

The Persistence of Memory

J. A. Tenreiro Machado · António M. Lopes

Abstract This paper analyzes several natural and man-made complex phenomena in the perspective of dynamical systems. Such phenomena are often characterized by the absence of a characteristic length-scale, long range correlations and persistent memory, which are features also associated to fractional order systems. For each system, the output, interpreted as a manifestation of the system dynamics, is analyzed by means of the Fourier transform. The amplitude spectrum is approximated by a power law function and the parameters are interpreted as an underlying signature of the system dynamics. The complex systems under analysis are then compared in a global perspective in order to unveil and visualize hidden relationships among them.

Keywords Complex systems · Dynamical systems · Fourier transform · Visualization

1 Introduction

Complexity is an underlining signature shared by many natural and man-made signals and systems. Examples include financial markets [21], transportation infrastructures [3], telecommunication facilities [43], world and country economies [57], social networks [6], geophysics [50], living organisms [52], computational systems [18], electro-mechanical structures [29] and others [5,44].

Complex systems (CS) are often composed of many distinct components, autonomous but interconnected, that interact dynamically with each other, as well as with their environments, at different scales in time and space [51]. From the small scale interactions between system parts, some kind of self-organization emerges, without any global control or supervision, giving rise to a large scale behavior that cannot be described by a single rule, nor reduced to just one level of description. CS exhibit much richer global dynamics than their composing individual parts. Those components share a dense network of interconnections that act as feedback loops and often originate semi-stable patterns at higher-level scales. Moreover, CS are sensitive to small changes in the system inputs, which make large scale dynamics difficult to analyze. Despite some mathematical and computational tools that have been used to understand CS behavior, many reveal to be inappropriate, or to have severe limitations. In fact, good methodologies for the systematic analysis of the CS are still

lacking. Common properties to CS are the absence of characteristic length-scale, long range correlations and persistent memory, which are also associated to fractional order systems [4, 19, 20, 33], meaning that CS reveal power-law (PL) behavior [14, 40].

In this paper we study several CS in the perspective of dynamical systems. From such a viewpoint, a CS has an output that results from stimuli, being that signal interpreted as a manifestation of the global system dynamics. The system output is then analyzed by means of the Fourier transform (FT) and the amplitude spectrum is approximated by a PL function. In this way, no a priori mathematical assumptions are made about the system. This methodology does not follow the classical modeling approach, since it is supported only by the signals emerging from the CS, but it is sufficiently simple and robust to be adopted in a large number of different CS.

The analyzed systems are: stock markets (ST), musical sounds (MU), Internet bytes rate (IB), Internet packets rate (IP), atmospheric temperatures (AT), solar data (SD), river flow (RF), US tornadoes (TO), earthquakes (EQ), ozone concentration (OZ), tectonic plates motion (TP), electroencephalograms (EEG), electrocardiograms (ECG), arterial pressure (AT), heart RR interval (RR), human gait (HG), seismic waves (SW), human deoxyribonucleic acid (DNA), lake level (LK), life expectancy (LE), international trade openness (ITO), gross domestic product (GDP), rainfall (RA), population density (PD) and CO₂ emissions (CO₂).

In a complementary approach, we visualize similarities among the cases studied by means of hierarchical clustering and multidimensional scaling (MDS) tools. Our approach contributes to better understand CS dynamics and their underlying ruling principles, as well as to expose hidden correlations between systems from very distinct areas.

In this line of thought, in Sect. 2 we briefly describe the CS adopted in this study. In Sect. 3 we present the mathematical fundamentals used for processing the data. In Sect. 4 we present and discuss the results obtained. Finally, in Sect. 5 we outline the main conclusions.

2 Dataset

This section describes briefly the phenomena studied in this work and characterizes the time-series for further

numeric analysis. We analyze several instances of 25 CS categories, in a total of 167 time-series. For each category, $i = 1, \dots, 25$, we adopt $4 \leq p \leq 10$ instances. It should be noted that all data are freely available on the Internet and the time-series are characterized by distinct sampling intervals, h , and time-lengths, P (Table 1).

2.1 Stock markets (ST)

ST are CS originated by different types of dynamical processes and characterized by universal PLs. The analysis of financial time-series has attracted the interest of many researchers [17, 27, 37], as large amounts of data are freely available for study. Moreover, the universality exhibited by financial markets is shared with other CS. In the sequel we analyze nine indexes at a daily time horizon: S&P 500, NYSE Composite, Nikkei 225, NASDAQ Composite, HANG SENG, FTSE 100, Dow Jones Industrial Average, DAX and CAC 40. The datasets are available at the Yahoo Finance website (<https://finance.yahoo.com/>). The time period of analysis, P , comprises years 1992–2013, with $h = 1$ day sampling of stock closing values. Occasional missing values as well as values corresponding to closing days are estimated using linear interpolation.

2.2 Musical sounds (MU)

A musical composition consists of a finite time sequence of sounds, organized in a complex dynamical structure [28, 34]. In this study we analyze musical compositions in the perspective of dynamic systems. A given original musical analogue source is represented as a time sequenced digital data stream, sampled at $h = 22.7$ μ s, and subsequently converted to a single (mono) digital time-series. Each sample is represented as a 32-bit signed floating value [33]. Five compositions are tackled: ‘My Way’ (Frank Sinatra), ‘Saturday Night Fever’ (Bee Gees), ‘We are the champions’ (Queen), ‘Elevation’ (U2) and ‘Bohemian rhapsody’ (Queen).

2.3 Internet traffic rate (IB, IP)

Internet traffic results from millions of interactions between users, applications and protocols. Such traf-

Table 1 Summary of the main characteristics presented by the 25 CS studied

System	Acronym	System, i	Instances, p	Sampling period, h	Signal time-length, P
Stock markets	ST	1	9	1 day	Years 1992–2013
Musical sounds	MU	2	5	22.7 μ s	Music time-length
Internet bytes rate	IB	3	5	0.1 s	900 s
Internet packets rate	IP	4	5	0.1 s	900 s
Atmospheric temperatures	AT	5	5	1 day	Years 1882–2012
Solar data	SD	6	4	64 s (240 s magnetic field)	Year 2012
River flow	RF	7	5	1 day	Years 1982–2013
US tornadoes	TO	8	5	1 min	1950–2012
Earthquakes	EQ	9	5	1 s	Years 1904–2013
Ozone concentration	OZ	10	6	1 day	Years 2004–2013
Tectonic plates motion	TP	11	5	1 day	2006–2013
Electroencephalogram	EEG	12	6	6.25 ms	61 s
Electrocardiogram	ECG	13	5	2.78 ms	60 min
Arterial pressure	AP	14	5	2.78 ms	60 min
Heart RR interval	RR	15	7	1 s	24 h
Human gait	HG	16	5	1 s	6 min/15 min
Seismic waves	SW	17	5	50 ms	Main shock time-length
Human deoxyribonucleic acid	DNA	18	7	–	–
Lake level	LL	19	8	10 day	Years 1992–2013
Life expectancy	LE	20	10	1 day	Years 1960–2012
International trade openness	ITO	21	10	1 day	Years 1960–2012
Gross domestic product	GDP	22	10	1 day	Years 1960–2012
Rainfall	RA	23	10	1 day	Years 1910–2013
Population density	PD	24	10	1 day	Years 1960–2012
CO ₂ emissions	CO ₂	25	10	1 day	Years 1960–2012

fic exhibits complex behavior [2,39], affected by the diversity and unpredictability of human activities, as well as technical issues, like congestion and anomalies. We study Internet traffic as the flow of data observed at a given measuring point, during a certain period of time. Time-series sampled at $h = 0.1$ s, representing the total byte rate (IB) and packet rate (IP) of information are generated from data collected at the Internet traffic data repository [22], maintained by the measurement and analysis on the WIDE Internet (MAWI) working group of the WIDE Project (<http://www.wide.ad.jp/project/wg/mawi.html>). The MAWI working group focuses on traffic measurement analysis, in particular on the long-term measurement on wide-area global Internet. We present the results obtained for the analysis of six traces consisting of anonymized packet headers collected in both directions on a 150 Mb Ethernet

external link which connects the WIDE backbone and its upstream. The traces are dated from March 2008, resulting from traffic monitoring during approximately $P = 900$ s.

2.4 Atmospheric temperatures (AT)

Weather conditions are changing and global warming is a major concern nowadays. Negative impacts on social, economic and health aspects of the human life are expected in medium to long term scenarios. Usually, human activity is pointed out as one important cause of global warming. In fact, since the beginning of the industrial revolution, the non-natural emissions of carbon dioxide have increased, potentiating the greenhouse effect and heating Earth's atmosphere. We use temperature time-series from five cities as representa-

tive signals of the global warming dynamics, namely Berlin, Bombay, Tokyo, Los Angeles and Washington, for the time period, P , composed by years 1882–2012. The data are available at the National Aeronautics and Space Administration (NASA) website (http://data.giss.nasa.gov/gistemp/station_data/). Each data record consists of the average temperatures per month, expressed in Celsius degrees. Some occasional gaps of one month in the data (represented on the original data by the value 999.9) are substituted by a linear interpolation between the two adjacent values. Moreover, although of minor influence, the distinct number of days of each month and the leap years are also taken into account. Before computing the FT, the data were interpolated linearly in order to get daily, $h = 1$ day, temperatures [30,53].

2.5 Solar data (SD)

Solar activity varies with time and position of the Sun. In particular, solar wind is the flow of charged particles, namely protons, electrons and heavier ionized atoms, from the Sun, spreading in all directions at speeds of about 400 km/s and temperatures close to 10^6 K [58]. The exact mechanism of solar wind formation is not well known, but it varies regularly with the 27-day rotation period of the Sun, as well as irregularly in response to certain phenomena in Sun's corona. The solar wind is responsible for the heliosphere (resulting from the deflection of the wind charged particles when interacting with Earth's magnetic field), the geomagnetic storms, the aurora and the plasma tails of comets, always pointing away from the sun. In this work we analyze four time-series of solar data: wind particles' (proton) density (cm^{-3}), temperature (K), speed (km/s) and interplanetary magnetic field magnitude (nano Tesla). The datasets are freely available at the Advanced Composition Explorer (ACE) Science Center (<http://www.srl.caltech.edu/ACE/ASC/index.html>). The ACE spacecraft was launched in August 1997. It occupies an orbit about the L1 point, located about 1.5×10^6 km from the Earth, where the gravitational attraction of the Sun and Earth is equal and opposite. We use ACE SWEPAM Level 2 Data, corresponding to $h = 64$ s averages ($h = 240$ s for interplanetary magnetic field) and for year $P = 2012$. Occasional missing values are filled in using linear interpolation.

2.6 River flow (RF)

Analysis of RF time-series is particularly important in hydrology. It can help preventing extreme events, like floods or water supply failures, delineating policies for water storage and drainage, and managing environmental flows and river ecosystems, among others. RF dynamics is affected by meteorological and hydrological coupled processes, as well as human activities, that difficult system analysis by means of traditional mathematical tools [41,48]. Data for US RF is freely available at the USGS National Water Information System (<http://waterdata.usgs.gov/nwis>). The time-series contain the daily discharge mean values, $h = 1$ day, expressed in cubic feet per second, measured on a given location (station). The period of observation, P , includes years 1982–2013. Five river/station datasets are considered in this study: Arkansas (USGS 06888350 Kansas R NR Belvue, KS, USA), Colorado (USGS 09427520 Colorado River below Parker Dam, AZ–CA, USA), Mississippi (USGS 07010000 Mississippi River at St. Louis, MO, USA), Missouri (USGS 06893000 Missouri River at Kansas City, MO, USA) and Rio Grande (USGS 08276500 RIO Grande BLW Taos Junction Bridge near Taos, NM, USA).

2.7 US tornadoes (TO)

A tornado is a violently rotating column of air extending from a cumulonimbus cloud to the ground. Tornadoes may assume different shapes and sizes, but they typically appear as a funnel with the narrower end touching the ground. Tornadoes' damaging paths can surpass 1.5 km wide and 80 km long and the most violent events can cause huge destruction and fatalities [47]. Tornadoes are classified according to their intensity and damage. Earlier events use the Fujita scale, while most recent tornadoes are often classified using the enhanced Fujita scale. In both systems, the least damaging tornadoes are rated as F0, while the most damaging tornadoes are rated as F5 [10]. We use the US tornado database compiled by the National Oceanic and Atmospheric Administration (NOAA), National Weather Service, Storm Prediction Center. Data are available online at (<http://www.spc.noaa.gov/>), containing all reported US tornadoes for years $P = 1950, \dots, 2012$. Each event includes, among other features, information about date and time

(with 1-min time resolution, $h = 1$ min), Fujita scale, total number of fatalities and injuries, geographic location (i.e., touchdown and liftoff latitude and longitude coordinates), tornado width and path length. In this study, five time series are generated by modeling the events as sequences of Dirac impulses with amplitude proportional to the tornadoes path length (measured in km). Events for US states Florida, Kansas, Nebraska, Tennessee and Texas are considered.

2.8 Earthquakes (EQ)

Earthquakes are natural events caused by the relative displacement between adjacent tectonic plates. This phenomenon leads to strain energy accumulation around the fault surfaces and, when the stress is sufficiently high to break through the surfaces asperities, a sudden motion of the plates occurs, causing shaking and ground displacement. Earthquakes unveil complex correlations in space, time and magnitude. Those are characterized by self-similarity and absence of characteristic length-scale, meaning that seismic parameters exhibit PL behavior, as predicted by the Gutenberg–Richter and Omori laws [13, 46]. The Bulletin of the International Seismological Centre (ISC), available online at (<http://www.isc.ac.uk/>), is used in this study. The data catalog covers the period of years $P = 1904$ –2013. Each record contains information about the date and time of the event (with 1-s time resolution, $h = 1$ s), geographic location and magnitude. We consider five Flinn–Engdahl (FE) seismic regions and, for each region, we generate the corresponding time-series by modeling the events as sequences of Dirac impulses with amplitude proportional to the moment magnitude of the events. The FE regions are: Alaska-Aleutian arc, Andean South America, Fiji Islands area, Taiwan area and Middle East-Crimea-Eastern Balkans.

2.9 Ozone concentration (OZ)

Ozone is formed in the lower atmosphere (troposphere) as a result of reactions between oxides of nitrogen and volatile organic components. At the lower atmosphere, ozone is a pollutant that harms plants and animals and is itself a greenhouse gas. On the other hand, at the higher atmosphere (stratosphere), the ozone concentration is much higher than at the troposphere. Usually, such ozone concentration is known as the

‘ozone layer’ and serves as a filter to the ultraviolet radiation, which would otherwise reach the surface of the Earth. The global warming felt in the lower atmosphere has led to cooling in the stratosphere, favoring the destruction the ozone layer. Total ozone concentration involves both troposphere and stratosphere processes and weather variables, making ozone concentration time-series highly complex and non-linear [25, 55]. We present the analysis of data regarding the tropospheric concentration of ozone. The data are available at the NOAA’s Earth System Research Laboratory (ESRL) (<http://www.esrl.noaa.gov/gmd/grad/neubrew/OmiDataTimeSeries.jsp>) in a daily basis, $h = 1$ day, covering the period of years $P = 2004$ –2013. Six time-series are considered, corresponding to large cities around the world: Tokyo, London, São Paulo, Los Angeles, Bombay and Sydney.

2.10 Tectonic plates motion (TP)

The global positioning system (GPS) enables the measurement of tectonic plates’ displacement with high accuracy. These measurements are used to analyze the spatio-temporal evolution of the tectonic plates’ motion and its relation with earthquakes and other geological phenomena [42]. We examine GPS datasets representative of the tectonic plates’ dynamical behavior. The data are available at the UNAVCO website (<http://www.unavco.org/unavco.html>). UNAVCO is a non-profit university-governed consortium for geoscience research and education using geodesy, sponsored by the National Science Foundation (NSF) and NASA. Time-series from five stations are considered: ChernaburaAK2008 (ID AC12; location Sand Point, AK, USA), BEMT-SCGN-CS2001 (ID BEMT; location Twentynine Palms, CA, USA), Wickum-RnchMT2006 (ID P050; location Chester, MT, USA), Quinlt014GWA2008 (ID P400; Olympic NP, WA, USA) and GammaGulchCS2006 (ID P599; Yucca Valley, CA, USA). The north-displacement daily values, measured in mm, are used in our study (the sampling interval is $h = 1$ day and the period of analysis includes years $P = 2006$ –2013).

2.11 Electroencephalogram (EEG)

The brain is a complex system that exhibits rich spatiotemporal dynamics [38]. Active nerve cells in the

brain produce electrical currents that generate voltage differences on the scalp. The EEG is a neurophysiological examination used to evaluate the brain's function through the analysis of the spontaneous cerebral electrical activity. Electroencephalography is used in the diagnosis of a variety of neurological disorders, as epilepsy, Alzheimer's disease (and different kinds of dementia) and central nervous system infections. EEGs data are freely available at the PhysioBank (<http://www.physionet.org/physiobank/>). PhysioBank is a collection of over 50 databases serving as an archive of well-characterized digital recordings of physiologic signals and related data for use by the biomedical research community. We present results obtained for the analysis of six EEG time-series. The data are part of the dataset created by the developers of the BCI2000 instrumentation system, which was used to collect the EEGs [12]. The time-series are sampled at $h = 6.25$ ms time interval and last $P = 61$ s. The samples correspond to channel FC5 voltage, expressed in μV .

2.12 Electrocardiogram (ECG)

The ECG is a non-invasive diagnostic exam, used to monitor heart electrical activity that is generated by the polarization and depolarization of the cardiac tissue. The ECG signal consists of waveform components, which indicate electrical events during each heart-beat. These waveforms are known as P wave, QRS complex, T wave and U wave. The P wave is the first small upward movement observed in the ECG signal, indicating that the atria are contracting. The QRS complex begins with a downward deflection, Q, followed by a larger upwards excursion, R, and then a downwards refraction, S. The QRS complex has much larger amplitude than the P wave and represents ventricular depolarization and contraction. The T wave represents the repolarization of the ventricles and appears in the ECG signal as a small upwards waveform. The U wave has either low amplitude or is absent. It always follows the T wave and has the same direction in amplitude. We analyze ECG time-series available at the PhysioBank (<http://www.physionet.org/physiobank/>) Massachusetts General Hospital/Marquette Foundation (MGH/MF) waveform database [12]. The typical recording includes three ECG leads, arterial pressure, pulmonary arterial pressure, central venous pressure, respiratory impedance, and airway CO_2 waveforms. In this study we

consider five ECG time-series, generated at $h = 2.78$ ms sampling time interval and total time length of about $P = 1$ h. The samples correspond to ECG lead 1 voltage, expressed in mV.

2.13 Arterial pressure (AP)

Human heart acts like a pump, impelling blood across blood vessels. AP refers to the pressure exerted by circulating blood upon the walls of blood vessels, measured at the upper arm brachial artery, during systemic circulation, i.e., when oxygenated blood is pumped away from the heart to the body, returning deoxygenated back to the heart. When the left ventricle ejects blood into the aorta, the aortic pressure rises. The maximal aortic pressure following ejection is termed the systolic pressure. As the left ventricle is relaxing and refilling, the pressure in the aorta falls. The lowest pressure in the aorta, which occurs just before the ventricle ejects blood into the aorta, is termed the diastolic pressure. Typically, both pressures are measured in mm of mercury (mmHg). AP is one of the most important vital signs. Abnormal values are a major cause of severe diseases such as stroke, heart attack, heart failure, arterial aneurysm and renal failure. For this study, five arterial pressure time-series are extracted from the same records presented above (<http://www.physionet.org/physiobank/>), using the same (h, P) parameters.

2.14 Heart RR interval (RR)

RR is the time interval between two repeated R waves as observed in an ECG signal. At rest, human heart RR interval is between 60 and 100 beats per min. However, changes of RR interval may be related to specific cardiovascular and noncardiovascular diseases. The PhysioBank's Cast RR-Interval Sub-study Database (<http://www.physionet.org/physiobank/>): Dataset E is tackled in this study. Six datasets are considered, corresponding to subjects who were randomly assigned to receive Encainide (i.e., a class of antiarrhythmic agent). RR is measured in seconds over a $P = 24$ h time-period and $h = 1$ s.

2.15 Human gait (HG)

HG refers to the patterns exhibited by humans while walking in an upright posture over a plane surface. Such

patterns result from complex interactions between body systems as brain, nerves, muscles and bones. During human gait, body balance and stability are maintained while body weight is continually moved between the left and right leg. Abnormal gait patterns can indicate medical problems, namely in feet, legs, hips, spinal cord or brain. Five HG time-series from the PhysioBank's Gait in Aging and Disease Database are studied (<http://www.physionet.org/physiobank/>). Data corresponds to subjects walking on level ground, around an obstacle-free path, from $P = 6$ up to $P = 15$ min. The time between foot-strikes is automatically computed based on instrumentation [12] and $h = 1$ s.

2.16 Seismic waves (SW)

Earthquakes generate SW of energy that are measured by seismometers. There are various types of SW, namely (1) body waves, which travel through the inner layers of the Earth; and (2) surface waves, that propagate along the Earth's surface. SW travel through the Earth's layers with different velocities, depending on the physical characteristics of the propagation medium, namely density and elasticity. Body waves are faster than surface waves. Earthquakes result from sudden energy release that has been previously stored at neighboring fault surfaces in the course of plates' slow motion. That energy release is accompanied by strong ground motion across the fault surfaces, which exhibit asperities, causing friction and stick-slip behavior. The asperities on the fault surfaces are like fractals sliding over each other, explaining the complex behavior and fractal-like scaling observed in earthquake phenomena [7,31,54]. In this work we analyze seismograms registered by the Global Seismographic Network (GSN), which is a network of seismological and geophysical sensors, formed in partnership among the USGS, the National Science Foundation and the Incorporated Research Institutions for Seismology (IRIS, <http://www.iris.edu/hq/>). Currently, the GSN comprises about 150 seismic stations distributed globally. All data from the GSN are archived at the IRIS Data Management Center (DMC). Five time-series sampled at $h = 50$ ms time interval are studied. Those correspond to the large earthquakes: offshore Maule, Chile, 2010/02/27, magnitude 8.8; near the east coast of Honshu, Japan, 2011/03/11, magnitude 9.0; off the west coast of northern Sumatra, 2012/04/11, magnitude

8.6; southern Sumatra, Indonesia, 2007/09/12, magnitude 8.5 and Haiti, 2010/01/12, magnitude 7.0. The seismograms were captured by the stations: IU LVC, Limon Verde, Chile; IU MAJO, Matsushiro, Japan; IU CHTO, Chiang Mai, Thailand; IU GUMO, Guam, Mariana Islands and IU TEIG, Tepich, Yucatan, Mexico.

2.17 Human deoxyribonucleic acid (DNA)

DNA represents the hereditary material in humans (and almost all living organisms), encompassing all information related to species evolution at distinct scales and complexity levels. Its structure is made up of two polymers forming a double helix and containing four nitrogenous bases: thymine (T), cytosine (C), adenine (A) and guanine (G). Each base on one side bonds with just one type of base on the other side, forming the base pairing A–T and C–G. The first seven human chromosomes are used in this study. Data are available at the Human-Genome Reference Consortium website (<http://www.ncbi.nlm.nih.gov/projects/genome/assembly/grc/>). For processing the data we first convert the four symbols alphabet T, C, A, G into a numerical value. A fifth symbol available in the original data, N, is considered as 'zero' (i.e., it has no practical meaning for the decoding process). A symbol translation is adopted: $A = 1 + j0$, $C = -1 + j0$, $T = 0 + j$, $G = 0 - j$, $N = 0 + j0$, which follows a 'base pairing' leading to $A = -C$, $T = -G$ and A 'orthogonal' to T ($j = \sqrt{-1}$). Second, we consider that we move along the DNA strip, one symbol at a time. The resulting values are then interpreted as a 'time-series' [33,52].

2.18 Lake level (LK)

As defined by the World Meteorological Organization's International Glossary of Hydrology, lakes are inland bodies of water of considerable size [56]. Lake levels are complex indices that change over time in response to weather conditions (e.g., wind, precipitation and evaporation), human factors (e.g., water outflow for agriculture, domestic and industrial supply) and size and composition of the lake's watershed. We use data from the Open Altimeter Database (<http://openadb.dgfi.badw.de/>). Time series for lakes Balkhash, Erie, Guri, Ladoga, Malawi, Nicaragua, Tanganyika and Tchad are studied. The data consist of

$h = 10$ days level values, over a period of more than twenty years, $P = 1992\text{--}2013$, recorded using satellite altimetry.

2.19 Life expectancy (LE)

We analyze ten time-series of LE at birth for ten countries: Brazil, Canada, China, France, UK, India, Italy, Japan, Mexico and South Africa. The data, from the World Bank database (<http://data.worldbank.org/>), consist of the annual values of LE at birth (in years) for the period of years $P = 1960\text{--}2012$. Before computing the FT, the data are interpolated linearly in order to get daily values ($h = 1$ day). Life expectancy at birth is the average number of years that a new born could expect to live, if subjected to the age-specific mortality rates of a given period. It is an indicator of mortality and health conditions [9].

2.20 International trade openness (ITO)

ITO is an economic index given by the ratio of the total trade (exports plus imports) and the gross domestic product of a country. ITO has increased with globalization, potentiating growth but also contributing to higher exposure to international threats [26]. We analyze time-series of ITO data for the ten countries listed in the previous item. The data (<http://data.worldbank.org/>) consist of the annual value of the ITO index (in percentage of the GDP) for the time period of years $P = 1960\text{--}2012$. As before the data are first interpolated in order to get daily values ($h = 1$ day).

2.21 Gross domestic product (GDP)

GDP corresponds to the value of all final goods and services produced within a country over a given period of time, usually one year, expressed in US dollars. It is the primary indicator used to measure the health of a country's economy [23]. Ten time-series of GDP values for the countries listed previously are presented. The data source, sampling interval and period of analysis (h , P) are the ones listed in the previous paragraph.

2.22 Rainfall (RA)

Precipitation is any form of water falling from the atmospheric environment and reaching the earth's sur-

face. It includes rain, sleet, snow and hail. The formation of precipitation follows a complex process in which the moisture of the lower atmosphere is heated by conduction, becomes lighter than the surrounding air and undergoes an adiabatic ascent. The rising air causes cooling that can make it reach the saturation point. From that level, condensation of water vapor occurs in the form of tiny droplets that are held in suspension, as clouds or fog. These drops have not yet sufficient size to overcome the air resistance, thus being maintained in suspension until, by a growth process, they reach sufficient mass to precipitate [16]. We study time-series from ten UK locations: Armagh, Bradford, Durham, Eskdalemuir, Oxford, Sheffield, Stornoway, Tiree, Valley and Wickairport. The data are available at the UK MetOffice website (<http://www.metoffice.gov.uk/>). It is provided in a daily basis ($h = 1$ day), i.e., daily precipitation values measured in mm, covering years $P = 1910\text{--}2013$. Some occasional gaps in the data are substituted by a linear interpolation between the two adjacent values.

2.23 Population density (PD)

PD is the number of people per unit of area [11]. For a country, it is the number of inhabitants divided by the country's area. Most densely populated countries are Singapore, Bahrain and Bangladesh. The analyzed data are from the World Bank database (<http://data.worldbank.org/>) and consist of the annual values of population density (in inhabitants per square-km) for years $P = 1960\text{--}2012$. Ten countries are considered: Brazil, Canada, China, France, UK, India, Italy, Japan, Mexico and South Africa. Before computing the FT, the data are interpolated linearly in order to get daily values ($h = 1$ day).

2.24 Carbon dioxide emissions (CO₂)

CO₂ is naturally present in the atmosphere as part of the Earth's carbon cycle. However, industrialization has increased the non-natural emissions, potentiating the greenhouse effect and contributing to Earth's atmosphere heating [49]. Time-series consisting of the emissions annual values (expressed in kilo tones) in the ten countries listed above are analyzed. The data source, sampling interval and period of analysis (h , P) are the ones listed in the previous paragraph.

3 Materials and methods

This section presents the mathematical tools used to process the data, namely Fourier transform, mutual information, hierarchical clustering and multidimensional scaling.

3.1 Fourier transform

The FT is a classical, powerful and robust signal processing tool for the analysis of systems' dynamics. It converts the time-domain signal, $x(t)$, to the frequency-domain signal, $X(j\omega)$:

$$\mathcal{F}\{x(t)\} = X(j\omega) = \int_{-\infty}^{+\infty} x(t)e^{-j\omega t} dt \quad (1)$$

where $\mathcal{F}\{\cdot\}$ represents the Fourier operator and ω denotes the angular frequency.

FT has the advantage of being robust, allowing a direct interpretation of the processed data and being usable in a wide range of signals. In our analysis FT does not represent any system model, but just characterizes the 'system dynamical manifestation'. FT is a common analysis tool applied to signals generated by different types of CS, where signal occasional artifacts are 'diluted' into total result, and global embedded properties are unveiled.

3.2 Mutual information

Mutual information has been introduced in the context of information theory and then adopted as a common approach to the analysis of CS [32,45]. Mutual information measures the statistical dependence between two random variables, yielding the amount of information that one random variable 'contains' about the other. Mathematically, for the discrete random variables (X_m, X_n) , the mutual information, $I(X_m, X_n)$ is given by:

$$I(X_m, X_n) = \sum_{x_n \in X_n} \sum_{x_m \in X_m} p(x_m, x_n) \cdot \log \left(\frac{p(x_m, x_n)}{p(x_m) \cdot p(x_n)} \right) \quad (2)$$

where, $p(x_m)$ and $p(x_n)$ represent the marginal probability distribution functions of X_m and X_n , respectively, and $p(x_m, x_n)$, is the joint probability distribution function. The mutual information is always nonnegative and symmetric, meaning that $I(X_m, X_n) = I(X_n, X_m)$.

Often the normalized mutual information, $I_N(X_m, X_n) \in [0, 1]$, is adopted:

$$I_N(X_m, X_n) = \frac{I(X_m, X_n)}{H(X_m, X_n)} \quad (3)$$

with $H(X_m, X_n)$ representing the joint entropy:

$$H(X_m, X_n) = - \sum_{x_n \in X_n} \sum_{x_m \in X_m} p(x_m, x_n) \cdot \log [p(x_m, x_n)]. \quad (4)$$

In that case $I_N(X_m, X_n) = 1$, for $m = n$. The normalized mutual information has the advantage of simplifying comparison across different conditions and improving sensitivity.

3.3 Hierarchical clustering

Clustering is a common statistical technique for data analysis. It has been widely adopted, namely in data mining, machine learning, pattern recognition, image analysis, information retrieval and bioinformatics [15]. Clustering analysis groups objects in such a way that those similar to each other, in some sense, are placed in the same group (cluster). In particular, for hierarchical clustering, a hierarchy of object clusters is built, based on two alternative algorithms. In agglomerative clustering each object starts in its own singleton cluster and, at each step, the two most similar clusters are greedily merged. The agglomerative algorithm iterates until there is a single cluster containing all objects. In divisive clustering, all objects start in one single cluster and, at each step, the algorithm removes the 'outsiders' from the least cohesive cluster. The iterations stop when each object is in its own singleton cluster. The clusters are combined (split) for agglomerative (divisive) clustering, based on a measure of dissimilarity between clusters. This is achieved using a metric (i.e., a measure of the distance between pairs of objects) and a linkage criterion (i.e., the definition of the dissimilarity between clusters as a function of the pairwise distances between objects). In analytic terms, given the clusters, R and S , any metric can be used to measure the distance, $d(x_R, x_S)$, between objects $x_R \in R$ and $x_S \in S$. The Euclidean and Manhattan distances [1] are often adopted. Based on these metrics, the maximum, minimum and average linkages are commonly used, being, respectively,

$$d_{\max}(R, S) = \max_{x_R \in R, x_S \in S} d(x_R, x_S) \quad (5)$$

$$d_{\min}(R, S) = \min_{x_R \in R, x_S \in S} d(x_R, x_S) \quad (6)$$

Fig. 1 Normalized time series, $\tilde{x}_6^1(t)$, corresponding to wind speed time-series

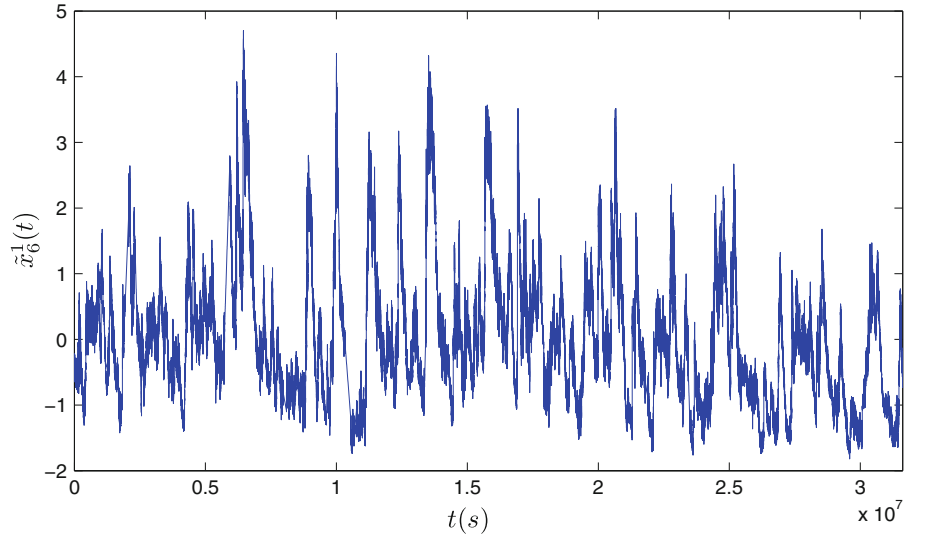
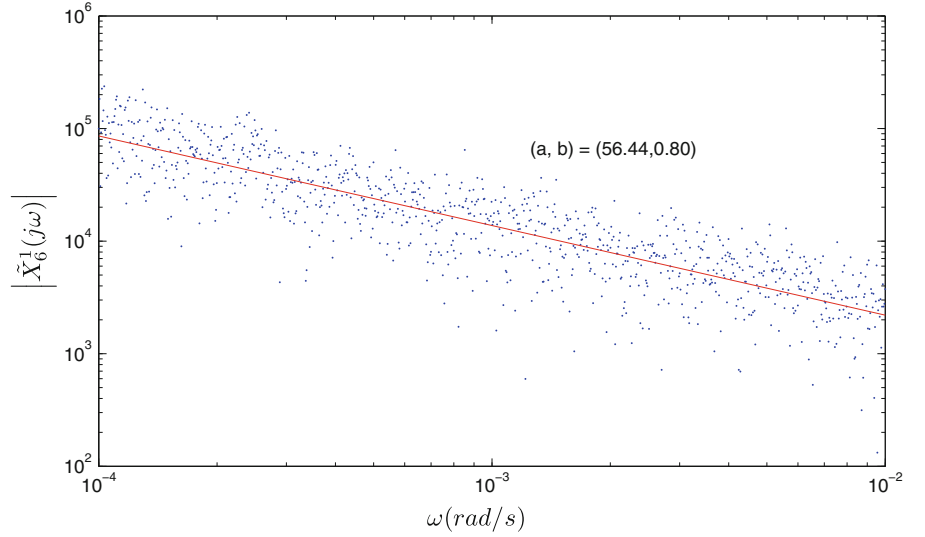


Fig. 2 Amplitude of the FT, $|\tilde{X}_6^1(j\omega)|$, and PL approximation corresponding to the signal shown in Fig. 1



$$d_{\text{ave}}(R, S) = \frac{1}{\|R\| \|S\|} \sum_{x_R \in R, x_S \in S} d(x_R, x_S). \quad (7)$$

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

3.4 Multidimensional scaling

MDS is a technique for visualizing information in data [8]. The main idea is to allow the researcher to observe similarities between the objects. MDS requires the definition of a measure of similarity and the construction of a $s \times s$ symmetric matrix C of item to item similarities, where s is the total number of objects. In classical

MDS, matrix C is symmetric and its main diagonal is composed of '1'. MDS assigns a point to each item in a multi-dimensional space and arranges the set in order to reproduce the observed similarities. By rearranging the object's positions in space, MDS tries to arrive at a configuration that best approximates the observed similarities. MDS uses a minimization algorithm that evaluates different configurations for maximizing the goodness-of-fit. A common measure that is used to evaluate how a particular configuration reproduces the observed distance matrix C is the raw stress:

$$\mathcal{S} = [d_{ij} - f(\delta_{ij})]^2 \quad (8)$$

Fig. 3 Global map of the (a, b) parameters

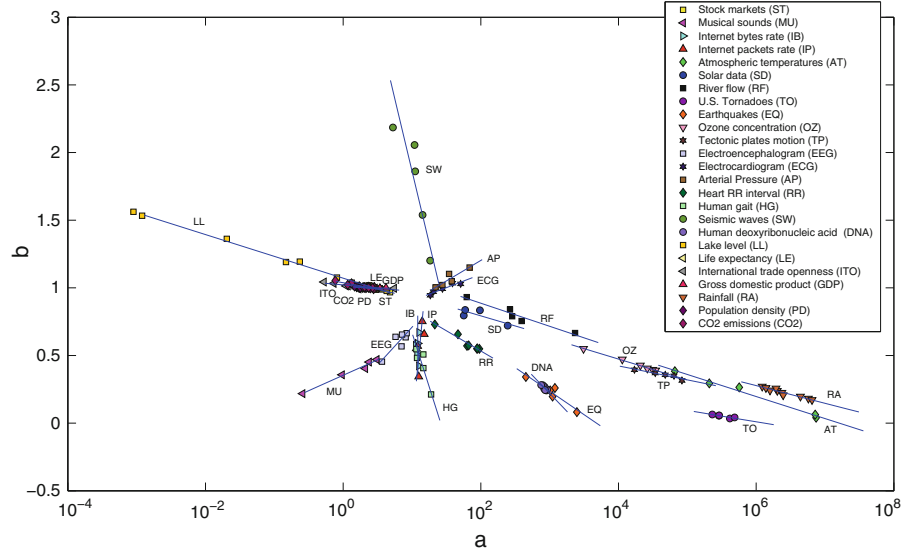
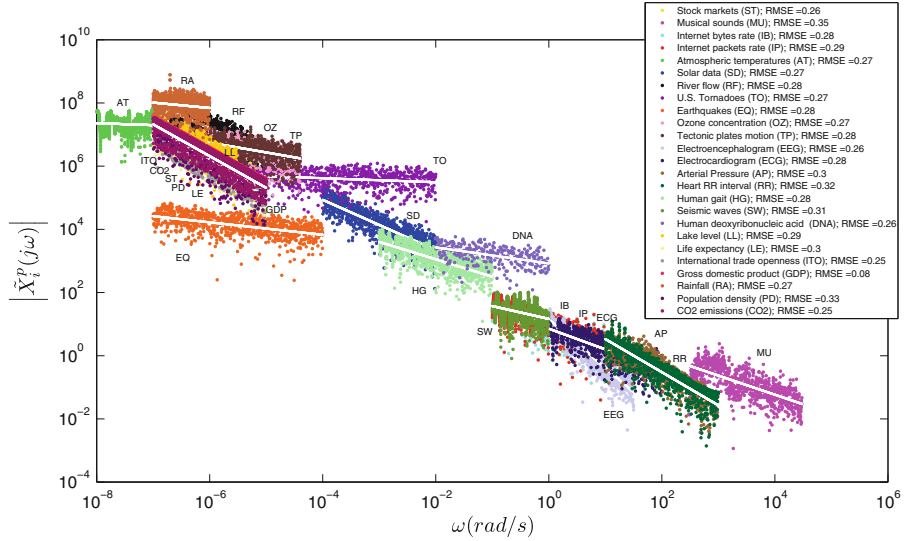


Fig. 4 Amplitudes of the FT and PL approximations for the first instance of all system categories, $|\tilde{X}_i^1(j\omega)|, i = 1, \dots, 25$



where d_{ij} stands for the reproduced distances, given the respective number of dimensions, δ_{ij} represents the observed distances and $f(\cdot)$ indicates some type of transformation. The smaller the stress, \mathcal{S} , the better is the fit between the observed and the reproduced matrices. Therefore, the actual orientation of axes in the final solution is arbitrary. We can rotate or translate the MDS map and the distances between points remain the same. The final orientation of axes in the space is to be decided by the researcher that chooses the one that produces a clearer visualization. For accessing the quality of the MDS procedure, the stress and Shepard plots are adopted. The stress plot represents \mathcal{S} versus the number of dimensions k of the

MDS map and leads to a monotonic decreasing chart. We choose the ‘best’ dimension k as a compromise between stress reduction and number of dimensions for the MDS map. The Shepard diagram compares the reproduced distances, for a particular value k of the MDS map, versus the observed distances. Therefore, a narrow/large scatter around a 45° line indicates a good/bad fit of the distances to the dissimilarities for that value k . The measure for constructing matrix C depends on the researcher’s choice and, therefore, we can have multiple criteria. The MDS interpretation is based on the emerging clusters and relative distances, rather than on the absolute coordinates, or shapes.

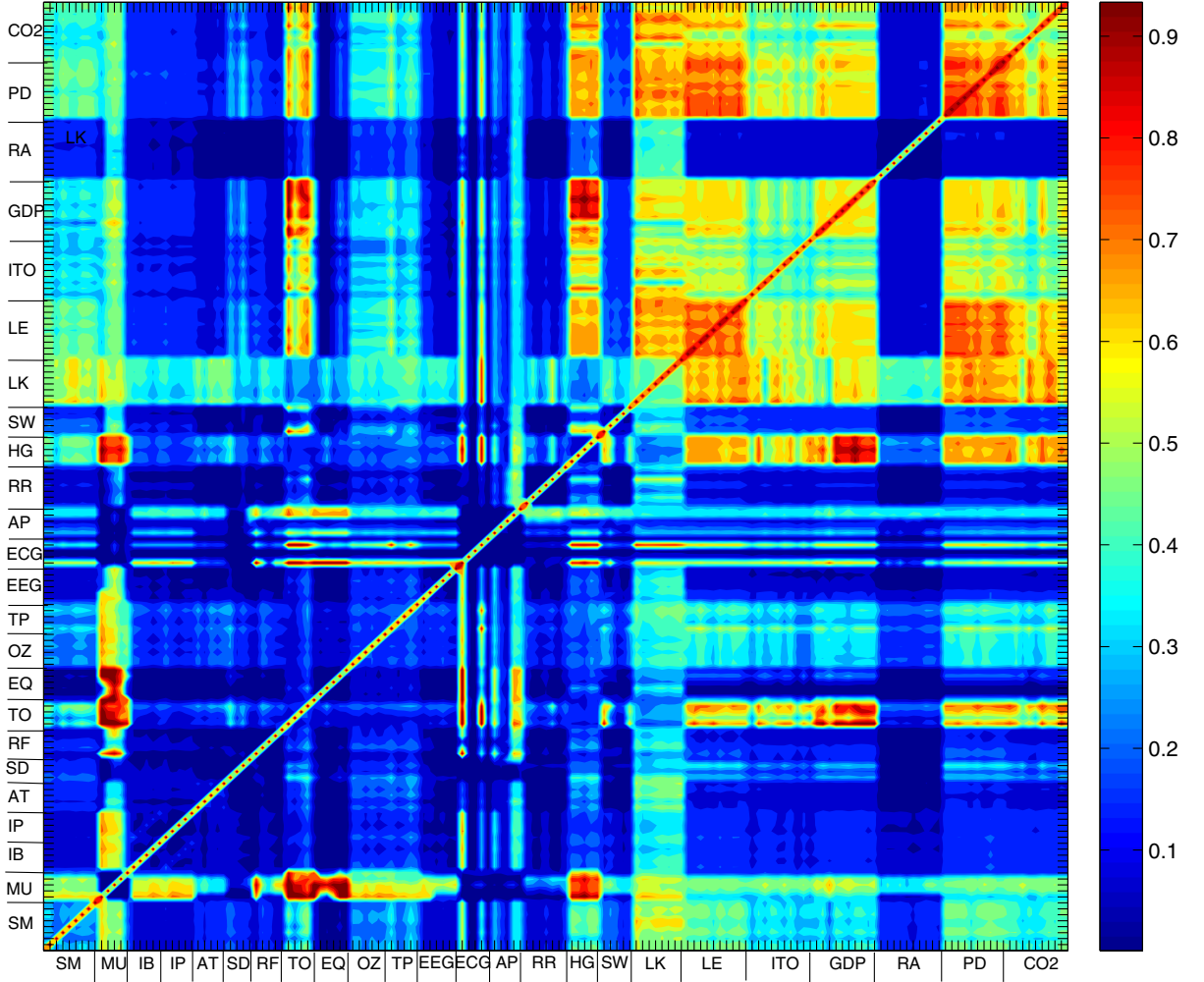


Fig. 5 Contour map representing matrix C

4 Results

In this section we study 25 complex systems, each with several instances. For each category, $i = 1, \dots, 25$, are adopted $4 \leq p \leq 10$ instances. In Sect. 4.1 a total of 167 time-series are studied and compared by means of hierarchical clustering and MDS. In Sect. 4.2 the residuals associated to the PL approximations of the FT data are further analyzed.

4.1 Time-series analysis and comparison

Each time-series, $x_i^p(t)$, is first normalized and processed by the FT. In a second step, the resulting amplitude

spectrum is approximated by a PL function and the PL parameters are compared.

In analytical terms, a time-series $x_i^p(t)$, is represented by:

$$x_i^p(t) = \sum_{k=1}^{T_i^p} A_{ik}^p \delta(t - t_{ik}^p) \quad (9)$$

where, A_{ik}^p represents the amplitudes of the data points, t_{ik}^p are time instants, t represents time and T_i^p is the total time-length of the time-series.

The time-series $x_i^p(t)$, are normalized by:

$$\tilde{x}_i^p(t) = \frac{x_i^p(t) - \mu_i^p}{\sigma_i^p} \quad (10)$$

A phylogenetic tree illustrating the relationships between 100 bacterial strains. The tree is rooted on the left and branches out to the right. Each strain is represented by a horizontal line (branch) ending in a colored triangle. The colors of the triangles correspond to the strain names, which are listed on the right side of the tree. The strains are grouped into several major clusters, each represented by a different color: purple (top), yellow (middle), green (lower middle), and blue (bottom). The branches are labeled with bootstrap values, indicating the confidence in the branching order. The tree shows a high degree of genetic differentiation between the major clusters, with some internal branching within clusters also showing high bootstrap support.

Strain names (from top to bottom):

- SD3
- SW1
- SW5
- ECG3
- ECG4
- ECG5
- ECG6
- ECG7
- ECG8
- ECG9
- ECG10
- ECG11
- ECG12
- ECG13
- ECG14
- ECG15
- ECG16
- ECG17
- ECG18
- ECG19
- ECG20
- ECG21
- ECG22
- ECG23
- ECG24
- ECG25
- ECG26
- ECG27
- ECG28
- ECG29
- ECG30
- ECG31
- ECG32
- ECG33
- ECG34
- ECG35
- ECG36
- ECG37
- ECG38
- ECG39
- ECG40
- ECG41
- ECG42
- ECG43
- ECG44
- ECG45
- ECG46
- ECG47
- ECG48
- ECG49
- ECG50
- ECG51
- ECG52
- ECG53
- ECG54
- ECG55
- ECG56
- ECG57
- ECG58
- ECG59
- ECG60
- ECG61
- ECG62
- ECG63
- ECG64
- ECG65
- ECG66
- ECG67
- ECG68
- ECG69
- ECG70
- ECG71
- ECG72
- ECG73
- ECG74
- ECG75
- ECG76
- ECG77
- ECG78
- ECG79
- ECG80
- ECG81
- ECG82
- ECG83
- ECG84
- ECG85
- ECG86
- ECG87
- ECG88
- ECG89
- ECG90
- ECG91
- ECG92
- ECG93
- ECG94
- ECG95
- ECG96
- ECG97
- ECG98
- ECG99
- ECG100

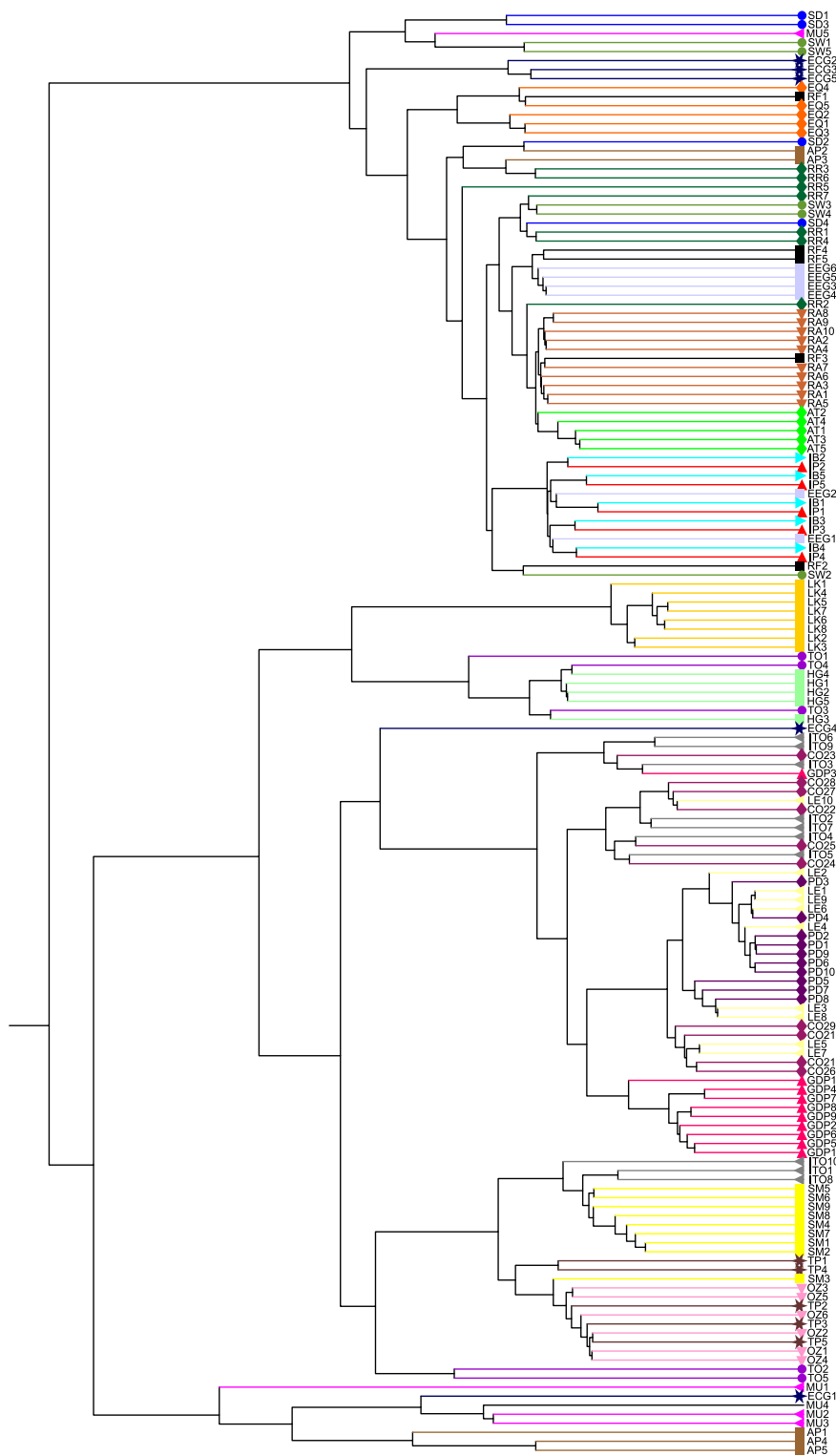


Fig. 8 Two dimensional MDS map representing similarities between time-series

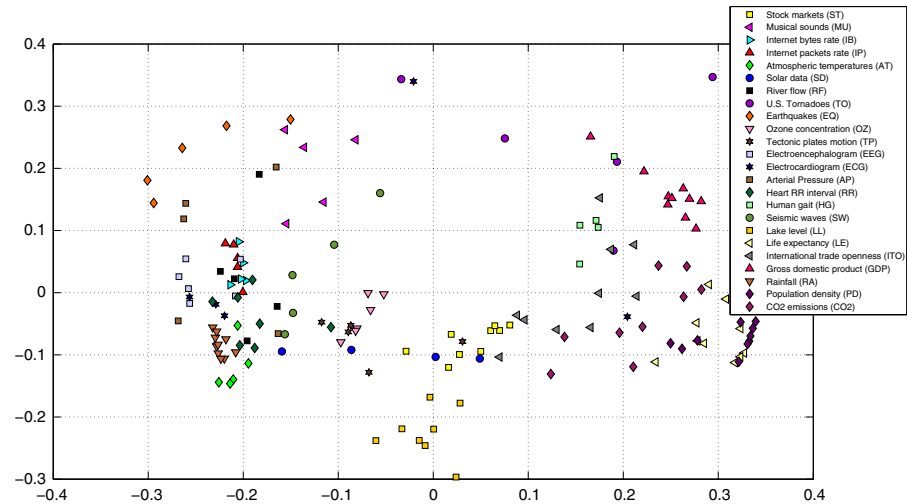


Fig. 9 Three dimensional MDS map representing similarities between time-series

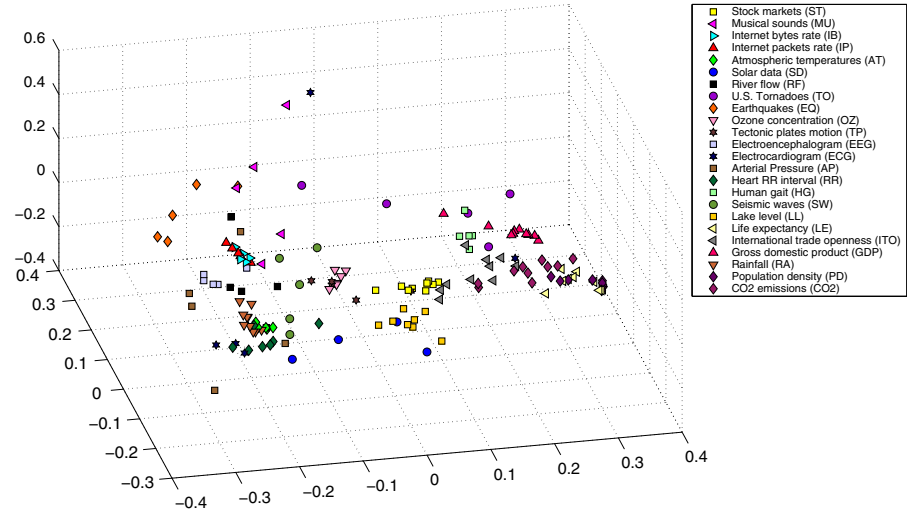


Fig. 10 Histogram and beta-function approximation for the time-series $\tilde{x}_6^1(t)$, which corresponds to solar wind speed

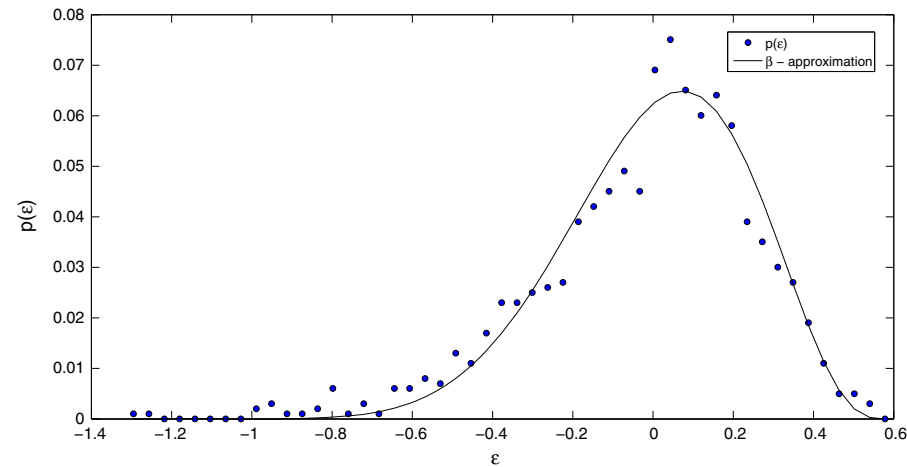
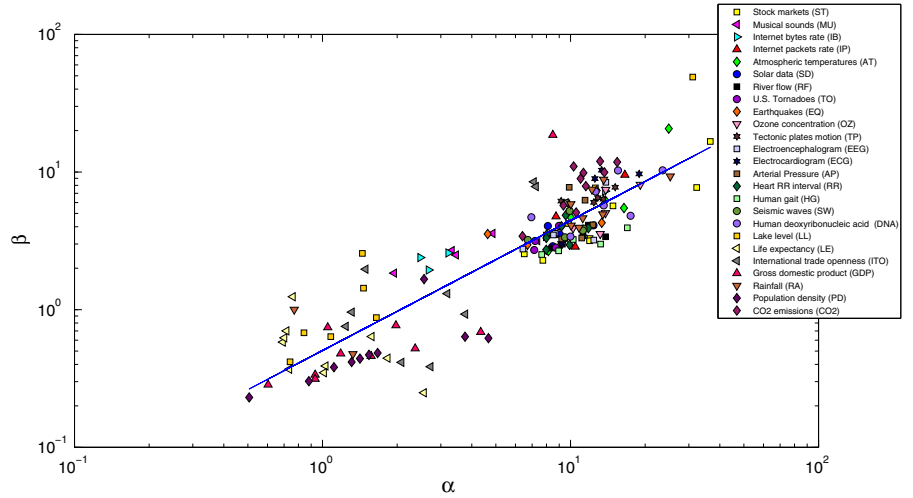


Fig. 11 Global map of (α, β) parameters



ters led to a fractional value of parameter b , being $(a, b) = (56.44, 0.80)$.

The parameters (a, b) are computed for the whole set of time-series (167 in total) and the corresponding map is depicted in Fig. 3. It can be seen that the instances of each system category are placed along a line in the (a, b) space. Some categories point in different directions, but further analysis is needed to unveil the meaning of such observed features.

In Fig. 4 we represent the amplitudes of the FT and PL approximations for the first instance of all system categories, $|\tilde{X}_i^1(j\omega)|$, $i = 1, \dots, 25$. For each case the root mean squared error of the PL fit is given. As can be seen, for each case the frequency range is conditioned by instrumentation issues (i.e., sampling rate) and acquisition time length: lower sampling periods result in higher frequencies in the Fourier domain, while longer time-series allow smaller frequencies in the Fourier spectrum.

In the sequel all time-series are compared (except DNA, which is not a true time-series). First, we calculate the normalized mutual information, $I(X_m, X_n)$, $(m, n) = 1, \dots, 160$, between each pair of signals. The probabilities are estimated from histograms constructed using $M = 200$ bins. Second, we compute a 160×160 similarity matrix $C = [I_N(X_m, X_n)]$, which is used to feed the hierarchical clustering and MDS algorithms.

Figure 5 depicts C as a contour map. The map reveals strong correlations between certain signals, corresponding to extreme values of $I_N(X_m, X_n)$. However, the comparison requires a considerable work and is more appropriate for pairwise comparisons.

As alternative methods to visualize and to compare results, a hierarchical clustering algorithm is adopted. A visualization tree and dendrogram are generated (Figs. 6, 7, respectively) using the successive (agglomerative) clustering and average-linkage method. The software PHYLIP was used for generating both graphs (<http://evolution.genetics.washington.edu/phylip.html>). Instances of the same system category are represented using the same color. As can be seen, most instances of each system are placed in the same cluster.

The MDS maps for $k = 2$ and $k = 3$ dimensions are depicted in Figs. 8 and 9, respectively. It should be noted that a shorter (larger) distance between two points on the maps means that the corresponding objects are more similar (distinct). As before, instances of each system are represented using the same color. Both visualization trees and MDS maps allow good interpretation of the results. The MDS maps have advantage when dealing with large number of objects, being more intuitive.

4.2 Analysis of residuals

The residuals, ε , associated with the PL approximations to the FT amplitude spectra are analyzed:

$$\varepsilon = |\tilde{X}_i^p(j\omega)| - a\omega^{-b}. \quad (13)$$

First, the histograms of the residuals are calculated adopting $Q = 50$ bins. The probabilities distributions, $p(\varepsilon)$ are then approximated by beta-functions and the corresponding parameters are depicted on a

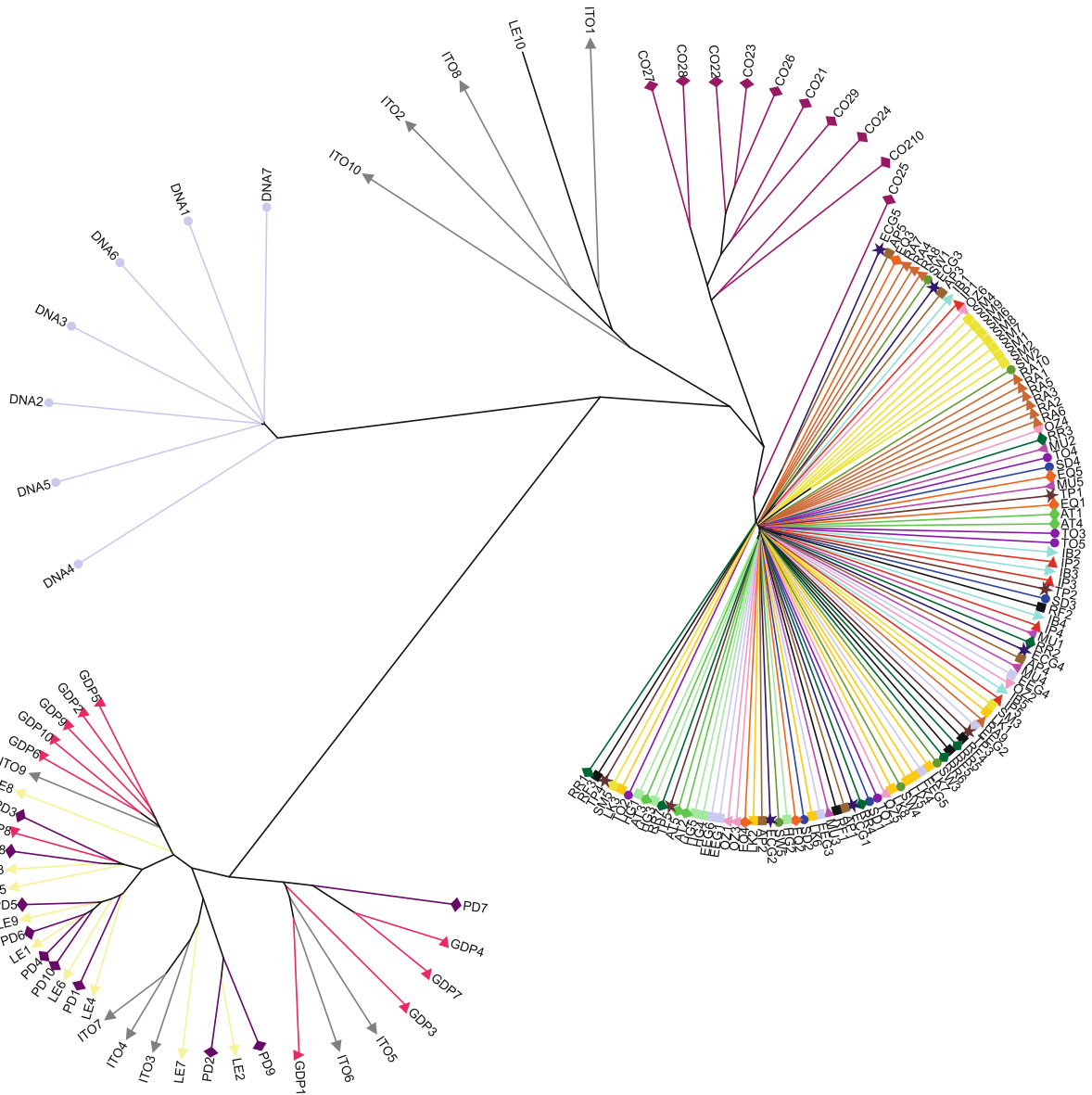


Fig. 12 Visualization tree comparing the time-series residuals

map. Second, we use the tools of Sect. 4.1 to compare the data.

Analytically, the beta-function is given by:

$$f(\varepsilon) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha) + \Gamma(\beta)} \varepsilon^{\alpha-1} (1 - \varepsilon)^{\beta-1} \quad (14)$$

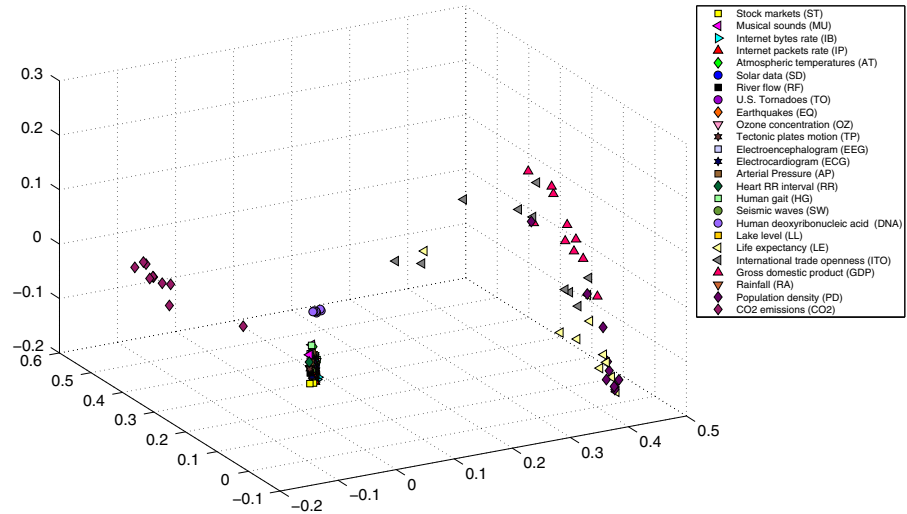
where, $\Gamma(\cdot)$ represent the gamma function and $(\alpha, \beta) \in \mathcal{R}$.

Figure 10 represents the histogram and beta-function approximation for the time-series $\tilde{x}_6^1(t)$, which corresponds to solar wind speed. The parameters $(\alpha, \beta) =$

$(9.2, 4.0)$. The global map of (α, β) parameters is depicted in Fig. 11.

In Figs. 12 and 13 we compare the whole time-series residuals by means of visualization tree and 3D MDS map, respectively. In all cases we observe the emergence of clusters. Nevertheless, the Fourier analysis proves to be superior since the map reveals a ‘symmetry’ that usually characterizes mathematical methodologies capable of leading to a deeper insight of the underlying phenomena. In conclusion, the Fourier

Fig. 13 Three dimensional MDS map comparing the time-series residuals



analysis proved to be a solid scientific approach in the endeavor of characterizing a large set of distinct phenomena.

5 Conclusions

Several natural and man-made complex systems were studied and compared in a global perspective. For each system, output signals representative of the systems dynamics were analyzed by means of the Fourier transform and PL trendlines. The time-series were compared and similarities were depicted in visualization trees and MDS maps. The proposed analysis and corresponding findings can further contribute to better understanding CS behavior and reveal similarities between systems of different nature.

Acknowledgments The authors acknowledge the inspiration of the paintings ‘The Persistence of Memory’ (http://en.wikipedia.org/wiki/The_Persistence_of_Memory) and ‘The Disintegration of the Persistence of Memory’ (http://en.wikipedia.org/wiki/The_Disintegration_of_the_Persistence_of_Memory), by Salvador Dali (http://en.wikipedia.org/wiki/Salvador_Dal%C3%AD).

References

- Aggarwal, C.C., Hinneburg, A., Keim, D.A.: On the Surprising Behavior of Distance Metrics in High Dimensional Space. Springer, New York (2001)
- Albert, R., Jeong, H., Barabási, A.L.: Error and attack tolerance of complex networks. *Nature* **406**(6794), 378–382 (2000)
- Amaral, L.A.N., Scala, A., Barthélemy, M., Stanley, H.E.: Classes of small-world networks. *Proc. Natl. Acad. Sci.* **97**(21), 11,149–11,152 (2000)
- Baleanu, D.: *Fractional Calculus: Models and Numerical Methods*, vol. 3. World Scientific, London (2012)
- Bar-Yam, Y.: *Dynamics of Complex Systems*, vol. 213. Addison-Wesley Reading, MA (1997)
- Barabási, A.L., Jeong, H., Nédá, Z., Ravasz, E., Schubert, A., Vicsek, T.: Evolution of the social network of scientific collaborations. *Phys. A Stat. Mech. Appl.* **311**(3), 590–614 (2002)
- Carlson, J.M., Langer, J.S., Shaw, B.E.: Dynamics of earthquake faults. *Rev. Mod. Phys.* **66**(2), 657 (1994)
- Cox, T.F., Cox, M.A.: *Multidimensional Scaling*. CRC Press, USA (2000)
- Denton, F.T., Feaver, C.H., Spencer, B.G.: Time series analysis and stochastic forecasting: an econometric study of mortality and life expectancy. *J. Popul. Econ.* **18**(2), 203–227 (2005)
- Edwards, R., LaDue, J.G., Ferree, J.T., Scharfenberg, K., Maier, C., Coulbourne, W.L.: Tornado intensity estimation: Past, present, and future. *Bull. Am. Meteorol. Soc.* **94**(5), 641–653 (2013)
- Ehrlich, P.R., Ehrlich, A.H.: *The population explosion*. Simon and Schuster, New York (1990)
- Goldberger, A.L., Amaral, L.A., Glass, L., Hausdorff, J.M., Ivanov, P.C., Mark, R.G., Mietus, J.E., Moody, G.B., Peng, C.K., Stanley, H.E.: Physiobank, physiotoolkit, and physiomet components of a new research resource for complex physiologic signals. *Circulation* **101**(23), e215–e220 (2000)
- Gutenberg, B., Richter, C.F.: Frequency of earthquakes in California. *Bull. Seismol. Soc. Am.* **34**(4), 185–188 (1944)
- Guzzetti, F., Malamud, B.D., Turcotte, D.L., Reichenbach, P.: Power-law correlations of landslide areas in central Italy. *Earth Planet. Sci. Lett.* **195**(3), 169–183 (2002)
- Hartigan, J.A.: *Clustering Algorithms*. Wiley, New York (1975)

16. Houze, R.A.: Structures of atmospheric precipitation systems: a global survey. *Radio Sci.* **16**(5), 671–689 (1981)
17. Huang, W.Q., Zhuang, X.T., Yao, S.: A network analysis of the Chinese stock market. *Phys. A Stat. Mech. Appl.* **388**(14), 2956–2964 (2009)
18. Huberman, B.A., Pirolli, P.L., Pitkow, J.E., Lukose, R.M.: Strong regularities in world wide web surfing. *Science* **280**(5360), 95–97 (1998)
19. Ionescu, C.M.: *The Human Respiratory System: An Analysis of the Interplay Between Anatomy, Structure, Breathing and Fractal Dynamics*. Springer, New York (2013)
20. Ionescu, C.M., De Keyser, R.: Relations between fractional-order model parameters and lung pathology in chronic obstructive pulmonary disease. *IEEE Trans. Biomed. Eng.* **56**(4), 978–987 (2009)
21. Johnson, N.F., Jefferies, P., Hui, P.M.: *Financial market complexity*. University Press, Oxford (2003)
22. Kato, A., Murai, J., Katsuno, S., Asami, T.: An internet traffic data repository: the architecture and the design policy. In: *Proceedings of INET'99* (1999)
23. Kenwood, A.G., Graff, M., Lougheed, A.: *Growth of the International Economy*, pp. 1820–2015. Routledge, London (2013)
24. Keshner, M.S.: $1/f$ Noise. *Proc. IEEE* **70**(3), 212–218 (1982)
25. Lee, C.K., Juang, L.C., Wang, C.C., Liao, Y.Y., Yu, C.C., Liu, Y.C., Ho, D.S.: Scaling characteristics in ozone concentration time series (octs). *Chemosphere* **62**(6), 934–946 (2006)
26. Lee, J.W.: International trade, distortions and long-run economic growth. *Staff. Pap. Int. Monet. Fund.* **40**(2), 299–328 (1993)
27. Lim, G., Kim, S., Kim, J., Kim, P., Kang, Y., Park, S., Park, I., Park, S.B., Kim, K.: Structure of a financial cross-correlation matrix under attack. *Phys. A Stat. Mech. Appl.* **388**(18), 3851–3858 (2009)
28. Lima, M.F., Machado, J., Costa, A.C.: A multidimensional scaling analysis of musical sounds based on pseudo phase plane. In: *Abstract and Applied Analysis*, vol. 2012. Hindawi Publishing Corporation (2012)
29. Lopes, A.M., Machado, J.T.: Dynamical behaviour of multi-particle large-scale systems. *Nonlinear Dyn.* **69**(3), 913–925 (2012)
30. Lopes, A.M., Tenreiro Machado, J.: Analysis of temperature time-series: embedding dynamics into the mds method. *Commun. Nonlinear Sci. Numer. Simul.* **19**(4), 851–871 (2014)
31. Lopes, A.M., Tenreiro Machado, J., Pinto, C., Galhano, A.M.: Fractional dynamics and mds visualization of earthquake phenomena. *Comput. Math. Appl.* **66**(5), 647–658 (2013)
32. Machado, J.A.T., Lopes, A.M.: Analysis and visualization of seismic data using mutual information. *Entropy* **15**(9), 3892–3909 (2013)
33. Machado, J.T.: And I say to myself: “What a fractional world”. *Fract. Calc. Appl. Anal.* **14**(4), 635–654 (2011)
34. Machado, J.T., Costa, A.C., Lima, M.F.: Dynamical analysis of compositions. *Nonlinear Dyn.* **65**(4), 399–412 (2011)
35. Mandelbrot, B.B.: *The Fractal Geometry of Nature*, vol. 173. Macmillan, New York (1983)
36. Mandelbrot, B.B., Van Ness, J.W.: Fractional brownian motions, fractional noises and applications. *SIAM Rev.* **10**(4), 422–437 (1968)
37. Mantegna, R.N., Stanley, H.E.: *Introduction to Econophysics: Correlations and Complexity in Finance*. Cambridge University Press, Cambridge (2000)
38. Nan, X., Jinghua, X.: The fractal dimension of eeg as a physical measure of conscious human brain activities. *Bull. Math. Biol.* **50**(5), 559–565 (1988)
39. Park, K., Willinger, W.: *The Internet as a Large-Scale Complex System*, vol. 3. Oxford University Press, Oxford (2005)
40. Pinto, C., Lopes, A.M., Machado, J.: A review of power laws in real life phenomena. *Commun. Nonlinear Sci. Numer. Simul.* **17**(9), 3558–3578 (2012)
41. Porporato, A., Ridolfi, L.: Nonlinear analysis of river flow time sequences. *Water Resour. Res.* **33**(6), 1353–1367 (1997)
42. Prawirodirdjo, L., Bock, Y.: Instantaneous global plate motion model from 12 years of continuous gps observations. *J. Geophys. Res. Solid Earth* (1978–2012) **109**(B8) (2004)
43. Robertazzi, T.G., Sarachik, P.: Self-organizing communication networks. *Commun. Mag. IEEE* **24**(1), 28–33 (1986)
44. Rose-Anderssen, C., Allen, P., Tsinopoulos, C., McCarthy, I.: Innovation in manufacturing as an evolutionary complex system. *Technovation* **25**(10), 1093–1105 (2005)
45. Shannon, C.E.: A mathematical theory of communication. *ACM SIGMOBILE Mob. Comput. Commun. Rev.* **5**(1), 3–55 (2001)
46. Shaw, B.E.: Generalized omori law for aftershocks and foreshocks from a simple dynamics. *Geophys. Res. Lett.* **20**(10), 907–910 (1993)
47. Simmons, K.M., Sutter, D., Pielke, R.: Normalized tornado damage in the United States: 1950–2011. *Environ. Hazards* **12**(2), 132–147 (2013)
48. Sivakumar, B., Jayawardena, A., Fernando, T.: River flow forecasting: use of phase-space reconstruction and artificial neural networks approaches. *J. Hydrol.* **265**(1), 225–245 (2002)
49. Solomon, S., Plattner, G.K., Knutti, R., Friedlingstein, P.: Irreversible climate change due to carbon dioxide emissions. *Proc. Natl. Acad. Sci.* **106**(6), 1704–1709 (2009)
50. Sornette, D., Pisarenko, V.: Fractal plate tectonics. *Geophys. Res. Lett.* **30**(3) (2003)
51. Strogatz, S.H.: Exploring complex networks. *Nature* **410**(6825), 268–276 (2001)
52. Tenreiro Machado, J.: Accessing complexity from genome information. *Commun. Nonlinear Sci. Numer. Simul.* **17**(6), 2237–2243 (2012)
53. Tenreiro Machado, J., Lopes, A.M.: Dynamical analysis of the global warming. *Math. Probl. Eng.* (2012)
54. Turcotte, D.L., Malamud, B.D.: Earthquakes as a complex system. *Int. Geophys. Ser.* **81**(A), 209–228 (2002)
55. Vazotsos, C., Kirk-Davidoff, D.: Long-memory processes in ozone and temperature variations at the region 60 s–60 n. *Atmos. Chem. Phys.* **6**(12), 4093–4100 (2006)
56. Vuglinskiy, V.: Assessment of the status of the development of the standards for the terrestrial essential climate variables, t4: Water level (water level in lakes and reservoirs, water storage). *Glob. Terr. Obs. Syst.* (2009)

57. Zhang, W.B.: Theory of complex systems and economic dynamics. *Nonlinear Dyn. Psychol. Life Sci.* **6**(2), 83–101 (2002)
58. Zhao, L., Zurbuchen, T., Fisk, L.: Global distribution of the solar wind during solar cycle 23: ACE observations. *Geophys. Res. Lett.* **36**(14) (2009)