Local multifractal analysis of marked spatial point processes

Stéphane G Roux, Janka Lengyel, Patrice Abry, François Sémécurbe, Stéphane Jaffard

▶ **To cite this version:**

Stéphane G Roux, Janka Lengyel, Patrice Abry, François Sémécurbe, Stéphane Jaffard. Local multifractal analysis of marked spatial point processes. XXVIIIème Colloque Francophone de Traitement du Signal et des Images, GRETSI'22, Sep 2022, Nancy, France. hal-03871578

HAL Id: hal-03871578
https://hal.science/hal-03871578

Submitted on 25 Nov 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Local multifractal analysis of marked spatial point processes

Stéphane G. ROUX[1], Janka LENGYEL[1,4], Patrice ABRY[1], François SÉMÉCURBE[2], Stéphane JAFFARD[3]

[1] ENS de Lyon, CNRS, Laboratoire de Physique, F-69342 Lyon, France

[2] Université de Franche-Comté, THÉMA, F-25030 Besançon, France

[3]Université Paris-Est - Créteil Val-de-Marne, LAMA, F-94010 Créteil, France

[4] Université Gustave Eiffel, LVMT, IFSTTAR, ENPC, F-77454 Marne-la-Vallée, France

stephane.roux@ens-lyon.fr, janka.lengyel@ens-lyon.fr,
patrice.abry@ens-lyon.fr, francoissemecurbe@gmail.com, jaffard@u-pec.fr

Résumé – Nous proposons une méthode d'estimation locale des propriétés multifractales. Cette methode, qui repose sur une moyenne d'incrément à une dimension, est applicable à tout type de processus ponctuel marqué pour lesquels l'information n'est pas accessible partout. Nous montrons la robustesse d'estimation sur trois types de supports différents.

Abstract – In this paper, we develop a methodology for the local estimation of multifractal properties in random 2D fields. The main novelty of our approach lies in introducing a local average of one-dimensional increments, rendering the analysis applicable not only for fully defined images but also for any marked point process where information is not ubiquitously available, e.g. in the context of geospatial data analysis and modeling. We demonstrate the robustness of the estimation by deploying the methodology on a multifractal random field defined as a marked 2D point pattern with three different underlying supports: an equidistant grid (or image), a self-similar and a multifractal Sierpinski carpet. We show that the estimation of obtained scaling characteristics is statistically concurrent on these three spatial distributions. We conclude by presenting a real-world application using geospatial data.

1 Introduction

Scaling or multifractal analysis [1, 2, 3] is now a standard procedure for estimating local regularities in empirical data. Scale invariance is indeed a property that has been extensively observed in numerous phenomena of very different nature [4, 5]. In general, multifractal analysis consists in measuring scaling exponents, whose values are then commonly involved in various detection, identification or classification tasks. Despite its increasing popularity, existing methods for multifractal estimation are only able to deal with signals or images that are defined everywhere on a full support or where non-existent values are replaced by 0. Even though the latter approach is reliable if the density of missing values is very small, it can not be used for datasets where access to information is irregular, e.g. in the context of geospatial point patterns. Even though an original methodology to derive (multi-) fractal properties of such heterogeneous geographical distributions was recently developed [5] it focused exclusively on unmarked points so far. Against this background, we propose a procedure based on an average of one-dimensional increments allowing for the targeted extension of these methods for *marked 2D point processes* with arbitrary underlying spatial patterns. Building upon a brief overview of the classical multifractal procedure, we introduce the two main contributions of this work : the multiscale quantity specifically tailored for irregular 2D distributions and the local multifractal analysis. We conclude by providing a real-world example using geographical data.

2 Classical Multifractal Analysis

Multifractal analysis aims at characterizing the local regularity of a process V. Different pointwise regularity exponents can be used, most commonly the Hölder exponent $h(x_0) > 0$. It is defined as the largest α so that there exists, in a neighborhood of x_0, a constant $C > 0$ and a polynomial P (of order smaller than α) satisfying $|V(x_0 - r) - P(x_0 - r)| < C|r|^\alpha$. The multifractal spectrum $D(h)$ is defined as the collection of Hausdorff dimensions ($\dim_H$) of the sets of points x, where $h(x)$ takes the values h. $D(h) = \dim_H \{x \text{ such that } h(x) = h\}$. The shape and the width of $D(h)$ provide information on the intensity of fluctuations in local regularity.

Practical estimation of the multifractal spectrum requires the use of multifractal formalism, a procedure originally inspired by thermodynamics (cf., e.g., [2]). It is based on estimating the different moments of certain multiscale quantities and observing their evolution across scales. Historically, multiscale coefficients were defined as the increments of the process, however these were later generalized to wavelet coefficients (defined as the high pass filtering of the process [6]) or to some non-linear transformations of these [7, 8].

Let $T(x,r)$ denote the multiscale coefficient at point x and scale r. A process V defined on $\mathbb{R}^d$ is said to possess scale-invariance or scaling properties if - for some statistical orders q - the time averages of $|T(x,r)|^q$ in a fixed scale display power law behaviors with respect to radii r

$$S_q(r) = \mathbb{E}\{|T(x,r)|^q\} = \lim_{N\to+\infty} \frac{1}{N}\sum_{x=1}^{N} |T(x,r)|^q \sim F_q |r|^{\zeta_q},$$

where N is the total number of points x. The scaling exponents ζ_q are related by a Legendre transform to its multifractal spectrum $D(h) = \min_q(d + qH - \zeta_q)$.

If we suppose that ζ_q is smooth at 0, we can use a Taylor expansion of $\zeta_q = \sum_{p\leq 1} c_p \frac{q^p}{p!}$ and show that the log-cumulant of multiscale coefficients behaves linearly with respect to scales r. For the two first orders one obtains :

$$C_1(r) = \mathbb{E}\{\log|T(x,r)|\} = c_1 \log(|r|) + d_1 \quad (1)$$
$$C_2(r) = \mathbb{E}\{\log|T(x,r)|^2\} - C_1^2(r) = c_2 \log(|r|) + d_2 \quad (2)$$

which give a quadratic approximation of the singularity spectrum $D(h) \sim 1 + \frac{(h-c_1)^2}{2c_2}$.

3 Local Multifractal Analysis

Here, we propose a novel methodology that is able to measure *local* scaling properties of random fields. It involves two essential steps : the definition of multiscale quantities and their local weighting functions.

Multiscale coefficient. We start by introducing a multiscale coefficient, defined as the average of all one-dimensional increments originating from x, i.e.

$$T(x,r) = \frac{1}{N_x(r)} \sum_{x',\, ||x'-x||<r} (V(x') - V(x)) \quad (3)$$

where $N_x(r)$ is the number of points in radius r that is centered on x. In a two-dimensional space, this coefficient can be seen as a convolution of the field V with the high-pass filter $\psi_r(x) = \pi r^2 \delta_0 - \mathbb{1}_{B_0(r)}$ which verifies the admissibility condition $\int_{\mathbb{R}^2} \psi_r(x)dx = 0$ for all r. This filter can be regarded as a wavelet and the scaling in equations 1, 1 and 2 are still valid as long as $0 < H < 1$ [6].

Local weighting function. To carry out the local multifractal analysis, we first introduce a local environment L, that is larger than the radius (or radii) r. Second, we introduce a regular grid of arbitrary size with points x_g, which will serve as the focal points for local estimations. Moving on, we compute a weighted average based on the distance between a chosen estimation site x_g and the locations of the original points x :

$$\begin{aligned} S_q(x_g,r) &= \sum_x w_{x_g,x} |T(x,r)|^q \\ C_1(x_g,r) &= \sum_x w_{x_g,x} \log|T(x,r)| \\ C_2(x_g,r) &= \sum_x w_{x_g,x} \left(\log|T(x,r)| - C_1(x_g,r)\right)^2 \end{aligned} \quad (4)$$

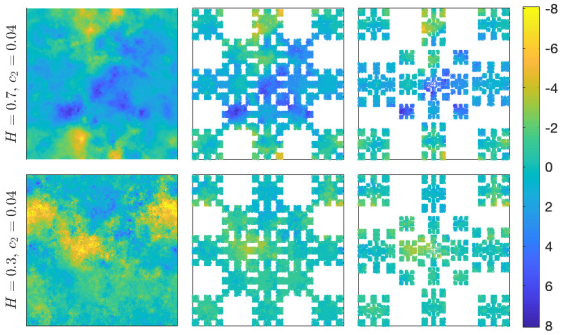

FIGURE 1 – Images showing the same MRW process with a full support (right), a self-similar Sierpinski carpet (middle) and a multifractal Sierpinski carpet. The top row displays a realization of a MRW with $H = 0.7$ and $c_2 = 0.04$, the bottom row with $H = 0.3$ and $c_2 = 0.04$.

where weights are defined as $w_{x_g,x} = f(||x_g - x||/L)$. We use a quartic (biweight) kernel defined as $f(x) = (1-x^2)^2$ if $||x|| < 1$ and 0 otherwise. For better estimation quality, we restrict our analysis to estimation sites where $N_{x_g}(\sqrt{2}l) \neq 0$, l standing for the distance between points x_g. Locally we recover the scaling

$$\begin{aligned} \log S_q(x_g,r) &\sim \zeta_q(x_g)\log(|r|) + F_q(x_g) \\ C_1(x_g,r) &\sim c_1(x_g)\log(|r|) + d_1(x_g) \\ C_2(x_g,r) &\sim c_2(x_g)\log(|r|) + d_2(x_g) \end{aligned} \quad (5)$$

Finally, global estimation can be defined by choosing a collection of estimation sites $N(x_g)$ and derive the average over this set of measurement points :

$$\begin{aligned} \overline{\log S_q(r)} = \frac{1}{N_{x_g}} \sum_{x_g} \log S_q(x_g,r) &\sim \zeta_q \log(|r|) + F_q \\ \overline{C_1(r)} = \frac{1}{N_{x_g}} \sum_{x_g} C_1(x_g,r) &\sim c_1 \log(|r|) + d_1 \\ \overline{C_2(r)} = \frac{1}{N_{x_g}} \sum_{x_g} C_2(x_g,r) &\sim c_2 \log(|r|) + d_2 \end{aligned} \quad (6)$$

4 Numerical simulations

In this section, first, the procedure for generating the analyzed 2D marked point processes is described ; second, they are used to assess the performance of the local multifractal analysis introduced in Section 3.

Multifractal random walk. The multifractal random walk (MRW) [9, 10], is a popular and representative member of the group of multiplicative cascade processes. It constitutes one of the most prominently used class of multifractal models for applied research. More precisely, MRW is a non-stationary, non-Gaussian process with stationary increments. Its multifractal properties are those of the well-known multiplicative log-normal cascade of Mandelbrot. In one dimension, It is defined as $V(0) = 0$ and $V(n) = \sum_{k_1=1}^{n} \omega_H(k_1) e^{\omega(k_1)}$, where n is an integer, $\omega_H(k)$ are the increments of a fractional Brownian

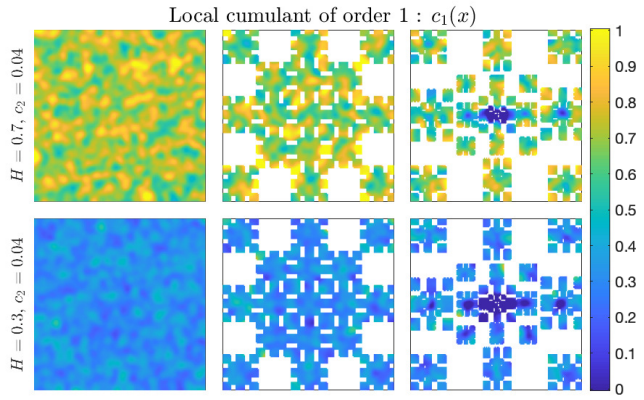

FIGURE 2 – Local analysis. Cumulant of order 1 (eq. 5) obtained on an image (left), a self-similar (middle) and a multifractal Sierpinski carpet (right). The parameters used are $H = 0.7$ and $c_2 = 0.04$ (top) and $H = 0.3$ and $c_2 = 0.04$ (bottom).

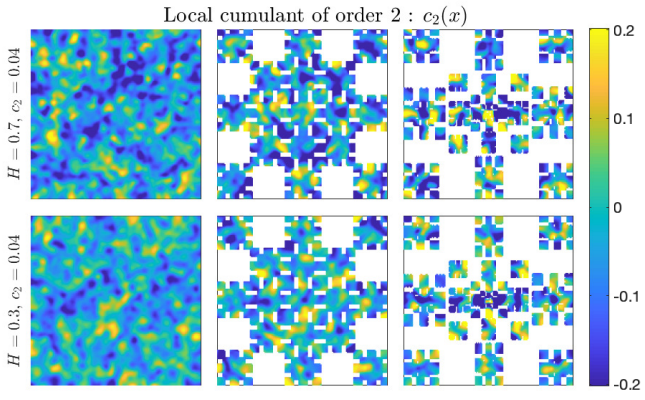

FIGURE 3 – Local analysis. Cumulant of order 2 (eq. 5) obtained on an image (left), a self-similar (middle) and a multifractal Sierpinski carpet (right). The parameters used are $H = 0.7$ and $c_2 = 0.04$ (top row) and $H = 0.3$ and $c_2 = 0.04$ (bottom row).

motion with parameter H (thus a fractional Gaussian Noise of parameter H) ; ω is a Gaussian random process, independent of ω_H with auto-covariance $\mathrm{Cov}(\omega(k_1), \omega(k_2)) = c_2 \ln(\frac{T}{|k_1 - k_2|+1})$ when $|k_1 - k_2| < T$ and 0 otherwise. MRW has the following scaling properties[10] : $\zeta_q = c_1 q - \frac{c_2}{2} q^2$ and $D(h) = 1 + \frac{(1-c_1)^2}{2c_2}$ with $c_1 = H + c_2$.

The two-dimensional isotropic MRW field is obtained using the generalization of the 1D case proposed in [11, 12]. For the purposes of this study, we generate a full image of an MRW process of size $M = 4096$ with parameters H and c_2. In the followings, this image will be construed as a marked point process with spatial positions at $[0,\ M-1]^2$.

Mono- and multifractal support. In order to emphasize general applicability that is independent of underlying point patterns, we construct two Sierpinski carpets using the IFS fractal generator "GenFrac" [15]. We deploy one ($d_0 = 1.76$) and then four different ($d_0 = 1.56$) reduction factors to obtain a mono- and multifractal structure, with different capacity dimensions $d_0 = -\log(N_r)/\log(r)$. Since our methodology may be of interest to research on geospatial systems, it is compelling to observe such rather different point patterns characterized by distinct variations in their point densities.

For each centroid point of the 2D carpet elements, we assign the value of the MRW at the rounded point position. As a result, we obtain three marked point process $\Gamma^i = x_i, y_i, V_i$, $i = 1, 2$ or 3. Fig. 1 shows the support structure of the three processes and their assigned MRW values that were generated with two different H exponents (0.7 and 0.3) and a single shared $c_2 = 0.04$. We note that restricting the fully defined image of the multifractal random field to a fractal set of points may be mathematically ill-defined. However, this does not constitute a problem here as the methodology from Section 3 operates exclusively on a finite collection of scales.

Performance assessment. In accordance with (3), we compute the multiscale coefficients at each location x_g (see Fig 1). For the purposes of this analysis, we apply ten scales logarithmically spaced between $r = 2^2$ and 2^7. A non-weighted linear regression is then carried out at each point x_g (spaced with a resolution of 25 pixels), making use of all available scales to obtain $\zeta_2(x_g)$, $c_1(x_g)$ and $c_2(x_g)$. Final results form an image of size 162X162 : In Fig. 2, we show the resulting slopes obtained for the cumulant of order one, $c_1(x_g)$, computed on the processes shown in Fig. 1, whilst Fig. 3 displays the associated cumulant of order two $c_2(x_g)$.

In general, obtained values do fluctuate locally from point to point but the fluctuations are homogeneous for the three supports. Except from some very sparse areas in the center of the multifractal Sierpinski carpet, the estimation does not significantly depend on its support (as can be clearly seen in Fig. 2). Note that this situation is an additional motivation for computing $c_1(x_g)$ and $c_2(x_g)$, i.e. exponents derived from values of $D(h)$ close to the maximum of the spectrum. Indeed, sets of small dimensions might become "invisible" when their trace on a fractal set is considered. Moving on to consider local estimation results, Fig. 2 demonstrates that the range of locally obtained H exponents is evidently distinct for the two generated multifractal random fields : Darker blue and green colors are dominating in the bottom row ($H = 0.3$) whilst bright green and yellow colors prevail in the top (predefined $H = 0.7$). In contrast, the same color range is observable for all cases in Fig. 3, as they all share one common intermittency coefficient ($c_2 = 0.04$). The quality of the regression is very good with a coefficient of determination R^2 for $\zeta_2(x_g)$ above 0.95 for all three point processes : 0.997%, 0.984%, 0.795% for the image, the self-similar and the multifractal carpet respectively. For $c_1(x_g)$, we get the same quality for the image and the self-similar carpet (with 0.993% and 0.917%) but a poorer estimation for the multifractal support with only 0.52% of the points above $R^2 > 0.96$. For c_2 the quality of the fit degrades strongly.

Finally, the top row in Fig. 4 shows the statistics for the three local slopes, $\zeta_2(x_g)$, $c_1(x_g)$ and $c_2(x_g)$, and the bottom line the corresponding global scaling for $S_2(r)$, $C_1(r)$ and $C_2(r)$ as an average over the entire 2D field x_g. What can be clearly seen is the strong correspondence between the predefined theoreti-

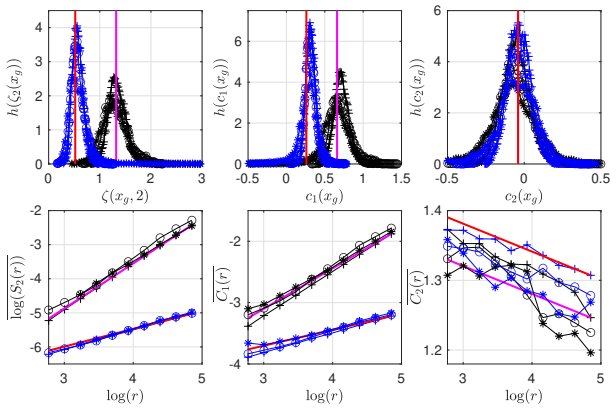

FIGURE 4 – The first row shows the histogram of the local scaling $\zeta_2(x_g)$ (left), $c_1(x_g)$ (middle) and $c_2(x_g)$ obtained from the data on the full support in Fig. 1 left. The second row displays the global average of the scaling, see (6), for the latter three exponents. The parameters are $H = 0.7$ and $c_2 = 0.04$ (black symbols) and $H = 0.3$ and $c_2 = 0.04$ (blue symbols). The estimation is done for the image ($+$), the self-similar Sierpinski carpet ($\circ$) and the multifractal Sierpinski carpet ($\star$).

cal values (red) and the results of the local (top) and global (bottom) estimations obtained with the proposed methodology (blue and black). It has to be emphasized that the latter holds true for all three here-observed supports (Fig. 4 bottom).

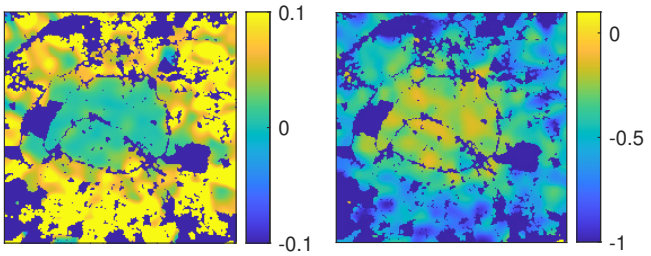

FIGURE 5 – Cumulant of order 1 (left) and order 2 (right) of housing price [16] in and around Paris, France.

5 Application to geographical data

Due to its local nature and broad range of applicability, the methodology may find a number of important applications in settings where information is commonly rather sparse, e.g in the domains of human or environmental geography. In Fig. 5, we deployed the methodology to analyze housing prices ("demande de valeurs foncières" [16]) in and around the city of Paris, France. Small-scale multifractality of land price is of great potential urban planning and policy relevance and it has already been associated with more classical measures of socioeconomic segregation and inequalities [13, 14]. We must emphasize that data distribution is highly irregular, it marks the centroids of cadastral areas where purchases were being made in our observed time window between 2014 and 2020. Notwithstanding, we were able to derive both local parameters, $c_1(x_g)$ and $c_2(x_g)$ with a sufficient regression quality : the average R^2 for the entire observed 2D field depicted in Fig. 5 lies at 0.84 for the cumulant of order one $c_1(x_g)$ and at 0.93 for the cumulant of order two $c_2(x_g)$.

Conclusions. To conclude, this paper set out to illustrate how the introduced methodology is suitable for the local multifractal analysis of marked spatial point processes whose underlying support structure may be strongly non-uniform. More precisely, it could recover - with sufficient regression quality - certain predefined characteristic parameters of multifractal random fields, namely the Hurst exponent H and the intermittency coefficient c_2. It has to be stressed that the latter analysis was possible *both on the global and local levels* and that estimations were also statistically satisfactory on real-world geospatial data.

Références

[1] S. Jaffard, SIAM J. of Math. Anal., vol. 28, no. 4, pp. 971–998, 1997.

[2] P. Abry, P. Flandrin, M. Taqqu, and D. Veitch, K. Park and W. Willinger, Eds., Wiley, pp. 39–88, 2000.

[3] E. Bacry, J. Muzy, and A. Arneodo, J. Stat. Phys., vol. 70, pp. 635–674, 1993.

[4] S. Jaffard, M. Lapidus and M. van Frankenhuijsen, Eds., Proc. Symposia in Pure Mathematics. 2004, vol. 72(2), pp. 91–152, AMS.

[5] F. Sémécurbe, C. Tannier, & S. G. Roux (2019). Journal of Geographical Systems, 21(2), 271-293.

[6] J. F. Muzy, E. Bacry, and A. Arneodo, Phys. Rev. E, vol. 47, no. 2, pp. 875, 1993.

[7] R. Leonarduzzi, H. Wendt, S. G. Roux, M. E. Torres, C. Melot, S. Jaffard, and P. Abry, Physica A, vol. 448, pp. 319–339, 2016.

[8] H. Wendt, S.G. Roux, P. Abry, S. Jaffard, Signal Proces., vol. 89, no. 6, pp. 1100-1114, 2009.

[9] E. Bacry, J. Delour, and J. R. Muzy, Physical Review E 64 (2001), no. 2, 026103.

[10] E. Bacry and J.F. Muzy, Phys. Rev. E, vol. 66, 2002.

[11] R. Robert and V. Vargas, Communications In Mathematical Physics 284 (2008), no. 3, 649–673.

[12] R. Robert and V. Vargas, Ann. Proba., 38 (2010), pp. 605–631.

[13] H. Salat, R. Murcio, K. Yano, E. Arcaute PloS one. 2018 Apr 30;13(4) :e0196737.

[14] S. Hu, Q. Cheng, L. Wang, & S. Xie, (2012). Applied Geography, 34, 161-170.

[15] GenFrac, Fractal Generator. URL : https ://source-sup.renater.fr /www/genfrac/.

[16] Demande de Valeurs Foncières. URL : https ://cerema.app.box.com/v/dvfplus-opendata.