INFORMATION LOSS UNDER PROJECTIONS FROM SEISMIC DATA

Esquivel, Javier
University of Granada, Department of Statistics and Operations Research
Faculty of Sciences, Campus Fuentenueva s/n
Granada, E-18071, Spain
E-mail: jesquivel@ugr.es

Angulo, Jose M.
University of Granada, Department of Statistics and Operations Research
Faculty of Sciences, Campus Fuentenueva s/n
Granada, E-18071, Spain
E-mail: jmangulo@ugr.es

Analysis of discrete events occurring in space-time is often performed by projecting the available historical data on space or time, thus reducing dimensionality to enable studying certain characteristics of interest such as spatial distribution or temporal frequency patterns. Another case is the projection from 3D coordinates to 2D maps, to facilitate visualization and analysis. Necessarily, data projections imply a certain loss of information which will depend on the structural properties of the underlying generating process. In this paper, we are interested in the study of this effect in relation to real phenomena displaying multifractal behavior such as seismic activity. More specifically, 3D and 2D multifractal analyses, based on singularity and multifractality spectra, are applied to study the information loss derived from projection of spatio-temporal seismic data by removing the time component. A similar study is applied to compare the sets of seismic data obtained with and without considering the magnitude attribute to each event. The analysis is performed on a seismic sequence in the Agron area located at southern Spain, within the period 1988-1989.

Introduction

The aim of this work is to study the information loss derived from projection on space or on time of available data of discrete events occurring in space-time. In recent works, earthquake distribution in space has been studied showing fractal properties on various scales. Besides, a recent work reveals multifractal patterns of seismicity in Greece seismicity (Dimitriu *et al.* 2000).

In this work, we apply multifractal analysis techniques (multifractality and singularity spectra), jointly with entropic analysis, to test information loss comparing the data before and after projection.

Methodological aspects

- Multifractal analysis

  Scale invariance is the main feature of the fractal theory. This term indicates that specific characteristics of a system are independent from the size of the magnitude of the scale at which it is analyzed. Fractals are also known as monofractals and it have only one dimension. On the one hand, when the system behavior can be described with only one scaling dimension this behavior is denominated monofractal, this is the information dimension. On the other hand, when the system behavior needs more than one dimension for being described it is denominated multifractal.

Multifractality and singularity spectra are the most extended multifractal analysis tools in applications, both techniques being based on a similar approach. There are several procedures to obtain both spectra. In this work we adopt the box-counting procedure, which is based on obtaining a probability mass function reflecting the space-time and/or magnitude distribution of the data set. For that purpose, the area under study is covered by $N(\varepsilon)$ boxes of size ε, and p_i is defined as the ratio of the number of events that fall into box i to the total number of events in the set.

For obtaining both spectra we use the generalized "q" representation of the partition function:

$$Z_q(\varepsilon) = \sum_{i=1}^{N(\varepsilon)} p_i^q(\varepsilon).$$

Its average value behaves at the origin, that is, with $\varepsilon \to 0$, as

$$E[Z_q(\varepsilon)] \sim \varepsilon^{\tau(q)}.$$

The generalized dimensions are defined as

$$D_q = \frac{1}{q-1}\tau(q), \quad \text{where} \quad \tau(q) = \lim_{\varepsilon \to 0} \frac{\log E[Z_q(\varepsilon)]}{\log \varepsilon}.$$

The multifractality spectrum is the representation of D_q vs. q.

The singularity spectrum $f(\alpha)$ is given by the Legendre transform of the Hölder exponents $\alpha(q) = \tau'(q)$:

$$f(\alpha(q)) = q\alpha(q) - \tau(q).$$

To calculate $f(\alpha)$ we use the formalism exposed by Chhabra and Jensen (1989), given by the forms

$$\alpha(q) = \lim_{\varepsilon \to 0} \frac{\log E\left[\sum_{i=1}^{N(\varepsilon)} \tilde{p}_i \log p_i(\varepsilon)\right]}{\log \varepsilon}, \qquad f(q) = \lim_{\varepsilon \to 0} \frac{\log E\left[\sum_{i=1}^{N(\varepsilon)} \tilde{p}_i \log \tilde{p}_i(\varepsilon)\right]}{\log \varepsilon},$$

with

$$\tilde{p}_i(\varepsilon) = \frac{p_i^q(\varepsilon)}{\sum_{j=1}^{N(\varepsilon)} p_j^q(\varepsilon)}.$$

These limits do not have a direct solution but, recently, several methods based on log-log fit have been proposed for their calculation. This methodology consists in carrying out a linear fit to the log-log representation. In the case of the multifractality spectrum, the procedure is:

$$\tau(q) = \lim_{\varepsilon \to 0} \frac{\log E[Z_q(\varepsilon)]}{\log \varepsilon} \longrightarrow \tau(q) = \frac{\log E[Z_q(\varepsilon)]}{\log \varepsilon} + error(\alpha) \longrightarrow$$

$$\tau(q)\log\varepsilon - error(\alpha)\log\varepsilon = \log[Z_q(\varepsilon)].$$

The spectra interpretations are:

i) *Multifractality spectrum and the generalized dimensions*

The multifractality spectrum indicates the main characteristics of the multifractal behavior. The behavior of the data set is multifractal if $D_q > D_{q'}$ for $q' > q$. The larger the difference between the maximum and the minimum the stronger the multifractal behavior of the data set. In the case where D_q is constant the data set presents monofractal behavior.

The fractal dimensions indicate different multifractal characteristics of interest. The most important ones are:

* Capacity dimension, D_0. It shows how the points of the data set fill the area under study. The larger the value of this dimension the better the space of the studied area is covered.
* Information dimension, D_1. It is a measure of order-disorder of the points of the data set in the area under study. Large values indicate high disorder.
* Correlation dimension, D_2. It indicates how is the clustering-inhibition patterns of the points of the data set in the studied area. Low values correspond to high level of clustering.
* Multifractal step, $D_{-\infty} - D_{\infty}$. It shows the level of the multifractal behavior. Large values indicate a strong multifractal behavior. On the contrary, low values correspond to mainly monofractal behavior.

ii) *Singularity spectrum, $f(\alpha)$ vs. α*

The singularity spectrum shows different aspects. Its range indicates the level of multifractal behavior. In the case of a monofractal structure we find scale-invariance, the spectrum concentrates on. On the contrary, a wide range indicates high multifractal behavior. The maximum of the curve coincides with the capacity dimension. In general, the left-hand branch corresponds to $q > 0$ and the right-hand branch to $q < 0$. When increasing the value of q the effect of the larger probabilities also increases. For this reason, the left-hand branch is related with the degree of spatial clustering, and the curves whose left-hand branch has a slow decay correspond to strong spatial clustering patterns.

- Entropy analysis

The concept of entropy can be understood as the uncertainty associated to a system. This uncertainty can be explained as the incapacity of predicting one event by means of one occurred before. The most commonly used measure of entropy is the Shannon entropy. High values of the Shannon entropy correspond to high levels of uncertainty and disinformation. Its expression for a given discrete distribution $(p_1, ..., p_n)$ is:

$$H = -\sum_{i=1}^{n} p_i \ln p_i.$$

In our case, n is the number of boxes that cover the area under study, and p_i is the proportion of events falling into the box i.

The entropy is maximized when all probabilities are equal, $p_i = \frac{1}{n}$. In this case,

$$H_{max} = -\sum_{i=1}^{n} \frac{1}{n} \ln \frac{1}{n} = -\ln \frac{1}{n} = \ln n.$$

The value of the Shannon entropy increases with the number of boxes. For this reason, we work with the configuration entropy, which is defined as

$$H_{config} = \frac{H}{H_{max}} = -\frac{\sum_{i=1}^{n} p_i \ln p_i}{\ln n}.$$

A configuration entropy value close to 1 indicates a highly uniform distribution od points. Conversely, small values of the configuration entropy correspond to a high degree of heterogeneity, that is, to strong spatial clustering.

Real data application

The above tools are applied to a real seismic data sequence, which comprises a seismic swarm of about 256 earthquakes of magnitude between 1 and 4.1 in the Richter scale occurred during 241 days within the period November 1988 to April 1989 near the village of Agron (Granada province, Andalusia, southern Spain).

Firstly, we perform the analysis considering the geometrical and temporal components (3D), and the results are compared the those obtained by considering only the geographical coordinates (2D), without the effect of the time component. Also, to check for possible temporal heterogeneity, we divide the data set into three subsets corresponding to the days 1-77 (133 earthquakes), 78-160 (64 earthquakes) and 161-241 (59 earthquakes), and obtain results restricted to each one of the three clusters.

Secondly, we compar the results derived by performing the analysis on geographical and magnitude components (3D) to those obtained considering only the geographical coordinates. Again, the set is further divided into three clusters corresponding to the 154 earthquakes with magnitude lower than 1.8, the 73 earthquakes with magnitude above 2.4, and 29 earthquakes with magnitude above 2.4. In all cases, we study the loss of information caused by projection.

- Case 1: Geographical and temporal components

  In Figure 1 we can see both spectra,the singularity and multifractality spectra. These show that the first cluster has stronger multifractal behaviour than other ones. In the singularity spectrum we can see that the second and the third clusters cover well the space because their maximum value, that is, the capacity dimension, is large and present a high degree of homogeneity reflected in the sudden drop of the left-hand branch of the singularity spectrum. These aspects can be also seen in table 1 containing the values of the generalized multifractal dimensions. The values of the steps show that the first cluster has the stronger multifractal behavior. The capacity dimension of the first cluster indicates that it does not cover well the space, contrary to the second and third clusters. The low values of the information and correlation dimensions indicate that this cluster has a strong spatial clustering pattern and is heterogeneous.

  These remarks are corroborated by the entropic analysis. In Table 2 we observe that the value of the configuration entropy of the three variables jointly is not sufficiently high and, for that

Figure 1. Multifractality and singularity spectra. Case 1

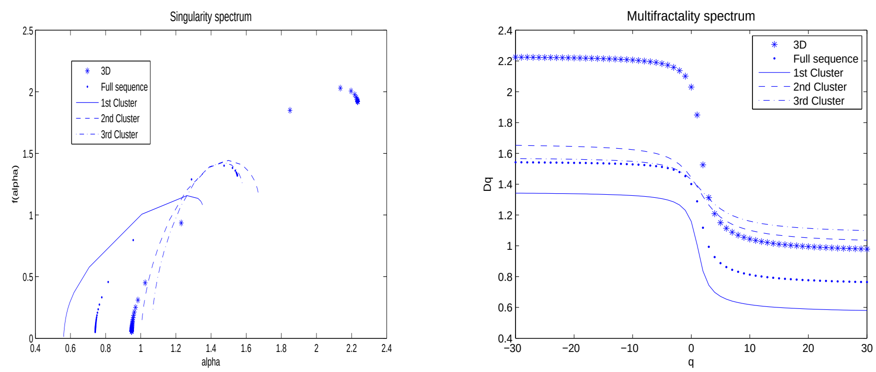

Table 1. Generalized fractal dimensions. Case 1

	D_0	D_1	D_2	Step
3D Analysis	2.0302	1.8493	1.5253	1.2458
Full sequence	1.4008	1.2888	1.1174	0.7778
1^{st} Cluster	1.1579	1.0061	0.8359	0.7620
2^{nd} Cluster	1.4436	1.3840	1.3209	0.6164
3^{rd} Cluster	1.4298	1.3907	1.3478	0.4681

reason, in the case of carrying out the projection will be lost important information. The entropic analysis also indicates that, in the case of performing a projection the information loss is minimum when the temporal variable is removed instead of any of geographical coordinates.

Table 2. Configuration entropies. Case 1

X-Y-T	X-Y	X-T	Y-T
0.7286	0.7477	0.8675	0.8345

- Case 2: Geographical and magnitude components

  In this case, the singularity and multifractality spectra displayed in Figure 2 show that the first cluster has a slightly stronger multifractal behavior than the other ones. On the other hand, the singularity spectrum shows that the second cluster covers better the space and presents a higher degree of homogeneity than the other ones. These remarks are also confirmed by the generalized multifractal dimension values given in Table 3.

  Furthermore, from entropic analysis, the values of configuration entropy in Table 4 indicates that significant information is loss under projection, with the minimum being obtained in the case where the geographical coordinates is removed.

Conclusions

In this paper, based on the singularity and multifractality spectra, we show the effects of projec-

Figure 2. Multifractality and singularity spectra. Case 2

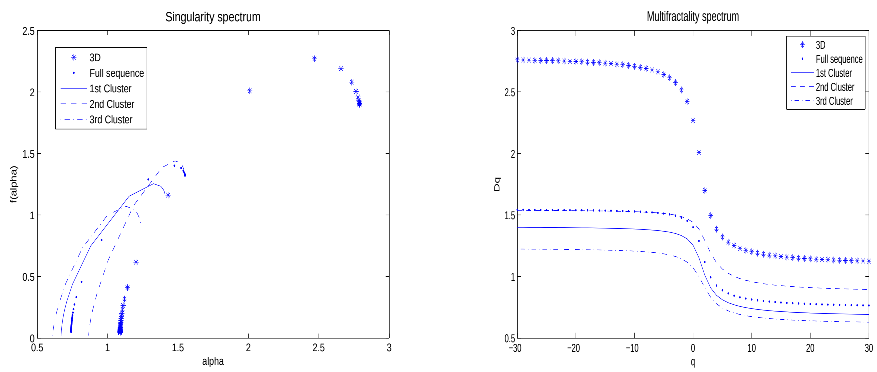

Table 3. Generalized fractal dimensions. Case 2

	D_0	D_1	D_2	Step
3D Analysis	2.2690	2.0088	1.6979	1.6357
Full sequence	1.4008	1.2888	1.1174	0.7778
1^{st} Cluster	1.2543	1.1524	1.0113	0.7081
2^{nd} Cluster	1.4397	1.3855	1.2928	0.6433
3^{rd} Cluster	1.0714	0.9995	0.9090	0.5944

tion of a multidimensional set of data consisting of earthquake events identified in terms of geographical coordinates and the time of occurrence, as well as the magnitude, by removing one of the components.

In particular, significant heterogeneities associated to the time or the magnitude dimension, which may be of potential interest for explaining the structuring of the data, are ignored as a consequence of projection. The entropic analysis allows to quantify the information loss derived.

Acknowledgements

This paper has been partially supported by projects MTM2009-13250 of the SGPI, and P08-FQM-03834 of the Andalusian CICE, Spain. The authors also thank the Andalusian Institute of Geophysics for providing the data and related review results.

REFERENCES

Blachers, S., Frouers, F. and Ananthakrishna, G., 1992, Multifractal analysis, a method to investigate the morphology of materials. Physica A, 185, 28-34.

Chhabra, A., Jensen, R.V., 1989, Direct determination of the $f(\alpha)$ singularity spectrum. Physics Review Letter, 62, 1327-1330.

Dimitriu, P.P., Scordilis, E.M. and Karacostas, V.G., 2000, Multifractal analysis of the Arnea, Greece, seismicity with potential implications for earthquake prediction. Natural Hazards, 21, 277-295.

Saccorotti, G., Carmona, E., Ibáñez, J.M. and Del Pezzo, E., 2002, Spatial characterization of Agron, southern Spain, 1988-1989 seismic series. Physics of the Earth and Planetary Interiors, 129, 13-29.

Table 4. Configuration entropies. Case 2

X-Y-M	X-Y	X-M	Y-M
0.6959	0.7910	0.7549	0.7128