University of Granada

Department of Statistics and Operations Research

Statistical Analysis of Extreme Values in Spatio-Temporal Processes

Ana Esther Madrid García

Thesis supervised by:

Prof. José Miguel Angulo Ibáñez

Granada 2010

Editor: Editorial de la Universidad de Granada
Autor: Ana Esther Madrid García
D.L.: GR 616-2011
ISBN: 978-84-694-0153-8

This thesis work has been developed by Ana Esther Madrid García under the supervision of Professor José Miguel Angulo Ibáñez, and submitted to the department of Statistics and Operations Research of the University of Granada, for the degree of Doctor of Philosophy at University of Granada.

Ana Esther Madrid García

Certified by

Prof. José Miguel Angulo Ibáñez

Granada, 2010

Acknowledgements

Firstly, I am heartily thankful to my supervisor Prof. José Miguel Angulo, who has helped me to achieve all the knowledge collected in this thesis, and who has provided me the illusion of wanting to be a researcher.

I also deeply thank Prof. María Dolores Ruiz Medina. Starting without knowing anything is hard, but she helped me to give my first steps and she has continued doing it.

I am especially grateful to Prof. Jorge Mateu for having invited me to work with him in the department of Mathematics of University Jaume I of Castellon. Thanks also to all my partners there for making me feel welcome.

I am indebted to Prof. Carlo Gaetan for receiving me in the department of Statistics of the University Ca'Foscari of Venice, affording me the opportunity to obtain a European doctorate and live a great experience.

I must also acknowledge the department of Statistics and Operations Research of the University of Granada, from the beginning of my studies it has supplied me wonderful teachers, later colleagues and now friends. Thanks to all of them for providing me a friendly environment in which to develop my research.

Finally, I would like to show my gratitude to all the people who have shared my life during the development of this thesis and who have helped me in the bad times, and especially to my family. To my sister, first she encouraged me to come to study to Granada and later to do a doctorate, supporting me in all moments. To my brother, for his good advices and protection since I was a child. To my parents, for the example, education, and life that they have provided me. And to my husband who shares my life since more than eleven years, he always has helped me to get up and follow walking when I have fallen with his love.

I wish to dedicate this thesis to my family and to those who help me from heaven.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

This thesis work has been developed in the framework of granted projects P05-FQM-09900 and P08-FQM-03834 of Consejería de Innovación, Ciencia y Empresa, Junta de Andalucía, and MTM2009-13250 of Ministerio de Ciencia e Innovación, Spain. The title of the thesis, "Statistical Analysis of Extreme Values in Spatio-Temporal Processes", makes reference to a broad context, of increasing relevance in current statistical and modeling research, in which various aspects of interest have been investigated. Specifically, the following topics are the main object of this research:

- Analysis of geometrical characteristics of spatial threshold exceedance sets. Effect of deformation and blurring transformations.

- Application of (marked) point process techniques to the analysis of distribution patterns in spatial threshold exceedance sets.

- Wavelet analysis of temporal intermittency and the effect of time deformation.

- Block thresholding techniques based on entropy criteria for spatial correlated shrinkage.

- Extended formulation of blur-generated space-time models involving dynamical deformation.

In particular, differential aspects in relation to underlying model properties such as local variability and dependence ranges, as well as to threshold specifications, are explored.

Extremal events constitute a natural manifestation inherent to most real phenomena. Their intrinsic rareness and complexity, jointly with the fact that, in many cases, they lead to important consequences in terms of human or economical costs, has attracted the interest of many scientists with the aim to explain their genesis, structural characteristics and possible predictability. The analysis of extremal events has a long history in statistical literature, and has lead to properly recognized theories from various perspectives, including the (Generalized) Extreme Value Theory (GEVT), the Theory of Large Deviations, among others.

Risk indicators used in practical studies are commonly formulated in relation to the possibility of occurrence of extremal events defined by exceedance over a given threshold, with such a threshold representing a certain critical level for the magnitude of interest. Part of the GEVT is then devoted to the probabilistic and statistical study of threshold exceedances. Parallelly, in the context of stochastic processes (or random fields), the structural characteristics of excursion sets defined by threshold exceedances (hereafter referred to as threshold exceedance sets), defined by the coordinates where the realizations or sample paths exceed a given threshold, are intrinsically related to the generating process properties. In particular, different scenarios of local variability and dependence ranges, jointly with the threshold considered, are expected to provide different configurations of the critical episodes or areas at risk. Such information is useful for estimation of certain risk probabilities and pattern-related indicators (e.g. recurrence, persistence, etc.). In this thesis work, we have considered two main approaches of analysis in this context. The first one addresses the geometrical analysis of threshold exceedance sets; in particular, the Euler characteristic and the hypervolume (two special cases of the *intrinsic volumes*) can be related, respectively, to the probability of having at least one point in a critical state and with the extent (in time and/or space) of a critical event. The second one is based on (marked) point process techniques, applied to the study of the temporal/spatial distribution of the isolated connected components within a critical episode represented by a threshold exceedance set. Specific aspects investigated under both approaches are described below.

As mentioned above, one of the key aspects in this context is to analyze in which form certain properties of the underlying process of interest are inherited by threshold exceedance sets in terms of their structural characteristics. In the last decades, there has been an increasing interest and active research in random field modeling; in particular, es-

pecially in the recent years, in the spatio-temporal context. The main aim is to formulate flexible model families, suitable to represent features observed in real phenomena, whose behavior is commonly subject to complex dynamics. This area of research has lead to advanced developments in relation to ranges of dependence, local variability, multiscale properties, spatio-temporal interaction, and heterogeneities, among other aspects.

The contributions derived in this research are described next, with reference to the original motivation and aspects investigated. A synthetic view, as well as some related problems and lines for continuing research, are given in the final chapter.

Transformations such as deformation and blurring are meaningful in many applications, mainly from two points of view. First, in many cases the process dynamics or its observation physically involves such type of transformations, which can be explained in various forms; for instance, deformation can be related to unstability of the medium, or to the effect of some covariates on an expected regular behavior; blurring can be associated to diffusivity; both, deformation and blurring, can account for technical characteristics of observation devices, etc. Second, in some cases deformation, by inversion, can be used as a means to approximate heterogeneously behaved processes in terms of simpler homogeneous models, in warping techniques, etc.; blurring may be used to smooth out local variation energy not useful for certain study purposes, etc. In Chapter 3, we study the effect of deformation and blurring transformation on geometrical characteristics — namely, the Euler characteristic and the hypervolume— of threshold exceedance sets. In particular, we analyze through simulation differential effects in relation to dependence ranges and local variability, as well as to increasing values of the threshold. For deformation, two distinct cases are considered: 'level' variables, where deformation only implies coordinate translations, and 'flow' variables, in which aggregation leads to local change of the variable states.

In Chapter 4, we make a complementary analysis on pattern characteristics of threshold exceedance sets, using (marked) point process techniques. The fragmented nature of threshold exceedance sets, particularly for high thresholds, makes this approach particularly useful to study certain characteristics related to the spatial and/or temporal distribution of the isolated connected components within. As mentioned before, certain risk indicators can be formulated in terms of such characteristics. Given a random field model and a certain subregion in its domain, a family of point processes can be defined by the centroids of the connected components of threshold exceedance sets, each one

associated to a fixed threshold, with various possible marks such as size, orientation, etc. The study has been addressed to evaluation of distribution patterns, including aggregation/inhibition effects, degree of homogeneity, and possible anisotropy, depending on the underlying model properties and the threshold considered.

A feature of interest in risk analysis is intermittency. The concept of intermittency has different interpretations in the literature, the most common and closer being the pseudoperiodic occurrence of high level or variation episodes in the realization of a process, but in other contexts it refers directly to a multifractal scaling behavior. In Chapter 5, based on the Continuous Wavelet Transform applied to a temporal signal, we first analyze the effect of inter- and intra-scale transfer of energy derived from time deformation, according to its local contraction/dilation properties. Then, we study the aggregated effect through a specific scale-dependent intermittency indicator, for different types of real and random process generated signals.

A variety of thresholding techniques based on the Discrete Wavelet Transform have been introduced for filtering signals affected by additive observation random noise. In this context, the exceedance of wavelet coefficients over a conveniently specified threshold is interpreted as evidence of the presence of signal energy, with respect to pure noise. Most of these techniques have been developed for deterministic signals. In Chapter 6, we propose a correlated nonlinear shrinkage estimation methodology for spatial random signals, where block hard thresholding is performed based on mutual information criteria. Specifically, a scale-dependent hard thresholding rule for selection of the most informative wavelet blocks is defined based on the cumulative empirical distribution of the mutual information ratios at each resolution level.

The above mentioned aspects treated in this research establish various methodological approaches which constitute the basis for continuing work, in particular, in the spatio-temporal context. In Chapter 7, based on a well-known discrete-time continuous-space blur-generated spatio-temporal non-separable homogeneous model, we introduce an extended formulation involving a dynamical deformation of the spatial domain. Such an extended model formulation provides much higher flexibility to incorporate certain dynamical heterogeneities observed in real phenomena which can be interpreted, for instance, in terms of unstability if the medium, or to account for the effect of certain covariates. Self-consistency conditions for an underlying continuous-time model are investigated in terms of the existence of an infinitesimal generator for the deformation

field.

Computational and graphical work developed in this thesis has been carried out using different software environments. Most procedures have been implemented creating our own font code in **MATLAB** language, especially appropriate for matrix manipulation, function and data representation, as well as algorithm implementation. Random field realizations have been generated using the **R** library *RandomFields* written by Martin Schlather. Computations related to the analysis of spatial point processes have been also performed in **R**, using the *Spatstat* library written by Adrian Baddeley and Rolf Turner, as well as adapted font codes designed by Jorge Mateu, which occasionally call some **FORTRAN** subroutines to save computing time. The main font codes originally developed in this thesis are included in the Appendix.

Capítulo 1

Introducción

Este trabajo de tesis ha sido desarrollado en el marco de los proyectos P05-FQM-09900 y P08-FQM-03834 de la Consejería de Innovación, Ciencia y Empresa, Junta de Andalucía, y MTM2009-13250 del Ministerio de Ciencia e Innovación. El título de la tesis, "Análisis Estadístico de Valores Extremos en Procesos Espacio-Temporales", hace referencia a un contexto amplio, de relevancia creciente en la investigación actual en estadística y modelización, en el que se han investigado varios aspectos de interés. Específicamente, los siguientes temas constituyen el objeto principal de esta investigación:

- Análisis de características geométricas de conjuntos de excedencias de umbrales espaciales. Efecto de transformaciones de deformación y borrosidad.

- Aplicación de técnicas de procesos puntuales (marcados) al análisis de patrones de distribución espacial en conjuntos de excedencias de umbrales espaciales.

- Análisis mediante *wavelets* de intermitencia temporal y efecto de la deformación del tiempo.

- Técnicas *block thresholding* basadas en criterios de entropía para *shrinkage* correlado espacial.

- Formulación extendida de modelos en espacio-tiempo generados mediante borrosidad involucrando deformación.

En particular, se exploran aspectos diferenciales en relación con propiedades del modelo subyacente tales como variabilidad local y rangos de dependencia, así como con

especificaciones del umbral. Sucesos extremos constituyen una manifestación natural inherente a la mayor parte de los fenómenos reales. Su rareza y complejidad intrínsecas, junto al hecho de que, en muchos casos, conducen a importantes consecuencias en términos de costes humanos o económicos, ha atraído el interés de muchos científicos con el propósito de explicar su génesis, características estructurales y posible predecibilidad. El análisis de sucesos extremos tiene una larga historia en la literatura estadística, y ha dado lugar a teorías propiamente reconocidas desde varias perspectivas, incluyendo la Teoría de Valores Extremos (Generalizada) (TVEG), la Teoría de Grandes Desviaciones, entre otras.

Indicadores de riesgo usados en estudios prácticos se formulan comúnmente en relación con la posibilidad de ocurrencia de sucesos extremos definidos por excedencias sobre un umbral dado, representando tal umbral un cierto nivel crítico para la magnitud de interés. Parte de la TVEG se dedica entonces al estudio probabilístico y estadístico de excedencias de umbrales. Paralelamente, en el contexto de procesos estocásticos (o campos aleatorios), las características estructurales de conjuntos de excursión definidos por excedencias de umbrales (a partir de ahora referidos como conjuntos de excedencias de umbrales), definidos por las coordenadas donde las realizaciones o trayectorias exceden un umbral dado, están intrínsecamente relacionados con las propiedades del proceso generador. En particular, se espera que diferentes escenarios de variabilidad local y rangos de dependencia, junto con el umbral considerado, proporcionen diferentes configuraciones de los episodios críticos o áreas en riesgo. Tal información es útil para la estimación de ciertas probabilidades de riesgo e indicadores relativos a patrones (e.g. recurrencia, persistencia, etc.). En este trabajo de tesis, hemos considerado dos enfoques principales en este contexto. El primero está dirigido al análisis geométrico de conjuntos de excedencias de umbrales; en particular, la característica de Euler y el hipervolumen (dos casos especiales de los *volúmenes intrísecos*) pueden relacionarse, respectivamente, con la probabilidad de tener al menos un punto en un estado crítico y con la extensión (en tiempo y/o espacio) de un suceso crítico. El segundo se basa en técnicas de procesos puntuales (marcados), aplicadas al estudio de la distribución temporal/espacial de las componentes conexas aisladas, dentro de un episodio crítico representado por un conjunto de excedencias de umbral. Más abajo se describen aspectos específicos investigados bajo ambos enfoques.

Como se ha mencionado más arriba, uno de los aspectos clave en este contexto consiste en analizar en qué forma ciertas propiedades del proceso de interés subyacente son heredadas por conjuntos de excedencias de umbrales en términos de sus características estructurales. En las últimas décadas, ha habido un interés creciente y una investigación activa en modelización de campos aleatorios; en particular, especialmente en los años recientes, en el contexto espacio-temporal. El principal propósito es formular familias flexibles de modelos, adecuadas para representar rasgos observados en fenómenos naturales, cuyo comportamiento está comúnmente sujeto a dinámicas complejas. Este área de la investigación ha dado lugar a desarrollos avanzados en relación con rangos de dependencia, variabilidad local, propiedades multiescalares, interación espacio-temporal y heterogeneidades, entre otros aspectos.

A continuación se describen las contribuciones derivadas de esta investigación, con referencia a la motivación original y aspectos tratados. Una visión sintética, así como algunos problemas relacionados y líneas de continuación de la investigación, se dan en el capítulo final.

Transformaciones tales como deformación y borrosidad son significativas en muchas aplicaciones, principalmente desde dos puntos de vista. Primero, en muchos casos la dinámica del proceso o su observación involucran físicamente tal tipo de transformaciones, que pueden explicarse de varias formas; por ejemplo, la deformación puede relacionarse con inestabilidad del medio, o con el efecto de algunas covariables sobre un comportamiento regular esperado; la borrosidad puede asociarse a difusividad; ambas, deformación y borrosidad, pueden justificarse por características técnicas de dispositivos de observación, etc. Segundo, en algunos casos la deformación, por inversión, puede usarse como un medio de aproximar procesos con comportamiento heterogéneo en términos de modelos más simples homogéneos, en técnicas *warping*, etc.; la borrosidad puede usarse para eliminar por suavizado energía de variación local no útil para ciertos propósitos de estudio, etc. En el Capítulo 3, estudiamos el efecto de transformaciones de borrosidad y deformación sobre características geométricas —concretamente, la característica de Euler y el hipervolumen— de conjuntos de excedencias de umbrales. En particular, analizamos a través de simulación efectos diferenciales en relación con rangos de dependencia y variabilidad local, así como con valores crecientes del umbral. Para la deformación, se consideran dos casos distintos: variables 'nivel', en la cuales la deformación sólo implica traslaciones de coordenadas, y variables 'flujo', en las que la agregación conduce al

cambio local de los estados de la variable.

En el Capítulo 4, realizamos un análisis complementario sobre características de patrones en conjuntos de excedencias de umbrales, usando técnicas de procesos puntuales (marcados). La naturaleza fragmentada de los conjuntos de excedencias de umbrales, particularmente para umbrales altos, hace este enfoque particularmente útil para estudiar ciertas características relativas a la distribución espacial y/o temporal de las componentes conexas aisladas incluídas. Como se mencionó antes, ciertos indicadores de riesgo pueden formularse en términos de tales características. Dado un modelo de campo aleatorio y una cierta subregión de su dominio, puede definirse una familia de procesos puntuales mediante los centroides de las componentes conexas de los conjuntos de excursión, cada uno asociado a un umbral fijado, con varias marcas posibles tales como tamaño, orientación, etc. El estudio ha sido dirigido a la evaluación de patrones de distribución, incluyendo efectos de agregación/inhibición, grado de homogeneidad, y posible anisotropía, dependiendo de las propiedades del modelo subyacente y del umbral considerado.

Un rasgo de interés en análisis de riesgos es la intermitencia. El concepto de intermitencia tiene diferentes interpretaciones en la literatura, siendo la más común y próxima la ocurrencia pseudoperiódica de episodios de alto nivel o variabilidad en la realización de un proceso, pero en otros contextos se refiere directamente a un comportamiento de escalamiento multifractal. En el Capítulo 5, basándonos en la Transformada Wavelet Continua aplicada a una señal temporal, primero analizamos el efecto de transferencia inter- e intra-escalar de energía derivado de la deformación del tiempo, de acuerdo con sus propiedades locales de contracción/dilatación. Entonces, estudiamos el efecto agregado a través de un indicador específico de intermitencia dependiente de la escala, para diferentes tipos de señales reales o generadas mediante procesos aleatorios.

Una variedad de técnicas basadas en la Transformada Wavelet Discreta han sido introducidas para el filtrado de señales afectadas por ruido aleatorio aditivo de observación. En este contexto, la excedencia de un coeficiente wavelet sobre un umbral especificado se interpreta como evidencia de la presencia de energía de la señal, con respecto a ruido puro. La mayor parte de esas técnicas han sido desarrolladas para señales determinísticas. En el Capítulo 6, proponemos una metodología de estimación no lineal correlada *shrinkage* para señales aleatorias espaciales, donde se realiza *block hard thresholding* basado en criterios de información mutua. Específicamente, se define una regla *hard thresholding*

dependiente de la escala para la selección de los bloques de coeficientes *wavelets* más informativos, basada en la distribución empírica acumulativa de las razones de información mutua en cada nivel de resolución.

Los aspectos mencionados más arriba tratados en esta investigación establecen varios enfoques metodológicos que constituyen la base para la continuación de la investigación, en particular, en el contexto espacio-temporal. En el Capítulo 7, basándonos en un conocido modelo espacio-temporal, en tiempo discreto y espacio continuo, homogéneo y no separable, generado mediante borrosidad, introducimos una formulación extendida involucrando una deformación dinámica del dominio espacial. Tal formulación extendida del modelo proporciona una flexibilidad mucho mayor para incorporar ciertas heterogeneidades dinámicas observadas en fenómenos reales que pueden interpretarse, por ejemplo, en términos de inestabilidad del medio, o explicando el efecto de ciertas covariables. Se investigan condiciones de auto-consistencia para un modelo subyacente en tiempo continuo, en términos de la existencia de un generador infinitesimal para el campo de deformación.

El trabajo computacional y gráfico desarrollado en esta tesis ha sido llevado a cabo usando diferentes entornos de *software*. La mayor parte de los procedimientos han sido implementados creando código fuente propio en lenguaje **MATLAB**, especialmente apropiado para manipulación de matrices, representación de funciones y de datos, así como implementación de algoritmos. Realizaciones de campos aleatorios han sido generadas usando la librería *RandomFields* escrita en **R** por Martin Schlather. Cálculos relacionados con el análisis de procesos puntuales espaciales han sido también realizados en **R**, usando la librería *Spatstat* escrita por Adrian Baddeley y Rolf Turner, así como adaptaciones de códigos fuente diseñados por Jorge Mateu, que ocasionalmente llaman a algunas subrutinas **FORTRAN** para ahorrar tiempo de cálculo. Se incluyen en el Apéndice los principales códigos fuente originales desarrollados en esta tesis.

Chapter 2

Preliminaries

In this chapter we introduce the main basic elements and techniques which are used through this thesis. To define properly each context, and for the sake of completeness, the contents are presented in a relatively larger extent than strictly needed for the developments in the following chapters.

The chapter is organized as follows. Given that the fundamental topics dealt in this thesis are related with extreme events, in Section 2.1 we present the classical Extreme Value Theory, with reference to the main approaches developed. Section 2.2 is devoted to random fields; in particular, the models considered in the different chapters are specified. Section 2.3 introduces spatial point processes, an approach based on which is considered in Chapter 4. Section 2.4 presents different wavelet elements and tools, which are used in Chapters 5 and 6. Finally, in Section 2.5 we define some entropy measures, which are involved in criteria formulations given in Chapter 6.

2.1 Extreme values

The classical Extreme Value Theory is well recognized as an established scientific discipline. This section gives an introduction about its fundamental developments, with specific reference to the three main existing approaches.

2.1.1 Introduction

Extremal analysis is mainly concerned with the formal study of the abnormal behavior observed in real (natural or derived from human activity) phenomena, quantified in terms of large deviations for appropriate indicators. The extreme value theory plays an important role in many areas, mainly in relation to risk analysis, as, for instance, Geophysics, Environmental Sciences, Engineering, Astrophysics, Finance or Insurance.

Historically, the first work on the extreme value problems was presented by Nicolas Bernouilli in 1709. In that work he discussed the mean largest distance from the origin given n points lying at random on a straight line of a fixed length t. However, the concept of distribution of largest value was not introduced until 1922, when von Bortkiewicz (1922)[28] dealt with the distribution of range in random samples from a normal distribution. Fréchet (1927)[48] found one possible limiting distribution for the largest order statistic and Fisher and Tippet (1928)[47] proved that there are only three types of limiting distributions. Afterwards, these distributions were unified by a reparametrization given by von Mises (1954)[90] and Jenkisson (1955)[67], leading to the generalized extreme value distributions. In the meantime, Gnedenko (1943)[52] introduced a rigorous basis for the extreme value theory giving necessary and sufficient conditions for the weak convergence of the extreme order statistics. Gumbel (1958)[60] was the first to pay the attention to possible applications of the formal extreme value theory from the statistical point of view, his book being still relevant today.

An extension of the previous theory, in which only the maxima are considered, consists of using several of the largest order statistics. This was proposed as a statistical procedure by Weissman (1978)[121] and developed further by Gomes (1981)[57], Smith(1986)[107] and Tawn (1988)[112], although in this model the probabilistic foundations are based on the work of Leadbetter *et al.* (1983)[76].

'Peaks over threshold' (named POT) constitutes an alternative approach to the classical extreme value theory, that looks to study exceedances over high thresholds. Todorovic and Zelenhasic (1970)[114] and Todorovic and Rouselle (1971)[115] gave the first systematic development, although the key result is the theorem of Balkema-Hans-Pickands (Balkema and Hans 1974[24]; Pickands 1975[95]) which states a distribution for a wide class of distribution losses which exceed high enough thresholds.

Therefore, the extreme values theory pays attention to higher values of the distribu-

tion instead of the average, that is, the interest is in the events associated with the tail of the distribution. The way of identifying such events determines the three existing approaches. The first approach considers the maximum that one variable takes in successive periods. These selected observations constitute the extremal events, called block maxima. The second approach deals with several of the largest order statistics. Finally, the third approach focuses on the realizations which exceed a given high threshold. The fundamental theoretical results underlying the three approaches are presented next, although this thesis has been focused in the context of threshold exceedances.

2.1.2 Classical Extreme Value Theory

Assume that the given data y_i are maxima, that is,

$$y_i = \max\{x_{i,1}, ..., x_{i,n}\}, \quad i = 1, ..., m, \tag{2.1}$$

where the $x_{i,j}$ may not be observable. If the $x_{i,j}$ in (2.1) can be observed, then a possibility to extract upper extreme values from a set of data is to get maxima out blocks. This method is called annual maxima, blocks or Gumbel method.

For independent random variables $X_1, ..., X_n$ having a common distribution function F, the distribution of maxima, $M_n = \max\{X_1, ..., X_n\}$, can be computed as

$$P\{M_n \leq z\} = P\{X_1 \leq z, ..., X_n \leq z\} = \prod_{i=1}^{n} P\{X_i \leq z\} = [F(z)]^n.$$

Thus, the y_i's in (2.1) are governed by F^n if the $x_{i,j}$'s are governed by F. The problem is that, in practice, the distribution function F is usually unknown. A common approach is to look for approximate families of models for F^n which can be estimated on the basis of the extreme data only. However, for any z less than the upper end-point of F, $z_+ = \inf\{z : F(z) = 1\}$, the distribution of M_n degenerates to a point mass on z_+, that is,

$$\lim_{n\to\infty} F^n(z) = \begin{cases} 0, & z < z_+, \\ 1, & z \geq z_+. \end{cases}$$

To avoid this difficulty, a linear renormalization of the variable M_n is used:

$$M_n^* = \frac{M_n - b_n}{a_n},$$

where $\{a_n > 0\}$ and $\{b_n\}$ are appropriate sequences of constants which stabilize the location and scale of M_n^* as n increases and avoid the degeneration of M_n^*,

$$\begin{aligned} P\{M_n^* \leq z\} &= P\left\{\frac{M_n - b_n}{a_n} \leq z\right\} = P\{M_n \leq a_n z + b_n\} \\ &= F^n(a_n z + b_n) \xrightarrow[n\to\infty]{} G(z). \end{aligned}$$

Definition 2.1.1. Let $F(z)$ and $G(z)$ be two non-degenerate distribution functions. $F(z)$ is said to be *max-stable* or to *belong to the maximal domain of attraction of* $G(z)$, $F \in \mathcal{D}(G)$, if

$$\lim_{n\to\infty} F^n(a_n z + b_n) = G(z), \quad \forall z, \tag{2.2}$$

for a suitable choice of constants $\{a_n > 0\}$ and $\{b_n\}$.

Theorem 2.1.1. The only non-degenerate families of max-stable distributions are:

$$I\text{: } G(z) = \exp\left\{-\exp\left\{-\left(\tfrac{z-b}{a}\right)\right\}\right\}, \quad -\infty < z < \infty;$$

$$II\text{: } G(z) = \begin{cases} 0, & z \leq b, \\ \exp\left\{-\left(\tfrac{z-b}{a}\right)^{-\alpha}\right\}, & z > b; \end{cases}$$

$$III\text{: } G(z) = \begin{cases} \exp\left\{-\left[-\left(\tfrac{z-b}{a}\right)^{\alpha}\right]\right\}, & z < b, \\ 1, & z \geq b; \end{cases}$$

for parameters $a > 0$, b and, in the cases II and III, $\alpha > 0$.

Theorem 2.1.1 implies that if there are two sequences of constants $\{a_n > 0\}$ and $\{b_n\}$ such that $F^n(a_n z + b_n)$ converges to a non-degenerate distribution $G(z)$, then the only possible limiting distributions for the rescaled maxima, $M_n^* = (M_n - b_n)/a_n$, belong to one of the families labeled I, II and III, which are designated jointly *'extreme value distributions'* (EVD), with types I, II and III corresponding to Gumbel, Fréchet and Weibull families, respectively. Each family has a location and a scale parameter, b and a, respectively; additionally, the Fréchet and Weibull families have a shape parameter

α. However, although the three types of EVD have a different behavior, the Fréchet and Weibull distributions attain the shape of a Gumbel distributions when the shape parameter goes to infinity and minus infinity, respectively.

Von Mises (1954)[90] and Jenkinson (1955)[67] showed, separately, that taking the reparameterization $\xi = 1/\alpha$ of EVD and employing appropriate values for the location and scale parameters, a continuous and unified model termed *'generalized extreme value'* (GEV) family of distributions can be obtained, which is given by

$$G(z) = \exp\left\{-\left[1+\xi\left(\frac{z-\mu}{\sigma}\right)\right]^{-1/\xi}\right\}, \tag{2.3}$$

defined on the set $\{z : 1+\xi(z-\mu)/\sigma > 0\}$, where $-\infty < \mu < \infty$, $\sigma > 0$ and $-\infty < \xi < \infty$ are the location, scale, and shape parameters, respectively. The type II and type III classes of EVD respectively correspond to the cases $\xi > 0$ and $\xi < 0$ in this reparameterization. The subset of GEV family with $\xi = 0$ is interpreted as the limiting of (2.3) as $\xi \to 0$, leading to the Gumbel family. In applications, the model unification has the advantage of avoiding to have to choose which model of three families is most appropriate for the data.

The previous results can be formally re-stated in the next theorem (see Coles, 2001[37]).

Theorem 2.1.2. Let $X_1, X_2, ..., X_n$ be a sequence of independent and identically distributed random variables and let $M_n = \max\{X_1, ..., X_n\}$. If there exist sequences of constants $\{a_n > 0\}$ and $\{b_n\}$ such that

$$P\left\{\frac{M_n - b_n}{a_n} \leq z\right\} \to G(z), \quad \text{as } n \to \infty,$$

for a non-degenerate distribution function G, then G is a member of the GEV family

$$G(z) = \exp\left\{-\left[1+\xi\left(\frac{z-\mu}{\sigma}\right)\right]^{-1/\xi}\right\}, \tag{2.4}$$

defined on $\{z : 1+\xi(z-\mu)/\sigma > 0\}$, where the parameters satisfy $-\infty < \mu < \infty$, $\sigma > 0$ and $-\infty < \xi < \infty$.

In some cases, as for instance in problems of system failure models, the interest is in the minima rather than maxima observation. Results for minima can be deduced by an analogous argument as applied above for maxima, or observing that

$$m_n = \min\{X_1, ..., X_n\} = -\max\{-X_1, ..., -X_n\} = -M_n.$$

Then, for large n,

$$\begin{aligned} P\{m_n \leq z\} = P\{-M_n \leq z\} = P\{M_n \geq -z\} &= 1 - P\{M_n \leq -z\} \\ &\approx 1 - \exp\left\{-\left[1 + \xi\left(\frac{-z-\mu}{\sigma}\right)\right]^{-1/\xi}\right\} \\ &= 1 - \exp\left\{-\left[1 - \xi\left(\frac{z-\tilde{\mu}}{\sigma}\right)\right]^{-1/\xi}\right\} \end{aligned}$$

on $\{z : 1 - \xi(x - \tilde{\mu})/\sigma > 0\}$, where $\tilde{\mu} = \mu$.

2.1.3 The r largest order statistic model

Suppose that $X_1, X_2, ..., X_n$ is a sequence of independent and identically distributed random variables with distribution F and let $X_{(1)} \leq X_{(2)} \leq ... \leq X_{(n)}$ be an arrangement in an increasing order. The previous result can be extended to other extreme order statistics, by defining

$$M_n^{(k)} = X_{(n-k)} = k\text{-th largest element of } \{X_1, ...X_n\}$$

and identifying the limiting behavior of this variable, for fixed k, as n goes to infinity.

Theorem 2.1.3. If there exist sequences of constants $\{a_n > 0\}$ and $\{b_n\}$ such that

$$P\left\{\frac{M_n - b_n}{a_n} \leq z\right\} \to G(z), \quad \text{as } n \to \infty,$$

then, for fixed k,

$$P\left\{\frac{M_n^{(k)} - b_n}{a_n} \leq z\right\} \to G_k(z),$$

as $n \to \infty$, on $\{z : 1 + \xi(z - \mu)\sigma > 0\}$, where

$$G_k(z) = \exp\{\tau(z)\} \sum_{s=0}^{k-1} \frac{\tau(z)^s}{s!}, \tag{2.5}$$

with

$$\tau(z) = \left[1 + \xi\left(\frac{z-\mu}{\sigma}\right)\right]^{-1/\xi}.$$

This theorem states that, if the k-th largest order statistic in a block is normalized in exactly the same way as the maximum, then its limiting distribution is of the form given by (2.5), the parameters of which correspond to the parameters of the limiting GEV distribution of the block maximum.

Frequently, in applications, it is necessary to use the joint distribution of several order statistics. This distribution is given by the following theorem.

Theorem 2.1.4. If there exist sequences of constants $\{a_n > 0\}$ and $\{b_n\}$ such that

$$\lim_{n\to\infty} F^n(a_n z + b_n) = G(z), \quad \forall z,$$

then, for fixed r, the limiting distribution as $n \to \infty$ of

$$\tilde{\mathcal{M}}_n^{(r)} = \left(\frac{M_n^{(1)} - b_n}{a_n}, ..., \frac{M_n^{(r)} - b_n}{a_n}\right)$$

falls within the family having joint probability density function

$$\begin{aligned} f(z^{(1)}, ..., z^{(r)}) = &\exp\left\{-\left[1 + \xi\left(\frac{z^{(r)} - \mu}{\sigma}\right)\right]^{-1/\xi}\right\} \\ &\times \prod_{k=1}^{r} \sigma^{-1}\left[1 + \xi\left(\frac{z^{(k)} - \mu}{\sigma}\right)\right]^{-1/\xi - 1}, \end{aligned}$$

where $-\infty < \mu < \infty$, $\sigma > 0$, $-\infty < \xi < \infty$, $z^{(r)} \le z^{(r-1)} \le ... \le z^{(1)}$, and $z^{(k)} : 1 + \xi(z^{(k)} - \mu/\sigma > 0)$, for $k = 1, ..., r$.

If $r = 1$, then the annual maximum is modeled and the limiting distribution coincides with GEV distributions.

2.1.4 Threshold models

Another method to extract upper extremes from a set of data $x_1, ..., x_n$ is to take the exceedances over a predetermined, high threshold u. Thus, threshold model methods are much more flexible than methods based on maxima, given that much more data are used. Exceedances over a fixed high threshold u are those x_i with $x_i > u$. Label these exceedances by $x_{(1)}, ..., x_{(k)}$, and define threshold excesses by $y_j = x_{(j)} - u$, for $i = 1, ..., k$.

Denoting an arbitrary term in the X_i sequence by X, it follows that a description of the stochastic behavior of extreme events is given by the conditional probability

$$P\{X > u + y | X > u\} = \frac{1 - F(u + y)}{1 - F(u)}, \quad y > 0.$$

Usually, the distribution F is unknown and therefore, to obtain the distribution of threshold exceedances, it is necessary to seek approximations, as in the case of the distribution of maxima. The following theorem gives the asymptotic model for the exceedances over a large enough threshold.

Theorem 2.1.5. Let $X_1, X_2, ...$ be a sequence of independent and identically distributed random variables, and let $M_n = \max\{X_1, ..., X_n\}$. Denote an arbitrary term in the sequence by X, and suppose that F satisfies theorem 2.1.2, so that for large n,

$$P\{M_n \leq z\} \approx G(z),$$

where

$$G(z) = \exp\left\{ - \left[1 + \xi \left(\frac{z - \mu}{\sigma} \right) \right]^{-1/\xi} \right\},$$

for some μ, $\sigma > 0$ and ξ. Then, for large enough u, the distribution function of $(X - u)$, conditional on $X > u$, is approximately

$$H(y) = 1 - \left[1 + \frac{\xi y}{\tilde{\sigma}} \right]^{-1/\xi}, \tag{2.6}$$

defined on $\{y : y > 0 \text{ and } (1 + \xi y/\tilde{\sigma}) > 0\}$, where

$$\tilde{\sigma} = \sigma + \xi(u - \mu).$$

A precise proof of theorem 2.1.5 is given in Leadbetter *et al.* (1983)[76]. This theorem states that the distribution of threshold excesses belongs to the family defined by (2.6), termed *'generalized Pareto'* (GP) family, if the distribution of block maxima belongs to the GEV family. In that case, the parameters of the GP distribution of threshold excesses are uniquely determined by those of the associated GEV distribution of block maxima. In (2.6) the case $\xi = 0$ is interpreted as the limit as $\xi \to 0$, i.e., the exponential distribution with mean σ. The case $\xi > 0$ is just a reparameterization of one of several forms of the usual Pareto distributions, and the extension to $\xi \leq 0$ was given by Pickands (1975)[95].

2.2 Random fields

The literature on spatial, temporal and spatio-temporal random field is very vast. Here we mention some well-known references of particular interest, such as Chilès and Delfiner (1999)[31], Christakos (1992[33], 2000[34]), Cressie (1993)[38], Stein (1999)[108], Wackernagel (1998)[120], Yaglom (1987)[124], among others. In this section we introduce some basic notions, formulate some specific models used in the following chapters, and define some geometrical characteristics of interest in relation to the structure of extremal events.

2.2.1 Basic definitions

Definition 2.2.1. Let $D \subset \mathbb{R}^d$, generally $d \geq 2$, be a domain with a positive volume, and $(\Omega, \mathcal{A}, P)$ be a probability space. A *random field* (RF), also called *random function* or *random process*, is a set of variables $\{X(\boldsymbol{s}, \omega) : \boldsymbol{s} \in D, \omega \in \Omega\}$ such that for each $\boldsymbol{s} \in D$ the section $X(\boldsymbol{s}, \cdot)$ is a random variable on $(\Omega, \mathcal{A}, P)$. Each of the functions $X(\cdot, \omega)$ defined on D as the section of the RF at $\omega \in \Omega$ is a realization of the RF. For short the RF is simply denoted by $X(\boldsymbol{s})$.

Thus a RF on D can be viewed as a family of random variables indexed by the elements of D, defined on a common probability space. If the set D is finite then $X(\boldsymbol{s})$ can be regarded as a single multidimensional random variable and it is specified by its multidimensional distribution. However, if the set D is infinite, then the RF $X(\boldsymbol{s})$ is described by the multidimensional distributions corresponding to all finite configurations of random variables of the RF. Let $\boldsymbol{s}_1, \boldsymbol{s}_2, ..., \boldsymbol{s}_n$ be n points of D. The finite-dimensional

distributions of the field $X(\boldsymbol{s})$ are defined as

$$F_{\boldsymbol{s}_1,\boldsymbol{s}_2,...,\boldsymbol{s}_n}(x_1,x_2,...,x_n) = P\{X(\boldsymbol{s}_1) \leq x_1, X(\boldsymbol{s}_2) \leq x_2, ..., X(\boldsymbol{s}_n) \leq x_n\}. \tag{2.7}$$

These functions must satisfy Kolmogorov's conditions of symmetry and compatibility (Yaglom 1987[124]):

a) Symmetry condition:

$$F_{\boldsymbol{s}_{i_1},\boldsymbol{s}_{i_2},...,\boldsymbol{s}_{i_n}}(x_{i_1},x_{i_2},...,x_{i_n}) = F_{\boldsymbol{s}_1,\boldsymbol{s}_2,...,\boldsymbol{s}_n}(x_1,x_2,...,x_n),$$

where $i_1, i_2, ..., i_n$ is any permutation of the index $1, 2, ..., n$.

b) Compatibility condition:

$$F_{\boldsymbol{s}_1,\boldsymbol{s}_2,...,\boldsymbol{s}_m,\boldsymbol{s}_{m+1},...,\boldsymbol{s}_n}(x_1,x_2,...,x_m,\infty,...,\infty) = F_{\boldsymbol{s}_1,\boldsymbol{s}_2,...,\boldsymbol{s}_m}(x_1,x_2,...,x_m),$$

for any value of m.

Also, it is usual to assume the existence of density functions corresponding to (2.7),

$$f_{\boldsymbol{s}_1,\boldsymbol{s}_2,...,\boldsymbol{s}_n}(x_1,x_2,...,x_n) = \frac{\delta^n}{\delta x_1 ... \delta x_n} F_{\boldsymbol{s}_1,\boldsymbol{s}_2,...,\boldsymbol{s}_n}(x_1,x_2,...,x_n).$$

An independent RF, so called *white noise RF*, is completely characterized by the univariate density functions $f_{\boldsymbol{s}}(x)$, since in this case

$$f_{\boldsymbol{s}_1,\boldsymbol{s}_2,...,\boldsymbol{s}_n}(x_1,x_2,...,x_n) = \prod_{i=1}^{n} f_{\boldsymbol{s}_i}(x_i).$$

Definition 2.2.2. The *mean* of the RF is the expected value of the random variable $X(\boldsymbol{s})$ at the point $\boldsymbol{s}$,

$$\mu(\boldsymbol{s}) = E[X(\boldsymbol{s})].$$

Definition 2.2.3. A RF $X(\boldsymbol{s})$ on D is said to be of *second-order* if $E[X(\boldsymbol{s})^2] < \infty$ for all $\boldsymbol{s} \in D$.

Definition 2.2.4. The *covariance* of the random variables $X(\boldsymbol{s}_i)$ and $X(\boldsymbol{s}_j)$ is defined

as

$$Cov[X(\boldsymbol{s}_i), X(\boldsymbol{s}_j)] = E[X(\boldsymbol{s}_i) - \mu(\boldsymbol{s}_i)][X(\boldsymbol{s}_j) - \mu(\boldsymbol{s}_j)].$$

This function describes the structural component of the RF, that is, the second-order dependence between the variables of the RF.

Consider that $X(\boldsymbol{s})$ indicates the value taken for the RF in a spatial location $\boldsymbol{s} \in \mathbb{R}^d$; then, $X(\boldsymbol{s})$ will be denominated *spatial RF*. In other cases, we can consider positions of the form $\boldsymbol{s} = (s_1, ..., s_d, t) \in \mathbb{R}^d \times \mathbb{R}$; then, the RF will be denominated *spatio-temporal RF*. As it is indicated by Gneiting *et al.* (2002)[53], physically there is a distinction between the spatial and temporal dimensions, and a statistical study should take this fact into account. However, from the mathematical viewpoint, in which $\mathbb{R}^d \times \mathbb{R} = \mathbb{R}^{d+1}$, there are no differences between the coordinates, and hence all definitions and results given for spatial RFs are applicable to the spatio-temporal case too. In many cases, to emphasize the spatio-temporal nature, the following notation is used:

$$\{X(\boldsymbol{s}, t) : (\boldsymbol{s}, t) \in \mathbb{R}^d \times \mathbb{R}\}.$$

Moreover, some specific definitions have been established for the spatio-temporal context, as, for example, the concepts of separability, full symmetry, or compact support.

Definition 2.2.5. A spatio-temporal RF $X(\boldsymbol{s}, t)$ on $\mathbb{R}^d \times \mathbb{R}$ is said to have *separable covariance function* if there exist purely spatial and purely temporal covariance functions Cov_S and Cov_T, respectively, such that

$$Cov[X(\boldsymbol{s}_1, t_1), X(\boldsymbol{s}_2, t_2)] = Cov_S[\boldsymbol{s}_1, \boldsymbol{s}_2]Cov_T[t_1, t_2]$$

for all space-time coordinates $(\boldsymbol{s}_1, t_1)$ and $(\boldsymbol{s}_2, t_2)$ in $\mathbb{R}^d \times \mathbb{R}$.

Definition 2.2.6. A spatio-temporal RF is said to have *fully symmetric covariance function* if

$$Cov[X(\boldsymbol{s}_1, t_1), X(\boldsymbol{s}_2, t_2)] = Cov[X(\boldsymbol{s}_1, t_2), X(\boldsymbol{s}_2, t_1)]$$

for all $(\boldsymbol{s}_1, t_1)$ and $(\boldsymbol{s}_2, t_2)$ in $\mathbb{R}^d \times \mathbb{R}$.

Definition 2.2.7. A spatio-temporal RF is said to have *compactly supported covariance*

function if $Cov[X(\boldsymbol{s}_1, t_1), X(\boldsymbol{s}_2, t_2)] = 0$ whenever $\|\boldsymbol{s}_1 - \boldsymbol{s}_2\|$ and/or $|t_1 - t_2|$ are sufficiently large.

Properties 2.2.1. 1. A covariance function is positive definite. That is, for all $a_i \in \mathbb{R}$ and for all $\boldsymbol{s}_i \in D$, $i = 1, ..., n$,

$$\sum_{i,j=1}^{n} a_i a_j Cov[X(\boldsymbol{s}_i), X(\boldsymbol{s}_j)] \geq 0.$$

2. $Cov[\boldsymbol{s}_i, \boldsymbol{s}_j] = Cov[\boldsymbol{s}_j, \boldsymbol{s}_i]$.

The positive definiteness property is a relevant condition to characterize functions of RFs as covariance or autocovariance functions (defined below). The following properties are satisfied by positive definite functions:

1. If F_1 and F_2 are positive definite, then $a_1 F_1 + a_2 F_2$ is positive definite, for all non-negative a_1 and a_2.

2. If $F_1, F_2, ...$ are positive definite and $\lim_{n\to\infty} F_n(\boldsymbol{s}) = F(\boldsymbol{s})$, for all $\boldsymbol{s} \in \mathbb{R}^d$, then F is positive definite.

3. If F_1 and F_2 are positive definite, then $F_1 F_2$ is positive definite

Stationarity

A common simplifying assumption to restrict the class of RFs is that the probabilistic structure in some sense looks similar in different parts of D.

Definition 2.2.8. A RF is *strictly stationary* if for all finite n, and $\boldsymbol{s}_1, ..., \boldsymbol{s}_n \in D$,

$$F_{\boldsymbol{s}_1,...,\boldsymbol{s}_n}(x_1, ..., x_n) = F_{\boldsymbol{s}_1+\boldsymbol{h},...,\boldsymbol{s}_n+\boldsymbol{h}}(x_1, ..., x_n), \quad \forall \boldsymbol{h} \in \mathbb{R}^d.$$

Thus stationarity can be thought of as an invariance property under an arbitrary translation of the points by a vector. A less restrictive type of stationarity is defined in terms of the first two moments of $X(\boldsymbol{s})$.

Definition 2.2.9. A secon-order RF $X(\boldsymbol{s})$ is *second-order* or *weakly stationary* if it satisfies:

1. $X(\boldsymbol{s}) = \mu$, for all $\boldsymbol{s} \in D$,

2. $Cov[X(\boldsymbol{s}+\boldsymbol{h}), X(\boldsymbol{s})] = C(\boldsymbol{h})$, for all $\boldsymbol{s} \in D$.

The function $C(\cdot)$ is called *autocovariance*[1] or *covariogram*.

Strict stationarity of a second-order RF implies second-order stationarity, while the reverse is not true. The second-order stationarity is the type of stationarity most frequently assumed, and hence commonly a second-order stationary RF is called simply stationary RF (SRF).

Properties 2.2.2. 1. If $X(\boldsymbol{s})$ is a SRF with autocovariance $C(\cdot)$ then

   a) $C(\mathbf{0}) = Var(X(\boldsymbol{s})) \geq 0$.

   b) $C(\boldsymbol{s}) = C(-\boldsymbol{s})$ (symmetry).

   c) $|C(\boldsymbol{s})| \leq C(\mathbf{0})$.

2. $C(\cdot)$ is the autocovariance function of a SRF if and only if it is symmetric and positive definite.

Given that a Gaussian distribution is completely defined by its first two moments, knowledge of the mean and the covariance function suffices to determine the distribution of a Gaussian RF. Therefore, for a Gaussian RF second-order stationarity is equivalent to strict stationarity.

Definition 2.2.10. A RF is *Gaussian* if all its finite dimensional distributions are multivariate Gaussian.

Gaussian RFs play a central role in modeling data, since they have a rich, detailed and very well understood theory.

In the spatio-temporal context a covariance is said to be '*spatially stationary*' if $Cov[X(\boldsymbol{s}_1, t_1), X(\boldsymbol{s}_2, t_2)]$ depends on the observation locations $\boldsymbol{s}_1$ and $\boldsymbol{s}_2$ only through the spatial separation vector, $\boldsymbol{s}_1 - \boldsymbol{s}_2$. Analogously, it is said to be '*temporally stationary*' if

[1] In the literature the terms covariance (for example, Chilès and Delfiner 1999[31]) and autocovariance (Stein 1999[108]) are used equivalently.

the covariance depends on the observation times t_1 and t_2 only through the temporal lag, $t_1 - t_2$. If a spatio-temporal RF has both spatially and temporally stationary covariance, it is said to have '*stationary covariance*', and then there exists a function C such that

$$Cov[X(\boldsymbol{s}_1, t_1), X(\boldsymbol{s}_2, t_2)] = C(\boldsymbol{s}_1 - \boldsymbol{s}_2, t_1 - t_2)$$

for all $(\boldsymbol{s}_1, t_1)$ and $(\boldsymbol{s}_2, t_2)$ in $\mathbb{R}^d \times \mathbb{R}$. This function is called '*autocovariance*' of the RF, as in spatial case, and its restrictions $C(\cdot, 0)$ and $C(\mathbf{0}, \cdot)$ are purely spatial and purely temporal covariance functions, respectively. Mitchell *et al.* (2005)[91] give a condition for a stationary spatio-temporal RF to be separable, and Gneiting (2002)[53] to be fully symmetric.

Proposition 2.2.3. A stationary space-time autocovariance function is separable if there exist stationary, purely spatial and purely temporal autocovariance functions C_S and C_T, respectively, such that $C(\boldsymbol{h}, u) = C_S(\boldsymbol{h})C_T(u)$ for all $(\boldsymbol{h}, u) \in \mathbb{R}^d \times \mathbb{R}$, or equivalently if the space-time covariance function can be factorized as

$$C(\boldsymbol{h}, u) = \frac{C(\boldsymbol{h}, 0) \cdot C(\mathbf{0}, u)}{C(\mathbf{0}, 0)}$$

for all $(\boldsymbol{h}, u) \in \mathbb{R}^d \times \mathbb{R}$.

Proposition 2.2.4. A stationary space-time autocovariance function is *fully symmetric* if

$$C(\boldsymbol{h}, u) = C(\boldsymbol{h}, -u) = C(-\boldsymbol{h}, u) = C(-\boldsymbol{h}, -u) \tag{2.8}$$

for all $(\boldsymbol{h}, u) \in \mathbb{R}^d \times \mathbb{R}$.

Gneiting *et al.* (2007)[54] summarize the relationships between separable, fully symmetric, stationary and compactly supported covariance functions of RFs as given in Figure 2.1.

Isotropy

We have already mentioned that stationarity can be seen as an invariance property under translation. It is also possible to define another condition about the invariance under rotations and reflections.

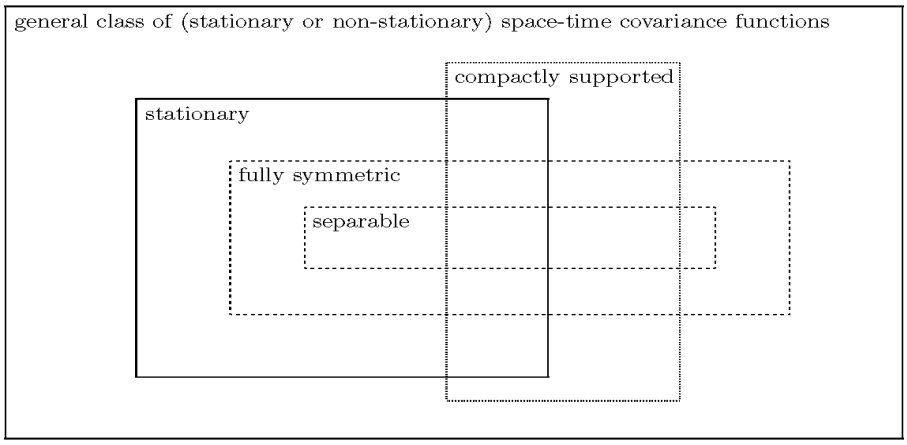


Figure 2.1: Schematic illustration of the relationships between separable, fully symmetric, stationary and compactly supported covariance functions within the general class of (stationary or non-stationary) space-time covariance functions (taken from Gneiting *et al.* 2007[54]).

Definition 2.2.11. A RF $X(\boldsymbol{s})$ on $\mathbb{R}^d$ is *strictly isotropic* if its finite dimensional joint distributions are invariant under all rigid motions. That is, for any orthogonal $d \times d$ matrix H and any $\boldsymbol{s} \in \mathbb{R}^d$,

$$P\{X(H\boldsymbol{s}_1 + \boldsymbol{s}) \leq x_1, ..., X(H\boldsymbol{s}_n + \boldsymbol{s}) \leq x_n\} = P\{X(\boldsymbol{s}_1) \leq x_1, ..., X(\boldsymbol{s}_n) \leq x_n\},$$

for all finite n, $\boldsymbol{s}_1, ..., \boldsymbol{s}_n \in \mathbb{R}^d$ and $x_1, ..., x_n \in \mathbb{R}$.

As in the case of stationarity, this condition is very hard to check, and another type of isotropy is defined.

Definition 2.2.12. A RF $X(\boldsymbol{s})$ on $\mathbb{R}^d$ is *weakly isotropic* if there exist a constant m and a function K on $[0, \infty)$ such that

1. $m(\boldsymbol{s}) = m$ and
2. $Cov[X(\boldsymbol{s}), X(\boldsymbol{s}')] = K(|\boldsymbol{s} - \boldsymbol{s}'|)$ for all $\boldsymbol{s}, \boldsymbol{s}' \in \mathbb{R}^d$, where $|\cdot|$ indicates the Euclidean distance.

This type of isotropy is the most commonly assumed and when a RF satisfies this condition is simply called isotropic RF.

Definition 2.2.13. A RF $X(\boldsymbol{s})$ is called *homogeneous* if it is intrinsically stationary and isotropic.

Continuity and differentiability

Continuity and differentiability are significant characteristics of RFs since they give information about the structure and regularity of the RF. Assume that the RF is of the second-order with null mean.

Definition 2.2.14. A RF $X(\boldsymbol{s})$ on $\mathbb{R}^d$ is *mean-square continuous* at $\boldsymbol{s}_0$ if and only if

$$\lim_{\boldsymbol{s}\to\boldsymbol{s}_0} E[X(\boldsymbol{s}) - X(\boldsymbol{s}_0)]^2 = 0.$$

For a SRF $X(\boldsymbol{s})$ with autocovariance function C, $E[X(\boldsymbol{s}), X(\boldsymbol{s}_0)]^2 = 2(C(\mathbf{0}) - C(\boldsymbol{s} - \boldsymbol{s}_0))$, hence $X(\boldsymbol{s})$ will be mean-square continuous if and only if C is continuous at the origin (Stein 1999[108]).

Definition 2.2.15. A RF $X(\boldsymbol{s})$ is *almost surely continuous* at $\boldsymbol{s}_0$ if and only if

$$X(\boldsymbol{s}) \to X(\boldsymbol{s}_0) \text{ a.s., as } \boldsymbol{s} \to \boldsymbol{s}_0.$$

If the RF is almost surely continuous for every $\boldsymbol{s}_0 \in D$, then the RF is said to have *continuous realizations*.

Definition 2.2.16. A RF $X(t)$ on $\mathbb{R}$ is *mean-square differentiable* if there exists a RF $X'(t)$ such that

$$\lim_{\delta\to 0} E[X_\delta(t) - X'(t)]^2 = 0,$$

where

$$X_\delta(t) = \frac{X(t+\delta) - X(t)}{\delta}.$$

It can be shown that the SRF $X_\delta(t)$ has an autocovariance $C_\delta(h)$ such that $\lim_{h\to 0} C_\delta(h) = -C''(h)$ provided that $C(h)$ is twice differentiable. This also shows that $-C''(h)$ is positive definite. Stein (1999)[108] proves that $X(t)$ is m-times mean-square differentiable if it is $(m-1)$-times mean-square differentiable and $X^{(m-1)}$ is mean-square differentiable. That is, there is a strong relation between the mean square differentiability of a RF and the derivative of its autocovariance function.

Spectral and convolution representation

The Fourier transform is a powerful tool to study RFs. In this context, it is usual to consider complex-valued RFs instead of real-valued ones, which are a particular type of the first. It is known that a complex-valued SRF $X(\boldsymbol{s})$ can be written as $X(\boldsymbol{s}) = U(\boldsymbol{s}) + iV(\boldsymbol{s})$, where U y V are real-valued RFs jointly weakly stationary. Following Yaglom (1987)[124] and Stein (1999)[108], a spectral representation, in the continuous case, can be obtained as follows:

$$X(\boldsymbol{s}) = \int_{\mathbb{R}^d} e^{i\omega^T \boldsymbol{s}} dY(\omega)$$

where $Y(\omega)$ is a RF with orthogonal increments, called '*random measure*' corresponding to $X(\boldsymbol{s})$. It has the following properties:

- $E[Y(A)] = 0$ for every measurable set A.
- $E[Y(A, B)] = 0$ for disjoint measurable sets A and B.
- $Y(A \cup B) = Y(A) + Y(B)$ for disjoint measurable sets A and B.

If $E[|X(A)|^2] = F(A)$ for some positive finite measure F, then this is called '*spectral measure*' of $X(\boldsymbol{s})$ and the autocovariance function associated with $X(\boldsymbol{s})$ can be expressed as

$$C(\boldsymbol{h}) = \int_{\mathbb{R}^d} e^{i\omega^T \boldsymbol{h}} dF(\omega). \tag{2.9}$$

It is easy to prove that this function is positive definite for any positive measure F:

$$\begin{aligned}\sum_{j,k=1}^{n} c_j \overline{c_k} C(\boldsymbol{h}_j, \boldsymbol{h}_k) &= \sum_{j,k=1}^{n} c_j \overline{c_k} \int_{\mathbb{R}^d} e^{i\omega^T (\boldsymbol{h}_j - \boldsymbol{h}_k)} dF(\omega) \\ &= \int_{\mathbb{R}^d} \left| \sum_{j=1}^{n} c_j e^{i\omega^T (\boldsymbol{h}_j - \boldsymbol{h}_k)} \right|^2 dF(\omega) \geq 0\end{aligned}$$

Therefore, all continuous positive positive definite complex-valued functions are of the form (2.9) with F being a positive measure of finite mass. This result is formalized in the following theorem given by Bochner (1955)[26].

Theorem 2.2.5. (Bochner's Theorem) A complex-valued function $C(\cdot)$ on $\mathbb{R}^d$ is the autocovariance function for a mean-square continuous SRF on $\mathbb{R}^d$ if and only if it can be represented as in (2.9), where F is a positive finite measure.

If the spectral measure has a density with respect to Lebesgue measure, this density is the spectral density f, defined as the Fourier transform of the autocovariance function:

$$f(\omega) = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} e^{-i\omega^T \boldsymbol{h}} C(\boldsymbol{h}) d\boldsymbol{h}.$$

Inference on the frequency domain pass through estimation of the spectral density. The following theorem, given in Gneiting *et al.* (2007)[54], specializes Bochner's theorem for fully symmetric covariances.

Theorem 2.2.6. Suppose that C is a continuous function on $\mathbb{R}^d \times \mathbb{R}$. Then C is a stationary, fully symmetric autocovariance if and only if it is of the form

$$C(\boldsymbol{h}, u) = \int \int \cos(\boldsymbol{h'\omega}) \cos(u\tau) dF(\boldsymbol{\omega}, \tau), \quad (\boldsymbol{h}, u) \in \mathbb{R}^d \times \mathbb{R},$$

where F is a finite, non-negative measure on $\mathbb{R}^d \times \mathbb{R}$.

If C is fully symmetric and the spectral density f exists, then f is fully symmetric too, that is,

$$f(\boldsymbol{\omega}, \tau) = f(\boldsymbol{\omega}, -\tau) = f(-\boldsymbol{\omega}, \tau) = f(-\boldsymbol{\omega}, -\tau),$$

for all $(\boldsymbol{\omega}, \tau) \in \mathbb{R}^d \times \mathbb{R}$.

Another useful representation is the convolution representation. To generate or represent a SRF $X(\boldsymbol{s})$ over $D \subset \mathbb{R}^d$, we can consider a white noise process $\epsilon(\boldsymbol{s})$ such that $E[\epsilon(\boldsymbol{s})] = \mu$, $Var[\epsilon(\boldsymbol{s})] = \sigma^2$ and $Cov[\epsilon(\boldsymbol{s}), \epsilon(\boldsymbol{s} + \boldsymbol{h})] = 0$. Then it is possible to describe a Gaussian stationary RF by convolving $\epsilon(\boldsymbol{s})$ with a square-integrable smoothing kernel function $k(\boldsymbol{s})$,

$$X(\boldsymbol{s}) = \int_{\mathbb{R}^d} k(\boldsymbol{u} - \boldsymbol{s}) \epsilon(\boldsymbol{u}) d\boldsymbol{u}.$$

This representation is useful since the autocovariance associated depends only on the kernel.

2.2.2 Some random field models

Here we present some RF models which will be used in the following chapters of this thesis.

Cauchy class

Stationary Gaussian fields with generalized Cauchy covariance have been introduced by Gneiting and Schlather (2004)[55]; its anisotropic counterpart has been studied by Lim and Teo (2008)[77]. This class of RFs is widely used in modeling of physical and geological phenomena since it is particularly useful to characterize the local variability and the dependence range properties separately. For this reason, we will consider RFs belonging to Cauchy class several times throughout this thesis, especially in Chapters 3 and 4.

Definition 2.2.17. A RF $X(\boldsymbol{s})$ on $\mathbb{R}^d$ is called a *Gaussian field with generalized Cauchy covariance* if it is a stationary Gaussian field with zero mean and autocovariance given by

$$C(\boldsymbol{h}) = \sigma^2(1 + |\boldsymbol{h}|^{\alpha})^{-\beta/\alpha}, \tag{2.10}$$

where $\alpha \in (0, 2]$ and $\beta > 0$.

We will refer to these RFs as the Cauchy class; the special case where $\alpha = 2$ and $\beta = 0.2$ is known as the '*Cauchy model*'. As justified below, parameter α is related with the local variability and parameter β determines the dependence range.

Fractal dimension, D, is a property that measures the local variability, while long-memory dependence is a global characteristic often referred to as Hurst effect, H. Both notions, fractal dimension and Hurst coefficient, are frequently linked in scientific literature although they are independent of each other. Long-memory dependence, or persistence, is associated with the case $H \in (1/2, 1)$ and linked to surfaces with low fractal dimensions. Rougher surfaces with higher fractal dimensions occur for antipersistent processes with $H \in (0, 1/2)$.

The following proposition is useful to compute separately the fractal dimension and Hurst coefficient for the Cauchy class (see Gneiting and Schlater 2004[55], and references therein).

Proposition 2.2.7. Let $X(\boldsymbol{s})$ be a stationary Gaussian RF on $\mathbb{R}^d$ with autocovariance function C. Then,

- If $1 - C(\boldsymbol{h}) \sim \mid \boldsymbol{h} \mid^{\alpha}$ as $\mid \boldsymbol{h} \mid \rightarrow 0$ for some $\alpha \in (0, 2]$, then the realizations of the RF have fractal dimension $D = d + 1 - \frac{\alpha}{2}$.

- If $C(\boldsymbol{h}) \sim \mid \boldsymbol{h} \mid^{-\beta}$ as $\mid \boldsymbol{h} \mid \to \infty$ for some $\beta \in (0,1)$, then the process is said to have *long-memory* with Hurst coefficient $H = 1 - \frac{\beta}{2}$.

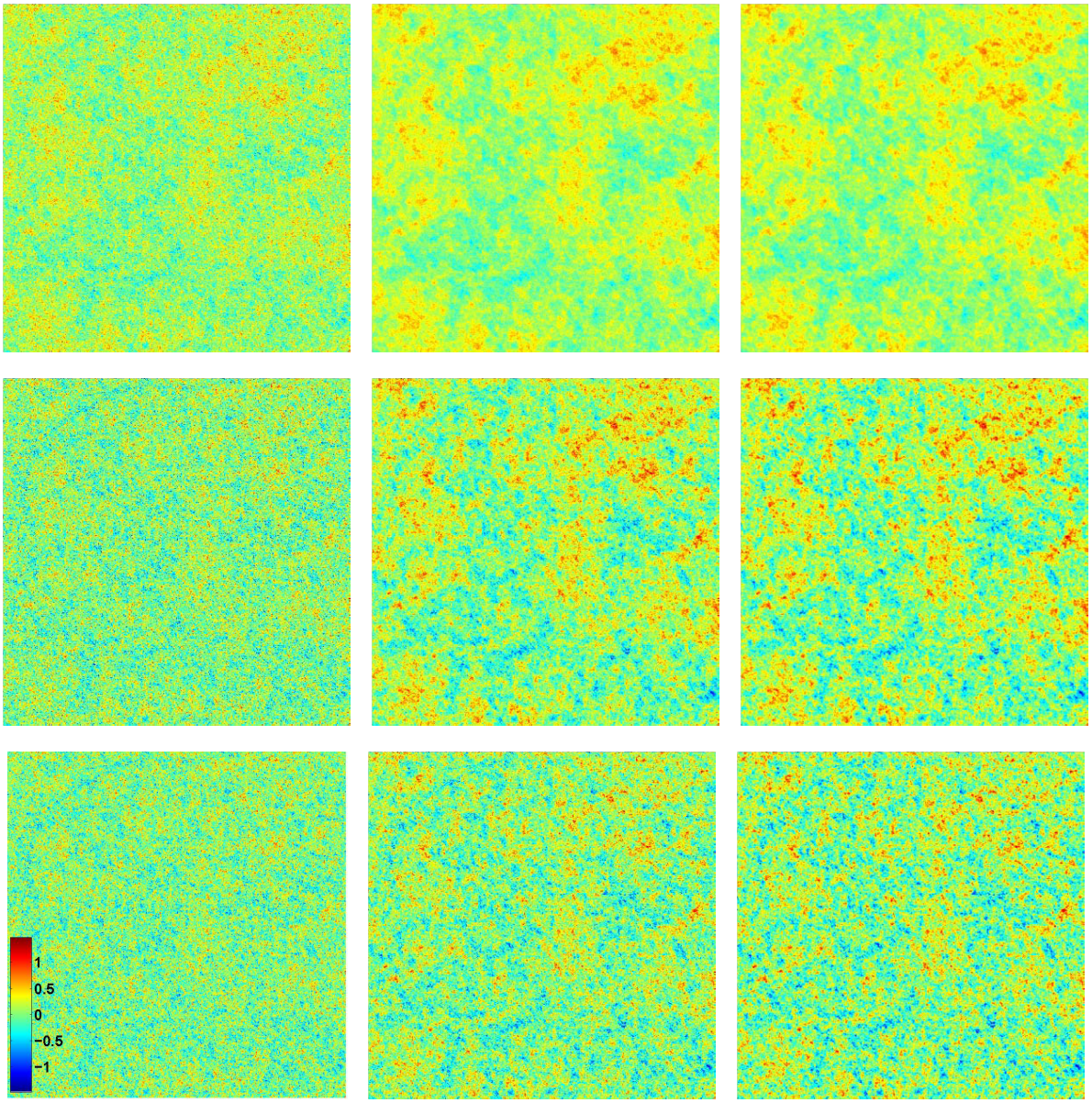


Figure 2.2: Simulated realizations of Cauchy class with $\alpha = 0.1, 1, 2$ (from left to right) and $\beta = 0.1, 0.5, 1$ (from top to bottom).

Figure 2.2 displays nine simulated realizations on the square $[0, 400]^2 \in \mathbb{R}^2$ of a Gaussian RF of Cauchy class (using the same pseudorandom numbers) based on crossed

combinations of parameters values $\alpha = 0.1,\ 1,\ 2$ (from left to right) and $\beta = 0.1,\ 0.5,\ 1$ (from top to bottom), with common $\sigma^2 = 0.1$. The images show clear differences regarding local variation and larger scale structuration. By row, the Hurst coefficient is constant but fractal dimension decreases, and this is reflected in the roughness of the images, that is, the images with lower fractality are smoother. By column, the fractal dimension is fixed and we can observe the effect of a reduction in the dependence range, that is, the process fluctuates more and more rapidly becoming more homogeneous.

Linnik model

The Linnik covariance function, for a zero-mean spatial process X, is given by

$$C(\boldsymbol{s}, \boldsymbol{s}+\boldsymbol{h}) = E[X(\boldsymbol{s})X(\boldsymbol{s}+\boldsymbol{h})] = \frac{1}{1+\|\mathbf{h}\|^{\alpha}}, \quad 0 < \alpha < d, \quad \boldsymbol{s} \in D \subset \mathbb{R}^d,$$

where α provides information on the local singularity of the model, i.e., $\alpha/2$ represents the second-order Hölder exponent of the model, since its spectral density has the following asymptotic behavior for high frequencies:

$$f_X(\boldsymbol{\lambda}) = c_1|\boldsymbol{\lambda}|^{-d-\alpha}(1+o(1)), \quad |\boldsymbol{\lambda}| \longrightarrow \infty,$$

for a certain positive constant c_1 (see, for example, Ivanov and Leonenko 2004[66]). On the other hand, the asymptotic behavior of f in the zero frequency is given by

$$f_X(\boldsymbol{\lambda}) = c(\alpha)|\boldsymbol{\lambda}|^{\alpha-d}(1-\theta(\boldsymbol{\lambda})), \quad |\boldsymbol{\lambda}| \longrightarrow 0,$$

where $\theta(\boldsymbol{\lambda}) \longrightarrow 0$ as $|\boldsymbol{\lambda}| \longrightarrow 0$. Here,

$$c(\alpha) = \frac{1}{2\Gamma(\alpha)\cos(\alpha\pi/2)}.$$

Thus, parameter α is also involved in the specification of the heavy-tailed behavior of the covariance model, since the spectral density f_X presents a singularity of order $2-\alpha$ at the zero frequency. Hence, when α increases the local singularity of the model decreases, and the covariance tails are slighter, i.e., small values of α correspond to a higher level of local singularity and long-range dependence. Covariances of this type will be used in

Function	Parameters
$\varphi(t) = \exp(-ct^{\gamma})$	$c > 0$, $0 < \gamma \leq 1$
$\varphi(t) = (2^{\nu-1}\Gamma(\nu))^{-1}(ct^{1/2})^{\nu}K_{\nu}(ct^{1/2})$	$c > 0$, $\nu > 0$
$\varphi(t) = (1 + ct^{\gamma})^{-\nu}$	$c > 0$, $0 < \gamma \leq 1$, $\nu > 0$
$\varphi(t) = 2^{\nu}(\exp(ct^{1/2} + \exp(-ct^{1/2}))^{-\nu}$	$c > 0$, $\nu > 0$

Function	Parameters
$\psi(t) = (at^{\alpha} + 1)^{\beta}$	$a > 0$, $0 < \alpha \leq 1$, $0 \leq \beta \leq 1$
$\psi(t) = \ln(at^{\alpha} + b)/\ln(b)$	$a > 0$, $b > 1$, $0 < \alpha \leq 1$
$\psi(t) = (at^{\alpha} + b)/(b(at^{\alpha} + 1))$	$a > 0$, $0 < b \leq 1$, $0 < \alpha \leq 1$

Table 2.1: Some possible choices for φ and ψ in equation (2.11) given by Gneiting (2002)[53].

Chapter 6.

Some non-separable space-time autocovariances

The first models used in spatio-temporal applications have been separable models built using (2.8). They have computational gains since the matrix associated to this kind of autocovariance can be factorized in the Kronecker product of the two matrices associated to the spatial and temporal autocovariances. However, a disadvantage of these models is their limitation to describe the space-time interaction. Gneiting (2002)[53] gives the following expression to construct a class of non-separable space-time autocovariance models:

$$C(\boldsymbol{h}, u) = \frac{\sigma^2}{(\psi(|u|^2))^{d/2}}\varphi\left(\frac{\|\boldsymbol{h}\|^2}{\psi(|u|^2)}\right), \tag{2.11}$$

where φ is a completely monotone function, ψ a positive function with a completely monotone derivative, σ^2 the variance, and d denotes the dimension of the spatial domain. This class of autocovariances is particularly important, as it constitutes one of the main contributions to stationary non-separable autocovariances for space-time data. Table 2.1 shows the proposal functions given by Gneiting (2002)[53] for the previous expression.

In Chapter 7, we consider a spatio-temporal RF, $X(\boldsymbol{s}, t)$ on $\mathbb{R}^2 \times \mathbb{R}$, with autocovariance function given by

$$C(\boldsymbol{h}, u) = 0.1(1 + u^2)^{-1.25}\left(1 + \frac{\boldsymbol{h}^2}{(1 + u^2)^{0.1}}\right)^{-0.1}. \tag{2.12}$$

The expression (2.12) is a well defined non-separable autocovariance function on $\mathbb{R}^{25} \times \mathbb{R}$ since it is obtained choosing the functions $\varphi(t) = (1 + t^2)^{-0.1}$, $\psi(t) = (t^2 + 1)^{0.05}$ and substituting in (2.11), with $\sigma^2 = 0.1$. Therefore, it is also a valid autocovariance function when we consider it restricted to $\mathbb{R}^2 \times \mathbb{R}$. Figure 2.3 shows nine simulated spatial cross-sections of this RF on the spatial domain defined by the square $D = [0, 400]^2 \subset \mathbb{R}^2$, with a sampling time interval $\Delta = 1/5$ to see more clearly the space-time interaction effect.

2.2.3 Geometrical characteristics

Threshold exceedances of RFs define a special case of excursion sets, the basis of theoretical study on the geometrical properties of RF sample paths. A relevant reference is the text by Adler and Taylor (2007)[4]; this reviews the recent technical advances in the theory of random field geometry, in relation to probabilistic aspects. A synthesis and an overview of various applications can be found in Adler (2008)[3].

Definition 2.2.18. Let $X(\boldsymbol{s})$ be a RF on $D \subset \mathbb{R}^d$ and let S be a measurable subset of D. Then, for each threshold $u \in \mathbb{R}$,

$$A_u(X, S) = \{\boldsymbol{s} \in S : X(\boldsymbol{s}) \geq u\}$$

is called the '*excursion set of* $X(\boldsymbol{s})$ *in* S *over the level* u'.

Figure 3.1, plot (a), represents a simulated realization of a Gaussian RF of Cauchy model on the square $[0, 400]^2 \in \mathbb{R}^2$; the corresponding excursion set for threshold $u = 0.3$ is shown in plot (b).

The geometry of excursion sets has been studied through numerical quantifiers including intrinsic volumes or Lipschitz-Killing curvatures and Minkowski functionals, which provide useful information about variational and structural properties. Since the differences between these quantifies are only of ordering and scaling, we shall consider the first. There are a number of ways to define them, one of which is by means of a basic result of integral geometry known as Steiner's formula. This says that for a convex subset A of $\mathbb{R}^d$ the volume in $\mathbb{R}^d$ of a tube of radius ρ around A has a finite expansion in powers of ρ.

Definition 2.2.19. Consider a convex subset $A \subset \mathbb{R}^d$ and $\rho > 0$. The *tube of radius* ρ

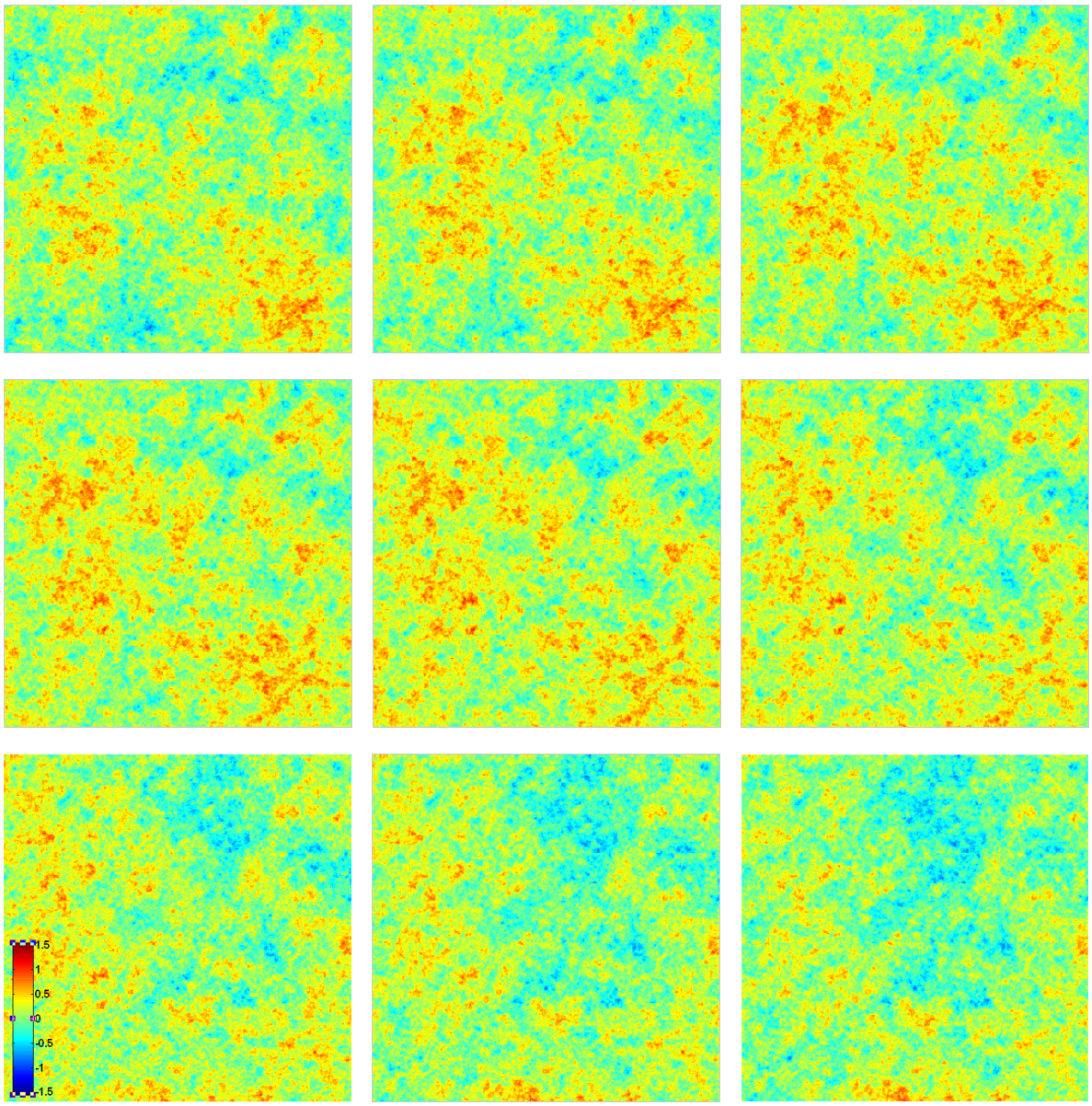
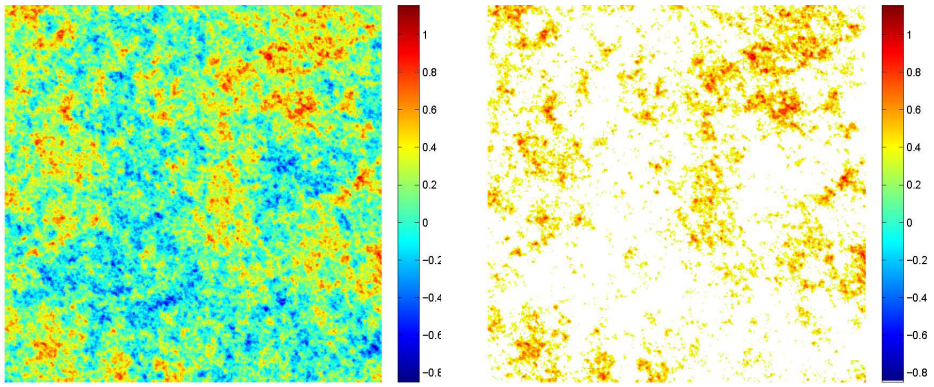


Figure 2.3: Spatial cross-sections of a spatio-temporal random field with covariance function given by 2.12 (see from left to right and from top to bottom).



(a) (b)

Figure 2.4: (a) Simulated spatial realization of Cauchy random field; (b) excursion set for $u = 0.3$.

or *ρ-tube around A* is defined by

$$\text{Tube}(A, \rho) = \left\{ x \in \mathbb{R}^d : d(x, A) := \inf_{y \in A} |x - y| \leq \rho \right\}.$$

Proposition 2.2.8. Denoting the Lebesgue measure in $\mathbb{R}^d$ by λ_d, Steiner's formula states that

$$\lambda_d(\text{Tube}(A, \rho)) = \sum_{j=1}^{d} \omega_{d-j} \rho^{d-j} \mathcal{L}_j(A),$$

where

$$\omega_j = \lambda_j(B(0, 1)) = \frac{\pi^{j/2}}{\Gamma(\frac{j}{2} + 1)}$$

is the volume of the unit ball in $\mathbb{R}^j$.

The quantifies $\mathcal{L}_j(A)$, known as '*intrinsic volumes*' or '*Lipschitz-Killing curvatures*', satisfy the scaling property $\mathcal{L}_j(\delta A) = \delta^j \mathcal{L}_j(A)$, for $\delta > 0$, and hence can be understood, in a certain sense, as related to the j-dimensional measure of A, for each j. If the ambient space is $\mathbb{R}^2$, then $\mathcal{L}_2$ measures area, $\mathcal{L}_1$ measures boundary length, while $\mathcal{L}_0$ gives the Euler characteristic. In $\mathbb{R}^3$, $\mathcal{L}_3$ measures volume, $\mathcal{L}_2$ measures surface area, $\mathcal{L}_1$ is a measure of cross-sectional diameter and $\mathcal{L}_0$ is again the Euler characteristic. In higher dimensions, $\mathcal{L}_d$ and $\mathcal{L}_{d-1}$ measure hypervolume and hypersurface area, while

$\mathcal{L}_0(A)$ is always the Euler characteristic which measures structural connectivity. In particular, for $d = 2$, $\mathcal{L}_0(A)$ is the number of connected components minus the number of 'holes' in A; and for $d = 3$, it is the number of connected components minus the number of 'handles' plus the number of 'holes'.

The hypervolume ($\lambda_d(A) := \mathcal{L}_d$) and the Euler characteristic ($\varphi(A) := \mathcal{L}_0(A)$) have a special interest for us since through them we can compute or approximate some key probabilities in the risk analysis context, and thus, they can be used to define certain risk indicators (see Chapters 3 and 7). Specifically, the hypervolume $\lambda_d(A_u(X, S))$ directly reflects the total size of the regions in a critical state within S. Hence, the expected proportion $(\lambda_d(S))^{-1}E[\lambda_d(A_u(X, S))]$ gives the (average) probability that a generic point within S enters into a critical state, which also gives a global risk indicator for S. On the other hand, for a large threshold u, the expected value of the Euler characteristic for the excursion set $A_u(X, S)$ of a smooth RF provides a good approximation of the level crossing probability $P\{\sup_{\boldsymbol{s}\in S} X(\boldsymbol{s}) \geq u\}$. Therefore, there is a strong relationship between the probabilistic and geometric characteristics of RFs.

The idea for approximating the level crossing probability by the expected value of the Euler characteristic is based on the structure of a excursion set. Specifically,

$$\begin{aligned}\sup_{\boldsymbol{s}\in D} X(\boldsymbol{s}) \geq u &\Leftrightarrow A_u(X, D) \neq \emptyset \\ &\Leftrightarrow \text{the number of connected components of } A_u \geq 1,\end{aligned}$$

and when u is sufficiently large, the structure of the excursion set A_u becomes simple, containing only simply connected components, likely to be a "nice ball-like" and, so will have Euler characteristic equal to one. Therefore, the excursion set A_u over a high level u, if non-empty, is very likely to have Euler characteristic equal to 1. In particular,

$$E[\varphi(A)] = E[\varphi(A)\mathbf{1}(A_u \neq \emptyset)] \approx P\{A_u \neq \emptyset\} = P\{\sup_{\boldsymbol{s}\in D} X(\boldsymbol{s}) > u\}$$

The previous heuristic explanation and the rigorous result for constant variance Gaussian fields, which we show below, is given in Taylor *et al.* (2005)[113].

Proposition 2.2.9. Let D be a 'nice' manifold, and let $\{X(\boldsymbol{s}); \boldsymbol{s} \in D\}$ be a smooth zero mean constant variance σ^2 Gaussian random field on D. Then, there is a number

$a > 0$ such that, for $u > 0$ large enough,

$$\left| P\left\{ \sup_{\boldsymbol{s} \in D} X(\boldsymbol{s}) > u \right\} - E[\varphi(A_u)] \right| \leq e^{-\frac{(1+a)u^2}{2\sigma^2}}.$$

2.3 Spatial point processes

An approach based on marked point processes to analyze structural properties of threshold exceedance sets is considered in Chapter 4. Here we give a characterization of such processes and introduce some fundamental first- and second-order characteristics, as well as tools to analyze possible anisotropy effects.

2.3.1 Characterization of spatial point processes

Informally speaking, a point pattern is defined as an unordered finite set of points or events $\{\boldsymbol{s}_i\}$ observed inside a bounded region W, called the observation window. More formally, a '*point process*' is a stochastic model governing the locations of events $\{\boldsymbol{s}_i\}$. Although W could be any locally compact Hausdorff space whose topology has a countable basis (Cressie 1993[38]), we here consider W to be a subset of $\mathbb{R}^d$, since we work with spatial point processes or, in the case in which the events contain associated measurements or marks $z \in F$, W will be the product space $D \times F$ with $D \subset \mathbb{R}^d$, being then these refereed as '*marked point processes*'. Another extension is to consider spatio-temporal point processes, which can be viewed as a family of spatial point processes indexed by time or directly as a dynamic spatial point process with W defined as a subset of $\mathbb{R}^d \times \mathbb{R}$, possibly with associated marks in F.

There are two ways of characterizing a point process: through the spatial locations of events $\{\boldsymbol{s}_i\}$, or defining a counting measure ϕ on W. Both characterizations are given in Cressie (1993)[38]. Here we follow the characterization through a counting measure for a marked point process given also in Mateu (2000)[88]; the unmarked case is similar, conveniently adapting the component elements related to the marks. Let χ be the Borel sets of D and let $\mathcal{B}$ be the Borel sets of F. Then, for a counting measure ϕ, $\phi(A \times B)$ is defined as the number of events in A whose marks belong to B, where $A \in \chi$ and $B \in \mathcal{B}$. Let (Ω, Λ, P) be a probability space and let Φ be a collection of locally finite counting measures on $D \times F$. On Φ define $\mathcal{N}$ as the smallest σ- algebra generated by

sets of the form $\{\phi \in \Phi : \phi(A \times B) = n\}$, $A \in \chi$, $B \in \mathcal{B}$ and all $n \in \{0, 1, 2, ...\}$. Then, a marked point process N is a measurable mapping of (Ω, Λ, P) into $(\Phi, \mathcal{N})$. A marked point process defined over (Ω, Λ, P) induces a probability measure $\Pi_N(Y) = P(N \in Y)$ on $(\Phi, \mathcal{N})$ for all $Y \in \mathcal{N}$. The marked point process will be denoted by $N = [\boldsymbol{s}, Z(\boldsymbol{s})] = \{[\boldsymbol{s}_1, Z_1], [\boldsymbol{s}_2, Z_2], ...[\boldsymbol{s}_n, Z_n]\}$ and the corresponding unmarked process by $N = [\boldsymbol{s}] = \{[\boldsymbol{s}_1], [\boldsymbol{s}_2], ...[\boldsymbol{s}_n]\}$.

Two important concepts in the context of point processes are stationarity and isotropy. A point process is said to be '*stationary*' if it is invariant under translations, that is, $\Pi_N(A) = \Pi_{N_h}(A)$, for all $A \in \chi$ and all $\boldsymbol{h} \in \mathbb{R}^d$, with $N_h = [\boldsymbol{s} + \boldsymbol{h}, Z(\boldsymbol{s} + \boldsymbol{h})]$. A stationary point process is '*isotropic*' if it is invariant under rotations. If the process is stationary and isotropic then it is '*rigid-motion invariant*'.

In the following subsections, we refer to spatial point processes on the plane, that is, with $d = 2$. The contents can be conveniently extended for higher dimensions.

2.3.2 Complete spatial randomness

Definition 2.3.1. A spatial point process is said to be *complete spatial random* if it satisfies:

1. the number of events in any planar region A follows a Poisson distribution with mean $\lambda v(A)$, where $v(\cdot)$ represents the Lebesgue measure, and

2. given n events $\boldsymbol{s}_i$ in a region A, the $\boldsymbol{s}_i$ are an independent random sample from the uniform distribution on A.

Condition 1 means that the expected number of points per unit area is constant, that is, it does not vary over the plane. Condition 2 implies that there is no interaction amongst the events. Cressie (1993)[38] states that 'complete spatial randomness' is the 'white noise' of spatial point processes since it characterizes the absence of structure in the data. For several reasons, according to Diggle (1983)[41], an analysis of point processes should begin with a test of complete spatial randomness. A test to look for evidence against complete spatial randomness can be carried out by nonparametric techniques such as '*quadrat counting*'. In this technique, the window is divided into subregions ('quadrats') of equal area. Then, the numbers of points falling into each quadrat is counted. Obviously, there are several types of alternatives to complete spatial randomness; for example, there can

exist a mechanism inherent to the process favoring or inhibiting the presence of other events near an existing event. These types of processes are known as cluster Poisson processes and hard-core processes, respectively.

2.3.3 First- and second-order characteristics

Definition 2.3.2. Let $N(A)$ be the number of points of a spatial point process N in the Borel set $A \subset D$. The *first-moment measure* or *mean measure* on (D, χ) is given by

$$\mu(A) = E[N(A)] = \int_{\Phi} \phi(A)\Pi(d\phi).$$

Definition 2.3.3. The *second-moment measure* of N is defined by

$$\mu^{(2)}(A_1 \times A_2) = E[N(A_1)N(A_2)] = \int_{\Phi} \phi(A_1)\phi(A_2)\Pi(d\phi),$$

where $A_1, A_2 \in \chi$.

Note that $\mu^{(2)}$ is a measure on (D^2, χ^2), with χ^2 being the smallest σ-algebra containing the product sets $A_1 \times A_2$.

Definition 2.3.4. The *covariance measure* of a spatial point process N is given by

$$C(A_1 \times A_2) = \mu^{(2)}(A_1 \times A_2) - \mu(A_1)\mu(A_2),$$

where $A_1, A_2 \in \chi$.

Now, let us consider the behavior of the moment measures as $A \downarrow \{\boldsymbol{s}\}$, where $\boldsymbol{s} \in D$. Let $d\boldsymbol{s}$ and $d\boldsymbol{t}$ be small regions located at $\boldsymbol{s}$ and $\boldsymbol{t}$, respectively.

Definition 2.3.5. The *first-order intensity function* is defined as

$$\lambda(\boldsymbol{s}) = \lim_{v(d\boldsymbol{s})\to 0} \frac{\mu_(d\boldsymbol{s})}{v(d\boldsymbol{s})},$$

provided that the limit exists.

Definition 2.3.6. A spatial point process is said to be *homogeneous* if its intensity function is constant, otherwise it is said to be inhomogeneous.

A homogeneous Poisson process is a complete spatial random point process. The non-homogeneous Poisson process provides a more general description of randomly occurring points in a subset of the space.

Definition 2.3.7. A spatial point process N is said to be a *non-homogeneous Poisson process* with intensity $\lambda(\cdot)$ if it satisfies:

1. The number of events in any planar region A follows a Poisson distribution with mean $\Lambda(A)$, where $\Lambda(A) = \int_A \lambda(x)dx$.

2. Given n events $\boldsymbol{s}_i$ in a region A, the $\boldsymbol{s}_i$ are an independent random sample from the uniform distribution on A.

Definition 2.3.8. The *second-order intensity* is defined as

$$\lambda_2(\boldsymbol{s}, \boldsymbol{t}) = \lim_{v(d\boldsymbol{s})\to 0 v(d\boldsymbol{t})\to 0} \frac{\mu^{(2)}(d\boldsymbol{s} \times d\boldsymbol{t})}{v(d\boldsymbol{s})v(d\boldsymbol{t})},$$

provided the limit exists (Ripley 1977[98]).

Definition 2.3.9. The *covariance density* of a spatial point process is given by

$$\beta(\boldsymbol{s}, \boldsymbol{t}) = \lambda_2(\boldsymbol{s}, \boldsymbol{t}) - \lambda(\boldsymbol{s})\lambda(\boldsymbol{t}),$$

where $\boldsymbol{s}, \boldsymbol{t} \in D$.

Ordinary moment measures of point processes are often unsatisfactory because the A_is in the product space are not necessarily disjoint. This means that the events of the process can occur in two or more A_is causing a possible redundancy of information and singularity of the moment measures. Factorial moment measures of simple point processes (that is, point processes in which no more than one event can occur at each location) do not suffer from this problem, since they exclude the cases where two or more points belong to the same set.

Definition 2.3.10. The *second factorial measure* of N is defined by

$$\begin{aligned}\alpha^{(2)}(A_1 \times A_2) &= \int_{\Phi} \sum_{\boldsymbol{s}_1 \neq \boldsymbol{s}_2 \in \phi} I(\boldsymbol{s}_1 \in A_1) I(\boldsymbol{s}_2 \in A_2) \Pi(d\phi) \\ &= E\left[\sum_{\boldsymbol{s}_1 \neq \boldsymbol{s}_2 \in \phi} I(\boldsymbol{s}_1 \in A_1) I(\boldsymbol{s}_2 \in A_2) \right],\end{aligned}$$

where $A_1, A_2 \in \chi$, with $I(\cdot)$ being the indicator function.

The term 'factorial moment measure' comes from the fact that for an arbitrary k, $\alpha^{(k)} = E[N(A)(N(A)-1)...(N(A)-k+1)]$.

Definition 2.3.11. The *second-order product density*, $\ell^{(2)}$, is (assuming continuity) the density function of $\alpha^{(2)}$, that is,

$$\alpha^{(2)}(A_1 \times A_2) = \int_{A_1} \int_{A_2} \ell^{(2)}(\boldsymbol{s}, \boldsymbol{t}) d\boldsymbol{s} d\boldsymbol{t}.$$

For infinitesimal $d\boldsymbol{s}$ and $d\boldsymbol{t}$, $\ell^{(2)}(\boldsymbol{s}, \boldsymbol{t}) d\boldsymbol{s} d\boldsymbol{t}$ can be interpreted as the probability that there is one point of the point process in each of two specified infinitesimal sets with areas $d\boldsymbol{s}$ and $d\boldsymbol{t}$, respectively. If the point process is homogeneous, then the product density depends only on the (undirected) difference $\boldsymbol{h} = \boldsymbol{t} - \boldsymbol{s}$. If we use polar coordinates, $\boldsymbol{h}$ is given by $\boldsymbol{h} = (r, \varphi)$, with $r \geq 0$ and $0 \leq \varphi \leq \pi$, and then, in the homogeneous case, $\ell^{(2)}(\boldsymbol{s}, \boldsymbol{t}) = \ell^{(2)}(\boldsymbol{h}) = \ell^{(2)}(r, \varphi)$. Moreover, if the point process is isotropic, the product density only depends on the distance r between the points $\boldsymbol{s}$ and $\boldsymbol{t}$. This argument will be applied to other characteristics below.

Definition 2.3.12. The *pair correlation function* is defined by the second-order product density normalized by the square of the first-order intensity, that is,

$$g(\boldsymbol{s}, \boldsymbol{t}) = \frac{\ell^{(2)}(\boldsymbol{s}, \boldsymbol{t})}{\lambda^2}.$$

For a complete spatial random point process the pair correlation function takes the value 1. In the isotropic case, values of $g(r)$ larger than 1 show that point pairs of distance r are relatively more frequent compared to those in a complete spatial random process, which is typical of an aggregation process, and conversely, values of $g(r)$ smaller than

1 indicate that the corresponding distances are rare and this is typical of an inhibition process. In the anisotropic case, it is sufficient to consider only values of φ between 0 and π, as $g(r,\varphi) = g(r,\varphi+\pi)$ for $0 \leq \varphi \leq \pi$. The form of $g(r,\varphi)$ for different φ reflects the form of the anisotropy, and those φ values for which the differences are greater should be found.

An alternative characterization of the second-order properties of a stationary and isotropic point process is through the reduced second-order moment measure (Ripley 1977[98]; Diggle 1983[41]).

Definition 2.3.13. The *reduced second-order moment measure* or *Ripley's K-function* of a stationary and isotropic point process N is defined as

$$K(r) = \frac{E[\text{number of further events within distance } r \text{ of an arbitrary event}]}{\lambda}. \quad (2.13)$$

Stoyan and Stoyan (1994)[111] show the following link between $K(r)$ and $g(r)$:

$$g(r) = \frac{1}{2\pi r}\frac{d}{dr}K(r), \quad r \geq 0.$$

Thus, for a complete spatial random process, $K(r) = \pi r^2$.

Definition 2.3.14. The *L-function* is the transformation of the K-function given by

$$L(r) = \sqrt{\frac{K(r)}{\pi}}. \quad (2.14)$$

This function transforms the Poisson K-function to the straight line $L(r) = r$, making visual assessment of the graph much easier. In addition, the square root transformation stabilizes the variance of the estimator. In spatial statistics, usually the pair correlation function is used for exploratory analysis of data, whilst the K-function and L-function are used for hypothesis testing.

A generalization of the K-function for inhomogeneous second-order stationary processes has been introduced by Baddeley *et al.* (2000)[23].

Definition 2.3.15. Let N be a second-order stationary spatial point process. The *in-*

homogeneous K-function of N is given by

$$K_{inhom}(r) = \frac{1}{|A|} E\left[\sum_{\boldsymbol{s}_i \in N \bigcap A} \sum_{\boldsymbol{s}_j \in A \setminus \boldsymbol{s}_i} \frac{I(||\boldsymbol{s}_i - \boldsymbol{s}_j|| \leq r)}{\lambda(\boldsymbol{s}_i)\lambda(\boldsymbol{s}_j)}\right], \quad r \geq 0, \tag{2.15}$$

for any $A \in \chi$ with $|A| > 0$, where $|A|$ denotes the area (Lebesgue measure) of A, and we take $t/0 = 0$ for $t \geq 0$.

This expression is a generalization of the usual K-function, since a stationary (or at least first-order stationary) point process has $\lambda(\boldsymbol{s}) \equiv \lambda$ and in this case the right-hand side of (2.15) reduces to the definition of $K(r)$ in Ripley (1977)[98]. Similarly to the stationary case, here $L = (K/\pi)^{1/2}$, and for an inhomogeneous Poisson processes, $L(r) = r$.

In addition to all previous characteristics, there are some other specific ones which take into account the marks of the point processes in order to describe the correlations between them. Let $f(z_i, z_j)$ be a non-negative test function depending on the marks $z_i = Z(\boldsymbol{s}_i)$ and $z_j = Z(\boldsymbol{s}_j)$ of two points $\boldsymbol{s}_i$ and $\boldsymbol{s}_j$. The test function has to be chosen according to the problem and the type of mark. We will use the next functions:

- $f(z_i, z_j) = z_i z_j$, when the marks are the sizes of objects represented by the marked points;
- $f(z_i, z_j) = d(z_i, z_j) = \min\{|z_i - z_j|, \pi - |z_i - z_j|\}$, when the marks are orientations of such objects between 0° and 180°.

Definition 2.3.16. The *second-factorial moment measure* with respect to the mark-sum measure is given by

$$\alpha_f^{(2)}(A_1 \times A_2) = E\left[\sum_{\boldsymbol{s}_1 \neq \boldsymbol{s}_2 \in \phi} f(Z(\boldsymbol{s}_1), Z(\boldsymbol{s}_2)) I(\boldsymbol{s}_1 \in A_1) I(\boldsymbol{s}_2 \in A_2)\right].$$

Definition 2.3.17. The *f-second-order product density*, $\ell_f^{(2)}$, is (assuming continuity) the density function of $\alpha_f^{(2)}$, that is,

$$\alpha_f^{(2)}(A_1 \times A_2) = \int_{A_1} \int_{A_2} \ell_f^{(2)}(\boldsymbol{s}, \boldsymbol{t}) d\boldsymbol{s} d\boldsymbol{t}.$$

Definition 2.3.18. The *mark correlation function* (Stoyan and Stoyan 1994[111]) is given by the quotient

$$\kappa_f(\boldsymbol{s},\boldsymbol{t}) = \frac{\ell_f^{(2)}(\boldsymbol{s},\boldsymbol{t})}{\ell^{(2)}(\boldsymbol{s},\boldsymbol{t})}. \tag{2.16}$$

The latter can be interpreted as the mean of $f(z_i, z_j)$ given that there is a pair of points of N at both locations $\boldsymbol{s}_i$ and $\boldsymbol{s}_j$ with marks z_i and z_j, respectively. If $f(z_i, z_j) = d(z_i, z_j)$, we write $\kappa_d(\boldsymbol{s},\boldsymbol{t})$ instead of $\kappa_f(\boldsymbol{s},\boldsymbol{t})$ and it is called orientation correlation function since it can be interpreted as the mean of the difference of directions corresponding to the orientations marks of two points.

When the process is homogeneous and isotropic, then $\kappa_f(\boldsymbol{s},\boldsymbol{t}) = \kappa_f(\boldsymbol{h})$, and using polar coordinates $\kappa_f(r,\varphi) = \kappa_f(r)$, with r being the distance between $\boldsymbol{s}$ and $\boldsymbol{t}$. In order to get still more of the character of a correlation function, the normalized version

$$\widehat{\hbar}_f(r) = \frac{\kappa_f(r)}{\kappa_f(\infty)}, \quad r \geq 0, \tag{2.17}$$

is commonly used, with $\kappa_f(\infty)$ being calculated by means of the mark distribution, $\mathcal{M}$, as follows. Assuming that for large distances the marks are independent,

$$\kappa_f(\infty) = \int\int f(z_1, z_2)\mathcal{M}(dz_1)\mathcal{M}(dz_2).$$

Large values of $\widehat{\hbar}_f(r)$ compared to the value 1 suggest positive correlation between the marks at distance r. Conversely, values of $\widehat{\hbar}_f(r)$ smaller than 1 suggest negative correlation at such a distance. In the case where the mark variable is uncorrelated, $\widehat{\hbar}_f(r) = 1$.

2.3.4 Anisotropy characteristics

Many homogeneous points fields are anisotropic. There is a wide variety of forms of anisotropy; for instance, they include:

- anisotropic arrangement of the points;

- anisotropic behavior of marks if they describe orientations;

- combination of anisotropic point distribution and anisotropic mark behavior.

Two of the characteristics introduced in the previous section, the pair correlation function and the mark correlation function in their anisotropic versions, can be used as a tool for analyzing anisotropy, depending of the point process being unmarked or marked, respectively. Here, we introduce some further characteristics to describe the anisotropic behavior of a point process.

Let N be a marked point process in which each event has a pair of marks: one of size and another one of orientation; then every event is described by the triplet

$$[\boldsymbol{s} = \text{location}, Z(\boldsymbol{s}) = \{l = \text{size mark}, \alpha = \text{orientation mark}\}] = [\boldsymbol{s}, l, \alpha].$$

The sequence of triplets $[\boldsymbol{s}_i, l_i, \alpha_i]$, $i = 1, 2, ..., n$ is interpreted as a sample of a homogeneous marked point process. An associated segment process can be considered as a point process with realization defined by segments with the centers $\boldsymbol{s}_i$, lengths l_i and directions θ_i, for $i = 1, ..., n$.

Definition 2.3.19. The *rose of directions* of a marked point process (Stoyan *et al.* 1987[110]) is the length-weighted orientation distribution of the segments of unit length centred at the points and having directions given by the angle marks. The corresponding distribution function gives the fraction of the length elements with an orientation between 0 and β, $0 \leq \beta \leq \pi$ and is given by

$$R(\beta) = \frac{1}{\bar{l}} \int \int I_{[0,\beta]}(\theta) l \mathcal{M}(dl, d\theta), \tag{2.18}$$

where $\bar{l}$ is the mean length and $\mathcal{M}$ is the joint mark distribution.

2.4 Wavelets

Wavelet-based techniques, which allow to analyze and describe location-scale related features of deterministic or random signals, are used in the developments presented in Chapters 5 and 6.

2.4.1 Introduction

The first mention of wavelets can be found in an appendix of the thesis of Alfred Haar, in the early twentieth century. Later, he discovered an orthonormal system of functions such that, for any continuous functions $f(x)$, the series

$$\sum_{j=0}^{\infty}\sum_{k=0}^{2^j-1} a_{2^{j+k}} w(2^j x - k), \quad \text{for } 0 \leq x < 1,$$

converges to $f(x)$ uniformly over the interval $0 \leq x < 1$. The main drawback of Haar basis functions is that they are inefficient in modeling smooth signals because of theirs discontinuous nature. In spite of this fact, Paul Lévy showed that the Haar basis function was more appropriate than Fourier basis functions to study small complicated details when he analyzed Brownian motion.

Gabor (1946)[49] introduced the first time-frequency wavelet. His idea consisted in splitting a wavelet into segments, in such a way that each of which was well defined in frequency and time, to analyze the individual segments of the wavelets. Shortly later, other approaches and transformations appeared. The first synthesis of these theories are attributed to Grossmann and Morlet(1984)[58] and Morlet (1982)[92], introducing the concept of using wavelets to analyze non-stationary signals in the eighties. A wavelet was defined, in the Fourier domain, as an $\mathbb{L}^2(\mathbb{R})$ function for which the Fourier transform[2] $\Psi(\omega)$ satisfies

$$\int_0^{\infty} |\Psi(t\omega)|^2 \frac{dt}{t} = 1, \text{ for almost all } \omega.$$

Vidakovic (1998a)[118] affirms that the previous definition is quite broad and over time the meaning of the term wavelet became narrower; currently, the term wavelet is usually associated with a function $\psi \in \mathbb{L}^2(\mathbb{R})$ such that the translations and dyadic dilations of ψ,

$$\psi_{jk}(x) = 2^{j/2}\psi(2^j x - k), \quad j, k \in \mathbb{Z},$$

constitute an orthonormal basis of $\mathbb{L}^2(\mathbb{R})$.

[2]The Fourier transformation of a function $f \in \mathbb{L}_1(\mathbb{R})$ is defined by

$$\hat{f}(\omega) = \langle f(x), e^{i\omega x}\rangle = \int_{\mathbb{R}} f(x)\overline{e^{i\omega x}}dx = \int_{\mathbb{R}} f(x)e^{-i\omega x}dx.$$

2.4.2 Continuous Wavelet Transform

Definition 2.4.1. *Wavelets* are functions generated from translation and re-scaling of a single function called *mother wavelet*, $\psi \in \mathbb{L}^2(\mathbb{R})$, defined as

$$\psi_{a,b}(x) = \frac{1}{\sqrt{|a|}}\psi\left(\frac{x-b}{a}\right),$$

where a and b are the scale and translation factors.

Definition 2.4.2. The *continuous wavelet transform* ($\mathcal{CWT}$) at the scale a and time localization b of any function $f(x) \in \mathbb{L}^2(\mathbb{R})$ is defined by correlating f with a wavelet,

$$\mathcal{CWT}_f(a,b) = \langle f, \psi_{ab}\rangle = \int_{\mathbb{R}} f(x)\psi_{ab}(x)dx. \tag{2.19}$$

By varying a and b, the original wavelet is compressed or dilated and translated, respectively, and, in this way, low and high frequencies can be selectively analyzed.

To guarantee that a wavelet transform is invertible, it is necessary that the mother wavelet satisfies the *admissibility condition*

$$C_\psi = \int_{\mathbb{R}} \frac{|\Psi(\omega)|^2}{|\omega|}d\omega < \infty,$$

where $\Psi(\omega)$ is the Fourier transform of $\psi(x)$. Then, the *reconstruction formula* is

$$f(x) = \frac{1}{C_\psi}\int_{\mathbb{R}^2}\int_{\mathbb{R}} \mathcal{CWT}_f(a,b)\psi_{ab}(x)\frac{dadb}{a^2}. \tag{2.20}$$

Mallat (1998)[85] shows that the $\mathcal{CWT}$ can be a valuable tool for detecting irregularities and analyzing fractals. Suppose that ψ is real. Since ψ has a zero average, a wavelet coefficient $\mathcal{CWT}_f(a,b)$ measures the variation of f in a neighborhood of b. Thus, following across scales the local maxima of the wavelet transform, the singularities are detected. If at location b the value of $\mathcal{CWT}_f(a,b)$ increases for small scales, a short range of singularity can be expected, such as noise. Large values at coarse scales indicate a long range singularity, typically an important signal feature.

2.4.3 Discrete Wavelet Transform and multiresolution analysis

The $\mathcal{CWT}$ is a highly redundant representation, whose redundancy is characterized by a reproducing kernel equation. Inserting the reconstruction formula (2.20) into the definition of the wavelet transform (2.19) yields

$$\mathcal{CWT}_f(a_0, b_0) = \int_{\mathbb{R}} \left(\frac{1}{C_\psi} \int_{\mathbb{R}^2} \int_{\mathbb{R}} \mathcal{CWT}_f(a, b) \psi_{ab}(x) da \frac{db}{b^2} \right) \overline{\psi_{a_0 b_0}}(x) dx.$$

Interchanging these integrals gives

$$\mathcal{CWT}_f(a_0, b_0) = \frac{1}{C_\psi} \int K(a, a_0, b, b_0) \mathcal{CWT}_f(a, b) da \frac{db}{b^2},$$

with

$$K(a, a_0, b, b_0) = \langle \psi_{ab}, \psi_{a_0 b_0} \rangle.$$

The '*reproducing kernel*' $K(a, a_0, b, b_0)$ measures the correlation of two wavelets ψ_{ab} and $\psi_{a_0 b_0}$. Given the redundancy of $\mathcal{CWT}$ it is advisable to select discrete values of a and b which minimize the transformation and allow that it continues being invertible.

Definition 2.4.3. A *discrete wavelet transform* ($\mathcal{DWT}$) can be obtained by choosing a and b according to a rule known as the critical sampling where $a = 2^{-j}$ and $b = k2^{-j}$, for $j, k \in \mathbb{Z}$.

Other discretization schemes are possible; see Vidakovic 1998a[118].

Under mild conditions on the wavelet function ψ, such a sampling strategy produces an orthogonal basis of $\mathbb{L}^2(\mathbb{R})$ called '*wavelet basis*',

$$\{\psi_{jk} = 2^{j/2} \psi(2^j x - k), \quad j, k \in \mathbb{Z}\},$$

where j and k represent the set of discrete dilations and translations, respectively. Orthogonal wavelets dilated by 2^j carry signal variations at the resolution 2^{-j}. For $f \in \mathbb{L}^2(\mathbb{R})$, the partial sum of wavelet coefficient $\sum_{k \in \mathbb{Z}} \langle f, \psi_{jk} \rangle \psi_{jk}$ can be interpreted as the difference between two approximations of f at the resolutions 2^{-j+1} and 2^{-j}. The approximation of a function at resolution 2^{-j} is defined as an orthogonal projection on a space $V_j \subset \mathbb{L}^2(\mathbb{R})$ which regroups all possible approximations at the resolution 2^{-j}. Mallat (1989)[84] and Meyer (1992)[87] introduced the definition of multiresolution spaces.

Definition 2.4.4. A *multiresolution analysis* (MRA) is a sequence of closed subspaces $\{V_j, j \in \mathbb{Z}\}$ in $\mathbb{L}^2(\mathbb{R})$ satisfying the following properties:

i) $\forall j \in \mathbb{Z}$, $V_j \subset V_{j+1}$;

ii) $\overline{\cup_j V_j} = \mathbb{L}^2(\mathbb{R})$;

iii) $\bigcap_j V_j = \{0\}$;

iv) $\forall j \in \mathbb{Z}$, $f(2^j x) \in V_j \Leftrightarrow f(x) \in V_0$;

v) There exists a scaling function $\phi \in V_0$, called *father wavelet*, whose integer-translates span the space V_0,

$$V_0 = \left\{ f \in \mathbb{L}^2(\mathbb{R}) | f(x) = \sum_{k \in \mathbb{Z}} c_k \phi(x-k) \right\}, \tag{2.21}$$

for which the set $\{\phi(\cdot - k),\, k \in \mathbb{Z}\}$ is an orthonormal basis of V_0.

Condition (i) means that the signal to be analyzed at the given resolution contains all the information of the signal at coarser resolutions. Condition (ii) means that any signal can be approximated with arbitrary precision. Condition (iii) indicates that the function 0 is the only object common to all spaces V_j. Condition (iv) shows that there is only one space, e.g. V_0, all the other spaces being scaled versions of the prototype V_0. The orthogonality requirement of Condition (v) can be relaxed, it is sufficient to assume that the system of functions $\{\phi(\cdot - k),\, k \in \mathbb{Z}\}$ constitutes a Riesz basis[3].

Since $V_j \subset V_{j+1}$, the space V_{j+1} is generated by the basis of the coarser scale V_j plus other basis functions which generate a space W_j of detail functions. Therefore, all

[3]A Riesz basis of a Hilbert space H is a basis satisfying:

i) there exist constants $C' > C > 0$ such that, for every sequence of scalars $\{c_m : m \in \mathbb{N}\}$,

$$C \sum_m |c_m|^2 \leq \left\| \sum_m c_m \phi_m \right\|_H^2 \leq C' \sum_m |m|^2, \text{ and} \tag{2.22}$$

ii) the vector space of finite sums $\sum_m c_m \phi_m$ (on which (2.22) is tested) is dense in H.

functions in V_{j+1} can be represented in terms of the union of the basis of V_j and the basis of W_j:

$$V_{j+1} = V_j \oplus W_j.$$

This complementary space W_j is generated by translations of one mother wavelet ψ, stretched to meet the current scale j, and thus it has a similar structure to V_j.

Scaling and wavelet equation and filter banks

Condition (i) of MRA states that $V_0 \subset V_1$, therefore the scaling function can be represented as a linear combination of functions from V_1 leading to the *scaling equation*

$$\phi(x) = \sum_{k\in\mathbb{Z}} h_k \sqrt{2}\phi(2x-k), \tag{2.23}$$

where $h_k = \int_{\mathbb{R}} \phi(x)\overline{\phi_{1k}(x)}dx = \langle \phi(x), \sqrt{2}\phi(2x-k)\rangle$; these are called '*high pass filters*'.

An analogous argument about the mother wavelet gives the *wavelet equation*

$$\psi(x) = \sum_{k\in\mathbb{Z}} g_k \sqrt{2}\phi(2x-k), \tag{2.24}$$

where $g_k = \int_{\mathbb{R}} \psi(x)\overline{\phi_{1k}(x)}dx = \langle \psi(x), \sqrt{2}\phi(2x-k)\rangle$; these are called *low pass filters*.

Wavelets can be obtained by iteration of filters with rescaling. The $\mathcal{DWT}$ is computed by successive low pass and high pass filtering of the discrete time-domain function. At each level, the high pass filter produces detail information, while the low pass filter associated with the scaling function produces coarse approximations. The resolution of the function is determined by the filtering operations, and the scale is determined by upsampling and downsampling (subsampling) operations.

2.4.4 Two-dimensional wavelet transforms

In some applications, as in Chapter 6, it is necessary a two dimensional transform. This can be obtained through a generalization of the theory of MRA introduced in the previous section. Here we consider a separable two-dimensional wavelet transform which can be directly designed from their one-dimensional counterparts. Then a two-dimensional scaling function is a product of two one-dimensional functions, $\phi(x,y) =$

$\phi(x)\phi(y)$, and the scaling equation takes the form

$$\phi(x,y) = \phi(x)\phi(y) = 2\sum_{k\in\mathbb{Z}}\sum_{l\in\mathbb{Z}} h_{kl}\phi(2x-k, 2y-l).$$

In a similar way, we can construct the wavelets. However, we have now three wavelet functions: $\psi^H(x,y) = \phi(x)\psi(y)$, $\psi^V(x,y) = \psi(x)\phi(y)$ and $\psi^D(x,y) = \psi(x)\psi(y)$, the corresponding wavelet equations being

$$\psi^H(x,y) = 2\sum_{k\in\mathbb{Z}}\sum_{l\in\mathbb{Z}} h_k g_l \phi(2x-k, 2y-l),$$

$$\psi^V(x,y) = 2\sum_{k\in\mathbb{Z}}\sum_{l\in\mathbb{Z}} g_k h_l \phi(2x-k, 2y-l),$$

$$\psi^D(x,y) = 2\sum_{k\in\mathbb{Z}}\sum_{l\in\mathbb{Z}} g_k g_l \phi(2x-k, 2y-l).$$

The superscripts 'H', 'V' and 'D' indicate that these finctions extract, respectively, horizontal, vertical and diagonal details of the two-dimensional signal.

2.4.5 Shrinkage estimators

Let $f \in \mathbb{L}^2(\mathbb{R}^d)$ be a deterministic signal. Assume that f admits the wavelet representation

$$\sum_{j\in\mathbb{Z}}\sum_{\boldsymbol{k}\in\mathbb{Z}^n} f_{j\boldsymbol{k}}\psi_{j\boldsymbol{k}}(\cdot),$$

where $\{\psi_{j\boldsymbol{k}} : j \in \mathbb{Z}, \boldsymbol{k} \in \mathbb{Z}^n\}$ denotes a wavelet basis of $\mathbb{L}^2(\mathbb{R}^d)$, and $\{f_{j\boldsymbol{k}} : j \in \mathbb{Z}, \boldsymbol{k} \in \mathbb{Z}^n\}$ its wavelets coefficients, that is, $f_{j\boldsymbol{k}} = \langle f, \psi_{j\boldsymbol{k}}\rangle_{\mathbb{L}^2(\mathbb{R}^d)}$.

In the white-noise setting, the observed process Y is defined as

$$Y_\sigma(d\boldsymbol{z}) = f(\boldsymbol{z})d\boldsymbol{z} + \sigma W(d\boldsymbol{z}),$$

where W is a Brownian motion, and σ denotes the noise level, for $0 < \sigma$.

The purpose of shrinkage estimators is to recover the unknown signal from the observed process; more specifically, an estimate of f is obtained which minimizes the mean

squared error (MSE). The shrinkage estimator family is then defined as

$$\mathcal{F} = \left\{ \hat{f}(\cdot) = \sum_{j\in\mathbb{Z}} \sum_{\boldsymbol{k}\in\mathbb{Z}^n} \gamma_{j\boldsymbol{k}} Y_{j\boldsymbol{k}} \psi_{jk}(\cdot) : \quad \gamma_{j\boldsymbol{k}} \in [0,1] \right\},$$

where $Y_{j\boldsymbol{k}}$ are the empirical (observed) wavelet coefficients, that is, the wavelet coefficients of process Y_σ, and $\{\gamma_{jk} : j \in \mathbb{Z}, \boldsymbol{k} \in \mathbb{Z}^n\}$ are the weights defining the shrinkage function, i.e., the estimator $\hat{f}$ of the signal of interest f.

The simplest wavelet non-linear shrinkage technique is named '*thresholding*'. This technique consists of comparing the wavelet coefficients with a threshold λ and deciding which ones are significant. The two most common policies are hard and soft; their analytical expression are, respectively,

$$\delta^h_\lambda(\hat{f}_{j\boldsymbol{k}}, \lambda) = \begin{cases} 0 & \text{if } |\hat{f}_{j\boldsymbol{k}}| \leq \lambda \\ \hat{f}_{j\boldsymbol{k}} & \text{if } |\hat{f}_{j\boldsymbol{k}}| > \lambda \end{cases}$$

and

$$\delta^s_\lambda(\hat{f}_{j\boldsymbol{k}}, \lambda) = \begin{cases} 0 & \text{if } |\hat{f}_{j\boldsymbol{k}}| \leq \lambda \\ \hat{f}_{j\boldsymbol{k}} - sgn(\hat{f}_{j\boldsymbol{k}})\lambda & \text{if } |\hat{f}_{j\boldsymbol{k}}| > \lambda. \end{cases}$$

In Chapter 6 we use hard thresholding, although other techniques could be considered. The choice of λ is a key aspect in thresholding procedures. A small threshold yields a result close to the input but this result may contain noise. On the other hand, a large threshold produces a signal with a large amount of zero wavelet coefficients, which may remove some of the signal singularities. Many methods have been developed to look for a threshold which minimizes the error of the result as compared with the noise-free data, but since these data are unknown, the error cannot be computed or minimized exactly.

Minimax and maxiset approaches

Two main approaches have been adopted in the definition of optimality of shrinkage estimators: The minimax and the maxiset approaches. Specifically, an estimator $\hat{f}^*$ is said to be '*minimax*' in the class $\mathcal{F}(M)$ of functions in the normed space $\mathcal{F}$ having norms bounded above by a fixed constant M if (see, for instance, Donoho and Johnstone

1994a[42], 1994b[43], 1995[44] and Donoho, Johnstone, 1998[45])

$$\sup_{f\in\mathcal{F}(M)} R_n(\hat{f}_n^*, f) = \inf_{\hat{f}} \sup_{f\in\mathcal{F}(M)} R_n(\hat{f}_n, f),$$

where $R_n(\hat{g}_n, g) = E\|\hat{g}_n - g\|_{\mathcal{F}}$ denotes the risk in the $\mathcal{F}-$norm, and $\hat{g}_n$ denotes, in our case, a wavelet shrinkage estimator. Thus, for a given class $\mathcal{F}(M)$ of deterministic signals, the optimality is measured in terms of the best convergence rate achieved.

In the maxiset approach, the quality of the estimators is measured in terms of the largest space where a given convergence rate is achieved (see Kerkyacharian and Picard 2000[70], 2002[71]; Vidakovic 1998a[118]). Thus, the '*maxiset*' is defined as

$$MS(\hat{f}_n, \rho, \alpha_n)(T) = \left\{\theta \in \Theta : \sup_n E_\theta^n \rho(\hat{f}_n, f(\theta))(\alpha_n)^{-1} \leq T\right\},$$

for a fixed constant T, a loss function ρ, a convergence rate α_n, and a sequence of estimators $\hat{f}_n$. Here, $M_n = \{P_\theta^n : \theta \in \Theta\}$ defines the underlying sequence of models, with $\{P_\theta^n : n \in \mathbb{N}, \theta \in \Theta\}$ being probability distributions on $\{\Omega_n, : n \in \mathbb{N}\}$, and Θ the parameter space (see also Autin 2008[22], on $\mu-$thresholding rules, under the maxiset approach).

2.5 Entropy measures

Entropy and related information measures are useful to quantify uncertainty and dependence. In Chapter 6, we develop a shrinkage estimation methodology based on mutual information criteria for scale-adaptive selection of the threshold.

2.5.1 Introduction

Entropy was defined first in the context of thermodynamics and statistical mechanics. However, we refer here to the concept of entropy introduced by Shannon (1948)[106] in the context of Information Theory. Assuming a set of possible events whose probabilities of occurrence are $p_1, p_2, ..., p_n$, the aim is to define a measure of how much 'choice' is involved in the selection of the event, or how uncertain is the outcome. Such a measure, denoted $H(p_1, p_2, ..., p_n)$, should satisfy three requirements:

i) H is continuous in the p_is;

ii) if all the p_i are equally probable, $p_i = 1/n$, then H is a monotonically increasing function of n;

iii) H is additive.

Shannon proved that the only H satisfying these three assumptions is of the form

$$H(p_1, p_2, ..., p_n) = -K \sum_{i=1}^{n} p_i \ln p_i,$$

where K is a constant which depends on a unit of measure and on the base of the logarithm chosen.

2.5.2 Shannon's entropy and related measures for discrete random variables

Definition 2.5.1. The *Shannon entropy* of a discrete random variable X with alphabet χ and probability mass function $p(x) = P\{X = x\}$, $x \in \chi$, is defined as

$$H(X) = -\sum_{x \in \chi} p(x) \log p(x).$$

The log can be considered in any base and, depending on the base, the entropy is expressed in different units. In the information-theoretical context of entropy, typical units of measure are bits and the base of the logarithm is 2. Commonly, the convention $0 \log 0 = 0$ is used, which is easily justified by continuity since $x \log x \to 0$ as $x \to 0$.

Note that the entropy of X can be interpreted as the expected value of the random variable $\log \frac{1}{p(X)}$, where X is drawn according to the probability mass function $p(x)$. Thus,

$$H(X) = E_p \left[\log \frac{1}{p(X)} \right].$$

Properties 2.5.1. 1. $H(X)$ is a function only of the probabilities $p(x_i)$s, independent of their ordering or labeling.

2. $H(X) \geq 0$, with equality if and only if all except one of the p_is are 0, i.e., $p(x)$ is a point mass at some $x_0 \in \chi$.

3. $H(X) \leq \log |\chi|$, with equality if and only if $p_i = p_2 = ... = p_n$, i.e., $p(x)$ is uniform.

4. H is a strictly concave function of X, i.e., for $0 \leq \lambda \leq 1$ and random variables X and Y,
$$H(\lambda X + (1-\lambda)Y) \geq \lambda H(X) + (1-\lambda)H(Y),$$
with equality if and only if $\lambda = 0$, or $\lambda = 1$, or $X = Y$.

Considering two or more random variables as a single vector-valued random variable, the previous definition can be extended to define the joint entropy, and then the entropy of a random variable given another.

Definition 2.5.2. Let (X, Y) be a pair of discrete random variables with a joint distribution $p(x, y)$. The *joint entropy* is defined as

$$H(X, Y) = -\sum_{x,y} p(x, y) \log p(x, y) = -E_p[\log p(x, y)].$$

Proposition 2.5.2. $H(X + Y) \leq H(X) + H(Y)$, with equality if and only if X and Y are independent.

Definition 2.5.3. Let (X, Y) be a pair of discrete random variables with a joint distribution $p(x, y)$. The *conditional entropy* is defined as

$$H(X|Y) = -\sum_{x,y} p(x, y) \log p(x|y) = -E_p[\log p(x|y)].$$

Theorem 2.5.3. (Entropy Chain Rule)

$$H(X_1, X_2, ..., X_n) = H(X_1) + \sum_{i=1}^{n} H(X_n|X_1, ..., X_{n-1}).$$

Kullback and Leibler (1951)[73] introduced an information measure similar to Shannon's entropy to measure the distance between two distributions (although not a distance in the true metric sense since it is not symmetric and does not satisfy the triangle inequality).

Definition 2.5.4. The *relative entropy* or *Kullback-Leibler divergence* between two probability mass functions $p(x)$ and $q(x)$ is defined as

$$D(p,q) = \sum_{x\in\chi} p(x)\log\frac{p(x)}{q(x)} = E_p\left[\log\frac{p(x)}{q(x)}\right],$$

where the conventions $0\log\frac{0}{0} = 0$, $0\log\frac{0}{q}$ and $p\log\frac{p}{0} = \infty$ are used.

The Kullback-Leibler divergence is a measure of inefficiency of assuming the distribution $q(x)$ when the true distribution is $p(x)$.

Proposition 2.5.4. Let $p(x)$ and $q(x)$ be two probability mass functions. Then,

$$D(p,q) \geq 0,$$

with equality if and only if $p(x) = q(x)$ for all x.

The next measure quantifies the mutual dependence of two variables and gives the amount of information that one variable contains about another.

Definition 2.5.5. Let (X,Y) be a pair of discrete random variables with a joint distribution $p(x,y)$. The *mutual information* between X and Y is defined as

$$\begin{aligned} I(X;Y) &= \sum_{x,y} p(x,y)\log\frac{p(x,y)}{p(x)p(y)} \\ &= D(p(x,y), p(x)p(y)) \\ &= E_{p(x,y)}\left[\log\frac{p(x,y)}{p(x)p(y}\right], \end{aligned} \tag{2.25}$$

with $p(x)$ and $p(y)$ being the marginal probability mass functions of X and Y, respectively.

Equivalent expressions for the mutual information are:

$$\begin{aligned} I(X;Y) &= H(X) + H(Y) - H(X,Y) \\ &= H(X) - H(X|Y) \\ &= H(Y) - H(Y|X) = I(Y;X). \end{aligned} \tag{2.26}$$

Therefore, we can see the mutual information from two different points of view, as is showed in Whittaker (1990)[122]. On the one hand, seeing (2.25), the mutual information measures the divergence between the joint distribution of X and Y and the 'independent' distribution given by the product of the marginal probability mass functions of X and Y. In this sense it quantifies the departure from independence. On the other hand, seeing (2.26), the mutual information is closely related to the problem of predicting the value of one variable from the value of another. That is, it is a natural measure of the information held by one random variable about another.

Definition 2.5.6. The *conditional mutual information* of random variables X, Y given Z is defined as

$$\begin{aligned} I(X;Y|Z) &= H(X|Z) - H(X|Y,Z) \\ &= E_{p(x,y,z)}\left[\log \frac{p(x,y|z)}{p(x|z)p(y|z)}\right]. \end{aligned}$$

Theorem 2.5.5. (Chain Rule for mutual information)

$$I(X_1, X_2, ..., X_n; Y) = \sum_{i=1}^{n} I(X_i; Y|X_{i-1}, X_{i-2}, ..., X_i).$$

2.5.3 Shannon's entropy and related measures for continuous random variables

For continuous random variables, an analogous measure to the entropy has been defined, this is called 'differential entropy'. Although, in a first view, it can be seen as an extension of the first, there are some important differences between them. We use a lowercase h to indicate that this is the differential entropy rather than discrete entropy H.

Definition 2.5.7. The *differential entropy* of a continuous random variable X with density function $f(x)$ is defined as

$$h = -\int_S f(x)\log f(x)dx = -E_f[\log f(x)],$$

where S is the support of X.

Properties 2.5.6. 1. $h(X + c) = h(X)$, for all $c \in \mathbb{R}$,

2. $h(aX) = h(X) + \log|a|$, for all $a \in \mathbb{R}$.

An important difference with respect to the entropy of a discrete random variable is that the differential entropy of a continuous random variable can take values less than 0.

Related measures, as defined previously in the discrete case, can be formulated for continuous random variables, replacing sums with integrals.

Definition 2.5.8. Let (X, Y) be a pair of variables with a joint density function $f(x, y)$. The *joint differential entropy* is defined as

$$h(X, Y) = -\int f(x, y) \log f(x, y) dxdy = -E_f[\log f(x, y)].$$

Definition 2.5.9. Let (X, Y) be a pair of random variables with a joint density function $f(x, y)$. The *conditional differential entropy* is defined as

$$h(X|Y) = -\int_{x,y} f(x, y) \log f(x|y) = -E_f[\log f(x|y)].$$

In general, $f(x|y) = f(x, y)/f(y)$, hence we can write

$$h(X|Y) = h(X, Y) - h(Y),$$

assuming that both terms in the right-hand side are not equal to infinity simultaneously.

Theorem 2.5.7. (Entropy of a multivariate normal distribution) Let $X_1, X_2, ..., X_n$ have a multivariate normal distribution with mean ν and covariance matrix Σ. Then,

$$h(X_1, X_2, ..., X_n) = \frac{1}{2} \log(2\pi e)^n |\Sigma|,$$

where $|\Sigma|$ denotes the determinant of Σ.

Definition 2.5.10. The *relative entropy* or *Kullback-Leibler divergence* between two density functions $f(x)$ and $g(x)$ is defined as

$$D(f, g) = \int f(x) \log \frac{f(x)}{g(x)} dx = E_f \left[\log \frac{f(x)}{g(x)}\right].$$

Definition 2.5.11. The *mutual information* between two random variables X and Y with joint density function $f(x, y)$ is defined as

$$\begin{aligned} I(X;Y) &= \int f(x,y) \log \frac{f(x,y)}{f(x)f(y)} dxdy \\ &= E_{f(x,y)} \left[\log \frac{f(x,y)}{f(x)f(y} \right], \end{aligned}$$

with marginal probability density functions $f(x)$ and $f(y)$.

Similarly to the discrete case, other ways to express the mutual information of continuous random variables are:

$$\begin{aligned} I(X;Y) &= D(f(x,y), f(x)f(y)) \\ &= h(X) + h(Y) - h(X,Y) \\ &= h(X) - h(X|Y) = I(Y;X), \end{aligned}$$

2.5.4 Some extensions of Shannon's entropy

Extensions of Shannon's work have resulted in many alternative measures of information or entropy. Here we refer to two widely known families.

Renyi (1960)[97], relaxing the requirement of additivity, extends the Shannon's entropy to a continuous family of entropies parameterized by a continuous parameter q, defined by

$$H_q(X) = \frac{1}{1-q} \log \sum_{i=1}^{n} p_i^q, \quad q \geq 0.$$

Some particular cases of interest are:

- $H_0(X)$, which is the logarithm of the cardinality of X, named '*Hartley entropy*'.

- $q \to 1$, since for any probability distribution X we have $H_q(X) \to H(X)$, that is, it converges to the Shannon entropy.

- $H_2(X)$, which is the negative logarithm of the likelihood of two independent random variables with the same probability distribution to have the same value. This is called '*collision entropy*', or just '*Rényi entropy*'.

- $H_\infty(X)$, which is a function only of the highest probability, called '*min-entropy*'.

Tsallis (1998)[116] introduced the unique non-extensive entropy as

$$H_q(X) = \frac{1 - \sum_{i=1}^{n} p_i^q}{q-1}, \quad q \in \mathbb{R},$$

where q is the entropic index that characterizes the degree of non-extensivity, since given two random variables X and Y,

$$H_q(X+Y) = H_q(X) + H_q(Y) + (1-q)H_q(X)H_q(Y).$$

Note that, for $q = 1$, the Tsallis entropy coincides with the Shannon entropy.

Chapter 3

Structural Analysis of Spatial Threshold Exceedances

Due both to the complexity of real systems and to the technical difficulties inherent to extremal analysis, the statistics of extremes in spatio-temporal processes has become one of the most challenging research areas in relation to the increasingly demanding interest on risk assessment tools in many fields of application. Recent advances in spatio-temporal statistical analysis are focused, in particular, on the formulation and study of new model families, flexible to represent such real complexities and, at the same time, suitable for technical treatment and interpretation, as well as on related system dynamics problems.

In this chapter, significant characteristics of threshold exceedances with reference to structural properties of the processes generating such events, particularly in the context of input/output systems, are analyzed. Specifically, the effect of spatial deformation and blurring transformations on the second-order structure and the geometrical properties of excursion sets is studied and illustrated through some simulated cases.

3.1 Introduction

During the last two decades, there has been an important effort in formulating and studying sophisticated models for spatio-temporal processes suitable to represent, in an appropriate form, structural characteristics of interest such as local variability orders, memory ranges, spatio-temporal interaction, heterogeneities, multiscale properties, etc.

Among the increasing number of contributions, in some cases giving a synthetic view of different approaches in this context, we shall mention Brown *et al.* (2000)[29], Christakos (2000)[34], Cressie and Huang (1999)[39], De Iaco *et al.* (2002)[64], Gneiting (2002)[53], Gneiting and Schlather (2004)[55], Kolovos *et al.* (2004)[72], Kyriakidis and Journel (1999)[74], Ma (2003)[79], (2008)[80], Porcu *et al.* (2008)[96] and Stein (2005)[109]. The theory of generalized random fields provides a suitable framework to formulate and study model families which incorporate some of the mentioned structural characteristics in terms of appropriate test function spaces (see, for instance, Ruiz-Medina *et al.* 2008[104]).

The problems of detection and prediction of extremal events in input/output systems randomly evolving in space and time brings us necessarily to very primary questions on how such occurrences are generated, which are their effects, then to the analysis of their characteristics in terms of the system structural properties, namely local regularity/singularity orders, ranges of memory, heterogeneities, multiscale properties, spatio-temporal interaction, among others. The first question has long been a matter of interest particularly in engineering oriented research (see, for example, Aldous 1989[5]; Vanmarcke 2010[117]), but also in environmental applications. For instance, abnormal persistence or, what we can call, concentration of 'energy' in input variables, even under critical point levels, can eventually lead to and explain observed output extremal episodes. Conversely, extremal input deviations may originate abnormal behavior in an output variable, or just be absorbed by the system with no significant effect, either observed or in terms of risk. Relatedly, public health policies are often based, sometimes with no well-founded justification, on alert indicators involving a combination of smoothing operations (e.g. averaging, thresholding) which damp and hide small space or time scale exceptional events (see, for instance, Le and Zidek 2006[75]).

Among other significant transformations eventually involved in a system, and hence affecting its structural properties, we here refer to deformation and blurring, meaningful for a variety of problems and applications. In particular, Angulo and Mardia (1998)[18], and Mardia *et al.* (2006)[86], propose an entropy-based approach for evaluation of the structural distortion effects from spatial deformation. Goitía *et al.* (2004)[56] develop a functional approach, in a generalized random field framework, to estimate deformation and blurring in a system, based on entropy criteria.

We analyze the implications of transformations of this type in relation to the properties of threshold exceedances in input/output systems (see Angulo and Madrid 2008[11],

2009a[12], 2009b[13] and 2010a[15]). Regarding spatial deformation, even for one-to-one smooth transformations, the difference between the cases where deformation affects the whole system or intervenes only at the input level is critical in relation to the genesis and effects of extremal behavior. In the first case, the underlying input-output correspondence remains invariant, except for the point displacements and the change in the space-time local measure. However, properties of threshold exceedances based on natural temporal and geographical coordinates, such as structural patterns and related characteristics, as well as those derived for local (in space and/or time) averages, etc., are consequently affected by the system deformation. Spatial heterogeneity characteristics are also modified, which has been used, in particular, to generate or approximate non-homogeneous models based on deformation of homogeneous random fields (among others, see Sampson and Guttorp 1992[105]; Anderes and Stein 2008[8]; and the references therein). Further interest, from structural point of view, has the second case, since deformation at the input level actually involves a transformation of the system filter, hence affecting the output probabilistic structure and, in particular, the genesis of eventual output threshold exceedances from the altered spatio-temporal distribution of the input 'energy'. A similar analysis is performed as to blurring transformations. Here, the main interest lies in the change of local variability orders, multiscale characteristics and space-time interaction (see, for example, Ruiz-Medina and Angulo 2007[103]), again with significant consequences in relation to the genesis and effects of threshold exceedances in input/output systems.

In Section 3.2 we define 'deformed' and 'blurred' random fields, deriving some formal implications of the transformations involved. Section 3.3 illustrates through simulated examples the effect of spatial deformation and blurring on the second-order structure of a random field and on some of the geometrical characteristics of its excursion sets, namely the Euler characteristic and the hypervolume.

3.2 Spatial deformation and blurring

Spatial deformation has had two main significant applications in the random field context. First, in image warping, where a template image is used as a reference for exact or approximate correspondence with related observed images based on a deformation of the spatial domain, generally subject to suitable regularity conditions; different approaches to define such deformations include landmark-based thin-plate splines (see, for example,

Glasbey and Mardia 1998[50], 2001[51]; Mardia *et al.* 2006[86]), viscoelastic fluid kinematics SPDE models (see, for instance, Amit *et al.* 1991[7]; Christensen *et al.* 1996[35]), etc. Second, spatial deformation has been used to fit homogeneous, possibly isotropic random field models to spatial data which exhibit a certain lack of homogeneity which can be explained or corrected in terms of such type of transformations (see Sampson and Guttorp 1992[105]; Perrin and Senoussi 1999[94]; Clerc and Mallat 2003[36]; Anderes and Stein 2008[8], among others). The latter can be seen as an approach to definition of flexible classes of non-homogeneous random fields from homogeneous ones. Angulo and Mardia (1998)[18], and Mardia *et al.* (2006)[86], propose an entropy-based approach for evaluation of the structural distortion effects from spatial deformation.

Formally, let X be a random field on $D \subseteq \mathbb{R}^d$. Let $\Phi : D \to D' \subseteq \mathbb{R}^d$ be a spatial deformation, such that Φ is a C^1-diffeomorphism with positive Jacobian, $|J_\Phi| > 0$. By $X[\Phi]$ we denote the 'deformed' random field, defined by

$$X[\Phi](\mathbf{s}) = X(\Phi^{-1}(\mathbf{s})), \quad \forall \mathbf{s} \in D'. \tag{3.1}$$

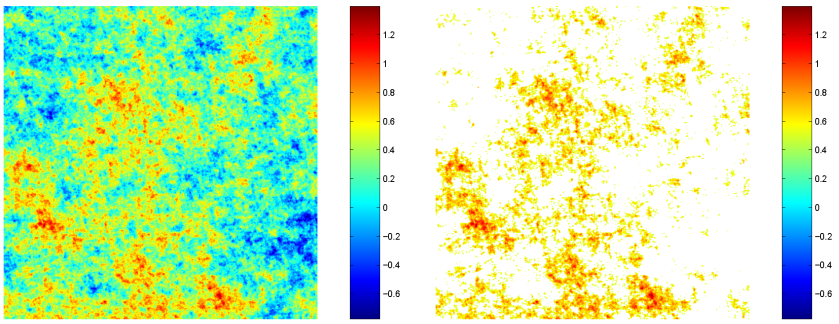
This common interpretation of a random field deformation means that the state values remain as original, but reallocated. Hence, for a second-order random field, the point displacement and derived local change of measure induce a distortion effect on the random field second-order structure. However, variances remain as originally in displaced points. Regarding the effects of spatial deformation on excursion set characteristics, it is evident that shapes are deformed and hypervolumes change also because of local change of measure; since regularity of Φ preserves the connectivity properties of excursion sets, Euler characteristic remains invariant in the transformed space, but obviously changes as referred to the original domain.

The situation is substantially different if deformation should be interpreted as acting at the stage of generation of a 'flow' random field from purely random inputs. In this case, the effect of deformation of the base space on the aggregation of the input energy leads to the following state representation of the deformed random field, $X[\tilde{\Phi}]$:

$$X[\tilde{\Phi}](\mathbf{s}) = X(\Phi^{-1}(\mathbf{s}))|J_\Phi(\Phi^{-1}(\mathbf{s}))|^{-1} = X(\Phi^{-1}(\mathbf{s}))|J_{\Phi^{-1}}(\mathbf{s})|, \quad \forall \mathbf{s} \in D'. \tag{3.2}$$

As a consequence, in this case, the distortion effect on the second-order structure reflects

a local change of the state-scale (hence, of the variances) derived from the Jacobian factor. More interesting, excursion sets display related geometrical variations, not just in terms of shape deformation, but also in structure due to the locally heterogeneous modification of the level scales. Thus, for $u > 0$, a value of $|J_\Phi(\Phi^{-1}(\mathbf{s}))| < 1$, which means local contraction, induces a possible extension in $A_u[\tilde{\Phi}]$ with respect to $A_u[\Phi]$, and conversely, for $|J_\Phi(\Phi^{-1}(\mathbf{s}))| > 1$, which means local dilation, there will be a possible reduction in $A_u[\tilde{\Phi}]$.



(a) (b)

Figure 3.1: (a) Simulated spatial realization of Cauchy random field; (b) excursion set for $u = 0.4793$.

To illustrate graphically these effects we consider a simulated realization on the square $[-200, 200]^2 \in \mathbb{R}^2$ of Cauchy random field model (see Section 2.2.2) with $\sigma^2 = 0.1$, displayed in plot (a) of Figure 3.1; plot (b) shows the excursion set corresponding to the threshold $u = 0.4793$. This specific value for the threshold u was simply selected, from a dense set of points partitioning the range of variation of the realization of X, to illustrate an intermediate level where the corresponding excursion set is structured with a high number of connected components.

Based on this realization, in Figures 3.2 and 3.3 we depict the effect of two different spatial deformations, named D1 and D2. These are defined in terms of landmark-based thin-plate splines (Bookstein 1989[27]), to have different contraction and dilation properties, particularly at the central area of the square domain considered. In each of both cases, sixteen landmark points are considered, sharing a subset of eight boundary anchor points (that is, points forced to remain fixed under the deformation), consisting of the four corners and the four edge middle points, to prevent significant distortion of the square borderlines. For D1, the eight interior points $(0, -100)$, $(0, 100)$, $(-100, 0)$,

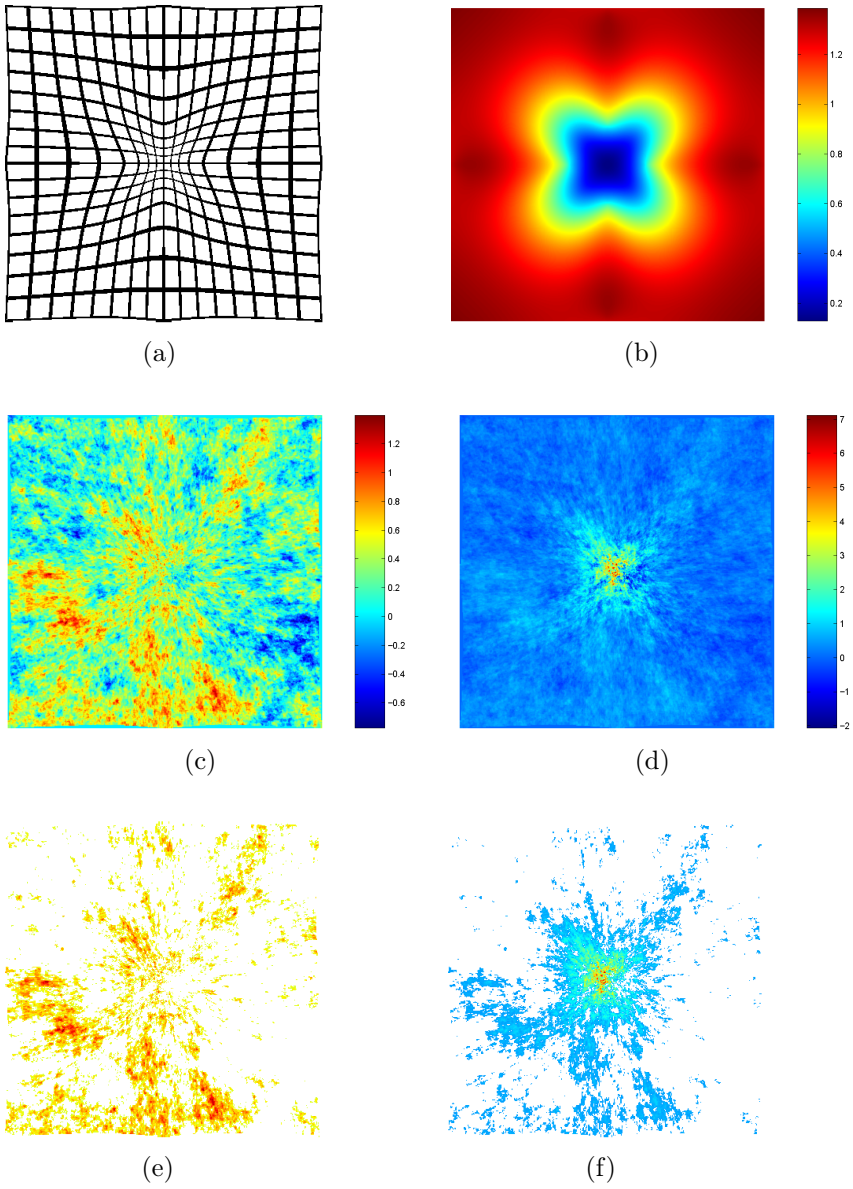


Figure 3.2: Deformation D1: (a) Effect on a regular mesh; (b) Jacobian determinant values; (c)-(d) effect on simulated realization of Cauchy random field, considered as a 'level' or as a 'flow' variable, resp.; (e)-(f) corresponding excursion sets for $u = 0.4793$.

$(100, 0)$, $(-50, -50)$, $(50, 50)$, $(-50, 50)$, and $(50, -50)$ are forced to move half distance towards the square central point, i.e. the origin (0,0); that is, they are respectively transformed into the points $(0, -50)$, $(0, 50)$, $(-50, 0)$, $(50, 0)$, $(-25, -25)$, $(25, 25)$, $(-25, 25)$, $(25, -25)$. On the other hand, D2 is based on the reciprocal assignments. Visualization

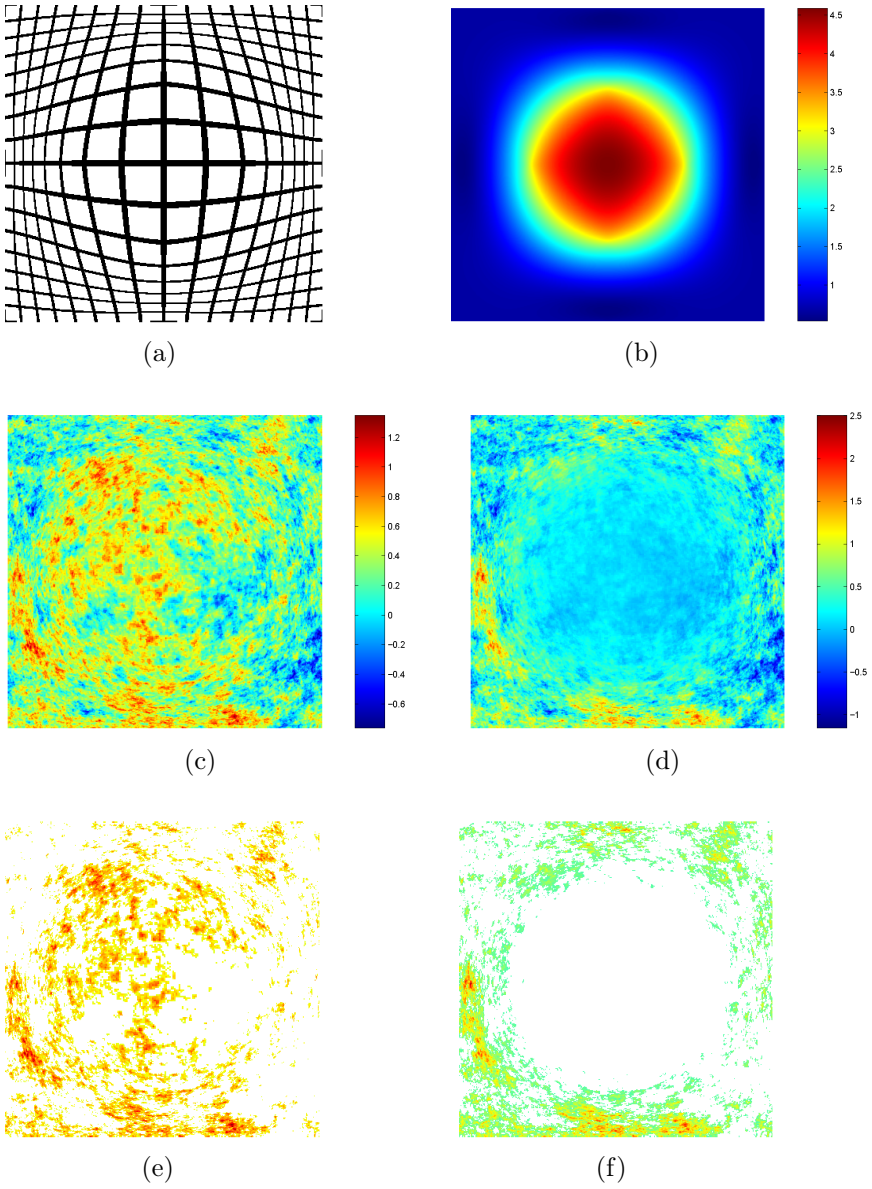


Figure 3.3: Deformation D2: (a) Effect on a regular mesh; (b) Jacobian determinant values; (c)-(d) effect on simulated realization of Cauchy random field, considered as a 'level' or as a 'flow' variable, resp.; (e)-(f) corresponding excursion sets for $u = 0.4793$.

of contraction and dilation effects is helped by applying both deformations on a regular mesh template, in plot (a), and by displaying the corresponding Jacobian determinant values, in plot (b), for each figure. The random field is considered either as a 'level' or as a 'flow' variable (see, respectively, plots (c) and (d) of each figure). Plots (e) and (f)

of both figures show the excursion sets based on the threshold $u = 0.4793$ corresponding to the transformed realizations displayed in the respective plots (c) and (d). The local change of scale due to the Jacobian determinant factor becomes evident in the 'flow' case, with significant structural effects on the excursion sets.

Another case of random field transformation important for modeling and applications is blurring, which can be properly formalized in terms of convolution operators. Blurring can be directly related to physical effects derived, for example, from the way input signals are collected and registered by technical devices. But it has been also used as a mathematical means for model generation; in particular, as a local smoothing operation and, in the spatio-temporal context, to represent the diffusive transmission of information between different times (see, for instance, the approach by Brown *et al.* 2000[29], later referred in Chapter 7).

For simplicity, here we assume that random field X is defined on the whole space $\mathbb{R}^d$ (otherwise, specifications would be required for convolution in relation to the domain boundary). Let $h : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ be a kernel such that $h \geq 0$ and $\int_{\mathbb{R}^d} h(\mathbf{s}, \mathbf{s}')d\mathbf{s}' = 1$, $\forall \mathbf{s} \in \mathbb{R}^d$. The 'blurred' random field $X[h]$ is defined by

$$X[h](\mathbf{s}) = h * X(\mathbf{s}) = \int_{\mathbb{R}^d} h(\mathbf{s}, \mathbf{s}')X(\mathbf{s}')d\mathbf{s}'.$$

In many cases, blurring can be considered to occur homogeneously in space, that is, kernel h depends on the difference vector (or, even, isotropically on the distance), being then defined as $h : \mathbb{R}^d \to \mathbb{R}$, with $h \geq 0$ and $\int_{\mathbb{R}^d} h(\mathbf{s})d\mathbf{s} = 1$. In this case, the 'blurred' random field is formulated as

$$X[h](\mathbf{s}) = h * X(\mathbf{s}) = \int_{\mathbb{R}^d} h(\mathbf{s} - \mathbf{s}')X(\mathbf{s}')d\mathbf{s}'.$$

The smoothing effect of blurring on the random field second-order structure is better understood in the isotropic case, leading to variance reduction, increased dependence ranges, and increased local regularity of the field. As for the geometrical properties of excursion sets, we will expect smoother shapes and boundaries, and reduced expected Euler characteristic and hypervolume, among other consequences.

In Figure 3.4, a Gaussian kernel, h, with three different variance values ($\sigma^2_{h_1} = 10$, $\sigma^2_{h_2} = 100$, and $\sigma^2_{h_3} = 1000$), is used to generate blurred images based on the simulated

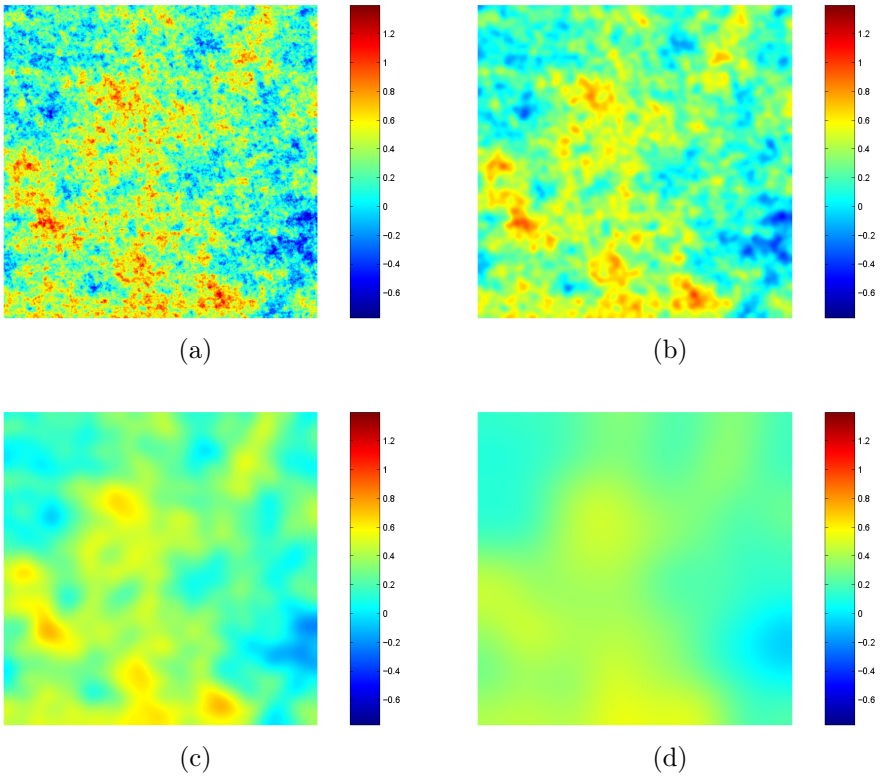


Figure 3.4: Gaussian blurring on simulated realization of Cauchy random field: (a) Original field; (b)-(c)-(d) blurred field for $\sigma^2_{h_1} = 10$, $\sigma^2_{h_2} = 100$, and $\sigma^2_{h_3} = 1000$, resp.

realization of Cauchy random field model previously considered. Threshold exceedance sets for $u = 0.374$ shown in Figure 3.5 illustrate the above mentioned structural differences. Again, selection of this specific value for threshold u was intentional just for visualization of significant contrasting features.

In some applications, a combination of both blurring and deformation transformations may be present in the model; see, for example, Goitía *et al.* 2004[56], where a functional approach is developed, in a generalized random field framework, to estimate deformation and blurring in a system, based on entropy criteria. Taking, for simplicity, $D = D' = \mathbb{R}^d$, the deformed and (then) blurred random field $X[\Phi; h]$ is defined by

$$X[\Phi; h](\mathbf{s}) = h * X[\Phi](\mathbf{s}) = \int_{\mathbb{R}^d} h(\mathbf{s}, \mathbf{s}') X(\Phi^{-1}(\mathbf{s}')) d\mathbf{s}', \quad \forall \mathbf{s} \in \mathbb{R}^d.$$

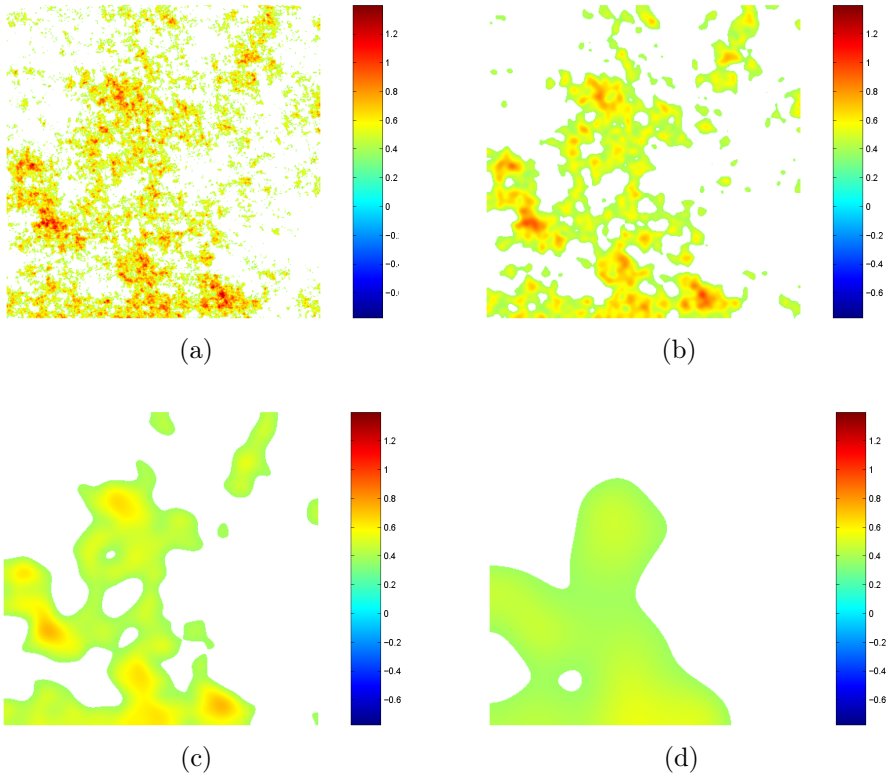


Figure 3.5: Threshold exceedance sets for $u = 0.374$ based on original and blurred simulated realization of Cauchy random field (see Figure 3.4).

Equivalently, $X[\Phi; h] = X[h_\Phi]$, with

$$h_\Phi(\mathbf{s}, \mathbf{v}') = h(\mathbf{s}, \Phi(\mathbf{v}'))|J_\Phi(\mathbf{v}')|.$$

Similarly, in the 'flow' random field case,

$$X[\tilde{\Phi}; h](\mathbf{s}) = h * X[\tilde{\Phi}](\mathbf{s}) = \int_{\mathbb{R}^d} h(\mathbf{s}, \mathbf{s}')X(\Phi^{-1}(\mathbf{s}'))|J_{\Phi^{-1}}(\mathbf{s}')|d\mathbf{s}', \quad \forall \mathbf{s} \in \mathbb{R}^d,$$

also defined as $X[\tilde{\Phi}; h] = X[h_{\tilde{\Phi}}]$, with

$$h_{\tilde{\Phi}}(\mathbf{s}, \mathbf{v}') = h(\mathbf{s}, \Phi(\mathbf{v}')).$$

It is important to note that, according to the order in which both transformations are applied here, blurring has a certain 'hiding' effect on deformation, which raises an in-

teresting problem of identifiability in applications. The joint effect of deformation and blurring is considered in Chapter 7 as the basis for a spatio-temporal model formulation which extends the blur-generated model introduced in Brown *et al.* (2000)[29] (see also Angulo and Madrid 2009b[13] and 2010a[15]).

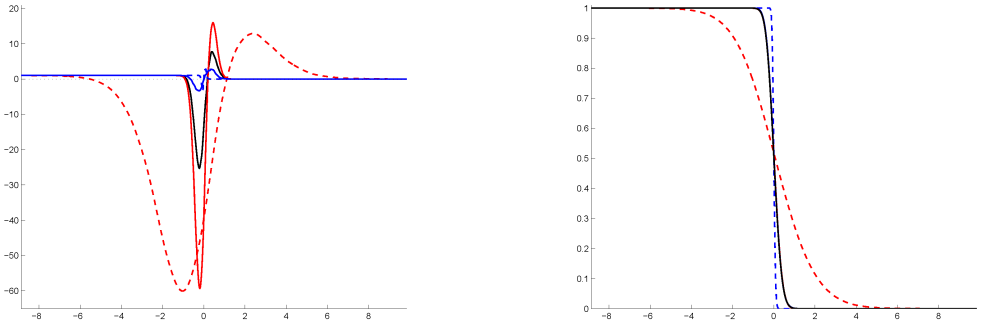
3.3 Evaluation of geometrical characteristics

To have more insight into several significant effects of deformation and blurring on the structure of spatial threshold exceedance sets, we calculate average values of the Euler characteristic and the hypervolume (i.e., the area, in this case) for a dense sequence of thresholds u, based on 300 simulated independent replicates of Cauchy random field on $\mathbb{R}^2$, with $\sigma^2 = 0.1$, $\alpha = 2$, and $\beta = 0.2$. The study is restricted to the subdomain $S = [-25, 25]^2$, a central area within the square $[-200, 200]^2$. The results are drawn and compared in a series of curves displayed in Figure 3.6, for deformation, and Figure 3.7, for blurring. In both cases, we use the example transformations previously introduced.

In plot (a) of Figure 3.6 we can see, for both the 'level' (Φ) and the 'flow' ($\tilde{\Phi}$) cases, the increase or reduction effect on the degree of disgregation of threshold exceedance sets, measured in terms of the Euler characteristic, caused by spatial deformation, depending on whether this transformation acts as a local contraction or dilation, respectively. This effect is clearly affected by extra local change of scale in the 'flow' case. As commented in Section 2.2.3, for large u the different rates of decay towards 0 of the Euler characteristic can be interpreted, by approximation, in terms of the probability

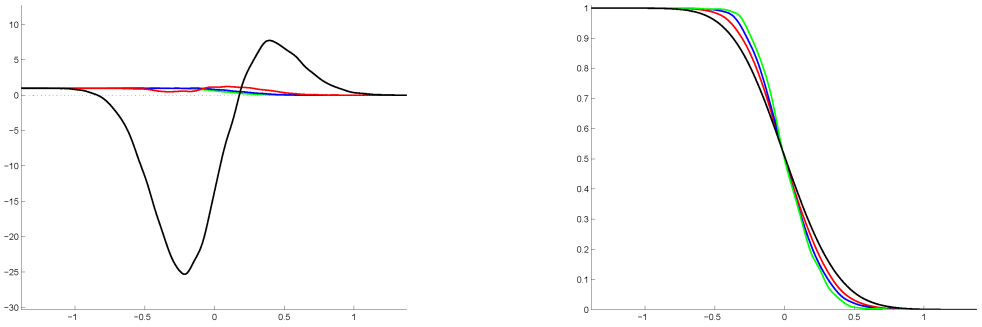
$$P\{\sup_{\boldsymbol{s}\in D} X(\boldsymbol{s}) \geq u\} \tag{3.3}$$

of the system entering into alert on some location in S; for instance, in the 'flow' case and for a contracting deformation, such a probability becomes significantly higher. As for hypervolumes, plot (b), there is no significant effect of deformation in the 'level' case (note that the homogeneity of the random field considered is also reflected in the average proportion of areas above and below threshold u; such a proportion hence remains invariant in expectation by a 'level' deformation). However, in the 'flow' case, the change of scale due to the Jacobian determinant factor modifies, as before, the probabilities of a generic point in S entering into a critical state.



(a) (b)

Figure 3.6: Curves of (a) Euler characteristic, and (b) hypervolume for: (black) original field; deformed field under D1, (red) 'level' case, (dashed red) 'flow' case; deformed field under D2, (blue) 'level' case, (dashed blue) 'flow' case.

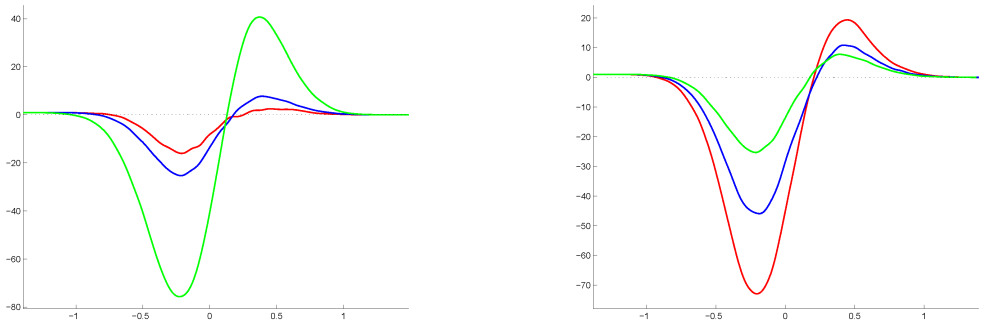


(a) (b)

Figure 3.7: Curves of (a) Euler characteristic, and (b) hypervolume for: (black) original field; blurred field with kernel (red) h_1, (blue) h_2, and (green) h_3.

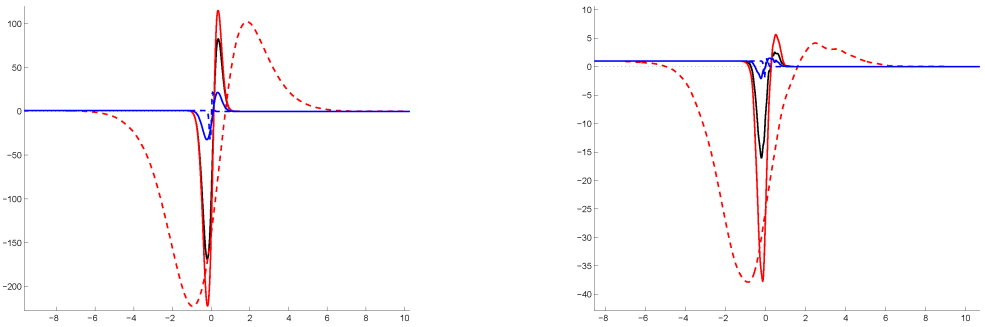
The smoothing effect of blurring (for usual 'well-behaved' kernels) is translated into a reduction in structural complexity of threshold exceedance sets, particularly in the degree of disgregation, as shown in Figure 3.7, plot (a), for the three Gaussian kernels previously considered. Again, for large u, the probability (3.3) is progressively reduced as the blurring kernel variance σ_h^2 increases. Similarly, the range of u values where the average proportion of points in a critical state significantly changes from 1 to 0, as u increases, gets reduced by the blurring transformation of the field.

Although the main features of these effects are a consequence of the nature of both types of transformations considered, and hence would be expected to be similar for general random fields, it is interesting to analyze differences derived from certain spe-



(a) (b)

Figure 3.8: Curves of Euler characteristic: plot (a), for fixed $\alpha = 2$ and varying β =0.1 (red), 0.2 (blue), 0.9 (green); plot (b), for fixed $\beta = 0.2$ and varying α =0.5 (red), 1 (blue), 2 (green).



(a) (b)

Figure 3.9: Effect of deformation on Euler characteristic curves: plot (a), for $\alpha = 0.5$ and $\beta = 0.9$; plot (b), for $\alpha = 2$ and $\beta = 0.1$. In each plot, curves correspond to: (black) original field; deformed field under D1, (red) 'level' case, (dashed red) 'flow' case; deformed field under D2, (blue) 'level' case, (dashed blue) 'flow' case.

cific characteristics of the model, namely in relation to local variability and dependence ranges. We consider the Cauchy class, particularly useful for this purpose considering different combinations of parameter values (see Section 2.2.2). Here, we restrict the study to the Euler characteristic. Figure 3.8 displays average values of the Euler characteristic from 300 replicates, calculated for a dense set of u thresholds, for fixed $\alpha = 2$ and varying $\beta = 0.1$, 0.2 and 0.9, in plot (a), and for fixed $\beta = 0.2$ and varying $\alpha = 0.5$, 1 and 2, in plot (b). We observe that higher Hurst coefficient values correspond to more homogeneous structuring. The opposite happens when the fractal dimension increases,

which leads to larger variability in the Euler characteristic.

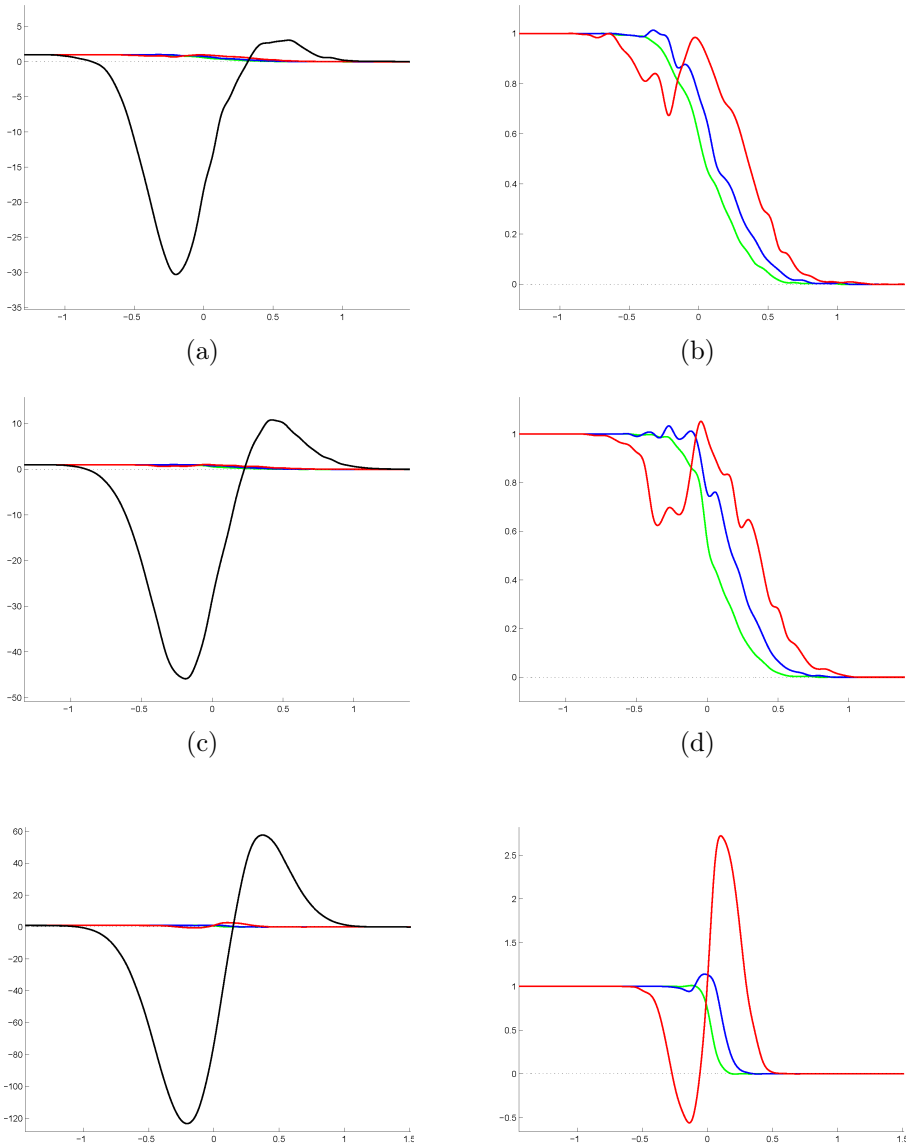


Figure 3.10: Effect of blurring on Euler characteristic curves, for fixed $\alpha = 1$ and varying β =0.1 (plots (a)-(b)), 0.2 (plots (c)-(d)), and 0.9 (plots (e)-(f)); plots (b)-(d)-(f) are just plots (a)-(c)-(e) rescaled without the curve for the original process for clearer visualization. In each plot, curves correspond to: (black) original field; blurred field with kernel (red) h_1, (blue) h_2, and (green) h_3.

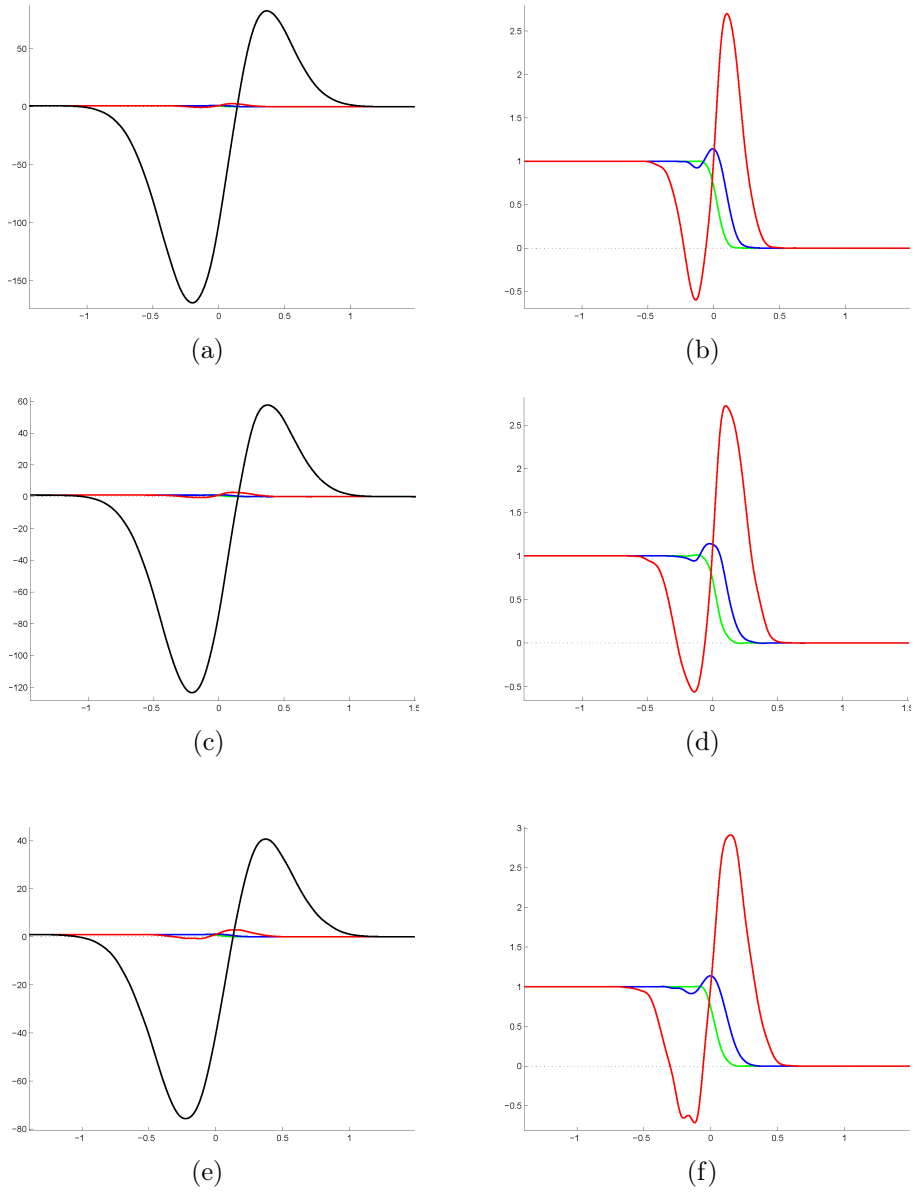


Figure 3.11: Effect of blurring on Euler characteristic curves, for fixed $\beta = 0.9$ and varying α =0.5 (plots (a)-(b)), 1 (plots (c)-(d)), and 2 (plots (e)-(f)); plots (b)-(d)-(f) are just plots (a)-(c)-(e) rescaled without the curve for the original process for clearer visualization. In each plot, curves correspond to: (black) original field; blurred field with kernel (red) h_1, (blue) h_2, and (green) h_3.

The deformation and blurring transformations previously considered have been applied in all cases, and average results for the Euler characteristic have been calculated. Significant differential effects observed in relation to the parameter specifications considered are shown, for some selected cases, in Figure 3.9 for deformation, and in Figures 3.10 and 3.11 for blurring. Plots (a) and (b) of Figure 3.9, respectively corresponding to the cases $(\alpha, \beta) = (0.5, 0.9)$ and $(\alpha, \beta) = (2, 0.1)$, show notably different magnitudes of relative increase and decrease of the Euler characteristic, more evident in the middle range of u thresholds where maxima and minima of the curves are attained. Particularly meaningful interpretations are derived for the blurring case. On one hand, structural differences associated with the larger scales of dependence observed for fixed α and varying β result in significantly different behavior between such cases in the Euler characteristic curves after blurring, as shown in Figure 3.10, in particular, for $\alpha = 1$ and $\beta = 0.1$, 0.2 and 0.9; here, plots (a), (c) and (e) include the corresponding curves for the original process, whilst plots (b), (d) and (f) represent only the curves after blurring transformation for clearer visualization in a more appropriate scale. In contrast, for fixed β the resulting curves of the Euler characteristic do not show significant differences in shape and scale when varying α (note, however, the scale differences in the curves for each original field), which is consistent with the fact that, for the regular type of kernel considered, blurring has a stronger smoothing effect at smaller scales, taking also into account the kernel variance values used. This is shown in Figure 3.11 for the cases $\beta = 0.9$ and $\alpha = 0.5$, 1 and 2.

Briefly, in this chapter we investigate significant aspects related to extremal analysis in relation to systemic structural characteristics in spatial modeling (see Chapter 7 for spatio-temporal case). In particular, the effect of common transformations such as spatial deformation and blurring, meaningful in many environmental applications, on geometrical properties of threshold exceedance sets —specifically, the Euler characteristic and the hypervolume— which have useful probabilistic interpretations in the context of risk assessment, is analyzed and illustrated through simulated examples. The study performed includes the evaluation of some differential aspects in relation to local variability and larger scale dependence properties of the processes subject to transformations.

Chapter 4

Spatial Threshold Exceedance Analysis Through Marked Point Processes

Indicators of recurrence, persistence, and in general, distribution patterns of extremal events defined by random field threshold exceedances provide relevant information on critical phenomena for risk assessment. Such indicators are directly related to geometrical properties describing the structure of the corresponding excursion sets. Given the intrinsic nature of the latter, marked point processes provide a natural approach to analyze distribution patterns of such extremal events in relation to specific characteristics of interest.

Here, based on simulations from a flexible model separating memory and fractality effects, we analyze the structure of threshold exceedances in terms of various second-order characteristics. In particular, we focus on the variations in size and distance heterogeneities in the components of excursion sets, as well as in clustering/inhibition patterns, depending on both the underlying model parameters and the threshold specifications.

4.1 Introduction and motivation

Analysis of environmental phenomena for risk assessment usually involves the construction of indicators related to structural characteristics of extremal events defined by exceedances over critical thresholds. Recurrence and persistence, among others, are examples of such characteristics, which provide information about the distribution patterns of

extremal events. Formally, these concepts are intimately related to the geometrical characteristics of the excursion sets defined by threshold exceedances over a given (bounded) domain. In particular, as mentioned in Section 2.2.3, useful mathematical descriptions can be given in terms of the Lipschitz-Killing curvatures or, equivalently, the intrinsic volumes of such sets (see Adler and Taylor 2007[4]). In Chapter 3, the effect of blurring and deformation transformations on the structure of a random field is studied in terms of the modifications implied on the Euler characteristic and hypervolume of threshold exceedance sets, depending on fractality and long-range dependence parameters, as well as on the threshold considered (see also Angulo and Madrid 2008[11], 2009a[12], 2009b[13] and 2010a[15]). As introduced in Section 2.2.3 and mentioned later in Section 3.3, under suitable conditions, the expected value of the Euler characteristic approximates, for high thresholds, the probability of exceedance in at least one point of the domain considered (Adler and Taylor 2007[4]), whilst the hypervolume provides an estimate of the probability of having an exceedance in a generic point of such a domain.

Given the fragmented nature of threshold exceedance sets, depending on the variation properties inherited by sample paths from the probabilistic structure of a random field and the threshold considered, marked point processes provide a powerful framework for the analysis of their structural properties. In fact, this approach can be exploited to help establishing the bridge between the construction and interpretation of risk indicators and the properties of the underlying random field generating critical events. More specifically, connected components of a threshold exceedance set can be treated as single, isolated events, with some geometrical properties such as size, contour length, relative intensity of dominant orientation, etc., being considered as possible marks of interest for complementary analysis of diverse forms of heterogeneity and anisotropy. Hence, a variety of marked point process characteristics can be used to describe some features of interest, in particular for risk assessment purposes (see Madrid *et al.* 2010b[82], and 2010c[83]).

The chapter is organized as follows. Section 4.2 is devoted to the development of the methodology to construct a family of point processes from the excursion sets of a random field and to study related properties. Then, in Section 4.3 we illustrate and discuss, based on simulation, the main characteristics of the observed point patterns obtained from the realizations of a Cauchy random field.

4.2 Methodology

Let X be a random field on $\mathbb{R}^d$. Considering any fixed bounded subdomain D and different thresholds u we can define a family of excursion sets A_u. The fragmented nature of the threshold exceedances allows to see the connected components as isolated events which can be depicted by a representative point. Thus, to each excursion set A_u we can associate a set of points $\{\boldsymbol{s}_1, \boldsymbol{s}_2, ..., \boldsymbol{s}_n\}$ which define a spatial point process. In this way, associated to X on D we have a family of spatial point processes, indexed by the threshold u. Here, we propose to represent each connected component by its centroid; if additionally the size l_i and orientation angle θ_i of each component are associated to its centroid as marks, marked point processes are considered. As mentioned before, the objective is to study the properties of such spatial point processes in relation to the structural characteristics of the original random field X.

Let $\{\boldsymbol{s}_1, \boldsymbol{s}_2, ..., \boldsymbol{s}_n\}$ be n points observed in a window $D \subset \mathbb{R}^d$, as described. In general, the first task in analyzing a point process is to find evidence against complete spatial randomness (CSR) by examining the intensity. We estimate the intensity function or intensity measure by the 'quadrat count' technique consisting of dividing the observation window in subregions ('quadrats') of equal area and counting the number of points falling into each quadrat. Based on this technique, a χ^2 test is carried out: Under the null hypothesis of CSR, the numbers of events within the subregions come from independent identically distributed Poisson random variables with the same expected value.

Then, second-order characteristics are computed. One widely used is the K-function (2.13), since for a stationary point process $\lambda K(r)$ defines the expected number of other points of the process within a distance r of a typical point of the process. Here we prefer using the L-function (2.14), given that this is a transformation of K that approximately stabilizes the variance of the estimator and provides an easier visual assessment of the graph. The L-function is estimated from the K-function as $\widehat{L}(r) = \sqrt{\widehat{K}(r)/\pi}$. In the homogeneous and isotropic case,

$$\widehat{K}(r) = \frac{1}{\widehat{\lambda}^2 |D|} \sum_{i=1}^{n} \sum_{j \neq i=1}^{n} \frac{I(\|\boldsymbol{s}_j - \boldsymbol{s}_i\| \leq r)}{\nu(D_{\boldsymbol{s}_j} \cap D_{\boldsymbol{s}_i})},$$

where $\widehat{\lambda} = \frac{n}{|D|}$ is the estimate of the intensity, whereas when the homogeneity hypothesis

is rejected,

$$\widehat{K'} = \frac{1}{|D|}\sum_{i}^{n}\sum_{j\neq i}^{n}\frac{I(\|\boldsymbol{s}_i - \boldsymbol{s}_j\| \leq r)}{\widehat{\lambda}(\boldsymbol{s}_i)\widehat{\lambda}(\boldsymbol{s}_j)\nu(D_{\boldsymbol{s}_j}\cap D_{\boldsymbol{s}_i})},$$

where the intensity function is estimated now by $\widehat{\lambda}(\boldsymbol{u}) = \sum_{i=1}^{n}\frac{k_\epsilon(\boldsymbol{u}-\boldsymbol{s}_i)}{\nu(D_{\boldsymbol{s}_j}\cap D_{\boldsymbol{s}_i})}$. We use the Epanechnikov kernel given by

$$k_\varepsilon(h) = \begin{cases} (1-h^2/\varepsilon^2)\frac{3}{4\varepsilon}, & -\varepsilon \leq h \leq \varepsilon, \\ 0, & \text{otherwise.} \end{cases}$$

It is known that the key aspect in kernel-based estimation is the choice of the bandwidth parameter, rather than the type of the kernel function used (see Stoyan and Stoyan 1994[111]). Here, we use the quantiles of the empirical distribution of the distances between points as a criterion for bandwidth selection. In both cases, $1/\nu(D_{\boldsymbol{s}_j}\cap D_{\boldsymbol{s}_i})$ is the translational edge correction with $D_{\boldsymbol{s}_j}$ meaning the window D translated by $\boldsymbol{s}_j$ and, if D is a rectangle of sides a and b, then $\nu(D_{\boldsymbol{s}_j}\cap D_{\boldsymbol{s}_i}) = |a-|x_i-x_j||\cdot|b-|y_i-y_j||$ (see Illian *et al.* 2008[65]). Other edge corrections have been introduced, but different choices do not produce significant differences.

Although the L-function is primarily intended for exploratory purposes, it is also possible to use it as a basis for statistical inference. In the context of hypothesis testing, the null hypothesis is that the data point pattern is a realization of CSR. The alternative hypothesis is that the data pattern is a realization of another, unspecified point process. Then, a Monte Carlo test is run based on simulations from the null hypothesis and envelopes are generated from the simulations.

To get information on anisotropy and use the appropriate estimator for the characteristics of marked point processes it is necessary to test the hypothesis of isotropy. However, here we consider a stronger test, for independence plus uniformity, proposed by Mateu (2000)[88], since as there it is commented, the possible spatial correlation makes it difficult to perform the first one. This test states that the orientations θ_i are independent for different components and uniformly distributed. The test uses a cumulative counterpart of $\kappa_f(r)$ (see (2.16) and (2.17)) defined through

$$\mathcal{K}_f(r) = \int_0^r k_f(u)g(u)2\pi u du = \int_0^r \frac{\ell_f^{(2)}(u)}{\lambda_0^2}2\pi u du, \quad r>0,$$

and its scaled version given by

$$L_f(r) = \sqrt{\mathcal{K}_f(r)/(\pi k_f(\infty))}, \quad r > 0.$$

The procedure consists of the following steps:

1. Fix the locations $\boldsymbol{s}_i$, $i = 1, ..., n$, of the given sample in the window D.

2. Generate new orientation marks by simulation independently and uniformly on $[0, \pi]$.

3. Repeat this procedure M times leading to M new marked point patterns of size n in D.

   (a) For each of them, calculate the estimators of $\mathcal{K}_d(r)$ and $L_d(r)$ through

   $$\widehat{\mathcal{K}}_d(r) = \frac{1}{\widehat{\lambda}_0^2} \sum_{0<\|\boldsymbol{s}_i - \boldsymbol{s}_j\| \le r} \frac{\min\{|\theta_i - \theta_j|, \pi - |\theta_i - \theta_j|\}}{\nu(D_{\boldsymbol{s}_j} \cap D_{\boldsymbol{s}_i})}, \tag{4.1}$$

   where $\widehat{\lambda}^2 = n(n-1)/|D|^2$ and

   $$\widehat{L}_d(r) = \sqrt{\frac{\widehat{\mathcal{K}}_d(r)}{\pi \frac{\pi}{4}}}. \tag{4.2}$$

   (b) Determine

   $$\widehat{L}_d^u(r) = \max_{i=1,...,M} \widehat{L}_{d,i}(r),$$
   $$\widehat{L}_d^l(r) = \min_{i=1,...,M} \widehat{L}_{d,i}(r).$$

4. If the empirical $\widehat{L}_d(r)$ for the originally given pattern is not completely inside the band given by $\widehat{L}_d^u(r)$ and $\widehat{L}_d^l(r)$, the hypothesis is rejected.

If additionally we want to take into account information about length marks, l_i, then the test can be slightly modified to state also that the pairs (l_i, θ_i) are independent for different components and, for each pair, l_i is independent of θ_i. To check this, in the step 2 new size marks are taken in addition from the empirical distribution of the size marks

of the sample. Then, in step 3, instead of (4.1) and (4.2), the next estimators are used:

$$\widehat{\mathcal{K}}_d^l(r) = \frac{1}{\widehat{\lambda}^2} \sum_{0<\|\boldsymbol{s}_i-\boldsymbol{s}_j\|\leq r} \frac{\min\{|\theta_i-\theta_j|, \pi - |\theta_i-\theta_j|\} l_i l_j}{\nu(D_{\boldsymbol{s}_j} \cap D_{\boldsymbol{s}_i})}$$

and

$$\widehat{L}_d^l(r) = \sqrt{\frac{\widehat{\mathcal{K}}_d^l(r)}{\pi \frac{\pi}{4} \bar{l}^2}},$$

where $\bar{l} = \sum_{i=1}^n l_i/n$.

As we just commented, this test is stronger than the one for anisotropy, and then when the empirical curves are outside the band curves we assume that there is anisotropy. For simplicity, in the following we refer to this test simply as the anisotropy test.

Considering the results of the test, the mark correlation function is used to detect spatial dependencies between the marks using the following non-negative test functions:

- $f(z_i, z_j) = l_i l_j$ if the marks are sizes, and
- $f(z_i, z_j) = d(\theta_i, \theta_j) = \min\{|\theta_i - \theta_j|, \pi - |\theta_i - \theta_j|)\}$ if the marks are angles of orientation

Then, in the isotropic case, the normalized mark correlation function (2.17) is estimated by

$$\widehat{\hbar}_f(r) = \frac{\widehat{\kappa}_f(r)}{\widehat{\kappa}_f(\infty)} = \frac{\widehat{\ell}_f^{(2)}(r)/\widehat{\ell}^{(2)}(r)}{\widehat{\kappa}_f(\infty)}, \tag{4.3}$$

where

$$\widehat{\ell}^{(2)}(r) = \frac{1}{2\pi r} \sum_{i=1}^n \sum_{i\neq j=1}^n \frac{k_\varepsilon(r - \|\boldsymbol{s}_i - \boldsymbol{s}_j\|)}{\nu(D_{\boldsymbol{s}_j} \cap D_{\boldsymbol{s}_i})} \tag{4.4}$$

is an estimate of the second-order product density (from which dividing by the square of intensity an estimation of the pair correlation function is obtained),

$$\widehat{\ell}_f^{(2)}(r) = \frac{1}{2\pi r} \sum_{i=1}^n \sum_{i\neq j=1}^n \frac{f(z_i, z_j) k_\varepsilon(r - \|\boldsymbol{s}_i - \boldsymbol{s}_j\|)}{\nu(D_{\boldsymbol{s}_j} \cap D_{\boldsymbol{s}_i})} \tag{4.5}$$

is an estimate of the f-product density, and $\widehat{\kappa}_f(\infty)$ is estimated by means of the mark distribution.

Then, substituting (4.4) and (4.5) in (4.3), an estimator of the normalized mark correlation function is expressed as

$$\widehat{\hbar}_f(r) = \frac{\sum_{i=1}^{n}\sum_{i\neq j=1}^{n} \dfrac{f(z_i,z_j)k_\varepsilon(r-\|\boldsymbol{s}_i-\boldsymbol{s}_j\|)}{\nu(D_{\boldsymbol{s}_j}\cap D_{\boldsymbol{s}_i})}}{\widehat{\kappa}_f(\infty)\sum_{i=1}^{n}\sum_{i\neq j=1}^{n} \dfrac{k_\varepsilon(r-\|\boldsymbol{s}_i-\boldsymbol{s}_j\|)}{\nu(D_{\boldsymbol{s}_j}\cap D_{\boldsymbol{s}_i})}}.$$

In the anisotropic case, the non-normalized version (2.16) is used. For (r,φ) being a point given in polar coordinates $(r\geq 0; 0\leq\varphi<\pi)$,

$$\widehat{\kappa}_f(r,\varphi) = \frac{\widehat{\ell}_f^{(2)}(r,\varphi)}{\widehat{\ell}^{(2)}(r,\varphi)},$$

with the estimator of the anisotropic second-order product density being formulated as

$$\widehat{\ell}^{(2)}(r,\varphi) = \frac{1}{2}\sum_{i=1}^{n}\sum_{j\neq i=1}^{n} \frac{\mathbf{K}(\boldsymbol{s}_j-\boldsymbol{s}_i,(r,\varphi)) + \mathbf{K}(\boldsymbol{s}_j-\boldsymbol{s}_i,(r,\varphi+\pi))}{\nu(D_{\boldsymbol{s}_j}\cap D_{\boldsymbol{s}_i})};$$

in a similar way, an estimation of the anisotropic f-product density is given by

$$\widehat{\ell}_f^{(2)}(r,\varphi) = \sum_{i=1}^{n}\sum_{i\neq j=1}^{n} \frac{\mathbf{K}(\boldsymbol{s}_j-\boldsymbol{s}_i,(r,\varphi))f(z_i,z_j)}{\nu(D_{\boldsymbol{s}_j}\cap D_{\boldsymbol{s}_i})},$$

where in both expressions

$$\mathbf{K}(\boldsymbol{s}_j-\boldsymbol{s}_i,(r,\varphi)) = \frac{1}{r}k_{\varepsilon_r}(\|\boldsymbol{s}_i-\boldsymbol{s}_j\|-r)k_{\varepsilon_\varphi}(\alpha(\boldsymbol{s}_i,\boldsymbol{s}_j)-\varphi),$$

with $k_{\varepsilon_r}(.)$ and $k_{\varepsilon_\varphi}(.)$ being two Epanechnikov kernels with suitable bandwidths, $\alpha(\boldsymbol{s}_i,\boldsymbol{s}_j)$ denoting the angle between the line from $\boldsymbol{s}_i$ to $\boldsymbol{s}_j$ and the x-axis, and $0\leq\varphi\leq\pi$. Here, the bandwidths are respectively selected from the empirical distribution corresponding to the distances and the angles.

Finally, when the processes are anisotropic, the main objective is the estimation of the directional distribution, i.e., the length-weighted orientation distribution of the segments, or rose of directions, which usually differs from the uniform distribution on $[0°, 180°]$. The

rose of directions (2.18) of the associated segment process is estimated by

$$\widehat{R}(\beta) = \frac{\sum_{i=1}^{n} I(0 \leq \theta_i \leq \beta) l_i}{\sum_{i=1}^{n} l_i}, \quad 0 \leq \beta < \pi.$$

4.3 Spatial structure analysis of exceedance patterns

Here, we again consider the Cauchy class with covariance function defined by $C(h) = \sigma^2(1 + h^\alpha)^{-\beta/\alpha}$, $\alpha \in (0, 2]$ and $\beta > 0$, with the same parameter values as in the previous chapter, that is, $\alpha = 0.5$, 1, and 2, and $\beta = 0.1$, 0.2, and 0.9, to guarantee long-range dependence (recall that for this it is necessary that $0 < \beta < 1$, see Section 2.2.2). Figure 4.1 shows simulated realizations on the square $[-100, 100] \in \mathbb{R}^2$ (using the same pseudo-random numbers) based on crossed combinations of such parameter values, with $\sigma^2 = 0.1$.

To analyze the effects of the parameters regarding the structural characteristics of the excursion sets, since the different values of the parameters define random fields with different ranges of variation, we determine the thresholds from the empirical distribution functions of the simulated realizations. Specifically, we consider several extreme percentiles (80th, 90th, 97th) which define the different thresholds. This practice is commonly applied in areas such as Ecology, Geophysics and Environmental Sciences, and it means that the size of the area at risk is fixed.

Some significant characteristics observed in the realizations in relation to the properties of fractality and dependence ranges shown in Figure 4.1, commented in Section 2.2.2, continue being visible when we consider the exceedances over high thresholds. In Figure 4.2 we display, based on the simulated realizations shown in Figure 4.1, the excursion sets for the thresholds corresponding to the 90th percentile, indicating by points the centroids of separate connected components, from which the point processes are defined. By row, from left to right, we can observe that the number of connected components decreases, increasing their sizes since the lower fractality leads to less fragmentation. By rows, from top to bottom, the ranges of dependence are reduced, increasing the levels of disaggregation and hence the number of components.

The cases corresponding to the combinations of the values of parameter (α, β) equal to $(0.5, 0.1)$, $(2, 0.1)$, $(0.5, 0.9)$ and $(2, 0.9)$ are particularly interesting, as it can be seen in the plots located at the corners of Figure 4.2 (also of Figure 4.1). The top-right $(2, 0.1)$

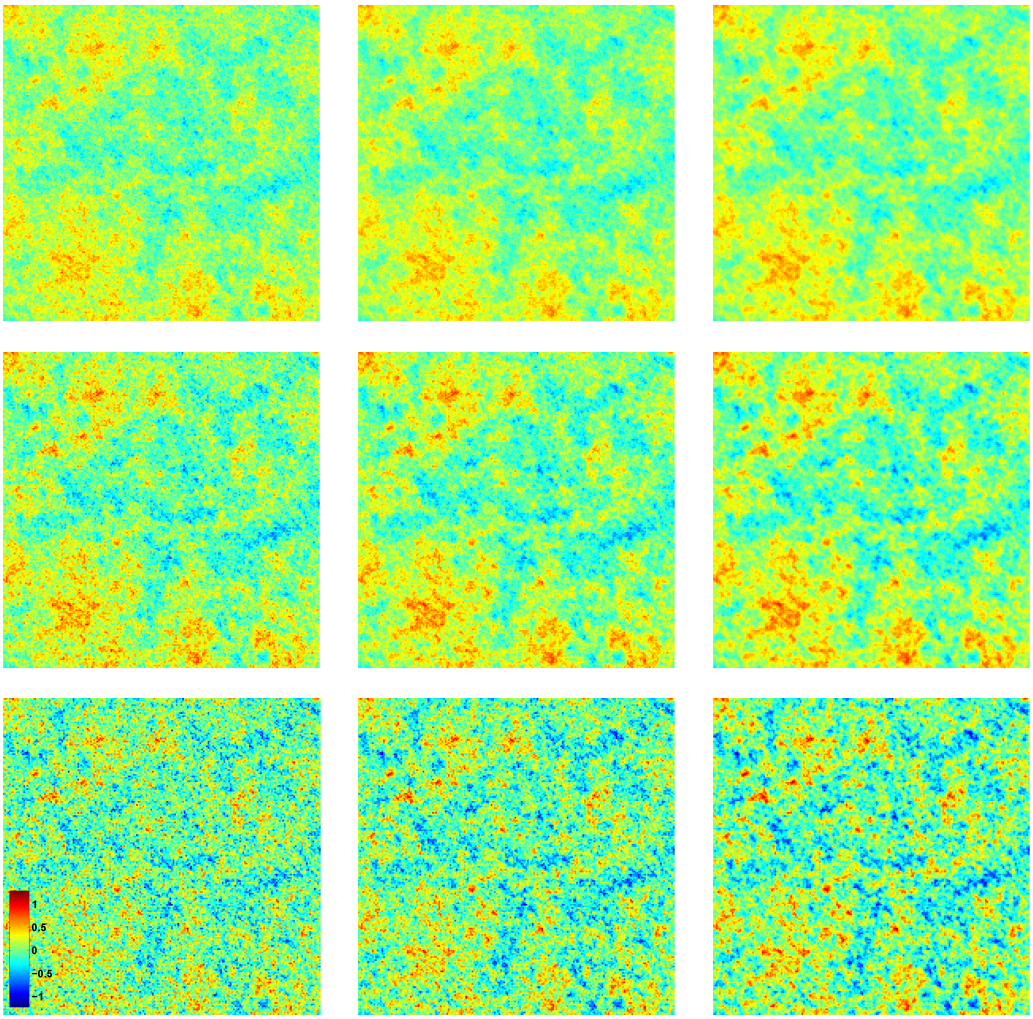


Figure 4.1: Simulated realizations of Cauchy class with $\alpha = 0.5, 1, 2$ (from left to right) and $\beta = 0.1, 0.2, 0.9$ (from top to bottom).

and bottom-left $(0.5, 0.9)$ corners correspond to the more distant cases in number of connected components, as well as their structuring in sizes and inter-distances. On the other hand, the opposite corners, $(0.5, 0.1)$ and $(2, 0.9)$, have a similar number of connected components since there is a trade off between fractality and Hurst effect, but there are significant differences in the structuring. Considering higher thresholds, the number and the size of the components diminish but the structuring behavior is similar, as it can be observed in Table 4.1, which contains the number of connected components for different crossed combinations of parameter values and specifications of the thresholds; see also

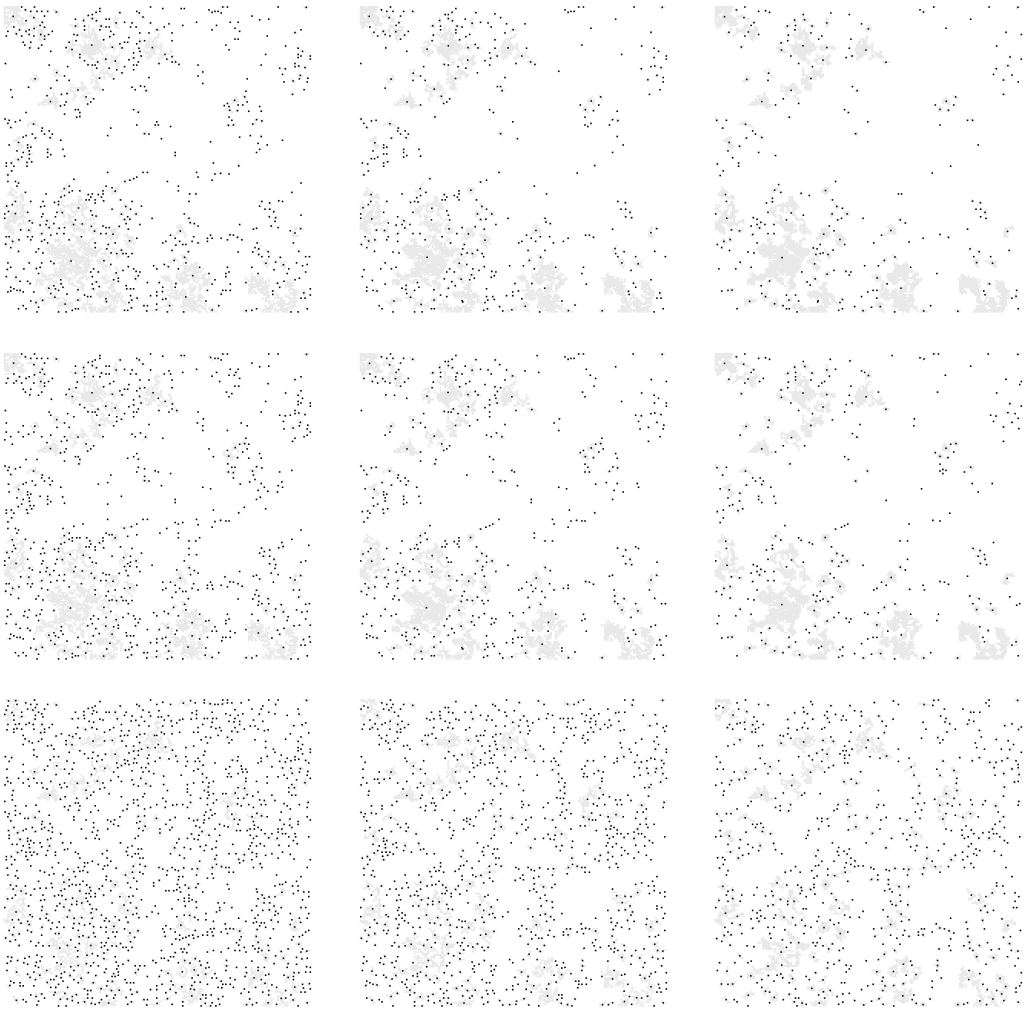


Figure 4.2: Excursion sets for threshold corresponding to 90th percentile, based on realizations of Figure 4.1. Black dots represent centroids of connected components.

Figures 4.3 and 4.4, which display the empirical cumulative distribution functions for the size mark of the point processes corresponding to connected components of the threshold exceedance sets for the 80th and 97th percentiles, respectively, for these four cases.

In Figures 4.5 and 4.6 we illustrate the distributions of events by quadrats for the mentioned cases and thresholds corresponding to the 80th and 97th percentiles, respectively. Table 4.2 collects the results of χ^2 homogeneity tests based in quadrat counting for the excursion sets corresponding to different thresholds, emphasizing in bold when there is no evidence to reject the null hypothesis at 5% of significance. We observe that increasing the threshold, as well as the range of dependence, both produce a rupture of

80th Percentile			
Parameters	$\alpha = 0.5$ $(D = 2.75)$	$\alpha = 1$ $(D = 2.5)$	$\alpha = 2$ $(D = 2)$
$\beta = 0.1$ $(H = 0.95)$	850	548	343
$\beta = 0.2$ $(H = 0.9)$	949	614	385
$\beta = 0.9$ $(H = 0.55)$	1628	1204	833

90th Percentile			
Parameters	$\alpha = 0.5$ $(D = 2.75)$	$\alpha = 1$ $(D = 2.5)$	$\alpha = 2$ $(D = 2)$
$\beta = 0.1$ $(H = 0.95)$	693	417	233
$\beta = 0.2$ $(H = 0.9)$	785	489	318
$\beta = 0.9$ $(H = 0.55)$	1437	1080	744

97th Percentile			
Parameters	$\alpha = 0.5$ $(D = 2.75)$	$\alpha = 1$ $(D = 2.5)$	$\alpha = 2$ $(D = 2)$
$\beta = 0.1$ $(H = 0.95)$	384	228	142
$\beta = 0.2$ $(H = 0.9)$	396	277	181
$\beta = 0.9$ $(H = 0.55)$	759	579	$3M$

Table 4.1: Number of connected components for different crossed combinations of parameter values and threshold specifications.

80th Percentile			
Parameters	$\alpha = 0.5$ $(D = 2.75)$	$\alpha = 1$ $(D = 2.5)$	$\alpha = 2$ $(D = 2)$
$\beta = 0.1$ $(H = 0.95)$	0.01203	0.01886	0.00347
$\beta = 0.2$ $(H = 0.9)$	0.02520	**0.09130**	**0.09134**
$\beta = 0.9$ $(H = 0.55)$	**0.56094**	**0.63398**	**0.88333**

90th Percentile			
Parameters	$\alpha = 0.5$ $(D = 2.75)$	$\alpha = 1$ $(D = 2.5)$	$\alpha = 2$ $(D = 2)$
$\beta = 0.1$ $(H = 0.95)$	$1.85e^{-23}$	$3.11e^{-23}$	$1.06e^{-15}$
$\beta = 0.2$ $(H = 0.9)$	$2.23e^{-14}$	$5.73e^{-14}$	$2.27e^{-11}$
$\beta = 0.9$ $(H = 0.55)$	**0.27894**	**0.18329**	**0.1305**

97th Percentile			
Parameters	$\alpha = 0.5$ $(D = 2.75)$	$\alpha = 1$ $(D = 2.5)$	$\alpha = 2$ $(D = 2)$
$\beta = 0.1$ $(H = 0.95)$	$2.14e^{-56}$	$3.54e^{-38}$	$1.71e^{-22}$
$\beta = 0.2$ $(H = 0.9)$	$1.24e^{-53}$	$4.46e^{-38}$	$1.90e^{-24}$
$\beta = 0.9$ $(H = 0.55)$	$4.71e^{-14}$	$3.98e^{-15}$	$1.31e^{-7}$

Table 4.2: Results of Pearson's χ^2 test based on quadrat counts for the excursion sets corresponding to different thresholds.

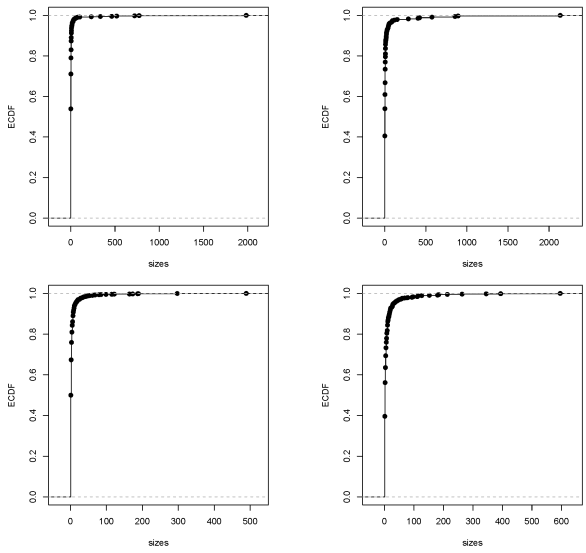


Figure 4.3: Empirical cumulative distribution function for the size mark of the point processes defined by centroids of connected components of the threshold exceedance sets corresponding to 80th percentile and parameter values $\alpha = 0.5, 2$ (from left to right) and $\beta = 0.1, 0.9$ (from top to bottom).

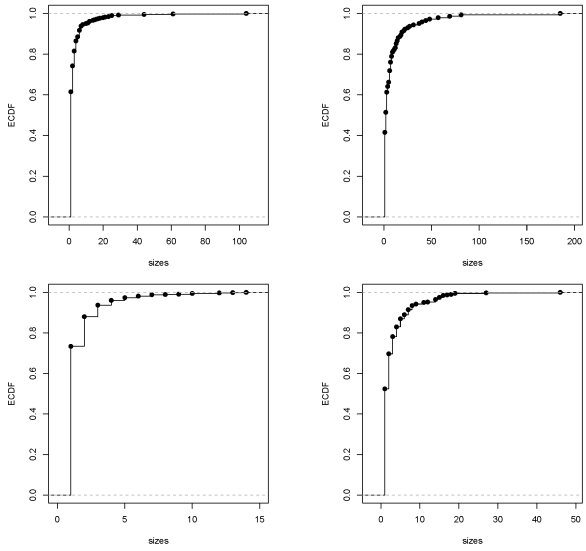


Figure 4.4: Empirical cumulative distribution function for the size mark of the point processes defined by centroids of connected components of the threshold exceedance sets corresponding to 97th percentile and parameter values $\alpha = 0.5, 2$ (from left to right) and $\beta = 0.1, 0.9$ (from top to bottom).

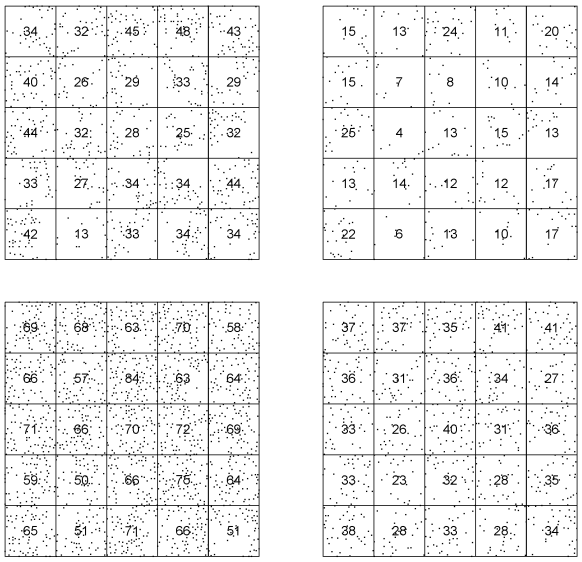


Figure 4.5: Distribution of events by quadrats for the excursion sets corresponding to 80th percentile and parameter values $\alpha = 0.5, 2$ (from left to right) and $\beta = 0.1, 0.9$ (from top to bottom).

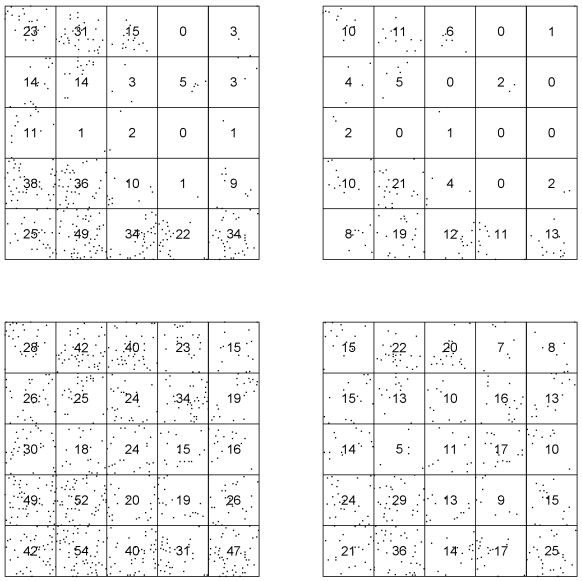


Figure 4.6: Distribution of events by quadrats for the excursion sets corresponding to 97th percentile and parameter values $\alpha = 0.5, 2$ (from left to right) and $\beta = 0.1, 0.9$ (from top to bottom).

the homogeneity.

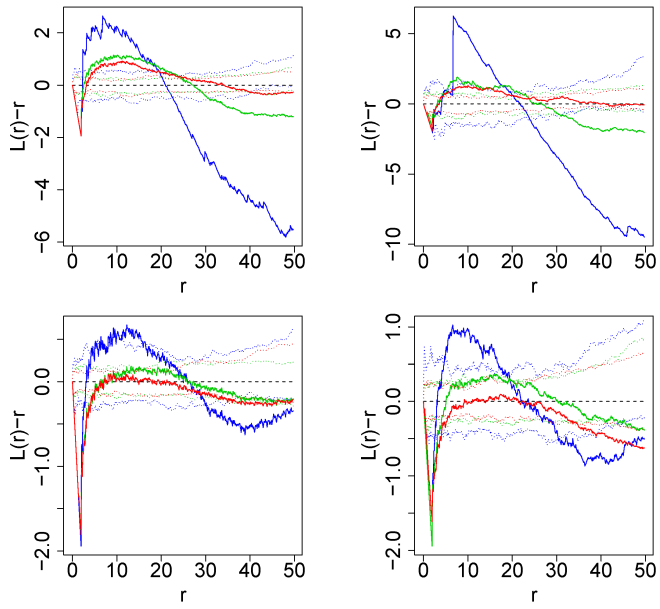


Figure 4.7: $L(r) - r$ function for the point processes defined by centroids of connected components, for the excursion sets corresponding to percentiles 80th (red), 90th (green) and 97th (blue), based on the realizations of Cauchy class with parameter values $\alpha = 0.5, 2$ (from left to right) and $\beta = 0.1, 0.9$ (from top to bottom).

Taking into account the possible homogeneity or inhomogeneity, the L-function is computed for the different point processes. In Figure 4.7 we display the values $L(r) - r$, to make easier a visual assessment comparing with respect to the theoretical value 0. The observed common pattern in the different curves shows an inhibition behavior at short distances which can be explained by the construction of the processes: We are considering points defined by the centroids of connected components, and the distances are measured between them (not between the edges of components), thus the minimum inter-point distance will be $r_0 = 2$ (in pixel units) and it will increase with the sizes of the components. Comparing now the curves, we can see that considering more restrictive thresholds the levels of aggregation or inhibition both increase; a similar effect occurs when we consider values of β corresponding to stronger ranges of dependence.

The next step has been to carry out the test of anisotropy. When we only consider the orientation mark, there is no evidence, in any of the cases studied, to reject the null hypothesis, see Figures 4.8 and 4.9. That is, we can assume that the orientations are

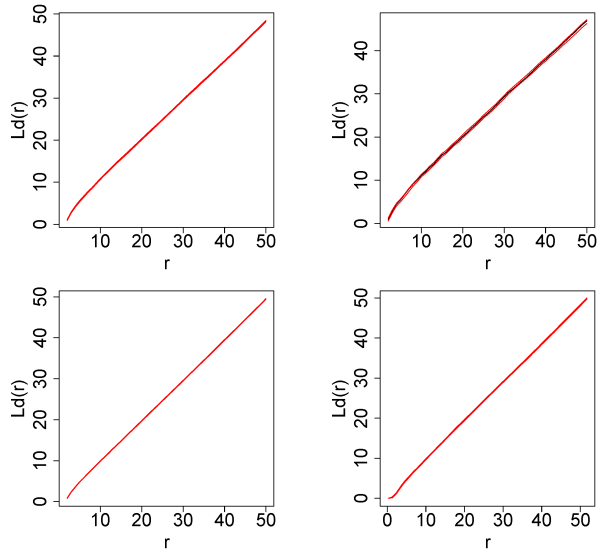


Figure 4.8: Anisotropy test using orientation mark for the point processes defined by centroids of connected components of the threshold exceedance sets corresponding to 80th percentile and parameter values $\alpha = 0.5, 2$ (from left to right) and $\beta = 0.1, 0.9$ (from top to bottom).

independent of the locations and that they are uniformly distributed. This agrees with Figures 4.10 and 4.11, which show the empirical cumulative distribution functions for the orientation mark. Note that, except for certain jumps, the curves suggest a uniform distribution. The occurrence of such jumps is explained by the fact that two pixels are considered to be connected it they have a common side or a common corner; hence, in particular, the angles 0°, 45°, 90° and 135° are more likely to occur due to the high relative number of two-pixel size components, which is reflected by the empirical distribution. The isotropic correlation functions for the 80th and 90th percentiles are represented in Figures 4.12 and 4.13, these showing that there is no significant correlation between the orientation marks in any case and independently of the distance, since the curves are very close to the value 1 in all cases.

When additionally we consider information of the size mark, the null hypothesis of isotropy is rejected in some cases, see Figures 4.14 and 4.15. In the cases corresponding to lower thresholds the empirical curves are outside the bands, whilst for higher thresholds the results depend on the parameter values, with those corresponding to long-range dependence being the most significant. This seems to indicate that, from an isotropic

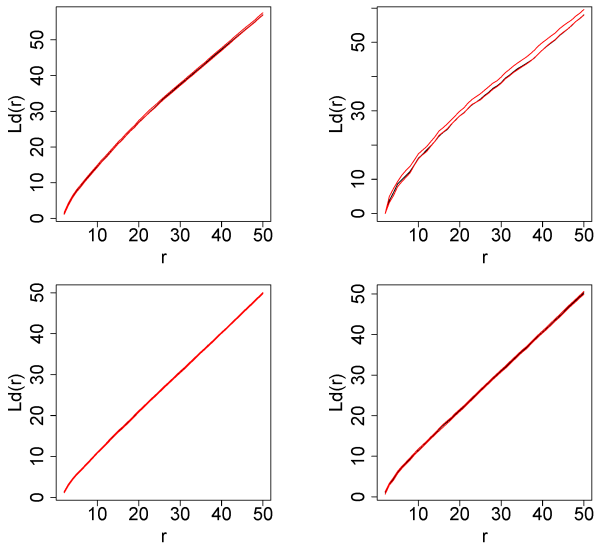


Figure 4.9: Anisotropy test using orientation mark for the point processes defined by centroids of connected components of the threshold exceedance sets corresponding to 97th percentile and parameter values $\alpha = 0.5, 2$ (from left to right) and $\beta = 0.1, 0.9$ (from top to bottom).

realization, by gradually increasing the threshold we can get anisotropic point patterns, and finally obtain isotropic point patterns again. The parameter values influence the magnitude of the threshold at which the changes of behavior occur. In spite of the fact that in some cases the tests suggest isotropy in almost all distances, we choose to represent the anisotropic version in all cases. Comparing Figures 4.14 and 4.16, for a threshold corresponding to the 80th percentile, we can observe that similarities between the plots in the first one are also present in those included in the second one. Furthermore, the peaks within the range 150° - 180° in the cases with $\beta = 0.1$ for the anisotropic mark correlation indicate that there is significant correlation between the sizes for such orientations, which can be related to the presence of anisotropy according to the test. Now, comparing Figures 4.15 and 4.17, derived for the 97th percentile, we observe that the plots at the top-right corner, corresponding to $\alpha = 2$ and $\beta = 0.1$, are particularly significant since the test indicate isotropy except for short distances. This is reflected in the representation of the anisotropic mark correlation, which is flat except for the peak appearing at distances less than 20 and orientations close to zero.

Given that we have supposed that the point processes are anisotropic only when the

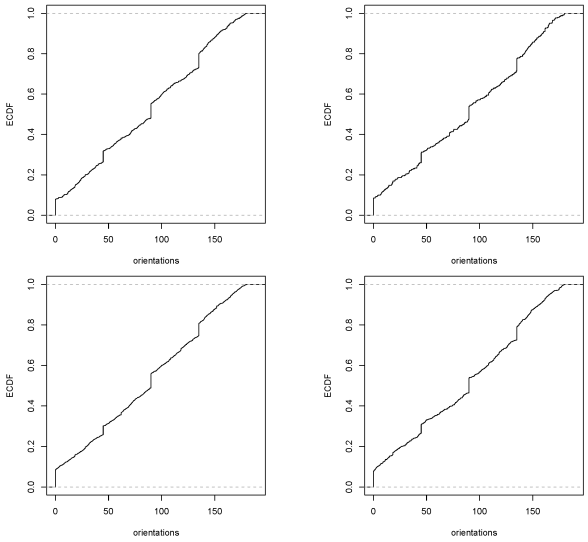


Figure 4.10: Empirical cumulative distribution function for the orientation mark of the point processes defined by centroids of connected components of the threshold exceedance sets corresponding to 80th percentile and parameter values $\alpha = 0.5, 2$ (from left to right) and $\beta = 0.1, 0.9$ (from top to bottom).

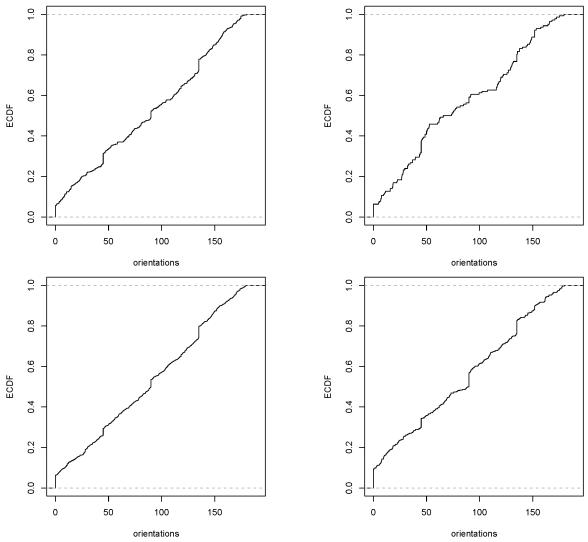


Figure 4.11: Empirical cumulative distribution function for the orientation mark of the point processes defined by centroids of connected components of the threshold exceedance sets corresponding to 97th percentile and parameter values $\alpha = 0.5, 2$ (from left to right) and $\beta = 0.1, 0.9$ (from top to bottom).

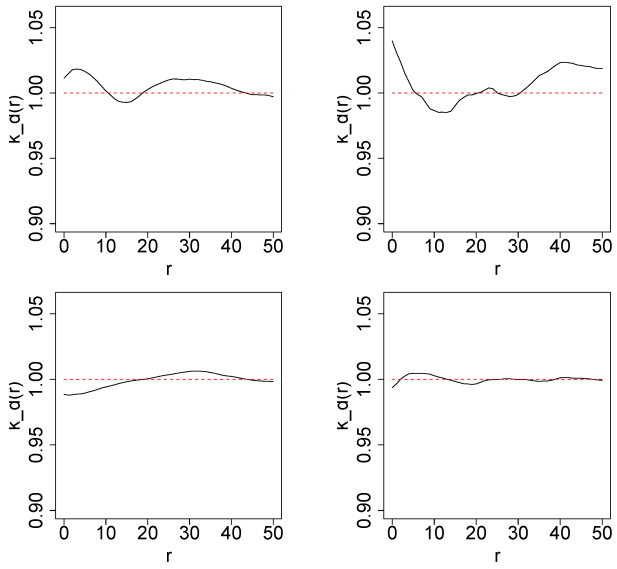


Figure 4.12: Orientation correlation function for the point processes defined by centroids of connected components of the threshold exceedance sets corresponding to 80th percentile and parameter values $\alpha = 0.5, 2$ (from left to right) and $\beta = 0.1, 0.9$ (from top to bottom).

size and orientation marks are considered, we compute the length-weighted orientation distributions of the segments, that is, the rose of directions with weights the sizes. Figures 4.18 and 4.19 show the rose of directions corresponding to the 80th and 97th percentiles, respectively. We observe that in the cases corresponding to the 80th percentile and $\beta = 0.1$ (top, left and right), there is a predominant angle around $100° - 110°$, whilst in the remaining cases the curves are very similar to the uniform distribution. This is not surprising since in these cases the empirical curves of the anisotropy test are very near the bands, but we should remind that this test is stronger than the one specifically addressed to assess anisotropy, then these point processes may actually be isotropic.

In summary, we have focused on the analysis of structural changes in marked point processes based on excursion sets corresponding to different thresholds, depending on fractality and long-range dependence properties of the generating random field. Specifically, for thresholds corresponding to various high percentiles in the empirical distribution of sample-path values, we analyze size heterogeneities of isolated events defined by connected components, distance ranges where the spatial distribution of the centroids

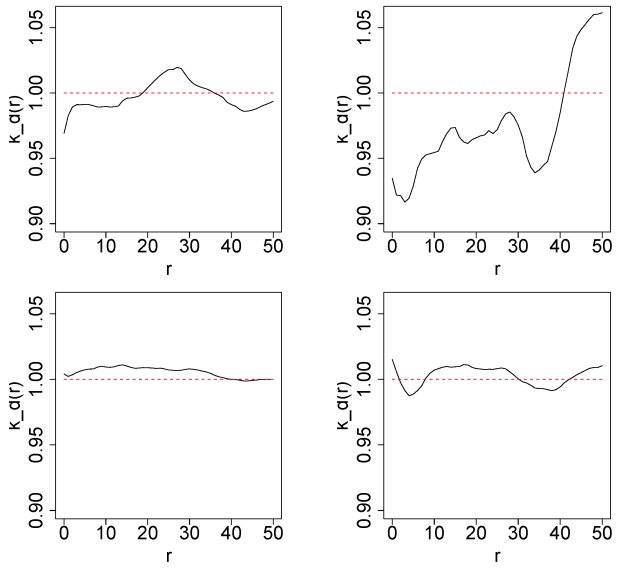


Figure 4.13: Orientation correlation function for the point processes defined by centroids of connected components of the threshold exceedance sets corresponding to 97th percentile and parameter values $\alpha = 0.5, 2$ (from left to right) and $\beta = 0.1, 0.9$ (from top to bottom).

representing such components display clustering/inhibition patterns, as well as the level of anisotropy. The results show significant differences, depending on the scenario determined by the model parameters, which have interesting interpretations related to the underlying random field probabilistic structure as well as in terms of risk indicators. It is shown that both an increase in the threshold and a decrease in the dependence range make the pattern inhomogeneous with more variability in the sizes and distances among components. In addition, the degree of clustering or inhibition is notoriously increased. Finally, we have seen that increasing the threshold yields a variation in the degree of local anisotropy of the point patterns associated to the corresponding threshold exceedance sets.

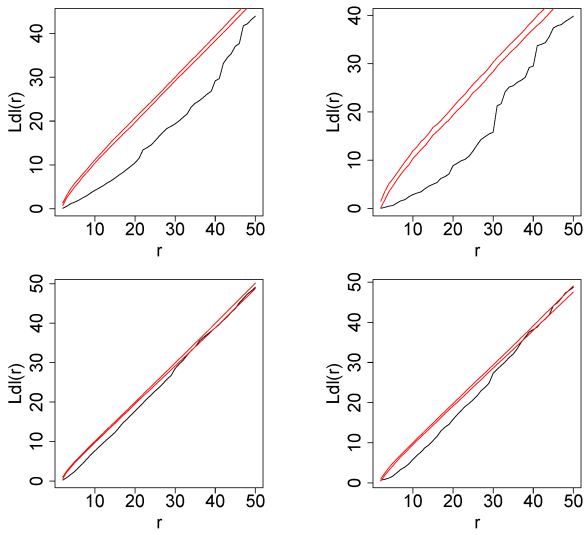


Figure 4.14: Anisotropy test using orientation and size marks for the point processes defined by centroids of connected components of the threshold exceedance sets corresponding to 80th percentile and parameter values $\alpha = 0.5, 2$ (from left to right) and $\beta = 0.1, 0.9$ (from top to bottom).

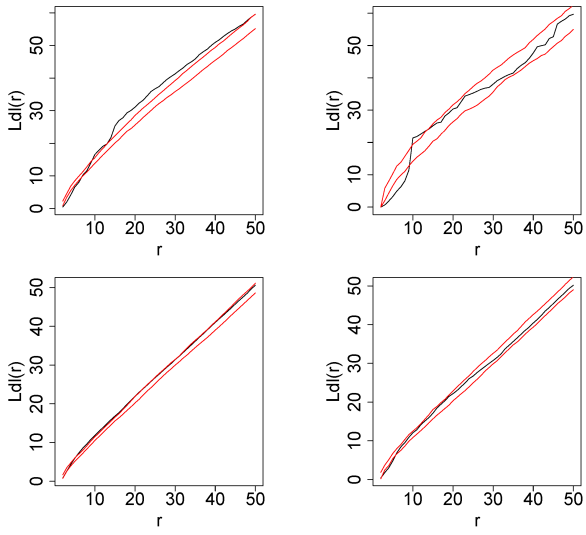


Figure 4.15: Anisotropy test using orientation and size marks for the point processes defined by centroids of connected components of the threshold exceedance sets corresponding to 97th percentile and parameter values $\alpha = 0.5, 2$ (from left to right) and $\beta = 0.1, 0.9$ (from top to bottom).

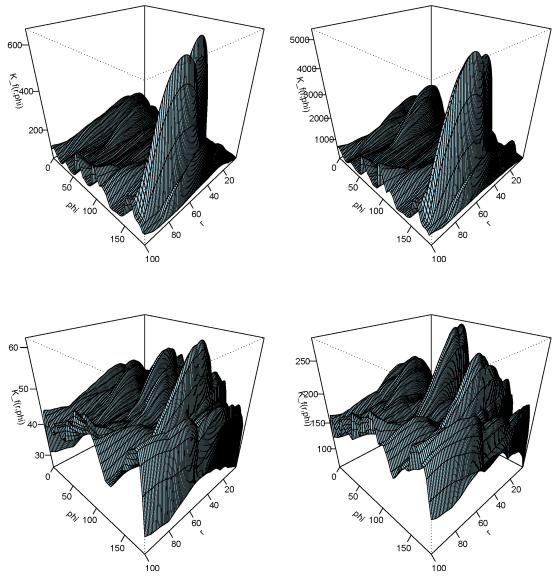


Figure 4.16: Anisotropic mark correlation function for the point processes defined by centroids of connected components of the threshold exceedance sets corresponding to 80th percentile and parameter values $\alpha = 0.5, 2$ (from left to right) and $\beta = 0.1, 0.9$ (from top to bottom).

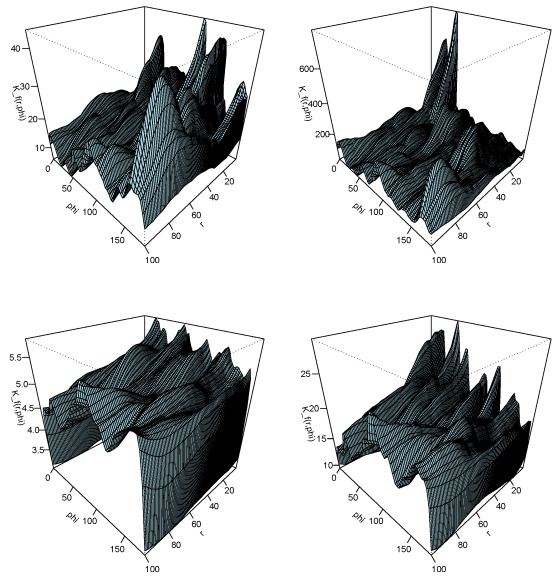


Figure 4.17: Anisotropy mark correlation function for the point processes defined by centroids of connected components of the threshold exceedance sets corresponding to 80th percentile and parameter values $\alpha = 0.5, 2$ (from left to right) and $\beta = 0.1, 0.9$ (from top to bottom).

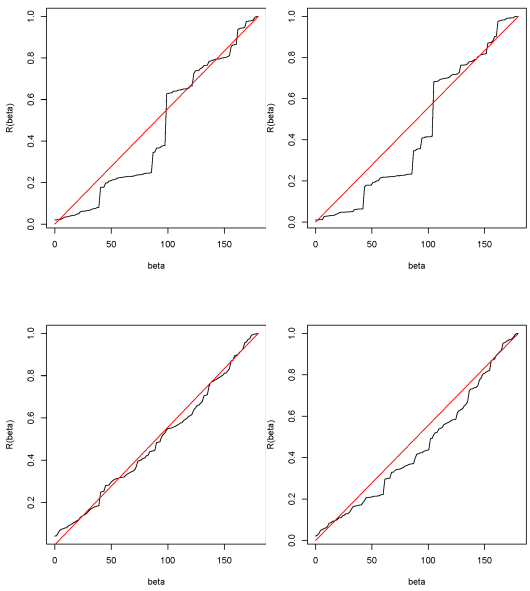


Figure 4.18: Rose of directions for the point processes defined by centroids of connected components of the threshold exceedance sets corresponding to 80th percentile and parameter values $\alpha = 0.5, 2$ (from left to right) and $\beta = 0.1, 0.9$ (from top to bottom).

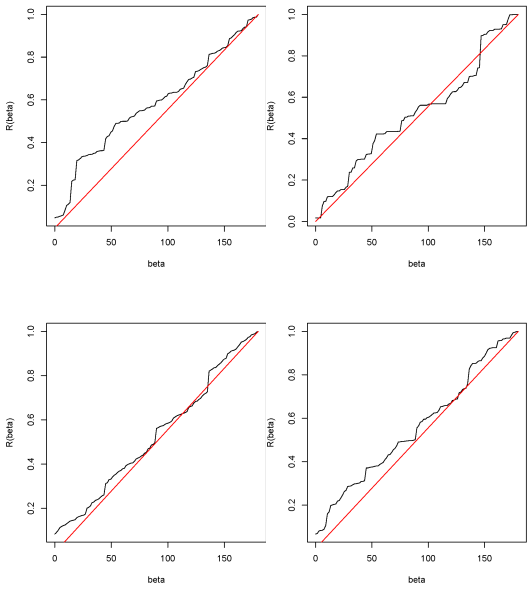


Figure 4.19: Rose of directions for the point processes defined by centroids of connected components of the threshold exceedance sets corresponding to 97th percentile and parameter values $\alpha = 0.5, 2$ (from left to right) and $\beta = 0.1, 0.9$ (from top to bottom).

Chapter 5

Wavelet-Based Multiscale Intermittency Analysis

Essential aspects intrinsic to the concept of intermittency such as 'scales of variation' and 'localization', and related interactive dynamics, have led to the use of wavelets as a suitable functional tool for technical analysis. Here the study is focused on the analysis of significant aspects related to random field deformations, meaningful both from physical considerations and for methodological purposes, and structural implications of such transformations on intermittency, as well as concerning wavelet-related methods.

5.1 Introduction

Intermittency, generally understood as pseudo-periodic occurrence of high level or variation episodes within a certain regular behavior, is considered a phenomenon of interest in very diverse fields of applications (e.g. Seismology, Turbulence, Hydrology, Astronomy, Finance, Insurance, Epidemiology, etc.), related to which a variety of specific formalizations and measures have been derived. Structural characteristics associated with such effect, in relation to the underlying generating process, often constitute a primary objective in environmental studies, as they provide relevant information for detection and prediction of critical events and for risk assessment.

There are different manifestations of intermittency, which have lead to various interpretations and formal definitions of this concept. Essentially, they are related to the

genesis and, as a consequence, the nature of its effects, as a certain form of heterogeneous behavior. Depending on the domain where such heterogeneities occur, it is usual to distinguish between 'isolated' type of intermittency, for heterogeneities in the spatial/temporal domain, and 'non-isolated' type of intermittency, associated to heterogeneous scaling, closer to the concept of multifractality (see, for example, Nicolleau and Vassilicos, 1999[93]).

A common feature, then, to these approaches is the heterogeneity in the energy distribution of a signal over space/time and/or scales. This fact justifies the use of wavelet functions and related tools to analyze intermittency. In particular, we here use, for the analysis of the effect of deformation and the structure of intermittency, a scale-dependent indicator of heterogeneity introduced in Farge *et al.* (1990)[46].

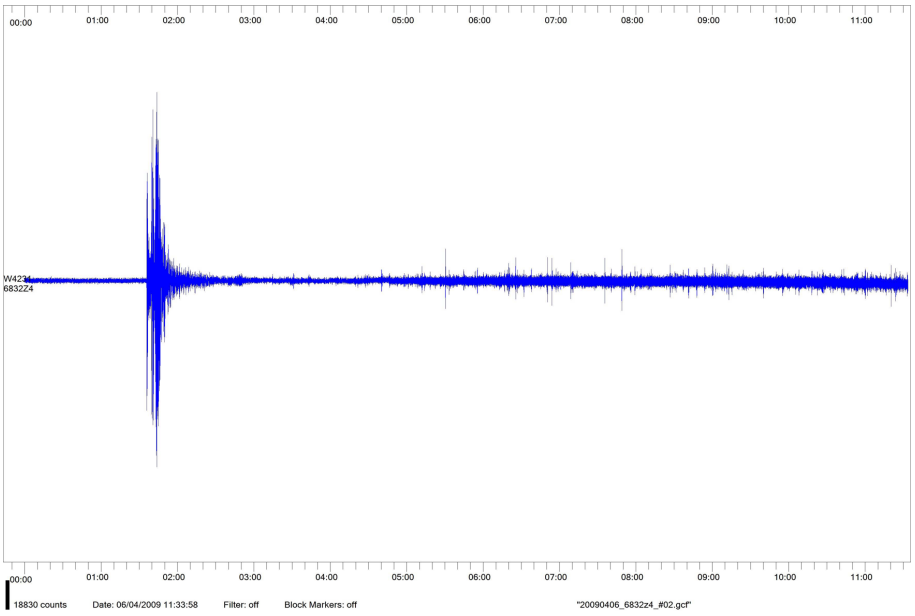


Figure 5.1: Signal registered in Granada of L'Aquila earthquake (April 6, 2009).

For illustration, we consider the following examples of real and simulated signals. Figure 5.1 shows the registered signal in Granada corresponding to the earthquake happened on the day April 6, 2009, with epicenter near L'Aquila (Italy), with magnitude of 6.3 on the Ritcher scale. Figure 5.2 displays four subperiods with different behaviors: plot (a) a regular period, and plots (b), (c) and (d) three activity periods. Note that each period is separated in the time from the other ones, however there is a significant

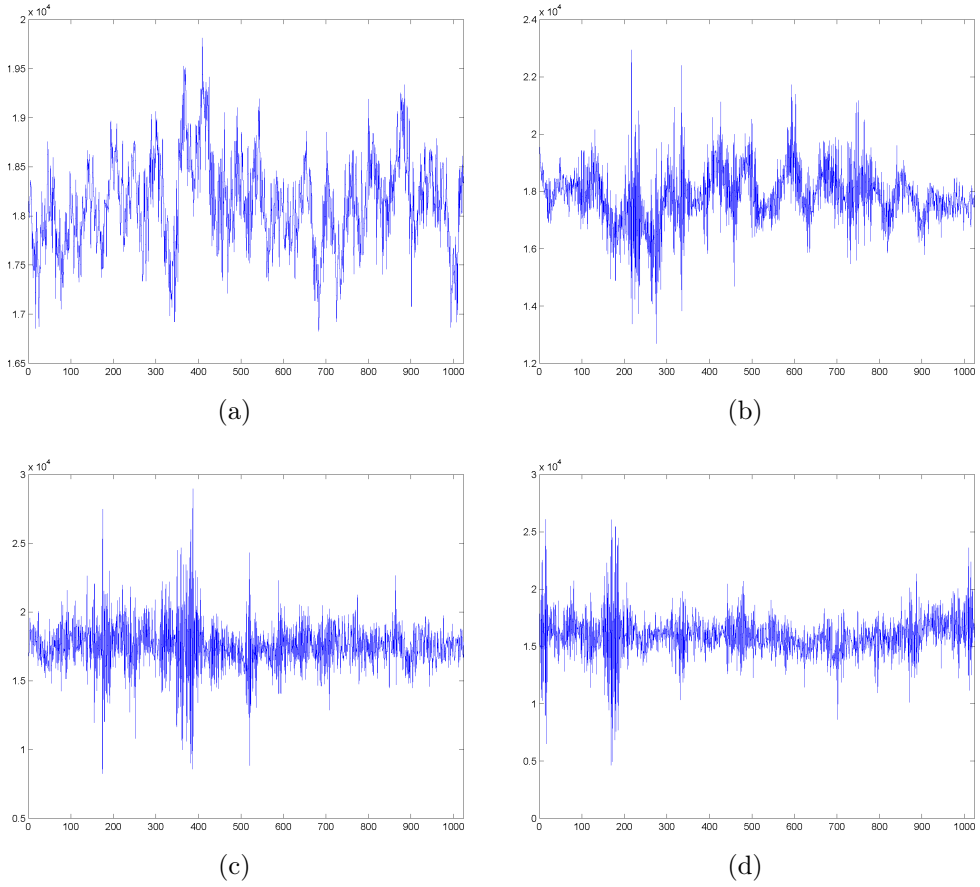
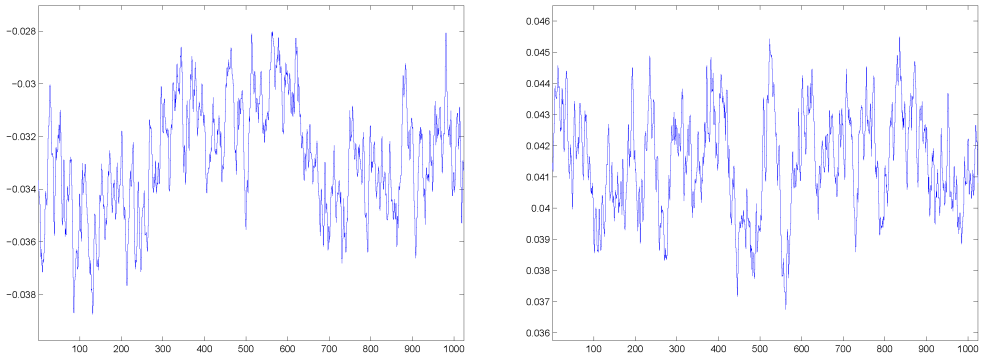


Figure 5.2: A regular phase and three activity periods of L'Aquila earthquake showed in Figure 5.1, beginning at: (a) 01:29:00 (b) 03:31:10 (c) 09:00:18 (d) 11:24:03.

similarity between the plots (c) and (d). In Figure 5.3, two time series corresponding to different locations, generated from a certain multifractal spatio-temporal model are plotted. Finally, a simulated realization of an ARMA(1,1) model with Cauchy white noise, which exhibits isolated-type of intermittent behavior, is shown in Figure 5.4.

5.2 Intermittency analysis: wavelet approaches

Wavelet-based techniques constitute a powerful tool for studying intermittencies in signals, and have been applied in many studies, in particular in relation to environmental areas. An interesting related insight was given, for example, by Farge *et al.* (1990)[46]. Let $x(t)$ be a signal. As we saw in Section 2.4.2, the Continuous Wavelet Transform is



(a) (b)

Figure 5.3: Two time series generated at different locations from a heterogeneous space-time model.

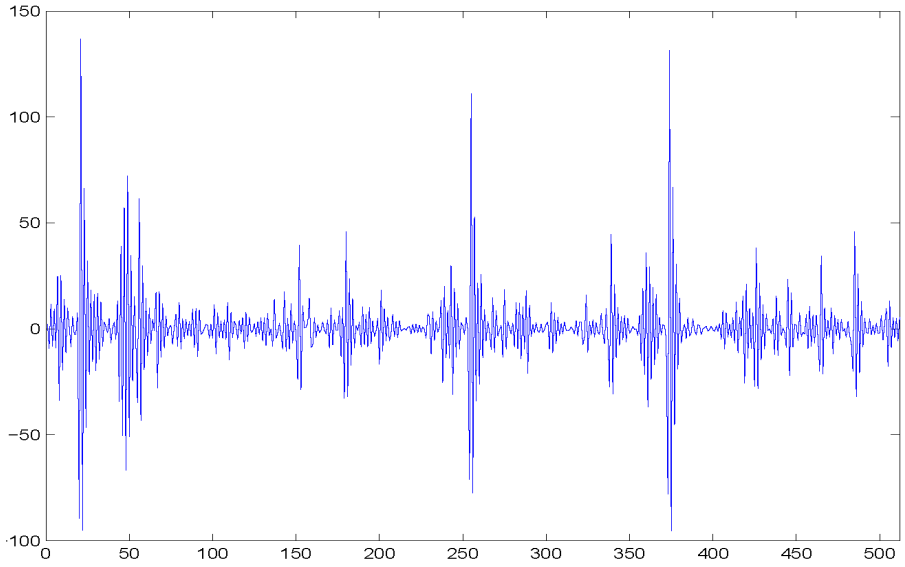


Figure 5.4: Simulated realization of ARMA(1,1) model, with $\phi = 0.7$, $\theta = 0.9$, and Cauchy white noise.

obtained through convolution between $x(t)$ and all the possible versions, scaled and translated, of a basic wavelet function ψ. Using the terms in Meneveau (1991)[89], a wavelet transform coefficient $W(a, b)$ can then be seen as a measure of the 'correlation' between the signal $x(t)$ and the wavelet function $\psi(t)$ at scale a and position translation b. The total energy of $x(t)$ can be decomposed in terms of the wavelet coefficient energies, as

$$E_x = \int_{-\infty}^{\infty} x(t)^2 dt = \int_0^{\infty} E_w(\lambda) d\lambda = \int_0^{\infty} \left(\frac{1}{2\pi C_\psi^2} \int_{-\infty}^{\infty} |W_x(a(\lambda), b)|^2 db \right) d\lambda$$
$$= 2\pi \int_0^{\infty} \frac{E_w(a)}{a^2} da = \frac{1}{C_\psi} \int_0^{\infty} \int_{-\infty}^{\infty} \frac{|W_x(a,b)|^2}{a^2} dbda,$$

where $\lambda = 2\pi/a$. That is, a density for the scale-time distribution of the energy density of the signal $x(t)$, through scales a and locations b, is obtained from the square of $W_x(a,b)$, named 'scalogram'. Taking into account that intermittency occurs when the energy is not evenly distributed, Farge *et al.* (1990)[46] introduced the so-called 'local intermittency measure' (LIM), defined by normalization of the scalogram with respect to the average local energy observed within each scale:

$$LIM_x(a,b) = \frac{|W_x(a,b)|^2}{\mathsf{Ave}_b \left[|W_x(a,b)|^2 \right]}. \tag{5.1}$$

If $LIM_x(a,b) = 1$ for all a and b, then the energy of the function is equally distributed in time within each scale and the signal does not exhibit intermittent behavior. On the contrary, a coefficient $LIM_x(a,b) > 1$ indicates that this given scale-location pair (a,b) contributes significantly, in relative terms, to the average. Therefore, the $LIM_x(a,b)$ map can be seen as the fundamental analysis of energy information which can be further explored and processed to look for certain patterns or characteristics related to the concept of intermittency (inter-scale coherence, recurrence, persistence, event distribution patterns, etc.).

An immediate indicator of intermittency is represented by the temporal average of $LIM_x^2(a,b)$, which defines the flatness $F_x(a)$ of the wavelet coefficients within each scale a (see Meneveau 1991[89]):

$$F_x(a) = \mathsf{Ave}_b \left[LIM_x^2(a,b) \right] = \frac{\mathsf{Ave}_b \left[|W_x(a,b)|^4 \right]}{\left(\mathsf{Ave}_b \left[|W_x(a,b)|^2 \right] \right)^2}.$$

The flatness is then a scale dependent measure of the kurtosis of the input signal. A relatively 'high' $F_x(a)$ value is taken as a possible indication of intermittency at scale a. Locally, the condition $LIM_x^2(a,b) > 3$ can be used as a basic criterion to identify those events contributing to departure from Gaussianity in the distribution of wavelet

coefficients. Nicolleau and Vassilicos (1999)[93] explain that often a signal is interpreted as intermittent when $F_x(\cdot)$ increases for decreasing a since an intermittent signal displays activity over only a fraction of time, and this portion decreases with the scale a under consideration. A disadvantage of $F_x(\cdot)$ is that it does not discriminate intermittencies of isolated from non-isolated type.

A key question in the analysis through wavelets is the choice of the mother wavelet, since the main task of this type of study is to retain certain characteristics of interest of a signal. Therefore, the wavelet function should be selected depending on the features to be detected. For example, it is known that Haar wavelet is appropriate to detect a sustained change in signal level, while Morlet wavelet is more appropriate to seek concentration in time-frequency of signal energy (see Guilliam *et al.* 2000[59]).

Figure 5.5 illustrates, for the seismic segment displayed in Figure 5.2 (b), the analysis of intermittency using the above quantifiers based on Morlet wavelet. Specifically, plots (b) and (c) represent the corresponding scalogram and LIM_x maps; plot (d) shows the F_x curve, indicating significantly high values at the lower scales, and a clear decay within that range, which can be interpreted as the presence of intermittency. Finally, plots (e) and (f) emphasize the exceedances of $LIM_x^2(a,b)$ values over the base threshold 3 and a higher threshold, showing the structure of energy concentrations over time and the degree of coherence between different scales.

The results of a similar analysis using Haar wavelet are displayed in Figure 5.6. In this case, the F_x curve (plot (d)) has significantly high values and a global decay for almost the whole range of scales considered. Threshold exceedances for LIM_X^2 values show clearly different structure with respect to the previous case, indicating more frequent concentrations of energy, but during shorter time periods, and a higher degree of inter-scale coherence.

Finally, the same study using Morlet wavelet is performed for the simulated time series based on the $ARMA(1,1)$ model with Cauchy white noise represented in Figure 5.4. The choice of this wavelet function is justified by the presence of a heterogeneous distribution of variation, with no significant level changes. Here, a global intermittency behavior is clear from the F_x curve, and a well defined inter-scale coherence of energy concentrations over time can be observed (see Figure 5.2).

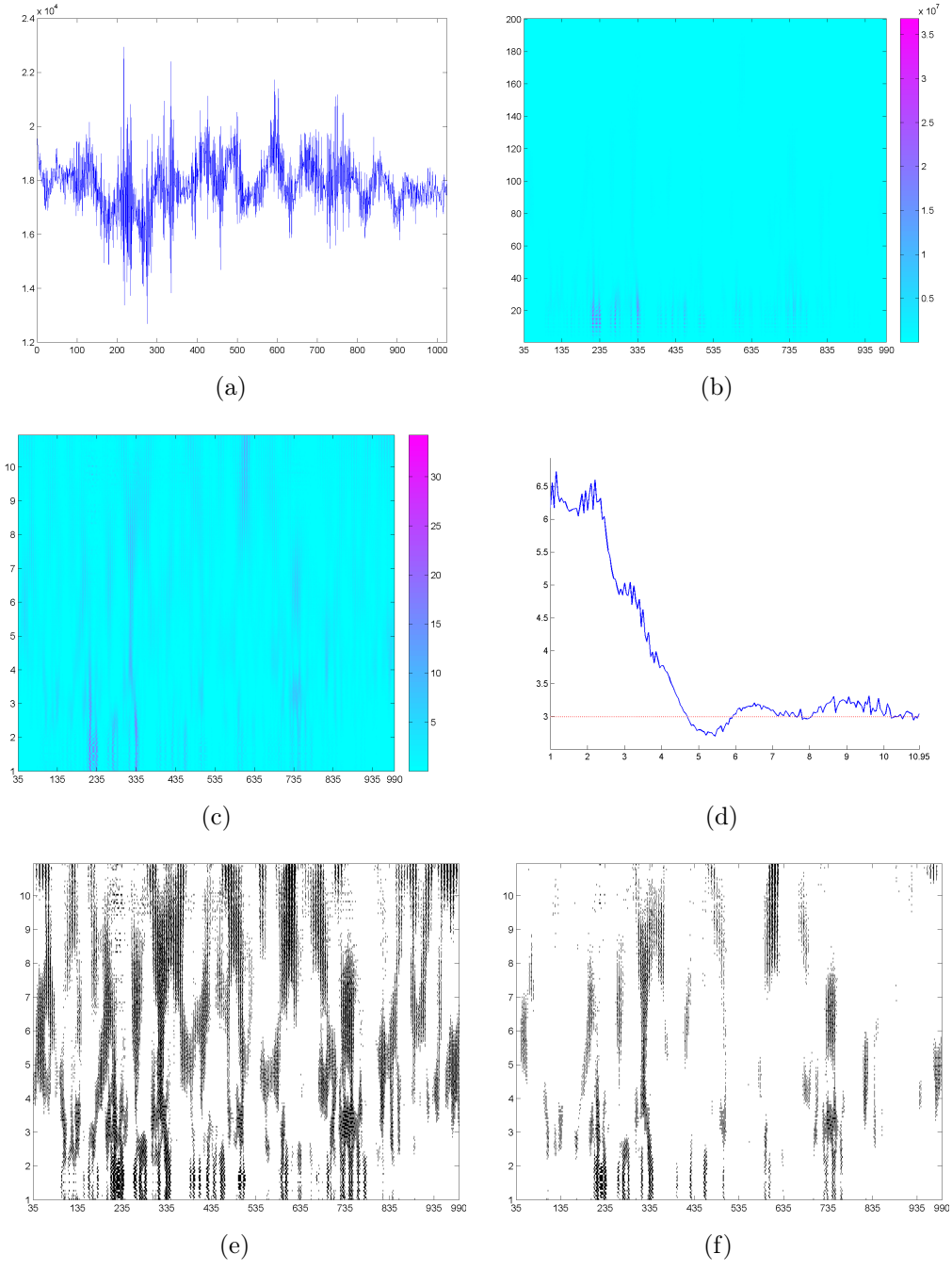


Figure 5.5: (a) Seismic signal (seismic segment 2), (b) scalogram $W_x^2(a,b)$, (c) LIM_x map, (d) F_x curve, (e) threshold exceedance set for $LIM_x^2(a,b) > 3$, (f) threshold exceedance set for $LIM_x^2(a,b) > 16$, using Morlet wavelet.

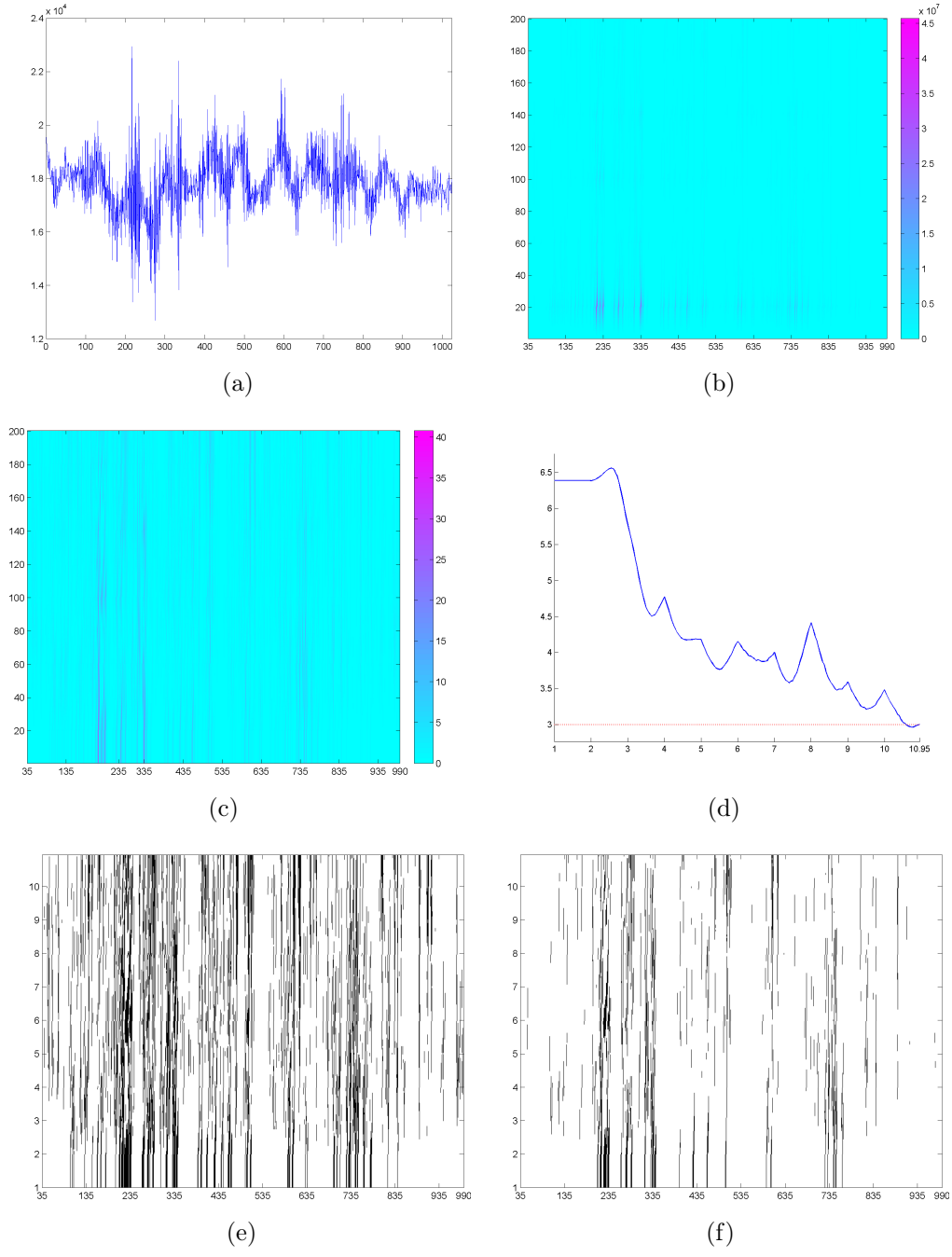


Figure 5.6: (a) Seismic signal (seismic segment 2), (b) scalogram $W_x^2(a,b)$, (c) LIM_x map, (d) F_x curve, (e) threshold exceedance set for $LIM_x^2(a,b) > 3$, (f) threshold exceedance set for $LIM_x^2(a,b) > 16$, using Haar wavelet.

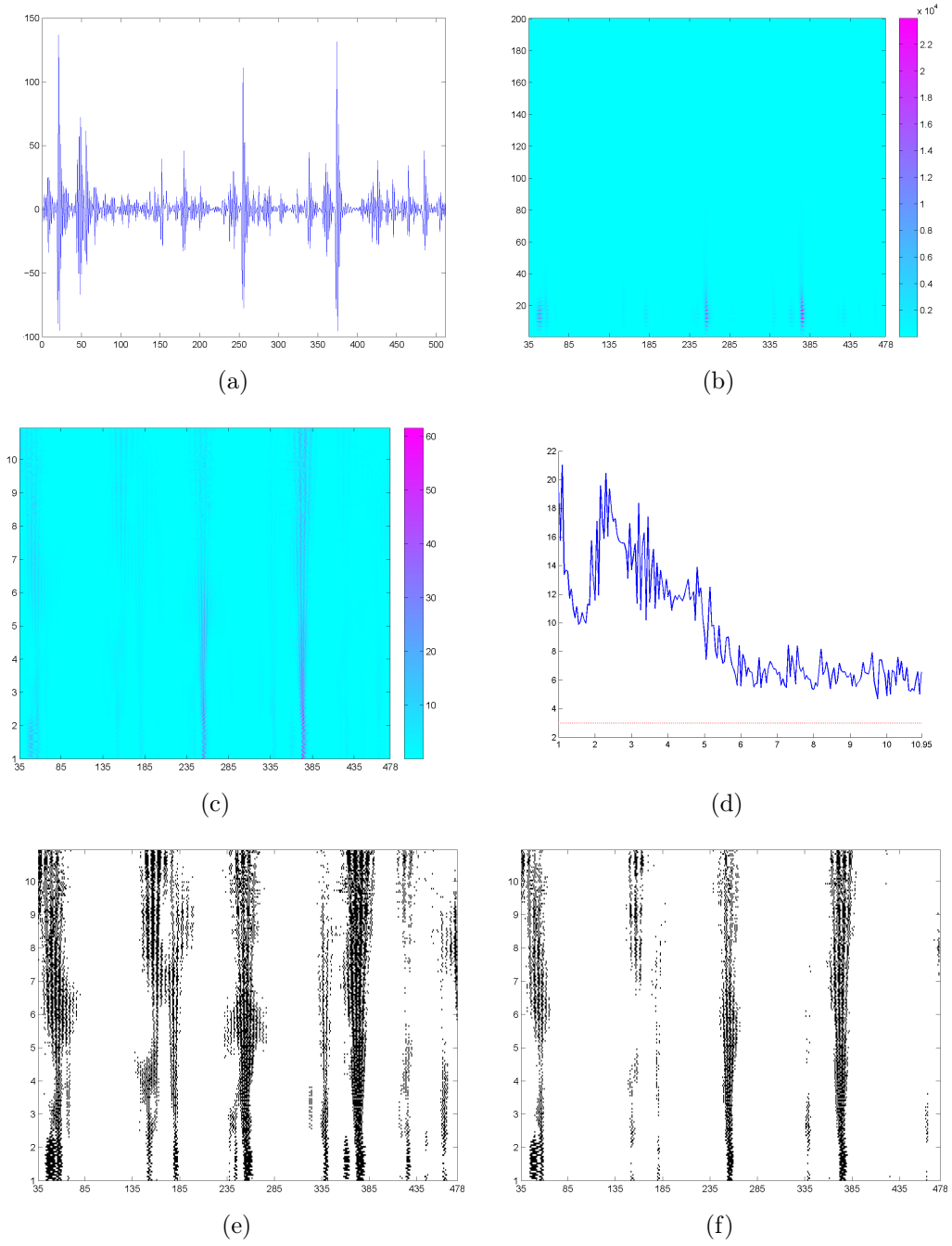


Figure 5.7: (a) Simulated realization of ARMA(1,1) model, with $\phi = 0.7$, $\theta = 0.9$, and Cauchy white noise, (b) scalogram $W_x^2(a,b)$, (c) LIM_x map, (d) F_x curve, (e) threshold exceedance set for $LIM_x^2(a,b) > 3$, (f) threshold exceedance set for $LIM_x^2(a,b) > 16$, using Morlet wavelet.

5.3 Deformation and intermittency

Let x be a signal on $[0, T]$. In order to explain induced heterogeneity and intermittency changes derived from deformation we consider a simple linear transformation $\Phi : [0, T] \to [0, T']$, defined by $\Phi(t) = ct$, with $c > 0$ a real constant, and $T' = cT$. Hence, $\Phi(t)$ has a contraction effect for $c < 1$ and a dilation effect for $c > 1$.

We distinguish the cases where x represents either a 'level' or a 'flow' magnitude, that is, depending on whether the states represent values intrinsic to specific time points or accumulated during successive periods. In correspondence, the deformed signals $x[\Phi]$ and $x[\tilde{\Phi}]$ are defined on $[0, T']$, respectively by

$$x[\Phi](t') = x(\Phi^{-1}(t')) = x\left(\frac{t'}{c}\right)$$

and

$$x[\tilde{\Phi}](t') = x(\Phi^{-1}(t'))(\Phi')^{-1} = x\left(\frac{t'}{c}\right)\frac{1}{c},$$

with $t' \in [0, T']$ (see Section 3.2).

In Section 5.3.1, we analyze how the deformation of a signal produces a transfer of energy between different scales. The effect of both types of deformation on the flatness coefficient is studied in Section 5.3.2.

5.3.1 Inter- and intra-scale transfer of energy

Recall that the wavelet coefficients are given by the convolution of the signal with a given rescaled and translated wavelet function, that is,

$$W_x(a, b) = \langle x(t), \psi_{ab}(t) \rangle = \frac{1}{a^{1/2}} \int_0^T x(t)\psi\left(\frac{t-b}{a}\right) dt.$$

Then, the wavelet coefficients of the deformed signal $x[\Phi]$ (level case) are obtained substituting $x[\Phi](u)$ for $x(t)$ in the previous expression:

$$\begin{aligned} W_{x[\Phi]}(a,b) &= \langle x[\Phi](u), \psi_{ab}(u)\rangle = \frac{1}{a^{1/2}}\int_0^{cT} x[\Phi](u)\psi\left(\frac{u-b}{a}\right)du \\ &= \frac{1}{a^{1/2}}\int_0^{cT} x\left(\frac{u}{c}\right)\psi\left(\frac{u-b}{a}\right)du = \frac{1}{a^{1/2}}\int_0^{T} x(t)\psi\left(\frac{ct-b}{a}\right)cdt \\ &= \frac{c}{a^{1/2}}\int_0^{T} x(t)\psi\left(\frac{t-b/c}{a/c}\right)dt = \frac{c^{1/2}}{(a/c)^{1/2}}\int_0^{T} x(t)\psi\left(\frac{t-b/c}{a/c}\right)dt \\ &= c^{1/2}W_x\left(\frac{a}{c},\frac{b}{c}\right), \end{aligned} \tag{5.2}$$

where we have used the change of variable $t = u/c$ in the fourth equality. This relationship between the wavelet coefficients of the original and deformed signals indicates that, without considering the change of domain, there is a transfer of energy from scale a to scale ca. In fact, averaging the squared wavelet coefficients over b, using the change of variable $b' = b/c$ and then renaming $b' \to b$, we get

$$\begin{aligned} \mathsf{Ave}_b\left[W^2_{x[\Phi]}(a,b)\right] &= \frac{1}{cT}\int_0^{cT} W^2_{x[\Phi]}(a,b)db = \frac{1}{cT}\int_0^{cT} (c^{1/2})^2 W_x^2\left(\frac{a}{c},\frac{b}{c}\right)db \\ &= \frac{1}{T}\int_0^{T} W_x^2\left(\frac{a}{c},b'\right)cdb' = \frac{c}{T}\int_0^{T} W_x^2\left(\frac{a}{c},b\right)db \\ &= c\mathsf{Ave}_b\left[W_x^2\left(\frac{a}{c},b\right)\right]. \end{aligned} \tag{5.3}$$

Substituting the expressions (5.2) and (5.3) in (5.1), the LIM of the deformed signal is computed as

$$LIM_{x[\Phi]}(a,b) = \frac{W^2_{x[\Phi]}(a,b)}{\mathsf{Ave}_b\left[W^2_{x[\Phi]}(a,b)\right]} = \frac{cW_x^2\left(\frac{a}{c},\frac{b}{c}\right)}{c\mathsf{Ave}_b\left[W_x^2\left(\frac{a}{c},b\right)\right]} = LIM_x\left(\frac{a}{c},\frac{b}{c}\right). \tag{5.4}$$

Hence, we can say that if the deformation has a local effect of dilation ($c > 1$), then the relative energy value of scale a/c at localization b/c is transferred to the higher scale $a > a/c$ at localization b. On the other hand, if the deformation has a local effect of contraction ($c < 1$), there is a transfer of relative energy from scale a/c at localization b/c to the lower scale $a < a/c$ at localization b. Note that, due to the change of measure

between the original and transformed physical spaces, the total energy is expanded or shortened by deformation depending on the properties of dilation or contraction, as it can be seen using the changes of variables $a' = a/c$ and $b' = b/c$:

$$\begin{aligned} E_{x[\Phi]} &= \frac{1}{C_\psi}\int_0^\infty\int_{-\infty}^\infty \frac{|W_{x[\Phi]}(a,b)|^2}{a^2}dadb = \frac{1}{C_\psi}\int_0^\infty\int_{-\infty}^\infty \frac{c|W_x\left(\frac{a}{c},\frac{b}{c}\right)|^2}{a^2}dadb \\ &= \frac{c}{C_\psi}\int_0^\infty\int_{-\infty}^\infty \frac{|W_x(a',b')|^2}{(ca')^2}cda'cdb' = \frac{c}{C_\psi}\int_0^\infty\int_{-\infty}^\infty \frac{|W_x(a',b')|^2}{a'^2}da'db' \\ &= cE_x. \end{aligned} \tag{5.5}$$

For a general deformation Φ, depending on where we have local contraction or dilation effects, there will be transfer of energy between scales in different directions.

A similar analysis can be developed for x representing a flow signal. In this case, we obtain the following expressions:

$$W_{X[\tilde{\Phi}]}(a,b) = \langle x[\tilde{\Phi}](u), \psi_{ab}(u)\rangle = \frac{1}{c}W_{x[\Phi]}(a,b) = \frac{1}{c^{1/2}}W_x\left(\frac{a}{c},\frac{b}{c}\right),$$

$$\mathsf{Ave}_b\left[W^2_{x[\tilde{\Phi}]}(a,b)\right] = \mathsf{Ave}_b\left[\frac{1}{c^2}W^2_{x[\Phi]}(a,b)\right] = \frac{1}{c}\mathsf{Ave}_b\left[W^2_x\left(\frac{a}{c},b\right)\right],$$

$$LIM_{x[\tilde{\Phi}]}(a,b) = \frac{|W_{x[\tilde{\Phi}]}(a,b)|^2}{\mathsf{Ave}_b\left[|W_{x[\tilde{\Phi}]}(a,b)|^2\right]} = \frac{\frac{1}{c}W^2_x\left(\frac{a}{c},\frac{b}{c}\right)}{\frac{1}{c}\mathsf{Ave}_b\left[W^2_x\left(\frac{a}{c},b\right)\right]} = LIM_x\left(\frac{a}{c},\frac{b}{c}\right). \tag{5.6}$$

Comparing (5.4) and (5.6), we can see that the LIM values coincide in both cases. However, as it is shown below, using the same changes of variables as before, now the total energy change is

$$\begin{aligned} E_{x[\tilde{\Phi}]} &= \frac{1}{C_\psi}\int_0^\infty\int_{-\infty}^\infty \frac{|W_{x[\tilde{\Phi}]}(a,b)|^2}{a^2}dadb = \frac{1}{C_\psi}\int_0^\infty\int_{-\infty}^\infty \frac{|W_x\left(\frac{a}{c},\frac{b}{c}\right)|^2}{ca^2}dadb \\ &= \frac{1}{C_\psi}\int_0^\infty\int_{-\infty}^\infty \frac{|W_x(a',b')|^2}{c(ca')^2}cda'cdb' = \frac{1}{C_\psi}\int_0^\infty\int_{-\infty}^\infty \frac{|W_x(a',b')|^2}{ca'^2}da'db' = \frac{E_x}{c}. \end{aligned}$$

Hence, when the signal is of flow type the energy is reduced where the deformation has local dilation properties ($c > 1$), and enhanced where the deformation has local

contraction properties ($c < 1$). Note that in the case of a level signal, just the opposite effect was proved (see (5.5)).

5.3.2 Further effects on intermittency characteristics

As introduced in Section 5.2, a global multiscale quantification of intermittency is given in terms of the flatness coefficient F, which averages squared values of the local intermittency measures (LIM) obtained within each single scale. In the case of a level signal x, and for the deformation that we are considering, $\Phi(t) = ct$, using the results in Section 5.3.1, we have

$$\begin{aligned} F_{x[\Phi]}(a) &= \mathsf{Ave}_b\left[LIM^2_{x[\Phi]}(a,b)\right] = \frac{1}{cT}\int_0^{cT} LIM^2_{x[\Phi]}(a,b)\,db \\ &= \frac{1}{cT}\int_0^{cT} LIM^2_x\left(\frac{a}{c},\frac{b}{c}\right)db = \frac{1}{cT}\int_0^{T} LIM^2_x\left(\frac{a}{c},b'\right)cdb' \\ &= \frac{1}{T}\int_0^{T} LIM^2_x\left(\frac{a}{c},b\right)db = F_x\left(\frac{a}{c}\right). \end{aligned}$$

This shows that the shape of the flatness curve is preserved, but on different scales.

For a flow signal x, since the LIM values coincides with those obtained in the level case, the flatness coefficient values coincide too,

$$F_{x[\tilde{\Phi}]}(a) = \mathsf{Ave}_b\left[LIM^2_{x[\tilde{\Phi}]}(a,b)\right] = \mathsf{Ave}_b\left[LIM^2_{x[\Phi]}(a,b)\right] = F_x\left(\frac{a}{c}\right).$$

5.4 Illustrative cases

In this section we study the effect of deformation of the time domain (both 'level' and 'flow' cases are considered) on the distribution of energy concentrations over time at different scales, measured in terms of the quantifiers introduced in Section 5.2.

Signals analyzed are considered to be observed at times $t = 1, 2, ..., 1024$. For simplicity, we apply a smooth deformation with increasing contraction and dilation effects respectively towards the left and right ends of the time interval, but preserving the

domain. Formally, Φ is defined by its inverse, as

$$\Phi^{-1}(t') = \frac{0.7 \times 1024}{\pi} \sin\left(\frac{\pi}{1024} t'\right) + t', \quad \text{for } u \in [0, 1024]$$

(see Figure 5.8).

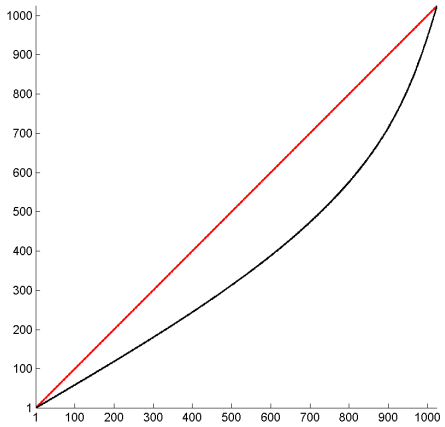


Figure 5.8: Deformation Φ.

First, we compare the effect of deformation ('level' and 'flow' cases) on the seismic segments 1 and 3 (Figure 5.2 (a) and (c)), which respectively correspond to a regular phase and an activity period. Figure 5.9 shows both original signals and their transformations after deformation. Using Morlet wavelet, the corresponding scalograms, LIM maps and F curves are displayed in Figures 5.10, 5.11 and 5.12, respectively. In particular, looking at the F curves the seismic segment 1, which originally shows lightly significant values only for some lower scales, increases intermittency levels under deformation in the flow case. A similar effect can be observed for the seismic segment 3, though in this case with the original signal showing highly significant values at lower scales.

Secondly, the different effects of deformation with respect to variation and level change properties of the signal are analyzed comparing the results obtained using Morlet and Haar wavelets applied to the simulated signal displayed in Figure 5.3 (b). For both wavelets, a clear increase of intermittency levels is obtained after deformation, particularly in the flow case, as it can be seen comparing Figures 5.13, 5.14 and 5.15.

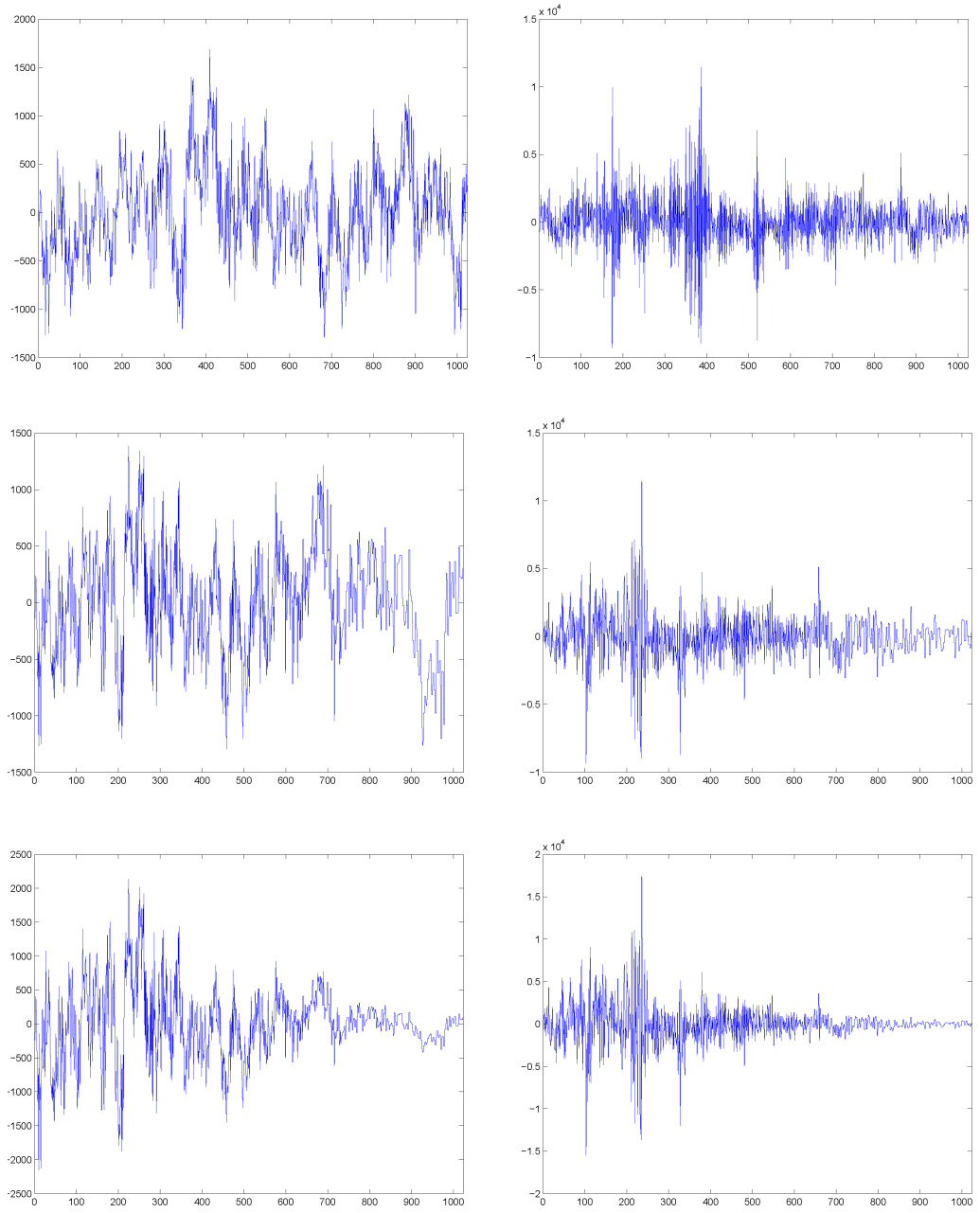


Figure 5.9: From top to bottom: Original signal x, deformed signal of level type $x[\Phi]$, and deformed signal of flow type $x[\tilde{\Phi}]$; for seismic segment 1 (left), seismic segment 3 (right).

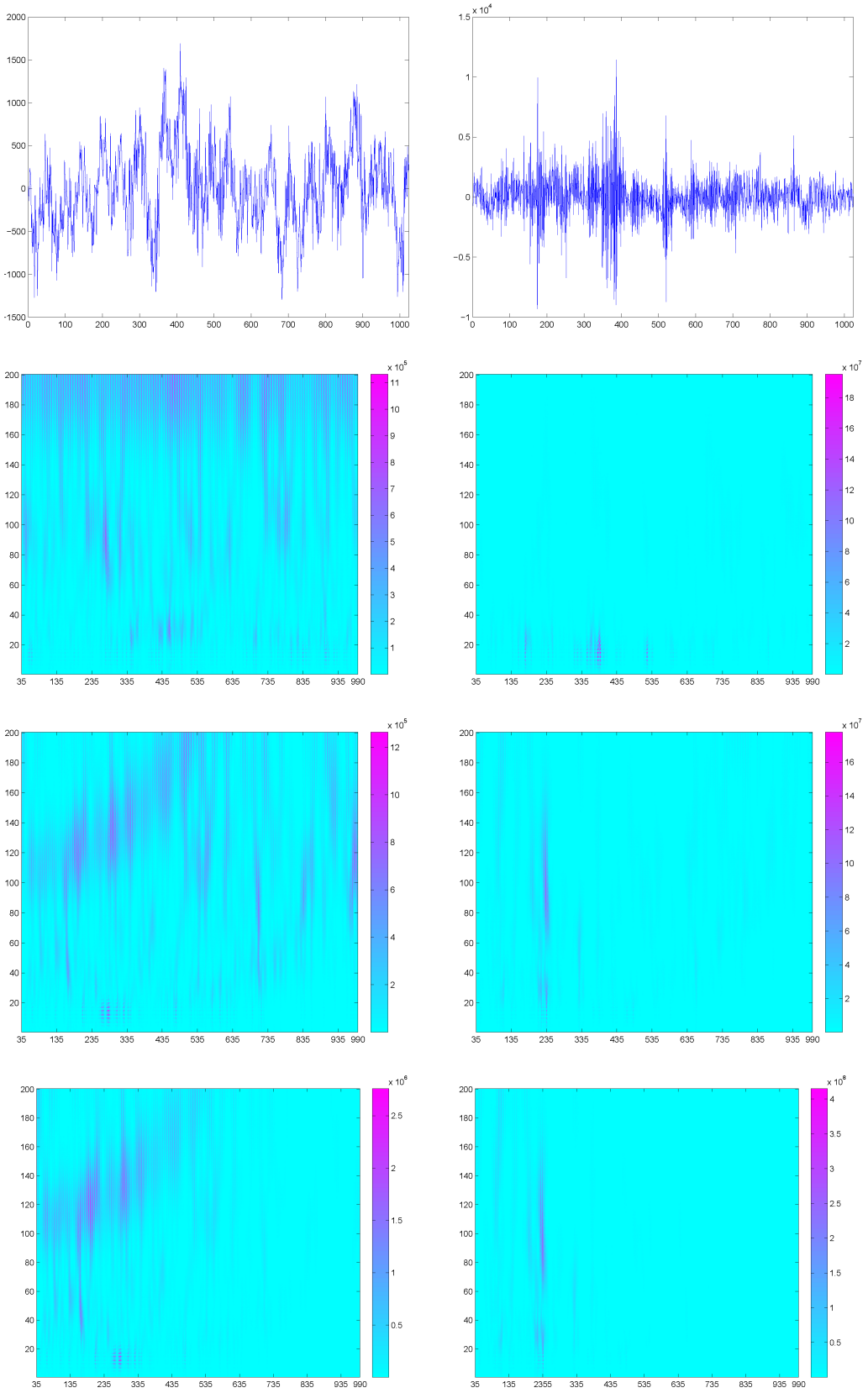


Figure 5.10: From top to bottom: Original signal x, scalogram W_x^2, scalogram $W_{x[\Phi]}^2$, and scalogram $W_{x[\tilde{\Phi}]}^2$; for seismic segment 1 (left), seismic segment 3 (right).

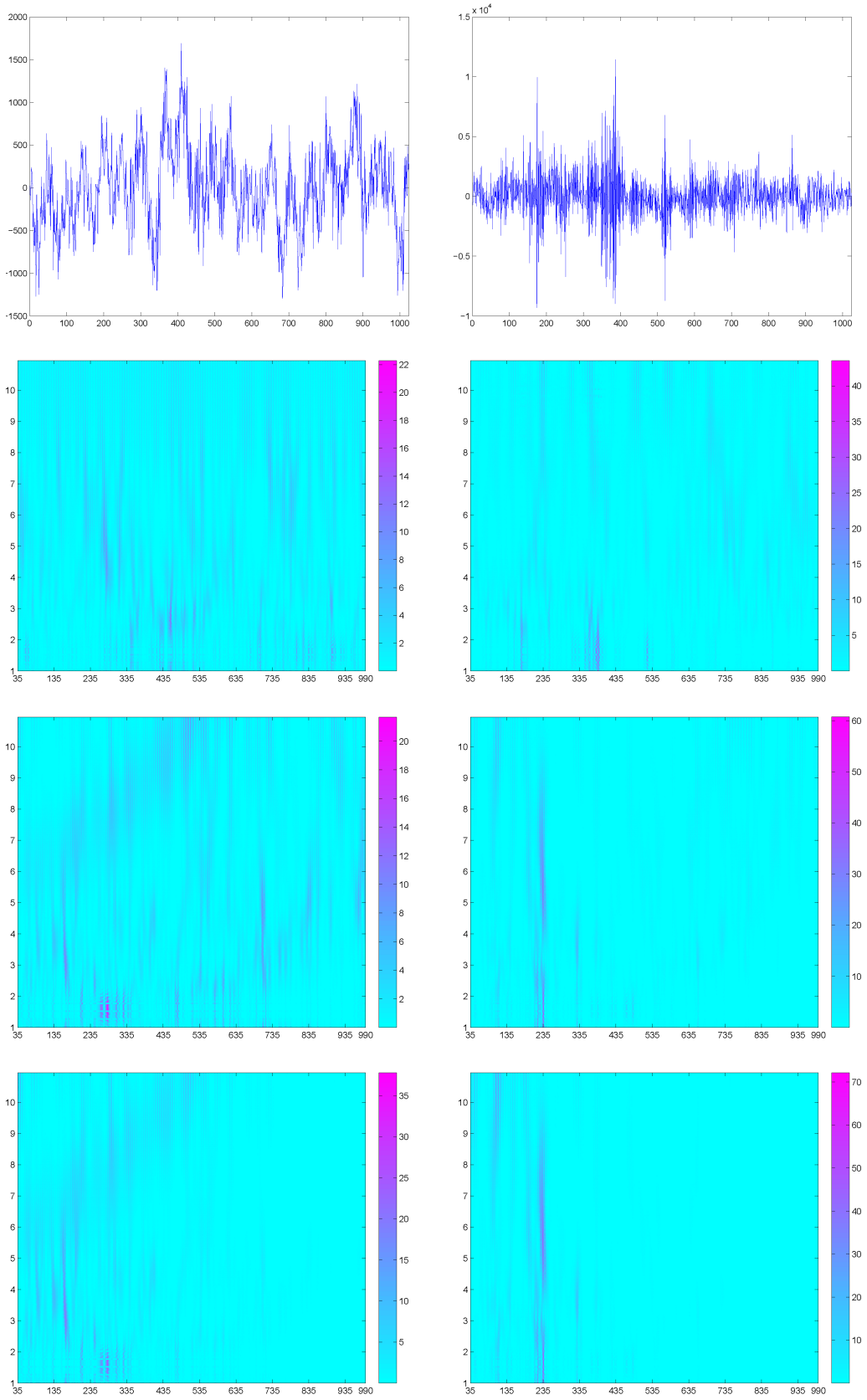


Figure 5.11: From top to bottom: Original signal x, LIM_x map, $LIM_{x[\Phi]}$ map, and $LIM_{x[\tilde{\Phi}]}$ map; for seismic segment 1 (left), seismic segment 3 (right).

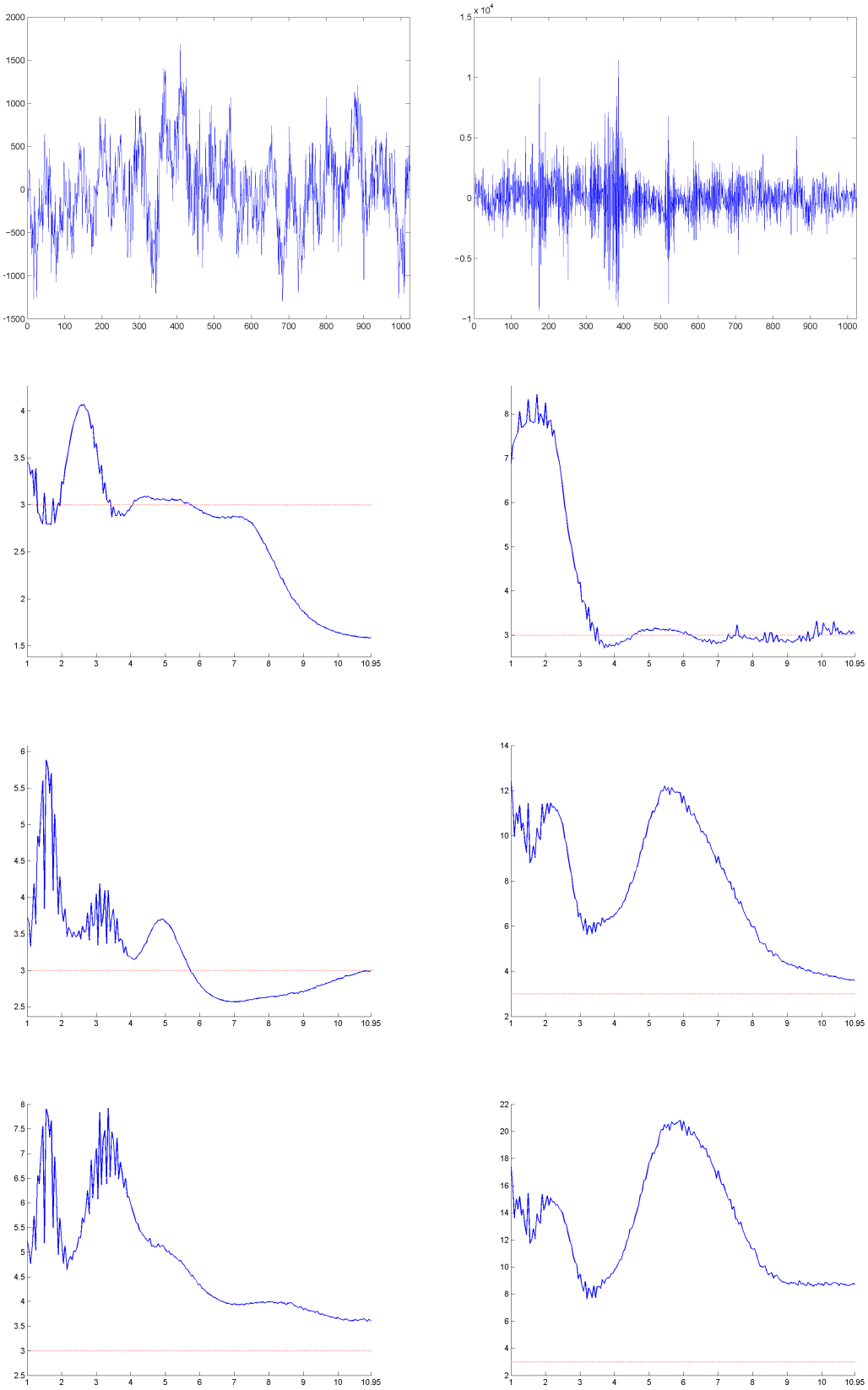


Figure 5.12: From top to bottom: Original signal x, F_x curve, $F_{x[\Phi]}$ curve, and $F_{x[\tilde{\Phi}]}$ curve; for seismic segment 1 (left), seismic segment 3 (right).

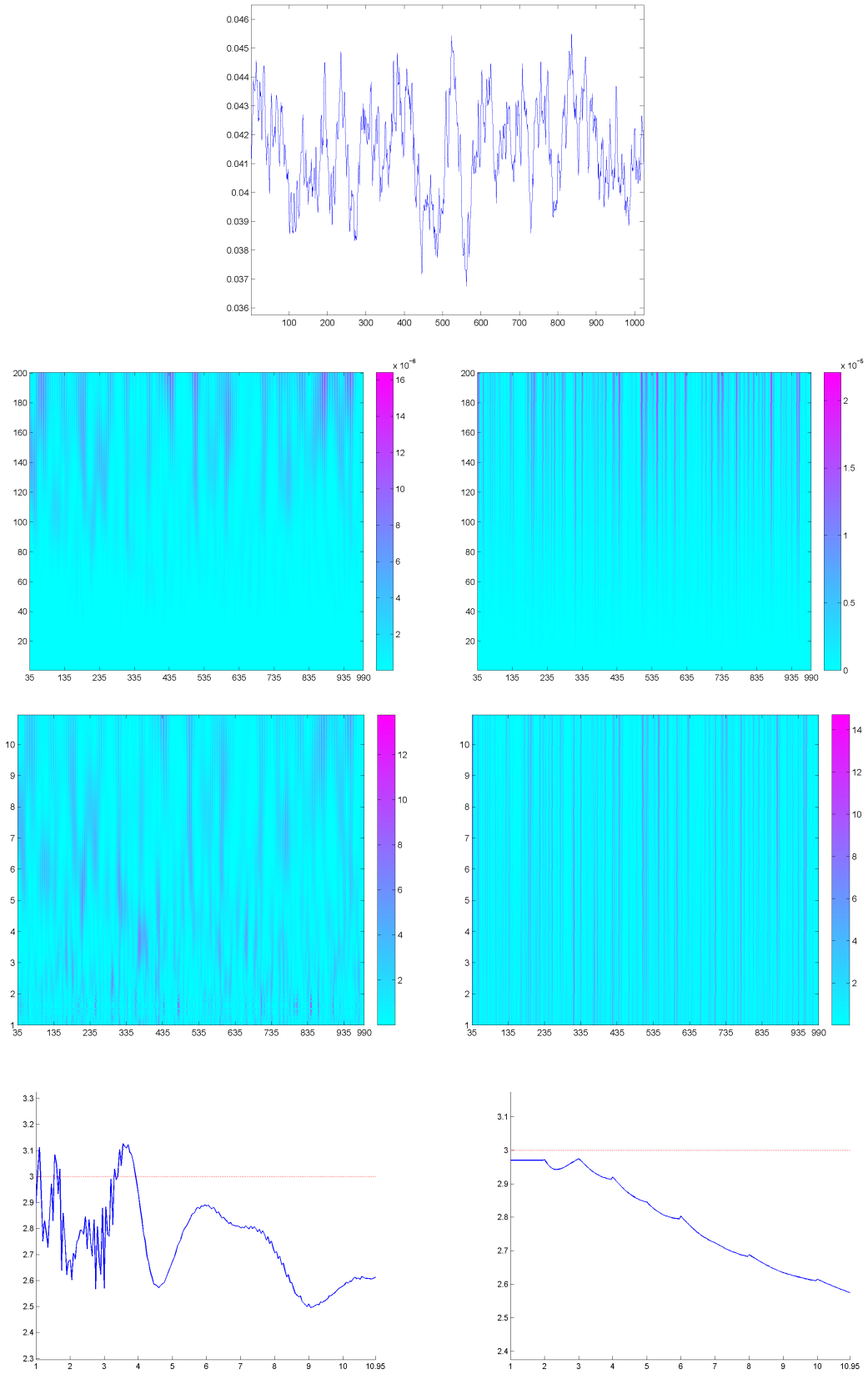


Figure 5.13: From top to bottom: time series x generated at a fixed location from a heterogeneous space-time model x, scalogram W_x^2, LIM_x map, and F_x curve; using Morlet wavelet (left), Haar wavelet (right).

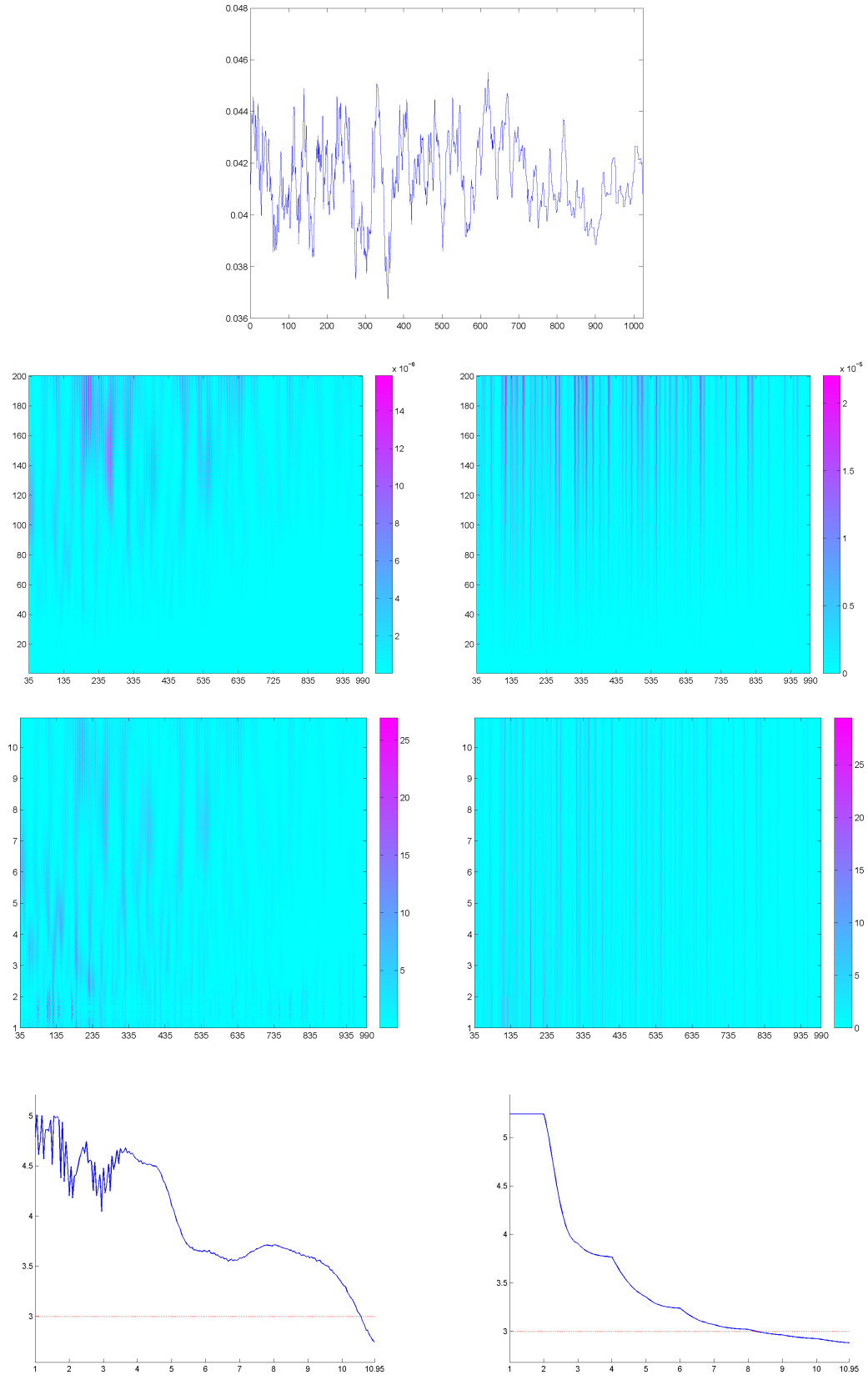


Figure 5.14: From top to bottom: level type deformed signal $x[\Phi]$ based on signal x shown in Figure 5.13, scalogram $W^2_{x[\Phi]}$, $LIM_x[\Phi]$ map, and $F_x[\Phi]$ curve; using Morlet wavelet (left), Haar wavelet (right).

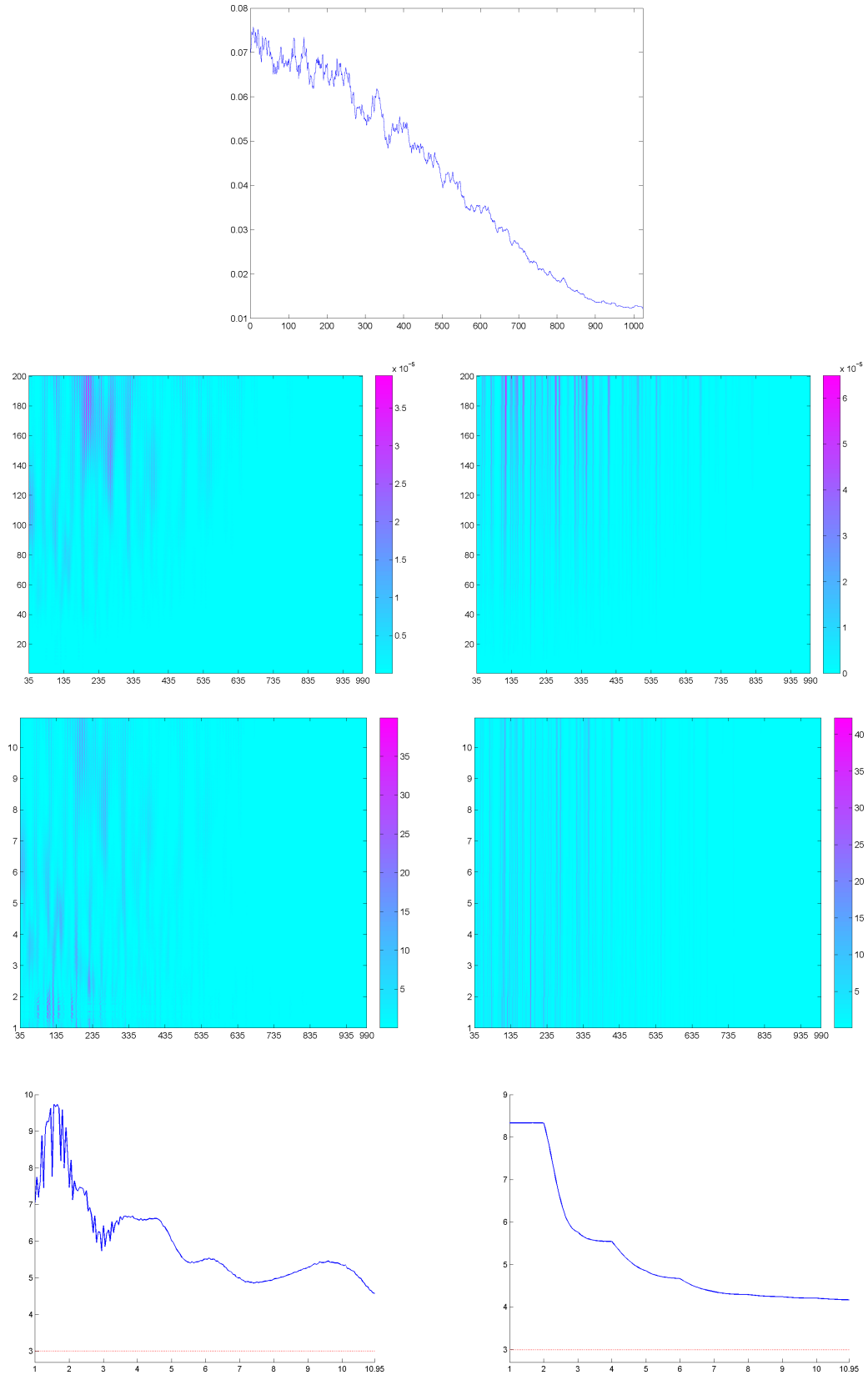


Figure 5.15: From top to bottom: flow type deformed signal $x[\tilde{\Phi}]$ based on signal x shown in Figure 5.13, scalogram $W^2_{x[\tilde{\Phi}]}$, $LIM_x[\tilde{\Phi}]$ map, and $F_x[\tilde{\Phi}]$ curve; using Morlet wavelet (left), Haar wavelet (right).

Briefly, in this chapter, we study some measures based on the continuous wavelet transform useful to detect heterogeneous energy concentrations in a signal, which can be interpreted as intermittency in a certain sense. In particular, we examine the effect of a time deformation on the energy distribution of a signal. It is shown that such a deformation induces generally higher intermittency intensities as well as structural space-scale changes in the intermittency regimes. In this way, some forms of intermittency behavior can be explained in terms of physical space and state deformation.

Chapter 6

Entropy-Based Correlated Shrinkage of Spatial Random Processes

Shrinkage estimation constitutes an active area of research in signal processing. This estimation methodology has usually been designed in the wavelet domain in terms of thresholding rules, providing a powerful tool in the discrimination between signal and noise energy. In the last and a half decade, an extensive literature has been developed in the context of correlated shrinkage, rising the possibility of incorporation of the dependence structural properties between wavelet coefficients.

This framework is adopted in the present chapter, considering the entropy-based design of block-thresholding rules in a first local correlated stage of the estimation algorithm proposed. The thresholding design is adaptive to each resolution level, since it depends on the empirical distribution function of the mutual information ratios between empirical wavelet blocks and the random variables of interest, at each scale. In a second stage, the global correlation structure of the selected wavelet coefficients is incorporated in the filter defining the final estimation of the original random signal. The resulting double correlated shrinkage procedure also allows the detection of high local variability episodes, since local singularity properties of the original random signal are preserved, in the first stage, and then introduced as input information for global reconstruction of micro-scale signal properties, in the second stage. A simulation study is developed, in the Gaussian context, to analyze the sensitivity, measured by empirical stochastic ordering, of the entropy-based block hard thresholding stage in relation to the parameters

characterizing local variability (fractality) and dependence range of the spatial process of interest, the noise level, and the design of the region of interest.

6.1 Introduction

The wavelet transform is usually considered as a whitening filter for optimal processing of random signals. Shrinkage estimation techniques have often been developed under the assumption of independence of the wavelet coefficients. For example, Bayes shrinkage is commonly developed under the consideration of independent a-priori distributions for the wavelet coefficients (see, for example, Abramovich *et al.* 1998[1]; Chipman *et al.* 1997[32]; Johnstone *et al.* 2005[68]; Vidakovic 1998b[119], among others). However, wavelet coefficients of natural environmental phenomena display significant dependencies (see, for example, Anh *et al.* 1998[21]). The performance of wavelet-based filtering algorithms is then improved when such inter- and/or intra-scale dependencies are incorporated to the design of shrinkage functions. Tree-based, e.g. hidden Markov models (see, for example, Crouse *et al.* 1998[40]; Romberg *et al.* 2001[99]), and covariance-based approaches (see, for instance, Amiri *et al.* 2007[6]) constitute the main alternatives in correlated shrinkage. Liu and Moulin (2001)[78] use mutual information to measure dependence between wavelet coefficients.

This chapter provides a new formulation of correlated non-linear shrinkage estimation, based on entropy, for spatial random processes affected by additive noise (see also Angulo *et al.* 2007a[16], 2010b[17] and 2007b[20]). Since the pioneering work by Donoho and Johnstone (1994a)[42], the classical approach of shrinkage estimation considers that the signal of interest to be estimated is deterministic. In the shrinkage approach presented here, the signal of interest to be approximated is considered to be random. As we commented in Section 2.4.5, two main approaches have been adopted for assessing the optimality of shrinkage estimators: The minimax approach (see Donoho and Johnstone 1994a[42], 1994b[43], 1995[44]), and the maxiset approach (see Kerkyacharian and Picard 2000[70], 2002[71]; Vidakovic 1998a[118]). The optimality of the entropy-based correlated shrinkage method presented here is assessed in terms of the loss of information caused after applying the block hard thresholding rule, in the first stage, as well as in terms of the mean quadratic error of the reconstruction obtained in the second stage. Since the local variation (fractality) and the spatial dependence range parameters involved in the

definition of the structure of the process of interest, as well as the parameters related to the noise level of the data and the design of the region of interest, will affect the quality of the shrinkage estimation, a simulation study, in the Gaussian context, is performed for testing in terms of stochastic ordering criteria the influence of such parameters.

In the context of spatial and spatio-temporal sampling design in the wavelet domain, the shrinkage estimation provides a suitable methodology for selecting the most significant observed wavelet coefficients (or wavelet blocks) in the approximation of the values of interest of the underlying spatial process, which here is assumed to be corrupted by additive noise. It is well known that the problem of optimal selection of spatial locations where to collect observations related to the natural process of interest constitutes one of the most important research issues in the Environmental Sciences. Different formulations have been adopted in the literature, generally in terms of characteristics associated to the geostatistical/geophysical problem, as well as in terms of the specifications given by the spatial or spatio-temporal statistical model representing the environmental phenomenon of interest (see Christakos 2000[34]; Cressie and Huang 1999[39]; Gneiting 2002[53]; De Iaco *et al.* 2001[63], 2002[64]; Kiriakidis and Journel 1999[74]; Ruiz-Medina and Angulo 2002[102], 2007[103] ; Ruiz-Medina *et al.* 2003[101]; Ruiz-Medina *et al.* 2008[104]; Stein 2005[109]; Wikle and Cressie 1999[123], among others). In this chapter we refer to the entropy-based approach, where the optimality criterion is defined as to maximize the mutual information between the variables of interest (at the region of interest) and the observable variables to be selected from a set of candidates (see, for instance, Angulo and Bueso 2001[9]; Angulo *et al.* 2000[10], 2005[19]; Bueso *et al.* 1998[30]; Guttorp *et al.* 1993[61]; Le and Zidek 2006[75]).

In Section 6.2 the two fundamental stages of the correlated shrinkage methodology proposed are described. Some numerical examples are studied in Section 6.3, illustrating the influence of the model and noise parameters, and the configuration of the region of interest, on the entropy-based block hard thresholding design.

6.2 Shrinkage methodology

The shrinkage methodology proposed in this chapter consists of two main steps: Stage I performs a local correlated entropy-based shrinkage. Stage II provides a global correlated shrinkage. Specifically, the main steps of stage I are described as follows:

- The set of candidates in the spatial sampling design problem is defined in terms of the discrete wavelet transform of the observable process, with the wavelet coefficients being grouped into blocks. In this step, different block designs can be considered (e.g. nearest-neighbor blocks, tree blocks, square blocks by scales, mixed detail blocks, etc.), according to the local correlation structure of the wavelet coefficients, as well as to the geometrical configuration of the region of interest in relation to the set of candidates.

- The mutual information between wavelet blocks and the variables of interest is then computed. Related intra-scale, cumulated-scale and global information ratios are obtained. The empirical cumulative distribution function of the mutual information ratios is obtained for each scale.

- The hard thresholding rule for selection of the most informative wavelet blocks is defined in terms of the percentiles of the empirical cumulative distribution of the mutual information ratios at each resolution level. The threshold design then depends on the quality (noise level) of the observed wavelet blocks, measured in terms of the magnitude of the information ratio they provide with respect to the variable of interest. Thus, an increasing noise level in the data will provide a decreasing stochastic ordering of the corresponding mutual information ratio random variables.

After suitable selection of the most informative wavelet blocks, stage II is defined as follows:

- A filter defined in terms of the conditional expectation of the values of interest to the values estimated in stage I is applied. In the Gaussian case, the normalized version of such a filter leads to its definition in terms of the inverted global correlation function between selected empirical wavelet blocks, and the cross-correlation function between the values of the process of interest and the selected empirical wavelet blocks. This filter is applied to the selected empirical wavelet coefficients to estimate the values of the process of interest.

The entropy sampling design of wavelets blocks is given in Section 6.2.1. Section 6.2.2 introduces the approximation of the values of the spatial process of interest.

6.2.1 Entropy-based block hard thresholding rule

The first stage solves the problem of entropy-based spatial sampling design in the wavelet domain.

Let $\left\{X(\boldsymbol{s}) : \boldsymbol{s} \in D \subseteq \mathbb{R}^d\right\}$ be the spatial random process of interest. Assume that the observed process is defined as

$$Y(\boldsymbol{s}) = X(\boldsymbol{s}) + \epsilon(\boldsymbol{s}), \quad \boldsymbol{s} \in D,$$

where ϵ represents additive observation noise uncorrelated with X. We are focused on the approximation of the values of X on a region of interest Λ.

We start from the discrete wavelet transform of Y, given by

$$\{Y_{\mathbf{k}} : \mathbf{k} \in \Gamma_0\} \cup \left\{Y_{j,\mathbf{k}} : \mathbf{k} \in \widetilde{\Gamma}_j,\ j \in \mathbb{N}\right\},$$

where

$$\begin{aligned} Y_{\mathbf{k}} &= \int_D Y(\boldsymbol{s})\phi_{\mathbf{k}}(\boldsymbol{s})d\boldsymbol{s}, \quad \mathbf{k} \in \Gamma_0, \\ Y_{j,\mathbf{k}} &= \int_D Y(\boldsymbol{s})\psi_{j,\mathbf{k}}(\boldsymbol{s})d\boldsymbol{s}, \quad \mathbf{k} \in \widetilde{\Gamma}_j,\ j \in \mathbb{N}, \end{aligned} \tag{6.1}$$

with $\{\phi_{\mathbf{k}} : \mathbf{k} \in \Gamma_0\}$ representing the system of compactly supported scaling functions generating the space V_0, and $\left\{\psi_{j,\mathbf{k}}, : \mathbf{k} \in \widetilde{\Gamma}_j\right\}, j \in \mathbb{N}$, representing the compactly supported wavelet bases generating the spaces $W_j,\ j \in \mathbb{N}$. Note that, as commented before, we have assumed that Y satisfies the necessary conditions to ensure that the integrals in equation (6.1) are well defined in some stochastic sense. For example, in the Gaussian case, the integrals are defined in the mean-square sense. In other cases, the integrals are usually defined in the sample-path sense. In practice, the wavelet transform of Y is approximated by the wavelet transform of the data on a region of candidate observation sites Π.

The scaling coefficients $\{Y_{\mathbf{k}} : \mathbf{k} \in \Gamma_0\}$, and the wavelet coefficients $\left\{Y_{j,\mathbf{k}} : \mathbf{k} \in \widetilde{\Gamma}_j, j \in \mathbb{N}\right\}$ are respectively grouped into blocks $\mathcal{B}_1, \ldots, \mathcal{B}_{M_0}$, and $\mathcal{B}_1^j, \ldots, \mathcal{B}_{M_j}^j,\ j \in \mathbb{N}$. The design of the blocks depends on the local spatial correlation properties of the wavelet coefficients (homogeneous vs. heterogeneous, isotropic

vs. anisotropic models), and on the configuration of the locations defining the set Λ. For example, a regular network design can be considered, consisting of defining blocks with the same support size and geometrical form for each resolution level $j \in \mathbb{N}$, covering the region Λ, and defining a regular block network (each block is constituted by wavelet coefficients corresponding to averaged information in adjacent supports). Such a design is suitable for the case where the spatial locations defining the region Λ configure a regular grid. Mixed detail blocks constituted by vertical, horizontal and diagonal details are considered for covering anisotropic local correlated wavelet coefficients, providing homogeneous information on the variables of interest. Tree block designs are considered when localized averaged information referred to each point of interest must be kept within each block, etc.

In the first stage, the approximation on the values of interest of X is given by the formula

$$\widetilde{X}(\boldsymbol{s}) = \sum_{\mathbf{k}\in\Gamma_0} \gamma_{\mathbf{k}} Y_{\mathbf{k}} \phi_{\mathbf{k}}(\boldsymbol{s}) + \sum_{j\in\mathbb{N}} \sum_{\mathbf{k}\in\widetilde{\Gamma}_j} \gamma_{j,\mathbf{k}} Y_{j,\mathbf{k}} \psi_{j,\mathbf{k}}(\boldsymbol{s}), \tag{6.2}$$

where the weights $\gamma_{\mathbf{k}}$, $\mathbf{k} \in \Gamma_0$, and $\gamma_{j,\mathbf{k}}$, $\mathbf{k} \in \widetilde{\Gamma}_j$, $j \in \mathbb{N}$, are defined as follows:

$$\begin{aligned}
\gamma_{\mathbf{k}} &= \begin{cases} 1 & \text{if } Y_{\mathbf{k}} \in \mathcal{B}_l \text{ for some } l \in \{1,\ldots,M_0\} \text{ with } \mathbf{I}(X_\Lambda, \mathcal{B}_l) > \lambda_0 \\ 0 & \text{otherwise,} \end{cases} \\
\gamma_{j,\mathbf{k}} &= \begin{cases} 1 & \text{if } Y_{j,\mathbf{k}} \in \mathcal{B}_l^j \text{ for some } l \in \{1,\ldots,M_j\} \text{ with } \mathbf{I}(X_\Lambda, \mathcal{B}_l^j) > \widetilde{\lambda}_j \\ 0 & \text{otherwise,} \end{cases}
\end{aligned} \tag{6.3}$$

with $\mathbf{I}(X_\Lambda, \mathcal{B}_l)$ (respectively $\mathbf{I}(X_\Lambda, \mathcal{B}_l^j)$) denoting the mutual information between the random variables of interest, located at region Λ, and the random variables defining the wavelet coefficients that constitute the blocks $\mathcal{B}_l$, for $l \in \{1,\ldots,M_0\}$ (respectively $\mathcal{B}_l^j$, for $l \in \{1,\ldots,M_j\}$). To select the magnitude of the thresholds $\lambda_0, \widetilde{\lambda}_j$, $j \in \mathbb{N}$, the empirical cumulative distribution of the information ratios at each scale is fitted. Thus, an increasing stochastic ordering of the information ratio variables, for a fixed scale, means an increasing quality of the data at such a scale (regarding information provided on the variable of interest). This fact obviously leads to an increasing magnitude of

the percentiles of the corresponding empirical cumulative distributions. Therefore larger values of the threshold are allowed leading to a more selective thresholding design, and then to a more effective dimension reduction approach. On the other hand, in the opposite case of decreasing stochastic ordering of the information ratio variables (i.e., decreasing quality of the data), discrimination between significant wavelet blocks is harder than in the previous case, and the concentration of the corresponding empirical cumulative distributions with respect to small values leads to a smaller magnitude of the percentiles, i.e., of the thresholds selected at each resolution level, $\lambda_0, \widetilde{\lambda}_j,\ j \in \mathbb{N}$. In the spatial sampling design problem, both situations correspond to the sound strategy of decreasing or increasing the number of spatial locations selected, according to the increasing or decreasing quality of the data.

As it can be interpreted from equation (6.3), an adaptative threshold design through scales is adopted, since the thresholds $\lambda_0, \widetilde{\lambda}_j,\ j \in \mathbb{N}$, depend on the resolution level. Note that the proposed threshold design is more sensitive at lower scales (i.e., coarse resolution levels) when macro-scale properties are studied (e.g. dependence range in the Gaussian case, dispersion of spatial locations in the region Λ, etc.). Conversely, the threshold design is more sensitive at higher scales (i.e., finest resolution levels) when micro-scale properties are studied (e.g. local variability properties, intensity of the observation noise, etc.). This is highlighted in the simulation study developed in Section 6.3.3.

6.2.2 Global correlated shrinkage

We now briefly introduce the second stage of the shrinkage methodology proposed. Specifically, the following estimates are defined, for each $\boldsymbol{s} \in \Lambda$:

$$\hat{X}(\boldsymbol{s}) = E[X(\boldsymbol{s}) \mid \{\widetilde{X}(\mathbf{y}),\ \mathbf{y} \in \Lambda\}], \tag{6.4}$$

where, for each $\mathbf{y} \in \Lambda$, $\widetilde{X}(\mathbf{y})$ is computed as in equation (6.2) in stage I.

Note that, in the case where the spatial process of interest X is also defined in the wavelet domain, denoting by $\{X_{\mathbf{k}} :\ \mathbf{k} \in \Gamma_0\}$ and $\left\{X_{j,\mathbf{k}} :\ \mathbf{k} \in \widetilde{\Gamma}_j,\ j \in \mathbb{N}\right\}$ the wavelet coefficients involved in the definition of its wavelet transform, the following estimates

can be computed in stage II:

$$\begin{aligned}\hat{X}_{\mathbf{k}} &= E[X_{\mathbf{k}} \mid \{\gamma_{\mathbf{k}} Y_{\mathbf{k}} : \ \mathbf{k} \in \Gamma_0;\ \gamma_{j,\mathbf{k}} Y_{j,\mathbf{k}} : \ \mathbf{k} \in \widetilde{\Gamma}_j,\ j \in \mathbb{N}\}], \quad \mathbf{k} \in \Gamma_0, \\ \hat{X}_{j,\mathbf{k}} &= E[X_{j,\mathbf{k}} \mid \{\gamma_{\mathbf{k}} Y_{\mathbf{k}} : \ \mathbf{k} \in \Gamma_0;\ \gamma_{j,\mathbf{k}} Y_{j,\mathbf{k}} : \ \mathbf{k} \in \widetilde{\Gamma}_j,\ j \in \mathbb{N}\}], \quad \mathbf{k} \in \widetilde{\Gamma}_j,\ j \in \mathbb{N}. \end{aligned} \tag{6.5}$$

The next scheme describes in synthesis our entropy-based correlated shrinkage algorithm:

$$\left.\begin{matrix} X \\ \epsilon \end{matrix}\right] \longrightarrow Y \xrightarrow{\mathcal{W}} \{Y_{\mathbf{k}} : \ \mathbf{k} \in \Gamma_0;\ Y_{j,\mathbf{k}} : \ \mathbf{k} \in \widetilde{\Gamma}_j,\ j \in \mathbb{N}\}$$

$$\text{Stage 1} \downarrow$$

$$\{\widetilde{X}(\boldsymbol{s}), \boldsymbol{s} \in \Lambda\}$$

$$\text{Stage 2} \downarrow$$

$$\{\hat{X}(\boldsymbol{s}), \boldsymbol{s} \in \Lambda\},$$

where $\mathcal{W}$ denotes the wavelet transform.

6.3 Numerical examples

A Gaussian scenario is considered in the numerical examples analyzed in this section, under which the second stage of our shrinkage methodology becomes linear. Two spatial dependence structures for the process of interest are analyzed: The exponential covariance model, and the Linnik-type covariance model. In the exponential case, the influence of the dependence range, the design of the region of interest, and the noise intensity is analyzed in relation to the entropy-based block hard thresholding design. In the Linnik model, the influence of the local singularity of the model, measured in terms of a fractality parameter, as well as of the noise intensity is studied in relation to the design of the first stage of the methodology proposed. In both cases, we are assuming that the variance of the process is equal to 1.

The exponential covariance function, for a zero-mean spatial process X, is defined as

$$C(\boldsymbol{s}, \boldsymbol{s}+\mathbf{h}) = E[X(\boldsymbol{s})X(\boldsymbol{s}+\mathbf{h})] = \exp\left\{-\frac{3\|\mathbf{h}\|}{a}\right\}, \quad a > 0, \quad \boldsymbol{s} \in D \subset \mathbb{R}^2,$$

where the parameter a represents the dependence range.

The Linnik covariance function is given by

$$B_X(\boldsymbol{s}, \boldsymbol{s}+\mathbf{h}) = E[X(\boldsymbol{s})X(\boldsymbol{s}+\mathbf{h})] = \frac{1}{1+\|\mathbf{h}\|^{\alpha}}, \quad 0 < \alpha < 2, \quad \boldsymbol{s} \in D \subset \mathbb{R}^2,$$

where α provides information on the local singularity of the model (see Section 2.2.2).

The observation model is defined as

$$Y(\boldsymbol{s}) = X(\boldsymbol{s}) + \epsilon(\boldsymbol{s}), \quad \boldsymbol{s} \in \Pi \subseteq D \subset \mathbb{R}^2,$$

where ϵ represents additive observation noise, which is assumed to be Gaussian white noise with intensity σ^2, uncorrelated with the process of interest X. Here, Π defines the region of candidate locations for observation, which can be discrete or continuous. The region of interest Λ is constituted by the spatial locations where we are interested in the knowledge of X (i.e., the spatial region on which the thematic map is constructed).

In the examples, we consider the case where the region of candidates Π is defined in terms of a discrete finite set of spatial locations, given by a regular $2^6 \times 2^6$ grid. We then apply to the potentially observable random variables, located on Π, the two-dimensional Discrete Wavelet Transform (DWT), based on the tensorial product of the Haar wavelet basis, with 6 resolution levels (hereafter named 'scales'). From scale 2, the wavelet coefficients are grouped into three-term blocks, each consisting of related vertical, horizontal and diagonal details.

Regarding the block hard threshold design in stage I, its dependence on the covariance range is illustrated in Section 6.3.1, for the exponential covariance model. Dependence on the distribution of the spatial locations of region Λ is also shown in this section. The effect of the local singularity order of the process of interest on the thresholding design of stage I is illustrated in Section 6.3.2, under the Linnik covariance model. Section 6.3.3 provides the approximation of the values of the process of interest from stage II for the latter case.

6.3.1 Exponential model

Sensitivity of the entropy-based block hard thresholding design in relation to the dependence range parameter a is first established considering the scale 3, since the wavelet blocks provide information on the large-scale properties of the variables of interest at such a scale. The empirical cumulative distribution function of the information ratios associated with the wavelet blocks is computed. The percentiles of this distribution then define different options in the threshold design. Information ratios are represented vs. empirical percentiles and thresholds, as well as empirical percentiles vs. thresholds, for the dependence ranges $a = 1/6$, $1/3$, and $1/2$, in Figure 6.1, and for $a = 5$, 10, and 15, in Figure 6.2, in all cases considering $\sigma^2 = 0.1$, and a uniformly distributed spatial configuration of Λ with 64 locations of interest. Significant differences are observed in the threshold design for the values $a = 1/6$, $1/3$, and $1/2$, but not for $a = 5$, 10, and 15 due to redundancy, since the latter values are too large in relation to the relative distance between the spatial locations in regions Λ and Π. Within the range of values $a = 1/6$, $1/3$, and $1/2$, a higher amount of information is provided by the largest dependence range, $a = 1/2$, where more concentration on higher information ratio values can be seen in the empirical cumulative distribution function. This fact is reflected in Figure 6.1 (top right, information ratios vs. thresholds), where the information ratio variable associated with $a = 1/2$ is the largest variable, i.e., the dominant variable (maximum quality of observed blocks in terms of mutual information), in terms of the usual stochastic ordering that can be established between the three variables displayed, associated with the empirical cumulative distributions corresponding to the cases $a = 1/6$, $1/3$, and $1/2$.

To study the influence of the noise intensity parameter σ^2, we consider the values $\sigma^2 = 0.1$, 0.25, and 0.50, for a fixed dependence range $a = 15$, and a uniformly distributed configuration of region Λ. In this case, the threshold design is studied at scales 4 and 5, where the codified information is related to the local variability properties of the observed process Y. We show here the results corresponding to scale 4 (results are qualitatively similar for scale 5). Figure 6.3 displays the information ratios vs. empirical percentiles and corresponding thresholds, as well as the relationship between percentiles and threshold values at scale 4. Increasing the intensity of the noise, a higher concentration on small values of the empirical cumulative distribution is obtained, and smaller information ratios are provided by each of the wavelet blocks on the variable of interest. Hence, lower

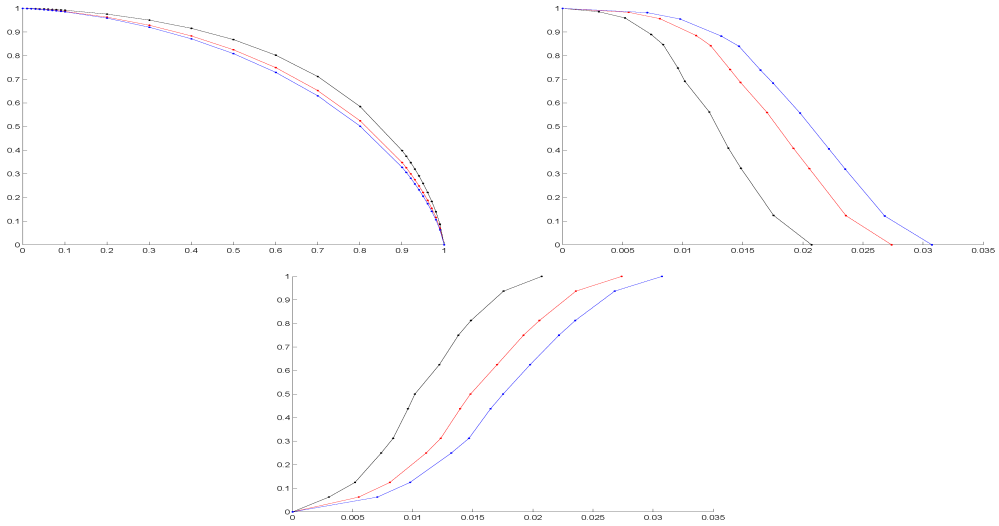


Figure 6.1: Exponential model. Region Λ with 64 locations, uniformly distributed; scale 3; $\sigma^2 = 0.1$; $a = 1/6$ (black), $a = 1/3$ (red), $a = 1/2$ (blue). Information ratios vs. empirical percentiles (top left), information ratios vs. thresholds (top right), and empirical percentiles vs. thresholds (bottom).

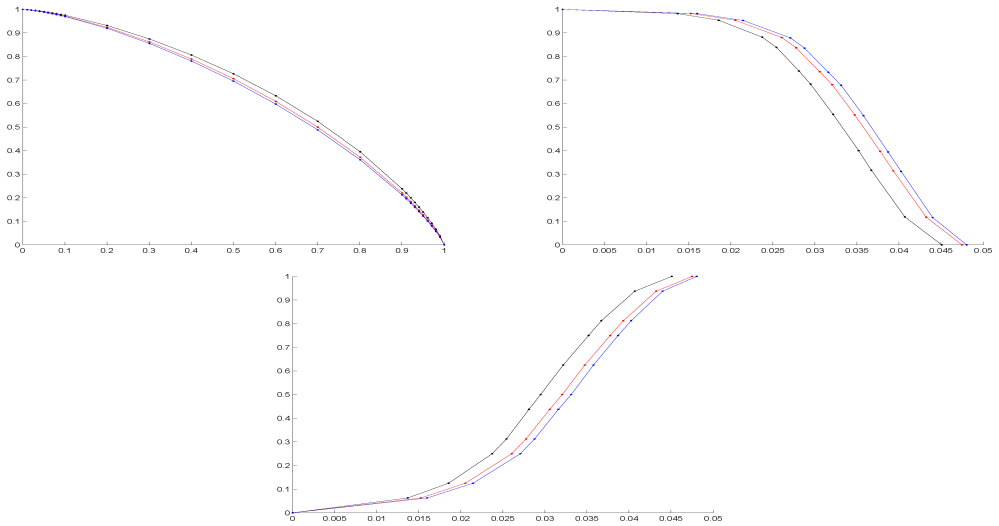


Figure 6.2: Exponential model. Region Λ with 64 locations, uniformly distributed; scale 3; $\sigma^2 = 0.1$; $a = 5$ (black), $a = 10$ (red), $a = 15$ (blue). Information ratios vs. empirical percentiles (top left), information ratios vs. thresholds (top right), and empirical percentiles vs. thresholds (bottom).

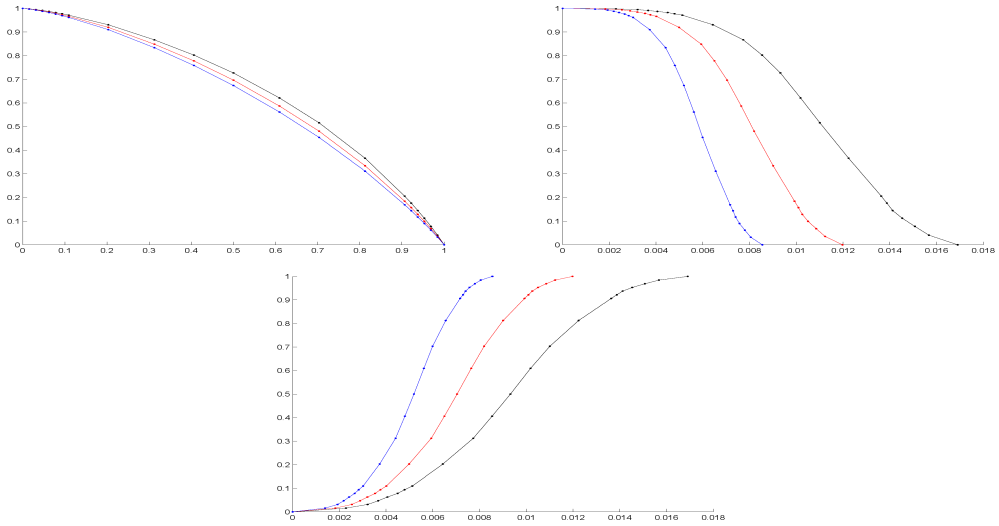


Figure 6.3: Exponential model. Region Λ with 64 locations, uniformly distributed; scale 4; $a = 15$; $\sigma^2 = 0.1$ (black), $\sigma^2 = 0.25$ (red), $\sigma^2 = 0.5$ (blue). Information ratios vs. empirical percentiles (top left), information ratios vs. thresholds (top right), and empirical percentiles vs. thresholds (bottom).

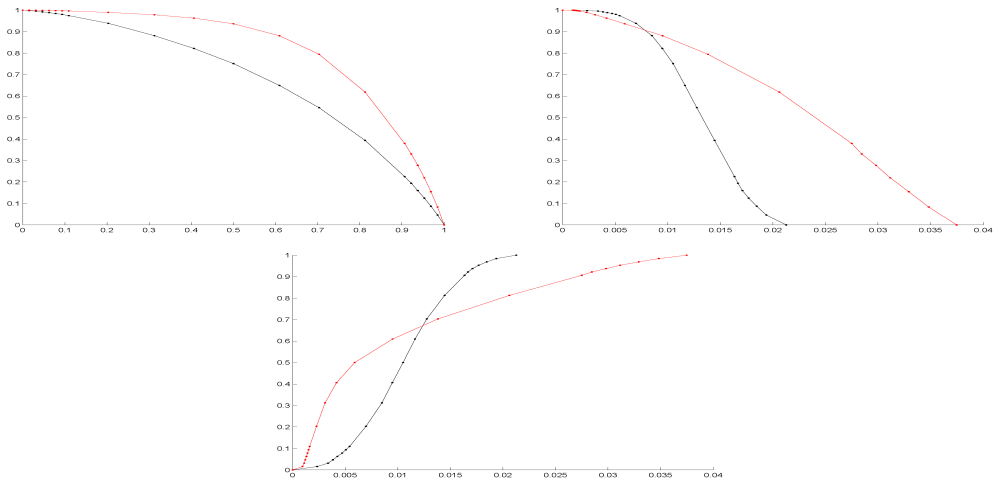


Figure 6.4: Exponential model. Region Λ with 64 locations, uniformly distributed (black) and normally distributed (red); scale 4; $a = 5$; $\sigma^2 = 0.1$; . Information ratios vs. empirical percentiles (top left), information ratios vs. thresholds (top right), and empirical percentiles vs. thresholds (bottom).

thresholds should be considered in the design. Equivalently, a larger number of blocks is needed to obtain a fixed amount of information when the intensity of the noise increases. Again, the stochastic ordering of the three corresponding mutual information variables is reflected in Figure 6.3 (top right), where the most informative case $\sigma^2 = 0.1$ provides the dominant or largest empirical mutual information variable.

Finally, the influence of the spatial configuration of region Λ is studied considering the uniform and normal distributions to generate 50 spatial locations for region Λ. The threshold design is analyzed at scales 4 and 5 for $a = 5$ and $\sigma^2 = 0.1$ (see Figure 6.4). Information ratios associated with the wavelet blocks are more concentrated on small values in the case of the uniformly distributed region of interest. Hence, smaller thresholds than in the normal case must be considered to provide a fixed amount of information on the variable of interest. The normal design of region Λ is then more informative than the uniform one, as it can be seen in Figure 6.4 (top right), where the dominant or largest empirical mutual information ratio variable, in the stochastic ordering sense, is given by the normal case.

6.3.2 Linnik model

In this section, we consider the Linnik-type spatial dependence model and study the effect of the fractality and long-range dependence parameter α on the block hard thresholding design in stage I. In the analysis, the empirical cumulative distribution functions of the information ratios associated with wavelet blocks at scale 6 are considered under different scenarios. Specifically, 64 spatial locations are generated from the uniform distribution to define Λ, and different levels of fractality and long-range dependence are analyzed, corresponding to the values $\alpha = 0.25$, 0.5, and 0.75. In Figure 6.5, the information ratios vs. empirical percentiles, the information ratios vs. thresholds and the empirical percentiles vs. thresholds are displayed for scale 6. In the three cases considered, the noise level is given by $\sigma^2 = 0.1$. We observe that when α increases the observed wavelet blocks are more informative when thresholds approximately less than 0.5×10^{-3} are established. The stochastic ordering of the three cases considered, $\alpha = 0.25$, 0.5, and 0.75, does not completely reflect the most informative case, corresponding to the lowest order of local singularity $\alpha = 0.75$. The reason is that large values of α also provide short-range dependence, which corresponds to a less informative case as we observed in

the exponential model. This can be interpreted as a disadvantage of considering a model with information on local singularity and dependence range determined by a unique parameter.

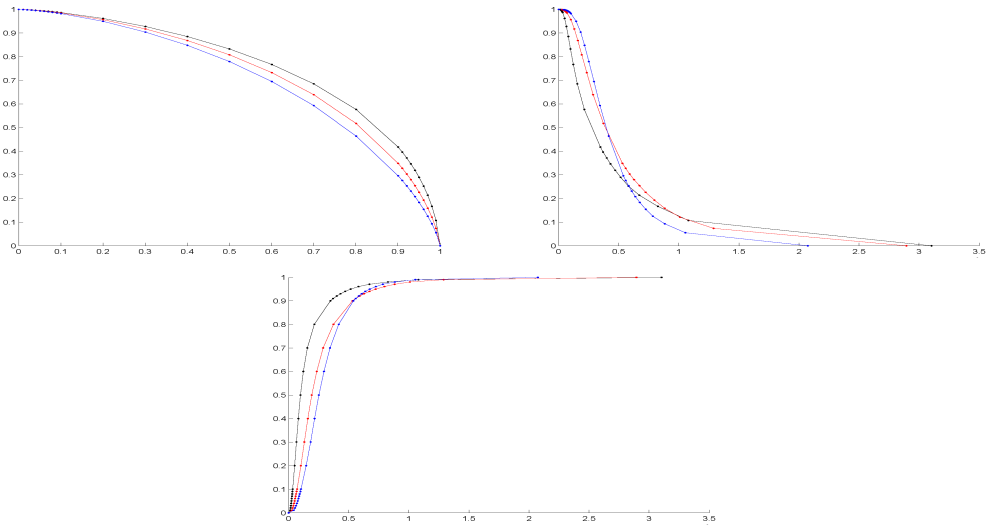


Figure 6.5: Linnik model. Region Λ with 64 locations, uniformly distributed; scale 6; $\sigma^2 = 0.1$; $\alpha = 0.25$ (black), $\alpha = 0.5$ (red), $\alpha = 0.75$ (blue). Information ratios vs. empirical percentiles (top left), information ratios vs. thresholds (top right), and empirical percentiles vs. thresholds (bottom).

In Figures 6.6-6.8, the influence of the noise level of data in the wavelet block hard thresholding design, is analyzed. Specifically, as it can be seen in the top-right plots of these figures, a perfect stochastic ordering can be observed from the least ($\sigma^2 = 0.5$) to the most ($\sigma^2 = 0.1$) informative case. This differences in the stochastic ordering are stronger in the smoothest case ($\alpha = 0.75$), where local variability in the observation model is essentially introduced by the observation noise.

6.3.3 Shrinkage estimation

The shrinkage estimation resulting from stage II of the proposed methodology is now illustrated for the Linnik-type spatial dependence model. Specifically, the performance of the correlated shrinkage estimation procedure formulated is shown, starting from the entropy-block hard thresholding stage computed in the previous section. The mutual

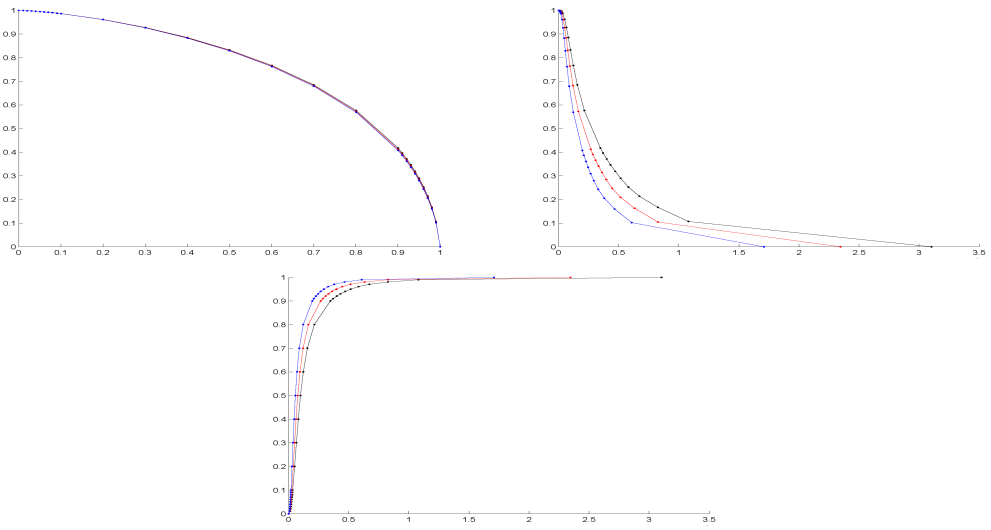


Figure 6.6: Linnik model. Region Λ with 64 locations, uniformly distributed; scale 6; $\alpha = 0.25$; $\sigma^2 = 0.1$ (black), $\sigma^2 = 0.25$ (red), $\sigma^2 = 0.5$; (blue). Information ratios vs. empirical percentiles (top left), information ratios vs. thresholds (top right), and empirical percentiles vs. thresholds (bottom).

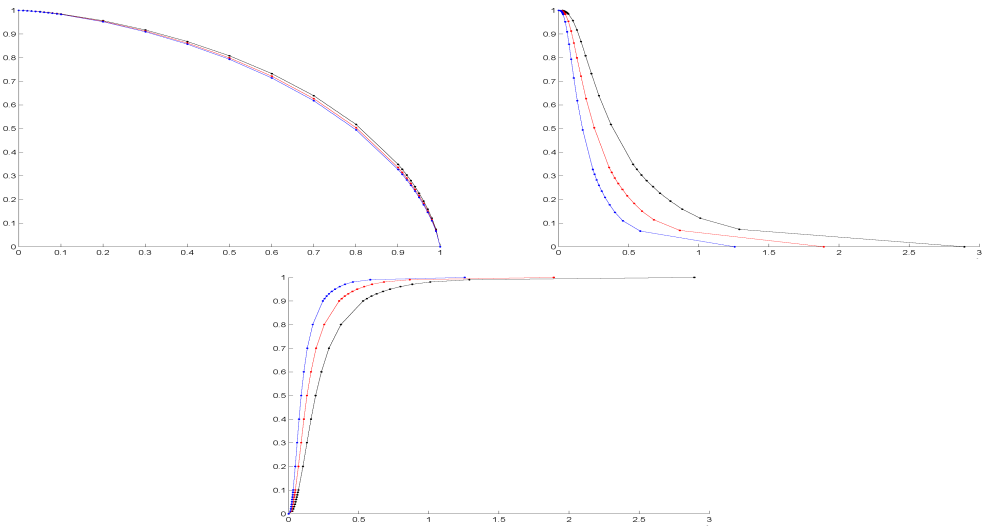


Figure 6.7: Linnik model. Region Λ with 64 locations, uniformly distributed; scale 6; $\alpha = 0.5$; $\sigma^2 = 0.1$ (black), $\sigma^2 = 0.25$ (red), $\sigma^2 = 0.5$; (blue). Information ratios vs. empirical percentiles (top left), information ratios vs. thresholds (top right), and empirical percentiles vs. thresholds (bottom).

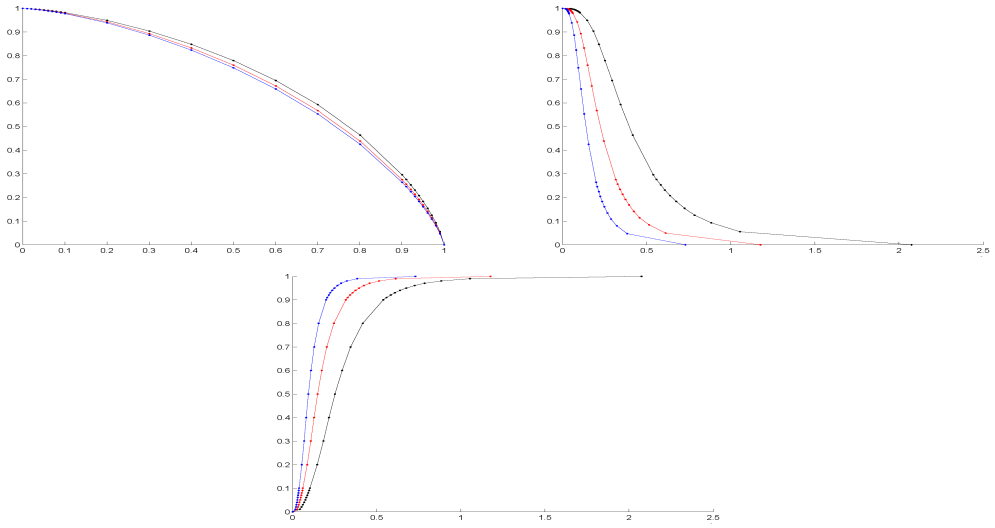


Figure 6.8: Linnik model. Region Λ with 64 locations, uniformly distributed; scale 6; $\alpha = 0.75$; $\sigma^2 = 0.1$ (black), $\sigma^2 = 0.25$ (red), $\sigma^2 = 0.5$; (blue). Information ratios vs. empirical percentiles (top left), information ratios vs. thresholds (top right), and empirical percentiles vs. thresholds (bottom).

information ratio empirical percentile is selected according to the signal-to-noise ratio. This choice provides the corresponding threshold via the empirical cumulative distribution function. The influence of the local singularity of the model of interest on the quality of the estimates (discrimination between signal and noise energy in the wavelet coefficients) is also analyzed under a strong-dependence model ($\alpha = 0.15$) for the process of interest, displaying high local singularity behavior.

The two upper plots of Figures 6.9 and 6.10 represent the values of the process of interest, generated with $\alpha = 0.15$, at 64 uniformly distributed locations of interest, and the corresponding entropy-based shrinkage estimates obtained from observed values in the cases where the intensity of the noise is $\sigma^2 = 0.1$ and $\sigma^2 = 0.5$, respectively. The thresholds selected from the empirical distribution of the information ratios at scale 6, in the two cases considered, are the values corresponding to: the empirical percentile 0.05 for $\sigma^2 = 0.1$, and the empirical percentile 0.4 for $\sigma^2 = 0.5$. The plots at the bottom of the two figures represent local mean-square error estimates based on 100 independent realizations of the interest-observation model in each case.

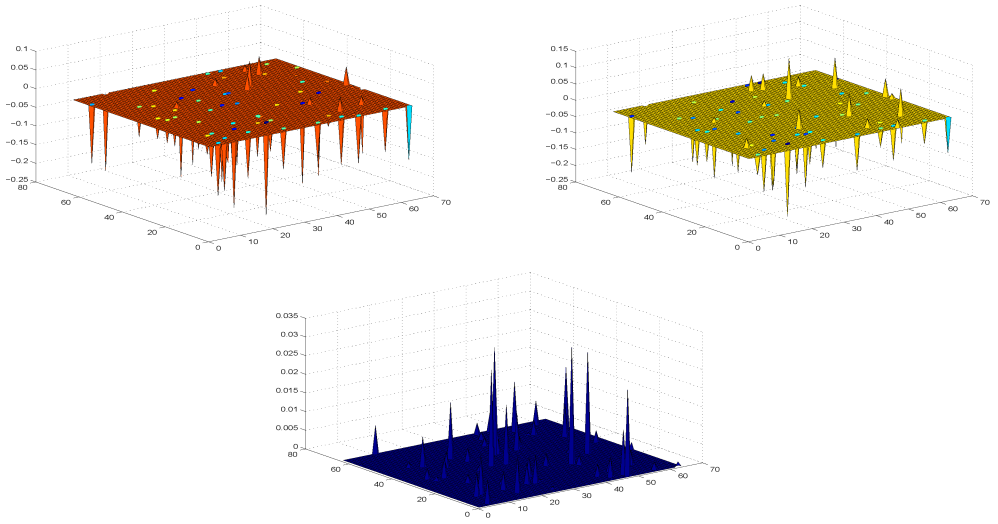


Figure 6.9: Linnik model. Region Λ with 64 locations, uniformly distributed; scale 6; $\alpha = 0.15$, and $\sigma^2 = 0.1$. Original values of the process of interest (top left), and entropy-based shrinkage estimates (top right); local mean-square error estimates based on 100 independent replicates (bottom).

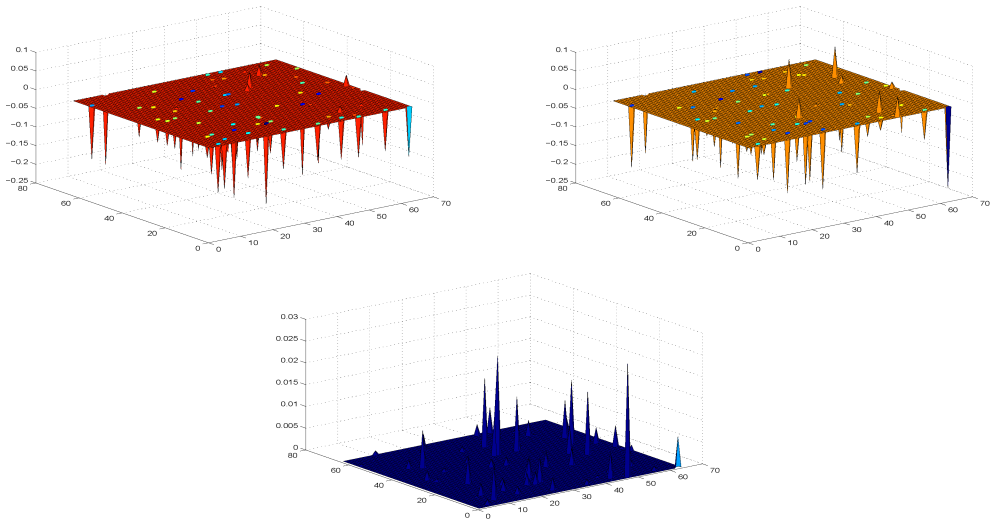


Figure 6.10: Linnik model. Region Λ with 64 locations, uniformly distributed; scale 6; $\alpha = 0.15$, and $\sigma^2 = 0.5$. Original values of the process of interest (top left), and entropy-based shrinkage estimates (top right); local mean-square error estimates based on 100 independent replicates (bottom).

To summarize, in this chapter, a two stage correlated non-linear shrinkage estimation methodology is proposed. The first stage consists of a locally correlated entropy-based block hard thresholding rule. In the second stage, a global correlated shrinkage function is considered. The main contribution lies on the design of the first stage, based on computing mutual information quantities, since random spatial signals can be filtered with this methodology, in contrast with the deterministic character of the process of interest in the classical formulation of shrinkage rules. The first stage can be interpreted in the context of spatial sampling design based on entropy, with the region of candidates being given in terms of blocks defined from the two-dimensional wavelet transform of the observed random variables. Macro- and micro-scale properties of the variables of interest can then be analyzed separately, in terms of the respective low and high resolution levels. Finally, the numerical examples computed illustrate the performance of the estimation methodology formulated, depending on the parameter values of the spatial dependence model of interest, the noise level of the observation and the design of the region of interest.

Chapter 7

A Deformation/Blurring-Based Spatio-Temporal Model

In Chapters 3 and 5, the effect of space or time deformation on the structure of threshold exceedance sets has been investigated in relation to different aspects and under different methodological perspectives. In the spatio-temporal context, where one of the key objectives, as mentioned before, consists of providing an appropriate representation of complexity in interactive spatio-temporal dynamics inherent to real phenomena, spatial deformation and its propagated effect through time provides a meaningful way to describe certain heterogeneous dynamics; in particular, in relation to unstable media or to account for the possible effect of covariates, to mention some significant interpretations.

In this chapter, a spatio-temporal model, in continuous space and discrete time, based on spatial deformation and blurring, which provides a suitable representation for a variety of environmental applications, is formulated. Some significant aspects regarding the temporal structure of threshold exceedance sets under different scenarios are explored and discussed based on simulation. Formal consistency of the model towards an underlying continuous time representation, from which the discrete-time version derives in terms of interval sampling, is investigated.

7.1 Model formulation

In the last years, there has been an increasing interest and effort to formulate and study new families of space-time statistical models suitable to represent a variety of features encountered in real phenomena. A first stage in these developments was to consider second-order models with non-separable space-time covariance, then involving a proper spatio-temporal interaction dynamics. Most proposals were restricted to the stationary case (jointly in space and time); see, for instance Brown *et al.* (2000)[29], Christakos (2000)[34], Cressie and Huang (1999)[39], De Iaco *et al.* (2002)[64], Gneiting (2002)[53], Gneiting and Schlather (2004)[55], Kyriakidis and Journel (1999)[74], Ma (2003)[79], Porcu *et al.* (2008)[96] and Stein (2005)[109]; see also Kolovos *et al.* (2004)[72] and Ma (2008)[80] for a synthetic review and discussion on different approaches adopted in this context. More recently, in a second stage, the interest has been focused on providing new generalized versions and methods with the aim to represent heterogeneities in different forms, hence giving a more sophisticated representation for real complexity. This generally involves a higher analytical and statistical complexity, as well as a parallelly higher computational cost derived from more demanding implementation.

In this chapter, we propose an approach for extended spatio-temporal dynamics modeling based on continuous spatial deformation affecting the evolution of the process. This has several possible interpretations. On one hand, it can be viewed as a means to represent the spatio-temporal dynamics of processes evolving in unstable media, subject to deformation over time. On the other hand, it can be used, in a statistical model setting, to account for the effect of certain covariables on the regular behavior of the process of interest. In both cases, the common idea is to introduce in the model heterogeneities which can be explained under a mechanics perspective: the progressive transformation of coordinate points, producing distortion on the regular structure of the process, or, in a dual sense, departure of the dependence structure from a reference simpler regular model, which can be explained in terms of a certain transformation of coordinates. Essentially, in our extended formulation, deformation affects the transfer of information between successive times.

In our study, we start from the blur-generated non-separable space-time model introduced by Brown *et al.* (2000)[29], noting that the same methodology can be applied to other formulations. This model was defined, first, as a continuous-space discrete-time

process following the state equation

$$Y(\boldsymbol{s},t) = aY[h](\boldsymbol{s},t-1) + Z(\boldsymbol{s},t), \tag{7.1}$$

with $Y(\boldsymbol{s},0) = Y_0$ being a spatial random field on $\mathbb{R}^d$, Z a spatio-temporal random field on $\mathbb{R}^d \times \mathbb{R}$, h a blurring kernel, and a a constant such that $0 < a < 1$.

As mentioned by the authors, a physical interpretation is that the model represents a diffusive dynamics given by the blurring effect in the time first-order autoregression, leading to non-separability. The model formulation proposed by Brown *et al.* (2000)[29] was restricted to the stationary case, generated, among other assumptions mentioned above, by the consideration of a homogeneous blurring kernel h.

A natural way, from our point of view, to extend this model to represent possible heterogenous dynamics would be to consider a non-homogeneous kernel h in model (7.1) (see Angulo and Madrid 2008[11], 2009a[12], 2009b[13] and 2010a[15]). However, to make the model analytically tractable and physically interpretable, the specification on h must be subject to appropriate conditions. Here, we remind that in Chapter 3 we saw how the iteration of spatial deformation and blurring with a homogeneous kernel can be interpreted as a certain form of non-homogeneous blurring, with a high flexibility in the resulting transformation in relation to the potential choices of deformation. Our proposal here, then, consists of generalizing the formulation of model (7.1) in the following way. For a spatial deformation Φ, and depending on whether Y represents a 'level' or a 'flow' magnitude, our deformation/blurring-based space-time model is formulated as

$$Y(\boldsymbol{s},t) = aY[\Phi;h](\boldsymbol{s},t-1) + Z(\boldsymbol{s},t), \tag{7.2}$$

or

$$Y(\boldsymbol{s},t) = aY[\tilde{\Phi};h](\boldsymbol{s},t-1) + Z(\boldsymbol{s},t), \tag{7.3}$$

respectively, where all the elements involved, besides Φ and $\tilde{\Phi}$, are interpreted as before. The differences between both the level and flow cases regarding the spatio-temporal dynamics and, as a consequence, the geometrical characteristics of excursion sets are particularly noticeable due to cumulative propagation effects under the autoregressive model scheme. Now, the physical interpretation given by Brown *et al.* (2000)[29] is extended in the sense that the evolution mechanism of the process involves a movement

of coordinates during each time interval which can be either associated to unstability of the medium or, alternatively, to properly heterogeneous diffusion, possibly derived from the effect of certain covariates on the system. In fact, Brown *et al.* (2000)[29] suggested a simple extension of their model considering a non-centered homogeneous kernel, with the constant mean vector accounting, for example, for the effect of wind with a constant speed and direction. Our model clearly gives a more flexible and realistic way to implement such a type of effect.

A somehow related model was proposed by Huang and Hsu (2004)[62], which incorporates the magnitude and direction of wind as covariates using an "ad hoc" parametric setup, but maintaining the expression of the blurring operator in the original coordinates. In our case, we act directly on Y_{t-1} in terms of deformation, which is equivalent to making a change of spatial coordinates (deformation), jointly with a rescaling of states, in terms of the deformation Jacobian, in the case of 'flow' type variables.

In the following sections, we explore and study different aspects related to the analytical properties of model (7.2), and to the behavior of sample path in the sense described in Chapter 3. Specifically, in Section 7.2 we show by simulation how model (7.2) can lead to heterogeneity in the geometrical characteristics associated to threshold exceedance sets, under different scenarios. In particular, how spatial deformation induces locally heterogeneous temporal dynamics in terms of the threshold exceedance patterns. In Section 7.3, we study some analytical aspects, particularly in relation to possible self-consistency of the model in terms of an underlying continuous-time interpretation, leading to the concept of infinitely divisible deformation with an infinitesimal generator.

7.2 An exploration on geometrical characteristics under different scenarios

The following examples based on simulated realizations under model (7.2) illustrate various significant aspects of geometrical characteristics in relation to the information transfer in input-output systems. We consider two types of input random fields. In the first type, $Z^{[1]}$ is assumed to be a zero-mean Gaussian spatio-temporal random field, with second-order structure defined in terms of Gneiting model (see Section 2.2.2), specifically

as

$$C(r,\tau) = 0.1(1+\tau^2)^{-1.25}\left(1+\frac{r^2}{(1+\tau^2)^{0.1}}\right)^{-0.1}. \tag{7.4}$$

The marginal time cross-sections of model (7.4) (i.e. for each fixed t) correspond to Cauchy spatial model. The difference with the second type is that now $Z^{[2]}$ is uncorrelated in time. The spatial domain is restricted to the square $D = [0, 400]^2 \subset \mathbb{R}^2$. Deformation Φ ($\tilde{\Phi}$ when the Jacobian changes the state scale) is defined by a nonhomogeneous clockwise rotational movement with a certain attraction towards the central point of the square, $(200, 200)$; specifically, through homothetic transformation of D onto the unit square $[-1/2, -1/2]^2$, the deformation is given by the following change of polar coordinates:

$$\begin{aligned} \theta' &= \theta - \gamma(1-2r)^{\beta}, \\ r' &= r(1-\delta(1-2r)^{\alpha}), \end{aligned}$$

restricted to the circle $B((0,0), 1/2)$. Here, we take $\gamma = 0.1$, $\beta = 2$, $\delta = 0.25$, and $\alpha = 5$. We can see the clear contraction effect in the central area observing Figure 7.1, which shows the Jacobian determinant values of this deformation.

A Gaussian kernel h with different variances is used to produce blurring. For convenience, the time unit considered is $1/5$. Starting from a zero field, the model is run through 300 times.

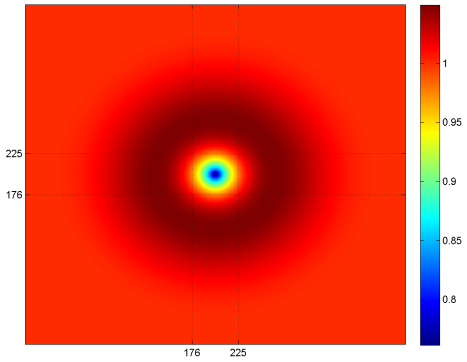
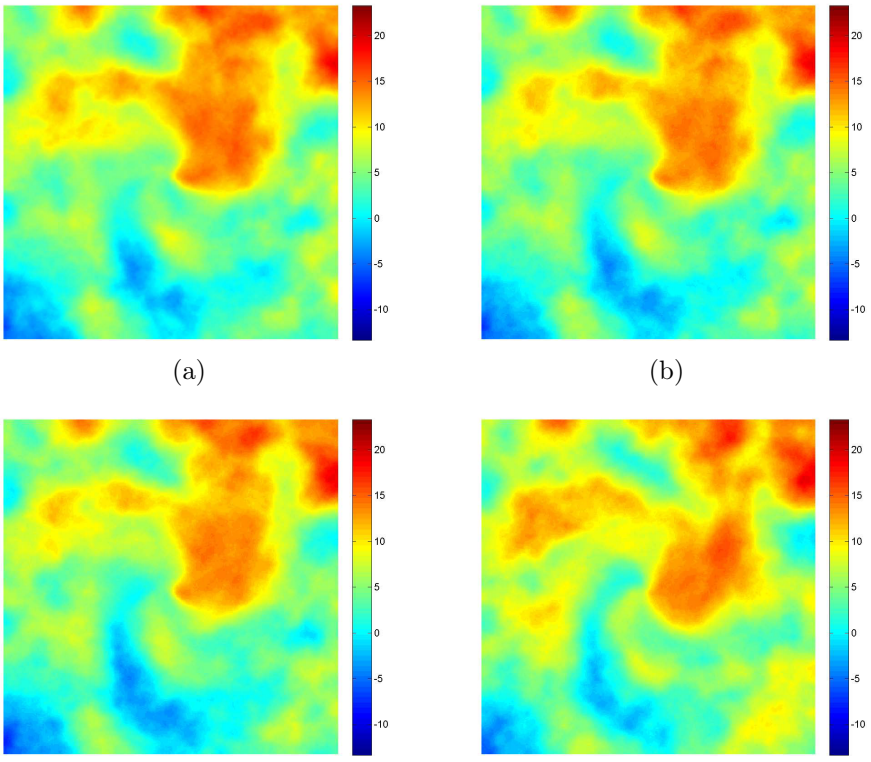


Figure 7.1: Jacobian determinant values, $|J_{\Phi}(\Phi^{-1}(s))|$.

Figure 7.2 shows four spatial cross-sections of the output Y corresponding to times $t = 100$, 103, 106, and 116, using a Gaussian kernel with $\sigma_h^2 = 0.625 \times 10^{-4}$; $a = 1 - 10^{-6}$,



(a) (b)

(c) (d)

Figure 7.2: Four spatial cross-sections of a simulated realization of the spatio-temporal random field Y, for: (a) $t = 15$, (b) $t = 18$, (c) $t = 21$, and (d) $t = 30$, with $Z^{[1]}$, $\sigma_h^2 = 0.625 \times 10^{-4}$ and $a = 1 - 10^{-6}$.

close to 1, to produce a smooth transition between subsequent times; the input random field is $Z^{[1]}$. The first three times give an impression of the inertial spatial changes over time; the fourth is included for further vision of their cumulative effect. In Figure 7.3 we represent four spatial cross-sections of the output Y, again corresponding to same times and parameter values but now with the input random field being $Z^{[2]}$. Comparing both figures, we can see that the effect of deformation is clear in all plots, but the spatial cross-sections corresponding to $Z^{[1]}$ as input suggest a slower movement over time, and lower local variability is observed in this case due to the more inertial behavior of the input random field.

To explore differences in the recurrence and duration patterns of the spatio-temporal threshold exceedance sets in relation to the model components, we calculate the Euler

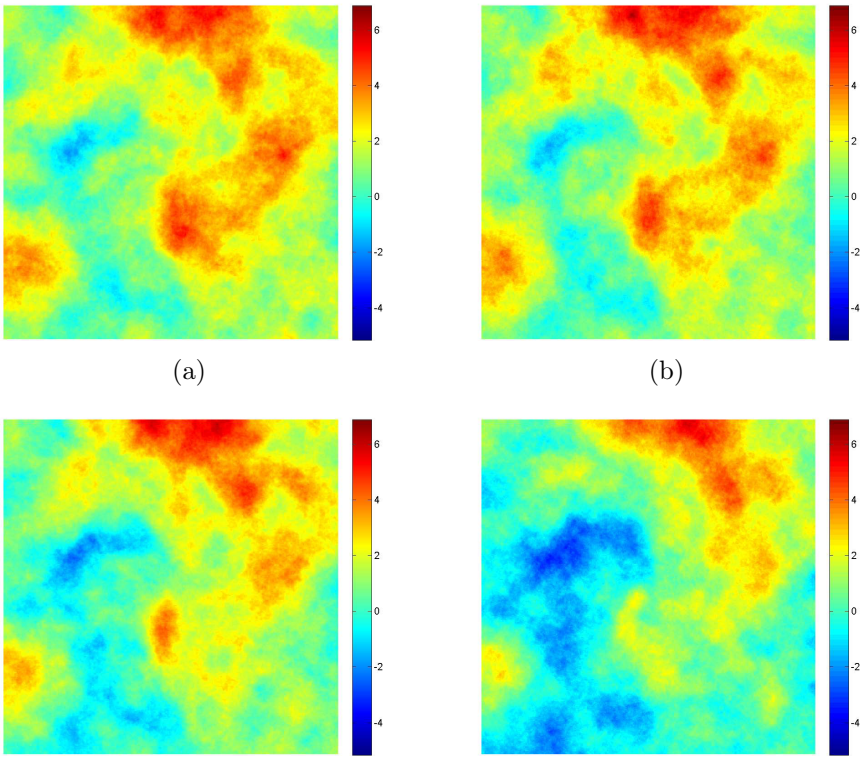
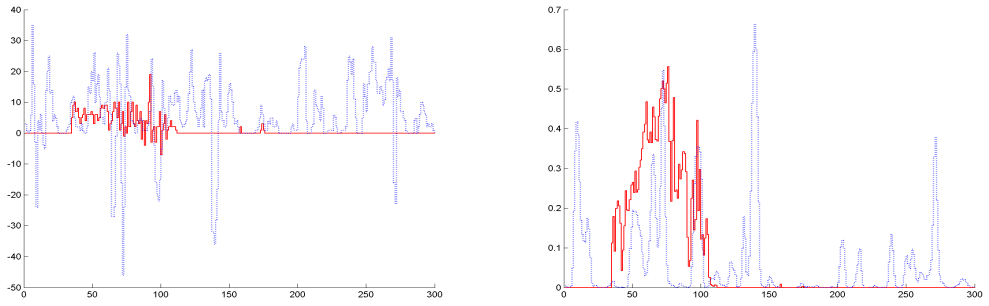


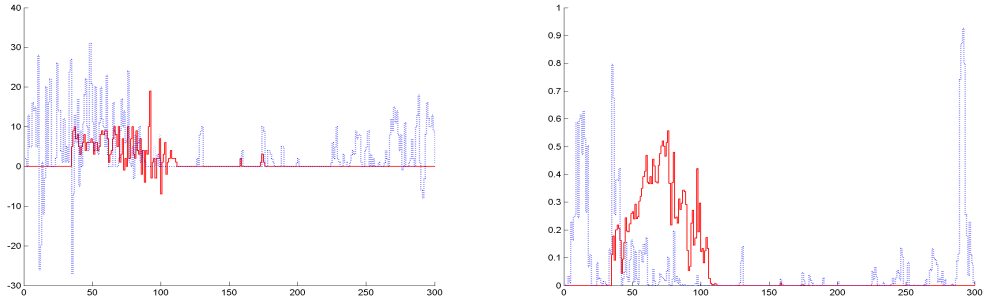
Figure 7.3: Four spatial cross-sections of a simulated realization of the spatio-temporal random field Y, for: (a) $t = 15$, (b) $t = 18$, (c) $t = 21$, and (d) $t = 30$, with $Z^{[2]}$, $\sigma_h^2 = 0.625 \times 10^{-4}$ and $a = 1 - 10^{-6}$.

characteristic and the hypervolume (i.e., area) for the 300 spatial cross-sections on the central subdomain $S = [176, 225]^2$ (see Figure 7.1). Since the input and output random fields evolve with different scales of variation, we take for comparison thresholds related to the sample standard error $\hat{\sigma}$ based on the set of data for the corresponding random field. The results obtained for random field $Z^{[1]}$ and for random field Y with input $Z^{[1]}$, $\sigma_h^2 = 0.625 \times 10^{-4}$ and $a = 1 - 10^{-6}$, are displayed in Figure 7.4. Apart from $Z^{[1]}$ showing a generally higher structural complexity than Y in the corresponding threshold exceedance sets (see scales in plot (a)), we can observe stronger inertial behavior in Y reflected in the occurrence of more persistent critical episodes, compared to the number of critical episodes with shorter duration in $Z^{[1]}$, as well as a certain temporal delay in the input-output transfer in this context (better noticed in plot (b)).



(a) (b)

Figure 7.4: Sequences of (a) Euler characteristic and (b) hypervolume values for spatial cross-sections over time and threshold $u = 1.5\hat{\sigma}$, for processes $Z^{[1]}$ (dotted blue), and corresponding output Y (red) with $\sigma_h^2 = 0.625 \times 10^{-4}$ and $a = 1 - 10^{-6}$.



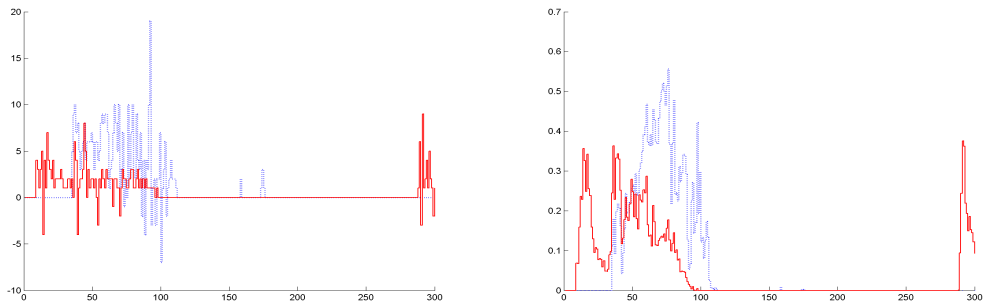
(a) (b)

Figure 7.5: Sequences of (a) Euler characteristic and (b) hypervolume values for spatial cross-sections over time and threshold $u = 1.5\hat{\sigma}$, for process Y, with $Z^{[2]}$, $\sigma_h^2 = 0.625 \times 10^{-4}$ and $a = 1 - 10^{-1}$ (dotted blue), $a = 1 - 10^{-6}$ (red).

In Figure 7.5 we compare the results for Y based on the input $Z^{[2]}$, with $\sigma_h^2 = 0.625 \times 10^{-4}$ and $a = 1 - 10^{-6}$, to the results obtained for the same model using $a = 1 - 10^{-1}$ and the same remaining parameter values and realization of $Z^{[2]}$. The latter case implies a lesser inertial behavior of Y over time, and a higher influence of the contemporaneous value of the input $Z^{[2]}$ at every time. Although the differences in the structural complexity of threshold exceedance sets (see scales in plot (a)) are not as clear as in the previous study, the recurrence pattern for critical episodes of the output Y gets closer to that of the input Z, still with a certain delay, in the latter case, which makes evident the

significance of the system parameter a in this respect.

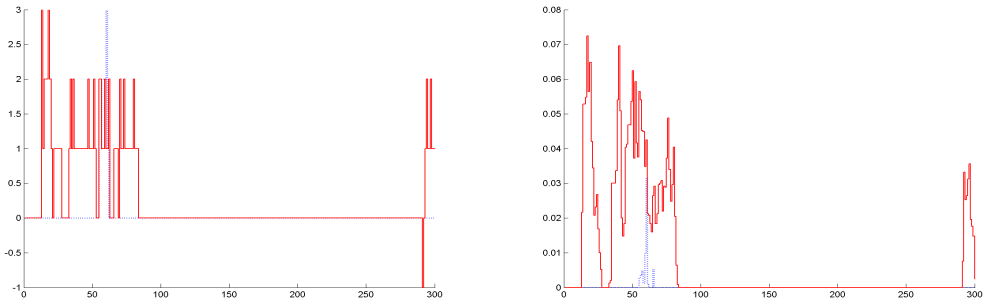
Now we study the deformation effects depending on the type of magnitude ('level' and 'flow' cases). On the one hand, when a 'flow' magnitude is considered, since the deformation has properties of contraction on the subregion S, the Jacobian factor enhances higher-state values (see Section 3.2), thus contributing to the occurrence of events above a certain threshold. On the other hand, given the autoregressive nature of the model, we expect that the influence of the past times be translated into a larger persistence of such events with a stronger structuring. These properties are visible at the plots of Figure 7.6, where we represent the Euler characteristic (plot a)) and hypervolume (plot b)) for the model using both types of magnitudes. In addition, since the Jacobian increases the possibly of higher values, the events over higher thresholds are more probable in the flow than in the level case (see Figure 7.7). That is, the excess and hypervolume in risk tend to be larger in the case of a 'flow' magnitude.



(a) (b)

Figure 7.6: Sequences of (a) Euler characteristic and (b) hypervolume values for spatial cross-sections over time and threshold $u = 1.5\hat{\sigma}$, for process Y, with $Z^{[2]}$, $\sigma_h^2 = 0.625 \times 10^{-4}$, $a = 1 - 10^{-6}$, using $\tilde{\Phi}$ (red), and Φ (dotted blue).

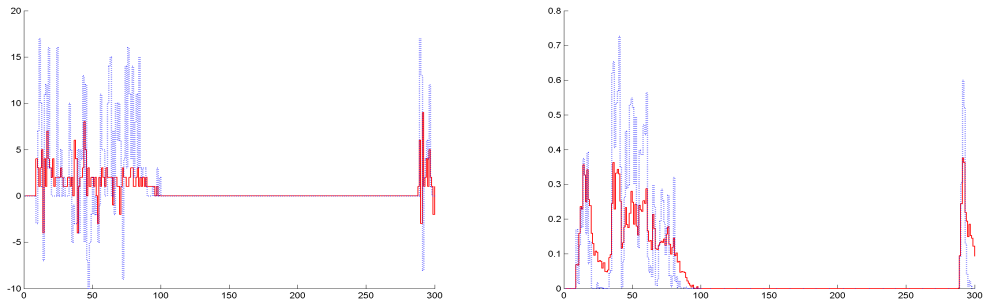
To analyze the influence of the blurring operator we consider two kernels with different variances. The blurring acts on the output variable by transferring spatially averaged past information. If the average is computed taking into account locations in a larger neighborhood, with a higher smoothing effect, then the possible past extreme events lose influence through the time transfer and the current input variable has a higher direct effect on the output. This means that processes with more disperse kernels will have a lower structuring, in the sense of more variability in the value of Euler characteristic and



(a) (b)

Figure 7.7: Sequences of (a) Euler characteristic and (b) hypervolume values for spatial cross-sections over time and threshold $u = 2.5\hat{\sigma}$, for process Y, with $Z^{[2]}$, $\sigma_h^2 = 0.625 \times 10^{-4}$, $a = 1 - 10^{-6}$, using $\tilde{\Phi}$ (red), and Φ (dotted blue).

area at risk due to the randomness of Z. These effects can be emphasized if, in addition, a 'flow' type magnitude is considered (see figure 7.8).

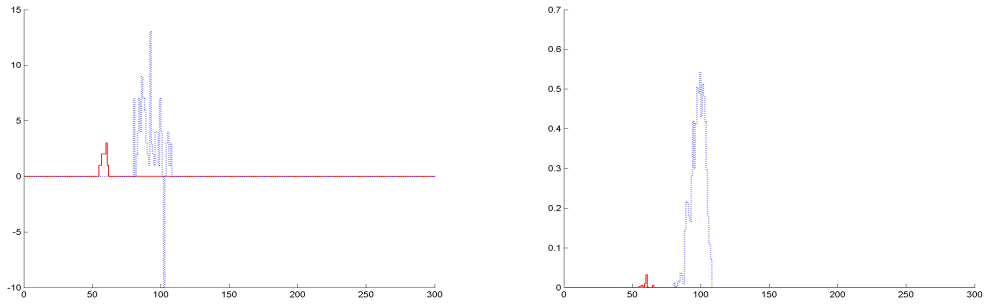


(a) (b)

Figure 7.8: Sequences of (a) Euler characteristic and (b) hypervolume values for spatial cross-sections over time and threshold $u = 1.5\hat{\sigma}$, for process Y, with $Z^{[2]}$, $a = 1 - 10^{-6}$, and $\sigma_h^2 = 0.625 \times 10^{-4}$ (red), and $\sigma_h^2 = 0.625 \times 10^{-3}$ (dotted blue).

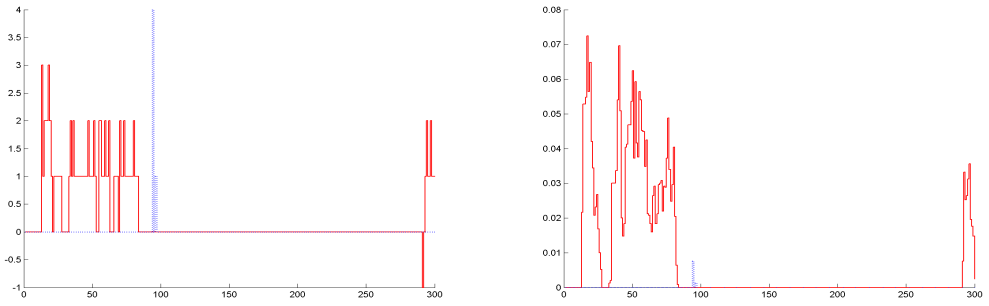
Finally, in Figures 7.9 and 7.10 we study the behavior of model Y in two subregions where the deformation has different contraction/dilation properties: the central square $[176, 225] \times [176, 225]$ considered above, and the square $[176, 225] \times [111, 160]$ (see Figure 7.1). If the magnitude is of 'level' type and deformation has local contraction properties, the blurring operator will have stronger smoothing effect, because of spatial accumu-

lation of past state values in the average. Therefore, in Figure 7.9 we can see a lower complexity in the structuring, and reduced areas at risk in central region compared to the corresponding in the alternative region. On the other hand, if the magnitude is of 'flow' type, the Jacobian effect leads to amplification/reduction of the extreme state according to the local contraction/dilation properties of the deformation. Thus, in the region where the deformation has contraction properties, there is a higher probability of observing events over the threshold (see Figure 7.10).



(a) (b)

Figure 7.9: Sequences of (a) Euler characteristic and (b) hypervolume values for spatial cross-sections over time and threshold $u = 1.5\hat{\sigma}$, for process Y, with $Z^{[2]}$, $a = 1 - 10^{-6}$, and $\sigma_h^2 = 0.625 \times 10^{-4}$ in central region (dotted blue), and in alternative region (red).



(a) (b)

Figure 7.10: Sequences of (a) Euler characteristic and (b) hypervolume values for spatial cross-sections over time and threshold $u = 1.5\hat{\sigma}$, for process Y, with $Z^{[2]}$, $a = 1 - 10^{-6}$, and $\sigma_h^2 = 0.625 \times 10^{-4}$ in central region (dotted blue), and in alternative region (red).

7.3 Self-consistency for a continuous-time generating process

Model (7.2), as formulated, is well defined in the sense that, starting from an initial random field Y_0, the autoregressive scheme leads to specification of the states Y_t for successive times $t \in \mathbb{N}$. Although the model can be used in this form for statistical fitting, the idea of having an underlying continuous-time process generating this type of recursion for discrete-time sampling has interest in relation to possible physical interpretations. In this respect, the key point, commonly adopted in simulation algorithms, is to establish conditions under which the discrete-time recursive scheme is reproducible under proper sampling refinements. Such an approach is adopted in Brown *et al.* (2000)[29] for model 7.1, with the conclusion that, in order to have self-consistency, kernel h must be infinitely divisible, that is, it must rescale preserving its form under convolution of any order.

In this section, we study self-consistency of model (7.2) according to the same idea of reproducibility, and show that an extra condition of infinite divisibility (in the sense of iterated composition) is required for the deformation Φ, which is satisfied under existence of an infinitesimal generator field for such a transformation (see also Angulo and Madrid 2009b[13]). For simplicity, we here restrict technical details to the case of a 'level' type magnitude. A similar development can be derived for a 'flow' type variable, involving the iterated product of Jacobian terms.

Let model (7.2) be written for any non-necessarily unit discrete-time sampling interval Δ in the following form:

$$Y(\boldsymbol{s},t) = a_\Delta Y[\Phi_\Delta; h_\Delta](\boldsymbol{s}, t-\Delta) + Z_\Delta(\boldsymbol{s},t). \tag{7.5}$$

Now, taking $\delta = \Delta/n$ as the new, refined sampling interval, iteration leads to

$$\begin{aligned}
Y(\boldsymbol{s},t) &= a_\delta Y[\Phi_\delta; h_\delta](\boldsymbol{s}, t-\delta) + Z_\delta(\boldsymbol{s},t) \\
&= a_\delta h_\delta * Y(\Phi_\delta^{-1}(\boldsymbol{s}), t-\delta) + Z_\delta(\boldsymbol{s},t) \\
&= a_\delta h_\delta * \left[a_\delta h_\delta * Y(\Phi_\delta^{-1}(\Phi_\delta^{-1}(\boldsymbol{s})), t-2\delta) + Z_\delta(\Phi_\delta^{-1}(\boldsymbol{s}), t-\delta)\right] + Z_\delta(\boldsymbol{s},t) \\
&= a_\delta^2 h_\delta^{2*} * Y(\Phi_\delta^{-1} * \Phi_\delta^{-1}(\boldsymbol{s}), t-2\delta) + a_\delta h_\delta * Z_\delta(\Phi_\delta^{-1}(\boldsymbol{s}), t-\delta) + Z_\delta(\boldsymbol{s},t) \\
&= ...
\end{aligned}$$

Iterating substitution $n-1$ times, and imposing preservation of the model structure, we must have:

i) $a_\Delta = a_\delta^n$.

ii) $h_\Delta = h_\delta^{n*}$.

iii) $\Phi_\Delta = \Phi_\delta \circ \Phi_\delta \circ \ldots^{n)} \circ \Phi_\delta$.

iv) $Z_\Delta(\boldsymbol{s},t) = \sum_{k=1}^{n-1} a_\delta^k h_\delta^{k*} * Z(\Phi_\delta^{k[-1]}(\boldsymbol{s}), t-k\delta) + Z_\delta(\boldsymbol{s},t)$, where $\Phi_\delta^{k[-1]}$ stands for $\Phi_\delta^{-1} \circ \Phi_\delta^{-1} \ldots^{k)} \circ \Phi_\delta^{-1}$.

Note that conditions (i), (ii), and (iv) coincide with those given in Brown *et al.* (2000)[29], with the exception that (iv) now involves the inverse deformation. The interesting point comes from condition (iii), which means that Φ must be infinitely divisible by iterated composition (or, shortly, infinitely decomposable). Following the introduction of vector fields from particle displacement given in Rossmann (2002, Ch. 1)[100], in our case, we can define a vector field $X : \mathbb{R}^2 \to \mathbb{R}$ as a function which assigns to each point $\boldsymbol{u} \in \mathbb{R}^2$ the velocity vector $X(\boldsymbol{u})$ which represents the infinitesimal change of coordinates at $\boldsymbol{u}$ by deformation, that is,

$$X(\boldsymbol{u}) = \frac{d\Phi_\delta}{d\delta}(\boldsymbol{u}).$$

This can be viewed as the "current of the flow" (Rossmann 2002[100]); in particular, since Φ does not depend on the time, it corresponds to the case of a steady flow (see Joseph 1990[69]). Hence, X can be interpreted as an infinitesimal generator of deformation, with the equation

$$\Phi_\Delta(\boldsymbol{u}_0) = \int_0^\Delta X(\Phi_\delta(\boldsymbol{u}_0))d\delta + \boldsymbol{u}_0,$$

giving the change of coordinates from $\boldsymbol{u}_0$ provided by effect of Φ over a time interval Δ. Hence, any transformation Φ having an associated velocity field X (in the above sense) satisfies the infinitely decomposibility condition (iii), and can lead to a self-consistent deformation/blurring-based model formulation (7.2).

In summary, in this chapter a deformation/blurring-based spatio-temporal model, suitable, for instance, for applications involving dynamical effects derived from unstability of the medium and possible covariates, is formulated, giving, in particular, an

extension of the blur-generated model introduced by Brown *et al.* (2000)[29]. Certain aspects related to the geometrical properties of threshold exceedance sets are explored in this context; in particular, the influence of systemic input-output parameters on the transfer of recurrence and duration patterns of critical episodes. Analytical conditions on the deformation to possibilitate a consistent underlying continuous-time representation are investigated.

Chapter 8

Conclusions and Future Research Lines

In this thesis work, we have addressed various aspects related to the analysis of threshold exceedance sets for spatial/temporal processes.

Specifically, in Chapter 3 we define 'deformed' and 'blurred' random fields, analyze some formal implications of the transformations involved, and illustrate the effects of such transformations on the second-order structure and on some geometrical characteristics of the threshold exceedance sets. Among possible extensions of this study, we first look at the implementation of stochastic deformations, a fundamental step to incorporate statistical information related to random media and from random covariates in many applications, as well as consideration of time adaptive thresholds, or thresholds defined in relation to certain functionals of interest.

In Chapter 4, we design a methodology based on associating a family of spatial (marked) point processes to the excursion sets for different thresholds, considering the fragmentation of such sets into connected components. Then, we apply (marked) point process techniques to analyze certain significant characteristic of threshold exceedance sets. Here we have taken as marks the sizes and orientations of the corresponding connected components. This study can be immediately extended considering new marks of interest, as well as random fields in higher dimensions. A main aspect for continuing work is the application of this methodology to study further the effects of deformation and blurring transformations, as well as considering more general thresholds.

In both of the previous approaches, we have focused on the exploration of different aspects in relation to the underlying model properties such as local variability and

dependence ranges, and we show that many of the observed features in the threshold exceedance sets are inherited from the random field from which they are generated. For this purpose, in Chapters 3 and 4 we use random fields belonging to the Cauchy class. Other more general models can be considered. For instance, given that we have seen that from a realization of an isotropic random field, by gradually increasing the threshold, the degree of local anisotropy of threshold exceedance sets is modified, consideration of properly anisotropic models can be of particular interest. A main line for extending this study is the consideration of more complex forms of heterogeneity, jointly with generalized thresholds.

In Chapter 5 we consider quantifiers based on the Continuous Wavelet Transform for detecting heterogeneous concentrations of energy in a signal at different scales, which may indicate a certain degree of intermittency. Through these quantifiers, we examine the effects of time deformation, according to its local contraction/dilation properties, on the distribution of energy at different scales. In fact, certain forms of intermittent behavior can be expected to be approximately explained in terms of deformations based on more regular reference models. In this context, further research towards inverse estimation methods for such deformations constitutes one of the main aspects of interest. Also, during the development of this chapter, we have considered the potencial design of new measures which reflect, in a more precise way, the presence of intermittency; for instance, taking into account the possible coherence among the energy distributions at different scales. In this respect, we are studying the use of generalized entropies, as normalized Tsallis entropy, applied to the Discrete Wavelet Transform coefficients.

In Chapter 6, we have developed a correlated non-linear shrinkage estimation methodology appropriate for filtering random spatial signals. The procedure uses in a first stage a scale-dependent block hard thresholding rule, based on mutual information criteria, which allows to analyze, separately, macro- and micro-scale properties of the variable of interest, considering low and high resolutions levels, respectively. The second stage considers a global correlated shrinkage function. Some research lines remain open, particularly in relation to the asymptotic properties of the class of estimators proposed. In this respect, the Gaussian case can be easily addressed, since the second stage becomes linear, and numerical projection methods in terms of suitable bases can be applied.

Finally, in Chapter 7, we study the dynamical joint effect of deformation and blurring transformations on the threshold exceedance sets in a spatio-temporal context, continuing

the work initiated in Chapter 3, and providing an extension of the blur-generated non-separable space-time model introduced by Brown *et al.* (2000)[29]. Some formal aspects in relation to the properties the input process must satisfy for self-consistency and a possible continuous-time representation still remain open for continuing research. Other important extensions include, as before, the consideration of time-varying and stochastic deformations, dynamical inverse estimation, as well as application of this approach for incorporation of deformation to other existing space-time model formulations.

The above mentioned contents of this research are presented here in thematic separate chapters. However, they have clear interconnections according to various aspects; namely, the effect of deformation, analysis of structural characteristics, influence of threshold specifications, effect of model parameters, among others. In this sense, the extensions proposed chapter by chapter should be viewed from a synthetic point of view, in a more general framework, particularly aimed at various generalizations concerning deformations and thresholds, as well as extensions in the spatio-temporal context and exploitation in applications in relation to risk indicators.

Capítulo 8

Conclusiones y Futuras Líneas de Investigación

En este trabajo de tesis, hemos abordado varios aspectos relativos al análisis de conjuntos de excedencias de umbrales para procesos espaciales/temporales.

Específicamente, en el Capítulo 3 definimos campos aleatorios afectados por deformación y borrosidad, analizamos algunas implicaciones formales de dichas transformaciones e ilustramos los efectos de las mismas sobre la estructura de segundo orden y sobre algunas características geométricas de los conjuntos de excedencias de umbrales. Entre las posibles extensiones de este estudio, contemplamos en primer lugar la implementación de deformaciones estocásticas, un paso fundamental para la incorporación de información estadística relacionada con medios aleatorios y de covariables en muchas aplicaciones, así como la consideración de umbrales adaptativos en el tiempo, o de umbrales definidos en relación con ciertos funcionales de interés.

En el Capítulo 4, diseñamos una metodología basada en asociar una familia de procesos puntuales (marcados) a los conjuntos de excursión para diferentes umbrales, considerando la fragmentación de tales conjuntos en componentes conexas. Entonces, aplicamos técnicas de procesos puntuales (marcados) al análisis de ciertas características significativas de los conjuntos de excedencias de umbrales. Aquí hemos tomado como marcas los tamaños y las orientaciones de las correspondientes componentes conexas. Este estudio puede extenderse de forma inmediata considerando nuevas marcas de interés, así como campos aleatorios en dimensiones más altas. Un aspecto principal para la

continuación de esta línea es la aplicación de esta metodología también al estudio de los efectos de transformaciones de deformación y borrosidad, así como bajo la consideración de umbrales más generales.

En ambos planteamientos previos, hemos enfocado el trabajo a la exploración de diferentes aspectos en relación con propiedades del modelo subyacente tales como variabilidad local y rangos de dependencia, y mostramos cómo muchos de los rasgos observados en los conjuntos de excedencias de umbrales son heredados del campo aleatorio a partir del cual son generados. Con ese propósito, en los Capítulos 3 y 4 utilizamos campos aleatorios pertenecientes a la clase de Cauchy. Otros modelos más generales podrían considerarse. Por ejemplo, dado que hemos comprobado que a partir de una realización de un campo isotrópico, incrementando gradualmente el umbral, se modifica el grado de anisotropía local de los conjuntos de excedencias de umbrales, la consideración de modelos propiamente anisotrópicos puede ser de particular interés. Una línea principal para la extensión de este estudio consiste en la consideración de formas más complejas de heterogeneidad, junto con umbrales generalizados.

En el Capítulo 5 consideramos ciertos cuantificadores basados en la Transformada *Wavelet* Continua para la detección de concentraciones heterogéneas de energía en una señal a diferentes escalas, lo que puede indicar un cierto grado de intermitencia. A través de dichos cuantificadores, examinamos los efectos de la deformación del tiempo, de acuerdo con sus propiedades locales de contracción/dilatación, sobre la distribución de la energía a diferentes escalas. De hecho, es esperable que ciertos comportamientos de intermitencia puedan explicarse de forma aproximada en términos de deformación basándose en modelos de referencia más regulares. En este contexto, la investigación de métodos de estimación inversa para tales deformaciones constituye uno de los principales aspectos de interés. Asimismo, durante el desarrollo de este capítulo, hemos considerado el diseño potencial de nuevas medidas que reflejen, de una forma más precisa, la presencia de intermitencia; por ejemplo, teniendo en cuenta las posibles coherencias entre las distribuciones de la energía a diferentes escalas. A este respecto, estamos estudiando el uso de entropías generalizadas, como la entropía de Tsallis, aplicadas a los coeficientes de la Transformada *Wavelet* Discreta.

En el Capítulo 6, hemos desarrollado una metodología de estimación no lineal correlada *shrinkage* apropiada para el filtrado de señales espaciales aleatorias. El procedimiento usa en una primera fase una regla *block hard thresholding* dependiente de la

escala, basada en criterios de información mutua, que permite analizar, de forma separada, propiedades a macro- y micro-escala de la variable de interés, considerando niveles de resolución bajos y altos, respectivamente. En la segunda fase, se considera una función *shrinkage* correlada global. Permanecen abiertas algunas líneas, particularmente en relación con las propiedades asintóticas de la clase de estimadores propuesta. En este contexto, el caso gaussiano puede abordarse de forma sencilla, puesto que la segunda fase se convierte en lineal y pueden aplicarse métodos de proyección numérica en términos de bases apropiadas.

Finalmente, en el Capítulo 7 estudiamos el efecto dinámico conjunto de transformaciones de deformación y borrosidad sobre los conjuntos de excedencias de umbrales en un contexto espacio-temporal, continuando el trabajo iniciado en el Capítulo 3, y proporcionando una extensión del modelo espacio-temporal no separable generado mediante borrosidad introducido por Brown *et al.* (2000)[29]. Algunos aspectos formales en relación con las propiedades que debe satisfacer el proceso input para auto-consistencia del modelo y una posible representación en tiempo continuo aún permanecen abiertos para la continuación de la investigación. Otras extensiones importantes incluyen, como anteriormente, la consideración de deformaciones variables en el tiempo y estocásticas, estimación inversa dinámica, así como aplicaciones de este enfoque para la incorporación de la deformación a otras formulaciones existentes de modelos en el espacio-tiempo.

Los contenidos antes mencionados de esta investigación son presentados en esta memoria en capítulos temáticos separados. Sin embargo, éstos tienen claras interconexiones atendiendo a varios aspectos; concretamente, el efecto de la deformación, el análisis de características estructurales, la influencia de especificaciones del umbral, el efecto de los parámetros del modelo, entre otros. En este sentido, las extensiones propuestas capítulo por capítulo deben verse desde un punto de vista sintético, en un marco más general, particularmente dirigidas a varias generalizaciones concernientes a deformaciones y umbrales, así como a extensiones en el contexto espacio-temporal y a la explotación y aplicaciones en relación con indicadores de riesgo.

Bibliography

[1] Abramovich F, Sapatinas T, Silverman BW (1998) Wavelet thresholding via a Bayesian approach. *Journal of the Royal Statistical Society, Series B* **60**: 725-749

[2] Adler RJ (1981) *The Geometry of Random Fields*. Wiley, Chichester, UK

[3] Adler RJ (2008) Some new random field tools for spatial analysis. *Stochastic Environmental Research and Risk Assessment* **22**: 802-822

[4] Adler RJ, Taylor JE (2007) *Random Fields and Geometry*. Springer, New York

[5] Aldous D (1989) *Probability Approximations via the Poisson Clumping Heuristic*. Springer, New York

[6] Amiri M, Azimifar Z, Fieguth P (2008) Correlated non-linear wavelet shrinkage. *Proceedings of the International Conference on Image Processing* (ICIP 2008), San Diego, California, October 12-15, 2008, 2348-2351

[7] Amit Y, Grenander U, Piccioni M (1991) Structural image restoration through deformable templates. *Journal of the American Statistical Association* **86**: 376-387

[8] Anderes EB, Stein ML (2008) Estimating deformations of isotropic Gaussian random fields on the plane. *The Annals of Statistics* **36**: 719-741

[9] Angulo JM, Bueso MC (2001) Random perturbation methods applied to multivariate spatial sampling design. *Environmetrics* **12**: 631-646

[10] Angulo JM, Bueso MC, Alonso FJ (2000) A study on sampling design for optimal prediction of space-time stochastic processes. *Stochastic Environmental Research and Risk Assessment* **14**: 412-427

[11] Angulo JM, Madrid AE (2008) Analysis of spatio-temporal threshold exceedances. *Statistics for Spatio-Temporal Modelling. Proceedings of the 4th International Workshop on Spatio-Temporal Modelling*(METMA4), Alghero, Italy, September 24-26, Cocchi D, Mateu J, Montes F, Otranto E, Porcu E, Usai A (ed), ISBN: 88-6025-098-6, pp. 1-9

[12] Angulo JM, Madrid AE (2009a) Análisis del efecto de transformaciones de borrosidad y deformación sobre excesos de umbrales en procesos espaciales. *Proceedings of the XXXI Congreso Nacional de Estadística e Investigación Operativa y V Jornadas de Estadística Pública* (SEIO-2009), Murcia, Spain, Febrery 10-13, Editum Publicaciones, ISBN: 978-84-691-81, pp. 27

[13] Angulo JM, Madrid AE (2009b) Un modelo de interaccion espacio-temporal con heterogeneidad inducida por deformación. *Proceedings of the XXXI Congreso Nacional de Estadística e Investigación Operativa y V Jornadas de Estadística Pública* (SEIO-2009), Murcia, Spain, Febrery 10-13, Editum Publicaciones, ISBN: 978-84-691-81, pp. 120

[14] Angulo JM, Madrid AE (2009c) Wavelet-based multiscale intermittency analysis in environmental applications. *Proceedings of TIES 2009 - the 20th Annual Conference of the International Environmetrics Society and GRASPA Conference 2009*, Bologna, Italy, July 5-9, pp. 19

[15] Angulo JM, Madrid AE (2010) Structural analysis of spatio-temporal threshold exceedances. *Environmetrics* **21**: 415-438

[16] Angulo JM, Madrid AE, Ruiz-Medina MD (2007a) Entropy-based thresholding for filtering and prediction. *Proceedings of the International Environmetrics Society-TIES2007*, Mikulov, Czech Republic, August 16-20, Horová I, Hrebícek J (ed), ISBN: 978-80-210-4333-6, pp. 7

[17] Angulo JM, Madrid AE, Ruiz-Medina MD (2010) Entropy-based correlated shrinkage of spatial random processes. *Stochastic Environmental Research and Risk Assessment*, accepted.

[18] Angulo JM, Mardia KV (1998) Measuring the effect of image deformation. *Research Report No.* STAT 98/07, Department of Statistics, University of Leeds, UK.

[19] Angulo JM, Ruiz-Medina MD, Alonso FJ, Bueso MC (2005) Generalized approaches to spatial sampling design. *Environmetrics* **16**: 523-534

[20] Angulo JM, Riz-Medina MD, Madrid AE (2007b) Spatial analysis from macroscale to microscale based on mutual information. *Proceedings of the XXX Congreso Nacional de Estadística e Investigación Operativa y IV Jornadas de Estadística Pública* (SEIO-2007), Valladolid, Spain, September 25 -28, Barrio E, Josa R, César P (eds), ISBN: 978-84-690-7249-3, pp. 114

[21] Anh VV, Angulo JM, Ruiz-Medina MD, Tieng Q (1998) Long-range dependence and second-order intermittency of two dimensional turbulence. *Environmental Modelling & Software* **13**: 233-238

[22] Autin F (2008) Maxisets for $\mu-$thresholding rules. *Test* **17**: 332-349

[23] Baddeley AJ, Moller J, Waagepetersen R (2000) Non- and semi-parametric estimation of interaction in inhomogeneous point patterns. *Statistica Neerlandica* **54**: 329-350

[24] Balkema AA, Haan L (1974) Residual life time at great age. *Annals of Probability* **2**: 792-804

[25] Beirlant J, Goegebeur Y, Teugles J, Segers J, Wall D, Ferro C (2004) *Statistics of Extremes. Theory and Applications.* Wiley, New York

[26] Bochner S (1955) *Harmonic Analysis and the Theory of Probability.* University of California Press, Berkeley and Los Angeles

[27] Bookstein FL (1989) Principal warps: thin-plate splines and the descomposition of deformations. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **11**: 567-585

[28] Von Bortkiewicz L (1922) Variationsbreite und mittlerer Fehler. *Sitzungsberichte der Berliner Mathematischen Gesellschaft* **21**: 3-11

[29] Brown PE, Karesen KF, Roberts GO, Tonellato S (2000) Blur-generated non-separable space-time models. *Journal of the Royal Statistical Society, Series B* **62**: 847-860

[30] Bueso MC, Angulo JM, Alonso FJ (1998) A state-space-model approach to optimal spatial sampling design based on entropy. *Environmental and Ecological Statistics* **5**: 29-44

[31] Chilès JP, Delfiner P (1999) *Geostatistics, Modelling Spatial Uncertainty.* Wiley-Interscience, New York

[32] Chipman HA, Kolaczyk ED, Mcculloch RE (1997) Adaptive Bayesian wavelet shrinkage. *Journal of the American Statististical Association* **92**: 1413-1421

[33] Christakos G (1992) *Random Field Models in Earth Sciences.* Academic Press, San Diego

[34] Christakos G (2000) *Modern Spatiotemporal Geostatistics.* Oxford University Press, New York

[35] Christensen GE, Rabbit RD, Miller MI (1996) Deformable templates using large deformation kinematics. *IEEE Transations on Image Processing* **5**: 1435-1447

[36] Clerc M, Mallat S (2003) Estimating deformations of stationary processes. *Annals of Statistics* **31**: 1772-1821

[37] Coles S (2001) *An Introduction to Statistical Modeling of Extreme Values.* Springer, London

[38] Cressie N (1993) *Statistics for Spatial Data.* Wiley, New York

[39] Cressie N, Huang H-C (1999) Classes of nonseparable, spatio-temporal stationarity covariance functions. *Journal of the American Statistical Association* **94**: 1330-1340

[40] Crouse M, Nowak R, Baraniuk R (1998) Wavelet-based statistical signal processing using hidden Markov models. *IEEE Transactions on Signal Processing* **46**: 886-902

[41] Diggle PJ (1983) *Statistical Analysis of Spatial Point Patterns.* Academic Press, London

[42] Donoho DL, Johnstone IM (1994a) Ideal spatial adaptation via wavelet shrinkage. *Biometrika* **81**: 425-455

[43] Donoho DL, Johnstone IM (1994b) Minimax risk over l_p-balls for l_q-error. *Probability Theory and Related Fields* **99**: 277-303

[44] Donoho DL, Johnstone IM (1995) Adapting to unknown smoothness via wavelet shrinkage. *Journal of the American Statistical Association* **90**: 1200-1224

[45] Donoho DL, Johnstone IM (1998) Minimax estimation via wavelet shrinkage. *Annals of Statistics* **26**: 879-921

[46] Farge M, Holschneider M, Colonna JF (1990) Wavelet analysis of coherent structures in two dimensional turbulent flow. *Topological Fluid Mechanics* (ed. H. K. Moffatt and A. Tsinober). Cambridge University Press, 765-766

[47] Fisher RA, Tippett LCH (1928) Limiting forms of frequency distributions of the largest or smallest member of a sample. *Proceedings of the Cambridge Philosophical Society* **24**: 180-190

[48] Fréchet M (1927) Sur la loi de probabilité de l'écart maximum. *Annales de la Société Polonaise de Mathématique, Cracovie* **6**: 93-116

[49] Gabor D (1946) Theory of communication. *Journal of the Institute of Electrical Engineers* **93**: 429-457

[50] Glasbey CA, Mardia KV (1998) A review of image warping methods. *Journal of Applied Statistics* **25**: 155-171

[51] Glasbey CA, Mardia KV (2001) A penalized likelihood approach to image warping. *Journal of the Royal Statistical Society, Series B* **63**: 465-514

[52] Gnedenko B (1943) Sur la distribution limite du terme maximum d'une série aléatorie. *Annals of Mathematics* **44**: 423-453. Translated and reprinted in: *Breakthroughs in Statistics*, Vol **1**, 1992, eds. S. Kotz and N.L. Johnson, Springer-Verlag, 195-225

[53] Gneiting T (2002) Nonseparable, stationary covariance functions for space-time data. *Journal of the American Statistical Association* **97**: 590-600

[54] Gneiting T, Genton MG, Guttorp P (2007) Geostatistical space-time models, stationarity, separability and full symmetry. *Statistical Methods for Spatio-Temporal Systems*, (ed. B. Finkenstadt, L. Held, and V. Isham). Boca Raton: Chapman and Hall/CRC, 151-175

[55] Gneiting T, Schlather M (2004) Stochastic models that separate fractal dimension and the Hurst effect. *SIAM Review* **46**: 269-282

[56] Goitía A, Ruiz-Medina MD, Angulo JM (2004) Joint estimation of spatial deformation and blurring in environmental data. *Stochastic Environmental Research and Risk Assessment* **19**: 1-7

[57] Gomes MI (1981) An i-dimensional limiting distribution function of largest values and its relevance to the statistical theory of extremes. *Statistical Distributions in Scientific Work* **6**: 389-410

[58] Grossmann A, Morlet J (1984) Descomposition of Hardy functions into square integrable wavelets of constant shape. *SIAM Journal on Mathematical Analysis* **15**: 723-736

[59] Gilliam X, Dunyak J, Doggett A, Smith D (2000) Coherent structure detection using wavelet analysis in long time-series. *Journal of Wind Engineering and Industrial Aerodynamics* **88**: 183-195

[60] Gumbel EJ (1958) *Statistics of Extremes*. Columbia University Press, New York

[61] Guttorp P, Le ND, Sampson PD, Zidek JV (1993) Using entropy in the redesign of an environmental monitoring network. In *Multivariate Environmental Statistics*, Patil GP, Rao CR (eds). North Holland/Elsevier Science, New York, 175-202

[62] Huang HC, Hsu NJ (2004) Modelling transport effects on ground-level ozone using a non-stationary space-time model. *Environmetrics* **15**: 251-268

[63] De Iaco S, Myers D, Posa D (2001) Space-time analysis using a general product-sum model. *Statistics & Probability Letters* **52**: 21-28

[64] De Iaco S, Myers D, Posa D (2002) Nonseparable space-time covariance models: some parametric families. *Mathematical Geology* **34**: 23-42

[65] Illian J, Penttinen A, Stoyan H, Stoyan D (2008) *Statistical Analysis and Modelling of Spatial Point Patterns* Wiley, Chichester, UK

[66] Ivanov AV, Leonenko NN (2004) Asymptotic theory of nonlinear regression with long-range dependence. *Mathematical Methods in Statistics* **13**: 153-178

[67] Jenkinson AF (1955) The frequency distribution of annual maximum values of meteorogical elements. *Quarterly Journal of the Royal Meteorological Society* **81**: 158-171

[68] Johnstone IM, Bernard W, Silverman B (2005) Empirical Bayes selection of wavelet thresholds. *The Annals of Statistics* **33**: 1700-1752

[69] Joseph DD (1990) *Fluids Dynamics of Viscoelastic Liquids*. Springer-Verlag, New York

[70] Kerkyacharian PC, Picard D (2000) Thresholding algorithms, maxisets and well concentrated bases. *Test* **9**: 283-345

[71] Kerkyacharian PC, Picard D (2002) Minimax or Maxisets. *Bernoulli* **8**: 219-253

[72] Kolovos A, Christakos G, Hristopulos DT, Serre ML (2004) Methods for generating non-separable spatiotemporal covariance models with potential environmental applications. *Advances in Water Resources* **27**: 815-830

[73] Kullback S, Leibler RA (1951) On information and sufficiency. *Annals of Mathematical Statistics* **22**: 79-86

[74] Kyriakidis PC, Journel AG (1999) Geostatistical space-time models: A review. *Mathematical Geology* **31**: 651-684

[75] Le ND, Zidek JV (2006) *Statistical Analysis of Environmental Space-Time Processes.* Springer, New York

[76] Leadbetter MR, Lindgren G, Retztzen A (1983) *Extremes and Related Properties of Random Sequences and Series.* Springer-Verlag, New York

[77] Lim SC, Teo LP (2008) Gaussian fields and Gaussian sheets with generalized Cauchy covariance structure. *Stochastic Processes and their Applications* **119**: 1325-1356

[78] Liu J, Moulin P (2001) Complexity-regularized image denoising. *IEEE Transactions on Image Processing* **10**: 841-851

[79] Ma C (2003) Families of spatio-temporal stationary covariance models. *Journal of Statistical Planning and Inference* **116**: 489-501

[80] Ma C (2008) Recent developments on the construction of spatio-temporal covariance models. *Stochastic Environmental Research and Risk Assesment* **22**: 39-47

[81] Madrid AE, Angulo JM (2010a) Multiscale entropy-based analysis of the effect of deformation on intermittency. *Proceedings of the Fifth International Workshop on Applied Probability*(IWAP-2010), Madrid, Spain, July 5-8, Arribas A, Glaz J, Jiménez R, Romo, J (ed), pp. 139

[82] Madrid AE, Angulo JM, Mateu J (2010b) Spatial threshold exceedances analysis through marked point processes. *Proceedings of the 5th International Workshop on Spatio-Temporal Modelling (METMA5).* González-Manteiga et al. (eds). Publisher Unidixital. ISBN: 978-84-691-81,

[83] Madrid AE, Angulo JM, Mateu J (2010c) Características anisotrópicas de eventos extremos. Un análisis mediante procesos puntuales marcados. *Proceedings of the XXXII Congreso Nacional de Estadística e Investigación Operativa y VI Jornadas de Estadística Pública* (SEIO-2010), A Coruña, Spain, September 14-17, Costa J, Fernández R, Presedo MA, Vilar JM (eds), ISBN: 978-84-693-6152-8, pp. 100

[84] Mallat S (1989) Multiresolution approximations and wavelet orthonormal bases of $\mathbb{L}_2(\mathbb{R})$. *Transactions of the American Mathematical Society* **315**: 69-87

[85] Mallat S (1998) *A Wavelet Tour of Signal Processing*. Academic Press, San Diego

[86] Mardia KV, Angulo JM, Goitía A (2006) Synthesis of image deformation strategies. *Image and Vision Computing* **23**: 35-47

[87] Meyer Y (1992) *Wavelets and Operators*. Advanced Mathemathics, Cambridge University Press, Cambridge

[88] Mateu J (2000) Second-order characteristics of spatial marked process with applications. *Nonlinear Analysis: Real World Applications* **1**: 145-162

[89] Meneveau C (1991) Analysis of turbulence in the orthonormal wavelet representation. *Journal of Fluids Mechanics* **323**: 469-520

[90] Von Mises R (1954) La distribution de la plus grande de n valeurs. *Revision Mathématique Union Interbalcanique.* **1:** 141-160. Reproduced in *Selected Papers of Richard von Mises, II* , 271-294, American Mathematical Society, Providence

[91] Mitchell M, Genton MG, Gumpertz M (2005) Testing for separability of space-time covariances. *Environmetrics* **16**: 819-831

[92] Morlet J, Arens G, Fourgeau E, Giard D (1982) Wave propagation and sampling theory. *Geophysics* **47**: 203-236

[93] Nicolleau F, Vassilicos JC (1999) Wavelets for the study of intermittency and its topology. *Philosophical Transaction of the Royal Society London A* **357**: 2439-2457

[94] Perrin O, Senoussi R (1999) Reducing non stationary stochastic processes to stationarity by a time deformation. *Journal of Mathematical Imaging and Vision* **26**: 59-71

[95] Pickands J (1975) Statistical inference using extreme value order statistics. *Annals of Statistics* **3**: 119-131.

[96] Porcu E, Mateu J, Saura F (2008) New classes of covariance and spectral density functions for spatio-temporal modelling. *Stochastic Environmental Research and Risk Assessment* **22**: 65-79

[97] Renyi A (1960) On measures of entropy and information. *Proceedings of the 4th Berkeley Synposium on Mathematics, Statistics and Probability, 1960*, 547-561

[98] Ripley BD (1977) The second-order analysis of stationary point processes. *Journal of Applied Probability* **13**: 255-266

[99] Romberg J, Choi H, Baraniuk R (2001) Bayesian tree-structured imaged modeling using wavelet-domain hidden Markov models. *IEEE Transactions on Image Processing* **10**: 1056-1068

[100] Rossman W (2002) *Lie Groups. An Introduction Through Linear Groups*. Oxford University Press, New York

[101] Ruiz-Medina MD, Alonso FJ, Angulo JM, Bueso MC (2003) Functional stochastic modeling and prediction of spatio-temporal processes. *Journal of Geophysical Research – Atmospheres* **108**: (D24) 9003, doi:10.1029/2003JD003416

[102] Ruiz-Medina MD, Angulo JM (2002) Spatio-temporal filtering using wavelets. *Stochastic Environmental Research and Risk Assessement* **16**: 241-266

[103] Ruiz-Medina MD, Angulo JM (2007) Functional estimation of spatiotemporal heterogeneities. *Environmetrics* **18**: 775-792

[104] Ruiz-Medina MD, Angulo JM, Anh VV (2008) Multifractality in space-time statistical models. *Stochastic Environmental Research and Risk Assessement* **22**(S1): 81-86

[105] Sampson PD, Guttorp P (1992) Nonparametric estimation of nonstationary spatial covariance structure. *Journal of the American Statistical Association* **87**: 108-119

[106] Shannon CE (1948) The mathematical theory of communications. *Bell System Technical Journal* **27**: 379-423, 623-656

[107] Smith RL (1986) Extreme value theory based on the r largest anual events. *Journal of Hydrology* **86**: 27-43

[108] Stein ML (1999) *Interpolation of Spatial Data. Some Theory for Kriging*. Springer, New York

[109] Stein M (2005) Space-time covariance functions. *Journal of the American Statistical Association* **100**: 310-321

[110] Stoyan D, Kendall WS, Mecke J (1987) *Stochastic Geometry and its Applications*. Akademie-Verlag: Berling / Wiley: Chichester, UK

[111] Stoyan D, Stoyan H (1994) *Fractals, Random Shapes and Points Fields. Methods of Geometrical Statistics*. Wiley, Chichester, UK

[112] Tawn JA (1988) An extreme value theory model for dependent observations. *Journal of Hydrology* **101**: 227-250

[113] Taylor JE, Takemura A, Adler RJ (2005) Validity of the expected Euler characteristic heuristic. *Annals of Probability* **33**: 1362-1396

[114] Todorovic P, Zelenhasic E (1970) A stochastic model for flood analysis. *Water Resources Research* **6**: 1641-1648

[115] Todorovic P, Rousselle J (1971) Some problems of flood analysis. *Water Resources Research* **5**: 1144-1150

[116] Tsallis C (1988) Possible generalization of Boltzmann-Gibbs statistics. *Journal of Statistical Physics* **52**: 479-487

[117] Vanmarcke E (2010) *Randoms Fields: Analysis and Synthesis* (Revised and expanded new edition). World Scientific Publishing, Singapore

[118] Vidakovic B (1998a) *Statistical Modeling by Wavelets*. Wiley Series in Probability and Statistics. Wiley, New York

[119] Vidakovic B (1998b) Wavelet-based nonparametric Bayes methods. Practical nonparametric and semiparametric Bayesian statistics. *Lecture Notes in Statistics* **133**: 133-155. Springer, New York

[120] Wackernagel H (2003) *Multivariate Geostatistics. An Introduction with Applications* (3rd ed). Springer, Berlin

[121] Weissman I (1978) Estimation of parameters and large quantiles based on the k largest observations. *Advances in Applied Probability* **10**: 331-332

[122] Whittaker J (1990) *Graphical Models in Applied Multivariate Statistics*. Wiley, New York

[123] Wikle CK, Cressie N (1999) A dimension-reduction approach to space-time Kalman filtering. *Biometrika* **86**: 815-829

[124] Yaglom AM (1987) *Correlation Theory of Stationary and Related Random Functions I. Basic Results*. Springer, New York

Appendix

This appendix collects a selection of the most relevant **MATLAB** codes developed for the elaboration of the thesis.

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Program  to apply a deformation with/without Jacobian   %
% to several realizations of a spatial RF using landmarks %
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

clear all
clc

%%%% Definition of arrays and parameters %%%%

cd C:\variables
load Z % realizations
dim=size(Z);
dim_x=dim(1);
dim_y=dim(2);
n=dim(3); % number of realizations
% interest domain D=[DXA,DXB]x[DYA,DYB]
DXA=0;
DXB=1;
DYA=0;
DYB=1;
```

```
cd C:\MATLAB6p5\work

% landmarks
marcas=[0,0;1,1;0,1;1,0; 0,0.5; 0.5,1; 1,0.5; 0.5,0;3/8,3/8;1/4,1/2;...
       3/8,5/8;5/8,3/8;3/4,1/2;5/8,5/8;1/2,1/4;1/2,3/4];
save marcas marcas

% transformed landmarks
marcas_trans=[0,0;1,1;0,1;1,0; 0,0.5; 0.5,1; 1,0.5; 0.5,0; 7/16, 7/16;...
      3/8,1/2; 7/16,9/16; 9/16,7/16; 5/8,1/2; 9/16,9/16; 1/2,3/8; 1/2,5/8];
save marcas_trans marcas_trans

minimo=0;

% parameters for Newton-Raphson's algorithm
delta=10^(-6);
eps=10^(-6);
ite_max=500;

% matrix of coefficients for the system (inv(L)*Y)
coef=coefgamma(marcas,marcas_trans);
save coef coef

% deformation of first realization
for i=1:dim_x
    for j=1:dim_y
        u = [(DXB-DXA)*(i-0.5)/dim_x,(DYB-DYA)*(j-0.5)/dim_y];
        save u u
        [u,it,er_abs]=raphson('tgamma','j_gamma',u,delta,eps,ite_max);
        V1(i,j)=fix((dim_x*u(1)/(DXB-DXA))+1);
        V2(i,j)=fix((dim_y*u(2)/(DYB-DYA))+1);
        jacobian(i,j)=det(j_gamma([u(1),u(2)]));
```

```
        if V1(i,j)<1 | V2(i,j)<1 | V1(i,j)>dim_x| V2(i,j)>dim_y
           Y_def(i,j,1)=minimo;
           Y_def_jac(i,j,1)=minimo;
        else
           Y_def(i,j,1)=Z(V1(i,j),V2(i,j),1);
           Y_def_jac(i,j,1)=Z(V1(i,j),V2(i,j),1)*(jacobian(i,j))^(-1);
        end
    end
end

% deformations of remaining realizations
for r=2:n
   for i=1:dim_x
      for j=1:dim_y
         if V1(i,j)<1 | V2(i,j)<1 | V1(i,j)>dim_y | V2(i,j)>dim_y
           Y_def(i,j,r)=minimo;
           Y_def_jac(i,j,r)=minimo;
         else
           Y_def(i,j,r)=Z(V1(i,j),V2(i,j),r);
           Y_def_jac(i,j,r)=Z(V1(i,j),V2(i,j),r)*(jacobian(i,j))^(-1);
         end
      end
   end
end

cd C:\variables
save Y_def Y_def
save Y_def_jac Y_def_jac
```

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Program  to apply blurring to several realizations of spatial RF %
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

clear all
clc

%%%% Definition of arrays and parameters %%%%
cd C:\metma4\conexos\variables
load Y % realizations

%%% Dimension of spatial domain
dim=size(Y);
dim_x=dim(1);
dim_y=dim(2);
n=dim(3);

cd C:\MATLAB6p5\work

% Blurring kernel
M=[401,401];
S1=eye(2)*10;
u=10^(-6);
[H1,cola1]=H_restringida(M,S1,u);

for i=1:n
  % unfolding edges
  Ydesd(:,:)=[flipud(fliplr(Y(1:100,1:100,i))),flipud(Y(1:100,1:400,i)),...
   flipud(fliplr(Y(1:100,301:400,i)));fliplr(Y(1:400,1:100,i)),Y(:,:,i),...
   fliplr(Y(1:400,301:400,i));flipud(fliplr(Y(301:400,1:100,i))),...
   flipud(Y(301:400,1:400,i)),flipud(fliplr(Y(301:400,301:400,i)))];
  %convolution-blurring
  Y_aux(:,:)=conv2(Ydesd(:,:),H1,'same');
```

```
  %reduction
  Y_conv(:,:,i)=Y_aux(101:500,101:500);
end

cd C:\variables
save Y_conv Y_conv
```

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Function to define a restricted matrix %
% from a bivariate Gaussian density      %
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

function [HR]=H_restringida(M1,S1,u)

% M1: mean
% S1: covariance
% u: threshold

% HR: Restricted Gaussian matrix

for k=1:M1(1)*2-1
    for l=1:M1(2)*2-1
        H(k,l)=mvnpdf([k,l],M1,S1);
    end
end

H=ismenor(H,u,0);
aux=H(M1(1),1:M1(1));
aux2=aux;
borrar=find(aux==0);
aux2(borrar)=[];
minimo=min(aux2);
indice1=find(aux==minimo);

aux3=H(1:M1(2),M1(2));

aux4=aux3;
borrar=find(aux3==0);
aux4(borrar)=[];
minimo=min(aux4);
```

```
indice2=find(aux3==minimo);

HR=H(indice2:M1(1)*2-1-indice2+1,indice1:M1(2)*2-1-indice1+1);
suma=sumamatriz(HR);
HR=HR/suma;
```

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Function to determine the connected components %
% above a threshold in a matrix                  %
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

function [num, ind, etiq]=conexas(A,u)

%'connected-component': pixels with common sides and/or corners

% A: matrix
% u: threshold

% num: number of connected components
% ind: index of element above the threshold
% etiq: label of corresponding component

dim=size(A);
ind=find(A>u);

if length(ind)<1
    num=0;
else
    et=1;
    etiq(1)=et;

    for i=2:length(ind)

    if (length(find(ind==ind(i)-(dim(1)+1)))==1 & ...
       mod(ind(i)-1,dim(1))~=0&...
       length(find(ind==ind(i)-(dim(1)-1)))==0)|...
       (length(find(ind==ind(i)-(dim(1)+1)))==1 & ...
       mod(ind(i)-1,dim(1))~=0&...
       mod(ind(i)+1,dim(1))==1)
```

```
    etiq(i)=etiq(find(ind==ind(i)-(dim(1)+1)));

else if length(find(ind==ind(i)-(dim(1)+1)))==1 & ...
        mod(ind(i)-1,dim(1))~=0&...
        length(find(ind==ind(i)-(dim(1)-1)))==1&...
        mod(ind(i)-dim(1),dim(1))~=0
        w=find(ind==ind(i)-(dim(1)+1));
        z=find(ind==ind(i)-(dim(1)-1));

        if etiq(w)<etiq(z)
           etiq(i)=etiq(w);
           etiq(find(etiq==etiq(z)))=etiq(w);
        else
           etiq(i)=etiq(z);
           etiq(find(etiq==etiq(w)))=etiq(z);
        end
        clear z w
        else if length(find(ind==ind(i)-dim(1)))==1
            etiq(i)=etiq(find(ind==ind(i)-dim(1)));
        else if (length(find(ind==ind(i)-(dim(1)-1)))==1 & ...
                mod(ind(i),dim(1))~=0 & length(find(ind==ind(i)-1))==0)|...
                (length(find(ind==ind(i)-(dim(1)-1)))==1 & ...
                mod(ind(i),dim(1))==1)
                etiq(i)=etiq(find(ind==ind(i)-(dim(1)-1)));
            else if length(find(ind==ind(i)-(dim(1)-1)))==1 & ...
                mod(ind(i),dim(1))~=0 &...
                length(find(ind==ind(i)-1))==1&...
                mod((ind(i)-1),dim(1))~=0
                w=find(ind==ind(i)-(dim(1)-1));
                z=find(ind==ind(i)-1);
                if etiq(w)<etiq(z);
                   etiq(i)=etiq(w);
                   etiq(find(etiq==etiq(z)))=etiq(w);
```

```
                else
                    etiq(i)=etiq(z);
                    etiq(find(etiq==etiq(w)))=etiq(z);
                end
                clear z w
                else if length(find(ind==ind(i)-1))==1 &...
                        mod((ind(i)-1),dim(1))~=0
                        etiq(i)=etiq(find(ind==ind(i)-1));
                    else
                        et=et+1;
                        etiq(i)=et;
                    end
                end
            end
        end
end
end
end

restar=0;
for i=1:et
    if length(find(etiq==i))==0
        restar=restar+1;
        borrar(restar)=i;
    end
end
num=et-restar;
if exist('borrar')~=0
    n_fic=length(borrar);
    if n_fic>0
        borradas=0;
        for i=n_fic:-1:1
            empezar=borrar(i)+1;
```

```
                ultima=et-borradas;
                for k=etiq(find(etiq==empezar)):etiq(find(etiq==ultima))
                    etiq(find(etiq==k))=k-1;
                end
                borradas=borradas+1;
            end
        end
    end
 etiq=etiq';
end
end
```

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Function to determine the hole components above a threshold in a matrix %
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

function [num]=anulares(A,u)
% A: matrix
% u: threshold

% num: number of 'hole' components

dim=size(A);
frontera=sort([2:dim(1)-1,...
               dim(1)*dim(2)-dim(1)+1:dim(1)*dim(2),...
               1:dim(1):dim(1)*dim(2)-dim(1),...
               dim(1):dim(1):dim(1)*dim(2)-1]);

x=find(A<u);
if length(x)<1
    num=0;
else
    et=1;
    y(1)=et;

for i=2:length(x)
    if length(find(x==x(i)-dim(1)))==1
       y(i)=y(find(x==x(i)-dim(1)));
    else if length(find(x==x(i)-1))==1 & mod((x(i)-1),dim(1))~=0
            y(i)=y(find(x==x(i)-1));
         else
            et=et+1;
            y(i)=et;
         end
     end
```

```
end

for i=length(x):-1:1
    if length(find(x==x(i)-1))==1 & mod((x(i)-1),dim(1))~=0 & i-1>0
       y(i-1)=y(i);
    end
end

for j=1:length(x)
    if length(find(frontera==x(j)))==1
        z=find(y==y(j));
        y(z)=0;
    end
end

restar=0;
for i=1:et
    if length(find(y==i))==0
        restar=restar+1;
    end
end
    num=et-restar;
end
```

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Program to compute the Euler characteristic and hypervolume  %
% of several realizations of a spatial random field            %
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

clear all
clc

cd C:\variables
load Y

cd C:\funciones
dim=size(Y);
sum=0;
for t=1:dim(3)
    for j=1:dim(2)
        for i=1:dim(1)
            sum=sum+Y(i,j,t)^2;
        end
    end
end
varianza=sum/(dim(1)*dim(2)*dim(3));
sigma=varianza^(1/2);
umbral=1.5*sigma;
tam=dim(1)*dim(2);

for r=1:dim(3)
   for i=1:length(umbral)
      x=find(Y(:,:,r)>=umbral(i));
      area(r,i)=length(x)/tam;
      sum=0;
      Aux=Y(:,:,r);
      for j=1:length(x)
```

```
            sum=sum+Aux(x(j))-umbral(i);
        end
        volumen(r,i)=sum/tam;

        [num_com]=conexas(Aux,umbral(i));
        cconexas(r,i)=num_com;
        [num_anu]=anulares(Aux,umbral(i));
        aanulares(r,i)=num_anu; % number of holes
        euler(r,i)=cconexas(r,i)-aanulares(r,i);
    end
end

cd C:\variables
save area area
save volumen volumen
save cconexas cconexas
save aanulares aanulares
save euler euler
save umbral umbral
save varianza varianza
```

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Program  to apply a deformation with/without Jacobian  and blurring %
% to several realizations of a spatial RF using landmarks             %
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

cd C:\variables
load Z1
cd C:\MATLAB6p5\work
dim=size(Z1);
dim_x=dim(1);
dim_y=dim(2);
M=[401,401];
S=eye(2)*10;
u=10^(-6);
[HR,cola]=H_restringida(M,S,u);
a=0.9;
b=1;
DXA=0;
DXB=1;
DYA=0;
DYB=1;
delta=10^(-6);
eps=10^(-6);
ite_max=500;

Y1(:,:,1)=Z1(:,:,1);
Y1_jac(:,:,1)=Z1(:,:,1);
for i=1:dim_x
        for j=1:dim_y
            u = [(DXB-DXA)*(i-0.5)/dim_x,(DYB-DYA)*(j-0.5)/dim_y];
            u=u-[1/2,1/2];
            [v,w]=cart2pol(u(1),u(2));
            coor_pol=[v,w];
```

```
            save coor_pol coor_pol
            jacobiano(i,j)=1;
            if coor_pol(2)<1/2
               [coor_pol,it,er_abs]=raphson('tcircular','j_circular',...
                                    coor_pol,delta,eps,ite_max);
                jacobiano(i,j)=det(j_circular(coor_pol));
            end
            [r,s]=pol2cart(coor_pol(1),coor_pol(2));
            coor_cart=[r,s]+[1/2,1/2];
            V1(i,j)=fix( dim_x*coor_cart(1)/(DXB-DXA) +1);
            V2(i,j)=fix( dim_y*coor_cart(2)/(DYB-DYA) +1);
            if V1(i,j)<1 | V2(i,j)<1 | V1(i,j)>dim_x | V2(i,j)>dim_y
               Ydef(i,j)=minimo;
               Ydef_jac(i,j)=minimo;
            else
               Ydef(i,j)=Y1(V1(i,j),V2(i,j),1);
             Ydef_jac(i,j)=Y1_jac(V1(i,j),V2(i,j),1)*(jacobiano(i,j))^(-1);
            end
        end
end

clear u v w coor_pol r s coor_cart

Yd(:,:)=[flipud(fliplr(Ydef(1:100,1:100))),flipud(Ydef(1:100,1:400)),...
           flipud(fliplr(Ydef(1:100,301:400)));fliplr(Ydef(1:400,1:100))...
           , Ydef(:,:),fliplr(Ydef(1:400,301:400));...
           flipud(fliplr(Ydef(301:400,1:100))),flipud(Ydef...
           301:400,1:400)),flipud(fliplr(Ydef(301:400,301:400)))];
Yc(:,:)=conv2(Yd(:,:),HR,'same');
Y1(:,:,2)=a*Yc(101:500,101:500) + b*Z1(:,:,2);
clear Ydef Yd Yc

Yd_jac(:,:)=[flipud(fliplr(Ydef_jac(1:100,1:100))),flipud(Ydef_jac...
```

```
           (1:100,1:400)),flipud(fliplr(Ydef_jac(1:100,301:400)));...
           fliplr(Ydef_jac(1:400,1:100)), Ydef_jac(:,:),fliplr(Ydef_jac...
           (1:400,301:400));flipud(fliplr(Ydef_jac(301:400,1:100))),...
           flipud(Ydef_jac(301:400,1:400)),flipud(fliplr(...
           Ydef_jac(301:400,301:400)))];
Yc_jac(:,:)=conv2(Yd_jac(:,:),HR,'same');
Y1_jac(:,:,2)=a*Yc_jac(101:500,101:500) + b*Z1(:,:,2);
clear Ydef_jac Yd_jac Yc_jac

for t=3:100
    for i=1:dim_x
        for j=1:dim_y
             if V1(i,j)<1 | V2(i,j)<1 | V1(i,j)>dim_x | V2(i,j)>dim_y
                    Ydef(i,j)=minimo;
                    Ydef_jac(i,j)=minimo;
             else
                    Ydef(i,j)=Y1(V1(i,j),V2(i,j),t-1);
                    Ydef_jac(i,j)=Y1_jac(V1(i,j),V2(i,j),t-1)*...
                                    (jacobiano(i,j))^(-1);
             end
        end
    end

    Yd(:,:)=[flipud(fliplr(Ydef(1:100,1:100))),flipud(Ydef(1:100,1:400))...
           ,flipud(fliplr(Ydef(1:100,301:400)));fliplr(Ydef(1:400,1:100))...
           , Ydef(:,:),fliplr(Ydef(1:400,301:400));flipud(fliplr(...
           Ydef(301:400,1:100))),flipud(Ydef(301:400,1:400)),flipud...
           (fliplr(Ydef(301:400,301:400)))];
    Yc(:,:)=conv2(Yd(:,:),HR,'same');
    Y1(:,:,t)=a*Yc(101:500,101:500) + b*Z1(:,:,t);
    clear Yd Ydef Yc

    Yd_jac(:,:)=[flipud(fliplr(Ydef_jac(1:100,1:100))),flipud(Ydef_jac...
```

```
            (1:100,1:400)),flipud(fliplr(Ydef_jac(1:100,301:400)));...
            fliplr(Ydef_jac(1:400,1:100)), Ydef_jac(:,:),fliplr(Ydef_jac...
            (1:400,301:400));flipud(fliplr(Ydef_jac(301:400,1:100))),...
            flipud(Ydef_jac(301:400,1:400)),flipud(fliplr...
            (Ydef_jac(301:400,301:400)))];
    Yc_jac(:,:)=conv2(Yd_jac(:,:),HR,'same');
    Y1_jac(:,:,t)=a*Yc_jac(101:500,101:500) + b*Z1(:,:,t);
    clear Yd_jac Ydef_jac Yc_jac
end

clear Z1
cd C:\variables
save Y1 Y1
save Y1_jac Y1_jac
save jacobiano jacobiano
```

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Function to generate a Linnik model  covariance %
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

function [CRI,CovXTW2,CIOG,infG]=covarianzas(niveles,sep,TI2,alpha,lambda)

% niveles= number of scales
% sep= noise intensity
% alpha= Linnik model parameter
% TI2 = number of locations in interest region

% CRI= covariance values on interest region
% CovXTW2= 2DTW2 of first component of the covariance on candidates region
% CIOG= Cross-covariance between interest and candidate regions
% infG= Information of candidate region on interest variable

N=2^niveles;
N2=N^2;

for I=1:N
    P(I)=I/N;
end

for K=1:TI2
    COORG(K,:)=lambda(K,:);
end

H=TI2+1;
M=TI2+1;
for L=1:N
    for K=1:N
        COORG(H,1)=L/N;
        COORG(M,2)=K/N;
```

```
        H=H+1;
        M=M+1;
    end
end

% Covariance values on candidate region
for K=1:TI2
    for I=1:TI2
        z1=(lambda(K,1)-lambda(I,1));
        z2=(lambda(K,2)-lambda(I,2));
        z=(z1^2+z2^2);
                CRI(K,I)= 1/(1+sqrt(z)^alpha);
    end
end

for K=1:N
    for L=1:N
        for I=1:N
            for J=1:N
                z1=(P(K)-P(I));
                z2=(P(L)-P(J));
                z=(z1^2+z2^2);
                CRC4(K,L,I,J)= 1/(1+sqrt(z)^alpha);
            end
        end
    end
end

for I=1:TI2
for K=1:N
    for L=1:N
        z1=(COORG(I,1)-P(K));
        z2=(COORG(I,2)-P(L));
```

```
        z=(z1^2+z2^2);
        CRC3(I,K,L)= 1/(1+sqrt(z)^alpha);
    end
end
end

for L=1:TI2
    for I=1:N
        for J=1:N
            CIC1D(I,J)=CRC3(L,I,J);
        end
    end
    [C,S]=wavedec2(CIC1D,niveles,'haar');
    for I=1:N2
        TWDCIC(L,I)=C(1,I);
    end
end

clear CIC1D CIC C S

for K=1:N
    for L=1:N
        for I=1:N
            for J=1:N
               CovX1(I,J)=CRC4(K,L,I,J);
            end
        end
        [C,S]=wavedec2(CovX1,niveles,'haar');
        for I=1:N2
           CovX1TW(K,L,I)=C(1,I);
        end
    end
end
```

```
clear CovX1 CI4 C S

for K=1:N2
    for I=1:N
        for J=1:N
            CovX1(I,J)=CovX1TW(I,J,K);
        end
    end
    [C,S]=wavedec2(CovX1,niveles,'haar');
    for I=1:N2
        CovXTW2(K,I)=C(1,I);
    end
end

clear CovX1 COVX1TW C S

COG=CovXTW2+sep*eye(N2);
CIOG=TWDCIC;
CCG=CRI-CIOG*inv(COG)*CIOG';

infG=1/2*log(det(CRI))-1/2*log(det(CCG));
```

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Function  to compute different information ratios %
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

function [RBG,RIE,RBE,infe]=razones(niveles,sep,TI2,CI,CovXTW2,CIOG,infG)

% niveles: number of scales
% sep: noise intensity
% TI2: number of locations of candidate region
% CI:covariance values on interest region
% CovXTW2: TW2 of first component of the covariance on candidates region
% CIOG:Cross-covariance between interest and candidate regions
% infG: Information of candidate region on interest variable

% RBG= Information ratios wavelet blocks-global
% RIE= Information ratios scales-global
% RBE= Information ratios wavelet blocks-scales

N=2^niveles;
N2=N^2;

for M=1:niveles
    if M==1
        TB(M)=4;
        NB(M)=1;
        TE(M)=NB(M)*TB(M);
    else
        TB(M)=3;
    end
    NB(M)=2^(2*M-2);
    TE(M)=NB(M)*TB(M);
end
```

```
for M=1:niveles
     if M==1
         J=1;
         H=1;
         for F=1:TI2
             for I=1:4
                 CIO(F,H,J,M)=CIOG(F,I);
                 H=H+1;
             end
             H=1;
         end
         for P=1:4
             for Q=1:4
                 CBTW2(P,Q,J,M)=CovXTW2(P,Q);
                 COB(P,Q,J,M)=CBTW2(P,Q,J,M);
                 if P==Q
                    COB(P,Q,J,M)=CBTW2(P,Q,J,M)+sep;
                end
             end
         end
         CCBI(:,:,J,M)=CI-CIO(:,:,J,M)*inv(COB(:,:,J,M))*CIO(:,:,J,M)';
         entcb(J,M)=TI2/2*(1+log(2*pi))+1/2*log(det(CCBI(:,:,J,M)));
         infcb(J,M)=1/2*log(det(CI))-1/2*log(det(CCBI(:,:,J,M)));
         RBG(J,M)=infcb(J,M)/infG;
         Sum=4;
         Sum1=4;
         Sum2=4;
     else
         H=1;
         for J=1:NB(M)
             for F=1:TI2
             for I=1:NB(M):TE(M)
                     CCIO(F,H,J,M)=CIOG(F,Sum+I+J-1);
```

```
                H=H+1;
            end
            H=1;
        end
        H=1;
    end
    Sum=Sum+TE(M);

    P=1;
    Q=1;
    for J=1:NB(M)
        for K=1:NB(M):TE(M)
            for I=1:NB(M):TE(M)
                CCOB(P,Q,J,M)=CovXTW2(Sum2+J-1+K,Sum1+J-1+I);
                if P==Q
                   CCOB(P,Q,J,M)=CovXTW2(Sum2+J-1+K,Sum1+J-1+I)+sep;
                end
                Q=Q+1;
            end
            Q=1;
            P=P+1;
        end
        P=1;
        Q=1;
    CCBI(:,:,J,M)=CI-CCIO(:,:,J,M)*inv(CCOB(:,:,J,M))*CCIO(:,:,J,M)';
        entcb(J,M)=TI2/2*(1+log(2*pi))+1/2*log(det(CCBI(:,:,J,M)));
        infcb(J,M)=1/2*log(det(CI))-1/2*log(det(CCBI(:,:,J,M)));
        RBG(J,M)=infcb(J,M)/infG;
    end
    Sum1=Sum1+TE(M);
    Sum2=Sum2+TE(M);
  end
end
```

```
for M=1:niveles
    if M==1
         Sum=0;
         H=1;
         for F=1:TI2
             for I=1:4
                 CIOE(F,H)=CIOG(F,I);
                 H=H+1;
             end
             H=1;
         end

         H=1;
         L=1;
         for K=1:4
             for I=1:4
                 COBE(L,H)= CovXTW2(K,I);
                 if I==K
                    COBE(L,H)=CovXTW2(K,I)+sep;
                 end
                 H=H+1;
             end
             H=1;
             L=L+1;
         end
         H=1;
         L=1;
         CCE=CI-CIOE*inv(COBE)*CIOE';
         entce(M)=TI2/2*(1+log(2*pi))+1/2*log(det(CCE));
         infe(M)=1/2*log(det(CI))-1/2*log(det(CCE));
         RIE(M)= infe(M)/infG;
         Sum=4;
```

```
        Sum1=4;
        Sum2=4;

    else

        for F=1:TI2
            for I=1:TE(M)
                CIOE(F,H)=CIOG(F,Sum+I);
                H=H+1;
            end
            H=1;
        end
        Sum=Sum+TE(M);

        H=1;
        L=1;
        for K=1:TE(M)
            for I=1:TE(M)
                COBE(L,H)=CovXTW2(Sum2+K,Sum1+I);
                if I==K
                   COBE(L,H)=CovXTW2(Sum2+K,Sum1+I)+sep;
                end
                H=H+1;
            end
            H=1;
            L=L+1;
        end
        H=1;
        L=1;
        Sum1=Sum1+TE(M);
        Sum2=Sum2+TE(M);
        CCE=CI-CIOE*inv(COBE)*CIOE';
        entce(M)=TI2/2*(1+log(2*pi))+1/2*log(det(CCE));
```

```
          infe(M)=1/2*log(det(CI))-1/2*log(det(CCE));
          RIE(M)= infe(M)/infG;
      end
end

for M=1: niveles
    for J=1:NB(M)
        RBE(J,M)=infcb(J,M)/infe(M);
    end
end
```

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Function to select blocks from the cumulative %
% distribution of information ratios            %
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

function [percentil, umbral, cont_emp, bloq_selec]=umbrales(RBG,probabilidad)

%RBG:Information ratios wavelet blocks-global
%probabilidad: probabilities for percentiles

%percentil: value of percentile
%umbral: threshold
%cont_emp-> cont_emp(c,d,e): number of wavelet blocks > threshold
%bloq_selec->bloq_selec(c,d,e):c-th block > d-th threshold for scale e

dim=size(RBG);
niveles=dim(2);
for n=1:niveles
    if n==1
        NB(n)=1;
    else
        NB(n)=2^(2*n-2);
    end
end
cont_emp=zeros(niveles,length(probabilidad));

for n=1:niveles


    r=RBG(:,n);
    r(NB(n)+1:1:NB(niveles))=[];
    [g,y]=ecdf(r);

    y(1)=0;
```

```
    for p=1:length(probabilidad)
        for i=1:size(g)
            if g(i)>=probabilidad(p)
                percentil(n,p)=g(i);
                umbral(n,p)=y(i);
                break
            end
        end
    end

    for j=1:NB(n)
        for k=1:length(umbral)
            if r(j)>umbral(n,k)
                cont_emp(n,k)=cont_emp(n,k)+1;
                bloq_selec(cont_emp(n,k),k,n)=j;
            else
                break
            end
        end
    end
end
```

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Function to compute the information provided by selected blocks %
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

function [infbse]=informacion_umbrales_escalas(sep,umbral,cont_emp,...
                  bloq_selec,CIOG,CI,CovXTW2)

% sep: noise intensity
% umbral: threshold
% cont_emp:cont_emp(c,d,e): number of wavelet blocks > threshold
% bloq_selec:->bloq_selec(c,d,e): block c for threshold d in scale e
% CIOG:Cross-covariance between interest and candidate regions
% CI:covariance values on interest region
% CovXTW2: 2D-TW of first component of the covariance on candidate region

% infbse:information of selected wavelet blocks of interest variable

dim=size(umbral);
niveles=dim(1);
n_umbrales=dim(2);

for M=1:niveles
    NB(M)=2^(2*M-2);
    if M==1
        TB(M)=4; %
        NB(M)=1;
        ucn(M)=NB(M)*TB(M);
    else
        TB(M)=3;
        ucn(M)=ucn(M-1)+NB(M)*TB(M);
    end
end
```

```
for n=3:niveles
    for p=1:n_umbrales
        selec=[];
        for j=1:cont_emp(n,p)
            k=bloq_selec(j,p,n);
            if n==1&k==1
                selec=[1,2,3,4];
            end
            if n>1
                selec=[selec, k+ucn(n-1), k+ucn(n-1)*2, k+ucn(n-1)*3];
            end
        end
        selec=sort(selec);
        noselec=1:ucn(niveles);
        noselec(selec)=[];

        CCIBS=CIOG;
        CCIBS(:,noselec)=[];
        CBSTW2=CovXTW2;
        CBSTW2(noselec,:)=[];
        CBSTW2(:,noselec)=[];

        COBS=CBSTW2+sep*eye(size(CBSTW2));
        CCBS=CI-CCIBS*inv(COBS)*CCIBS';
        infbse(n,p)=1/2*log(det(CI))-1/2*log(det(CCBS));

        clear CCIBS CBSTW2 COBS CCBS
    end
end
```

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Function to compute the information obtained after reconstruction %
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

function [infbse,CCIBS,COBS,CCBS]=informacion_reconstruccion(sep,n_th,...
                            umbral,percentil,u,bloq_selec,CIOG,CI,CovXTW2)

% sep:  noise intensity
% n_th: level to thresholding
% umbral: threshold
% percentil: value of percentile
% u: threshold
% bloq_selec:->bloq_selec(c,d,e): block c for threshold d in scale e
% CIOG:Cross-covariance between interest and candidate regions
% CI:covariance values on interest region
% CovXTW2: 2D-TW of first component of the covariance on candidate region

% infbse: information
% CCIBS: cross-covariance between interest and candidate regions
% COBS: observation matrix of selected blocks
% CCBS: conditional cross covariance interest to selected block

dim=size(umbral);
niveles=dim(1);
n_umbrales=dim(2);

for M=1:niveles
    NB(M)=2^(2*M-2);
    if M==1
        TB(M)=4;
        NB(M)=1;
        ucn(M)=NB(M)*TB(M);
    else
```

```
        TB(M)=3;
        ucn(M)=ucn(M-1)+NB(M)*TB(M);
    end
end

bloques=bloq_selec(:,u,n_th);
r=find(bloques==0);
bloques(r)=[];

selec=1:ucn(n_th-1);

for j=1:length(bloques)
    k=bloques(j);
    if n_th>1
       selec=[selec, k+ucn(n_th-1), k+ucn(n_th-1)*2, k+ucn(n_th-1)*3];
    end
    if n_th<niveles
       selec2=ucn(n_th)+1:ucn(niveles);
       selec=[selec,selec2];
    end
end

selec=sort(selec);
noselec=1:ucn(niveles);
noselec(selec)=[];

CCIBS=CIOG;
CCIBS(:,noselec)=[];

CBSTW2=CovXTW2;
CBSTW2(noselec,:)=[];
CBSTW2(:,noselec)=[];
```

```
COBS=CBSTW2+sep*eye(size(CBSTW2));
CCBS=CI-CCIBS*inv(COBS)*CCIBS';
infbse=1/2*log(det(CI))-1/2*log(det(CCBS));
```