

Fractal characterization of rain-gauge networks and precipitations: an application in Central Italy

Valerio Capecchi · Alfonso Crisci · Samantha Melani · Marco Morabito · Paolo Politi

Received: 25 October 2010 / Accepted: 4 August 2011 / Published online: 18 August 2011
© Springer-Verlag 2011

Abstract The measuring stations of a geophysical network are often spatially distributed in an inhomogeneous manner. The areal inhomogeneity can be well characterized by the fractal dimension D_H of the network, which is usually smaller than the euclidean dimension of the surface, this latter equal to 2. The resulting dimensional deficit, $(2 - D_H)$, is a measure of precipitating events which cannot be detected by the network. The aim of the present study is to estimate the fractal dimension of a rain-gauge network in Tuscany (Central Italy) and to relate its dimension to the dimensions of daily rainfall events detected by a mixed satellite/radar methodology. We find that $D_H \simeq 1.85$, while typical summer precipitations are characterized by a dimension much greater than the dimensional deficit 0.15.

1 Introduction

The distribution of a geophysical network is a multi-stage decision process which mainly relies on economic and demographic interests and on access problems in remote areas. Although an ideal network of stations should be spatially homogeneous and sufficiently dense to discriminate the minimum wave-length of the investigated geophysical phenomena, the irregularity and sparsity of observation points imply interpolation errors when reporting data on a regular grid. The areal clustering of point-sets can be measured by statistical indices as pointed out by Ouchi and Uekawa (1986) or, when the inter-station distances are scale-invariant, it can be well characterized thanks to a fractal analysis (Mandelbrot 1982).

If the point-set is self-similar, i.e. any small part of it is the magnified version of the whole set, the set is a fractal and it can be characterized by its fractal dimension D_H , which is a real number with $D_H < D_E$, where D_E is the standard euclidean dimension of the embedding space (in our case, $D_E = 2$). In literature several works (Korvin et al. 1990; Lovejoy et al. 1986; Mazzarella and Tranfaglia 2000; Olsson and Niemczynowicz 1996; Tessier et al. 1994) deal with the fractal characterization of a single-point observation network and sometimes this analysis is used as a method to drive an optimal enlargement of the network (Mazzarella and Tranfaglia 2000). Lovejoy et al. (1986) state that any sufficiently sparsely distributed phenomena having a fractal dimension smaller than the dimensional deficit $\delta = 2 - D_H$ of the observing network cannot be detected by the network itself. Since the sparse precipitating phenomena are the most intense and potentially severe, they are of prominent interest,

V. Capecchi (✉) · S. Melani
Institute of Biometeorology,
Via Madonna del Piano 10, Sesto Fiorentino,
Florence, Italy
e-mail: v.capecchi@ibimet.cnr.it

A. Crisci
Institute of Biometeorology, Via Caproni 6, Florence, Italy

M. Morabito
Interdepartmental Center of Bioclimatology,
University of Florence, Piazzale delle Cascine 18,
Florence, Italy

P. Politi
Istituto dei Sistemi Complessi, Consiglio Nazionale delle
Ricerche, Via Madonna del Piano 10, Sesto Fiorentino,
Florence, Italy

particularly when the network of measuring stations are constantly used for civil protection purposes. The aim of this study was to compute the fractal dimension D_H of the rain-gauge network belonging to the Centro Funzionale Regionale in Tuscany (Central Italy) and to compare D_H with the fractal dimension d of daily rainfall events occurring in the same area, using an independent network for rainfall, for the month of July 2010. We find that $D_H \simeq 1.85$, therefore giving a dimensional deficit $\delta \simeq 0.15$. On the other hand, all rain patterns give a fractal dimension $d > 0.6$, well above δ . A rough extrapolation of data for d as a function of 24-h rain thresholds suggests that our rain-gauge networks might fail to record precipitation events whose intensity is about 75 mm/day or more.

The paper is organized as follows. In Section 2 we detail the data used in this study and the methodology adopted to evaluate D_H . In Section 3 the computation of the fractal dimension of the rain-gauge network is presented and compared to those obtained for rainy days. Finally in Section 4 results are discussed with reference to the potentiality and limits of the applied methodology.

2 Data sources and methods

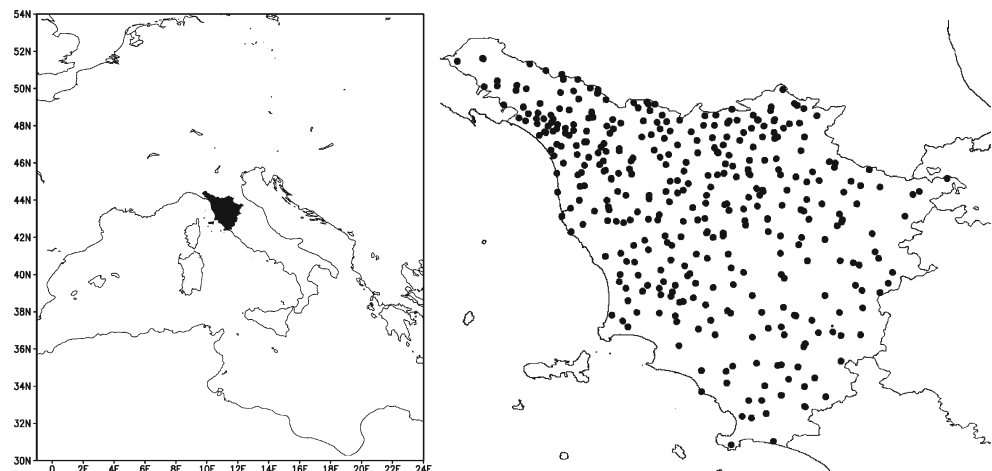
2.1 Data sources

The location of the rain-gauges belonging to the Centro Funzionale Regionale (CFR) in Tuscany (Central Italy) is shown in Fig. 1. Its establishment has been a long-term decision process involving several local institutions over more than 20 years. The network comprises 377 stations and encloses several basins over an area of about 23,000 km² (yielding a density of

about one station every 60 km²). More than 90% of the stations are located below 810 m. The biggest inter-station distance (in other words the size or diameter of the point-set) is about 250 km. These data make the geography of our network similar to that studied by Mazzarella and Tranfaglia (2000).

Satellite imagery acquired by the Meteosat Second Generation (MSG-2) satellite in the infrared (IR) channel centered at 10.8 μm was used in this study as a proxy to detect and monitor cold clouds systems. The study period is July 2010 while the spatial and temporal data resolutions are $4.5 \times 4.5 \text{ km}^2$ and 15 min, respectively. A brightness temperature (T_B) threshold was used to identify cold cloud systems that are most likely to be associated with convective activity. Kolios and Feidas (2007, 2010) used a T_B of 228 K to best identify convective systems in the Mediterranean area based on a set of lighting data. The same temperature threshold of 228 K was used by Morel and Senesi (2002) for assessing the climatology of the European MCSs, this value being very close to that of 221 K used by Garcia-Herera et al. (2005) for Spain. This low-temperature threshold allows to investigate mostly anvil regions and embedded areas of active deep convection (Johnson et al. 1990). In this study, a 228 K T_B threshold value was chosen for identifying very deep convective events over Tuscany. Furthermore, discrimination between precipitating and non-precipitating cold cloud systems, previously subjected to T_B threshold test, was performed using RADAR data provided by the DPCN (National Civil Protection Department) radar network. The data consist of a mosaic of instantaneous surface rainfall intensities (SRI) with a spatial and temporal resolution of 1 km and 15 min, respectively. Several daily precipitation amounts were tested; thresholds values of 1, 2, 5, 10, 15, 20, 25, 30, 35 mm per day were

Fig. 1 Location of rain gauges (right side of the picture)—belonging to the Centro Funzionale Regionale network in Tuscany (Central Italy, on left side of the picture)



used to analyze the different phenomenology linked to precipitation, from weak to moderate regimes.

2.2 Methodology

While euclidean geometry deals with ideal geometric forms and assigns dimension 0 to points, 1 to lines and so on, fractal geometry deals with non-integer dimensions. The fractal, or Hausdorff, dimension D_H has been the most common used measure of the *strangeness* of attractors of dissipative dynamical systems that exhibit chaotic behavior (Grassberger and Procaccia 1983b). Since for experimental data the value of D_H is difficult to determine using the box-counting algorithm (Strogatz 1994), we computed the fractal dimension D_2 of point-set using the method proposed in Grassberger and Procaccia (1983a, b), as also found in the literature (Korvin et al. 1990; Lovejoy et al. 1986; Mazzarella and Tranfaglia 2000; Olsson and Niemczynowicz 1996). In the case under examination we choose to use D_2 as a good approximation of D_H , since as stated in Grassberger and Procaccia (1983b), $D_2 \leq D_H$ and inequalities are rather tight in most cases.

In the present study we compute the correlation dimension in a 2-dimensional space but in general, to obtain D_2 given a point-set $\{\mathbf{X}_i\}_{i=1}^N$ with $\mathbf{X}_i \in \mathbf{R}^n$, we have to consider the correlation integral $C(R)$ that counts the number of pairs $\{\mathbf{X}_i, \mathbf{X}_j\}$ such that $\|\mathbf{X}_i - \mathbf{X}_j\|$ is smaller than a given threshold $R > 0$, with $\|\cdot\|$ being the standard euclidean distance in \mathbf{R}^n . In formulas:

$$C(R) = \frac{2}{N(N-1)} \sum_{i=1}^N \sum_{\substack{j=1 \\ j \neq i}}^N \Theta(R - \|\mathbf{X}_i - \mathbf{X}_j\|), \quad (1)$$

where Θ is the Heaviside function and where $\frac{2}{N(N-1)}$ is the normalization factor so that $C(R)$ tends to 1 for R tending to infinite.

If the rain-gauge network is a fractal then $C(R)$ grows like a power:

$$C(R) \propto R^{D_2}, \quad (2)$$

that is

$$\log(C(R)) \propto D_2 \log(R). \quad (3)$$

Therefore, one can derive D_2 from the regression coefficient of relationship (3).

In order to determine the correlation dimension D_2 of the rain-gauge network described previously, we computed the correlation function defined in Eq. 1, as described in Lovejoy et al. (1986), i.e. we determined

the cumulative frequency distribution of the inter-station distances for the total number of 377 stations. The distances were determined by spherical trigonometry, using geographic coordinates and ignoring elevations owing to the smallness of the elevation with respect to the two horizontal dimensions.

For what concerns the values of the parameter R , as done in Mazzarella and Tranfaglia (2000), we started in computing the inter-station distances defined in Eq. 1 from 1 km. This value was gradually increased by a factor of 1.1 up to 250 km since, as expected by definition of $C(R)$ given in Eq. 1, for all $R \geq$ size (area of interest) the correlation integral $C(R)$ saturates to 1 and $\log(C(R))$ saturates to 0.

For experimental data the linear behavior of $\log(C(R))$ on $\log(R)$ is limited to a scaling region S_R , i.e. only for R belonging to the interval $S_R = [R_{\min}, R_{\max}]$ (Strogatz 1994). This happens because $C(R)$ is underestimated from those points near the edge of the set so that the criteria to determine the bounds of S_R need to be analyzed in each singular case (Liebovitch and Toth 1989). According to the literature (Forrest and Witten 1979; Grassberger and Procaccia 1983a; Korvin et al. 1990), the upper limit R_{\max} is chosen equal to one third of the diameter of the area (about 80 km). In order to choose the lower limit R_{\min} , we didn't perform any statistical significance computation, since in our case

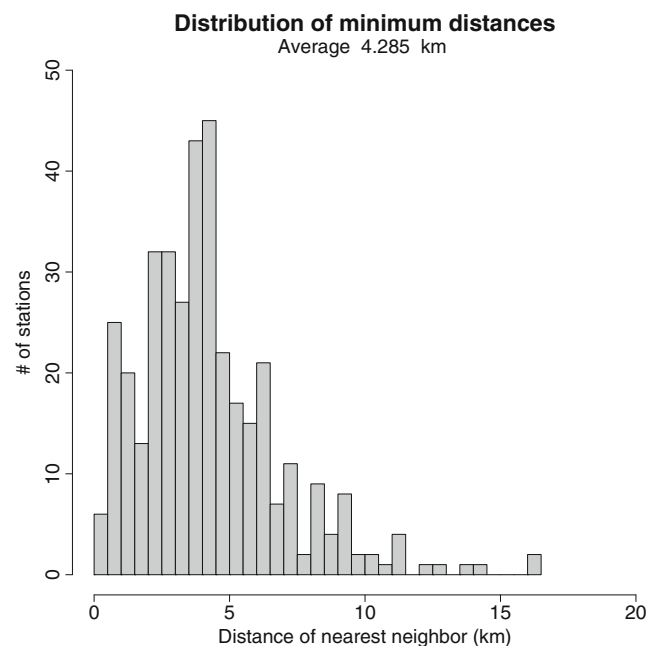


Fig. 2 Distribution of nearest neighbor's distances for each station in the point-set. Average value is 4.2 km which is taken as R_{\min} , the lower limit for the regression of $\log(C(R))$ on $\log(R)$

the correlation coefficients are statistically significant at 99% confident level for all $R \geq 1$ km. Rather, for each station we computed the distance of the nearest neighbor and took the average of this distribution as the meaningful index of the points separation.

In Fig. 2 we plot this distribution; the average of nearest neighbor's distances is about 4.2 km and this value is considered as the lower limit R_{\min} meaningful for the regression.

3 Results

The linear fitting between $\log(C(R))$ and $\log(R)$ within the scaling region S_R bounded by $R_{\min} = 4.2$ km and $R_{\max} = 80.2$ km yields a slope, and thus a correlation dimension value D_2 of 1.85. Figure 3 shows the results of this regression. The dimensional deficit δ of the network, defined as the difference between the dimension of the embedded space and D_2 , is $(2 - D_2) = 0.15$.

The value of δ should be related to the dimension d of rainfall phenomena. From a climatological point of view, the area of interest (central Italy) is mainly affected by convective storms or frontal systems, depending on the seasonality. Convective storms are of uppermost interest for our analysis since they are smaller, more or less separate rainfall areas displaying a considerable spatial variability and thus suitable for fractal analysis. They are typical of the warm season (from June to September roughly). The frontal storms are characterized by continuous rainfall areas of large

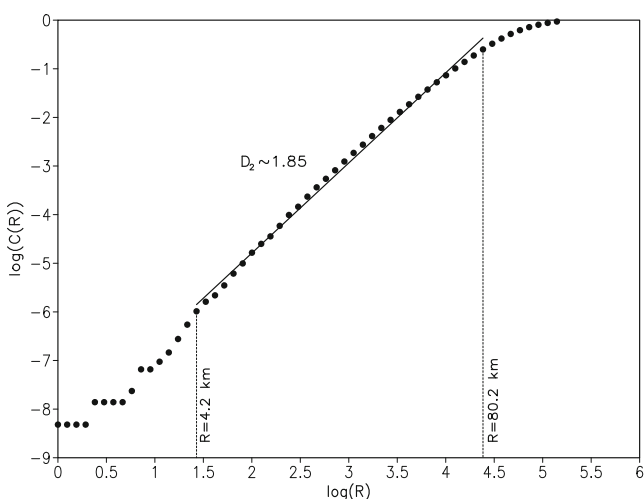


Fig. 3 Log-log plot of correlation integral $C(R)$ on R with scaling region delimited by $R_{\min} = 4.2$ km and $R_{\max} = 80.2$ km. The corresponding slope of the regression line which determines the correlation dimension D_2 is equal to 1.85

spatial extensions and are typical in autumn and winter seasons.

Using remote sensing and ground instruments described in Section 2 we collect data for every day in July 2010. First step is to select all the pixel in the MSG-2 15-min dataset having a brightness temperature below 228 K so that we can obtain a point-set (i.e. pixel-set) of potential precipitation cells (Kolios and Feidas 2010; Morel and Senesi 2002). Secondly, to assign a rain amount to the selected pixels we use the radar data. For each selected pixel in the MSG 15-min dataset we retrieve the surface rainfall intensities (SRI) as estimated by the RADAR data provided by the DPCN (National Civil Protection Department). Finally for each day of July 2010 we add all the 96 daily images (for each day we have one image every 15 min) and obtain a daily estimate of precipitation amount. Rain estimates were processed in order to compute the correlation dimension using the method detailed in Section 2. In Fig. 4 we plot the correlation dimensions of rainfall events registered in the month of July 2010 in Tuscany. Daily rainfall events were divided on the basis of prescribed thresholds, chosen equal to 1, 2, 5, 10, 15, 20, 30, and 35 mm. For each threshold, Fig. 4 shows the average and standard deviation values of correlation dimension of the rainy pixel-set for those days having a significant number of points that registered an amount of precipitation above the threshold.

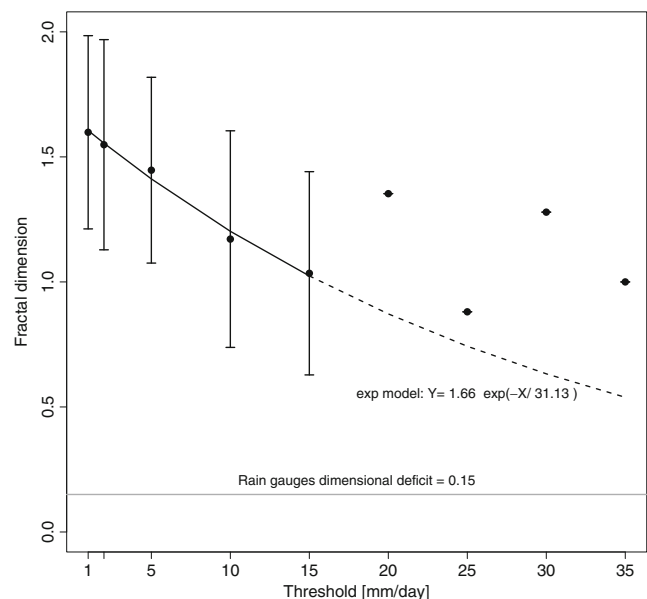


Fig. 4 Correlation dimensions of rainfall events (average values and standard deviations) registered in Tuscany in the month of July 2010 for different thresholds. Horizontal gray line is δ , the dimensional deficit of rain-gauge network

4 Discussion

The present study achieved the issue to estimate the areal sparseness of the monitoring rain-gauge network belonging to the CFR owned by Tuscany Administration by means of the fractal (correlation) dimension D_2 . In Table 1, we compare this value with dimension D_2 found in other, similar studies in the literature. Except the cases of Australia and Canada, where the dominance of inhabited areas along the coast lowers the value of the fractal dimension, our D_2 value is in good agreement with the others.

However, we have to point out that the computed correlation dimension D_2 must be handled with care because, according to the Tsonis criterion (Tsonis et al. 1994), the minimum number N_{\min} of points required to produce a correlation integral with no more than an error Err (normally $Err = 0.05 D_E$) is approximately

$$N_{\min} \propto 10^{2+0.4D_2}$$

which, in our case, means $N_{\min} \simeq 600$ whereas we have 377 stations.

In order to analyze the dimensional deficit $\delta = 0.15$, we need to compare it with the fractal dimension of rainfall events, as done in the previous section. As for the precipitations fallen in July 2010, even the more intense ones were characterized by a fractal dimension $d > 0.6$, much greater than the dimensional deficit $\delta = 0.15$. We stress that the used rain data (see Fig. 4) are *independent* from the rain-gauge network used to determine the fractal dimension $D_H = 1.85$. Therefore, these preliminary results allow us to state with a good confidence that our rain-gauge network was precise enough to record all precipitation events occurred in July 2010. Empirically we can suppose that the fractal dimension goes to zero as the threshold increases; this

Table 1 Comparison of D_2 computed in the present study with other correlation dimensions found in literature

Reference	# of points	Area coverage	D_2
Lovejoy et al. (1986)	9,563	Global land	1.75
	3,593	France	$\simeq 1.8$
	414	Canada	$\simeq 1.5$
Korvin et al. (1990)	$\simeq 65,000$	Australia	1.42
Tessier et al. (1994)	7,983	Global land	1.79
Olsson and Niemczynowicz (1996)	230	$\simeq 10,000 \text{ km}^2$	$\simeq 2$
Mazzarella and Tranfaglia (2000)	215	$\simeq 38,000 \text{ km}^2$	1.84
	300	$\simeq 38,000 \text{ km}^2$	1.89
Present study	377	$\simeq 23,000 \text{ km}^2$	1.85

It is also reported the bibliography references, the number of points taken into account and the extent of geographical area

because intense precipitation events (at least thermo-convective ones) are more scattered. On the other hand light rains are more homogeneous and then associated to a (decreasing) linear trend for small thresholds. We then choose to fit the values plotted in Fig. 4 with an exponential model which is almost linear for small values of x and decreases to zero as x tends to infinity. We can then suppose that

$$y = Ae^{-x/x_0}$$

where the independent variable x on the x -axis represents the daily amount of rain, y represents the fractal dimensions and the parameters are $A = 1.657$ and $x_0 = 31.133$. The model is calibrated using the first values of x (thresholds from 1 mm/day up to 15 mm/day, continuous line in Fig. 4), since above we don't have any significant statistics (just two days registering at least 20 pixels with a precipitation above 20 mm and one day registering at least 20 pixels with a precipitation above 25 mm).

The exponential regression intersects $\delta = 0.15$ for

$$x_{\delta} = -x_0 \ln \left(\frac{\delta}{A} \right)$$

that is $x_{0.15} \simeq 75 \text{ mm/day}$. In other words our data suggests that rainfalls with daily amount equal or above 75 mm/day might correspond to a fractal dimension $d < \delta$, so that these events could not be detected. This value is based, by construction, on the remote sensed data and ground instruments and on the phenomenology of rainfall events. Ongoing efforts are directed toward the improvements of the accuracy of instruments and toward the calibration of the algorithms.

Further studies are required to investigate the relationship between correlation dimension D_2 of the observing network (but we are rather interested in the dimensional deficit δ) and the dimensions d of rainfall events. Firstly the most important improvement of the research is to expand the statistics of precipitating events considering several months for, at least, a couple of years. Moreover it would be interesting to take into consideration the fall/winter precipitations, which are mainly associated with cold and warm fronts, to evaluate the different behavior of dimension d and check if, for some thresholds, it drops below δ .

Acknowledgements MSG imagery is copyright of EUMETSAT and was made available by the EUMETSAT on-line Archive. We thank National Department of Civil Protection (DPCN) for providing weather radar data. The authors are grateful to Andrea Antonini and Stefano Romanelli for preprocessing and providing remote sensed data. In addition special thanks go to Melissa Morris for the revision of the text. Partial financial support by Regione Toscana is gratefully acknowledged.

References

- Forrest SR, Witten TA (1979) Long-range correlations in smoke-particle aggregates. *J Phys A (Math Gen)* 12:L109–L117
- Garcia-Herera R, Hernandez E, Paredes D, Barriopedro D, Correoso J, Prieto L (2005) A mascote-based characterization of MCSs over Spain, 2000–2002. *Atmos Res* 73:261–282
- Grassberger P, Procaccia I (1983a) Characterization of strange attractors. *Phys Rev Lett* 50:346–349
- Grassberger P, Procaccia I (1983b) Measuring the strangeness of strange attractors. *Physica D* 9:189–208
- Johnson RH, Gallus Jr WA, Vescio MD (1990) Near-tropopause vertical motion within the trailing stratiform region of a mid-latitude squall line. *J Atmos Sci* 47:2200–2210
- Kolios S, Feidas H (2007) Correlation of lightning activity with spectral features of clouds in Meteosat-8 imagery over the Mediterranean basin. In: *Proceedings of the 8th Pan-Hellenic geographical congress, Athens, Greece*
- Kolios S, Feidas H (2010) A warm season climatology of mesoscale convective systems in the Mediterranean basin using satellite data. *Theor Appl Climatol* 102:29–42
- Korvin G, Boyd DM, O'Dowd R (1990) Fractal characterization of the south australian gravity station network. *Geophys J Int* 100:535–539
- Liebovitch LS, Toth T (1989) A fast algorithm to determine fractal dimensions by box counting. *Phys Lett A* 141:386–390
- Lovejoy S, Schertzer D, Ladoy P (1986) Fractal characterization of inhomogeneous geophysical measuring networks. *Nature* 319:43–44
- Mandelbrot BB (1982) *The fractal geometry of nature*. Freeman, San Francisco
- Mazzarella A, Tranfaglia G (2000) Fractal characterisation of geophysical measuring networks and its implication for an optimal location of additional stations: an application to a rain-gauge network. *Theor Appl Climatol* 65:157–163
- Morel C, Senesi S (2002) A climatology of mesoscale convective systems over europe using satellite infrared imagery. II. Characteristics of European mesoscale convective systems. *Q J R Meteorol Soc* 128:1973–1995
- Olsson J, Niemczynowicz J (1996) Multifractal analysis of daily spatial rainfall distributions. *J Hydrol* 187(1–2):29–43
- Ouchi T, Uekawa T (1986) Statistical analysis of the spatial distribution of earthquakes-variation of the spatial distribution of earthquakes before and after large earthquakes. *Phys Earth Planet Inter* 44:211–225
- Strogatz SH (1994) *Nonlinear dynamics and chaos: with applications to physics, biology, chemistry and engineering*. Perseus Books, Reading, pp 409–412
- Tessier Y, Lovejoy S, Schertzer D (1994) Multifractal analysis and simulation of the global meteorological network. *J Appl Meteorol* 33(12):1572–1586
- Tsonis AA, Triantafyllou GN, Elsner JG (1994) Searching for determinism in observed data: a review of the issue involved. *Nonlinear Process Geophys* 1:12–25