A New Estimator of Intrinsic Dimension Based on the Multipoint Morisita Index

Jean GOLAY and Mikhail KANEVSKI

Institute of Earth Surface Dynamics, Faculty of Geosciences and Environment, University of Lausanne, Switzerland. Contact: jean.golay@unil.ch.

Abstract

The size of datasets has been increasing rapidly both in terms of number of variables and number of events. As a result, the empty space phenomenon and the curse of dimensionality complicate the extraction of useful information. But, in general, data lie on non-linear manifolds of much lower dimension than that of the spaces in which they are embedded. In many pattern recognition tasks, learning these manifolds is a key issue and it requires the knowledge of their true intrinsic dimension. This paper introduces a new estimator of intrinsic dimension based on the multipoint Morisita index. It is applied to synthetic data sets of varying complexities and comparisons with other existing estimators are carried out. The proposed estimator turns out to be fairly robust to sample size and noise, unaffected by edge effects, able to handle large datasets and computationally efficient.

Keywords: Intrinsic dimension, Multipoint Morisita index, Fractal dimension, Multifractality

1. Introduction

The 21st century is more and more data-dependent and, in general, when collecting data for a particular purpose, it is not known which variables matter the most. This lack of knowledge leads to the emergence of high-dimensional data sets characterized by redundant features which artificially increase the volume of data to be processed. As a result, the empty space phenomenon [1] and the curse of dimensionality [2] make it challenging to conduct pattern recognition tasks such as clustering and classification.
The goal of dimensionality reduction (DR) \cite{3, 4}, also called manifold learning in the non-linear case, is to address this issue by mapping the data into the lower dimensional space where they truly lie. Such a space is often considered as a manifold of intrinsic dimension (ID) \( M \) embedded in a Euclidean space of dimension \( E \) with \( E \geq M \). \( E \) equals the number of variables of a dataset and the ID of a manifold is equal to the ID of the data when the number \( N \) of sampled data points tends to \( \infty \), i.e. the available points are sampled from a manifold whose ID is the theoretical ID of the data. If a point set is space-filling, the dimension of the data manifold \( M \approx E \). In contrast, if the Euclidean space is not entirely covered, \( M < E \).

The optimality of DR greatly depends on the accuracy of ID estimates. An underestimation of the theoretical ID will result in the implosion of the data manifold and information will be irreparably lost. On the contrary, an overestimation will lead to noise in the final mapping. From an application perspective, DR can be used to produce low dimensional syntheses of high dimensional datasets \cite{5} and as a preprocessing tool for supervised learning \cite{6, 7} and data visualization \cite{8}.

DR methods perform variable transformations to capture the complex dependencies generating redundancy. Nevertheless, it is often important not to recast data. The fractal dimension reduction (FDR) algorithm \cite{9, 10, 11, 12} was proposed to take such a requisite into account. The fundamental idea is to drop all the variables which do not affect the fractal-based ID estimate of a dataset. FDR can also be adapted to supervised feature selection methods \cite{13}. The goal is then to reject irrelevant or redundant variables (or features) according to a prediction task (i.e. regression or classification). Although ID estimation lies at the core of FDR, more traditional unsupervised \cite{14, 15, 16} and supervised \cite{17, 18, 19, 20, 21, 22} feature selection methods do not rely heavily on it. It has, however, a great potential in speeding up search strategies \cite{23, 24, 25, 26}.

These different approaches highlight that ID estimation is a fundamental problem when dealing with high-dimensional datasets. This consideration is supported by the statistical learning theory which states that the power of a classifier depends on ID \cite{27, 28}. Unfortunately, ID estimators \cite{29, 30} suffers from the curse of dimensionality as well. Their overall performance depends on many factors (to various degree), such as the number of data

\footnote{In Physics, mainly, \( M \) is often referred to as the degrees of freedom of data}
points, the theoretical ID of data and the shape of manifolds. This present research deals with a new ID estimator in order to provide a solution to the problems raised by these factors. It is based on the recently introduced multipoint Morisita index \((m\text{-Morisita})\)\(^{31,32,33}\). The \(m\text{-Morisita} \) index is a measure of global clustering closely related to the concept of multifractality and, so far, it has been successfully applied within the framework of (2-dimensional) spatial data analysis \(^{31,32}\).

The paper is organized as follows: In Section 2, traditional fractal-based and maximum likelihood methods of ID estimation are presented. Section 3 derives a new ID estimator from the \(m\text{-Morisita} \) index and introduces a new algorithm for its applications to high-dimensional datasets. Section 4 is devoted to comparisons of the proposed estimator and those of Section 3. A characterization of their behaviour regarding sample sizes, noise and the dimension of manifolds is analysed. A special attention is also paid to their bias and variance using Monte-Carlo simulations. Finally, conclusions are drawn in Section 5.

2. Existing Methods

Many ID estimation methods have been proposed \(^{29,30,34,35}\) and they can be roughly divided into projection (e.g. PCA) and geometric methods (e.g. fractal, nearest-neighbor and MLE methods). This section focuses on fractal-based and maximum likelihood estimators. They are commonly used in a wide range of applications and they generally provide non-integer values as ID estimates.

2.1. Fractal-Based Estimation Methods

The word \textit{fractal} was first coined by B. Mandelbrot \(^{36}\) to describe scale-invariant sets with abrupt and tortuous edges. At small scales \(\delta\), for a given point pattern, one has that:

\[
n_{box}(\delta) \propto \delta^{-D_0}
\]

where \(n_{box}(\delta)\) is the number of grid cells necessary to cover the whole pattern and \(D_0\) is known as the box-counting dimension \(^{36,37,38}\). In practical applications, due to its simplicity, \(D_0\) often replaces the Hausdorff dimension \(D\) (or fractal dimension) and, mathematically, it can be proved that \(D_0\) is an upper bound of \(D\) \(^{39}\). From the perspective of pattern recognition,
$D_0$ can be regarded as the original fractal-based estimator of ID, since it is purely geometric.

In complex cases, the scaling behaviour of all the moments of the point distributions cannot be fully characterized by only one fractal dimension and a full spectrum of generalized dimensions, $D_q$, is required to describe these so-called multifractal sets [10, 11, 12, 13]. $D_q$ is generally obtained using a generalization of the box-counting method [10, 11, 12, 14] based on Rényi’s information, $RI_q(\delta)$, of $q^{th}$ order [45]. The central power law of this approach can be written as follows for $q \neq 1$:

$$\exp(RI_q(\delta)) \propto \delta^{-D_q}$$

(2)

where

$$RI_q(\delta) = \frac{1}{1 - q} \log \left( \sum_{i=1}^{n_{\text{box}}(\delta)} p_i(\delta)^q \right)$$

(3)

In this last equation, $p_i(\delta) = n_i/N$ is the value of the probability mass function in the $i^{th}$ grid cell of size $\delta$ ($n_i$ is the number of points falling into the $i^{th}$ cell) and $q \in \mathbb{R}\setminus\{-1\}$. Finally, one has that:

$$D_q = \lim_{\delta \to 0} \frac{RI_q(\delta)}{\log \left( \frac{1}{\delta} \right)}$$

(4)

and

$$\lim_{q \to 1} D_q = d_{fi}$$

(5)

$$D_2 = d_{cor}$$

(6)

where $d_{fi}$ and $d_{cor}$ are, respectively, the information dimension [46, 40] and the correlation dimension [47].

Usually, $d_{cor}$ is computed using the Grassberger-Procaccia (GP) algorithm [47]. This algorithm is designed to better take advantage of the range of available pairwise distances between points. It can be introduced as follows: at small scales, for a point set, $X_N = \{x_1, \ldots, x_N\}$, one has that

$$C(\delta) \propto \delta^{d_{cor}}$$

(7)

where

$$C(\delta) = \frac{2}{N(N-1)} \sum_{i=1}^{N} \sum_{j=i+1}^{N} 1\{\|x_i - x_j\| \leq \delta\}$$

(8)
and

\[ df_{cor} = D_2 = \lim_{\delta \to 0} \frac{\log(C(\delta))}{\log(\delta)} \]  \hspace{1cm} (9)

The available values of \( RI_q(\delta) \) depend on the data resolution and a commonly used method for estimating \( -D_q \) consists in plotting \( RI_q(\delta) \) vs \( \log(\delta) \) for a chosen scale interval. The final estimate is then the slope of the linear regression fitting at best the linear part of the resulting chart. The procedure is the same for the GP algorithm, except that \( df_{cor} \) and \( \log(C(\delta)) \) replace, respectively, \( -D_q \) and \( RI_q(\delta) \). Eventually, both \( D_q \) (in general \( 0 \leq q \leq 2 \)) and \( df_{cor} \) can be used as ID estimators.

Although these methods may entail some disadvantages due to the finiteness of datasets \cite{35}, they are widely spread and have been successfully applied in various fields, such as spatial \cite{48,49} and time series \cite{50} analysis, cosmology \cite{51}, climatology \cite{52} and pattern recognition \cite{53,54}. They have also been used in different procedures improving their overall performance \cite{55,56}.

2.2. Maximum Likelihood Estimation Methods

The maximum likelihood estimation (MLE) of ID was introduced in \cite{30}. The proposed method relies on the assumption that the \( k \)-nearest neighbors (\( k \)-NN) of any point \( x_i \) of a point set \( X_N = \{x_1, \ldots, x_N\} \) are stemming from a uniform probability density function \( f(x_i) \). As a consequence, for a fixed \( x_i \), the observations are treated as a homogeneous Poisson process within a small sphere \( S_{x_i}(R) \) of radius \( R \) centered at \( x_i \). On this basis, the inhomogeneous binomial process \( \{N(t,x_i), 0 \leq t \leq R\} \) with

\[ N(t, x_i) = \sum_{j=1}^{N} 1\{x_j \in S_{x_i}(t)\} \]  \hspace{1cm} (10)

counts the number of observations within distance \( t \) of \( x_i \) and can be approximated as a Poisson process. The rate of this process is:

\[ \lambda(t, x_i) = f(x_i) \ V(m(x_i)) \ m(x_i) \ t^{m(x_i)-1} \]  \hspace{1cm} (11)

where \( m(x_i) \) is the dimension of the manifold on which \( x_i \) lies and \( V(m(x_i)) \) is the volume of the unit sphere in \( \mathbb{R}^{m(x_i)} \) centered at \( x_i \). The log-likelihood function of \( N(t, x_i) \) can then be expressed as:

\[ L(m(x_i), \theta(x_i)) = \int_{0}^{R} \log(\lambda(t, x_i)) \ dN(t, x_i) - \int_{0}^{R} \lambda(t, x_i) \ dt \]  \hspace{1cm} (12)
where \( \theta(x_i) = \log(f(x_i)) \). Finally, the MLE for \( m(x_i) \) provides a local estimator of ID [30, 57, 58]:

\[
\hat{m}_k(x_i) = \left[ \frac{1}{k-2} \sum_{j=1}^{k-1} \log \left( \frac{T_k(x_i)}{T_j(x_i)} \right) \right]^{-1} \tag{13}
\]

where \( k > 2 \) is the number of NN taken into account and \( T_k(x_i) \) is the distance between \( x_i \) and its \( k \)th NN. If it is assumed that all the observations belong to the same manifold, one has that:

\[
\hat{m}_k = \frac{1}{N} \sum_{i=1}^{N} \hat{m}_k(x_i) \tag{14}
\]

which is simply an average over the whole dataset and, for \( k \in \{k_1, k_1 + 1, ..., k_2\} \) with \( k_1 > 2 \), the final estimate of ID is provided by [30]:

\[
\hat{m} = \frac{1}{k_2 - k_1 + 1} \sum_{k=k_1}^{k_2} \hat{m}_k \tag{15}
\]

A delicate issue which arises from Equations 13, 14 and 15 is the range of \( k \) values to be chosen. In practical applications, this is similar to the choice of the scale interval in the fractal-based methods. Here as well, the finiteness of datasets may greatly influence the final estimate of ID if the considered \( k \) values are not carefully selected. In [30], it is advocated to retain a range of small to moderate values, so that each \( S_{x_i}(R) \) is small enough to ensure \( f(x_i) \approx \text{const} \) and great enough to contain sufficiently many points.

In [59], a modified version of the MLE algorithm is proposed. It consists in averaging the inverse of the \( N \) estimators \( \hat{m}_k(x_i) \) of Equation 14, so that the final estimator of Equation 15 is replaced with:

\[
\hat{m} = \frac{1}{k_2 - k_1 + 1} \sum_{k=k_1}^{k_2} \left[ \frac{1}{N(k-1)} \sum_{i=1}^{N} \sum_{j=1}^{k-1} \log \left( \frac{T_k(x_i)}{T_j(x_i)} \right) \right]^{-1} \tag{16}
\]

Although the second version of the algorithm is better for small values of \( k \), both of them yield similar results [58] and have been successfully applied in various studies [30, 59, 57]. Finally, notice that, in the remainder of this paper, the estimators of Equations 15 and 16 will be named after their authors, \( \hat{m}_{LB} \) and \( \hat{m}_{KG} \).
3. A New Estimator of Intrinsic Dimension

3.1. The $m$-Morisita Index

The $m$-Morisita index $I_{m,\delta}$ is a global measure of clustering. It is a generalization of the Morisita index $I_{\delta}$ and it was first proposed by [32] for the analysis of population distributions in ecology. It was later modified by [31] to take into account the notion of scale in spatial data analysis and...
a relationship to multifractality was established.

For its computation, $I_{m,\delta}$, requires the studied datasets to be covered with a grid of $Q$ quadrats (or cells) of changing size $\delta$ (see Figure 1). For a fixed $\delta$, $I_{m,\delta}$ measures how many times more likely it is that $m$ ($m \geq 2$) randomly selected data points will be from the same quadrat than it would be in the case of a random distribution generated from a Poisson process. Mathematically, it is calculated as follows:

$$I_{m,\delta} = Q^{m-1} \sum_{i=1}^{Q} n_i (n_i - 1) (n_i - 2) \cdots (n_i - m + 1)$$

$$N (N - 1) (N - 2) \cdots (N - m + 1)$$

(17)

where $n_i$ is the number of points in the $i^{th}$ quadrat and $N$ is the total number of points. The computation of the index starts with a relatively big quadrat size $\delta$. It is then reduced until it reaches a minimum value and a plot relating every $I_{m,\delta}$ to its matching $\delta$ can be drawn.

Figure 1 illustrates the computation of the index in two dimensions for three benchmark point distributions (or patterns), for three different scales and for $m = 2$ and $m = 3$. For the highest possible $\delta$, when only one quadrat is considered, $I_{m,\delta}$ returns the same value for each pattern and $m$. Then, as the number of quadrats increases, $I_{m,\delta}$ adopts a behaviour discriminating between the three benchmark patterns. If the points are distributed at random, every computed $I_{m,\delta}$ oscillates around the value of 1. If the points are clustered, the value of the index increases as $\delta$ decreases and, finally, if the points are dispersed, the index approaches 0 at small scales [31, 60]. Further, as $m$ increases, $I_{m,\delta}$ becomes more and more sensitive to the structure of the pattern under study. In complex situations, $I_{m,\delta}$ computed with small $m$ may miss structures which are detected with higher $m$.

### 3.2. The Morisita Estimator of Intrinsic Dimension

Several parallels can be drawn between the $m$-Morisita index and Rényi’s information of $q^{th}$ order. In particular, it was established, for fractal point sets, that [31]:

$$\lim_{\delta \to 0} \log \left( \frac{I_{m,\delta}}{\log \left( \frac{1}{\delta} \right)} \right) \frac{1}{m - 1} \approx E - D_m = C_m$$

(18)

where $m \in \{2, 3, 4, \cdots\}$, $C_m$ is the codimension of order $q = m$, $E$ is the dimension of the Euclidean space where the data set is embedded (i.e. $E$
equals the number of variables) and $D_m$ is Rényi’s generalized dimension of order $q = m$. In practical applications, for finite datasets, it can be shown that Equation 18 is verified only under the condition that $H := \max_i (n_i) \gg m \ [31]$. If so, $C_m$ can be estimated from the slope, $S_m$, of the straight line fitting the linear part of the plot relating $\log (I_{m,\delta})$ to $\log \left( \frac{1}{\delta} \right)$. Then, $C_m \approx S_m / (m - 1)$ and one has that:

$$D_m \approx E - \left( \frac{S_m}{m - 1} \right)$$  (19)

where $m \in \{2, 3, 4, \cdots \}$.

In high-dimensional spaces, the condition $H \gg m$ is hardly ever met at small scales. In such situations, it is important to notice that the major difference between $\log (I_{m,\delta})$ and $RI_m(\delta)$ lies in the following inequality:

$$n_i^m > n_i(n_i - 1) \cdots (n_i - m + 1)$$  (20)

Consequently, unlike $\log (I_{m,\delta})$, $RI_m(\delta)$ seriously overestimates the probability of randomly drawing $m$-tuples of points from grid cells characterized by a small $n_i$. Such cells are numerous at small scales or when the sample size is limited and can greatly affect the accuracy of $D_m$. From this perspective, it is possible to suggest a new ID estimator based on $\log (I_{m,\delta})$ (for $m \geq 2$):

$$M_m = E - \left( \frac{S_m}{m - 1} \right)$$  (21)

which should be more robust to sample size than the traditional estimator based on Equation 4. Notice that $M_m$ should only be computed under the condition that $H > m$ at all considered scales and if the studied point distributions are not dispersed and space-filling at once. In the remainder of this paper, $M_m$ will be referred to as the Morisita estimator of ID. It will be thoroughly tested in Section 4 with synthetic datasets of various complexities.

3.3. An Algorithm for Large Data sets

Algorithms used for handling large datasets should be affected as little as possible by the amount of main memory available. Regarding the Morisita Estimator of ID, a delicate issue concerns the way the number of points per quadrat is counted. A good algorithm must be able to effectively
Algorithm 1 MINDID

**INPUT:** a $N \times E$ matrix, $D$, with $E$ features and $N$ points; a vector $L$ of values $\ell$; a vector $M$ of contiguous values $m$. **OUTPUT:** a vector containing $M_m$ for each value $m$.

1: Rescale each variable to $[0, 1]$
2: **for all** values $\ell$ do
3: Divide each element of $D$ by $\ell$ and round the result to the next lowest whole number in a $N \times E$ matrix called $D\ell$
4: Count the number, $\text{nbr}_\ell$, of different lines in $D\ell$ and store their frequency in a vector, $ni$, of size $\text{nbr}_\ell$
5: **for all** values $m$ do
6: Compute $\log (I_{m,\delta})$ (See Equation 23) using the vector $ni$ and store the result in a $|L| \times |M|$ matrix called $\text{logMindex}$
7: **end for**
8: **end for**
9: Optional: Compute the values $\delta$ (See Equation 24) using the vector $L$ and store the result in a vector $\Delta$ of cardinality $|L|$
10: **for all** values $m$ do
11: Compute $M_m$ (See Equation 21) using $\text{logMindex}$ and $\Delta$ (or $L$)
12: **end for**

disregard empty quadrats. It is also more appealing if its implementation is straightforward in most programming environments (e.g. R and matlab). In order to take into account these requisites, MINDID (see Algorithm 1) incorporates a version of an algorithm proposed in [10]. It accepts as input a matrix with as many columns as there are variables and as many lines as there are data events (or points). It rescales each variable to $[0, 1]$ and, then, takes advantage of the interesting properties of the square cells of the hyper-grid covering the data: for a given cell size $\delta$, each value of the dataset is divided by $\ell$ (see Figure 1) and rounded to the next lowest integer. In this way, in the resulting matrix, all the data points falling into the same cell are matched by as many equal lines.

Another issue concerns $Q_{m-1}$, since it is often given the value $\infty$ for small values of $\ell$ when $E \gg 1$. A way to overcome this problem is to resort to $\log (I_{m,\delta})$ instead of $I_{m,\delta}$. First, $Q_{m-1}$ is related to $\ell$ and $E$ through:

\[
Q_{m-1} = \left( \frac{1}{\ell} \right)^{E(m-1)}
\] (22)
and

\[
\log (I_{m,\delta}) = E(m - 1) \log \left( \frac{1}{\ell} \right) + \log \left( \frac{\sum_{i=1}^{Q} n_i (n_i - 1) \cdots (n_i - m + 1)}{N(N - 1) \cdots (N - m + 1)} \right)
\]  

(23)

This solution is satisfactory, since the computation of \(M_m\) (See Equation 21) only requires \(\log (I_{m,\delta})\). Consequently, the second part of the MINDID algorithm is devoted to the computation of Equation 23.

Finally, the computation of \(M_m\) is carried out using either \(\log(\ell)\) or \(\log(\delta)\). In the rest of this paper, the second option will be preferred. Notice that \(\ell\) and \(\delta\) are related as follows:

\[
\delta = \sqrt{E\ell^2}
\]  

(24)

4. Assessment of the Morisita Estimator of ID

4.1. Synthetic Data

Several data sets were built, so that each of them lies on a known manifold (or near a known manifold in the case of noise injection). They can be divided into four categories (see Figure 2):

1. Swiss rolls (e.g. [3]) of 500, 1000 and 5000 points. The theoretical ID of the data is equal to 2.
2. Noisy Swiss rolls of 1000, 5000 and 10000 points. The noise is modelled as a Gaussian variable \(G \sim N(0, \sigma^2)\) where \(\sigma\) varies from 0 to 0.5.
3. Uniform clouds of 500, 5000 and 10000 points. Each of the \(N\) distinct points \(x_i\), with \(i \in \{1, 2, \cdots, N\}\), is an \(E\)-dimensional vector \([x_{1i}, x_{2i}, \cdots, x_{Ei}]^T \in \mathbb{R}^E\), the components of which are i.i.d. variables following a uniform distribution. \(E\) is gradually increased from 1 to 7 and is equal to the theoretical ID of the data.
4. This last category is based on the properties of the Cartesian product of some fractals [39, 61]. One-dimensional Cantor sets of 512, 8192 and 65536 points are first created. For each number, \(N\), of points, the resulting vector are shuffled seven times to generate as many variables. An Euclidean space \(\mathbb{R}^E\) can then be constructed and \(E\) can be gradually increased from 1 to 7. The dimension of the data manifold (i.e. the theoretical ID) is equal to \(\frac{\log(3)}{\log(2)} E\), where \(\frac{\log(2)}{\log(3)}\) is the Hausdorff dimension of a one-dimensional Cantor set.
4.2. The Morisita Estimator $M_m$ Versus $D_m$

Two categories of datasets were used to compare the Morisita estimator of ID, $M_m$, with $D_m$: the uniform point clouds and the Cantor sets. They were employed with varying $N$ and $E$ (see Subsection 4.1) and for each combination of these two parameters, 100 sets were generated. The results are displayed in Figures 3 and 4 and in Tables 1 and 2.
Figure 3: The results of the application of $I_{2,\delta}$ and $RI_{2}(\delta)$ to 100 uniform point clouds of 10000, 5000 and 500 points. In the bottom-right table, the corresponding $D_2$ and $M_2$ are provided as follows: the mean (computed over the 100 sets) ± the standard deviation.

Regarding the uniform point clouds, $\log \left( RI_{2}(\delta) \right)$ and $\log \left( I_{2,\delta} \right)$ were computed using an interval of parameter $\ell$ ranging from 1 to 15 (see Equation 24). Figure 3 shows the results for $N = 10000$, $N = 5000$ and $N = 500$ and for increasing $E$. The points of each plot correspond to the average values (computed over the 100 sets) yielded by the two indices and are associated with an error bar calculated from the standard deviation.
Figure 4: The results of the application of $I_{2,5}$ (left) and $RI_2(\delta)$ (right) to 100 Cantor sets of 8192 and 65536 points. In the bottom table, the corresponding $D_2$ and $M_2$ are provided as follows: the mean (computed over the 100 sets) ± the standard deviation.
If the value of $H$ (see Subsection 3.2) was smaller than 2 at certain scales $\delta$ for a given dimension $E$, the entire plots describing the behaviour of $\log (RI_2(\delta))$ and $\log (I_2,\delta)$ in the corresponding $E$-dimensional space was not drawn. This was motivated by the condition that $H$ must be greater than $m$ at all scales for $M_m$ to be computed.

In spite of this limitation, $M_m$ provides better estimates of ID than $D_q$ when the same scale interval is considered. In Figure 3 this is highlighted, for the largest $E$, by the steady state reached by $\log (RI_2(\delta))$ at small scales, which shows a departure from the power law of Equation 2. As a consequence, $D_2$ cannot be derived from linear regressions calculated over the whole scale range. In contrast, $\log (I_2,\delta)$ follows, on average and throughout the scales, the power law which underlies Equation 21 and all the plots are superimposed on a constant mean level of 0 as expected from a Poisson distribution (i.e. a space-filling set of constant density). The cost of this near absence of bias is an increase in the variability of the values provided by $(I_2,\delta)$ as $\delta$ decreases, but it has only a small impact on the variability of the final estimates of ID which stays low, as indicated in the accompanying table. The means and standard deviations of ID estimates were calculated only if the dependence between $\log (I_2,\delta)$ or $\log (RI_2(\delta))$ and $\delta$ could be reasonably approximated by using a linear regression over all the scales (i.e if only one slope could be distinguished in the different plots). The results show that $D_2$ becomes unreliable for $E > 2$, while $M_2$ works better and can even be used up to $E = 6$ for $N = 10000$. Finally, the number of $E$-dimensional spaces, for which $M_2$ can be calculated, decreases as $N$ is reduced, since it becomes less likely that at least two points fall into the same cell at small scales. Nevertheless, whatever $N$, the bias affecting $\log (RI_2(\delta))$ is always noticeable for the largest $E$ and tends to lead to an underestimation of ID for the others (e.g. $E = 2$ for $N = 500$).

Similar comments can be made about the results obtained for the Cantor sets (see Figure 4). For this second category of data, the interval of parameter $\ell$ follows a geometric series with ratio $r = 3$ and ranges from 1 to 81. In this way, the grid used for the computation of both $\log (I_2,\delta)$ and $\log (RI_2(\delta))$ is in accordance with the mathematical construction of a Cantor set. For the same arguments as those previously set out, $M_2$ turns out

\footnote{The linear slope of the plots must be multiplied by $-1$ to yield $S_m$ and $D_q$, since the x-axis represents $\delta$ instead of $\frac{1}{\delta}$.
Table 1: The results of the application of $M_m$ and $D_m$ to 100 Cantor sets of 65536 points for $m = 2$, $m = 3$ and $m = 5$. The mean ± the standard deviation is provided for each $E$-dimensional space.

| $N = 65536$ | $M_2$ | $D_2$ | $M_3$ | $D_3$ | $M_5$ | $D_5$ |
|-------------|-------|-------|-------|-------|-------|-------|
| $E = 1$     | $0.63 \pm 0.00$ | $0.63 \pm 0.00$ | $0.63 \pm 0.00$ | $0.63 \pm 0.00$ | $0.63 \pm 0.00$ | $0.63 \pm 0.00$ |
| $E = 2$     | $1.26 \pm 0.00$ | $1.26 \pm 0.00$ | $1.26 \pm 0.00$ | $1.26 \pm 0.00$ | $1.26 \pm 0.00$ | $1.26 \pm 0.00$ |
| $E = 3$     | $1.89 \pm 0.00$ | $1.88 \pm 0.00$ | $1.89 \pm 0.00$ | $1.88 \pm 0.00$ | $1.89 \pm 0.00$ | $1.87 \pm 0.00$ |
| $E = 4$     | $2.52 \pm 0.00$ | $2.39 \pm 0.00$ | $2.52 \pm 0.00$ | $2.37 \pm 0.00$ | $2.53 \pm 0.01$ | $2.33 \pm 0.00$ |
| $E = 5$     | $3.16 \pm 0.00$ | -                | $3.16 \pm 0.02$ | -                | -                | -                |
| $E = 6$     | $3.79 \pm 0.02$ | -                | -                | -                | -                | -                |
| $E = 7$     | $4.43 \pm 0.07$ | -                | -                | -                | -                | -                |

Table 2: The results of the application of $M_m$ and $D_m$ to 100 Cantor sets of 8192 points for $m = 2$, $m = 3$ and $m = 5$. The mean ± the standard deviation is provided for each $E$-dimensional space.

| $N = 8192$ | $M_2$ | $D_2$ | $M_3$ | $D_3$ | $M_5$ | $D_5$ |
|------------|-------|-------|-------|-------|-------|-------|
| $E = 1$    | $0.63 \pm 0.00$ | $0.63 \pm 0.00$ | $0.63 \pm 0.00$ | $0.63 \pm 0.00$ | $0.63 \pm 0.00$ | $0.63 \pm 0.00$ |
| $E = 2$    | $1.26 \pm 0.00$ | $1.26 \pm 0.00$ | $1.26 \pm 0.00$ | $1.25 \pm 0.00$ | $1.26 \pm 0.00$ | $1.25 \pm 0.00$ |
| $E = 3$    | $1.90 \pm 0.00$ | $1.82 \pm 0.00$ | $1.90 \pm 0.00$ | $1.80 \pm 0.00$ | $1.90 \pm 0.01$ | $1.76 \pm 0.00$ |
| $E = 4$    | $2.53 \pm 0.01$ | -                | $2.53 \pm 0.03$ | -                | -                | -                |
| $E = 5$    | $3.16 \pm 0.04$ | -                | -                | -                | -                | -                |
| $E = 6$    | -            | -                | -                | -                | -                | -                |
| $E = 7$    | -            | -                | -                | -                | -                | -                |

The means to be a more reliable estimator of ID than $D_2$: here as well, a bias affects the behaviour of $\log (RI_2(\delta))$ when the number of points contained in occupied cells is low. It is also interesting to notice that $M_2$ can be computed up to $E = 5$ for $N = 8192$, although the lowest considered scale $\delta$ is much smaller than in the case of the uniform point clouds. This is due to the dimension of the data manifold that is systematically smaller than $E$. As a consequence, Cantor sets are not space-filling and it is more likely that two points fall into the same cell than in the case of a Poisson distribution. Put differently, the curse of ID\cite{12} is a central issue when studying high-dimensional spaces.

The Cantor sets were also used to assess the accuracy of $M_m$ and $D_m$ for $m$ greater than 2. Tables\cite{1} and\cite{2} show the results for, respectively, $N = 65536$ and $N = 8192$. In both cases, $M_m$ provides results closer to the theoretical ID than those yielded by $D_m$ for $E > 2$. Nevertheless, the difference between the two estimators is lessened as the value of $m$ is increased. This follows from Equation\cite{20}, when $m$ ($m = q$) is high, the impact of the inequality on the difference between the two estimators is negligible and the implementation of $M_2$ requires more data points than that of $D_m$. 

16
4.3. The Morisita Estimator $M_2$ Versus $d_{cor}$, $\hat{m}_{LB}$ and $\hat{m}_{KG}$

$M_2$ was also compared to the other estimators of ID presented in Section 2, namely $d_{cor}$, $\hat{m}_{LB}$ and $\hat{m}_{KG}$. The Swiss rolls and the uniform point clouds were used for this task. The parameters of each estimator were set on the Swiss rolls of 500 points and they stayed unchanged in the whole subsection. At each step, it was made sure that these parameters were close to the ideal ones. Regarding $M_2$, the interval of parameter $\ell$ was chosen, so that it ranged from 5 to 15 and the two MLE estimators were computed with $k$ going from 10 to 20. More challenging, the interval of parameter $\delta$ of $d_{cor}$ turned out to be relatively complicated to set, since it tended to deviate from the ideal values as both $N$ and $E$ were increased. It was finally decided to resort to percentiles and the problem was empirically solved as follows: (a) 1 percent of the pairwise distances between points had to be lower than the smallest $\delta$ and 7 percent of them had to be lower than the largest one; (b) the range of the interval was divided by 100 to produce intermediate values of $\delta$. It is also worth mentioning that each variable was rescaled, so that it ranged from 0 to 1. Such a transformation is mandatory when working with data of different nature. Regarding $d_{cor}$, $\hat{m}_{LB}$ and $\hat{m}_{KG}$, it amounts to using the Mahalanobis distance and, if the data are not noisy, it has no influence on $M_m$ (or $D_q$), since the grid employed in their computation is also transformed.

Tables 3 and 4 show the results (mean ± sd computed over 100 sets) provided by the four estimators for, respectively, the Swiss rolls and the uniform point clouds. The values yielded by the four estimators are similar if the data points are densely distributed on their manifold. This situation is encountered for $N = 5000$ in the case of the Swiss rolls and for relatively low $E$ and high $N$ in the case of the uniform distributions. When the data points are sparse (low $N$ and/or high theoretical ID), $M_2$ provides better

| Swiss Roll | $N = 500$ | $N = 1000$ | $N = 5000$ |
|------------|-----------|-----------|-----------|
| $M_2$      | 2.04 ± 0.06 | 2.03 ± 0.03 | 2.03 ± 0.01 |
| $d_{cor}$  | 1.94 ± 0.03 | 1.95 ± 0.01 | 1.94 ± 0.00 |
| $\hat{m}_{LB}$ | 1.95 ± 0.02 | 1.95 ± 0.02 | 1.98 ± 0.01 |
| $\hat{m}_{KG}$ | 1.94 ± 0.03 | 1.95 ± 0.02 | 1.98 ± 0.01 |

Table 3: The results of the application of the four estimators to Swiss rolls of $N = 500$, $N = 1000$ and $N = 5000$ points (100 sets for each $N$). The mean ± the standard deviation of the estimators, computed over the 100 sets, is provided for each $N$. 

| Swiss Roll | $N = 500$ | $N = 1000$ | $N = 5000$ |
|------------|-----------|-----------|-----------|
| $M_2$      | 2.04 ± 0.06 | 2.03 ± 0.03 | 2.03 ± 0.01 |
| $d_{cor}$  | 1.94 ± 0.03 | 1.95 ± 0.01 | 1.94 ± 0.00 |
| $\hat{m}_{LB}$ | 1.95 ± 0.02 | 1.95 ± 0.02 | 1.98 ± 0.01 |
| $\hat{m}_{KG}$ | 1.94 ± 0.03 | 1.95 ± 0.02 | 1.98 ± 0.01 |
|                | \(M_2\)  | \(d_{cor}\) | \(\hat{m}_{LE}\) | \(\hat{m}_{KG}\) |
|----------------|---------|-------------|-----------------|-----------------|
| **\(N = 500\)**|         |             |                 |                 |
| \(E = 1\)     | 1.00 ± 0.01 | 0.99 ± 0.01 | 1.00 ± 0.01 | 1.00 ± 0.01 |
| \(E = 2\)     | 2.00 ± 0.03 | 1.91 ± 0.01 | 1.93 ± 0.02 | 1.92 ± 0.03 |
| \(E = 3\)     | 3.01 ± 0.1  | 2.74 ± 0.04 | 2.79 ± 0.04 | 2.78 ± 0.04 |
| \(E = 4\)     | -        | 3.52 ± 0.05 | 3.58 ± 0.05 | 3.57 ± 0.06 |
| \(E = 5\)     | -        | 4.24 ± 0.06 | 4.34 ± 0.07 | 4.33 ± 0.07 |
| \(E = 6\)     | -        | 4.93 ± 0.07 | 5.08 ± 0.07 | 5.07 ± 0.08 |
| \(E = 7\)     | -        | 5.60 ± 0.08 | 5.79 ± 0.09 | 5.80 ± 0.09 |
| **\(N = 5000\)**|         |             |                 |                 |
| \(E = 1\)     | 1.00 ± 0.00 | 0.99 ± 0.00 | 1.00 ± 0.00 | 1.00 ± 0.00 |
| \(E = 2\)     | 2.00 ± 0.00 | 1.91 ± 0.01 | 1.98 ± 0.01 | 1.98 ± 0.01 |
| \(E = 3\)     | 3.00 ± 0.01 | 2.74 ± 0.01 | 2.90 ± 0.01 | 2.90 ± 0.01 |
| \(E = 4\)     | 4.00 ± 0.04 | 3.51 ± 0.01 | 3.78 ± 0.02 | 3.77 ± 0.02 |
| \(E = 5\)     | 5.02 ± 0.13 | 4.24 ± 0.01 | 4.61 ± 0.02 | 4.60 ± 0.02 |
| \(E = 6\)     | -        | 4.93 ± 0.01 | 5.22 ± 0.01 | 5.40 ± 0.03 |
| \(E = 7\)     | -        | 5.58 ± 0.01 | 5.94 ± 0.01 | 6.18 ± 0.03 |
| **\(N = 10000\)**|          |             |                 |                 |
| \(E = 1\)     | 1.00 ± 0.00 | 0.99 ± 0.00 | 1.00 ± 0.00 | 1.00 ± 0.00 |
| \(E = 2\)     | 2.00 ± 0.00 | 1.91 ± 0.00 | 1.98 ± 0.00 | 1.98 ± 0.01 |
| \(E = 3\)     | 3.00 ± 0.00 | 2.74 ± 0.01 | 2.93 ± 0.01 | 2.92 ± 0.01 |
| \(E = 4\)     | 4.00 ± 0.02 | 3.51 ± 0.01 | 3.82 ± 0.01 | 3.81 ± 0.01 |
| \(E = 5\)     | 5.00 ± 0.08 | 4.24 ± 0.01 | 4.66 ± 0.02 | 4.66 ± 0.02 |
| \(E = 6\)     | 6.06 ± 0.24 | 4.93 ± 0.01 | 5.42 ± 0.01 | 5.47 ± 0.02 |
| \(E = 7\)     | -        | 5.58 ± 0.01 | 6.19 ± 0.02 | 6.27 ± 0.02 |

Table 4: The results of the application of the four estimators to uniform distribution clouds of \(N = 500\), \(N = 5000\) and \(N = 10000\) points (100 sets for each \(N\)). The mean ± sd, computed over the 100 sets, is provided for each \(N\).
ID estimates if it can be calculated. Lastly, $d_{cor}$, $\hat{m}_{LB}$ and $\hat{m}_{KG}$ are always able to provide ID estimates, but, when the data are simply too sparse for $M_2$ to be computed, they tend to seriously underestimate the true ID of the data. The resulting difference reduces as $N$ increases and is induced (at least partially) by edge effects. Several Edge effect corrections have been proposed and thoroughly studied in spatial data analysis \cite{62}, but the problem has often been overlooked in ID estimation methods. $d_{cor}$ appears to be more affected by this problem than the other estimators. Nevertheless, the case of $d_{cor}$ is difficult to deal with, since the parameter $\delta$ is very sensitive to $N$ and $E$. Consequently, concerning the uniform distribution clouds, it was decided that a new series of 10 estimations would be carried out for $N = 5000$ and $N = 10000$ with improved parameters. The results appear in blue in Table \ref{tab:results} and reveal that the modifications have improved the estimates, even if $\hat{m}_{LB}$ remains the best distance-based estimator tested in this study.

Figure 5: (left) Application of $M_2$, $\hat{m}_{LB}$ and $\hat{m}_{KG}$ to noisy Swiss rolls; (right) Application of $M_2$ to noisy Swiss rolls of $N = 1000$, $N = 5000$ and $N = 10000$. The x-axis represents the standard deviation of the noise and for each level 100 sets are considered.

4.4. Noise Injection

The presence of noise implies that the data points are located near a manifold instead of being exactly on it \cite{30}. Consequently, a good ID estimator should be as insensitive to noise as possible. In order to test the robustness of $M_m$, the noisy Swiss rolls, presented in Subsection \ref{subsec:noise} were
used and the results are shown in Figure 5. First, on the left, a comparison between $M_2$, $\hat{m}_{LB}$ and $\hat{m}_{KG}$ was conducted and one can observe how the mean and the standard deviation of the estimates (computed over 100 sets) change as the noise increases. Leaving aside the initial difference, the sensitivity of the three estimators appear to be similar. Secondly, on the left hand-side of the same figure, it is highlighted that the number of points $N$ has a low influence on the responsiveness of $M_2$ to noise. On the basis of these results, the behaviour of $M_2$ in presence of noise is not better or worse than that of the other two estimators. Finally, notice that this subsection does not discuss the impact of data scaling in presence of noise, since it would be out of the scope of this paper.

5. Conclusion

The Morisita estimator, $M_m$, is a new tool for estimating the intrinsic dimension of data and it is related to Rényi’s generalized dimensions, $D_q$, for $m = q \geq 2$. $M_m$ tended to provide better results than $D_q$ on the synthetic data used in this study. This turned out to be particularly true for order 2 (i.e. $m = q = 2$) when the data points were sparsely distributed. Consequently, from the perspective of pattern recognition, $M_2$ might be of great interest, since it could be a good replacement for $D_2$ in algorithms such as the fractal dimension algorithm [9, 12]. It might also open a new door to fractal supervised feature selection [13] of large data sets (our current work in progress), since its accuracy is coupled with a high computational efficiency.

$M_2$ was also compared with three distance-based estimators, namely $d_{cor}$, $\hat{m}_{LB}$ and $\hat{m}_{KG}$ and it yielded good results, since it does not suffer from edge effects. Of course, the Morisita estimator of ID is based on a grid and further studies involving real data should be conducted to investigate how it performs when several scaling behaviours are expected at small scales.

Finally, it is also worth mentioning that the multipoint Morisita index is a ratio of probabilities deep-rooted in the field of spatial clustering analysis. Consequently, $M_m$ can be viewed from a dual perspective: a fractal one and a simple statistical one.
6. Acknowledgements

The research was partly supported by the Swiss NSF project No. 200021-140658: ’Analysis and modelling of space-time patterns in complex regions’. The authors also want to thank Michael Leuenberger for many fruitful discussions.

References

[1] D. W. Scott, J. R. Thompson, Probability density estimation in higher dimensions, in: J. R. Gentle (Ed.), Proceedings of the Fifteenth Symposium on the Interface, Elsevier Science Publishers, North-Holland, 1983, pp. 173–179.
[2] R. Bellman, R. E. Bellman, Adaptive Control Processes: A Guided Tour, Princeton University Press, Princeton (NJ), 1961.
[3] J. A. Lee, M. Verleysen, Nonlinear Dimensionality Reduction, Springer, New-York, 2007.
[4] C. J. C. Burges, Dimension reduction: A guided tour, Foundations and Trends in Machine Learning 2 (4) (2009) 275–365.
[5] J. B. Tenenbaum, V. de Silva, J. C. Langford, A global geometric framework for nonlinear dimensionality reduction, Science 290 (5500) (2000) 2319–2323.
[6] A. Lendasse, J. A. Lee, V. Wertz, M. Verleysen, Time series forecasting using CCA and Kohonen maps - application to electricity consumption, in: M. Verleysen (Ed.), Proceedings of ESANN 2000, 8th European Symposium on Artificial Neural Networks, Bruges, 2000, pp. 329–334.
[7] M. Kanevski, A. Pozdnoukhov, V. Timonin, Machine Learning for Spatial Environmental Data: Theory, Applications and Software, EPFL Press, Lausanne, 2009.
[8] S. Kaski, J. Peltonen, Dimensionality reduction for data visualization, Signal Processing Magazine 28 (2) (2011) 100–104.
[9] C. Traina, A. Traina, L. Wu, C. Faloutsos, Fast feature selection using fractal dimension, in: Proceedings of the XV Brazilian Symposium on Databases (SBBD), 2000, p. 158–171.
[10] H. Zhang, C. Perng, Q. Cai, An improved algorithm for feature selection using fractal dimension, in: Proceedings of the Second International Workshop on Databases, Documents, and Information Fusion, 2002.
[11] E. P. M. De Sousa, C. Traina Jr., A. J. M. Traina, L. Wu, C. Faloutsos, A fast and effective method to find correlations among attributes in databases, Data Mining and Knowledge Discovery 14 (2007) 367–407.
[12] C. Traina Jr., A. J. M. Traina, C. Faloutsos, Fast feature selection using fractal dimension - Ten years later, Journal of Information and Data Management 1 (1) (2010) 17–20.
[13] D. Mo, S. H. Huang, Fractal-based intrinsic dimension estimation and its application in dimensionality reduction, IEEE Transactions on Knowledge and Data Engineering 24 (1) (2012) 59–71.
[14] J. G. Dy, C. E. Brodley, Feature Selection for Unsupervised Learning, Journal of Machine Learning Research 5 (2004) 845–889.
[15] J. G. Dy, Unsupervised feature selection, in: H. Liu, H. Motoda (Eds.), Computational Methods of Feature Selection, Chapman Hall/CRC, London/Boca Raton, 2013, pp. 29–55.

[16] S. Alelyani, J. Tang, H. Liu, Feature Selection for Clustering: A Review, in: C. C. Aggarwal, C. K. Reddy (Eds.), Data Clustering: Algorithms and Applications, Chapman Hall/CRC, London/Boca Raton, 2013, pp. 29–55.

[17] D. W. Aha, R. L. Bankert, A comparative evaluation of sequential feature selection algorithms, in: D. Fisher, H. J. Lenz (Eds.), Learning from Data: AI and Statistics, Springer, New-York, 1996, pp. 199–206.

[18] A. K. Jain, D. Zongker, Feature selection: Evaluation, application and small sample performance, IEEE Transactions on Pattern Analysis and Machine Intelligence 19 (2) (1997) 153–158.

[19] A. Blum, P. Langley, Selection of relevant features and examples in machine learning, Artificial Intelligence 97 (1-2) (1997) 245–271.

[20] L. C. Molina, L. Belanche, A. Nebot, Feature selection algorithms: A survey and experimental evaluation, in: Proceedings of 2002 IEEE International Conference on Data Mining (ICDM’02), Japan, 2002, pp. 306–313.

[21] I. Guyon, A. Elisseeff, An introduction to variable and feature selection, Journal of Machine Learning Research 3 (2003) 1157–1182.

[22] I. Guyon, S. Gunn, M. Nikravesh, L. A. Zadeh, Feature extraction: Foundations and Applications, Springer, Berlin, 2006.

[23] T. Marill, D. M. Green, On the effectiveness of receptors in recognition systems, IEEE Transactions on Information Theory 9 (1) (1963) 11–17.

[24] P. Pudil, J. Novovičová, J. Kittler, Floating search methods in feature selection, Pattern Recognition Letters 15 (11) (1963) 1119–1125.

[25] S. M. Vieira, M. C. Sousa, T. A. Runkler, Ant colony optimization applied to feature selection in fuzzy classifiers, Lecture Notes in Computer Science 4529 (2007) 778–788.

[26] X. Wang, J. Yang, X. Teng, W. Xia, F. Richard, Feature selection based on rough sets and particle swarm optimization, Pattern Recognition Letters 28 (4) (2007) 459–471.

[27] V. N. Vapnik, Statistical Learning Theory, Wiley, New-York, 1998.

[28] V. Cherkassky, F. Mulier, Learning From Data: Concepts, Theory and Methods, Second Edition, Wiley, Hoboken (USA), 2007.

[29] F. Camastra, Data dimensionality estimation methods: a survey, Pattern Recognition 36 (12) (2003) 2945 – 2954.

[30] E. Levina, P. J. Bickel, Maximum likelihood estimation of intrinsic dimension 17, in: Advances in Neural Information Processing Systems, Vol. 17, The MIT Press, Cambridge (USA), 2004.

[31] J. Golay, M. Kanevski, C. D. Vega Orozco, M. Leuenberger, The multipoint Morisita index for the analysis of spatial patterns, Physica A 406 (2014) 191–202.

[32] S. H. Hurlbert, Spatial Distribution of the Montane Unicorn, Oikos 58 (3) (1990) 257–271.

[33] M. Morisita, Measuring of the Dispersion of Individuals and Analysis of the Distributional Patterns, Memoires of the Faculty of Science (Serie E), Kyushu University 2 (4) (1959) 215–235.
[34] A. K. Jain, R. C. Dubes, Algorithms for Clustering Data, Prentice-Hall, New-Jersey, 1988.
[35] J. Theiler, Estimating fractal dimension, Journal of the Optical Society of America 7 (6) (1990) 1055 – 1073.
[36] B. B. Mandelbrot, The Fractal Geometry of Nature, W.H. Freeman, San Francisco, 1983.
[37] S. Lovejoy, D. Schertzer, P. Ladoy, Fractal Characterization of Inhomogeneous Geophysical Measuring Networks, Nature 319 (6048) (1986) 43–44.
[38] T. G. Smith, W. B. Marks, G. D. Lange, W. H. Sheriff, E. A. Neale, A Fractal Analysis of Cell Images, Journal of Neuroscience Methods 27 (2) (1989) 173–180.
[39] K. Falconer, Fractal Geometry: Mathematical Foundations and Applications, 2nd Edition, Wiley, Chichester (UK), 2003.
[40] H. G. E. Hentschel, I. Procaccia, The Infinite Number of Generalized Dimensions of Fractals and Strange Attractors, Physica D 8 (3) (1983) 435–444.
[41] P. Grassberger, Generalized Dimensions of Strange Attractors, Physics Letters A 97 (6) (1983) 227–230.
[42] G. Paladin, A. Vulpiani, Anomalous Scaling Laws in Multifractal Objects, Physics Reports 156 (4) (1987) 147–225.
[43] T. Tel, A. Fülöp, T. Vicsek, Determination of Fractal Dimensions for Geometrical Multifractals, Physica A 159 (1989) 155–166.
[44] T. Vicsek, Fractal Growth Phenomena, World Scientific, Singapore, 1993.
[45] A. Rényi, Probability Theory, Akadémiai Kiadó, Budapest, 1970.
[46] J. Balatoni, A. Rényi, On the Notion of Entropy, Publ. Math. Inst. Hungarian Acad. Sci (1) (1956) 5–40, english translation in Selected Papers of A. Rényi, Budapest, vol. 1 (1976), 558.
[47] P. Grassberger, I. Procaccia, Characterization of strange attractors, Physical Review Letters 50 (5) (1983) 346–349.
[48] L. Seuront, Fractals and Multifractals in Ecology and Aquatic Science, CRC Press, Boca Raton (USA), 2010.
[49] Y. Chen, Multifractals of central place systems: Models, dimension spectrums, and empirical analysis, Physica A 402 (2014) 266–282.
[50] M. M. Dubovikov, N. V. Starchenko, M. S. Dubovikov, Dimension of the minimal cover and fractal analysis of time series, Physica A 339 (3-4) (2004) 591–608.
[51] S. Borgani, G. Murante, A. Provenzale, R. Vaklarnini, Multifractal Analysis of the Galaxy Distribution: Reliability of Results from Finite Data Sets, Physical Review E 47 (6) (1993) 3879–3888.
[52] S. Lovejoy, D. Schertzer, The Weather and Climate, Cambridge University Press, New-York, 2013.
[53] Q. Huang, J. R. Lorch, R. C. Dubes, Can the fractal dimension of images be measured?, Pattern Recognition 27 (3) (1994) 339–349.
[54] J. B. Florindo, O. M. Bruno, Fractal descriptors based on the probability dimension: A texture analysis and classification approach, Pattern Recognition Letters 42 (2014) 107–114.
[55] F. Camastra, A. Vinciarelli, Estimating the intrinsic dimension of data with a fractal-based method, IEEE Transactions on Pattern Analysis and Machine Intelligence 24 (10) (2002) 1404–1407.
[56] B. Kégel, Intrinsic dimension estimation using packing numbers, in: Advances in Neural Information Processing Systems, Vol. 14, The MIT Press, Cambridge (USA), 2002.

[57] J. Eberhardt, Estimating Intrinsic Dimension, University of Minnesota, Duluth, 2007.

[58] A. Asensio Ramos, H. Socas-Navarro, A. López Ariste, M. J. Martínez González, The intrinsic dimensionality of spectropolarimetric data, The Astrophysical Journal 660 (2007) 1690–1699.

[59] J. C. MacKay, D. Z. Ghahramani, Comments on "maximum likelihood estimation of intrinsic dimension" by E. Levina and P. Bickel @ONLINE [Jan. 2005]. URL http://http://www.inference.phy.cam.ac.uk/mackay/dimension/

[60] M. Kanevski, M. Maignan, Analysis and Modelling of Spatial Environmental Data, EPFL Press, Lausanne, 2004.

[61] Y. Shi, C. Gong, Multifractality of a cartesian product of two fractals, Communications in Theoretical Physics 23 (1995) 245–248.

[62] D. B. Ripley, Spatial Statistics, Wiley, New-York, 1981.