Università degli Studi di Palermo

AREA QUALITÀ, PROGRAMMAZIONE E SUPPORTO STRATEGICO
SETTORE STRATEGIA PER LA RICERCA
U. O. DOTTORATI

Corso di Dottorato in Scienze Economiche e Statistiche
Dipartimento di Scienze Economiche, Aziendali e Statistiche
Settore Scientifico Disciplinare: SECS-S/01

LOCAL METHODS FOR COMPLEX SPATIO-TEMPORAL POINT PROCESSES

IL DOTTORE
NICOLETTA D'ANGELO

IL COORDINATORE
ANDREA CONSIGLIO

IL TUTOR
GIADA ADELFIO

CICLO XXXV
ANNO CONSEGUIMENTO TITOLO: 2022

“Se la scienza ci insegna qualcosa,
ci insegna ad accettare i nostri fallimenti,
come i nostri successi,
con calma, dignitá, e classe...”
- Dott. Frederick von Frankenstein (Gene Wilder)
from Mel Brooks’ “Young Frankenstein”

Acknowledgements

What probably makes me the happiest about wrapping up this thesis work is realizing the amazing group of people I surrounded myself with during these years. They all contributed to my writing of this research in different manners.

To start with, I express my gratitude to my (super!) supervisor professor Giada Adelfio for introducing me to the local point processes topic for my master thesis. The trust she placed in me since then motivated me to apply for PhD and to continue studying this topic at my best. With her passion, optimism, and altruism, she made the whole doctoral experience not just possible but also much more pleasurable than I could ever expect. Her wisdom and counsel surely contributed to shaping the way I learned to approach research. I am glad I got to share this irreplaceable experience with her, with teachings going much beyond the mere results of these pages.

Special thanks go to professors Jorge Mateu and Ottmar Cronie, who allowed me to spend time working with them abroad. In particular, Jorge's enthusiasm to start working together even from far apart has been a ray of hope during the first period of my doctoral course, and his supervision since then never failed to make me progress. I am particularly grateful to Ottmar for having shared with me not only his immense statistical knowledge but also all the doubts I am sure every researcher skilled as he has gone through.

I also want to thank all the statistics professors in Palermo, whose kindness never ceases to amaze me. They surely helped me gain the technical skills to finish this research and also nourished my self-confidence. Professor Marcello Chiodi deserves particular mention for the opportunities offered and for always showing unconditional trust in my abilities. Above all, nothing makes me more grateful than having shared these years' experiences with my colleagues Ale, Prilli, Andri, Furio, and Toti. They are by far the best mates one could ever wish for.

All of this has been possible thanks to my family and friends love. My parents, who never doubted my choices so far, taught me the importance of celebrating every success, as well as failure in life. Without this in mind, I would not have enjoyed these years' challenges as I did. Among my friends, a special thank goes to my very best friend Agnese, which not only believed in me but forced me to believe it myself.

All these people have been sources of inspiration for both my academic and personal life, and I genuinely hope they will continue to be so in future.

Contents

Chapter 1

Introduction

Point processes are stochastic processes defining a natural and convenient formal tool to describe the process of discrete events that occur in a continuous space, time or a space-time domain. Examples, spanning many scientific branches, are found with forest fires, crimes, earthquakes, diseases, tree locations, animal locations or communication network failures, to name just a few. Depending on the domain where the events occur, we can talk about spatial, temporal or spatio-temporal point processes. When time is present, the process has an evolution in time, and the events can be sorted according to their chronological order, sharing some common features with time series. When a property or a characteristic can also be attached to each event, such as the magnitude of an earthquake or the burned area of a wildfire, the point process is then called a marked point process.

Modelling real problems through space-time point processes is crucial in many scientific and engineering fields, such as environmental sciences, meteorology, image analysis, seismology, astronomy, epidemiology and criminology. Indeed, specifying a statistical model is a very effective way compared to analyzing data by calculating summary statistics. Formulating and fitting a statistical model to the data allows for taking into account effects that otherwise could introduce distortion in the analysis. In parallel, the growing data availability leads scientific research to look for more detailed information through statistical methodologies suitable for describing complex phenomena.

This thesis aims at providing new local methods and models for such complex spatio-temporal point processes. Going "local" aims to obtain greater detail in the study of space-time phenomena, identifying outlying points or sub-areas that require greater attention and could be the subject of further studies. Methods such as dependence models or second-order characteristics, useful for describing point processes, are typical tools used in the global context, which can serve as the basis for building new local tools.

Therefore, the contribution of this thesis includes several aspects of local characteristics of spatial and spatio-temporal point processes, and the main aims addressed are (1) to individuate sub-regions where point patterns cluster, (2) to take into account individual contributions of points into different statistical procedures such as hypothesis testing, point estimation, and diagnostics, and (3) to include accessory available information, such as the area and points.

It is important to highlight that the context of application is very broad, as local methods are of interest in describing any phenomenon with a complex spatio-temporal dependence. Some examples addressed in this thesis include seismic events, GPS data, traffic accidents, and crimes. Nevertheless, the motivating problems arising from the several contexts of applications serve as starting point to develop different methods ascribable to specific research goals. The methods and models that are developed in this thesis can hopefully be applied to different scientific fields and could be suitable for all those phenomena for which it makes sense to hypothesize an interdependence in space and time. Furthermore, this thesis covers a wide range of topics, including modelling, statistical inference, and simulation issues on spatial and spatio-temporal point processes, point processes on linear networks, and non-Euclidean spaces.

Indeed, in the last decade, spatial statistics has undergone an extraordinary methodological and computational advancement focused on generalizing and extending the foundations of the theories to more complex geometric spaces that allow fairer statistical analysis of new kinds of spatial data. In particular, spatial and spatio-temporal point process methodologies have been addressed to consider non-classical geometric supports for analyzing events on linear networks, such as traffic accidents in a city.

1.1 Complex point process models

Specifying a statistical model is a very effective way compared to analyzing data by calculating summary statistics. Formulating and adapting a statistical model to the data allows taking into account effects that otherwise could introduce distortion in the analysis (Baddeley et al., 2015).

In the context of spatial point processes, for example, it may be suitable to assume a log-linear Poisson model depending on a set of explanatory variables. An extension of these global models to the local case has been proposed by Baddeley (2017), introducing models based on local likelihood, known in this context as "geographically weighted regression models". Indeed, it might be unrealistic to assume that the model parameters are constant across the analyzed region, and it might be useful to specify a model where parameter estimates vary in space. For example, regarding the study of seismic events, a log-linear relationship could be assumed between the intensity of earthquakes and the distance from the nearest seismic source. In this example, we could assume that the model's coefficients vary as a function of space due to the different genesis of events. The use of local approaches is beneficial for describing areas with inhomogeneous characteristics induced by underlying complex spatial processes.

It is essential to specify that the description of real point processes often requires more complex models than a Poisson model since the latter is based on the assumption of the absence of interdependence between events. Two large classes of alternative models for the analysis of spatial point processes are the Gibbs and Cox processes (Van Lieshout, 2000; Moller and Waagepetersen, 2003; Illian et al., 2008). Gibbs processes typically exhibit negative association, while Cox processes always exhibit positive association. Moreover, for these large classes of models Baddeley (2017) has proposed a local version, thus allowing to obtain estimates of the parameters that vary in space, not only representing the effect of the explanatory variables included in the model but also the interdependence structure between the points.

Models used for studying and predicting space-time phenomena, when these have a clustered structure, are the log-Gaussian Cox processes (Møller et al., 1998) models. These models can be referred to as *empirical models*, and they are based on relatively simple approximations, on observation rather than theory. Alternatively, *mechanistic models* are based on mathematical descriptions of mechanical, chemical, biological (etc.) phenomena or processes. The former describes the pattern in a dataset without referring to any particular underlying mechanism (e.g. by a Gaussian Random Field), while in the latter, the interaction between events is expressed considering the physical knowledge on the process, typically through a conditional intensity. A typical example of mechanistic models is the Self-Exciting models, like the ETAS models (Ogata and Katsura, 1988). The general Self-Exciting models are applied for the description of space-time phenomena and are often used to model events that appear to be clustered over time. They were introduced for point-of-time processes by Hawkes and Adamopoulos (1973) and then extended to the space-time context.

A point process is called Self-Exciting if the occurrence of an event increases the probability of observing other points, while it is called Self-Correcting if the points take on an inhibitory effect (Schoenberg, 1999). The Self-Exciting process can be defined in terms of the conditional intensity function (Daley and Vere-Jones, 2007) as the sum of the spatial density of the background component and the density function of the induced component. Methods for describing such processes are typically based on the arrangement of points in space and time without additional information. Only recently, some models that allow the inclusion of explanatory variables have been implemented (Meyer et al., 2012; Adelfio and Chiodi, 2020).

Log-Gaussian Cox process models, on the other hand, are modifications of Poisson processes that allow the incorporation of random influences in the models. Indeed, a Cox process model is often chosen when it is believed that the arrangement of points is influenced by unobserved or unobservable explanatory variables. Specifically, a log-Gaussian Cox process model is a Cox process whose intensity is taken as the exponential of a Gaussian random field. Log-Gaussian Cox process models are therefore made up of a deterministic and a stochastic part. In the deterministic component, it is possible to consider the effect of the explanatory variables, exactly as in the linear predictor of the log-linear Poisson models. Thanks to the stochastic component, on the other hand, it is possible to describe the interdependence structure among the points through parameters of scale and variance.

Both Self-Exciting and log-Gaussian Cox processes are mainly used for the study of space-time phenomena, and they are particularly appropriate for the purposes of this thesis, as they allow us to consider explanatory variables suitable for describing the phenomenon studied. Therefore, the inclusion of explanatory variables in the model can be linked to the possibility of accounting for the local effects of some covariates in a space or space-time domain.

1.2 Outline of the thesis

Chapter 2 aims at introducing the preliminaries for point processes and their local characteristics. As most spatio-temporal models and methods have been developed as a series of extensions of spatial methods, we present the methodology starting from purely spatial point processes definitions, moving then to space and time. Moreover, we outline the main differences when dealing with global and local characteristics of point processes. Finally, we turn to some definitions of the global framework of point processes occurring on the non-Euclidean space of linear networks.

Chapter 3 presents novel fitting procedures for local complex spatio-temporal point processes, applied to seismic data. The motivation comes from seismic data analysis, and the aim is to take into account individual contributions of points into the statistical inference of different models. This allows for obtaining varying parameters and identifying sub-regions of the analysed areas reflecting different first- and second-order characteristics. The methods in this chapter are developed and applied to real seismic datasets.

In detail, section 3.1 provides the motivating problem through an application of the local (purely) spatial models. We propose the use of advanced and flexible statistical models to describe the spatial displacement of earthquake data. The section aims to account for the external geological information in the description of complex seismic point processes, through the estimation of models with space-varying parameters. A local version of the log-Gaussian Cox processes (LGCPs) is introduced and applied for the first time, exploiting the local Palm likelihood inferential tool. We provide methods and approaches accounting for the interaction among points, typically described by LGCP models through the estimation of the covariance parameters of the Gaussian Random Field (GRF) that, in this local version, are allowed to vary in space, providing a more realistic description of the clustering feature of seismic events. Furthermore, we contribute to the framework of diagnostics, outlining suitable methods for the local context and proposing a new step-wise approach addressing the particular case of multiple covariates. Overall, we show that local models provide good inferential results and could serve as the basis for future spatio-temporal local model developments, peculiar for the description of the complex seismic phenomenon.

After that, in section 3.2, the local method of *minimum contrast*, for estimating the second-order parameters of a general Cox process, is extended to space and time, and is used for fitting local spatio-temporal LGCPs. We propose a local version of spatio-temporal LGCPs using Local Indicators of Spatio-Temporal Association (LISTA) functions into the minimum contrast procedure to obtain space as well as time-varying parameters. We resort to the joint minimum contrast fitting method to estimate the set of second-order parameters. This approach has the advantage of being suitable in both separable and non-separable parametric specifications of the correlation function of the underlying GRF.

In chapter 4, two new modelling approaches on linear networks are proposed. The main aim is to include accessory available information on the area, i.e. constrained supports and/or spatial covariates, and/or points, through marks.

In section 4.1 the Gibbs models with random effects are adapted to the network structure and to the peculiar problem of describing cruise passengers' data. We analyse the spatio-temporal distribution of visitors' stops by touristic attractions in Palermo (Italy) using the theory of stochastic point processes living on linear networks. We first propose an inhomogeneous Poisson point process model, with a separable parametric spatio-temporal first-order intensity. We account for the spatial interaction among points on the given network, fitting a Gibbs point process model with mixed effects for the purely spatial component. This allows us to study first-order and second-order properties of the point pattern, accounting both for the spatio-temporal clustering and interaction and for the spatio-temporal scale at which they operate. Due to the strong degree of clustering in the data, we then formulate a more complex model, fitting a spatio-temporal LGCP to the point process on the linear network, and addressing the problem of the choice of the most appropriate distance metric.

In section 4.2, we formalize a Self-Exciting network model for the analysis of crime data. Motivated by the analysis of crime data in Bucaramanga (Colombia), we propose a spatio-temporal Hawkes point process model adapted to events living on linear networks. We first consider a non-parametric modelling strategy, for which we follow a non-parametric estimation of both the background and the triggering components. Then we consider a semi-parametric version, including a parametric estimation of the background based on covariates, and a non-parametric one of the triggering effects. Our model can be easily adapted to multi-type processes. Our network model outperforms a planar version, improving the fitting of the Self-Exciting point process model.

In chapter 5, we combine the two frameworks, i.e. local and network, by proposing the use of Local Indicators of Spatio-Temporal Association functions onto the networks. The aim is to take into account individual contributions of points in some statistical procedures, such as diagnostics, and hypothesis testing, in order to individuate sub-regions where point patterns behave differently in terms of clustering.

Specifically, in section 5.1, the proposed LISTA functions are used to build a statistical test for assessing local differences between the spatio-temporal second-order structure of two point patterns occurring on the same linear network. This allows to identify differences in the spatio-temporal clustering behaviour of two point patterns, a point pattern of interest and a background one, both occurring on the same linear network. We assess the performance of the testing procedure for local second-order structure through simulation studies under a variety of scenarios that also account for different generating point processes. We show that the proposed local test is able to correctly identify the spatio-temporal difference in the local second-order structure of two point patterns occurring on the same linear network, finding that the power of the test increases as the number of points and the clustered structure of the point pattern of interest increase as well.

Moreover, in section 5.2 we put particular emphasis on the local K-functions, deriving some theoretical results showing that LISTA functions are useful for the diagnostics of models specified on networks and showing that they can be used to assess the goodness-of-fit of different spatio-temporal models fitted to point patterns occurring on linear networks. Our methods do not rely on any particular model assumption on the data, and thus they can be applied to whatever the underlying model of the process.

Details on the simulation schemes implemented and used in this chapter come in Appendix A. We finally use the proposed methodologies for the analysis of two traffic-related problems.

Chapter 6 focuses on local characteristics of functional marked point processes. Examples of point patterns with associated functional data include forest patterns where for each tree, we have a growth function, curves representing the incidence of an epidemic over a period of time, and the evolution of distinct economic parameters such as unemployment and price rates, all for distinct spatial locations.

In section 6.1, we introduce a family of local inhomogeneous mark-weighted summary statistics, of order two and higher, for general marked point processes. Depending on how the involved weight function is specified, these summary statistics capture different kinds of local dependence structures. We first derive some basic properties and show how these new statistical tools can be used to construct most existing summary statistics for (marked) point processes. We then propose a local test of random labelling. This procedure allows us to identify points, and consequently, regions, where the random labelling assumption does not hold, e.g. when the (functional) marks are spatially dependent. Through a simulation study, we show that the test is able to detect local deviations from random labelling. We also provide an application to an earthquake point pattern with functional marks given by seismic waveforms.

In section 6.2 we only focus on the functional aspect of seismic data. In particular, we propose a novel picking algorithm for the automatic P- and S- waves onset time determination. Our algorithm is based on the variance piecewise constant models of the earthquake waveforms. The effectiveness and robustness of our picking algorithm are tested both on synthetic seismograms and real data. We simulate seismic events with different magnitudes (between 2 and 5) recorded at different epicentral distances (between 10 and 250 km). For the application to real data, we analysed waveforms from the seismic sequence of L'Aquila (Italy) in 2009. The obtained results are compared with those obtained by the application of the classic STA/LTA picking algorithm. Although the two algorithms lead to similar results in the simulated scenarios, the proposed algorithm results in greater flexibility and automation capacity, as shown in the real data analysis. Indeed, our proposed algorithm does not require testing and optimization phases, resulting potentially very useful in earthquakes routine analysis for novel seismic networks or in regions whose earthquakes characteristics are unknown.

In section 7, we exploit some theoretical results, from which we know the expected value of the local spatio-temporal K-function weighted by the true first-order intensity function. This theoretical result can serve as an estimation method for obtaining the parameters' estimates of a specific model assumed for the data. The motivation is to generally avoid dealing with the complex likelihoods of some point process models (like the ETAS process) and their maximization. By further considering the local second-order characteristics, we can obtain the whole set of parameters assumed for the fitted model, one for each point of the analysed point pattern.

Finally, in chapter 8, general conclusions are drawn, and some paths for future work are outlined.

Codes for the analyses were written using the software R Core Team (2022), whose description is provided in Appendix B.

Each section broadly corresponds to a scientific paper, referred in Appendix C together with authors' contributions.

Chapter 2

Preliminaries on point processes and their local characteristics

Following Cressie (2015), we introduce point processes by a mathematical approach that uses the definition of a counting measure on a set $X \subseteq \mathbb{R}^d, d \geq 1$, with positive values in $\mathbb{Z}$: for each Borel set B this $\mathbb{Z}_+$-valued random measure gives the number of events falling in B.

Definition 2.1. ***Point process***
Let $(\Omega, \mathcal{A}, P)$ be a probability space and Φ a collection of locally finite counting measures on $X \subset \mathbb{R}^d$. Define $\mathcal{X}$ as the Borel σ-algebra of X and let $\mathcal{N}$ be the smallest σ-algebra on Φ, generated by sets of the form $\{\phi \in \Phi : \phi(B) = n\}$ for all $B \in \mathcal{X}$. A point process N on X is a measurable mapping of $(\Omega, \mathcal{X})$ into $(\Phi, \mathcal{N})$. A point process defined on $(\Omega, \mathcal{A}, P)$ induces a probability measure $\Pi_N(Y) = P(N \in Y), \forall Y \in \mathcal{N}$.

Then, for any set $B \in \mathcal{X}$, $N(B)$ represents the number of points falling in B, such that if B is the union of disjoint sets $\tilde{B}_1, \tilde{B}_2, \ldots$, then $N(B) = \sum N(\tilde{B}_i)$.

Given a point process N defined on the space $(X, \mathcal{X})$ and a Borel set B, the number of points $N(B)$ in B is a random variable with first moment defined by:

$$\mu_N(B) = \mathbb{E}[N(B)] = \int_\Phi \phi(B)\Pi_N(d\phi)$$

that is a measure on $(X, \mathcal{X})$. The measure μ_N is called the mean measure or first moment measure of N (Cressie, 2015).

Higher-order moments can be defined: the k-th moment measure of N is given by

$$\mu_N^{(k)}(B_1 \times \ldots \times B_k) = \mathbb{E}[N(B_1) \ldots N(B_k)] = \int_\Phi \phi(B_1) \ldots \phi(B_k)\Pi_N(d\phi)$$

where $B_1, B_2, \ldots, B_k \in \mathcal{X}$.

Note that $\mu_N^{(k)}$ is a measure on $(X^k, \mathcal{X}^{(k)})$ and that $\mathcal{X}^{(k)}$ is the smallest σ-algebra found by the product sets $B_1 \times \ldots \times B_k$ where $B_i \in \mathcal{X}$.

The process is *second-order* if $\mu_N^{(2)}(B_1 \times B_2)$ is finite in $\mathcal{X}^2$.

As the covariance of random variables the covariance measure of a point process is:

$$C_N(B_1 \times B_2) = \mu_N^{(2)}(B_1 \times B_2) - \mu_N(B_1)\mu_N(B_2)$$

where $B_1, B_2 \in \mathcal{X}$: C_N is a signed measure on $(X^2, \mathcal{X}^2)$.

The k-th factorial moment measure of N is defined by:

$$\alpha_N^{(k)}(B_1 \times \ldots \times B_k) = \int_\Phi \sum_{s_1,\ldots,s_k \in \Phi}^{\neq} \mathbf{1}(s_1 \in B_1) \ldots \mathbf{1}(s_k \in B_k) \Pi_N(d\phi)$$

Note that $\alpha_N^{(k)}(B^k) = \mathbb{E}[(N(B)(N(B)-1)\ldots N(B)-k+1)]$ is the k-th factorial moment of the random variable $N(B)$.

For $k = 2$, $\alpha_N^{(2)}(B_1 \times B_2) = \mu_N^{(2)}(B_1 \times B_2) - \mu_N(B_1 \cap B_2)$ is the second-order factorial moment measure. Factorial moments are often used since ordinary moment measures of point process may be unsatisfactory. Indeed, the B_i's in $B_1 \times \cdots \times B_k$ are not necessary disjoint. This means that the events could occur in more of the B_i's causing redundancy of information and singularity of the moment measures.

Definition 2.2. ***Simple or orderly point process.***
Let N be a point process defined on $X \subseteq \mathbb{R}^d, d \geq 1$, and s a point of X. N is said simple if

$$P\{N(s) = 0 \textit{ or } 1 \textit{ for all } s\} = 1$$

that is if $\phi(\{s\}) \in \{0, 1\}$ for all s and almost all $\phi \in \Phi$.

The homogeneous Poisson process is the basis of the theory of point processes. Given the space $X \subseteq \mathbb{R}^d$, the homogeneous Poisson process on X is defined by the following postulates:

1. for some $\lambda > 0$ and any bounded Borel subset B of $\mathcal{X}$, $N(B)$ follows a Poisson distribution with mean $\lambda v(B)$, where $v(\cdot)$ indicates the Lebesgue measure;

2. given $N(B) = n$, the n events in B form an independent random sample from the uniform distribution on B;

3. the random variables $N(B_i)$ and $N(B_j)$, representing the number of events occurring in disjoint Borel sets B_i and B_j, are independent (independent increments).

The homogeneous Poisson process is used to describe events with a completely random occurrence on the considered space and, usually, it represents the basis for a qualitative comparison.

If phenomena under study show a variable intensity function over the state space, the hypothesis of stationarity has to be relaxed and the constant intensity of the Poisson process is replaced by a varying intensity function $\eta(s)$.

Given the space $X \subseteq \mathbb{R}^d$ and for any Borel set B of $\mathcal{X}$, the inhomogeneous Poisson process can be introduced providing the following conditions:

1. $N(B)$ has a Poisson distribution with mean $\int_B \eta(s)ds$.
2. Given $N(B) = n$, the n events in B form an independent random sample from a distribution on B with density function proportional to $\eta(s)$, with $s \in X$.
3. The process has independent increments.

An important point process often used in the theory of seismic process analysis is the marked point process. It consists of the locations of events x_i in a studied region $X \subseteq \mathbb{R}^d$ and associated random measurements or marks m_i in a set $\mathcal{M}$, simply assumed to be $\mathbb{R}$ or a subset of $\mathbb{R}$. For any point $x \in X$ marked by a quantity $m \in \mathcal{M}$, then (x, m) is a point in $X \times \mathcal{M}$ and, therefore, a marked point process on X can be represented as a point process on the product space $X \times \mathcal{M}$.

The process of locations $\{x_i\}$ is named ground process, denoted by N_g.

Definition 2.3. ***Simple and Stationary marked point process***
Let $N = \{(x_i, m_i)\}$ be a marked point process on $X \times \mathcal{M}$, with ground process N_g.

(a) The marked point process N is simple if the ground process N_g is simple.

(b) It is stationary on X if the probability structure of the process is invariant under shifts in X.

If the ground process is not simple, it can be considered as a cluster process, in which clusters centres are the single locations x_i and clusters members are all pairs (x_i, m_{ij}) where m_{ij} are the marks of points with common location x_i.

2.1 Spatial point processes

A spatial point pattern N is an unordered set $\mathbf{x} = \{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$ of points $\mathbf{x}_i$ where $n(\mathbf{x}) = n$ denotes the number of points, not fixed in advance. If $\mathbf{x}$ is a point pattern and $D \subset \mathbb{R}^2$ is a bounded region, we write $\mathbf{x} \cap D$ for the subset of $\mathbf{x}$ consisting of points that fall in D and $n(\mathbf{x} \cap D)$ for denoting the number of points of $\mathbf{x}$ falling in D. A point process model assumes that $\mathbf{x}$ is a realization of a finite point process N in D without multiple points. A point location in the plane is denoted by a lowercase letter like $\mathbf{u}$. Any location $\mathbf{u}$ can be specified by its Cartesian coordinates $\mathbf{u} = (u_1, u_2)$ in such a way that we do not need to mention the coordinates explicitly. The first-order property of N is described by the intensity function, defined as

$$\lambda(\mathbf{u}) = \lim_{|d\mathbf{u}| \to 0} \frac{\mathbb{E}[N(d\mathbf{u})]}{|d\mathbf{u}|}$$

where $d\mathbf{u}$ is an infinitesimal region that contains the point $\mathbf{u} \in D$, $|\mathrm{d}\mathbf{u}|$ is its area and $\mathbb{E}[N(\mathrm{d}\mathbf{u})]$ denotes the expected number of events in $\mathrm{d}\mathbf{u}$. When the intensity is constant, the process is called homogeneous. In the inhomogeneous case, the intensity is not constant in the study area but may depend, for instance, on the coordinates of points. A point process model, assuming independence, is completely described by its intensity function $\lambda(\mathbf{u})$ (Daley and Vere-Jones, 2007). The inter-point interaction between events is measured by the second-order moment characteristics, studied through the second-order intensity function, defined as

$$\lambda_2(\mathbf{u}, \mathbf{v}) = \lim_{|d\mathbf{u}||d\mathbf{v}| \to 0} \frac{\mathbb{E}[N(d\mathbf{u})N(d\mathbf{v})]}{|d\mathbf{u}||d\mathbf{v}|}$$

where $\mathbf{u} \neq \mathbf{v}$. Refer to Daley and Vere-Jones (2007), Preposition 3.3.I, for existence and convergence property of the point processes intensity.

Several functional summary statistics are used to study the second-order characteristics of a point pattern and to measure dependence.

2.1.1 Second-order characteristics of spatial point processes

A widely used summary statistic for descriptive analysis and diagnostics is Ripley's K-function (Ripley, 1976, 1988), which is defined as

$$K(r) = 2\pi\lambda^{-2} \int_0^r \lambda_2(x) x \mathrm{d}x,$$

for a stationary and isotropic process for which $\lambda_2(r) = \lambda^2$ (Schabenberger and Gotway, 2017). It is also known as the reduced second-moment measure (Cressie, 2015), as the second reduced moment function (Chiu et al., 2013), and as the second order reduced moment measure (Moller and Waagepetersen, 2003).

Given that in a stationary process the distribution of N is the same as the distribution of the shifted process $N + \mathbf{v}$, for any vector $\mathbf{v}$, the K-function is defined as

$$K(r) = \frac{1}{\lambda}\mathbb{E}\left[\text{number of r-neighbours of } \mathbf{u} | N \text{ has a point at location } \mathbf{u}\right]$$

for any distance $r \geq 0$ and any location $\mathbf{u}$. Being a measure of the distribution of the inter-point distances, $K(r)$ captures the spatial dependence between different regions of a point process. Under the homogeneous Poisson assumption, the following theoretical relation holds $K_{pois}(r) = \pi r^2$. When the process is stationary, the most commonly used estimator is introduced in Ripley (1976):

$$\hat{K}(r) = \frac{|D|}{n(n-1)} \sum_{i=1}^{n} \sum_{i \neq j} I\{d_{ij} \leq r\} e(\mathbf{x}_i, \mathbf{x}_j, r) \tag{2.1}$$

where $I\{\cdot\}$ is the indicator function, n is the number of points in the pattern, $|D|$ is the area of the window, $e(\mathbf{x}_i, \mathbf{x}_j, r)$ is the Ripley's edge correction and $d_{ij} = ||\mathbf{x}_i - \mathbf{x}_j||$ is the pairwise distances between all distinct pairs of points $\mathbf{x}_i$ and $\mathbf{x}_j$ in the pattern. When a process has $\hat{K}(r) < \pi r^2$, it means that the points tend to cluster, and when $\hat{K}(r) > \pi r^2$, it means that the points tend to inhibit each other.

Let $\lambda(\mathbf{x}_i)$ be the intensity function characterizing the generating process N, then each point $\mathbf{x}_i$ can be weighted by the inverse of the intensity function, i.e. $1 \backslash \lambda(\mathbf{x}_i)$, the reciprocal of the intensity at $\mathbf{x}_i$, and each pair of points $\mathbf{x}_i, \mathbf{x}_j$ will be weighted by $1/(\lambda(\mathbf{x}_i)\lambda(\mathbf{x}_j))$. The inhomogeneous K-function is defined as:

$$K_{inhom}(r) = \mathbb{E}\left[\sum_{\mathbf{x}_j \in \mathbf{X}} \frac{1}{\lambda(\mathbf{x}_j)} I\{0 < \|\mathbf{x}_i - \mathbf{x}_j\| \leq r\} \middle| \mathbf{x_i} \in \mathbf{X}\right]$$

assuming that this does not depend on the location $\mathbf{x}_i$ (Baddeley et al., 2000). The standard estimator of the K-function in Equation (2.1) can be extended to the inhomogeneous K-function as follows:

$$\hat{K}_{inhom}(r) = \frac{1}{D} \sum_{i} \sum_{j \neq i} \frac{I\{\|\mathbf{x}_i - \mathbf{x}_j\| \leq r\ \}}{\hat{\lambda}(\mathbf{x}_i)\hat{\lambda}(\mathbf{x}_j)} e(\mathbf{x}_i, \mathbf{x}_j; r) \tag{2.2}$$

where $e(\mathbf{x}_i, \mathbf{x}_j, r)$ is an edge correction weight as before, and $\hat{\lambda}(\mathbf{u})$ is an estimate of the intensity function $\lambda(\mathbf{u})$. If $\hat{\lambda}(\mathbf{u})$ is the intensity estimated through a fitted model, then the inhomogeneous version of the K-function can be used as a diagnostic tool.

The inhomogeneous K-function is useful for interpreting the local features of data since if the estimated intensity $\hat{\lambda}(\mathbf{u})$ is close to the generating one $\lambda(\mathbf{u})$, the estimated $\hat{K}_{inhom}(r)$ should behave like the corresponding function under a Poisson model. Moreover, $\hat{K}_{inhom}(r)$ values greater than the expected value under the Poisson model $K_{inhom,pois}(r) = \pi r^2$ (exactly as for the homogeneous case) indicate that the fitted model is not appropriate, since the distances computed among points exceed the Poisson theoretical ones, suggesting those locations where departures are more evident. In practice, given the fitted competitive models, the model with the estimated inhomogeneous K-function which most resembles the theoretical Poisson one has to be preferred.

2.1.2 Spatial Poisson models

A general spatial log-linear Poisson model (Cox, 1972), generalizes both homogeneous and inhomogeneous models, such that:

$$\lambda(\mathbf{u}) = \lambda(\mathbf{u}, \boldsymbol{\theta}) = \exp(B(\mathbf{u}) + \boldsymbol{\theta}^\top \mathbf{Z}(\mathbf{u})), \tag{2.3}$$

where $\mathbf{u} \in D$, $B(\mathbf{u})$ and $\mathbf{Z}(\mathbf{u}) = (Z_1(\mathbf{u}), \ldots, Z_p(\mathbf{u}))$ are known functions (covariates), $\boldsymbol{\theta}^\top = (\theta_1, \ldots, \theta_p)$ are unknown parameters, and therefore $\boldsymbol{\theta}^\top \mathbf{Z}(\mathbf{u}) = (\theta_1 Z_1(\mathbf{u}) + \ldots + \theta_p Z_p(\mathbf{u}))$. These models have an especially convenient structure, since the log intensity is a linear function of the parameters, and covariates can be quite general functions, making them a very wide class of models. The estimation of the point process parameters is carried out through the maximization of the log-likelihood, defined by:

$$\mathrm{logL}(\boldsymbol{\theta}) = \sum_i \mathrm{log}\lambda(\mathbf{x}_i; \boldsymbol{\theta}) - \int_D \lambda(\mathbf{u}; \boldsymbol{\theta}) d\mathbf{u} \tag{2.4}$$

where the sum is over all point $\mathbf{x}_i$ in the point process $\mathbf{x}$ (Daley and Vere-Jones, 2007).

2.1.2.1 Spatial log-Gaussian Cox processes

The main feature of the Poisson point process is independence and stationarity, that is, a theoretical assumption often inappropriate for describing real data. However, considering more general point process models may be a complex issue in terms of inference and diagnostics (Adelfio and Schoenberg, 2009; Adelfio and Chiodi, 2015a), but necessary if dealing with real contexts where dependence structure among data is of main interest (Siino et al., 2017, 2018a). Two large classes of alternative models for spatial point processes are the Gibbs and Cox processes (Van Lieshout, 2000; Moller and Waagepetersen, 2003; Illian et al., 2008). For clustered point patterns, the most commonly-used models are Cox and cluster processes (Baddeley et al., 2015, Chapter 12).

The Cox processes are a generalisation of the inhomogeneous Poisson processes, where the intensity is a realisation of a random field. For this reason, they are also called doubly stochastic Poisson processes. The point process N is said to be a Cox process driven by Λ, if the conditional distribution of the point process N given a realisation $\Lambda(\mathbf{u}) = \lambda(\mathbf{u})$ is a Poisson process on D with intensity function $\lambda(\mathbf{u})$.

The most used Cox point processes are the shot-noise Cox processes (Moller and Waagepetersen, 2003) and the log-Gaussian Cox processes (Møller et al., 1998). In a shot-noise Cox process, offspring are observed around unseen parents representing the random unobservable variability.

Following the inhomogeneous specification in Diggle et al. (2013), the log-Gaussian Cox process for a generic point in space has the following intensity

$$\Lambda(\mathbf{u}) = \lambda(\mathbf{u}) \exp(S(\mathbf{u})) \tag{2.5}$$

where S is a Gaussian Random Field (GRF) with mean function $\mu = \mathbb{E}(S(\mathbf{u})) = -\frac{\sigma^2}{2}$ so that $\mathbb{E}[\exp(S(\mathbf{u}))] = 1$. The covariance function is $\mathbb{C}(S(\mathbf{u}), S(\mathbf{v})) = \mathbb{C}(||\mathbf{u}-\mathbf{v}||) = \sigma^2 \rho(r)$ under the stationary assumption, i.e. only depends on the variance parameter σ^2, and distance $r = \|\mathbf{u} - \mathbf{v}\|$ between locations $\mathbf{u}$ and $\mathbf{v}$.

$\rho(\cdot)$ is the correlation function of the GRF, completely specified by its first and second moments. Following the notation used in Baddeley et al. (2015), the random intensity $\Lambda(\mathbf{u}) = \exp(S(\mathbf{u}))$ is a log-Gaussian or log-normal distribution. If $Z \sim N(\mu, \sigma^2)$ and $\Lambda = \exp(Z)$ then Λ has mean $\exp(\mu + \sigma^2/2)$. In this thesis, we assume the exponential structure for the covariance function as in Brix and Diggle (2001),

$$\mathbb{C}(r) = \sigma^2 \exp\left(-\frac{r}{\alpha}\right) \tag{2.6}$$

depending only on α and σ^2 that represent the scale parameter for the spatial distance and the variance, respectively. The effect of increasing σ^2 is to generate higher peaks in the surface intensity which leads to clusters of points. Increasing the spatial scale parameter α, the underlying GRF presents a stronger spatial correlation, and it corresponds to a diffuse aggregation of points of the LGCP. Other basic stationary and isotropic family models commonly used for the spatial component are the Matérn, Cauchy, Gaussian and Spherical covariance functions (Gelfand et al., 2010). In a spatial log-Gaussian Cox process, the intensity model can be expressed as a function of further covariates, according to a log-linear model as in Equation (2.3). We know that $\lambda(\mathbf{u}) = \mathbb{E}(\Lambda(\mathbf{u})) = \mathbb{E}(\exp(S(\mathbf{u}))) = \exp((\mu(\mathbf{u}) + \sigma^2/2))$, where

$$\mu(\mathbf{u}) = -\frac{\sigma^2}{2} + B(\mathbf{u}) + \boldsymbol{\theta}^T \mathbf{Z}(\mathbf{u}) \tag{2.7}$$

is the mean function of the random intensity function. This peculiarity, together with the possibility of detecting the clustered structure of the point process, makes the LGCP model attractive for modelling seismic data.

2.1.3 Gibbs point processes

It has become widely accepted in spatial statistics that Gibbs models are suitable primarily for modelling regularity or inhibition between points, and are not able to model strong clustering (Illian et al., 2008; Moller and Waagepetersen, 2003) although moderate clustering can be achieved. However, this conclusion is based on experience with the simple Gibbs models that have predominated in the literature, in particular with *pairwise interaction processes* (Baddeley et al., 2013).

Definition 2.4. *A finite point process whose probability density f is hereditary is called a (finite) Gibbs point process. For a Gibbs process, the (Papangelou) conditional intensity at a location $u \in W$ given a configuration $\boldsymbol{x} \in \mathcal{X}$ may be defined as*

$$\lambda(u, \boldsymbol{x}) = \frac{f(\boldsymbol{x} \cup \{u\})}{f(\boldsymbol{x} \backslash \{u\})},$$

provided the numerator and denominator are nonzero, and $\lambda(u, \boldsymbol{x}) = 0$ otherwise.

The most common models considered in the literature are *pairwise interaction processes* of the general form

$$h(\mathbf{x}) = \left[\prod_{i=1}^{n} b(x_i)\right]\left[\prod_{i<j} c(x_i, x_j)\right], \tag{2.8}$$

where $b : W \rightarrow [0, \infty)$ and $c : W \times W \rightarrow [0, \infty)$ are given functions. The second bracketed term is a product over all unordered pairs of points in $\mathbf{x}$. If b is constant and c is translation equivariant, i.e., if $c(u, v) = c(u - v)$, then we shall call the point process *stationary*. A simple special case of (2.8) is the Strauss process (Strauss, 1975). Many other models of the class of (finite) Gibbs point processes can be fitted rapidly to real data sets containing large numbers of points. We refer to Baddeley et al. (2000, 2006b); Diggle (2013); Geyer (1999); Moller and Waagepetersen (2003) for model-fitting, prediction and simulation details.

2.1.4 Local spatial Poisson models

In this section, we report some basic definitions and results about the local composite likelihood applied to Poisson point processes, following the notation used in Baddeley (2017). In the local context, given a spatial point pattern $\mathbf{x}$, a realization of a finite point process N_1 in D, the *template model* is defined as a point process N_2, governed by a parameter $\boldsymbol{\theta} \in \Theta \subseteq \mathbb{R}^p$. N_2 is defined as the "template" model rather than the "homogeneous" one, because the potential dependence of the distribution of N_2 on some spatial covariates would make N_2 a spatially inhomogeneous point process. The template model is assumed to provide good local approximations to N_1, in the sense that, in any small region $B \subset D$, the distribution of $N_1 \cap B$ is well-approximated by the distribution of $N_2 \cap B$ of parameter $\boldsymbol{\theta}$ which may depend on B.

When spatial log-linear models are used, the model parameters are usually assumed to be constant across the entire study region. This assumption may be too simplistic in contexts that are characterised by multi-scale and fractal features, like the seismic one, where the relationship between the intensity of earthquakes and other possible characteristics of the area where events occur. In this context, the features can be described by a Poisson process with intensity

$$\lambda(\mathbf{u}) = \lambda(\mathbf{u}; \boldsymbol{\theta}(\mathbf{u})) = \exp(B(\mathbf{u}) + \boldsymbol{\theta}^\top(\mathbf{u})\mathbf{Z}(\mathbf{u})) \tag{2.9}$$

with local coefficients, where $\boldsymbol{\theta}(\cdot)$ is a function of the spatial location. This is the "inhomogeneous" alternative to the template log-linear intensity (2.3). For estimating local models as in Equation (2.9), the likelihood at $\mathbf{u}$ is used, being a weighted version of the original likelihood in Equation (2.4), with the greatest weight on locations close to $\mathbf{u}$.

Therefore, in the local context, the local log-likelihood associated with location $\mathbf{s}$ (Loader et al., 1999) is:

$$\log \mathrm{L}(\mathbf{s}; \boldsymbol{\theta}) = \sum_{i=1}^{n} w_h(\mathbf{x}_i - \mathbf{s}) \log \lambda(\mathbf{x}_i; \boldsymbol{\theta}) - \int_D \lambda(\mathbf{u}; \boldsymbol{\theta}) w_h(\mathbf{u} - \mathbf{s}) d\mathbf{u} \tag{2.10}$$

where $w_h(\mathbf{u}) = h^{-d} w(\mathbf{u}/h)$ is a weight non-parametric function, and $h > 0$ is a smoothing bandwidth. It is not necessary to assume that w_h is a probability density. Throughout the thesis, we will consider a kernel of fixed bandwidth h. In the local likelihood in Equation (2.10) this is usually chosen by the cross-validation criterion (Loader et al., 1999). The optimal bandwidth h_{opt} maximises

$$\mathrm{LCV}(h) = \sum_i \log \lambda(\mathbf{x}_i; \hat{\boldsymbol{\theta}}_i(\mathbf{x}_i)) - \int_D \lambda(\mathbf{u}; \hat{\boldsymbol{\theta}}(\mathbf{u})) d\mathbf{u} \tag{2.11}$$

where $\hat{\boldsymbol{\theta}}(\mathbf{u}) = \hat{\boldsymbol{\theta}}(\mathbf{u}, h)$ is the local estimate of $\boldsymbol{\theta}$ at location $\mathbf{u}$ using bandwidth $h > 0$, and $\hat{\boldsymbol{\theta}}_i(\mathbf{x}_i) = \hat{\boldsymbol{\theta}}_i(\mathbf{x}_i; h)$ is the corresponding "leave-one-out" estimate at the location $\mathbf{x}_i$ computed from the data omitting $\mathbf{x}_i$.

Maximizing the local likelihood for each fixed $\mathbf{u}$ gives local parameter estimates $\hat{\boldsymbol{\theta}}(\mathbf{u})$ of the original model, confidence intervals, hypothesis tests and other standard tools (Baddeley, 2017). These fitted coefficients can be plotted as a function of spatial location $\mathbf{u}$. The fitted intensity can be obtained as a function of $\mathbf{u}$ too, as $\hat{\lambda}(\mathbf{u}) = \lambda(\mathbf{u}; \hat{\boldsymbol{\theta}}(\mathbf{u}))$. It is worth recalling that the local likelihood approach assumes that the template model is a good local approximation to the data. On the one hand, this allows using existing statistical techniques and theory for the template model to be easily modified in local estimation and inference. On the other hand, there is no explicit description of the "true" model, but only of the template model, which is the local approximation of the true one.

The key assumption is that, for each spatial location $\mathbf{u} \in D$, there is an unobserved spatial subdomain $D(\mathbf{u})$ containing $\mathbf{u}$ where the template model is *exactly* true and parameter $\boldsymbol{\theta}(\mathbf{u})$. If the support of the kernel centred at $\mathbf{u}$ falls entirely inside $D(\mathbf{u})$, then the statistical properties of $\hat{\boldsymbol{\theta}}(\mathbf{u})$ are the same as they would be if the entire process followed the template model with constant parameter value $\boldsymbol{\theta} = \boldsymbol{\theta}(\mathbf{u})$ (Baddeley, 2017).

2.1.4.1 Local spatial log-Gaussian Cox processes

In this thesis, we formalize the characteristics of the local version of the LGCP model, also allowing the additional parameters σ and α to vary in space (denoted by $\sigma(\mathbf{u})$ and $\alpha(\mathbf{u})$ throughout the rest of the thesis), as well as the effects of the considered covariates, according to a log-linear model as in Equation (2.9). Recalling the specification of the global LGCP, we know that now, as the interaction parameters can vary in space, the GRF and its moments will also depend on the locations $\mathbf{u}$. In other words, for each location $\mathbf{u}$, we obtain a local GRF. Basically, while the global LGCP implies a unique generating process, the local LGCP implies that events are generated from multiple processes, i.e. with different correlation patterns, that can yet be estimated through a unique model. Therefore, the local LGCP represents an interesting model to analyse and describe those observed regions for which it is reasonable to assume different generating GRFs, characterized by different covariances. Furthermore, this implies that a local version of an LGCP model can be useful to spot regions with different correlation structures, and the mean of the local process in Equation (2.7), also depends on the space varying parameters $\boldsymbol{\theta}^T(\mathbf{u})$ and $\sigma^2(\mathbf{u})$. Therefore, if in the global model the mean of the process is influenced just by the linear predictor and by the constant σ^2, in the local LGCP, the mean varies also because of the local variation of $\boldsymbol{\theta}^T(\mathbf{u})$ and $\sigma^2(\mathbf{u})$ parameters. Then, the covariance function, assumed in Equation (2.6), now depends on the local variance parameter $\sigma(\mathbf{u})$ and spatial scale parameter $\alpha(\mathbf{u})$. A visual display of the space-varying estimates of the interaction parameters may suggest the presence of a space-varying clustered structure.

2.1.4.1.1 Local spatial LGCP model estimation through local Palm likelihood In this Section, the model estimation procedure based on the local version of the Palm likelihood of Ogata and Katsura (1991), and developed by Baddeley (2017), is reported. The Cox models are usually estimated by a two-step procedure involving both the intensity and the cluster or correlation parameters. In the first step, a Poisson model is fitted to the point pattern data, providing the estimates of the coefficients of all the terms in the model formula characterizing the intensity, as the one in the most general log-linear model in Equation (2.3). This intensity is then considered as the true one in the second step, and the cluster or correlation parameters are estimated by either the method of minimum contrast (Pfanzagl, 1969; Eguchi, 1983; Diggle, 1979; Diggle and Gratton, 1984; Siino et al., 2018a), Palm likelihood (Ogata and Katsura, 1991; Tanaka et al., 2008) or composite likelihood (Guan, 2006). Hereafter we will denote by $\boldsymbol{\theta}$ the vector of (first-order) intensity parameters and by $\boldsymbol{\psi}$ the cluster parameters, also denoted as correlation or interaction parameters by some authors. In the case of a spatial Log-Gaussian Cox process with exponential covariance as the one in (2.6), the cluster parameters correspond to $\boldsymbol{\psi} = (\alpha, \sigma)$.

In order to estimate the local version of the LGCP models, we refer to the adaptation of the Palm likelihood to a general nonstationary point process, considered by Baddeley (2017) because of its formal similarity to the Poisson likelihood. Ogata and Katsura (1991) developed a surrogate likelihood function called Palm likelihood for the analysis of stationary point processes.

The Palm distribution formalises the concept of conditioning on a point of the process. Assume that the process has the intensity function $\lambda(\mathbf{u};\boldsymbol{\phi})$ for $\mathbf{u} \in D$ and second moment intensity $\lambda_2(\mathbf{u},\mathbf{v};\boldsymbol{\phi})$ for $\mathbf{u},\mathbf{v} \in D$, depending on parameters $\boldsymbol{\phi} \in \Phi$. Define the intensity of the Palm distribution at $\mathbf{u}$ as

$$\lambda_p(\mathbf{u}|\mathbf{v};\boldsymbol{\phi}) = \lambda_2(\mathbf{u},\mathbf{v};\boldsymbol{\phi})/\lambda(\mathbf{u};\boldsymbol{\phi}). \tag{2.12}$$

The Palm log-likelihood can be written as

$$\log \mathrm{PAL}(\boldsymbol{\phi}) = \sum_i \left[\sum_{j \neq i} Q(\mathbf{x}_i,\mathbf{x}_j) \log \lambda_p(\mathbf{x}_j|\mathbf{x}_i;\boldsymbol{\phi}) - \int_D Q(\mathbf{x}_i,\mathbf{u}) \lambda_p(\mathbf{u}|\mathbf{x}_i;\boldsymbol{\phi}) d\mathbf{u} \right] \tag{2.13}$$

where $Q(\mathbf{u},\mathbf{v})$ is a 0/1 valued constraint function designed to simplify computation and optimise statistical properties, typically taken to be $Q(\mathbf{u},\mathbf{v}) = I\{\|\mathbf{u}-\mathbf{v}\|\} \leq R$ where $I\{\cdot\}$ is the indicator function, such that $I\{x\} = 1$ if x is true, and $R > 0$ is an upper bound to the correlation distance of the model. The quantity in square braces is recognisable as the log-likelihood of a Poisson point process with intensity $\lambda_p(\cdot|\mathbf{x}_i;\boldsymbol{\phi})$ restricted to the set where $Q(\mathbf{x}_i,\cdot) = 1$.

Dealing with the Cox processes, it is usual to work with models in which the intensity and interaction parameters are "separable"

$$\lambda_2(\mathbf{u},\mathbf{v};\boldsymbol{\phi}) = \lambda(\mathbf{u};\boldsymbol{\theta})\lambda(\mathbf{v};\boldsymbol{\theta})\rho(\mathbf{u},\mathbf{v};\boldsymbol{\psi}) \tag{2.14}$$

where $\boldsymbol{\phi} = (\boldsymbol{\theta},\boldsymbol{\psi})$, with $\boldsymbol{\theta}$ representing the intensity parameters and $\boldsymbol{\psi}$ the interaction parameters, and ρ is the correlation function. As stated by Baddeley (2017), examples include log-Gaussian Cox processes and certain inhomogeneous Neyman-Scott processes (Waagepetersen, 2007; Waagepetersen and Guan, 2009). Recalling what was previously introduced, specifying a local Log-Gaussian Cox process, $\boldsymbol{\theta}$ is the vector of parameters representing the effects of the covariates included in the intensity, while $\boldsymbol{\psi}$ is the vector of the interaction parameters σ and α, representing the spatial scale parameter and the variance of the underlying GRF.

The Palm log-likelihood (2.13) involves summation and integration over pairs of spatial points. To define a local version of the Palm likelihood, it is possible to apply local weights to the second element of each pair. That is, the weights are applied just to the points for which "predictions" are made. Summing over the first element

$\mathbf{x}_i$, we get:

$$\log \mathrm{PAL}(\mathbf{s};\boldsymbol{\phi}) = \sum_i \sum_{j\neq i} w_h(\mathbf{x}_j - \mathbf{s}) Q(\mathbf{x}_i,\mathbf{x}_j) \log \lambda_p(\mathbf{x}_j|\mathbf{x}_i;\boldsymbol{\phi}) \\ - \sum_i \int_D w_h(\mathbf{u}-\mathbf{s}) Q(\mathbf{x}_i,\mathbf{u}) \lambda_p(\mathbf{u}|\mathbf{x}_i;\boldsymbol{\phi}) d\mathbf{u}$$

that is basically the same Equation (2.13) obtained replacing $Q(\mathbf{u},\mathbf{v})$ by $w_h(\mathbf{v}-\mathbf{s})Q(\mathbf{u},\mathbf{v})$. This must be maximised numerically. We assume a separable isotropic model as in Equation (2.14). Recalling Equation (2.12), then we have

$$\lambda_p(\mathbf{v}|\mathbf{u};\boldsymbol{\phi}) = \lambda(\mathbf{v};\boldsymbol{\theta})\rho(||\mathbf{u}-\mathbf{v}||;\boldsymbol{\psi})$$

so that the local Palm log-likelihood to be maximised is

$$\log \mathrm{PAL}(\mathbf{s};\boldsymbol{\phi}) = \sum_i \sum_{j\neq i} w_h(\mathbf{s}-\mathbf{x}_i) Q(\mathbf{x}_i,\mathbf{x}_j) \{\log \lambda(\mathbf{x}_j : \boldsymbol{\psi}) + \log \rho(||\mathbf{x}_i - \mathbf{x}_j||;\boldsymbol{\theta})\} \\ - \sum_i \int_D w_h(\mathbf{s}-\mathbf{u}) Q(\mathbf{x}_i,\mathbf{u}) \lambda(\mathbf{u};\boldsymbol{\psi}) \rho(||\mathbf{x}_i - \mathbf{x}_j||;\boldsymbol{\theta}) d\mathbf{u}. \tag{2.15}$$

A two-step estimation procedure is used as follows:

- The intensity parameters $\boldsymbol{\theta}$ are estimated by maximising the Palm likelihood of the Poisson process with intensity $\lambda(\cdot;\boldsymbol{\theta})$, that is equivalent to a weighted Poisson likelihood.
- The interaction parameters $\boldsymbol{\psi}$ are estimated separately at each location $\mathbf{u}$ by maximising the local Palm likelihood (2.15) using the simplex algorithm.

If the Palm likelihood is taken in the form (2.13), the natural analogue of the Poisson likelihood cross-validation criterion (2.11) for Palm likelihood is

$$\mathrm{LCV}(h) = \sum_i \sum_{j\neq i} Q(\mathbf{x}_i,\mathbf{x}_j) \log \lambda_p(\mathbf{x}_j|\mathbf{x}_i;\hat{\boldsymbol{\phi}}_{-j}(\mathbf{x}_j)) - \sum_i \int_D Q(\mathbf{x}_i,\mathbf{u}) \lambda_p(\mathbf{u}|\mathbf{x}_i;\hat{\boldsymbol{\phi}}(\mathbf{u})) d\mathbf{u} \tag{2.16}$$

for bandwidth h, where $\hat{\boldsymbol{\phi}}_{-j}(\mathbf{x}_j)$ is the "leave-one-out" estimate, that is, the local parameter estimate at $\mathbf{s} = \mathbf{x}_j$ based on the data $\mathbf{x} \setminus \mathbf{x}_j$.

Alternative approaches for fitting local Cox and cluster processes include the methods of minimum contrast (Pfanzagl, 1969; Diggle and Gratton, 1984) and model-based clustering (Banfield and Raftery, 1993; Dasgupta and Raftery, 1998; Walsh and Raftery, 2002, 2005). In particular, a local version of the minimum contrast approach is developed using the local K-functions or the local pair correlation functions by Baddeley (2017), bearing a very close resemblance to the local Palm likelihood approach.

2.1.4.2 Diagnostics for local models

This section outlines relevant methods for carrying out diagnostics for local models. A particular focus is given to the model selection, dealing with multiple covariates. Some well-known methods for diagnostics are reviewed and proposed as model selection approaches in the context of local models. Basically, these concern intensity-based methods. After that, we focus on local tests, which are specific diagnostic tools for models with space-varying parameters, testing the significance of parameters and the properness of a local model.

2.1.4.2.1 Global diagnostics methods for local model selection This section reviews some methods widely used for diagnostics of global spatial point processes models, here proposed as diagnostic methods for spatial models in the local framework. As stated in some previous papers (Adelfio et al., 2020; Adelfio and Schoenberg, 2009), the main problem when dealing with residual analysis for point processes is to find a correct definition of residuals since the one used in dependence models cannot be used for point processes. Two of the most used methods for diagnostics of spatial point processes are the inhomogeneous K-function and the smoothed raw residuals, here reported.

For an inhomogeneous Poisson process model, with fitted intensity $\hat{\lambda}(\mathbf{u})$, the predicted number of points falling in any region D is $\int_D \hat{\lambda}(\mathbf{u})\mathrm{d}\mathbf{u}$. Hence, the residual in each region $D \subset \mathbb{R}^2$ is the "observed minus predicted" number of points falling in D (Alm, 1998), that is $R(D) = n(\mathbf{x} \cap D) - \int_D \hat{\lambda}(\mathbf{u})d\mathbf{u}$, where $\mathbf{x}$ is the observed point pattern, $n(\mathbf{x} \cap D)$ the number of points of $\mathbf{x}$ in the region D, and $\hat{\lambda}(\mathbf{u})$ is the intensity of the fitted model. A simple residual visualization can be obtained by smoothing them. The "smoothed residual fields" are defined as

$$s(\mathbf{u}) = \tilde{\lambda}(\mathbf{u}) - \lambda^{\dagger}(\mathbf{u}) \tag{2.17}$$

where $\tilde{\lambda}(\mathbf{u}) = e(\mathbf{u}) \sum_{i=1}^{n(\mathbf{x})} \kappa(\mathbf{u}-\mathbf{x}_i)$ is the non-parametric, kernel estimate of the fitted intensity $\hat{\lambda}(\mathbf{u})$, while $\lambda^{\dagger}(\mathbf{u})$ is a correspondingly-smoothed version of the (typically parametric) estimate of the intensity of the fitted model, $\lambda^{\dagger}(\mathbf{u}) = e(\mathbf{u}) \int_W \kappa(\mathbf{u} - \mathbf{v})\hat{\lambda}(\mathbf{v})d\mathbf{v}$. Here, κ is the smoothing kernel and $e(\mathbf{u})$ is the edge correction. The smoothing bandwidth for the kernel estimation of the raw residuals is selected by cross-validation as the value that minimises the Mean Squared Error criterion defined by Diggle (1985), by the method of Berman and Diggle (1989). See Diggle (2013) for further details. The difference in Equation (2.17) should be approximately zero when the fitted model is close to the real one. Therefore, the best model is the one with the lowest values of the smoothed raw residuals.

Throughout the thesis, these residuals are used for diagnostics, both for their flexibility and their simplicity of implementation, and the complexity of model estimation when the parameters vary in the space domain.

2.1.4.2.2 Local diagnostics methods: testing the local departure from "homogeneity" In this section, we introduce local methods for local model diagnostics, namely the tests of local departure from the "homogeneity" assumption, as proposed in Baddeley (2017). For further confuting this hypothesis, still retaining local features, and for assessing the contribution of each covariate to the rejection of the null hypothesis, dealing with multiple linear predictors, we propose a bootstrap and step-wise procedure in Sections 3.1.1 and 3.1.1.1, respectively.

Let us introduce the following system:

$$\begin{cases} \mathcal{H}_0: & \boldsymbol{\theta} \in \Theta_0 \\ \mathcal{H}_1: & \boldsymbol{\theta} \in \Theta \end{cases}$$

where $\Theta_0 \subset \Theta$, a "local test" can be introduced assessing for parameters dependence on the spatial location $\mathbf{u} \in D$. In the simplest case, suppose $\boldsymbol{\theta} = (\theta_1, ..., \theta_p)$ and assume that we want to verify

$$\begin{cases} \mathcal{H}_0: & \theta_j = 0 \\ \mathcal{H}_1: & \theta_j \neq 0 \end{cases}$$

The local version of the Wald test can be computed to assess the significance of the effects of the spatial covariates as a function of $\mathbf{u}$. This test is based on the standardised local coefficient estimate or "t-statistic"

$$t_j(\mathbf{u}) = \hat{\theta}_j(\mathbf{u}) / se(\hat{\theta}_j(\mathbf{u})), \tag{2.18}$$

whose asymptotic null distribution is the standard Normal. For asymptotic properties of local tests, refer to Baddeley (2017).

Local tests can also be used for cluster detection, defining a formal hypothesis test of local departure from "homogeneity". In this context, the pattern is considered "homogeneous" if it is generated by the template model, called the "global model", with a constant parameter vector $\boldsymbol{\theta}$. Therefore, let us consider the null hypothesis $\mathcal{H}_0$ of a Poisson process with the template log-linear global model as in Equation (2.3). The "inhomogeneous" alternative $\mathcal{H}_1$ is a Poisson process with the intensity function as in Equation (2.9), called "local model", with $\boldsymbol{\theta}(\cdot)$ function of the spatial location. So the test of local departure from "homogeneity" refers to the following hypotheses system:

$$\begin{cases} \mathcal{H}_0: & \lambda(\mathbf{u}) = \lambda(\mathbf{u}, \boldsymbol{\theta}) = \exp(B(\mathbf{u}) + \boldsymbol{\theta}^\top \mathbf{Z}(\mathbf{u})) \\ \mathcal{H}_1: & \lambda(\mathbf{u}) = \lambda(\mathbf{u}; \boldsymbol{\theta}(\mathbf{u})) = \exp(B(\mathbf{u}) + \boldsymbol{\theta}^\top(\mathbf{u}) \mathbf{Z}(\mathbf{u})) \end{cases}$$

To assess this hypothesis, a Monte Carlo test can be carried out for a locally-fitted Poisson point process model, using either the local likelihood ratio test statistic or, as in this thesis, the local score test statistic:

$$T^2(\mathbf{u}) = U(\mathbf{u}; \dot{\boldsymbol{\theta}}(\mathbf{u})) I(\mathbf{u}; \dot{\boldsymbol{\theta}}(\mathbf{u}))^{-1} U(\mathbf{u}; \dot{\boldsymbol{\theta}}(\mathbf{u}))$$

where $\dot{\boldsymbol{\theta}}(\mathbf{u})$ is the maximizer of $\log L(\mathbf{u};\boldsymbol{\theta})$ subject to $\boldsymbol{\theta} \in \Theta_0$. These statistics require a separate computation at each spatial location $\mathbf{u}$, and their asymptotic null distribution is χ^2_d. By visual inspection of the maps of the resulting local p-values (one value for each spatial location $\mathbf{u}$), we individuate those regions of the analysed area where the hypothesis is confuted. Therefore, it can be possible to identify the regions where the local model is more appropriate than the global one, gathering information on the areas with the most inhomogeneous behaviour, that is, where the effect of the considered covariates varies. Of course, if the hypothesis is rejected in the majority of the area under study, the local model should be preferred to the global one.

2.2 Spatio-temporal point processes

We consider a spatio-temporal point process with no multiple points as a random countable subset X of $\mathbb{R}^2 \times \mathbb{R}$, where a point $(\mathbf{u}, t) \in X$ corresponds to an event at $\mathbf{u} \in \mathbb{R}^2$ occurring at time $t \in \mathbb{R}$. A typical realisation of a spatio-temporal point process X on $\mathbb{R}^2 \times \mathbb{R}$ is a finite set $\{(\mathbf{u}_i, t_i)\}_{i=1}^n$ of distinct points within a bounded spatio-temporal region $W \times T \subset \mathbb{R}^2 \times \mathbb{R}$, with area $|W| > 0$ and length $|T| > 0$, where $n \geq 0$ is not fixed in advance. In this context, $N(A \times B)$ denotes the number of points of a set $(A \times B) \cap X$, where $A \subseteq W$ and $B \subseteq T$. As usual (Daley and Vere-Jones, 2007), when $N(W \times T) < \infty$ with probability 1, which holds e.g. if X is defined on a bounded set, we call X a finite spatio-temporal point process.

For a given event $(\mathbf{u}, t)$, the events that are close to $(\mathbf{u}, t)$ in both space and time, for each spatial distance r and time lag h, are given by the corresponding spatio-temporal cylindrical neighbourhood of the event $(\mathbf{u}, t)$, which can be expressed by the Cartesian product as

$$b((\mathbf{u}, t), r, h) = \{(\mathbf{v}, s) : ||\mathbf{u} - \mathbf{v}|| \leq r, |t - s| \leq h\}, \qquad (\mathbf{u}, t), (\mathbf{v}, s) \in W \times T,$$

where $||\cdot||$ denotes the Euclidean distance in $\mathbb{R}^2$. Note that $b((\mathbf{u}, t), r, h)$ is a cylinder with centre $(\mathbf{u}, \mathrm{t})$, radius r, and height $2h$.

Product densities $\lambda^{(k)}, k \in \mathbb{N}$ and $k \geq 1$, arguably the main tools in the statistical analysis of point processes, may be defined through the so-called Campbell Theorem (see Daley and Vere-Jones (2007)), which states that given a spatio-temporal point process X, and for any non-negative function f on $(\mathbb{R}^2 \times \mathbb{R})^k$, we have

$$\mathbb{E}\left[\sum_{\zeta_1, \ldots, \zeta_k \in X}^{\neq} f(\zeta_1, \ldots, \zeta_k)\right] = \int_{\mathbb{R}^2 \times \mathbb{R}} \cdots \int_{\mathbb{R}^2 \times \mathbb{R}} f(\zeta_1, \ldots, \zeta_k) \lambda^{(k)}(\zeta_1, \ldots, \zeta_k) \prod_{i=1}^{k} \mathrm{d}\zeta_i,$$

that constitutes an essential result in spatio-temporal point process theory. In particular, for $k = 1$ and $k = 2$, these functions are respectively called the *(first-order) intensity function* λ and the *(second-order) product density* $\lambda^{(2)}$. Broadly speaking, the intensity function describes the rate at which the events occur in the given spatio-temporal region, while the second-order product densities are used when the interest is in describing spatio-temporal variability and correlations between pairs of points of a pattern. They represent the point process analogues of the mean function and the covariance function of a real-valued process, respectively. Then, the intensity function is defined as

$$\lambda(\mathbf{u}, t) = \lim_{|\mathrm{d}\mathbf{u} \times \mathrm{d}t| \to 0} \frac{\mathbb{E}[N(\mathrm{d}\mathbf{u} \times \mathrm{d}t)]}{|\mathrm{d}\mathbf{u} \times \mathrm{d}t|},$$

where $\mathrm{d}\mathbf{u} \times \mathrm{d}t$ defines a small region around the point $(\mathbf{u}, t)$ and $|\mathrm{d}\mathbf{u} \times \mathrm{d}t|$ is its volume. The second-order intensity function is

$$\lambda^{(2)}((\mathbf{u}, t), (\mathbf{v}, s)) = \lim_{|\mathrm{d}\mathbf{u} \times \mathrm{d}t|, |\mathrm{d}\mathbf{v} \times \mathrm{d}s| \to 0} \frac{\mathbb{E}[N(\mathrm{d}\mathbf{u} \times \mathrm{d}t) N(\mathrm{d}\mathbf{v} \times \mathrm{d}s)]}{|\mathrm{d}\mathbf{u} \times \mathrm{d}t||\mathrm{d}\mathbf{v} \times \mathrm{d}s|}.$$

Finally, the pair correlation function

$$g((\mathbf{u},t),(\mathbf{v},s)) = \frac{\lambda^{(2)}((\mathbf{u},t),(\mathbf{v},s))}{\lambda(\mathbf{u},t)\lambda(\mathbf{v},s)}$$

can be interpreted formally as the standardised probability density that an event occurs in each of two small volumes, $\mathrm{d}\mathbf{u} \times \mathrm{d}t$ and $\mathrm{d}\mathbf{v} \times \mathrm{d}s$, in the sense that for a Poisson process, $g((\mathbf{u},t),(\mathbf{v},s)) = 1$.

2.2.1 Second-order characteristics of spatio-temporal point processes

In this thesis, the focus is on second-order characteristics of spatio-temporal point patterns, with an emphasis on the K-function (Ripley, 1976). This is a measure of the distribution of the inter-point distances and captures the spatio-temporal dependence of a point process. Gabriel and Diggle (2009) extend the second-order methods provided by Baddeley et al. (2000) to the spatio-temporal setting, defining the spatio-temporal inhomogeneous K-function and proposing a non-parametric estimator.

Definition 2.5. *(Gabriel and Diggle, 2009) A spatio-temporal point process is second-order intensity reweighted stationary and isotropic if its intensity function is bounded away from zero and its pair correlation function depends only on the spatio-temporal difference vector (r,h), where $r = ||\boldsymbol{u} - \boldsymbol{v}||$ and $h = |t-s|$.*

Definition 2.6. *(Gabriel and Diggle, 2009) For a second-order intensity reweighted stationary, isotropic spatio-temporal point process, the space-time inhomogeneous K-function takes the form*

$$K(r,h) = 2\pi \int_0^r \int_0^h g(r',h')r'\,dr'\,dh' \tag{2.19}$$

where $g(r,h) = \lambda^{(2)}(r,h)/(\lambda(\boldsymbol{u},t)\lambda(\boldsymbol{v},s)), r = ||\boldsymbol{u} - \boldsymbol{v}||, h = |t-s|$.

The simplest expression of an estimator of the spatio-temporal K-function is given as

$$\hat{K}(r,h) = \frac{1}{|W||T|}\sum_{i=1}^{n}\sum_{j>i}\mathbf{1}(||\mathbf{u}_i - \mathbf{u}_j|| \leq r, |t_i - t_j| \leq h). \tag{2.20}$$

For a homogeneous Poisson process $\mathbb{E}[\hat{K}(r,h)] = \pi r^2 h$, regardless of the intensity λ.

Both the K-function and the pair correlation function can be used as a measure of spatio-temporal clustering and interaction (Gabriel and Diggle, 2009; Møller and Ghorbani, 2012). Usually, $\hat{K}(r,h)$ is compared with the theoretical $\mathbb{E}[\hat{K}(r,h)] = \pi r^2 h$. Values $\hat{K}(r,h) > \pi r^2 h$ suggest clustering, while $\hat{K}(r,h) < \pi r^2 h$ points to a regular pattern.

Inhomogeneous second-order statistics can be constructed and used for assessing the goodness-of-fit of fitted first-order intensities. Nevertheless, it is a widespread practice in the statistical analysis of spatial and spatio-temporal point pattern data to focus primarily on comparing the data with a homogeneous Poisson process, which is generally the null model in applications rather than the fitted model. Indeed, when dealing with diagnostics in point processes, often two steps are needed: the transformation of data into residuals (thinning or rescaling (Schoenberg, 2003)) and the use of tests to assess the consistency of the residuals with the homogeneous Poisson process (Adelfio and Schoenberg, 2009). Usually, second-order statistics estimated for the residual process (i.e. the result of a thinning or rescaling procedure) are analysed. Essentially, to each observed point, a weight inversely proportional to the conditional intensity at that point is given. This method was adopted by Veen and Schoenberg (2006) in constructing a weighted version of the K-function of Ripley and Kelly (1977). The resulting weighted statistic is in many cases more powerful than residual methods (Veen and Schoenberg, 2006).

The spatio-temporal inhomogeneous version of the K-function in (2.20) is given by Gabriel and Diggle (2009) as

$$\hat{K}_I(r,h) = \frac{|W||T|}{n(n-1)} \sum_{i=1}^{n} \sum_{j>i} \frac{\mathbf{1}(||\mathbf{u}_i - \mathbf{u}_j|| \leq r, |t_i - t_j| \leq h)}{\hat{\lambda}(\mathbf{u}_i, t_i)\hat{\lambda}(\mathbf{u}_j, t_j)}, \tag{2.21}$$

where $\lambda(\cdot,\cdot)$ is the first-order intensity at an arbitrary point.
We know that $\mathbb{E}[\hat{K}_I(r,h)] = \pi r^2 h$, which is the same as the expectation of $\hat{K}(r,h)$ in (2.20), when the intensity used for the weighting is the true generator model. This is a crucial result that allows using the weighted estimator $\hat{K}_I(r,h)$ as a diagnostic tool for assessing the goodness-of-fit of spatio-temporal point processes with generic first-order intensity functions. Indeed, if the weighting intensity function is close to the true one $\lambda(\mathbf{u},t)$, the expectation of $\hat{K}_I(r,h)$ should be close to $\mathbb{E}[\hat{K}(r,h)] = \pi r^2 h$ for the Poisson process. For instance, values $\hat{K}_I(r,h)$ greater than $\pi r^2 h$ indicate that the fitted model is not appropriate since the distances computed among points exceed the Poisson theoretical ones.

2.2.2 Local second-order characteristics of spatio-temporal point processes

In the Euclidean context, both spatial and spatio-temporal second-order measures, such as the K-function and the pair correlation function (Ripley, 1977; Illian et al., 2008; Gabriel and Diggle, 2009), have been proposed to detect deviations from the Poisson assumption. While the use of global spatio-temporal second-order summary statistics is a well-established practice to describe global interaction structures between points in a point pattern, the use of local tools was first advocated by Siino et al. (2018b). They introduce Local Indicators of Spatio-temporal Association functions as an extension of the purely spatial Local Indicators, whose definition was introduced by Anselin (1995).

Definition 2.7. ***Local Indicators of Spatio-Temporal Association***
The Local Indicators of Spatio-Temporal Association (LISTA) functions are a set of functions that are individually associated with each of the points of the point pattern and can provide information about the local behaviour of the pattern.

We note that several second-order characteristics of point processes are considered for building LISTA functions, such as the K-function, the pair correlation function, and the product density function Baddeley et al. (2017). Getis and Franklin (1987) use a local version of the Ripley K-function estimator, i.e., an estimator of the individual K-function at each point of Ψ_G, to show that points may exhibit different behaviours when examined at different scales of analysis. Cressie and Collins (2001) develop second-order product density LISA functions to examine the behaviour of the individual points in a point pattern in terms of their relation to the neighbouring points at several scales simultaneously. In Siino et al. (2018b), the authors extend local indicators of spatial association to the spatio-temporal context (LISTA functions) based on the product density. These local functions have been used to define a proper statistical test for clustering detection. Then, Adelfio et al. (2020) introduce local versions of the homogeneous and inhomogeneous spatio-temporal K-functions and use them as diagnostic tools while retaining local information.

Successively, Adelfio et al. (2020) introduced local versions of both the homogeneous and inhomogeneous spatio-temporal K-functions on the Euclidean space and used them as diagnostic tools while also retaining local information.

Defining an estimator of the overall intensity by $\hat{\lambda} = n/(|W||T|)$, they propose the local version of (2.20) for the i-th event $(\mathbf{u}_i, t_i)$ as

$$\hat{K}^i(r,h) = \frac{1}{\hat{\lambda}^2|W||T|} \sum_{(\mathbf{u}_i,t_i)\neq(\mathbf{v},s)} \mathbf{1}(||\mathbf{u}_i - \mathbf{v}|| \leq r, |t_i - s| \leq h) \tag{2.22}$$

and the local version of (2.21) as

$$\hat{K}_I^i(r,h) = \frac{1}{|W||T|} \sum_{(\mathbf{u}_i,t_i)\neq(\mathbf{v},s)} \frac{\mathbf{1}(||\mathbf{u}_i - \mathbf{v}|| \leq r, |t_i - s| \leq h)}{\hat{\lambda}(\mathbf{u}_i, t_i)\hat{\lambda}(\mathbf{v}, s)}, \tag{2.23}$$

with $(\mathbf{v}, s)$ being the spatial and temporal coordinates of any other point. The authors extended the spatial weighting approach of Veen and Schoenberg (2006) to spatio-temporal local second-order statistics, proving that the inhomogeneous second-order statistics behave as the corresponding homogeneous ones, basically proving that the expectation of both (2.22) and (2.23) is equal to $\pi r^2 h$.

2.2.3 Spatio-temporal point process models

As introduced in section 1.1, point process models may be broadly categorized as mechanistic or empirical. Here we only review the most relevant spatio-temporal point process models for the purposes of this thesis work, i.e. the log-Gaussian Cox processes and the Self-Exciting point processes, which include both the Hawkes and the ETAS models. For a complete review of statistical models for the analysis of spatio-temporal point processes, we refer to González et al. (2016).

2.2.3.1 Spatio-temporal log-Gaussian Cox processes

In the Euclidean context, LGCPs are one of the most prominent clustering models. By specifying the intensity of the process and the moments of the underlying GRF, it is possible to estimate both the first and second-order characteristics of the process. Following the inhomogeneous specification in Diggle et al. (2013), an LGCP for a generic point in space and time has the intensity

$$\Lambda(\mathbf{u},t) = \lambda(\mathbf{u},t)\exp(S(\mathbf{u},t))$$

where S is a Gaussian process with $\mathbb{E}(S(\mathbf{u},t)) = \mu = -0.5\sigma^2$ and so $\mathbb{E}(\exp S(\mathbf{u},t)) = 1$ and with variance and covariance matrix $\mathbb{C}(S(\mathbf{u}_i,t_i),S(\mathbf{u}_j,t_j)) = \sigma^2\gamma(r,h)$ under the stationary assumption, with $\gamma(\cdot)$ the correlation function of the GRF, and r and h some spatial and temporal distances. Following Møller et al. (1998), the first-order product density and the pair correlation function of an LGCP are $\mathbb{E}(\Lambda(\mathbf{u},t)) = \lambda(\mathbf{u},t)$ and $g(r,h) = \exp(\sigma^2\gamma(r,h))$, respectively. In this thesis, we first consider a separable structure for the covariance function of the GRF (Brix and Diggle, 2001) that has exponential form for both the spatial and the temporal components,

$$\mathbb{C}(r,h) = \sigma^2 \exp\left(\frac{-r}{\alpha}\right)\exp\left(\frac{-h}{\beta}\right), \tag{2.24}$$

where σ^2 is the variance, α is the scale parameter for the spatial distance and β is the scale parameter for the temporal one. The exponential form is widely used in this context and nicely reflects the decaying correlation structure with distance or time. Moreover, we may consider a non-separable covariance of the GRF useful to describe more general situations. Gneiting et al. (2006) review parametric non-separable space-time covariance functions for geostatistical models. Following the parametrisation in Schlather et al. (2015), the Gneiting covariance function can be written as

$$\mathbb{C}(r,h) = (\psi(h)+1)^{-d/2}\varphi\left(\frac{r}{\sqrt{\psi(h)+1}}\right) \qquad r \geq 0, \quad h \geq 0,$$

where $\varphi(\cdot)$ is a complete monotone function associated to the spatial structure, and $\psi(\cdot)$ is a positive function with a completely monotone derivative associated with the temporal structure of the data. For example, the choice $d = 2$, $\varphi(r) = \sigma^2 \exp(-(\frac{r}{\alpha})^{\gamma_s})$ and $\psi(h) = ((\frac{h}{\beta})^{\gamma_t} + 1)^{\delta/\gamma_t}$ yields to the parametric family

$$\mathbb{C}(r, h) = \frac{\sigma^2}{((\frac{h}{\beta})^{\gamma_t} + 1)^{\delta/\gamma_t}} \exp\left(- \frac{(\frac{r}{\alpha})^{\gamma_s}}{((\frac{h}{\beta})^{\gamma_t} + 1)^{\delta/(2\gamma_t)}} \right), \tag{2.25}$$

where $\alpha > 0$ and $\beta > 0$ are scale parameters of space and time, δ takes values in $(0, 2]$, and σ^2 is the variance. Another parametric covariance belongs to the Iaco-Cesare family (De Cesare et al., 2002; De Iaco et al., 2002), and there is a wealth of covariance families that could well be used for our purposes.

As stated by Gabriel and Diggle (2009), a single realisation of a stationary Cox process with stochastic intensity $\Lambda(\mathbf{u}, t)$ is indistinguishable from a realisation of an inhomogeneous Poisson process whose first-order intensity function coincides with the unobserved realisation of $\Lambda(\mathbf{u}, t)$. To resolve this ambiguity, one can make the working assumption that first-order effects are separable, meaning that $\lambda(\mathbf{u}, t)$ can be factorised as $\lambda(\mathbf{u}, t) = \lambda(\mathbf{u})\lambda(t)$. Under this assumption, any non-separable effects are interpreted as second-order rather than first-order. Suitable estimates of $\lambda(\mathbf{u})$ and $\lambda(t)$ depend on the characteristics of the specific application.

This formulation further allows interpreting separately the temporal and spatial components peculiar to the description of our data. Indeed, as stated throughout the thesis, we are able to identify cycles in the occurrences in time, marginally with respect to the space component. Moreover, we can identify the main determinants of the spatial displacement of points, as well as assess the significance of both fixed and random effects, regardless of the temporal component.

2.2.3.1.1 Spatio-temporal LGCP model estimation through minimum contrast Driven by a GRF, controlled in turn by a specified covariance structure, the implementation of the LGCP framework in practice requires a proper estimate of the intensity function. In general, the Cox model is estimated by a two-step procedure, involving first the intensity and then the cluster or correlation parameters. First, a Poisson process with a particular model for the log intensity is fitted to the point pattern data, providing the estimates of the coefficients of all the terms that characterise the intensity. Then, the estimated intensity is taken as the true one, and the cluster or correlation parameters are estimated using either the *method of minimum contrast* (Pfanzagl, 1969; Eguchi, 1983; Diggle, 1979; Diggle and Gratton, 1984; Møller et al., 1998; Davies and Hazelton, 2013; Siino et al., 2018a), *Palm likelihood* (Ogata and Katsura, 1991; Tanaka et al., 2008), or *composite likelihood* (Guan, 2006). The most common technique is the *minimum contrast*, and it is the method which we shall refer to here.

Let the function J represent either the pair correlation function g or the K-function, and $\hat{J}$ stands for the corresponding non-parametric estimate. The minimum contrast estimates $\hat{\sigma}^2$ and $\hat{\alpha}$ are found minimising

$$M_J\{\sigma^2,\alpha\} = \int_{r_0}^{r_{max}} \phi(r)\{\nu[\hat{J}(r)] - \nu[J(r;\sigma^2,\alpha)]\}^2 \mathrm{d}r \tag{2.26}$$

where r_0 and r_{max} are the lower and upper lag limits of the contrast criterion, ϕ denotes some scalar weight associated with each spatial lag r, ν represents some transformation of its argument and the approximation of M_J is obtained by summing over a fine sequence of lags $R = \{r_0, r_1, \ldots, r_{max}\}$ equally spaced, so that $R_{diff} = r_b - r_a, b > a$

$$M_J\{\sigma^2,\alpha\} \approx R_{diff}^{-1} \sum_{r\in R} \phi(r)\{\nu[\hat{J}(r)] - \nu[J(r;\sigma^2,\alpha)]\}^2. \tag{2.27}$$

Estimation of β is performed minimising the squared discrepancy between the covariance function between the expected frequency of observations at two points in time $C(t,t-h;\beta)$ and its natural estimator, the empirical autocovariance function $\hat{C}(t,t-h)$, for a finite sequence of temporal lags. The contrast criterion for the temporal part is

$$M_C\{\beta\} = \sum_{h=1}^{h_{max}} \sum_{t=h+1}^{T} [\hat{C}(t,t-h) - C(t,t-h;\beta)]^2, \tag{2.28}$$

for some user-specified value of $h_{max} < T$, that it is the temporal counterpart of r_{max}. The theoretical version of the temporal covariance depends upon the spatial parameters. Thus the spatial parameters must be estimated first and then plugged into the temporal minimum contrast procedure.

Universally preferable options for r_0, r_{max}, $w(\cdot)$ and $v[\cdot]$ do not exist. In this thesis, r_0 is chosen as the smallest interpoint distance in the observed data, r_{max} is set to $min(x_W, y_W)/4$, as a good rule of thumb supported by Diggle (2013), where x_W and y_W represent the maximum width and height of the observation region W, respectively. We choose to consider W instead of L because, as already stated, we are referring to the most classical LGCPs defined in the Euclidean context. Indeed, the choice of the upper limit r_{max} can be considered less important given that we are generally not interested in the second-order properties at large distances. Furthermore, Diggle (2013) advocates setting $v[\cdot] = \sqrt[4]{\cdot}$ and $w(\cdot) = 1$ when using the K-function and $v[\cdot] = \log[\cdot]$ for the inhomogeneous pair correlation function. The arbitrariness of the minimum contrast procedure can be criticised, however, the relative computational simplicity with respect to other estimation procedures (such as likelihood or Bayesian estimation procedures) makes this method suitable for estimating Cox process parameters (Siino et al., 2018a). In this thesis, we have chosen $J(r)$ as the pair correlation function (Baddeley et al., 2015).

Alternatively, Siino et al. (2018a) proposed a new fitting method to estimate the set of second-order parameters for the class of LGCPs with constant first-order intensity function. Hereafter we will denote by $\boldsymbol{\theta}$ the vector of (first-order) intensity parameters and by $\boldsymbol{\psi}$ the cluster parameters, also denoted as correlation or interaction parameters by some authors. For instance, in the case of a spatio-temporal LGCP with exponential covariance, as the one in Equation (2.24), the cluster parameters correspond to $\boldsymbol{\psi} = (\sigma, \alpha, \beta)$. The second-order parameters $\boldsymbol{\psi}$ are found by minimising

$$M_J\{\boldsymbol{\psi}\} = \int_{h_0}^{h_{max}} \int_{r_0}^{r_{max}} \phi(r,h)\{\nu[\hat{J}(r,h)] - \nu[J(r,h;\boldsymbol{\psi})]\}^2 \mathrm{d}r\mathrm{d}h, \tag{2.29}$$

where $\phi(r,h)$ is a weight that depends on the space-time distance and ν is a transformation function. With simulations, Siino et al. (2018a) show that the *joint minimum contrast procedure*, based on the spatio-temporal pair correlation function, provides reliable estimates. Its main advantage is that it can be used in the case of both separable and non-separable parametric specifications of the correlation function of the underlying GRF. Therefore, it represents a more flexible method with respect to other currently available methods, and it is the method that we chose to extend into the local context.

2.2.3.1.2 Diagnostics for the spatio-temporal LGCPs For diagnostics of the proposed LGCP models and the corresponding estimation procedure, a test for spatio-temporal clustering can be employed as in Tamayo-Uria et al. (2014); Siino et al. (2018a) and section 4.1 of this thesis. Q realisations from spatio-temporal LGCPs are computed as follows:

1. Generate a realisation from a GRF $S(\mathbf{u},t)$, with covariance function $\mathbb{C}((\mathbf{u},t),(\mathbf{v},s))$ and mean function $\mu(\mathbf{u},t)$;
2. Define the generating intensity function $\lambda_0(\mathbf{u},t) = \hat{\lambda}(\mathbf{u},t)\exp(S(\mathbf{u},t))$;
3. Set an upper bound λ_{max} for $\lambda_0(\mathbf{u},t)$;
4. Simulate a homogeneous Poisson process $\mathbf{x}$ with intensity λ_{max} and denote by N the number of generated points, with coordinates $(\mathbf{u}',t')$;
5. Compute $p(\mathbf{u}',t') = \frac{\lambda(\mathbf{u}',t')}{\lambda_{max}}$ for each point $(\mathbf{u}',t')$ of a homogeneous Poisson process;
6. Generate a sample $\mathbf{p}$ of size N from the uniform distribution on $(0,1)$;
7. Thin the simulated homogeneous Poisson process $\mathbf{x}$ retaining the $n \leq N$ locations for which $\mathbf{p} \leq p(\mathbf{u}',t')$.

The GRF is generated using the function `RFsim` of the R package `CompRandFld` (Padoan and Bevilacqua, 2015). Having simulated $Q > 1$ realisations of spatio-temporal point patterns $\mathbf{x}_1, \dots, \mathbf{x}_Q$, these simulated processes can be used for computing Q inhomogeneous spatio-temporal K-functions as given by Equation (2.21). Given the Q inhomogeneous K-functions, their corresponding mean and variance, denoted by E_K and V_K, respectively, are computed. The overall test statistic is

$$T_q = \int_{r_0}^{r_{max}} \int_{h_0}^{h_{max}} \frac{\hat{K}_q(r,h) - E_K(r,h)}{\sqrt{V_K(r,h)}} \mathrm{d}h\mathrm{d}r,$$

one for each $\hat{K}_q(r,h)$, obtaining $T_1, \dots, T_Q$. r_{max} and h_{max} are the maximum spatial and temporal distances considered for the inhomogeneous K-functions. Then, the same test statistic is also computed for the empirical point pattern and denoted by T^*. The p-value is then defined as

$$\frac{1 + \sum_{q=1}^{Q} \mathbf{1}(T_q > T^*)}{Q+1}, \tag{2.30}$$

basically counting how many times the K-functions computed over the simulated processes are higher than the one computed on the original point process. Then, if the obtained p-value is smaller than a significance level α there is evidence against the null hypothesis. That is, the analysed spatio-temporal point process still presents globally some clustering behaviour.

The resulting inhomogeneous K-functions can also be used to obtain upper and lower envelopes at a chosen significance level α, and so to visually assess the possible residual clustered structure of the analysed point pattern, unexplained by the proposed model. In particular, if the estimated K-function lies above the obtained envelopes, then there is still a clustering behaviour of points that is not completely described by the proposed model. Furthermore, by a visual assessment of the results, we can also get indications of possible ranges needing more complex models, able to explain and take into account the residual spatio-temporal dependence of the data.

2.2.3.2 Self-Exiting spatio-temporal point processes

Point processes can be formally specified in several ways, for instance, by considering the joint distribution of the counts of points in arbitrary sets or by defining a conditional intensity function. Let N be a point process $\{(t_i, x_i, y_i) : i = 1, \dots, n\}$ on a spatio-temporal domain $X = W \times T \subseteq \mathbb{R}^2 \times \mathbb{R}_+$, with area $|W| > 0$ and length $|T| > 0$, and with t representing the time, and x and y the two spatial coordinates. Its conditional intensity function is defined by

$$\lambda(t,x,y|\mathcal{H}_t) = \lim_{\Delta t, \Delta x, \Delta y \to 0^+} \frac{\mathbb{E}[N((t,t+\Delta t) \times (x, x+\Delta x) \times (y, y+\Delta y)|\mathcal{H}_t)]}{|\Delta t \Delta x \Delta y|} \tag{2.31}$$

where $\mathcal{H}_t$ is the space-time occurrence history of the process up to time t, i.e. the σ-algebra of events occurring at times up to but not including t. $\Delta t, \Delta x, \Delta y$ are time and space increments, respectively, and $\mathbb{E}[N((t, t+\Delta t) \times (x, x+\Delta x) \times (y, y+\Delta y)|\mathcal{H}_t)]$ is the history-dependent expected value of occurrence in the volume $\{(t, t+\Delta t) \times (x, x+\Delta x) \times (y, y+\Delta y)\}$. Assuming such a limit exists for each point (t, x, y) in the space-time domain and that the point process N is simple, the conditional intensity process uniquely characterises the finite-dimensional distributions of the point process (Daley and Vere-Jones, 2007).

To model events that are clustered, the Self-Exciting point processes are often used. These models are largely used to describe earthquakes characteristics, assuming that the occurrence of an event increases the probability of occurrence of others events in time and space. Examples of Self-Exciting point processes include Hawkes models (Hawkes, 1971a,b; Hawkes and Adamopoulos, 1973) and ETAS models (Ogata and Katsura, 1988; Adelfio and Chiodi, 2015b). The conditional intensity function of a linear Self-Exciting process is defined by

$$\lambda(t, x, y|\mathcal{H}_t) = \mu(t, x, y) + \int_{-\infty}^{t} \int_{W} g(t-t^{'}, x-x^{'}, y-y^{'}) N(\mathrm{d}x^{'} \times \mathrm{d}y^{'} \times \mathrm{d}t^{'}), \quad (2.32)$$

being the sum of two non-negative functions: $\mu(t, x, y) > 0$, that describes the large-scale variation of $\lambda(t, x, y|\mathcal{H}_t)$, and $g()$, such that $\iint g(v)dv < 1$, which describes its small-scale variation due to the interaction with the events in the past. This process can be interpreted as a generalised Poisson cluster process associating to centres, of rate μ, a branching process of descendants. The spatio-temporal Hawkes process has a conditional intensity of the form (Hawkes, 1971a,b)

$$\lambda(t, x, y) = \mu(t, x, y) + \sum_{i:t_i<t} g(t-t_i, x-x_i, y-y_i), \quad (2.33)$$

where $\mu(t, x, y)$ is the background rate, and $g(t, x, y)$ is the rate of occurrence triggered by an event at time 0 and location at the origin. The triggering density governs the spatial-temporal distance of triggered events from their antecedent events and is usually modelled to decay with distance from the origin over time and space (Park et al., 2021). In Mohler et al. (2011), the background rate $\mu(t, x, y)$ was assumed to be a function of space and time and they used kernel functions to smooth the estimates of both μ and g, introducing the idea of the stochastic reconstruction algorithm (Zhuang et al., 2004; Zhuang, 2006; Marsan and Lengline, 2008) for the analysis of crime data, by means of relaxing parameters and periodic components in the background rates.

2.2.3.2.1 Spatio-temporal Hawkes model estimation through the Estimation-Maximization algorithm We consider the Self-Exciting model in Equation (2.32), following the semi-parametric specification proposed by Zhuang and Mateu (2019) for a spatio-temporal Hawkes process. Therefore, the full model that we consider here is specified as follows

$$\lambda(t,x,y) = \mu_0\mu_t(t)\mu_w(t)\mu_b(x,y) + A\int_{-\infty}^{t_-}\int_W g(t-s)h(x-u,y-v)N(\mathrm{d}u\times\mathrm{d}v\times\mathrm{d}s), \tag{2.34}$$

where $\mu_t(t)$ and $\mu_w(t)$ represent the trend term and the weekly periodicity in the temporal components of the background rate, $\mu_b(x,y)$ represents the spatial background rate, and $g(t-s)h(x-u,y-v)$ represents the sub-process triggered by an event previously occurring at time s and location (u,v). Note that this model enables the background rate to include a spatial background pattern that can be separated from the periodicity effects and the long-term temporal trend. We estimate the two relaxation coefficients A and μ_0, normalise to 1 the average values of $\mu_t(t)$, $\mu_w(t)$ and $\mu_b(x,y)$, and define the probability density functions g and h such that $\int_0^\infty g(s)\mathrm{d}s = 1$, and $\int\int_X h(u,v)\mathrm{d}u\mathrm{d}v = 1$.

In the following, we outline the modified Estimation-Maximization (E-M) algorithm employed by Zhuang and Mateu (2019), as we do not take into account the daily temporal resolution, not available for the data presented in section 4.2.1 and further analysed in section 4.2.3.

The E-M estimation algorithm is as follows:

Step 1. Set initial values for $\{\mu_0,\mu_t,\mu_w,\mu_b,A,g,h\}$

Step 2. Compute background and excitation components by reconstructing for each i-th event $(i=1,\dots,n)$

$$\begin{aligned}
w_i^{(t)} &= \mu_t(t_i)\mu_b(x_i,y_i)/\lambda(t_i,x_i,y_i),\\
w_i^{(w)} &= \mu_w(t_i)\mu_b(x_i,y_i)/\lambda(t_i,x_i,y_i),\\
\varphi_i &= \mu_0\mu_t(t)\mu_w(t)\mu_b(x,y)/\lambda(t_i,x_i,y_i),\\
\rho_{i,j} &= g(t_j-t_i)h(x_j-x_i,y_j-y_i)/\lambda(t_i,x_i,y_i),
\end{aligned}$$

and by estimating the background terms

$$\begin{aligned}
\hat{\mu}_t(t) &\propto \sum_i w_i^{(t)} K(t-t_i;\omega_t)\Big/\int_0^T K(u-t_i;\omega_t)du,\\
\hat{\mu}_w(t) &\propto \sum_i w_i^{(w)} \sum_{k=0}^{\lfloor T/7\rfloor} K(t-t_i+7\lfloor t_i/7\rfloor-7k;\omega_w)\Big/\int_0^T K(u-t_i;\omega_t)du,\\
\hat{\mu}_b(x,y) &\propto \sum_i \varphi_i \frac{K(x-x_j+x_i;\omega_x)K(y-y_j+y_i;\omega_y)}{\int\int_X K(x-x_j+x_i;\omega_x)K(y-y_j+y_i;\omega_y)\mathrm{d}u\mathrm{d}v},
\end{aligned}$$

where $K(x;\omega) = \frac{1}{\sqrt{(2\pi)\omega}}\exp\left(-\frac{x^2}{2\omega^2}\right)$ is the Gaussian Kernel, and $\lfloor x\rfloor$ is the largest integer not bigger than x.

The triggering terms are estimated as follows

$$\hat{g}(t) \propto \frac{\sum_{i,j} \rho_{i,j} \times K(t - t_j + t_i; \omega_g) / \int_0^{T-t_i} K(u - t_j; \omega_g) \mathrm{d}u}{\sum_i \mathbf{1}(t_i + t \leq T)},$$

$$\hat{h}(x, y) \propto \frac{\frac{\sum_{i,j} \rho_{i,j} \times K(x - x_j + x_i; \omega_{h_x}) K(y - y_j + y_i; \omega_{h_y})}{\int\int_S K(u - x_j + x_i; \omega_{h_x}) K(v - y_j + y_i; \omega_{h_y}) \mathrm{d}u \mathrm{d}v}}{\sum_i I\{(x_i + x, y_i + y) \in S\}},$$

Step 3. Estimate the relaxation parameters

$$A^{(k+1)} = \frac{n - \sum_{i=1}^{n} \varphi_i^{(k)}}{G},$$

$$\mu_0^{(k+1)} = \frac{n - \sum_{i=1}^{n} A^{(k)} G}{U},$$

where

$$\varphi_i^{(k)} = \frac{\mu_0^{(k)} \mu_t(t_i) \mu_w(t_i) \mu_b(x_i, y_i)}{\mu_0^{(k)} \mu_t(t_i) \mu_w(t_i) \mu_b(x_i, y_i) + A^{(k)} \sum_{j: t_j < t_i} g(t_j - t_i) h(x_j - x_i, y_j - y_i)}.$$

and G and U are the arguments that maximise the likelihood function with respect to μ_0 and A, respectively.

Indeed, since the likelihood function takes the form

$$\log L = \sum_{i=1}^{n} \log \lambda(t_i, x_i, y_i) - \int_0^T \iint_X \lambda(t, x, y) \, \mathrm{d}x \mathrm{d}y \, \mathrm{d}t$$

where

$$\lambda(t, x, y) = \mu_0 \mu(t, x, y) + A \sum_{i:\, t_i < t} g(t_i - t, x_i - x, y_i - y),$$

and denoting $U = \int_0^T \iint_X \mu(t, x, y) \, \mathrm{d}x \mathrm{d}y \, \mathrm{d}t$ and $G = \int_0^T \iint_X \sum_{i:\, t_i < t} g(t - t_i, x - x_i, y - y_i) \, \mathrm{d}x \mathrm{d}y \, \mathrm{d}t$, then the equations $\frac{\partial}{\partial \mu_0} \log L = 0$ and $\frac{\partial}{\partial A} \log L = 0$ can be written as

$$\begin{aligned} \sum_{i=1}^{n} \frac{\mu(t_i, x_i, y_i)}{\lambda(t_i, x_i, y_i)} - U &= 0 \\ \sum_{i=1}^{n} \frac{\sum_{j: t_j < t_i} g(t_j - t_i, x_j - x_i, y_j - y_i)}{\lambda(t_i, x_i, y_i)} - G &= 0. \end{aligned}$$

Step 4. If convergence is reached, we obtain the estimates of μ_0 and A. Otherwise, we go back to Step 2.

2.2.3.2.2 Spatio-temporal Epidemic Type Aftershock Sequence point process According to a branching structure, the conditional intensity function of the Self-Exciting model is defined as the sum of a term describing the large-time scale variation (spontaneous activity or background, generally assumed homogeneous in time but not in space) and one relative to the small-time scale variation due to the interaction with the events in the past (induced or triggered activity). Following Adelfio and Chiodi (2020), the ETAS model can be written as

$$\lambda_\theta(t, \mathbf{u}|\mathcal{H}_t) = \mu f(\mathbf{u}) + \sum_{t_j < t} \frac{\kappa_0 \exp(\eta_j)}{(t - t_j + c)^p} \{(\mathbf{u} - \mathbf{u}_j)^2 + d\}^{-q}, \tag{2.35}$$

where $\mathcal{H}_t$ is the past history of the process up to time t, μ is the large-scale general intensity, and $f(\mathbf{u})$ is the spatial density. Concerning the triggered component, $\eta_j = \boldsymbol{\beta}'\mathbf{Z}_j$ is a linear predictor, with $\mathbf{Z}_j$ the external known covariate vector, including the magnitude, and $\boldsymbol{\theta} = (\mu, \kappa_0, c, p, d, q, \boldsymbol{\beta})$ are the parameters to be estimated. In particular: κ_0 is a normalising constant, c and p are characteristic parameters of the seismic activity of the given region, and d and q are two parameters related to the spatial influence of the mainshock. In the usual ETAS model, with the only external covariate representing the magnitude, $\boldsymbol{\beta} = \alpha$, as $\eta_j = \boldsymbol{\beta}'\mathbf{Z}_j = \alpha(m_j - m_0)$, where m_j is the magnitude of the j^{th} event and m_0 the threshold magnitude, that is, the lower bound for which earthquakes with higher values of magnitude are surely recorded in the catalogue.

2.3 Point processes on linear networks

Point processes on linear networks are recently considered to analyse events occurring on particular network structures, such as traffic accidents on a road network that we analyse in this thesis (see Figure 5.20). They were firstly introduced in the spatial context and then extended to the spatio-temporal case, focusing on the analysis of first- and second-order summary statistics (Ang et al., 2012; McSwiggan et al., 2017; Rakshit et al., 2017; Moradi et al., 2019; Rakshit et al., 2019a; Moradi and Mateu, 2020). For a general treatment of spatial and spatio-temporal point processes on the Euclidean space, the reader is referred to Diggle (2013).

A linear network $L = \cup_{i=1}^{n} l_i \subset \mathbb{R}^2$ is commonly taken as a finite union of line segments $l_i \subset \mathbb{R}^2$ of positive length. A line segment is defined as $l_i = [\mathbf{u}_i, \mathbf{v}_i] = \{k\mathbf{u}_i + (1-k)\mathbf{v}_i : 0 \leq k \leq 1\}$, where $\mathbf{u}_i, \mathbf{v}_i \in \mathbb{R}^2$ are the endpoints of l_i. For any $i \neq j$, the intersection of l_i and l_j is either empty or an endpoint of both segments.

Following (Ang et al., 2012; Baddeley et al., 2020), we consider a linear network as the union of a finite number of line segments on the plane, namely $L = \bigcup_{i=1}^{n} l_i$, such that each $l_i = [u_i, v_i] = \{w : w = tu_i + (1-t)v_i, 0 \leq t \leq 1\}$, and u_i, v_i are the endpoints of the segment l_i. The total length of the network L is denoted by $|L|$. Additionally, a path between two points u and v in a linear network L is a sequence $x_0, x_1, \ldots, x_m$ of points in L so what $x_0 = u$, $x_m = v$ and $[x_i, x_{i+1}] \subset L$ for each $i = 0, \ldots, m-1$. The total length of the path is the sum of the segments' lengths that make it up. The shortest path distance $d_L(u,v)$ between u and v in L is the minimum of the lengths of all paths from u to v. If there are no paths from u to v (implying that the network is not connected), then the distance is defined by $d_L(u,v) = \infty$.

We also assume that the disc of radius $r > 0$ and centre point $u \in L$ is the set of all points v in the network that lies no more than a distance r from the location u, that is $b_L(u,r) = \{v \in L : d_L(u,v) \leq r\}$. The relative boundary of the disc is the set of points lying exactly r units away from u, $\partial b_L(u,r) = \{v \in L : d_L(u,v) = r\}$. The circumference $m(u,r)$ is the number of points of L there is in $\partial b_L(u,r)$. The circumference is finite for all $r < \infty$, and we set $m(u,\infty) = \infty$ by convention. The circumradius of the network is the radius of the smallest disc that contains the entire network, $R = R(L) = \inf\{r : m(u,r) = 0, \text{ for some } u \in L\} = \min_{u \in L} \max_{v \in L} d_L(u,v)$. If L is not path-connected, then R is the minimum of the circumradius of the connected components of L. For more details, see (Ang et al., 2012; Baddeley et al., 2020).

In the same way as in Rakshit et al. (2021), we use the definition of a finite, simple) point pattern X on L as a finite set $X = \{x_1, x_2, \ldots, x_n\}$ of distinct points $x_i \in L$, where $n \geq 0$. Further, for any set $B \subset L$, $N_X(B) = N(X \cap B)$ is the number of points on X lying in B. Thus for a finite and simple point possess with intensity function $\lambda(u)$, $u \in L$, is satisfied that $\mathbb{E}[N_X(B)] = \Lambda(B) = \int_B \lambda(u) \mathrm{d}_1 u$, for all measurable B in L, where $\mathrm{d}_1 u$ denotes integration with respect to arc length

(Federer, 2014). If λ is constant, then X is said to be homogeneous. Otherwise, X is said to be inhomogeneous. For a homogeneous point process, λ is the mean number of points per unit length. A Poisson point process on L with intensity $\lambda > 0$ is characterized by the properties that, for any line segment $B \subset L$, the number of points falling in B has a Poisson distribution with mean $\lambda|B|$, while events occurring in disjoint line segments $B_1, \ldots, B_m \subset L$ are independent (Ang et al., 2012).

2.3.1 Spatio-temporal linear network point processes

A spatio-temporal linear network point process is a point process on the product space $L \times T$, where L is a linear network and T is a subset (interval) of $\mathbb{R}^+$. We hereafter focus on a spatio-temporal point process X on a linear network L with no overlapping points $(\mathbf{u}, t)$ where $\mathbf{u} \in L$ is the location of an event and $t \in T (T \subseteq \mathbb{R}^+)$ is the corresponding time occurrence of $\mathbf{u}$. Note that the temporal state space T might be either a continuous or a discrete set. A realisation of X with n points is represented by $\mathbf{x} = (\mathbf{u}_i, t_i), i = 1, \ldots, n$ where $(\mathbf{u}_i, t_i) \in L \times T$. A spatio-temporal disc with centre $(\mathbf{u}, t) \in L \times T$, network radius $r > 0$ and temporal radius $h > 0$ is defined as

$$b((\mathbf{u}, t), r, h) = \{(\mathbf{v}, s) \in L \times T : d_L(\mathbf{u}, \mathbf{v}) \le r \quad \text{and} \quad |t - s| \le h\},$$

where $|\cdot|$ is a numerical distance, and $d_L(\cdot, \cdot)$ stands for the appropriate distance in the network. The cardinality of any subset $A \subseteq L \times T, N(X \cap A) \in 0, 1, \ldots$, is the number of points of X restricted to A, whose expected number is denoted by

$$\nu(A) = \mathbb{E}[N(X \cap A)], \quad A \subseteq L \times T,$$

where ν, the intensity measure of X, is a locally finite product measure on $L \times T$ (Baddeley et al., 2006a).

We now recall Campbell's theorem for point processes on linear networks (Cronie et al., 2020). Assuming that the product densities/intensity functions $\lambda^{(k)}$ exist, for any non-negative measurable function $f(\cdot)$ on the product space L^k, we have

$$\mathbb{E}\left[\sum_{\zeta_1, \ldots, \zeta_k \in X}^{\neq} f(\zeta_1, \ldots, \zeta_k) \right] = \int_{L^k} f(\zeta_1, \ldots, \zeta_k) \lambda^{(k)}(\zeta_1, \ldots, \zeta_k) \prod_{i=1}^{k} \mathrm{d}\zeta_i, \tag{2.36}$$

where $\neq$ indicates that the sum is over distinct values. Assume that X has an intensity function $\lambda(\cdot, \cdot)$, hence Equation (2.36) reduces to

$$\mathbb{E}[N(X \cap A)] = \int_A \nu(d(\mathbf{u}, t)) = \int_A \lambda(\mathbf{u}, t) d_2(\mathbf{u}, t), A \subseteq L \times T,$$

where $d_2(\mathbf{u}, t)$ corresponds to integration over $L \times T$.

2.3.2 Spatio-temporal Poisson processes on linear networks

The Poisson point process model is crucial for defining other more complicated models and represents a benchmark model in exploratory data analysis. For a spatio-temporal Poisson point process X, the points $(\mathbf{u}, t)$ on $L \times T$ have to satisfy the following conditions:

- In any bounded set $A \subseteq L \times T, N(X \cap A)$ follows a Poisson distribution with expected number $\int_A \lambda(\mathbf{u}, t_u) d_2(\mathbf{u}, t_u)$. For instance, if we assume $A = A_L \times A_T \subseteq L \times T$, then the expected value is the expected number of points in the sub-network $A_L \subseteq L$ and within the time interval $A_T \subseteq T$.
- For any k arbitrary disjoint subsets of $L \times T$, say $A_1, \ldots, A_k$, their cardinality $N(X \cap A_1), \ldots, N(X \cap A_k)$ are independent variables.

Moreover, for any time interval $S = [s_1, s_2] \subset T$, the projection of X onto L defines a spatial Poisson process on L with intensity function $\int_{s_1}^{s_2} \lambda(\mathbf{u}, t) dt$. Similarly, for any set of segments $A \subset L$, the projection of X onto T defines a temporal Poisson process with intensity function $\int_A \lambda(\mathbf{u}, t) d_1 \mathbf{u}$ (see Illian et al., 2008, Chapter 6; Diggle, 2013, Chapter 10).

2.3.2.1 Inhomogeneous Poisson processes on linear networks

The description of the observed point pattern intensity is a crucial issue in dealing with spatio-temporal data. For linear network point patterns, non-parametric estimators of the intensity function $\lambda(\cdot, \cdot)$ have been proposed (Mateu et al., 2020), suggesting any variation of the distribution of the process over its state-space $L \times T$. When dealing with intensity estimation for spatio-temporal point processes, it is quite common to assume that the intensity function $\lambda(\mathbf{u}, t)$ is separable (Diggle, 2013; Gabriel and Diggle, 2009). Under this assumption, the intensity function is given by the product $\lambda(\mathbf{u}, t) = \lambda(\mathbf{u})\lambda(t)$ where $\lambda(\mathbf{u})$ and $\lambda(t)$ are non-negative functions on L and T, respectively (Gabriel and Diggle, 2009). A kernel-based intensity estimator for spatio-temporal linear network point processes, based on the first-order separability assumption, considered in Moradi and Mateu (2020), is given by

$$\hat{\lambda}(\mathbf{u}, t) = \frac{\hat{\lambda}(\mathbf{u})\hat{\lambda}(t)}{n}, \quad (\mathbf{u}, t) \in L \times T \tag{2.37}$$

that is an unbiased estimator for the expected number of observed points. The spatial intensity, given the network, denoted by $\lambda(\mathbf{u})$, is usually estimated by non-parametric approaches mostly based on kernel-based estimators. For instance, spatial smoothing can be performed using a kernel based on the path distances in the network, using the ‘equal-split continuous’ rule described in Okabe and Sugihara (2012). Computation is rapidly performed by solving the classical heat equation on the network, as described in McSwiggan et al. (2017). The smoothing parameter for the kernel estimator of the temporal intensity function $\lambda(t)$ can be the usual bandwidth suggested by Silverman's rule of thumb (Silverman, 2018). In Mateu et al. (2020) a pseudo-separable estimator is proposed, accounting for spatio-temporal interactions while relaxing the separability assumption and the specification of a metric on $L \times T$. In this thesis, we introduce, for the first time in the context of spatio-temporal point processes, parametric estimates of the spatial and temporal components following a dependence model with a mixed effects approach (see section 4.1). We estimate the spatial intensity through a Gibbs point process model with mixed effects, and the temporal intensity by a Poisson harmonic model. The latter can be defined as follows

$$\log \lambda(t) = \delta + \sum_{k \in K_\kappa} \{\alpha_k \cos(k\omega t) + \beta_k \sin(k\omega t)\}$$

where $\omega = 2\pi\kappa$, with κ depending on the unit of analysis, e.g. it is 4 for quarterly data, 12 for monthly data, 52.25 for weekly data, and 365.25 for daily data, if considering a 1-year period. Moreover, K_κ is the subset of time instants we use to split our observation period. In our application, we chose $\kappa = 24$, corresponding to the hours in a day, and $K_\kappa = \{1, 2, 4, 8\}$, determining temporal cycles up to 3 hours. Finally, δ, α_k and β_k ($\forall k \in K_\kappa$) are parameters to be estimated.

While the temporal component can be fitted without any alteration arising for the accounting of the network, the spatial component is the one needing careful attention and should be changed in order to take into account the structure of the network. In the next section, we address this problem.

2.3.2.2 Intensity estimation on linear networks

A natural first step in analysing spatial patterns is to form a kernel density estimate (Silverman, 2018) of the spatially-varying events rate. However, this is not straightforward on a linear network.

Concerning intensity estimation, several kernel-based methods (Borruso, 2005, 2008; Xie and Yan, 2008; Okabe et al., 2009; Okabe and Sugihara, 2012; McSwiggan et al., 2017; Moradi et al., 2018; Rakshit et al., 2019a) and a resample-smoothing technique applied to Voronoi intensity estimators (Moradi et al., 2019; Mateu et al., 2020) have been proposed.

Obtaining good estimates for intensity functions of point processes on linear networks has been a challenging task due to geometrical complexities and unique methodological problems (Cronie et al., 2020). Nevertheless, there have been a few particularly interesting proposals. Baddeley et al. (2020) reviews the most recent approaches for spatial kernel density estimation on linear networks. We summarise here the main contributions.

Approaches based on sums of kernels (Xie and Yan, 2008) basically assume that the statistical basis for kernel estimation can be transferred from the real line to the linear network. However, several authors (Okabe et al., 2009; Okabe and Sugihara, 2012; McSwiggan et al., 2017) have highlighted that translating the kernel density estimate on the one-dimensional real line directly to the linear network is a fallacious estimate because it does not conserve mass. This basically means that the kernel $K(\cdot, x_i) = k(d_L(\cdot, x_i))$ is not a probability density on the linear network, and therefore the corresponding density estimate is not a probability density. The true probability density would be roughly overestimated in the denser parts of the network.

Then, Borruso (2005, 2008) and Okabe et al. (2009) proposed several modifications of the kernel sums estimator. Moradi et al. (2018) and McSwiggan et al. (2020) pointed out that the bias can be removed by adapting classical edge-corrections from spatial statistics, thus proposing some *edge-corrected classical kernel-based intensity estimators*.

As per *path enumeration methods*, (Okabe et al., 2009; Sugihara et al., 2010; Okabe and Sugihara, 2012) considered computational algorithms which start with a kernel k on the real line and progressively redistribute the mass of this kernel over the network. These authors proposed two kernel estimators satisfying many desired properties: (a) the "equal-split continuous" rule, which has excellent properties (symmetric, conserves mass, and is unbiased when the true intensity is uniform) but it is extremely slow to compute; (b) the "equal-split discontinuous" rule, which is faster but has less desirable properties (see Okabe and Sugihara (2012) and McSwiggan et al. (2017) for the algorithms).

Another proposal is represented by the *heat kernel*: it is the network counterpart of the Gaussian kernel. It was developed by McSwiggan et al. (2017) by exploiting the connection between kernel smoothing and diffusion. Since the heat kernel estimator is mathematically equivalent to the "equal-split continuous" estimator extended to the Gaussian kernel, it has the same properties.

Finally, a *fast kernel smoothing using two-dimensional convolutions* has been recently proposed by Rakshit et al. (2019b). It is a computationally efficient and statistically principled method for kernel smoothing of point pattern data on a linear network. The point locations, and the network itself, are convolved with a two-dimensional kernel and then combined into an intensity function on the network. This can be computed rapidly using the Fast Fourier Transform (FFT), even on large networks and for large bandwidths, and is robust against errors in network geometry. The estimator is consistent, and its statistical efficiency is only slightly suboptimal.

A crucial problem with kernel estimation is that if there are wide variations in intensity across the spatial domain, it may be impossible to find a single fixed bandwidth value which is satisfactory for smoothing every part of the spatial domain. Consequently, the bandwidth can be spatially-varying, giving rise to a spatially "adaptive" kernel estimator at the cost of increased complexity. Recently, there has been an increasing interest in point patterns on linear networks. Here the matter of kernel estimation is even more delicate due to the geometry of the underlying network. For example, street crimes and traffic accidents tend to happen particularly in busy streets, which may be surrounded by quiet neighbourhoods. In such cases, the classical kernel estimation approach is often unsuitable for such types of data, and alternatives based on *resample-smoothing of adaptive Voronoi intensity estimators* have been proposed (Moradi et al., 2019).

2.3.3 Second-order characteristics of point processes on linear networks

The second-order Campbell's theorem is obtained from (2.36) with $k = 2$

$$\begin{aligned}\mathbb{E}\Bigg[\sum_{(\mathbf{u},t),(\mathbf{v},s)\in X}^{\neq} f\big((\mathbf{u},t),(\mathbf{v},s)\big)\Bigg] &= \\ = \int_{L\times T}\int_{L\times T} f\big((\mathbf{u},t),(\mathbf{v},s)\big)\lambda^{(2)}\big((\mathbf{u},t),(\mathbf{v},s)\big)\mathrm{d}_2(\mathbf{u},t)\mathrm{d}_2(\mathbf{v},s).\end{aligned} \tag{2.38}$$

Assuming that X has a second-order product density function $\lambda^{(2)}(\cdot,\cdot)$, we then obtain

$$\mathbb{E}[N(X\cap A)N(X\cap B)] = \int_A\int_B \lambda^{(2)}((\mathbf{u},t),(\mathbf{v},s))d_2(\mathbf{u},t)d_2(\mathbf{v},s), \quad A,B\subseteq L\times T.$$

Finally, an important result concerns the conversion of the integration over $L\times T$ to that over $\mathbb{R}\times\mathbb{R}$ (Rakshit et al., 2017). For any measurable function $f: L\times T\to\mathbb{R}$

$$\int_{L\times T} f(\mathbf{u},t)\mathrm{d}_2(\mathbf{u},t) = \int_0^\infty\int_0^\infty \sum_{\substack{(\mathbf{u},t)\in L\times T:\\ d_L(\mathbf{u},\mathbf{v})=r,\\ |t-s|=h}} f(\mathbf{u},t)\mathrm{d}r\mathrm{d}h. \tag{2.39}$$

Letting $f(\mathbf{u},t) = \eta(d_L(\mathbf{u},\mathbf{v}),|t-s|)$ then

$$\int_{L\times T}\eta(d_L(\mathbf{u},\mathbf{v}),|t-s|)\mathrm{d}_2(\mathbf{u},t) = \int_0^\infty\int_0^\infty \eta(r,h)M((\mathbf{u},t),r,h)\mathrm{d}r\mathrm{d}h$$

where $M((\mathbf{u},t),r,h)$ is the number of points lying exactly at the shortest-path distance $r\geq 0$ and the time distance $h\geq 0$ away from $(\mathbf{u},t)$.

Following the work of Moradi and Mateu (2020), we first recall the concept of second-order pseudostationary point processes, and then we turn to the inhomogeneous case.

Definition 2.8. *(Moradi and Mateu, 2020) Assume X is a spatio-temporal point process on $L \times T$ with constant intensity function $\lambda \geq 0$. Let the homogeneous K-function be given by*

$$K_L((\boldsymbol{u}, t), r, h) = \frac{1}{\lambda}\mathbb{E}\left[\sum_{(\boldsymbol{u},t)\in X} \frac{\mathbf{1}(d_L(\boldsymbol{u}, \boldsymbol{v}) \leq r, |t-s| \leq h)}{M((\boldsymbol{u}, t), d_L(\boldsymbol{u}, \boldsymbol{v}), |t-s|)} \middle| (\boldsymbol{u}, t \in X) \right].$$

Then, X is called second-order pseudostationary and isotropic if $K_L((\boldsymbol{u}, t), r, h)$ does not depend on the point $(\boldsymbol{u}, t)$, and we then write $K_L((\boldsymbol{u}, t), r, h) = K_L(r, h)$.

Definition 2.9. *(Moradi and Mateu, 2020) For a homogeneous Poisson point process on $L \times T$ with constant intensity function λ, $K_L(r, h) = rh$.*

The authors proposed a non-parametric estimator of both the homogeneous K-function

$$\hat{K}_L(r, h) = \frac{|L||T|}{n(n-1)} \sum_{i=1}^{n} \sum_{i \neq j} \frac{I\{d_L(\mathbf{u}_i, \mathbf{u}_j) < r, |t_i - t_j| < h\}}{M((\mathbf{u}_i, t_i), d_L(\mathbf{u}_i, \mathbf{u}_j), |t_i - t_j|)},$$

and pair correlation function

$$\hat{g}_L(r, h) = \frac{|L||T|}{n(n-1)} \sum_{i=1}^{n} \sum_{i \neq j} \frac{\kappa(d_L(\mathbf{u}_i, \mathbf{u}_j) - r)\kappa(|t_i - t_j| - h)}{M((\mathbf{u}_i, t_i), d_L(\mathbf{u}_i, \mathbf{u}_j), |t_i - t_j|)},$$

where $|L| > 0$ and $|T| > 0$ are the total length of the network L and of the time interval T, respectively, and κ is a one-dimensional kernel function.

As in the Euclidean case, for a Poisson process on the linear network L, these second-order statistics can be used to indicate clustering and inhibition, if compared to their Poisson process values rh. Both the global homogeneous K-function and pair correlation functions are used in Moradi and Mateu (2020) for detecting departure from the homogeneity of spatio-temporal point patterns occurring on different linear networks.

Motivated by practical situations where homogeneity is not a realistic assumption, Moradi and Mateu (2020) further proposed the corresponding inhomogeneous second-order characteristics, namely the inhomogeneous K-function

$$\hat{K}_{L,I}(r, h) = \frac{1}{|L||T|} \sum_{i=1}^{n} \sum_{i \neq j} \frac{I\{d_L(\mathbf{u}_i, \mathbf{u}_j) < r, |t_i - t_j| < h\}}{\hat{\lambda}(\mathbf{u}, t)\hat{\lambda}(\mathbf{v}, s) M((\mathbf{u}, t), d_L(\mathbf{u}, \mathbf{v}), |t - s|)}, \tag{2.40}$$

and the inhomogeneous pair correlation function

$$\hat{g}_{L,I}(r, h) = \frac{1}{|L||T|} \sum_{i=1}^{n} \sum_{i \neq j} \frac{\kappa(d_L(\mathbf{u}_i, \mathbf{u}_j) - r)\kappa(|t_i - t_j| - h)}{\hat{\lambda}(\mathbf{u}_i, t_i)\hat{\lambda}(\mathbf{u}_j, t_j) M((\mathbf{u}_i, t_i), d_L(\mathbf{u}_i, \mathbf{u}_j), |t_i - t_j|)}, \tag{2.41}$$

where $\hat{\lambda}(\cdot,\cdot)$ is an estimate of the intensity function, and $M((\mathbf{u},t),r,h)$ is the number of points lying exactly at the shortest-path distance $r \geq 0$ and the time distance $h \geq 0$ away from the point $(\mathbf{u},t)$.

Moradi and Mateu (2020) recommend using some normalisation factor of the form $D(X) = \frac{1}{(|L||T|)^2}\sum_{i=1}^{n}\sum_{i\neq j}\frac{1}{\hat{\lambda}(\mathbf{u}_i,t_i)\hat{\lambda}(\mathbf{u}_j,t_j)}$ in such a way that the updated estimates of the inhomogeneous K- and pair correlation functions $\frac{1}{D(\mathbf{x})}\hat{K}(r,h)$ and $\frac{1}{D(\mathbf{x})}\hat{g}(r,h)$ provide estimators with low bias and variance. They also prove that for Poisson processes, and for any $r,h > 0$, it holds that $\hat{K}_{L,I}(r,h) = rh$ and $\hat{g}_{L,I}(r,h) = 1$, which is useful in model selection and hypothesis testing.

Therefore, a relevant result is that the expectation of $\hat{K}_{L,I}(r,h)$, when the intensity used for the weighting is the true generator model, is the same as the expectation of $\hat{K}_L(r,h)$ for the Poisson process, that is $\mathbb{E}[\hat{K}_L(r,h)] = rh$. Therefore, as in the Euclidean case, the weighted estimator $\hat{K}_{L,I}(r,h)$ can be used as a diagnostic tool for assessing the goodness-of-fit of spatio-temporal point processes occurring on a linear network with generic first-order intensity functions. Consequently, if the weighting intensity function is close to the true one $\lambda(\mathbf{u},t)$, the estimated value $\hat{K}_{L,I}(r,h)$ should be close to $\mathbb{E}[\hat{K}_L(r,h)] = rh$.

Chapter 3

Local space and space-time point processes for seismic data

Interest in methods for analysing spatial point processes is increasing across many fields of science, notably in ecology, epidemiology, geoscience, astronomy, econometrics, and crime research (Baddeley et al., 2015; Diggle, 2013). When the structure of a given point pattern is observed, it is assumed as a realization of an underlying generating process, whose properties are estimated and then used to describe the structure of the observed pattern. In analysing a spatial point process, the first step is to learn about its first-order characteristics, studying the relationship of the points with the underlying environmental variables that describe the observed heterogeneity. When the purpose of the analysis is to describe the possible interaction among points, that is, if the given data exhibit spatial inhibition or aggregation, the second-order properties of the process are analysed. However, in the analysis of spatial point process data, it can be difficult to disentangle the two previous aspects, i.e. the heterogeneity corresponding to spatial variation of the intensity and the dependence structure amongst the points (Illian et al., 2008; Diggle, 2013). For this reason, it is attractive and motivating to define and estimate models that account simultaneously for dependence structure among events, including also the effect of the observed covariates. The spatial seismic process belongs to the class of point pattern data where the locations of events in space are the observations of interest. The aim is typically to learn about the mechanism that generates these events (Møller et al., 1998; Diggle, 2013; Illian et al., 2008).

The generator process of earthquakes is quite complex since it is characterized by multidimensional and multiscale dynamics, dependence features of points both in space and time, spatial anisotropy and spatial variation of the intensity according to the seismogenic sources. Several spatial and spatio-temporal models have been proposed in the literature for describing the evolution of the earthquake process in space, in time and in both dimensions. The temporal clustered structure of earthquake data has been described by Self-Exciting point processes, i.e. the Hawkes model (Hawkes and Adamopoulos, 1973) or the Epidemic Type Aftershock Sequence (ETAS) model (Ogata, 1988; Adelfio and Chiodi, 2015b). On the other hand, when an inhibitive pattern is observed for earthquake data, self-correcting processes based on the strain-release models (Vere-Jones, 1978) are used, assuming the fault fracture generating earthquakes reduces the amount of strain present at the breakpoint along the fault. In the context of spatial point patterns, an important example of self-correcting processes is the Strauss point process model (Strauss, 1975) with a negative association, which is also the simplest inhibitive Gibbs point process (Baddeley and Møller, 1989). Gibbs point processes are generally applied in the analysis of spatial interaction and inhomogeneity depending on covariate information. For instance, a multitype Strauss process and geological spatial covariates are combined to describe sequences of events in Anwar et al. (2012); Ye et al. (2015). In Siino et al. (2017), Hybrid of Gibbs models (Baddeley et al., 2013) are used to describe the Hellenic seismicity that shows interactions at different spatial scales (multiscale structure), characterising the spatial inhomogeneity of the processes as a function of the geological information available in the study area (the presence of volcanoes, plate boundaries and faults). As for the Cox processes, they are models used to describe environmentally driven processes (Cox, 1955; Møller et al., 1998; Moller and Waagepetersen, 2003). Møller and Toftaker (2014) propose a spatial model framework to estimate geometric anisotropy in spatial point process. In particular, they characterise the anisotropy of the main seismicity around Los Angeles (California), estimating an inhomogeneous log-Gaussian Cox process. Furthermore, in the spatio-temporal context, Siino et al. (2018a) describe earthquake sequences comparing several Cox model specifications (with separable and non-separable spatio-temporal covariance functions) estimating parameters through the minimum contrast method.

All the aforementioned models can be referred to as “global” models, as they are globally defined, and the process properties and estimated parameters are assumed to be constant in the whole study area. However, a model with constant parameters may not represent detailed local variations in the data adequately, since the pattern may present spatial variation in the influence of covariates, in the scale or spacing between points and in the abundance of points. Indeed, a different way of analysing a point pattern can be based on local techniques identifying specific and undiscovered local structures, for instance, sub-regions characterised by different interactions among points, intensity and influence of covariates.

Throughout the chapter, we shall distinguish between “global” models, in which the parameters are assumed constant as in regular regression models, and “local” models, in which the parameters are allowed to vary with location.

For earthquake data, Ogata and Katsura (1988), Ogata (1989) used a space-adaptive space-time point process model. In Ogata and Katsura (1988) a method is developed for estimating both the spatial intensity of the point locations and the spatial variation of a characteristic parameter of the distributions for the attached marks, analysing seismological and ecological data. In Ogata (1989) the occurrence rate of earthquakes is extended to include multiple aftershock sequences.

For spatial point process, Baddeley (2017) presents a general framework based on the local composite likelihood to detect and model gradual spatial variation in any parameter of a spatial stochastic model (such as Poisson, Gibbs and Cox processes). In particular, the parameters in the model that govern the intensity, the dependence of the intensity on the covariates and the spatial interaction between points are estimated locally. Moreover, this approach has the advantage of detecting and modelling spatial variation in any property of a point process within a formal likelihood framework providing space-varying parameter estimates, confidence intervals and hypothesis tests.

Section 3.1 aims at using these recent results for the local composite likelihood for spatial point processes to describe the spatial displacement of earthquake data, accounting for external geological information. We show that models with parameters that vary in space can be suitable for describing both seismic catalogues and sequences.

Then, in section 3.2, we introduce the local method of *minimum contrast* for estimating the second-order parameters of a general Cox process, extended to space and time, and used for fitting local spatio-temporal log-Gaussian Cox processes for analysing the same real dataset of Greek seismic data.

3.1 Analysis of seismic catalogues and sequences by local spatial log-Gaussian Cox processes

D'Angelo, N., Siino, M., D'Alessandro, A., and Adelfio, G. (2022g). Local spatial log-gaussian cox processes for seismic data. *Advances in Statistical Analysis*. *https://doi.org/10.1007/s10182-022-00444-w*

Seismicity *catalogues* are datasets with statistical units describing earthquakes, and variables consisting in their location, origin time, and magnitude, together with additional metadata such as associated uncertainties and focal mechanism information (Woessner et al., 2010). Earthquakes typically occur in *sequences* that may include foreshocks, the mainshock (the largest event or events), and aftershocks. Aftershocks are smaller earthquakes following the mainshock. They typically occur on or near the rupture plane of the mainshock, resulting from changes in stress and frictional properties of the fault zone caused by the mainshock. The duration of aftershock sequences is typically a few years for earthquakes at plate boundaries, but can last much longer within stable continental interiors (Liu and Stein, 2019).

In this Section, we consider two case studies, analysing datasets concerning seismic activity in Greece and Italy.

The Central Mediterranean area is one of the most geodynamically complex areas and active regions of the world. In this framework are embedded the Italian peninsula and the Greek region and its Hellenic arc, which are considered the most tectonically complex and seismically active regions of Mediterranean, object of study in this work. The tectonic evolution of Greek region, is strongly related to the active subduction along the Hellenic arc which is the boundary between the African and the Anatolian-Aegean microplates (Le Pichon and Angelier, 1979; Hall et al., 2005). The overall tectonism of the region can be attributed to several events, such as the collision of African and Eurasian plates, convergence between Arabian and Eurasian plates and displacement of the Anatolian-Aegean microplate (Ryan et al., 1982; Taymaz et al., 1991). The stress regime in Hellenic arc region is a multi-component one (Oya, 2016). There are subduction zones governing the seismicity in the region, back arc events, and tension areas induced by the interference of continental and oceanic basins (Spakman et al., 1988; Taymaz et al., 1991; Papazachos et al., 1995). The shape and seismicity distribution of the Italian peninsula and Sicily island are consequence of the convergence between the African and Eurasian plates, active since at least 65 Ma and especially on the Quaternary evolution of Central Mediterranean (Amato et al., 1997; Di Stefano et al., 1999; Sgroi et al., 2012). In the Italian peninsula and Sicily, at least 5 tectonic districts can be identified: i) the Alpine arc, with evidence of past subduction; ii) the North Apennines, where subduction of the oceanic lithosphere was followed by continental collision; iii) the Calabrian arc, with clear evidence of subducting slab of oceanic lithosphere; iv) the Southern Appenines, where a detached slab is revealed by seismicity and seismic tomography; v) the Sicilian region that represents a portion of the Apennine-Maghrebide fold-and-thrust belt (Amato et al., 1997; Di Stefano et al., 1999; Sgroi et al., 2012).

Hence, when analysing seismic catalogues, the analysis of the dependence of the first-order intensity function on external covariates is a crucial issue. This is accounted for properly by the local Poisson model, describing large-scale variation by including external covariates and letting their effects vary along the studied area. This is shown with a case study for the description of the seismic events that occurred in Greece between 2005 and 2014, an area of high seismic hazard that has been characterised by many destructive earthquakes in the last century. Furthermore, the local version of the Poisson model, fitted with a spatial varying effect of covariates, is compared to a global inhomogeneous Poisson model in terms of fitting and interpretability of results.

Describing seismic sequences, the analysis of the small-scale variation is of interest, focusing on the second-order characteristics of the process. Therefore, we introduce and fit the local log-Gaussian Cox process to study a seismic sequence. To the best of our knowledge, our proposal is a novelty for this field of application and represents a first step for suggesting the use of local complex models for describing the fractal seismic phenomena. The interaction among points is typically taken into account by the log-Gaussian Cox processes models through the estimation of the covariance parameters of the Gaussian Random Field. In their local version, these parameters vary in space, allowing the description and characterization of the study area through multiple underlying processes, meaning that it would make sense to assume a generating process of the observed point pattern that is not unique. The peculiarity of the new local spatial log-Gaussian Cox processes lies in the inhomogeneity of the point process being simultaneously addressed via spatial covariates, a latent error process, and their estimated local coefficients. Therefore, after providing an application of the local Poisson models to the Greek seismic catalogue and identifying the regions that display the most inhomogeneous behaviour, a local log-Gaussian Cox process model is fitted to describe a selected Greek seismic sequence, with space varying effects of both intensity and interaction parameters. Indeed, one crucial issue in the local analysis of spatial as well as spatio-temporal point patterns is the identification of subregions where points behave differently in terms of clustering. In addition, based on the assumptions on the underlying processes, local estimates provide further insight into the relationship of the intensity of the process on external covariates, as well as on the correlation structure, that in this setting is allowed to vary with location. Local models are also applied to the Italian data, since as Greece, Italy is an area of high seismic risk, and in particular, Central Italy, where our analysis focuses, has been characterised by a strong series of earthquakes between 2009 and 2017. The results are reported in section 3.1.3 for the sake of brevity. The key assumption stating that for each spatial location, there is an unobserved spatial subdomain inside which the template model is exactly true allows us to exploit the statistical properties of the spatially varying parameters, as they were the ones of a model with constant parameters in that specific subregion. This feature is particularly appealing as it allows us to build diagnostic tools and interpret parameters as we are used in classical log-linear models. Besides this, we also test on the two case studies that local models provide good inferential results if compared to the semi-parametric competitors.

Moreover, this work represents a contribution to the framework of diagnostics for local models.

Indeed, when a model is fitted to a set of random points, diagnostic measures are necessary to assess the goodness-of-fit and to evaluate the ability of that model to describe the random point pattern behaviour (Adelfio et al., 2020). However, tools for checking or criticizing the fitted model are quite limited. There is currently no analogue for spatial point patterns of the comprehensive strategy for model criticism in the linear model, which uses tools such as residual plots and influence diagnostics to identify unusual or influential observations, for assessing the model assumptions sequentially and to individuate any departures from the model (Baddeley et al., 2005).

It is a widespread practice in the statistical analysis of spatial point pattern data to focus primarily on comparing the data with a homogeneous Poisson process ("complete spatial randomness"), which is generally the null model in applications rather than the fitted model. This approach considers a stationary Poisson residual process by randomly rescaling (Meyer, 1971; Schoenberg, 1999) or thinning (Schoenberg, 2003) and investigates whether the second-order properties of the observed residuals are consistent with those of the stationary Poisson process.

An alternative approach is to define a weighted second-order statistic, where essentially, to each observed point, a weight inversely proportional to the conditional intensity at that point is given. This method was adopted by Veen and Schoenberg (2006) in constructing a weighted version of the spatial K-function of Ripley (1977), and Veen (2006), although the spatially weighted analogue of Ripley's K-function which was first introduced by Baddeley and Turner (2000).

Therefore, establishing a coherent strategy for model criticism in spatial point process models, resembling the existing methods for the linear model, depends crucially on finding the right definition of residuals for a spatial point process model fitted to point pattern data.

In this Section, we propose the use of some diagnostics tools generally considered for global point processes, e.g. the smoothed raw residuals and the inhomogeneous K-function (Baddeley et al., 2005), to the local context, and properly review local tests and the test of departure from "homogeneity" (Baddeley, 2017), meaning that the observed pattern is generated by a template model with constant parameters. Our contribution concerns a bootstrap procedure carried out to support the model selection process and also retain local information. Furthermore, a stepwise procedure is also developed, assessing the single local effects of the considered covariates on the rejection of the given "homogeneity" hypothesis. The method provides a ranking of the covariates that most contribute to the local features, together with p-values maps, useful for progressively identifying the variables' effects within the study area.

Therefore, this section contributes to the framework of the local spatial processes in two main directions: providing a formalization and characterization of local LGCP for earthquake data and contributing to diagnostic methods for this class of models.

A review of spatial point processes and log-linear Poisson models, in terms of both global and local methodology, was already reported in section 2.1.4. The spatial local log-Gaussian Cox processes are introduced, and their estimation procedure is discussed, together with diagnostics methods for local models. In particular, it contains global diagnostics tools, well-established in spatial point pattern analysis, and suitable for local models. Then, section 3.1 is structured as follows. In Section 3.1.1, we propose two local diagnostics procedures based on the test of local departure from "homogeneity" testing. In section 3.1.2, local models are applied to describe seismic activity in Greece, describing both the entire Greek seismic catalogue and a particular seismic sequence. In section 3.1.3, an Italian seismic sequence is analysed by means of local models, including different covariates coming from a seismic catalogue. Section 3.1.4 is devoted to conclusions and final remarks.

3.1.1 A bootstrap procedure for "homogeneity" testing

Recall the local models and diagnostic methods introduced in section 2.1.4. When the rejection of the null hypothesis, based on the tests reviewed in section 2.1.4.2.2, is particularly due to some specific areas, it becomes interesting to individuate those regions with inhomogeneous behaviour, that is, splitting the study area into subregions requiring further study or different models. Therefore, to confute the hypothesis of "homogeneity" while retaining information about the local features, a bootstrap procedure is proposed and carried out in this chapter. Indeed, the bootstrap procedure (Efron, 1982) is often used as an alternative to the statistical inference based on the assumption of a parametric model when that assumption is in doubt. The procedure used in this chapter can be outlined as follows:

1. Set M as the number of bootstrap samples to be considered.
2. Bootstrap M times from the analysed point process, that is, random sample with replacement the points of the process, obtaining M new point processes.
3. The candidate local model is fitted to each of the M bootstrapped samples.
4. For each of the M resulting local models, the test of "homogeneity" is carried out, obtaining M sets of local p-values for the given confidence level a.

5. The local p-values are averaged over the M samples.

The resulting averaged local p-values can be plotted as a function of the spatial location $\mathbf{u}$, together with their standard deviations. This map visually identifies the regions with inhomogeneous behaviour, i.e. where there is evidence against the hypothesis of "homogeneity". Contrary to the global p-value, computed as the average of the local p-values of one "homogeneity" test, this bootstrap procedure contributes to the model selection process, also retaining local information. Fixing a confidence level a, e.g. 0.05, the areas where the local p-values are less than a confute the hypothesis of "homogeneity", while values greater than a provide evidence in favour of the null hypothesis, that is, where the global model could be more preferable than the local one. An average of the resulting local p-values can also be computed, contributing to the model selection procedure, empirically comparing this value with the fixed level a (Diggle and Gratton, 1984).

3.1.1.1 Assessing the effect of covariates on the local departure from "homogeneity"

When the previously discussed local tests provide evidence against the null hypothesis of "homogeneity", we can describe the available point pattern by a local model. At this point, we wonder which covariates most contribute to the rejection of this hypothesis, that is to say, what covariates' effects really vary in space. Indeed, the "homogeneity" test compares the linear predictor of the local model to the one of the corresponding global model, and therefore it does not provide information on each covariate individually.

In this section, we contribute to the framework of diagnostic tools for local models by proposing a stepwise procedure that highlights the effect of each covariate on the rejection of the hypothesis of "homogeneity". The procedure that we describe in this section can be used for local Poisson models with multiple linear predictors, that is, dealing with multiple covariates.

Suppose that the chosen model is a local Poisson model with spatial varying intercept $\theta_0(\mathbf{u})$ and parameters $\boldsymbol{\theta}(\mathbf{u}) = (\theta_1(\mathbf{u}), ..., \theta_p(\mathbf{u}))$ corresponding to the spatially varying effects of the p covariates $\mathbf{Z}(\mathbf{u}) = (Z_1(\mathbf{u}), ..., Z_p(\mathbf{u}))$. Several local Poisson models are fitted, starting from the one in which only the space-varying intercept is added into the linear predictor, and adding one spatial covariate at a time. The covariates are added starting from the most significant in explaining the intensity marginally, based, for instance, on the information obtained from results of Berman's tests (Berman, 1986). This test is used to assess the dependence of a point process on a spatial continuous covariate $Z(\mathbf{u})$. It can be carried out before formulating and fitting a model, to get a first indication of the variables influencing the intensity of the process, marginally. Then, the test of "homogeneity" is carried out in sequence to each of the local Poisson models, basically comparing each time the local Poisson model to its global counterpart. This procedure can be outlined as follows:

1. The Berman tests are computed on each available covariate, marginally. The covariates $\mathbf{Z}(\mathbf{u}) = (Z_1(\mathbf{u}), ..., Z_p(\mathbf{u}))$ are then ranked from the most significant in explaining the intensity of the analysed process (therefore the one with the lowest p-value) to the least significant, obtaining the ordered vector $\mathbf{Z}^*(\mathbf{u}) = (Z_1^*(\mathbf{u}), ..., Z_p^*(\mathbf{u}))$. The corresponding parameters are also ranked accordingly.
2. The model $\lambda(\mathbf{u}) = \exp(\theta_0(\mathbf{u}))$ is fitted and the first test of local departure from "homogeneity" is carried out for comparing

   $$\begin{cases} \mathcal{H}_0: & \lambda(\mathbf{u}) = \exp(\theta_0) \\ \mathcal{H}_1: & \lambda(\mathbf{u}) = \exp(\theta_0(\mathbf{u})) \end{cases}$$

   obtaining the first map of p-values, where values smaller than 0.05 indicate the regions where the local model should be preferred to the global one.
3. The second local model is fitted, adding the covariate that resulted in the most significant in explaining the intensity of the process marginally, to the linear predictor $\lambda(\mathbf{u}) = \exp(\theta_0(\mathbf{u}) + \theta_1^*(\mathbf{u})Z_1^*(\mathbf{u}))$. Therefore, the second test of local departure from "homogeneity" is carried out to compare

   $$\begin{cases} \mathcal{H}_0: & \lambda(\mathbf{u}) = \exp(\theta_0 + \theta_1^* Z_1^*(\mathbf{u})) \\ \mathcal{H}_1: & \lambda(\mathbf{u}) = \exp(\theta_0(\mathbf{u}) + \theta_1^*(\mathbf{u})Z_1^*(\mathbf{u})). \end{cases}$$

4. Repeat the last step for each $Z(\mathbf{u})$ following the ranking given in step 2 until the test of local departure from "homogeneity" is carried out on the chosen model

   $$\begin{cases} \mathcal{H}_0: & \lambda(\mathbf{u}) = \exp(\theta_0 + \theta_1^* Z_1^*(\mathbf{u}) + \cdots + \theta_k^* Z_k^*(\mathbf{u})) \\ \mathcal{H}_1: & \lambda(\mathbf{u}) = \exp(\theta_0(\mathbf{u}) + \theta_1^*(\mathbf{u})Z_1^*(\mathbf{u}) + \cdots + \theta_k^*(\mathbf{u})Z_k^*(\mathbf{u})). \end{cases}$$

The result is a set of maps, as many as the performed tests, that is, the number of covariates in the chosen model, plus the intercept. Each map will display the p-values associated with the test of "homogeneity" that compares the local model considered at that step with its global counterpart. Looking at each map individually, we obtain information about the regions in which the "homogeneity" hypothesis is rejected. Of course, the wider the area leading to rejection, the stronger the evidence against the global model. By comparing the rejection areas obtained through the subsequent tests, carried out by adding one variable at a time, we are able to identify the most contributing variables to the overall rejection of the "homogeneity" hypothesis in the final chosen model. In particular, if adding one variable to the model, say $Z_j(\mathbf{u})$, the rejection area of the corresponding map gets wider than the one relative to the model without $Z_j(\mathbf{u})$, then $Z_j(\mathbf{u})$ strongly contributes to the overall rejection, that is, the effect of $Z_j(\mathbf{u})$ varies in space. Otherwise, if no significant differences are evident between subsequent maps, then, the variable $Z_j(\mathbf{u})$ does not contribute significantly to the overall rejection, suggesting that the effect of $Z_j(\mathbf{u})$, though considered as local in the model, could actually be considered as constant. This procedure is validated on the basis of the chosen local Poisson model fitted to the Greek seismic catalogue data in section 3.1.2.1.

3.1.2 Analysis of the Greek seismicity

In this section, local models are applied to study the recent seismicity in Greece, assessing their goodness of fit by the diagnostics tools previously introduced.

Global models, the global diagnostic tools reviewed in section 2.1.4.2.1 and Berman's tests are computed with functions implemented in the `spatstat` package (Baddeley and Turner, 2005). As for the local tests, and the fitting of the local models we refer to the `spatstat.local` package (Baddeley, 2019).

The analysed data (1105 events) concern earthquakes that occurred in Greece between 2005 and 2014. Only seismic events with a magnitude larger than 4 are considered in this study, and the analyses in this section are marginal with respect to time, focusing on the spatial dependence of events. In Figure 3.1 (a) earthquakes are reported together with the location of the five active volcanoes of the area (in green), the plate boundary (in red), and the faults (in blue). Data about the seismic events come from the Hellenic Unified Seismic Network (H.U.S.N.), while covariate information comes from the Greek Database of Seismogenic Sources (GreDaSS).

In Figure 3.1 (b) the observed K-function, is represented. The dotted red line represents the values of the K-function for a homogeneous Poisson process, while the black one represents the K-function of the observed point process. The K statistic of the observed point pattern, with larger values than the theoretical one of a Poisson process, suggests that distances in the observed pattern are shorter than the Poisson ones. In other words, events are more clustered than a homogeneous Poisson pattern, as is also evident in Figure 3.1 (a).

Starting from the information of the area under study, we are interested in assessing if the presence of seismic sources affects the intensity of the process, that is if the effect of the available covariates can improve the fitting of the model, by fitting a global inhomogeneous Poisson process. Therefore, the considered explanatory variables are *Distance from the faults* ($D_f(\mathbf{u})$), *Distance from the plate boundary* ($D_{pb}(\mathbf{u})$) and *Distance from volcanoes* ($D_v(\mathbf{u})$), computed as the Euclidean distances from the spatial location $\mathbf{u}$ of events and the map of geological information (Baddeley et al., 2015). In particular, only the information about the top segment of the fault is used to compute the distances. The covariate surfaces are displayed in Figure 3.2.

The Berman's tests are computed for assessing the dependence of the intensity

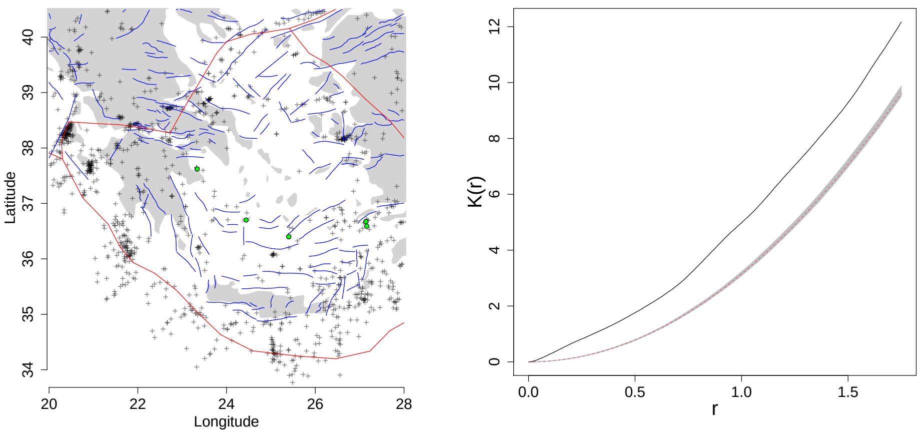


Figure 3.1: Earthquakes occurred in Greece between 2005 and 2014. Volcanoes, in green, faults, in blue, and plate boundary, in red (a); K-function of the observed point process of the Greek data (b). The dotted red line represents the values of the K-function for a homogeneous Poisson process, while the black one represents the K-function of the observed point process.

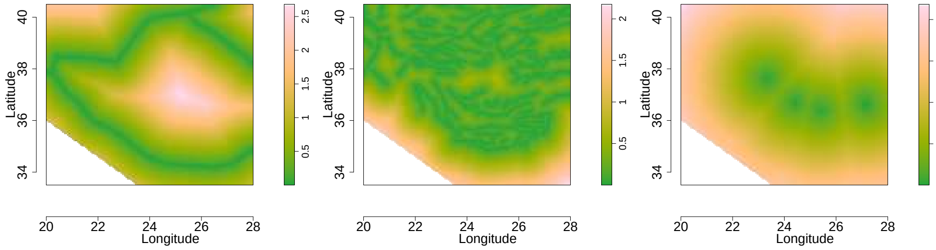


Figure 3.2: The available covariates for the Greek data: Distance from the faults ($D_f(\boldsymbol{u})$), Distance from the plate boundary ($D_{pb}(\boldsymbol{u})$) and Distance from volcanoes ($D_v(\boldsymbol{u})$), computed as the Euclidean distances from the spatial location $\boldsymbol{u}$ of events and the map of geological information.

of the process from the continuous spatial covariates. The corresponding p-values are displayed in Table 3.1.

Table 3.1: Berman's Tests carried out for the available covariates for the Greek data

Variable	p-value
Distance from the faults	0.0000
Distance from the plate boundary	0.0000
Distance from the volcanoes	0.6862

Therefore, the marginal effects of $D_{pb}(\mathbf{u})$ and $D_f(\mathbf{u})$ seem significant. Otherwise, $D_v(\mathbf{u})$ does not significantly affect the earthquakes' intensity (p-value 0.68). To assess the variables' influence on the intensity of the process, we plot the marginal smoothed intensity function for each covariate. Let $\lambda(\mathbf{u})$ be the intensity of the process under study, and $Z(\mathbf{u})$ the considered covariate. Then, assuming $\lambda(\mathbf{u}) = f(Z(\mathbf{u}))$ where f is a non-parametric estimate of the intensity of the analysed point process, we wonder if the observed intensity depends on each spatial covariate, at least marginally. Smooth estimators of $f(Z(\mathbf{u}))$ were proposed by Baddeley et al. (2012), and the smoothing procedure we consider is based on fixed-bandwidth kernel density estimation. The Poisson confidence bands are also computed.

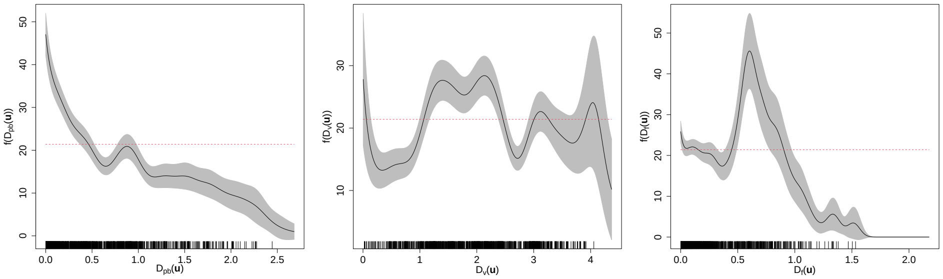


Figure 3.3: Non-parametric estimate of the intensity of the analysed point process, as a function of the available variables for the Greek data: Distance from the plate boundary (a), Distance from the volcanoes (b) and Distance from the faults (c). The considered smoothing procedure is based on fixed-bandwidth kernel density estimation.

The analysis of the smoothed functions confirms the results of the Berman's tests. Indeed, $D_v(\mathbf{u})$ does not seem to affect the intensity, and Figure 3.3 (b) shows a quite constant relationship with the intensity. Looking at the smoothed functions of $D_{pb}(\mathbf{u})$ and $D_f(\mathbf{u})$, it is evident that the effect of both covariates varies as a function of the scale. In detail, the intensity exponentially decreases moving away from the plate boundary, while it has a piece-wise trend with respect to $D_f(\mathbf{u})$.

In section 3.1.2.1, a local Poisson model is fitted to the Greek catalogue data and results are compared to those referred to a global model fitting, assuming an inhomogeneous Poisson model. In section 3.1.2.2, a local LGCP model is fitted to a seismic sequence selected from this catalogue of events.

3.1.2.1 Local Poisson model for the Greek catalogue

On the basis of the previous results and for the given data, we want to assess if the local features cannot be ignored, fitting an inhomogeneous Poisson model as in Equation (3.1). Starting from the available variables, the choice of the final model is based on diagnostic approaches, as reported in Sections 2.1.4.2.1 and 2.1.4.2.2. The chosen model is the one including the *Distance from the faults* (D_f), the *Distance from the plate boundary* (D_{pb}) and the *Distance from the volcanoes* (D_v), that is

$$\lambda(\mathbf{u}) = \exp(\theta_0(\mathbf{u}) + \theta_1(\mathbf{u})D_f(\mathbf{u}) + \theta_2(\mathbf{u})D_{pb}(\mathbf{u}) + \theta_3(\mathbf{u})D_v(\mathbf{u})) \tag{3.1}$$

where the smoothing parameter h, obtained maximizing the LCV in Equation (2.11) used in the local likelihood in Equation (2.10), is 0.73. The maps of the varying coefficients $\hat{\theta}_1(\mathbf{u})$, $\hat{\theta}_2(\mathbf{u})$ and $\hat{\theta}_3(\mathbf{u})$ are shown on the top three panels in Figure 3.4.

The summary statistics of the local coefficients are reported in Table 3.2. As expected, the estimated coefficients are negative quite in the whole area, since usually, as the distance from the seismic source increases the intensity decreases, but it is important to notice that the coefficients take quite different values along the whole area, reaching the lowest values of -10.39 for $\hat{\theta}_1(\mathbf{u})$. The lowest values correspond to the earthquakes that occurred along the seismic sources.

Table 3.2: Summary of the space varying coefficients of the local Poisson model (3.1)

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
Intercept	-0.8465	3.7821	4.6536	4.7818	5.4126	13.7942
D_f	-10.3862	-4.0474	-2.3888	-2.4529	-0.6133	4.6906
D_{pb}	-4.735	-1.925	-1.413	-1.448	-1.04	2.17
D_v	-5.3552	-0.6595	-0.2799	-0.2987	0.2351	2.9546

To study the effect of each covariate separately and to assess if the estimated values of the spatial varying coefficients are significant, T-tests of the parameters of the model are computed, as in Equation (2.18). The T-test output is shown in the bottom panels of Figure 3.4, where the level curves correspond to the ±1.96 threshold, associated with a 0.05 confidence level. Therefore, in the darkest (and lighter) regions the null hypothesis is rejected, that is, the coefficients corresponding to these locations are significantly different from zero. It is interesting to notice that the coefficients of $D_f(\mathbf{u})$ take negative values along the whole study area but for a wide area on the bottom-left of the window in which coefficients are positive and

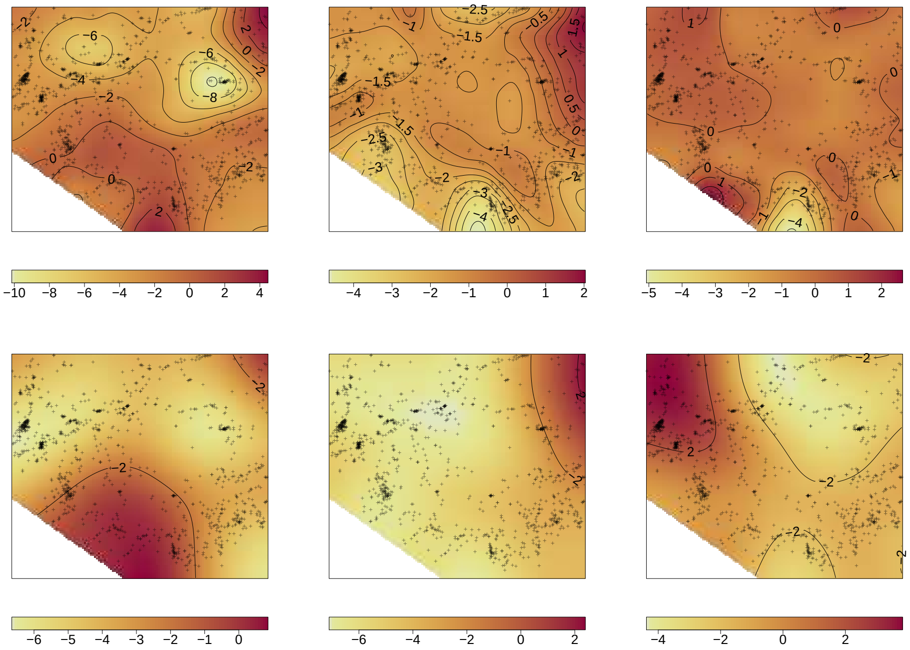


Figure 3.4: On panels (a), (b), and (c): Spatial varying coefficients $\hat{\theta}_1(\boldsymbol{u})$, $\hat{\theta}_2(\boldsymbol{u})$ and $\hat{\theta}_3(\boldsymbol{u})$ of the local Poisson Model in Equation (3.1), the coefficients of the variables $D_f(\boldsymbol{u})$, $D_{pb}(\boldsymbol{u})$, and $D_v(\boldsymbol{u})$, respectively. On panels (d), (e), and (f): the corresponding T-tests. T-level curves correspond to the ±1.96 threshold, associated with a 0.05 confidence level. The darkest (and lighter) regions identify where the null hypothesis is rejected, that is where the local coefficients are significantly different from zero.

not significant. The coefficients of $D_{pb}(\mathbf{u})$ are significant and negative basically in the whole study area, but for the two areas at the boundary of the window. Finally, the coefficients of $D_v(\mathbf{u})$ are significant only in three bounded areas of the window, taking both positive and negative values.

These results suggest that a local model might be suitable for describing the features of the study area. To get more evidence supporting this hypothesis, the test of the “homogeneity” is carried out, bootstrapping 100 datasets from the original one, and computing the average of the resulting p-values obtained for each location, as discussed in section 3.1.1. These values are shown in Figure 3.5, together with their standard deviations, suggesting the inadequacy of the null hypothesis for the whole study area: indeed, few local p-values slightly greater than the fixed level 0.05 are observed, but less than 0.1, in the southern region, corresponding also to the sub-area with the largest standard errors. Besides, the mean of all the previous p-values is 0.01, with a standard deviation equal to 0.05, further corroborating the hypothesis of the absence of “homogeneity” in these data.

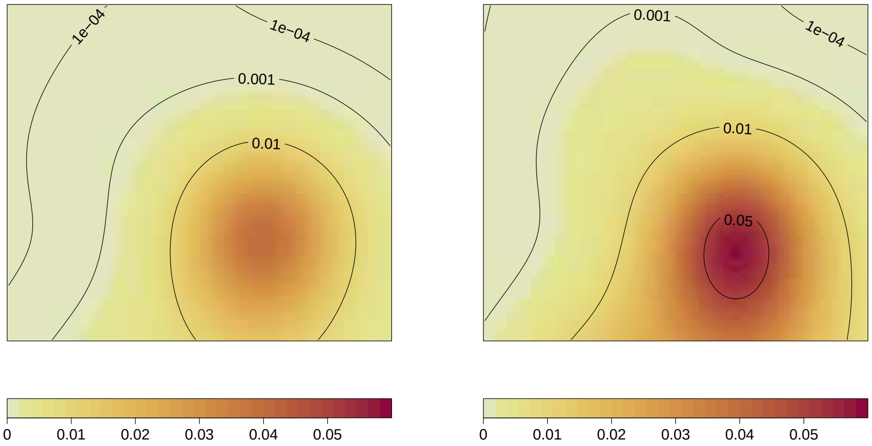


Figure 3.5: Averaged local p-values (left panel) and their standard deviations (right panel) for the test of “homogeneity”, computed over 100 bootstrapped datasets

This result suggests that the proposed local model in Equation (3.1) should be preferred to its global counterpart

$$\lambda(\mathbf{u}) = \exp(\theta_0 + \theta_1 D_f(\mathbf{u}) + \theta_2 D_{pb}(\mathbf{u}) + \theta_3 D_v(\mathbf{u})).$$

To understand the effect of each variable in rejecting this hypothesis, we propose the stepwise comparison between the different ‘homogeneity’ settings, introduced in section 3.1.1.1. Therefore, four local Poisson models are fitted, starting from the one in which only the local intercept is considered, and adding one spatial covariate at a time. The covariates are ranked from the most significant in explaining the intensity marginally, based on the results of Berman’s tests in Table 3.1, such that $\mathbf{Z}^*(\mathbf{u}) = (D_f(\mathbf{u}), D_{pb}(\mathbf{u}), D_v(\mathbf{u}))$. These are subsequently added to the linear predictor and the test of “homogeneity” is carried out each time comparing the template global Poisson model to its local counterpart. Values smaller than 0.05 in Figure 3.6 are plotted in lighter colours, and they are found in those regions where the hypothesis of “homogeneity” is rejected. Of course, we may notice that the more covariates with local effects, the more areas ‘needing’ a local model. In particular, as the greater differences between subsequent maps are in correspondence with the addition of $D_f(\mathbf{u})$ and $D_{pl}(\mathbf{u})$, this means that these are the covariates with spatially varying effects. On the contrary, $D_v(\mathbf{u})$ slightly contributes to the rejection of the hypothesis and this may suggest that its effect could be considered constant. Overall, we may notice that the best fitting for the whole area would require an even more complex model, but introducing local coefficients makes the model more flexible and more useful for a deeper interpretation.

A feasible trade-off between the flexibility of a local model and the parsimony of a global one could be the model with a linear predictor that includes a non-parametric term for spatial coordinates and parametric expression for the spatial covariates, as

$$\lambda(\mathbf{u}) = \exp(f(\mathbf{u}) + \theta_1 D_f(\mathbf{u}) + \theta_2 D_{pb}(\mathbf{u}) + \theta_3 D_v(\mathbf{u})) \tag{3.2}$$

where $f(\cdot)$ is a non-parametric function for $\mathbf{u} \in D$, estimated here through thin plate regression splines (Wood, 2003) with 30 knots. Thin plate splines are a type of smoothing spline used for the visualization of complex relationships between continuous predictors and response variables. Thin plate splines are ideal for examining the combined effect of two continuous predictors on a single outcome, because of their multi-dimensional appearance. Therefore, they are suitable for accounting for the spatial coordinates effect on the first-order intensity function of a point pattern. The choice of the number of knots is performed by comparing models by means of the AIC (Hastie and Tibshirani, 1990). Comparisons between the models (3.1) and (3.2) are first carried out by the inspection of the smoothed raw residuals, as defined in Equation (2.17). The smoothing bandwidth used for the kernel estimation of the raw residuals is 0.025. The smoothed raw fields of the two models are shown in Figure 3.7 and their ranges are reported in Table 3.3.

Table 3.3: Range of the smoothed raw fields of the local Poisson Model (3.1) *and the global Poisson Model* (3.2)

	Min. range	Max. range
GlobalPoisson	-1.27	1.41
LocalPoisson	-3.51	3.42

From these results, we see that both models achieve a good fitting to the data, even if the global model slightly presents smaller residuals than the local one.

Secondly, a comparison is carried out in terms of the inhomogeneous K-function introduced in Equation (2.2). In Figure 3.8 (a) we report the inhomogeneous K-function for the local model (3.1), while in Figure 3.8 (b), the one of the global inhomogeneous semi-parametric model (3.2) is displayed. In both Figures, the solid black line represents the inhomogeneous K-function $\hat{K}_{inhom}(r)$ estimated with the intensity of the fitted model, while the dotted red line represents the theoretical one $K_{inhom,pois}(r)$. The corresponding envelopes, that are obtained considering the inhomogeneous K-function for a given number of simulated patterns, assuming the same intensity of the observed patterns, are displayed in grey. For their computation, we simulated 100 realizations from the fitted model in Equation (3.1). The inhomogeneous K-function of both models lay very close to the theoretical Poisson

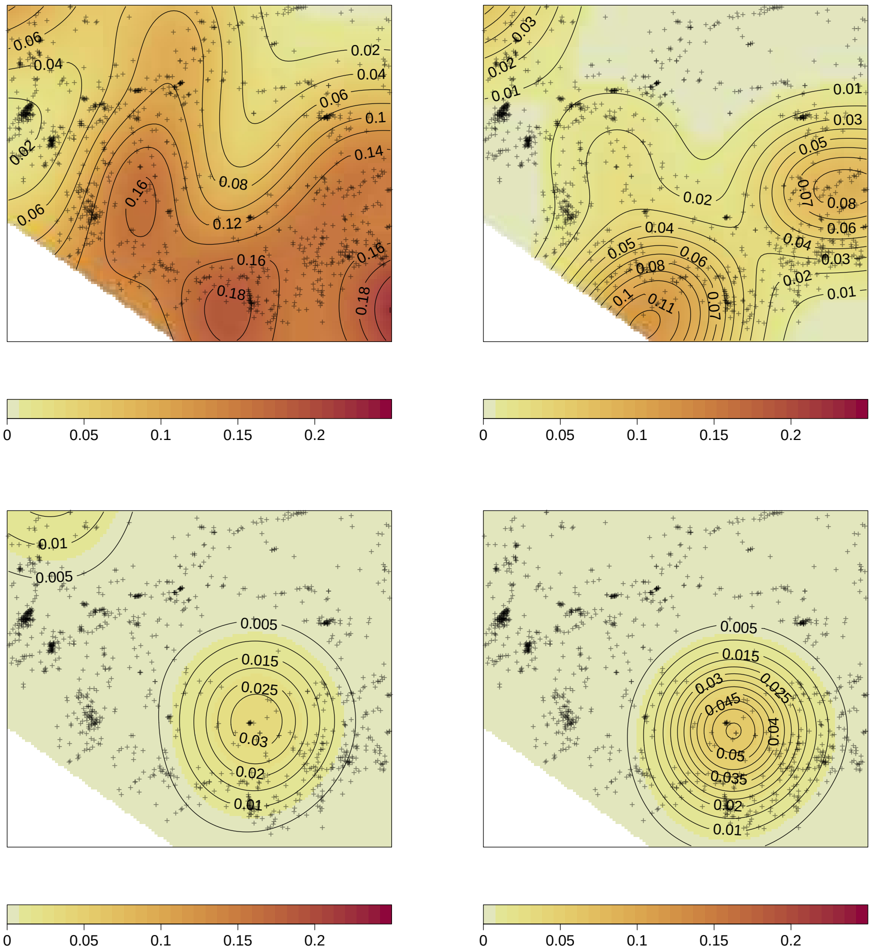


Figure 3.6: Test of "homogeneity" change: p-values of the test of "homogeneity" are plotted, in reference to the four local Poisson models fitted for describing Greek data. Starting from the one which includes only $\hat{\theta}_0(\boldsymbol{u})$ in panel (a), the covariates are added in the models gradually. The p-values are computed for the tests of "homogeneity" carried out on the models including, step-by-step, the Distance from the faults (b), the Distance from the plate boundary (c) and the Distance from the volcanoes (d)

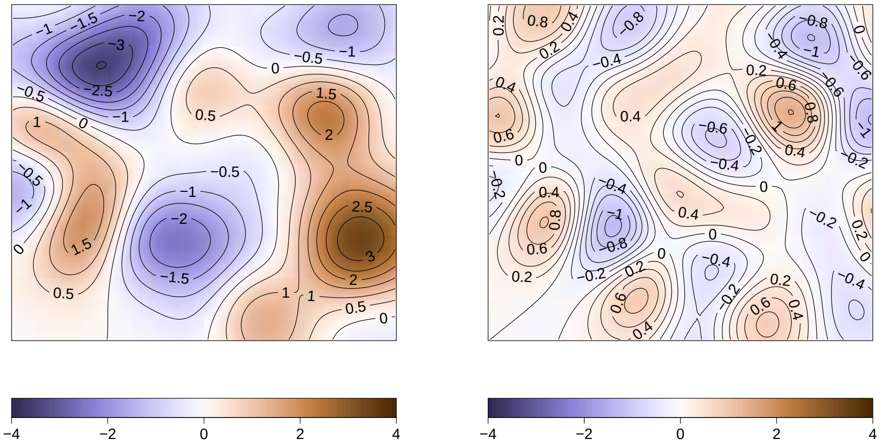


Figure 3.7: Smoothed raw residuals of the local inhomogeneous Poisson Model (3.1) *and the global inhomogeneous semi-parametric Poisson Model* (3.2)

curves, even if the one of the global model is a bit closer to its theoretical one, even for larger distances. Based on these results, the global model in Equation (3.2) should be preferred to the local one in Equation (3.1).

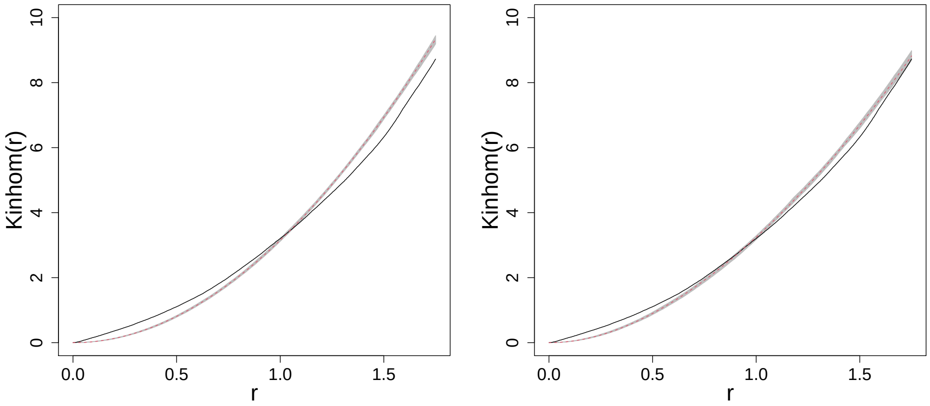


Figure 3.8: Inhomogeneous K-functions of the local inhomogeneous Poisson Model (3.1) *(a) and the global Inhomogeneous semi-parametric Poisson Model* (3.2) *(b). In both panels, the inhomogeneous K-function estimated on the observed pattern* $\hat{K}_{inhom}(r)$ *is the solid black line and the theoretical Poisson one* $\hat{K}_{inhom,pois}(r)$ *is the dotted red line, together with the corresponding envelopes, in grey.*

Overall, we might now conclude that both the global model (3.2), in which the spatial component is modelled non-parametrically, and the local model (3.1) achieve good fitting to the analysed data. Nevertheless, with both models, it is not possible to get a concise comment about the effect of the spatial covariates, earning in terms of fitting but losing interpretation. On the one hand, the great advantage of local models is that one can get a whole map of parameters, one for each covariate added in the linear predictor, and this allows us to detect those regions in which covariates have a different effect of the phenomenon under study. Therefore, if we wish to get a general overview of the effect of covariates a global model is to be preferred. On the other hand, if we are interested in spotting those areas in which covariates may have a different effect, and therefore, highlighting the multi-scale nature of the observed generating process, local models could be more appropriate. Moreover, as we add more covariates into the linear predictor, complexity increases, as each one of them carries a whole map of parameters. Finally, local models are useful also when covariates are not available. Indeed, since the test of "homogeneity" can be carried out to any kind of linear predictor, this may spot those regions that really need a local varying intercept, that is those areas that need more attention and further analysis.

3.1.2.2 Local log-Gaussian Cox process model for a Greek seismic sequence

When analysing seismic sequences, the analysis of the small-scale variation is of crucial interest (Giunta et al., 2009). For this purpose, a log-Gaussian Cox process is a candidate model, accounting for the interaction among points through a Gaussian Random Field. In their local version, parameters of the Gaussian Random Field vary in space, allowing to describe the complex dependence structure of the generating process. Starting from the Greek catalogue, we selected a smaller area in order to emphasize the dependence features at short distances. The chosen area is situated in the Ionian Sea, and it consists of three of 'the Seven Islands', namely Ithaki, Kefalonia and Zakynthos, from North to South. In this area, earthquakes appear to be clustered in the Western area of Kefalonia island and South to Zakynthos island. For this analysis, only $D_f(\mathbf{u})$ and the $D_{pb}(\mathbf{u})$ are considered, since the volcanoes are very far from this analysed area. In Figure 3.9 earthquakes, together with plate boundary (in red) and faults (in blue), are displayed. Given that previous analyses have highlighted the high inhomogeneity of the chosen subregion, and therefore the need for more complex models than the inhomogeneous Poisson, we fit the log-Gaussian Cox process, as our objective here is to find the model that better describes the clustered structure of the point pattern.

After choosing the following multiplicative local Poisson model for the deterministic component, then, it is used to fit the local LGCP model:

$$\lambda(\mathbf{u}) = \exp(\theta_0(\mathbf{u}) + \theta_1(\mathbf{u})D_f(\mathbf{u}) + \theta_2(\mathbf{u})D_{pb}(\mathbf{u}) + \theta_3(\mathbf{u})D_f(\mathbf{u})D_{pb}(\mathbf{u})) \tag{3.3}$$

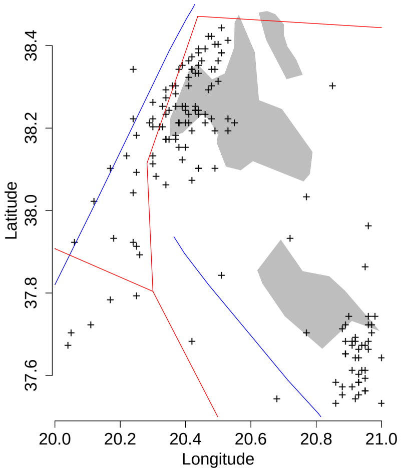


Figure 3.9: Earthquakes occurred in Ithaki, Kefalonia and Zakynthos between 2005 and 2014. The plate boundary is displayed in red while the faults are displayed in blue

where the local weight function w in the local Palm likelihood is an isotropic Gaussian density, with a smoothing parameter equal to 0.135, chosen by the cross-validation criterion (2.16). The summary statistics of the estimated parameters of the deterministic component of the local LGCP model are reported in Table 3.4.

Table 3.4: Summary statistics of the local coefficients of the deterministic component of the local LGCP model for the Greek data

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
Intercept	-11.264	2.372	5.797	2.943	6.396	10.052
D_f	-28.259	1.318	8.095	8.303	14.777	51.888
D_{pb}	-23.820	-8.782	-0.575	6.843	25.126	55.091
$D_f D_{pb}$	-306.57	-84.03	-45.48	-47.90	-12.10	63.96

The estimates of the local interaction parameters $\hat{\sigma}^2(\mathbf{u})$ and $\hat{\alpha}(\mathbf{u})$ are displayed in Figure 3.10 and their summary statistics are reported in Table 3.5.

The results in Table 3.5 are quite reasonable, since the estimates of the coefficients $\hat{\sigma}^2$ and $\hat{\alpha}$ of the global counterpart of the LGCP model, estimated maximizing the (global) Palm Likelihood, are respectively equal to 1.92 and 0.2.

We know that the effect of increasing σ^2 is to generate higher peaks in the surface intensity, inducing a clustering effect. Increasing the spatial scale parameter

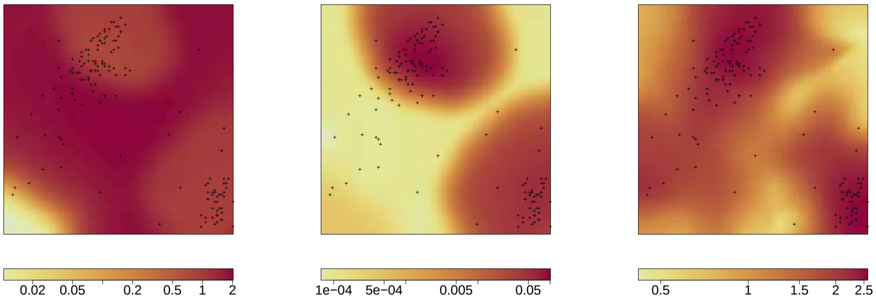


Figure 3.10: Local coefficients $\hat{\sigma}(\boldsymbol{u})$ and $\hat{\alpha}(\boldsymbol{u})$ of the global LGCP model, respectively (a) and (b) and the mean of the process $\hat{\mu}(\boldsymbol{u})$ (c)

Table 3.5: Summary statistics of the local coefficients of the covariance function of the local LGCP model for the Greek data

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
σ^2	0.000105	0.802612	2.451644	2.192861	3.627219	4.157830
α	0.00001	0.00009	0.00012	0.01988	0.03764	0.11281

α, the underlying GRF presents a strong spatial correlation, and it corresponds to a diffuse aggregation of points of the LGCP. Therefore, from the visual inspection of the maps of the interaction parameters in Figure 3.10, we are able to identify regions with different underlying Gaussian processes, driven by different covariance structures. Figure 3.10 (a) shows the map of the estimates $\hat{\sigma}(\mathbf{u})$. We may notice that, unexpectedly, the highest values are found where events are not clustered. Looking at Figure 3.10 (b), the highest values of $\hat{\alpha}(\mathbf{u})$ correspond to those regions exhibiting clustered points. From the exponential expression of the covariance chosen for the GRF in Equation (2.6) we know that the values of σ^2 contribute to the covariance in a multiplicative way, while the values of α affect the covariance exponentially. Then, a unit increase of α has a greater effect on the computation of the covariance, if compared to a unit increase of σ^2. Therefore, inspecting Figure 3.10, we know that the difference in the correlation among points is evident, and this is higher in those areas in which clusters are observed, i.e. on the top and on the bottom-right of the region.

A further application of the proposed local models is provided in section 3.1.3, focusing on the study of an important seismic sequence of Abruzzo, a region of central Italy, characterized by different geological features with respect to the Greek ones.

3.1.3 Analysis of the Italian seismicity

In this section, the local LGCP model is used to describe an important Italian seismic sequence. The aim of specifying an LGCP model is to take into account the existing interactions among events, that are not accounted for by a Poisson model, even local. The analysis of the Greek seismic events has been carried out assuming that the intensity of earthquakes could depend only on spatial covariates, such as the distance to the seismic sources. This has been useful in order to test local models and make different comparisons, but these variables only, all representing distances from some source of seismic activity, may not be enough in order to explain the phenomenon. Indeed, one could assume that other relevant variables could be useful to explain the intensity of the observed earthquakes, such as variables related to network dislocation and measurement errors.

The Italian catalogue considered in this section is downloaded from the Istituto Nazionale di Geofisica e Vulcanologia (INGV) archive[1], and it describes the space-time Italian seismicity, from May 5th 2012 to May 7th 2016 with 2.5 as the threshold magnitude, and reports 2226 events and several variables, including hypocentral coordinates (longitude, latitude, depth, time), magnitude, hypocentral uncertainty, distance from the nearest station, rms, which is a measure of the quality of the location and the number of stations that recorded the event. As the location of faults along the study area is also available, the distance from the nearest fault is computed, in the same way as for the Greek data, and it is denoted by $D_f(\mathbf{u})$.

The Italian data is quite different with respect to the Greek one. While the Greek covariates have been obtained from information external to the analysed point pattern, the Italian catalogue carries more complete information about seismic events. Indeed, Greek covariates are defined for the whole region under study, while Italian covariates refer only to those seismic events already occurred. While the structure of the Italian catalogue may lead to thinking of the available variables as marks of the point process, some of these variables can be actually thought of as spatial covariates, if interpolated along the whole study area. Indeed, it makes sense to think that variables such as the Hypocentral uncertainty, the RMS and the Distance to the nearest station (Err, Rms and D_{ns}) could depend on the location, and therefore it makes sense to get an interpolation of these variables along all the study area. On the other hand, variables such as Magnitude, Depth and Time, are characteristics of the points, known as ‘marks’ in spatial point process theory. The spatial covariates obtained from the interpolation, through the application of Cubic spline interpolation, are treated as potential covariates in the fitted models and denoted by $Err(\mathbf{u})$, $Rms(\mathbf{u})$ and $D_{ns}(\mathbf{u})$.

It is worth noticing that, for computational reasons, it was not possible to describe the other Italian areas, such as the one corresponding to the Emilia sequence. Indeed, analysing the given point process would lead to both unsatisfactory inter-

---

[1] https://istituto.ingv.it/it/risorse-e-servizi/archivi-e-banche-dati.html

polation and computational problems when fitting local models, due to the strong irregularity of data. Moreover, as already stated by Baddeley (2017), the local composite likelihood works well if kernel smoothing performs well. Dealing with earthquake data, this point can not be guaranteed, since their well-known features of inhomogeneity and clustering in space.

Consequently, we focus on a seismic sequence. The selected region is Abruzzo, sadly well-known for its seismic history. In April 2009, in the mountainous Abruzzo region, a 6.3-magnitude quake killed nearly 300 people and destroyed or seriously damaged thousands of buildings. During a period of three months in 2016, Central Italy was hit by three large earthquakes and numerous aftershocks. The first one, with a magnitude of 6.0, occurred in Amatrice on August 24 and killed 298 people. On October 26 evening, a series of shocks with the strongest one having a magnitude of about 6, took place in approximately the same region (Lutikov et al., 2018). Those sequences are actually missing from the data analysed in this section.

The earthquakes that occurred between May 2012 and May 2016 in Abruzzo are displayed in Figure 3.11 (a), consisting of 85 events with 2.5 as the threshold magnitude. As we may notice, a concentration of events is observed on the top-left of the region, in correspondence with the city of L'Aquila. So, even if the events occurred in 2009 are not analysed in this section, their consequences on the following sequences can be observed.

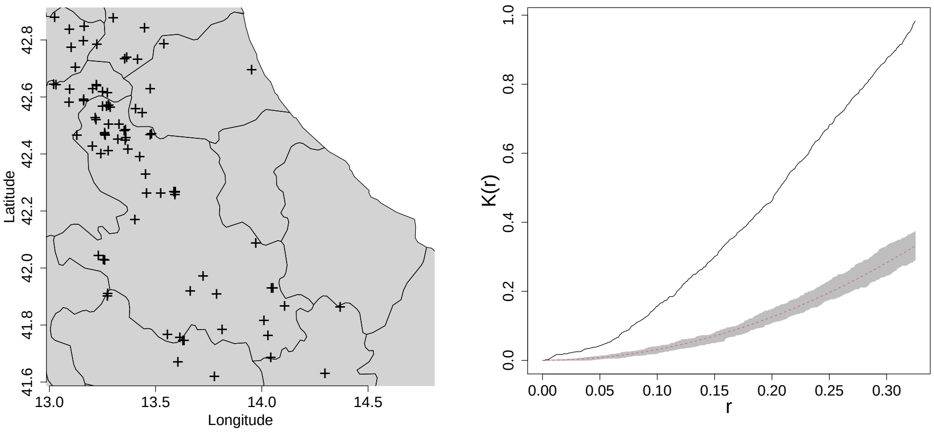


Figure 3.11: Earthquakes occurred in Abruzzo between May 2012 and May 2016 (a), consisting of 85 events with 2.5 as the threshold magnitude; K-function of the observed (b) Italian seismic sequence. The dotted red line represents the values of the K-function for a homogeneous Poisson process, while the black one represents the K-function of the observed point process.

The K-function of the analysed point process is shown in Figure 3.11 (b), highlighting a clear clustered pattern of the events. The Berman's tests are carried out in order to check if these variables influence the intensity of earthquakes. As we may notice from Table 3.6, the variables that seem to influence the spatial intensity are $D_f(\mathbf{u})$, $D_{ns}(\mathbf{u})$, $Err(\mathbf{u})$, while $Rms(\mathbf{u})$ doesn't influence the intensity values significantly.

Table 3.6: Berman's Tests for the available spatial variables in the Italian catalogue

Variable	p value
Distance from the faults	0.002958
Distance from the nearest station	0.0002208
Hypocentral error	0.001353
RMS	0.1441

3.1.3.1 Model proposal

Starting from the Italian data, a local Poisson model is estimated to describe the local features of data, including the available covariates, assuming that they may influence the intensity as a function of the spatial location. After having compared different models through the inhomogeneous K-function and the smoothed raw residuals, the chosen local Poisson model is the following

$$\lambda(\mathbf{u}) = \exp(\theta_0(\mathbf{u}) + \theta_1(\mathbf{u})D_f(\mathbf{u}) + \theta_2(\mathbf{u})D_{ns}(\mathbf{u})) \tag{3.4}$$

where only the Distance from the faults and the Distance from the nearest station are involved. The local coefficients of the model are shown in Figure 3.12 and their summary statistics are reported in Table 3.7. Negative values of the coefficients of both variables are reasonable. Indeed, usually, as the distance from the faults increases, the intensity decreases. Moreover, as the distance from the nearest station recording the event increases, the intensity decreases, as detection of earthquakes is usually more accurate if they occur not far from the network station. Looking at the left panels of Figure 3.12 both variables take different values of their coefficients along the whole study area, though looking at their T-test, they are not significant along the whole area. Indeed, it seems that the area could be divided into two subregions (right panels of Figure 3.12): one, on the top-right, in which the variable affecting the intensity is $D_{ns}(\mathbf{u})$, and one, that is all the remaining area, in which only $D_f(\mathbf{u})$ affects the intensity.

For testing local departure from "homogeneity", the stepwise approach already used for the Greek data is here considered. In order to assess the influence of each

Table 3.7: Summary statistics of the coefficients of the local Poisson model (3.4) *for the Italian data*

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
Intercept	3.17	4.04	4.51	4.52	4.99	5.82
D_{ns}	-0.10	-0.08	-0.06	-0.06	-0.05	-0.03
D_f	-7.78	-5.91	-2.67	-2.90	0.10	3.03

variable on the lack of "homogeneity", the space-varying p-values of the "homogeneity" tests are visually inspected. In Figure 3.13 (a) the p-values for the "homogeneity" test in the Null Model are less than 0.05 in two separate areas, namely the one on the top-left and the one on the right of the window, corresponding to those areas where events appear to be clustered or are missing. Looking at Figure 3.13 (b) we see that including the variable $D_{ns}(\mathbf{u})$ in the model, the "homogeneity" hypothesis is rejected in the middle of the window. Finally, in Figure 3.13 (c), including also $D_f(\mathbf{u})$, the necessity of a local model is more evident. Therefore, as for the Greek data, adding covariates into the model makes the local model more adequate for describing observed data, with respect to a global model with constant regression coefficients.

Table 3.8: Summary statistics of the spatial varying coefficients of the local LGCP model for the Italian data

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
σ^2	0.000026	2.666119	3.455831	3.055668	3.889292	4.343586
α	0.0005174	0.0094810	0.0113980	0.0194697	0.0157996	0.0770704

As already stated in the analysis of the Greek data, we fit a local LGCP model, with the same linear predictor selected from the above local Poisson model.

In Figure 3.14 the estimates of the two additional parameters $\hat{\sigma}(\mathbf{u})$ and $\hat{\alpha}(\mathbf{u})$ are shown, and their summary statistics are reported in Table 3.8. The results are coherent with the estimates of the coefficients $\hat{\sigma}^2$ and $\hat{\alpha}$ of the global LGCP model which are respectively equal to 1.47 and 0.21.

The results indicate that these two parameters have different behaviour on the top-left of the selected window, with respect to the remaining region. This area corresponds to the region where events are more clustered, suggesting, as obvious, that in this area the correlation among points is higher and therefore the corresponding regions should be treated differently from the rest.

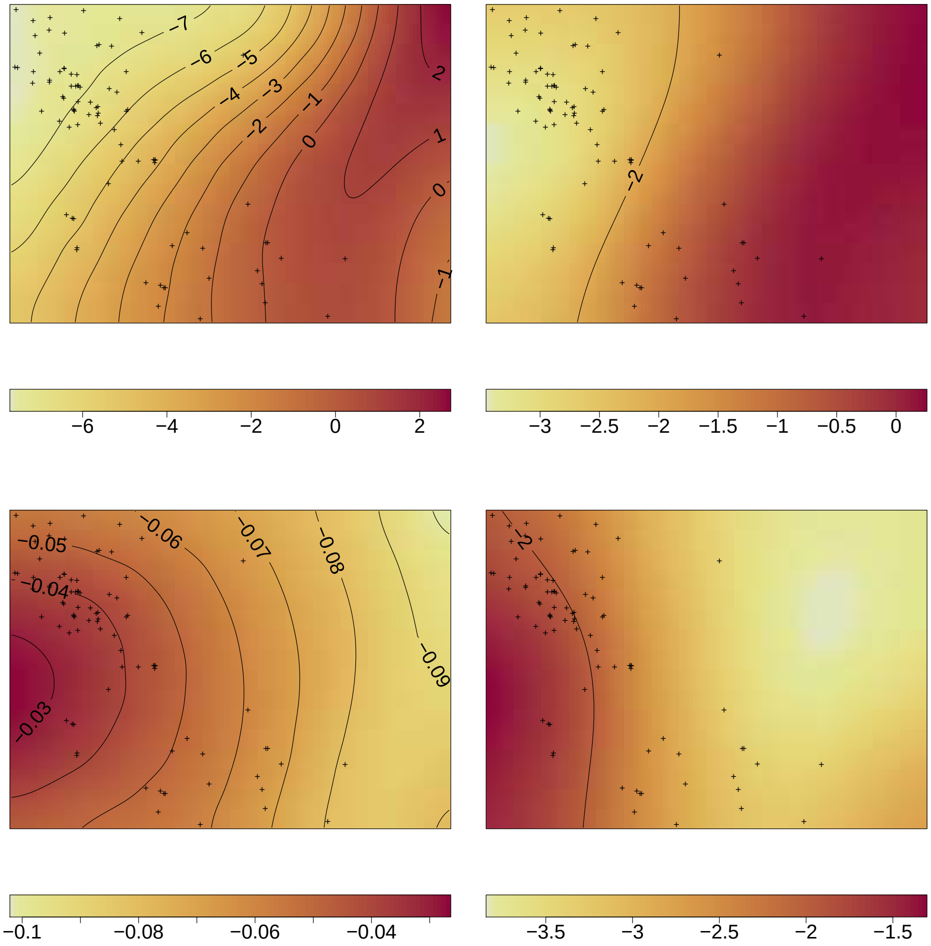


Figure 3.12: Coefficients of the local Poisson model (3.4) *and their T-tests for the Italian data. The coefficients of the Distance from the nearest station are displayed on the top-left panel while the coefficients of the Distance from the faults are displayed on the bottom-left panel*

3.1.4 Conclusions and discussion

In this section, we have introduced models with space-varying parameters, estimated by the composite likelihood for spatial point processes (Baddeley, 2017), for the description of complex earthquake data characterized by multiple and multifractal dependence structures. Analyzing the recent seismicity of a highly seismic active area, local models seem to be suitable for accounting for the complexity of the seismic process since they describe its local features fitting a unique model to the

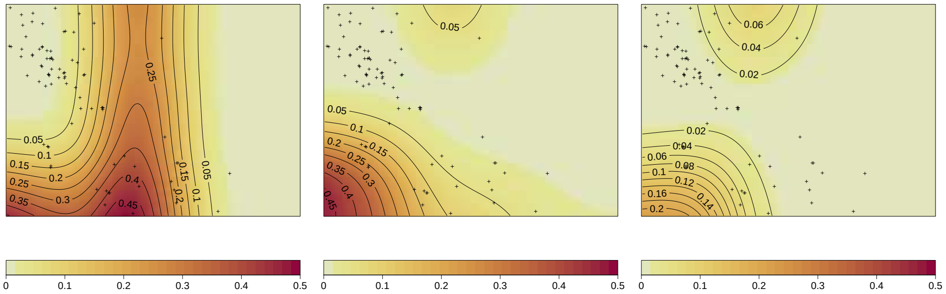


Figure 3.13: Test of "homogeneity" change for the model (3.4) fitted for describing Italian data: p-values of the test of "homogeneity" are plotted, for the Null Model (a) that involves only $\hat{\theta}_0(\boldsymbol{u})$ into the linear predictor, the model in which also the Distance from the nearest station is added (b) and the model in which both the Distance from the nearest station and the Distance from the faults are added into the linear predictor (c)

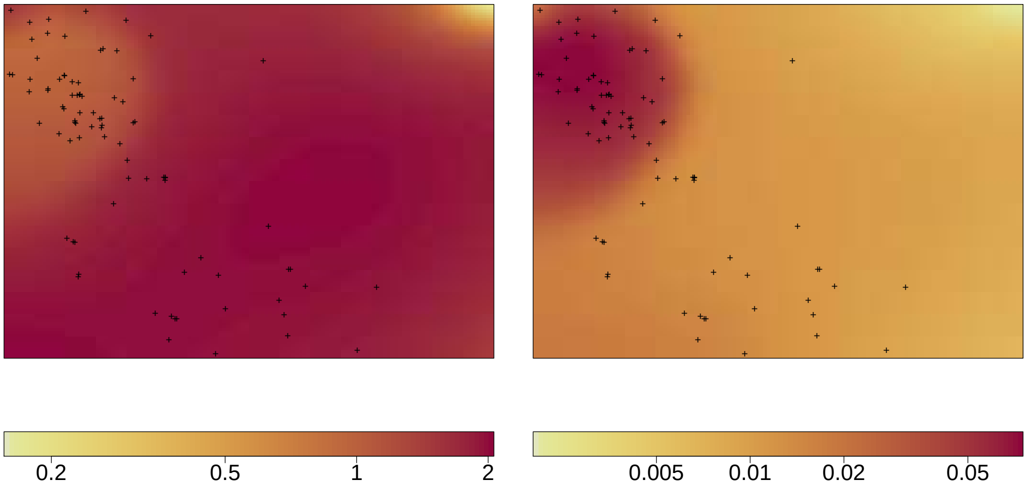


Figure 3.14: Spatial varying additional parameters $\hat{\sigma}(\boldsymbol{u})$ and $\hat{\alpha}(\boldsymbol{u})$ of the local LGCP model fitted for describing the Italian seismic sequence

whole area under study.

First, a log-linear relationship between the earthquakes' intensity and the distances from the nearest seismic sources has been modelled, fitting a local inhomogeneous Poisson process model. From the related obtained results, it seems reasonable to assume local coefficients, that is, location-depending effects of the considered covariates, reflecting the complex seismic activity of the given areas. Furthermore, we have found that although the local models are useful in finding spatially varying covariate effects, they may not be optimal for summarizing these different effects. However, this information could be used in order to split the study region into different areas and separately interpret the estimated local coefficients, following multiple global approaches. Secondly, interaction among points, crucial in the context of the analysis of seismic events, has been taken into account by fitting a local LGCP model.

This represents the first main contribution of the section, as the local log-Gaussian Cox process is comprehensively addressed and applied here for the first time. We use the LGCP for describing a seismic sequence rather than an entire catalogue since there is the possibility to account for the interaction among points by the estimation of the parameters of the underlying Gaussian Random Field. This makes the LGCP particularly suitable for the description of highly inhomogeneous areas, such as the one chosen in the application provided in this section. In addition, by fitting a local version of the LGCP, it becomes possible to account both for the effect of covariates and for the interaction among points as a function of the spatial location. Results of the reported application confirm the crucial role of the proposed approach for describing and characterising the study area through multiple underlying processes, with both the first- and second-order characteristics varying with location.

Furthermore, the second main innovation of this section refers to our contribution to the framework of diagnostics for local models, outlining all the relevant methods for diagnostics and model selection, also proposing a bootstrap procedure for driving the model selection process, and a stepwise procedure, that is able to assess single local effects of covariates on the rejection of the hypothesis of "homogeneity". These procedures have been validated by the application of the proposed methods to real seismic data, showing that, as the number of covariates in the linear predictor increases, the number of areas actually needing a local model increases too. The use of existing diagnostic tools for global models in the context of local fitting and the proposed stepwise approach provide valid results and useful hints for assessing the estimated models and selecting the best one, testing the need for a local model in the analysed area.

Overall, starting from the results of the proposed application, local models provide good inferential results in the seismic point processes context. The complex earthquake phenomenon is characterized by multiple and multifractal dependence structures, and its description demands local tools and models to address its peculiar features. It seems that fitting a model with locally varying parameters is reasonable when analysing both seismic catalogues and sequences. In particular, the local LGCP models succeed more in identifying regions where the events are clustered than the global LGCP models that, through the estimation of constant interaction parameters, can just identify the presence of a global clustered structure.

Moreover, local models provide a good alternative in terms of fitting global models, even non-parametric ones. A further advantage of the application of local models to the seismic context proposed in this section is the possibility of fitting a unique fully parametric model, considering in a proper way the interaction between covariates, peculiar in describing the complex seismic phenomenon, compared to the most classical semi-parametric model proposed in the literature, and reviewed in the introduction of this section. Furthermore, as already pointed out by Baddeley (2017), some methodological problems still remain, such as the need for a theory of misspecified models and for variance estimators for Cox and cluster processes. Earthquake data typically show extremely high concentrated points in small zones of high seismicity. Alternative approaches include the division of the study area into Dirichlet cells, fitting a proper model to the union of such cells (e.g the spatially-adaptive space-time point process models for earthquakes (Ogata and Katsura, 1988; Ogata, 1989)), with the drawback of providing high dimensional models. Spatial clustering techniques may work differently in different tectonic environments. In this section, we have applied the proposed technique to seismicity typical of compressive tectonics and in particular in an area such as the Greek one where, as is well known, a lithospheric subduction process is underway. It is certainly desirable in the future to test the spatial clustering procedure here proposed also to seismicity recorded in areas characterized by extensional or strike-slip tectonics, as well as in areas characterized by seismicity linked to volcanic processes. This would make it possible to verify the performance of the proposed technique in different tectonic regimes and it is possible fine-tuning aimed at optimizing the spatial clustering.

Finally, a relevant drawback of this work is that we have not accounted for the temporal dimension of the seismic events, whose realization depends on their past history, as proved by the existence of aftershocks. Given the results of this section on the application of spatial local models to the seismic context, we believe that these models could be the basis for future development for introducing local models and methods both for spatial and temporal dimensions, referring to the theory of the space-time point processes. Indeed, few classical models widely used in the spatio-temporal seismic context account for external information such as spatial or temporal covariates (a recent proposal for the Epidemic Type Aftershock sequence model is provided by Adelfio and Chiodi (2020)). Therefore, it is interesting to develop spatio-temporal extensions of the spatial local models used in this section, for instance, exploiting the alternative estimation procedure for the local LGCP models with minimum contrast based on local second-order summary statistics, as done in the next section.

3.2 Locally weighted minimum contrast estimation for space-time log-Gaussian Cox processes

D'Angelo, N., Adelfio, G., and Mateu, J. (2022c). Locally weighted minimum contrast estimation for spatio-temporal log-gaussian cox processes. *https://arxiv.org/abs/2209.07153 Computational Statistics & Data Analysis. Major revisions*

Section 3.1 showed that Baddeley's purely local models provide good inferential results by applying them to earthquake data. However, that work did not account for the temporal dimension of the seismic events, whose realisations depend on their past history, as proved by the existence of aftershocks. This leaves a gap in the literature about the availability of local spatio-temporal point process models. The challenge of extending Baddeley's work would be to enrich the literature on this topic. Additionally, it is well documented in recent literature that many spatial point patterns are just snapshots of real spatio-temporal ones, and thus they should be treated as such. Our paper enlarges the statistical tools available for spatio-temporal patterns. Motivated by these reasons, we propose a local version of spatio-temporal LGCPs employing LISTA functions plugged into the minimum contrast procedure to obtain space as well as time-varying parameters. For the parameters of the deterministic part, we follow Baddeley and Turner (2000), using a quadrature scheme for the locally weighted log-linear Poisson regression. For the parameters of the covariance structure, we resort to the joint minimum contrast fitting method proposed by Siino et al. (2018a) to estimate the set of second-order parameters of the spatio-temporal LGCPs. This approach has the advantage of being suitable in both separable and non-separable parametric specifications of the correlation function of the underlying GRF. The advantage of the minimum contrast procedure stands in providing parameters' estimates, avoiding the computational demand needed for maximum likelihood and Bayesian inferential methods. For this reason, it is useful for the rapid exploration of a range of possible models and provides graphical methods for assessing model fit. Moreover, it is quite a flexible approach since it only depends on the knowledge of the involved summary statistics and not on the separability structure of the covariance function.

Davies and Hazelton (2013) firstly evaluated minimum contrast parameter estimation for both the spatial and spatio-temporal LGCPs. However, their estimation setting does not fit to all the contexts and types of observed data (such as when recorded in a continuous time resolution or when presenting a high cluster structure such that counts by units of time have several zeros), but most importantly, it cannot be used if a non-separable correlation structure of the GRF is assumed. This limitation was overcome by the proposal of Siino et al. (2018a). The possibility of

fitting models based on non-separable covariances is crucial when separability cannot be assumed.
We point out that our proposal represents the first attempt to fit a localised version of a log-Gaussian Cox process. Therefore, the main contribution of this paper stands in the novel availability of methods for fitting spatio-temporal point process models involving the estimation of local covariance parameters. Indeed, the interaction among points is typically taken into account by the LGCPs models through the estimation of the covariance parameters of the GRF. In their local version, these parameters vary in space and time, allowing the description and characterisation of the study area through multiple underlying processes, meaning that it would make sense to assume a generating process of the observed point pattern that is not unique. The peculiarity of the new local LGCPs, already exploited in their purely spatial version for the description of both seismic catalogues and sequences, lies in the inhomogeneity of the point process being simultaneously addressed via spatial covariates, a latent error process, and their estimated local coefficients.

We, therefore, introduce the local estimation, obtaining a whole set of parameters for each point of the analysed dataset, and we refer to this new procedure as *locally weighted spatio-temporal minimum contrast*. We show some simulations, finding promising results as the estimates tend to be quite precise on average if compared to the 'global' counterpart while also reflecting the assumed variability in space and time. Moreover, our proposal presents further advantages than global alternatives, since it allows to obtain and interpret local parameters. To enforce these results, we apply the proposed methodology to a real dataset in seismicity, considering the quite convenient separable covariance function of the LGCP, proving that local LGCPs can provide a better fit to the data than their global counterparts. Overall, our proposal poses the basis for many further investigations and applications.

The structure of the section is as follows. In section 3.2.1, the proposed local minimum contrast estimation based on the LISTA functions is presented. Section 3.2.2 reports the diagnostic procedure used to assess the goodness-of-fit of a global LGCP and a proposed modification to deal with local LGCPs. Then, the performance of the proposed local estimation method is assessed in section 3.2.3 through a simulation study. An application to real seismic data comes in Section 3.2.4. Finally, the overall section ends with some conclusions in section 3.2.5.

3.2.1 Model estimation proposal for the local spatio-temporal LGCPs

We first review and extend the statistical theory and computational strategies for local estimation and inference for spatio-temporal Poisson point processes, in section 3.2.1.1. Then, in section 3.2.1.2, we extend the joint minimum contrast procedure reviewed in section 2.2.3.1.1, to the local context.

3.2.1.1 Estimating the first-order intensity function through local likelihood

We assume that the template model is a Poisson process, with a parametric intensity or rate function

$$\lambda(\mathbf{u}, t; \theta), \quad X \in W, \quad t \in T, \quad \theta \in \Theta.$$

The log-likelihood is

$$\log L(\theta) = \sum_i \lambda(\mathbf{u}_i, t_i; \theta) - \int_W \int_T \lambda(\mathbf{u}, t; \theta) \mathrm{d}t \mathrm{d}u \tag{3.5}$$

up to an additive constant, where the sum is over all points $\mathbf{u}_i$ in X. We often consider intensity models of log-linear form

$$\lambda(\mathbf{u}, t; \theta) = \exp(\theta Z(\mathbf{u}, t) + B(\mathbf{u}, t)), \quad \mathbf{u} \in W, \quad t \in T \tag{3.6}$$

where $Z(\mathbf{u}, t)$ is a vector-valued covariate function, and $B(\mathbf{u}, t)$ is a scalar offset.

3.2.1.1.1 Quadrature scheme Following Berman and Turner (1992), we use a finite quadrature approximation to the log-likelihood, as implemented in the `spatstat` package (Baddeley and Turner, 2005). Renaming the data points as $\mathbf{x}_1, \ldots, \mathbf{x}_n$ with $(\mathbf{u}_i, t_i) = \mathbf{x}_i$ for $i = 1, \ldots, n$, then generate m additional "dummy points" $(\mathbf{u}_{n+1}, t_{n+1}) \ldots, (\mathbf{u}_{m+n}, t_{m+n})$ to form a set of $n + m$ quadrature points (where $m > n$). Then we determine quadrature weights $a_1, \ldots, a_m$ so that integrals in (3.5) can be approximated by a Riemann sum

$$\int_W \int_T \lambda(\mathbf{u}, t; \theta) \mathrm{d}t \mathrm{d}u \approx \sum_{k=1}^{n+m} a_k \lambda(\mathbf{u}_k, t_k; \theta) \tag{3.7}$$

where a_k are the quadrature weights such that $\sum_{k=1}^{n+m} a_k = l(W \times T)$ where l is the Lebesgue measure.

Then the log-likelihood (3.5) of the template model can be approximated by

$$\begin{aligned} \log L(\theta) &\approx \sum_i \log \lambda(\mathbf{x}_i; \theta) + \sum_j (1 - \lambda(\mathbf{u}_j, t_j; \theta)) a_j = \\ &= \sum_j e_j \log \lambda(\mathbf{u}_j, t_j; \theta) + (1 - \lambda(\mathbf{u}_j, t_j; \theta)) a_j \end{aligned}$$

where $e_j = 1\{j \leq n\}$ is the indicator that equals 1 if u_j is a data point. Writing $y_j = e_j / a_j$ this becomes

$$\log L(\theta) \approx \sum_j a_j (y_j \log \lambda(\mathbf{u}_j, t_j; \theta) - \lambda(\mathbf{u}_j, t_j; \theta)) + \sum_j a_j. \tag{3.8}$$

Apart from the constant $\sum_j a_j$, this expression is formally equivalent to the weighted log-likelihood of a Poisson regression model with responses y_j and means $\lambda(\mathbf{u}_j, t_j; \theta) = \exp(\theta Z(\mathbf{u}_j, t_j) + B(\mathbf{u}_j, t_j))$. This can be maximised using standard GLM software. For more details see Berman and Turner (1992); Baddeley et al. (2000, 2005) section 9.8.

We define the spatio-temporal quadrature scheme by defying a spatio-temporal partition of $W \times T$ into cubes C_k of equal volume ν, assigning the weight $a_k = \nu/n_k$ to each quadrature point (dummy or data) where n_k is the number of points that lie in the same cube as the point u_k (Raeisi et al., 2021).

The number of dummy points should be sufficient for an accurate estimate of the likelihood. Following Baddeley et al. (2000) and Raeisi et al. (2021), we start with a number of dummy points $m \approx 4n$, increasing it until $\sum_k a_k = l(W \times T)$.

3.2.1.1.2 Local Poisson models' estimation in space and time The local log-likelihood associated with the spatio-temporal location $(\mathbf{v}, s)$ is given by

$$\begin{aligned} \log L((\mathbf{v}, s); \theta) = & \sum_i w_{\sigma_s}(\mathbf{u}_i - \mathbf{v}) w_{\sigma_t}(t_i - s) \lambda(\mathbf{u}_i, t_i; \theta) + \\ & - \int_W \int_T \lambda(\mathbf{u}, t; \theta) w_{\sigma_s}(\mathbf{u}_i - \mathbf{v}) w_{\sigma_t}(t_i - s) \mathrm{d}t \mathrm{d}u \end{aligned} \tag{3.9}$$

where w_{σ_s} and w_{σ_t} are weight functions, and $\sigma_s, \sigma_t > 0$ are the smoothing bandwidths. It is not necessary to assume that w_{σ_s} and w_{σ_t} are probability densities. For simplicity, we shall consider only kernels of fixed bandwidth, even though spatially adaptive kernels could also be used. Note that if the template model is the homogeneous Poisson process with intensity λ, then the local likelihood estimate $\hat{\lambda}(\mathbf{v}, s)$ reduces to the kernel estimator of the point process intensity (Diggle, 2013) with kernel proportional to $w_{\sigma_s} w_{\sigma_t}$.

We now use in (3.9) a similar approximation as in (3.8) for the local log-likelihood associated with each desired location $(\mathbf{v}, s) \in W \times T$

$$\log L((\mathbf{v}, s); \theta) \approx \sum_j w_j(\mathbf{v}, s) a_j (y_j \log \lambda(\mathbf{u}_j, t_j; \theta) - \lambda(\mathbf{u}_j, t_j; \theta)) + \sum_j w_j(\mathbf{v}, s) a_j, \tag{3.10}$$

where $w_j(\mathbf{v}, s) = w_{\sigma_s}(\mathbf{v} - \mathbf{u}_j) w_{\sigma_t}(s - t_j)$.

Basically, for each desired location $(\mathbf{v}, s)$, we replace the vector of quadrature weights a_j by $a_j(\mathbf{v}, s) = w_j(\mathbf{v}, s) a_j$ where $w_j(\mathbf{v}, s) = w_{\sigma_s}(\mathbf{v} - \mathbf{u}_j) w_{\sigma_t}(s - t_j)$, and use the GLM software to fit the Poisson regression. The local likelihood is defined at any location $(\mathbf{v}, s)$ in continuous space. In practice, it will be enough to consider a grid of points $(\mathbf{v}, s)$. The choice of the grid depends on the computing resources, the computations required at each location, and the spatial resolution required.

Bandwidth selection is always an unavoidable topic with kernel estimation. In principle, the bandwidth should be selected according to data resolution, which is at the order of 1-10 times the nearest neighbouring distance (Zhuang, 2020). This general consideration holds for both the spatial and temporal components. A natural first step in analysing spatial patterns is to form a kernel density estimate (Silverman, 2018) of the spatially-varying events rate. The smoothing parameter for the kernel estimator of the temporal intensity function is usually selected by the bandwidth of Silverman's rule of thumb. However, the simple kernel estimate with a fixed bandwidth has a serious disadvantage, especially for spatial data. Indeed, for a (spatially) clustered point dataset, a small bandwidth gives a noisy estimate for the sparsely populated area, whereas a large bandwidth mixes up the boundaries between the densely populated and sparsely populated areas. Therefore, instead of the kernel estimates w with fixed bandwidth σ, other proposals in the literature include variable bandwidth by $\sigma_j = max\{\epsilon, \inf(r : N[B(\mathbf{u}_j; r)] > n_p)\}$, where ϵ is a small number, $B(\mathbf{u}_j; r)$ is the disk centered at $\mathbf{u}_j$ with a radius of r, and n_p is a positive integer, i.e. σ_j is the distance to n_p-th closest event. In other words, once a suitable integer between 10 and 100 for the parameter n_p is chosen, we calculate a bandwidth value σ_j, of each event j, as the radius of the smallest circle centred at the location of the jth event that includes at least n_p other events. See Silverman (2018) for similar locally dependent estimates, Zhuang et al. (2002) for their use in a seismic context, and Musmeci and Vere-Jones (1986); Choi and Hall (1999) for similar ideas.

3.2.1.2 Proposal: Locally weighted spatio-temporal minimum contrast

In the purely spatial context, a localised version of minimum contrast is developed using the local K-functions or local pair correlation functions by Baddeley (2017), bearing a very close resemblance to the local Palm likelihood approach, whose implementation is provided by the function `locmincon()` of the R package `spatstat.local` (Baddeley, 2019).

Combining the *joint minimum contrast* (Siino et al., 2018a) and the *local minimum contrast* (Baddeley, 2017) procedures, we can obtain a vector of parameters $\boldsymbol{\psi}_i$ for each point i, by minimising

$$M_{J,i}\{\boldsymbol{\psi}_i\} = \int_{h_0}^{h_{max}} \int_{r_0}^{r_{max}} \phi(r,h)\{\nu[\bar{J}_i(r,h)] - \nu[J(r,h;\boldsymbol{\psi})]\}^2 \mathrm{d}r\mathrm{d}h, \tag{3.11}$$

where $\bar{J}_i(r,h)$ is the average of the local functions $\hat{J}_i(r,h)$, weighted by some pointwise kernel estimates. This procedure not only provides individual estimates, but it does also account for the vicinity of the observed points, and therefore the contribution of their displacement on the estimation procedure. This conceptually resembles the methodology used for the local log-likelihood in section (3.2.1.1.2). Following Siino et al. (2018a), we suggest using $\phi(r,h) = 1$ and ν as the identity function. Then, r_{max} and h_{max} are selected as $1/4$ of the maximum observable spatial and temporal distances.

Thus, consider again the weights $w_i(\mathbf{v}_i, s_i) = w_{\sigma_s}(\mathbf{v}_j - \mathbf{u}_i) w_{\sigma_t}(s_j - t_i)$ given by some kernel estimates. The same considerations hold for the choice of the bandwidth in the local log-likelihood. Then, the averaged weighted local statistics $\bar{J}_i(r, h)$ in (3.11), for each point i, is

$$\bar{J}_i(r,h) = \frac{\sum_{j=1}^{n} \hat{J}_j(r,h) w_{ij}}{\sum_{j=1}^{n} w_{ij}} \qquad i, j = 1, \dots, n.$$

In particular, we consider $\hat{J}_i(\cdot)$ as the local spatio-temporal pair correlation function (Gabriel et al., 2021), evaluated for each ith event $(\mathbf{u}_i, t_i)$,

$$\hat{J}_i(r,h) = \hat{g}_i(r,h) = \frac{1}{4\pi r |W \times T| \hat{\lambda}^2} \sum_{j \neq i} \frac{\kappa_{s,t}(||\mathbf{u}_i - \mathbf{u}_j|| - r, |t_i - t_j| - h)}{\omega(\mathbf{u}_i, \mathbf{u}_j)\omega(t_i, t_j)} \tag{3.12}$$

where ω is the edge correction factor.

The kernel function κ has a multiplicative form $\kappa_{s,t}(||\mathbf{u}_i - \mathbf{u}_j|| - r, |t_i - t_j| - h) = \kappa_s(||\mathbf{u}_i - \mathbf{u}_j|| - r)\kappa_t(|t_i - t_j| - h)$ where κ_s and κ_t are kernel functions with bandwidths ϵ_s and ϵ_t, respectively. Both of them are computed using the Epanechnikov kernel (Illian et al., 2008) and the bandwidths are estimated with a direct plug-in method (Sheather and Jones, 1991) using the `dpik` function of the package `Kernsmooth` (Wand, 2020).

3.2.2 Diagnostics of the local spatio-temporal LGCPs

To deal with local LGCPs, we provide a slightly modified diagnostics method. Indeed, given the local covariance parameters, diagnostics can be carried out following the same procedure outlined above, but simulating a *local* GRF from the estimates of the fitted model. This will reflect the variability of the estimated covariance parameters. From the local GRF, we can then generate a local spatio-temporal LGCP. The algorithm is basically the same as the one in section 2.2.3.1.2, except that we substitute step 1 with the following additional steps. The rest remains the same.

1.a Generate a realisation from a GRF $S(\mathbf{u}, t)$, with covariance function $\mathbb{C}((\mathbf{u}, t), (\mathbf{v}, s))$ and mean function $\mu(\mathbf{u}, t)$;

1.b Define a spatio-temporal grid, whose breaks are evenly spaced (the definition of the grid depends on the level of detail to be given), for the simulation of the *local* GRF, and for each point $(\mathbf{u}_i, t_i)$, obtain a GRF $S(\mathbf{u}_i, t_i)$ simulated from the individual set of estimates $\boldsymbol{\psi}_i$;

1.c Compute the average of the GRF values obtained in 1.a within the sub-grids defined in 1.b;

1.d Fill the empty sub-grids with the global GRF values;

The procedure 1.a - 1.d ends obtaining a *local* GRF $S'(\mathbf{u}, t)$, for which the information of the local estimates is exploited.

3.2.3 Simulation study

A simulation study with a number of scenarios is carried out to assess and compare the performances of the global estimation method Siino et al. (2018a) and the local one proposed in section 3.2.1.2.

In the first set of scenarios, we assume a stationary and isotropic LGCP with a separable structure of the covariance of the underlying GRF, with an exponential model both in space and time as in Equation (2.24), the vector of parameters given by $\boldsymbol{\psi} = \{\sigma^2, \alpha, \beta\}$. For each scenario, 200 point patterns are generated with $n = 1000$ expected number of points in the spatio-temporal window $W \times T = [0,1]^2 \times [0,50]$, with a constant first-order intensity equal to $b = \log(n/|W \times T|)$. We consider several degrees of clustering in the process with variance $\sigma^2 = \{5,8\}$ and scale parameters in space and time, $\alpha = \{0.005, 0.10, 0.25\}$ and $\beta = \{2,5,10\}$. The mean of the GRF is fixed $\mu = -0.5\sigma^2$. These sets of parameters are the same used in the simulation study in Siino et al. (2018a). Table 3.9 contains mean and quartiles of the distributions of the estimated local parameters $\hat{\boldsymbol{\psi}}_i = \{\hat{\sigma}_i^2, \hat{\alpha}_i, \hat{\beta}_i\}$, averaged over the 200 simulated point patterns.

The results obtained are quite promising: indeed, even considering fixed bandwidths for the weights in the proposed locally weighted minimum contrast, the procedure manages to provide quite precise estimates. This is particularly evident if compared to the results in Siino et al. (2018a) and, even before, in Davies and Hazelton (2013), where the authors provide a number of simulations studies to assess the overall performance of the minimum contrast procedure under different aspects, concluding that the estimates of the variance σ^2 strongly tend to be underestimated. However, our main goal here is not to provide an alternative to the classical global minimum contrast procedure, but instead to estimate local parameters. This objective is clearly achieved as we manage to obtain a whole distribution for each parameter (of each analysed point pattern). Overall, we do not aim at achieving the most precise result (i.e. MSE close to zero), but with some degree of variability, that is to say, with an MSE not largely exceeding the mean value of its estimates' distribution.

The same results hold for the scenarios simulated from 200 LGCPs with $n = 1000$ points each and the non-separable covariance function in (2.25), as evident from Tables 3.9 and 3.10. Note that we have chosen $\delta = \{0.3, 1.8\}$, and kept γ_t and γ_s equal to 1 in all the scenarios.

3.2.4 Application to the Greek seismic data

We analyse the same seismic data in section 3.1, related to 1111 earthquakes that occurred in Greece between 2005 and 2014, coming from the Hellenic Unified Seismic Network (H.U.S.N.). Time has been converted into days, and only seismic events with a magnitude larger than 4 are considered in this study. The earthquakes are

Table 3.9: Mean (m) and quartiles of the distributions of the local parameters $\hat{\boldsymbol{\psi}}_i = \{\hat{\sigma}_i^2, \hat{\alpha}_i, \hat{\beta}_i\}$, averaged over the 200 simulated point patterns generated assuming an exponential form in both the spatial and temporal dimensions for the GFR with covariance as in (2.24).

True			$\hat{\sigma}^2$				$\hat{\alpha}$				$\hat{\beta}$			
σ^2	α	β	25%	50%	m (mse)	75%	25%	50%	m (mse)	75%	25%	50%	m (mse)	75%
5	0.05	2	5.27	6.30	6.45(2.38)	7.60	0.05	0.07	0.14(5.26)	0.09	1.77	2.26	2.63(5.19)	2.97
	0.10		4.64	5.51	5.67(2.56)	6.47	0.09	0.11	0.13(0.91)	0.15	1.80	2.32	2.61(4.80)	3.08
	0.25		3.68	4.39	4.63(3.11)	5.37	0.19	0.24	0.34(4.47)	0.32	1.69	2.22	2.50(5.39)	2.92
	0.05	5	4.36	5.43	5.54(2.64)	6.58	0.05	0.07	0.12(4.02)	0.09	3.36	4.43	5.03(4.74)	5.93
	0.10		4.13	4.96	5.14(2.95)	5.97	0.09	0.11	0.14(1.93)	0.15	3.40	4.45	5.09(4.61)	6.17
	0.25		3.29	4.10	4.27(4.20)	5.02	0.17	0.24	0.40(6.56)	0.34	3.04	4.21	4.89(6.81)	5.90
	0.05	10	4.08	5.03	5.20(3.09)	6.12	0.05	0.06	0.10(4.07)	0.08	5.69	7.85	8.61(7.75)	10.54
	0.10		3.66	4.44	4.66(4.19)	5.50	0.08	0.11	0.16(4.23)	0.14	5.64	8.00	8.85(8.58)	11.07
	0.25		3.05	3.73	3.97(4.63)	4.70	0.16	0.22	0.35(6.07)	0.30	5.15	7.09	8.15(11.02)	9.98
8	0.05	2	7.26	8.23	8.29(3.05)	9.37	0.05	0.06	0.07(1.47)	0.08	2.27	2.85	3.36(4.77)	3.87
	0.10		6.32	7.26	7.40(2.80)	8.36	0.08	0.10	0.12(2.39)	0.13	2.25	2.84	3.17(4.53)	3.72
	0.25		5.07	5.97	6.16(3.30)	7.13	0.17	0.22	0.32(5.51)	0.29	1.97	2.62	2.83(5.81)	3.43
	0.05	5	6.72	7.64	7.76(2.96)	8.83	0.05	0.06	0.08(2.56)	0.08	3.35	4.43	5.05(5.17)	6.11
	0.10		5.73	6.74	6.91(3.05)	7.99	0.08	0.10	0.13(3.16)	0.13	3.29	4.29	4.82(5.58)	5.81
	0.25		4.79	5.61	5.81(3.39)	6.69	0.15	0.19	0.29(5.84)	0.26	3.01	4.17	4.58(5.88)	5.59
	0.05	10	6.11	7.06	7.14(2.96)	8.14	0.05	0.06	0.08(2.32)	0.08	5.37	7.50	8.30(7.99)	10.22
	0.10		5.15	6.19	6.33(3.35)	7.24	0.08	0.10	0.12(2.68)	0.12	4.93	7.04	7.88(8.25)	9.84
	0.25		4.19	5.05	5.23(3.71)	6.19	0.14	0.19	0.28(5.45)	0.26	4.57	6.51	7.60(9.81)	9.54

Table 3.10: Mean (m) and quartiles of the distributions of the local parameters $\hat{\boldsymbol{\psi}}_i = \{\hat{\sigma}_i^2, \hat{\alpha}_i, \hat{\beta}_i\}$, averaged over the 200 simulated point patterns generated assuming an exponential form in both the spatial and temporal dimensions for the GFR with covariance as in (2.25).

True				$\hat{\sigma}^2$				$\hat{\alpha}$				$\hat{\beta}$				$\hat{\delta}$			
σ^2	α	β	δ	25%	50%	m (mse)	75%	25%	50%	m (mse)	75%	25%	50%	m (mse)	75%	25%	50%	m (mse)	75%
5	0.05	2	1.80	4.99	7.25	9.60(6.69)	13.94	0.03	0.05	0.12(0.51)	0.07	0.16	0.38	0.71(3.33)	0.96	1.01	1.84	1.49(1.07)	2.00
	0.10	5		2.89	3.77	3.79(3.30)	4.62	0.05	0.07	0.09(0.47)	0.10	4.99	5.55	7.22(6.32)	7.78	0.01	0.02	0.23(1.21)	0.11
	0.05	2	0.30	4.23	4.87	4.67(2.66)	4.99	0.03	0.04	0.05(0.30)	0.05	1.95	2.01	4.23(5.85)	2.43	0.01	0.02	0.08(1.24)	0.09
	0.10	5		3.16	4.03	3.86(3.22)	4.55	0.04	0.06	0.10(0.55)	0.09	4.96	5.05	6.93(6.26)	5.55	0.01	0.02	0.10(1.23)	0.09
8	0.05	2	1.80	5.07	6.89	6.50(2.38)	7.82	0.02	0.03	0.05(0.32)	0.05	1.90	2.89	4.01(4.36)	4.44	0.01	0.01	0.16(1.24)	0.06
	0.10	5		3.77	4.48	4.88(2.76)	5.64	0.04	0.06	0.08(0.51)	0.08	3.29	4.94	6.42(6.07)	7.76	0.01	0.01	0.20(1.22)	0.09
	0.05	2	0.30	5.31	6.70	6.36(2.12)	7.50	0.02	0.03	0.04(0.16)	0.05	2.17	2.97	4.78(5.00)	5.14	0.01	0.01	0.13(1.23)	0.05
	0.10	5		4.94	6.29	5.99(2.23)	7.25	0.03	0.05	0.06(0.18)	0.07	5.00	5.18	6.55(5.30)	6.09	0.01	0.03	0.15(1.20)	0.13

depicted in Figure 3.15(a), while Figure 3.15(b) shows the pair correlation function, the global version of (3.12) (Gabriel et al., 2021).

The bandwidths ϵ_s and ϵ_t, used in kernel estimation within the pair correlation function, are computed using the direct plug-in method of Sheather and Jones (1991) as 0.15 and 28.49, for space and time, respectively. Recall that the pair correlation function under complete randomness is a constant equal to one. Figure 3.15(b) shows that the empirical pair correlation function takes values much larger than one for short distances indicating a tendency to cluster far from a Poisson process.

Indeed we can also observe such clustered structures in Figure 3.15(a).

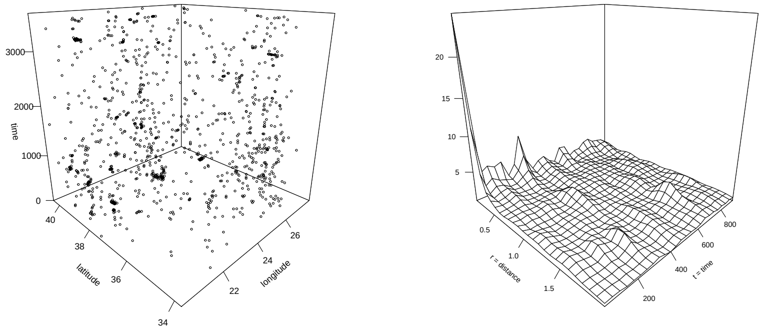


Figure 3.15: Earthquakes occurred in Greece between 2005 and 2014 (a); Pair correlation function of the observed point pattern (b).

We first consider a global LGCP with exponential covariance as in (2.24) using a joint estimation method as explained in section 2.2.3.1.1. The employed spatial and temporal bandwidths ϵ_s and ϵ_t are the ones used for the global pair correlation function displayed in Figure 3.15. We assume a constant intensity function estimated as $\hat{\lambda}(\mathbf{u}, t) = \hat{\lambda} = \frac{n}{|W \times T|} = 0.006$. The estimates of the covariance parameters are $\hat{\boldsymbol{\psi}} = \{\hat{\sigma}^2, \hat{\alpha}, \hat{\beta}\} = \{6.14, 0.27, 449.55\}$, which indicate a quite clustered underlying process, given the high estimated variance, and the relatively small scale parameters. We have also fitted other global LGCPs with non-separable covariances on the analysed data, but these provided unsatisfactory parameter estimates. Therefore, we decided not to show them in the text, discussing the separable case only.

To assess the goodness-of-fit of the proposed (global) model, a residual analysis is carried out by means of the Monte Carlo test based on the inhomogeneous K-functions outlined in section 3.2.2. Figure 3.16(a) displays an example of an LGCP point pattern simulated with the estimated parameters $\hat{\boldsymbol{\psi}}$, which is used to compute the envelopes for the test. Then, Figure 3.16(b) shows the empirical K-function for the earthquake data, following (2.21), as well as the envelopes computed on 39 simulations. An overall p-value of 0 is obtained from (2.30): being equal to zero indicates that the empirical pattern is not compatible with the simulated ones (that come from an LGCP with estimated parameters $\hat{\boldsymbol{\psi}}$). Therefore, the assumed model is not a good fit and an alternative should be searched, able to take into account the residual clustered behaviour of the points. Furthermore, Figure 3.16(b) confirms this result, as the observed K-function does not lie within the envelopes.

We then consider a local spatio-temporal LGCP for the seismic data and follow the *locally weighted minimum contrast* proposed in section 3.2.1.2. The bandwidths

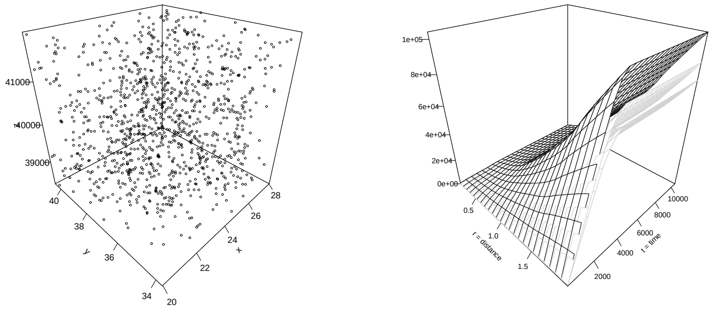


Figure 3.16: (a) An example of LGCP point pattern simulated with the estimated parameters $\hat{\boldsymbol{\psi}}$*; (b) In black: the estimated weighted K-function for the Greek seismic data. In light grey: envelopes based on* 39 *simulations from the global spatio-temporal LGCP at a significance level of 0.05.*

ϵ_s and ϵ_t, for the kernel in the pair correlation function (3.12), are the same used for the global fitting, while the bandwidths $\{\sigma_x, \sigma_y, \sigma_t\}$ for the local weighting are estimated by a Gaussian kernel with Silverman's bandwidth as $2.58, 1.79$ and 1140.31, for x, y, and t coordinates, respectively. The results of the local fitting are shown in Table 3.11 and in Figure 3.17.

Table 3.11: Summary statistics of the estimated covariance parameters for the local LGCPs fitted to the Greek seismic data.

	$\hat{\sigma}^2$	$\hat{\alpha}$	$\hat{\beta}$
Min.	1.28	0.13	42.34
1st Qu.	5.19	0.25	366.43
Median	6.03	0.28	534.86
Mean	5.77	0.39	535.52
3rd Qu.	6.60	0.35	597.59
Max.	12.47	1.94	1411.45

As shown in Table 3.11, the proposed procedure provides a whole distribution for each parameter. Then, as evident from Figure 3.17, the estimated parameters allow to clearly distinguish different sub-areas where points behave differently from each other. Indeed, by inspection of the left panels of Figure 3.17, we can clearly spot two main spatial regions: the one on the top and on the left with higher $\hat{\sigma}_i^2$ values but lower estimates for $\hat{\alpha}_i$ and $\hat{\beta}_i$, and the area on the bottom-right which displays the opposite situation (low variance but large scale parameters).

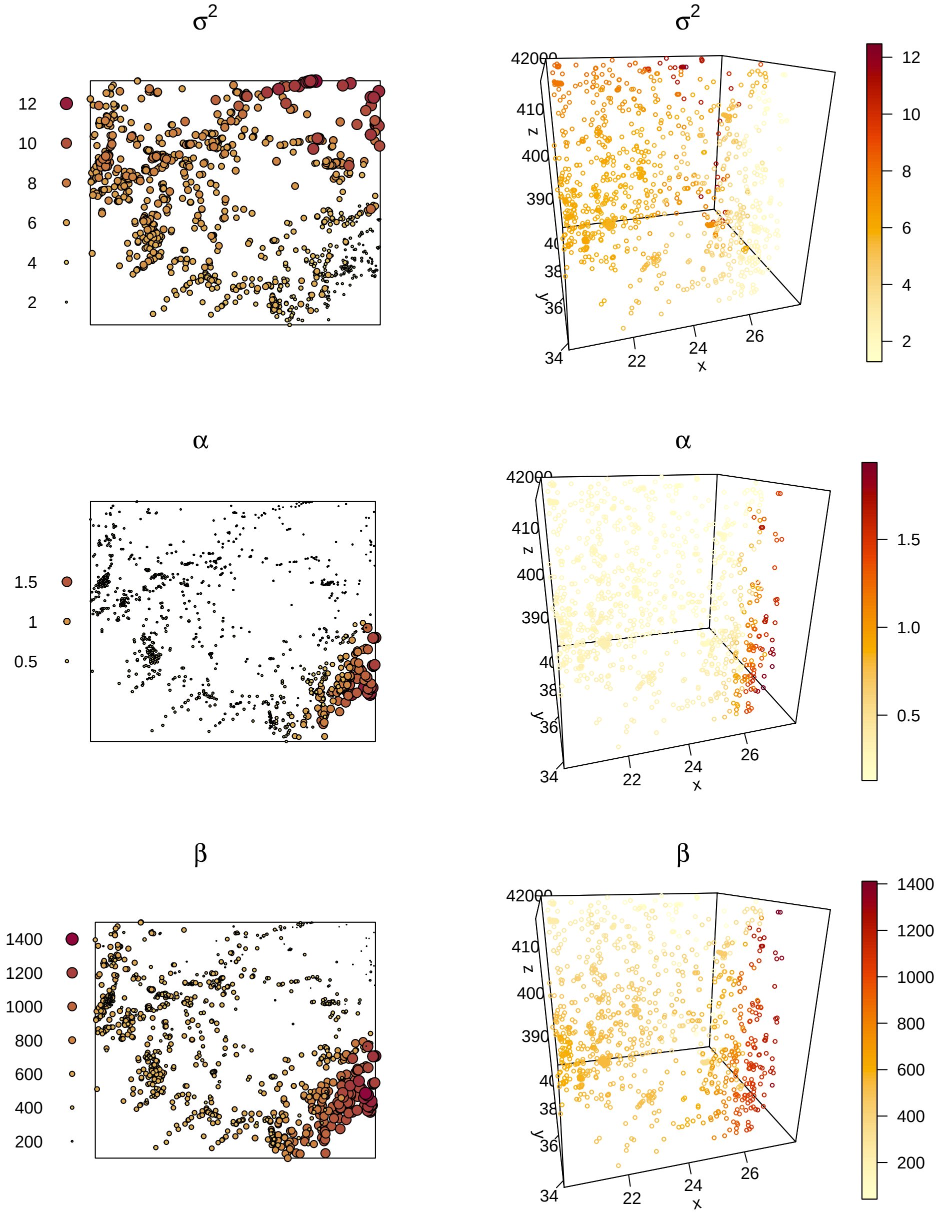


Figure 3.17: Local estimates of the local LGCP fitted to the Greek seismic data. We show a spatial representation on the left and a spatio-temporal one on the right.

This result is particularly appealing because it gives us more insight into the local behaviour of the seismic phenomenon. Indeed, from the global estimates, we are only able to draw the conclusion that the analysed process is overall clustered. Conversely, the local application shows that earthquakes that occurred on the top-left of the analysed area are grouped in smaller clusters, while the ones on the bottom-left have a way more diffuse clustering behaviour.

Furthermore, by looking at the right panels of Figure 3.17, we can assess also the time-varying behaviour of the estimates and therefore on the underlying process: the most relevant result concerns the variability of the covariance parameters $\hat{\sigma}_i$ in time, indicating that the number of earthquakes tends to increase in time, and in particular, in the top-left spatial region.

To assess whether the local model is a better fit to the data, also compared to the global counterpart, we proceed with the residual analysis as outlined in section 3.2.2. For the local LGCP, the procedure is based on the modification of the GRF to be used to generate the point patterns, allowing the inclusion of information on the local estimates. To show the difference between the global and local GRFs, Figure 3.18 displays the spatial GRFs for three temporal instants, global and local, in the upper and lower panels, respectively. As evident, the local GRFs contain information about the local estimates, which in turn result in local variations.

Finally, Figure 3.19(a) displays an example of a point pattern simulated with the estimated local parameters $\hat{\boldsymbol{\psi}}_i$, while Figure 3.19(b) depicts the estimated K-function and the envelopes computed from 39 local simulations. It is worth noticing how the point pattern simulated from the local parameters, following our procedure, better mimics the spatio-temporal arrangement of the original points (Figure 3.15(a)), if compared to the globally simulated point patterns (Figure 3.16(a)). An overall p-value of 0.28 is obtained using the Monte Carlo local test, indicating that the patterns simulated from the assumed model (with the local estimated parameters) do not present any residual clustering behaviour. Therefore, we can conclude that the local model is a good fit to the analysed data. This result is further confirmed by Figure 3.19 (b), as now the observed K-function lies almost entirely between the envelopes obtained from the local simulation. Moreover, the local model also represents a better fit than its previously fitted global counterpart.

3.2.5 Conclusions

We have introduced a novel local fitting procedure for obtaining space-time local parameters for a log-Gaussian Cox process. From a methodological point of view, we have resorted to the joint minimum contrast procedure (which is appealing for its flexibility in dealing also with non-separable covariances), extending it to the local context and therefore allowing to obtain a whole set of covariance parameters for each point of the analysed process. The motivating problem came from the seismic application, where of course it is of interest to study the characteristics of the process in relation to both the spatial and the temporal occurrence of points.

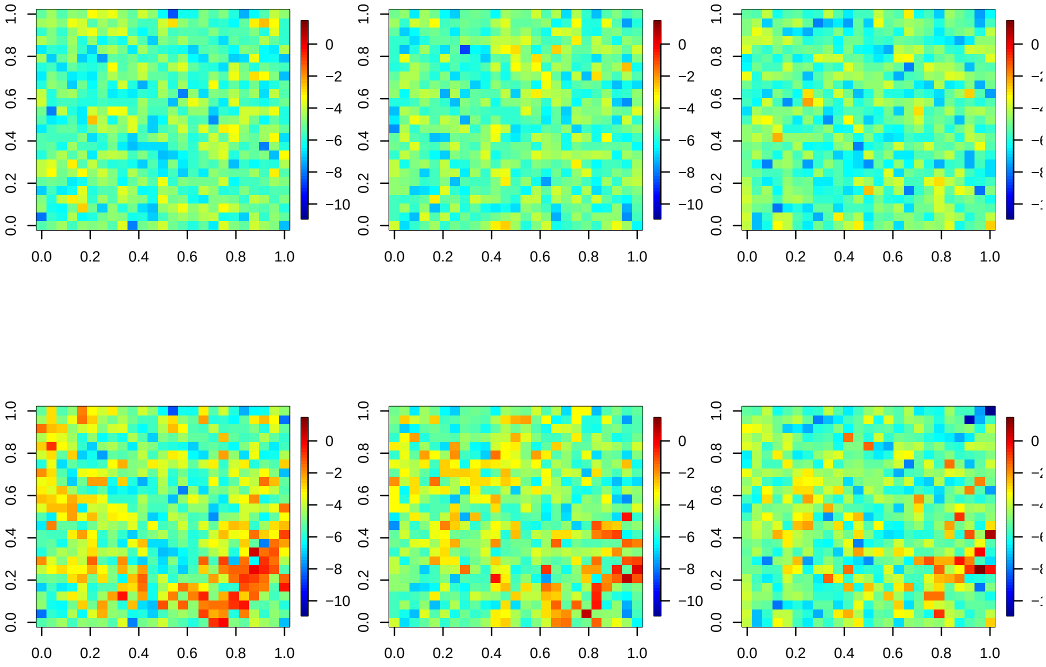


Figure 3.18: GRFs for three temporal instants: Top panels: global; Bottom panels: local.

By simulations, we have shown that the local proposal provides good estimates on average if compared to the global fitting alternatives. Focusing on the application to real seismic data we have been able to assess that a local LGCP can be a better fit to the data if compared to its global counterpart. This is an expected result as the local fitting is based on a weighting given by a non-parametric estimate. However, our proposal presents further advantages than non-parametric alternatives, as it allows to obtain and interpret local parameters.

Our proposal poses the basis for many further investigations and applications. Indeed, in the future, the proposed methodology could be applied to different real spatio-temporal point patterns, where it is of interest to study the characteristics of the underlying process, in relation to the spatial displacement and the temporal occurrence of points. Some examples may include seismology, forestry, criminology, or epidemiology.

Concerning the local LGCPs, it would be interesting to develop software to include external covariates into the first-order intensity and to propose diagnostic

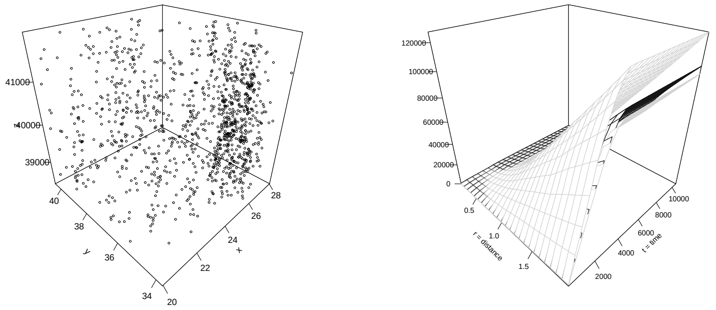


Figure 3.19: (a) A point pattern simulated with the estimated parameters $\hat{\boldsymbol{\psi}}_i$*; (b) In black: the estimated inhomogeneous* K*-function for the Greek seismic data. In light grey: envelopes based on 39 simulations from the local spatio-temporal LGCP at a significance level of 0.05.*

methods for the particular case of multiple covariates, as an extension to section 3.1. Indeed, few spatio-temporal point process models account for external covariates: see Adelfio and Chiodi (2020) for the ETAS model, and section 4.2 for the spatio-temporal Hawkes point process model adapted to events living on linear networks. However, to the best of our knowledge, none of the recent proposals provides local estimates of the model regression-type parameters.

The inclusion of covariates would further allow to explore and compare their effects both in global and local models. See for instance Fotheringham and Sachdeva (2022), which discuss some examples of the spatial variant of Simpson's paradox, by local models for areal data. Indeed, there might be some cases where the spatial variations observed in the local estimates of some covariates could simply be due to some misspecification of the model form (such as omitting informative covariates), rather than to spatially inhomogeneous behaviour. The same issue could hold in some spatio-temporal contexts.

Moreover, future works could also focus on the use of other summary statistics, such as the K-function, in the local minimum contrast procedure. Suitable simulation studies could be carried out, and the performance of different local second-order summary statistics could be compared, as discussed in Davies and Hazelton (2013).

A related issue concerns the weights given to the averaged summary statistic, which strictly speaking could need not be kernel estimates nor Gaussian ones. For instance, Zhuang (2015) (which applies the weighted likelihood estimator to the spatio-temporal ETAS model to study the spatial variations of seismicity characteristics in the Japan region) chooses a two-dimensional step-wise kernel function based on concentric disjoint octagon rings. The optimal kernel function and bandwidth, not discussed in detail in this study, could be investigated in future research.

Furthermore, other Cox models estimation could be carried out by exploiting the proposed minimum contrast procedure based on local second-order summary statistics. For instance, Neyman-Scott cluster processes are popular models for spatially clustered patterns of points. Baddeley et al. (2022) proposed a new parametrisation of the model involving an index of clustering strength in such a way that the Poisson process is included in the model by setting the clustering strength to zero. The parametrisation leads to improved performance of fitting algorithms. In future, we want to explore the possibility of implementing the same parametrisation in space and time, but also in the local context. Indeed, once established that the local model is a better fit compared to the global counterpart, we might wonder if this statement holds for the whole analysed region or only for a sub-area. For the log-Gaussian Cox processes, this would be more straightforward as the Poisson process corresponds to the case where the variance parameter of the Gaussian random field is equal to zero.

Chapter 4

New modelling approaches on linear networks

In this Chapter, two new modelling approaches on linear networks are proposed, with the aim of explicitly accounting for the network structure in the fitting procedure, as well as including external covariates represented by characteristics of the roads and socio-economic spatial covariates.

In section 4.1 the Gibbs models with random effects (Illian and Hendrichsen, 2010) are adapted to the network structure and to the peculiar problem of describing cruise passengers data. This purely spatial proposal already introduced novelties such as the possibility of accounting for the interaction among points, the road network structure, and external information. Moreover, taking inspiration from Tamayo-Uria et al. (2014) we consider the spatial intensity obtained as one component of a separable spatio-temporal model fitted to the data. At the time of writing, this represents the first attempt to consider a parametric specification of a model for point processes occurring on linear networks. Finally, to take into account the residual clustered behaviour of points, not completely captured by the models fitted previously, we consider a log-Gaussian Cox process.

In section 4.2, we formalize a Self-Exciting network model for the analysis of crime data. Motivated by the analysis of crime data in Bucaramanga (Colombia), we propose a spatio-temporal Hawkes point process model adapted to events living on linear networks, considering a non-parametric modelling strategy for both the background and the triggering components. We further perform a parametric estimation of the background based on covariates. We show that our network model outperforms a planar version, improving the fitting of the Self-Exciting point process model.

The goodness of fit of both the models proposed in this chapter is assessed through specifically tailored diagnostic tools.

4.1 Inhomogeneous spatio-temporal point process on linear networks for visitors' stops data

D'Angelo, N., Adelfio, G., Abbruzzo, A., and Mateu, J. (2022a). Inhomogeneous spatio-temporal point processes on linear networks for visitors' stops data. *The Annals of Applied Statistics*, 16(2):791 – 815

Global Positioning System (GPS) devices can record units with their travel times and spatial coordinates of locations to a high degree of temporal precision. Nowadays, GPS devices are compact, equipped with significant autonomy, and, of the utmost importance, they can store the geographical coordinates of a statistical unit at a given time. The widespread availability of GPS data allows the analysis of social behaviour across time and geographic scales (Butz and Torrey, 2006). Given the wide interest in understanding human mobility, applications of GPS tracking technologies in social science research are widespread (see e.g., Elgethun et al. (2003); Zenk et al. (2011)).

In this section, we focus on GPS data recording places that tourists are often visiting while moving in a constrained space (linear network) given by the topological structure of the study area (i.e. streets, squares, etc.). In particular, data concerns cruisers' stops in Palermo (Italy) in a day, whose spatial and temporal positions along the given network are recorded. Although original data actually constitutes individual trajectories of visitors, we summarise the behaviour of aggregated units. We note that this crucial assumption could lead to biased analyses. From a methodological perspective, a variety of approaches have been used to model intra-destination tourist movements. Identifying stop locations is essential for summarising the information in tracking data since they may indicate touristic areas of interest for the individuals. Gong et al. (2015) review the main research results on stop location identification. In tourism, stop identification may reveal popular places, such as tourist attractions, restaurants, or shopping centres. After identifying attractions, several studies used graph-based methods to analyse tourist movement patterns (Kurashima et al., 2010; Yang et al., 2017; Hu et al., 2019), intending to reconstruct the network of the relationships among the various attractions. Other authors instead focused their attention on trajectory similarity. Shoval et al. (2015), for example, used the sequence alignment method to segment tourists according to their trajectory similarity. Zheng et al. (2019) propose a combination of the Dynamic Time Warping with Earth Mover Distance to compare the trajectories of 56 tourists in the Xiamen Campus, and Petry et al. (2019) provide a comparative approach among various similarity measures for comparing trajectories. Due to the complexity of modelling stop location intensity, most studies undertake descriptive approaches, generally based on visual examination of stop location maps.

At the same time, a modelling-based approach may help to reveal essential tourist behaviours.
From an empirical perspective, similar to recent studies (De Cantis et al., 2016; Ferrante et al., 2018; Domènech et al., 2020a,b; Navarro-Ruiz et al., 2020; Shoval et al., 2020; Casado-Díaz et al., 2021), the proposed research is focused on cruise tourism segments for several reasons. First, the single entry-exit point that characterises cruise visits at the destination makes implementing a GPS-based survey feasible. Second, the limited duration of the visit allows for a complete picture of cruise passengers' experience at the destination. Third, the cruise tourism segment has been largely debated due to their impact on destinations, determining to overcrowding of places in the face of a lower expenditure than traditional tourists (Brida and Zapata, 2010; Larsen et al., 2013). However, none of the previous studies has explicitly focused on cruise passengers' stop activities, to the best of our knowledge. Given the above, we assume that the visitors' stops may represent a spatio-temporal point pattern and therefore, we propose a modelling framework for inhomogeneous spatio-temporal point processes. In particular, we analyse first-order characteristics of the process through the estimation of the spatio-temporal intensity, given the linear network. Then, we consider log-Gaussian Cox processes to analyse the spatio-temporal structure inherent in the data. The obtained results may be useful for identifying both the most visited places and the relationships among the various points of interest, with interesting implications for policymakers' management.

We thus aim at describing visitors' stops proposing new parametric approaches in the context of spatio-temporal point processes on linear networks, contributing to this framework with different aspects. We first consider parametric estimation of the spatio-temporal first-order intensity on the given network. Previous papers have estimated parametric separable spatio-temporal intensities, such as Tamayo-Uria et al. (2014), but, to the authors' knowledge, our approach represents a novelty in the context of point processes on linear networks, since most of the literature about point processes on networks is concerned with the non-parametric estimation of the spatio-temporal intensity (Moradi and Mateu, 2020; Moradi et al., 2019; Mateu et al., 2020). Of course, even if first-order characteristics allow to describe the spatio-temporal displacement of point patterns, a dependence modelling approach allows accounting for effects that might affect the spatio-temporal structure (Baddeley et al., 2015), both related to the characteristics of individuals, referring to the marked point processes methodology, and to the characteristics of the linear network, represented by spatial covariates. Indeed, conversely to the spatial context, spatio-temporal dependence models on linear networks represent a new research field to investigate.

In the second place, and to take into account the spatial interaction among points, since visitors stop several times during the day, we consider Gibbs point process models with mixed effects (Illian and Hendrichsen, 2010; Baddeley et al., 2000) on the networks. Indeed, since the trajectories of each tourist are available, we can track all the individual movements, studying the subject-specific behaviour. We consider the Berman-Turner device for maximum pseudolikelihood (Baddeley et al., 2000), using a quadrature scheme generated on the network. Then, we fit a parametric model incorporating both spatial inhomogeneity and interpoint interactions.

Finally, a first proposal in fitting a log-Gaussian Cox process to spatio-temporal point processes on linear networks is also provided. Our analyses have been carried out using the spatio-temporal package `stlnpp` (Moradi et al., 2020) of the software R (R Core Team, 2022) for point pattern analysis on linear networks, and also the packages `mgcv`, `stpp` and `CompRandFld` (Wood, 2003; Gabriel et al., 2021; Padoan and Bevilacqua, 2015) for intensity estimation and spatio-temporal simulations.

The structure of the section is as follows. Section 4.1.1 presents the data and the motivating problem. Section 2.2.3.2 provided an overview of spatio-temporal point processes on linear networks, referring to both first- and second-order characteristics and the methods used throughout the rest of the section are detailed in section 4.1.2. Section 4.1.3 presents the data analysis, and section 4.1.4 is devoted to conclusions.

4.1.1 Data and motivating problem: GPS visitors' stops in Palermo

The available data refer to a survey carried out in Palermo (Italy) in 2014 for cruiser passengers disembarking in the harbour of Palermo. GPS was used to record the location of the participant tourists and their arrival and departure times. In particular, right before the embarkation of visitors, a survey is carried out (Abbruzzo et al., 2021), collecting information regarding individual characteristics also related to their stay, such as the visit to any specific touristic site. The purpose of this survey is to study the behaviour of cruise passengers in urban contexts, defining profiles of visitors depending on the city and the touristic attractions. Procedures and other survey characteristics may be found in De Cantis et al. (2016) and Shoval et al. (2020).

We consider the GPS trajectories at an aggregated level by analysing individual-level points of interest. Indeed, the analysed data are obtained by using the density-based spatial clustering with noise (*dbscan*) algorithm, which is a well-known non-parametric data clustering algorithm proposed by Ester et al. (1996). Given a set of points in some space, it clusters points that are closely packed together (points with many nearby neighbours), marking as outliers those points lying alone in low-density regions (whose nearest neighbours are too far away). This, applied to a set of trajectories, produces a set of clusters (whose centroids can be seen as stops) and a set of noise points (that can be removed from the analysis) (Abbruzzo et al., 2021). Each cluster is therefore identified by a spatial and a temporal location, and this information is used to build the point pattern. Since the aim is to study the spatio-temporal displacement of the *visitors' stops*, we here ignore the trajectories and time is discretised in hours in order to identify the densest hours during a day. Indeed, we assume that the stops identify the points of interest visited by tourists. The complete spatio-temporal point pattern is displayed in Figure 4.1. Information about the spatial support, that is the road network L, has been obtained from the OpenStreetMap data.

Due to interpretational and computational reasons, we focus our analysis on a subset of the original data. As the complete dataset actually includes some noise clusters and most of the visitors' stops occur in the historical downtown, the analysis is restricted to the spatio-temporal point pattern consisting of 159 stops of 44 visitors, stopping 4 times on average during the day, in the downtown of Palermo city on the 18^{th} April 2014. The linear network has 990 vertices and 1180 lines. The data consist, for each visitor stop, of the spatial location on the road network and the occurrence time during the day, with $t \in T = [8\ AM, 6\ PM]$. In the left panel of Figure 4.2, the locations of the events are displayed, as obtained from the *dbscan* algorithm. On the right panel, the locations considered in this analysis are displayed. Indeed, when computing the `stlnpp` object in R, points that lay outside the linear support are relocated to the closest location from the original one, but on the network.

Since we expect that people stop more often near well-known touristic attractions, we computed $Z(\mathbf{u})$ as the *distance from the nearest touristic attraction*, where the metric used was the shortest-path distance on the network from the location of these known attractions:

- Quattro Canti
- Cathedral
- Teatro Massimo
- Fontana Pretoria
- Teatro Politeama
- Piazza Bellini
- Chiesa Valdese
- Via Principe di Belmonte

Note that $Z(\mathbf{u})$ is defined on a support that is slightly smaller than the originally considered network, as the latter is not completely connected. The location of the touristic attractions on the linear network and the values of the resulting distance variable are displayed in Figure 4.3.

4.1.2 Model specification proposal

In this section, the new modelling approaches proposed are introduced. In particular, section 4.1.2.1 deals with a spatial point process model, that is a Gibbs point process model with mixed effects, and its estimation on a linear network is introduced. Then, section 4.1.2.2 develops log-Gaussian Cox processes on a linear network.

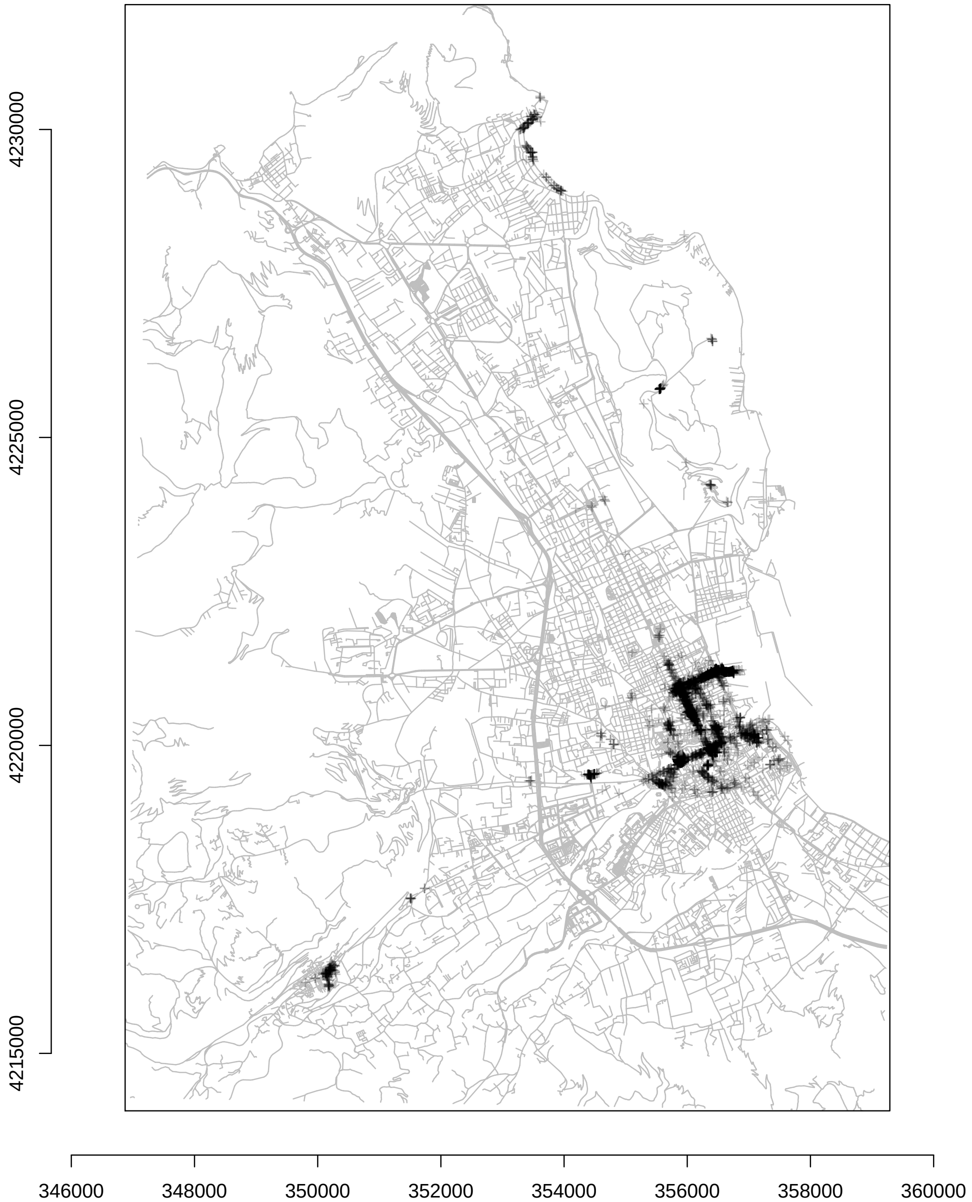


Figure 4.1: The grey segments represent the streets of Palermo city, obtained from OpenStreetMap. In black, the points represent the visitors' stops.

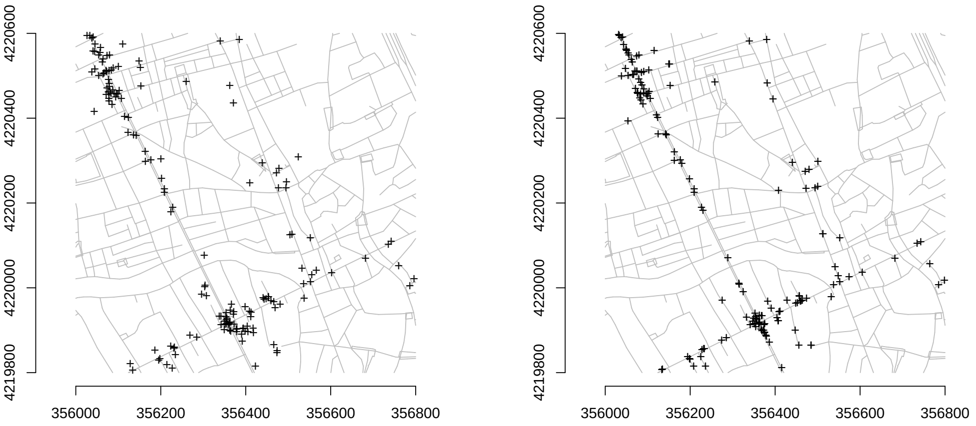


Figure 4.2: Original locations of points and the locations relocated on the network support.

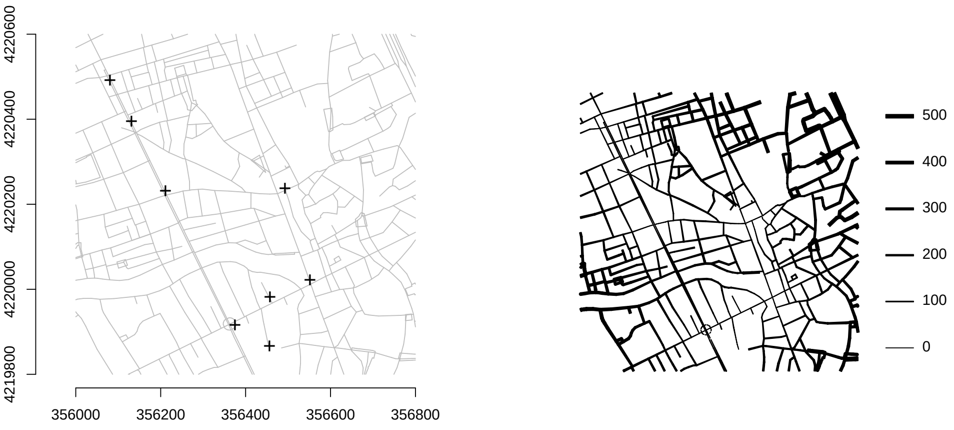


Figure 4.3: Left*: Location of the points of interest, on the network.* Right*: Values of the variable distance from the nearest touristic attraction.*

4.1.2.1 Gibbs point process models with mixed effects on linear networks

We here introduce a new modelling approach for describing the spatio-temporal behaviour of the visitors. The aim is to fit a parametric model to the points (representing the visitors' stops) accounting for the individual choices during the day by introducing a random subject-specific effect.

At this aim we refer to Gibbs point process models with mixed effects (Bell and Grunwald, 2004), conforming the procedure in Illian and Hendrichsen (2010) to the context of the linear networks. More specifically, we focus on a pairwise term interaction process, where the conditional intensity is a log-linear model and the interaction terms are obtained as the sum of fixed and random effects, modelling the variation within each individual. The considered model reflects both the inter-group differences and interaction strength, proposing a flexible interaction function. The proposed approach refers to the application of the Berman-Turner device (Baddeley et al., 2000) to a generalised linear mixed model with log link and Poisson outcome for large datasets. The model is fitted using the functions of the package `mgcv` of `R`.

Let M be the number of visitors on a linear network L, each generating the point patterns $\mathbf{x}_1, \ldots, \mathbf{x}_M$ that can be thought of as the individual pattern of stops. This flexible procedure allows accounting for the individual characteristics information both by suitable random and fixed factors and by external covariates.

Here we only assume that visitors' stops interact within each individual's path and that the intensity of the pattern varies among the individuals. Then, the `ID` of the visitor is introduced as a random effect in the model. This resolves in assuming, for each $\mathbf{x}_m$ with $m = 1, \ldots, M$, a pairwise interaction process (Van Lieshout, 2000), with conditional intensity (Kallenberg, 1984) given by

$$\lambda_{\theta,\phi_m}(\mathbf{u}; \mathbf{x}_m) = b_{\theta,\phi_m}(\mathbf{u}) \prod_{i=1, \mathbf{x}_{mi} \neq \mathbf{u}}^{n(\mathbf{x}_m)} h_{\theta,\phi_m}(\mathbf{u}, \mathbf{x}_{mi}) \tag{4.1}$$

where $n(\mathbf{x}_m)$ is the number of points in $\mathbf{x}_m$, that is the number of stops per visitor, and $b_{\theta,\phi_m}(\mathbf{u})$ and $h_{\theta,\phi_m}(\mathbf{u}, \mathbf{v})$ are

$$b_{\theta,\phi_m}(\mathbf{u}) = \exp(\theta^T B_1(\mathbf{u}) + \phi_m^T B_2(\mathbf{u}))$$

and

$$h_{\theta,\phi_m}(\mathbf{u}, \mathbf{v}) = \exp(\theta^T H_1(\mathbf{u}, \mathbf{v}) + \phi_m^T H_2(\mathbf{u}, \mathbf{v}))$$

i.e., two functions that model the intensity and the interaction, respectively.

Each of these functions consists of the sum of a term describing the general intensity (or interaction) and another term describing the intensity (or interaction) specific to this point, which is assumed to be an independent sample from some probability distribution. Parameter θ is a vector of fixed effects and the random effects ϕ_m are assumed to come from $\Phi \sim N(0, \sigma_\phi^2 I)$. The vectors $B_k(\mathbf{u})$ and $H_k(\mathbf{u}, \mathbf{v})$

are defined for each $\mathbf{u}, \mathbf{v} \in L$ and $H_k(\cdot,\cdot)$ is a symmetric function of $\mathbf{u}$ and $\mathbf{v}$. Note that $B(\mathbf{u})$ is constant, assuming spatial homogeneity. Indeed, it is important to notice that the general form assumed for b_θ embraces not only parametric models but also generalised additive models (Hastie and Tibshirani, 1990) in which $B(\mathbf{u})$ would be a vector of spline basis functions. However, this apparently does not extend to GAM-type models for h_θ, hence we focus on purely parametric models for interpoint interaction.

Typically, the function H would be a pairwise interaction function, while $B(\mathbf{u}) = (B_1(\mathbf{u}), B_2(\mathbf{u}))^T$ would be a vector of convenient scalar functions such as polynomials or orthonormal functions of coordinates. Furthermore, this model can accommodate various kinds of interpoint interactions. For example, in many applications, interaction strength can be assumed to gradually decrease with distance. In situations where simple models like the Strauss process might be too simplistic and inappropriate, an interaction function which can be used to model either attraction or repulsion is preferable (Illian et al., 2008). We consider a more general but still quite simple smooth interaction function as the one proposed in Illian and Hendrichsen (2010), which we assume to only depend on the shortest-path distance between points, rather than on Euclidean distances. For two points with location $\mathbf{u}$ and $\mathbf{v}$, we define

$$H_k(\mathbf{u},\mathbf{v}) = \begin{cases} \left(1 - \left(\frac{d(\mathbf{u},\mathbf{v})}{R}\right)^2\right)^2 & \text{if} \quad 0 < d(\mathbf{u},\mathbf{v}) \leq R \\ 0 & \text{else} \end{cases} \tag{4.2}$$

where $d(\mathbf{u},\mathbf{v})$ is computed as the shortest-path distance and $R \geq 0$ defines the radius of interaction.

For example, as we later show in our application, we can consider a function $Z(\cdot)$ defined on each point (or quadrature point) of the network L, as a particular covariate of interest obtained as the minimum shortest-path distance from location $\mathbf{u}$ to the nearest point of another point pattern Y.

Therefore, Gibbs point processes with mixed effects considered in this section have an intensity of the form

$$\lambda_{\theta,\phi_m}(\mathbf{u},\mathbf{x}_m) = \exp(\theta^T S_1(\mathbf{u},\mathbf{x}_m) + \phi_m^T S_2(\mathbf{u},\mathbf{x}_m))$$

where

$$S_k(\mathbf{u},\mathbf{x}_m) = B_k(\mathbf{u}) + \sum_{i=1,\mathbf{x}_{mi}\neq\mathbf{u}}^{n(\mathbf{x}_m)} H_k(\mathbf{u},\mathbf{x}_{mi}), \qquad \text{for} \quad k = 1,2$$

is a vector of spatial covariates defined at each point $\mathbf{u}$ in L. It is easy to introduce spatial trends or dependence on spatial covariates, by simply adding more terms to the linear predictor S in the associated Poisson log-linear regression model.

Finally, the conditional intensity for $\mathbf{x}_m$ in (4.1) becomes

$$\lambda_{\theta,\phi_m}(\mathbf{u},\mathbf{x}_m) = \exp\Bigg(\theta^T B_1(\mathbf{u}) + \phi_m^T B_2(\mathbf{u}) + \theta^T \sum_{i=1,\mathbf{x}_{mi}\neq\mathbf{u}}^{n(\mathbf{x}_m)} H_1(\mathbf{u}, x_{mi}) + \phi_m^T \sum_{i=1,\mathbf{x}_{mi}\neq\mathbf{u}}^{n(\mathbf{x}_m)} H_2(\mathbf{u},\mathbf{x}_{mi})\Bigg)$$

and the density with respect to the homogeneous Poisson process is then

$$f(\mathbf{x}_m;\theta,\phi_m) = \alpha(\theta,\phi_m) \prod_{i=1}^{n(\mathbf{x}_m)} b_{\theta,\phi_m}(\mathbf{x}_{mi}) \prod_{i<j} h_{\theta,\phi_m}(\mathbf{x}_{mi},\mathbf{x}_{mj})$$

where $\alpha(\theta,\phi_m)$ is a normalising constant.

Assuming independence between the patterns observed for each individual m we can derive a density of a model of the overall pattern as

$$f(\mathbf{x}_1,\dots,\mathbf{x}_M;\theta,\phi_m) = \tilde{\alpha}(\theta,\phi) \prod_{m=1}^{M} \prod_{i=1}^{n(\mathbf{x}_m)} b_{\theta,\phi_m}(\mathbf{x}_{mi}) \prod_{i<j} h_{\theta,\phi_m}(\mathbf{x}_{mi},\mathbf{x}_{mj})$$

where $\phi = (\phi_1,\dots,\phi_M)$ and $\tilde{\alpha}(\theta,\phi)$ is a further normalising constant.

4.1.2.1.1 Model estimation through pseudolikelihood Adapting the pseudolikehood over the subset $A \subseteq W$ into the network, where W is the observation window, we define the pseudolikelihood over the subset $A \subseteq L$ as

$$\mathrm{PL}_A(\theta,\phi;\mathbf{x}_1,\dots,\mathbf{x}_M) = \prod_{m=1}^{M}\left(\prod_{\mathbf{x}_{mi}\in A} \lambda_{\theta,\phi_m} \exp\left(-\int_A \lambda_{\theta,\phi_m}(\mathbf{u};\mathbf{x}_m) d\mathbf{u}\right)\right)$$

that is,

$$\begin{aligned}\mathrm{PL}_A(\theta,\phi;\mathbf{x}_1,\dots,\mathbf{x}_M) = \prod_{m=1}^{M}\left(\prod_{i=1}^{n(\mathbf{x}_m)} b_{\theta,\phi_m}(\mathbf{x}_{mi}) \prod_{i\neq j}^{n(\mathbf{x}_m)} h_{\theta,\phi_m}(\mathbf{x}_{mi},\mathbf{x}_{mj})\right) \\ \times \exp\left\{-\int_A b_{\theta,\phi_m}(\mathbf{u}) \prod_{i=1}^{n(\mathbf{x}_m)} h_{\theta,\phi_m}(\mathbf{u},\mathbf{x}_{mi}) du\right\}\end{aligned} \tag{4.3}$$

Following Baddeley et al. (2000), we apply the Berman-Turner device. Based on some quadrature rule, we approximate the integral in (4.3) for each m by a finite sum over a set of points $\mathbf{u}_{mj}$ with $j = 1,\dots,l_m$, containing all the data points. It is shown that the log-pseudolikelihood is formally equivalent to the log-likelihood of independent weighted Poisson variables, i.e.

$$\log \mathrm{PL}_A(\theta, \phi; \mathbf{x}_1, \ldots, \mathbf{x}_M) \approx \sum_{m=1}^{M} \sum_{j=1}^{l_m} y_{mj} \log \lambda_{\theta,\phi_m}(\mathbf{u}_{mj}) - \lambda_{\theta,\phi_m}(\mathbf{u}_{mj}) w_{mj}$$

where w_{mj} are the quadrature weights, $y_{mj} = \frac{z_{mj}}{w_{mj}}$ and

$$z_{mj} = \begin{cases} 1 & \text{if} \quad \mathbf{u}_{mj} \in \mathbf{x}_{m1}, \ldots, \mathbf{x}_{mn(\mathbf{x}_m)} \\ 0 & \text{if} \quad \mathbf{u}_{mj} \notin \mathbf{x}_{m1}, \ldots, \mathbf{x}_{mn(\mathbf{x}_m)} \end{cases} \tag{4.4}$$

In general, the procedure for the Berman-Turner device starts with generating a set of dummy points and superimposing it with the data points to form the set of quadrature points. In this section, as we are analysing a point pattern on a linear network, the generation of the dummy points is restricted to the considered network segments. Furthermore, as we are dealing with a marked point pattern (the ID of the visitors being the mark), the procedure is slightly modified. The first task in fitting the proposed spatial model is to create the quadrature scheme. We then replicate the dummy points for each possible mark, and generate dummy marked points at the same locations as the data points but with different marks. Finally, the two dummy patterns and the data points are superimposed, and we compute the quadrature weights and form the indicators as in (4.4).

Having generated the quadrature points, we construct the smooth function as in (4.2) by computing

$$v_{mj} = \sum_{i=1}^{n(\mathbf{x}_m)} H(\mathbf{u}_{mj}, \mathbf{x}_{mi}). \tag{4.5}$$

We finally specify the model as a log-linear Poisson regression to be fitted to the responses y_{mj} and covariate values v_{mj} with weights w_{mj}. Also any spatial covariate $Z(\cdot)$, if included in $b_{\theta,\phi_m}(\mathbf{u})$, would need to be defined on each $\mathbf{u}_{mj}$.

Since the conditional intensity is expressed as the sum of fixed and random factors, we refer to a generalised linear mixed model with log-link and Poisson outcome. The numerical methods are designed for datasets containing upwards of several tens of thousands of data (Wood et al., 2015). The advantage is a much lower memory footprint, but it can also be much faster, for large datasets. An alternative fitting approach was provided by Wood et al. (2017); Li and Wood (2020), based on the discretisation of covariate values and C code level parallelisation.

4.1.2.2 Log-Gaussian Cox processes on linear networks

Since the estimation of the LGCP model for patterns on linear networks is introduced in this thesis for the first time, we face the question of choosing the appropriate distance metric. In the context of analysis on networks, Cronie et al. (2020) discuss the intensity reweighted moment pseudostationary (IRMPS) log-Gaussian Cox processes.

Let X be a log-Gaussian Cox process with random intensity measure $\Gamma(A) = \int_A \Gamma(\mathbf{u}) d_1\mathbf{u} = \int_A \exp\{S(\mathbf{u})\} d_1\mathbf{u}, \quad A \subseteq L$, where S is a Gaussian Random Field on L with mean function $\mu(\mathbf{u}) \in \mathbb{R}, \mathbf{u} \in L$, and covariance function $\mathbb{C}(\mathbf{u}, \mathbf{v}) = \mathbb{C}(d_L(\mathbf{u}, \mathbf{v}_1), d_L(\mathbf{u}, \mathbf{v}_2)) \in \mathbb{R}, \quad \mathbf{v}_1, \mathbf{v}_2 \in L$, for any $\mathbf{u} \in L$ and some function $\mathbb{C}$. In Cronie et al. (2020) it is proved that, if d_L is an arbitrary regular distance metric, $\mathbb{C}(\mathbf{u}, \mathbf{v}) = \mathbb{C}(d_L(\mathbf{u}, \mathbf{v}_1), d_L(\mathbf{u}, \mathbf{v}_2))$ is a proper covariance function on L which does not depend on the choice of $\mathbf{u}$.

Baddeley et al. (2020) briefly address the issue of the choice of the regular metric to be used. Baddeley et al. (2017) and Rakshit et al. (2017) investigated the statistical implications of using a distance metric other than the shortest-path distance. For any given metric δ, a point process is defined to be $\delta - correlated$ if its pair correlation function depends only on $\delta - distance$. Baddeley et al. (2017) showed there is a rich class of point process models which are $\delta - correlated$ with respect to Euclidean distance.

Nevertheless, since the fitting of the LCGP based on metrics different from the Euclidean distance is not yet formalized, we refer to the most classical LGCP defined in the Euclidean context. This choice is based on the assumption that the geometry of the network can be ignored, considering Euclidean distances instead of the shortest-path distances, and therefore considering the observation window W rather than the linear network L. This assumption may be relevant depending on the considered application. In our particular case, the structural differences between the Euclidean and shortest-path distances are certainly small, reinforcing the choice of the Euclidean distances when fitting an LGCP.

4.1.3 Data analysis

4.1.3.1 Exploratory analysis

Figure 4.4 shows the spatial and temporal displacements of visitors' stops. The point pattern in Figure 4.4 (left) displays some spatial inhomogeneity, reflecting the spatial distribution of the considered points of interest. Two clusters are evident in the bottom left and top left areas of the analysed window, corresponding to the two well-known touristic attractions, *Quattro Canti* and *Cathedral*. On the right panel of Figure 4.4, the cumulative number of visitors stopping at given hours of the day is shown, noting two peaks in the temporal distributions of visitors' stops, corresponding to 10 a.m. and 2 p.m.

For assessing if the spatial intensity $\lambda(\mathbf{u})$ might depend on a given spatial covariate $Z(\mathbf{u})$, Berman's test is carried out over each covariate. In particular, we focus on the possible dependence of the spatial intensity on the two spatial coordi-

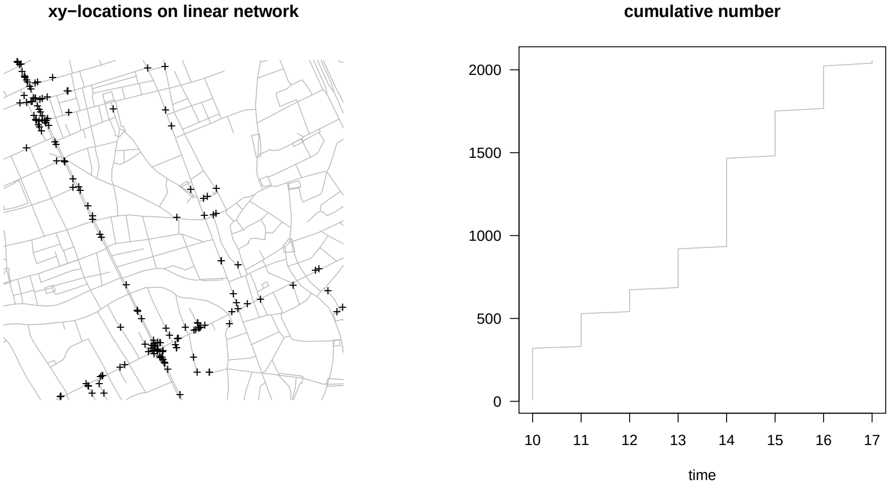


Figure 4.4: Analysed spatio-temporal point pattern on the linear network. Left*: Projection of the points onto the network.* Right*: Cumulative number of points.*

nates, latitude and longitude, and on the variable *distance from the nearest touristic attraction*. The corresponding p-values are reported in Table 4.1. The effect of explanatory variables on the intensity of the process can also be investigated using non-parametric curve estimation techniques. Although the techniques were developed for spatial point patterns in two dimensions, they depend only on the space of values of Z, so they apply to point patterns in any space (Baddeley et al., 2012), in particular to networks (Baddeley et al., 2015). So, to get further information on the type of dependence of the spatial intensity on the given continuous spatial covariates, the smoothed functions are computed as $f(\mathbf{u}) = \rho(Z(\mathbf{u}))$, and displayed in Figure 4.5. As in section 3.1.2, among the estimators of $f(Z(\mathbf{u}))$ proposed by Baddeley et al. (2012), we consider a smoothing procedure based on fixed-bandwidth kernel density estimation.

Table 4.1: Berman's Test for the available continuous spatial covariates

Spatial covariate	p-value
Latitude	0.9814
Longitude	0.9757
Distance from the nearest touristic attraction	0.0000

From Table 4.1 we see that only the *distance from the nearest touristic attraction* influences the spatial intensity of the process. However, by looking at Figure 4.5, we see that the effect of the spatial coordinates varies as a function of the scale and, therefore, in a regression model for the spatial intensity, this variable could be included non-parametrically. The distance from the nearest touristic attraction seems to have an effect on the intensity of the process only up to a certain distance, about 150 meters. For these small distances, the effect is negative, that is to say, as expected, the closer the tourists are to an attraction point, the more likely they are to stop.

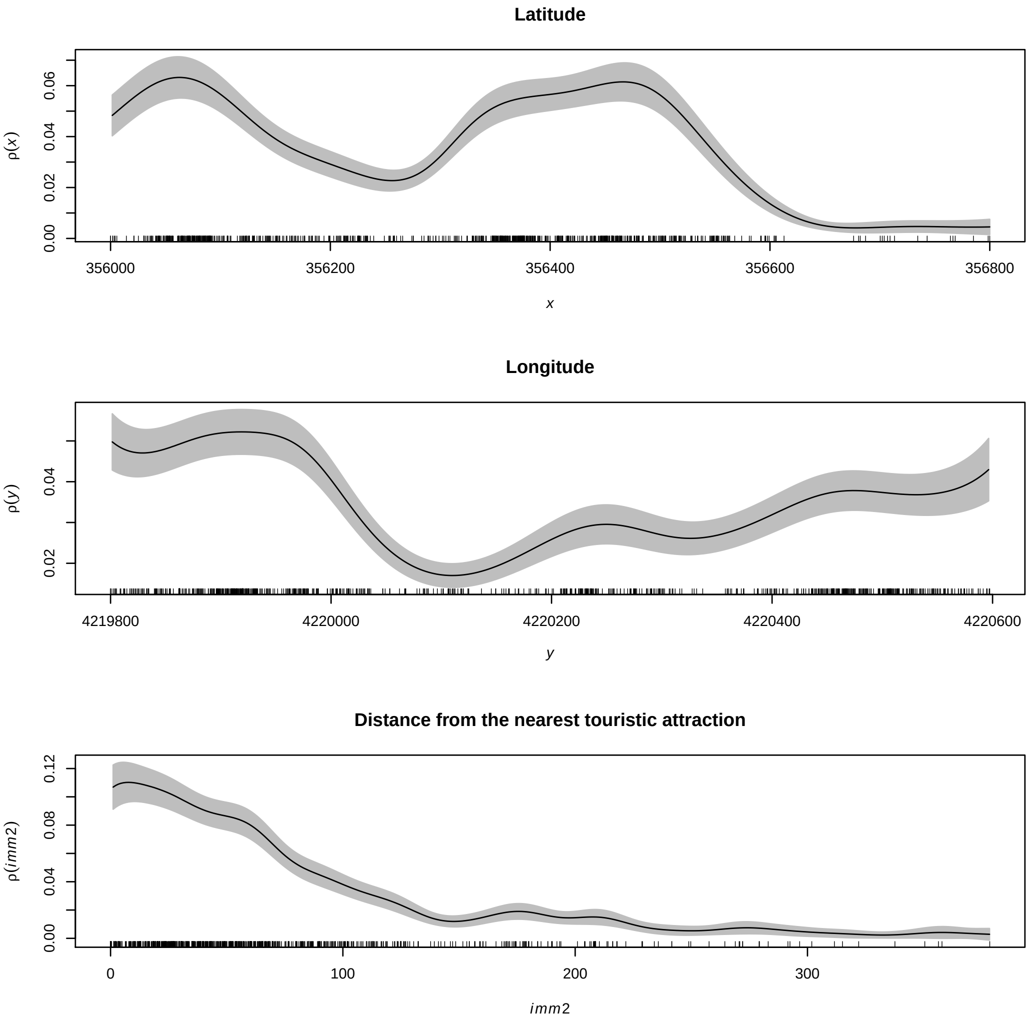


*Figure 4.5: Smoothed functions linking Latitude (**a**), Longitude (**b**) and the* distance from the nearest touristic attraction *(**c**) to the spatial intensity.*

4.1.3.2 Intensity estimation

Table 4.2: Parametric coefficients model and approximate significance of smooth terms of the Gibbs model

		Estimate	Std. Error	z value	$\Pr(>\|z\|)$	
θ_1	(Intercept)	-8.023371	0.616119	-13.022	0.000	***
θ_2	(v)	0.092422	0.020467	4.516	0.000	***
θ_3	(d)	-0.012348	0.003082	-4.006	0.000	***
		edf	Ref.df	Chi.sq	p-value	
$H(\mathbf{u}_{im}, \mathbf{v}_{im})$	(id)	23.03	43.00	71.06	0.000	***
$B_4(\mathbf{u}_{im})$	(lat,long)	16.92	21.07	66.00	0.000	***

4.1.3.2.1 Spatial intensity estimation We now fit the purely spatial model as introduced in section 4.1.2.1. We set $B_1(\mathbf{u}_{im}) = \mathbf{I}$, with $\mathbf{I}$ the identity function and $B_3(\mathbf{u}_{im}) = Z(\mathbf{u}_{im})$. In addition, $B_2(\mathbf{u}_{im})$ denotes the `ID` of the tourist, included as a random effect. $B_4(\mathbf{u}_{im})$ is a non-parametric function for $\mathbf{u}_{im} \in L$, estimated here through thin plate regression splines with 29 knots.

Therefore, we have

$$b_{\theta,\phi_m}(\mathbf{u}_{im}) = \exp(\theta_1 + \phi_{1m} B_2(\mathbf{u}_{im}) + \theta_3 B_3(\mathbf{u}_{im}) + B_4(\mathbf{u}_{im}))$$

and

$$h_{\theta,\phi_m}(\mathbf{u}_{im}, \mathbf{v}_{im}) = \exp(\theta_2 H(\mathbf{u}_{im}, \mathbf{v}_{im}) + \phi_{2m} H(\mathbf{u}_{im}, \mathbf{v}_{im}))$$

where $H(\cdot, \cdot)$ is defined as in Equation (4.2). In this application, the interaction radius is set to $R = 100$ meters, since it is reasonable to assume that visitors' stops interact with each other up to this distance. The inclusion of this smooth function in the model allows attractive or repulsive interaction indicated by positive or negative interaction parameters, respectively.

Thus, our model for the spatial intensity is given by

$$\log \lambda_{\theta,\phi_m}(\mathbf{u}_{im}) = \theta_1 + \phi_{1m} B_2(\mathbf{u}_{im}) + \theta_2 v_{im} + \theta_3 Z(\mathbf{u}_{im}) + B_4(\mathbf{u}_{im}) + \phi_{2m} v_{im}$$

where

1. θ_1 is the common intercept
2. θ_2 is the fixed effect of the smooth function in (4.5)
3. θ_3 is the fixed effect of the *distance from the nearest point of attraction*
4. ϕ_{1m} is the random effect of the `ID`

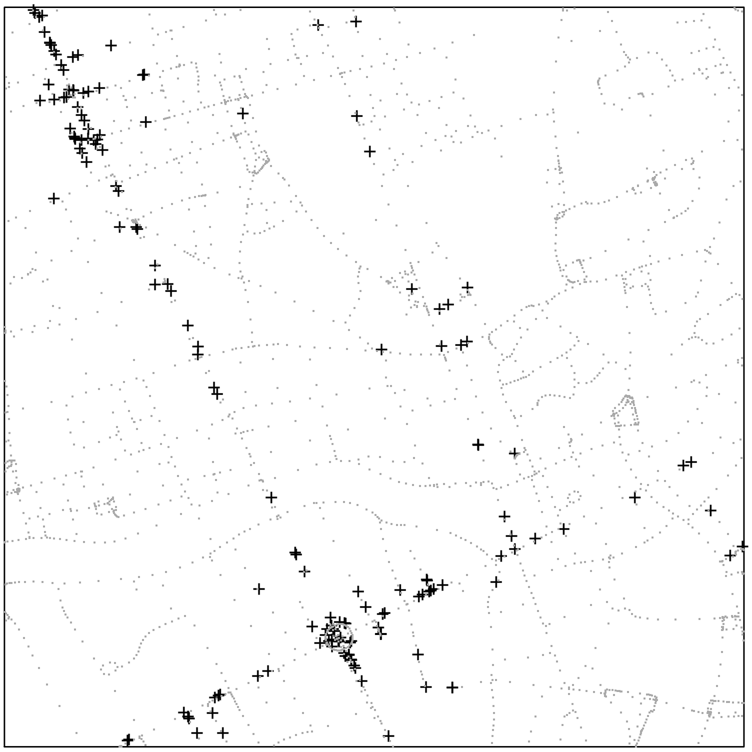


Figure 4.6: Data and dummy points.

5. ϕ_{2m} represents the random effects for the interaction smooth function

The final quadrature scheme used for model fitting consists of 159 data points and 2362 dummy points, displayed in Figure 4.6. This leads to a dataset of 110924 observations, which is equal to the number of data points plus the number of dummy points, all replicated for the number of marks.

Since the random effects ϕ_{2m} for the interaction smooth function are not significant, the interaction among points turns out to be fixed, meaning that visitors' stops attract each other in the same way for each visitor. The chosen model is

$$\log \hat{\lambda}_{\theta,\phi_m}(\mathbf{u}_{im}) = \hat{\theta}_1 + \hat{\phi}_{1m} B_2(\mathbf{u}_{im}) + \hat{\theta}_2 v_{im} + \hat{\theta}_3 B_3(\mathbf{u}_{im}) + B_4(\mathbf{u}_{im}).$$

In Table 4.2 we report the estimates of the fixed effects and the summary of the random effects. When $\exp(\hat{\theta}_1)$ is multiplied by the length of the network, the model estimates 8.082637 stops for each individual, higher than the original average stops. The positive interaction parameter $\exp(\hat{\theta}_2) = 1.096828$ indicates that overall the visitors' stops attract each other. Therefore, visitors tend to stop in the same spots. Furthermore, $\exp(\hat{\theta}_3) = 0.9877283$ indicates that moving away from any touristic attraction decreases the probability of visitors stopping. We notice that only the intensity varies among visitors and not the interaction.

In Figure 4.7 we represent the fitted intensity for both data and generated dummy points. The model clearly estimates higher intensities in correspondence to the most visited areas, that is the *Quattro Canti* and the *Cathedral* (zoomed in the bottom panels of Fig. 4.7). Therefore, the fitted purely spatial model adequately describes the intensity of the densest regions and predicts the expected number of cases.

4.1.3.2.2 Temporal intensity estimation with Poisson harmonic regression In the left panel of Figure 4.8, the total counts of stops per hour of the day are displayed. The temporal intensity $\lambda(t)$ is estimated parametrically as a Poisson Harmonic regression. This is an appealing formulation when data displays a seasonal or cyclic variation. Thus it seems to make sense to model the mean function as being periodic with a period of 24 hours or less and the obvious way to do that is to use trigonometric functions. After computing $\omega = 2\pi/24$, we fit the following Poisson model

$$\begin{aligned}\log \lambda(t) = {} & \delta + \alpha_1 \cos(\omega t) + \beta_1 \sin(\omega t) + \alpha_2 \cos(2\omega t) + \beta_2 \sin(2\omega t) + \alpha_3 \cos(4\omega t) \\ & + \beta_3 \sin(4\omega t)\big) + \alpha_4 \cos(8\omega t) + \beta_4 \sin(8\omega t)\end{aligned}$$

that models the number of visitors' stops per hour as a function of sinusoidal and cosinusoidal functions of different periods, namely 24, 12, 6 and 3 hours cycles.

We compare this model with all the nested ones, removing one period at a time, following an F-tests stepwise procedure. In the right panel of Figure 4.8, we plot the fitted values of the four nested models, starting from the one with only a 24-hour cycle, adding each time half cycle. We chose the final model based on the results of the F-tests. Indeed, the cycles corresponding to 24, 12 and 6 hours are significant in predicting the total number of visitors who stop during a day. Therefore the chosen model is

$$\begin{aligned}\log \hat{\lambda}(t) = {} & \hat{\delta} + \hat{\alpha}_1 \cos(\omega t) + \hat{\beta}_1 \sin(\omega t) + \hat{\alpha}_2 \cos(2\omega t) + \hat{\beta}_2 \sin(2\omega t) + \hat{\alpha}_3 \cos(4\omega t) \\ & + \hat{\beta}_3 \sin(4\omega t).\end{aligned}$$

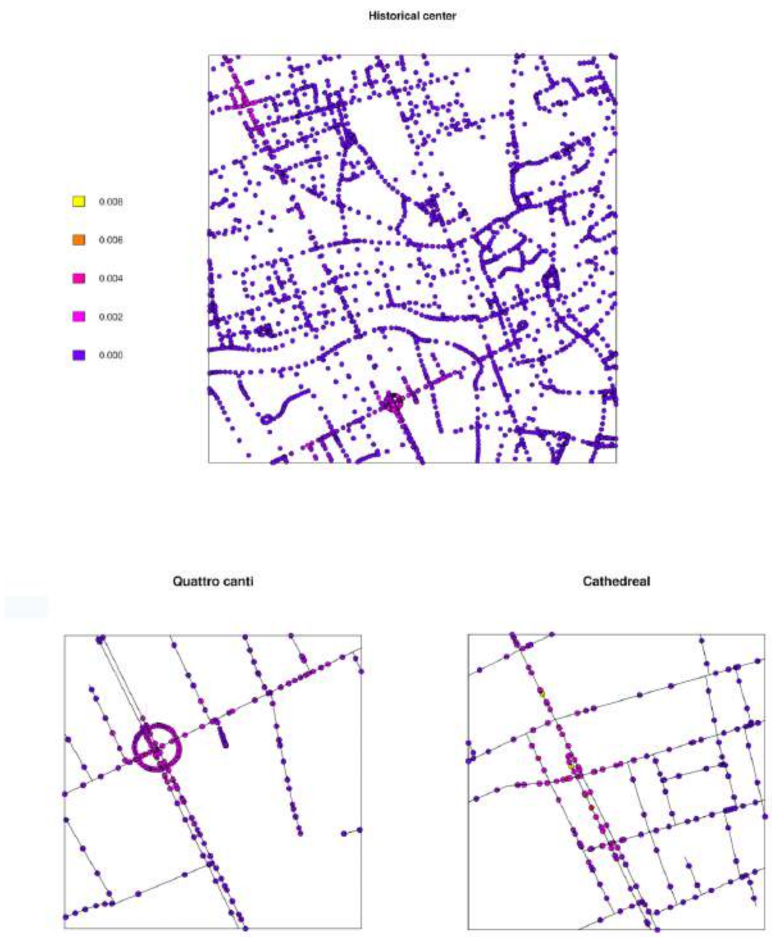


Figure 4.7: Fitted spatial intensity.

In Figure 4.8 we display, together with the temporal distribution on the left panel, the fitted intensities on the right panel. The straight line represents the fitted intensity of the chosen model, able to correctly estimate the number of visitors who stop during the day, as well as the two peaks in the densest hours.

4.1.3.3 Spatio-temporal intensity and second-order characteristics

The spatio-temporal intensity is obtained as in (2.37) by multiplying the purely spatial and purely temporal intensities previously fitted. The resulting intensity had to be normalised, in order to make the estimator unbiased, that is, giving the

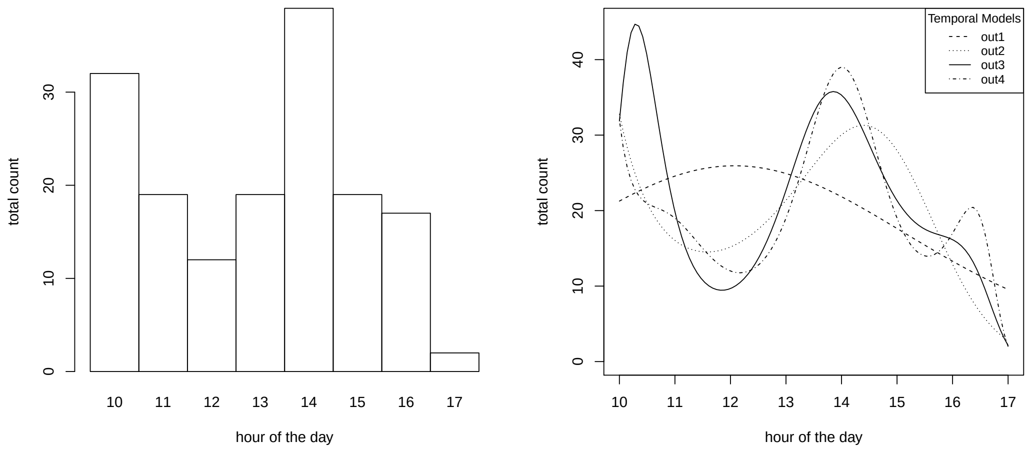


Figure 4.8: Left*: Histogram of the total count in each hour of the day for the visitors' stops data.* Right*: Predicted values of the four Poisson harmonic models.*

expected number of points

$$\mathbb{E}\left[\int_{L\times T}\hat{\lambda}(\mathbf{u},t)d_2(\mathbf{u},t)\right]=\int_{L\times T}\lambda(\mathbf{u},t)d_2(\mathbf{u},t)=n.$$

Therefore, the final intensity function is actually obtained as

$$\hat{\lambda}(\mathbf{u},t)=\frac{\hat{\lambda}(\mathbf{u})\hat{\lambda}(t)}{\int_{L\times T}\hat{\lambda}(\mathbf{u},t)d_2(\mathbf{u},t)}$$

and it will be used for simulations for model diagnostics.

To test for spatio-temporal clustering, we simulate $Q=99$ inhomogeneous spatio-temporal Poisson processes with intensity $\lambda_0(\mathbf{u},t)=\hat{\lambda}(\mathbf{u},t)$ (see Appendix A), and we then compute estimates $\hat{K}_q(r,h)$ by means of Equation (2.40). In Figure 4.9, the inhomogeneous K-function estimated from the data is displayed in white, and the envelopes obtained from $Q=99$ simulations and at a significance level of $\alpha=0.05$ are displayed in grey. As we note, the observed K-function does not lay within the envelopes for each spatial and temporal distance. This result, together with a borderline p-value of 0.05 of the test of clustering, suggests that the proposed model cannot completely catch the features of the observed point pattern. In particular, the presence of ranges where the observed K-function lays above the envelopes suggests that there is additional clustering induced from external causes, such as unobserved covariates. The findings in Figure 4.10 further show that the observed pattern of visitors' stops cannot be described only by an inhomogeneous spatio-temporal Poisson point process (see in particular Figure 4.10(d)). Therefore, an LGCP model is further used for describing the residual clustered structure.

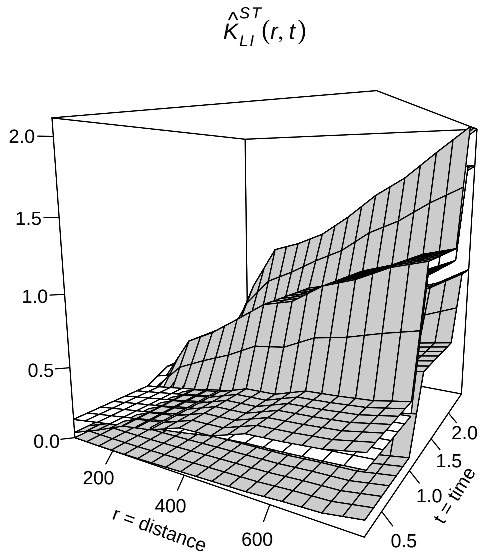


Figure 4.9: In white*: the estimated inhomogeneous* K*-function for visitors' stops data.* In grey*: envelopes based on 99 simulations from inhomogeneous spatio-temporal Poisson point processes at a significance level of 0.05.*

4.1.3.4 LGCP model fitting and diagnostics

Having established that an inhomogeneous process is not enough to accurately describe the analysed pattern, we turn to the LGCP fitting. Of course, when specifying a model for point patterns whose spatial locations are restricted to a linear network, one wishes to take into account the geometry of the network. Therefore, a crucial issue when LGCP is fitted on the network, as anticipated in section 4.1.2.2, is the choice of the most appropriate distance metric. To assess if there are differences between the Euclidean and the shortest-path distances computed for the point pattern under study, we visually compare their distributions, shown in Figure 4.11. As expected, the shortest-path distances are slightly overall larger than the Euclidean ones. Anyway, we chose to neglect these differences and consider the Euclidean dis-

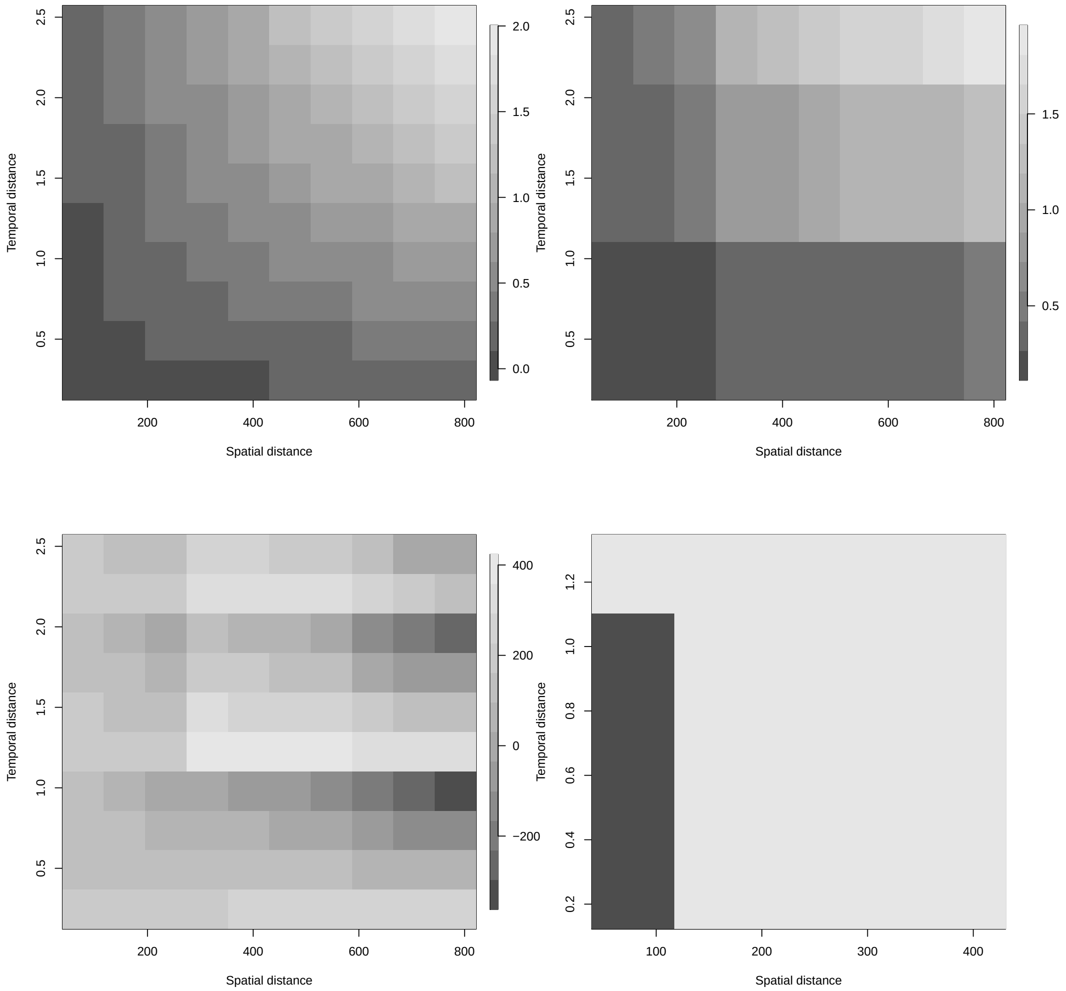


Figure 4.10: Top panels*: [a] Values of the observed K-function, [b] corresponding values under Poisson: the brighter the colour the higher the values.* Bottom panels*: [c] $\hat{K}(r,h) - rh$ values and [d] a close-up to the comparison between $\hat{K}(r,h) - rh$ and tolerance envelopes indicating spatio-temporal clustering. The darkest colour identifies the regions in (r,h) where $\hat{K}(r,h) - rh$ is greater than $\hat{K}^{0.95}(r,h)$. The bottom right panel [d] represents detail up to half of the spatial and temporal ranges.*

tances, to sustain the Gaussian Random Field and the overall fitting of the LGCP model.

The initial values for the spatial parameters estimated through M_J (2.27) are

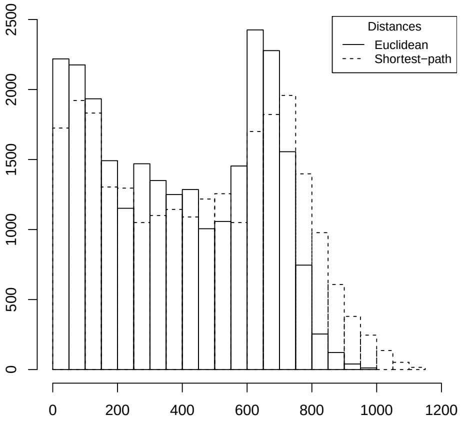


Figure 4.11: Distributions of Euclidean and shortest-path distances of the analysed point process.

set to $\alpha_{initial} = 8$ and $\sigma^2_{initial} = log(n)/2 = 2.53$, where n is the number of points in the point pattern. As far as the temporal interval in M_C (2.28), this is set equal to $[0, 10]$. The resulting estimates of the parameters are $(\hat{\alpha}, \hat{\sigma}^2, \hat{\tau}) = (5.70, 5.17, 0.51)$, obtained using the pair correlation function as a non-parametric estimate of $\hat{J}$ in M_J (2.27). To evaluate these estimates, in Figure 4.12 we plot the empirical covariances for the purely spatial and temporal components, together with their exponential estimates. These represent the best estimates since the minimum contrast estimation method actually selects the covariance parameters as the ones minimising the area between the empirical covariance (straight line) and the theoretical one (dashed line). While the exponential spatial estimate seems to adequately fit the empirical one, discrepancies in the temporal counterpart are due to the cyclic behaviour of the temporal intensity, which is reflected in the empirical autocovariance function estimating also negative values. Furthermore, in Figure 4.13 we represent empirical and theoretical spatial covariance functions at different time lags. We have chosen to consider two-time instants, namely 10 a.m and 2 p.m., which are the densest hours. For each of them, we select three temporal lags, namely 1, 2 and 3, and compute the spatial covariances. It is evident that the spatial covariance strongly depends on time. In particular, for the morning hours, the model seems to achieve a better fit as the temporal lag increases (conversely, for the afternoon hours). From an interpretational point of view, this is a reasonable result, indicating, as expected, that the spatial covariance is higher in the afternoon when visitors are less dispersed and tend to return to their gathering site. From a theoretical point of view, this confirms our hypothesis, that the phenomenon under study could not be explained only by a purely separable first-order intensity function $\lambda(\mathbf{u}, t)$ that does not take into account the spatio-temporal interactions, as $S(\mathbf{u}, t)$ does.

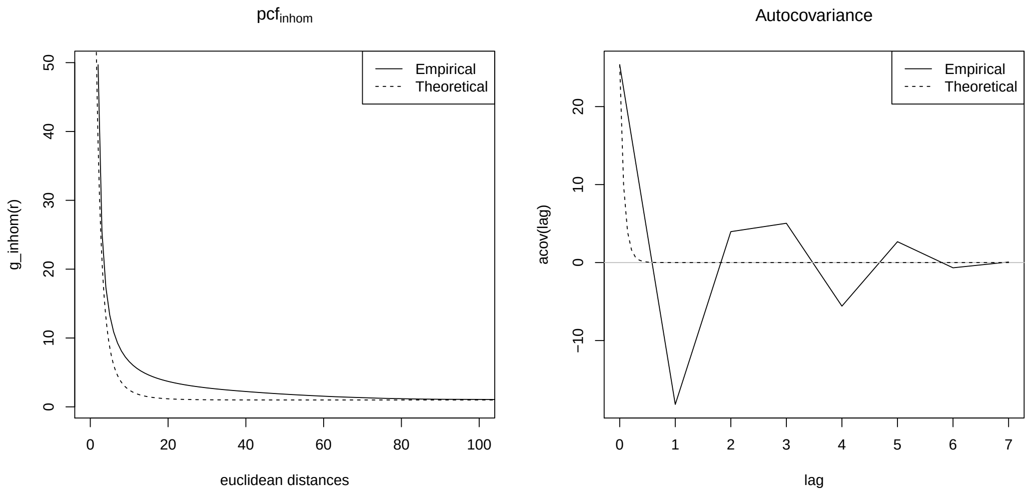


Figure 4.12: Empirical and theoretical purely spatial and temporal covariances. The range of the spatial lags is restricted up to 100 meters, and the temporal lags up to 7 days.

As for the inhomogeneous Poisson case, diagnostics of the fitted LGCP is performed, considering the modified algorithm in section 2.2.3.1.2, choosing a small value $\lambda_{max} = 0.001$ for the simulation of the spatio-temporal point patterns. In Figure 4.14, the inhomogeneous K-function estimated from the data is displayed in white, and the envelopes, coming from $Q = 99$ simulations of inhomogeneous LGCPs and at a significance level of $\alpha = 0.05$, are displayed in grey. We now see that the envelopes of the processes simulated from the fitted LGCP model include the empirical K-function, indicating a much better fit compared to that in Figure 4.9. This, together with a p-value of 0.19, indicates an improved fitting and supports the LGCP model as a good one to explain the spatio-temporal clustering of the visitors' stops.

4.1.4 Conclusions and discussion

In this section, we have described GPS referred to tourists visiting the downtown of Palermo, one of the most touristic cities in Italy, with its many touristic attractions, belonging to the UNESCO-World Heritage listing. Assuming that the visitors' stops may represent a spatio-temporal point pattern, we have estimated a spatio-temporal first-order intensity on the network represented by the streets of downtown Palermo. According to this approach, we describe the phenomenon under study taking into

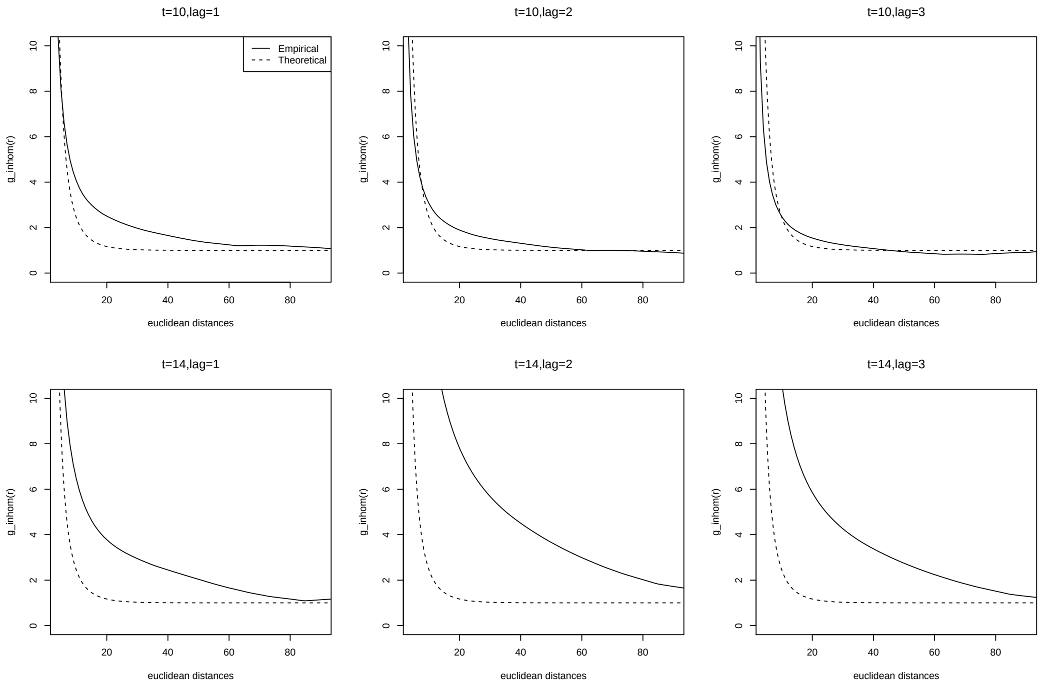


Figure 4.13: Empirical and theoretical spatial covariance functions at different time lags.

account both the support of the location displacements of points and the temporal component. This section contributes to the framework of spatio-temporal point processes on linear networks in different aspects.

First, we have described visitors' stops proposing new parametric approaches in the context of spatio-temporal point processes on the network, through a parametric spatio-temporal separable intensity function. For the purely spatial intensity, we have accounted for the characteristics of each individual considering a Gibbs process with mixed effects. We have chosen the ID of each visitor as a mark and we have fitted random effects corresponding to each individual. As the model is a pairwise interaction process, we have included a smooth interaction function depending on the shortest-path distance between pair of points, to identify the degree of interaction among points. Furthermore, we have also included in the point process description the effect of spatial covariates, e.g. the distance from the nearest point of touristic attraction. The results of the proposed analysis suggest that the random effects for the individuals have a significant variance, meaning that the probability of stopping at different spots in the city centre depends on the subject-specific individual characteristics. Fixing the interaction radius to 100 meters, we find that the interaction among points is positive, meaning that visitors tend to stop in the same locations.

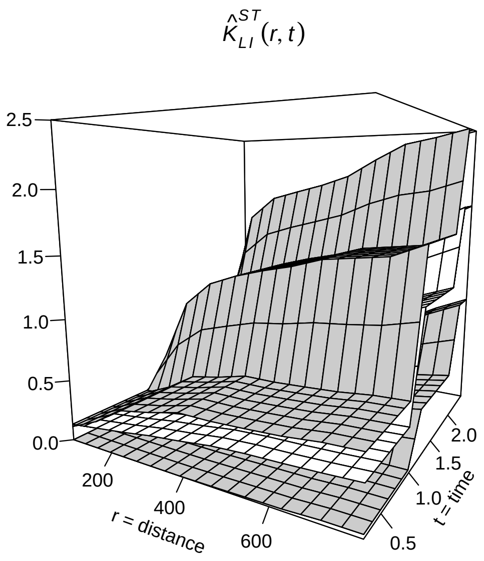


Figure 4.14: In white*: Estimated inhomogeneous K-function for visitors' stops data.* In grey*: envelopes based on 99 simulations from inhomogeneous LGCP spatio-temporal point processes at a significance level of 0.05.*

Finally, the effect of the distance from the nearest touristic attraction is significant and negative, that is to say, as expected, people tend to stop closer to well-known touristic attractions. For the temporal component, we have found cyclic behaviour during the day, with periodic peaks up to 6 hours.

After having estimated the separable spatio-temporal intensity we also assess the overall goodness-of-fit of the spatio-temporal model finding some residual clustered structures, not accounted for by the inhomogeneous model. In particular, we have identified the small spatial and temporal ranges for which the points of the inhomogeneous point pattern still show some clustering degree, and this has led us to consider a more complex model, an LGCP, driven by an underlying log Gaussian random field (log GRF), allowing the representation of point aggregation depending on the properties of the GRF. In detail, estimating the covariance parameters, we have found that the LGCP model achieves a better fitting to the analysed spatio-temporal point patterns, and correctly captures all the residual clustered structures.

A crucial issue when considering an LGCP on the network is the choice of the most appropriate distance metric. In this work, given the small difference between the observed Euclidean and shortest-path distances, we fit the LGCP model by minimum contrast, based on the intensity estimated on the whole plane, and restricting the results only on the network. Future works could deal with the development of methods able to fit LGCPs on networks. Indeed, even if the estimation of the covariance parameters by minimum contrast can be easily adapted to linear networks considering inhomogeneous versions of the pair correlation function and the K-function on the networks, the generation of the Gaussian Random Field on the network remains an open question.

Our approach results in the possibility of both improving the fitting to the data, as well possibility of identifying the main determinants of the spatial displacement of points. The main contribution of the latter to our application is to give a better knowledge of the determinants of spatial intensity of cruise passengers' stop locations during their visit in urban contexts, which may orient destination management policy.

Indeed, by knowing in advance (based on predictions) the behaviour of the visitor's stops in terms of their spatial arrangements and temporal counts, policymakers and police can anticipate these movements of visitors throughout the city.

Finally, the proposed modelling strategies, i.e. the purely spatial Gibbs point process model and log-Gaussian Cox processes fitted on the network, could be useful in analyzing different data settings and problems describing any phenomenon with a complex spatio-temporal dependence occurring on a linear network. Indeed, our methodology can be easily adapted to analysing crime data or public health diseases, as examples of phenomena evolving in space and time.

4.2 Self-exciting point processes on linear networks for crime data

D'Angelo, N., Payares, D., Adelfio, G., and Mateu, J. (2022f). Self-exciting point process modelling of crimes on linear networks. *Statistical Modelling.* https://doi.org/10.1177/1471082X221094146

A number of papers have dealt with the analysis of crime data using the Self-Exciting point process theory after the analogy drawn by Mohler et al. (2011) between aftershock ETAS models and crime. In particular, several papers have proposed a Hawkes-type point process modelling framework for crime data, as this type of data is usually clustered (Reinhart, 2018; Park et al., 2021). Recently, Zhuang and Mateu (2019) proposed a spatio-temporal Hawkes-type point process model, which includes a background component with daily and weekly periodisation and a clustering component that is triggered by previous events. Their model is used to describe the occurrences of violence or robbery cases in an urban environment during two years, and their results show that robbery crime is highly influenced by daily life rhythms, revealed by its daily and weekly periodicity, and that about 3% of such crimes can be explained by clustering.

As crime events are naturally constrained to occur on the street structure of a city, in this analysis, we advocate the use of the theory of point processes on linear networks. Most of the literature about spatial and spatio-temporal point processes on networks is concerned with the non-parametric estimation of the first-order intensity (Moradi et al., 2019; Moradi and Mateu, 2020; Mateu et al., 2020). However, only a recent example is given in section 4.1 of this thesis, dealing with parametric intensity specification of inhomogeneous first-order intensities on networks to analyse the spatio-temporal distribution of visitors' stops by touristic attractions in Palermo (Italy). This latter work only considered inhomogeneous Poisson models and did not take into account the potential Self-Exciting behaviour of points. Other recent papers that deal with the model building of non-Poisson models are the following. Baddeley et al. (2017) adapt to linear networks popular procedures for constructing a point process, such as the Switzer-type pseudostationary process, Cell process, and Cluster processes. Van Lieshout (2018) defines nearest-neighbour point processes on graphs with Euclidean edges and linear networks, which can be seen as analogues of renewal processes on the real line. Rasmussen and Christensen (2021) adapt the so-called conditional intensity function used for specifying point processes on the timeline to the setting of directed linear networks, considering specific classes of point process models as Poisson processes, Hawkes processes, non-linear Hawkes processes, self-correcting processes, and marked Hawkes processes, used in that paper to analyse simulated and neurological data.

Following these considerations, as none of the above-mentioned papers about point process modelling of crime has proposed models taking into account the road network geometry, the aim of this section is to analyse crime data with Self-Exciting point processes while also accounting for the underlying network structure where events occur. Statistical analysis of network data presents severe challenges (Baddeley et al., 2020). A network is not spatially homogeneous, which creates geometrical and computational complexities and leads to new methodological problems with a high risk of methodological error. Real network data, such as crime data, can also exhibit an extremely wide range of spatial scales. These problems pose a significant challenge to the classical methodology of spatial statistics based on stationary processes, which is largely inapplicable to data on a network. Note also that the choice of distance metric on the network is pivotal in the theoretical development and in the analysis of real data. This is thus a key aspect when treating robberies that happen on the streets of a city to better model and understand the true spatial and temporal structures underpinning such types of crimes. Indeed, regarding the intensity estimation issue addressed in this section, considering the geometry of the network in the fitting procedure represents a reasonable choice, given the increasing availability of road network data and computational resources nowadays. For instance, from a practical point of view, the predicted intensity could be obtained only where events are driven to occur. Of course, crimes are an obvious example of constrained point patterns, and therefore, the impact of considering the road network structure in the inference procedure potentially reflects on more detailed information, which may then drive policy actions. In detail, we analyse robbery crimes that occurred in the city of Bucaramanga (Colombia) in 2018. We fit a model similar to the semiparametric specification employed in Zhuang and Mateu (2019) into the network case. Therefore, the first main contribution of the current section regards the proposal of an extension of the Hawkes model proposed by Zhuang and Mateu (2019). We do not only include the network geometry in the fitting procedure but also draw conclusions on the scale of shortest-path distances, typically more appropriate when dealing with point processes occurring on linear networks. We find that our proposed model achieves a much better fit when compared to the planar counterpart, allowing us to better interpret the results.

Examples of applications incorporating the external information in Self-Exciting models can be found in Schoenberg (2016); Reinhart (2018); Adelfio and Chiodi (2020). In particular, Park et al. (2021) modelled gang-related violent crimes in Los Angeles (California) using spatio-temporal Hawkes processes, proposing an algorithm to estimate the spatio-temporal varying background rate non-parametrically as a function of demographic covariates. Therefore, the second contribution of this section concerns a further extension of our proposed model by including external covariates in the purely spatial background component, following the specification of Self-Exciting models that Meyer et al. (2012) proposed in an epidemiological context. In spatial point process theory, the spatial covariates are referred to as those variables with observable values, at least in principle, at each spatial location in the spatial window. For inferential purposes, their values must be known at each point of the data point pattern and at least at some other locations. Therefore, in the crime data context, those would comprehend socio-economic characteristics of the analysed regions, as in Park et al. (2021) such as education level, illiteracy, access to public services like water, electricity, sewerage, and unemployment, housing quality, socio-economic status. Other available covariates could be related to the distance to and density of facilities per street (such as police stations, hospitals, and schools), environmental characteristics (slope, PM10), and street characteristics (direction, type, structures). In this section, we consider spatial covariates in the background component, even though the specification of the proposed model would also allow for the inclusion of marks in the triggered component (see Adelfio and Chiodi (2020) for their proposal in the seismic context, and Chiodi et al. (2021) for an application).

We find that the inclusion of some spatial covariates in the background component further improves the fitting of the model, and therefore it lays the bases for future developments in this promising direction, such as the inclusion of individual-related covariates into the triggering component.

All the codes are available through the link `http://www.statmod.org/smij/archive.html`, together with the data analysed.

The structure of the section is as follows. Section 4.2.1 presents the data and the motivating problem. Section 2.2.3.2 provided an overview of Self-Exciting spatio-temporal point processes. The novel methods introduced in the work are detailed in section 4.2.2. Section 4.2.3 presents the data analysis, and section 4.2.4 is devoted to conclusions.

4.2.1 Data and motivating problem: crimes in Bucaramanga

Reports of crimes from January 2010 to September 2020 in Bucaramanga (Colombia) were collected by the city Home Secretariat and the Mayor's Office of Security and Defense. The reports contain spatial location and time of occurrence of crimes that happened in Bucaramanga, together with some further information such as type of crime, severity, victim characteristics, victimiser's crime weapon, and transportation. The data aims primarily to study the behaviour in space and time of reported crimes in Bucaramanga to define the location and population-based public policies to mitigate the city's reported crime rates. Further information about the data can be found at the Municipal Digital Observatory of Bucaramanga[1].

We note that in crime data, there are naturally unreported crimes and some falsely reported crimes. The latter is not our case (or at least there potentially are very few false cases) as the data we work with were double-checked by the Home Secretary before releasing this info to the public instances. It is an aspect that we

---

[1] Municipal Observatory of Bucaramanga: Crimes in Bucaramanga

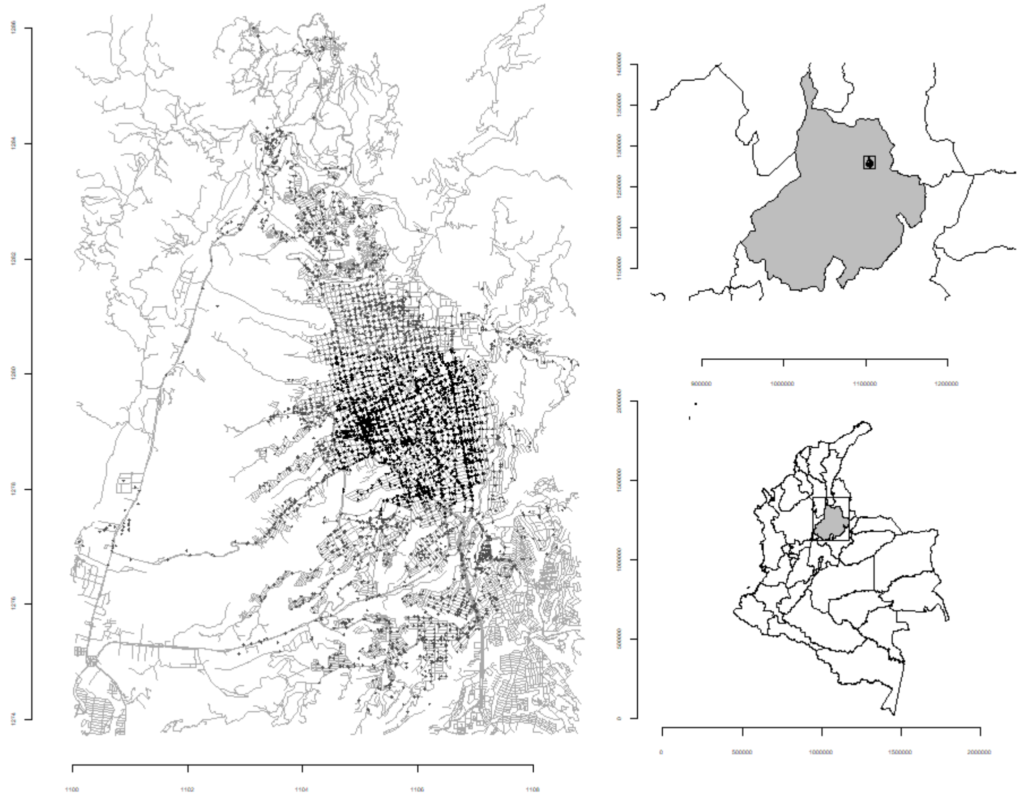


Figure 4.15: The black points represent the armed robberies in Bucaramanga, and the dark grey points the crimes in the city's downtown. In light grey, the segments of the streets of Bucaramanga city, obtained from OpenStreetMap

have to live with and assume that what we observe is just a noisy (subset) version of the real number of crimes. So we are dealing with reported crimes rather than with the total population of crimes. We also note that the events we focus on here (armed robberies) recorded in the datasets often respond to complex historical police-community interactions, which means that the data collection process is possibly influencing the reported cases. Again, this could pose a biased sampling problem and this comes back to the fact that what we observe is a noisy, biased version of reality.

Following Zhuang and Mateu (2019) analysis, we bound our data to only robbery-related reported crimes, as these types of crimes are often encountered and they make the population and society being afraid and scare of moving freely within the city. Furthermore, we specified spatial and temporal windows to filter the original data for computational cost and statistical representativeness purposes. The subset data of this study comprises a spatio-temporal point pattern consisting of 2671 armed robberies in downtown Bucaramanga city from the 1st of January 2018 to the 31st of December 2018. Armed robberies have sharply increased in Bucaramanga over the last ten years, becoming the leading crime affecting the city's inhabitants. Given this crime frequency and situational characteristics, the local police constantly struggle to combat and prevent their occurrence. Efforts to reduce the armed robberies burden in Bucaramanga have been developed mainly in commercial and residential areas where robbery rates are intimidating. For instance, the city's downtown area accounted for 40-60% of annual reported armed robberies in Bucaramanga between 2010 and 2020. By 2018, the figures in downtown reached almost 55% of all robbery reported crimes. We selected armed robberies in the downtown region for 2018 as a representative sample of robbery-related crime dynamics in Bucaramanga, as the downtown region contains most of the city shops and facilities. Figure 4.15 displays the armed robberies for the entire city of Bucaramanga (dark grey) and its downtown (black) in 2018. In this study, we focus on analysing the latter subregion.

Since the reported crime locations were slightly shifted from the street configuration, we relocated the points to the closest location on the linear network. Figure 4.16 shows the georeferenced coordinates of the crimes as reported in the original database (left panel) and the relocated coordinates matching the linear network L (right panel). Figure 4.17 displays the location of the armed reported crimes per month in 2018. We note how the number of reported crimes increases throughout the year and concentrates east of the city's downtown. The linear network L is composed of 3,136 nodes and 4,290 segments referring to Bucaramanga's downtown. The downtown area covers 23 commercial and ten residential neighbourhoods containing most of the city's shops and facilities. Likewise, the Mayor's Office of Bucaramanga estimates that at least 60% of its inhabitants commute daily to this area for professional, academic and touristic activities.

4.2.1.1 Spatial covariates

Demographic, socio-economic, geographical, and environmental variables were obtained at the block level for the city of Bucaramanga. The data was provided by the UN-Habitat Colombia, the United Nations program for human settlements and sustainable urban development in Colombia. The UN-Habitat Colombia develops multilayered indexes to assess socio-economic and demographic aspects related to urban planning. We used all 36 variables, including socio-economic factors such as unemployment rate, education level, literacy rate, public services coverage, socio-economic strata, and housing quality; demographic aspects, such as total population, gender ratio, and elderly population; environmental conditions, for instance, street slope and street type; and geographical covariates such as distances to facilities, the density of facilities per 10,000 inhabitants, among others. Some variables are of continuous nature (such as a percentage) or others are just informing of the presence /absence of a characteristic (for example, the existence of a tunnel in a street). For

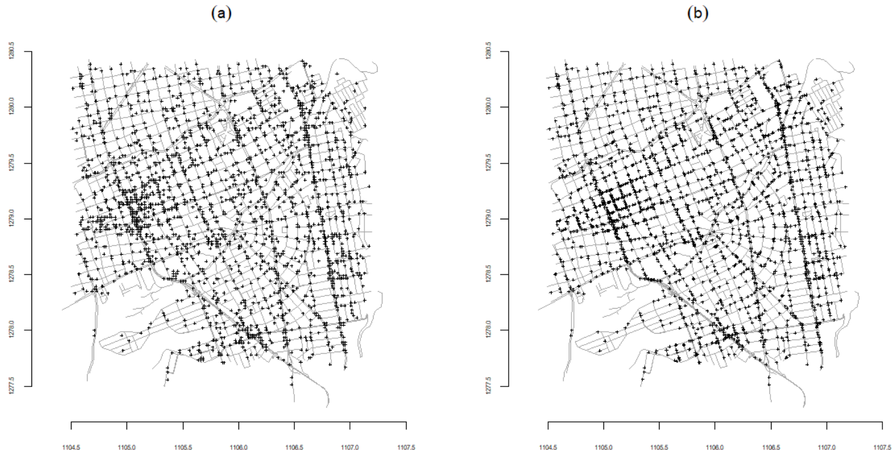


Figure 4.16: (a) Original locations of points and (b) locations relocated on the linear network

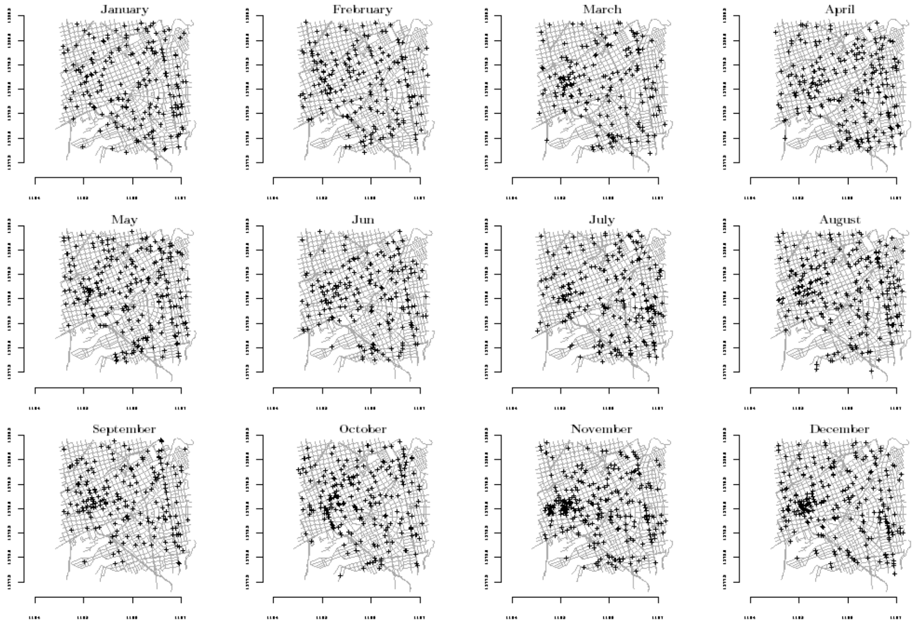


Figure 4.17: Armed robberies in Bucaramanga's downtown per month, 2018

reasons of space, Figure 4.18 displays six out of the thirty-six variables employed in our analysis. As shown in the figure, the covariate value is known in each segment of the linear network under analysis. Some of them are continuous variables, such as literacy rate in panel (a), and some are binary, such as bridges, in panel (f).

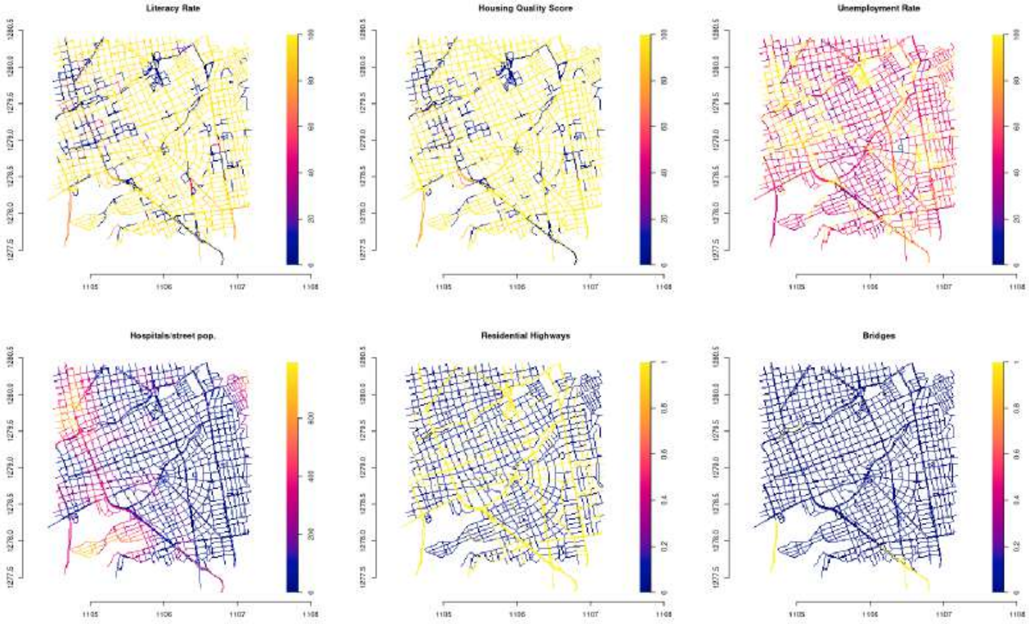


Figure 4.18: Some socio-economic, demographic and environmental spatial covariates used

As the covariates were georeferenced at the block level, except for the geographical ones, we translate the values into the network by assigning the median value of its neighbouring blocks to each segment. We manually computed the values for highly heterogeneous areas of the city. For example, the Northwestern area of Bucaramanga's downtown possesses the richest and poorest zones of the city. The streets of wealthy neighbourhoods can be connected to streets belonging to poverty-stricken blocks.

4.2.2 Model proposal

To adapt (2.34) on the underlying spatial network, the main issue is to choose estimators for the spatial components $\hat{\mu}_b(x,y)$ and $\hat{h}(x-u,y-v)$, taking properly into account the underlying network structure. In section 2.3.2.2, we have reviewed the most recent proposals for intensity estimation on linear networks.

4.2.2.1 A Hawkes point process model on linear networks

The spatio-temporal Hawkes process model on the linear network L that we propose has the following specification

$$\lambda(t,x,y)=\mu_0\mu_t(t)\mu_w(t)\mu_L(x,y)+A\int_{-\infty}^{t_-}\int\int_L g(t-s)h_L(x-u,y-v)N(\mathrm{d}u\times\mathrm{d}v\times\mathrm{d}s), \tag{4.6}$$

where $\mu_L(x,y)$ and $h_L(x-u,y-v)$ are computed using the 2D convolutional Gaussian Kernel of Rakshit et al. (2019b), which represents the best alternative for our purposes. Indeed, Rakshit et al. (2019b) proposed a kernel estimator on a linear network based on a 2D smoothing kernel. The original motivation was speed: this estimator can be expressed in terms of 2D convolutions of the kernel, so it can be computed very rapidly using the FFT.

Definition 4.1. ***Convolution Kernel Estimator***
Let $\boldsymbol{x}=\{\boldsymbol{x}_1,\dots,\boldsymbol{x}_n\}$ be a point pattern on a linear network L. Let κ denote a bivariate kernel function, that is, a probability density on $\mathbb{R}^2$. The convolution kernel estimator of the intensity is, with the uniform correction,

$$\hat{\lambda}^U(\boldsymbol{u})=\frac{1}{c_L(\boldsymbol{u})}\sum_{i=1}^{n}\kappa(\boldsymbol{u}-\boldsymbol{x}_i),\quad \boldsymbol{u}\in L, \tag{4.7}$$

and with Jones-Diggle correction, it becomes

$$\hat{\lambda}^{JD}(\boldsymbol{u})=\sum_{i=1}^{n}\frac{\kappa(\boldsymbol{u}-\boldsymbol{x}_i)}{c_L(\boldsymbol{x}_i)},\quad \boldsymbol{u}\in L, \tag{4.8}$$

where

$$c_L(\boldsymbol{u})=\int_L \kappa(\boldsymbol{u}-\boldsymbol{v})\,d_1\boldsymbol{v},\quad \boldsymbol{u}\in L. \tag{4.9}$$

The denominator $c_L(\mathbf{u})$, defined in Equation (4.9), is the convolution of the kernel with the arc-length measure on the network. This function is evaluated only at locations on the network.

It is theoretically possible to choose a kernel κ that is not isotropic, that is, not invariant under rotation. This seems undesirable in practice, except in situations where the coordinate system is not isometric, such as the latitude-longitude coordinates on a globe. We thus assume here that κ is isotropic. Unlike estimators of the intensity based on path distances in the network, the convolution estimators are robust against errors in the geometry of the network. If κ is uniformly continuous, the quantities (4.7) and (4.8) are continuous functions of the point pattern $\mathbf{x}$ and the linear network L, and the sums in (4.7) and (4.8), and the integral (4.9) can be recognised as convolutions of the kernel κ with different measures M on $\mathbb{R}^2$ (Rakshit et al., 2019b).

Therefore, the fitting procedure for our model in Equation (4.6) follows the same strategy outlined in section 2.2.3.2.1 with the main modification concerning the planar spatial intensities $\mu_b(x,y)$ and $h(x,y)$ that are substituted by their network counterparts $\mu_L(x,y)$ and $h_L(x,y)$, computed using the 2D convolutional Gaussian Kernel of Rakshit et al. (2019b), and implemented in Baddeley and Turner (2005) package.

4.2.2.2 Further extensions: dependence on external covariates

The model in Equation (4.6) can be extended by including external spatial covariates in $\mu_L(x,y)$. In the context of spatial point processes, if $Z(x,y)$ is referred to as a spatial covariate, this means that its value is assumed to be observable, at least in principle, at each location (x,y) in the region of interest. For inferential purposes, its values must be known at each point of the data point pattern and at least at some other locations. This specification gives rise to

$$\begin{aligned}\lambda(t,x,y) &= \mu_0\mu_t(t)\mu_w(t)\mu_L(x,y,\boldsymbol{\beta}_{back}) \\ &\quad + A\int_{-\infty}^{t-}\int\int_L g(t-s)h_L(x-u,y-\nu)N(\mathrm{d}u\times\mathrm{d}\nu\times\mathrm{d}s),\end{aligned} \tag{4.10}$$

where $\boldsymbol{\beta}_{back}$ denotes the parameters associated to the spatial covariates $Z(x,y)$ included in the model.

Note that if the main interest would be focused on the selection of the covariates, the algorithm in section 2.2.3.2.1 would have been modified in order to maximise the likelihood also with respect to the parameters of the external covariates. Otherwise, as we are interested in achieving the best fitting, we use the same algorithm, and the selection of the variables can be performed during the setting of the first guesses, and those can be held fixed in the E-M algorithm.

As all the available covariates are continuous in space, these can be included linearly by choosing a basis function. If $b_i(x)$ is the i^{th} such basis function, then the function f is assumed to have a representation

$$f(x) = \sum_{i=1}^{q} b_i(x)\beta_i, \tag{4.11}$$

for some values of the unknown parameters, β_i.

The degree of smoothing is controlled by the basis dimension. In this section, we keep the basis dimension fixed at a size a bit larger than it is believed could reasonably be necessary. After having compared different numbers of knots, these are chosen to be equal to 5 for the available external covariates, and equal to 30 for the spatial coordinates, estimated here through thin plate regression splines.

4.2.3 Data analysis

4.2.3.1 Model selection

As presented in section 4.2.1, we analyse 2671 armed robberies in the city of Bucaramanga, Colombia, in 2018. We fitted Zhuang and Mateu (2019) model (recall this is the (2.34) model on Euclidean plane) and our proposed extensions (Equations (4.6) and (4.10)) to the data. Table 4.3 displays the bandwidths of the background events and the windows of the triggered events that were adopted for the models. These values are identical in every model to ensure comparability and inference. Note that we omitted the daily periodicity term $\mu_t(t)$, and thus its bandwidth, as our data lacks hourly crime data information. The weekly periodicity and long-term bandwidths were selected to secure the resolution requirements of the temporal components. We chose the spatial and temporal extent (15 days and 2.5 km, respectively) to restrict the spatio-temporal domain in which we expect non-spontaneous events to occur.

Table 4.3: Temporal bandwidths for the background events and spatial and temporal windows for the triggered events

Background	Bandwidth
Weekly periodicity	0.4 days
Long-term	0.6 days
Triggered	**Window**
Temporal	15 days
Spatial	2.5 km

We estimated the relaxation coefficients $\hat{\mu}$ and $\hat{A}$ through a 40-loops iterative algorithm as in Zhuang and Mateu (2019). We also computed the log-likelihood to assess the fit of the three space-time point process models. Table 4.4 reports the estimates of the relaxation coefficients and the corresponding log-likelihood. $\hat{\mu}$ and $\hat{A}$ values reached convergence after the 25th iteration in all approaches (Figure 4.19). The convergence threshold was established as the iteration on which the difference between the values of the k^{th} and $k^{th}-1$ iteration is smaller than 0.0001 units.

The estimates of $\hat{\mu}$ have similar magnitudes in the planar model and linear network model without background covariates. The effect of the covariates in the estimation of $\hat{\mu}$ is evident. The value in the model with covariates, model in Equation (4.10) decreases significantly compared to its counterparts. A higher value of $\hat{\mu}$ for the linear network model in Equation (4.6) compared to the model in the planar case, reflects the change of spatial support, that is, the expected number of events increases when estimated per linear segment instead of per surface. Nonetheless, the inclusion of covariates adjusts for the change of support forcing $\hat{\mu}$ to drop. Note that the convergence rate of $\hat{\mu}$ for the model in Equation (4.10) is smoother than the one of the model without covariates.

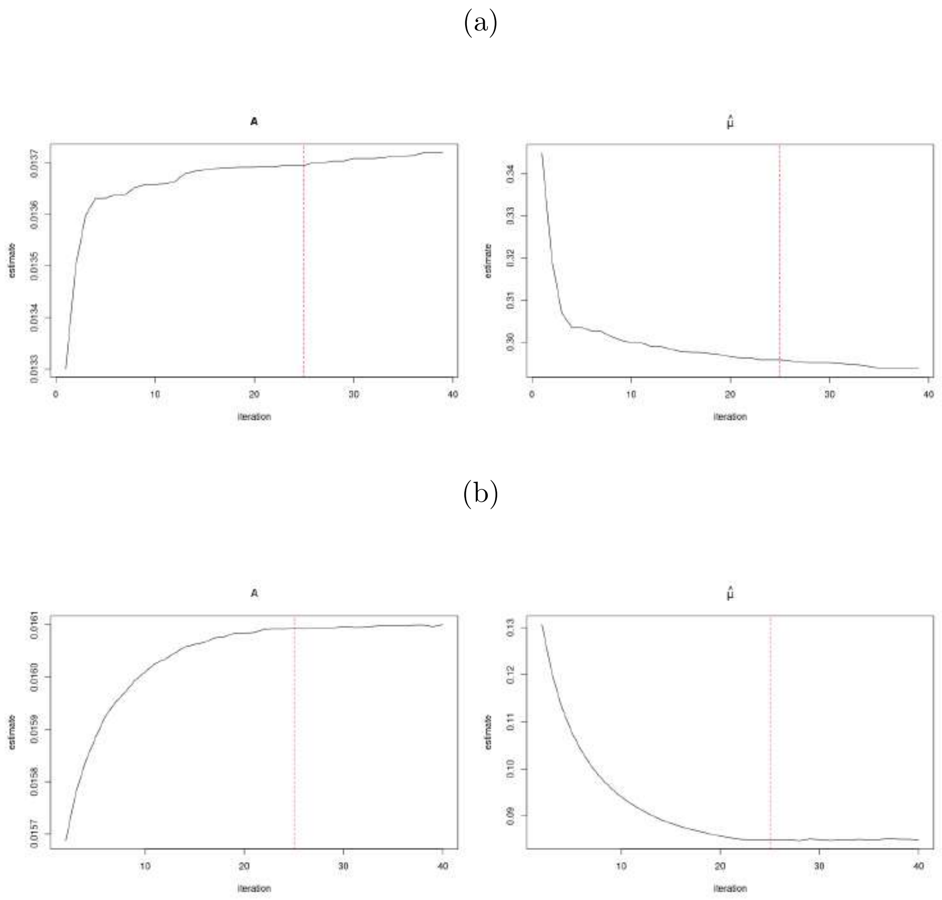


Figure 4.19: Relaxation coefficients convergence: (a) Model (4.6) *and (b) Model* (4.10)

$\hat{A}$ changes noticeably from specifications (2.34) to specifications (4.6) and (4.10). In the former, almost 64% of the crimes are triggered, while the latter informs that only 1.4% and 1.6% (respectively) of armed robberies are provoked by previous crimes. The $\hat{A}$ estimate of our models better resembles Bucaramanga's robbery offences scenario: robberies typically obey the crime opportunity theory, to wit, a robber's motivation arises from the context and factors involved in the environment they are situated (Larrota et al., 2017). One would expect a low triggering effect as the offenders rather exploit the situational factors (spontaneous events) than premeditate their crimes' circumstances (non-spontaneous events).

The log-likelihood shows that our models fit better the armed robberies data with a difference of 353.12 and 386.77 units for models (4.6) and (4.10), respectively. We fit both the background and the triggering effect directly on a linear network, and we get accurate estimates and model the data appropriately. Furthermore, introducing covariates in the background rate improves the estimation of the relaxation coefficients and the overall fitting of the model.

Table 4.4: Comparative results for the three fitted models

Model	$\hat{\mu}$	$\hat{A}$	$\log(L)$
Model on the plane (2.34)	0.204	0.639	-706.47
Model on the network (4.6)	0.292	0.014	-353.35
Model on the network plus covariates (4.10)	0.085	0.016	-319.70

To demonstrate the advantages of the linear network models (4.6) and (4.10), we compare their spatial results with the ones of the planar model (Equation (2.34)). Figure 4.20 displays the spatial background rate for the three Self-Exciting point process models. The spatial background rates share similarities in every model. Although the models' background rates present a close resemblance, the rates in the linear network models are higher than those in the planar one. As we discussed before, the rates increase in the linear network models due to the reduction of the area unit. The distribution of the intensities in Bucaramanga's city centre is equivalent in all models. However, the linear network models provide larger estimated intensities in the central and northwestern regions of the study area. In particular, the specification from Equation (4.10) displays larger and rougher high-intensity areas. This is primarily attributed to the effects of the covariates as these affect the estimates per segment rather than per unit of space. Also, note that the planar model indeed captures the "network" nature of the data, but the clustering effect extends to areas instead of street segments. This might lead to misinterpreting low-crime rate streets as dangerous ones when these connect to neighbouring streets with high crime rates. Furthermore, as the planar model is defined over continuous space, it generates artificial intensities in locations lacking events.

One of the main differences between Zhuang and Mateu (2019) model and our proposed approaches is the definition of the spatial triggering component. Figure 4.21 compares the spatial response function of the planar and network models. The spatial response function of the planar model remains constant across events. The occurrence of an armed robbery in a specific location will excite future armed robberies within a radial distance of approximately 100 meters. This is an unrealistic assumption for events located in a linear network as the extent of the spatial triggering effect depends intrinsically on the street topology. Figure 4.21(b) shows the triggering function for four armed robberies. In our approach, $h_L(x, y)$ adapts to the street configuration on which each event lies. The triggering window on the linear network varies from event to event, preserving the geometrical properties of

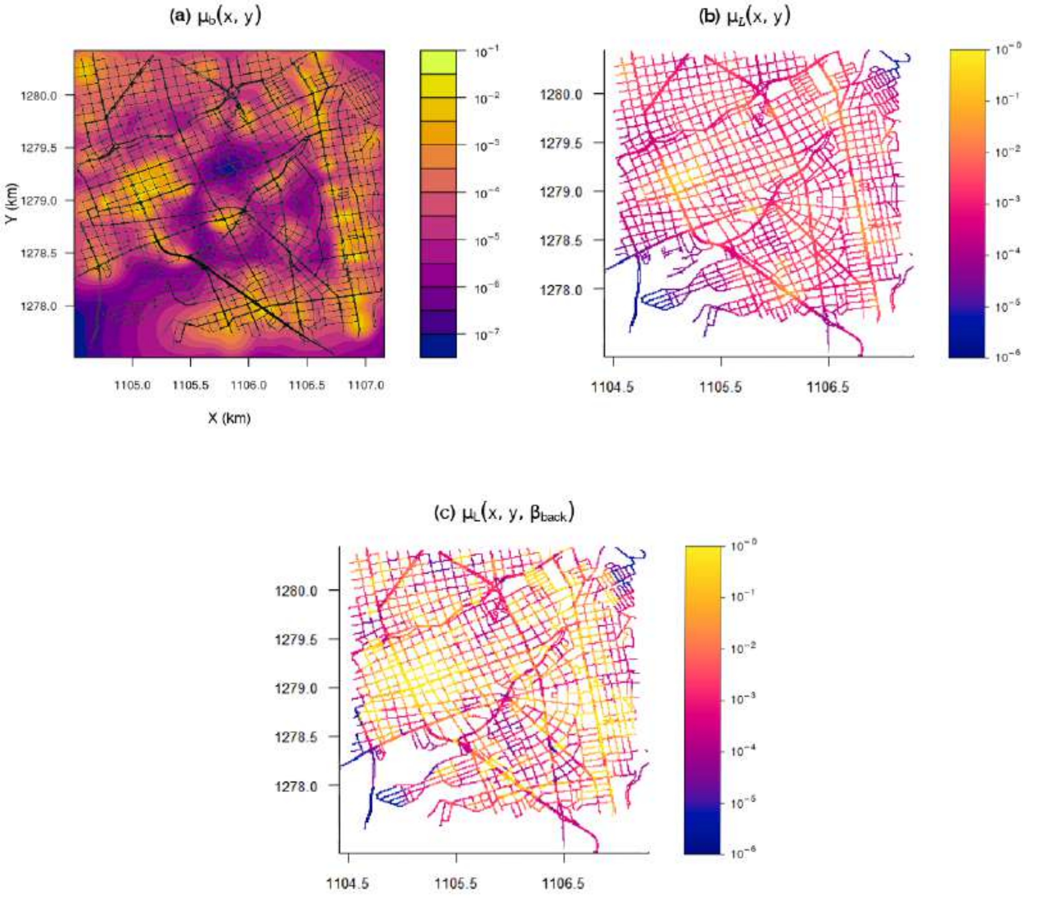


Figure 4.20: Spatial background rate: (a) planar model $\mu_b(x, y)$ (Equation (2.34)), (b) linear network model $\mu_L(x, y)$ (Equation (4.6)) and (c) linear network model with background covariates $\mu_L(x, y, \boldsymbol{\beta}_{back})$ (Equation (4.10)). Rates are reported on the log-scale

network structure, for example, the shortest path distance between events.

Figure 4.22 displays the temporal components of the linear network models. The temporal results in both models were identical. The long-term trend (Figure 4.22(a)) indicates that armed robberies occur primarily in the second semester of the year. The crime rates increase from June onwards and reach their maximum value in December. In Colombia, the period from June to December comprises most of the annual public and school holidays. As the number of tourists increases in the commercial and touristic areas of the country's main cities (such as Bucaramanga), local surveillance and security weaken, promoting higher crime rates. The weekly periodicity implies that robberies occur mainly on Wednesdays, Fridays, and Saturdays. The original data also reports these days as the days with the highest frequencies of armed robberies. On Wednesdays, the city centre's main venues such as restaurants, bars, and cinemas offer discounts on their services, attracting many locals. Fridays and Saturdays are naturally the days when most people commute to the city centre for recreational activities. People leaving venues in late hours and limited public transport provide the offenders with the ideal circumstances to take advantage of. The temporal response function (Figure 4.22(c)) suggests that after an armed robbery occurs, subsequent armed robberies are triggered within the coming ten days. However, the triggering effect lessens after one week.

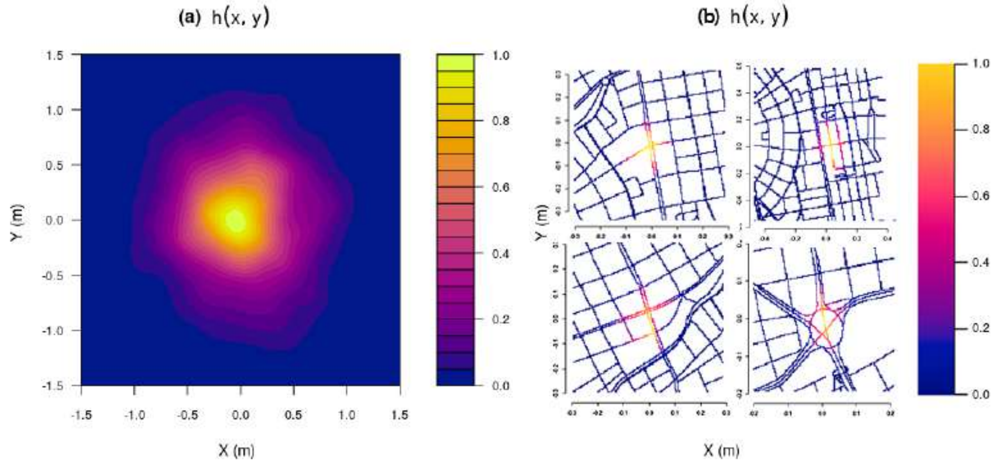


Figure 4.21: Spatial triggering function: (a) planar model $h(x, y)$ and (b) linear network model $h_L(x, y)$

4.2.3.2 Covariates results

As previously introduced, we select the covariates to include in the model in Equation (4.10) prior to the E-M procedure. Among the significant external covariates affecting the background spatial intensity, we have cycleway, pedestrian, primary, residential, secondary, service, and tertiary highways, bridges and steep slopes. Indeed, all of these binary variables are significant in explaining the crime pattern and their estimated coefficients are all positive meaning that the crimes are most likely to occur in those places. Only one-way street and flat slope coefficients are estimated as negative values, meaning that crimes are less likely to occur there.

In Figure 4.23, we show the marginal relationship among the significant contin-

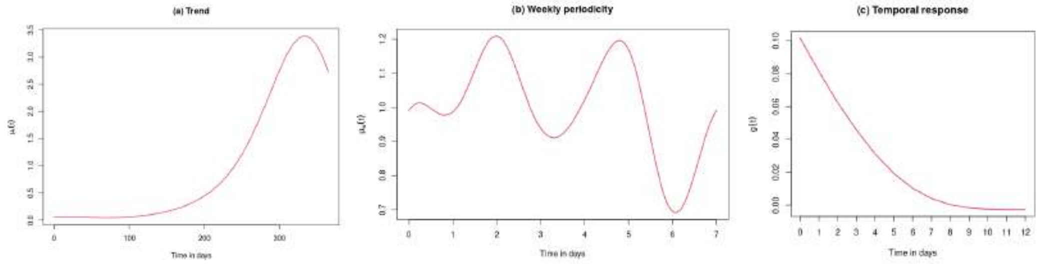


Figure 4.22: Temporal outputs: (a) Long-term function $\mu_t(t)$; (b) weekly periodicity $\mu_w(t)$; (c) temporal response function $g(t)$

uous covariates and the intensity of the point pattern under study. We consider a smoothing procedure based on fixed-bandwidth kernel density estimation (Baddeley et al., 2012), in accordance with sections 3.1.2 and 4.1.3. The non-parametric combination of these variables plus the event coordinates best explains (based on the Akaike criteria) the linear network point pattern.

4.2.3.3 Residual analysis

As stated in some previous papers (Adelfio et al., 2020; Adelfio and Schoenberg, 2009), the main problem when dealing with residual analysis for point processes is to find a correct definition of residuals since the one used in dependence models cannot be used for point processes.

For the temporal domain diagnostics of the fitted model, the marginal time process can be obtained by integrating the estimated intensity function with respect to the observed spatial region (Adelfio and Chiodi, 2015b). Indeed, given a point process $N = \{(t_i, x_i, y_i), i = 1, \ldots, n\}$ which is determined by a conditional intensity $\lambda(t, x, y|\mathcal{H}_t)$, the transformation

$$t_i \to \tau_i = \int_0^{t_i} \int_X \lambda(u, x, y|\mathcal{H}_t) \mathrm{d}x\mathrm{d}y\mathrm{d}u$$

transforms N into a stationary Poisson process N' with unit rate (Vere-Jones and Schoenberg, 2004). The resulting process is called the transformed time sequence (Ogata and Katsura, 1988). Then, a plot of τ_i versus i can give insight into the quality of the fitting in time, concluding that the model has a good fit to the data if the transformed time sequence does not deviate significantly from the standard Poisson process. In particular, this plot, together with a plot of the estimated time intensities, informs on the time at which departures from model assumptions are more evident. See Schoenberg (2002) for details on the bands' construction.

We thus consider this type of residual analysis for the best-fitted model in the previous section, i.e. the model based on Equation (4.10). Figure 4.24 depicts both

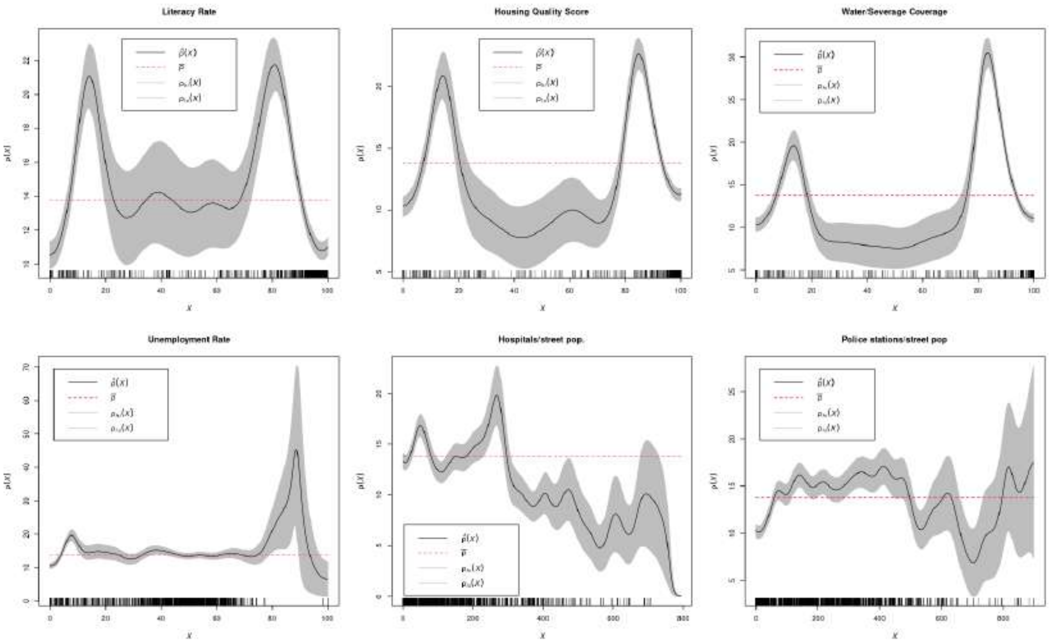


Figure 4.23: Non-parametric estimate of the intensity of the analysed point pattern, as a function of the available variables for the crime data: literacy rate (a), housing quality score (b), water coverage (c), unemployment rate (d), number of hospitals per street population (e), and number of police stations per street population (f). The considered smoothing procedure is based on fixed-bandwidth kernel density estimation

the cumulative frequencies of the original time sequence, and the transformed time for the data for all three models for comparison purposes. We, in particular, follow the recommendation by Zhuang and Mateu (2019) which compare the real and the transformed rates of event occurrences.

Inspecting Figure 4.24, we note that model (4.10) provides the best fit, with the cumulative frequencies lying within the confidence interval at 95%. Although the transformed times deviate from the average rate of occurrence at some points, the overall trend remains consistent. Transformed times diverge from the actual time, mostly in low crime rate periods, for example, in the first months of the year when just a few armed robberies were reported. Furthermore, the fluctuation in the rate of occurrence of the transformed times could be attributed to the kernel estimation.

Similarly to the time component, one of the most used methods for diagnostics of spatial point processes is the smoothed raw residuals, which follow the same line of reasoning (see section 2.1.4.2.1). We employ the same reasoning, knowing that $\lambda^{\dagger}(\mathbf{u})$ is the smoothed version of the estimate of the intensity of the fitted models, and of course $\mathbf{u} \in L$. The difference in Equation (2.17) should be approximately

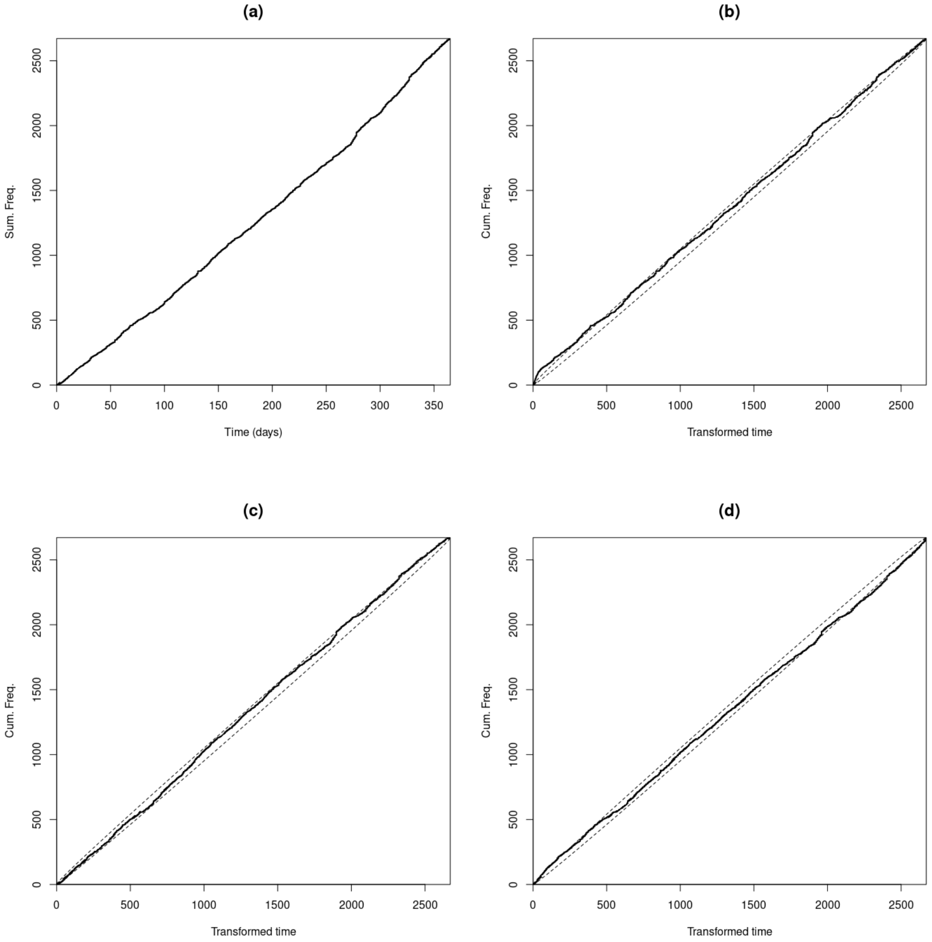


Figure 4.24: Cumulative frequencies of armed robberies events: (a) original occurrence times; (b) transformed times for model (2.34)*; (c) transformed times for model* (4.6)*; and (d) transformed times for model* (4.10)*. The 95% confidence bands under a Poisson process are also shown in dashed lines*

zero when the fitted model is close to the real one. Therefore, the best model is the one with the lowest values of the smoothed raw residuals.

Figure 4.25 depicts the smoothed raw residuals for the two models fitted on the network, and for the model in the planar case, along with their distributions. Even though smoothed raw residuals are most appropriate for parametric specifications of fitted models, we still manage to obtain useful information. Indeed, it appears evident that both models on the network achieve a good fit to the data if compared to their planar version, which tends to overestimate the intensity in several areas.

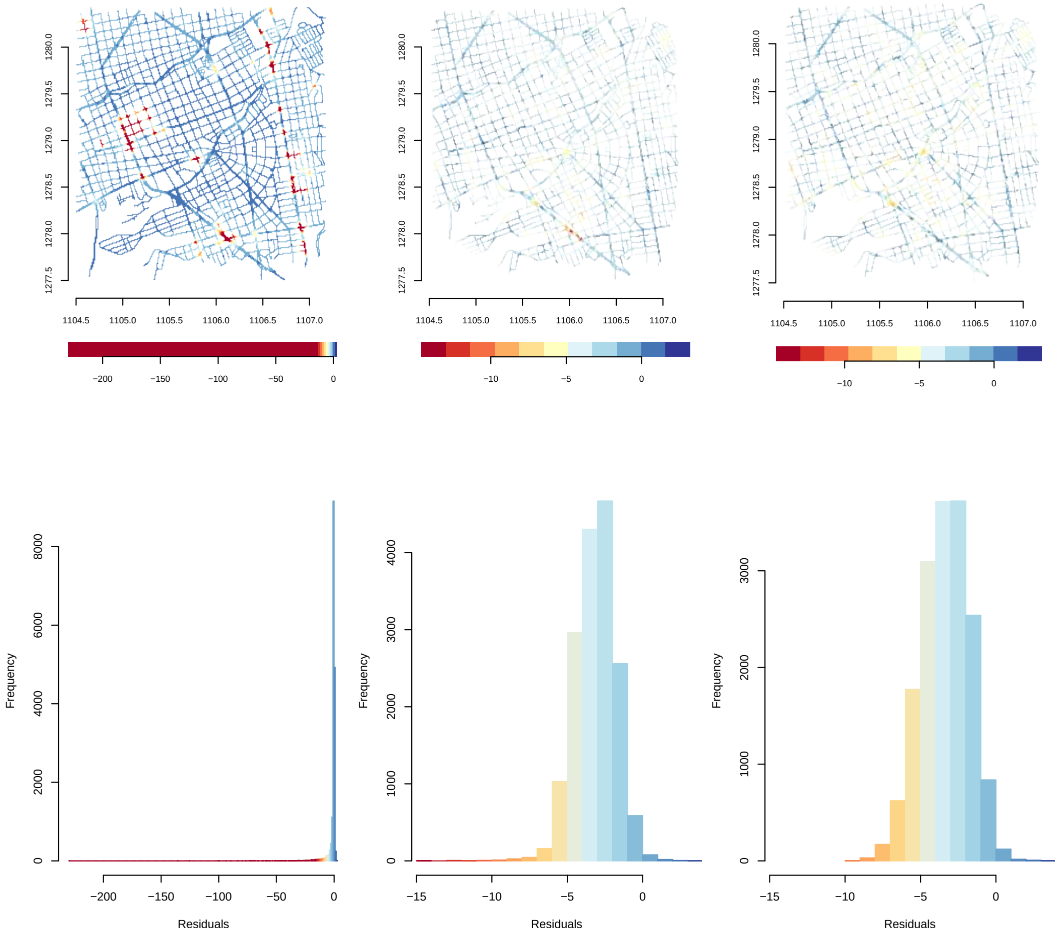


Figure 4.25: Maps and histograms of smoothed raw residuals for models (2.34)*,* (4.6)*, and* (4.10)*, respectively*

Furthermore, the model with covariates provides residuals with a lower range and magnitude compared to the model with no covariates. All in all, the smoothed raw residuals of the model with covariates are overall better than the ones of the model with no covariates, even though both models still represent good competitors if compared to the planar version. We note that there were other options for carrying out diagnostics in this analysis. First, super thinned residuals (Clements et al., 2012), which have the disadvantage of being less straightforwardly applicable, as they require a tuning parameter to be chosen. Another possible option could have been using the weighted spatio-temporal second-order statistics, as in Adelfio et al. (2020), which do not require any transformation of the data. However, we opted for diagnostic procedures separated in space and time, due to the separable specification of the models employed in the analysis. Nevertheless, the chosen diagnostics allows us to interpret separately the contribution of the space specification into the network, as evident from Figure 4.25.

4.2.4 Conclusions and discussion

In this section, we analyse robbery crimes as events of a spatio-temporal point pattern living on a linear network structure. We first fit a Euclidean planar model following Zhuang and Mateu (2019) providing two parameters related to the background and triggered occurrence rates. We further propose an extension of that model in order to take into account the linear network. This allows us both to include the network geometry into the fitting procedure, but most importantly, to draw conclusions on the scale of shortest-path distances, more appropriate when dealing with point processes occurring on linear networks. Furthermore, starting from the specification of this second model, we are also able to include the dependence on external covariates rightly constrained onto the spatial support of the road network. We find that our proposed models on the network achieve a much better fit when compared to the planar counterpart.

As future work, we draw different lines of research, both methodological and applied. From a methodological perspective, the proposed Self-Exciting model could be extended by including external covariates in the triggered components, as proposed by Meyer et al. (2012) in an epidemiological context, and Adelfio and Chiodi (2020) in the seismic one. Indeed, the model could include individual-specific covariates in the triggered component of the conditional intensity function, by means of an exponential linear predictor. This would allow for including the dependence on some specific characteristics of the event i.e. the crime, such as the gender, or the age of the victim.

Another main issue is related to the fitting procedure. Indeed, the proposed algorithm could be modified in order to include the estimation of the covariates parameters at each iteration of the Expectation-Maximization procedure. This would allow considering only the crimes most likely to be background events when selecting the spatial covariates, as done by Park et al. (2021), since, in principle, only background crimes should be used. The best model, and therefore the best set of covariates, could be selected by comparing some information criterion of the fitted competitor models.

Finally, we could follow a multivariate specification analysing several types of crimes in only one modelling strategy.

Chapter 5

Local Indicators of Spatio-Temporal Association on linear networks

In this Chapter, we introduce the Local Indicators of Spatio-Temporal Association functions on linear networks, focusing on the ones based on the local K-functions.

While most of the literature about point processes on networks is concerned with the non-parametric estimation of the spatio-temporal intensity (Moradi et al., 2019; Mateu et al., 2020; Moradi and Mateu, 2020), some papers have recently addressed second- and higher-order characteristics of point patterns occurring on linear networks (Moradi and Mateu, 2020; Cronie et al., 2020). However, none of these approaches has considered local properties of the introduced second-order statistics to measure local first- or second-order structure.

As a starting point, we consider global second-order statistics for spatio-temporal point processes occurring on networks, namely the K-function, developed in the non-Euclidean space of linear networks by Moradi and Mateu (2020), and reviewed in section 2.3.3. We, therefore, consider here the local counterpart of this second-order statistics to measure the local structure of point patterns occurring on network-based structures, proposing Local Indicators of Spatio-Temporal Association (LISTA) functions based on the K-function on linear networks.

The aim of this chapter is to take into account individual contributions of points in some statistical procedures, such as diagnostics, and hypothesis testing, in order to individuate sub-regions where point patterns behave differently in terms of clustering. Specifically, in section 5.1, the proposed LISTA functions are used to assess local differences between the spatio-temporal second-order structure of two point patterns occurring on the same linear network. Moreover, in section 5.2 we show a number of simulations to test their performance in diagnosing spatio-temporal models fitted to point pattern data on networks. Details on the simulation schemes implemented and used in this chapter come in section A.

Therefore, we propose the *local spatio-temporal inhomogeneous K-function for the i-th event $(\boldsymbol{u}_i, t_i)$ on a linear network* as

$$\hat{K}^i_{L,I}(r,h) = \frac{1}{|L||T|} \sum_{(\mathbf{u}_i,t_i)\neq(\mathbf{v},s)} \frac{I\{d_L(\mathbf{u}_i,\mathbf{v}) < r, |t_i - s| < h\}}{\hat{\lambda}(\mathbf{u}_i,t_i)\hat{\lambda}(\mathbf{v},s)M((\mathbf{u}_i,t_i), d_L(\mathbf{u}_i,\mathbf{v}), |t_i - s|)}, \tag{5.1}$$

and the corresponding *local pair correlation function*

$$\hat{g}^i_{L,I}(r,h) = \frac{1}{|L||T|} \sum_{(\mathbf{u}_i,t_i)\neq(\mathbf{v},s)} \frac{\kappa(d_L(\mathbf{u}_i,\mathbf{v}) - r)\kappa(|t_i - s| - h)}{\hat{\lambda}(\mathbf{u}_i,t_i)\hat{\lambda}(\mathbf{v},s)M((\mathbf{u}_i,t_i), d_L(\mathbf{u}_i,\mathbf{v}), |t_i - s|)}, \tag{5.2}$$

weighted by the reciprocal of the normalising factor

$$D(X) = \frac{n-1}{|L||T|} \sum_{i=1}^{n} \sum_{i \neq j} \frac{1}{\hat{\lambda}(\mathbf{u}_i,t_i)\hat{\lambda}(\mathbf{u}_j,t_j)}, \tag{5.3}$$

obtaining $\frac{1}{D(X)}\hat{K}^i_{L,I}(r,h)$ and $\frac{1}{D(X)}\hat{g}^i_{L,I}(r,h)$. Basically, we avoid summing up all the points as in the global statistics counterparts, and we denote the individual contribution to the global statistics with the index i.

The homogeneous versions are computed by weighting the second-order summary statistics by the constant intensity $\hat{\lambda} = n/(|L||T|)$, obtaining

$$\hat{K}^i_{L}(r,h) = \frac{1}{\hat{\lambda}^2|L||T|} \sum_{(\mathbf{u}_i,t_i)\neq(\mathbf{v},s)} \frac{I\{d_L(\mathbf{u}_i,\mathbf{v}) < r, |t_i - s| < h\}}{M((\mathbf{u}_i,t_i), d_L(\mathbf{u}_i,\mathbf{v}), |t_i - s|)}, \tag{5.4}$$

and

$$\hat{g}^i_{L}(r,h) = \frac{1}{\hat{\lambda}^2|L||T|} \sum_{(\mathbf{u}_i,t_i)\neq(\mathbf{v},s)} \frac{\kappa(d_L(\mathbf{u}_i,\mathbf{v}) - r)\kappa(|t_i - s| - h)}{M((\mathbf{u}_i,t_i), d_L(\mathbf{u}_i,\mathbf{v}), |t_i - s|)}. \tag{5.5}$$

Following the Euclidean case, we provide the operational definition based on the local second-order spatio-temporal summary statistics on a linear network.

Definition 5.1. *Any local second-order spatio-temporal summary statistic $\lambda^i_{2,L}((\boldsymbol{u},t),(\boldsymbol{v},s))$ computed on an observed point pattern $\boldsymbol{x} = (\boldsymbol{u}_i,t_i), i = 1,\dots,n$ on a linear network L, where $(\boldsymbol{u}_i,t_i) \in L \times T$, that satisfies the following operational definition*

$$\hat{\lambda}_{2,L}((\boldsymbol{u},t),(\boldsymbol{v},s)) = \frac{1}{n-1} \sum_{(\boldsymbol{u}_i,t_i)\in\boldsymbol{x}} \hat{\lambda}^i_{2,L}((\boldsymbol{u}_i,t_i),(\boldsymbol{v},s)), \tag{5.6}$$

where $\hat{\lambda}_{2,L}$ is an estimator of the global second-order intensity function, can be called a LISTA function on a linear network.

Corollary 5.1. *Knowing that the sum of the individual contributions given by the local estimator of the K-function in Equation* (5.4) *are actually equal to the global statistics, Equation* (5.6) *is satisfied, and therefore it can be called LISTA function on a linear network.*

Throughout the thesis, we will only focus on the K-function, but of course, also the pair correlation function can serve as an alternative.

5.1 Assessing local differences between the second-order structure of two point patterns on a network

D'Angelo, N., Adelfio, G., and Mateu, J. (2021b). Assessing local differences between the spatio-temporal second-order structure of two point patterns occurring on the same linear network. *Spatial Statistics*, 45:100534

As local second-order characteristics provide information on how events relate to nearby events, in this Section, we aim at employing the proposed LISTA functions to build a local test to identify differences between the local spatio-temporal second-order structure of two patterns occurring on the same linear network.

Indeed, spatio-temporal clustering analysis is a key aspect of the practical analysis of spatio-temporal point patterns, and the detection of clustering structure in a spatio-temporal point pattern occurring on a linear network is still an unexplored topic.

Thanks to the proposed local test, we are able to explore how individual points are related to their neighbouring events and classify points with similar spatio-temporal local second-order structures, identifying the areas that differ in terms of clustering behaviour. This classification is not possible by using global estimates as they average out all local structures. This chance is particularly relevant when dealing with Self-Exciting point processes specified through a conditional intensity function as the sum of a background component and a triggered one. The estimation procedure for such models commonly requires the need of declustering the analysed events into a background point pattern, which we shall call large-scale term henceforth to avoid confusion, and the triggered component one, typically displaying a more clustered behaviour. Some examples include the analyses of seismic data (Zhuang et al., 2002, 2004; Zhuang, 2006), in which events can be declustered into mainshocks and aftershocks, or crime data in which events can be declustered into primary and secondary events (Mohler et al., 2011; Zhuang and Mateu, 2019), as well as traffic data (Lim et al., 2016; Fan et al., 2018; Song et al., 2018; Kalair et al., 2021). Other recent applications of Self-Exciting processes concern phenomena such as epidemics (Schoenberg et al., 2019; Lesage, 2020), in which both the declustering of primary and secondary events, as well as the accounting for the underlying linear network could be relevant.

Therefore, motivated by the availability of information about the underlying network in which these patterns occur, it appears suitable to develop novel statistical methodologies and computational techniques that take into account the geometry of the linear network. Indeed, as stated by Baddeley et al. (2020), due to geometrical complexities, the analysis of individual data observed along a network of lines is extremely challenging. Moreover, the intrinsic lack of homogeneity in a network militates against the traditional methods of spatial statistics.

In this Section, we provide simulation studies to assess the performance of the testing procedure of local second-order structures. In particular, we consider homogeneous, inhomogeneous, and clustered spatio-temporal point processes to study how the test behaves under these different settings. We expect that the local test performs better as the number of points increases and the clustered structure of the pattern strengthens. Note that we show some inhomogeneous scenarios to reinforce that our test is also helpful and reliable in these cases, even though we are only considering second-order structure.

Nevertheless, real data analysis often requires more complex model specifications. In some real data, such as seismic, traffic or crime events, a clear separation cannot be performed between the so-called "parents" and "children" points, as in clustered point processes. Therefore, in our simulation study, we also consider more complex spatio-temporal point processes, such as Self-Exciting processes, where each point of the pattern can potentially produce offspring.

We also provide an application of the proposed methodology to two traffic-related datasets that are of particular importance for society.

Section 5.1 is structured as follows. Section 5.1.1 is devoted to the proposal of the test for local spatio-temporal second-order structure, and in section 5.1.2 its performance is assessed through simulation studies. Note that the simulation schemes implemented and used for our purposes are reported in Appendix A. An application to two case studies on traffic data comes in section 5.1.3. Conclusions are drawn in section 5.1.4.

5.1.1 Testing for local second-order structure on a network

We now formally define the local null hypothesis $\mathcal{H}_0$ we aim at testing, and outline the statistical testing procedure for local interaction, motivated by Siino et al. (2018b) for the Euclidean context, adapting it to the network case. In this particular context, ignoring the network would result in bias when computing the K-functions, as without the network correction $M((\mathbf{u}, t), r, h)$ (introduced by Ang et al. (2012) for the purely spatial case and extended by Moradi and Mateu (2020) to the spatio-temporal context) that basically counts the number of points lying exactly at the shortest-path distance $r \geq 0$ and the time distance $h \geq 0$ away from the point $(\mathbf{u}, t)$, we would overestimate the number of points in any spatio-temporal lag considered. Note that our test is based on a subtraction of two K functions which are adapted to the geometry of the network. Although the network is the same for both terms of the difference, the events are different, and $M((\mathbf{u}, t), r, h)$ provides different corrections. Thus it is important to consider K-functions over the network rather than on the Euclidean plane.

The aim is to detect significant differences in the local second-order structure of two given point patterns $\mathbf{x}$ and $\mathbf{z}$, both occurring on the same linear network, and being realisations of the point processes X and Z, respectively. Therefore, the proposed testing procedure can be useful for comparing the local behaviour of different types of events occurring on the same network. We test the null hypothesis of no differences in the spatio-temporal second-order structure of the two observed point patterns $\mathbf{x}$ and $\mathbf{z}$ with respect to the i^{th} point $\mathbf{x}_i = (\mathbf{u}_i, t_i) \in \mathbf{x}$, with $i = 1, \dots, n$. Formally, for each point $(\mathbf{u}, t) \in \mathbf{x}$, the hypothesis test reads as

$$\begin{cases} \mathcal{H}_0: & \text{no difference in the local second-order structure of } (\mathbf{u}, t) \\ & \qquad \text{w.r.t} \quad \{\{\mathbf{x} \setminus (\mathbf{u}, t)\} \cup \mathbf{z}\} \\ \mathcal{H}_1: & \text{significant difference in the local second-order structure of } (\mathbf{u}, t) \\ & \qquad \text{w.r.t} \quad \{\{\mathbf{x} \setminus (\mathbf{u}, t)\} \cup \mathbf{z}\} \end{cases}$$

Rejecting $\mathcal{H}_0$ for most of the points in the point pattern $\mathbf{x}$ means that the local second-order structures of $\mathbf{x}$ and $\mathbf{z}$ around those arbitrary points $(\mathbf{u}, t)$ are different, and therefore the second-order global structure of $\mathbf{x}$ and $\mathbf{z}$ is different, meaning that they are most likely coming from different underlying point processes.

Nevertheless, in addition to this global conclusion, the most interesting results concern the rejection of $\mathcal{H}_0$ for specific points, allowing for the identification of points for which the spatio-temporal local behaviour varies with the spatio-temporal location. This could also result in the identification of areas where the spatio-temporal local second-order structure of the points in the pattern of interest $\mathbf{x}$ differs the most and needs to be further analysed. If, for instance, the significant points tend to cluster in time and space, this could indicate multiple underlying generating processes, in contrast to a unique one.

Furthermore, the identification of a significant point may already be an indication of outlying behaviour for the remaining points.

To take into account the geometry of the underlying network, we develop a testing procedure based on the proposed homogeneous LISTA functions. A permutation test approach is considered, where the null distribution of a given statistical test is estimated by randomly permuting the labels attached to the two point patterns. The steps of the testing procedure are as follows

1. Define the point pattern of interest $\mathbf{x}$ and the background pattern $\mathbf{z}$, assumed to be realisations of the point processes X and Z, with $N(X) = n$ and $N(Z) = m$. Typically, $\mathbf{x}$ is the point pattern with the most clustered structure, between the two;

2. Set k as the number of permutations;

3. For each point $(\mathbf{u}_i, t_i) \in \mathbf{x}$

   (a) Estimate the LISTA function $\hat{K}^i_L(r, h)$, for fixed spatial and temporal ranges r and h;

(b) Define $\mathbf{q}^{-i} = \{\{\mathbf{x} \setminus (\mathbf{u}_i, t_i)\} \cup \mathbf{z}\}$ as the point pattern obtained as the superposition of all the points in $\mathbf{x}$ except the i-th one, and all the points in $\mathbf{z}$;

(c) Define lab_x and lab_z as the set of labels attached to the generated points in the new pattern $\mathbf{q}^{-i}$, identifying the points coming from the observed point patterns $\{\mathbf{x} \setminus (\mathbf{u}_i, t_i)\}$ and $\mathbf{z}$, of cardinality $n-1$ and m, respectively;

(d) Generate k point patterns $\mathbf{q}^{-i,j}$ by independently and randomly sampling from the same population, where $j = 1, \dots, k$. That is, randomly permuting the labels of the $n + m - 1$ points in $\mathbf{q}^{-i}$ and retaining only those $n - 1$ points with labels lab_x;

(e) Define k point patterns $\mathbf{q}^{i,j}$ under the null hypothesis, by re-adding the point $(\mathbf{u}_i, t_i)$ to $\mathbf{q}^{-i,j}$; Each of the resulting point patterns $\mathbf{q}^{i,j}$ will be of size n, as the original point pattern of interest $\mathbf{x}$;

(f) Compute the proposed local *deviation test* (Illian et al., 2008)

$$T^i = \int_0^{t_0} \int_0^{r_0} \left(\hat{K}_L^i(r,h) - \hat{K}_{L,H_0}^{-i}(r,h) \right)^2 \mathrm{d}r\mathrm{d}h$$

where $\hat{K}_{L,H_0}^{-i}(r,h)$ is the LISTA function for the i^{th} point, averaged over the $j = 1, \dots, k$ permutations, and where r_0 and t_0 are the maximum spatial and temporal ranges considered in the K-functions;

(g) Finally, a p-value is computed for each point of the pattern $\mathbf{x}$ as

$$p^i = \sum_{j=1}^{k} \frac{\mathbf{I}(T_{H_0}^{i,j} \geq T^i)}{k},$$

where

$$T_{H_0}^{i,j} = \int_0^{t_0} \int_0^{r_0} \left(\hat{K}_L^{i,j}(r,h) - \hat{K}_{L,H_0}^{-i,j}(r,h) \right)^2 \mathrm{d}r\mathrm{d}h,$$

with $\hat{K}_L^{i,j}(r,h)$ being the LISTA functions on the k patterns generated from the null hypothesis, and $\hat{K}_{L,H_0}^{-i,j}(r,h)$ being the LISTA function averaged over the points of the k^{th} permutation.

The testing procedure ends with providing a vector p^i of $N(X) = n$ p-values. The null hypothesis is not accepted if $p^i \leq \alpha$, where α is the fixed nominal value of the type I error.

5.1.2 Simulation studies

A simulation study is carried out to assess the performance of the testing procedure. We simulate from both the null and the alternative hypothesis to compute the empirical level and the power of the test. In particular, the linear network considered for the simulations comes from the dataset `chicago` of the package `spatstat.data` (Baddeley et al., 2021) of `R` (R Core Team, 2022). This dataset is a record of spatial locations of crimes that occurred in 2002 in Chicago (Illinois, USA) close to the University of Chicago. We note that alternative networks could have been used, but this represents a general network structure similar to our datasets. Finally, we set the number of permutations $k = 99$ for the simulation studies and the applications in the following section.

5.1.2.1 Empirical level

To assess the empirical level we simulate from the *null hypothesis* of no difference in the spatio-temporal second-order structure of two given point patterns $\mathbf{x}$ and $\mathbf{z}$ with respect to the ith point $(\mathbf{u}_i, t_i) \in X$. So the generating intensity is the same for each of the two point patterns. We run 100 simulations, and compute the rejection rates at a significant level $\alpha = 0.05$, for spatio-temporal homogeneous Poisson point processes with an expected number of overall events $\mathbb{E}[n+m]$ equal to $\{150, 300, 450\}$. An example of a spatio-temporal homogeneous Poisson point process is represented in Figure 5.1. Results of the simulations are in Table 5.1.

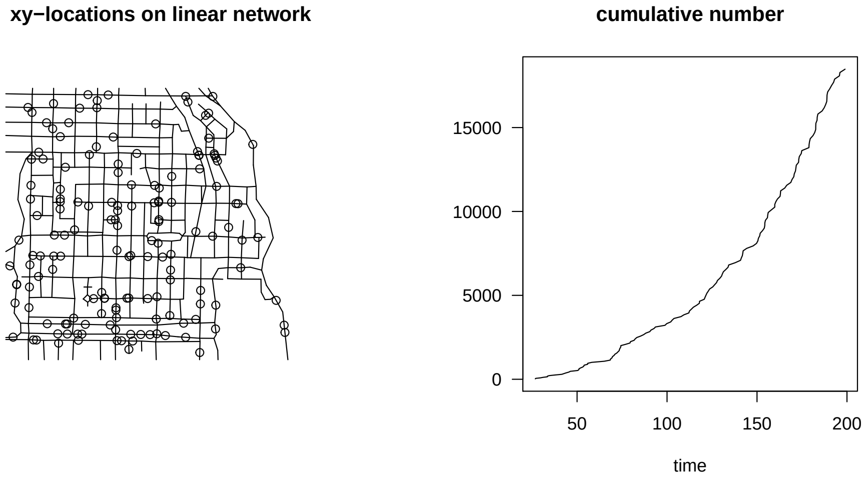


Figure 5.1: Spatio-temporal homogeneous Poisson process. Left panel: The projection of data onto the network. Right panel: The cumulative number of data points versus occurrence time.

The p-values under the null hypothesis should behave like a standard uniform random variable, with mean and variance approximately equal to 0.5 and 0.083, respectively. Increasing the number of points results in a slight improvement in the rejection rate, although there is no difference moving from 300 to 450 points.

Table 5.1: Rejection rates at $\alpha = .05$, with means and variances of the p-values for the test of local second-order structure. The point patterns of interest $\boldsymbol{x}$ and the background one $\boldsymbol{z}$ are simulated from spatio-temporal homogeneous Poisson processes with parameter $\lambda = \mathbb{E}[n+m]/2$.

X	Z	$\mathbb{E}[n+m]$	rate	mean	var
Homogeneous	Homogeneous	150	0.087	0.438	0.085
Homogeneous	Homogeneous	300	0.066	0.474	0.085
Homogeneous	Homogeneous	450	0.066	0.469	0.085

5.1.2.2 Power

For evaluating the power of the test, we simulate from the *alternative hypothesis* of the difference in the local second-order structure between $\mathbf{x}$ and $\mathbf{z}$. Therefore, the two generating intensities are different. In particular, the point pattern $\mathbf{x}$ is assumed to be more inhomogeneous than the pattern $\mathbf{z}$.

We emphasise that we consider inhomogeneous patterns for our power analysis to place ourselves in the worst case (as we are using homogeneous versions of the K-function) and evaluate the power under these conditions.

We consider different scenarios for the point pattern of interest $\mathbf{x}$, simulating from

1. an inhomogeneous Poisson process with intensity

$$\lambda(x, y, t) = \exp(8.25 - 4y - 2t), \tag{5.7}$$

   with $t \in [0, 1]$. Figure 5.2 shows an example of an inhomogeneous Poisson process with intensity function in (5.7).

2. a clustered Poisson process (Gabriel et al., 2013) in which parents form a Poisson process with intensity (5.7) and with time ranging from 0 to 200. An example of a clustered Poisson process with 15 parents and 10 children per parent is reported in Figure 5.3.

3. an ETAS model with CIF as in (2.35) and with homogeneous large-scale term with spatial intensity function $f(\cdot)$, and $t \in [0, 200]$. An ETAS process, simulated with parameters
   $\boldsymbol{\theta} = (\mu, \kappa_0, c, p, \alpha, d, q) = (0.079, 0.004, 0.013, 1.2, 0.5, 0.424, 1.165)$ and no external covariates comes in Figure 5.4.

4. an ETAS model with the same parameters and time vector, but with the large-scale term with intensity function $f(\cdot)$ generated by an inhomogeneous spatial intensity, given by
$$\lambda(x, y) = \exp(a - 4y), \tag{5.8}$$
where a is taken as the logarithm of the number of background points.

Note that for the first three scenarios, the background pattern $\mathbf{z}$ is always simulated from a homogeneous Poisson process with a constant intensity equal to $\lambda = \mathbb{E}[n + m]/2$, on the same network of $\mathbf{x}$.

The results are shown in Tables 5.2 and 5.3 for the point processes of interest coming from inhomogeneous and clustered processes, respectively. Tables 5.4 and 5.5 report the results of the patterns simulated from the ETAS model. In particular, Table 5.4 reports the results of the ETAS simulated model (with large-scale term constant in time and space), partitioning points by the stochastic declustering approach (Zhuang et al., 2002, 2004; Zhuang, 2006) and then assigning to $\mathbf{x}$ the triggered points and to $\mathbf{z}$ the ones of the large-scale term. In Table 5.5 results refer to simulation from the ETAS model but with $\mathbf{x}$ representing the whole point pattern (both large-scale and triggered component) and $\mathbf{z}$ simulated from a homogeneous Poisson process.

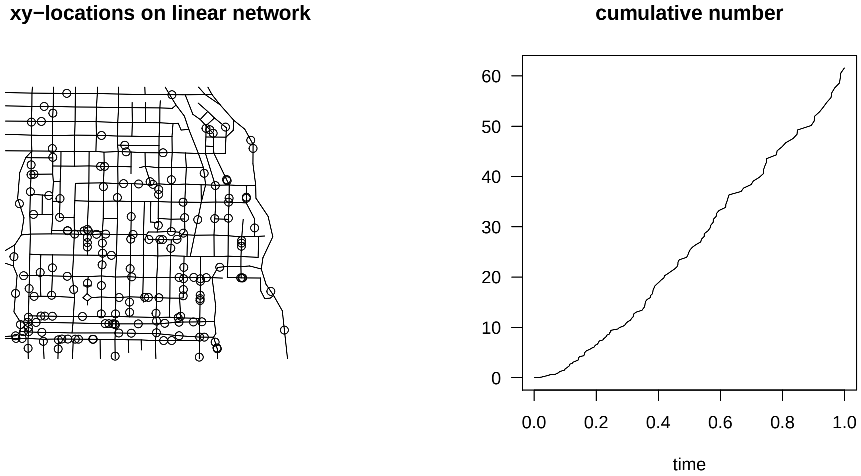


Figure 5.2: Spatio-temporal inhomogeneous Poisson process with intensity function as in (5.7) . Left panel: Projection of data onto the network. Right panel: Cumulative number of data points versus occurrence time.

The power of the test strongly depends on the similarity between the X and Z processes and the number of generated points, getting better results for increasing sizes. Based on these results, the test is able to correctly detect individual local

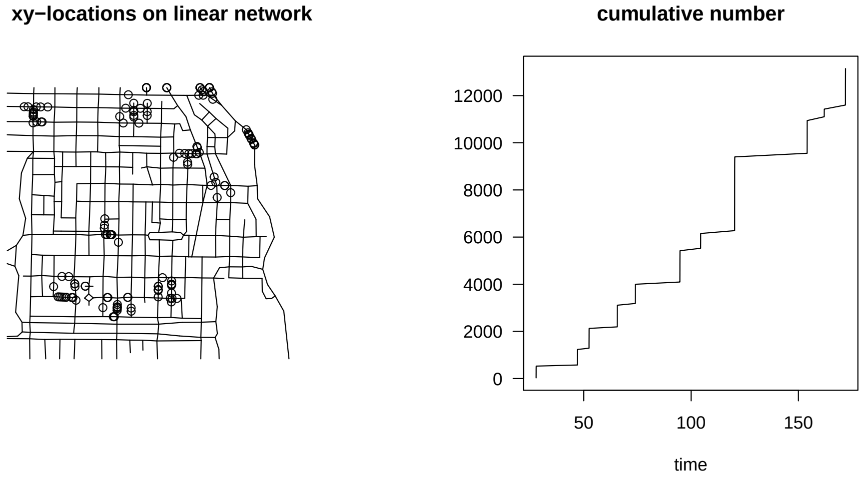


Figure 5.3: Spatio-temporal clustered Poisson process with 15 parents. Left panel: Projection of data onto the network. Right panel: Cumulative number of data points versus occurrence time.

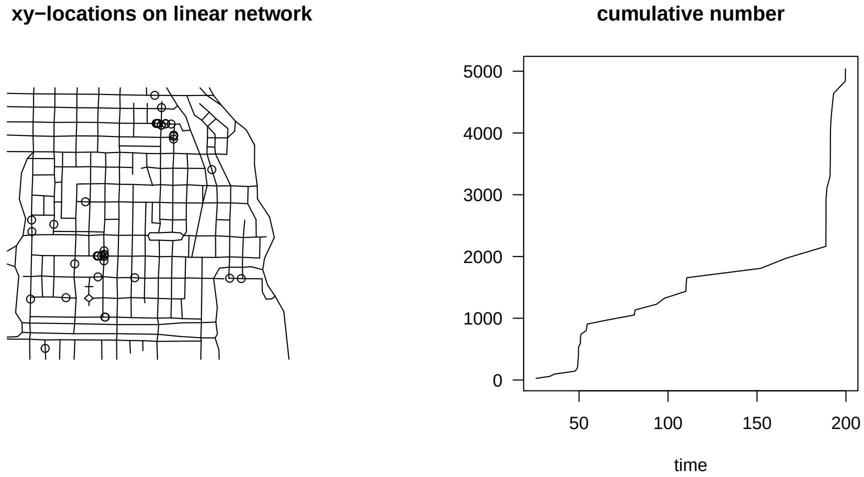


Figure 5.4: Spatio-temporal ETAS model with conditional intensity function as in (2.35) *and homogeneous large-scale term with spatial intensity function $f(\cdot)$. Left panel: Projection of data onto the network. Right panel: Cumulative number of data points versus occurrence time.*

differences when the two point patterns come from processes with different local

Table 5.2: Rejection rates at $\alpha = .05$ *for the test of local second-order structure. We simulate the patterns of interest from the inhomogeneous Poisson process with intensity as in* (5.7)*.*

X	Z	$\mathbb{E}[n+m]$	Power
Inhomogeneous	Homogeneous	150	0.644
Inhomogeneous	Homogeneous	300	0.769
Inhomogeneous	Homogeneous	450	0.855
Inhomogeneous	Homogeneous	600	0.948

Table 5.3: Rejection rates at $\alpha = .05$ *of the test of local second-order structure. We simulate the patterns of interest from a clustered Poisson process, with parent intensity coming from an inhomogeneous Poisson process, as in* (5.7)*.*

X	Z	$\mathbb{E}[n+m]$	parents	children per parent	Power
Clustered	Homogeneous	150	15	10	0.251
Clustered	Homogeneous	300	15	20	0.387
Clustered	Homogeneous	150	5	30	0.887
Clustered	Homogeneous	300	5	60	0.934

Table 5.4: Rejection rates at $\alpha = .05$ *of the test of local second-order structure. We simulate from the ETAS model in* (2.35) *and split the pattern into large-scale and triggered events through stochastic declustering.*

X	Z	$\mathbb{E}[n+m]$	μ	κ_0	Power
Triggered	Homogeneous spatial large-scale intensity	150	0.099	0.231	0.293
Triggered	Homogeneous spatial large-scale intensity	300	0.099	0.231	0.263

Table 5.5: Rejection rates at $\alpha = .05$ *of the test of local second-order structure. We simulate* X *from the ETAS model in* (2.35)*, whose large-scale term is simulated with a spatial intensity from a homogeneous Poisson process.*

X	Z	$\mathbb{E}[n+m]$	μ	κ_0	Power
ETAS with hom. spatial large-scale intensity	Hom.	150	0.118	0.185	0.469
ETAS with hom. spatial large-scale intensity	Hom.	300	0.236	0.185	0.494

second-order structures. This is the case in which the background pattern $\mathbf{z}$ is homogeneous, while the point pattern of interest $\mathbf{x}$ comes from an inhomogeneous point process or a clustered one. In this latter case, the lower the number of parents (the more clustered the structure of the pattern $\mathbf{x}$ is), the higher the power of the test. In both cases, the power increases as the expected number of points increases too.

It is interesting to look at the results when simulating from even more complex point processes, such as ETAS models. According to a branching structure, the conditional intensity function of the Self-Exciting model is defined as the sum of a term describing the large-time scale variation (spontaneous activity or background, generally assumed homogeneous in time but not in space) and one relative to the small-time scale variation due to the interaction with the events in the past (induced or triggered activity). In general, when simulating from the ETAS models, the power of the test is quite low. This is a reasonable result. Indeed, when assigning the points of the large-scale and of the triggered component to Z and X, of course, the local second-order structure is the same even if the global one might seem different, as the spatio-temporal location of the triggered events strongly depends on the events of the large-scale component. In addition, one main assumption of the ETAS model states that any event can be either an immigrant (parent) or an offspring, that is, all events (both large-scale and induced ones) may have offspring (Ogata and Katsura, 1988; Ogata, 1998). The dependence between the location of the large-scale and the triggered events can be elicited from Figure 5.5.

Moreover, when comparing a point pattern of interest $\mathbf{x}$ coming from an ETAS model and a background pattern $\mathbf{z}$ coming from a homogeneous pattern, the local second-order structure of the two point patterns does not differ significantly, as also the ETAS model consists of an independent homogeneous pattern represented by the events of the large-scale component.

In the last experiment, we indeed simulated the point pattern of interest $\mathbf{x}$ from an ETAS model with inhomogeneous spatial intensity as the large-scale term. The background pattern $\mathbf{z}$ is still a homogeneous spatio-temporal point process, generated independently of $\mathbf{x}$. The results come in Table 5.6.

Table 5.6: Rejection rates at $\alpha = .05$ of the test of local second-order structure. We simulate X from the ETAS model in (2.35), whose spatial large-scale term is simulated from the inhomogeneous Poisson process in (5.8).

X	Z	$\mathbb{E}[n+m]$	μ	κ_0	Power
ETAS with inh.spatial large-scale intensity	Inh.	150	0.026	0.028	0.710
ETAS with inh. spatial large-scale intensity	Inh.	300	0.053	0.028	0.952

As expected, the power increases as the model used for generating X is basically a superposition of a large-scale process, homogeneous in time but inhomogeneous in space and the triggered process.

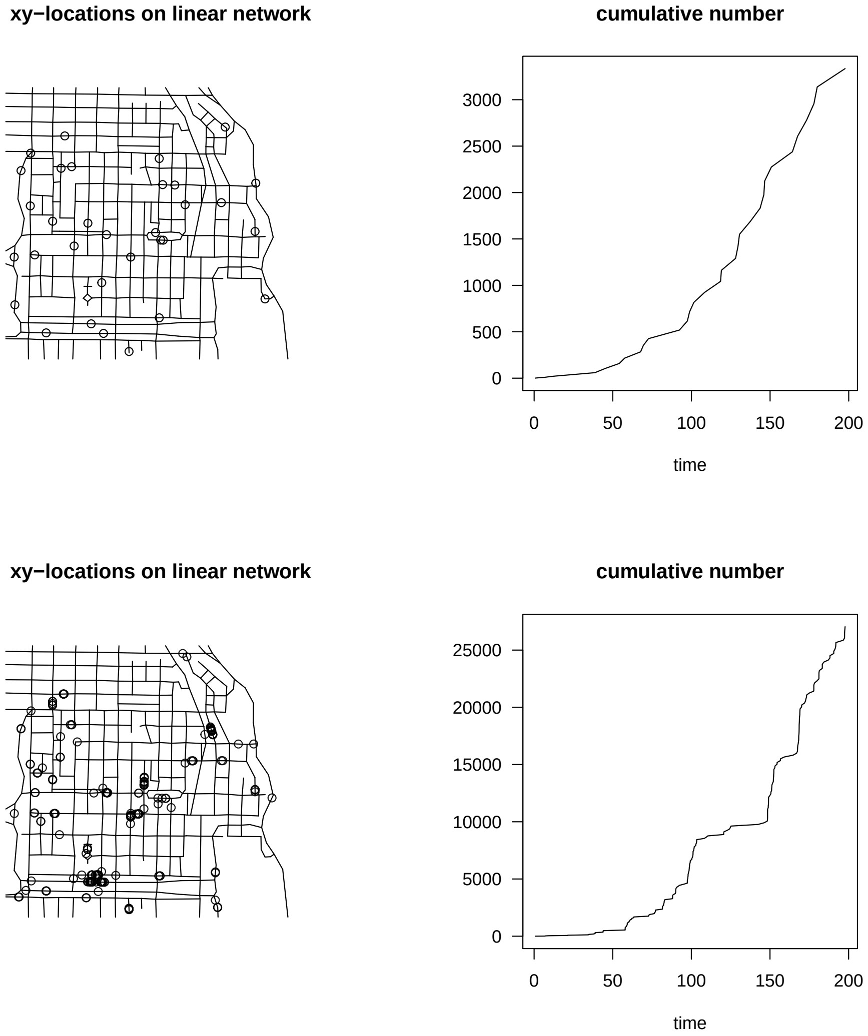


Figure 5.5: Spatio-temporal ETAS model with conditional intensity function as in (2.35). Upper panels: Events of the large-scale component. Bottom panels: Triggered events. Left panels: Projection of data onto the network. Right panels: Cumulative number of data points versus occurrence time.

5.1.3 Application to traffic data

We analyse two datasets on traffic accidents, one from Medellin (Colombia) and the other from Barcelona (Spain). For each dataset, we know the spatial and temporal locations of the traffic accidents occurring on the underlying linear network represented by the streets of the city under analysis. Both datasets actually consist of multitype point patterns, according to the type of traffic accident, i.e. the mark. Based on this knowledge, we split the datasets into two distinct point patterns, defining a point pattern of interest $\mathbf{x}$ and an alternative (background) pattern $\mathbf{z}$. We aim to investigate the difference in the spatio-temporal local second-order structure of the two resulting point patterns. Once we have applied the proposed local test, we can distinguish the significant points from the non-significant ones. In addition, by projecting these points onto the network, we can identify the points of the pattern of interest for which the spatio-temporal local second-order structure is more different from the background pattern, and identify those areas where these points occur more often.

5.1.3.1 Medellin traffic data

We analyse traffic accidents in the city of Medellin (Colombia), a dataset containing the locations of traffic accidents consisting of 1215 traffic accidents in 2019 in downtown Medellin. The entire data were published in the OpenData portal of Medellin Town Hall at
https://www.medellin.gov.co/geomedellin/index.hyg. The dataset actually consists of a multitype point pattern, categorised based on the damages caused by the accident (see a description in Table 5.7).

Table 5.7: Number of Medellin's traffic accidents by type.

Type	n
Damages only	675
Injuries to people	536
Deaths	4

In our analysis, we do not consider the four events representing car accidents that lead to deaths, and we define the accidents leading to damages as the point pattern of interest $\mathbf{x}$, and the accidents causing injuries to people as the alternative (background) pattern $\mathbf{z}$ (see Figure 5.6).

We thus test the hypothesis of the difference in the spatio-temporal local second-order structure between the two point patterns considered. The application of the test provides a p-value for each point in $\mathbf{x}$, and by comparing those to the confidence level $\alpha = 0.05$ we can identify *significant* events, i.e. those points for which $\mathcal{H}_0$ is not accepted and therefore the local second-order structure is significantly different from the background process, and *non-significant* events. The significant events are 100 out of the 675 events, and they are shown in Figure 5.7.

The results of the application of the local test support our hypothesis of the two point patterns behaving differently in terms of spatio-temporal local second-order structure. Furthermore, displaying the significant events and the non-significant

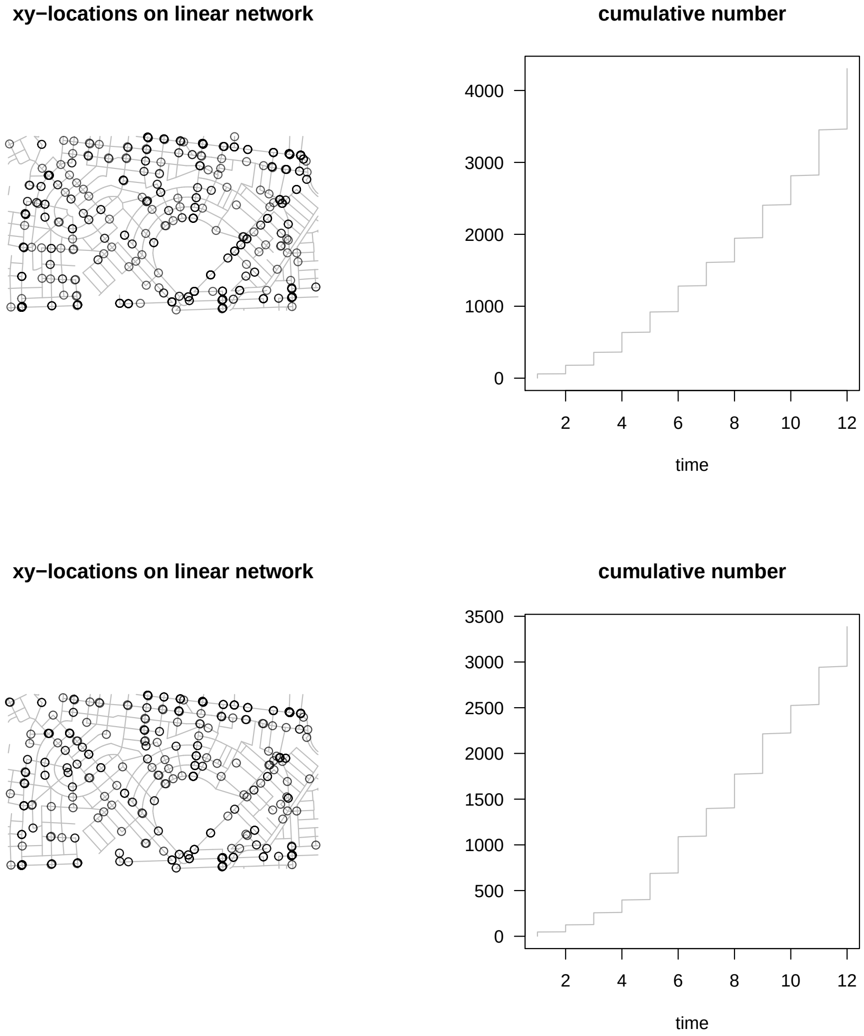


Figure 5.6: Top panel: Point pattern of damages. Bottom panel: Point patterns of injuries. Left panel: Projection of data onto the network. Right panel: Cumulative number of data points versus occurrence time.

events allows for identifying the areas containing those events whose spatio-temporal local second-order structure is different from the background pattern.

5.1.3.2 Barcelona traffic data

We now analyse traffic accidents in the city of Barcelona (Spain). This dataset represents a spatio-temporal point pattern consisting of 10536 traffic accidents recorded

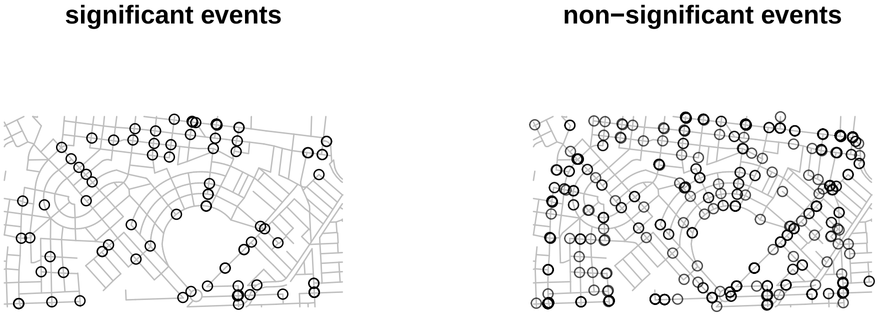


Figure 5.7: Left panel: Significant events. Right panel Non-significant events. These are obtained by applying the local test to the point pattern of interest $\boldsymbol{x}$, represented by damages, and the alternative patterns $\boldsymbol{z}$, represented respectively by injuries.

in 2019 (Moradi et al., 2021).

The dataset also reports detailed information about the traffic accidents recorded. Available covariates regard, of course, the spatial location, and the temporal location, both referring to the actual date (i.e. day of the week, month and year) and the hour during the day. For the purpose of our analysis, we constrain the network to some streets in downtown Barcelona, and we consider only the temporal coordinate as the hour of the day and only four types of accidents. This subsetting allows considering only types of events relevant for testing our proposed methodology, as they represent the point patterns with the most interesting spatio-temporal behaviour. Furthermore, restricting the spatial window prevents considering areas in which the displacement of points is too sparse and also eases the computational aspect of the running of the testing procedure. The number of events for each type of event considered comes in Table 5.8.

Table 5.8: Number of Barcelona's traffic accidents by type.

Type	n
Vehicles hitting pedestrians	223
Head-on and side-impact collisions	331
Collisions with non-moving objects	81
Head-on collisions	593

In our analysis, we define the “Vehicles hitting pedestrians” accidents as the point pattern of interest **x** (Figure 5.8), and the remaining three point patterns as the alternative (background) pattern **z** (Figure 5.9), namely: (1) both “Head-on and side-impact collisions” accidents, (2) “Collisions with non-moving objects” accidents, and (3) “Head-on collisions” accidents.

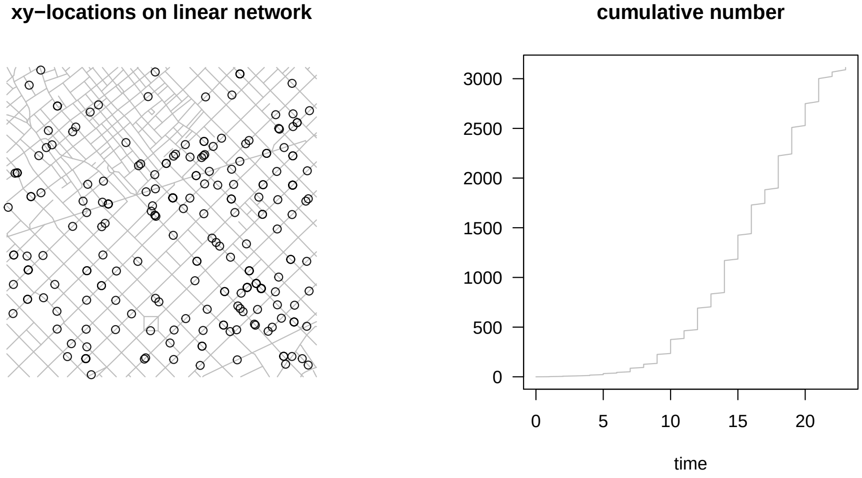


Figure 5.8: Point pattern of vehicles hitting pedestrians for Barcelona data. Left panel: Projection of data onto the network. Right panel: Cumulative number of data points versus occurrence time.

Again, the results of the application of the local test support our hypothesis of the different point patterns considered behaving differently in terms of spatio-temporal local second-order structure. In Table 5.9, the number of significant and non-significant events found after having applied the local test is reported for each considered combination.

Table 5.9: Number of significant and non-significant events found after having applied the local test.

x	**z**	sign.	non-sign.
Vehicles hitting pedestrians	Head-on and side-impact collision	35	188
Vehicles hitting pedestrians	Collisions with non-moving objects	34	189
Vehicles hitting pedestrians	Head-on collisions	40	183

Figure 5.10 depicts the significant events and the non-significant events, allowing us to draw interesting conclusions. It appears evident that there are some areas in

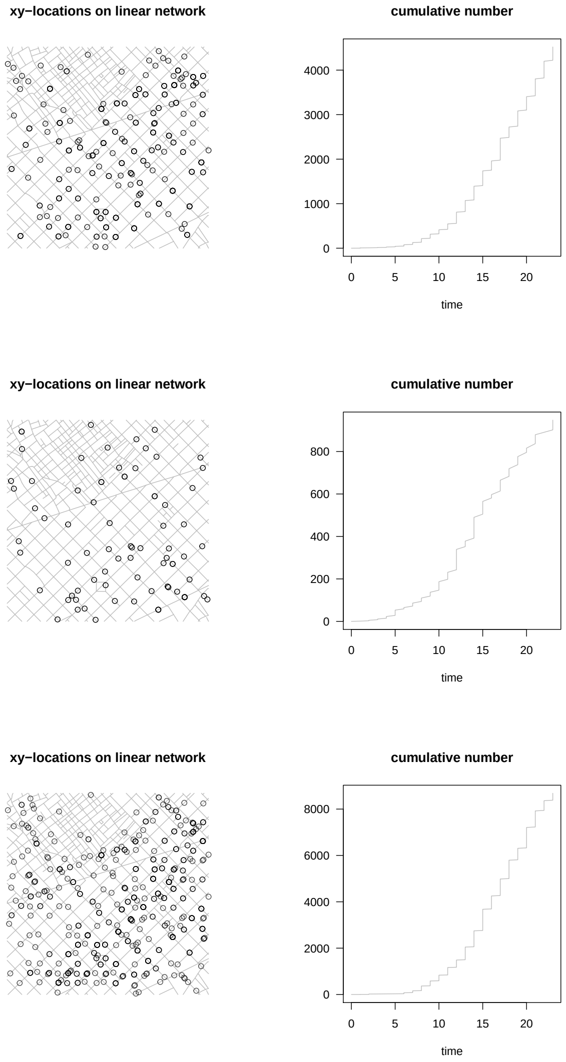


Figure 5.9: Point patterns of (a) head-on and side-impact collision, (b) collisions with non-moving objects, and (c) head-on collisions. Left panel: Projection of data onto the network. Right panel: Cumulative number of data points versus occurrence time.

which the significant events are clustered. This is even more evident in front-lateral collisions and lateral accidents. In addition, the spatial patterns of their significant events are quite similar, indicating that their difference in the local second-order structure with respect to Vehicles hitting pedestrians, does not differ that much. Finally, the events representing the accidents towards non-moving objects differ with respect to the Vehicles hitting pedestrians more randomly in space.

5.1.4 Conclusions

In this Section, we have proposed a local test to assess local differences between the spatio-temporal second-order structure of two point patterns occurring on the same linear network, based on the Local Indicators of Spatio-Temporal Association on linear networks.

We have shown that the local test is able to correctly identify the spatio-temporal difference in the local second-order structure of two point patterns occurring on the same linear network by simulating from different scenarios, with the point pattern of interest being generated from inhomogeneous, clustered, and Self-Exciting point processes. As expected, we have found that the test's power increases as the number of points and the clustered structure of the point pattern of interest increase too.

We have applied these features to two case studies, analysing traffic data represented by a marked point process occurring on the streets of Medellin and Barcelona, where the marks represent the type of car accident. In both cases, the local test has been used to test the hypothesis of difference among different point patterns, assumed to differ in their spatio-temporal local behaviour. Thanks to the proposed local test, we have been able to explore how individual points are related to their neighbouring events and classify points with similar spatio-temporal local structures, identifying the areas that differ in terms of clustering behaviour. Furthermore, displaying the significant events has allowed identifying the areas containing those events whose spatio-temporal local second-order structure is different from the background pattern.

The analysed point patterns are basically identified by characteristics of the events, i.e. the marks, recorded together with their spatial and temporal location. Future analyses could regard the application of the local test to patterns whose splitting is due to a statistical procedure, such as the stochastic declustering of the Self-Exciting point processes. Indeed, the topic of fitting parametric point process models while also accounting for the underlying network structure is still quite unexplored, and only a recent attempt has been carried out to fit Self-Exciting models to processes occurring on a network (Kalair et al., 2021). We believe that the latter represents an interesting research path for the future.

A final comment is in order. We have argued that we only use homogeneous versions of the K-function as we are interested in second-order structure. Additional simulation studies, not reported here for brevity, have shown that a local test based on the inhomogeneous K-function adds bias to the results, given that the contribution of the difference in the second-order characteristics is masked by the variability in the fitting of the first-order intensities. Indeed, when considering inhomogeneous Poisson patterns for X and homogeneous ones for Z with an expected number of points for both patterns of 450, the homogeneous test provided a power of 0.83, but

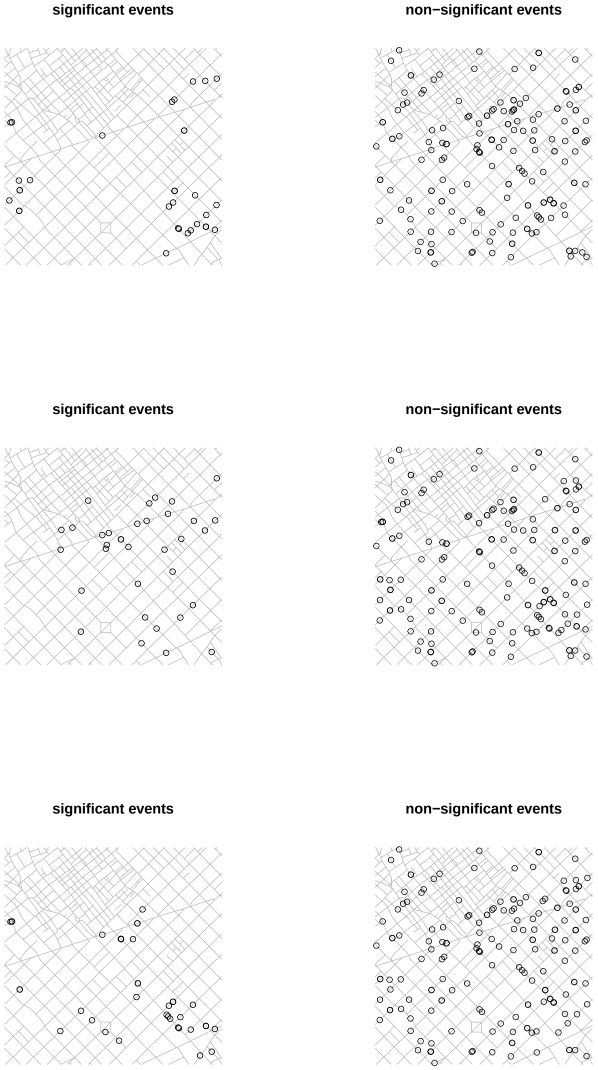


Figure 5.10: Left panel: Significant events. Right panel: Non-significant events. These are obtained by applying the local test to the point pattern of interest $\boldsymbol{x}$, represented by Vehicles hitting pedestrians, and the alternative patterns $\boldsymbol{z}$, represented respectively by: (a) Head-on and side-impact collision, (b) Collisions with non-moving objects, and (c) Head-on collisions.

when we correct by inhomogeneous K-functions, this power reduces down to 0.74. However, more exploration and deep analysis should be considered in this regard in order to better understand the effect of the first-order intensity on our statistical test.

5.2 Local space-time diagnostics on networks through LISTA functions

D'Angelo, N., Adelfio, G., and Mateu, J. (2022b). Local inhomogeneous second-order characteristics for spatio-temporal point processes occurring on linear networks. *Statistical Papers, http://doi.org/10.1007/s00362-022-01338-4*

At the beginning of this Chapter, we have extended local indicators of spatio-temporal association (Siino et al., 2018b), known as LISTA functions, for spatio-temporal point processes occurring on linear networks.

As the proposed local second-order statistics can be used for obtaining further insight into the local structure of the analysed point pattern, and on the characteristics of individual points, in section 5.1.1, we have used the LISTA homogeneous versions to build a local test allowing to assess local differences between the spatio-temporal second-order structure of two point patterns occurring on the same linear network.

The present section aims at developing inhomogeneous versions of two local second-order statistics for spatio-temporal point processes occurring on networks, namely the K-function and the pair correlation function (Moradi and Mateu, 2020). In particular, following Adelfio et al. (2020) for the Euclidean case, we use LISTA functions based on local inhomogeneous second-order statistics on networks to assess the goodness-of-fit of general spatio-temporal models. Indeed, the peculiar lack of homogeneity in a network structure discourages the use of traditional spatial and spatio-temporal methods based on stationary processes (Baddeley et al., 2020). Therefore, weighted second-order statistics are appropriate diagnostic tools since they directly apply to data without assuming homogeneity (Adelfio et al., 2020).

We provide several simulation studies by weighting the contribution of each observed point by the inverse of the intensity function coming from several types of inhomogeneous point processes on the network. In particular, we consider Poisson, Self-Exciting and log-Gaussian spatio-temporal processes, and detail the corresponding simulation procedures on a linear network.

The section is structured as follows. Section 5.2.1 develops some properties of the local weighted second-order statistics defined in this chapter. Then, through some simulation studies, in section 5.2.2 we show how the proposed LISTA functions can be used as diagnostic tools for different fitted models. Further, an application to a case study is presented in section 5.2.3. Conclusions are drawn in section 5.2.4.

5.2.1 Some properties of local weighted second-order statistics

In the following, we provide some properties of the proposed estimators.

Theorem 5.1. *For a Poisson process, we have* $\mathbb{E}[\hat{K}_L^i(r,h)] = rh$*.*

Proof.

$$\mathbb{E}[\hat{K}_L^i(r,h)] = \frac{1}{\lambda^2|L||T|}\mathbb{E}\left[\sum_{(\mathbf{u}_i,t_i)\neq(\mathbf{v},s)} \frac{I\{d_L(\mathbf{u}_i,\mathbf{v}) < r, |t_i - s| < h\}}{M((\mathbf{u}_i,t_i), d_L(\mathbf{u}_i,\mathbf{v}), |t_i - s|)}\right]$$

(By theorem (2.38) and since $\lambda^{(2)} \equiv \lambda^2$ for a Poisson process)

$$\begin{aligned}
&= \frac{1}{\lambda^2|L||T|}\int_{L\times T}\int_{L\times T} \frac{I\{d_L(\mathbf{u}_i,\mathbf{v}) < r, |t_i - s| < h\}}{M((\mathbf{u}_i,t_i), d_L(\mathbf{u}_i,\mathbf{v}), |t_i - s|)}\lambda^2 \mathrm{d}_2(\mathbf{v},s)\\
&= \frac{1}{\lambda^2|L||T|}\lambda^2\int_{L\times T}\int_{L\times T} \frac{I\{d_L(\mathbf{u}_i,\mathbf{v}) < r, |t_i - s| < h\}}{M((\mathbf{u}_i,t_i), d_L(\mathbf{u}_i,\mathbf{v}), |t_i - s|)} \mathrm{d}_2(\mathbf{v},s)\\
&= \frac{1}{|L||T|}\int_{L\times T}\int_{L\times T} \frac{I\{d_L(\mathbf{u}_i,\mathbf{v}) < r, |t_i - s| < h\}}{M((\mathbf{u}_i,t_i), d_L(\mathbf{u}_i,\mathbf{v}), |t_i - s|)} \mathrm{d}_2(\mathbf{v},s)
\end{aligned}$$

(By the change of variables in (2.39))

$$\begin{aligned}
&= \frac{1}{|L||T|}\int_0^\infty\int_0^\infty \sum_{\substack{(\mathbf{v},s)\in L\times T:\\ d_L(\mathbf{u}_i,\mathbf{v})=r,\\ |t_i-s|=h}} \frac{I\{d_L(\mathbf{u}_i,\mathbf{v}) < r, |t_i - s| < h\}}{M((\mathbf{u}_i,t_i), d_L(\mathbf{u}_i,\mathbf{v}), |t_i - s|)}\mathrm{d}r\mathrm{d}h\\
&= \frac{1}{|L||T|}\int_0^\infty\int_0^\infty \sum_{\substack{(\mathbf{v},s)\in L\times T:\\ d_L(\mathbf{u}_i,\mathbf{v})=r,\\ |t_i-s|=h}} \frac{I\{(||\mathbf{u}_i - \mathbf{v}||) < r, |t_i - s| < h\}}{M(\mathbf{u}_i,t_i)}\mathrm{d}r\mathrm{d}h\\
&= \frac{1}{|L||T|}\int_0^\infty\int_0^\infty \frac{I\{(||\mathbf{u}_i - \mathbf{v}||) < r, |t_i - s| < h\}M(\mathbf{u}_i,t_i)}{M(\mathbf{u}_i,t_i)}\mathrm{d}r\mathrm{d}h\\
&= \frac{1}{|L||T|}\int_0^\infty\int_0^\infty I\{(||\mathbf{u}_i - \mathbf{v}||) < r, |t_i - s| < h\}\mathrm{d}r\mathrm{d}h
\end{aligned}$$

(taking the conditional expectation with respect to its filtration)

$$\begin{aligned}
&= \frac{1}{|L||T|}|L||T|rh\\
&= rh
\end{aligned}$$

□

Theorem 5.2. *If the process is weighted by the true intensity function, the expectation of* $\hat{K}^i_{L,I}(r,h)$ *is the same as the expectation of* $\hat{K}^i_L(r,h)$.

Proof.

$$\mathbb{E}[\hat{K}^i_{L,I}(r,h)] =$$

$$= \frac{1}{|L||T|}\mathbb{E}\left[\sum_{(\mathbf{u}_i,t_i)\neq(\mathbf{v},s)} \frac{I\{d_L(\mathbf{u}_i,\mathbf{v})<r,|t_i-s|<h\}}{\lambda(\mathbf{u}_i,t_i)\lambda(\mathbf{v},s)M((\mathbf{u}_i,t_i),d_L(\mathbf{u}_i,\mathbf{v}),|t_i-s|)}\right]$$

(Second-order Campbell's theorem in (2.38))

$$= \frac{1}{|L||T|}\int_{L\times T}\int_{L\times T} \frac{I\{d_L(\mathbf{u}_i,\mathbf{v})<r,|t_i-s|<h\}}{\lambda(\mathbf{u}_i,t_i)\lambda(\mathbf{v},s)M((\mathbf{u}_i,t_i),d_L(\mathbf{u}_i,\mathbf{v}),|t_i-s|)}\lambda(\mathbf{v},s)\mathrm{d}_2(\mathbf{v},s)$$

$$= \frac{1}{|L||T|}\frac{1}{\lambda(\mathbf{u}_i,t_i)}\int_{L\times T}\int_{L\times T} \frac{I\{d_L(\mathbf{u}_i,\mathbf{v})<r,|t_i-s|<h\}}{M((\mathbf{u}_i,t_i),d_L(\mathbf{u}_i,\mathbf{v}),|t_i-s|)}\frac{1}{\lambda(\mathbf{v},s)}\lambda(\mathbf{v},s)\mathrm{d}_2(\mathbf{v},s)$$

$$= \frac{1}{|L||T|}\frac{1}{\lambda(\mathbf{u}_i,t_i)}\int_{L\times T}\int_{L\times T} \frac{I\{d_L(\mathbf{u}_i,\mathbf{v})<r,|t_i-s|<h\}}{M((\mathbf{u}_i,t_i),d_L(\mathbf{u}_i,\mathbf{v}),|t_i-s|)}\mathrm{d}_2(\mathbf{v},s)$$

(Change of variables in (2.39))

$$= \frac{1}{|L||T|}\frac{1}{\lambda(\mathbf{u}_i,t_i)}\int_0^\infty\int_0^\infty \sum_{\substack{(\mathbf{v},s)\in L\times T:\\ d_L(\mathbf{u}_i,\mathbf{v})=r,\\ |t_i-s|=h}} \frac{I\{d_L(\mathbf{u}_i,\mathbf{v})<r,|t_i-s|<h\}}{M((\mathbf{u}_i,t_i),d_L(\mathbf{u}_i,\mathbf{v}),|t_i-s|)}\mathrm{d}r\mathrm{d}h$$

$$= \frac{1}{|L||T|}\frac{1}{\lambda(\mathbf{u}_i,t_i)}\int_0^\infty\int_0^\infty \sum_{\substack{(\mathbf{v},s)\in L\times T:\\ d_L(\mathbf{u}_i,\mathbf{v})=r,\\ |t_i-s|=h}} \frac{I\{(||\mathbf{u}_i-\mathbf{v}||)<r,|t_i-s|<h\}}{M(\mathbf{u}_i,t_i)}\mathrm{d}r\mathrm{d}h$$

$$= \frac{1}{|L||T|}\frac{1}{\lambda(\mathbf{u}_i,t_i)}\int_0^\infty\int_0^\infty \frac{I\{(||\mathbf{u}_i-\mathbf{v}||)<r,|t_i-s|<h\}M(\mathbf{u}_i,t_i)}{M(\mathbf{u}_i,t_i)}\mathrm{d}r\mathrm{d}h$$

$$= \frac{1}{|L||T|}\frac{1}{\lambda(\mathbf{u}_i,t_i)}\int_0^\infty\int_0^\infty I\{(||\mathbf{u}_i-\mathbf{v}||)<r,|t_i-s|<h\}\mathrm{d}r\mathrm{d}h$$

(taking the conditional expectation with respect to its filtration)

$$= \frac{1}{|L||T|}\frac{1}{\lambda(\mathbf{u}_i,t_i)}\lambda(\mathbf{u}_i,t_i)|L||T|rh$$

$$= rh$$

□

We know that the expectation of $\hat{K}_{L,I}(r,h)$, when the intensity used for the weighting is the true generator model, is the same as the expectation of $\hat{K}_L(r,h)$ for the Poisson process, that is $\mathbb{E}[\hat{K}_L(r,h)] = rh$.

We have proved that the same result holds for the local version $\hat{K}^i_{L,I}(r,h)$, meaning that our proposed local estimator $\hat{K}^i_{L,I}(r,h)$ for a general point process, weighted by the true intensity function, has the same expectation of the local estimator $\hat{K}^i_{L}(r,h)$ under the Poisson case.

Based on this result, we can use our proposed local estimator $\hat{K}^i_{L,I}(r,h)$ as a diagnostic tool for general spatio-temporal point processes occurring on a linear network. Since the local inhomogeneous estimator behaves as the corresponding homogeneous one of a Poisson process, we can follow the approach for diagnostics in a local scale in Adelfio et al. (2020) and use our proposed LISTA functions based on the K-function for assessing the goodness-of-fit of spatio-temporal point processes on linear networks with any generic first-order intensity function $\lambda(\cdot,\cdot)$.

Basically, departures of the LISTA functions $\hat{K}^i_{L,I}(r,h)$ from the Poisson expected value rh directly suggest the unsuitability of the intensity function $\lambda(\cdot)$ used in the weighting of the LISTA functions. This means that, if the estimated intensity function used for weighting in our proposed LISTA functions (5.4) is the true one, then the LISTA functions should behave as the corresponding ones of a homogeneous Poisson process (5.1), that corresponds to the reference model.

5.2.2 Simulation studies: local space-time diagnostics on networks

This section is devoted to the use of the proposed LISTA functions to assess the goodness-of-fit of different spatio-temporal models fitted to point patterns occurring on linear networks.

We show some simulation studies by generating different spatio-temporal point patterns and then performing diagnostics on different fitted intensities. By comparing the values of the LISTA functions and their theoretical values, we evaluate whether the LISTA functions can correctly identify the true intensity when this is constrained on a network.

Since in simulations the weights are obtained by considering the real intensity function, the inhomogeneous statistics are expected to behave as the ones of a homogeneous Poisson process. If departures from such behaviour were observed, that would be an indication that the data comes from a model identified by a first-order intensity function different from the one used in the weighting procedure.

5.2.2.1 Simulation set up

We simulate 100 space-time point patterns from 3 different inhomogeneous processes with 150 points on average: (a) a spatio-temporal inhomogeneous Poisson process on a linear network; (b) a spatio-temporal ETAS model; (c) a log-Gaussian Cox process (see Appendix A for the details on the proposed simulation algorithms).

For assessing the usefulness of the weighted LISTA functions also in network domains, we carry out a simulation study considering three different scenarios, in which the K-function is weighted by:

- The real intensity function, known in simulations
- The estimated intensity by a nonparametric kernel density function
- The intensity function estimated by a homogeneous Poisson intensity (wrong model).

We expect that the expected value of the spatio-temporal LISTA K-function, when weighting by the true or estimated intensity, resembles that of a Poisson process, and therefore when a wrong model is used for the intensity, the expected value would be far from the Poisson one.

5.2.2.2 Results

For each of the three previously defined scenarios, we compute the chi-squared statistics between the LISTA functions and their theoretical values $\hat{K}^i_{L,Theo}(r,h) = rh$. In particular, we use the LISTA functions based on the inhomogeneous K-function, weighted each time with a different fitted intensity. The χ^2_i values are obtained following the expression

$$\chi^2_i = \int_L \int_T \left(\frac{\left(\hat{K}^i_{L,I}(r,h) - rh\right)^2}{rh} \right) \mathrm{d}h\mathrm{d}r, \tag{5.9}$$

one for each point in the point pattern. We weight the LISTA functions $\hat{K}^i_{L,I}(r,h)$ with different intensities: the true one, a kernel density function, and a constant one. The latter of course is not a good fit, as we have simulated from inhomogeneous processes. In Figure 5.11, as expected, the lowest values are found in correspondence with the true intensity and the non-parametric kernel-based estimation, while the largest values are encountered for the constant intensity. The same holds for the ETAS process, and for the log-Gaussian Cox process, in Figure 5.12 and Figure 5.13, respectively.

In summary, LISTA functions are able to correctly identify the true intensity and spot the best fit among candidate models.

We now report an example of an application on an ETAS process, simulated through the same procedure outlined in section 2.2.3.2.2, to show further advantages of using LISTA functions for carrying out diagnostics.

5.2.2.3 Motivating example: a simulated ETAS process

To show some advantages of using LISTA functions, instead of their global counterparts, we simulate an ETAS process on the network (under the same conditions as in Figure 5.12).

We then compute the LISTA functions weighting them with three different intensities: the true one, a fitted one, and a constant one. For assessing their goodness-of-fit, the standard approach (Baddeley et al., 2015) would require the computation of

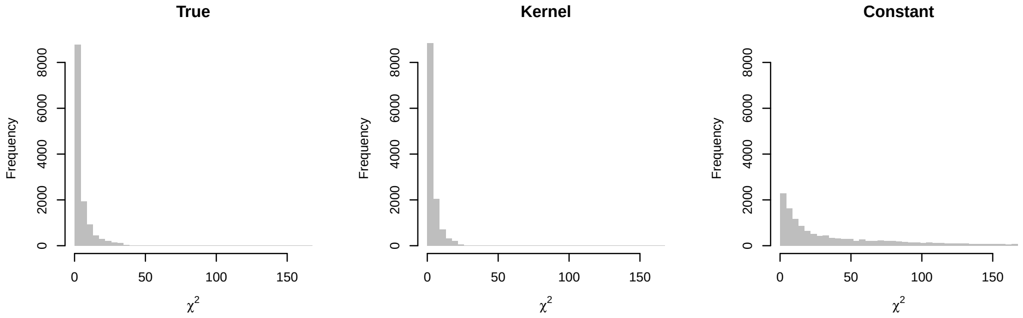


Figure 5.11: Distributions of the χ_i^2 statistics computed for assessing the differences between the local LISTA functions on the network and the theoretical expected value, calculated for each individual point from a simulation of an inhomogeneous Poisson process with intensity as in (3.2).

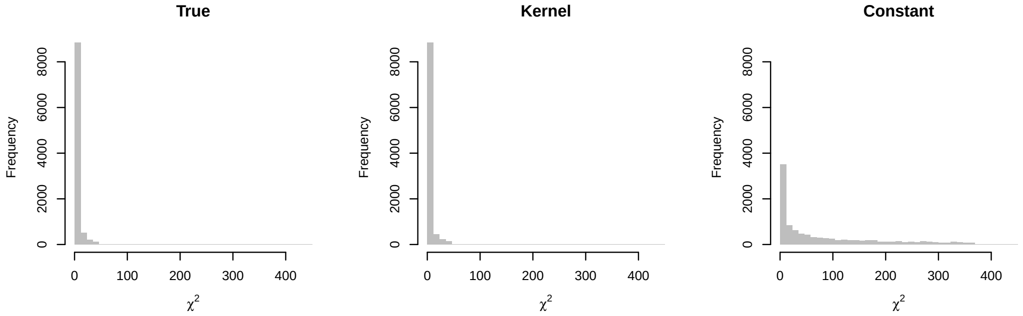


Figure 5.12: Distributions of the χ_i^2 statistics computed for assessing the differences between the local LISTA functions on the network and the theoretical expected value, calculated for each individual point from a simulation of an ETAS process with intensity as in (2.35).

the global inhomogeneous second-order summary statistics in section 2.3.3, weighted with their respective intensities. In Figure 5.14 we represent the difference between the global K-functions and their theoretical values, weighted by the three considered intensities. The surfaces whose values are closest to zero are the ones whose intensity best resembles the true one. According to these results, the global second-order characteristic is able to correctly identify the true intensity. This result in corroborated by the values of the χ^2, equal to 7.05, 8.32, and 19.26, for the true, the ETAS, and the constant intensities, respectively. However, the main reason for using LISTA functions is that the global summary statistics do not provide information on the individual points, and thus cannot identify which points influence the most

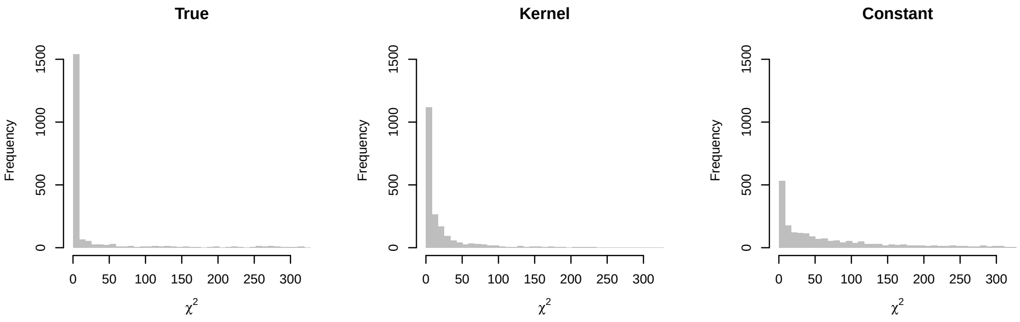


Figure 5.13: Distributions of the χ_i^2 statistics computed for assessing the differences between the local LISTA functions on the network and the theoretical expected value, calculated for each individual point from a simulation of a log-Gaussian Cox process following the parameters in section 2.2.3.1.

the goodness-of-fit of the model.

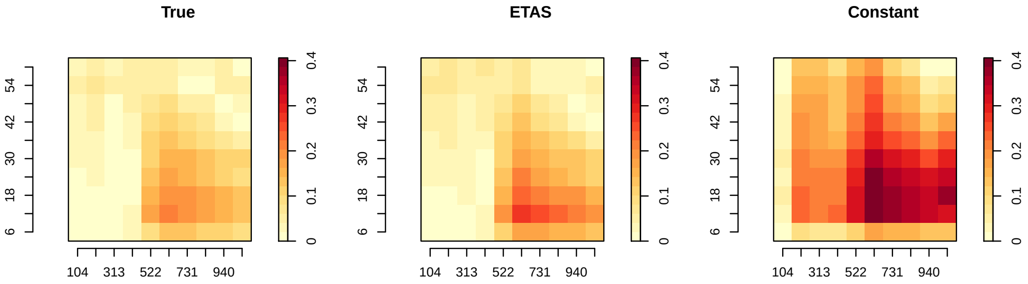


Figure 5.14: Difference between the global K-functions and their theoretical values, weighted by: (a) the true intensity; (b) the intensity estimated with the `etasFLP` *(Adelfio and Chiodi, 2015b) package of R; (c) the constant intensity given by $\hat{\lambda} = n/|L||T|$.*

Therefore we compute the LISTA functions, and again compare them to their theoretical values, by computing the χ_i^2 values (see Figure 5.15). Clearly, LISTA functions help to identify the true intensity as well as the best fit, in a similar manner to their global counterparts. Nonetheless, LISTA functions present further advantages. The main usefulness of carrying out diagnostics with LISTA functions is to get an insight into the local behaviour of the points of the analysed pattern, and therefore to identify the influential points. We indeed identify influential points by retaining only those points that exhibit χ_i^2 values larger than a fixed threshold. For instance, fixing the 75^{th} percentile of the distributions of the χ_i^2 values, and considering as influential points those with a χ_i^2 larger than this percentile, we obtain the black points in Figure 5.16.

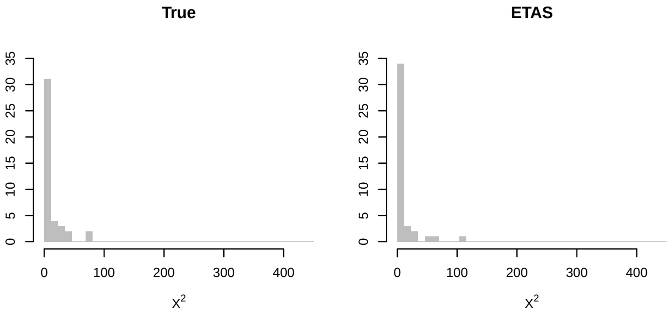


Figure 5.15: Distributions of the χ_i^2 statistics computed for assessing the differences between the local LISTA functions on the network and the theoretical expected value, calculated for each individual point of the simulated ETAS process.

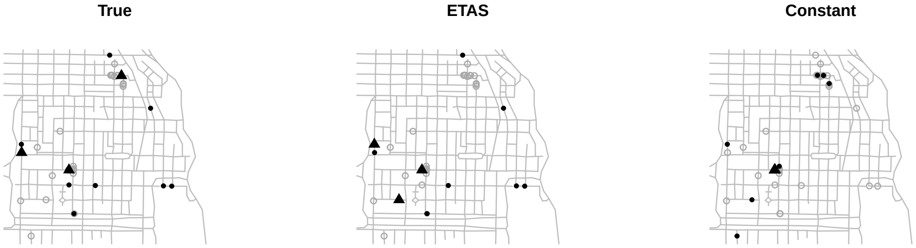


Figure 5.16: In black: Influential points for the three intensities, which exhibit the χ_i^2 values above the 75^{th} percentile. The triangles indicate the points with χ_i^2 values above the 95^{th} percentile.

We note that the influential points vary with the fitted intensity. In detail, we notice that the influential points obtained for the constant intensity are mainly placed in the clusters, and therefore indicate that we might need a better model in those areas, such as an inhomogeneous one. Furthermore, the resemblance between the true intensity and the kernel one (in terms of actual points, and not in the χ_i^2 values only), indicates that the kernel achieves a better fit if compared to the constant one. In Figure 5.17, we report the distributions of the χ_i^2 values of the influential points detected and shown in Figure 5.16. It is clear that the influential points of the wrong (constant) intensity are the ones contributing most to the bad fit of the model if compared to those of the true (or the most suitable) intensity.

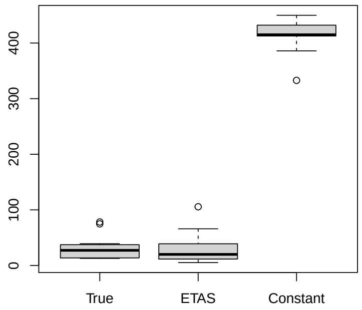


Figure 5.17: Distributions of the χ_i^2 *values above the* 75^{th} *percentile of the detected influential points detected, and shown in Figure 5.16 in black.*

Figure 5.18 shows an additional advantage of using LISTA functions. Indeed, their surfaces can be displayed and the different clustered structures of each individual point can be elicited. Of course, depending on the application, this can provide useful information. It can be interesting to plot the LISTA functions of the influential points, providing information on the clustered structure of individual points. Also, the degree of the identified clustered structure can be elicited and contextualised in the application. In particular, Figure 5.18 depicts the LISTA functions of the influential points, above the 95^{th} percentile, detected and shown in Figure 5.16 as triangles. Under the Poisson case, we expect the K-function to take rh values. In graphical terms, we expect the LISTA functions in Figure 5.18 to increase with those two distances, namely, to have constantly increasing values, from bottom-left to top-right. Points with different behaviour, such as the 41^{st}, indicate that the local structure around that point is somewhat deviating from the Poissonian case. In particular, for that point, we expect neighbouring points to start grouping together abruptly after a specific spatial distance. Overall, the temporal distance does not seem to influence the local structure.

5.2.3 Application to traffic data

We analyse traffic accidents in the city of Medellin (Colombia), a dataset containing the locations of traffic accidents consisting of 1215 traffic accidents in 2019 in downtown Medellin (see Figure 5.19).

We evaluate and compare an inhomogeneous intensity fitted with a kernel density with a constant one. Figure 5.20 depicts the distribution of the χ_i^2 statistics computed for assessing the differences between the local LISTA functions on the network and the theoretical expected value, weighted by an inhomogeneous inten-

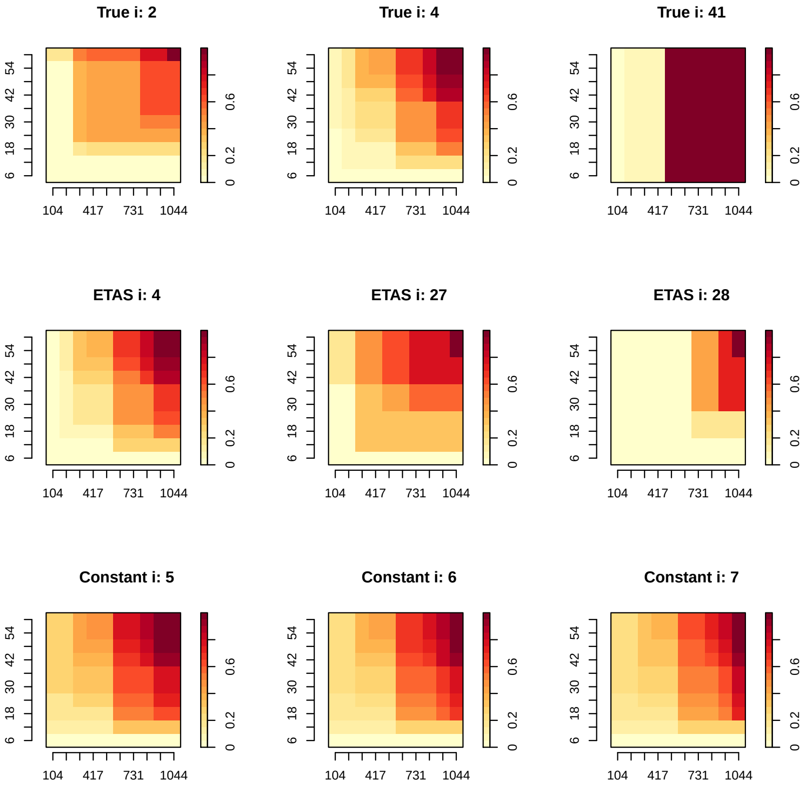


Figure 5.18: LISTA functions of the influential points, above the 95^{th} percentile, detected and shown in Figure 5.16 as triangles. From top to bottom: influential points of the true intensity, the ETAS one, and the constant one.

sity, and by a constant one. Clearly, the inhomogeneous one seems to report a much better fit.

We, therefore, separate the most influential points from the rest, retaining only those who exhibit chi-squared values larger that the 99^{th} percentile. We note in Figure 5.21 that the two fitted intensities generally identify different influential points. It is also possible to represent the individual LISTA functions and look at the structures that are more clustered (Figure 5.22). We indeed can notice different behaviours. For instance points 49 and 777 display a quite regular and common K-function with the lowest values at the lowest spatial and temporal ranges. However, other points display peculiar behaviours, with the largest values for some specific spatial and temporal ranges. The same considerations can be drawn for the LISTA functions weighted by the constant intensity (Figure 5.23).

So, on the one hand, the constant intensity is not appropriate and this means that an inhomogeneous intensity should be fitted to the data. Nevertheless, this information could have been elicited from the inspection of the global statistics. On the other hand, using LISTA functions has allowed identifying the most influential points and this result could lead to a further study on those particular points. This could be done by treating the point patterns as marked, considering some available characteristics of the events. Then, visual inspection of the areas where most of the influential points occur may guide further analysis of these spatio-temporal areas. For instance, one could think of fitting a parametric intensity that accounts for some characteristics of the network, by means of some available spatial covariates.

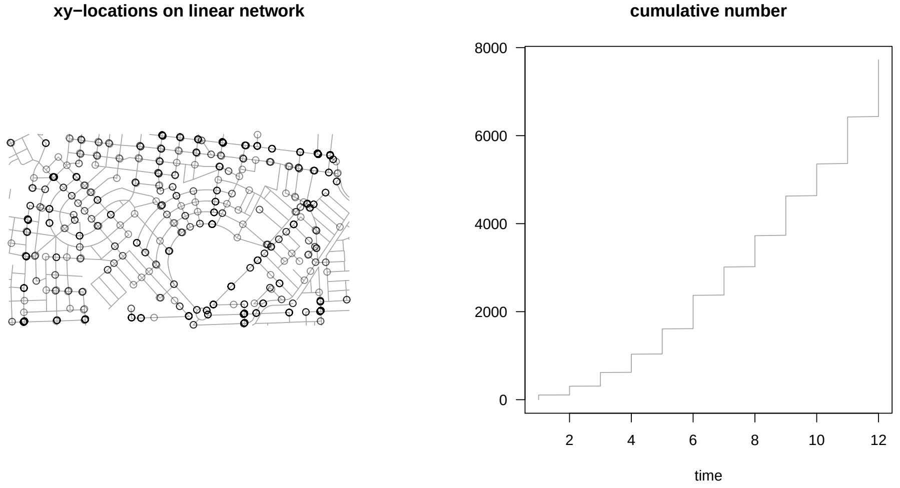


Figure 5.19: The dataset represents the spatio-temporal locations of traffic accidents in the downtown of Medellin (Colombia) in the period of 2019. Left panel: The projection of data onto the network. Right panel: The cumulative number of data points versus occurrence time.

5.2.4 Conclusions

In this Section, we have extended Local Indicators of Spatio-Temporal Association (LISTA) functions on a linear network to their inhomogeneous versions proposing the inhomogeneous version of the local spatio-temporal K and pair correlation functions on the networks, previously introduced. Through simulations, we have shown that the LISTA functions are useful for the diagnostics of models specified on the networks. We have proved, and also shown by simulations, that the proposed methods do not rely on any particular model assumption on the data, and thus they can be applied to whatever the generator model of the process. Therefore, an appealing feature of the method is that it can be used for assessing the goodness-of-fit of spatio-temporal point processes occurring on linear networks characterised by any

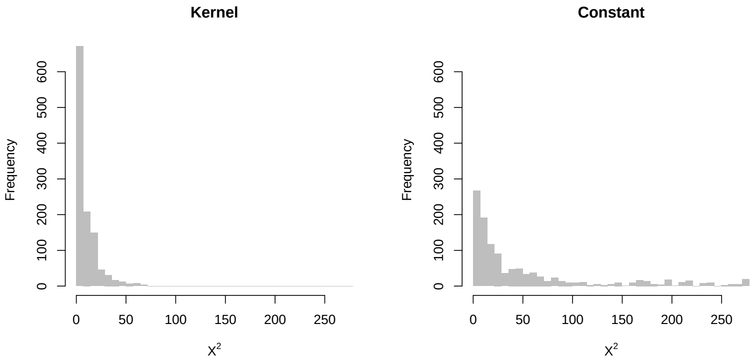


Figure 5.20: Distributions of the χ_i^2 statistics computed for assessing the differences between the local LISTA functions on the network and the theoretical expected value, for the Medellin dataset. Left panel: The LISTA functions are weighted by an inhomogeneous intensity. Left panel: The LISTA functions are weighted by a homogeneous intensity.

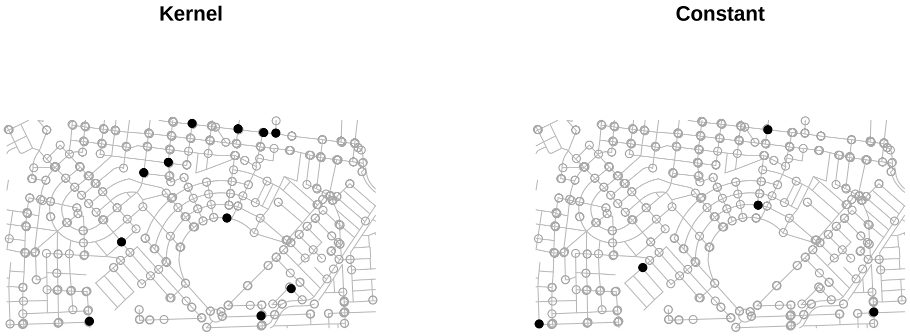


Figure 5.21: In black: Influential points for the Medellin dataset, identified as those that exhibit chi-squared values larger than the 99^{th} percentile. Left panel: The LISTA functions are weighted by an inhomogeneous intensity. Left panel: The LISTA functions are weighted by a homogeneous intensity.

first-order intensity function.

Given the growing availability of proposed models on the networks, we believe that the proposed diagnostic approach could be of great interest. Just to cite some

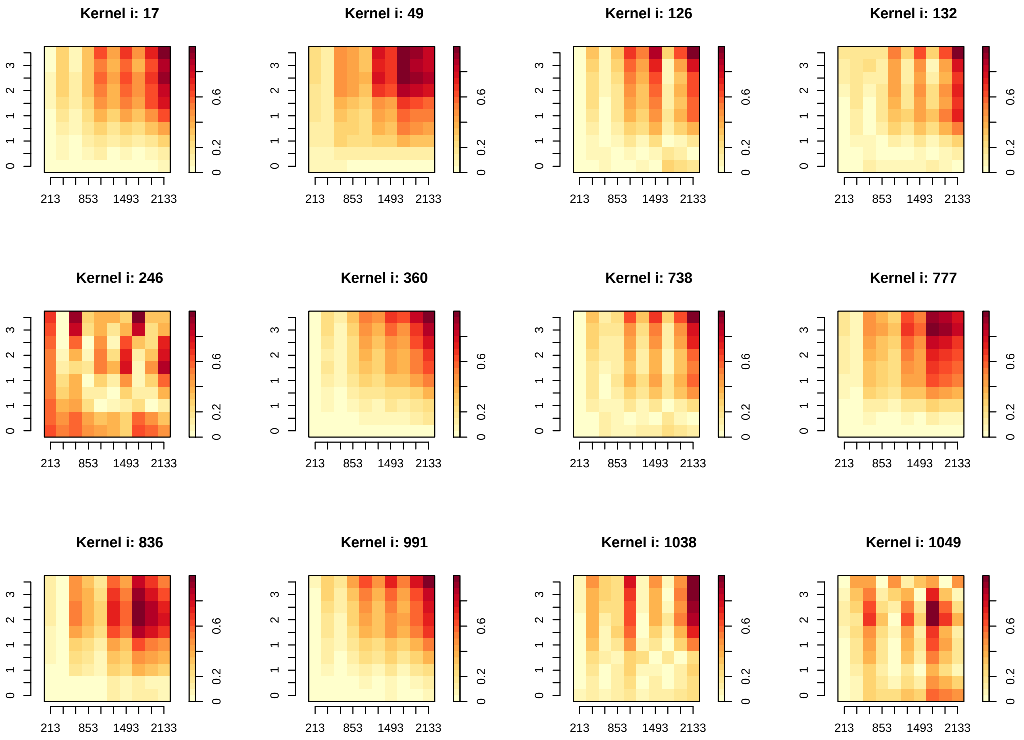
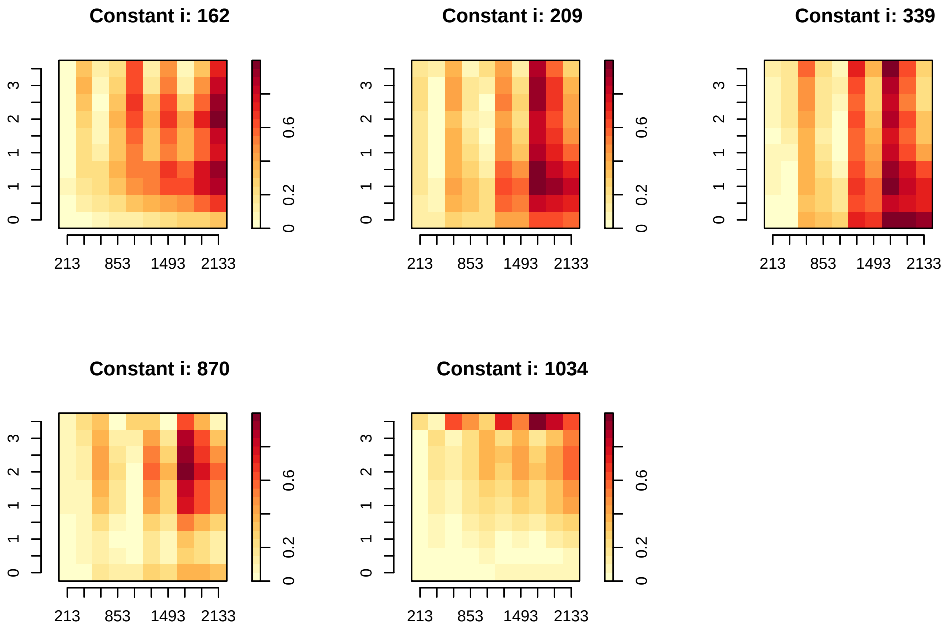


Figure 5.22: LISTA functions of the influential points associated to the inhomogeneous intensity, for the Medellin dataset.

recent works using this methodology, section 4.1 dealt with parametric intensity specification of inhomogeneous first-order intensities on networks to analyse the spatio-temporal distribution of visitors' stops by touristic attractions in Palermo (Italy). The authors fitted a Gibbs point process model with mixed effects for the purely spatial component, as well as a spatio-temporal log-Gaussian Cox process, adapting them to the underlying road network. Then, Gilardi et al. (2021) proposed a spatio-temporal model to analyse the ambulance interventions that occurred in the road network of Milan (Italy) from 2015 to 2017, adopting a non-separable first-order intensity function, based on a Poisson regression model for the temporal component, and a network kernel function for the spatial dimension. Finally, section 4.2 has taken into account the Self-Exciting behaviour of points, proposing a spatio-temporal Hawkes point process model adapted to linear networks to analyse crime data in Bucaramanga (Colombia).

We note that the LISTA functions can be used to fit the first-order intensity function. This option is explored in chapter 7 of this thesis. Namely, summary


Constant i: 162

Constant i: 209

Constant i: 339

Constant i: 870

Constant i: 1034

Figure 5.23: LISTA functions of the influential points associated to the homogeneous intensity, for the Medellin dataset.

statistics such as the K-function or the pair correlation function are commonly employed for some fitting procedures such as the minimum contrast for the estimation of the correlation parameters of Cox processes. Considering the individual contributions of such statistics could provide local estimates: see section 3.2 for a recent application to local log-Gaussian Cox processes. Consequently, our proposed LISTA functions could be used to fit intensities on linear networks, even though, to the authors' knowledge, no attempt has been made in this direction so far.

Chapter 6

Local characteristics of functional marked point processes and their application to seismic data

Seismic events can be characterized both by the spatio-temporal hypocentre locations and by the seismic waveform associated to each event. To correctly detect the arrival of the first P-wave, as well as to perform earthquake forecasting, it is crucial to understand the nature of the generating event. Indeed, while the spatial (and temporal) location of the epicentre of the earthquake is typically studied within the theory of point processes, the seismic waveforms are commonly investigated through Functional Data Analysis (FDA), for earthquake detection and characterization.

In this Section, we want to create a link between these sources of information, relating all the knowledge of the available seismograms to the point pattern locations, defining novel methods for earthquake analyses. This approach formally refers to the theory of Functional Marked Point Processes (FMPPs).

In section 6.1, we first aim at constructing a test of random labelling using the individual contributions of points into the summary statistics used to run the test. This represents an interesting context of application where we would expect waveforms (i.e. functional marks) to be similar to those of near points. Introducing a local test would allow us to identify where this assumption holds or not.

We believe that some interesting research questions could be answered by the analysis of the seismic phenomenon through the theory of FMPPs. With particular reference to the test of random labelling, we could investigate the following: do the waveforms and their characteristics depend on the spatial location? Are there some areas where the above hypothesis does not hold?

The achievement of this unification of data and statistical tools would result in building a framework where it would be possible to exploit the available information of the seismic phenomenon altogether, as well as use them in different contexts of analysis also characterised by a similar structure in terms of the phenomenon and available data.

Then, section 6.2 deals only with the functional aspect of seismic events and proposed a novel statistical picking algorithm for detecting the arrival times of the first P- and S-waves in a seismic waveform. The proposal is compared to a state-of-the-art picking algorithm, namely the STA/LTA approach, and its performance is further demonstrated through an application to real data.

6.1 Local inhomogeneous weighted summary statistics for marked point processes

D'Angelo, N., Adelfio, G., Mateu, J., and Cronie, O. (2022d). Local inhomogeneous weighted summary statistics for marked point processes. *https://arxiv.org/abs/2208.09233 Journal of Computational and Graphical Statistics. Major revisions*

The analysis of a point pattern, given as a collection of points in a region, typically begins with computing an estimate of some summary statistic which may be used to find specific structures in the data and suggest suitable models (Chiu et al., 2013; Daley and Vere-Jones, 2007; Gelfand et al., 2010; Illian et al., 2008; Van Lieshout, 2000). The choice of the summary statistic depends both on the pattern at hand and on the feature or hypothesis of interest.

A widely used summary statistic for descriptive analyses and diagnostics, which is obtained as an instance of the so-called reduced second-moment measure (Cressie and Collins, 2001; Chiu et al., 2013; Moller and Waagepetersen, 2003), is Ripley's K-function (Ripley, 1976), which is based on the assumption of a non-marked stationary and isotropic point process. In the marked case, assuming discrete marks and stationarity, cross versions of the K- or nearest neighbour distance distribution functions have been proposed (Diggle, 2013). For real-valued marks, the mark correlation type-functions in Penttinen and Stoyan (1989); Illian et al. (2008) are widely used and such second-order statistics have been studied in more detail and reformulated by Schlather (2001), in order to obtain a more rigorous formulation. However, although the assumption of stationarity is mathematically appealing, unfortunately, it can rarely be justified in practice, where, mostly, the intensity tends to change over the study region. This is to say that the underlying point process is inhomogeneous and, in the unmarked case, Baddeley et al. (2000) proposed an inhomogeneous extension of the K-function for a class of point processes, which are referred to as second-order intensity-reweighted stationary. Their ideas were extended to spatio-temporal point processes in Gabriel and Diggle (2009); Møller and Ghorbani (2012). Further, Moller and Waagepetersen (2003) proposed an extension of this K-function to second-order intensity-reweighted stationary multivariate point processes. As indicated in Cronie and van Lieshout (2016); Iftimi et al. (2019), this structure may be extended to K-functions for general marked point processes. To analyse higher-order interactions in general stationary marked point processes, Van Lieshout (2006) proposed marked versions of the nearest neighbour distance distribution functions, the empty space function and the J-function. These summary statistics, which allow us to study spatial interactions between different mark groupings of the points, were later extended to the inhomogeneous setting by Cronie and van Lieshout (2016); Iftimi et al. (2019). In particular, to test for random labelling, Cronie and van Lieshout (2016) proposed inhomogeneous Lotwick-Silverman-type Monte Carlo tests based on their new summary statistics, while Iftimi et al. (2019) proposed second-order Monte Carlo tests based on permuting the attached marks. Further details on the random shift-type testing considered in Lotwick-Silverman-type tests can be found in Mrkvička et al. (2021).

Despite the relatively long history of point process theory (see e.g. Diggle, 2013; Stoyan and Stoyan, 1994; Daley and Vere-Jones, 2007), few approaches have been proposed to analyse spatial point patterns where the features of interest are functions/curves instead of qualitative or quantitative variables. Examples of point patterns with associated functional data include forest patterns where for each tree, we have a growth function, curves representing the incidence of an epidemic over a period of time, and the evolution of distinct economic parameters such as unemployment and price rates, all for distinct spatial locations. The study of such configurations allows for analysing the effects of the spatial structure on individual functions. Illian et al. (2006) consider for each point a transformed Ripley (1976)'s K-function to characterise spatial point patterns of ecological plant communities, whilst Mateu et al. (2007) build new marked point processes formed by spatial locations and curves defined in terms of Local Indicators of Spatial Association (LISA) functions, which describe local characteristics of the points. They use this approach to classify and discriminate between points belonging to a clutter and those belonging to a feature. Finally, the idea of analysing point patterns with attached functions has been presented coherently by Comas et al. (2011); Ghorbani et al. (2021).

Ghorbani et al. (2021) introduced a very broad framework for the analysis of Functional Marked Point Processes (FMPPs), indicating how they connect the point process framework with both Functional Data Analysis (FDA; Ramsay and Silverman (2002)) and geostatistics. In particular, they defined a new family of summary statistics, so-called *weighted n-th order marked inhomogeneous K-functions*, together with their non-parametric estimators, which they exploited to analyse Spanish population structures, such as demographic evolution and sex ratio over time. This summary statistic family can be used to run a Monte Carlo test of random labelling, e.g. by means of global envelopes test (GET; Myllymäki et al. (2017)), to assess whether the functional marks of the analysed pattern are spatially dependent. However, this procedure is essentially global, since it does not provide information on the points which mostly contributed to the rejection of the random labelling hypothesis. Therefore, motivated by the need of detecting such points, and thus the regions in which they are located, where the functional marks really do depend on the surrounding structure, in this section we introduce a new class of summary statistics, *local t-weighted marked n-th order inhomogeneous K-functions*. These are used to propose a *local test of random labelling*. Here t refers to a function which governs how much weight we put on different aspects of the marked point process/pattern.

Further, we use the developed tools to analyse seismic data. Note that while the spatial (and temporal) locations of the epicentres of earthquakes are typically analysed within the framework of point processes, the associated seismic waveforms are commonly investigated in separate analyses through FDA. Applying the local test allows us to identify where one would expect waveforms (i.e. functional marks) to be similar to those of nearby points.

All the performed analyses are carried out through the R Core Team (2022) software, and the codes are available from the first author. Preliminary data manipulation is performed through the software Python (Van Rossum and Drake Jr, 1995).

The structure of the section is as follows. In section 6.1.1, the motivation of this work is presented, showing the dataset and problem that will be further analysed along the section. Section 6.1.2 contains some preliminaries on functional marked point processes. In section 6.1.3, we present our proposed local t-weighted n-th order inhomogeneous K-functions and their main properties, also relating them to their global counterparts. Section 6.1.4 outlines the main steps to run a local test of random labelling. In section 6.1.5, we present a motivating example to show the further advantages of a local test, compared to a global one. To have a comprehensive understanding of the performance of the proposed local test, we show simulation results under different scenarios. Section 6.1.6 provides an application to seismic data. Finally, conclusions are drawn in section 6.1.7.

6.1.1 Data and motivation

Earthquakes' detection provides a whole set of data which are usually studied separately, i.e. spatial (and temporal) occurrence of points through point process theory (Siino et al. (2017); Iftimi et al. (2019); D'Angelo et al. (2022g), to cite just a few recent works), and the analysis of waveforms through FDA (Adelfio et al., 2011, 2012; Chiodi et al., 2013).

A recently released set of data on Italian seismic activity encompasses both of these data types. *The Italian seismic dataset for machine learning (INSTANCE)* is a dataset of seismic waveforms data and associated metadata (Michelini et al., 2021), which includes 54008 earthquakes for a total of 1159249 3-channel waveforms. It also contains 132330 3-channel noise waveforms. For each of these waveforms, 115 metadata (i.e. statistical variables) are available, providing information on the station, trace, source, path and quality. Overall, the data are collected by 19 networks which consist of 620 seismic stations. The dataset is available on `http://www.pi.ingv.it/instance/`.

The earthquake list in the dataset is based on the Italian seismic bulletin (`http://terremoti.ingv.it/bsi`) of the "Istituto Nazionale di Geofisica e Vulcanologia", includes events which occurred between January 2005 and January 2020, and in the magnitude range between 0.0 and 6.5. The waveform data have been recorded

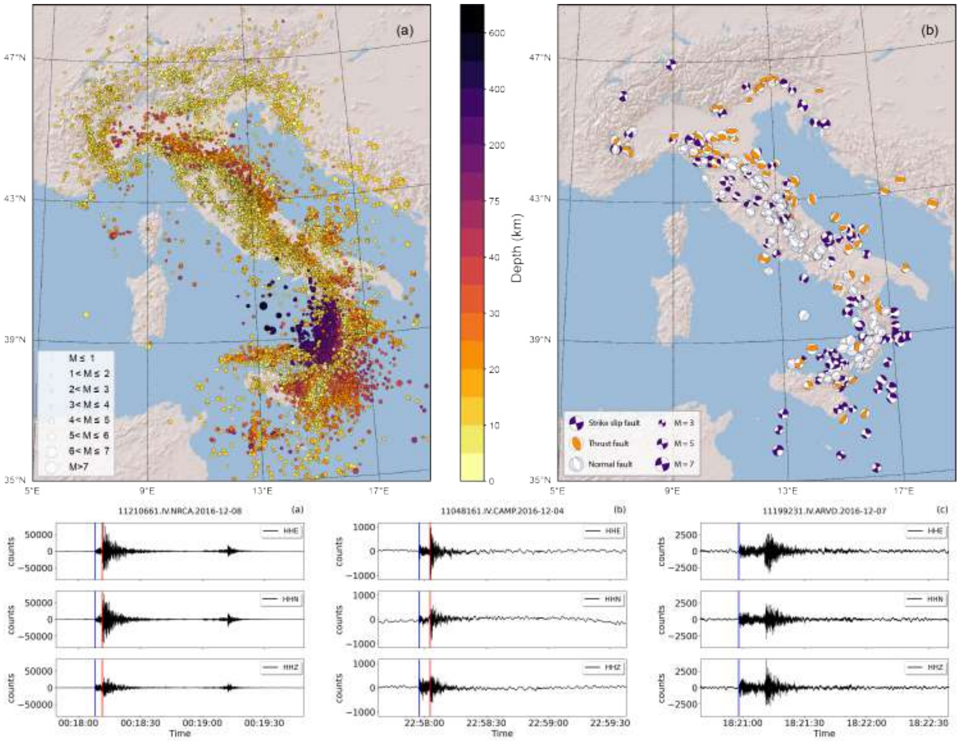


Figure 6.1: Top-left panel: Earthquake locations; Top-right panel: Seismic stations used for waveforms extraction. The symbol sizes are proportional to earthquake magnitude and the number of arrival phases recorded by stations, respectively; Bottom panels: Seismic waveforms of some events with magnitude in the range $[2, 4]$. *Source: Michelini et al. (2021).*

primarily by the Italian National Seismic Network. The top panels of Figure 6.1 depict the earthquake locations and the seismic stations which recorded the events.

In the bottom panels of Figure 6.1, some waveforms contained in the dataset are represented. All the waveform traces have a length of 120 seconds, are sampled at 100 Hz, and are provided both in counts and ground motion physical units after deconvolution of the instrument transfer functions. Details on this filtering procedure are found in Michelini et al. (2021). The waveform dataset is accompanied by metadata consisting of more than 100 variables providing comprehensive information on the earthquake source, the recording stations, the trace features, and other derived quantities.

6.1.2 Preliminaries on marked point processes

Throughout the section, we consider a marked point processe $Y = \{(x_i, m_i)\}_{i=1}^{N}$ (Daley and Vere-Jones, 2007, Definition 6.4.1), with ground points x_i in the d-dimensional Euclidean space $\mathbb{R}^d$, $d \geq 1$, which is equipped with the Lebesgue measure $|A| = \int_A \mathrm{d}z$ for Borel sets $A \in \mathcal{B}(\mathbb{R}^d)$. A closed Euclidean r-ball around $x \in \mathbb{R}^d$ will be denoted by $b[x, r]$. By definition, the ground process $Y_g = \{x_i\}_{i=1}^{N}$, obtained from Y by ignoring the marks, is a well-defined point process on $\mathbb{R}^d$ in its own right. We shall assume that Y is simple, that is, it almost surely (a.s.) does not contain multiple points. Note that, formally, Y is a random element in the measurable space $(N_{lf}, \mathcal{N})$ of locally finite point configurations/patterns $\mathbf{x} = \{((x_1, m_1), \ldots, (x_n, m_n))\}$, $n \geq 0$ (Daley and Vere-Jones, 2007; Van Lieshout, 2000). We assume that the mark space $\mathcal{M}$ is Polish and equipped with a finite reference measure ν on the Borel σ-algebra $\mathcal{B}(\mathcal{M})$. The Borel σ-algebra $\mathcal{B}(\mathbb{R}^d \times \mathcal{M}) = \mathcal{B}(\mathbb{R}^d) \otimes \mathcal{B}(\mathcal{M})$ is endowed with the product measure $A \times E \mapsto |A|\nu(E)$, $A \times E \in \mathcal{B}(\mathbb{R}^d \times \mathcal{M})$. We will let $Y(A \times E) = \sum_{(x,m) \in Y} \mathbf{1}\{(x, m) \in A \times E\}$, where $\mathbf{1}$ is the indicator function, denote the cardinality of the random set $Y \cap (A \times E)$.

Given this general setup, one may obtain various forms of marked point processes, most notably multivariate/multitype point processes with $\mathcal{M} = \{1, \ldots, k\}$ (Diggle, 2013) and functional marked point processes with $\mathcal{M}$ given by a suitable function space (Ghorbani et al., 2021).

6.1.2.1 Functional Marked Point Processes

In this section, we provide the definition of functional marked point processes following Ghorbani et al. (2021).

In classical FDA, one analyses a collection of functions $\{f_1(t), \ldots, f_n(t)\}$, $t \in \mathcal{T} \subset [0, \infty)$, $n \geq 1$, which take values in some Euclidean space $\mathbb{R}^k$, $k \geq 1$, and belong to some suitable function space, typically an L_2-space. Although t usually represents time, it could also represent some other quantity, for example, spatial distance. Classically, one would assume that such a collection of functions constitute realisations or samples $\{F_1(t), \ldots, F_n(t)\}$, $t \in \mathcal{T}$ of some collection of independent and identically distributed (iid) random functions or stochastic processes $F_1, \ldots, F_N$. Such an assumption may, however, be questioned in certain settings. For example, two functions f_i and f_j, which are spatially close to each other in $\mathbb{R}^k$, could gain (or lose) from being close to each other. Accordingly, it seems natural to relax the iid assumption for $F_1, \ldots, F_N$. A natural way to handle such a scenario is to generate $F_1, \ldots, F_N$ conditionally on some collection of (dependent) random spatial locations. Note that the conditional distribution of $F_1, \ldots, F_N$ could render them either dependent or independent. Moreover, note that we assume to work with some fully observed functions at an arbitrarily dense grid. This setting is often unrealistic, but mathematically convenient.

To facilitate such a setting, we consider a functional marked point process (Ghorbani et al., 2021), which is defined as a marked point process where the marks are random elements in some (Polish) function space, $\mathcal{M}$, most notably the space of L_2-functions $f : \mathcal{T} \to \mathbb{R}^k$. Realisations of FMPPs are called functional marked point patterns. It is noteworthy that the original formal construction of functional marked point processes by Ghorbani et al. (2021) also included an additional non-functional mark so that each ground process point would be marked by a pair which consists of a function and a non-functional variable. We here do not consider such auxiliary non-functional marks.

6.1.2.2 Product densities

Provided that it exists, the *n-th order intensity/product density* function $\rho^{(n)}$, $n \geq 1$, which is the density of the n-th order factorial moment measure $\alpha^{(n)}$, may be specified through the n-th order *Campbell formula*. It states that, for any non-negative measurable function h on $(\mathbb{R}^d \times \mathcal{M})^n$, the expectation of the random sum of h over n-tuples of distinct points of Y satisfies

$$\mathbb{E}\left[\sum_{(x_1,m_1),\dots,(x_n,m_n)\in Y}^{\neq} h((x_1,m_1),\dots,(x_n,m_n))\right] =$$
$$= \int \cdots \int h((x_1,m_1),\dots,(x_n,m_n))\rho^{(n)}((x_1,m_1),\dots,(x_n,m_n)) \prod_{i=1}^{n} \mathrm{d}x_i \nu(dm_i). \tag{6.1}$$

Heuristically, $\rho^{(n)}((x_1,m_1),\dots,(x_n,m_n))\mathrm{d}x_1\nu(dm_1)\cdots\mathrm{d}x_n\nu(dm_n)$ gives the probability that Y has points in infinitesimal neighbourhoods $d(x_i,m_i) \ni (x_i,m_i) \in \mathbb{R}^d \times \mathcal{M}$ with measures $\mathrm{d}x_i\nu(dm_i)$, $i = 1,\dots,n$. Moreover, we retrieve $\alpha^{(n)}((A_1 \times E_1) \times \cdots \times (A_n \times E_n))$, $(A_i \times E_i) \in \mathcal{B}(\mathbb{R}^d \times \mathcal{M})$, $i = 1,\dots,n$, by letting h be given by the indicator function $\mathbf{1}\{(x_1,m_1) \in (A_1 \times E_1),\dots,(x_n,m_n) \in (A_n \times E_n)\}$. It further follows that

$$\rho^{(n)}((x_1,m_1),\dots,(x_n,m_n)) = f_{x_1,\dots,x_n}(m_1,\dots,m_n)\rho_g^{(n)}(x_1,\dots,x_n),$$

where $\rho_g^{(n)}$ is the n-th order product density of Y_g and $f_{x_1,\dots,x_n}(\cdot)$ is a conditional density function on $\mathcal{M}^n$ which governs the joint distribution of n marks, given that their associated ground process points are given by $x_1,\dots,x_n \in \mathbb{R}^d$. These, in turn, yield the corresponding mark distributions

$$M^{x_1,\dots,x_n}(E_1,\dots,E_n) = \int_{E_1} \cdots \int_{E_n} f_{x_1,\dots,x_n}(m_1,\dots,m_n) \prod_{i=1}^{n} \nu(dm_i),$$

which govern the joint distribution on n marks, given the associated ground process locations.

The intensity measure of Y, which coincides with the first-order factorial moment measure, here satisfies

$$\begin{aligned}\alpha(A \times E) =&\mathbb{E}[Y(A \times E)] = \int_A \int_E \rho(x,m)\mathrm{d}x\nu(dm) \\ &= \int_A \int_E f_x(m)\rho_g(x)\mathrm{d}x\nu(dm) = \int_A M^x(E)\rho_g(x)\mathrm{d}x, \qquad (6.2)\end{aligned}$$

where the first order intensity functions $\rho = \rho^{(1)}$ and $\rho_g = \rho_g^{(1)}$ are typically referred to as *the intensity functions* of Y and Y_g. Note that ρ may be viewed as a "heat map" which reflects the infinitesimal chance of having a point of Y at/around an arbitrary location in $\mathbb{R}^d \times \mathcal{M}$. When the intensity function (of the ground process) is constant, we say that the (ground) process is homogeneous, otherwise, it is called inhomogeneous.

When conditional on the ground process, all marks have the same marginal univariate distribution, so that $M^z(E) = \int_E f_z(m)\mathrm{d}\nu(dm) = \int_E f(m)\mathrm{d}\nu(m) = M(E)$, we say that X has a common (marginal) mark distribution. This holds e.g. when Y is stationary, i.e. when its distribution is invariant under translations of the ground points. Here $\alpha(A \times E) = \rho_g M(E)|A|$ and $\rho_g > 0$ is the constant intensity of the ground process. We will see that, at times, it is particularly convenient to have here that the reference measure ν coincides with the common mark distribution M, which implies that the common mark density f is set to 1 and $\rho(x,m) = \rho_g(x)$.

When Y is independently marked, i.e. when the marks are independent conditional on the ground process, $f_{x_1,\dots,x_n}(m_1,\dots,m_n) = f_{x_1}(m_1)\cdots f_{x_n}(m_n)$ for any $n \geq 1$ and if, in addition, there is a common mark distribution, whereby the marks are iid conditional on the ground process, we say that Y is randomly labelled and note that $f_{x_1,\dots,x_n}(m_1,\dots,m_n) = f(m_1)\cdots f(m_n)$.

6.1.2.2.1 Intensity reweighted stationarity We next turn to the notion of a k-th order marked intensity reweighted stationary (k-MIRS) marked point process Y (Ghorbani et al., 2021). We say that Y is k-MIRS, $k \in \{1,2,\dots\}$, if ρ is bounded away from 0 and the correlation functions

$$\begin{aligned}&g^{(n)}((x_1,m_1),\dots,(x_n,m_n)) = \\ &= \frac{\rho^{(n)}((x_1,m_1),\dots,(x_n,m_n))}{\rho(x_1,m_1)\cdots\rho(x_n,m_n)} = \frac{f_{x_1,\dots,x_n}(m_1,\dots,m_n)}{f_{x_1}(m_1)\cdots f_{x_n}(m_n)} \frac{\rho_g^{(n)}(x_1,\dots,x_n)}{\rho_g(x_1)\cdots\rho_g(x_n)}, \qquad n \geq 1,\end{aligned}$$

satisfy $g^{(n)}((x_1,m_1),\dots,(x_n,m_n)) = g^{(n)}((z+x_1,m_1),\dots,(z+x_n,m_n))$ for any $z \in \mathbb{R}^d$ and any $n \leq k$. Note that $g^{(1)}(\cdot) \equiv 1$ and that the second ratio on the right-hand side is the n-th order correlation function, $g_g^{(n)}$, of the ground process. Provided that the product densities of all orders exist, stationarity implies k-MIRS for all orders $k \geq 1$. Note further that $g^{(n)}(\cdot) \equiv 1$, $n \geq 1$, for a Poisson process and when $g^{(n)}((x_1, m_1), \ldots, (x_n, m_n)) > 1$ points of Y_g in infinitesimal neighbourhoods of $x_1, \ldots, x_n$ with marks in infinitesimal neighbourhoods of $m_1, \ldots, m_n$ tend to cluster/aggregate. Similarly, $g^{(n)}((x_1, m_1), \ldots, (x_n, m_n)) < 1$ indicates inhibition/regularity.

6.1.2.3 Palm distributions

Let Y be a simple marked point process whose intensity function exists. Many of the summary statistics we will consider can be expressed in terms of *reduced Palm distributions*. These satisfy the *reduced Campbell-Mecke* formula which states that, for any non-negative measurable function h on the product space $(\mathbb{R}^d \times \mathcal{M}) \times N_{lf}$,

$$\begin{aligned}\mathbb{E}\left[\sum_{(z,m)\in Y} h((z,m), Y\backslash\{(z,m)\})\right] &= \int \mathbb{E}[h((x,m), Y^{!(x,m)})]\rho(x,m)\mathrm{d}x\nu(dm) \qquad (6.3)\\ &= \int \mathbb{E}^{!(x,m)}[h((x,m), Y)]\rho(x,m)\mathrm{d}x\nu(dm).\end{aligned}$$

Here $Y^{!(x,m)}$ is the reduced Palm process at $(x,m) \in \mathbb{R}^d \times \mathcal{M}$, which we interpret as Y conditioned on the null event that there is a point in (x,m), which is removed upon realisation. The probability distribution $P^{!(x,m)}(\cdot) = \mathbb{P}^{!(x,m)}(Y \in \cdot) = \mathbb{P}(Y^{!(x,m)} \in \cdot)$ on $(N_{lf}, \mathcal{N})$, which corresponds to $\mathbb{E}^{!(x,m)}$, is called the reduced Palm distribution at (x,m).

6.1.3 Local weighted marked summary statistics

Global summary statistics have had a prominent role in the statistical analysis of point processes. More precisely, their non-parametric estimators are typically used to characterise the degree of spatial interaction present in the underlying data-generating point process. In section 6.1, we have reviewed a few such examples, for instance, K-functions.

The individual contributions to a global statistic, which are commonly called Local Indicators of Spatial Association (LISA) functions, can be used to identify outlying components measuring the influence of each contribution to the global statistic (Anselin, 1995). This is the case of the scatter plot based on the local Moran index (Anselin, 1996). On the other hand, individual contributions can be used to test for specific local structures, such as spatial association and hot spot detection in areal data (Getis and Ord, 1992). Basically, the local statistics mentioned so far are often used to analyse areal data but Getis and Franklin (1987) introduced a local version of the K-function for spatial point processes to show that trees exhibit different kinds of heterogeneity when examined at different scales of analysis. The notion of individual functions for certain statistics has also been studied in Stoyan and Stoyan (1994) and Mateu et al. (2010) showed that the local product density function (Cressie and Collins, 2001) is more sensitive to identifying different local structures and unusual points than the local K-function. Applications of LISA functions range from the detection of features in images with noise (Mateu et al., 2007) to the detection of disease clusters (Moraga and Montes, 2011). In Siino et al. (2018b) the authors extend local indicators of spatial association to the spatio-temporal context (LISTA functions) based on the second-order product density, and these local functions have been used to define a proper statistical test for clustering detection. Recently, LISTA functions have been used both for diagnostic (Adelfio et al., 2020) and fitting purposes (D'Angelo et al., 2022c). Finally, D'Angelo et al. (2021b) extended LISTA functions to spatio-temporal point processes living on linear networks.

As we have clearly indicated, an alternative to studying the aforementioned global summary statistics for marked point processes is considering local summary statistics which describe the spatial interaction in the vicinity of a given marked point. In order to do so here in the marked context, we introduce the function

$$\begin{aligned} L((x,m),\mathbf{x}) &= L_n((x,m),\mathbf{x};\tilde{t},\tilde{\rho}) = \\ &= \sum_{(x_1,m_1),\dots,(x_{n-1},m_{n-1})\in\mathbf{x}}^{\neq} \frac{\tilde{t}((x,m),(x_1,m_1),\dots,(x_{n-1},m_{n-1}))}{\tilde{\rho}(x,m)\tilde{\rho}(x_1,m_1)\cdots\tilde{\rho}(x_{n-1},m_{n-1})}, \end{aligned} \tag{6.4}$$

for $(x,m)\in\mathbb{R}^d\times\mathcal{M}$, point pattern $\mathbf{x}\in N_{lf}$ and measurable $\tilde{t}:(\mathbb{R}^d\times\mathcal{M})^n\to\mathbb{R}$, $n\geq 2$. Note that, formally, the argument $\tilde{\rho}$ does not need to be the true intensity function ρ of Y, it could e.g. be a plug-in estimator. We will exploit Definition 6.1, and thereby (6.4), to define proper notions of (mark-weighted n-th order inhomogeneous) local summary statistics.

Definition 6.1. *Given a marked point process Y, we refer to the family of n-th order local marked cumulative summary statistics of Y associated with $\tilde{t}$ and $\tilde{\rho}$, as $L((x,m),Y\setminus\{(x,m)\};\tilde{t},\tilde{\rho})$, $(x,m)\in Y$.*

The construction of a specific local statistic is obtained by identifying when, for some function family $\{\tilde{t}_r\}$,

$$G(r,Y) = \sum_{(x,m)\in Y} L_n((x,m),Y\setminus\{(x,m)\};\tilde{t}_r,\tilde{\rho}) \tag{6.5}$$

forms an estimator of an existing global summary statistic.

Using n-th order local marked cumulative summary statistics to quantify local spatial interactions for a point pattern $\mathbf{x}$ entails inserting an estimate $\widehat{\rho}(x,m) = \widehat{f_z}(m)\widehat{\rho}_g(x)$ for the unknown intensity $\rho(x,m) = f_z(m)\rho_g(x)$, i.e. setting $\tilde{\rho} = \widehat{\rho}$. When we assume that there is a common mark distribution which coincides with the mark reference measure ν, we obtain that $\widehat{\rho}(x,m) = \widehat{\rho}_g(x)$, i.e. the intensity estimate does not depend on the mark values. Imposing this assumption is particularly convenient when dealing with functional marks since the estimation of the mark density, which here is a density on a function space, is rather challenging and beyond the scope of this work. Note that when Y is randomly labelled, it has a common mark distribution and in this setting the assumption $\widehat{\rho}(x,m) = \widehat{\rho}_g(x)$ thus makes sense.

Turning to the distributional properties of the n-th order local marked cumulative summary statistics, we next derive their expectations under the assumption of k-MIRS. Note, in particular, that the choice of $\tilde{t}$ plays a significant role here.

Theorem 6.1. *When Y is k-MIRS and $\tilde{\rho} = \rho$, for any $W \in \mathcal{B}(\mathbb{R}^d)$ the expectation of $L((x,m), Y \setminus \{(x,m)\} \cap W \times \mathcal{M}; \tilde{t}, \rho)$, $(x,m) \in Y \cap W \times \mathcal{M}$, is almost everywhere given by*

$$\int_{W-x} \cdots \int_{W-x} \left(\int_{\mathcal{M}} \cdots \int_{\mathcal{M}} \tilde{t}((x,m),(x_1+x,m_1),\ldots,(x_{n-1}+x,m_{n-1})) \times \right.$$
$$\left. \times \frac{f_{0,x_1,\ldots,x_{n-1}}(m,m_1,\ldots,m_{n-1})}{f_0(m) f_{x_1}(m_1)\cdots f_{x_{n-1}}(m_{n-1})} \nu(dm_1)\cdots\nu(dm_{n-1}) \right) \times$$
$$\times\, g_g^{(n)}(0,x_1,\ldots,x_{n-1})\, dx_1 \cdots dx_{n-1},$$

when $2 \le n \le k$. Moreover, the expectation of $G(r, Y \cap W \times \mathcal{M})$ is obtained by replacing $\tilde{t}$ by $\tilde{t}_r$ in the expression above and integrating it over $W \times \mathcal{M}$ with respect to the reference measure on $\mathbb{R}^d \times \mathcal{M}$.

Proof. Note first that the expectation coincides with

$$\mathbb{E}^{!(x,m)}[L_n((x,m), Y \cap W \times \mathcal{M}; \tilde{t}, \rho)] =$$
$$= \mathbb{E}^{!(x,m)} \left[\sum_{(x_1,m_1),\ldots,(x_{n-1},m_{n-1}) \in Y \cap W \times \mathcal{M}}^{\neq} \frac{\tilde{t}((x,m),(x_1,m_1),\ldots,(x_{n-1},m_{n-1}))}{\rho(x,m)\rho(x_1,m_1)\cdots\rho(x_{n-1},m_{n-1})} \right].$$

Hence, our starting point will be the reduced Campbell-Mecke formula. Consider an arbitrary bounded $A \times E \in \mathcal{B}(\mathbb{R}^d \times \mathcal{M})$. It follows that

$$\mathbb{E}\left[\sum_{(x,m)\in Y\cap A\times E} \sum_{(x_1,m_1),\ldots,(x_{n-1},m_{n-1}) \in Y\setminus\{(x,m)\}\cap W\times\mathcal{M}}^{\neq} \frac{\tilde{t}((x,m),(x_1,m_1),\ldots,(x_{n-1},m_{n-1}))}{\rho(x,m)\rho(x_1,m_1)\cdots\rho(x_{n-1},m_{n-1})} \right] =$$
$$= \int_{A\times E} \mathbb{E}^{!(x,m)} \left[\sum_{(x_1,m_1),\ldots,(x_{n-1},m_{n-1}) \in Y\cap W\times\mathcal{M}}^{\neq} \frac{\tilde{t}((x,m),(x_1,m_1),\ldots,(x_{n-1},m_{n-1}))}{\rho(x_1,m_1)\cdots\rho(x_{n-1},m_{n-1})} \right] \mathrm{d}x\nu(dm).$$

On the other hand, by the Campbell formula, we have that

$$
\begin{aligned}
&\mathbb{E}\left[\sum_{(x,m)\in Y\cap A\times E}\ \sum^{\neq}_{(x_1,m_1),\dots,(x_{n-1},m_{n-1})\in Y\setminus\{(x,m)\}\cap W\times\mathcal{M}} \frac{\tilde{t}((x,m),(x_1,m_1),\dots,(x_{n-1},m_{n-1}))}{\rho(x,m)\rho(x_1,m_1)\cdots\rho(x_{n-1},m_{n-1})}\right] = \\
&= \int_{A\times E}\int_{\mathbb{R}^d\times\mathcal{M}}\cdots\int_{\mathbb{R}^d\times\mathcal{M}} \mathbf{1}\{x_1,\dots,x_{n-1}\in W\}\tilde{t}((x,m),(x_1,m_1),\dots,(x_{n-1},m_{n-1}))\times \\
&\quad \times g^{(n)}((x,m),(x_1,m_1),\dots,(x_{n-1},m_{n-1}))\mathrm{d}x_1\nu(dm_1)\cdots\mathrm{d}x_{n-1}\nu(dm_{n-1})\mathrm{d}x\nu(dm) \\
&= \int_{A\times E}\int_{\mathbb{R}^d\times\mathcal{M}}\cdots\int_{\mathbb{R}^d\times\mathcal{M}} \prod_{i=1}^{n-1}\mathbf{1}\{u_i\in W-x\}\tilde{t}((x,m),(u_1+x,m_1),\dots,(u_{n-1}+x,m_{n-1}))\times \\
&\quad \times g^{(n)}((0,m),(u_1,m_1),\dots,(u_{n-1},m_{n-1}))\mathrm{d}u_1\nu(dm_1)\cdots\mathrm{d}u_{n-1}\nu(dm_{n-1})\mathrm{d}x\nu(dm)
\end{aligned}
$$

by the imposed k-MIRS and a change of variables, $u_i + x = x_i$. Hence, since $A\times E\in\mathcal{B}(\mathbb{R}^d\times\mathcal{M})$ was arbitrary, for almost every (x,m) we have that

$$
\begin{aligned}
&\mathbb{E}^{!(x,m)}[L_n((x,m),Y;\tilde{t},\rho)] = \\
&= \int_{(W-x)\times\mathcal{M}}\cdots\int_{(W-x)\times\mathcal{M}} \tilde{t}((x,m),(u_1+x,m_1),\dots,(u_{n-1}+x,m_{n-1}))\times \\
&\quad \times g^{(n)}((0,m),(u_1,m_1),\dots,(u_{n-1},m_{n-1}))\mathrm{d}u_1\nu(dm_1)\cdots\mathrm{d}u_{n-1}\nu(dm_{n-1}) \\
&= \int_{W-x}\cdots\int_{W-x}\Bigg(\int_{\mathcal{M}}\cdots\int_{\mathcal{M}} \tilde{t}((x,m),(u_1+x,m_1),\dots,(u_{n-1}+x,m_{n-1}))\times \\
&\quad \times \frac{f_{0,u_1,\dots,u_{n-1}}(m,m_1,\dots,m_{n-1})}{f_0(m)f_{u_1}(m_1)\cdots f_{u_{n-1}}(m_{n-1})}\nu(dm_1)\cdots\nu(dm_{n-1})\Bigg)\times \\
&\quad \times g_g^{(n)}(0,u_1,\cdots,u_{n-1})\mathrm{d}u_1\cdots\mathrm{d}u_{n-1},
\end{aligned}
$$

by Fubini's theorem.

□

The first thing we note is that when Y is independently marked than the density ratio in the expression for the expectation vanishes. In addition, if Y is a Poisson process on $\mathbb{R}^d\times\mathcal{M}$ which satisfies being a marked point process with mark space $\mathcal{M}$, then the expectation reduces to an integral with $\tilde{t}$ as integrand. These observations may be used as benchmarks for when Y exhibits mark (in)dependence and spatial interaction locally.

6.1.3.1 Special cases

We next illustrate how (6.5), through Definition 6.1 and (6.4), reduces to several existing summary statistic estimators by varying $\tilde{t}$ and $\tilde{\rho}$.

6.1.3.1.1 Ground K-functions First, set $n = 2$ and $\tilde{t}$ to $\tilde{t}_r((x, m), (x_1, m_1)) = w(x, x_1)\mathbf{1}\{x_1 \in x + C\}/|W|$, $r \geq 0$, where $W \subseteq \mathbb{R}^d$, $|W| > 0$, and $w(\cdot)$ is an edge correction term. If the ground process is stationary with intensity $\rho_g > 0$ and $\tilde{\rho}(x, m) \equiv \rho_g$, then (6.5) with Y set to $Y \cap W \times \mathcal{M}$ reduces to an estimator of Ripley's K-function when $x + C = x + b[0, r] = b[x, r]$ whereas if the ground process is inhomogeneous and we set $\tilde{\rho}(x, m) = \rho_g(x)$, it follows that (6.5) reduces to an estimator of the inhomogeneous K-function (Baddeley et al., 2000) for Y_g. The extension to space-time is straightforward: replace the Euclidean ball $b[0, r]$ by $C = \{(x, s) : \|x\| \leq r, |s| \leq t\} \in \mathcal{B}(\mathbb{R}^{d+1})$, where $\|\cdot\|$ denotes the Euclidean norm (Cronie and Van Lieshout, 2015; Gabriel and Diggle, 2009; Iftimi et al., 2019).

6.1.3.1.2 Marked K-functions When $n = 2$, if we let $\tilde{t}_r((x, m), (x_1, m_1)) = w(x, x_1)\mathbf{1}\{x_1 \in x+C\}\mathbf{1}\{m \in E, m_1 \in E_1\}/(|W|\nu(E)\nu(E_1))$ and $\tilde{\rho} = \rho$ in (6.4), using a suitable edge correction function $w(\cdot)$, then $G(r, Y \cap W \times \mathcal{M})$ in (6.5) reduces to an estimator of the *marked second-order reduced moment measure* $\mathcal{K}^{EE_1}(C)$ of Iftimi et al. (2019), which measures the intensity reweighted interactions between points with marks in E and points with marks in E_1, when their separation vectors belong to $C \in \mathcal{B}(\mathbb{R}^d)$. We note that measures of this kind are in general not symmetric, i.e. $\mathcal{K}^{EE_1}(\cdot) \neq \mathcal{K}^{E_1E}(\cdot)$ (Iftimi et al., 2019). Furthermore, choosing C to be the closed origin-centred ball $b[0, r]$ of radius $r \geq 0$, we consider the marked inhomogeneous K-function $K^{EE_1}_{inhom}(r)$ of Cronie and van Lieshout (2016), which measures pairwise intensity reweighted spatial dependence within distance r between points with marks in E and points with marks in E_1.

By additionally letting $n > 2$, we obtain a definition of a *marked n-th order reduced moment measure*, $\mathcal{K}^{E\times_{i=1}^{n-1}E_i}(C_1 \times \cdots \times C_{n-1})$, which measures the intensity reweighted spatial interaction between an arbitrary point with mark in E and distinct $(n-1)$-tuples of other points, where the separation vectors between the E-marked point and these $n-1$ points, which have marks in $E_1, \ldots, E_{n-1}$, belong to $C_1, \ldots, C_{n-1}$. We note that $C_i = b[0, r]$, $i = 1, \ldots, n-1$, $r \geq 0$, yields an n-point version of the marked inhomogeneous K-function $K^{E\times_{i=1}^{n-1}E_i}_{inhom}(r)$ of Cronie and van Lieshout (2016), which may be used to analyse intensity reweighted interactions between a point with mark in E and $n-1$ of its r-close neighbours, which have marks belonging to the respective sets $E_1, \ldots, E_{n-1}$.

6.1.3.1.3 Weighted marked reduced moment measures and K-functions Finally, by letting $\tilde{\rho} = \rho$ and $\tilde{t}((x, m), (x_1, m_1), \ldots, (x_{n-1}, m_{n-1}))$ be given by the product of

$$\begin{aligned}\tilde{t}(m, m_1, \ldots, m_{n-1}) =& t(m, m_1, \ldots, m_{n-1}) \frac{\mathbf{1}\{m \in E\}}{\nu(E)} \prod_{i=1}^{n-1} \frac{\mathbf{1}\{m_i \in E_i\}}{\nu(E_i)}, \\ \tilde{w}(x, x_1, \ldots, x_{n-1}) =& w(x, x_1, \ldots, x_{n-1}) \prod_{i=1}^{n-1} \mathbf{1}\{x_i \in (x + C_i)\},\end{aligned} \tag{6.6}$$

for $E \in \mathcal{B}(\mathcal{M})$, $\nu(E) > 0$, and $C_i \times E_i \in \mathcal{B}(\mathbb{R}^d) \times \mathcal{B}(\mathcal{M}) = \mathcal{B}(\mathbb{R}^d \times \mathcal{M})$, $\nu(E_i) > 0$, $i = 1, \ldots, n-1$, we obtain an unbiased estimator $\hat{\mathcal{K}}_t^{E \times_{i=1}^{n-1} E_i}(C_1 \times \cdots \times C_{n-1}) = G(r, Y \cap W \times \mathcal{M})$ of the t-weighted marked n-th order reduced moment measure of Ghorbani et al. (2021),

$$\begin{aligned}\mathcal{K}_t^{E \times_{i=1}^{n-1} E_i}(C_1 \times \cdots \times C_{n-1}) = & \frac{1}{|W| \nu(E) \prod_{i=1}^{n-1} \nu(E_i)} \times \\ \times \mathbb{E}\Bigg[\sum_{(x,m) \in Y \cap W \times E} & \sum_{(x_1,m_1),\ldots,(x_{n-1},m_{n-1}) \in Y \setminus \{(x,m)\}}^{\neq} \frac{t(m, m_1, \ldots, m_{n-1})}{\rho(x, m)} \times \\ \times \prod_{i=1}^{n-1} & \frac{\mathbf{1}\{x_i - x \in C_i\} \mathbf{1}\{m_i \in E_i\}}{\rho(x_i, m_i)} \Bigg]\end{aligned} \tag{6.7}$$

assuming that the edge correction function w is such that unbiasedness holds. Examples of such w include the minus sampling edge correction and the translational edge correction (Ghorbani et al., 2021). Note here that one just as well could have merged the scaled indicators in the expression for $\tilde{t}$ with t so that $\tilde{t} = t$. Ghorbani et al. (2021) included this mark set filtering to highlight that their summary statistic generalises previously proposed ones.

6.1.3.2 Local t-weighted nth-order marked inhomogeneous K-function

In this section, we provide the estimator corresponding to the local contributions of (6.7) and discuss its properties.

Definition 6.2. *Let $\tilde{t}$ be (up to indicator-scaling) as in (6.6) and consider*

$$\begin{aligned}&\hat{\mathcal{K}}_t^{(x,m) \times_{i=1}^{n-1} E_i}(C_1 \times \cdots \times C_{n-1}) = L_n((x, m), Y \setminus \{(x, m)\} \cap W \times \mathcal{M}; \tilde{t}, \tilde{\rho}) = \\ =& \frac{1}{\tilde{\rho}(x, m) \nu(E) \prod_{i=1}^{n-1} \nu(E_i)} \sum_{(x_1,m_1),\ldots,(x_{n-1},m_{n-1}) \in Y \setminus \{(x,m)\} \cap W \times \mathcal{M}}^{\neq} w(x, x_1, \ldots, x_{n-1}) \times \\ & \times t(m, m_1, \ldots, m_{n-1}) \prod_{i=1}^{n-1} \frac{\mathbf{1}\{x_i - x \in C_i\} \mathbf{1}\{m_i \in E_i\}}{\tilde{\rho}(x_i, m_i)}, \qquad (x, m) \in Y \cap W \times \mathcal{M},\end{aligned} \tag{6.8}$$

for some suitable edge correction w in (6.6), $W \in \mathcal{B}(\mathbb{R}^d)$, $E \in \mathcal{B}(\mathcal{M})$, $\nu(E) > 0$, and $C_i \times E_i \in \mathcal{B}(\mathbb{R}^d \times \mathcal{M})$, $\nu(E_i) > 0$, $i = 1, \dots, n-1$. We refer to $\hat{\mathcal{K}}_t^{(x,m)\times_{i=1}^{n-1} E_i}(r) = \hat{\mathcal{K}}_t^{(x,m)\times_{i=1}^{n-1} E_i}(b[0,r]^{n-1})$, $r \geq 0$, as a local t-weighted marked n-th order inhomogeneous K-function. In particular, $\hat{\mathcal{K}}_{t,n}^{(x,m)}(r) = \hat{\mathcal{K}}_t^{(x,m)\times \mathcal{M}^{n-1}}(r)$ does not perform any explicit mark set filtering.

Note first that when there is a common mark distribution which coincides with the reference measure on $\mathcal{M}$, setting $\tilde{\rho} = \rho$ we, for instance, obtain

$$\hat{\mathcal{K}}_{t,n}^{(x,m)}(r) =$$

$$= \sum_{(x_1,m_1),\dots,(x_{n-1},m_{n-1}) \in Y \setminus \{(x,m)\} \cap (b[x,r]\cap W)\times\mathcal{M}}^{\neq} \frac{t(m, m_1, \dots, m_{n-1}) w(x, x_1, \dots, x_{n-1})}{\rho_g(x)\rho_g(x_1)\cdots\rho_g(x_{n-1})}$$

since ν must be a probability measure here.

Regarding the distributional properties of (6.8), when Y is k-MIRS, Theorem 6.1 tells us that the expectation is given by

$$\frac{1}{\nu(E)} \prod_{i=1}^{n-1} \frac{1}{\nu(E_i)} \int_{\mathbb{R}^d} \cdots \int_{\mathbb{R}^d} w(x, x_1 + x, \dots, x_{n-1} + x) \prod_{i=1}^{n-1} \mathbf{1}\{x_i \in (x + C_i) \cap (W - x)\} \times$$

$$\times \left(\int_{E_1} \cdots \int_{E_{n-1}} t(m, m_1, \dots, m_{n-1}) \frac{f_{0,x_1,\dots,x_{n-1}}(m, m_1, \dots, m_{n-1})}{f_0(m) f_{x_1}(m_1) \cdots f_{x_{n-1}}(m_{n-1})} \nu(dm_1) \cdots \nu(dm_{n-1}) \right) \times$$

$$\times g_g^{(n)}(0, x_1, \dots, x_{n-1}) \mathrm{d}x_1 \cdots \mathrm{d}x_{n-1}.$$

In particular, under independent marking the mark-related integral within brackets reduces to $\int_{E_1} \cdots \int_{E_{n-1}} t(m, m_1, \dots, m_{n-1}) \nu(dm_1) \cdots \nu(dm_{n-1})$, whereby (6.8) is given by the product of this term and a term measuring intensity reweighted spatial interaction.

6.1.3.2.1 Test functions for FMPPs

Turning to the FMPP case, by choosing different test functions $t(\cdot)$ for the functional marks, we may extract different features. We here focus on pairwise interactions, i.e. $n = 2$.

The test function t is intended to reflect similarities between functions. Hence, a natural starting point would be a metric $t(f_1, f_2) = d(f_1, f_2)$ on the function space $\mathcal{M}$, which does not necessarily need to be the underlying assumed metric on $\mathcal{M}$. The first candidate that comes to mind is an L_p-distance:

$$t(f_1, f_2) = \left(\int_a^b |f_1(t) - f_2(t)|^p \mathrm{d}t \right)^{1/p}, \quad 1 \leq p \leq \infty, \tag{6.9}$$

where $p = \infty$ represents the supremum metric. For any choice of p in (6.9), similarity between functions implies a small value of the test function. Other tentative functions are semi-metrics based on the L_p distance between the s-th derivatives of the functions, for different combinations of p and s, with the L_1 and L_2 distances being particular cases, and semi-metrics based on functional principal component analysis.

A further alternative is the functional marked counterpart of the test function for the classical variogram, given by

$$t(f_1, f_2) = \int_a^b (f_1(t) - \bar{F}(t))(f_2(t) - \bar{F}(t))\mathrm{d}t, \tag{6.10}$$

with $\bar{F}(t) = (1/n)\sum_{i=1}^n f_i(t)$ being the average functional mark at time t for the observed functional part of the point pattern. Such averaging is motivated by the assumption of a common mark distribution.

6.1.4 Local test for random labelling

Simple hypotheses for spatial point patterns, such as Complete Spatial Randomness, are commonly tested using an estimator of a global summary statistic, e.g., Ripley's K-function. In this context, one typically resorts to Monte Carlo testing. The first step is then to generate Q simulations under the null hypothesis, and to estimate the chosen summary statistic for both the observed pattern and the simulations. In order to study whether there is *random labelling* in a (functional) marked point process, the simulations are obtained by permuting the (functional) marks, that is, randomly assigning them to the spatial points of the ground pattern, which are kept fixed. Then, the chosen summary statistic is estimated for each of these permutations and global envelopes at a given nominal level are generated based on them. The result of the test can be assessed graphically: if the summary statistic estimate for the observed pattern exits the envelopes, we proceed with the assumption that the underlying FMPP is not randomly labelled. Furthermore, it is possible to calculate a p-value based on the position of the observed summary statistic within the qth envelopes, following Myllymäki et al. (2017). We know, however, that the conclusion drawn from the application of the above-mentioned global test pertains to the whole analysed process, indicating whether all the functional marks are randomly labelled or not. Motivated by the will to further detect the specific points, and regions, where the functional marks really do depend on the other marked points, we propose a *local test for random labelling*. The main idea is to run a global envelope test on each point of the analysed pattern by means of the previously proposed *local t-weighted marked inhomogeneous K-functions*, to draw different conclusions about the individual points, based on the obtained p-values. In Algorithm 1 we outline the proposed local test. Note that we alternatively may use sampling without replacement in step 5 of Algorithm 1.

Algorithm 1 Local test of random labelling

```
Set a fixed nominal value $\alpha$ for type I error;
Consider a (functional) marked point pattern $\mathbf{x} = \{(x_j, m_j)\}_{j=1}^{k}$, $k \geq 1$;
Set a number of simulations, $Q \geq 1$;
for each $q = 1, \dots, Q$: do
    Randomly sample $k$ (functional) marks, with replacement, from the original $k$ ones;
    Denote the resulting point pattern by $\mathbf{x}_q = \{(x_j, m_j^q)\}_{j=1}^{k}$;
end for
for each $j = 1, \dots, k$, do
    Compute $L_n^{(j,q)} = \{\hat{\mathcal{K}}_t^{(x_j, m_j^q) \times_{i=1}^{n-1} E_i}(r; \mathbf{x}_q)\}_{r \in [0, r_{max}]}$ for all $q = 1, \dots, Q$;
    Apply global envelope testing, using the functions $L_n^{(j,q)}$, $q = 1, \dots, Q$, to generate the envelopes;
    Obtain a $p$-value $p_j$ from the test;
    Reject the null hypothesis for the $j^{th}$ point if $p_j \leq \alpha$.
end for
```

6.1.5 Motivating example and simulation study

This section is dedicated to simulation studies to assess the performance of our proposed local test. First, section 6.1.5.1 provides a motivating example of the use of such a test, by means of simulated data resembling seismic events, which in turn have motivated this work. In particular, this means simulating the functional marks as seismic waveforms, following the typical abrupt change in variance of the signal in correspondence with the arrivals of the first P- and S-waves. Then, section 6.1.5.2 presents an extensive simulation study, showing diverse and more general settings. Specifically, we assess the performance of the test by summarising the results in terms of classification rates.

6.1.5.1 The need for a local test

We simulate a homogeneous spatial point pattern with 250 points on the unit square, $W = [0,1] \times [0,1]$, which represents the ground pattern. For each ground point x_i, we simulate a functional mark of the form

$$\begin{aligned} f_i(t) =& y(t) = \mu(t) + \epsilon(t), \qquad t \in \mathcal{T} = [0,1], \\ \epsilon(t) \sim& N(0, \sigma(t)^2), \\ \sigma(t)^2 =& 0.2 + 7.5\mathbf{1}\{t > 0.4\} - 5\mathbf{1}\{t > 0.6\}, \end{aligned}$$

where the mean signal $\mu(t)$ is taken to be zero.

The spatial ground point pattern and the corresponding waveform for a given point are shown in Figure 6.2(a)-(b). Since the marks/waveforms are simulated from the same model, and independently of each other and the spatial locations of the points, we see that such a process is indeed randomly labelled.

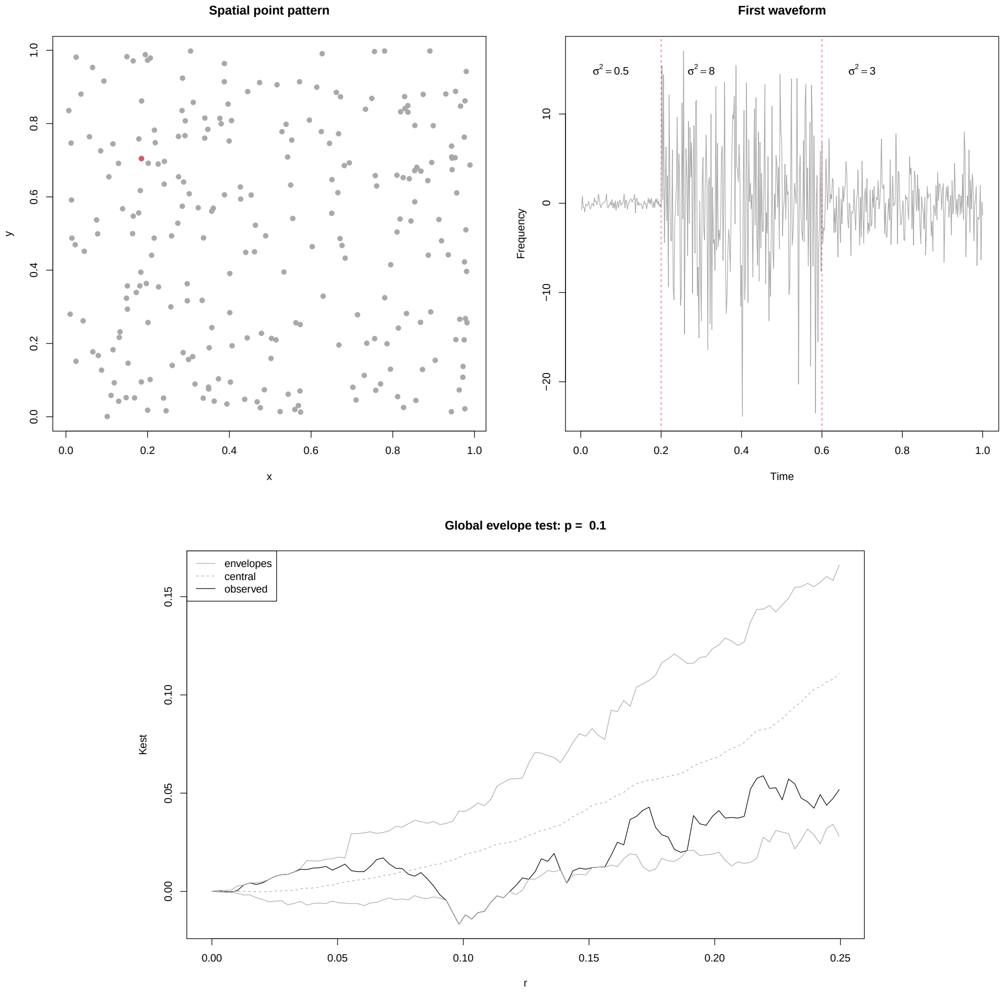


Figure 6.2: Simulated earthquake locations (a); Simulated waveform attached to the red point on the left (b); Result of the global test (c).

Having generated the data, we first run a *global envelope test for random labelling*, by randomly permuting the simulated waveforms, i.e. the functional marks, keeping the location of the points fixed. We run the test by means of the t-weighted marked n-th order inhomogeneous K-function of Ghorbani et al. (2021), with $n = 2$, making it a second-order summary statistic, and t given by the test function (6.10), i.e. the functional marked counterpart of the test function for the classical variogram. As previously mentioned, we assume that there is a common mark distribution which coincides with the reference measure on the mark space so that the intensity function is estimated by the ground process intensity estimate. To be as objective as possible, we do not use the homogeneous intensity estimator $\widehat{\rho}_g(\cdot) = Y_g(W)/|W|$ here but instead we use a kernel intensity estimator, as in practice it would be unknown to us whether the actual ground process is (in)homogeneous. We use a Gaussian kernel intensity estimator $\widehat{\rho}_g(\cdot)$, where we select the bandwidth, h, according to Cronie and Van Lieshout (2018). More specifically, we minimise the discrepancy between the area of the observation window and the sum of reciprocal estimated intensity values at the points of the point pattern, i.e. we minimise $CvL(h) = (|W| - \sum_i 1/\hat{\rho}_g(x_i; h))^2$, where the sum is taken over all the data points x_i and $\hat{\rho}_g(x_i; h)$ is the kernel intensity estimate with bandwidth h, evaluated in x_i. Then, once the bandwidth has been selected, the intensity estimate is corrected for edge effects through global edge correction (the option `diggle=FALSE` in the `spatstat` function `density.ppp`), i.e. dividing the estimate by the convolution of the Gaussian kernel with the window of observation (Diggle, 1985). Finally, for w we use Ripley's isotropic edge correction in the summary statistic to correct for edge effects. We repeated the procedure 39 times, obtaining the result depicted in Figure 6.2(c). We stress that our approach seems to be robust with respect to the bandwidth specification, i.e. the choice of bandwidth selection approach plays a minor role in the final result.

As evident from Figure 6.2(c), the observed summary statistic completely lies within the envelopes, and this confirms the expected result of lack of spatial dependence/structure of the functional marks. This result is further corroborated by the non-significant p-value, equal to 0.1.

6.1.5.1.1 Simulating spatially dependent functional marks To make the functional marks spatially dependent, we then superimpose a homogeneous spatial point pattern with 50 points, generated in the $[0, 0.5] \times [0, 0.5]$ square, i.e. the bottom left region of the entire study region W. For these additional points, we generate different functional marks than before, namely with the underlying trend $\mu(t) = 10 + 6\sin(3\pi z_t)$. Consequently, we have simulated an FMPP with spatially varying functional marks, i.e. not random labelled. We, therefore, expect a global test of random labelling to confirm this.

We first run the same global test of random labelling as before. We use $Q = 39$ and obtain a global p-value of 0.025. This, together with the observed K-function lying outside the envelopes (Figure 6.3(a)), indicates the ability of the global test to correctly detect the spatial dependence of the functional marks.

We know, however, that this conclusion should not be drawn for each point of the pattern if we consider local restrictions of it, but specifically for those in the vicinity of the $[0, 0.5] \times [0, 0.5]$ square. We, therefore, proceed by running our proposed local test, based on the proposed second-order local K-function $\hat{\mathcal{K}}_{t,2}^{(x,m)}(r)$, $r \in [0, r_{max}]$, in Definition (6.2), with the same choice of test function $t(\cdot)$ and the

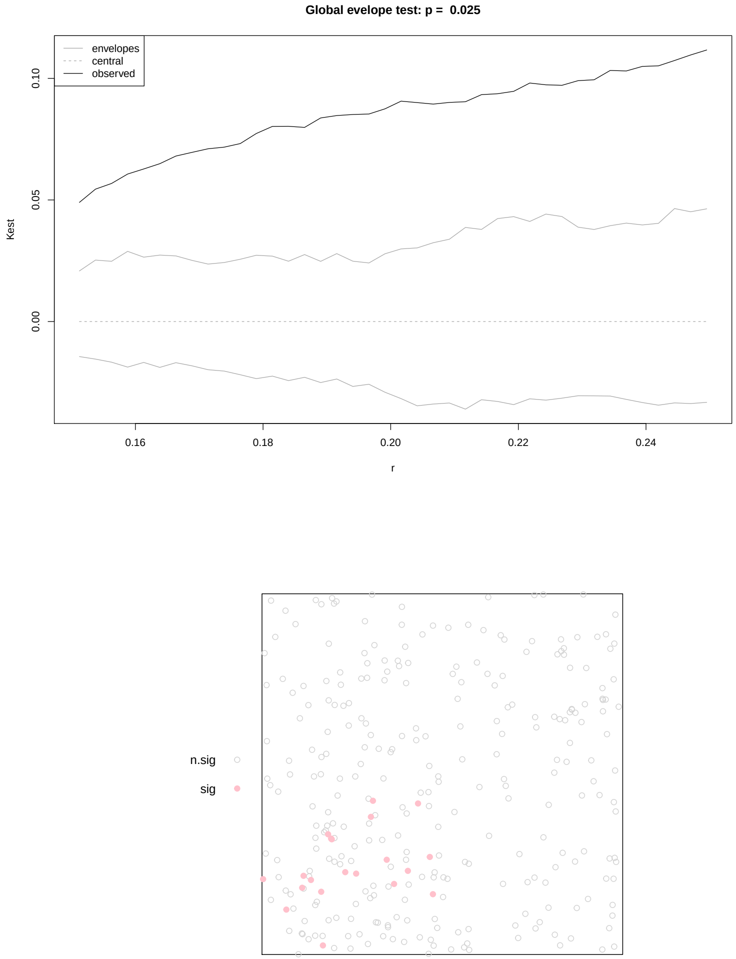


Figure 6.3: Result of global test for the spatially dependent simulated data (a); Output of the local test: significant points for which the hypothesis of random labelling is rejected are in pink (b).

same intensity estimation scheme as for the global one. Figure 6.3(b) depicts the points of the simulated point pattern, and it displays in pink those points for which the local test came out significant. Hence, this illustrates that the proposed local test is able to correctly identify some of the points and, consequently some parts of the region where the hypothesis of random labelling does not hold locally. Note that a universally preferable option for r_{max} does not exist. In this work, it is set to

$\min(x_W, y_W)/4$, where x_W and y_W represent the maximum width and height of the observation region W, respectively. Note that this rule of thumb is supported by Diggle (2013). Indeed, changing the value of r_{max} has an impact on the final results, and we found that our choice provided the best compromise among the options.

6.1.5.2 Extended simulation study

This section aims to study the proposed method's performance in terms of classification rates considering different scenarios concerning both the ground processes and the functional marks' structures. To this end, we simulate different such scenarios to obtain a comprehensive understanding of the results of the local test in different settings.

In detail, we consider three types of ground process structures, all with an expected point count of 200: (1) a homogeneous Poisson process; (2) an inhomogeneous Poisson process with intensity function $\rho_g(x) = \rho_g(x_1, x_2) = \exp(3.5 + 3x_2)$, $x \in W$; (3) a Thomas process, with the intensity of the Poisson process of cluster centres equal to 25, the standard deviation of random displacement of a point from its cluster centre equal to 0.05, and mean the number of points per cluster equal to 7. They are all generated in W, i.e. the unit square, and will be referred to as the *base patterns*. Then, we superimpose additional simulated patterns in the $[0, 0.5] \times [0, 0.5]$ square, coming from the same generating processes but with an expected number of points of 50. Hereby the expected total number of points on $[0, 0.5] \times [0, 0.5]$ is $50 + 200/4 = 100$, and on its complement, it is 150. These additional patterns will be referred to as *feature patterns*.

As for the functional marks, we consider the time domain $\mathcal{T} = [0, 10]$ and, practically, we sample each simulated mark function in 100 equally spaced time points in $\mathcal{T}$. We assume that each functional mark satisfies $f_i(t) = Z(x_i, t)$, where x_i is the ith ground point and

$$Z(x, t) = \mu + \xi(x, t), \quad (x, t) \in W \times \mathcal{T}, \tag{6.11}$$

for a zero-mean stationary Gaussian random field ξ with covariance function $C(h, u)$. Here h and u denote the spatial and the temporal lags, respectively. For the base patterns, we consider $\mu = 5$ and a pure nugget effect model with covariance function $C(h, u) = \sigma^2 \mathbf{1}\{h = 0\}$, $\sigma^2 = 0.01$. In other words, each f_i is random noise with mean 5 and variance 0.01 and all f_i's are iid. See the grey curves in the bottom panel of Figure 6.4. For the feature patterns, we consider three different marking models:

1. Shifted base model: We here let ξ have the same form as in the base model but let $\mu = 5.5$.
2. Decreased variance base model: We here let ξ have the same form as in the base model but let $\sigma^2 = 0.001$.

3. Non-separable space-time model: We here let $\mu = 5$ and consider a space isotropic covariance function given by $C(h,u) = (\psi(u)+1)^{-\delta/2}\phi(h/\sqrt{\psi(u)+1})$. Here, ϕ is a normal mixture and the corresponding covariance function only depends on the distance between two points, while ψ is a variogram model, which we choose according to a fractal Brownian motion with fractal dimension $\alpha = 1$. This is an intrinsically stationary isotropic variogram model.

We note that the first two of these scenarios represent independent but not identically distributed marks, whereas in the third scenario we additionally have that the marks are also dependent. A graphical representation of the scenarios comes in Figure 6.4. On the top panel, we display the three ground patterns. On the bottom ones, the corresponding functional marks are depicted.

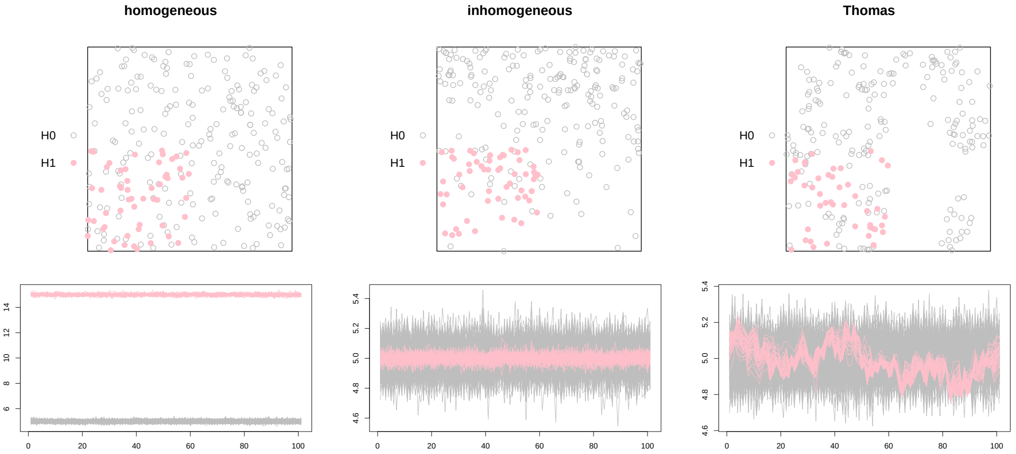


Figure 6.4: Simulation scenarios. Top panels: spatial ground patterns; Bottom panels: functional marks of model (6.11) *(in grey) and of the marking models in item* 1*,* 2*, and* 3*, from left to right (in pink).*

We show the results of the local test in terms of true-positive rate (TPR), false-positive rate (FPR), and accuracy (ACC), averaging over 100 simulated point patterns in Table 6.1. The rates are defined as

$$TPR = \frac{\text{true positives}}{\text{positives}}, \quad FPR = \frac{\text{false negatives}}{\text{negatives}}, \quad ACC = \frac{\text{true positives and negatives}}{\text{positives and negatives}}.$$

We of course wish to have TPR and ACC close to 1 and FPR close to 0.

As shown in Table 6.1, the performance of the local test in terms of classification rates strongly depends on the difference in the functional marks. Specifically, changing only the mean of the underlying random field is not enough to properly identify the points of the feature patterns.

This sufficiently improves when changing the variance only, but the best result is obtained when the whole model is changed, that is, changing the correlation structure. The effect of the type of ground pattern is less evident but still present. The inhomogeneous Poisson scenario reports the best classification rates, followed by the Thomas and homogeneous Poisson ones.

Finally, we found that the test function $t(\cdot)$ based on the L_2 distance in Equation (6.9) gave better results overall. To further explore how the choice of test function influences the test, we also compared it to a test function incorporating a derivative function accounting for the shape of the functional marks. This yielded similar results but turned out to be more computationally demanding.

Table 6.1: Results of the local test averaged over 100 simulated point patterns with an expected point count of 250 each.

Ground process	Marking model	TPR	FPR	ACC
Homogeneous Poisson	(1)	0.112	0.346	0.583
Homogeneous Poisson	(2)	0.583	0.066	0.820
Homogeneous Poisson	(3)	0.870	0.024	0.896
Inhomogeneous Poisson	(1)	0.032	0.585	0.449
Inhomogeneous Poisson	(2)	0.648	0.084	0.856
Inhomogeneous Poisson	(3)	0.895	0.023	0.932
Thomas	(1)	0.109	0.394	0.571
Thomas	(2)	0.637	0.088	0.846
Thomas	(3)	0.865	0.025	0.925

6.1.6 Real seismic data analysis

We analyse data coming from the *ISTANCE* dataset, presented in section 6.1.1. More specifically, we analyse a sample dataset provided at `http://www.pi.ingv.it/instance/`. The dataset contains 10000 records of 300 events, together with the associated metadata. Figure 6.5 shows the earthquake locations and time occurrences.

We first compute the proposed local K-function. The left panels of Figure 6.6 depict the estimated local summary statistics. In particular, the green lines represent the global statistics, while the grey ones represent the individual contributions. In blue we also represent the theoretical value. In the top-left panel, the K-function is based on a kernel intensity estimate whose bandwidth is selected by Diggle (2013)'s rule, while in the bottom-left panel the bandwidth is chosen as in Cronie and Van Lieshout (2018). We observe some relevant differences: while with Diggle (2013)'s rule we depict different local K-functions deviating from the global

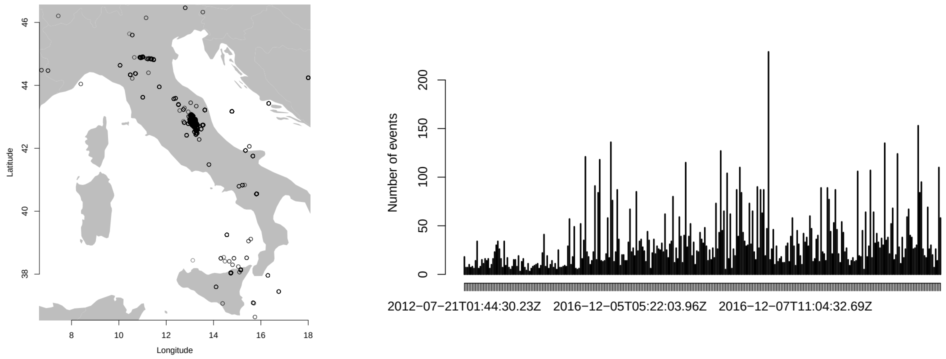


Figure 6.5: Earthquake locations and time occurrences

one, following Cronie and Van Lieshout (2018), we find a unique outlying local K-function. This may be explained by the fact that Cronie and Van Lieshout (2018)'s approach tends to yield a bit too large bandwidths when large parts of the study region contain no points, while Diggle (2013)'s approach tends to yield too small bandwidths in general. See Cronie and Van Lieshout (2018) for details. Note that by increasing the bandwidth we decrease the intensity estimate and, as a consequence, the summand denominators in (6.8) are decreased. Therefore, we run the proposed local test of random labelling with both options for the bandwidth selection and, as expected, the differences observed in the computation of the local K-functions are reflected in the results of the test.

Figure 6.6 displays the significant points (centre panels) and the non-significant ones (right panels). The top panels show the results with Diggle (2013)'s bandwidth while the bottom ones are obtained with Cronie and Van Lieshout (2018)'s bandwidth. For both choices, we selected a significance level of 0.1. We observe that the significant points tend to be similar in both cases, therefore the choice of bandwidth (selection method) does not seem to be crucial. We note that such bandwidth-induced differences were missing in the previously run simulation study. We attribute this sensitivity of the procedure to the shapes of the functional marks, which are obviously more variable if compared to the simulated ones.

Nevertheless, both bandwidths lead to significant events belonging to important well-known Italian seismic sequences. Of course, these sequences are likely generated by different underlying processes, giving rise to long-term and highly correlated aftershocks.

6.1.7 Conclusions

In this work, we have proposed a general form for local summary statistics for marked point processes, which has been exploited to define the family of local inhomogeneous

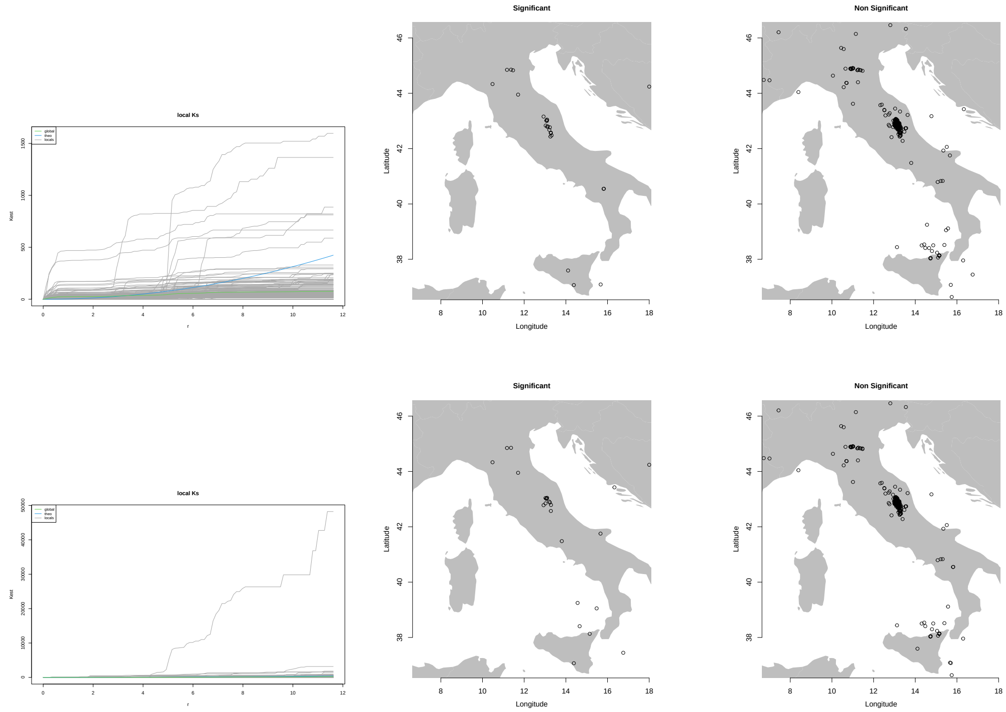


Figure 6.6: Left panels: Local K-functions. Center and right panels: Results of the local test at $\alpha = 0.1$. Top panels: The K-function is based on a kernel intensity estimate whose bandwidth is selected by Diggle (2013)'s rule. Bottom panels: The bandwidth is chosen as in Cronie and Van Lieshout (2018).

mark-weighted summary statistics for spatial point processes with functional marks, i.e. Functional Marked Point Processes. We have employed such local summary statistics to construct a local test for random labelling, that is, to identify points, as well as regions, where this hypothesis does not hold.

More specifically, we first introduce a general local function for marked point patterns. With this specification, we are able to show that this function may be exploited to generate most summary statistics established in the literature. With particular reference to the functional marked context, we define the family of local t-weighted marked n-th order inhomogeneous summary statistics based on the K-function, which is a local contribution to a global summary statistic estimator. We obtain a result for the expectation of the general local summary statistic and exploit it to derive an expression for the expectation of our t-weighted local statistics.

Having access to these tools, we have proposed a local test of random labelling, resorting to the second-order version of our proposed local estimator, obtaining a local test useful for identifying specific regions where a global test would not detect the atypical behaviour of the points.

To study the performance of the test in terms of classification rates, we have conducted a simulation study, considering a number of scenarios with different ground processes and structures for the functional marks. Such simulations have shown that in many settings, the local test performs well in identifying points of a pattern where the hypothesis of random labelling is not verified.

We can draw a number of future work paths. Nevertheless, the local functions proposed in this section can be considered as a very informative synthesis of the local second-order behaviour, useful for characterising the study area by an extended marked model, based on the FMPP theory. Incorporating local characteristics as functional marks would become part of the so-called *Constructed functional marks* (CFMs), which are marks reflecting the geometries of point configurations in neighbourhoods of the individual points.

Concerning the application to seismic data, we aim at including also auxiliary (non-functional) marks in the analysis. These could contain synthetic information about the waveforms, such as the arrival times of the seismic event, or the inter-time between the two. The achievement of the unification of earthquake data and the FMPP theory would result in building a framework where it would be possible to exploit the available information of the seismic point process altogether.

A final comment concerns the possible extension of this work's tools to spatio-temporal ground processes, which of course are of importance for processes which typically exhibit spatio-temporal interactions, such as the seismic one. Undoubtedly, such extension would be crucial for accounting for the temporal dimension of the seismic events, whose realization depends on their past history, as proved by the existence of aftershocks. From a methodological point of view, this would mean to consider a spatio-temporal marked point process $Y = \{(x_i, m_i)\}_{i=1}^{N}$, with ground points x_i in the 3-dimensional space $\mathbb{R}^2 \times \mathbb{R}^+$. Concerning the global summary statistics in equation (6.5), it would be sufficient to substitute the Euclidean ball by a cylinder with the same center and radius, and a temporal height. Furthermore, local summary statistics in space and time are well established, both theoretically (Siino et al., 2018b; Adelfio et al., 2020) and computationally (Gabriel et al., 2021). Therefore, it would only be necessary to employ the spatio-temporal LISTA functions in our proposed test. This is out of the scopes of this paper, but it surely represents an interesting path to cover in future.

6.2 A new picking algorithm based on the variance piecewise constant models

D'Angelo, N., Di Benedetto, A., Adelfio, G., D'Alessandro, A., and Chiodi, M. (2022e). A new picking algorithm based on the variance piecewise constant models. *Stochastic Environmental Research and Risk Assessment,* *https://doi.org/10.1007/s00477-022-02218-x*

Earthquakes may be generated by fracture processes in the Earth's crust, causing a partial release of the elastic strain energy stored by tectonic processes. The released energy is partially propagated away from its source as a wave field. There are three basic types of seismic waves - P-waves (also known as primary waves, travelling at the greatest velocity through the Earth), S-waves (transverse waves also known as secondary waves, slower than P-waves) and surface waves (similar in nature to water waves and travel just under the Earth's surface). P-waves and S-waves are sometimes collectively called body waves. The spatial sampling of the wave-field and recorded by a seismic network are the waveforms represented by seismograms. A correct registration and detection by the seismic station of the arrival of the first P-wave, as well as other relevant phases of the seismic event, is crucial for understanding the nature of the generating event (Adelfio et al., 2012). An earthquake monitoring network is a set of seismic stations (accelerometer and velocimeters) suitably distributed over the territory capable of detecting the occurrence of an earthquake. In addition to sensors capable of measuring the shaking generated by the earthquake, a seismic network includes data transmission and processing systems capable of determining in the shortest possible time the location of an earthquake (hypocenter) and its magnitude. When a seismic network is very efficient, i.e. able to automatically and quickly detect an earthquake, it can be used as an early warning tool. Both in the case of use for seismic monitoring and for early warning, the first step to be faced is the correct detection of the seismic event and the correct estimate of the arrival times of the main seismic phases. Given the great growth of seismic networks and the large amount of data that is collected during seismic sequences, the development of automatic picking algorithms capable of carrying out a precise and reliable identification of the arrival times of seismic phases has become increasingly important. These algorithms must be at the same robust but not computationally complex so that they can be executed in real-time and also used for early warning purposes. An accurate picking can allow precise hypocentral localization. Moreover, high-quality data can be used for tomographic reconstructions of the subsoil.

More in detail, as stated above, the first arrival times on seismograms coincide with the arrival of the first P-wave. The time of the phase-detection $\hat{T}_i$ at a station i is interpreted as the first P-phase arrival time, which is, of course, affected by an error ϵ_i. $\hat{T}_i$ may be written as $\hat{T}_i = T_0 + t_i + \epsilon_i$, where T_0 is the earthquake origin time and t_i is the travel time of a P-wave to station i. The coincidence trigger detects an event if for any combination of a minimum number of stations (typically three or four) the condition $|\hat{T}_i - \hat{T}_j| \leq \epsilon$ is met. ϵ is the maximum allowed difference between trigger times at neighbouring stations. This coincidence trigger works satisfactorily for local or regional networks, where the inter-distance among the seismic stations is not large. For global networks, this simple event detection algorithm has to be modified. Küperkoch et al. (2012) review the most widespread automatic picking algorithms. Comparative works among different pickers have been carried out in literature (Sleeman and Van Eck (1999); Aldersons (2004); Küperkoch et al. (2010)). Here we briefly outline the most known in the literature. Allen (1978, 1982) introduce the concept of characteristic function (CF), obtained by one or several non-linear transformations of the seismogram and should increase abruptly at the arrival time of a seismic wave. This allows both to estimate the arrival time from the CF and assess the quality estimation. Allen's picker is a fast and robust algorithm, which also accounts for automatic quality assessment. However, since this algorithm is just based on the amplitude information, it might miss emergent P-onsets. A comparative study by Küperkoch et al. (2010) shows that this algorithm tends to pick somewhat early compared to what an analyst would pick.

Another widely used picking algorithm is the one proposed by Baer and Kradolfer (1987). This algorithm is frequently applied, e.g. by 'Programmable Interactive Toolbox for Seismological Analysis' (PITSA, Scherbaum et al. (1999)) and the picking system MannekenPix (Aldersons (2004)). In contrast to Allen's squared envelope function, this CF is sensitive to changes in amplitude, frequency and phase.

The Baer and Kradolfer's picker is also very fast and robust and quite user-friendly, needing just four input parameters. A shortcoming of this algorithm is the missing automated quality assessment. Several comparative studies (Sleeman and Van Eck (1999); Aldersons (2004); Küperkoch et al. (2010)) show how this picking algorithm tends to be somewhat late compared to manual P-picks.

The statistical properties of the seismogram might be characterized by its distribution density function and by parameters like variance, skewness and kurtosis. The latter two are parameters of higher order statistics (HOS) and are defined by Hartung et al. (2014). Though just amplitude-based, higher-order statistics are quite sensitive to emergent P-onsets. In combination with a sophisticated picking algorithm (e.g. Küperkoch et al. (2010)), which exploits the entire information provided by the determined CF, it yields excellent results. If precisely tuned, the automated quality assessment proposed by Küperkoch et al. (2010) gives similar weights as the analysts. However, choosing the parameters for this sophisticated algorithm is quite difficult and needs a great experience.

Finally, the so called autoregressive-Akaike-Information-Criterion-piker (AR-AIC) proposed by Sleeman and Van Eck (1999) is based on the work by Akaike (1998, 1975), Morita (1984) and Takanami and Kitagawa (1988). It is a highly sophis-

ticated algorithm based on information theory. The algorithm is computationally quite expensive and hence much slower than the other reviewed pickers.
In this work, we advocate the usage of the algorithm proposed in Adelfio (2012) for the automated seismogram onset time determination. This considers the case of the changepoint detection procedure for changes in variation, assuming that the variance function can be described by a piecewise constant function with segments delimited by unknown change points. It is worth noticing that there exists a wide literature about changes in mean in a Gaussian model (Chernoff and Zacks, 1964; Gardner, 1969; Hawkins, 1992; Worsley, 1979), as well as the problem of variance change-point detection, mostly focusing on autoregressive time-series models (Wichern et al., 1976; Wang and Wang, 2006; Zhao et al., 2010).

In D'Angelo et al. (2020), a new automatic picking algorithm, based on the proposal of Adelfio (2012) and suitable for the implementation of an automatic seismic surveillance system, is proposed and tested on a set of 100 synthetic seismograms, showing that the model is always able to correctly detect the arrival of the first P-wave, as well as other relevant phases of the seismic event, such as the arrival of the first S-wave and the end of the seismic event. These simulated waveforms all presented the same true values of arrival times but different underlying noises.

In D'Angelo et al. (2021a) the performance of the proposed algorithm is tested on a set of simulated waveforms as generated by seismic events with different characteristics, such as the magnitude, and with different scenarios of detection, namely with different epicentral distances from the nearest seismic station that first recorded the event. This allows for assessing the performance of the algorithm with respect to the different characteristics of both the seismic event and the detection scenario, to identify the most suitable scenario for the application of our algorithm. Those preliminary experiments show that the algorithm performs well in identifying the arrival times of the first P- and S-waves. In particular, the arrival time of the first P-waves is detected more easily than the arrival time of the first S-waves. This is a relevant result because the arrival time of the first P-wave represents the beginning of the seismic event. Furthermore, it is noticed that the post-selection algorithm is not always able to correctly identify the relevant changepoints among the first estimated subset of possible values.

Following these results, in this section, we aim to present our proposed algorithm's methodology, suitable for the automatic identification of the two relevant phrases in a seismic waveform: the arrival times of the P- and S-waves. To assess the algorithm's performance in different scenarios, we simulate a new richer dataset of waveforms with different magnitudes and epicentral distances. Moreover, to show the advantages of our approach, we compare our results with that obtained, applying a standard Short Time Average over Long Time Average (STA/LTA) algorithm (Allen, 1978).

These two algorithms lead to similar results in terms of performance. However, the proposed algorithm is characterized by greater flexibility and automation capacity, as it does not require testing and optimization phases. This peculiarity makes it potentially very useful in earthquake routine analysis in the case of novel seismic networks, in particular in those areas where earthquake characteristics are unknown. Indeed, features like the window width, threshold and characteristic function may depend on the recording network and on the application. The proposed algorithm just requires setting the maximum number of potential changepoints, denoted as K^*. This may influence the computational time, as the larger is K^*, the more the time to estimate the corresponding changepoints, and most of all, the time to compare the set of the reduced models by the used *lars* procedure, for finding the best changepoints. Furthermore, our proposal provides automatic detection of the arrival time of the P- and S- waves, and therefore, no intervention is needed by the researcher to identify the arrivals. Finally, the proposed algorithm can be easily modified to allow the identification of further seismic phases, such as the end of the seismic event.

The structure of the section is as follows. Section 6.2.1 presents the new picking algorithm. Section 6.2.2 reports the testing of the algorithm on a dataset of simulated waveforms. An application to real data is presented in section 6.2.3. Section 6.2.4 contains the conclusions and future works.

6.2.1 Methodology: Variance piecewise constant models

This section proposes a new methodology for the automatic picking of arrival times based on the theory of the variance piecewise constant models.

Adelfio (2012) considers the case of changepoint detection procedure for changes in variation, assuming that the variance function can be described by a piecewise constant function with segments delimited by unknown change points.

Let y_i be the outcome and x_i be the observed sample, for $i = 1, 2, \ldots, n$ occasions. Let us assume that $y_i = \mu_i + \epsilon_i$, where μ_i is for instance a sinusoidal function representing the observed signal and $\epsilon_i \sim N(0, \sigma_i^2)$ is an error term. In this context, σ_i^2 is a variance function approximated by a piecewise constant regression function with $K_0 + 1$ segments. An example is shown in Figure 6.7.

For simplicity, the model for changes in variance after the k^*th observation is

$$y_i = \begin{cases} \mu_i + \lambda\epsilon_i & 1 \leq i \leq k^* \\ \mu_i + \tilde{\lambda}\epsilon_i & k^* \leq i \leq n \end{cases}$$

with λ, $\tilde{\lambda}$, and k^* unknown and

$$\begin{cases} H_0: & \lambda = \tilde{\lambda} \\ H_1: & \lambda \neq \tilde{\lambda} \end{cases}$$

Taking advantage of a generalized linear model formulation of the investigated problem, the test for stepwise changes in the variance of a sequence of Gaussian random

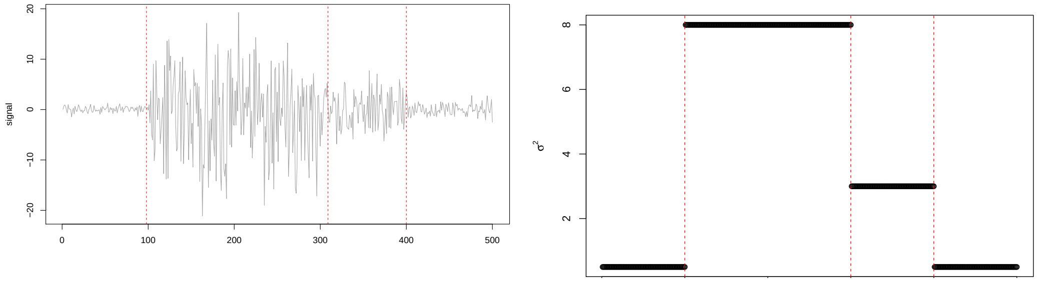


Figure 6.7: An example of simulated signal and its corresponding variance with jump points. The red dashed lines indicate the true changepoints.

variables may be transformed equivalently to the case of testing for changes in the mean of the squared residuals from an estimated linear model that accounts for the mean behaviour of the observed signal. The estimation of the mean signal $\hat{\mu}$ can be carried out by using a standard smoothing procedure, e.g., fitting a cubic smoothing spline to the data. Following a suggestion in Smyth et al. (2001), a gamma generalized linear model (GLM) is fitted with a log-link function, with the response given by the squared studentized residuals $s_i = (y_i - \hat{y}_i)^2/w_i$, with $\hat{y} = \hat{\mu}$ and weights $w_i = 1 - h_i$, where h_i is the ith diagonal element of the hat matrix H. According to this approach, testing H_0 against H_1 means that we are looking for a change in the mean of the residuals from a fitted linear model.

The proposed approach can be considered as a wider version of the *cumSeg* models proposed in Muggeo and Adelfio (2011) for independent normally distributed observations with constant variance and piecewise constant means to detect multiple changepoints in the mean of the gene expression levels in genomic sequences by the least-squares approach. The authors assume that the datum $y_i, \forall i$ is defined as the sum of the signal μ_i and noise $\epsilon_i \sim N(0, \sigma_i^2)$ and that μ_i is approximated by a piecewise constant regression function with $K_0 + 1$ segments, that is:

$$y_i = \beta_1 + \delta_1 \mathbf{1}(x_i > \psi_1) + \cdots + \delta_{K_0} \mathbf{1}(x_i > \psi_{K_0}) + \epsilon_i.$$

Here, $\mathbf{1}(\cdot)$ is the indicator function, such that $\mathbf{1}(x) = 1$ is x is true, $\boldsymbol{\psi}$ represents the K_0 locations of the changes on the observed phenomenon, β_1 is the mean level for $x_i < \psi_1$, and $\boldsymbol{\delta}$ is the vector of the differences in the mean levels at the change points. The authors proceed to take the cumulative sums of the jump-points model to get a convenient modelling expression that faces the discontinuities at the changepoints ψ_k assuming a piecewise linear or segmented relationship. Therefore, looking for changes in variance, the model is specified as

$$g(\theta_i) = \beta_1 x_i + \delta_1 (x_i - \psi_1)_+ + \cdots + \delta_{K_0} (x_i - \psi_{K_0})_+ \tag{6.12}$$

where the term $(x_i - \psi_k)_+$ for the changepoint k is defined as $(x_i - \psi_k)\mathbf{1}(x_i > \psi_k)$, and $\theta_i = \mathbb{E}[\sum_j^i s_j]$. This model specification has the advantage of an efficient estimating approach via the algorithm discussed in Muggeo (2003, 2008), fitting iteratively the generalized linear model:

$$g(\theta_i) = \beta_1 x_i + \sum_k \delta_k \tilde{U}_{ik} + \sum_k \gamma_k \tilde{V}_{ik}^-, \tag{6.13}$$

where $\tilde{U}_{ik} = (x_i - \tilde{\psi}_k)_+$, $\tilde{V}_{ik}^- = -\mathbf{1}(x_i > \tilde{\psi}_k)$. The parameters β_1 and $\boldsymbol{\delta}$ are the same of Equation (6.12), while the γ are the working coefficients useful for the estimation procedure Muggeo (2003). At each step, the working model in Equation (6.13) is fitted and new estimates of the changepoints are obtained via

$$\hat{\psi}_k = \tilde{\psi}_k + \frac{\hat{\gamma}_k}{\hat{\delta}_k}$$

iterating the process up to convergence. $K^*(< K)$ values are returned, producing the fitted model

$$g(\hat{\theta}_i^*) = \hat{\beta}_1 + \hat{\delta} V_{i1} + \cdots + \hat{\delta}_{K^*} V_{iK^*},$$

where $V_{ik} = \mathbf{1}(x_i > \hat{\psi}_k)$ for $k = 1, 2, \ldots, K^*$ and the squared residuals are modelled as the response of a gamma GLM with the logarithmic link function. Selecting the number of significant changepoints means selecting the significant variables among $V_1, \ldots, V_k$, where K^* is the number of estimated changepoints from model (6.12). The author solves the model selection problem by using the *lars* algorithm by Efron et al. (2004). Thus, the optimal fitted model with $\hat{K}^* < K^*$ changepoints, is selected by the generalized Bayesian Information Criterion (BIC_{C_n}), that is:

$$BIC_{C_n} = -2 \log L + edf \log(n) C_n$$

where L is the likelihood function, *edf* is the actual model dimension quantified by the number of estimated parameters (including the intercept, the δ and ψ vectors), and C_n is a known constant. The vector of the corresponding selected changepoints is denoted by $\hat{\boldsymbol{\psi}}^*$.

The first issue concerns the value of C_n to be used in the BIC_{C_n} criterion to select the changepoints. In D'Angelo et al. (2020), by simulation, the performance of different specifications of C_n is assessed and, among the different examined specifications of C_n, simulations reveal that $C_n = \log \log n$ has the best performance. Thus, we use this value for the provided analysis.

6.2.1.1 The proposed algorithm: Changepost

Based on the above methodology, we propose a further algorithm (denoted as *changepost*) to detect, among the estimated changepoints, the two corresponding to the arrival of the first P-wave, and the arrival of the first S-wave. Formally, we define the relevant changepoints to be identified as the true arrival times of the first P- and S-waves, denoted by ψ_1 and ψ_2, respectively (i.e. $K_0 = 2$). In particular, we compare the ratio between the variances of the subsequent phases identified by the $\hat{K}^*$ changepoints $\hat{\boldsymbol{\psi}}^*$ estimated by the main algorithm. The two relevant changepoints are selected as the two ones in correspondence to the two biggest variance ratios. The pseudo-code comes in Algorithm 1.

Algorithm 2 *changepost*

Input: $\hat{\boldsymbol{\psi}}^* = \{\hat{\psi}_1^*, \dots, \hat{\psi}_{\hat{K}^*}^*\}$; $s_i = (y_i - \hat{y}_i)^2/w_i$; $S = \{s_1, \dots, s_n\}$; $start = 1$; $end = n$

Output: $\hat{\boldsymbol{\psi}}$

```
 1: for ($i$  $in$  $1 : \hat{K}^*$) do
 2:     if ($i == 1$) then
 3:         $ratio[i] \leftarrow \frac{\texttt{var}(S[start:\hat{\psi}_i^*])}{\texttt{var}(S[start:\hat{\psi}_{i+1}^*])}$
 4:     end if
 5:     if ($1 < i < \hat{K}^*$) then
 6:         $ratio[i] \leftarrow \frac{\texttt{var}(S[\hat{\psi}_{i-1}^*:\hat{\psi}_i^*])}{\texttt{var}(S[\hat{\psi}_{i-1}^*:\hat{\psi}_{i+1}^*])}$
 7:     end if
 8:     if ($i == \hat{K}^*$) then
 9:         $ratio[i] \leftarrow \frac{\texttt{var}(S[\hat{\psi}_{i-1}^*:\hat{\psi}_i^*])}{\texttt{var}(S[\hat{\psi}_{i-1}^*:end])}$
10:     end if
11: end for
12: $ratio[ratio < 1] \leftarrow (ratio[ratio < 1])^{-1}$
13: $\hat{\boldsymbol{\psi}} \leftarrow \hat{\boldsymbol{\psi}}^*[\texttt{which}(\texttt{rank}(ratio) > \hat{K}^* - 2)]$
```

As shown in steps (4), (7), and (10), for each estimated changepoint $\hat{\psi}_i^*$, we compute the ratio between the variance of the interval delimited by the $\hat{\psi}_{i-1}^*$ and $\hat{\psi}_i^*$, and the variance of the interval between $\hat{\psi}_{i-1}^*$ and $\hat{\psi}_{i+1}^*$. Then, in step (14), the two highest ratios suggest which are the corresponding changepoints $\hat{\psi}_1$ and $\hat{\psi}_2$ leading to the most relevant changes in variance.

Figure 6.8 depicts the changepoints identified by *changepost*, on the simulated data of Figure 6.7. As the variance of the real signal is $\sigma_i^2 = 0.5 + 8\mathbf{1}(i > .2) + 3\mathbf{1}(i > .6) + \mathbf{1}(i > .8)$, it is evident that *changepost* correctly identifies the changepoints corresponding to the most abrupt changes in the variance.

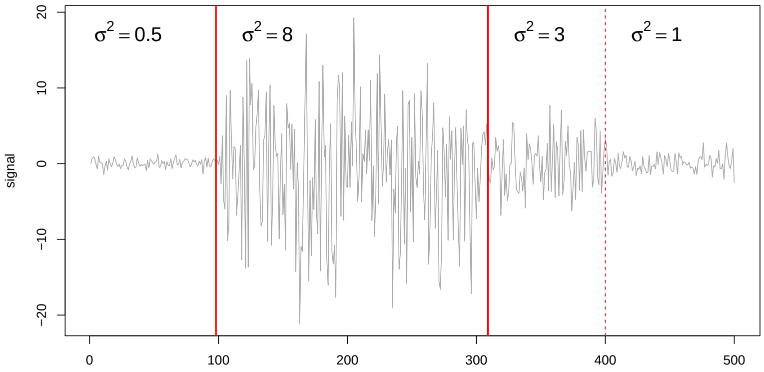


Figure 6.8: Simulated data of Figure 6.7. The red dashed lines indicate the change-points identified by the main algorithm, described in section 6.2.1. The red straight lines indicate the ones further identified by changepost.

6.2.2 Simulations: Evaluating the performance of the algorithm *changepost*

In this section, the proposed picking algorithm for the automatic seismogram onset time determination (simply denoted as *changepost*) is tested on a dataset of simulated waveforms. Simulated seismograms are used to have the maximum control over the arrival times of the P- and S- phases on the waveforms. This aspect is of fundamental importance for the correct validation of the algorithm, impossible with experimental seismograms. Indeed, when using real data, i.e. experimental seismograms, it is not possible to know with certainty the arrival times of the seismic phases. Experimental seismograms are recordings of ground motion, or seismic waves, generated at several kilometres in depth and distance. The identification of the arrival times of the seismic phases on experimental seismograms, or the best picking of the P and S waves, is carried out manually by expert seismologists. However, even an expert seismologist may introduce errors and uncertainties in the picking phase. Thus, for controlling the precision and accuracy of an automatic picking procedure, the best practice is to start from simulated data, with well-known seismic phases arrival times.

We aim at capturing the performance variations due to some characteristics of both the seismic event and its detection, which in turn affect some characteristics of the waveforms. Therefore, seismic events with different magnitudes are simulated, assuming different distances from the nearest seismic station.

Our tests allow highlighting the most general scenarios for the algorithm. Waveforms generated by earthquakes of small magnitude often have energy comparable to the background noise and allow for validation of the functioning of the algorithm in case of a low signal-to-noise ratio. Waveforms with different magnitudes and epicentral distances can also differ greatly in terms of frequency content. Events with small in magnitude and with small epicentral distances are generally rich in other frequencies. On the contrary, events with high magnitude and great epicentral distance are also rich in low frequencies. Variations in the epicentral distance also affect the nature of the seismic phases P and S.

The seismic phases, and more generally the shape of the seismograms, depend on the epicentral distance. At very small epicentral distances (from a few kilometres to a few tens of kilometres) the seismic waves travel inside the upper crust. The seismic phases coming first to the surface are not undergone refraction and reflection phenomena. They can be considered direct waves. At greater distances instead (typically over a hundred kilometres), the seismic phases first emerging on the surface are refracted critically from the upper mantle (Mohorovich discontinuity). Therefore, using three epicentral distances (10, 250 and 50 kilometres), simulations involve the recording of three different types of earthquakes, corresponding respectively to local events, whose first arrival seismic phases are direct waves, regional events, in which the first seismic phases are seismic waves critically refracted by the upper mantle, and transitional events.

6.2.2.1 The simulation setup

The waveforms are simulated as coming from seismic events with different characteristics, referring to altogether 12 scenarios, one for each combination of the following:

- 3 distances from the nearest station that recorded the seismic event: 10, 50 and 250 km;
- 4 magnitudes: 2, 3, 4 and 5.

For each scenario, 100 waveforms are simulated, all assumed to have impulsive onset of P- and S- waves and standard seismic noise. Moreover, the synthetic signals are generated with a sampling rate of 200 samples per second.

The seismic waveforms are simulated using the deterministic hybrid approach proposed by Mourhatch and Krishnan (2020). In detail, the low-frequency content (limited to a frequency of 0.5 Hz) of the ground motion is generated from a kinematic source model using the open-source seismic wave-propagation package SPECFEM3D (Komatitsch and Tromp, 1999; Komatitsch et al., 2004), that implements the spectral-element method, incorporating the regional 3-D wave-speed structure of the earth. Following Mourhatch and Krishnan (2020), low-frequency synthetic SPECFEM3D seismograms are combined with high-frequency seismograms generated using a variant of the classical EGF (Empirical Green's Function) approach.

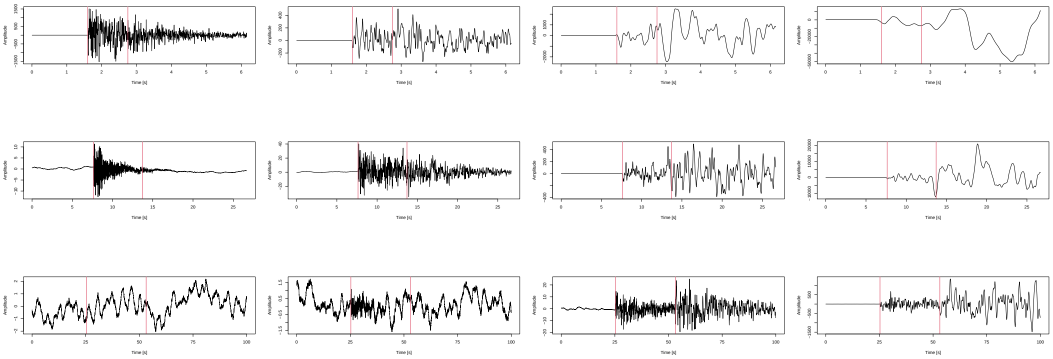


Figure 6.9: A simulated waveform for each scenario and true arrival times. From left to right: increasing magnitude levels. From top to bottom: increasing distance from the nearest seismic station.

For each seismic event, we generate all three components of motion (i.e. North-South, East-West and Vertical). In our analysis, we only report the results for the Vertical component, for the sake of brevity. In Figure 6.9, an example of waveforms (vertical movement component) for each of the considered scenarios with highlighted the true P- and S arrival times is shown.

6.2.2.2 Simulation results for *changepost*

Table 6.2 reports the empirical means (m) and Mean Squared Error values (s) of the two relevant changepoints estimated by the proposed algorithm over the 100 waveforms coming synthetic seismic events, for the four different Magnitudes and three epicentral distances. For each epicentral distance, we assume different true arrival times (in blue). Along with the mean and the mean squared error values, we also compute the percentage of waveforms where no changepoint is estimated.

Overall we may notice that, as expected, the *changepost* algorithm performs the best as the distance from the nearest seismic station that recorded the event decreases and as the magnitude of the seismic event increases. This is the case with the best signal-to-noise ratio. Indeed, the NA values are most likely to occur when the magnitude is small and the distance is large, that is basically when the P- and S-waves have comparable or lower energy with respect to the background noise, which is indiscernible from that. In such cases, the arrival times can not be estimated correctly. The scenario in which the distance from the nearest seismic stations is 50 km is the one reporting no NAs, regardless of the magnitude level. Nevertheless, this does not represent the best picking scenario, as the uncertainty of the estimates is larger than the performance in the 10 km scenario.

Table 6.2: Empirical means (m) and Mean Squared Error values (s) of the two relevant changepoints detected by changepost, over the 100 waveforms of each simulated dataset, with four different magnitudes and three epicentral distances.

M		ψ_1	ψ_2	NA%
		10 km		
true		41.6	42.75	
2	m	41.415	43.101	7
	s	0.107	2.256	
3	m	41.514	42.708	1
	s	0.017	1.604	
4	m	41.496	42.278	2
	s	0.012	1.355	
5	m	41.382	45.751	1
	s	0.055	1.081	
		50 km		
true		47.63	53.714	
2	m	47.592	53.394	0
	s	0.002	5.168	
3	m	47.453	57.686	0
	s	0.240	49.011	
4	m	47.229	53.194	0
	s	0.789	19.630	
5	m	47.118	55.537	0
	s	0.376	45.036	
		250 km		
true		75.26	103.15	
2	m	-	-	100
	s	-	-	
3	m	-	-	100
	s	-	-	
4	m	75.115	105.626	0
	s	7.904	80.144	
5	m	74.192	103.592	1
	s	1.322	434.865	

Table 6.3: STA/LTA settings

STA/LTA Parameters (s)	10 km	50 km	250 km
STA window length	0.1	0.5	2.5
LTA window length	1	5	25
Threshold Trigger on	5.0	5.0	5.0
Threshold Trigger off	2.5	2.5	2.5

6.2.2.3 Comparison with STA/LTA

In this paragraph, we compare the *changepost* picking algorithm, based on the variance piecewise constant models, introduced in section 6.2.1, with the Short-Term Average/Long-Term Average (STA/LTA) method (Allen, 1978).

The STA/LTA method is the simplest and most commonly picking technique used in earthquake seismology. The STA/LTA method computes the ratio of the continuously computed average energy (generally the waveforms envelope, the absolute amplitude, or other characteristic functions) of a recorded trace in two synchronous moving-time windows: a Short-Term window and a Long-Term window (STA/LTA ratio). The short-time window permits to highlight of sudden amplitude changes in the signal, while the long-time one estimates the current average of the seismic noise. Therefore, the STA/LTA ratio allows for highlighting variations in energy in the signal with respect to the background noise. These energy variations can be identified by setting thresholds: when the STA/LTA ratio exceeds a certain threshold, the arrival of a seismic phase is identified. The output of the STA/LTA algorithm is the characteristic function E_k, defined as:

$$E_k = x_k^2 + (x_k')^2 + C$$

where x_k is the seismic trace, x_k' is its derivative and C is an empirical weighting constant.

This method is undoubtedly computationally efficient, and its variants are widely used for the picking of seismic phases. However, it needs a calibration phase to identify both the best length of the STA and the LTA and the best threshold level. The optimal STA width depends on the frequency content of the seismic event and, therefore, on its magnitude and epicentral distance. Similarly, the width of the LTA should also be chosen according to the noise characteristics. The trigger threshold is also very important: values that are too high can lead to the failure to identify the arrival of the seismic phases, and values that are too low can provide false identifications. This method can therefore be inaccurate in the case of a low signal-to-noise ratio.

After several optimization tests for each earthquake class, we set the parameters reported in Table 6.3 for comparison. Once the parameters are set, we run the tests,

Table 6.4: Empirical means (m) and Mean Squared Error values (s) of the two relevant arrival times estimated by the STA/LTA algorithm over the 100 waveforms of each simulated dataset, with four different magnitudes and three distances.

M		ψ_1	ψ_2	NA%
		10 km		
true		41.6	42.75	
2	m	41.605	-	0
	s	0.000	-	
3	m	41.589	42.272	0
	s	0.003	0.413	
4	m	41.573	42.853	0
	s	0.000	0.140	
5	m	41.514	42.740	0
	s	0.009	0.291	
		50 km		
true		47.63	53.714	
2	m	47.682	49.148	1
	s	0.018	35.847	
3	m	47.620	47.642	0
	s	0.034	36.873	
4	m	47.617	53.632	0
	s	0.006	0.430	
5	m	47.589	53.472	0
	s	0.006	2.193	
		250 km		
true		75.26	103.15	
2	m	78.275	129.137	82
	s	9.145	989.601	
3	m	82.935	129.166	90
	s	58.905	922.363	
4	m	75.809	106.600	0
	s	0.655	13.357	
5	m	75.351	107.340	0
	s	0.033	21.231	

obtaining the results reported in Table 6.4. We have noticed that the number of the picked arrival times is generally lower using the STA/LTA, with respect to the *changepost* algorithm. Then, a further note is needed. Before computing Mean, MSE and NA%, we set the picking's STA/LTA results in this way: for each seismic event, we check if there are zero picking, one or more than one picking. If zero picking is observed, we only increment the NA%. Instead, if we find one or more than one picking, the closest picking to the real picking is determined. Just after finding all picking values for all events, we compute Mean and MSE for picked events, and NA%, as for the proposed algorithm. This specification is due since there are a number of cases where the STA/LTA algorithm picks a unique arrival time (see the first scenario – S-waves – in Table 6.4).

Nevertheless, if the number of the estimated picked arrival times $\hat{K}^*$ is large, the probability of having a picking close to the true one increases, resulting in smaller mean squared error values and then, influencing the results as mentioned above. Therefore, we define a probability index, computed for each waveform q: let $\hat{\boldsymbol{\psi}}^*_q$ be the vector of the $\hat{K}^*_q > 1$ estimated changepoints, the probability index is defined as follows:

$$\frac{\mathbf{1}(\hat{\boldsymbol{\psi}}^*{}_q \leq true \pm pick)}{\hat{K}^*_q}, \tag{6.14}$$

where $\mathbf{1}(\cdot)$ is the indicator function, such that $\mathbf{1}(x) = 1$ if x is true, i.e. counting how many times the changepoints $\hat{\boldsymbol{\psi}}^*_q$ estimated for the waveform q fall inside the interval $\pm pick$ around the true arrival time (which varies with the scenario considered).

In Tables 6.5 and 6.6 we report the computed probability index (6.14), for the proposed picking algorithm and for the STA/LTA algorithm, respectively. In particular, we set the pick to 0.2 seconds, and compute (6.14) for both the P- and S-arrivals, separately. Then, those values are averaged with respect to the 100 waveforms of each scenario, and the percentage of NAs is reported, to take into account both the waveforms where no changepoint is estimated (i.e. NAs in Tables 6.2 and 6.4) and those cases in which $\mathbf{1}(\hat{K}^*_q \leq true \pm pick) = 0$, i.e. no estimated changepoint for that specific waveform falls into the interval.

From Tables 6.5 and 6.6, we may notice that STA/LTA outperforms *changepost* in picking the P- Phases times. Almost in all the S- Phases, instead, STA/LTA provides the highest NA%: the scenario with a distance of 10 km and magnitude 2 is the only one where *changepost* provides 100% of NAs. Moreover, for the scenario 250km distance and magnitude 2, *changepost* does not find any changepoint and then provides the highest NA%. This percentage gradually decreases as the magnitude increases.

When the distance is 50km, even in lower magnitudes, the *changepost* algorithm provides a lower percentage of NA than STA/LTA. Overall, *changepost* outperforms STA/LTA in the S- Phases picking. Otherwise, the STA/LTA picking time is better for the P- Phases, being more precise in terms of tenths of a second.

Table 6.5: Percentage of the changepoints estimated by the changepost algorithm, lying within a ±0.2 *interval around the true value.*

	M	ψ_1	NA%	ψ_2	NA%
10 km	2	0.540	8	0.000	100
	3	0.740	1	0.020	96
	4	0.650	2	0.070	86
	5	0.200	69	0.115	77
50 km	2	0.500	0	0.020	96
	3	0.500	0	0.010	98
	4	0.300	43	0.020	96
	5	0.165	67	0.005	99
250 km	2	-	100	-	100
	3	-	100	-	100
	4	0.130	74	0.145	71
	5	0.010	98	0.010	98

Table 6.6: Percentage of the changepoints estimated by the STA/LTA algorithm, lying within a ±0.2 *interval around the true value.*

	M	ψ_1	NA%	ψ_2	NA%
10 km	2	1.000	0	0.000	100
	3	0.935	1	0.020	96
	4	0.433	0	0.253	38
	5	0.551	0	0.150	63
50 km	2	0.805	12	0.000	100
	3	0.935	5	0.000	100
	4	0.512	1	0.128	74
	5	0.464	0	0.112	73
250 km	2	0.000	100	0.000	100
	3	0.000	100	0.000	100
	4	0.170	81	0.000	100
	5	0.456	14	0.000	100

A comment on the computational cost is in order. Indeed, computation time is crucial in automated seismogram onset time determination, mostly accounting for its implications in seismic monitoring and in earthquake early warning systems. Even tough we have assessed that *changepost* is quite slower than STA/LTA, the computational time of the former does not represent a limitation. The only setting influencing the computational time is K^*, that is the maximum number of changepoints to be detected: the larger K^*, the higher the computational time, but also the more the estimated change points. Therefore, since *changepost* is able to process hundreds of waveforms within minutes, the researcher could even consider to reproduce the analyses with different values of K^*, depending on the available time and the complexity of the waveforms. Therefore, even tough STA/LTA has lower computational time, the automation of *changepost* counterbalances its higher computational cost.

6.2.3 Application to real data

The seismic events selected for showing an application to real data belong to the seismic sequence of L'Aquila (Italy) in 2009. The complete database is made up of 80 seismic events recorded by 12 stations with 3 components (identify the component of motion to which they refer: Up-Down, North-South and East-West), for a total of 2880 waveforms. They all exhibit a magnitude between 3 and 4.1. The waveforms were sampled at 100 Hz, and the length total of each of the waveforms is 100 seconds (10,000 samples). To increase the signal ratio noise, a bandpass filter was applied in the frequency band between 0.1 and 35 Hz. Such an operation was necessary to eliminate frequencies related to electronic and anthropic noise clearly not part of seismic signals. Also, the waveforms have been normalized with respect to the maximum amplitude.

Figures 6.10 and 6.11 contain five waveforms selected to show the results, and the arrival times identified by *changepost* and STA/LTA, respectively. For the proposed *changepost* procedure, K^* is set to 10. In Figure 6.10, the red dashed lines represent all the $\hat{K}^*$ change-points detected by the main algorithm, while the solid red ones identify the two selected change-points, representing the arrival times of the P- and S-waves, respectively. Figure 6.11 depicts the results of the application of STA/LTA: the solid red lines indicate the estimated arrival times. As evident from the two figures, while *changepost* is always able to identify two arrival times (most likely to represent the arrivals of the P- and S-waves), the STA/LTA algorithm either succeeds in identifying only very early arrival times (very likely to be arrival times of the P-waves) or completely fails to identify any arrival time.

This is an expected result for the seismic events considered in this experiment, in particular using the STA/LTA settings reported in Table 2. Better results for the STA/LTA algorithm, comparable with those just shown with synthetic seismograms, could be obtained only after a few rounds of optimization of the triggering

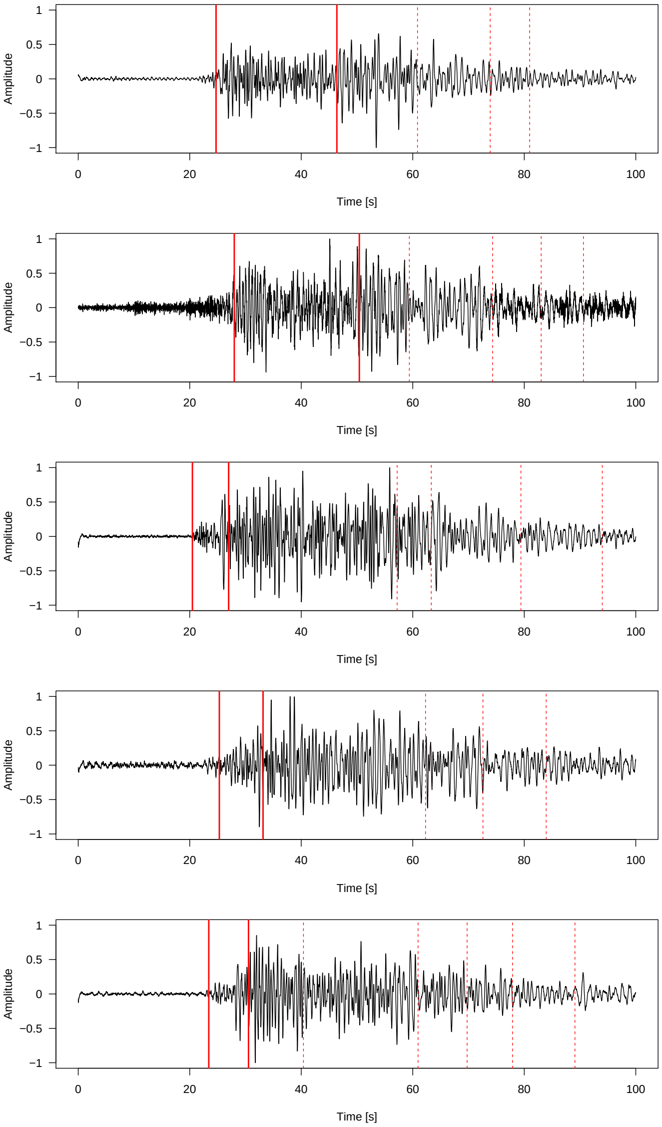


Figure 6.10: Vertical components of three seismic events detected by the seismic station BSSO - Busso (Italy) during the L'Aquila seismic sequence occurred in 2009. Red dashed lines: all the $\hat{K}^*$ change-points estimated by the main algorithm. Red straight lines: the $\hat{K} = 2$ change-points, among the $\hat{K}^*$ estimated ones, most likely to represent the true arrival times of the P- and S-waves.

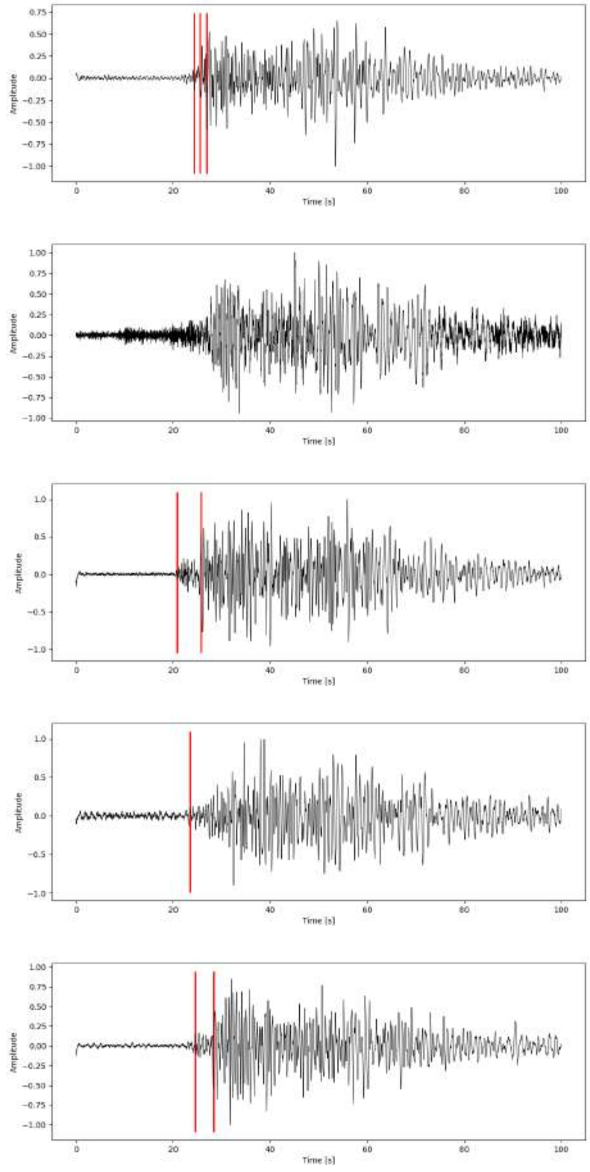


Figure 6.11: Vertical components of three seismic events detected by the seismic station BSSO - Busso (Italy) during the L'Aquila seismic sequence occurred in 2009. Red straight lines: the arrival times identified by STA/LTA.

parameters. These results confirm the conclusion drawn by the simulation study, that is the high flexibility of the proposed *changepost* algorithm. Indeed, it does not require either testing or optimization phase, and its accuracy is almost independent of the analyzed dataset.

6.2.4 Conclusions

The precise and quick determination of the arrival times of the main seismic phases is of fundamental importance for seismic surveillance and routine earthquake hypocenter determination. Clearly, to be suitable also for early warning, a picking algorithm must be computationally efficient, avoid false alarms, and time picking must be as accurate as possible.

With these premises, in this work, we proposed a novel picking algorithm for the automatic P- and S- waves onset time determination. The *changepost* algorithm is based on the variance piecewise constant models. The effectiveness and robustness of our picking algorithm were tested on synthetic seismograms. In order to make the STA/LTA algorithm work correctly, many tests are necessary to optimize the processing parameters. These parameters (window width, threshold, characteristic function) are clearly a function of the type of seismicity recorded by the network and must be optimized from time to time.

If compared to the well-established STA/LTA offline picking algorithm, the *changepost* opens a promising path. Indeed, *changepost* is entirely automatic, meaning that no choice of any parameters is needed to run the algorithm. This feature can be particularly important when, for example, the characteristics of the seismicity of a given area are not well known or when a new seismic monitoring network is set up. The only prior setting regards the maximum number of changepoints to be detected: the larger the number, the more the resulting estimated changepoints, but also the higher the computational time. Furthermore, *changepost* provides automatically the arrival time of the P- and S- waves, and therefore, no intervention is needed by the researcher to identify the arrivals among those possibly triggered.

Changepost can be easily modified to allow the identification of further seismic phases, such as the end of the seismic event. Certainly, interesting results can be obtained by applying the same technique to transforms of the original signal (integrated signal, derivative, frequency filtered, etc.). These future developments that we are foreseeing could certainly improve the performance of the proposed algorithm.

A possible extension of the proposed method is in progress, considering multidimensional signals. The proposed approach can be used as a new picking algorithm to automatically identify the P- and S-waves arrival times on different seismograms recording the same seismic event.

Chapter 7

Minimum contrast for first-order intensity estimation

D'Angelo, N. and Adelfio, G. Minimum contrast for the first-order intensity estimation of complex spatio-temporal point processes. *Work in progress*

In this section, we introduce some in-progress results for future research. We start from the theoretical result related to the expectation of the weighted K-function, using the true first-order intensity function. This theoretical result can serve as an estimation method for obtaining the parameters' estimates of a specific model assumed for the data. Indeed, parameters in spatial-temporal point process models are typically fit by maximum likelihood estimation or some of its variants. The motivation is to generally avoid dealing with the complex likelihoods of some point process models, like the ETAS process, and their maximization. By further considering the local second-order characteristics, we can obtain the whole set of parameters assumed for the fitted model, one for each point of the analysed point pattern.

In detail, section 7.1 provides the sketch of the proposed method, section 7.2 reports some provisional simulations, and section 7.3 concludes with some ideas to be developed in future.

7.1 Methodology

We advocate the use of the K-function for inferential tools, exploiting the theoretical results in Adelfio et al. (2020) for estimating the parameters of a model, assumed for the data. In a few words, the spatio-temporal K-function estimator (both local and global) for a general point process, weighted by the *true* intensity function, has the same expectation of the local spatio-temporal K-function under the Poisson case.

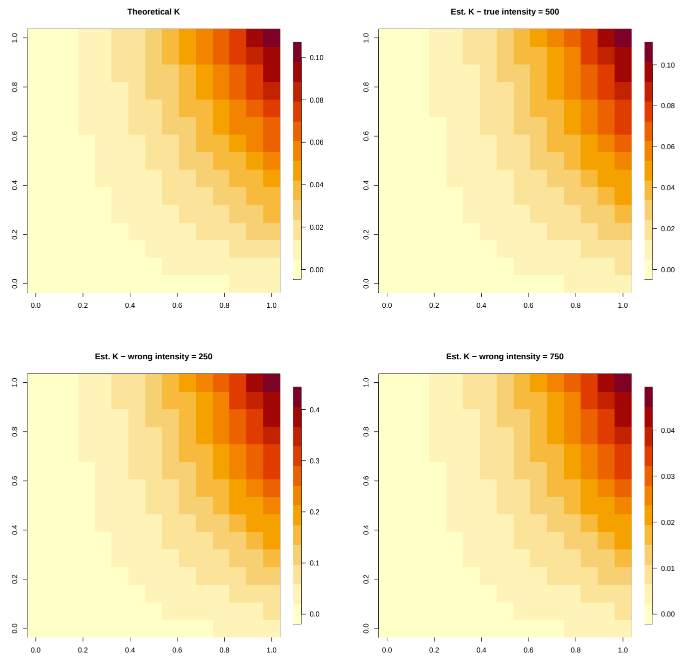


Figure 7.1: (a) Theoretical K-function of a simulated Poisson process with 500 points. (b) Estimated K-function, weighted by the true intensity function. (c)-(d) Estimated K-functions, weighted by some wrong intensities.

The idea is shown graphically in Figure 7.1. Panel (a) of Figure 7.1 displays the theoretical K-function of a simulated Poisson process with 500 points. As an example, we estimate three K-functions, weighted by: the true intensity (panel (b)) and two wrong intensities (250 in panel (c) and 750 in panel (d)). From these plots, we can observe that the overall behaviour of the K-function is the same, and the only difference is the values of the actual K-functions. In particular, the K-function weighted by the true intensity function reports the same values as the theoretical one.

Indeed, the theory suggests that the squared difference between the observed K-function and the theoretical one should approach zero, as the intensity used for weighting the observed K-function approaches the true one. This is shown in Figure 7.2, where the squared differences between the three K-functions of Figure 7.1 and the theoretical one are represented, together with the value of the sum of the squared differences, on the top. As expected, the lowest value is obtained when weighting by the true intensity function.

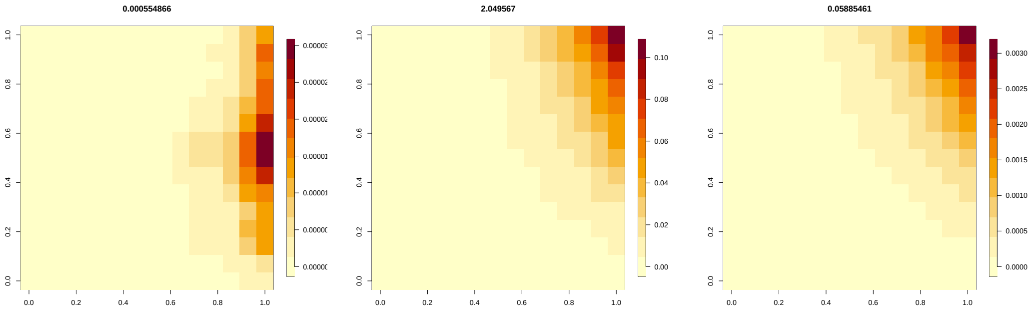


Figure 7.2: Squared differences between the three K-functions of Figure 7.1 and the theoretical one. On top, the value of the sum of the squared differences.

7.1.1 Proposal

Suppose then, that X is a point process with the intensity $\lambda(\mathbf{u}, t; \boldsymbol{\theta})$ (in brief: $\lambda(\boldsymbol{\theta})$) incorporates a vector of parameters $\boldsymbol{\theta} \in \Theta$. By minimizing

$$\mathcal{M}(\boldsymbol{\theta}) = \int_{h_0}^{h_{max}} \int_{r_0}^{r_{max}} \phi(r,h)\{(\hat{K}_I(r,h;\lambda(\boldsymbol{\theta})) - \pi r^2 h)\}^2 \mathrm{d}r\mathrm{d}h \tag{7.1}$$

with respect to $\boldsymbol{\theta}$, we obtain a vector of estimates $\hat{\boldsymbol{\theta}}$:

$$\hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} \mathcal{M}(\boldsymbol{\theta}). \tag{7.2}$$

We believe that the optimization procedure could be improved, using a weight as in Equation 7.1. This would basically correspond to giving more importance to some specific points or regions. In particular, if we weight the objective function by a $\phi(r,h)$ function, this would penalize some specific spatial and temporal lags.

For instance, Diggle (2013) suggested using $\phi(r,h)$ to weight the discrepancy measure by the inverse (approximate) variance of the K-function (in which Guan and Sherman (2007), used their sub-sampling method to achieve). The variance of the K-function however, is typically unknown. For a spatial Poisson process it is suggested to use $\phi(r) = r^{-2}$, but no other specific recommendations for other types of process are given.

7.1.1.1 Local extension

The proposed estimation approach can be extended also to the local case. Suppose that our model incorporates a vector of parameters $\boldsymbol{\theta}$. Let $\hat{K}_I^i(r,h;\lambda(\boldsymbol{\theta}))$ denote the local estimators calculated from the data. For each point indexed by i we consider

$$\mathcal{M}_{local}(\boldsymbol{\theta}_i) = \int_{h_0}^{h_{max}} \int_{r_0}^{r_{max}} \phi(r,h)\{(\hat{K}_I^i(r,h;\lambda(\boldsymbol{\theta})) - \pi r^2 h)\}^2 \mathrm{d}r\mathrm{d}h. \tag{7.3}$$

Then, we can obtain a vector of estimates $\hat{\boldsymbol{\theta}}_i$, one for each point i, as

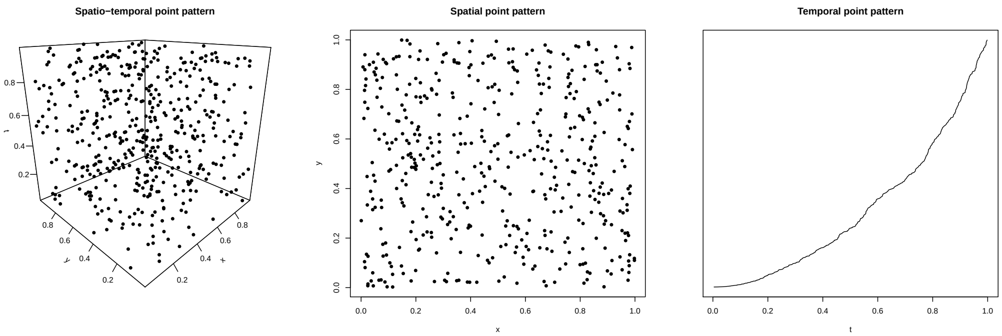
$$\hat{\boldsymbol{\theta}}_i = \arg\min_{\boldsymbol{\theta}} \mathcal{M}_{local}(\boldsymbol{\theta}_i). \tag{7.4}$$

7.2 Simulations

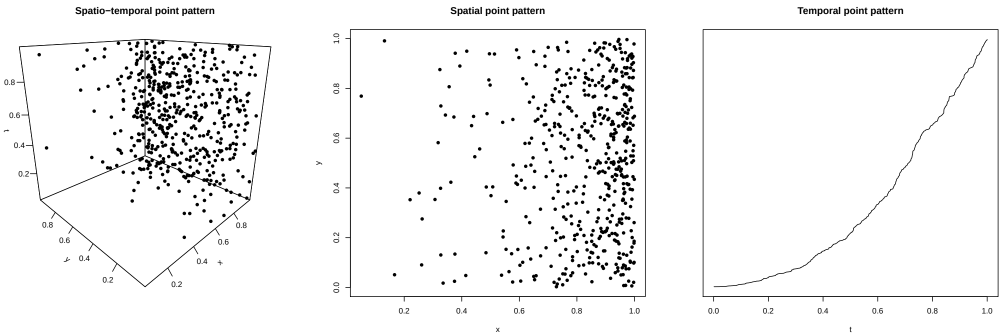
In this section, we report some provisional simulation results, for assessing the proposed estimation procedure. We start by considering simple processes, as described below. We simulate 1000 space-time point patterns in the unit cube with 500 number of points on average from the following point processes:

- homogeneous Poisson process with constant intensity λ;
- inhomogeneous Poisson process with intensity $\lambda(x, y, t) = \exp(\alpha + \beta x)$.

Examples of such simulated patterns are in Figure 7.3.



(a) Realization of a homogeneous point process with intensity $\lambda(x, y, t) = 500$



(b) Realization of an inhomogeneous point process with intensity $\lambda(x, y, t) = \exp(2 + 6x)$

Figure 7.3: Space and time locations of two examples of simulated homogeneous and inhomogeneous patterns.

Table 7.1 reports the *mean* and $sqrt(MSE)$ of the estimated of the two considered processes, averaged over 1000 simulations. In particular, the spatial and temporal distances in the observed weighted K-functions are 15 values ranging from 0 to r_{max} in Equation (7.3) equal to 1/4 of the maximum (space or time) distances.

We notice that the mean of the intensity function for the homogeneous scenario appears systematically overestimated. This might be due to different approximations in the computation of the space-time K-function.

For the inhomogeneous point processes, we employ a penalized procedure, not reported here for brevity, where the penalty enters like an additional data point which is used to "pull" in the parameter direction in which the data provides the least information. The mean of the (unpenalized) estimates for the α and β parameters are 4.3780 and 2.9476 (with 2.1560 and 2.3885 as standard errors, respectively). After having added the penalization (with tuning parameter $R = 2.5$), we note from Table 7.1 that the estimated parameters are pulled away towards the true values.

Table 7.1: Results over 1000 simulations.

Process	$\mathbb{E}[n]$	True par	mean	sqrt(MSE)	mean(s.e.)
Hom. Pois.	500	$\lambda = 500$	529.3286	39.1756	414.0448
Inhom. Pois.	500	$\alpha = 2$	2.8496	2.3815	1.1852
$R = 2.5$		$\beta = 6$	4.8887	2.3385	1.5173

7.3 Future work

The idea presented in this section represents a work in progress. Therefore, much can be done in future.

Section 7.1 introduced the further advantage of our proposal, showing the possibility of fitting local parameters. Preliminary analyses showed a systematic overestimation of the local parameters in the homogeneous case. As for the inhomogeneous case, the local procedure achieves similar results if compared to the global penalized version.

Moreover, we want to run extended simulation studies to assess the performance of the proposed procedure in more complex settings, for instance, with Self-Exciting models such as the ETAS ones.

As for the comparison with other estimation procedures, we put particular emphasis on the following. A similar idea is being developed by Kresin and Schoenberg (2022), who showed that parameters in spatial-temporal point process models, alternatively to MLE, can be estimated consistently, under general conditions, by instead minimizing the Stoyan-Grabarnik (SG) statistic. More specifically, the spatial-temporal region is divided up into cells, and the sum of squares of the SG statistic is minimized. The resulting estimator has desirable properties, is extremely easy and quick to compute, and does not require approximation of the pesky integral in the log-likelihood formula. Therefore, we wish to compare our proposal based on the K-function to both the Maximum Likelihood Estimation and Kresin and Schoenberg (2022)'s proposal.

Chapter 8

Conclusions and future work

In this thesis, we have covered a wide range of topics including modelling, statistical inference, and simulation issues on spatial and spatio-temporal point processes, point processes on linear networks, and non-Euclidean spaces.

In particular, we have provided new local methods and models for complex spatio-temporal point processes. In this way, we have been able to obtain greater detail in the study of space-time phenomena, by identifying outlying points or subareas that require greater attention and could be the subject of further studies. In particular, we have explored both dependence models and second-order characteristics.

Furthermore, we have proven the extent of applicability of our methods, by addressing different contexts of application, such as seismic events, GPS data, traffic accidents, and crimes.

Nevertheless, we believe that the methods and models developed in this thesis can be applied to different scientific fields, and could be suitable for all those phenomena for which it makes sense to hypothesize an interdependence in space and time.

We have drawn conclusions and outlined specific future works for each of the presented research, therefore, for a detailed discussion of the individual projects we refer to the sections' conclusions of the thesis.

In the following, we outline other projects related to the aims of this thesis work. These represent some ongoing attempts to include accessory information to the statistical procedures, namely: the linear network, local indicators of spatio-temporal association, and textual analysis tools, which are becoming more and more used given the increasing availability of data nowadays.

In Diaz-Sepulveda et al. (2022) we consider the problem of features detection, in the presence of clutter, in point processes on a linear network. For the pure spatial case, a previous study addresses the issue of nearest-neighbour clutter removal. We extend this methodology to a more complex geometric context, where the classical properties of a point process change and data visualization are not intuitive. As a result, the method is suitable for a feature with clutter as two superimposed Poisson processes on any linear network. We present simulations and one example of traffic accidents in Bogota-Colombia to illustrate the method.

Then, in Policastro et al. (2022), we combine `robin` and LISTA functions. `robin` is an R package to assess the robustness of the community structure of a network found by one or more methods to give indications about their reliability. We use it to propose a classification algorithm of events in a spatio-temporal point pattern, by means of the local second-order characteristics and the community detection procedure in network analysis. We demonstrate the proposed procedure on a real data analysis of seismic data.

Finally, in Simonetti et al. (2022), we propose to model retweet event sequences using a marked Hawkes process. The aim is to analyse Twitter data by combining temporal point processes theory and textual analysis. Since each retweet event carries a set of properties, we mark the process by different characteristics drawn from the textual analysis, finding that the tone of the description of the Twitter user is a good predictor of the number of retweets in a single cascade.

As a general conclusion, we note that, in this thesis, we have presented the methodology starting from purely spatial point processes definitions, moving then to space and time, starting from Baddeley et al. (2015). The reason is that most work on spatio-temporal statistical models and methods has developed as a series of extensions of spatial methods rather than temporal methods. Diggle (2021)'s suggestion asserts that the origin of the practice of extending spatial tools to space and time is the mathematical appeal of the parametric families of covariance functions for a real-valued spatio-temporal Gaussian process, which is completely (and therefore conveniently) determined by its first and second moments.

Nevertheless, conditioning on the past is a very natural way to build scientifically interpretable models for spatio-temporal phenomena, often with the advantage of a more straightforward statistical analysis than for their purely spatial counterparts. This aspect is not addressed in this dissertation, but it is believed to represent an interesting future research path to investigate.

Appendix A

Simulation schemes for point processes on networks

As we simulate patterns from different spatio-temporal point processes on networks, we detail here the corresponding simulation schemes we have used throughout chapter 5.

Spatio-temporal Poisson point processes

To generate a point pattern from a *homogeneous* Poisson point process in $L \times T$, we use a two-step procedure following Gabriel et al. (2013), but adapted onto the network:

1. Simulate the number of events n occurring in $L \times T$ according to a Poisson distribution with mean $\lambda|L||T|$.
2. Sample each of the n locations and n times according to a uniform distribution on L and T, respectively.

To generate a point pattern from an *inhomogeneous* spatio-temporal point process on a linear network, from the spatio-temporal intensity $\hat{\lambda}(\mathbf{u}, t)$, we proceed with the following steps:

1. Define the generating intensity function $\lambda_0(\mathbf{u}, t) = \hat{\lambda}(\mathbf{u}, t)$
2. Set an upper bound λ_{max} for $\lambda_0(\mathbf{u}, t)$.
3. Simulate a homogeneous Poisson process with intensity λ_{max} and denote by N the number of generated points, with coordinates $(\mathbf{u}', t')$.
4. Compute $p(\mathbf{u}', t') = \frac{\lambda(\mathbf{u}', t')}{\lambda_{max}(\mathbf{u}', t')}$ for each point $(\mathbf{u}', t')$ of an homogeneous Poisson process.

5. Generate a sample $\mathbf{p}$ of size N from the uniform distribution on $(0,1)$.

6. Thin the simulated homogeneous Poisson process $\mathbf{x}$ retaining the $n \leq N$ locations for which $\mathbf{p} \leq p(\mathbf{u}', t')$.

The algorithm provides a point pattern $\mathbf{x}$ with n points. This scheme follows the procedure for simulating inhomogeneous point patterns outlined in Gabriel et al. (2013). In the context of point patterns occurring on linear networks, the events are guaranteed to be constrained on the network since each point $(\mathbf{u}, t)$ of the generating intensity function $\hat{\lambda}(\mathbf{u}, t)$ occurs only on the network line segments as either a vector of values or a function.

The simulations of both the homogeneous and inhomogeneous Poisson point processes are performed through the function `rpoistlpp()` in the R package `stlnpp` (Moradi et al., 2020), by imputing the true intensity functions.

Spatio-temporal log-Gaussian Cox processes

Dealing with log-Gaussian Cox processes, the above diagnostics method has been slightly modified. Indeed, provided the covariance parameters, diagnostics can be carried out following the same procedure outlined above, but simulating point patterns from the intensity function of the fitted LGCP. Therefore, the generation of a point process from a spatio-temporal LGCP on a linear network is proposed and carried out, obtaining the algorithm discussed above and replacing step 1 by adding the two following additional steps:

1.a Generate a realisation from a Gaussian Random Field $S(\mathbf{u}, t)$, with covariance function $\mathbb{C}((\mathbf{u}, t), (\mathbf{v}, s))$ and mean function $\mu(\mathbf{u}, t)$.

1.b Define the generating intensity function $\lambda_0(\mathbf{u}, t) = \hat{\lambda}(\mathbf{u}, t) \exp(S(\mathbf{u}, t))$

The simulation is performed through the function `sim_LGCPnet` of the `R` (R Core Team, 2022) package `localpps` (see Appendix B).

Spatio-temporal clustered point processes

To generate a point pattern from a Poisson cluster point process in $L \times T$ we use the three-step procedure proposed in Gabriel et al. (2013), restraining the results on the network. Defining W as the Euclidean observation region, we proceed with the following steps:

1. Simulate a Poisson process of parent points with intensity $\lambda_p(\mathbf{u}, t)$ in $W' \times T'$, where $W' \subset W$ and $T' \subset T$, to avoid edge-effects.

2. For each simulated parent, generate a random number n_c of offspring from a Poisson distribution with mean m_c.
3. Generate the spatio-temporal displacements of the offspring from their parents as independent realisations from the trivariate distribution with density $f_X(x_1, x_2, x_3)$ in $\mathbb{R}^2 \times \mathbb{R}^+$.
4. Reallocate the points outside the linear support to the closest location on the network.

Clustered Poisson processes are generated on the plane with the `rpcp` function of the `R` (R Core Team, 2022) package `stpp` (Gabriel et al., 2021).

Space-time Epidemic Type Aftershock Sequence processes

We now outline the generating scheme for simulating a pattern from an Epidemic Type Aftershocks-Sequences (ETAS) process (Ogata and Katsura, 1988) with conditional intensity function as in Adelfio and Chiodi (2020) but adapting it to the network. We do this by defining the space-time region in a fixed temporal range and a linear network.

The steps for the simulations are the following:

1. Input the true parameters values μ, κ_0, c, p, d, q, the β parameters, related to covariates, other control parameters such as the boundaries of the space-time region, and parameter b of the Gutenberg-Richter distribution.
2. Generate a random number n_0 from a Poisson distribution with parameter $E(N) = \mu\Delta t$.
3. Generate n_0 triples of space-time uniform coordinates in the space-time region.
4. For each point $\mathbf{x}_j$, $j = 1, \ldots, n_0$:
   (a) Generate a random number n_j from a Poisson distribution with parameter proportional to $exp(\eta_j)$.
   (b) Generate n_j triples of space-time coordinates of aftershocks in the space-time region.
   (c) Add the n_j new points to the set of events.
5. Proceed from step (2) until all the events inside the region are involved in the simulation process as possible generators of further events.

The algorithm provides a point pattern $\mathbf{x}$ with n points. It also generates n magnitudes and other covariates values, if specified.

The simulation is performed through the function `sim_ETASnet` of the `R` (R Core Team, 2022) package `localpps` (see Appendix B). The simulation on the network is guaranteed by the homogeneous spatial Poisson processes being generated on the network.

Appendix B

R codes

Codes for the analyses were written using the software R Core Team (2022) by me, and sometimes partially inspired by some original codes of Marianna Siino and Giada Adelfio. All the codes are meant to form a package, that we shall call `localpps`, to be published on CRAN. The main dependencies of the package are:

- `spatstat` (Baddeley and Turner, 2005);
- `stpp` (Gabriel and Diggle, 2009);
- `stlnpp` (Moradi and Mateu, 2020).

We define a `stp` object as a dataframe with three columns: `x`, `y`, and `t`. If also the linear network `L`, of class linnet, is provided, a `stlp` object is created. The functions `summary`, `print`, and `plot` are implemented to work on these two classes of objects. See below and Figure B.1.

```
> stp1
Spatio-temporal point pattern
43 points
Enclosing window: rectangle = [0.044759, 5] x [0.519806, 5] units
Time period: [0.043, 4.93]

> stlp1
Spatio-temporal point pattern on a linear network
43 points
Linear network with 19 vertices and 26 lines
Enclosing window: rectangle = [-0.01, 5.1] x [-0.01, 5.1] units
Time period: [0.043, 4.93]
```

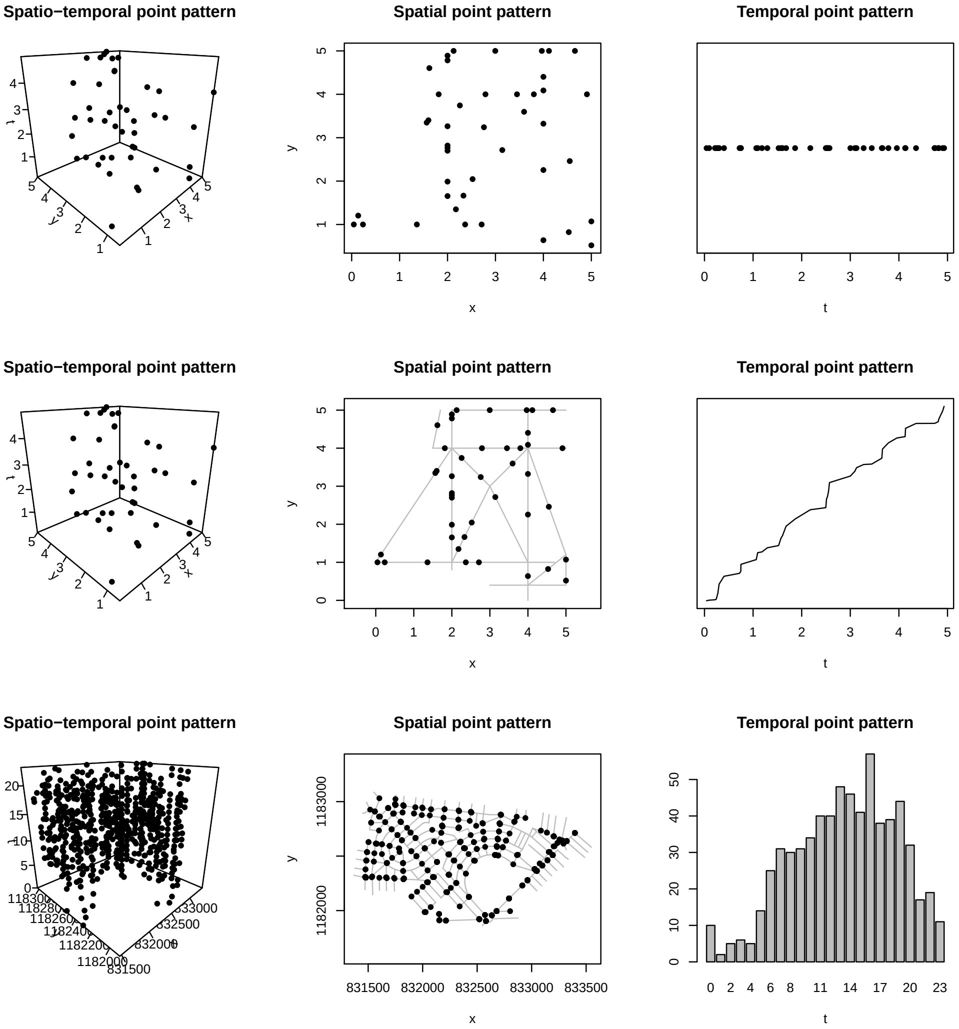


Figure B.1: Plots of the **`stp`** *and* **`stlp`** *objects. The temporal point pattern can be displayed cumulatively by specifying* `tcum = T`*, while a barplot is automatically plotted if there are repeated counts (typically with discrete times). If* `marg = F`*, only the spatio-temporal point pattern is plotted.*

Several functions to convert the classes of the `localpps` package to those of `spatstat`, `stpp`, and `stlnpp`, and vice versa, are provided.

The functions `STLKinhom_local` and `STLginhom_local` implement the inhomogeneous LISTA functions proposed in section 5.2.1, in Equations (5.1) and (5.2), respectively.

```
> STLKinhom_local(X, lambda = lambda, normalize = FALSE,
+ r = NULL, t = NULL, nxy = 10)
```

This function takes as first input a spatio-temporal point process on the network `X`, of class `stlp`. `lambda` must be a vector containing the pointwise first-order intensity. `normalize` indicates whether the normalisation in Equation (5.3) must be considered. If `r` and `t` are not provided, a grid is considered, starting from the `nxy` values. The default is a 10 × 10 grid, where r and t are chosen using the Diggle's rule (Diggle, 2013).
The function returns a `list` of `sumstlpp` objects (Moradi and Mateu, 2020), one for each point in `X`. Each object contains four slots: the $r \times t$ matrix with the estimated values; the $r \times t$ matrix with the theoretical values; the vector of the spatial ranges considered r; the vector of the temporal ranges t.

The homogeneous K-function and pair correlation functions, in Equations (5.4) and (5.5), can be obtained easily with `STLKinhom_local` and `STLKginhom_local`, by imputing a `lambda` vector of constant intensity values, the same for each point:

```
> lambda = rep(npoints(X) / (volume(X$domain) * (X$time[2] +
- X$time[1])), npoints(X))
```

The functions `sim_ETASnet` and `sim_LGCPnet` simulate ETAS processes and log-Gaussian Cox processes on a linear network, respectively. They both return an `stlp` object.

In the following, I list the main functions to perform analyses based on the methods proposed in this thesis. Each of those functions is equipped with its own class and methods: `summary`, `print`, and `plot`.

`jointlgcp` estimates the spatio-temporal log-Gaussian Cox processes by means of the *joint minimum contrast* procedure proposed in Siino et al. (2018a). It can also estimate the local lgcps, following the *locally weighted minimum contrast* procedure introduced in section 3.2. Three covariances are available: separable exponential, Gneiting, and DeIaco-Cesare.

If both `first` and `second` arguments are set to `"global"`, a log-Gaussian Cox process is fitted by means of the joint minimum contrast.

```
lgcp1 <- jointlgcp(stp2, formula = ~ 1, first = "global",
            second = "global")
> summary.jointlgcp(lgcp1)

Joint minimum contrast fit
for a log-Gaussian Cox process with
global first-order intensity and
```

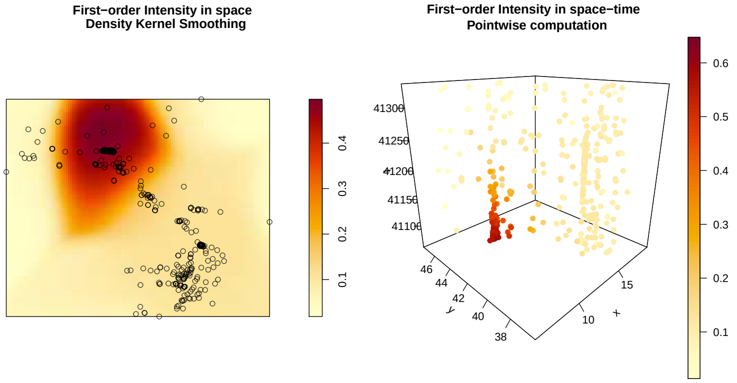
```
global second-order intensity
--------------------------------------------------
Homogeneous Poisson process
with Intensity: 0.01221

Estimated coefficients of the first-order intensity:
(Intercept)
     -4.406
--------------------------------------------------
Covariance function: separable

Estimated coefficients of the second-order intensity:
 sigma  alpha   beta
12.649  0.397  8.057
--------------------------------------------------
Model fitted in 0.147 minutes
```

If `first = "local"`, local parameters for the first-order intensity are provided. In this case, the `summary` function contains information on their distributions. The resulting intensity is used for weighting the second-order statistics used in the minimum contrast procedure, therefore the covariance parameters' estimates are different from the ones of the previous model.

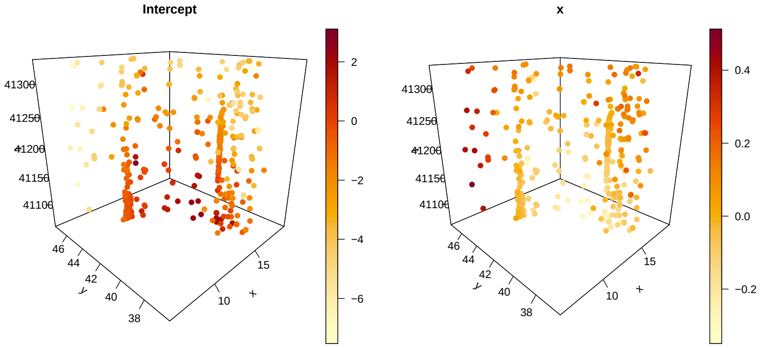
```
> lgcp2 <- jointlgcp(stp2, formula = ~ x, first = "local",
                second = "global")
> summary.jointlgcp(lgcp2)

Joint minimum contrast fit
for a log-Gaussian Cox process with
local first-order intensity and
global second-order intensity
--------------------------------------------------
Inhomogeneous Poisson process
with Trend: ~x

Summary of estimated coefficients of the first-order intensity
  (Intercept)            x
 Min.   :-7.5243   Min.   :-0.349146
 1st Qu.:-2.2680   1st Qu.:-0.062339
 Median :-1.0822   Median :-0.026079
 Mean   :-1.4498   Mean   :-0.015410
 3rd Qu.:-0.2572   3rd Qu.: 0.007207
```

```
 Max.   : 3.1075   Max.   : 0.512574
--------------------------------------------------
Covariance function: separable

Estimated coefficients of the second-order intensity:
sigma alpha  beta
2.598 0.001 2.132
--------------------------------------------------
Model fitted in 0.747 minutes
```

For local models, some functions are implemented, with outputs in Figures B.2 and B.3.

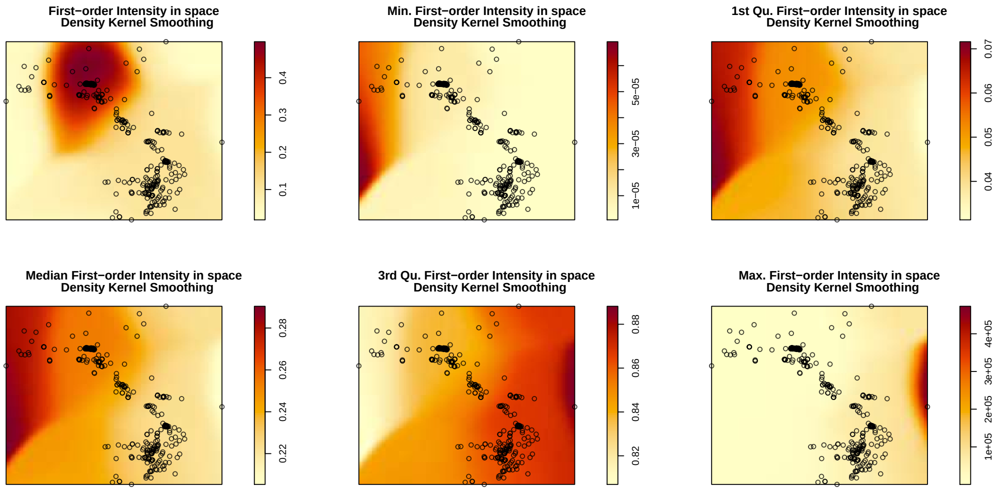


(a) `plot.jointlgcp`



(b) `localplot.jointlgcp`

Figure B.2: `plot.jointlgcp` returns the fitted intensity, displayed both in space (by means of a density kernel smoothing) and in space and time. `localplot.jointlgcp` displays the local estimates.



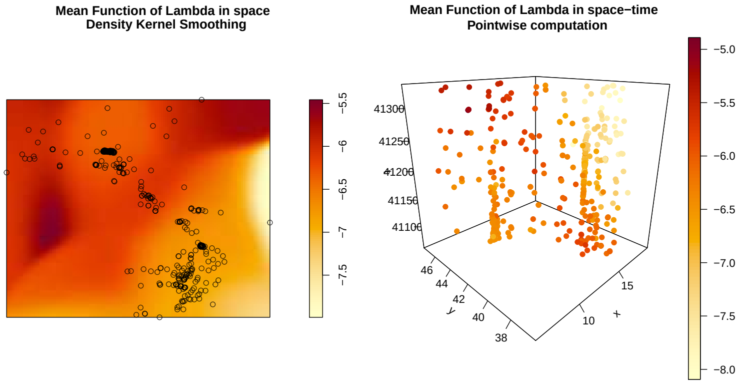
(a) `localsummary.jointlgcp`

Figure B.3: `localsummary.jointlgcp` breaks up the contribution of the local estimates to the fitted intensity, by plotting the overall intensity and the density kernel smoothing of some artificial intensities obtained by imputing the quartiles of the local parameters' distributions.

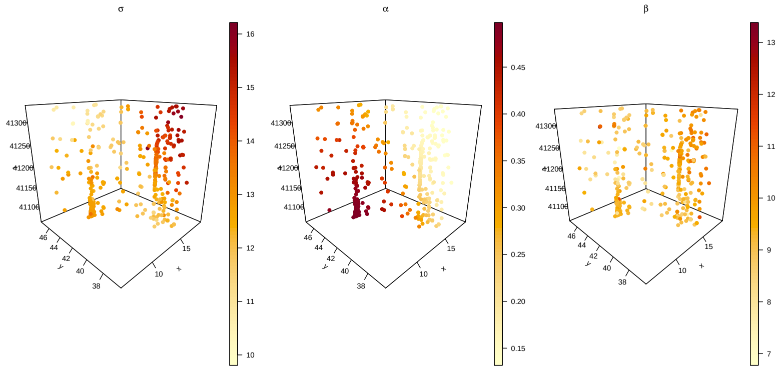
Finally, if `second = "local"` (default option), the model with local parameters for the covariance is fitted.

```
> lgcp3 <- jointlgcp(stp2, formula = ˜ 1, first = "global",
            second = "local")
> summary.jointlgcp(lgcp3)

Joint minimum contrast fit
for a log-Gaussian Cox process with
global first-order intensity and
local second-order intensity
--------------------------------------------------
Homogeneous Poisson process
with Intensity: 0.01221

Estimated coefficients of the first-order intensity:
(Intercept)
     -4.406
--------------------------------------------------
Covariance function: separable
```

```
Summary of estimated coefficients of the second-order intensity
     sigma            alpha             beta
 Min.   : 9.806   Min.   :0.1317   Min.   : 6.776
 1st Qu.:12.527   1st Qu.:0.1841   1st Qu.: 9.104
 Median :12.907   Median :0.2851   Median : 9.382
 Mean   :13.057   Mean   :0.3183   Mean   : 9.418
 3rd Qu.:13.630   3rd Qu.:0.4880   3rd Qu.: 9.836
 Max.   :16.213   Max.   :0.4975   Max.   :13.382
--------------------------------------------------
Model fitted in 3.276 minutes
```



(a) `plot.jointlgcp`



(b) `localplot.joinlgcp`

Figure B.4: In the case of local covariance parameters, `plot.jointlgcp` *returns the mean of the random intensity, displayed both in space (by means of a density kernel smoothing) and in space and time.* `localplot.joinlgcp` *displays the local estimates of the chosen covariance function.*

It is also possible to fit local parameters for both the first and second-order intensity, but we abstain from addressing this case here.

`localtest` runs the permutation test of the local structure of spatio-temporal point pattern data, proposed in Siino et al. (2018b). The network counterpart is also implemented, following section 5.1.

```
test <- localtest.stlp(X, Z, method = "K", k = 19)
> summary.localtest(test)
Test for local differences between two
spatio-temporal point patterns on a linear network
--------------------------------------------------
Backgound pattern X: 24
Alternative pattern Z: 32

7 significant points at alpha = 0.05
```

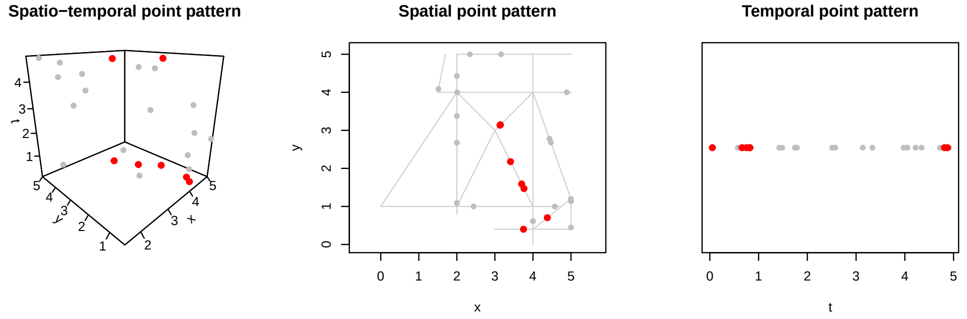


Figure B.5: Output of `summary.localtest` *function, with significant points in red.*

Two functions carry out diagnostics, both globally (`diagnoseG()`) and locally (`diagnose()`), as in Adelfio et al. (2020). Diagnostics of point processes on networks is also implemented, as in section 5.2.

The `diagnoseG` function computes the K-function weighted by the provided intensity to diagnose (given as a vector) and returns the sum of squared differences between the observed and theoretical K-function. It also plots both of them (either in the same or different scale by means of the argument `samescale`), and their difference.

```
> diagnoseG.stp(stp1, intensity = lPOIS)
> diagnoseG.stp(stp1, intensity = lPOIS, samescale = F)
```

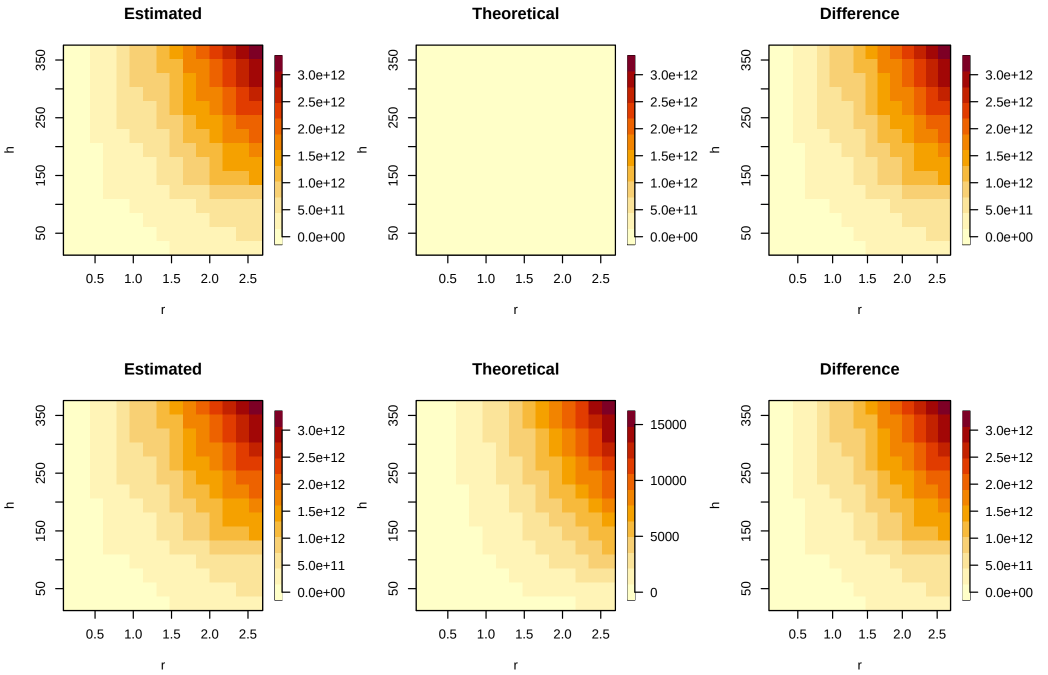


Figure B.6: Output of the `diagnoseG` *function.*

For the local diagnostics, a threshold for the points to be considered outlying must be specified, as follows. This corresponds to the quantile of the distribution of the χ_i^2 statistics computed as in Equation (5.9).

```
> resPOIS <- diagnose.stp(stp1, intensity = lPOIS, p = 0.965)
> resPOIS
Points outlying from the 0.965 percentile
of the anaysed spatio-temporal point pattern
--------------------------------------
Analysed pattern X: 200 points
7 outlying points
```

The plot of the created objects returns information on the outlying points, as displayed in Figure B.7

`infl()` returns the plots of the influential points returned by `diagnose`.

```
infl.localdiag(resPOIS)
```

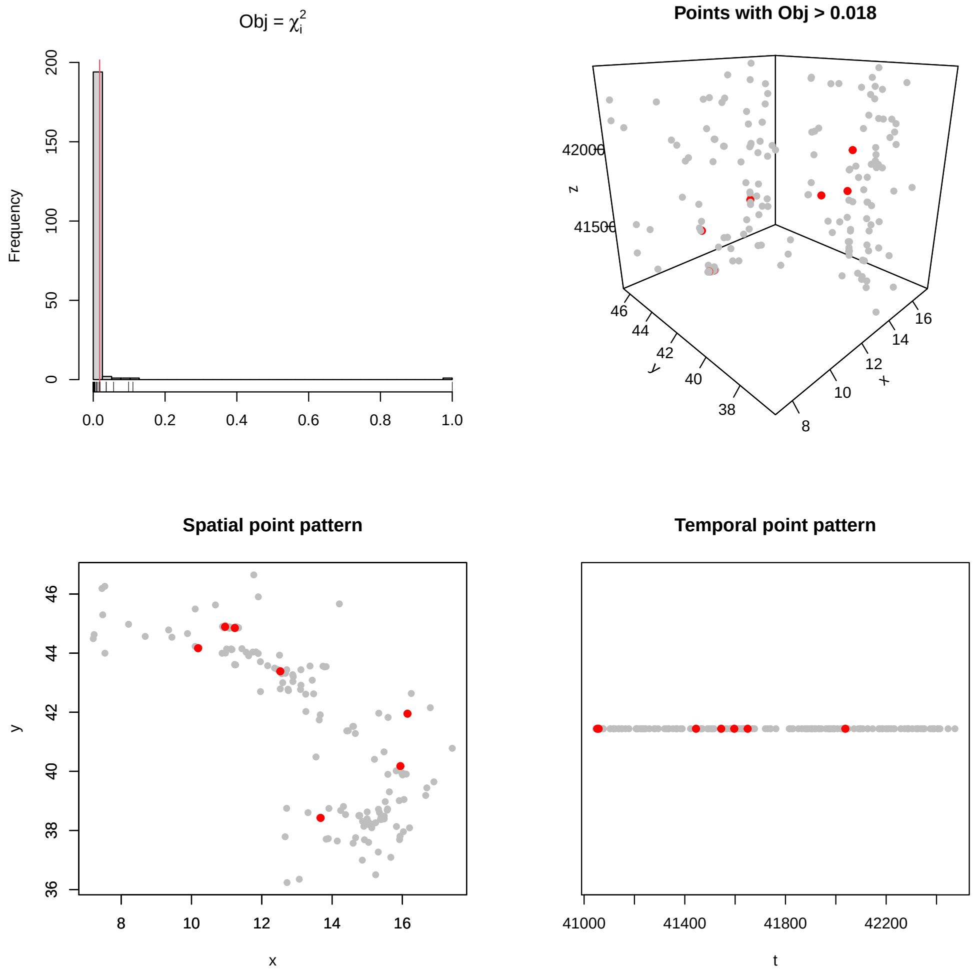


Figure B.7: Output of the `plot.localdiag` *function. If* `marg = F`*, only the top panels are plotted.*

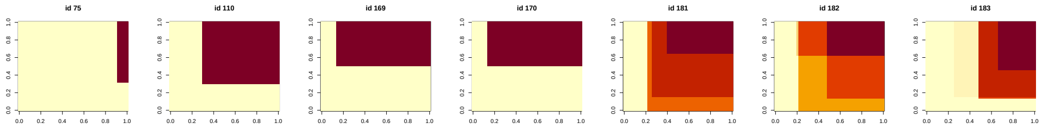


Figure B.8: Output of the `infl` *function: LISTA functions of the outlying points.*

Appendix C

Publications

Scientific publications associated with the thesis research are listed in the following, together with authors' contributions in the CRediT Author Statement format.

1. D'Angelo, N., Siino, M., D'Alessandro, A., and Adelfio, G. (2022g). Local spatial log-gaussian cox processes for seismic data. *Advances in Statistical Analysis.* *https://doi.org/10.1007/s10182-022-00444-w*

2. D'Angelo, N., Adelfio, G., and Mateu, J. (2022c). Locally weighted minimum contrast estimation for spatio-temporal log-gaussian cox processes. *https://arxiv.org/abs/2209.07153* *Computational Statistics & Data Analysis. Major revisions*

3. D'Angelo, N., Adelfio, G., Abbruzzo, A., and Mateu, J. (2022a). Inhomogeneous spatio-temporal point processes on linear networks for visitors' stops data. *The Annals of Applied Statistics*, 16(2):791 – 815

4. D'Angelo, N., Payares, D., Adelfio, G., and Mateu, J. (2022f). Self-exciting point process modelling of crimes on linear networks. *Statistical Modelling.* *https://doi.org/10.1177/1471082X221094146*

5. D'Angelo, N., Adelfio, G., and Mateu, J. (2021b). Assessing local differences between the spatio-temporal second-order structure of two point patterns occurring on the same linear network. *Spatial Statistics*, 45:100534

6. D'Angelo, N., Adelfio, G., and Mateu, J. (2022b). Local inhomogeneous second-order characteristics for spatio-temporal point processes occurring on linear networks. *Statistical Papers,* *http://doi.org/10.1007/s00362-022-01338-4*

7. D'Angelo, N., Adelfio, G., Mateu, J., and Cronie, O. (2022d). Local inhomogeneous weighted summary statistics for marked point processes. *https://arxiv.org/abs/2208.09233* *Journal of Computational and Graphical Statistics. Major revisions*

8. D'Angelo, N., Di Benedetto, A., Adelfio, G., D'Alessandro, A., and Chiodi, M. (2022e). A new picking algorithm based on the variance piecewise constant models. *Stochastic Environmental Research and Risk Assessment*, *https://doi.org/10.1007/s00477-022-02218-x*

9. D'Angelo, N. and Adelfio, G. Minimum contrast for the first-order intensity estimation of complex spatio-temporal point processes. *Work in progress*

CRediT Author Statement

1. Nicoletta D'Angelo: Conceptualization, Methodology, Software, Writing - original draft. Marianna Siino: Conceptualization, Writing - Reviewing and Editing, Data Curation. Antonino D'Alessandro: Writing - Reviewing and Editing. Giada Adelfio: Conceptualization, Methodology, Writing - Reviewing and Editing, Supervision.
2. Nicoletta D'Angelo: Conceptualization, Methodology, Software, Writing - original draft. Giada Adelfio: Conceptualization, Writing - Reviewing and Editing, Supervision. Jorge Mateu: Writing - Reviewing and Editing, Supervision.
3. Nicoletta D'Angelo: Conceptualization, Methodology, Software, Writing - original draft. Giada Adelfio: Conceptualization, Writing - Reviewing and Editing, Supervision. Antonino Abbruzzo: Writing - Reviewing and Editing, Data Curation. Jorge Mateu: Conceptualization, Writing - Reviewing and Editing, Supervision.
4. Nicoletta D'Angelo: Conceptualization, Methodology, Software, Writing - original draft. David Payares: Methodology, Data curation, Software, Writing - original draft. Giada Adelfio: Conceptualization, Writing - Reviewing and Editing, Supervision. Jorge Mateu: Conceptualization, Writing - Reviewing and Editing, Supervision.
5. Nicoletta D'Angelo: Conceptualization, Methodology, Software, Writing - original draft. Giada Adelfio: Writing - Reviewing and Editing, Supervision. Jorge Mateu: Writing - Reviewing and Editing, Supervision.
6. Nicoletta D'Angelo: Conceptualization, Methodology, Software, Writing - original draft. Giada Adelfio: Writing - Reviewing and Editing, Supervision. Jorge Mateu: Writing - Reviewing and Editing, Supervision.
7. Nicoletta D'Angelo: Conceptualization, Methodology, Software, Writing - original draft. Giada Adelfio: Writing - Reviewing and Editing, Supervision. Jorge Mateu: Writing - Reviewing and Editing, Supervision. Ottmar Cronie: Writing - Reviewing and Editing, Supervision.
8. Nicoletta D'Angelo: Conceptualization, Methodology, Software, Writing - original draft. Andrea Di Benedetto: Conceptualization, Software, Writing - original draft. Giada Adelfio: Conceptualization, Methodology, Writing - Reviewing and Editing, Supervision. Antonino D'Alessandro: Conceptualization, Writing - Reviewing and Editing, Supervision. Marcello Chiodi: Writing - Reviewing and Editing, Supervision.
9. Nicoletta D'Angelo: Conceptualization, Methodology, Software, Writing - original draft. Giada Adelfio: Conceptualization, Methodology, Writing - Reviewing and Editing, Supervision.

Bibliography

Abbruzzo, A., Ferrante, M., and De Cantis, S. (2021). A pre-processing and network analysis of gps tracking data. *Spatial Economic Analysis*, 16(2):217–240.

Adelfio, G. (2012). Change-point detection for variance piecewise constant models. *Communications in Statistics-Simulation and Computation*, 41(4):437–448.

Adelfio, G. and Chiodi, M. (2015a). Alternated estimation in semi-parametric space-time branching-type point processes with application to seismic catalogs. *Stochastic Environmental Research and Risk Assessment*, 29(2):443–450.

Adelfio, G. and Chiodi, M. (2015b). Flp estimation of semi-parametric models for space-time point processes and diagnostic tools. *Spatial Statistics*, 14:119–132.

Adelfio, G. and Chiodi, M. (2020). Including covariates in a space-time point process with application to seismicity. *Statistical Methods & Applications*, pages 1–25.

Adelfio, G., Chiodi, M., D'Alessandro, A., and Luzio, D. (2011). Fpca algorithm for waveform clustering. *Journal of Communication and Computer*, 8(6):494–502.

Adelfio, G., Chiodi, M., D'Alessandro, A., Luzio, D., D'Anna, G., and Mangano, G. (2012). Simultaneous seismic wave clustering and registration. *Computers & geosciences*, 44:60–69.

Adelfio, G. and Schoenberg, F. P. (2009). Point process diagnostics based on weighted second-order statistics and their asymptotic properties. *Annals of the Institute of Statistical Mathematics*, 61(4):929–948.

Adelfio, G., Siino, M., Mateu, J., and Rodríguez-Cortés, F. J. (2020). Some properties of local weighted second-order statistics for spatio-temporal point processes. *Stochastic Environmental Research and Risk Assessment*, 34(1):149–168.

Akaike, H. (1975). Markovian representation of stochastic processes by canonical variables. *SIAM Journal on Control*, 13(1):162–173.

Akaike, H. (1998). Autoregressive model fitting for control. In *Selected Papers of Hirotugu Akaike*, pages 153–170. Springer.

Aldersons, F. (2004). Toward three-dimensional crustal structure of the dead sea region from local earthquake tomography.

Allen, R. (1982). Automatic phase pickers: Their present use and future prospects. *Bulletin of the Seismological Society of America*, 72(6B):S225–S242.

Allen, R. V. (1978). Automatic earthquake recognition and timing from single traces. *Bulletin of the Seismological Society of America*, 68(5):1521–1532.

Alm, S. E. (1998). Approximation and simulation of the distributions of scan statistics for poisson processes in higher dimensions. *Extremes*, 1(1):111–126.

Amato, A., Chiarabba, C., Selvaggi, G., et al. (1997). Crustal and deep seismicity in italy (30 years after). *Annals of Geophysics.*

Ang, Q. W., Baddeley, A., and Nair, G. (2012). Geometrically corrected second order analysis of events on a linear network, with applications to ecology and criminology. *Scandinavian Journal of Statistics*, 39(4):591–617.

Anselin, L. (1995). Local indicators of spatial association-lisa. *Geographical analysis*, 27(2):93–115.

Anselin, L. (1996). Chapter eight the moran scatterplot as an esda tool to assess local instability in spatial association. *Spatial Analytical*, 4:121.

Anwar, S., Stein, A., and van Genderen, J. (2012). Implementation of the marked strauss point process model to the epicenters of earthquake aftershocks. In Leung, Y., editor, *Advances in Geo-Spatial Information Science.*, pages 125–140. Taylor & Francis, London.

Baddeley, A. (2017). Local composite likelihood for spatial point processes. *Spatial Statistics*, 22:261–295.

Baddeley, A. (2019). *spatstat.local: Extension to 'spatstat' for Local Composite Likelihood.* R package version 3.6-0.

Baddeley, A., Bárány, I., and Schneider, R. (2006a). *Stochastic Geometry: Lectures Given at the CIME Summer School Held in Martina Franca, Italy, September 13-18, 2004.* Springer.

Baddeley, A., Chang, Y.-M., Song, Y., and Turner, R. (2012). Nonparametric estimation of the dependence of a spatial point process on spatial covariates. *Statistics and its interface*, 5(2):221–236.

Baddeley, A., Davies, T. M., Hazelton, M. L., Rakshit, S., and Turner, R. (2022). Fundamental problems in fitting spatial cluster process models. *Spatial Statistics*, page 100709.

Baddeley, A., Gregori, P., Mateu, J., Stoica, R., and Stoyan, D. (2006b). *Case Studies in Spatial Point Pattern Modelling.* Lecture Notes in Statistics, 185. Springer, New York.

Baddeley, A. and Møller, J. (1989). Nearest-neighbour markov point processes and random sets. *International Statistical Review*, 57(2):89–121.

Baddeley, A., Nair, G., Rakshit, S., and McSwiggan, G. (2017). " stationary" point processes are uncommon on linear networks. *Stat*, 6(1):68–78.

Baddeley, A., Nair, G., Rakshit, S., McSwiggan, G., and Davies, T. M. (2020). Analysing point patterns on networks - a review. *Spatial Statistics*, page 100435.

Baddeley, A., Rubak, E., and Turner, R. (2015). *Spatial point patterns: methodology and applications with R.* Chapman and Hall/CRC.

Baddeley, A. and Turner, R. (2005). spatstat: An R package for analyzing spatial point patterns. *Journal of Statistical Software*, 12(6):1–42.

Baddeley, A., Turner, R., Mateu, J., and Bevan, A. (2013). Hybrids of gibbs point process models and their implementation. *Journal of Statistical Software.*

Baddeley, A., Turner, R., Møller, J., and Hazelton, M. (2005). Residual analysis for spatial point processes (with discussion). *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 67(5):617–666.

Baddeley, A., Turner, R., and Rubak, E. (2021). *spatstat.data: Datasets for 'spatstat' Family.* R package version 2.1-0.

Baddeley, A. and Turner, T. R. (2000). Pratical maximum pseudo likelihood for spatial point patterns (with discussion). *Australian & New Zealand Journal of Statistics*, 42(3):283–322.

Baddeley, A. J., Møller, J., and Waagepetersen, R. (2000). Non-and semi-parametric estimation of interaction in inhomogeneous point patterns. *Statistica Neerlandica*, 54(3):329–350.

Baer, M. and Kradolfer, U. (1987). An automatic phase picker for local and teleseismic events. *Bulletin of the Seismological Society of America*, 77(4):1437–1445.

Banfield, J. D. and Raftery, A. E. (1993). Model-based gaussian and non-gaussian clustering. *Biometrics*, pages 803–821.

Bell, M. L. and Grunwald, G. K. (2004). Mixed models for the analysis of replicated spatial point patterns. *Biostatistics*, 5(4):633–648.

Berman, M. (1986). Testing for spatial association between a point process and another stochastic process. *Applied Statistics*, 35(1):54–62.

Berman, M. and Diggle, P. (1989). Estimating weighted integrals of the second-order intensity of a spatial point process. *Journal of the Royal Statistical Society: Series B (Methodological)*, 51(1):81–92.

Berman, M. and Turner, T. R. (1992). Approximating point process likelihoods with glim. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 41(1):31–38.

Borruso, G. (2005). Network density estimation: analysis of point patterns over a network. In *International Conference on Computational Science and Its Applications*, pages 126–132. Springer.

Borruso, G. (2008). Network density estimation: a gis approach for analysing point patterns in a network space. *Transactions in GIS*, 12(3):377–402.

Brida, J. G. and Zapata, S. (2010). Economic impacts of cruise tourism: The case of costa rica. *Anatolia*, 21(2):322–338.

Brix, A. and Diggle, P. J. (2001). Spatiotemporal prediction for log-gaussian cox processes. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 63(4):823–841.

Butz, W. P. and Torrey, B. B. (2006). Some frontiers in social science. *Science*, 312(5782):1898–1900.

Casado-Díaz, A. B., Navarro-Ruiz, S., Nicolau, J. L., and Ivars-Baidal, J. (2021). Expanding our understanding of cruise visitors' expenditure at destinations: The role of spatial patterns, onshore visit choice and cruise category. *Tourism Management*, 83:104199.

Chernoff, H. and Zacks, S. (1964). Estimating the current mean of a normal distribution which is subjected to changes in time. *The Annals of Mathematical Statistics*, 35(3):999–1018.

Chiodi, M., Adelfio, G., D'Alessandro, A., and Luzio, D. (2013). Clustering and registration of multidimensional functional data. In *Statistical Models for Data Analysis*, pages 89–97. Springer.

Chiodi, M., Nicolis, O., Adelfio, G., D'Angelo, N., and Gonzalez, A. (2021). Etas space–time modeling of chile triggered seismicity using covariates: Some preliminary results. *Applied Sciences*, 11(19):9143.

Chiu, S. N., Stoyan, D., Kendall, W. S., and Mecke, J. (2013). *Stochastic geometry and its applications.* John Wiley & Sons.

Choi, E. and Hall, P. (1999). Nonparametric approach to analysis of space-time data on earthquake occurrences. *Journal of Computational and Graphical Statistics*, 8(4):733–748.

Clements, R. A., Schoenberg, F. P., and Veen, A. (2012). Evaluation of space–time point process models using super-thinning. *Environmetrics*, 23(7):606–616.

Comas, C., Delicado, P., and Mateu, J. (2011). A second order approach to analyse spatial point patterns with functional marks. *Test*, 20(3):503–523.

Cox, D. (1972). The statistical analysis of dependencies in point processes. *Stochastic Point Processes. Wiley: New York*, pages 55–66.

Cox, D. R. (1955). Some statistical methods connected with series of events. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 129–164.

Cressie, N. (2015). *Statistics for spatial data.* John Wiley & Sons.

Cressie, N. and Collins, L. B. (2001). Analysis of spatial point patterns using bundles of product density lisa functions. *Journal of Agricultural, Biological, and Environmental Statistics*, 6(1):118–135.

Cronie, O., Moradi, M., and Mateu, J. (2020). Inhomogeneous higher-order summary statistics for point processes on linear networks. *Statistics and computing*, 30(5):1221–1239.

Cronie, O. and Van Lieshout, M. (2015). A J-function for inhomogeneous spatio-temporal point processes. *Scandinavian Journal of Statistics*, 42(2):562–579.

Cronie, O. and van Lieshout, M. N. M. (2016). Summary statistics for inhomogeneous marked point processes. *Annals of the Institute of Statistical Mathematics*, 68(4):905–928.

Cronie, O. and Van Lieshout, M. N. M. (2018). A non-model-based approach to bandwidth selection for kernel estimators of spatial intensity functions. *Biometrika*, 105(2):455–462.

Daley, D. J. and Vere-Jones, D. (2007). *An Introduction to the Theory of Point Processes. Volume II: General Theory and Structure.* Springer-Verlag, New York, second edition.

D'Angelo, N., Adelfio, G., Abbruzzo, A., and Mateu, J. (2022a). Inhomogeneous spatio-temporal point processes on linear networks for visitors' stops data. *The Annals of Applied Statistics*, 16(2):791 – 815.

D'Angelo, N., Adelfio, G., D'Alessandro, A., and Chiodi, M. (2020). A fast and efficient picking algorithm for earthquake early warning application based on the variance piecewise constant models. In *International Conference on Computational Science and Its Applications*, pages 903–913. Springer.

D'Angelo, N., Adelfio, G., D'Alessandro, A., and Chiodi, M. (2021a). Evaluating the performance of a new picking algorithm based on the variance piecewise constant models. In *50th Meeting of the Italian Statistical Society, ISBN 9788891927361.*

D'Angelo, N., Adelfio, G., and Mateu, J. (2021b). Assessing local differences between the spatio-temporal second-order structure of two point patterns occurring on the same linear network. *Spatial Statistics*, 45:100534.

D'Angelo, N., Adelfio, G., and Mateu, J. (2022b). Local inhomogeneous second-order characteristics for spatio-temporal point processes occurring on linear networks. *Statistical Papers, http://doi.org/10.1007/s00362-022-01338-4.*

D'Angelo, N., Adelfio, G., and Mateu, J. (2022c). Locally weighted minimum contrast estimation for spatio-temporal log-gaussian cox processes. *https://arxiv.org/abs/2209.07153 Computational Statistics & Data Analysis. Major revisions.*

D'Angelo, N., Adelfio, G., Mateu, J., and Cronie, O. (2022d). Local inhomogeneous weighted summary statistics for marked point processes. *https://arxiv.org/abs/2208.09233 Journal of Computational and Graphical Statistics. Major revisions.*

D'Angelo, N., Di Benedetto, A., Adelfio, G., D'Alessandro, A., and Chiodi, M. (2022e). A new picking algorithm based on the variance piecewise constant models. *Stochastic Environmental Research and Risk Assessment, https://doi.org/10.1007/s00477-022-02218-x.*

D'Angelo, N., Payares, D., Adelfio, G., and Mateu, J. (2022f). Self-exciting point process modelling of crimes on linear networks. *Statistical Modelling. https://doi.org/10.1177/1471082X221094146.*

D'Angelo, N., Siino, M., D'Alessandro, A., and Adelfio, G. (2022g). Local spatial log-gaussian cox processes for seismic data. *Advances in Statistical Analysis. https://doi.org/10.1007/s10182-022-00444-w.*

Dasgupta, A. and Raftery, A. E. (1998). Detecting features in spatial point processes with clutter via model-based clustering. *Journal of the American statistical Association*, 93(441):294–302.

Davies, T. M. and Hazelton, M. L. (2013). Assessing minimum contrast parameter estimation for spatial and spatiotemporal log-gaussian cox processes. *Statistica Neerlandica*, 67(4):355–389.

De Cantis, S., Ferrante, M., Kahani, A., and Shoval, N. (2016). Cruise passengers' behavior at the destination: Investigation using gps technology. *Tourism Management*, 52:133–150.

De Cesare, L., Myers, D., and Posa, D. (2002). Fortran programs for space-time modeling. *Computers & Geosciences*, 28(2):205–212.

De Iaco, S., Myers, D. E., and Posa, D. (2002). Nonseparable space-time covariance models: some parametric families. *Mathematical Geology*, 34(1):23–42.

Di Stefano, R., Chiarabba, C., Lucente, F., and Amato, A. (1999). Crustal and uppermost mantle structure in italy fron the inversion of p-wave arrival times: geodynamic impliations. *Geophysical Journal International*, 139(2):483–498.

Diaz-Sepulveda, J. F., D'Angelo, N., Adelfio, G., Gonzalez, J., and Rodriguez-Cortez, F. J. (2022). Nearest-neighbor clutter removal for estimating features in linear point processes. *https://arxiv.org/abs/2209.14082*.

Diggle, P. (1985). A kernel method for smoothing point process data. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 34(2):138–147.

Diggle, P. J. (1979). On parameter estimation and goodness-of-fit testing for spatial point patterns. *Biometrics*, pages 87–101.

Diggle, P. J. (2013). *Statistical analysis of spatial and spatio-temporal point patterns*. Chapman and Hall/CRC.

Diggle, P. J. (2021). Conditional intensity: A powerful tool for modelling and analysing point process data. *Australian & New Zealand Journal of Statistics*, 63(1):83–92.

Diggle, P. J. and Gratton, R. J. (1984). Monte carlo methods of inference for implicit statistical models. *Journal of the Royal Statistical Society: Series B (Methodological)*, 46(2):193–212.

Diggle, P. J., Moraga, P., Rowlingson, B., Taylor, B. M., et al. (2013). Spatial and spatio-temporal log-gaussian cox processes: extending the geostatistical paradigm. *Statistical Science*, 28(4):542–563.

Domènech, A., Gutiérrez, A., and Anton Clavé, S. (2020a). Cruise passengers' spatial behaviour and expenditure levels at destination. *Tourism Planning & Development*, 17(1):17–36.

Domènech, A., Gutiérrez, A., and Clavé, S. A. (2020b). Built environment and urban cruise tourists' mobility. *Annals of Tourism Research*, 81:102–189.

Efron, B. (1982). *The jackknife, the bootstrap, and other resampling plans*, volume 38. Siam.

Efron, B., Hastie, T., Johnstone, I., Tibshirani, R., et al. (2004). Least angle regression. *The Annals of statistics*, 32(2):407–499.

Eguchi, S. (1983). Second order efficiency of minimum contrast estimators in a curved exponential family. *The Annals of Statistics*, pages 793–803.

Elgethun, K., Fenske, R. A., Yost, M. G., and Palcisko, G. J. (2003). Time-location analysis for exposure assessment studies of children using a novel global positioning system instrument. *Environmental Health Perspectives*, 111(1):115–122.

Ester, M., Kriegel, H.-P., Sander, J., Xu, X., et al. (1996). A density-based algorithm for discovering clusters in large spatial databases with noise. In *Kdd*, volume 96, pages 226–231.

Fan, Y., Zhu, X., She, B., Guo, W., and Guo, T. (2018). Network-constrained spatio-temporal clustering analysis of traffic collisions in jianghan district of wuhan, china. *PLoS one*, 13(4):e0195093.

Federer, H. (2014). *Geometric measure theory.* Springer.

Ferrante, M., De Cantis, S., and Shoval, N. (2018). A general framework for collecting and analysing the tracking data of cruise passengers at the destination. *Current Issues in Tourism*, 21(12):1426–1451.

Fotheringham, A. S. and Sachdeva, M. (2022). On the importance of thinking locally for statistics and society. *Spatial Statistics*, page 100601.

Gabriel, E. and Diggle, P. J. (2009). Second-order analysis of inhomogeneous spatio-temporal point process data. *Statistica Neerlandica*, 63(1):43–51.

Gabriel, E., Diggle, P. J., Rowlingson, B., and Rodriguez-Cortes, F. J. (2021). *stpp: Space-Time Point Pattern Simulation, Visualisation and Analysis.* R package version 2.0-5.

Gabriel, E., Rowlingson, B. S., and Diggle, P. J. (2013). stpp: An R package for plotting, simulating and analyzing Spatio-Temporal Point Patterns. *Journal of Statistical Software*, 53(2):1–29.

Gardner, L. (1969). On detecting changes in the mean of normal variates. *The Annals of Mathematical Statistics*, 40(1):116–126.

Gelfand, A. E., Diggle, P., Guttorp, P., and Fuentes, M. (2010). *Handbook of spatial statistics*. CRC press.

Getis, A. and Franklin, J. (1987). Second-order neighborhood analysis of mapped point patterns. *Ecology*, 68(3):473–477.

Getis, A. and Ord, J. K. (1992). The analysis of spatial association by use of distance statistics. *Geographical Analysis*, 24(3):189–206.

Geyer, C. J. (1999). Likelihood inference for spatial point processes. *Stochastic Geometry: Likelihood and Computation*, 80:79–140.

Ghorbani, M., Cronie, O., Mateu, J., and Yu, J. (2021). Functional marked point processes: a natural structure to unify spatio-temporal frameworks and to analyse dependent functional data. *Test*, 30(3):529–568.

Gilardi, A., Borgoni, R., and Mateu, J. (2021). A non-separable first-order spatio-temporal intensity for events on linear networks: an application to ambulance interventions. *arXiv preprint arXiv:2106.00457*.

Giunta, G., Luzio, D., Agosta, F., Calo, M., Di Trapani, F., Giorgianni, A., Oliveri, E., Orioli, S., Perniciaro, M., Vitale, M., Chiodi, M., and Adelfio, G. (2009). An integrated approach to investigate the seismotectonics of northern sicily and southern tyrrhenian. *Tectonophysics*, 476:13–21.

Gneiting, T., Genton, M. G., and Guttorp, P. (2006). Geostatistical space-time models, stationarity, separability, and full symmetry. *Monographs On Statistics and Applied Probability*, 107:151.

Gong, L., Sato, H., Yamamoto, T., Miwa, T., and Morikawa, T. (2015). Identification of activity stop locations in gps trajectories by density-based clustering method combined with support vector machines. *Journal of Modern Transportation*, 23(3):202–213.

González, J. A., Rodríguez-Cortés, F., Cronie, O., and Mateu, J. (2016). Spatio-temporal point process statistics: a review. *Spatial Statiscts*, 18:505–544.

Guan, Y. (2006). A composite likelihood approach in fitting spatial point process models. *Journal of the American Statistical Association*, 101(476):1502–1512.

Guan, Y. and Sherman, M. (2007). On least squares fitting for stationary spatial point processes. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 69(1):31–49.

Hall, J., Aksu, A., and Yaltirak, C. (2005). Miocene to recent tectonic evolution of the eastern mediterranean; new pieces of the old mediterranean puzzle. *Marine Geology*, 221(1-4):1–13.

Hartung, J., Elpelt, B., and Klösener, K.-H. (2014). *Statistik: Lehr-und Handbuch der angewandten Statistik*. Walter de Gruyter GmbH & Co KG.

Hastie, T. J. and Tibshirani, R. J. (1990). *Generalized additive models*, volume 43. CRC press.

Hawkes, A. and Adamopoulos, L. (1973). Cluster models for erthquakes-regional comparison. *Bulletin of the International Statistical Institute*, 45(3):454–461.

Hawkes, A. G. (1971a). Point spectra of some mutually exciting point processes. *Journal of the Royal Statistical Society: Series B (Methodological)*, 33(3):438–443.

Hawkes, A. G. (1971b). Spectra of some self-exciting and mutually exciting point processes. *Biometrika*, 58(1):83–90.

Hawkins, D. (1992). Detecting shifts in functions of multivariate location and covariance parameters. *Journal of Statistical Planning and Inference*, 33(2):233–244.

Hu, F., Li, Z., Yang, C., and Jiang, Y. (2019). A graph-based approach to detecting tourist movement patterns using social media data. *Cartography and Geographic Information Science*, 46(4):368–382.

Iftimi, A., Cronie, O., and Montes, F. (2019). Second-order analysis of marked inhomogeneous spatiotemporal point processes: Applications to earthquake data. *Scandinavian Journal of Statistics*, 46(3):661–685.

Illian, J., Benson, E., Crawford, J., and Staines, H. (2006). Principal component analysis for spatial point processes-assessing the appropriateness of the approach in an ecological context. In *Case studies in spatial point process modeling*, pages 135–150. Springer.

Illian, J., Penttinen, A., Stoyan, H., and Stoyan, D. (2008). *Statistical analysis and modelling of spatial point patterns*, volume 70. John Wiley & Sons.

Illian, J. B. and Hendrichsen, D. K. (2010). Gibbs point process models with mixed effects. *Environmetrics: The official journal of the International Environmetrics Society*, 21(3-4):341–353.

Kalair, K., Connaughton, C., and Alaimo Di Loro, P. (2021). A non-parametric hawkes process model of primary and secondary accidents on a uk smart motorway. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 70(1):80–97.

Kallenberg, O. (1984). An informal guide to the theory of conditioning in point processes. *International Statistical Review/Revue Internationale de Statistique*, 52(2):151–164.

Komatitsch, D., Liu, Q., Tromp, J., Suss, P., Stidham, C., and Shaw, J. H. (2004). Simulations of ground motion in the los angeles basin based upon the spectral-element method. *Bulletin of the Seismological Society of America*, 94(1):187–206.

Komatitsch, D. and Tromp, J. (1999). Introduction to the spectral element method for three-dimensional seismic wave propagation. *Geophysical journal international*, 139(3):806–822.

Kresin, C. and Schoenberg, F. (2022). Estimation of spatial-temporal point process models using a stoyan-grabarnik statistic. *METMA X*, page 13.

Küperkoch, L., Meier, T., and Diehl, T. (2012). Automated event and phase identification. *New Manual of Seismological Observatory Practice 2 (NMSOP-2)*, pages 1–52.

Küperkoch, L., Meier, T., Lee, J., Friederich, W., and Group, E. W. (2010). Automated determination of p-phase arrival times at regional and local distances using higher order statistics. *Geophysical Journal International*, 181(2):1159–1170.

Kurashima, T., Iwata, T., Irie, G., and Fujimura, K. (2010). Travel route recommendation using geotags in photo sharing sites. In *Proceedings of the 19th ACM international conference on Information and knowledge management*, pages 579–588.

Larrota, R., Gomez, M., and Beltran, C. (2017). Modus operandi en personas recluidas por el delito de hurto en prisiones de bucaramanga, colombia modus operandi in persons detained for the crime of theft in prisons of bucaramanga, colombia. *INFORMES Psicologicos REPORTS in Psychology*, 17(2):107–118.

Larsen, S., Wolff, K., Marnburg, E., and Øgaard, T. (2013). Belly full, purse closed: Cruise line passengers' expenditures. *Tourism Management Perspectives*, 6:142–148.

Le Pichon, X. and Angelier, J. (1979). The hellenic arc and trench system: a key to the neotectonic evolution of the eastern mediterranean area. *Tectonophysics*, 60(1):1–42.

Lesage, L. (2020). *A Hawkes process to make aware people of the severity of COVID-19 outbreak: application to cases in France.* PhD thesis, Université de Lorraine; University of Luxembourg.

Li, Z. and Wood, S. N. (2020). Faster model matrix crossproducts for large generalized linear models with discretized covariates. *Statistics and Computing*, 30(1):19–25.

Lim, K. W., Wang, W., Nguyen, H., Lee, Y., Cai, C., and Chen, F. (2016). Traffic flow modelling with point processes. In *Proceedings of the 23rd World Congress on Intelligent Transport Systems*, pages 1–12.

Liu, M. and Stein, S. (2019). Earthquake, aftershocks. *Encyclopedia of solid earth geophysics*, pages 1–4.

Loader, C. R. et al. (1999). Bandwidth selection: classical or plug-in? *The Annals of Statistics*, 27(2):415–438.

Lutikov, A., Rogozhin, E., Dontsova, G. Y., and Kuchai, M. (2018). Strong earthquakes in 2009–2016 in central italy: Tectonic position, seismic history, and aftershock processes. *Izvestiya, Physics of the Solid Earth*, 54(2):233–251.

Marsan, D. and Lengline, O. (2008). Extending earthquakes' reach through cascading. *Science*, 319(5866):1076–1079.

Mateu, J., Lorenzo, G., and Porcu, E. (2007). Detecting features in spatial point processes with clutter via local indicators of spatial association. *Journal of Computational and Graphical Statistics*, 16(4):968–990.

Mateu, J., Lorenzo, G., and Porcu, E. (2010). Features detection in spatial point processes via multivariate techniques. *Environmetrics*, 21(3-4):400–414.

Mateu, J., Moradi, M., and Cronie, O. (2020). Spatio-temporal point patterns on linear networks: Pseudo-separable intensity estimation. *Spatial Statistics*, 37:100400.

McSwiggan, G., Baddeley, A., and Nair, G. (2017). Kernel density estimation on a linear network. *Scandinavian Journal of Statistics*, 44(2):324–345.

McSwiggan, G., Baddeley, A., and Nair, G. (2020). Estimation of relative risk for events on a linear network. *Statistics and Computing*, 30(2):469–484.

Meyer, P.-A. (1971). Démonstration simplifiée d'un théorème de knight. *Séminaire de probabilités de Strasbourg*, 5:191–195.

Meyer, S., Held, L., and Hohle, M. (2012). Spatio-temporal analysis of epidemic phenomena using the r package surveillance. *Journal of Statistical Software*, 77(11).

Michelini, A., Cianetti, S., Gaviano, S., Giunchi, C., Jozinović, D., and Lauciani, V. (2021). Instance–the italian seismic dataset for machine learning. *Earth System Science Data*, 13(12):5509–5544.

Mohler, G. O., Short, M. B., Brantingham, P. J., Schoenberg, F. P., and Tita, G. E. (2011). Self-exciting point process modeling of crime. *Journal of the American Statistical Association*, 106(493):100–108.

Møller, J. and Ghorbani, M. (2012). Aspects of second-order analysis of structured inhomogeneous spatio-temporal point processes. *Statistica Neerlandica*, 66(4):472–491.

Møller, J., Syversveen, A. R., and Waagepetersen, R. P. (1998). Log gaussian cox processes. *Scandinavian journal of statistics*, 25(3):451–482.

Møller, J. and Toftaker, H. (2014). Geometric anisotropic spatial point pattern analysis and cox processes. *Scandinavian Journal of Statistics*, 41(2):414–435.

Moller, J. and Waagepetersen, R. P. (2003). *Statistical inference and simulation for spatial point processes*. CRC Press.

Moradi, M., Cronie, O., and Mateu, J. (2020). *stlnpp: Spatio-temporal analysis of point patterns on linear networks*. R package version 0.3-7.

Moradi, M., Mateu, J., and Comas, C. (2021). Directional analysis for point patterns on linear networks. *Stat*, 10(1):e323.

Moradi, M. M., Cronie, O., Rubak, E., Lachieze-Rey, R., Mateu, J., and Baddeley, A. (2019). Resample-smoothing of voronoi intensity estimators. *Statistics and computing*, 29(5):995–1010.

Moradi, M. M. and Mateu, J. (2020). First-and second-order characteristics of spatio-temporal point processes on linear networks. *Journal of Computational and Graphical Statistics*, 29(3):432–443.

Moradi, M. M., Rodríguez-Cortés, F. J., and Mateu, J. (2018). On kernel-based intensity estimation of spatial point patterns on linear networks. *Journal of Computational and Graphical Statistics*, 27(2):302–311.

Moraga, P. and Montes, F. (2011). Detection of spatial disease clusters with lisa functions. *Statistics in Medicine*, 30(10):1057–1071.

Morita, Y. (1984). Automatic detection of onset time of seismic waves and its confidence interval using the autoregressive model fitting. *Earthquake*, 37:281–293.

Mourhatch, R. and Krishnan, S. (2020). Simulation of broadband ground motion by superposing high-frequency empirical green's function synthetics on low-frequency spectral-element synthetics. *Geosciences*, 10(9):339.

Mrkvička, T., Dvořák, J., González, J. A., and Mateu, J. (2021). Revisiting the random shift approach for testing in spatial statistics. *Spatial Statistics*, 42:100430.

Muggeo, V. (2008). segmented: An r package to fit regression models with broken-line relationships. *R NEWS*, 8/1:20–25.

Muggeo, V. M. (2003). Estimating regression models with unknown break-points. *Statistics in medicine*, 22(19):3055–3071.

Muggeo, V. M. and Adelfio, G. (2011). Efficient change point detection for genomic sequences of continuous measurements. *Bioinformatics*, 27(2):161–166.

Musmeci, F. and Vere-Jones, D. (1986). A variable-grid algorithm for smoothing clustered data. *Biometrics*, pages 483–494.

Myllymäki, M., Mrkvička, T., Grabarnik, P., Seijo, H., and Hahn, U. (2017). Global envelope tests for spatial processes. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 79(2):381–404.

Navarro-Ruiz, S., Casado-Díaz, A. B., and Ivars-Baidal, J. (2020). Modelling the intra-destination behaviour of cruise visitors based on a three-dimensional approach. *Journal of Destination Marketing & Management*, 18:100470.

Ogata, Y. (1988). Statistical models for earthquake occurrences and residual analysis for point processes. *Journal of the American Statistical Association*, 83(401):9–27.

Ogata, Y. (1989). Statistical model for standard seismicity and detection of anomalies by residual analysis. *Tectonophysics*, 169(1-3):159–174.

Ogata, Y. (1998). Space-time point-process models for earthquake occurrences. *Annals of the Institute of Statistical Mathematics*, 50(2):379–402.

Ogata, Y. and Katsura, K. (1988). Likelihood analysis of spatial inhomogeneity for marked point patterns. *Annals of the Institute of Statistical Mathematics*, 40(1):29–39.

Ogata, Y. and Katsura, K. (1991). Maximum likelihood estimates of the fractal dimension for random spatial patterns. *Biometrika*, pages 463–474.

Okabe, A., Satoh, T., and Sugihara, K. (2009). A kernel density estimation method for networks, its computational method and a gis-based tool. *International Journal of Geographical Information Science*, 23(1):7–32.

Okabe, A. and Sugihara, K. (2012). *Spatial analysis along networks: statistical and computational methods*. John Wiley & Sons.

Oya, P. (2016). Geodynamic assessment of eastern mediterranean region: a joint gravity and seismic b value approach. *Arabian Journal of Geosciences*, 9(5):360.

Padoan, S. A. and Bevilacqua, M. (2015). Analysis of random fields using comprandfld. *Journal of Statistical Software*, 63(1):1–27.

Papazachos, C., Hatzidimitriou, P., Panagiotopoulos, D., and Tsokas, G. (1995). Tomography of the crust and upper mantle in southeast europe. *Journal of Geophysical Research: Solid Earth*, 100(B7):12405–12422.

Park, J., Schoenberg, F. P., Bertozzi, A. L., and Brantingham, P. J. (2021). Investigating clustering and violence interruption in gang-related violent crime data using spatial–temporal point processes with covariates. *Journal of the American Statistical Association*, pages 1–14.

Penttinen, A. and Stoyan, D. (1989). Statistical analysis for a class of line segment processes. *Scandinavian Journal of Statistics*, pages 153–168.

Petry, L. M., Ferrero, C. A., Alvares, L. O., Renso, C., and Bogorny, V. (2019). Towards semantic-aware multiple-aspect trajectory similarity measuring. *Transactions in GIS*, 23(5):960–975.

Pfanzagl, J. (1969). On the measurability and consistency of minimum contrast estimates. *Metrika*, 14(1):249–272.

Policastro, V., D'Angelo, N., and Adelfio, G. (2022). Community detection of seismic point processes. In *SIS2022. ISBN: 9788891932310.*

R Core Team (2022). *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna, Austria.

Raeisi, M., Bonneu, F., and Gabriel, E. (2021). A spatio-temporal multi-scale model for geyer saturation point process: application to forest fire occurrences. *Spatial Statistics*, 41:100492.

Rakshit, S., Baddeley, A., and Nair, G. (2019a). Efficient code for second order analysis of events on a linear network. *Journal of Statistical Software*, 90(1):1–37.

Rakshit, S., Davies, T., Moradi, M. M., McSwiggan, G., Nair, G., Mateu, J., and Baddeley, A. (2019b). Fast kernel smoothing of point patterns on a large network using two-dimensional convolution. *International Statistical Review*, 87(3):531–556.

Rakshit, S., McSwiggan, G., Nair, G., and Baddeley, A. (2021). Variable selection using penalised likelihoods for point patterns on a linear network. *Australian & New Zealand Journal of Statistics*, 64(1):1 – 38.

Rakshit, S., Nair, G., and Baddeley, A. (2017). Second-order analysis of point patterns on a network using any distance metric. *Spatial Statistics*, 22:129–154.

Ramsay, J. O. and Silverman, B. W. (2002). *Applied functional data analysis: methods and case studies*, volume 77. Springer.

Rasmussen, J. G. and Christensen, H. S. (2021). Point processes on directed linear networks. *Methodology and Computing in Applied Probability*, 23(2):647–667.

Reinhart, A. (2018). A review of self-exciting spatio-temporal point processes and their applications. *Statistical Science*, 33(3):299–318.

Ripley, B. D. (1976). The second-order analysis of stationary point processes. *Journal of Applied Probability*, 13:255–266.

Ripley, B. D. (1977). Modelling spatial patterns (with discussion). *Journal of the Royal Statistical Society Series B*, 39(2):172–212.

Ripley, B. D. (1988). *Statistical Inference for Spatial Processes.* Cambridge University Press.

Ripley, B. D. and Kelly, F. P. (1977). Markov point processes. *Journal of the London Mathematical Society*, 2(1):188–192.

Ryan, W. B., Kastens, K. A., and Cita, M. B. (1982). Geological evidence concerning compressional tectonics in the eastern mediterranean. *Tectonophysics*, 86(1-3):213–242.

Schabenberger, O. and Gotway, C. A. (2017). *Statistical methods for spatial data analysis.* Chapman and Hall/CRC.

Scherbaum, F., Johnson, J., and Rietbrock, A. (1999). Programmable interactive toolbox for seismological analysis.

Schlather, M. (2001). On the second-order characteristics of marked point processes. *Bernoulli*, pages 99–117.

Schlather, M., Malinowski, A., Menck, P. J., Oesting, M., and Strokorb, K. (2015). Analysis, simulation and prediction of multivariate random fields with package random fields. *Journal of Statistical Software*, 63(8):1–25.

Schoenberg, F. (1999). Transforming spatial point processes into poisson processes. *Stochastic Processes and their Applications*, 81(2):155–164.

Schoenberg, F. P. (2002). On rescaled poisson processes and the brownian bridge. *Annals of the Institute of Statistical Mathematics*, 54(2):445–457.

Schoenberg, F. P. (2003). Multidimensional residual analysis of point process models for earthquake occurrences. *Journal of the American Statistical Association*, 98(464):789–795.

Schoenberg, F. P. (2016). A note on the consistent estimation of spatial-temporal point process parameters. *Statistica Sinica*, 26:861–879.

Schoenberg, F. P., Hoffmann, M., and Harrigan, R. J. (2019). A recursive point process model for infectious diseases. *Annals of the Institute of Statistical Mathematics*, 71(5):1271–1287.

Sgroi, T., de Nardis, R., and Lavecchia, G. (2012). Crustal structure and seismotectonics of central sicily (southern italy): new constraints from instrumental seismicity. *Geophysical Journal International*, 189(3):1237–1252.

Sheather, S. J. and Jones, M. C. (1991). A reliable data-based bandwidth selection method for kernel density estimation. *Journal of the Royal Statistical Society. Series B (Methodological)*, 53(3):683–690.

Shoval, N., Kahani, A., De Cantis, S., and Ferrante, M. (2020). Impact of incentives on tourist activity in space-time. *Annals of Tourism Research*, 80:102846.

Shoval, N., McKercher, B., Birenboim, A., and Ng, E. (2015). The application of a sequence alignment method to the creation of typologies of tourist activity in time and space. *Environment and Planning B: Planning and Design*, 42(1):76–94.

Siino, M., Adelfio, G., and Mateu, J. (2018a). Joint second-order parameter estimation for spatio-temporal log-gaussian cox processes. *Stochastic environmental research and risk assessment*, 32(12):3525–3539.

Siino, M., Adelfio, G., Mateu, J., Chiodi, M., and D'alessandro, A. (2017). Spatial pattern analysis using hybrid models: an application to the hellenic seismicity. *Stochastic Environmental Research and Risk Assessment*, 31(7):1633–1648.

Siino, M., Rodríguez-Cortés, F. J., Mateu, J., and Adelfio, G. (2018b). Testing for local structure in spatiotemporal point pattern data. *Environmetrics*, 29(5-6):e2463.

Silverman, B. W. (2018). *Density estimation for statistics and data analysis*. Routledge.

Simonetti, A., D'Angelo, N., and Adelfio, G. (2022). Marked hawkes processes for twitter data. In *JADT2022, ISBN: 979-12-80153-31-9*.

Sleeman, R. and Van Eck, T. (1999). Robust automatic p-phase picking: an on-line implementation in the analysis of broadband seismogram recordings. *Physics of the earth and planetary interiors*, 113(1-4):265–275.

Smyth, G. K., Huele, A. F., and Verbyla, A. P. (2001). Exact and approximate reml for heteroscedastic regression. *Statistical modelling*, 1(3):161–175.

Song, J., Wen, R., and Yan, W. (2018). Identification of traffic accident clusters using kulldorff's space-time scan statistics. In *2018 IEEE International Conference on Big Data (Big Data)*, pages 3162–3167. IEEE.

Spakman, W., Wortel, M., and Vlaar, N. (1988). The hellenic subduction zone: a tomographic image and its geodynamic implications. *Geophysical research letters*, 15(1):60–63.

Stoyan, D. and Stoyan, H. (1994). *Fractals, Random Shapes and Point Fields: methods of geometrical statistics*. Wiley, Chichester.

Strauss, D. J. (1975). A model for clustering. *Biometrika*, 62(2):467–475.

Sugihara, K., Satoh, T., and Okabe, A. (2010). Simple and unbiased kernel function for network analysis. In *2010 10th International Symposium on Communications and Information Technologies*, pages 827–832. IEEE.

Takanami, T. and Kitagawa, G. (1988). A new efficient procedure for the estimation of onset times of seismic waves. *Journal of Physics of the Earth*, 36(6):267–290.

Tamayo-Uria, I., Mateu, J., and Diggle, P. J. (2014). Modelling of the spatio-temporal distribution of rat sightings in an urban environment. *Spatial Statistics*, 9:192–206.

Tanaka, U., Ogata, Y., and Stoyan, D. (2008). Parameter estimation and model selection for neyman-scott point processes. *Biometrical Journal: Journal of Mathematical Methods in Biosciences*, 50(1):43–57.

Taymaz, T., Jackson, J., and McKenzie, D. (1991). Active tectonics of the north and central aegean sea. *Geophysical Journal International*, 106(2):433–490.

Van Lieshout, M. (2000). *Markov point processes and their applications*. World Scientific.

Van Lieshout, M. (2006). A j-function for marked point patterns. *Annals of the Institute of Statistical Mathematics*, 58(2):235–259.

Van Lieshout, M. (2018). Nearest-neighbour Markov point processes on graphs with Euclidean edges. *Advances in applied probability*, 50(4):1275–1293.

Van Rossum, G. and Drake Jr, F. L. (1995). *Python reference manual*. Centrum voor Wiskunde en Informatica Amsterdam.

Veen, A. (2006). *Some methods of assessing and estimating point processes models for earthquake occurrences*. University of California, Los Angeles.

Veen, A. and Schoenberg, F. P. (2006). Assessing spatial point process models using weighted k-functions: analysis of california earthquakes. In *Case Studies in Spatial Point Process Modeling*, pages 293–306. Springer.

Vere-Jones, D. (1978). Earthquake prediction-a statistician's view. *Journal of Physics of the Earth*, 26(2):129–146.

Vere-Jones, D. and Schoenberg, F. (2004). Rescaling marked point processes. *Australian & New Zealand Journal of Statistics*, 46:133–143.

Waagepetersen, R. and Guan, Y. (2009). Two-step estimation for inhomogeneous spatial point processes. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 71(3):685–702.

Waagepetersen, R. P. (2007). An estimating function approach to inference for inhomogeneous neyman–scott processes. *Biometrics*, 63(1):252–258.

Walsh, D. C. and Raftery, A. E. (2005). Classification of mixtures of spatial point processes via partial bayes factors. *Journal of Computational and Graphical Statistics*, 14(1):139–154.

Walsh, D. C. I. and Raftery, A. E. (2002). Detecting mines in minefields with linear characteristics. *Technometrics*, 44(1):34–44.

Wand, M. (2020). *KernSmooth: Functions for Kernel Smoothing Supporting Wand & Jones (1995)*. R package version 2.23-17.

Wang, L. and Wang, J. (2006). Change-of-variance problem for linear processes with long memory. *Statistical Papers*, 47(2):279.

Wichern, D. W., Miller, R. B., and Hsu, D.-A. (1976). Changes of variance in first-order autoregressive time series models-with an application. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 25(3):248–256.

Woessner, J., Hardebeck, J., and Haukkson, E. (2010). What is an instrumental seismicity catalog. *Community Online Resource for Statistical Seismicity Analysis*.

Wood, S. N. (2003). Thin plate regression splines. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 65(1):95–114.

Wood, S. N., Goude, Y., and Shaw, S. (2015). Generalized additive models for large data sets. *Journal of the Royal Statistical Society: Series C: Applied Statistics*, pages 139–155.

Wood, S. N., Li, Z., Shaddick, G., and Augustin, N. H. (2017). Generalized additive models for gigadata: modeling the uk black smoke network daily data. *Journal of the American Statistical Association*, 112(519):1199–1210.

Worsley, K. (1979). On the likelihood ratio test for a shift in location of normal populations. *Journal of the American Statistical Association*, 74(366a):365–367.

Xie, Z. and Yan, J. (2008). Kernel density estimation of traffic accidents in a network space. *Computers, environment and urban systems*, 32(5):396–406.

Yang, L., Wu, L., Liu, Y., and Kang, C. (2017). Quantifying tourist behavior patterns by travel motifs and geo-tagged photos from flickr. *ISPRS international journal of geo-information*, 6(11):345.

Ye, X., Yu, J., Wu, L., Li, S., and Li, J. (2015). Open source point process modeling of earthquake. In *Geo-Informatics in Resource Management and Sustainable Ecosystem*, pages 548–557. Springer.

Zenk, S. N., Schulz, A. J., Matthews, S. A., Odoms-Young, A., Wilbur, J., Wegrzyn, L., Gibbs, K., Braunschweig, C., and Stokes, C. (2011). Activity space environment and dietary and physical activity behaviors: a pilot study. *Health & place*, 17(5):1150–1161.

Zhao, W., Tian, Z., and Xia, Z. (2010). Ratio test for variance change point in linear process with long memory. *Statistical Papers*, 51(2):397–407.

Zheng, W., Zhou, R., Zhang, Z., Zhong, Y., Wang, S., Wei, Z., and Ji, H. (2019). Understanding the tourist mobility using gps: How similar are the tourists? *Tourism Management*, 71:54–66.

Zhuang, J. (2006). Second-order residual analysis of spatiotemporal point processes and applications in model evaluation. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 68(4):635–653.

Zhuang, J. (2015). Weighted likelihood estimators for point processes. *Spatial Statistics*, 14:166–178.

Zhuang, J. (2020). Estimation, diagnostics, and extensions of nonparametric hawkes processes with kernel functions. *Japanese Journal of Statistics and Data Science*, 3(1):391–412.

Zhuang, J. and Mateu, J. (2019). A semiparametric spatiotemporal hawkes-type point process model with periodic background for crime data. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 182(3):919–942.

Zhuang, J., Ogata, Y., and Vere-Jones, D. (2002). Stochastic declustering of space-time earthquake occurrences. *Journal of the American Statistical Association*, 97(458):369–380.

Zhuang, J., Ogata, Y., and Vere-Jones, D. (2004). Analyzing earthquake clustering features by using stochastic reconstruction. *Journal of Geophysical Research: Solid Earth*, 109(B5).