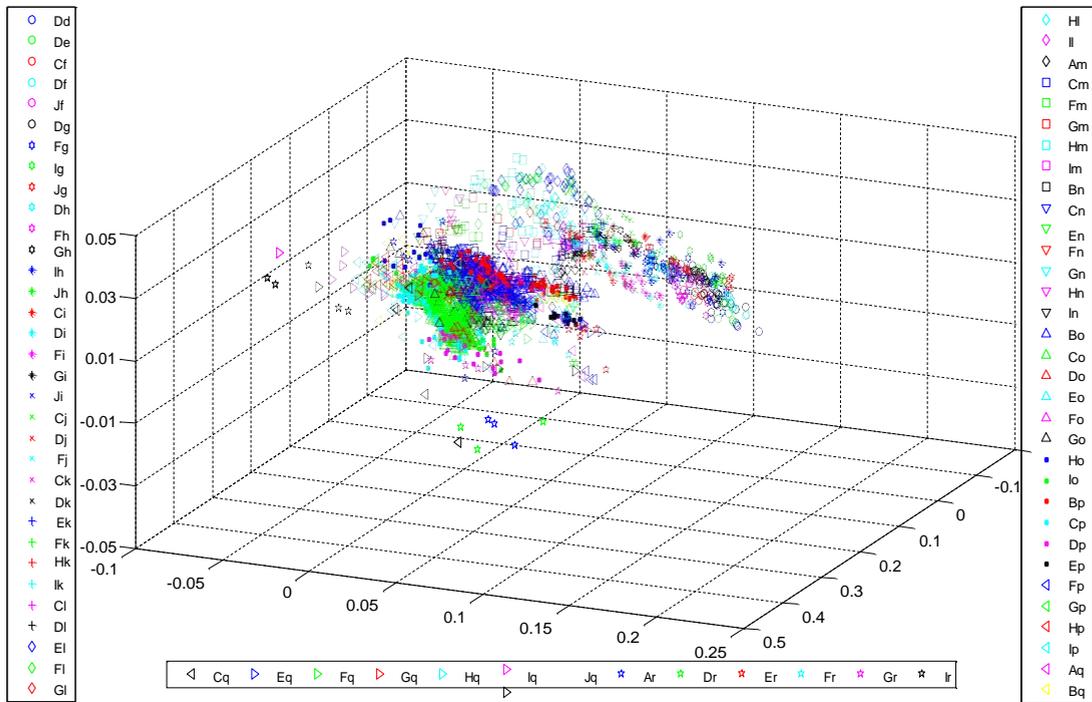


Fig. 16. Stress plot. The Euclidean distance c_{ij}^d is used and the period of analysis is 1951–2010.



(a)



(b)

Fig. 17. MDS maps: (a) 2D; (b) 3D. The Euclidean distance c_{ij}^d is used and the period of analysis is 1971–2000. Each point represents a meteorological station.

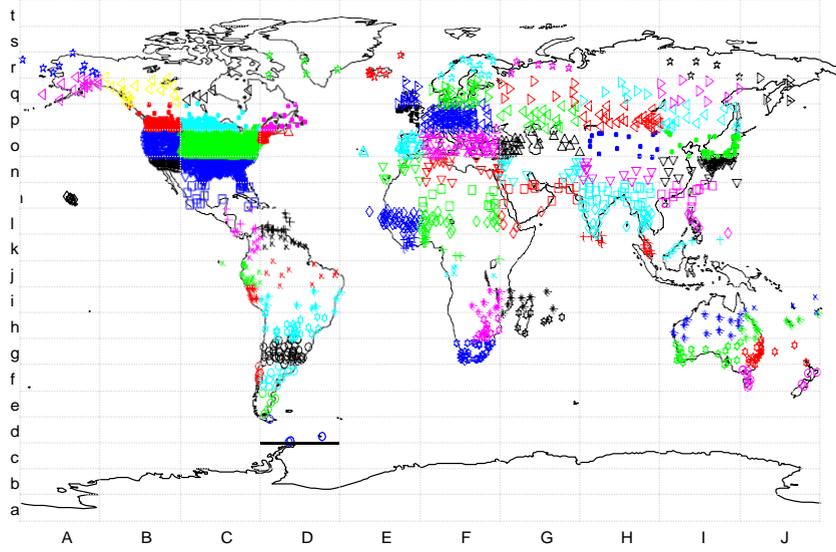


Fig. 18. Location of the stations on a geographic map. The Euclidean distance c_{ij}^E is used and the period of analysis is 1971–2000. The same symbols adopted for the MDS maps are used.

Firstly we adopt a correlation index, c_{ij} , representing a measurement of the similarity between stations i and j ($i, j = 1, \dots, s$). Then, based on that index, an $s \times s$ matrix, C , is computed. Finally, in order to reveal possible relationships between the stations under analysis, the MDS technique is used. In this perspective, while 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 optimise a cost function that describes how well the pairwise distances in a data set are preserved [50,51].

We begin the analysis using the cosine correlation index, mathematically given by Eq. (2):

$$c_{ij}^c = \left(\frac{\sum_{t=1}^p x_i(t) \cdot x_j(t)}{\sqrt{\sum_{t=1}^p x_i^2(t) \cdot \sum_{t=1}^p x_j^2(t)}} \right)^2 \quad (2)$$

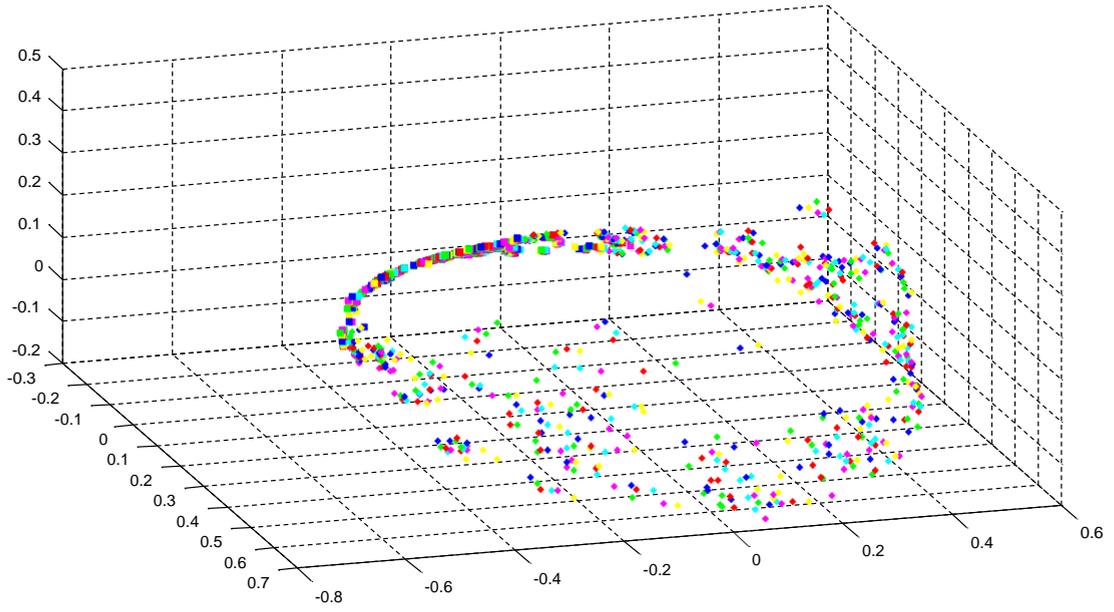
where t represents time in months and P is the total number of months in the analysed period, i.e., $p = 50 \times 12$ months.

Fig. 8 depicts the 2D and 3D locus of the stations given by MDS. Each point on the map represents a meteorological station. The symbol's code with two letters, displayed in the legend, corresponds to a cell of Fig. 9. As we are dealing with a large number of stations, it is somewhat difficult to analyse the results. To facilitate the comparison, a grid is defined based on the station's latitude/longitude location and identical (different) symbols are given to stations belonging to the same (different) cell(s). Moreover, stations belonging to cells that have less than a certain threshold number of stations are not represented. In Fig. 9 the station's locus is given on a geographic map using the same symbols adopted for the MDS maps.

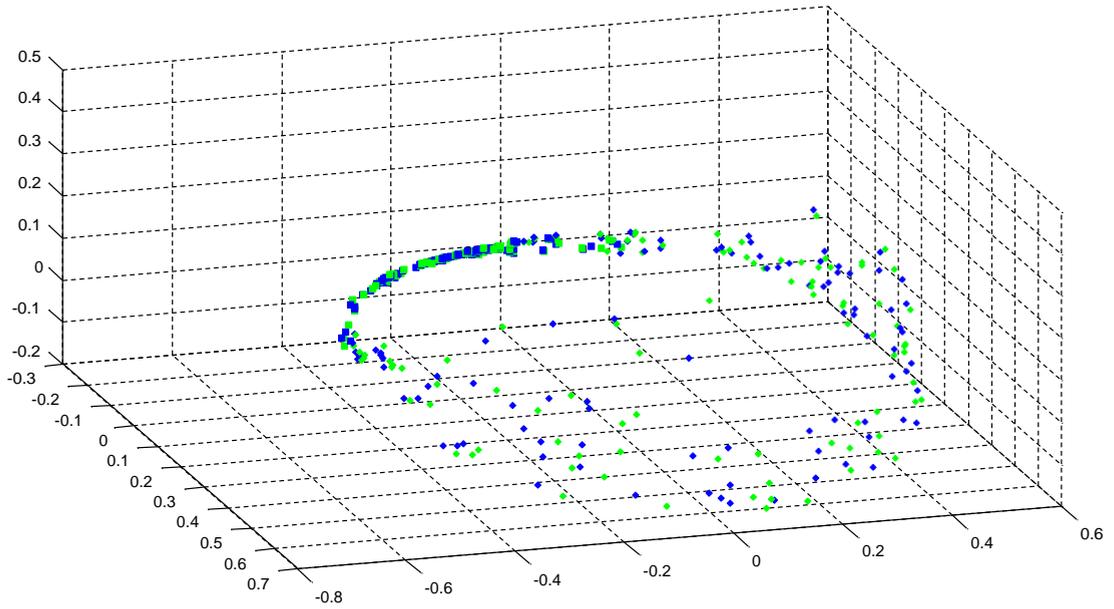
Looking at Fig. 8, we can notice that instead of clusters on the MDS maps, the stations form a large and contiguous cloud of points. This suggests interpreting the results in terms of similarities and dissimilarities among geographic areas, namely (i) the regions {Dg, Dh, El, Fl, Hk, Hl, Hm, Im, Cm} are similar to each other, which means that a large area in south America, central Africa, India plus south-eastern Asia and central Mexico plus Cuba present climatic similarities; (ii) on the opposite side (dissimilar) to the previous group are the regions {Cp, Dp, Gp}, corresponding to the northern USA and western Kazakhstan; (iii) the areas {Bo, Bp} are quite similar, representing western USA, and spreading along a vast area on the MDS map; (iv) similar to each other are also {Eo, Fn, Hn}, corresponding to the Mediterranean region, and (v) {Fp, Fq}, corresponding to northern and central Europe.

Figs. 10 and 11 depict the Shepard and the stress plots for the MDS, respectively. The Shepard plot shows a reasonable distribution of points around the 45 degree line, particularly when the dimensionality is three. The stress plot reveals that a three dimensional space describes well the locus of the s stations. In fact the stress diminishes strongly until the dimensionality is two, moderately towards dimensionality three and weakly from then on.

To complement the analysis, we adopted a standard clustering algorithm for comparing results. The phylogenetic analysis free software PHYLIP (<http://evolution.genetics.washington.edu/phylip.html>) was used. Based on the distance matrix $1 - c_{ij}^c$ a clustering tree was generated by successive (agglomerative) clustering, using an average-linkage method (Fig. 12). The leaves represent meteorological stations and are named according to the geographic grid described above (i.e., leaves with the same name belong to the same cell). While some small clusters can be noticed on the external branches, most stations concentrate in the middle of the tree, being almost impossible to identify. This also illustrates the superiority of MDS as a clustering and visualisation tool, mainly when dealing with large data sets.



(a)



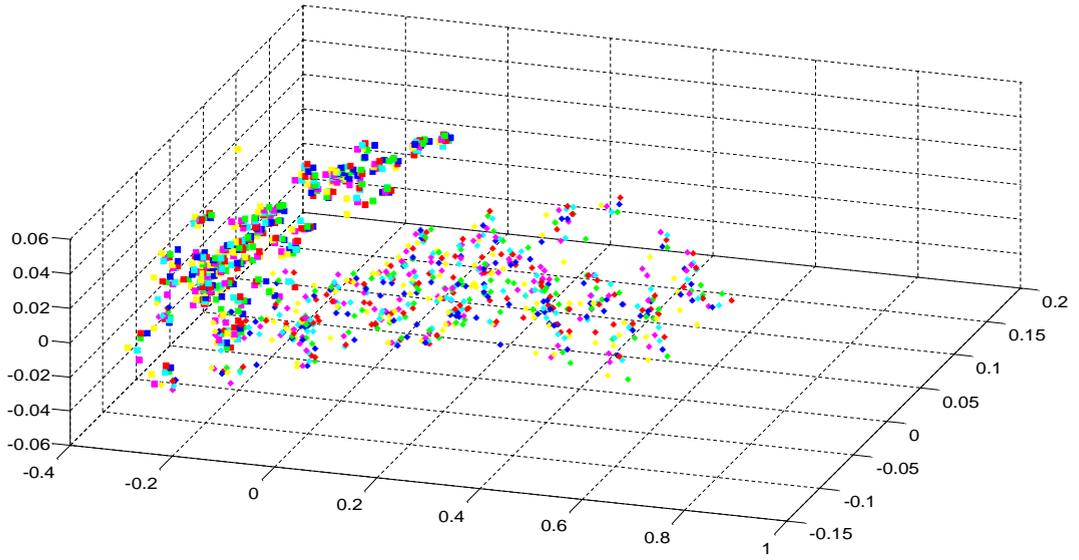
(b)

Fig. 19. MDS maps: (a) $l = 10$ years; (b) $l = 30$ years. The cosine correlation c_{ij}^c is used and the period of analysis is 1951–2010. Each point represents a station/slice object.

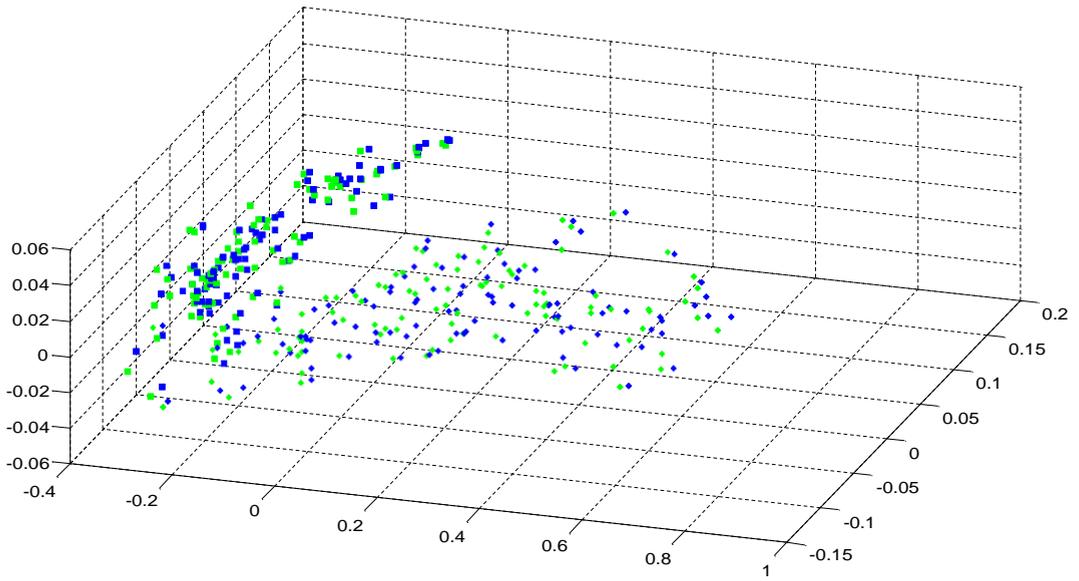
4.3. MDS analysis based on Euclidean distance correlation index

In order to complement the analysis, a different correlation index based on the Euclidean distance is tested. Mathematically, this index, c_{ij}^d is defined by expression (3):

$$d_{ij} = \sum_{t=1}^p (x_i(t) - x_j(t))^2 \quad (3a)$$



(a)



(b)

Fig. 20. MDS maps: (a) $l = 10$ years; (b) $l = 30$ years. The Euclidean distance c_{ij}^d is used and the period of analysis is 1951–2010. Each point represents a station/slice object.

$$c_{ij}^d = 1 - \frac{d_{ij}}{\max(d_{ij})} \quad (3b)$$

The 2D and 3D locus of the stations resulting from the MDS analysis are represented in Fig. 13. The location of the stations on a geographic map is shown in Fig. 14.

Fig. 13 suggests that the stations are organised on the MDS map forming four main clusters: (i) a big central cluster; (ii) a second cluster composed by the regions {Dg, Dh}; (iii) the third cluster includes the regions {Hk, El, Fl, Hl, Hm}; (iv) the fourth cluster comprises the regions {Gp, Hp, Ip}.

Taking a closer look at the clusters, we see that the cells {Dg, Dh}, on the second cluster, correspond to two neighbour regions of south America. Therefore, it is not surprising that they appear similar and near to each other on the MDS map. Regarding the third cluster, {Hk, Hl, Hm} represent a contiguous geographical area in India and south-eastern Asia. As before, they are near to each other on the MDS map. However, {El, Fl}, belonging to the third cluster, are located in Africa far away from {Hk, Hl, Hm}. The fourth cluster includes {Gp, Hp, Ip}, which correspond to a large region of Asia. The same latitude of the three regions might explain the observed similarities.

Looking at the MDS map from the point of view of dissimilarities, we can notice that the regions {Fp, Gp} are far from each other on the MDS, although they are geographically neighbours, corresponding to central Europe and eastern Europe (including parts of Russia, Ukraine and Kazakhstan), respectively. The same can be said with reference to {Bp} and {Cp, Dp}, which represent the north-western and north-eastern USA.

We should note that MDS is merely a mathematical visualisation tool and that a physical perspective of the reported results must be found in the light of the measuring index. For instance, certain neighbouring geographical regions, such as the north-western and north-eastern US appear far in the MDS map. Such result means merely that the cosine correlation finds the corresponding time series reasonably dissimilar within the total set of meteorological stations. No latitude or longitude concepts are included. Therefore, a further explanation about physical mechanisms associated with this disparity must be envisaged by standard complementary procedures.

The Shepard plots depicted in Fig. 15 shows that a good distribution of points around the 45 degree line is obtained. This distribution is even better than the one obtained for the cosine correlation index. The stress plot (Fig. 16) shows, as before, that a three dimensional space describes well the locus of the s stations.

While the cosine correlation and the Euclidean distance indices proved to be adequate for representing the similarities between the stations' temperature time-series, the latter is preferable as it leads to an easier interpretation of the MDS maps.

We consider now the analysis of a shorter time period of 1971–2000. In this case, the number of stations with valid data grows up to 2641 (which means a 50% increasing relative to the previous case), covering vast areas of the planet. It should be noted that three decades is usually the period considered by the meteorologists as the minimum necessary to capture the long-term weather variability. Therefore, the data is more volatile and the analysis is more affected by short-term periodic or irregular variations.

In this experiment we just use the Euclidean distance index as a measurement of the similarity between stations, as it gave superior results previously.

Figs. 17 and 18 show the MDS maps and the geographic location of the stations. Are not included the Sheppard and stress graphs as they are of the same type as those represented for the previous case. The results can be interpreted as before for Figs. 13 and 14. In this case we have (i) a large main cluster, mostly incorporating geographical areas located on the northern hemisphere (United States, Europe, India and Japan), (ii) several new small clusters that are located away from the border of the main one, and (iii) one new cluster that is appended to the main cluster.

We now have more stations than in Figs. 13 and 14, covering a larger region of Earth, namely areas at higher latitudes that correspond to colder regions (northern Canada and northern Europe and Asia). This explains the clusters mentioned in (ii). In fact, these points, belonging to colder regions, are clearly dissimilar from those of the main cluster. Identically, it is also justifiable the presence of the cluster (iii), appended to the main one. These points represent stations located on the southern hemisphere, essentially, on warm regions. This means that these regions are dissimilar from those that compose the main cluster.

In conclusion, both MDS maps visualise the stations forming comprehensive patterns that are easily interpretable under the light of similarities and surpassing the problems posed by the variable density of measuring stations along Earth's surface. These results may constitute a first step towards the formulation of preview methods based on the MDS scheme.

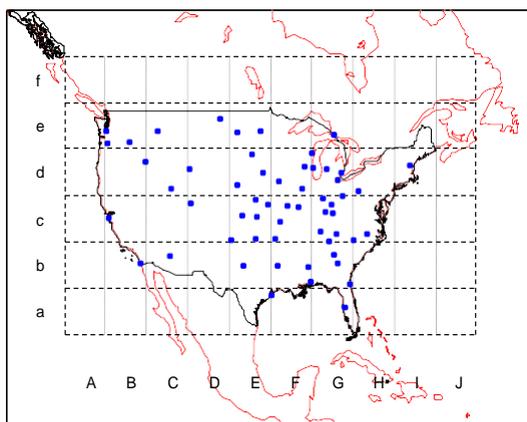
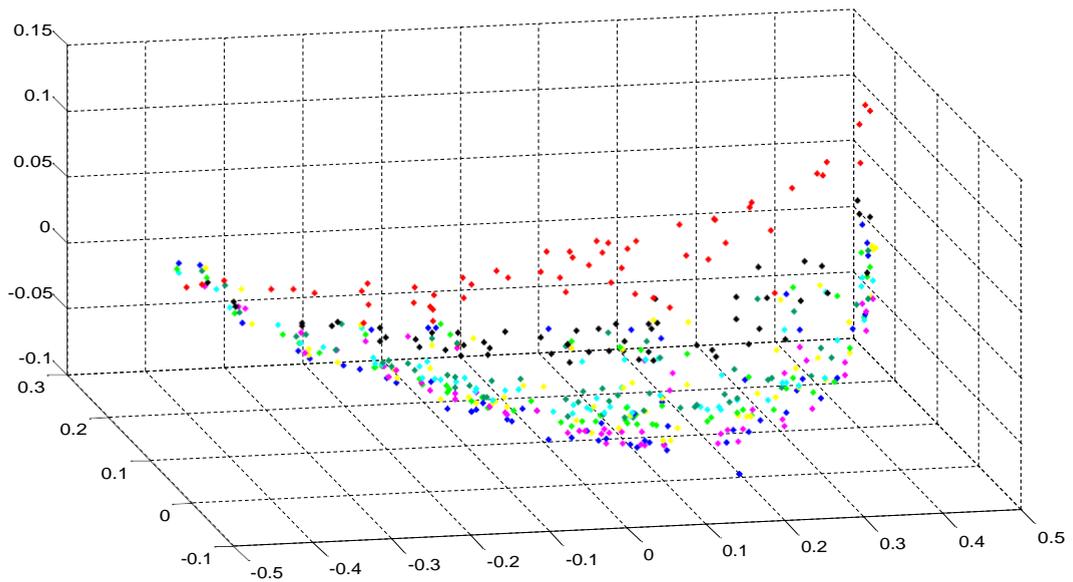
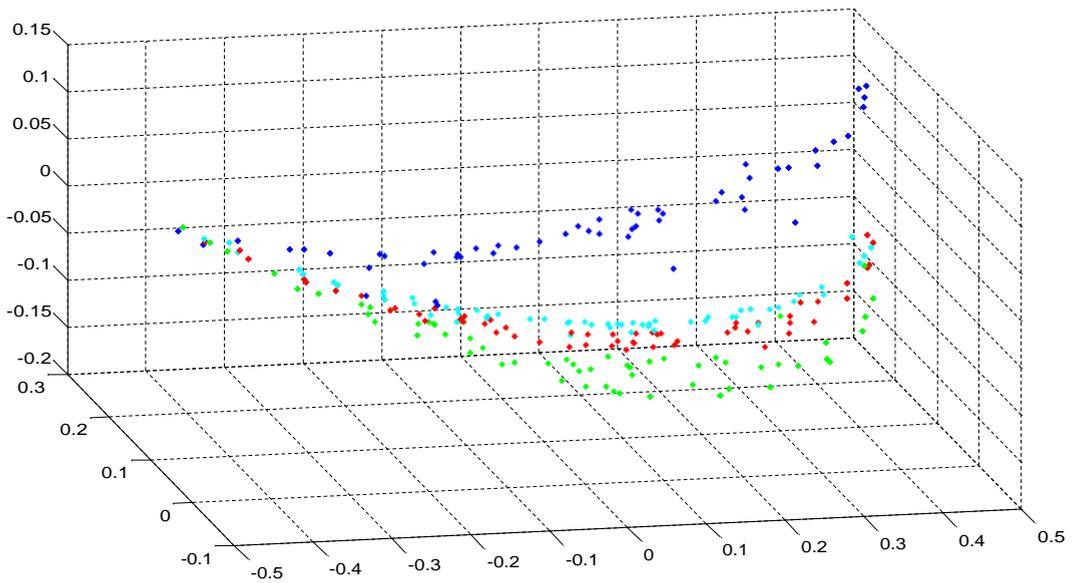


Fig. 21. United States meteorological stations with data series from 1891 up to 2010.



(a)

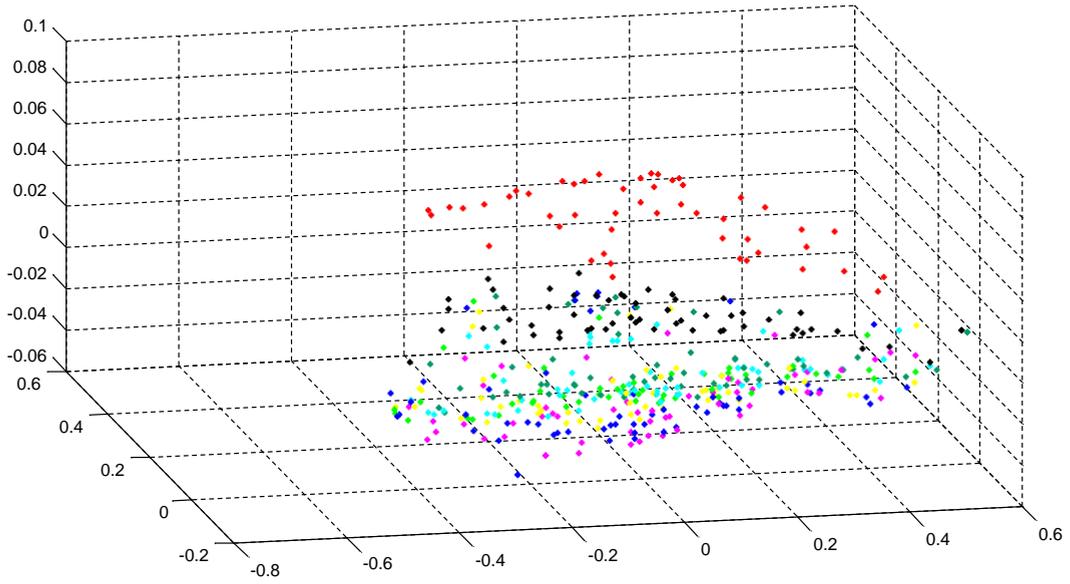


(b)

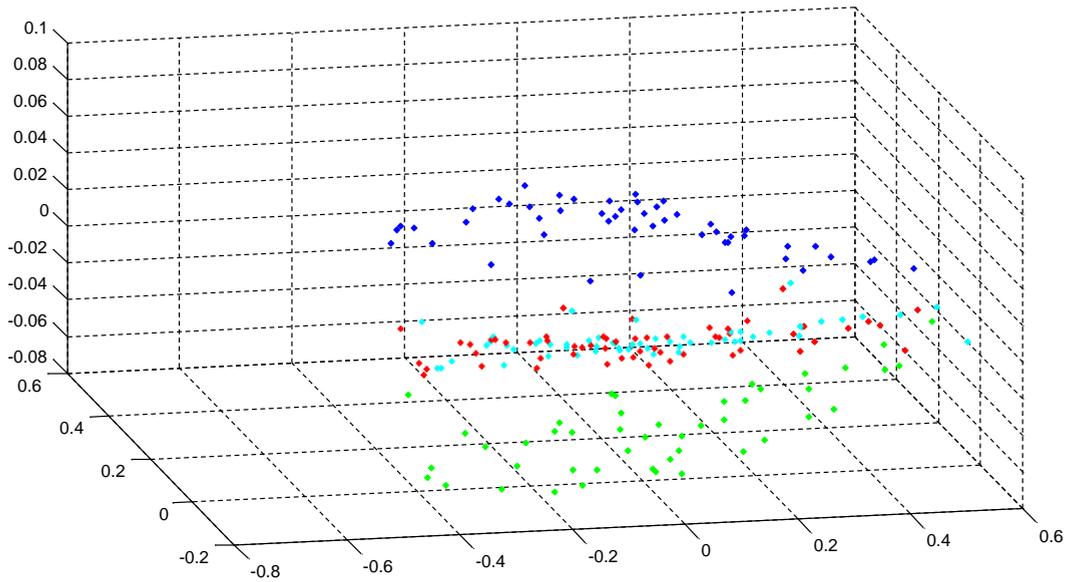
Fig. 22. MDS maps: (a) $l = 15$ years; (b) $l = 30$ years. The cosine correlation c_{ij}^c is used and the period of analysis is 1891–2010. Each point represents a station/slice object.

4.4. MDS analysis of world data and time dynamics

For the purpose of capturing time dynamics, the original data series of length P , corresponding to the meteorological stations, are “sliced” into n series of length $l = P/n$ each. The number of stations times the number of slices, $s \times n$, constituting now the objects for the MDS. Both periods 1951–2010 and 1971–2000 were analysed and several slice lengths, l , were adopted. A subset of the available stations was selected in order to keep the number of objects, $s \times n$, suitable for the MDS analysis. The stations were chosen distributed across the Earth. In this section we present the analysis for the 60-year period 1951–2010, as for 1971–2000 the results are similar. In the sequel only MDS maps are represented and Shepard and stress maps are not included since they do not add relevant new information.



(a)



(b)

Fig. 23. MDS maps: (a) $l = 15$ years; (b) $l = 30$ years. The Euclidean distance d_{ij}^d is used and the period of analysis is 1891–2010. Each point represents a station/slice object.

Fig. 19a represents the MDS map obtained using the cosine correlation index, defined in Eq. (2). The adopted series length is $l = 10$ years, corresponding to $n = 6$ time slices. A subset of $s = 193$ meteorological stations was considered, 99 above and 94 below the tropic of Cancer (Lat 23.5° , approximately). The total number of objects for the MDS totalizes 193×6 points. The tropic of Cancer separates the objects (stations/slices) into two clusters, the colours corresponding to the time slices. Since the objects below the tropic (square marks) are very close to each other, they belong to one cluster that can be characterised by weak time dynamics. On the contrary, objects above the tropic (diamond marks), are distant from each other and, therefore, unveil a strong variability in time; nevertheless, no clear pattern (e.g., periodicity) related to time dynamics emerges from the MDS map. The northern stations appear on the MDS map on the right side, whereas stations closer to the tropic are mainly located on the left side of the graph.

Fig. 19b depicts the MDS map for $l = 30$ years. In this case, no more than $n = 2$ time slices can be used. Both maps are similar, but no further conclusions can be taken.

Fig. 20a and b depict the MDS maps obtained using the Euclidean distance correlation index, defined by Eq. (3), and series lengths $l = 10$ and $l = 30$ years, respectively. The results are identical to those obtained with the cosine correlation index. Stronger time variability is observed for northern regions, which now appear on the bottom-right area of the MDS map, distant apart from the southern objects.

4.5. MDS analysis of US data and time dynamics

The analysis shown in the previous subsection reveals that the time variability is difficult to capture in a global perspective, being strongly constrained by the large number of points. To overcome this issue, in the sequel of this subsection we study a smaller region, corresponding to the continental United States (without Alaska). For this region, as there are many meteorological stations with long data records, several long slices of data can be extracted.

In the analysis, all the 57 stations with data records from 1891 up to 2010 were considered (Fig. 21). The length of slices was varied within the set $l = \{10, 15, 20, 24, 30, 40\}$ years, corresponding to $n = \{12, 8, 6, 5, 4, 3\}$ slices.

The result obtained using the cosine correlation is depicted in Fig. 22, where we observe graphical artefacts with “C” shape for each time period. Furthermore, the distinct colours represent the different time slices. Fig. 22a illustrates the case of 15-year slices. In this case, each 120-year record was sliced into 8 smaller series, being the total number of objects (stations/slices) 57×8 points. Fig. 22b depicts the MDS map of the stations/slices, now adopting 4 slices of 30-year each, which gives 57×4 objects. As can be noticed, four clusters emerge, each representing a time slice.

Similar results can be obtained using the Euclidean distance correlation index, as shown in Fig. 23.

In this case we observe graphical artefacts with “I” shape for each time period that, as previously, is represented with a distinct colour.

The “C” and “I” artefacts resulting for c_{ij}^c and c_{ij}^d , respectively, have no special meaning, as usually occurs in MDS analysis, being better or worse merely in the perspective of providing a better visualisation. However, we observe in both cases that we have, for each time slice, an evolution along the MDS object according with the latitude of the stations lying in the {E, F, G} cells. The rest of the stations are represented as separate points. Therefore, we verify that geographical information is embedded implicitly, but, again, with logic distinct from the one usual in classic meteorological portraits.

In conclusion, the MDS maps can complement standard geoscience tools and lead to a deeper understanding of all information involved.

5. Conclusions

Climate change is a major concern nowadays. Although several indicators of climate change can be used, the time evolution of Earth’s surface temperature is probably the easiest to understand. Temperature time-series characterise Earth as a slow dynamics spatiotemporal system, evidencing long memory behaviour and complex relationships. Such phenomena, typical of fractional order systems, are difficult to model and analyse, demanding for alternative approaches to those usually adopted in dynamical systems.

In this paper the Multidimensional scaling (MDS) analysis was proposed to visualise the climatic similarities among geographical regions of Earth. Temperature time-series registered by meteorological stations and available on-line at <http://www.ncdc.noaa.gov/ghcnm/v3.php> were used. Two time correlation indices to quantify the similarities between stations were adopted: the cosine correlation and the Euclidean distance indices. While both indices are suitable for representing similarities, the Euclidean distance index was found to be superior to the cosine correlation, as it leads to an easier interpretation of the MDS maps. MDS maps were proven as an intuitive and useful visual representation of the complex relationships that are present among temperature time-series, which are not perceived on traditional geographic maps.

At a different level this paper proposed also a new strategy for attacking the study of complex system. While the classical procedure is to model the phenomena and, based on the set of equations, to analyse their properties, the proposed method reformulates the scientific paradigm, by defining objects in a new space representation. The techniques and methods usual in dynamical systems are then productive shedding light to aspects untouched by more obvious direct approaches.

References

- [1] America’s Climate Choices. Board on atmospheric sciences and climate, division on earth and life studies, national research council of the national academies america’s climate choices. Washington, DC: The National Academies Press; 2011.
- [2] Meehl GA, Stocker TF, Collins WD, et al. Global climate projections. In: Solomon S, Qin D, Manning M, et al., editors. Climate change 2007: the physical science basis. Contribution of working group i to the fourth assessment report of the intergovernmental panel on climate change. Cambridge, United Kingdom and New York, USA: Cambridge University Press; 2007.
- [3] America’s Climate Choices. Panel on advancing the science of climate change, board on atmospheric sciences and climate, division on earth and life studies, national research council of the national academies, advancing the science of climate change. Washington, DC: The National Academies Press; 2010.
- [4] Alexiadis A. Global warming and human activity: a model for studying the potential instability of the carbon dioxide/temperature feedback mechanism. *Ecol Modell* 2007;203:243–56.
- [5] Allen MR, Frame DJ, Huntingford C, et al. Warming caused by cumulative carbon emissions towards the trillionth tonne. *Nature* 2009;458:1163–6.

- [6] IPCC, Climate Change 2001: Working group I – The Scientific Basis, <http://www.grida.no/publications/other/ipccetar/>; 2012 [accessed 1.04.12].
- [7] Dai A. Drought under global warming: a review. *Wiley Interdiscip Rev Clim Change* 2011;2:45–65.
- [8] You QL, Kang SC, Pepin N, et al. Climate warming and associated changes in atmospheric circulation in the eastern and central Tibetan Plateau from a homogenized dataset. *Global Planet Change* 2010;72:11–24.
- [9] Jevrejeva S, Moore JC, Grinsted A. Sea level projections to AD2500 with a new generation of climate change scenarios. *Global Planet Change* 2012;80–81:14–20.
- [10] Giannakopoulos C, Le Sager P, Bindi M, et al. Climatic changes and associated impacts in the Mediterranean resulting from a 2 C global warming. *Global Planet Change* 2009;68:209–24.
- [11] Chase TN, Pielke Sr RA, Kittel TGF, et al. Simulated impacts of historical land cover changes on global climate in northern winter. *Clim Dyn* 2000;16:93–105.
- [12] Oerlemans J, Anderson B, Hubbard A, et al. Modelling the response of glaciers to climate warming. *Clim Dyn* 1998;14:267–74.
- [13] Zhu Y, Wang H, Zhou W, Ma J. Recent changes in the summer precipitation pattern in East China and the background circulation. *Clim Dyn* 2011;36:1463–73.
- [14] Hansen J, Ruedy R, Sato M, Lo K. Global surface temperature change. *Rev Geophys* 2010;48:RG4004.
- [15] Brohan P, Kennedy JJ, Harris I, Tett SFB, Jones PD. Uncertainty estimates in regional and global observed temperature changes: a new dataset from 1850. *J. Geophys. Res.* 2006;111:D12106.
- [16] Menne MJ, Williams Jr CN. Homogenization of temperature series via pairwise comparisons. *J Clim* 2009;22:1700–17.
- [17] Wu Z, Huang NE, Wallace JM, et al. On the time-varying trend in global-mean surface temperature. *Clim Dyn* 2011;37:759–73.
- [18] Rohde R, Brillinger D, Curry J et al., Berkeley earth temperature averaging process. Submitted to *Econometrics*, 2011.
- [19] Lawrimore JH, Menne MJ, Gleason BE, et al. An overview of the Global Historical Climatology Network monthly mean temperature data set, version 3. *J Geophys Res* 2011;116:D19121.
- [20] Jones PD, Lister DH, Osborn TJ, et al. Hemispheric and large-scale land surface air temperature variations: An extensive revision and an update to 2010. *J Geophys Res* 2012. <http://dx.doi.org/10.1029/2011JD017139>.
- [21] <http://www.berkeleyearth.org/>; 2012 [accessed 10.04.12].
- [22] Capilla C. Time series analysis and identification of trends in a Mediterranean urban area. *Global Planet Change* 2008;63:275–81.
- [23] Grieser J, Trömel S, Schönwiese C-D. Statistical time series decomposition into significant components and application to European temperature. *Theor Appl Climatol* 2002;71:171–83.
- [24] Hughes G, Rao S, Rao T. Statistical analysis and time-series models for minimum/maximum temperatures in the Antarctic Peninsula. *Proc R Soc Ser A* 2007;463:241–59.
- [25] Viola FM, Paiva SLD, Savi MA. Analysis of the global warming dynamics from temperature time series. *Ecol Modell* 2010;221:1964–78.
- [26] Founda D, Papadopoulos KH, Petrakis M, Giannakopoulos C, Good P. Analysis of mean, maximum, and minimum temperature in Athens from 1897 to 2001 with emphasis on the last decade: trends, warm events, and cold events. *Global Planet Change* 1897;44(2004):27–38.
- [27] Ge Q-S, Zheng J-Y, Hao Z-X, et al. Temperature variation through 2000 years in China: an uncertainty analysis of reconstruction and regional difference. *Geophys Res Lett* 2010;37:L03703.
- [28] Deser C, Phillips AS, Alexander MA. Twentieth century tropical sea surface temperature trends revisited. *Geophys Res Lett* 2010;37:L10701.
- [29] Samko S, Kilbas A, Marichev O. Fractional integrals and derivatives: theory and application. London: Gordon and Breach Science Publishers; 1993.
- [30] Mainardi F. Fractional calculus and waves in linear viscoelasticity: an introduction to mathematical models. London: Imperial College Press; 2010.
- [31] Tenreiro Machado JA, Kiryakova V, Mainardi F. Recent history of fractional calculus. *Commun Nonlinear Sci Numer Simul* 2011;16:1140–53.
- [32] Tenreiro Machado JA, Costa António C, Quelhas Maria Dulce. Fractional dynamics in DNA. *Commun Nonlinear Sci Numer Simul* 2011;16:2963–9.
- [33] Tenreiro Machado JA, Duarte Fernando B, Monteiro Duarte Gonçalo. Analysis of stock market indices through multidimensional scaling. *Commun Nonlinear Sci Numer Simul* 2011;16:4610–8.
- [34] Polzella D, Reid G. Multidimensional scaling analysis of simulated air combat manoeuvring performance data. *Aviat Space Environ Med* 1989;60:141–4.
- [35] Costa AM, Machado JA, Quelhas MD. Histogram-based DNA analysis for the visualization of chromosome, genome and species information. *Bioinformatics* 2011;27:1207–14.
- [36] Machado JT, Duarte GM, Duarte FB. Identifying economic periods and crisis with the multidimensional scaling. *Nonlinear Dyn* 2011;63:611–22.
- [37] Glunt W, Hayden TL, Raydan M. Molecular conformation from distance matrices. *J Comput Chem* 1993;14:114–20.
- [38] Tenenbaum J, de Silva V, Langford J. A global geometric framework for nonlinear dimensionality reduction. *Science* 2000;290:2319–23.
- [39] Martinez-Torres M, Barrero Garcia F, Toral Marin S, Gallardo S. A digital signal processing teaching methodology using concept-mapping techniques. *IEEE Trans Educ* 2005;48:422–9.
- [40] Mao G, Fidan B. Localization algorithms and strategies for wireless sensor networks. Hershey: Igi-Global; 2009.
- [41] Tzagarakis C, Jerde TA, Lewis SM, Ugurbil K, Georgopoulos AP. Cerebral cortical mechanisms of copying geometrical shapes: a multidimensional scaling analysis of fMRI patterns of activation. *Exp Brain Res* 2009;194:369–80.
- [42] Oñate J, Pou A. Temperature variations in Spain since 1901: a preliminary analysis. *Int J Climatol* 1901;16(1996):805–15.
- [43] Stephenson D, Doblas-Reyes F. Statistical methods for interpreting Monte Carlo ensemble forecasts. *Tellus* 2000;52A:300–22.
- [44] Shepard R. The analysis of proximities: multidimensional scaling with an unknown distance function, I and II. *Psychometrika* 1962;27:125–40.
- [45] Kruskal J. Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika* 1964;29:1–27.
- [46] Kruskal J, Wish M. Multidimensional scaling. Newbury Park, CA: Sage Publications Inc; 1978.
- [47] Cox T, Cox M. Multidimensional scaling. 2nd ed. Chapman & Hall/CRC; 2001.
- [48] Borg I, Groenen P. Modern multidimensional scaling-theory and applications. 2nd ed. New York: Springer-Verlag; 2005.
- [49] Martínez W, Martínez A. Exploratory data analysis with MATLAB. UK: Chapman & Hall/CRC Press; 2005.
- [50] Ahrens B. Distance in spatial interpolation of daily rain gauge data. *Hydrol Earth Syst Sci* 2006;10:197–208.
- [51] Duarte FB, Machado JT, Duarte GM. Dynamics of the Dow Jones and the NASDAQ stock indexes. *Nonlinear Dyn* 2010;61:691–705.
- [52] Gakkhar S, Singh A. Complex dynamics in a prey predator system with multiple delays. *Commun Nonlinear Sci Numer Simul* 2012;17:914–29.
- [53] Bove R, Pelino V, De Leonibus L. Complexity in rainfall phenomena. *Commun Nonlinear Sci Numer Simul* 2006;11:678–84.
- [54] Zhang X, Zhang H, Chen B, Chen G, Zhao X. Water resources planning based on complex system dynamics: a case study of Tianjin city. *Commun Nonlinear Sci Numer Simul* 2008;13:2328–36.