MEMA International Workshop on Spatio-Temporal Modelling
Santiago de Compostela
30 Jun - 2 Jul 2010

Wavelet-clustering for stationary local approximation of intrinsic heterogeneous random fields

M.C. Bueso[1,*], M.D. Ruiz-Medina[2] and J.M. Angulo[2]

[1] *Department of Applied Mathematics and Statistics, Technical University of Cartagena, Campus Muralla del Mar, Cartagena, E-30202 Murcia, Spain; mcarmen.bueso@upct.es*
[2] *Department of Statistics and Operations Research, University of Granada, Campus Fuente Nueva s/n, E-18071 Granada, Spain; mruiz@ugr.es, jmangulo@ugr.es*
[*] *Corresponding author*

Abstract. *This paper addresses the problem of local approximation, in terms of stationary models, of intrinsic heterogeneous random fields, with increments displaying a local self-similar behavior characterized by a functional exponent. Clustering in the wavelet domain is performed for detecting the homogeneous patterns, with a common second-order Hölder exponent. The number of clusters is determined from the scalogram-based estimation of the Hölder exponent function.*

Keywords. *Cluster analysis; Heterogeneity; Hölder spectrum; Spatiotemporal data; Wavelet transform.*

1 Introduction

Heterogeneity analysis constitutes an important topic in spatiotemporal Geostatistics (see Christakos, 2000). A new class of spatiotemporal intrinsic heterogeneous random fields was introduced in Ruiz-Medina and Angulo (2007). The second-order local behavior of the increments of random fields in such a class is characterized by a functional Hölder exponent. The regularity assumptions satisfied by the Hölder exponent function (see Ruiz-Medina, Anh and Angulo, 2004) induce a slow local variation, reflected in the configuration of local homogeneous patterns in the covariance function. In this paper, we address the problem of detecting such patterns from the scalogram-based estimation of the Hölder exponent function, and the clustering performed in the highest resolution levels of the wavelet transformation of the data. A simulation study is developed to illustrate the results derived.

2 The methodology

The following definition provides the characterization of intrinsic heterogeneous random fields.

Definition 1 *Let X be a zero-mean second-order random field. X is said to be intrinsic heterogeneous if the following local behavior is displayed: For all $t \in T \subseteq \mathbb{R}$, and $\mathbf{z} \in D \subseteq \mathbb{R}^d$,*

$$E[X(t+\tau,\mathbf{z}+\boldsymbol{\xi}) - X(t,\mathbf{z})]^2 = o(\|(\tau,\boldsymbol{\xi})\|)^{2h(t,\mathbf{z})+d+1}, \qquad as \quad \|(\tau,\boldsymbol{\xi})\| \to 0,$$

where h denotes the Hölder exponent function.

In Ruiz-Medina, Anh and Angulo (2004), this class of random fields is introduced, in a generalized random field framework, in terms of the isomorphic relationship of its RKHS family with the multifractional fractional Sobolev space class. The log-wavelet transforms of the elements of this class of random fields display an asymptotic linear behavior with slopes proportional to the corresponding Hölder exponent functions (see Ruiz-Medina and Angulo, 2007). This fact motivates the methodology proposed.

Let X be a spatiotemporal process defined on the spatial domain $D \subset \mathbb{R}^2$. Assume that X is observed at N spatial locations and T time points. The method consists of the following steps:

- For each location, perform a nonparametric estimation, based on the wavelet transform, of the local Hölder spectrum. Apply k-means clustering on the N estimated values of the local Hölder spectrum. Select the optimal number of clusters using the silhouette coefficients and the average silhouette values for each number of clusters considered.
- The number of clusters determined in the previous step is considered for detecting the possible number of clusters in the highest resolution levels of the wavelet transformation of the data.
- For each location, apply the discrete wavelet transform to the observed time series.
- At each scale j, extract the detail coefficients and perform k-means clustering on the N objects defined by the N spatial locations. Compute the silhouette coefficients and the average silhouette values for each number of clusters.
- Determine the local homogeneous patterns in the covariance function from the clustering results obtained in the highest resolution levels.
- Repeat the procedure with different wavelet bases, selected according to the regularity properties of the Hölder function estimated, to discriminate the effect of the basis in the wavelet-clustering-based classification procedure.

3 A simulated illustration

In this section, the effect of the functional form and the parameter values characterizing the Hölder exponent is analyzed. The influence of the regularity properties and moment conditions of the wavelet basis selected is also studied. Note that in the Gaussian case, the methodology proposed also provides the detection of homogeneous patterns in the sample paths.

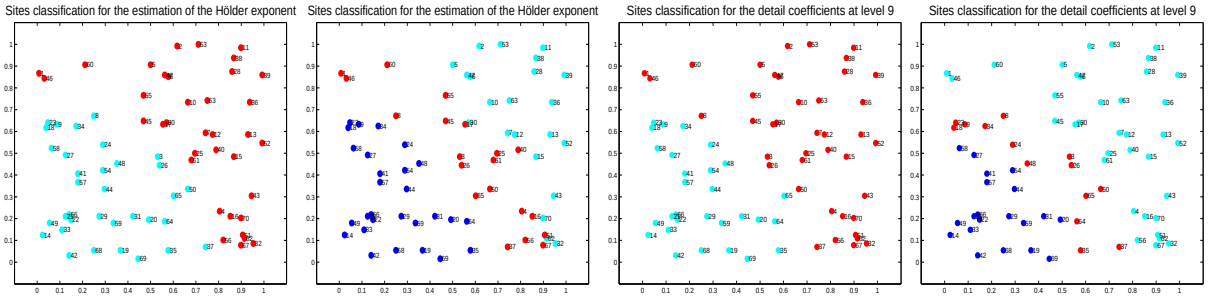

Figure 1: Case 1. Classification of the sites in 2 and 3 groups obtained by *k*-means clustering on the estimated values of the Hölder exponent and on the detail coefficients at level 9, applying the Haar wavelet transform (from left to right).

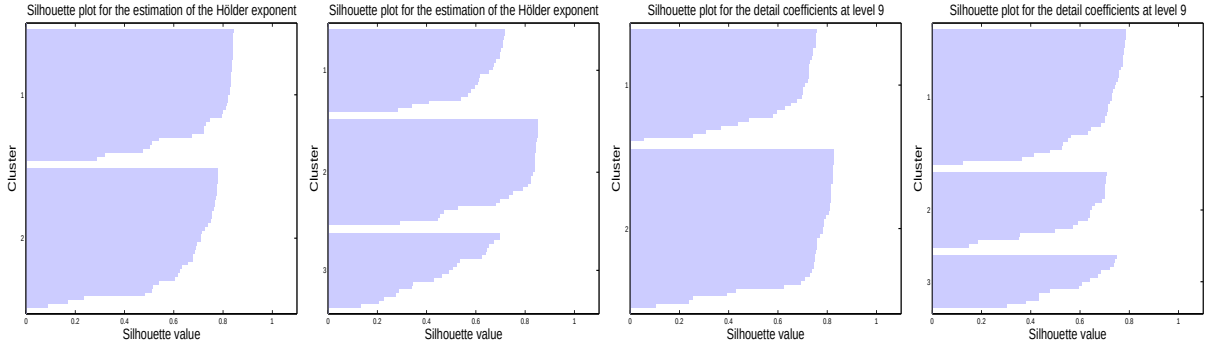

Figure 2: Case 1. Silhouette plots by *k*-means clustering on the estimated values of the Hölder exponent and on the detail coefficients at level 9, for 2 and 3 groups, applying the Haar wavelet transform (from left to right).

The intrinsic heterogeneous random field model considered is defined in terms of the convolution of an innovation process with the following spatiotemporal filter:

$$f(t_1,t_2;\mathbf{z}_1,\mathbf{z}_2;\alpha,\boldsymbol{\theta}) = \frac{\alpha}{1+(|t_1-t_2|^2+\|\mathbf{z}_1-\mathbf{z}_2\|^2)^{h_{\boldsymbol{\theta}}(\mathbf{z}_1)}},$$

with $\alpha > 0$, and $\boldsymbol{\theta} \in \Theta$, the parametric space associated with the Hölder function. The innovation process ε is assumed to be spatiotemporal Gaussian white noise with variance σ^2_ε. The methodology proposed is illustrated here for two functional forms for the Hölder exponent. In Case 1, h is given by

$$h_{\boldsymbol{\theta}}(\mathbf{z}) = \theta_1 \exp\left\{-\left((\theta_2+\|\mathbf{z}\|)^{\theta_3}\right)/\theta_4\right\},$$

and in Case 2, h is defined as

$$h_{\boldsymbol{\theta}}(\mathbf{z}) = \frac{\theta_1}{(\theta_2+\|\mathbf{z}\|)^{\theta_3}}.$$

The parameter values considered are $\alpha = 1$, $\theta_1 = 1$, $\theta_2 = 1$, $\theta_3 = 2$, $\theta_4 = 1$, $\sigma^2_\varepsilon = 0.01$, for Case 1, and $\alpha = 1$, $\theta_1 = 1$, $\theta_2 = 1$, $\theta_3 = 0.2$, $\sigma^2_\varepsilon = 0.01$, for Case 2. In both cases, the process X is simulated at 1024 equally spaced time points, at 70 spatial locations randomly selected from a regular 128×128 grid defined on the domain $D = [0,1]^2$. The wavelet transform is implemented in terms of Haar and Daubechies (order 6) wavelets, respectively for Cases 1 and 2. From the results displayed in Figures 1 to 4, it can be observed that the smoother behavior of the Hölder exponent function in Case 1 allows to fit

the optimal number of clusters within the resolution levels considered. However, the fractal behavior of the Hölder exponent function in Case 2 requires a higher number of resolution levels to be computed for estimation and, consequently, for appropriate determination of the optimal number of clusters required. Similar results can be appreciated regarding accuracy in estimation of the Hölder exponent function depending on the regularity conditions of the wavelet basis.

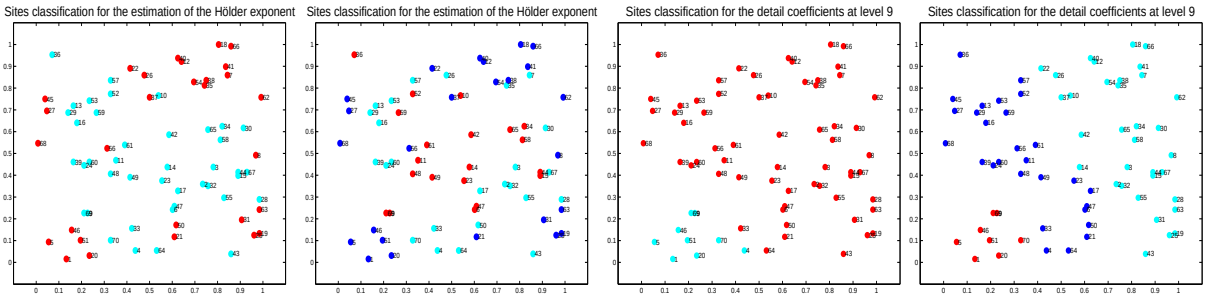

Figure 3: Case 2. Classification of the sites in 2 and 3 groups obtained by *k*-means clustering on the estimated values of the Hölder exponent and on the detail coefficients at level 9, applying the Daubechies (order 6) wavelet transform (from left to right).

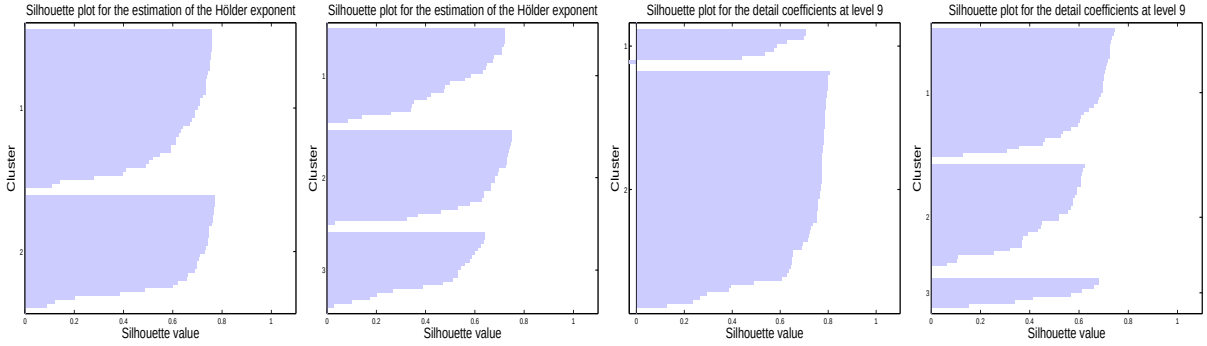

Figure 4: Case 2. Silhouette plots by *k*-means clustering on the estimated values of the Hölder exponent and on the detail coefficients at level 9, for 2 and 3 groups, applying the Daubechies (order 6) wavelet transform (from left to right).

Acknowledgments. Work partially supported by projects P08-FQM-3834 and P09-FQM-5052 of the Andalusian CICE, and projects MTM2009-13250 and MTM2009-13393 of the SGPI, MICINN, Spain.

References

[1] Christakos, G. (2000). *Modern Spatiotemporal Geostatistics*. Springer.

[2] Ruiz-Medina, M.D., and Angulo, J.M. (2007). Functional estimation of spatiotemporal heterogeneities. *Environmetrics* **18**, 775–792.

[3] Ruiz-Medina, M.D., Anh, V.V. and Angulo, J.M. (2004). Fractional generalized random fields of variable order. *Stochastic Analysis and Applications* **22**, 775–800.