

Analysis and Modeling of Spatio-Temporal Point Processes. Information Theory-Based Approaches and Risk Assessment

University of Granada

Doctoral Program in Mathematical and Applied Statistics



Adriana Medialdea Villanueva

Thesis supervised by:

Prof. José Miguel Angulo Ibáñez

Granada 2025

Editor: Universidad de Granada. Tesis Doctorales
Autor: Adriana Medialdea Villanueva
ISBN: 978-84-1195-951-3
URI: <https://hdl.handle.net/10481/108888>

Acknowledgements

First, I would like to express my sincere gratitude to my supervisor, Prof. José Miguel Angulo, for his constant support and guidance throughout my Ph.D. studies and related research. Thanks to him, I developed a strong interest in research and began to explore professional paths I had never previously considered.

I am especially grateful to Prof. Jorge Mateu for inviting me to work with him at the Department of Mathematics at the University Jaume I in Castellón. I truly appreciate the genuine interest he showed in my work, which allowed me to approach my research from new and valuable perspectives. I would also like to thank the colleagues I met during my stay in Castellón, who quickly became close friends. A special thanks to Andre Gilardi for his support with a specific part of my research.

I am also deeply grateful to Prof. Giada Adelfio for welcoming me to the Department of Economics, Business, and Statistical Sciences at the University of Palermo. Her support gave me the opportunity to pursue a European doctorate and to live a truly enriching experience.

I would like to acknowledge the Department of Statistics and Operational Research at the University of Granada, which has provided me with excellent professors, colleagues, and friends. I am also grateful to my colleagues at the Melilla campus for their support. In particular, I would like to thank Marta Castellar, who, having gone through a similar process, offered invaluable advice and assistance during the final stretch of my thesis.

Finally, I would like to express my gratitude to all the people who have shared my life during the development of this thesis and have supported me along the way. In particular, I thank my family - my parents, aunt, and sister - for their constant encouragement and unconditional support, especially during the most difficult times.

Abstract

This thesis work addresses the study of log-Gaussian Cox processes (LGCPs), with a particular focus on their structural characteristics. The research is framed within the context of information and complexity measures, leading to the development of a methodological approach for assessing the goodness of fit of point process models. Furthermore, a fully non-separable model for LGCPs is proposed, and its practical applicability is demonstrated through the generation of dynamic risk maps via conditional simulation.

First, we characterize the contribution to stochasticity of the two phases involved in the generation of an LGCP in relation to the transfer of information from the intensity field to the resulting point pattern. Both global and local analysis are conducted, the first based on entropy and complexity measures, and the latter based on divergence and relative complexity measures. The study is addressed by simulating a number of scenarios with varying parameter values of the Matérn covariance model for the underlying log-intensity field. We assess the sensitivity of the considered information and complexity measures to changes in the covariance parameters as well as in deformation parameters of generalized measures. Benchmark cases, including a homogeneous Poisson process and an LGCP with a white noise intensity, are also analyzed. Results reveal that information and complexity measures are sensitive to changes in structural information transfer, particularly in relation to variations in the mean, variance and scale parameters.

Subsequently, an approach to model LGCPs is proposed, combining a non-separable first-order intensity function to capture large-scale effects with a non-separable correlation structure for the latent random field that outlines small-scale effects, enabling an accurate representation of spatio-temporal dependencies in the process. To evaluate its predictive performance, the model is compared to a separable model and a mixed model that combines a separable first-order intensity with a non-separable covariance structure, by using local and global weighted second-order statistics. We use our modeling strategy

to analyze the behavior of a set of forest fires in Nepal, fitting global trends in the intensity function by using external covariate information. The separable and mixed models yield similar global fit; however, the non-separable model shows improved performance both globally and locally.

To extend the assessment of structural characteristics of LGCPs, a simulation study is conducted under nine different scenarios defined by varying the covariance parameters, each resulting in thousand realizations of point patterns . For each realization, intensity estimation is performed with different degrees of smoothing, also considering a constant intensity as a benchmark case. Shannon entropy, exponential LMC complexity, and Kullback-Leibler divergence are employed to compare the reference and estimated patterns and to study their sensitivity to changes in the covariance parameters, as well as to variations in the cell size of the quadrat scheme. Based on this simulation study, an entropy-based Monte Carlo goodness-of-fit test is proposed, enabling comparisons across multiple spatial resolutions. This test is evaluated through simulations based on an LGCP model, a simple sequential inhibition process model, and a Poisson process model with constant intensity. The test demonstrates strong power in discriminating LGCPs and inhibition processes from the null hypothesis of complete spatial randomness, while maintaining appropriate calibration under the homogeneous Poisson process. Finally, the proposed framework is applied to a spatial dataset of forest fires in Nepal to compare the fitting achieved by a non-parametric kernel density estimator and an LGCP model.

In summary, the results obtained provide relevant information on the fitting and inference of LGCPs, as well as their characterization through information and complexity measures. The applicability of the proposed fitting approach and the goodness-of-fit test is supported by an extensive simulation study and real data applications, demonstrating their practical relevance.

Resumen

Esta memoria de tesis aborda el estudio de los procesos de Cox log-gaussianos (LGCPs), con un enfoque particular en sus características estructurales. La investigación se enmarca en el contexto de las medidas de información y complejidad, conduciendo al desarrollo de una aproximación metodológica para evaluar la bondad de ajuste de modelos de procesos puntuales. Además, se propone un modelo completamente no separable para LGCPs, y se demuestra su aplicabilidad práctica mediante la generación de mapas de riesgo dinámicos a través de simulación condicionada.

En primer lugar, se caracteriza la contribución a la estocasticidad de las dos fases involucradas en la generación de un LGCP en base a la transferencia de información desde el campo de intensidad al patrón de puntos resultante. Se realiza un análisis global y local, el primero basado en medidas de entropía y complejidad, y el segundo basado en medidas de divergencia y complejidad relativa. El estudio se aborda mediante la simulación de varios escenarios con diferentes valores de los parámetros del modelo de covarianza de Matérn para el campo de log-intensidad subyacente. Se evalúa la sensibilidad de las medidas de información y complejidad consideradas ante cambios en los parámetros de covarianza, así como en los parámetros de deformación de medidas generalizadas. También se analizan procesos de referencia, incluyendo un proceso de Poisson homogéneo y un LGCP con intensidad de ruido blanco. Los resultados muestran que las medidas de información y complejidad son sensibles ante cambios en la transferencia de información estructural, en particular, cuando se producen variaciones en los valores de los parámetros de media, varianza y escala.

Posteriormente, se propone un modelo para LGCPs que combina una función de intensidad no separable de primer orden que captura los efectos a gran escala, con una estructura de correlación no separable para el campo aleatorio subyacente que describe los efectos a pequeña escala, permitiendo una representación precisa de las dependencias

espacio-temporales del proceso. Para evaluar el desempeño del modelo, se compara con un modelo separable y un modelo mixto, que combina una intensidad de primer orden separable y una estructura de covarianza no separable, mediante estadísticos ponderados de segundo orden locales y globales. Se emplea esta estrategia de ajuste para analizar el comportamiento de un conjunto de incendios forestales en Nepal, realizando el ajuste de las tendencias globales en la función de intensidad a través de covariables externas. Los modelos separable y mixto ofrecen ajustes similares a nivel global. Asimismo, se observa un mejor desempeño del modelo no separable tanto global como localmente.

Para extender el análisis de las características estructurales de los LGCPs, se realiza un estudio de simulación bajo nueve diferentes escenarios obtenidos a través de la variación de los parámetros de covarianza, resultando cada uno de ellos en mil realizaciones de patrones puntuales. Para cada realización, se estima la intensidad con diferentes grados de suavizado, considerando también una intensidad constante como referencia. Se utilizan la entropía de Shannon, la complejidad LMC exponencial y la divergencia de Kullback-Leibler para realizar comparaciones entre los patrones de referencia y los estimados, estudiando su sensibilidad ante cambios en los parámetros de covarianza así como a variaciones en el tamaño de las celdas de conteo. A partir de este estudio de simulación, se propone un test de bondad de ajuste mediante el método Monte Carlo basado en entropía, que permite realizar comparaciones a través de múltiples resoluciones espaciales. Este test se evalúa mediante simulaciones basadas en un modelo LGCP, un proceso de inhibición secuencial simple y un proceso de Poisson con intensidad constante. El test muestra una potencia significativa al discriminar LGCPs y procesos de inhibición respecto a la hipótesis nula de aleatoriedad espacial completa, manteniendo al mismo tiempo una calibración adecuada bajo un proceso de Poisson homogéneo. Finalmente, se aplica el marco metodológico propuesto a un conjunto de datos espaciales de incendios forestales en Nepal para comparar el ajuste obtenido por un estimador no paramétrico de intensidad tipo núcleo y un modelo LGCP.

En resumen, los resultados obtenidos proporcionan información relevante sobre el ajuste e inferencia de los LGCPs, así como sobre su caracterización mediante medidas de información y complejidad. La aplicabilidad del enfoque de ajuste propuesto y del test de bondad de ajuste se ha comprobado a través de amplio estudio de simulación y la aplicación a datos reales, demostrando su relevancia práctica.

Contents

1	Introduction	1
1	Introducción	7
2	Preliminaries	15
2.1	Random fields	15
2.1.1	Restrictions on the class of random fields	17
2.1.2	Continuity	18
2.1.3	Spatio-temporal random fields	19
2.1.4	Covariance models	20
2.2	Point processes	23
2.2.1	Properties and summary statistics for point processes	24
2.2.2	Estimation of first- and second-order properties	25
2.2.3	Poisson process	30
2.2.4	Cox processes	31
2.2.5	Goodness-of-fit testing for spatial point patterns	34
2.3	Information and complexity measures	36
2.3.1	Entropy measures	37
2.3.2	Complexity measures	38
2.3.3	Divergence and relative complexity measures	39
2.4	Risk measure theory	41
2.4.1	Basic concepts	41
2.4.2	Some relevant risk measure families	42

3	Structural Complexity and Informational Transfer in Spatial Log-Gaussian Cox Processes	45
3.1	Introduction	46
3.2	Methodological aspects	47
3.3	Analysis	52
3.3.1	Weak-structure reference processes	52
3.3.2	Marginal analysis	56
3.3.3	Joint analysis	62
3.4	Synthesis	65
4	An Approach to Model Log-Gaussian Cox Processes with Fully Non-Separable Structures	67
4.1	Introduction	68
4.2	Data: forest fires and meteorological covariates	70
4.3	Statistical model and inference	72
4.3.1	Deterministic component	75
4.3.2	Stochastic component	80
4.4	Goodness-of-fit analysis	81
4.4.1	Diagnosis results	83
4.5	Risk mapping	83
4.6	Synthesis	88
5	A Comparative Tool for Assessing the Goodness of Fit of Spatial Point Process Models	91
5.1	Introduction	92
5.2	Simulation study	93
5.3	Entropy-based Monte Carlo goodness-of-fit test	102
5.4	Data application: forest fires in Nepal	105
5.5	Synthesis	108
6	Results	111
6	Resultados	113

7 Conclusions and Future Research Lines	117
7 Conclusiones y Futuras Líneas de Investigación	121
Bibliography	125
Appendix	135

List of Figures

2.1	Realization of a homogeneous Poisson process with intensity 500 in the unit square.	31
2.2	Realization of an inhomogeneous Poisson process (right), and generating intensity function (left).	31
3.1	Example of quadrat-based counts of a point pattern.	49
3.2	Example of mean intensity raster (right) of an intensity field realization (left).	49
3.3	Mean Shannon entropy values for point patterns obtained with 100 replicates (left) and with only one generation (right) from each intensity random field realization.	51
3.4	Mean Kullback-Leibler divergence values for point patterns obtained with 100 replicates (left) and with only one generation (right) from each intensity random field realization.	51
3.5	Shannon entropy of 100 point patterns and their generating intensity field, with constant $\lambda = 10$ (left) and with white noise based intensity (right). .	53
3.6	Rényi entropy of 100 point patterns and their generating intensity field, with constant $\lambda = 10$ (left) and with white noise based intensity (right). .	54
3.7	LMC complexity of 100 point patterns and their generating intensity field, with constant $\lambda = 10$ (left) and with white noise based intensity (right). .	54
3.8	Kullback-Leibler divergence between 100 point patterns and their generating intensity field, with constant $\lambda = 10$ (red) and with white noise based intensity (blue).	55

3.9	Rényi divergence between 100 point patterns and their generating intensity field, with constant $\lambda = 10$ (left) and with white noise based intensity (right).	55
3.10	Generalized relative complexity between 100 point patterns and their generating intensity field, with constant $\lambda = 10$ (left) and with white noise based intensity (right).	56
3.11	Shannon entropy as a function of the variance, mean, smoothness and scale parameters.	57
3.12	Rényi entropy of log-Gaussian Cox processes as a function of the variance parameter (Scenario 1) and order $\alpha \in \{0, \ln 5, \ln 10, \ln 15, \ln 20\}$	58
3.13	Rényi entropy of log-Gaussian Cox processes as a function of the mean parameter (Scenario 2) and order $\alpha \in \{0, \ln 5, \ln 10, \ln 15, \ln 20\}$	58
3.14	Rényi entropy of log-Gaussian Cox processes as a function of the smoothness parameter (Scenario 3) and order $\alpha \in \{0, \ln 5, \ln 10, \ln 15, \ln 20\}$	59
3.15	Rényi entropy of log-Gaussian Cox processes as a function of the scale parameter (Scenario 4) and order $\alpha \in \{0, \ln 5, \ln 10, \ln 15, \ln 20\}$	59
3.16	Exponential LMC complexity as a function of the variance, mean, smoothness and scale parameters.	60
3.17	Generalized complexity as a function of deformation parameters α and β (with fixed parameter values according to each scenario, see Table 3.1).	61
3.18	Kullback-Leibler divergence between point patterns and intensity random fields as a function of the variance, mean, smoothness and scale parameters.	62
3.19	Rényi divergence between point patterns and random fields as a function of variance, mean, smoothness and scale parameters (Scenarios 1-4) and order $\alpha \in \{\log 0, \log 5, \log 10, \log 15, \log 20\}$	63
3.20	Generalized relative complexity as a function of variance parameter, for $\sigma^2 = 0$ (left), $\sigma^2 = 2.5$ (center) and $\sigma^2 = 5$ (right) and orders $\alpha, \beta \in \{0, 0.6, 1.2, 1.8, 2.4\}$ (fixed parameters as in Scenario 1).	64
3.21	Generalized relative complexity as a function of the mean parameter $\mu = 0$ (left), $\mu = 2$ (center) and $\mu = 4$ (right) and orders $\alpha, \beta \in \{0, 0.6, 1.2, 1.8, 2.4\}$ (fixed parameters as in Scenario 2).	64

3.22	Generalized relative complexity as a function of the smoothness parameter $\nu = 0$ (left), $\nu = 2.5$ (center) and $\nu = 5$ (right) and orders $\alpha, \beta \in \{0, 0.6, 1.2, 1.8, 2.4\}$ (fixed parameters as in Scenario 3).	65
3.23	Generalized relative complexity as a function of the scale parameter $\rho = 1$ (left), $\rho = 5$ (center) and $\rho = 11$ (right) and orders $\alpha, \beta \in \{0, 0.6, 1.2, 1.8, 2.4\}$ (fixed parameters as in Scenario 4).	65
4.1	Spatial locations of forest fires in Nepal by year.	70
4.2	Digital elevation model (DEM) of Nepal.	71
4.3	Time series of the daily number of forest fires in Nepal from February 2012 to June 2016.	71
4.4	Time series of the daily wind speed, precipitation and temperature in Nepal.	73
4.5	Median values of wind speed, precipitation and temperature in each district of Nepal.	74
4.6	Observed daily forest fires (black line) and fitted values (red line) using a generalized linear model.	76
4.7	Predicted spatial intensity from general linear model with Poisson response and meteorological covariates as regressors.	78
4.8	Daily autocorrelation function of forest fires with a lag of 1541 days.	79
4.9	Positive part of ACF (black line) and estimated weight function $w(h, t)$ (orange line).	80
4.10	95% envelope of the spatio-temporal K -function for separable (left) and non-separable (right) covariance functions, with the blue surface representing the estimated K -function for the observed data, and the red surfaces being the 95% envelopes obtained through simulations under the fitted model.	82
4.11	Surfaces of the mean differences between the estimated weighted K -function and the theoretical expected value for the separable (upper right), mixed (bottom left) and non-separable scenario (bottom right), and surface of the differences considering a constant intensity (upper left).	84

4.12	Values of the χ^2 statistic of the differences between the estimated weighted K -function and the theoretical expected value of each simulated field for the non-separable (red line), separable (green line) and mixed (blue line) scenarios.	84
4.13	95% quantile of the χ^2 statistic values of the difference between the estimated local weighted K -function and the theoretical expected value for each simulated field considering separable, mixed and non-separable scenarios (Figure 4.13(a)). Figure 4.13(b) displays the histogram of the χ^2 statistic values considering a constant intensity as the driving intensity field of the process. The vertical red line displays the 95% quantile of the χ^2 statistic values.	85
4.14	Risk maps of the predicted point patterns for the separable (left), mixed (center) and non-separable (right) scenarios from 21-04-2016 to 06-05-2016.	88
5.1	Examples of realizations from a log-Gaussian Cox process (LGCP) for different configurations of the covariance parameters σ^2 and ϕ	94
5.2	Shannon entropy and 95% confidence interval for observed (blue) and estimated patterns (red) for $[10 \times 10]$ (left), $[20 \times 20]$ (center left), $[30 \times 30]$ (center right) and $[50 \times 50]$ (right) quadrat scheme configurations.	98
5.3	Kullback-Leibler divergence and 95% confidence interval for each pair of observed (blue) and estimated patterns (red), for $[10 \times 10]$ (left), $[20 \times 20]$ (center left), $[30 \times 30]$ (center right) and $[50 \times 50]$ (right) quadrat scheme configurations.	99
5.4	Exponential LMC complexity and 95% confidence interval for observed (blue) and estimated patterns (red) for $[10 \times 10]$ (left), $[20 \times 20]$ (center left), $[30 \times 30]$ (center right) and $[50 \times 50]$ (right) quadrat scheme configurations.	100
5.5	Examples of realizations from simple sequential inhibition processes for different configurations of the packing density τ	104
5.6	Forest fires occurred in Nepal from 2012 to 2016.	106
5.7	Estimated intensities by using a kernel density estimator (left) and by considering a log-Gaussian Cox process model (right).	106

5.8	$\hat{H}_1(\mathbf{q}_k)$ (black line), $\bar{H}_1(\mathbf{q}_k)$ (dotted line), and upper and lower envelopes from 100 simulations (red dotted lines).	108
5.9	Histogram of discrepancy values u_i , and u_1 (red line)	108

List of Tables

2.1	Some examples of commonly used kernel smoothing functions.	25
3.1	Parameter configurations for the scenarios considered in the analysis. . .	48
3.2	Table of ‘Inter-Intra’ standard deviations of Shannon entropy and Kullback-Leibler divergence for point patterns generated under Scenario 1, with multiple and single replicates.	50
4.1	Deterministic first-order intensity functions, and covariance models of the Gaussian random fields used to model an LGCP for the separable, mixed and non-separable scenarios.	75
4.2	Estimated coefficients and associated p-values for the fitted generalized linear model for the daily temporal intensity.	77
4.3	Estimated coefficients and p-values of the fitted generalized linear model for the number of forest fires per day in Nepal using the meteorological covariates.	78
4.4	Parameter estimates of the covariance families for the separable and non-separable scenarios.	82
5.1	Parameter configurations for the scenarios considered in the analysis. . .	93
5.2	Significance level of Shannon entropy measure for each quadrat scheme q , intensity function $\hat{\lambda}_\tau, \hat{\lambda}_{2\tau}, \hat{\lambda}_{3\tau}, \hat{\lambda}_c$, and configuration of covariance parameters considered.	101
5.3	Significance level of Kullback-Leibler divergence measure for each quadrat scheme q , intensity function $\hat{\lambda}_\tau, \hat{\lambda}_{2\tau}, \hat{\lambda}_{3\tau}, \hat{\lambda}_c$, and configuration of covariance parameters considered.	101

5.4	Significance level of exponential LMC complexity measure for each quadrat scheme q , intensity function $\hat{\lambda}_\tau, \hat{\lambda}_{2\tau}, \hat{\lambda}_{3\tau}, \hat{\lambda}_c$, and configuration of covariance parameters considered.	101
5.5	Mean values of Shannon entropy indices for each quadrat scheme q , intensity function $\hat{\lambda}_\tau, \hat{\lambda}_{2\tau}, \hat{\lambda}_{3\tau}, \hat{\lambda}_c$, and configuration of covariance parameters considered.	102
5.6	Mean values of exponential LMC complexity indices for each quadrat scheme q , intensity function $\hat{\lambda}_\tau, \hat{\lambda}_{2\tau}, \hat{\lambda}_{3\tau}, \hat{\lambda}_c$, and configuration of covariance parameters considered.	102
5.7	Number of rejections of null hypothesis of complete spatial randomness against an LGCP with parameters σ^2 and ϕ by a 4% test.	104
5.8	Number of rejections of null hypothesis of complete spatial randomness against a simple sequential inhibition process with packing density τ by a 4% test.	105

Chapter 1

Introduction

This thesis work has been developed within the framework of granted projects PGC2018-098860-B-I00 and PID2021-128077NB-I00 from the Spanish Ministry of Science and Innovation, the Spanish State Research Agency, co-funded with FEDER, and A-FQM-345-UGR18 from the Economy, Knowledge, Enterprise and University council of the Andalusia Regional Government, co-funded with FEDER.

The central focus of this research is the study of log-Gaussian Cox processes; in particular, we analyze their structural characteristics within the framework of Information and Complexity measures. Additionally, the study is complemented with the proposal of a fully non-separable model for describing and inferring log-Gaussian Cox processes. Moreover, the theory of risk measures is applied to develop dynamical risk maps representing forest fires patterns based on conditional simulation.

Motivation and background

The theory of point processes rooted in the study of sequences of events occurring randomly in space and/or time. Originally developed in the early 20th century, this theory was closely linked to replacement processes, which represent intervals between successive replacements of system components. Key contributions in this field were made by Feller (1950)¹ and Smith (1958)² in the mid-20th century, who expanded the theory to include renewal processes in demographics and mortality tables, laying the groundwork for analyzing population dynamics over time and space.

One of the earliest examples of mortality tables, a precursor to point process analysis, can be traced back to the work of John Graunt (1662). This set the stage for the concept

of life expectancy. Over time, the theory of point processes evolved, with significant milestones such as the establishment of asymptotic normality in stochastic processes by Ellis (1844)³.

The counting of events across different intervals and regions played a fundamental role in the development of the theory of point processes. The earliest problems of this kind are attributed to Seidel (1876) and Abbé (1879), concerning the occurrence of storms and the number of white blood cells in hemocytometer, respectively. Later, Von Bortkiewicz (1898) conducted a systematic count of events by fitting a Poisson distribution.

Starting in 1936, there was a significant growth in the theory of stochastic processes, driven by the emergence of new applications and the effort to unify the field through the formulation of new theoretical concepts. A landmark contribution was made by Palm (1943)⁴, who generalized the Poisson process, introduced the concept of regeneration points, and proved the first limit theorem for point processes. The work of Palm also marked the first use of the term *point process*.

In recent decades, the development of point process analysis has accelerated, fueled by theoretical advancements and the increasing demand for practical applications in various scientific fields. These processes represent the behavior of structures formed by randomly distributed objects in one or more dimensions, where points represent the location of these objects and additional marks provide further information.

One fundamental type of point process is the Poisson process, which serves as a reference model for more complex processes. A stationary Poisson process is characterized by a constant intensity parameter, indicating the average number of points per unit volume, and the independence of point counts across disjoint regions. When the intensity varies with location, the process becomes non-stationary, and the intensity function plays a crucial role in describing the distribution of points.

Beyond Poisson processes, more complex models, such as Cox processes, account for dependencies between points. These doubly stochastic processes, first introduced by Cox (1955)⁴¹, are useful for modeling clustering phenomena. A Cox process is generated by randomizing the parameters of a Poisson process using a two-stage mechanism involving a random field, leading to a model where the intensity function itself is random.

Risk measures, closely related to stochastic processes, have evolved significantly since their inception. The modern concept of risk, defined as decision-making under uncertainty, emerged alongside the development of probability theory in the 17th and 18th

centuries. Laplace, in his *Essai philosophique sur les probabilités*, first published in 1814, introduced the concept of mathematical expectation, which laid the foundation for risk calculation by combining the probability of an event with its potential impact (see Angulo, 2016¹¹⁸).

Over time, more sophisticated risk measures were developed, including Value-at-Risk (VaR), introduced by JP Morgan in 1980 as a tool for financial risk management. The VaR estimates the minimum potential loss at a given confidence level, but it has limitations, particularly in capturing the magnitude of losses. As a result, alternative measures like Expected Shortfall (ES), also known as Average Value-at-Risk (AVaR), were proposed to address these shortcomings, offering a more comprehensive view of risk considering the tail distribution of losses.

In parallel, a probabilistic theory has been recently developed to unify various risk analysis paradigms across different fields. This theory encompasses both, comprehensive approaches, based on axiomatic principles, and extensive approaches, involving explicit models of risk measures. A key contribution to this theory was made by Artzner et al. (1999)⁵, who introduced the concept of coherent risk measures, characterized by properties such as translation invariance, subadditivity, positive homogeneity, and monotonicity.

Building on this foundation, researchers like Föllmer and Schied (2002)⁶, and Frittelli and Rosazza Gianin (2002)⁷, introduced convex risk measures, which extended the concept of coherence by accounting for non-linear growth in risk. Convex risk measures incorporate a penalty function in their dual representation, offering flexibility in modeling risk.

Among the various families of risk measures, spectral risk measures, proposed by Acerbi (2002)⁸, are particularly noteworthy. These measures incorporate risk aversion through a weighting function, with Expected Shortfall as a notable example. Another important family is entropic risk measures, which integrate elements from Information Theory and are defined using concepts like relative entropy or Kullback-Leibler divergence. These measures are convex and additive for independent variables, and coherent versions have been proposed for more general settings (see Föllmer and Knispel, 2011⁶²).

Information Theory, which originated from communication theory, provides tools like entropy, introduced by Shannon (1948)⁹, which quantifies the uncertainty in a system. Generalizations of Shannon entropy, such as Rényi (1961)¹⁰ and Tsallis (1988)¹¹ entropies, the latter first defined by Havrda and Charvát (1967)¹², offer a broader frame-

work for analyzing uncertainty by including a deformation parameter that acts on the probability distribution. Additionally, concepts like mutual information, which measures the reduction in uncertainty of one variable given another, are fundamental in understanding the dependencies within systems.

This background outlines the evolution of point process theory, risk measures, and their interplay with Information Theory, highlighting key developments and contributions that have shaped current methodologies in analyzing random events and associated risks.

Objectives and methodological aspects

As mentioned above, the primary goal of this research is the study of log-Gaussian Cox processes and their structural aspects. Specifically, this thesis addresses the following objectives:

- Analyze log-Gaussian Cox processes within the framework of Information and Complexity measures, focusing on the structural characteristics.
- Propose a fully non-separable model for describing and inferring log-Gaussian Cox processes.
- Develop a comparative tool based on Information and Complexity measures to assess the goodness of fit of point process models.
- Generate dynamic risk maps to display event concentration and dynamics of hotspots.

To achieve these objectives, various techniques are proposed for spatial and spatio-temporal point processes. Regarding the first objective, the methodological approach focuses on characterizing the transfer of information from the intensity field to the resulting point pattern using global and local analysis. The study employs an empirical method based on lattice box-counting, simulating log-Gaussian Cox processes under different configurations of covariance function and mean parameters. The study examines variations in entropy, complexity and divergence, analyzing both the sensitivity of these measures to changes in the parameters and the coherence between distributions at the cell level.

Regarding the modeling of spatio-temporal log-Gaussian Cox processes, it is common to assume a separable structure for the spatial and temporal dimensions. To model

spatio-temporal dependencies, we propose a model that incorporates a non-separable structure in both the first-order intensity function and the underlying random field. The deterministic component of the intensity function is constructed using a regression model with meteorological covariates for the temporal component and a combination of a non-parametric kernel estimator and spatial regression for the spatio-temporal dimension. A non-separable covariance function is employed, with parameters estimated via a minimum contrast procedure.

The effectiveness of this model is evaluated against other separable models by comparing predictive performance, using second-order statistics to assess the goodness of fit. Dynamic risk maps are generated to display the risk of occurrence of events and the dynamics of hotspots.

Other aspects of the study involve the use of Information and Complexity measures to evaluate the goodness of fit for spatial point process models, particularly log-Gaussian Cox processes. We conduct a simulation study across nine scenarios, each varying the covariance structure. Measures such as Shannon entropy, LMC complexity, and Kullback-Leibler divergence are employed to compare reference and simulated point patterns modeled under different degrees of accuracy, analyzing sensitivity to model parameters and quadrat scheme configuration. The impact of quadrat scheme selection on these measures is emphasized, particularly in relation to model parameters and the level of spatial clustering.

To illustrate the above technical and methodological aspects, they are thoroughly applied to a data set of forest fires occurred in Nepal. In addition, we include meteorological information for the same area and time intervals.

Numerical and graphical analysis has been performed using R language for data representation and function and algorithm implementation. We use several R libraries, such as, `RandomFields`, currently out of maintenance, for random fields realizations; `spatstat` for manipulating and simulating point patterns; `stpp`, which includes estimators of the second-order characteristics of spatio-temporal point patterns, and `lgcp`, specially useful for handling log-Gaussian Cox processes, including Bayesian inference via Markov Chain Monte Carlo.

Outline

This thesis report is structured as follows:

Chapter 2 reviews the preliminary concepts used throughout this thesis, divided into four sections. Section 2.1 covers aspects related to random fields. Section 2.2 introduces concepts related to point processes. Section 2.3 presents Information and Complexity measures. Finally, Section 2.4 discusses some preliminary aspects about Risk Measure Theory.

Chapter 3 conducts a global and local assessment of log-Gaussian Cox processes with the aim of characterizing the transfer of information from the intensity field to the resulting point pattern. This is done by exploring a number of scenarios with varying values of the covariance parameters. The transfer of information and the contribution to stochasticity from the two phases involved in the generation of the pattern are reflected in terms of the entropy, divergence and complexity estimates.

Chapter 4 proposes a model with a double non-separable structure to capture the interaction of spatial and temporal dimensions present in a pattern. We use global and local weighted second-order statistics to assess the improvement in predictive performance compared to traditional models with separable assumptions. Results are displayed using dynamic risk maps based on the Value-at-Risk measure.

Chapter 5 explores the use of Information and Complexity measures as a tool for comparing the goodness of fit of spatial point process models, extending the assessment conducted in Chapter 3 by examining the capability of these measures to capture the spatial structure of simulated patterns. Furthermore, we introduce an entropy-based Monte Carlo goodness-of-fit test, which is subsequently applied to the spatial dataset of forest fires occurrence in Nepal previously analyzed in Chapter 4.

Chapters 6 and 7 discuss key results of this thesis work and outline open lines of further research, including multifractal extensions, predictive risk evaluation, alternative spatial supports, and improved quadrat scheme selection.

Finally, the Appendix collects a selection of the most relevant R codes developed during this research.

Capítulo 1

Introducción

Este trabajo de tesis ha sido desarrollado en el marco de los proyectos PGC2018-098860-B-I00 y PID2021-128077NB-I00 del Ministerio de Ciencia e Innovación y la Agencia Estatal de Investigación, cofinanciado con FEDER y A-FQM-345-UGR18 de la Consejería de Economía, Conocimiento, Empresas y Universidad de la Junta de Andalucía, cofinanciado con FEDER.

El enfoque central de esta investigación es el estudio de los procesos de Cox log-gaussianos; en particular, se analizan sus características estructurales en el marco de las medidas de Información y Complejidad. Además, el estudio se complementa con la propuesta de un modelo completamente no separable para la descripción e inferencia de procesos de Cox log-gaussianos. Asimismo, se aplica la teoría de medidas de riesgo para desarrollar mapas dinámicos de riesgo que representan patrones de incendios forestales basados en simulación condicionada.

Motivación y antecedentes

La teoría de procesos puntuales tiene su origen en el estudio de secuencias de eventos ocurridos aleatoriamente en el espacio y/o el tiempo. Desarrollada inicialmente a comienzos del siglo XX, esta teoría estuvo estrechamente vinculada con los procesos de reemplazo, que representan los intervalos entre reemplazos sucesivos de los componentes de un sistema. Contribuciones clave en este campo fueron realizadas por Feller (1950)¹ y Smith (1958)² a mediados del siglo XX, expandiendo la teoría para incluir los procesos de renovación en demografía y las tablas de mortalidad, estableciendo las bases para el análisis de la dinámica poblacional en el tiempo y en el espacio.

Uno de los primeros ejemplos de tablas de mortalidad, precursoras del análisis de procesos puntuales, se remonta al trabajo de John Graunt (1662), quien introdujo el concepto de esperanza de vida. Con el tiempo, la teoría de procesos puntuales evolucionó, con hitos significativos como el establecimiento de la normalidad asintótica en procesos estocásticos por Ellis (1844)³.

El conteo de eventos en diferentes intervalos y regiones desempeñó un papel fundamental en el desarrollo de la teoría de procesos puntuales. Los primeros problemas de este tipo se atribuyen a Seidel (1876) y Abbé (1879), relacionados con la ocurrencia de tormentas y el número de glóbulos blancos en un hemocitómetro, respectivamente. Posteriormente, Von Bortkiewicz (1898) llevó a cabo un recuento sistemático de eventos mediante el ajuste de una distribución de Poisson.

A partir de 1936, la teoría de procesos estocásticos experimentó un notable crecimiento, impulsado por la introducción de nuevas aplicaciones y al intento de unificar la materia mediante la definición de nuevos conceptos teóricos. Una contribución clave fue la de Palm (1943)⁴, quien generalizó los procesos de Poisson, introdujo el concepto de puntos de regeneración y demostró el primer teorema límite para procesos puntuales. Su trabajo también marcó el uso inicial del término *proceso puntual*.

En las últimas décadas, el desarrollo del análisis de procesos puntuales se ha acelerado gracias a avances teóricos y a una creciente demanda de aplicaciones prácticas en diversas áreas científicas. Estos procesos permiten representar el comportamiento de estructuras formadas por objetos distribuidos aleatoriamente en una o más dimensiones, donde los puntos representan la localización de los objetos y las marcas aportan información adicional.

Un tipo fundamental de proceso puntual es el proceso de Poisson, que sirve como modelo de referencia para procesos más complejos. Un proceso de Poisson estacionario se caracteriza por una intensidad constante, que indica el número medio de puntos por unidad de volumen, y por la independencia del conteo de puntos en regiones disjuntas. Cuando la intensidad varía según la localización, el proceso es no estacionario, y la función de intensidad adquiere un papel clave en la descripción de la distribución de los puntos.

Más allá de los procesos de Poisson, modelos más complejos, como los procesos de Cox, permiten introducir dependencias entre puntos. Estos procesos doblemente estocásticos, fueron introducidos por Cox (1955)⁴¹ y son especialmente útiles para modelar

fenómenos de agrupamiento. Un proceso de Cox se genera aleatorizando los parámetros de un proceso de Poisson mediante un mecanismo de dos etapas que involucra un campo aleatorio, lo que da lugar a un modelo en el que la función de intensidad es, a su vez, aleatoria.

Las medidas de riesgo, estrechamente ligadas a los procesos estocásticos, han evolucionado notablemente desde su origen. El concepto moderno de riesgo, entendido como la toma de decisiones bajo incertidumbre, surgió paralelamente al desarrollo de la teoría de la probabilidad en los siglos XVII y XVIII. Laplace, en su *Essai philosophique sur les probabilités* (1814), introdujo el concepto de esperanza matemática, que sentó las bases para el cálculo del riesgo al combinar la probabilidad de un evento con el efecto de la ocurrencia de ese suceso (Angulo, 2016¹¹⁸).

Más adelante, se desarrollaron medidas de riesgo más sofisticadas, como el Valor en Riesgo (VaR), introducido por JP Morgan en 1980 como herramienta para la gestión del riesgo financiero. El VaR estima la mínima pérdida potencial a un determinado nivel de confianza, aunque presenta limitaciones, especialmente para captar la magnitud de las pérdidas. Por ello, se propusieron medidas alternativas como el Déficit Esperado (ES, *Expected Shortfall*) también llamado Valor en Riesgo Promedio (AVaR, *Average Value-at-Risk*), que ofrece una visión más completa del riesgo al considerar la cola de la distribución de las pérdidas.

Paralelamente, se ha desarrollado recientemente una teoría probabilística que unifica distintos paradigmas del análisis del riesgo. Esta teoría abarca tanto el enfoque comprensivo, basado principios axiomáticos, como el enfoque extensivo, centrado en la formulación de modelos explícitos de medidas de riesgo. Una contribución clave a esta teoría fue la de Artzner *et al.* (1999)⁵, quienes introdujeron el concepto de medidas de riesgo coherentes, caracterizadas por propiedades como la invariancia ante traslaciones, subaditividad, homogeneidad positiva y monotonía.

Sobre esta base, autores como Föllmer y Schied (2002)⁶, y Frittelli y Rosazza Gianin (2002)⁷, introdujeron las medidas de riesgo convexas, que extienden el concepto de coherencia al considerar el crecimiento no lineal del riesgo. Las medidas de riesgo convexas incorporan una función de penalización en su representación dual, ofreciendo mayor flexibilidad en la modelización del riesgo.

Entre las distintas familias de medidas de riesgo, destacan las medidas de riesgo espectrales, propuestas por Acerbi (2002)⁸, que introducen la aversión al riesgo mediante

una función de pesos, siendo un caso particular el Déficit Esperado. Otra familia relevante es la de las medidas de riesgo entrópicas, que integran elementos de la Teoría de la Información y se definen mediante conceptos como la entropía relativa o la divergencia de Kullback-Leibler. Estas medidas son convexas y aditivas para variables independientes, y existen versiones coherentes para contextos más generales (ver Föllmer and Knispel, 2011⁶²).

La Teoría de la Información, originada en el ámbito de la teoría de la comunicación, proporciona herramientas como la entropía, introducida por Shannon (1948)⁹, que cuantifica la incertidumbre en un sistema. Generalizaciones de la entropía de Shannon, como las entropías de Rényi (1961)¹⁰ y Tsallis (1988)¹¹, esta última definida por primera vez por Havrda and Charvát (1967)¹², proporcionan un marco más amplio para analizar la incertidumbre al incluir un parámetro de deformación que actúa sobre la distribución de probabilidad. Asimismo, conceptos como información mutua, que mide la reducción de la incertidumbre de una variable dada otra conocida, son fundamentales para entender las dependencias dentro de un sistema.

Este contexto resume la evolución de la teoría de procesos puntuales, las medidas de riesgo y su relación con la Teoría de la Información, destacando los principales desarrollos y contribuciones que han dado forma a las metodologías actuales en el análisis de eventos aleatorios y riesgos asociados.

Objetivos y aspectos metodológicos

Como se mencionó anteriormente, el objetivo principal de esta investigación es el estudio de los procesos de Cox log-gaussianos y sus aspectos estructurales. En concreto, esta tesis aborda los siguientes objetivos:

- Análisis de los procesos de Cox log-gaussianos en el marco de las medidas de Información y Complejidad, con especial atención en sus características estructurales.
- Propuesta de un modelo completamente no separable para describir y realizar inferencia sobre procesos de Cox log-gaussianos.
- Desarrollo de una herramienta comparativa basada en las medidas de información y complejidad para evaluar la bondad de ajuste de modelos de procesos puntuales.

- Generación de mapas dinámicos de riesgo con el objetivo de reflejar concentraciones de eventos y evaluar puntos críticos.

Para alcanzar estos objetivos, se proponen diversas técnicas aplicables a procesos puntuales espaciales y espacio-temporales. Respecto al primer objetivo, se adopta un enfoque metodológico centrado en caracterizar la transferencia de información desde el campo de intensidad hasta el patrón puntual resultante, utilizando un análisis global y local. El estudio emplea un método empírico basado en el conteo de cajas en una rejilla, simulando procesos de Cox log-gaussianos bajo distintas configuraciones de la función de covarianza y la media. Se examinan variaciones en la entropía, complejidad y divergencia, analizando la sensibilidad de estas medidas a los parámetros y la coherencia entre distribuciones a nivel de celda.

En cuanto a la modelización de procesos de Cox log-gaussianos espacio-temporales, es común asumir una estructura separable entre las dimensiones espacial y temporal. Para ajustar esta dependencia espacio-temporal, se propone un modelo que incorpora una estructura no separable tanto en la función de intensidad de primer orden como en el campo aleatorio subyacente. La componente determinista de la función de intensidad se construye utilizando un modelo de regresión con covariables meteorológicas para la dimensión temporal, y una combinación de un estimador de tipo núcleo no paramétrico y regresión espacial para la dimensión espacio-temporal. Se emplea una función de covarianza no separable, con parámetros estimados mediante el método de mínimos cuadrados.

La eficacia de este modelo para predecir eventos futuros se evalúa frente a otros modelos separables, utilizando estadísticos de segundo orden para evaluar la bondad de ajuste. Los resultados se representan mediante mapas dinámicos de riesgo basados en el Valor en Riesgo.

Otros aspectos del estudio implican el uso de medidas de información y complejidad para evaluar el ajuste de modelos para procesos puntuales espaciales, en particular, de procesos de Cox log-gaussianos. Se realiza un estudio de simulación considerando nueve escenarios mediante la variación de la estructura de covarianza. Se emplean medidas como la entropía de Shannon, la complejidad LMC y la divergencia de Kullback-Leibler para comparar patrones puntuales de referencia y simulados bajo diferentes grados de precisión. Se analiza la sensibilidad de las medidas a los parámetros del modelo y a la configuración de la cuadrícula de conteo, enfatizando el impacto de la selección del

tamaño de cuadrícula, particularmente en relación con los parámetros del modelo y el nivel de agrupamiento espacial.

Para ilustrar los aspectos técnicos y metodológicos mencionados, estos se aplican a un conjunto de datos de incendios forestales localizados en Nepal. Además, se incluye información meteorológica para las mismas áreas y periodos temporales considerados.

El análisis numérico y gráfico se ha realizado utilizando el lenguaje de programación R para la representación de datos y la implementación de funciones y algoritmos. Se han utilizado varias librerías de R, tales como `RandomFields` (actualmente sin mantenimiento), para la simulación de campos aleatorios; `spatstat`, para manipulación y simulación de patrones puntuales; `stpp`, que incluye estimadores de los estadísticos de segundo orden para patrones puntuales espacio-temporales, y `lgcp`, particularmente útil para el manejo de procesos de Cox log-gaussianos e inferencia bayesiana mediante el método de Monte Carlo basado en cadenas de Markov.

Líneas generales

Esta memoria de tesis está estructurado del siguiente modo:

El Capítulo 2 revisa los conceptos preliminares utilizados a lo largo de esta tesis, distribuidos en cuatro secciones. La Sección 2.1 cubre aspectos relacionados con campos aleatorios. La Sección 2.2 introduce conceptos relacionados con procesos puntuales. La Sección 2.3 presenta las medidas de Información y Complejidad. Por último, la Sección 2.4 discute algunos aspectos preliminares sobre la Teoría de Medidas de Riesgo.

El Capítulo 3 realiza una evaluación global y local de los procesos de Cox log-gaussianos con el objetivo de caracterizar la transferencia de información desde el campo de intensidad hasta el patrón de puntos resultante. Esto se realiza mediante la exploración de una serie de escenarios con diferentes valores de los parámetros de covarianza. La transferencia de información y la contribución a la aleatoriedad entre las dos fases involucradas en la generación del patrón se reflejan a través de estimaciones de las medidas de entropía, divergencia y complejidad.

El Capítulo 4 propone un modelo con una estructura doblemente no separable con el objetivo de capturar la interacción entre las dimensiones espacial y temporal presentes en el patrón. Se utiliza una versión con pesos de los estadísticos de segundo orden, tanto globales como locales, para evaluar la mejora en la predicción de eventos futuros en comparación con modelos tradicionales bajo supuestos de separabilidad. Los resultados

se muestran utilizando mapas de riesgo dinámicos basados en la medida del Valor en Riesgo.

El Capítulo 5 explora el uso de medidas de Información y Complejidad como herramienta para comparar la bondad de ajuste de modelos de procesos puntuales espaciales, ampliando el análisis realizado en el Capítulo 3 mediante la evaluación de la capacidad de estas medidas para capturar la estructura espacial de patrones simulados. Además, se introduce un test Monte Carlo de bondad de ajuste basado en entropía, que posteriormente se aplica al conjunto de datos espaciales sobre la ocurrencia de incendios forestales en Nepal, analizado previamente en el Capítulo 4.

En los Capítulos 6 y 7 se discuten los principales resultados de este trabajo de tesis y se proponen futuras líneas de investigación, incluyendo la extensión a procesos multifractales, evaluación predictiva de riesgos, extensión a otros soportes espaciales y mejoras en la selección del esquema de la cuadrícula de conteo.

Finalmente, el Apéndice recopila una selección del código más relevante desarrollado en R durante esta investigación.

Chapter 2

Preliminaries

In this chapter we introduce the fundamental concepts and techniques which are used throughout this thesis.

This chapter is organized as follows. Since the core topic of this thesis is related to the modeling of log-Gaussian Cox processes, Section 2.1 is focused on spatial and spatio-temporal random fields, presenting three covariance models used in the stochastic component of the models fitted in the later chapters. Section 2.2 provides an overview of point processes, including methods for estimating their first and second-order properties, with a particular emphasis on log-Gaussian Cox processes. In Section 2.3, we introduce information and complexity measures, relevant to structural information transfer and goodness-of-fit analysis. Finally, Section 2.4 outlines key concepts and measures useful for risk analysis and risk mapping.

2.1 Random fields

In this section preliminary concepts of random fields are introduced following some references such as Chilès and Delfiner (2012)¹³, Adler and Taylor (2007)¹⁴, Stein (1999)¹⁵, Christakos (1992)¹⁶ and Hristopulos (2020)¹⁷, among others.

Definition 2.1.1. Let $(\Omega, \mathcal{F}, \mathbb{P})$ a probability space, and $D \subset \mathbb{R}^d$ with $d > 1$ a non-empty set. A spatial random field $\{X(\mathbf{s}), \mathbf{s} \in D\}$ is a set of random variables defined on $(\Omega, \mathcal{F}, \mathbb{P})$. For simplicity, we denote the random field by $X(\mathbf{s})$.

The dependence structure of the random variables $X(\mathbf{s})$ can be defined through the joint cumulative distribution function (CDF) as

$$F_{\mathbf{s}_1, \mathbf{s}_2, \dots, \mathbf{s}_n}(x_1, x_2, \dots, x_n) = P[X(\mathbf{s}_1) \leq x_1, X(\mathbf{s}_2) \leq x_2, \dots, X(\mathbf{s}_n) \leq x_n], \quad \forall \mathbf{s}_1, \mathbf{s}_2, \dots, \mathbf{s}_n \in D, \\ n \in \mathbb{N}. \quad (2.1)$$

In the case of continuous variables the joint CDF can be expressed in terms of the joint probability density function as

$$F_{\mathbf{s}_1, \mathbf{s}_2, \dots, \mathbf{s}_n}(x_1, x_2, \dots, x_n) = \int_{-\infty}^{x_1} \int_{-\infty}^{x_2} \dots \int_{-\infty}^{x_n} f_{\mathbf{s}_1, \mathbf{s}_2, \dots, \mathbf{s}_n}(u_1, u_2, \dots, u_n) du_1 du_2 \dots du_n, \\ \forall \mathbf{s}_1, \mathbf{s}_2, \dots, \mathbf{s}_n \in D, \quad n \in \mathbb{N}. \quad (2.2)$$

A necessary and sufficient condition must be satisfied (see Christakos, 1992¹⁶) by the family of probability CDFs (2.1) for the existence of the random field $X(\mathbf{s})$:

a) *Symmetry condition*: It is required that

$$F_{\mathbf{s}_1, \mathbf{s}_2, \dots, \mathbf{s}_n}(x_1, x_2, \dots, x_n) = F_{\mathbf{s}_{i_1}, \mathbf{s}_{i_2}, \dots, \mathbf{s}_{i_n}}(x_{i_1}, x_{i_2}, \dots, x_{i_n}), \quad (2.3)$$

for any permutation $i_1, \dots, i_n$ of the indices $1, \dots, n$.

b) *Consistency condition*: For any $m < n$, it must be satisfied that

$$F_{\mathbf{s}_1, \dots, \mathbf{s}_m, \mathbf{s}_{m+1}, \dots, \mathbf{s}_n}(x_1, \dots, x_m, +\infty, \dots, +\infty) = F_{\mathbf{s}_1, \dots, \mathbf{s}_m}(x_1, \dots, x_m). \quad (2.4)$$

Definition 2.1.2. The mean, variance and marginal non-centered higher-order moment functions of the random field $\{X(\mathbf{s}), \mathbf{s} \in D\}$ are respectively defined as:

a) $\mu(\mathbf{s}) = E[X(\mathbf{s})]$

b) $\sigma^2(\mathbf{s}) = \text{Var}[X(\mathbf{s})]$

c) $\mu_r(\mathbf{s}) = E[X(\mathbf{s})^r], \quad \forall r \geq 1,$

for $\mathbf{s} \in D$, assuming that the expectations involved exists.

Definition 2.1.3. The spatial covariance and correlation functions of the random field $\{X(\mathbf{s}), \mathbf{s} \in D\}$ are respectively defined as

$$C(\mathbf{s}, \mathbf{s}') = E[(X(\mathbf{s}) - \mu(\mathbf{s}))(X(\mathbf{s}') - \mu(\mathbf{s}'))], \text{ and} \quad (2.5)$$

$$\rho(\mathbf{s}, \mathbf{s}') = \frac{C(\mathbf{s}, \mathbf{s}')}{\sigma(\mathbf{s})\sigma(\mathbf{s}')}, \quad (2.6)$$

for $\mathbf{s}, \mathbf{s}' \in D$ and any two random variables $X(\mathbf{s})$ and $X(\mathbf{s}')$, assuming that the expectation involved exists.

Definition 2.1.4. A random field $X(\mathbf{s})$ is said to be *Gaussian* if all its finite-dimensional distributions are multivariate Gaussian.

2.1.1 Restrictions on the class of random fields

In general, the joint CDF $F_{\mathbf{s}_1, \mathbf{s}_2, \dots, \mathbf{s}_n}(x_1, x_2, \dots, x_n)$ depends on the spatial coordinates of each location $\mathbf{s}_i$. In practice, this situation makes it difficult to perform inference about the probability laws characterizing the random field. Consequently, a common assumption is to restrict the class by considering that the probabilistic structure of the process is invariant under translations. This is called stationarity (or, equivalently, homogeneity). For formal simplicity, we consider here $D \equiv \mathbb{R}^d$.

Definition 2.1.5. A random field $X(\mathbf{s})$ is *weakly stationary* or *second-order stationary* if its second-order moment exists and

$$\text{a) } E[X(\mathbf{s})] = \mu, \quad \forall \mathbf{s} \in \mathbb{R}^d$$

$$\text{b) } C(\mathbf{s}, \mathbf{s}') = C(\mathbf{h}), \quad \forall \mathbf{s}, \mathbf{s}' \in \mathbb{R}^d$$

where $\mathbf{h} = \mathbf{s} - \mathbf{s}'$ is the spatial lag.

Property 2.1.1. If we assume that the random field $X(\mathbf{s})$ is weakly stationary, the following properties hold:

$$\text{a) } \textit{Uniform variance: } \text{Var}[X(\mathbf{s})] = \sigma^2, \quad \forall \mathbf{s} \in \mathbb{R}^d$$

b) *Translation invariant correlation function*: $\rho(\mathbf{s}, \mathbf{s}') = \rho(\mathbf{s} - \mathbf{s}') = \rho(\mathbf{h}) \quad \forall \mathbf{s}, \mathbf{s}' \in \mathbb{R}^d$

c) *Related semivariogram and covariance*: $\gamma(\mathbf{s} - \mathbf{s}') = \sigma^2 - C(\mathbf{s} - \mathbf{s}') \quad \forall \mathbf{s}, \mathbf{s}' \in D$,

where $\gamma(\mathbf{s} - \mathbf{s}') := \frac{1}{2}E[(X(\mathbf{s} - \mathbf{s}'))^2]$.

A stricter condition is *strong stationarity*, which involves the probabilistic structure of the random field, assuming that it is translation invariant on space.

Definition 2.1.6. A random field $X(\mathbf{s})$ is *strongly stationary* if, $F_{\mathbf{s}_1, \dots, \mathbf{s}_n}(x_1, \dots, x_n) \equiv F_{\mathbf{s}_1 + \mathbf{h}, \dots, \mathbf{s}_n + \mathbf{h}}(x_1, \dots, x_n)$, for all $\mathbf{s}_1, \dots, \mathbf{s}_n \in \mathbb{R}^d$ and for all $\mathbf{h} \in \mathbb{R}^d$.

Another restriction that may be considered is that the probabilistic structure of the random field is invariant under rotation. In that case, the random field is said to be isotropic.

Definition 2.1.7. A random field $X(\mathbf{s})$ is *weakly isotropic* if there exist a constant μ and a function K on $\mathbb{R}_0^+$ such that

a) $\mu(\mathbf{s}) = \mu, \quad \forall \mathbf{s} \in \mathbb{R}^d$

b) $C(\mathbf{s}, \mathbf{s}') = K(\|\mathbf{s} - \mathbf{s}'\|), \quad \forall \mathbf{s}, \mathbf{s}' \in \mathbb{R}^d$

where $\|\cdot\|$ is the Euclidean norm in $\mathbb{R}^d$.

A stricter condition is considering that the random field is strongly stationary and the finite-dimensional distributions are invariant under isometric transformations.

Definition 2.1.8. A random field $X(\mathbf{s})$ is *strictly isotropic* if for any orthogonal $d \times d$ matrix $\mathbf{H}$, any $\mathbf{k} \in \mathbb{R}^d$, and any $x_1, \dots, x_n \in \mathbb{R}$

$$F_{\mathbf{H}\mathbf{s}_1 + \mathbf{k}, \dots, \mathbf{H}\mathbf{s}_n + \mathbf{k}}(x_1, \dots, x_n) = F_{\mathbf{s}_1, \dots, \mathbf{s}_n}(x_1, \dots, x_n), \quad (2.7)$$

for all $\mathbf{s}_1, \dots, \mathbf{s}_n \in \mathbb{R}^d$ and $n \in \mathbb{N}$.

2.1.2 Continuity

Continuity is a characteristic that refers to the regularity observed in the dynamics of random fields. We expect that in spite of the randomness present in a natural process

it exhibits a certain degree of smoothness in its domain. This characteristic of random fields is expressed formally through local convergence. In that sense, we distinguish two types of continuity: almost sure continuity, related to the probabilistic structure of the field in relation to its realizations, and mean-square continuity, characterized by its first two moments.

Definition 2.1.9. A random field $X(\mathbf{s})$ is *almost surely continuous* at $\mathbf{s}_0$ if

$$P \left[\lim_{\mathbf{s} \rightarrow \mathbf{s}_0} |X(\mathbf{s}) - X(\mathbf{s}_0)| < \epsilon \right] = 1, \quad \forall \epsilon > 0. \quad (2.8)$$

A weaker condition for continuity is defined in terms of the first two moments of the random field.

Definition 2.1.10. A random field $X(\mathbf{s})$ is *mean-square continuous* at $\mathbf{s}_0$ if

$$\lim_{\mathbf{s} \rightarrow \mathbf{s}_0} E \left[\{X(\mathbf{s}) - X(\mathbf{s}_0)\}^2 \right] = 0 \quad (2.9)$$

The following theorems state necessary and sufficient conditions for mean-square continuity in terms of the random field covariance function:

Theorem 2.1.2. A random field $X(\mathbf{s})$ with continuous expectation $\mu(\mathbf{s})$ is mean-square continuous at $\mathbf{s}_0$, if and only if the covariance function $C(\mathbf{s}, \mathbf{s}')$ is continuous as $\mathbf{s} \rightarrow \mathbf{s}_0$ and $\mathbf{s}' \rightarrow \mathbf{s}_0$.

Theorem 2.1.3. A weakly stationary random field is continuous at every point $\mathbf{s} \in D$, if $C(\mathbf{r})$ is continuous at $\mathbf{r} = 0$

Definition 2.1.11. A random field $X(\mathbf{s})$ is almost surely continuous, or mean-square continuous, over D , if it is continuous in the respective sense at every point $\mathbf{s} \in D$.

2.1.3 Spatio-temporal random fields

A spatio-temporal random field is a collection of random variables $X(\mathbf{s}, t)$ defined in $(\Omega, \mathcal{F}, \mathbb{P})$ with $\mathbf{s}, t \in D \times T \subseteq \mathbb{R}^d \times \mathbb{R}$, where $\mathbf{s}$ represents a spatial location and t a time point.

Restrictions on the class of spatio-temporal random fields

Some restrictions on the class of spatial random fields described in Section 2.1.1 can be extended in a straightforward way to the spatio-temporal case. For instance:

Definition 2.1.12. A random field $X(\mathbf{s}, t)$, $(\mathbf{s}, t) \in \mathbb{R}^d \times \mathbb{R}$, is *weakly stationary* if the following conditions hold:

- a) $E[X(\mathbf{s}, t)] = \mu$, $\forall (\mathbf{s}, t) \in \mathbb{R}^d \times \mathbb{R}$
- b) $C(\mathbf{s}, t; \mathbf{s}', t') = C(\mathbf{s} - \mathbf{s}'; t - t') = C(\mathbf{h}, \tau)$, $\forall (\mathbf{s}, t), (\mathbf{s}', t') \in \mathbb{R}^d \times \mathbb{R}$

Definition 2.1.13. A random field $X(\mathbf{s}, t)$, $(\mathbf{s}, t) \in \mathbb{R}^d \times \mathbb{R}$ is *strongly stationary* if all its finite-dimensional probability distributions are invariant when the space-time points are translated by an arbitrary vector $(\mathbf{h}, \tau) \in \mathbb{R}^d \times \mathbb{R}$, *i.e.*,

$$F_{(\mathbf{s}_1, t_1), \dots, (\mathbf{s}_n, t_n)}(x_1, \dots, x_n) = F_{(\mathbf{s}_1 + \mathbf{h}, t_1 + \tau), \dots, (\mathbf{s}_n + \mathbf{h}, t_n + \tau)}(x_1, \dots, x_n), \quad \forall (\mathbf{s}_1, t_1), \dots, (\mathbf{s}_n, t_n) \in \mathbb{R}^d \times \mathbb{R}, \\ \forall (\mathbf{h}, \tau) \in \mathbb{R}^d \times \mathbb{R}, \quad (2.10)$$

2.1.4 Covariance models

In this section, some stationary models of covariance functions used in the next chapters are presented. In particular, the isotropic Whittle-Matérn spatial covariance model introduced by Matérn (1986)¹⁸, the separable spatio-temporal exponential covariance model, and two non-separable spatio-temporal models proposed by Delaco et al (2002)¹⁹ and Gneiting (2002)²⁰.

Let $X(\mathbf{s}, t)$, with $\mathbf{s} \in \mathbb{R}^d$ and $t \in \mathbb{R}$, denote a strictly stationary spatio-temporal random field. Its covariance function is given by $C(\mathbf{s}, t; \mathbf{s}', t') = C(\mathbf{s} - \mathbf{s}'; t - t') = C(\mathbf{h}, u)$.

A commonly adopted simplifying assumption for covariance models is space-time separability, namely that there is no interaction between the spatial and temporal dimensions. In that sense, the covariance function can be factorized (see, *e.g.*, Rodríguez-Iturbe and Mejía, 1974²¹), considering stationarity, as

$$C(\mathbf{h}, u) = C_S(\mathbf{h}) \cdot C_T(u), \quad \forall (\mathbf{h}, u) \in \mathbb{R}^d \times \mathbb{R}, \quad (2.11)$$

where $C_S(\cdot)$ and $C_T(\cdot)$ are purely spatial and purely temporal covariance functions, respectively. For the spatial component, isotropic models such as the exponential class

(Diggle, Tawn and Moyeed, 1998²²) or the Whittle-Matérn class (Whittle, 1954²³; Matérn 1986¹⁸) are frequently employed.

Another decomposition of the space-time covariance considered in the literature (see, *e.g.*, Myers and Journel, 1990²⁴) is based on the additive form

$$C(\mathbf{h}, u) = C_S(\mathbf{h}) + C_T(u), \quad \forall (\mathbf{h}, u) \in \mathbb{R}^d \times \mathbb{R}. \quad (2.12)$$

This model implies that the spatial behavior of the random field is invariant over time and, conversely, the temporal dynamics are location-independent. However, as mentioned by Cressie and Huang (1999)²⁶, this structure can lead to singular covariance matrices under certain configurations of space-time data.

Another key assumption is full symmetry (see Gneiting *et al.*, 2006⁹⁰). A stationary covariance function $C(\mathbf{h}, u)$ is called fully symmetric if it is spatially isotropic and temporally symmetric, i.e.

$$C(\mathbf{h}, u) = C(\mathbf{h}, -u) = C(-\mathbf{h}, u) = C(-\mathbf{h}, -u), \quad \forall (\mathbf{h}, u) \in \mathbb{R}^d \times \mathbb{R}. \quad (2.13)$$

In such case, the covariance function depends only on the distances between points. Note that separable models inherently possess full symmetry; however, the contrary does not hold.

Besides the simplicity of separable models as in Equation (2.11), they cannot describe properly the space-time interaction present in most real phenomena as the spatial and temporal components do not interact. To overcome this limitation, non-separable covariance models have been proposed, which allow for more flexible interactions between space and time. Non-separable, fully symmetric stationary covariance functions can be constructed as mixtures of separable covariances, with one of the simplest cases being the product-sum model proposed by De Cesare *et al.* (2001)²⁵:

$$C(\mathbf{h}, u) = a_0 C_S(\mathbf{h}) C_T(u) + a_1 C_S(\mathbf{h}) + a_2 C_T(u), \quad (\mathbf{h}, u) \in \mathbb{R}^d \times \mathbb{R}, \quad (2.14)$$

where $a_1 > 0$ and $a_2, a_3 \geq 0$, and C_S, C_T are purely spatial and purely temporal stationary covariance functions, respectively.

Whittle-Matérn model

The Whittle-Matérn is a flexible spatial class, defined by the homogeneous and isotropic covariance function

$$C(h) = \frac{\sigma^2}{2^{\nu-1}\Gamma(\nu)} \left(\frac{2\nu^{1/2}h}{\rho} \right)^\nu K_\nu \left(\frac{2\nu^{1/2}h}{\rho} \right), \quad (2.15)$$

where K_ν is the modified Bessel function of the second kind, $\sigma^2 \geq 0$ is the variance of the field, and $\nu > 0$ and $\rho > 0$ respectively represent the smoothness and scale parameters. A particular case within the Matérn class is the exponential covariance function, which can be obtained by setting $\nu = 1/2$.

Double exponential model

A very common separable covariance structure, assumed in Brix and Diggle (2001)⁴³, is constructed by the product of spatial and temporal exponential components,

$$C(h, u) = \sigma^2 \exp\left(-\frac{h}{\alpha}\right) \exp\left(-\frac{u}{\beta}\right), \quad (2.16)$$

where σ^2 is the variance parameter and α and β are, respectively, spatial and temporal scale parameters.

DeIaco-Cesare model

The DeIaco-Cesare family is obtained as a mixture of separate spatial and temporal covariances, written as

$$C(h, u) = \sigma^2 \left(1 + \left(\frac{h}{\alpha} \right)^{\gamma_s} + \left(\frac{u}{\beta} \right)^{\gamma_t} \right)^{-\delta}, \quad (2.17)$$

where $\alpha > 0$ and $\beta > 0$ are scale parameters of space and time, respectively, $\gamma_s \in (0, 2]$, $\gamma_t \in (0, 2]$ and, $\delta \geq 1$ determines the degree of interaction of the spatial and temporal dimensions.

Gneiting model

The Gneiting family is a very general class of non-separable space-time covariance models. It was introduced by Gneiting (2002)²⁰ based on the approach by Cressie and Huang (1999)²⁶.

Theorem 2.1.4. Let $\varphi(r)$, $r \geq 0$, be a completely monotone function and $\psi(r)$, $r \geq 0$, a positive function with a completely monotone derivative. Then, the expression

$$C(\mathbf{h}, u) = (\psi(u) + 1)^{-d/2} \varphi\left(\frac{\|\mathbf{h}\|}{\sqrt{\psi(u) + 1}}\right), \quad (\mathbf{h}, u) \in \mathbb{R}^d \times \mathbb{R} \quad (2.18)$$

defines a stationary covariance function.

Setting $d = 2$, $\varphi(h) = \sigma^2 \exp\left(-\left(\frac{h}{\alpha}\right)^{\gamma_s}\right)$ and $\psi(u) = \left(\left(\frac{u}{\beta}\right)^{\gamma_t} + 1\right)^{\delta/\gamma_t} - 1$ in (2.18) yields the parametric family,

$$C(h, u) = \frac{\sigma^2}{\left(\left(\frac{u}{\beta}\right)^{\gamma_t} + 1\right)^{\delta/\gamma_t}} \exp\left(-\frac{\left(\frac{h}{\alpha}\right)^{\gamma_s}}{\left(\left(\frac{u}{\beta}\right)^{\gamma_t} + 1\right)^{\delta/(2\gamma_t)}}\right) \quad (2.19)$$

where $\alpha, \beta > 0$ are spatial and temporal scale parameters, respectively, $\delta \in (0, 2]$, and σ^2 is the variance.

2.2 Point processes

A *spatial point pattern* is a set of events irregularly distributed within a region, being this set characterized by their observed spatial locations. Similarly, in a *spatio-temporal point pattern* the data is distributed within a spatial region and a time interval. Usually, the spatial region is a planar space, yet the point could be also distributed in spaces of lower or higher dimension than two or in different supports such as networks.

The dataset of locations that conform the point pattern can include additional information attached to each point, this attributes are called *marks* and the pattern is a *marked point pattern*. A spatio-temporal point pattern could be seen as a marked spatial point pattern whose marks are the times of occurrence of each event, or we can consider time as another dimension changing our support from two to three dimensions.

In order to describe and analyze the behavior of a point pattern we need the concept of a point process. A *point process* X is a stochastic mechanism that generates a countable set of events x_i in a pre-defined region.

2.2.1 Properties and summary statistics for point processes

Let $n(D)$ be the number of events in the region D with $|D|$ its area, $d\mathbf{x}$ the infinitesimal region that contains the point $\mathbf{x}$ and $|d\mathbf{x}|$ its area. We define the next *first-order* and *second-order* properties of a spatial point process:

- *Intensity function*: $\lambda(\mathbf{x}) = \lim_{|d\mathbf{x}| \rightarrow 0} \left(\frac{E[n(d\mathbf{x})]}{|d\mathbf{x}|} \right)$
- *Second-order intensity function*: $\lambda(\mathbf{x}, \mathbf{x}') = \lim_{|d\mathbf{x}|, |d\mathbf{x}'| \rightarrow 0} \left(\frac{E[n(d\mathbf{x})n(d\mathbf{x}')]}{|d\mathbf{x}||d\mathbf{x}'|} \right)$
- *Conditional intensity function*: $\lambda(\mathbf{x}|\mathbf{x}') = \lambda(\mathbf{x}, \mathbf{x}') / \lambda(\mathbf{x}')$

If we consider a stationary and isotropic process, the intensity function $\lambda(\mathbf{x})$ is a constant λ and the second-order intensity depends only on the distance $d(\mathbf{x}, \mathbf{x}') = \|\mathbf{x} - \mathbf{x}'\|$, *i.e.* $\lambda(\mathbf{x}, \mathbf{x}') = \lambda(\|\mathbf{x} - \mathbf{x}'\|)$.

A very commonly used statistic for analyzing the spatial correlation present in a stationary point pattern is the *K-function* proposed by Ripley (1977)²⁷. The $K(r)$ function of a point process X is defined as the expected number of r -neighbours of an arbitrary point of X divided by the intensity λ of the process,

$$K(r) = \lambda^{-1} E[n_0(r)], \quad (2.20)$$

where $n_0(r)$ is the number of extra events that lie within a distance r of an arbitrary point.

An alternative statistic for the K -function is the *pair correlation function* $g(r)$, defined as

$$g(r) = \frac{K'(r)}{2\pi r}, \quad (2.21)$$

with $K'(r)$ being the derivative of $K(r)$ with respect to r . The main difference between $K(r)$ and $g(r)$, is that the former is a cumulative function that takes its value from all the interpoint distances less than or equal to that r whilst $g(r)$ takes values from the points lying at distances equal to r .

2.2.2 Estimation of first- and second-order properties

First-order properties

The obvious estimator for the overall intensity λ is the observed number of points per unit area, $\hat{\lambda} = n/|A|$. However, when inhomogeneity is assumed the most usual non-parametric techniques used as a summary description of the intensity of the process are *quadrat counting* and *kernel estimation*. The former entails counting the events within sub-regions A of the domain D , resulting in the *quadrat count distribution* of the process,

$$p_n(A) = P[n(A) = n], \quad n = 0, 1, \dots \quad (2.22)$$

With regard to the *kernel estimation* of the intensity function $\lambda(\mathbf{x})$, a well known estimator proposed by Diggle (1985)²⁸ takes the form

$$\hat{\lambda}_\tau(\mathbf{x}) = \kappa_\tau(\|\mathbf{x} - \mathbf{x}_i\|) e(\mathbf{x}_i), \quad (2.23)$$

for a arbitrary location $\mathbf{x} \in D$, with $\kappa_\tau(\mathbf{x}) = \tau^{-1} \kappa(\tau^{-1}u)$ being a pdf symmetric about the origin with τ the smoothing bandwidth and $e(\mathbf{x}_i)$ an edge correction factor. The successful performance of the estimator relies on the selection of the bandwidth whereas the choice of the kernel function $\kappa(u)$ (see Table 2.1 for examples) is less important.

Kernel	$\kappa(u)$	Support
Uniform	$\frac{1}{2}$	$[-1, 1]$
Gaussian	$\frac{1}{\sqrt{(2\pi)}} \exp - (1/2) u^2$	$\mathbb{R}$
Epanechnikov	$\frac{3}{4} (1 - u^2)$	$[-1, 1]$
Quartic	$\frac{15}{16} (1 - u^2)^2$	$[-1, 1]$
Triangular	$(1 - u)$	$[-1, 1]$

Table 2.1: Some examples of commonly used kernel smoothing functions.

The smoothing bandwidth τ controls the degree of smoothing performed by the kernel. Various authors have proposed simple approaches for selecting the bandwidth pa-

parameter. Baddeley et al. (2015)²⁹ proposed taking τ equal to one-eighth of the shortest side length of the enclosing rectangle, while in the Scott's rule of thumb (Scott, 1992³⁰) states that τ is proportional to $n^{-1/(d+4)}$, where n is the number of points and d is the spatial dimensions.

Apart from some rules of thumb, there are other approaches for selecting the value of τ , such as likelihood cross-validation, minimization of the mean-square error, and the method proposed by Cronie and Van Lieshout (2018)³¹ that relies on Campbell's formula.

Loader (1999)³⁵ proposed selecting the bandwidth τ to maximize the log-likelihood function

$$LCV(\tau) = \sum_{i=1}^n \log(\hat{\lambda}(\mathbf{x}_i)) - n \left(\int_D \hat{\lambda}(\mathbf{u}) d\mathbf{u} - 1 \right), \quad (2.24)$$

where $\hat{\lambda}(\mathbf{x}_i)$ is the density estimate at $\mathbf{x}_i$ after removing this observation from the data.

Diggle (1985)²⁸, in the context of the log-Gaussian Cox process (see Section 2.2.4) and assuming stationarity, proposed the use of cross-validation, considering the mean-square error of $\hat{\lambda}(\mathbf{x})$ with respect to the distribution of the intensity process $\Lambda(\mathbf{x})$,

$$MSE(\tau) = E \left[\hat{\lambda}(\mathbf{x}) - \Lambda(\mathbf{x}) \right]^2, \quad (2.25)$$

where the bandwidth τ is chosen to minimize $MSE(\tau)$ by computing the quantity

$$M(\tau) = \frac{MSE(\tau)}{\lambda^2} - g(0) \quad (2.26)$$

as a function of τ , with λ the mean intensity and g the pair correlation function.

Cronie and Van Lieshout (2018)³¹ proposed an optimality criterion based on Campbell's formula where the bandwidth τ is chosen to minimise the squared difference between the area of the observation window and the sum of reciprocal estimated intensity function values at point locations,

$$CvL(\tau) = \left(|W| - \sum_i 1/\hat{\lambda}(\mathbf{x}_i) \right)^2. \quad (2.27)$$

Edge correction methods for kernel density estimation

Kernel density estimators have difficulties for accurately estimating the intensity near the boundaries of the observation domain, resulting in a bias called *edge effect*. This bias arises when the interpoint distance u_{ij} , computed as $u_{ij} = \sqrt{(\mathbf{x}_i - \mathbf{x}_j)^2 + (\mathbf{y}_i - \mathbf{y}_j)^2}$, exceeds the distance between the point i and the nearest boundary. When this occurs, the density near the edges becomes underestimated, as the kernel function extends beyond the boundary where no data is observed. Various correction methods for dealing with this bias have been proposed.

The classical approach to correcting edge effects involves the use of *edge kernels* near the boundary (for more details on edge kernels, see Gasser, 1979³²). These edge kernels are designed to taper the influence of the kernel function as it approaches the boundary, this compensating for the lack of data outside the observation domain. However, this strategy has fallen out of use in favor of newer, less complex methods.

One of such alternatives is the edge correction method proposed by Diggle (1985)²⁸, where the intensity estimate is corrected by dividing it by the convolution of the kernel with the observation domain. This leads to the following estimator:

$$\hat{\lambda}(\mathbf{x}) = \frac{1}{e(\mathbf{x})} \sum_{i=1}^n \kappa(\mathbf{x} - \mathbf{x}_i) \quad (2.28)$$

with $e(\mathbf{x}) = \int_D k(\mathbf{x} - \mathbf{y}) d\mathbf{y}$.

Another method was proposed by Jones (1993)³³ and further refined by Diggle (2010)³⁴. This method, known as Jones-Diggle edge correction method, introduces an estimator that adjusts for the edge effect at each data point,

$$\hat{\lambda}(\mathbf{x}) = \sum_{i=1}^n \frac{\kappa(\mathbf{x} - \mathbf{x}_i)}{e(\mathbf{x}_i)} \quad (2.29)$$

While the Jones-Diggle method has been shown to perform better in terms of bias reduction, the computational complexity is generally higher compared to Diggle correction.

Second-order properties

For the second-order properties, we first present a simple estimator for the K -function (2.20) defined in Section 2.2.1. Let X be a point process on D and let $\{\mathbf{x}_1, \dots, \mathbf{x}_n\}$ be a realization of the process. The estimator of $K(r)$ is defined as

$$\hat{K}(r) = \hat{\lambda}^{-1} \sum_{i=1}^n \sum_{j \neq i} \mathbf{I}(\|\mathbf{x}_i - \mathbf{x}_j\| \leq r) / n, \quad (2.30)$$

where $\hat{\lambda}$ is the estimator of the overall intensity of the process and $\mathbf{I}(\cdot)$ is the indicator function.

We can also consider the spatio-temporal version of (2.30). For a point process defined on $D \times T$ and a pair of spatial and temporal distances (r, h) , the estimator of the spatio-temporal Ripley's K -function is defined as

$$\hat{K}(r, h) = \frac{1}{|D \times T| \hat{\lambda}^2} \sum_{i=1}^n \sum_{j \neq i} \mathbf{I}[\|\mathbf{x}_i - \mathbf{x}_j\| \leq r, |t_i - t_j| \leq h]. \quad (2.31)$$

Regarding the local structure of the process (see Getis, 1984³⁶), given a specific location $\mathbf{x}_i$ and a time t_i , the local version of (2.31) can be written as

$$\hat{K}_i(r, h) = \frac{1}{|D \times T| \hat{\lambda}^2} \sum_{\substack{j=1 \\ j \neq i}}^n \mathbf{I}[\|\mathbf{x}_i - \mathbf{x}_j\| \leq r, |t_i - t_j| \leq h]. \quad (2.32)$$

Many procedures have been proposed for obtaining the estimator $\hat{g}(r, h)$ of the pair correlation function, in most of which the estimate of $g(r, h)$ is obtained by replacing the indicator function in $\hat{K}$ with a kernel $\kappa(\cdot)$, see Stoyan and Stoyan (1994)³⁷. Using this result in (2.21) the estimator takes the form

$$\hat{g}(r, h) = \frac{|W|}{4\pi r \hat{\lambda}^2} \sum_{i=1}^n \sum_{\substack{j=1 \\ j \neq i}}^n \kappa_\epsilon(\|\mathbf{x}_i - \mathbf{x}_j\| - r) \kappa'_\delta(|t_i - t_j| - h), \quad (2.33)$$

where κ_ϵ and κ'_δ are spatial and temporal kernel function with bandwidths ϵ and δ , respectively.

The non-parametric statistics defined above rely on the assumption that the point pattern is homogeneous. Baddeley *et al.* (2000)³⁸ extended Ripley's K -function (2.30)

for the class of non-stationary processes. Its estimator is given by

$$\hat{K}(r) = \frac{1}{|D|} \sum_{i=1}^n \sum_{j \neq i} \frac{w_{ij} \mathbf{I}(\|\mathbf{x}_i - \mathbf{x}_j\| \leq r)}{\lambda(\mathbf{x}_i) \lambda(\mathbf{x}_j)}. \quad (2.34)$$

Gabriel and Diggle (2009)³⁹ extended $\hat{K}$ and $\hat{g}$ statistics for inhomogeneous point patterns to the spatio-temporal domain. The estimator of the pair correlation function can be expressed as

$$\hat{g}(r, h) = \frac{1}{4\pi r} \sum_{i=1}^n \sum_{j \neq i} \frac{1}{w_{ij}} \frac{k_\epsilon(r - \|\mathbf{x}_i - \mathbf{x}_j\|) k_\delta(h - |t_i - t_j|)}{\hat{\lambda}(\mathbf{x}_i, t_i) \hat{\lambda}(\mathbf{x}_j, t_j)}, \quad (2.35)$$

where w_{ij} is an edge correction factor to deal with spatio-temporal edge effects and k_ϵ , k_δ are kernel functions with bandwidths ϵ and δ respectively.

The estimator of the inhomogeneous K -function for the spatio-temporal domain is expressed as

$$\hat{K}(r, h) = \frac{1}{|D \times T|} \sum_{i=1}^n \sum_{j \neq i} \frac{1}{w_{ij}} \frac{\mathbf{I}[\|\mathbf{x}_i - \mathbf{x}_j\| \leq r, |t_i - t_j| \leq h]}{\hat{\lambda}(\mathbf{x}_i, t_i) \hat{\lambda}(\mathbf{x}_j, t_j)}. \quad (2.36)$$

Moving towards local features, the inhomogeneous version of 2.32 for a generic event $(\mathbf{x}_i, t_i)$, is expressed as follows

$$\hat{K}(r, h) = \frac{1}{|D \times T|} \sum_{j \neq i}^n \frac{1}{w_{ij}} \frac{\mathbf{I}[\|\mathbf{x}_i - \mathbf{x}_j\| \leq r, |t_i - t_j| \leq h]}{\hat{\lambda}(\mathbf{x}_i, t_i) \hat{\lambda}(\mathbf{x}_j, t_j)}. \quad (2.37)$$

Baddeley et al. (2000)³⁸ introduced a weighted version of Ripley's K -function which is particularly useful for assessing the goodness of fit of point process models (see Veen, 2005¹¹⁹). For evaluating a given model λ_0 , it is defined as

$$\hat{K}^w(r, h) = \frac{1}{\lambda_{\inf}^2 |D \times T|} \sum_{i=1}^n \sum_{j \neq i} \frac{\mathbf{I}[\|\mathbf{x}_i - \mathbf{x}_j\| \leq r, |t_i - t_j| \leq h]}{\lambda(\mathbf{x}_i, t_i) \lambda(\mathbf{x}_j, t_j)}, \quad (2.38)$$

where $\lambda_{\inf} := \inf \{\lambda_0(\mathbf{x}, t); (\mathbf{x}, t) \in D \times T\}$ is the infimum of the conditional intensity over the observed region for the model to be assessed. Its local version, for the generic event $(\mathbf{x}_i, t_i)$ is given by (Adelfio et al., 2020⁴⁰)

$$\hat{K}_i^w(r, h) = \frac{1}{\lambda_{\inf}^2 |D \times T|} \sum_{j \neq i}^n \frac{\mathbf{I}[\|\mathbf{x}_i - \mathbf{x}_j\| \leq r, |t_i - t_j| \leq h]}{\lambda(\mathbf{x}_i, t_i) \lambda(\mathbf{x}_j, t_j)}. \quad (2.39)$$

2.2.3 Poisson process

Homogeneous Poisson process

The Poisson process is the most elementary stochastic mechanism that generates completely random point patterns. This process is characterized by its homogeneity and spatial independence. The importance of this process lies in that it can be used as a benchmark case for completely random patterns. The Poisson process can be characterized by the following properties:

Property 2.2.1. For some $\lambda > 0$ and a bounded spatial region D , $n(D)$ follows a Poisson distribution with mean $\lambda|D|$.

Property 2.2.2. If $n(D) = n_0$, the n_0 events in D constitute an independent random sample with uniform distribution on D .

Property 2.2.3. For $A \cap B = \emptyset$, the random variables $n(A)$ and $n(B)$ are independent.

Inhomogeneous Poisson process

The class of *inhomogeneous Poisson processes* is obtained when the intensity of a Poisson process is a spatially varying intensity function $\lambda(\mathbf{x})$. The next properties characterize this class of processes:

Property 2.2.4. $n(D)$ follows a Poisson distribution with mean $\int_D \lambda(\mathbf{x}) d\mathbf{x}$

Property 2.2.5. If $n(D) = n_0$, the n_0 events in D constitute an independent random sample from the distribution on D whose probability density function is proportional to $\lambda(\mathbf{x})$.

Consequently, the intensity function determines the number of events and their spatial distribution. This type of process can be described and analyzed by the introduction of covariates through the intensity function, $\lambda(\mathbf{x}) = \lambda(z_1(\mathbf{x}), \dots, z_k(\mathbf{x}))$.

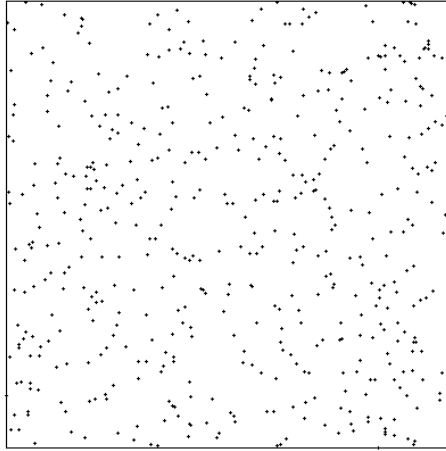


Figure 2.1: Realization of a homogeneous Poisson process with intensity 500 in the unit square.

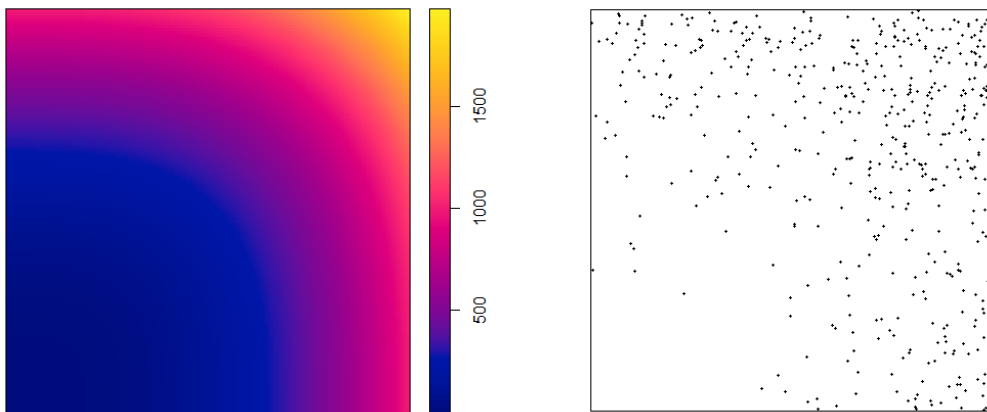


Figure 2.2: Realization of an inhomogeneous Poisson process (right), and generating intensity function (left).

2.2.4 Cox processes

The *doubly stochastic Poisson process*, also known as *Cox process*, so named by the seminal paper of Cox (1955)⁴¹, is a flexible model that allows the inclusion of spatial dependence between events, making it suitable for modeling clustered point patterns. This type of process is an extension of the Poisson process by introducing a stochastic intensity function, which incorporates environmental heterogeneity, leading to aggregation. However, observed aggregation in a pattern can arise from various sources and may lead to different types of processes. For instance, cluster processes involve an unseen

point pattern acting as an ancestor for generating the observed clusters (for more details on cluster processes, see Neyman and Scott, 1958⁴²).

Given a domain D , Cox defined a doubly stochastic Poisson process driven by a non-negative random field $\Lambda(\cdot)$, with realizations $\lambda(\mathbf{x})$ for all $\mathbf{x} \in D$, to be an inhomogeneous Poisson process with intensity function $\lambda(\cdot)$. This definition implies the following properties:

Property 2.2.6. $\{\Lambda(\mathbf{x}) : \mathbf{x} \in D\}$ is a non-negative valued stochastic process.

Property 2.2.7. Conditional on $\{\Lambda(\mathbf{x}) : \mathbf{x} \in D\}$, the events form an inhomogeneous Poisson process with intensity function $\lambda(\mathbf{x})$.

The fundamental assumption of non-negativity for the underlying random field is achieved, in particular, by defining $\Lambda(\cdot) = \exp\{Y(\cdot)\}$, where Y is a Gaussian process. These processes, called log-Gaussian Cox processes, are described in the next section.

Log-Gaussian Cox processes

As introduced in the previous section, a natural way to specify a Cox process is to consider the logarithm of the intensity surface to be a Gaussian process, which ensures the required non-negativity of the process. This approach is particularly useful in various applications, where it is often appropriate to model the intensity generating random field as a suitable function of a Gaussian random field. The advantage of this method lies in the well-known properties of Gaussian fields, where the probabilistic structure can be completely specified in terms of the first- and second-order moments. This allows for more straightforward inferential procedures and meaningful interpretations of the model.

Under this approach, the research has been primarily focused on the class of log-Gaussian Cox processes, for which the intensity generating random function is defined as

$$\Lambda(\mathbf{x}) = \exp\{G(\mathbf{x})\} \quad (2.40)$$

where $G(\mathbf{x})$ is a Gaussian random field, with the first- and second-order moments expressed, respectively, as

$$\begin{aligned}\mu(\mathbf{x}) &= E[G(\mathbf{x})] \\ C(\mathbf{x}, \mathbf{y}) &= C(G(\mathbf{x}), G(\mathbf{y})) = E[G(\mathbf{x})G(\mathbf{y})] - \mu(\mathbf{x})\mu(\mathbf{y}).\end{aligned}\tag{2.41}$$

In a spatio-temporal context, the intensity function can be extended to include a temporal component, resulting in a random process of the form $\Lambda(\mathbf{x}, t) = \exp\{G(\mathbf{x}, t)\}$, where $G(\mathbf{x}, t)$ is a Gaussian process over both space $\mathbf{x} \in \mathbb{R}^d$ and time $t \in [0, \infty)$. With the aim of performing inference over these models, Brix and Diggle (2001)⁴³ proposed a class of models that include a deterministic component $\lambda(\mathbf{x}, t)$, leading to the following formulation:

$$\Lambda(\mathbf{x}, t) = \lambda(\mathbf{x}, t) \exp\{G(\mathbf{x}, t)\},\tag{2.42}$$

where $\lambda(\mathbf{x}, t)$ represents the deterministic spatio-temporal variation in the mean number of events, while $\exp\{G(\mathbf{x}, t)\}$ is the residual variation at location $\mathbf{x}$ and time t , with $G(\mathbf{x}, t)$ being typically modeled as a stationary Gaussian process.

This modeling framework is particularly useful as it can capture a wide range of dependency effects and observed heterogeneities in real-world data, being commonly applied to fields such as environmental science, epidemiology and spatial ecology. Moreover, the incorporation of a deterministic component $\lambda(\cdot)$ provides a flexible way to introduce external covariates or trends.

Bayesian inference for log-Gaussian Cox processes

Let the Gaussian process G , be discretized on a regular grid across space and time, and let X be a realization over the spatial grid. According to Brix and Diggle (2001)⁴³, to infer and predict the log-intensity function G , samples from the conditional distribution of G given the observations X are required, *i.e.*

$$\pi(G|X) \propto \pi(X|G)\pi(G).\tag{2.43}$$

The Metropolis-adjusted Langevin algorithm (MALA) is commonly used for simulating from $\pi(G|X)$ (see Møller *et al.*, 1998⁴⁴, Brix and Møller, 2001⁴⁵ and Taylor *et al.*, 2013⁴⁶) given the covariance parameters σ, ϕ and θ (for more details on ad-hoc methods for parameter estimation, see, for example, Brix and Diggle, 2001⁴³ and Taylor *et al.*,

2013⁴⁶).

To ensure a valid covariance structure, the grid must be extended. For a grid of size M , the typical extension is a $2M \times 2M$ grid. We denote the Gaussian process on this extended grid as G^{ext} . To reduce computational complexity, we perform a linear transformation $\Gamma = \Sigma_{ext}^{-1/2} (G^{ext} - \sigma^2/2)$ where Σ is the covariance matrix of G (see Møller *et al.*, 1998⁴⁴). This leads to the expression

$$\pi(\Gamma|X) \propto \pi(X|G^{ext}) \pi(\Gamma) \quad (2.44)$$

Sampling from the conditional distribution given in Equation (2.44) requires its density f and gradient ∇f . An efficient method, proposed by Roberts and Tweedie (1996)⁴⁹, is to use the Langevin proposal for MCMC. In this method, the proposal distribution $q(\Gamma, \Gamma')$ is generated from a multivariate normal distribution with mean $\Gamma + \frac{1}{2} \nabla \log(\pi(\Gamma|X))$ and variance $h^2 \mathbb{I}$, where $\mathbb{I}$ is the identity matrix and $h > 0$ is a scale parameter chosen to optimize algorithm convergence.

2.2.5 Goodness-of-fit testing for spatial point patterns

Diggle (1983)¹²⁰ introduced a simulation-based approach for goodness-of-fit analysis, following the methodology of Bessag and Diggle (1977)¹²², to test a simple null hypothesis using a Monte Carlo framework.

Let $\hat{F}_1$ be a summary description of the observed point pattern, and F be the corresponding theoretical summary description under the null hypothesis H that the data are a realization of a specified process. Define $d(\cdot)$ as a discrepancy measure between functions. Ideally, F should be known analytically; if not, we use:

$$\bar{F} = (m-1)^{-1} \sum_{i=2}^m \hat{F}_i, \quad (2.45)$$

where $\hat{F}_i$ is computed from each of $m-1$ simulated realizations under the null hypothesis. Note that $\bar{F}$ depends only on the simulations and not on the original data, ensuring it provides an unbiased estimate of F .

Let $u_1 = d(\hat{F}_1, F)$ denote the discrepancy between the observed and theoretical summary description. For each of the $m-1$ simulations, we compute $u_i = d(\hat{F}_i, F)$. If

F is unknown, an alternative is:

$$\bar{F}_i = (m-2)^{-1} \sum_{\substack{j=2 \\ j \neq i}}^m \hat{F}_j \text{ and } u_i = d(\hat{F}_i, \bar{F}_i), \quad (2.46)$$

If $u_{(i)}$ denote the ordered values of u_i (with no ties assumed), the probability under H that u_1 takes the i -th rank is:

$$P[u_1 = u_{(i)}] = m^{-1}, \quad i = 2, \dots, m. \quad (2.47)$$

Thus, the rank of u_1 provides the exact significance level for the Monte Carlo test of H .

Diggle (1979)¹²¹ proposed the use of the following discrepancy measures $d(\cdot)$, each defined over a suitable range of x . One option is to use the supremum of the absolute difference between the empirical and theoretical function:

$$d(\hat{F}, F) = \sup |\hat{F}(x) - F(x)|. \quad (2.48)$$

Alternatively, the discrepancy can be expressed as the integral of the absolute difference:

$$d(\hat{F}, F) = \int |\hat{F}(x) - F(x)| dx, \quad (2.49)$$

or as the integral of the squared difference:

$$d(\hat{F}, F) = \int \left\{ \hat{F}(x) - F(x) \right\}^2 dx. \quad (2.50)$$

A more generic version, especially relevant when using K -function, incorporates a power transformation c and a tuning parameter t_0 :

$$d(\hat{F}, F) = \int_0^{x_0} \left\{ \hat{F}^c(x) - F^c(x) \right\}^2 dx \quad (2.51)$$

where the power transformation c controls the sampling fluctuation in $\hat{F}(x)$ and reduces the influence of large values of x , thereby reducing the sensitivity to the choice of x_0 .

Regarding the number of simulations, Marriot (1978)¹²⁵ recommends $m = 100$ as a practical choice. Beyond this, the power of the test increases only marginally with

increasing m (see Hope, 1968¹²³, for more details on power loss in specific contexts).

A key component of the inferential procedure is the choice of the summary description F of a spatial point process and its empirical counterpart $\hat{F}$ from the observed data. Diggle (1979)¹²¹ emphasized that a minimal requirement for a good summary description is that it should be capable of detecting a wide range of deviations from complete spatial randomness (CSR), particularly both clustered and regular patterns.

A commonly used test statistic was proposed by Clark and Evans (1954)¹²⁴, based on the sum of nearest-neighbor distances y_i from each of n events in a region A , defined as:

$$c = \sum_{i=1}^n y_i. \quad (2.52)$$

Another frequently used test statistic is the quadrat count index of dispersion, originally proposed by Fisher et al. (1992)¹²⁶.

$$q = \frac{\sum_{i=1}^{k^2} (z_i - \bar{z})^2}{\bar{z}}, \quad (2.53)$$

where the z_i represent the counts of events within each cell of a $k \times k$ partition over the region A . The statistic q is mathematically equivalent to Pearson's χ^2 goodness-of-fit criterion for assessing uniformity over A .

For summarizing spatial structure, Ripley (1977)²⁷ suggests the use of the K -function, defined in Equation (2.20). Its empirical counterpart involves the distribution of distances between all ordered pairs of events in A , weighted to account for edge effects.

2.3 Information and complexity measures

The structural properties of a dynamical system can be characterized by means of informational and complexity measures. The concept of entropy, first defined by Shannon (1948)⁹, and generalized by Rényi (1961)¹⁰, as the uncertainty contained in a probability distribution, can be used to quantify the degree of disorder in the system. Other informational measures such as Kullback and Leibler (1951)⁵⁰ divergence and the corresponding generalization proposed by Rényi (1961)¹⁰ are useful to determine the degree of dissimilarity between two systems.

A similar analysis can be performed in the context of complexity, for instance, with the exponential LMC complexity proposed by Catalán et al. (2002)⁵¹. This is indeed an extension of LMC complexity, see López-Ruiz et al. (1995)⁵², which solves the problems that arise for continuous distributions, but it is also appropriate in the discrete case as it can be interpreted in terms of diversity, see Angulo et al. (2021)⁵³ and Campbell (1996)⁵⁴. A related two-parameter generalization, in terms of Rényi entropies of different deformation orders, was formulated by López-Ruiz et al. (2009)⁵⁵. Under a similar product-type structure, the two parameter generalized relative complexity measure was introduced by Romera et al. (2011)⁵⁶, based on Rényi divergences.

2.3.1 Entropy measures

Information Theory arose in the context of Communication Theory for solving the emerging problems in message transmission through noisy channels. Based on the seminal concept of *information content* introduced by Hartley (1928)⁵⁷, as a measure of the amount of information provided by the knowledge of the state in a finite system, Shannon (1948)⁹ formulated *entropy* (or *information entropy*) as a measure of the uncertainty intrinsic to a given discrete probability distribution, $\mathbf{p} = (p_1, \dots, p_N)$, in terms of the expectation.

Definition 2.3.1. Shannon entropy of a discrete probability distribution, $\mathbf{p} = (p_1, \dots, p_N)$, is defined as

$$H(\mathbf{p}) := -KE [\log(\mathbf{p})] = -K \sum_{i=1}^N p_i \log(p_i) \quad (2.54)$$

where K is a positive constant.

The constant K is a scale term and the base of the logarithm is arbitrary, being the most commonly used base e and $K = 1$, then

$$H(\mathbf{p}) := -E [\ln(\mathbf{p})] = - \sum_{i=1}^N p_i \ln(p_i). \quad (2.55)$$

The maximum entropy value is obtained for a system consisting on N equiprobable states, with $H_{\max} = \ln(N)$.

Shannon entropy takes values in $\mathbb{R}_0^+$. Because it lacks an upper bound a normalized version, called *configuration entropy*, is sometimes used. It is expressed as

$$H_{\text{norm}}(\mathbf{p}) = \frac{H(\mathbf{p})}{H_{\text{max}}} = \frac{-\sum_{i=1}^N p_i \ln(p_i)}{\ln(N)}. \quad (2.56)$$

Among various well-known generalizations, formally derived by a certain relaxation of the intrinsic axiomatic, are Rényi¹⁰ and Tsallis¹¹ entropies of order α (with α being a *deformation parameter*).

Definition 2.3.2. Rényi entropy of order α of a discrete probability distribution, $\mathbf{p} = (p_1, \dots, p_N)$, is defined as

$$H_\alpha(\mathbf{p}) := \frac{1}{1-\alpha} \ln \left(\sum_{i=1}^N p_i^\alpha \right) = \frac{1}{1-\alpha} \ln (E[\mathbf{p}^{\alpha-1}]) \quad (\alpha \neq 1). \quad (2.57)$$

This entropy is a decreasing function of the order α , and Shannon entropy is obtained as the limiting case for $\alpha \rightarrow 1$ (hence being also denoted as $H_1(\mathbf{p})$). Rényi entropy is also maximum, for any α , in the case of equiprobability, again with $H_{\alpha, \text{max}} = \ln(N)$.

Definition 2.3.3. Tsallis entropy of order α of a discrete probability distribution, $\mathbf{p} = (p_1, \dots, p_N)$, is defined as

$$T_\alpha(\mathbf{p}) := \frac{1}{\alpha-1} \left(1 - \sum_{i=1}^N p_i^\alpha \right) = \frac{1}{\alpha-1} (1 - E[\mathbf{p}^{\alpha-1}]) \quad (\alpha \neq 1). \quad (2.58)$$

The maximum Tsallis entropy is obtained for the equiprobable distribution and depends on the parameter α ,

$$T_{\alpha, \text{max}} = \frac{1 - \sum_{i=1}^N \left(\frac{1}{N}\right)^\alpha}{\frac{1-N^{1-\alpha}}{\alpha-1}} = \frac{1 - N^{1-\alpha}}{\alpha-1}. \quad (2.59)$$

2.3.2 Complexity measures

Another important concept for structural assessment of a random system is complexity, which, among other approaches introduced in the literature, has been specifically understood, in a probabilistic informational sense, as departure from both degeneracy into one

single state (*perfect order*) and equiprobability (*complete disorder*). In this direction, López-Ruiz et al. (1995) proposed the following formulation of a complexity measure (usually referred as the *LMC complexity*): for a given discrete probability distribution $\mathbf{p}$,

$$C_{\text{LMC}}(\mathbf{p}) = H(\mathbf{p}) \cdot D(\mathbf{p}), \quad (2.60)$$

with the first factor being the Shannon entropy, and the second one representing the disequilibrium defined as the quadratic distance

$$D(\mathbf{p}) = \sum_{i=1}^N \left(p_i - \frac{1}{N} \right)^2. \quad (2.61)$$

López-Ruiz et al. (1995)⁵² proposed a modification of LMC complexity by replacing the Shannon entropy for the configuration entropy (2.56)

$$\bar{C}_{\text{LMC}} = H_{\text{norm}} \cdot D = \left(\frac{\sum_{i=1}^N p_i \ln(p_i)}{\ln(N)} \right) \left(\sum_{i=1}^N \left(p_i - \frac{1}{N} \right)^2 \right). \quad (2.62)$$

Extending this formulation to the continuous case, to ensure the positivity of complexity values, Catalán et al. (2002)⁵¹ proposed a modified *exponential* version of the LMC complexity, which, in the discrete case is given by

$$C_{\text{LMC}}^{\text{exp}} = e^{H(\mathbf{p})} \cdot D(\mathbf{p}). \quad (2.63)$$

This idea was the basis for the formulation of a two-parameter generalized complexity measure, proposed later by López-Ruiz et al. (2009)⁵⁵ for continuous distributions. In fact, the latter is perfectly meaningful also in the discrete case, see Angulo et al (2021)⁵³, for instance in relation to Campbell (1966)⁵⁴ diversity index, for which it takes the form

$$C_{\alpha,\beta}(\mathbf{p}) := \exp^{H_{\alpha}(\mathbf{p}) - H_{\beta}(\mathbf{p})}, \quad \text{with } 0 < \alpha, \beta < \infty. \quad (2.64)$$

2.3.3 Divergence and relative complexity measures

While entropy and complexity measures enable a comparison in global terms (marginally) of two given probability distributions, a proper assessment of their structural (state-by-state) dissimilarity, or lack of mutual coherence, is achieved by means of divergence and

relative complexity measures. For two given probability distributions, $\mathbf{p} = (p_1, \dots, p_N)$ and $\mathbf{q} = (q_1, \dots, q_N)$, on a system with N possible states, Kullback and Leibler (1951)⁵⁰ defined the divergence of $\mathbf{p}$ from $\mathbf{q}$ as

$$KL(\mathbf{p}||\mathbf{q}) := \sum_{i=1}^N p_i \ln \left(\frac{p_i}{q_i} \right) = E_p \left[\ln \left(\frac{\mathbf{p}}{\mathbf{q}} \right) \right]. \quad (2.65)$$

This is a non-symmetric (*directed*), non-negative measure, vanishing if and only if $\mathbf{p} = \mathbf{q}$. A corresponding generalization based on a *deformation parameter* is also given by Rényi (1961)¹⁰ divergence of order α of $\mathbf{p}$ from $\mathbf{q}$, defined as

$$H_\alpha(\mathbf{p}||\mathbf{q}) := \frac{1}{\alpha - 1} \ln \left(\sum_{i=1}^N p_i \left(\frac{p_i}{q_i} \right)^{\alpha-1} \right) = \frac{1}{\alpha - 1} \ln \left(E_p \left[\left(\frac{\mathbf{p}}{\mathbf{q}} \right)^{\alpha-1} \right] \right) \quad (\alpha \neq 1). \quad (2.66)$$

For fixed argument distributions, Rényi divergence is a non-decreasing function of the deformation parameter. For $\alpha \rightarrow 1$, $H_\alpha(\mathbf{p}||\mathbf{q})$ tends to $KL(\mathbf{p}||\mathbf{q})$, hence with the latter being also denoted as $H_1(\mathbf{p}||\mathbf{q})$.

In the special case of the divergence of order α from equiprobability, *i.e.* with $\mathbf{q} \equiv [\frac{1}{N}]$, its value can be calculated as

$$H_\alpha \left(\mathbf{p} || \left[\frac{1}{N} \right] \right) = H_{\alpha, \max} - H_\alpha(\mathbf{p}) = \ln N - H_\alpha(\mathbf{p}), \quad (2.67)$$

hence being known as the *information difference* of order α for $\mathbf{p}$.

A divergence-based formulation of a product-type generalized relative complexity measure was introduced for continuous distributions by Romera et al. (2011)⁵⁶. In the discrete case, it takes the form (see Angulo *et al.*, 2021⁵³, in relation to a concept of *relative diversity*),

$$C_{\alpha, \beta}(\mathbf{p}||\mathbf{q}) := \exp^{H_\alpha(\mathbf{p}||\mathbf{q}) - H_\beta(\mathbf{p}||\mathbf{q})}, \quad \text{with } 0 < \alpha, \beta < \infty. \quad (2.68)$$

In particular, for $q \equiv [\frac{1}{N}]$, the generalized complexity and relative complexity measures (2.64) and (2.68) are reciprocal in the following sense:

$$C_{\alpha, \beta} \left(\mathbf{p} || \left[\frac{1}{N} \right] \right) = \frac{1}{C_{\alpha, \beta}(\mathbf{p})} = C_{\beta, \alpha}(\mathbf{p}). \quad (2.69)$$

A synthetic review, providing connective relations and interpretation of the above summarized information-complexity concepts, is given in Angulo et al. (2021)⁵³.

2.4 Risk measure theory

Risk is a broad concept that involves many disciplines, such as finance, insurance and engineering fields, among others. In the last decades, motivated by the increasing frequency of extreme natural disasters, spatial risk measures have become prominent with the aim of identify areas of risk; see, for example, Koch (2017)⁵⁸, Romero et al. (2018)⁵⁹ and Koch (2019)⁶⁰. In this Section we present some notation and preliminary concepts related to risk measures and introduce the important family of quantile-based and entropic risk measures.

2.4.1 Basic concepts

Definition 2.4.1. Let $(\Omega, \mathcal{A}, P)$ be a probability space and $\mathcal{M}_{\mathcal{B}}(\Omega, \mathcal{A}, P)$ be the space of real-valued Borel-measurable functions defined on $(\Omega, \mathcal{A}, P)$. A *risk measure* is a map:

$$\rho : \mathcal{M}(\Omega, \mathcal{A}, P) \rightarrow \mathbb{R} \cup \{-\infty, \infty\}, \quad (2.70)$$

satisfying certain suitable conditions, where $\mathcal{M}(\Omega, \mathcal{A}, P)$ is a linear subspace of $\mathcal{M}_{\mathcal{B}}(\Omega, \mathcal{A}, P)$.

Föllmer and Schied (2004)⁶¹ proposed the following set of properties which define the class of *coherent* and *convex risk measures*:

- a. *Monotonicity*: If $X \leq Y$, then $\rho(X) \leq \rho(Y)$.
- b. *Traslation invariance*: If $m \in \mathbb{R}$, then $\rho(X + m) = \rho(X) + m$.
- c. *Convexity*: $\rho(\lambda X + (1 - \lambda)Y) \leq \lambda \rho(X) + (1 - \lambda) \rho(Y)$, for $0 \leq \lambda \leq 1$.
- d. *Positive homogeneity*: $\rho(\lambda X) = \lambda \rho(X)$, for $\lambda \geq 0$.
- e. *Subadditivity*: $\rho(X + Y) \leq \rho(X) + \rho(Y)$.

Definition 2.4.2. A risk measure ρ is a *convex risk measure* if it satisfies a, b and c.

A weaker requirement for convexity is *quasi-convexity*, which is obtained by relaxing the property c as follows:

$$\rho(\lambda X + (1 - \lambda)Y) \leq \rho(X) \vee \rho(Y), \quad \text{for } 0 \leq \lambda \leq 1. \quad (2.71)$$

Definition 2.4.3. A convex risk measure ρ is as *coherent risk measure* if it satisfies d and e.

2.4.2 Some relevant risk measure families

Quantile-based risk measures

Risk measures based on certain characteristics of the probability distribution, particularly the quantiles, are frequently used because of its easy interpretation. One of the most accepted measures of risk in the last decade on XX century was the Value-at-Risk (VaR), introduced by JP Morgan as a risk management control system and defined as the minimum loss expected in the $100\alpha\%$ of the worst scenarios ($1 - \alpha$ quantile of the distribution).

Definition 2.4.4. Let X be a loss random variable, with cumulative distribution function F_X , and let $\alpha \in (0, 1]$. The *Value-at-Risk* at a confidence level $1 - \alpha$ of the variable X is defined as

$$VaR_{1-\alpha}(X) := \inf \{x \in \mathbb{R} : F_X(x) \geq 1 - \alpha\}. \quad (2.72)$$

Value-at-Risk has certain limitations such as it does quantify the magnitude of potential losses and it does not fulfill the subadditivity property, therefore it is not coherent. For that reason, other alternative measures were proposed such as the Average Value-at-Risk (AVaR), also known as Expected Shortfall.

Definition 2.4.5. Let X be a loss random variable, with cumulative distribution function F_X , and let $\alpha \in (0, 1]$. The *Average Value-at-Risk* at a confidence level $1 - \alpha$ of X is defined as

$$AVaR_{1-\alpha}(X) := \frac{1}{\alpha} \int_{1-\alpha}^1 VaR_p(X) dp. \quad (2.73)$$

The AVaR is a coherent risk measure and represents the average lost that will occur in the $100\alpha\%$ of the worst scenarios. The AVaR belongs to the family of spectral risk measures, proposed by Acerbi (2002)⁸ and defined as follows

Definition 2.4.6. The *spectral risk measure* $M_\phi(X)$ with weight function ϕ is defined as

$$M_\phi(X) = \int_0^1 VaR_p(X) \phi(p) dp, \quad (2.74)$$

with $\phi(\cdot) \geq 0$ and $\int_0^1 \phi(p) dp = 1$. M_ϕ is a coherent risk measure for monotonically non-decreasing weight functions (e.g. $AVaR_{1-\alpha}$).

Entropic risk measures

Another relevant family of risk measures is defined by *entropic risk measures*, which incorporate concepts of Information Theory.

Definition 2.4.7. The class of *entropic risk measures* is defined as

$$e_\gamma(X) := \frac{1}{\gamma} \log E_{\mathbf{p}} [\exp^{\gamma X}] = \sup_{\mathbf{q} \ll \mathbf{p}} \left\{ E_{\mathbf{q}}[X] - \frac{1}{\gamma} H(\mathbf{q} \parallel \mathbf{p}) \right\}, \quad (2.75)$$

where $\gamma \in [0, \infty)$, $H(\mathbf{q} \parallel \mathbf{p})$ is the divergence of $\mathbf{q}$ with respect to $\mathbf{p}$, with $e_0(X) = E[X]$.

Within this class, Fölmer and Knispel (2011)⁶² proposed a coherent version of an entropic risk measure.

Definition 2.4.8. For each $c > 0$, the *coherent entropic risk measure* at level c is defined as

$$\rho_c(X) := \sup_{\mathbf{q}: H(\mathbf{q} \parallel \mathbf{p}) \leq c} E_{\mathbf{q}}[X] \quad (2.76)$$

The coherent entropic risk measure ρ_c is law-invariant, continuous from above and below and increasing in c .

Chapter 3

Structural Complexity and Informational Transfer in Spatial Log-Gaussian Cox Processes

The doubly stochastic mechanism generating the realizations of spatial log-Gaussian Cox processes is empirically assessed in terms of generalized entropy, divergence and complexity measures. The aim is to characterize the contribution to stochasticity from the two phases involved, in relation to the transfer of information from the intensity field to the resulting point pattern, as well as regarding their marginal random structure. A number of scenarios are explored regarding the Matérn model for the covariance of the underlying log-intensity random field. Sensitivity with respect to varying values of the model parameters, as well as of the deformation parameters involved in the generalized informational measures, is analyzed on the basis of regular lattice partitionings. Both a marginal global assessment based on entropy and complexity measures, and a joint local assessment based on divergence and relative complexity measures, are addressed. A Poisson process and a log-Gaussian Cox process with white noise intensity, the first providing an upper bound for entropy, are considered as reference cases. Differences regarding the transfer of structural information from the intensity field to the subsequently generated point patterns, reflected by entropy, divergence and complexity estimates, are discussed according to the specifications considered. In particular, the magnitude of the decrease of marginal entropy estimates between the intensity random fields and the corresponding

point patterns quantitatively discriminates the global effect of the additional source of variability involved in the second phase of the double stochasticity.

3.1 Introduction

Log-Gaussian Cox processes define a class of doubly stochastic Poisson processes (Cox, 1955⁴¹) where the Gaussian intensity function is transformed through exponentiation. These processes (see Møller *et al.*, 1998⁴⁴, for a formal definition and properties of log-Gaussian Cox processes) allow the generation of point patterns through a stochastic two-step procedure, where the clustering structure observed in the pattern is due to the inclusion of random heterogeneities in the intensity function. The first applications of these processes are due to Coles and Jones (1991)⁶³ who use a lognormal random field as a model of galaxies distribution, and Rathbun (1996)⁶⁴ that models the effect of external variables to describe the patterns formed by the location of organisms. Cox process models fit naturally in geosciences and ecology fields (see Illian *et al.*, 2013⁶⁵) as the resulting point processes are considered to be driven by environmental variables. There are also contributions in the context of epidemiology (Diggle *et al.*, 2005⁶⁶), ecology (see, for example, Illian *et al.*, 2009⁶⁷; Waagepetersen and Schweder, 2006⁶⁸), crime data analysis (Rodrigues and Diggle, 2012⁶⁹) and seismology (Siino *et al.*, 2017⁷⁰), among others.

The structural properties of random fields and point patterns can be characterized by means of informational and complexity measures. The concept of entropy, first defined by Shannon (1948)⁹, and generalized by Rényi (1961)¹⁰, as the uncertainty contained in a probability distribution, can be used to quantify the degree of inhomogeneity of each phase of the process. Other informational measures such as Kullback and Leibler (1951)⁵⁰ divergence and the corresponding generalization proposed by Rényi are useful to determine the probabilistic local coherence of the phases in the sense of the structural information transferred from the intensity field to the point pattern. A similar analysis can be performed in the context of complexity, for instance, with the exponential LMC complexity proposed by Catalán *et al.* (2002)⁵¹. This is indeed an extension of LMC complexity, see López-Ruiz *et al.* (1995)⁵², which solves the problems that arise for continuous distributions, but it is also appropriate in the discrete case as it can be interpreted in terms of diversity (Campbell, 1996⁵⁴; Angulo *et al.*, 2021⁵³). A related

two-parameter generalization, in terms of Rényi entropies of different deformation orders, was formulated by López-Ruiz *et al.* (2009)⁵⁵. Under a similar product-type structure, the two-parameter generalized relative complexity measure introduced by Romera *et al.* (2011)⁵⁶, based on Rényi divergences, is used here to describe the local coherence in terms of complexity between the two phases of the doubly stochastic process mechanism.

In the last decades, since Papangelou (1978)⁷¹ work defining entropy rate for continuous point processes in the real line, many other studies introduce theoretical concepts in the context of information theory and complexity for the analysis of point processes. Baratpour *et al.* (2007)⁷² assessed the properties of non-homogeneous Poisson processes in terms on entropy, Daley and Vere-Jones (2004)⁷³ extended the definition of entropy for a point process in a d -dimensional space, and more recently Angulo *et al.* (2021)⁵³ introduce approaches for the analysis of spatial point patterns regarding informational and complexity aspects and focusing on a multifractal context. However, to our knowledge, complexity and information transfer between the two phases of generation of log-Gaussian Cox processes has not been explored.

The main objective of this chapter is to analyze the structural transfer of information from the intensity random field to the subsequently generated point pattern in a log-Gaussian Cox process. A marginal global assessment is performed in terms of entropy and complexity measures, and, complementarily, local correspondence is evaluated based on divergence and relative complexity measures. The study, addressed by simulation under a variety of selected scenarios, is primarily focussed on sensitivity in relation to the configuration of model parameters, as well as concerning the specification of deformation parameters involved in generalized informational measures.

3.2 Methodological aspects

As mentioned, the analysis is aimed at characterizing, both in a global (marginal) and a local sense, the information transfer from the intensity field to the point pattern. To that end, an empirical approach based on the lattice box-counting methodology is adopted.

Formally, the analysis is based on the simulation of log-Gaussian Cox processes for different scenarios, under different varying configurations of the covariance function and mean parameters of the intensity-generating Gaussian random field; see details in Table 3.1. For each specific configuration, in the first stage, 100 independent replicates of the

intensity field are simulated on the square $D = [0, 10]^2$, based on a 180×180 pixel window. From each realization, in the second stage, one or multiple point patterns are independently generated, as discussed below according to the objective of the analysis.

For assessment, the information-complexity measures are applied considering, in particular, a 10×10 lattice overlaid on the domain D . Each intensity realization is converted to a quadrat-based ('raster') mean intensity field by averaging its values within each cell, and further transformed into a discrete probability distribution by normalization, hence representing the relative intensity for the quadrat design adopted. As for the point patterns, the relative frequencies obtained from intra-cell event counting provide the quadrat-based reference discrete probability distribution.

Entropy (Shannon, Rényi) and complexity (LMC, generalized) measures are marginally evaluated on the related probability distributions derived, for the assessment regarding the characterization and global transfer of information between the two phases of stochasticity. In particular, according to the scenarios considered, the analysis of results is focused on sensitivity with respect to variations in the model and deformation parameters.

Additionally, the degree of local (cell-based) coherence between the distributions corresponding to the two phases is analyzed on the basis of divergence (Kullback-Leibler, Rényi) and relative complexity (generalized) measures.

	Variable parameters	Fixed parameters		
Scenario 1	$\sigma^2 = [0, 0.25, \dots, 5]$	$\nu = 1$	$\rho = 1$	$\mu = 0$
Scenario 2	$\mu = [0, 0.2, \dots, 4]$	$\sigma^2 = 1$	$\nu = 1$	$\rho = 1$
Scenario 3	$\nu = [0.25, 0.375, \dots, 2.75]$	$\sigma^2 = 1$	$\rho = 1$	$\mu = 0$
Scenario 4	$\rho = [1, 1.5, \dots, 11]$	$\sigma^2 = 1$	$\nu = 1$	$\mu = 0$

Table 3.1: Parameter configurations for the scenarios considered in the analysis.

As mentioned, from each simulation of the intensity field, either just one or multiple independent point pattern replicates can be analyzed. The second strategy specifically seeks to discriminate the relative contribution from the two phases of stochasticity to the structural variability of the process. For a preliminary assessment regarding this issue, 100 independent patterns are generated from each one of 100 replicates of the intensity random field (hence having in total 100×100 patterns), for varying values

1	2	1	3	1	2	5	2	1	2
1	1	1	1	4	10	1	0	1	0
0	0	0	0	4	6	0	2	7	0
3	14	9	0	0	6	6	11	6	0
1	0	1	1	0	0	5	3	0	4
1	1	2	0	0	0	0	7	0	0
0	0	2	1	1	0	0	7	4	0
0	0	5	0	0	0	0	0	6	1
0	2	13	0	1	1	1	0	0	2
5	1	0	4	0	1	0	0	0	6

Figure 3.1: Example of quadrat-based counts of a point pattern.

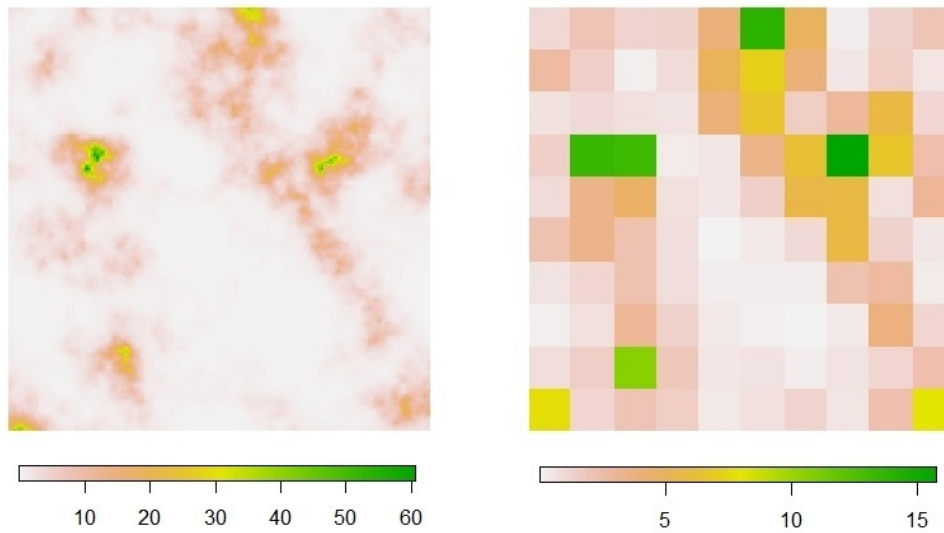


Figure 3.2: Example of mean intensity raster (right) of an intensity field realization (left).

of the variance parameter according to Scenario 1 (see Table 3.1). Different standard deviation values of Shannon entropy and Kullback-Leibler divergence, summarized in Table 3.2, are computed for comparison as follows:

- ‘Mean Intra SD’: The average of the internal standard deviations of the entropy and divergence values obtained within each of the 100 sets of point patterns.

- ‘Inter-Mean SD’: The standard deviation of the internal average entropy and divergence values obtained within each of the 100 sets of point patterns.
- ‘Total SD’: The standard deviation of the entropy and divergence values obtained for the 10000 patterns.
- ‘Inter-Single SD’: The standard deviation of entropy and divergence values based on only one single pattern being generated from each of the 100 replicates of the intensity field.

Parameter value	Shannon entropy SD				Kullback-Leibler divergence SD			
	Multiple patterns			Single pattern	Multiple patterns			Single pattern
	Mean Intra SD	Inter-Mean SD	Total SD	Inter-Single SD	Mean Intra SD	Inter-Mean SD	Total SD	Inter-Single SD
$\sigma^2 = 0$	0.0782	0.0076	0.0784	0.0820	0.0782	0.0076	0.0784	0.0820
$\sigma^2 = 0.25$	0.0765	0.0552	0.0943	0.0914	0.0698	0.0561	0.0895	0.0911
$\sigma^2 = 0.5$	0.0737	0.0607	0.0954	0.0911	0.0600	0.0707	0.0927	0.0912
$\sigma^2 = 0.75$	0.0749	0.0849	0.1133	0.1245	0.0532	0.0717	0.0894	0.0892
$\sigma^2 = 1$	0.0733	0.0966	0.1211	0.1279	0.0469	0.0655	0.0807	0.1051
$\sigma^2 = 1.25$	0.0743	0.1127	0.1349	0.1364	0.0429	0.0806	0.0917	0.1064
$\sigma^2 = 1.5$	0.0709	0.1497	0.1654	0.2081	0.0358	0.0778	0.0860	0.0865
$\sigma^2 = 1.75$	0.0702	0.1590	0.1734	0.1767	0.0326	0.0749	0.0820	0.0767
$\sigma^2 = 2$	0.0685	0.1432	0.1584	0.2025	0.0290	0.0666	0.0729	0.0914
$\sigma^2 = 2.25$	0.0695	0.1856	0.1977	0.2141	0.0282	0.0812	0.0863	0.0686
$\sigma^2 = 2.5$	0.0650	0.2227	0.2312	0.2758	0.0249	0.0788	0.0830	0.0759
$\sigma^2 = 2.75$	0.0619	0.2187	0.2266	0.3048	0.0201	0.0716	0.0746	0.0696
$\sigma^2 = 3$	0.0607	0.3478	0.3516	0.2188	0.0184	0.0632	0.0661	0.0649
$\sigma^2 = 3.25$	0.0625	0.2711	0.2774	0.2982	0.0177	0.0598	0.0626	0.0638
$\sigma^2 = 3.5$	0.0574	0.3561	0.3591	0.2585	0.0150	0.0491	0.0516	0.0512
$\sigma^2 = 3.75$	0.0560	0.2575	0.2627	0.2949	0.0143	0.0530	0.0551	0.0574
$\sigma^2 = 4$	0.0516	0.3702	0.3722	0.3422	0.0117	0.0493	0.0510	0.0572
$\sigma^2 = 4.25$	0.0519	0.3046	0.3079	0.4179	0.0116	0.0490	0.0505	0.0575
$\sigma^2 = 4.5$	0.0495	0.4012	0.4025	0.3855	0.0100	0.0370	0.0384	0.0567
$\sigma^2 = 4.75$	0.0493	0.3999	0.4013	0.3825	0.0103	0.0543	0.0556	0.0387
$\sigma^2 = 5$	0.0471	0.3920	0.3131	0.4212	0.0088	0.0384	0.0396	0.0428

Table 3.2: Table of ‘Inter-Intra’ standard deviations of Shannon entropy and Kullback-Leibler divergence for point patterns generated under Scenario 1, with multiple and single replicates.

In general terms, comparing the results for ‘Total SD’ and ‘Inter-Single SD’, we can see that the total variability of the entropy and divergence measures obtained from multiple patterns can be well captured with just one single pattern generated from each realization of the intensity field. Therefore, in the analyses performed in Section 3.3, we simulate patterns without replicates.

Nevertheless, from the ‘inter-intra’ analysis based on multiple replicates, it is interesting to observe that, as the variance parameter σ^2 increases, the average internal variability of entropy and divergence values for the patterns derived from each intensity

realization progressively decreases compared to a significant increase of the corresponding variability for the intra-averaged entropy values, and a more stable behavior for the divergence values. This is also visualized in the 95% confidence-level bands in curves represented in the plots of Figures 3.3 and 3.4, respectively for the mean entropy and mean divergence values obtained with and without multiple replicates.

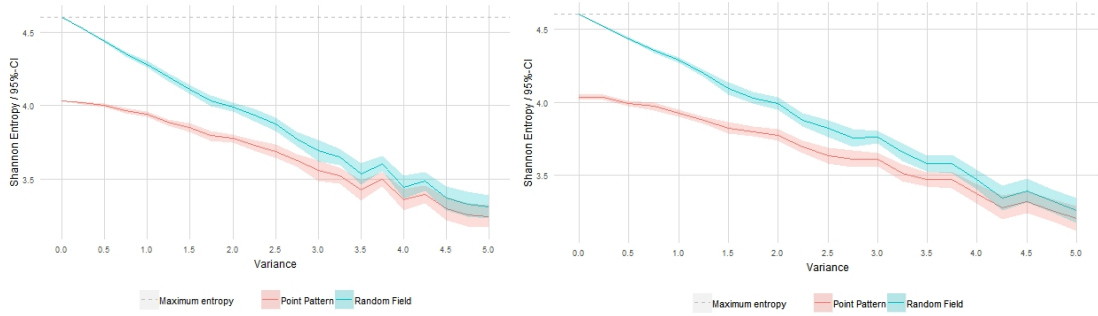


Figure 3.3: Mean Shannon entropy values for point patterns obtained with 100 replicates (left) and with only one generation (right) from each intensity random field realization.

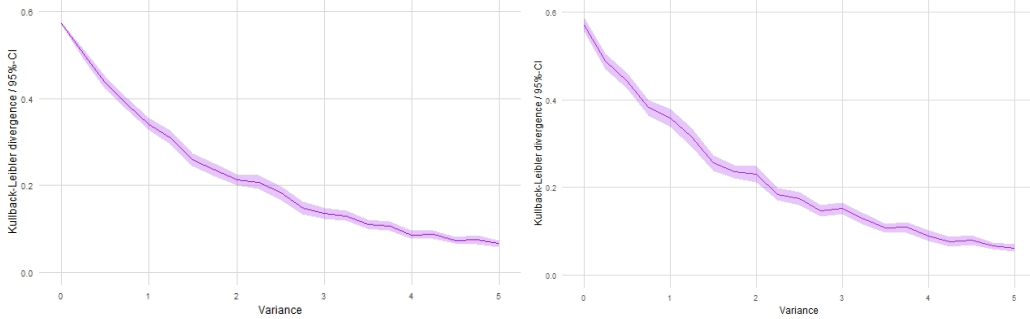


Figure 3.4: Mean Kullback-Leibler divergence values for point patterns obtained with 100 replicates (left) and with only one generation (right) from each intensity random field realization.

With the aim of providing a benchmark case, 100 independent realizations of a homogeneous Poisson process with constant intensity $\lambda = 10$, and similarly for a log-Gaussian Cox process with (pixel-based) white noise intensity, are simulated and analyzed. In the second case, the intensity generating the Gaussian random field has a local normal distribution with mean $\mu = 0$ and variance $\sigma^2 = 2 \ln(10)$, generating an average intensity close to 10. These cases result in patterns with low structuring, establishing certain upper and lower bounds for entropy and complexity, respectively.

The study is carried out using R software, in particular `spastat` and `RandomFields` packages for simulating the processes using `rLGCP` function; `raster` package is used for discretization of the intensity random field.

3.3 Analysis

3.3.1 Weak-structure reference processes

Due to the trivial structure of the reference homogeneous and white noise intensity-based Poisson processes, significant inhibition or aggregation effects are not expected to be observed in the pattern realizations, *i.e.* points should be mostly evenly distributed throughout the spatial domain; as a consequence, Shannon and Rényi entropies (Figures 3.5 and 3.6) take high values, still showing a certain global gain in structuring, slightly increased in the white noise intensity case. Obviously, as the intensity field is constant for the homogeneous Poisson process, its quadrat-based distribution is perfectly uniform and entropy reaches the maximum possible value $H_{\alpha, \max} = \ln(10^2) = 4.60517$, for all α , whilst the entropy of the quadrat-based distribution for the white noise intensity is always close to this maximum.

On the other hand, LMC complexity of the pattern realizations, in both cases, is close to the minimum value 0; see Figure 3.7. This is expected since $H \sim H_{\max}$ and $D \sim 0$. However, in terms of this complexity measure, the pattern structuring is still slightly higher in the white noise intensity case.

The structuring effect from the additional source of variability present in the generation of the point patterns is also assessed by the local discrepancies measured between the point pattern and intensity quadrat-based distributions, as shown in Figure 3.8 for Kullback-Leibler divergence, Figure 3.9 for Rényi divergence, and Figure 3.10 in terms of the generalized relative complexity. The initial variability present in the white noise intensity field is reflected as well in the generation of the patterns.

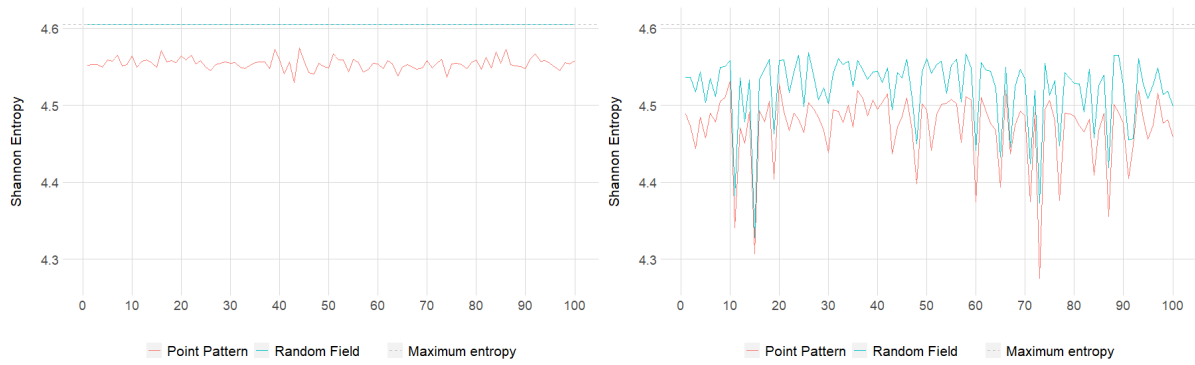


Figure 3.5: Shannon entropy of 100 point patterns and their generating intensity field, with constant $\lambda = 10$ (left) and with white noise based intensity (right).

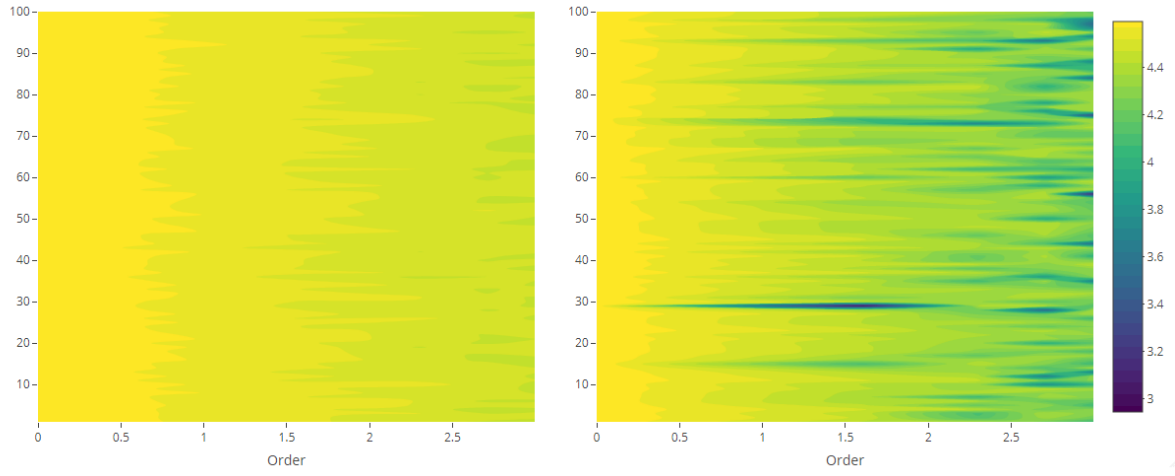


Figure 3.6: Rényi entropy of 100 point patterns and their generating intensity field, with constant $\lambda = 10$ (left) and with white noise based intensity (right).

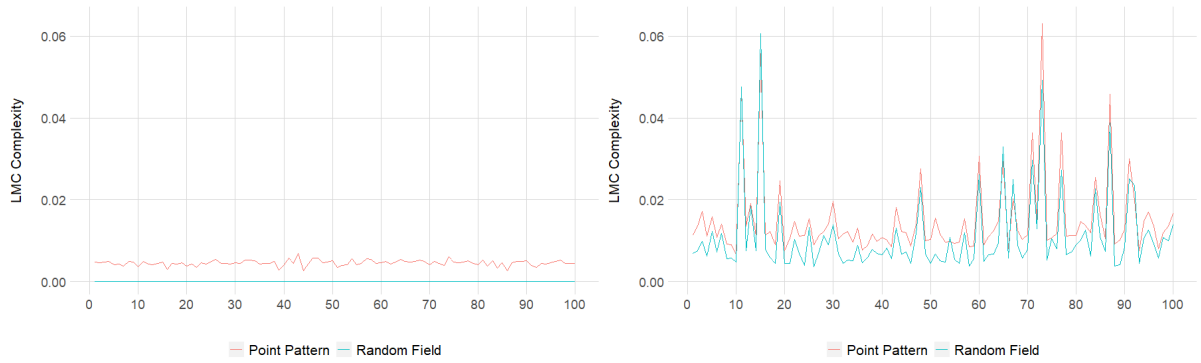


Figure 3.7: LMC complexity of 100 point patterns and their generating intensity field, with constant $\lambda = 10$ (left) and with white noise based intensity (right).

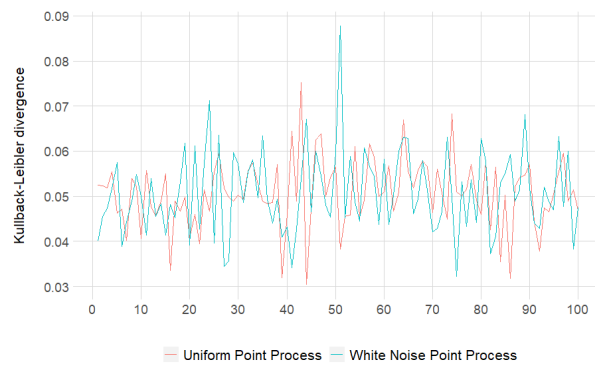


Figure 3.8: Kullback-Leibler divergence between 100 point patterns and their generating intensity field, with constant $\lambda = 10$ (red) and with white noise based intensity (blue).

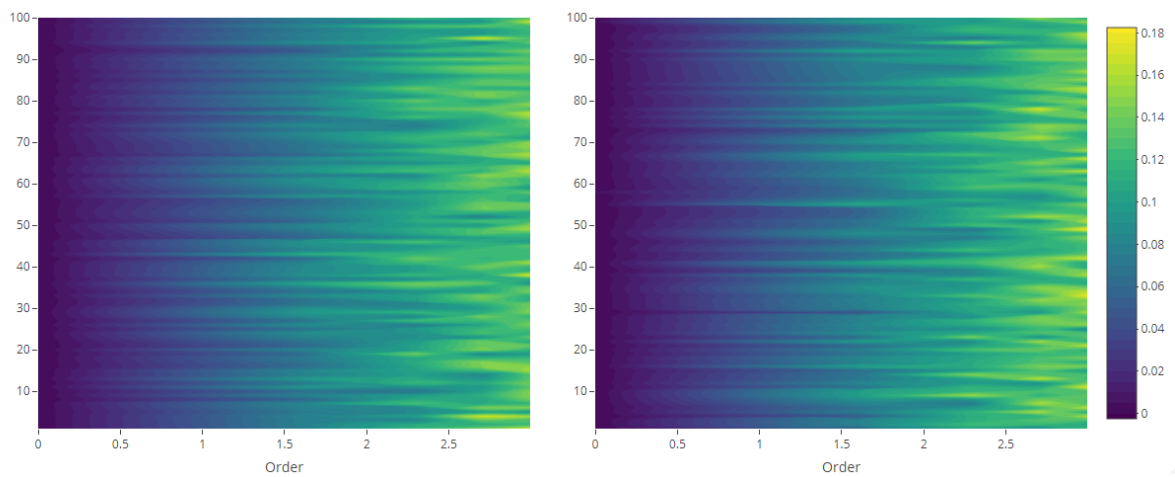


Figure 3.9: Rényi divergence between 100 point patterns and their generating intensity field, with constant $\lambda = 10$ (left) and with white noise based intensity (right).

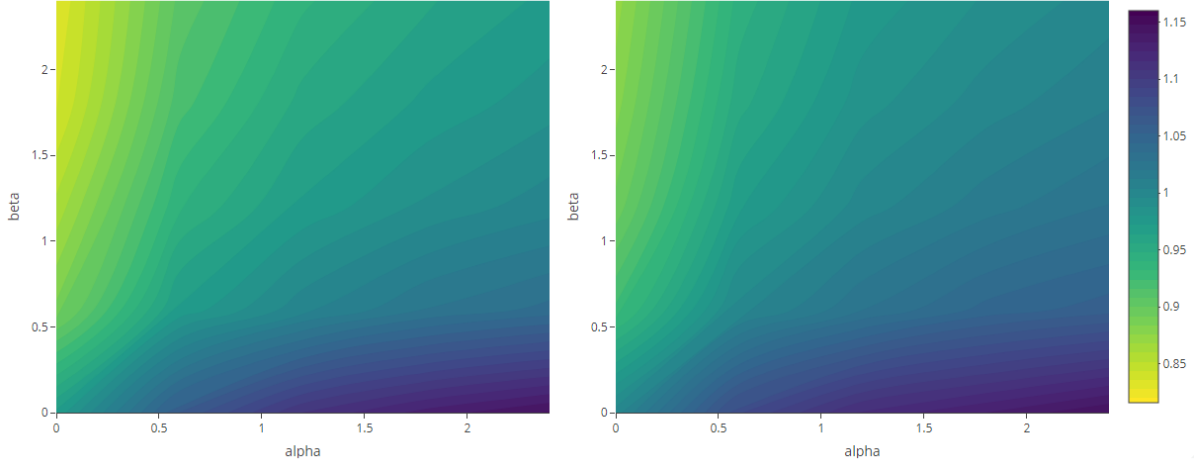


Figure 3.10: Generalized relative complexity between 100 point patterns and their generating intensity field, with constant $\lambda = 10$ (left) and with white noise based intensity (right).

3.3.2 Marginal analysis

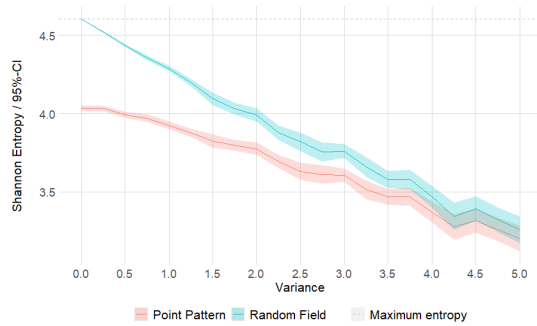
The behavior of entropy is highly influenced by the isolated variation of each parameter. The increase of the variance parameter results in clusters with a high concentration of points, which implies a decrease of entropy values as the distributions move further away from equiprobability (Figure 3.11(a)). Large values of the variance parameter increase the local variability in the magnitude of clustering.

The increase of the mean parameter is reflected in a larger mean intensity of the process, which derives in an exponential increase of the mean number of points in the pattern. As a result, while the entropy values of the quadrat-based probability distribution from the intensity field essentially remain at a constant level below H_{max} (related to the specification of a fixed variance $\sigma^2 = 1$), the entropy values for the corresponding patterns progressively tend from below to this level (Figure 3.11(b)).

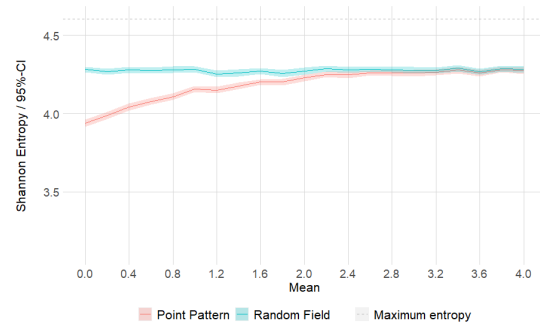
The ν parameter controls the differentiability of the intensity field and determines its smoothness. In Figure 3.11(c) the entropies of the two phases of the processes show a slight gradual decrease as the parameter increases, while the differences between them remain steady. This is partly influenced by the adopted box size, taking into account the local nature of the smoothness effect.

The scale parameter ρ measures the decay of covariance in the intensity field as a function of the distance. The covariance decays more slowly as ρ increases, which results in more homogeneous fields limiting the formation of clusters. As a result, larger values of entropy are obtained for the intensity fields as the scale parameter increases, whilst for the point patterns it can be observed an increase in the variability of the measure (Figure 3.11(d)).

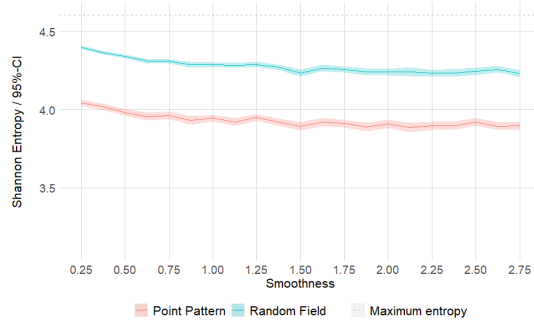
The maps of Rényi entropy (Figures 3.12 - 3.15) show the evolution of the global heterogeneity in each scenario for different magnitudes of the deformation parameter which produces a distortion effect on the probability distributions.



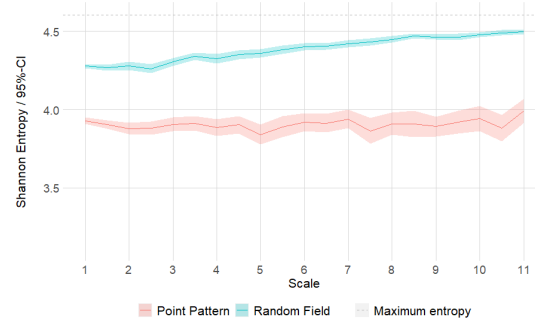
(a) Shannon Entropy as a function of the variance parameter (Scenario 1).



(b) Shannon entropy as a function of the mean parameter (Scenario 2).



(c) Shannon entropy as a function of the smoothness parameter (Scenario 3).



(d) Shannon entropy as a function of the scale parameter (Scenario 4).

Figure 3.11: Shannon entropy as a function of the variance, mean, smoothness and scale parameters.

The behavior of the processes in terms of the exponential LMC complexity (Fig-

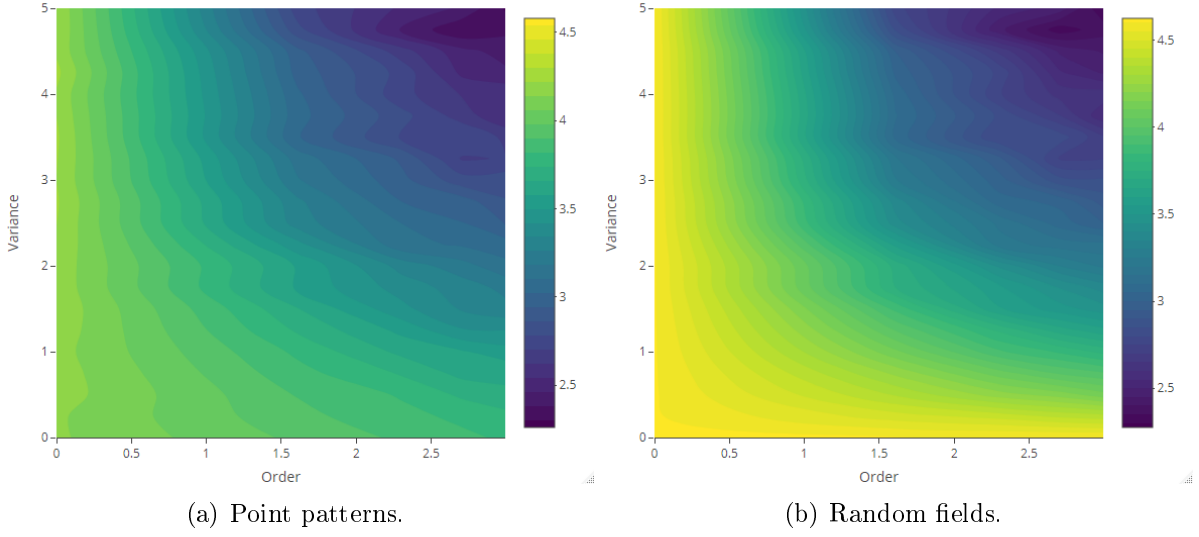


Figure 3.12: Rényi entropy of log-Gaussian Cox processes as a function of the variance parameter (Scenario 1) and order $\alpha \in \{0, \ln 5, \ln 10, \ln 15, \ln 20\}$.

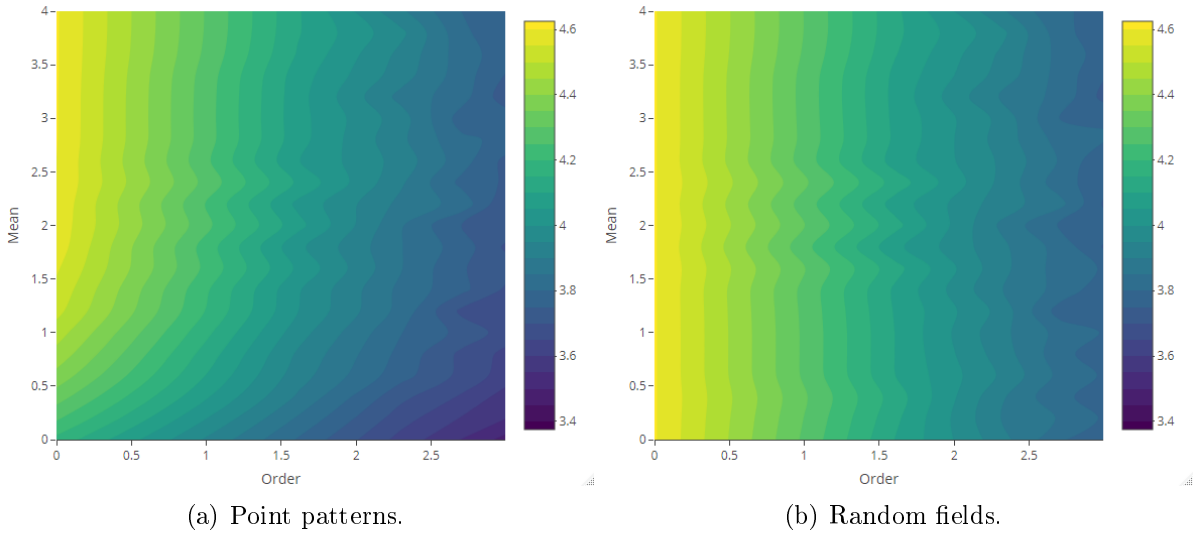


Figure 3.13: Rényi entropy of log-Gaussian Cox processes as a function of the mean parameter (Scenario 2) and order $\alpha \in \{0, \ln 5, \ln 10, \ln 15, \ln 20\}$.

ure 3.16) reflects, in general, an increase in structuring from the intensity field to the generated point pattern (the inversion observed in Figure 3.16(a) in relation to the variance parameter is related to the quadrat-counting procedure and the low mean value

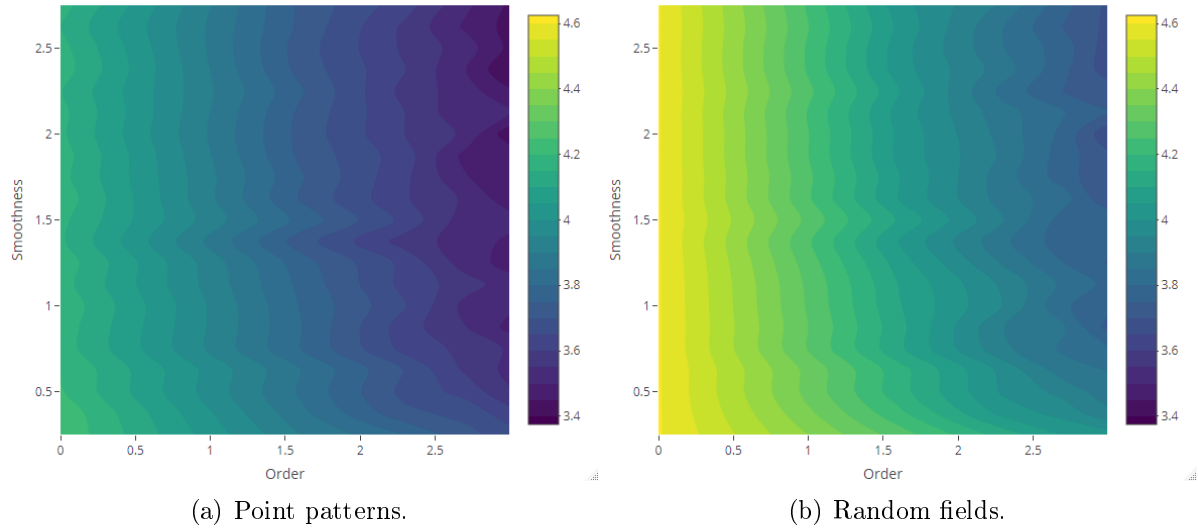


Figure 3.14: Rényi entropy of log-Gaussian Cox processes as a function of the smoothness parameter (Scenario 3) and order $\alpha \in \{0, \ln 5, \ln 10, \ln 15, \ln 20\}$.

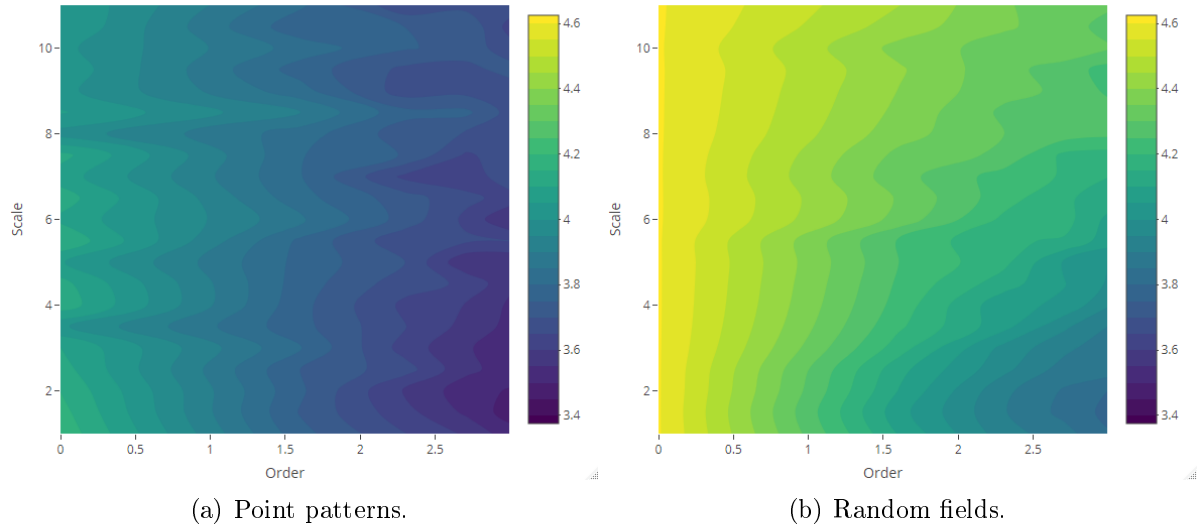
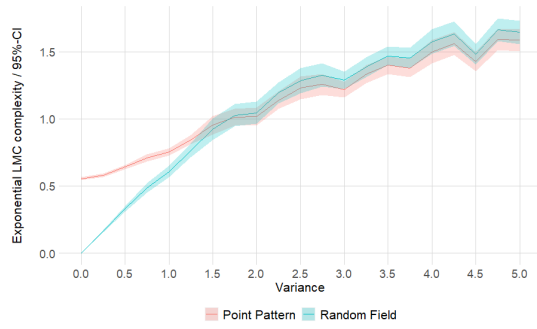


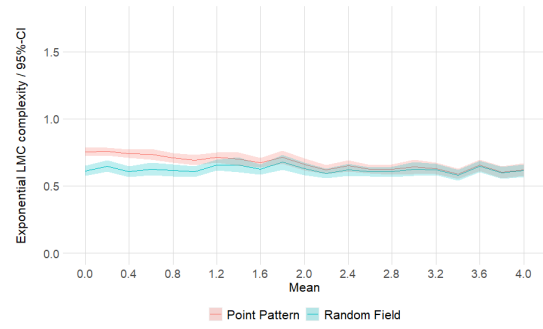
Figure 3.15: Rényi entropy of log-Gaussian Cox processes as a function of the scale parameter (Scenario 4) and order $\alpha \in \{0, \ln 5, \ln 10, \ln 15, \ln 20\}$.

specified).

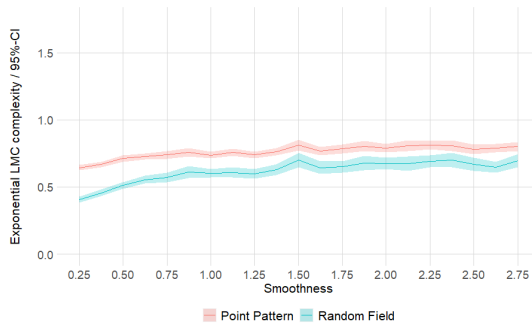
The maps of the generalized complexity (Figure 3.17) show the structural variation of each system for different combinations of the deformation parameters α and β .



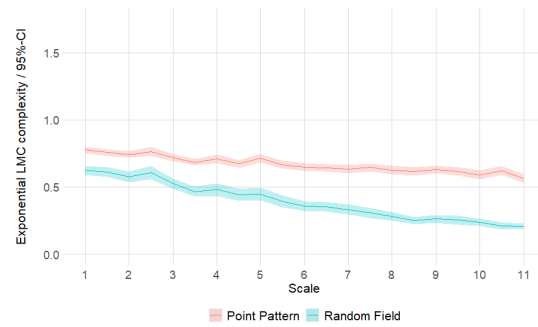
(a) Exponential LMC complexity as a function of the variance parameter (Scenario 1).



(b) Exponential LMC complexity as a function of the mean parameter (Scenario 2).

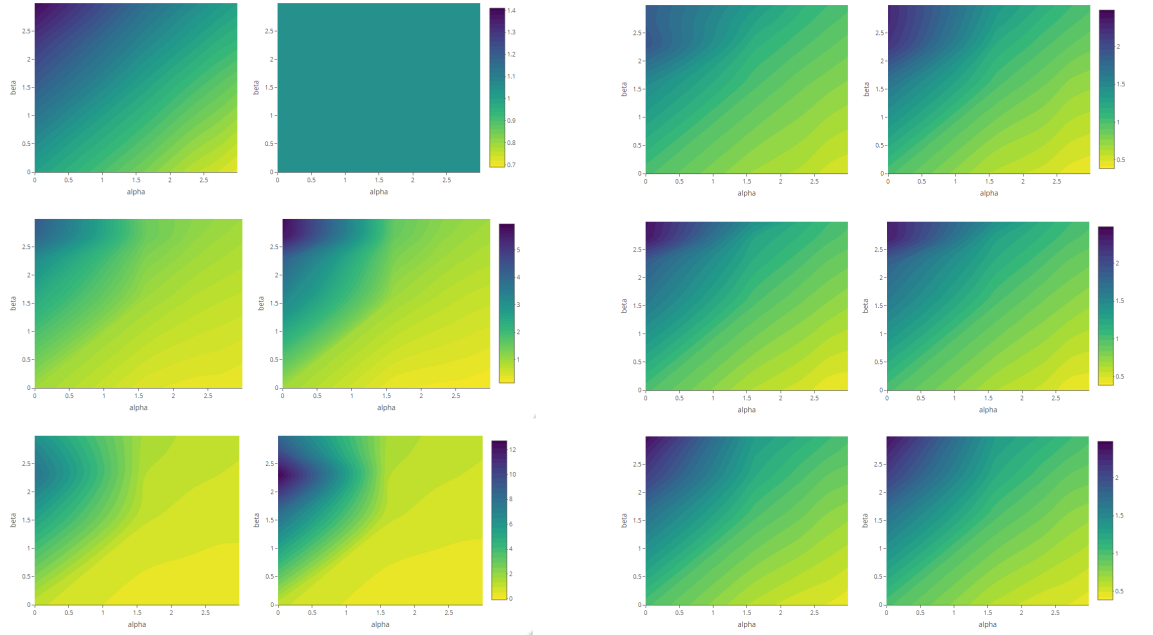


(c) Exponential LMC complexity as a function of the smoothness parameter (Scenario 3).



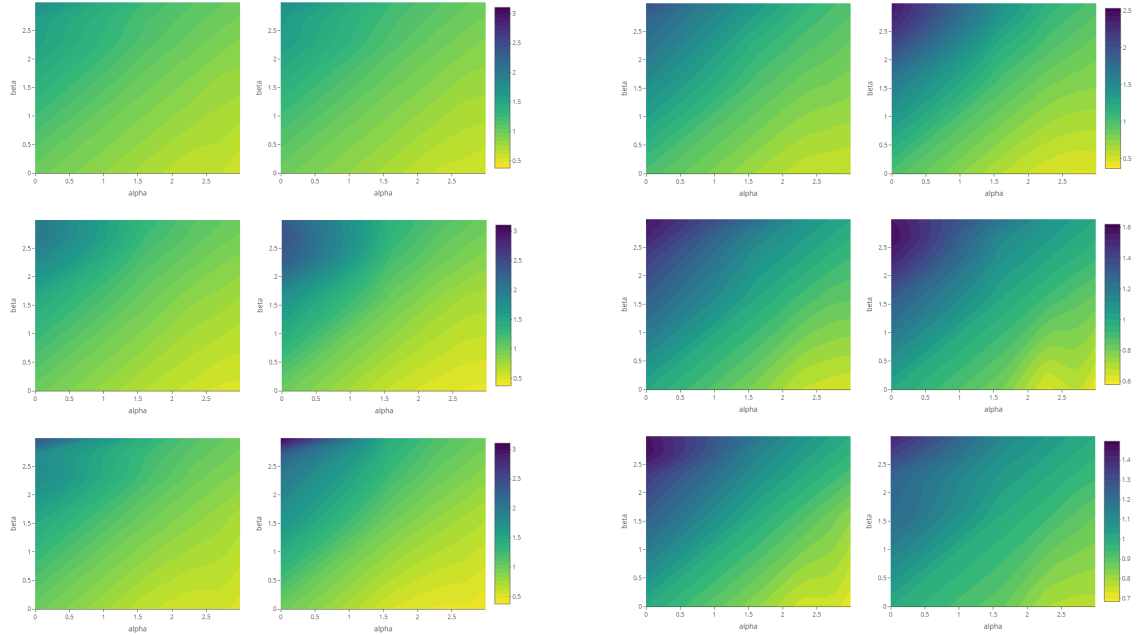
(d) Exponential LMC complexity as a function of the scale parameter (Scenario 4).

Figure 3.16: Exponential LMC complexity as a function of the variance, mean, smoothness and scale parameters.



(a) Generalized complexity with $\sigma^2 = 0$ (top), $\sigma^2 = 2.5$ (center), $\sigma^2 = 5$ (bottom).

(b) Generalized complexity with $\mu = 0$ (top), $\mu = 2$ (center), $\mu = 4$ (bottom).



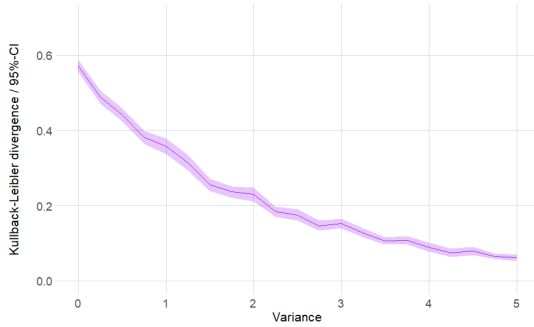
(c) Generalized complexity with $\nu = 0.25$ (top), $\nu = 1.375$ (center), $\nu = 2.75$ (bottom).

(d) Generalized complexity with $\rho = 1$ (top), $\rho = 5.5$ (center), $\rho = 11$ (bottom).

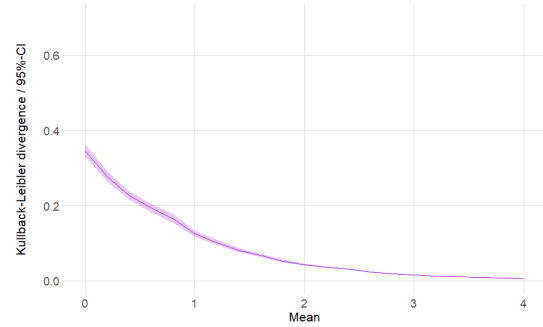
Figure 3.17: Generalized complexity as a function of deformation parameters α and β (with fixed parameter values according to each scenario, see Table 3.1).

3.3.3 Joint analysis

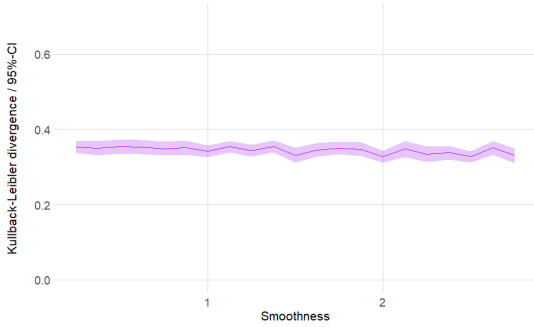
In this section we focus on the divergence-based measures which allow the local comparison of two probability distributions in each possible state of the system. In particular we aim to compare the distributions of the intensity fields and the point patterns to assess the state-by-state information transfer, and hence structural contribution, between the two phases.



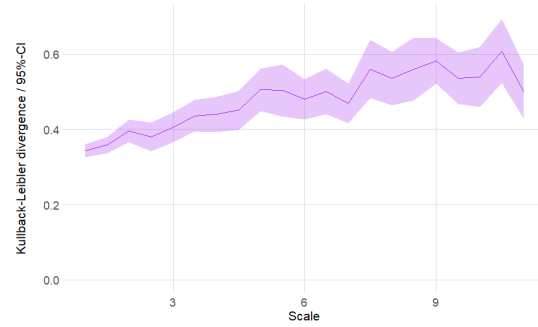
(a) Kullback-Leibler divergence as a function of the variance parameter (Scenario 1).



(b) Kullback-Leibler divergence as a function of the mean parameter (Scenario 2).



(c) Kullback-Leibler divergence as a function of the smoothness parameter (Scenario 3).



(d) Kullback-Leibler divergence as a function of the scale parameter (Scenario 4).

Figure 3.18: Kullback-Leibler divergence between point patterns and intensity random fields as a function of the variance, mean, smoothness and scale parameters.

As we have seen in the previous section, Shannon entropy for the distributions of intensity fields and corresponding point patterns in Scenarios 1 and 2 (Figures 3.11(a) and 3.11(b)) converges globally as the variance and mean parameter values increase, respectively. The Kullback-Leibler divergence shows that there exists as well an increasing local coherence in the transfer of information (Figures 3.18(a), 3.18(b)). On the other

hand, in Scenario 3, divergence essentially remains steady with respect to changes in the smoothness parameter (Figure 3.18(c)), which suggests that, under the quadrat-based approach adopted, the structural transfer has low sensitivity to variations in the smoothness parameter. In Scenario 4, the loss of the structure in the intensities induced by high magnitudes of the scale parameter increases the stochasticity in the generation of the patterns in the second phase (Figure 3.18(d)).

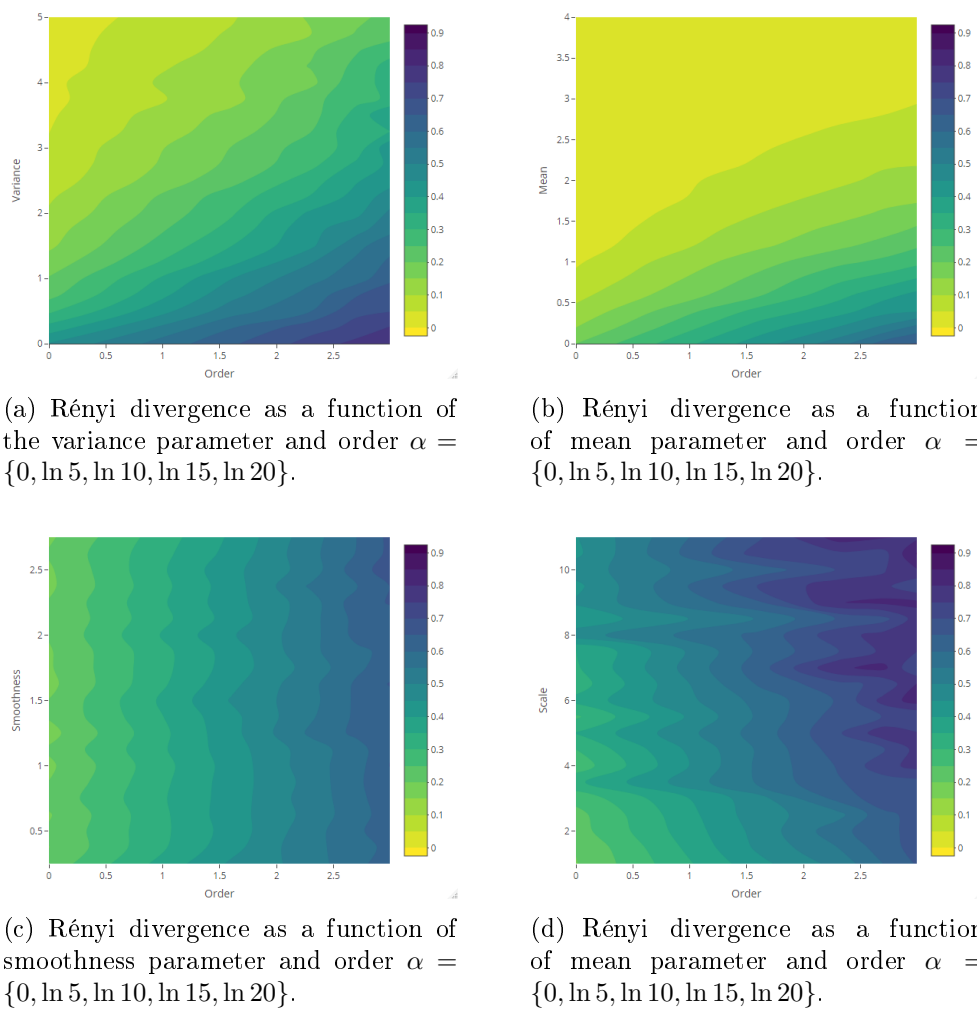


Figure 3.19: Rényi divergence between point patterns and random fields as a function of variance, mean, smoothness and scale parameters (Scenarios 1-4) and order $\alpha \in \{\log 0, \log 5, \log 10, \log 15, \log 20\}$.

The maps representing the Rényi divergences in Figure 3.19) allow to visualize how

the distortion induced by the deformation parameter is reflected in the structural local departure of the point patterns from the intensity fields. Similarly, those referring to the generalized relative complexity show the sensitivity of Rényi divergences with respect to incremental changes in the deformation parameter (Figures 3.20 - 3.23).

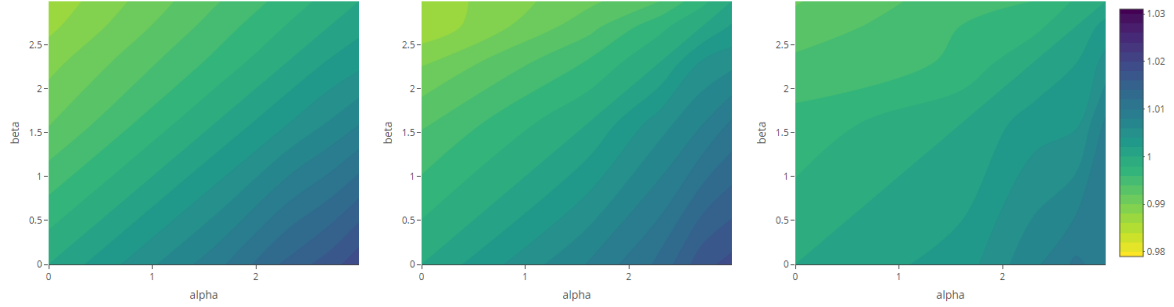


Figure 3.20: Generalized relative complexity as a function of variance parameter, for $\sigma^2 = 0$ (left), $\sigma^2 = 2.5$ (center) and $\sigma^2 = 5$ (right) and orders $\alpha, \beta \in \{0, 0.6, 1.2, 1.8, 2.4\}$ (fixed parameters as in Scenario 1).

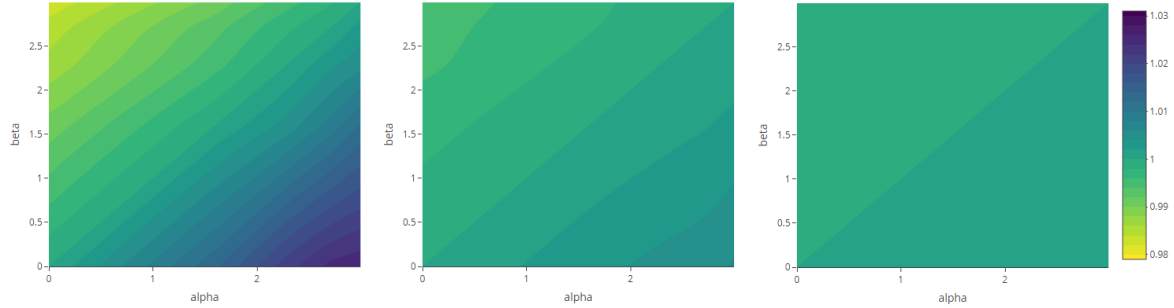


Figure 3.21: Generalized relative complexity as a function of the mean parameter $\mu = 0$ (left), $\mu = 2$ (center) and $\mu = 4$ (right) and orders $\alpha, \beta \in \{0, 0.6, 1.2, 1.8, 2.4\}$ (fixed parameters as in Scenario 2).

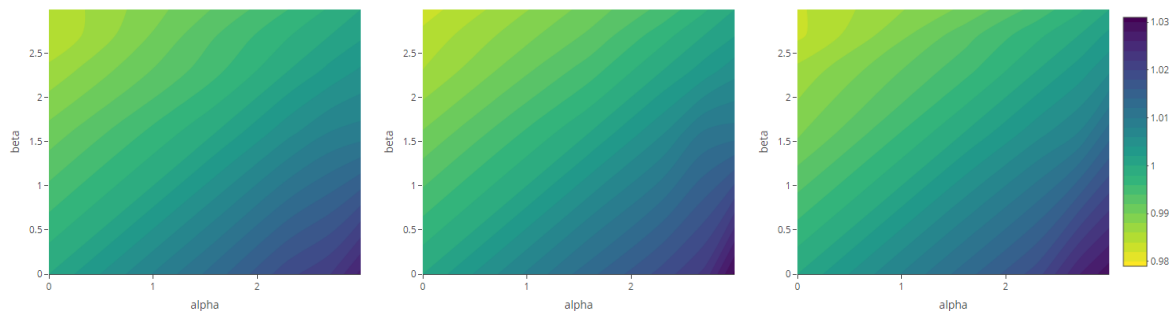


Figure 3.22: Generalized relative complexity as a function of the smoothness parameter $\nu = 0$ (left), $\nu = 2.5$ (center) and $\nu = 5$ (right) and orders $\alpha, \beta \in \{0, 0.6, 1.2, 1.8, 2.4\}$ (fixed parameters as in Scenario 3).

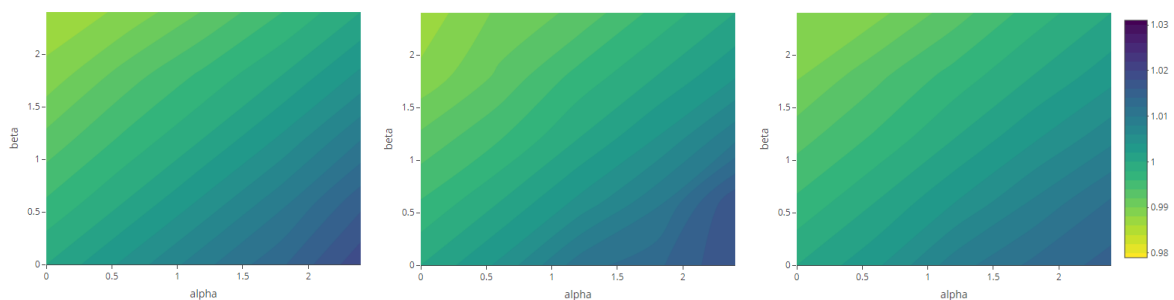


Figure 3.23: Generalized relative complexity as a function of the scale parameter $\rho = 1$ (left), $\rho = 5$ (center) and $\rho = 11$ (right) and orders $\alpha, \beta \in \{0, 0.6, 1.2, 1.8, 2.4\}$ (fixed parameters as in Scenario 4).

3.4 Synthesis

As a result from the study performed, it can be emphasized that the information transfer and, in general, the structure of the processes are mainly determined by the values of the mean, variance and scale parameters. Conversely, the value of the smoothness parameter does not have a perceptible effect on the structural information transferred between the phases.

In particular, among other aspects, the analysis shows a significant increase of the system complexity and a loss of diversity for large magnitudes of the variance parameter (both, marginally, for the intensity field and point pattern). Regarding the local

coherence measured in terms of divergence, it can be observed an enlargement of the structural information transferred as a result of the increase of the mean and variance parameter values. On the other hand, the increase of the scale parameter results in the loss of structure of the intensity field, enlarging the stochasticity inherent to the pattern generation, with an associated increase of the divergence between both phases.

Chapter 4

An Approach to Model Log-Gaussian Cox Processes with Fully Non-Separable Structures

Log-Gaussian Cox processes provide a flexible class of spatio-temporal models which allow the description of a wide variety of complex dependence structures in form of clustering in point patterns. The clustering effect observed in these patterns can be described by the inclusion of random heterogeneities through an unobservable intensity function. Typically, spatio-temporal point patterns have been modeled considering a separable structure in space and time, following pragmatic reasons, that in many cases were non-realistic. In this chapter, we overcome this way of proceeding by proposing a model which combines a non-separable structure for the first-order intensity function, with a non-separable correlation structure for the underlying random field. This methodology allows to underpin the interaction of the spatial and temporal components, while better mimicking the intrinsic behavior of the events in space and time. Global and local weighted second-order statistics are used to assess the improvement in the predictive performance of the proposed model compared to that obtained with a separable assumption. We use our modeling strategy to analyze the behavior of forest fires in Nepal, where global trend in the intensity function is modeled using external covariate information.

4.1 Introduction

Wildfires are considered one of the major threats to forest areas, as they are responsible for several harms into ecosystems by reducing biodiversity, damaging soils and landscapes, increasing desertification and inducing floods. Forest fires are strongly influenced by weather conditions; rainfall and humidity reduce the likelihood of fires, while increases in temperature and wind speed have a strong opposite effect (Flannigan and Wotton, 2001⁷⁴). In Nepal, where the data here analyzed comes from, the main season for forest fires is the pre-monsoon season between March and May (Parajuli *et al.*, 2015⁷⁵), which is the hottest period in the region. Here, the number of fires increases year after year, with a notable peak in 2016, when the number of fires and burned surface was greater than in previous years (Bhujel *et al.*, 2017⁷⁶). Future scenarios with an increasing temporal trend of forest fires can be explained by the cause of climate change (Pechony and Shindell, 2010⁷⁷). In order to prevent their occurrence, a good understanding of the spatio-temporal dynamics of fire patterns is the key to identifying areas at risk of wildfire.

Spatio-temporal point processes have raised as one of the most common approaches to studying wildfires (Juan *et al.*, 2012⁷⁸; Pereira *et al.*, 2013⁷⁹; Møller and Díaz-Avalos, 2010⁸⁰), as these are usually represented by their spatial coordinates and temporal instants. These processes can also be complemented by information about factors that can influence the occurrence of events, which are introduced in the form of spatial and temporal covariates (Serra *et al.*, 2013⁸¹, 2014⁸²). This is the case of vegetation, meteorology or topography, when dealing with forest fires. In addition, these processes are designed to depict and model the interaction structure between the spatial and temporal components, through non-separable (sometime complex) covariance structures (see for example Pereira *et al.*, 2013⁷⁹ and Gabriel *et al.*, 2017⁸³).

The clustering structure of wildfires can be attributed to the uneven distribution of risk factors across space and time, leading to an inhomogeneous latent intensity that drives the occurrence of the events. Consequently, wildfire modeling requires the inclusion of covariates to effectively capture the influencing factors. However, relying solely on observed covariates may not fully describe the spatio-temporal pattern. It is therefore useful to consider the inclusion of spatio-temporal random effects in the model to capture the residual variability.

The above aspects suggests that, in order to deal with modeling of this type of events, the use of an inhomogeneous point pattern model is clearly appropriate. In particular, the class of log-Gaussian Cox processes is often used to model these cases of environmentally-driven aggregated events, as they allow capturing a wide variety of dependence effects through the inclusion of random heterogeneities in a non-observable, underlying intensity field (Serra *et al.*, 2014⁸²; Illian *et al.*, 2010⁶⁷; Gabriel *et al.*, 2017⁸³; Pereira *et al.*, 2013⁷⁹).

Serra *et al.* (2014)⁸² studied wildfires in Catalonia by using spatio-temporal log-Gaussian Cox processes. Opitz *et al.* (2020)⁸⁴ fitted a spatio-temporal log-Gaussian Cox process using Bayesian inference, introducing a stratified subsampling technique in order to deal with the high dimensionality of the data. Kelling and Haran (2020)⁸⁵ also use this type of processes to measure the effect of multiple types of marks and spatial covariates in the occurrence of forest fires in the region of Castilla-La Mancha.

A common approach in all these references is that they assume separability in space and time, meaning that the spatial structure is repeated in time, under a certain independence structure. Just some few authors have considered non-separable models for the description and prediction of point processes. Gilardi *et al.* (2021)⁸⁶ proposed a non-separable first-order intensity function for the analysis of ambulance interventions, while Siino *et al.* (2018)⁸⁷ developed a method for estimating the covariance parameters of spatio-temporal log-Gaussian Cox processes applied to the description of earthquake series. In a more theoretical context, Kolovos *et al.* (2004)⁸⁸ presented new methods for the construction of non-separable space-time covariance models with applications to environmental processes. However, to our knowledge, a fully non-separable structure that takes into account space-time interactions in both the long-term trend and the local structure has not been developed in general and, in particular, in modeling wildfires.

In this chapter, starting from the assumption that the occurrence of forest fires can be modelled using the class of log-Gaussian Cox processes, we propose a model with a double non-separable structure, combining both a non-separable first-order intensity function and a non-separable correlation structure for the underlying random field. The non-separable deterministic intensity is constructed as the product of a temporal component modeled by linear regression using meteorological covariates, and a spatio-temporal term obtained by combining a weighted non-parametric kernel estimator with a spatial regression model. Furthermore, for the stochastic component we consider a non-separable DeIaco-Cesare covariance function and estimate its parameters using a minimum contrast

procedure. We apply the proposed model to a dataset of forest fires in Nepal, generating risk maps for future events by performing predictions, and we probe its predictive performance by comparing the results with those obtained through a model with a separable structure and a mixed model which combines a separable deterministic component and a non-separable covariance structure.

4.2 Data: forest fires and meteorological covariates

The main dataset is provided by ICIMOD (International Centre for Integrated Mountain Development), and covers forest fires that occurred in Nepal from February 2012 to June 2016. Each event is spatially georeferenced together with the temporal instant, having in total 13781 fires, distributed as 3149 in 2012, 1999 in 2013, 2386 in 2014, 489 in 2015, and 5758 in 2016.

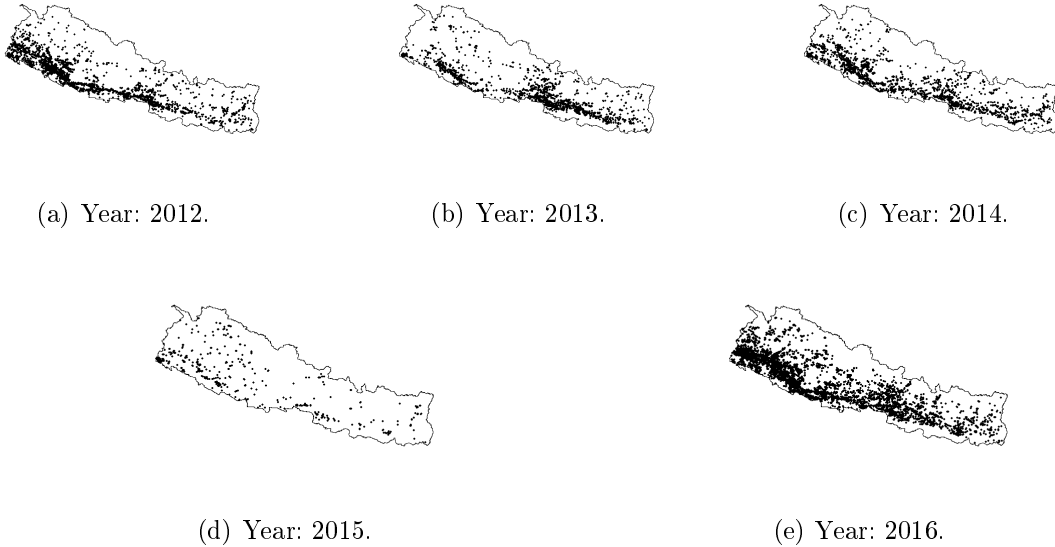


Figure 4.1: Spatial locations of forest fires in Nepal by year.

The study area corresponds to the whole country of Nepal, with an extension of 885 km from east to west and 241 km from north to south. The spatial distribution of the forest fires is shown in Figure 4.1, with a clear larger incidence in areas of low altitude areas (see Figure 4.2).

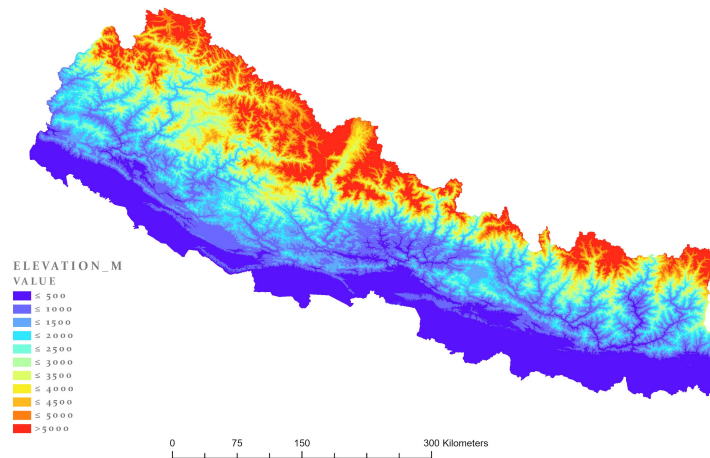


Figure 4.2: Digital elevation model (DEM) of Nepal.

The temporal dynamics of the data (figure 4.3) shows a decreasing trend until 2016, when forest fires reach a global maximum. Also, the temporal structure seems to show seasonality, with an increasing number of fires from April to July, and maximum peaks in May.

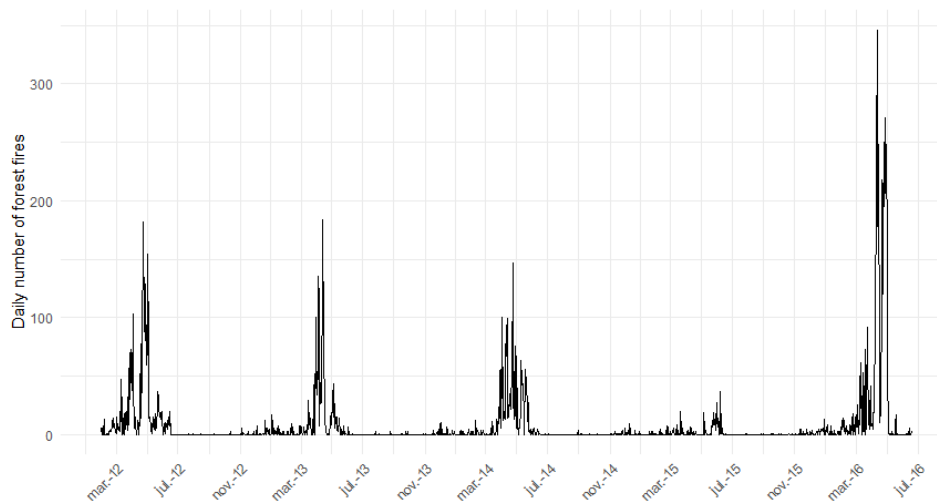


Figure 4.3: Time series of the daily number of forest fires in Nepal from February 2012 to June 2016.

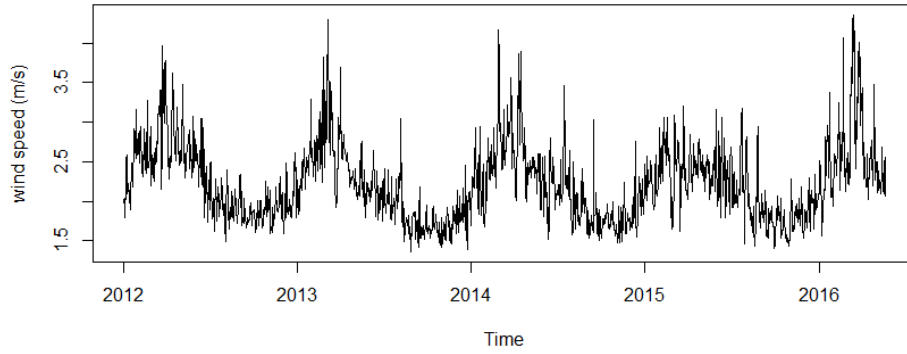
Weather is a major determinant of wildfire incidence, with regions experiencing high temperatures, strong winds, and low precipitation being particularly susceptible to forest

fires (Opitz *et al.*, 2020⁸⁴). We use data obtained from *opendatanepal.com*, that contains weather information for each of Nepal's 62 districts from January 1981 to December 2019. Fields in the dataset include date by day, district, precipitation (mm), temperature (C) and wind speed (m/s). Figure 4.4 shows the daily time series of the meteorological covariates, while Figure 4.5 shows their corresponding spatial distribution. The temporal trend of the covariates runs accordingly to the behavior observed in the time series of the fires (see Figure 4.3). Peaks in wind speed in May contribute to the observed seasonality in the fires. In addition, a decreasing trend can be observed from 2013 to 2015, followed by a notable increase in wind speed in 2016, as depicted in Figure 4.3. Conversely, precipitation shows a decreasing trend, reaching its minimum value in 2016. Temperature exhibits annual seasonality, with milder temperatures coinciding with fire season. Spatially, fires are mainly concentrated in regions characterized by high and mild temperatures, lower precipitation, and areas with high wind speeds with the exception of mountainous areas.

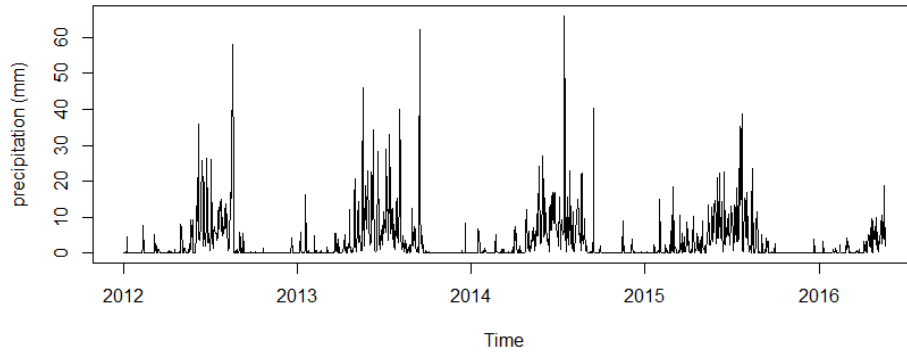
4.3 Statistical model and inference

As mentioned in Section 2.2.4, log-Gaussian Cox Processes (LGCP) define a powerful class of point processes that extend the traditional Cox process by incorporating a log-Gaussian random field as the underlying intensity function. Originally developed for the spatial context (Møller *et al.*, 1998⁴⁴), LGCPs have since been extended to the spatio-temporal domain, making them suitable for scenarios where the dynamics of events evolve both in time and space (Diggle *et al.*, 2013). In a general Cox process, events occur at spatio-temporal position $(\mathbf{u}, t) \in W \times T$ according to an inhomogeneous spatio-temporal Poisson process with a stochastic intensity function $\Lambda(\mathbf{u}, t)$. This spatio-temporal intensity function can be expressed as the product of a deterministic component, accounting for spatial and temporal effects, and a log-Gaussian stochastic process that captures the uncertainty and variability in the event occurrence. Furthermore, integration of the intensity function across the domain $W \times T$ should yield the expected number of events in that region. To formalize this, and following the specifications in Diggle *et al.* (2013)⁸⁹, the spatio-temporal intensity function $\Lambda(\mathbf{u}, t)$ can be expressed as

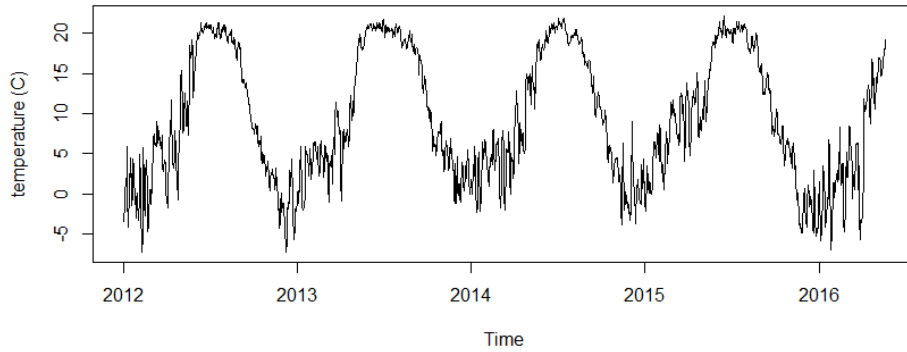
$$\Lambda(\mathbf{u}, t) = \lambda(\mathbf{u}, t) \exp \{G(\mathbf{u}, t)\}, \quad (4.1)$$



(a) Wind speed.



(b) Precipitation.



(c) Temperature.

Figure 4.4: Time series of the daily wind speed, precipitation and temperature in Nepal.

where G is a Gaussian process with $E[G(\mathbf{u}, t)] = -0.5\sigma^2$, so that $E[\exp\{G(\mathbf{u}, t)\}] = 1$ and variance-covariance matrix is given by $\text{Cov}(G(\mathbf{u}, t), G(\mathbf{v}, s)) = C(\|\mathbf{u} - \mathbf{v}\|, |t - s|) = C(r, h)$. Here, the stochastic intensity $\Lambda(\mathbf{u}, t)$ is split into a deterministic term $\lambda(\mathbf{u}, t)$ and a residual spatio-temporal variation given by $G(\mathbf{u}, t)$.

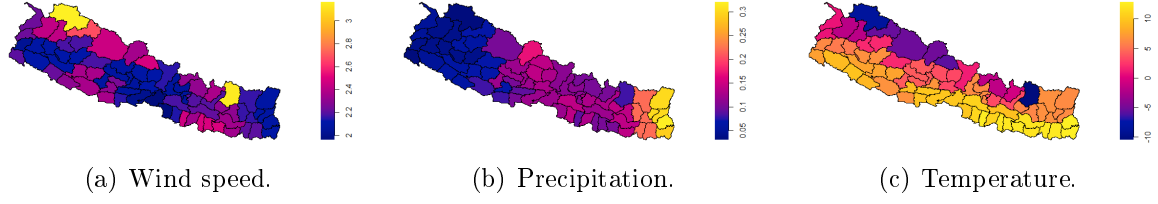


Figure 4.5: Median values of wind speed, precipitation and temperature in each district of Nepal.

We assume three scenarios regarding the degree of interaction between the spatial and temporal dimensions (see table 4.1), where the spatio-temporal intensity can be decomposed as the product of two terms as shown in (4.1) where $\lambda(\mathbf{u}, t)$ is the intensity of a realization of a non-homogeneous Poisson process in the region $\mathbf{u}$ at a time $t \in T$ and $\exp\{G(\mathbf{u}, t)\}$ is the residual spatio-temporal variation.

Table 4.1 depicts different scenarios we consider here in terms of separability, non-separability and mixed cases acting on $\lambda(\mathbf{u}, t)$ and $G(\mathbf{u}, t)$. Note that the temporal intensity $\mu(t)$ represents the expected number of events in W at time t , while $\lambda(\mathbf{u})$ is the overall aggregated spatial variation of the intensity in the region $\mathbf{u}$ at the time interval T . For the separable and mixed scenarios we assume, following a widely considered assumption for point processes, that the spatial component of the deterministic intensity function is independent of the temporal one. The first-order intensity of the non-separable scenario is intended to represent the space-time interaction present in the data, and it is decomposed as the product of $\mu(t)$ and a spatio-temporal density $\lambda_t(\mathbf{u})$ which reflects the spatial variation of the process in the region $\mathbf{u}$ at each time $t \in T$.

For the stochastic component, we consider several parametric families of covariance functions. For the separable case, the covariance function is typically assumed to be the product of a spatial and temporal covariances of the form $C(h, r) = \sigma^2 C(h) C(r)$. The most commonly used families include the exponential, Matérn and Gaussian (see, for example, Brix and Diggle, 2001⁴³; Serra *et al.*, 2014⁸²). For the non-separable cases, we consider the Delaco-Cesare covariance (see, for instance, Gneiting *et al.*, 2006⁹⁰; Delaco *et al.*, 2002¹⁹ and DeCesare *et al.*, 2001²⁵). Inference of unknown parameters follows the methodology outlined in Siino *et al.* (2018)⁸⁷.

In particular, the separable covariance function, considering an exponential form for

the spatial the temporal marginals, takes the form

$$C(h, r) = \sigma^2 \exp\left(-\frac{h}{\alpha}\right) \exp\left(-\frac{r}{\beta}\right), \quad (4.2)$$

with σ^2 the variance parameter, $\alpha > 0$ the scale parameter for the spatial component, and $\beta > 0$ the scale parameter for the temporal component. The non-separable Delacoe-Cesare covariance function can be written as

$$C(h, r) = \sigma^2 \left(1 + \left(\frac{h}{\alpha}\right)^{\gamma_s} + \left(\frac{r}{\beta}\right)^{\gamma_t}\right)^{-\delta}, \quad (4.3)$$

with $\alpha > 0$, $\beta > 0$ scale parameters of space and time respectively, $\gamma_s \in (0, 2]$, $\gamma_t \in (0, 2]$, and $\delta \geq 1$ determines the degree of interaction of the spatio-temporal dimensions.

Scenario	Deterministic intensity	Covariance model
Separable	$\lambda(\mathbf{u}, t) = \lambda(\mathbf{u}) \mu(t)$	$C(h, r) = \sigma^2 \exp\left(-\frac{h}{\alpha}\right) \exp\left(-\frac{r}{\beta}\right)$
Mixed	$\lambda(\mathbf{u}, t) = \lambda(\mathbf{u}) \mu(t)$	$C(h, r) = \sigma^2 \left(1 + \left(\frac{h}{\alpha}\right)^{\gamma_s} + \left(\frac{r}{\beta}\right)^{\gamma_t}\right)^{-\delta}$
Non-separable	$\lambda(\mathbf{u}, t) = \lambda_t(\mathbf{u}) \mu(t)$	$C(h, r) = \sigma^2 \left(1 + \left(\frac{h}{\alpha}\right)^{\gamma_s} + \left(\frac{r}{\beta}\right)^{\gamma_t}\right)^{-\delta}$

Table 4.1: Deterministic first-order intensity functions, and covariance models of the Gaussian random fields used to model an LGCP for the separable, mixed and non-separable scenarios.

4.3.1 Deterministic component

The temporal intensity $\mu(t)$ represents the expected number of forest fires that occurred in the spatial domain W during the time interval T . In all considered scenarios, this intensity is estimated through a generalized linear model with Poisson response $n(t)$, using precipitation, wind speed and temperature as linear covariates. We use the month of the year as a dummy variable taking January as reference, and a seasonal component corresponding to a periodicity of a year. The estimated coefficients and their significance are detailed in Table 4.2). If $n(t)$ stands for the observed number of forest fires in W at time t , then $\mu(t) = E(n(t))$, and the log-linear regression with a Poisson response can be expressed as:

$$\begin{aligned} \log(\mu(t)) = & \beta_0 + \text{month}(t) + \text{wind}(t) + \text{temp}(t) + \text{prec}(t) + \\ & + \sin\left(\frac{2\pi t}{365}\right) + \sin\left(\frac{4\pi t}{365}\right) + \cos\left(\frac{2\pi t}{365}\right) + \cos\left(\frac{4\pi t}{365}\right) \end{aligned} \quad (4.4)$$

where $\text{month}(t)$ is a factor representing the month of the year, $\text{wind}(t)$, $\text{temp}(t)$ and $\text{prec}(t)$ are the median value of wind speed, temperature and precipitation over the spatial districts spanning Nepal at time t , $\sin(\cdot)$ and $\cos(\cdot)$, correspond to annual cycles, and t comes in days. Figure 4.6 depicts the observed and predicted number of forest fires.

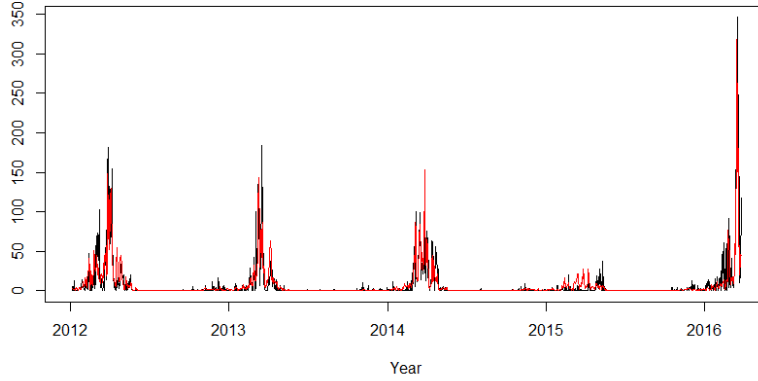


Figure 4.6: Observed daily forest fires (black line) and fitted values (red line) using a generalized linear model.

On the other hand, the spatial component $\lambda(\mathbf{u})$ describes the spatial variation in the intensity of forest fires, and it should satisfy the following conditions:

1. $\lambda(\mathbf{u}) > 0, \forall \mathbf{u} \in W$
2. $\int_W \lambda(\mathbf{u}) d\mathbf{u} = 1$

Its estimation for the separable and mixed scenarios is performed by using a generalised linear model with Poisson response representing the number of forest fires in each cell of a 128×128 pixel grid in W , using as regressors some weather covariate maps (Figure 4.5(a)-4.5(c)) containing the median values of wind speed, precipitation

	Coefficients	Prob. z
intercept	0.0552	0.8189
February	0.3469	0.0453
March	0.5169	0.0249
April	0.9497	< 0.001
May	0.5272	0.0317
June	1.0432	< 0.001
July	-0.3423	0.5317
August	-0.5282	0.3960
September	-1.6885	0.0067
October	-3.4002	< 0.001
November	-1.6442	< 0.001
December	0.1906	0.3661
wind	0.8903	< 0.001
precipitation	-0.0872	< 0.001
temperature	-0.1732	< 0.001
$\sin(2\pi t/365)$	-0.1123	0.6646
$\sin(4\pi t/365)$	0.5304	< 0.001
$\cos(2\pi t/365)$	-0.7993	< 0.001
$\cos(4\pi t/365)$	-1.1490	< 0.001

Table 4.2: Estimated coefficients and associated p-values for the fitted generalized linear model for the daily temporal intensity.

and temperature. These covariates are evaluated across the pixel grid on W and over the whole temporal window. The log-linear regression is given by

$$\log(\lambda(\mathbf{u})) = \beta_0 + \text{wind}(\mathbf{u}) + \text{temp}(\mathbf{u}) + \text{prec}(\mathbf{u}), \quad (4.5)$$

where $\text{wind}(\mathbf{u})$, $\text{temp}(\mathbf{u})$ and $\text{prec}(\mathbf{u})$ are the median value of wind speed, temperature and precipitation respectively at location $\mathbf{u}$. The coefficients of the fitted model are shown in table 4.3 and the predicted spatial intensity is displayed in figure 4.7.

Finally, the estimation of $\lambda_t(\mathbf{u})$ for the non-separable scenario is performed in two steps. First, $\lambda_{t_{(1)}}(\mathbf{u})$ is obtained through a weighted kernel density estimation following the method proposed by Gilardi *et al.* (2021)⁸⁶. For a time period T , a day t and a

	Coefficients	Prob. z
intercept	9.1459	< 0.001
wind	−1.1987	< 0.001
precipitation	−8.6253	< 0.001
temperature	0.1821	< 0.001

Table 4.3: Estimated coefficients and p-values of the fitted generalized linear model for the number of forest fires per day in Nepal using the meteorological covariates.

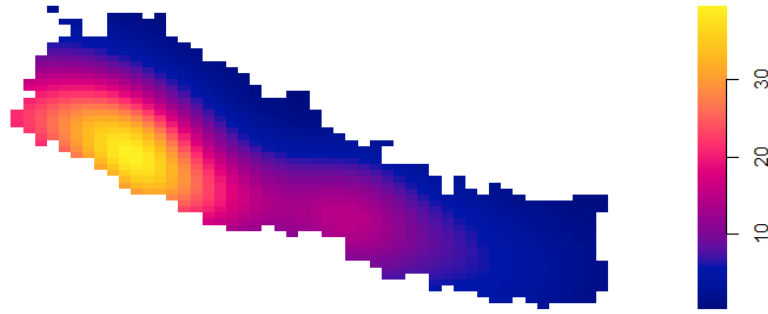


Figure 4.7: Predicted spatial intensity from general linear model with Poisson response and meteorological covariates as regressors.

location $\mathbf{u}$, we have

$$\lambda_{t(1)}(r) = \frac{\sum_{r \in T} \sum_{i=1}^{n(r)} w(r, t) K(\mathbf{u}, \mathbf{u}_{r,i}, s)}{\sum_{r \in T} \sum_{i=1}^{n(r)} w(r, t)}, \quad (4.6)$$

where $w(r, t)$ is a weight vector for the $\mathbf{u}_{r,i}$ forest fire and $K(\mathbf{u}, \mathbf{u}_{r,i}, s)$ is the Jones-Diggle kernel function with bandwidth s .

The weight function $w(r, t)$ is modeled as a function of the temporal lags $m = r - t$ as follows:

$$w(r, t) = w(r - t) = w(m) = \rho_1^{(m)} + \rho_2^{(m)} \rho_3^{\sin^2(\frac{\pi m}{365})} \rho_4^{\sin^2(\frac{\pi m}{730})}, \quad (4.7)$$

where ρ_1 represents a short-term correlation, ρ_3 and ρ_4 control for the annual and biennial seasonality, respectively (see Figure 4.8) and ρ_2 is a discount factor. The estimation of the parameters $\rho_1, \dots, \rho_4$ is obtained by minimizing

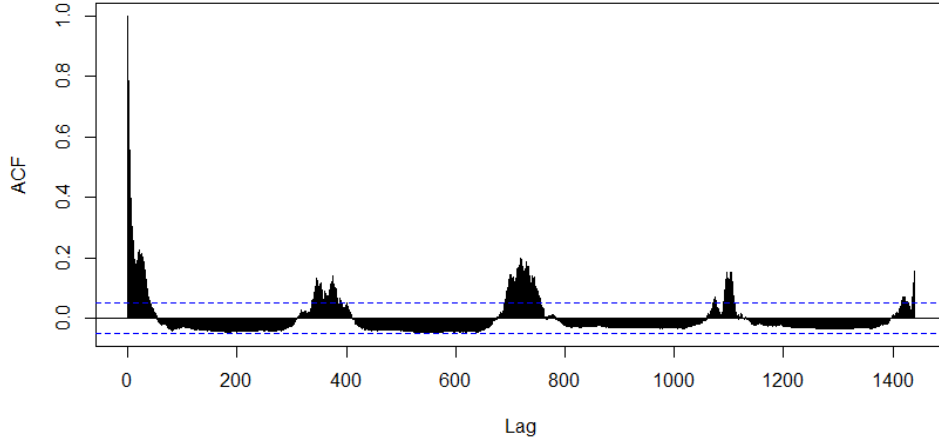


Figure 4.8: Daily autocorrelation function of forest fires with a lag of 1541 days.

$$\sum_{m=1}^M (ACF^+(m) - \rho_0 w_1(m))^2 \quad (4.8)$$

with $ACF^+(m)$ the positive part of the autocorrelation function for a lag m and $\rho_j \in [0, 1]$ for all $j = 0, \dots, 4$, with ρ_0 and additional discount factor without any practical interpretation, it is introduced to scale $w(\cdot)$ between 0 and 1, ensuring consistency with the ordinate range of the ACF^+ .

The estimated coefficients, using the algorithm proposed by Zhou *et al.* (2015)⁹¹, result in a high short-term correlation $\hat{\rho}_1 = 0.6864$ with a daily discount factor $\hat{\rho}_0 = 0.2250$. Also, the seasonal components equal $\hat{\rho}_3 = 0.029$ and $\hat{\rho}_4 = 0.5741$, and annual and biennial seasonalities vanish slowly with $\hat{\rho}_2 = 0.9996$. The observed ACF^+ and the estimated weighted function $w(h, t)$ are shown in Figure 4.9.

In a second step, we obtain $\lambda_{t(2)}(\mathbf{u})$ using the coefficients of the general linear model fitted in Table 4.3 for each temporal instant t . Then, the deterministic intensity for the non-separable scenario takes the form

$$\lambda_t(\mathbf{u}) = \left(\lambda_{t(i)}^*(\mathbf{u}) \cdot \lambda_{t(2)}^*(\mathbf{u}) \right)^{1/2},$$

where $\lambda_{t(i)}^*(\mathbf{u})$ for $i = 1, 2$ are the normalized intensities.

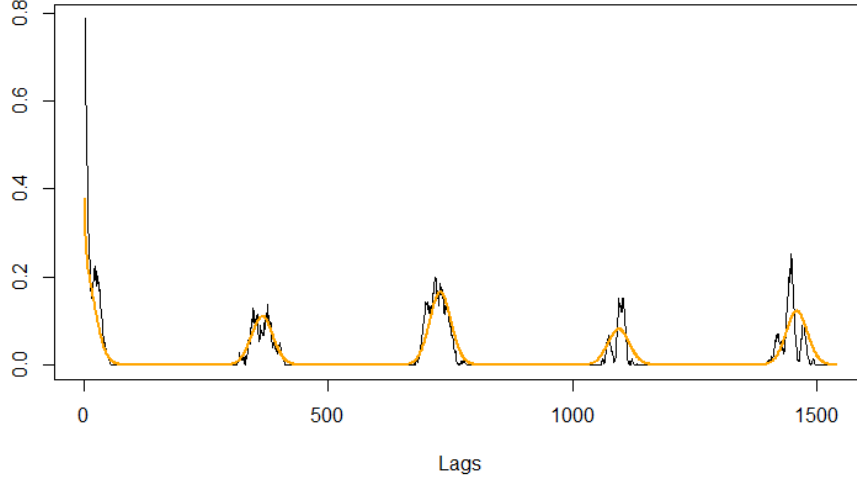


Figure 4.9: Positive part of ACF (black line) and estimated weight function $w(h, t)$ (orange line).

4.3.2 Stochastic component

The Gaussian process $G(\mathbf{u}, t)$ depends on the covariance structure given by $C(r, h)$, and whose parameters are estimated following the procedure proposed in Siino *et al.* (2018)⁸⁷. We indeed minimize the quadratic distance between the empirical and theoretical parametric forms of the pair correlation function $g(r, h)$ with respect to the vector of parameters θ , as follows:

$$\int_{r_0}^{r_{\max}} \int_{h_0}^{h_{\max}} \{\hat{g}(r, h) - g(r, h; \theta)\}^2 dh dr, \quad (4.9)$$

where $r_{\max} = 1/4 \cdot \text{lag}_{\max_s}$, $h_{\max} = 1/4 \cdot \text{lag}_{\max_t}$, with $\text{lag}_{\max_s}$ and $\text{lag}_{\max_t}$ being the maximum spatial and temporal lags, respectively (see Gabriel, 2012⁹²). We use an Epanechnikov kernel to compute the spatial component and a box kernel for the temporal component, setting r_0 and h_0 slightly larger than the spatial and temporal bandwidths used to estimate $\hat{g}(r, h)$. The estimator of the pair correlation function, proposed by Gabriel and Diggle (2009)³⁹, can be expressed as

$$\hat{g}(r, h) = \frac{1}{4\pi r} \sum_{i=1}^n \sum_{j \neq i} \frac{1}{w_{ij}} \frac{k_{\epsilon}(r - \|\mathbf{u}_i - \mathbf{u}_j\|) k_{\delta}(h - |t_i - t_j|)}{\hat{\lambda}(\mathbf{u}_i, t_i) \hat{\lambda}(\mathbf{u}_j, t_j)}, \quad (4.10)$$

where $\hat{\lambda}(\cdot)$ is the estimated first-order intensity, w_{ij} is an edge correction factor for spatial-temporal edge effects, and k_{ϵ} , k_{δ} are kernel functions with bandwidths ϵ and δ , respectively. These bandwidths are estimated using `dpik` function of the **R** package `kernsmooth` following the procedure in Siino *et al.* (2018)⁸⁷ and Sheather and Jones (1991)⁹³.

The estimated parameters come in Table 4.4. The assumption of non-separability leads to a higher value of σ^2 which induces the formation of point clusters. In addition, we obtain larger values for the scale parameter α compared to β , indicating a stronger spatial correlation than temporal correlation. Parameters γ_s , γ_t control the spatial and temporal smoothness, respectively, while δ determines the degree of interaction between the spatial and temporal dimensions. In the separable case, we also note a larger value of σ^2 , while in this case the spatial and temporal components have similar correlation scales.

To assess the goodness of fit of the stochastic component, we use envelopes based on the non-parametric second-order spatio-temporal statistic $\hat{K}(r, h)$. These envelopes are derived as the 95% critical values of the second-order statistic based on 100 simulated point patterns using the estimated parameters detailed in Table 4.4. These envelopes are compared with the empirical surface obtained by the values of $\hat{K}(r, h)$ from the observed data.

Figures 4.10(a) and 4.10(b) depict the envelopes based on K -functions for the separable and non-separable cases. We note that the non-separable model describes slightly better the second-order structure of the data, as most of the surface lies within the envelopes. Conversely, the separable model fails to adequately describe the observed data for a larger number of space and time distance combinations.

4.4 Goodness-of-fit analysis

The goodness of fit of the model is assessed according to the procedure outlined in Adelfio *et al.* (2020)⁴⁰, which is based on the behavior of the weighted second-order statistic K .

Covariance	$\hat{\sigma}^2$	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\gamma}_s$	$\hat{\gamma}_t$	$\hat{\delta}$
Exp-Exp	4.6674	0.0141	0.0099			
DeIaco-Cesare	7.4555	0.6619	0.0317	0.3733	1.3124	2.1900

Table 4.4: Parameter estimates of the covariance families for the separable and non-separable scenarios.

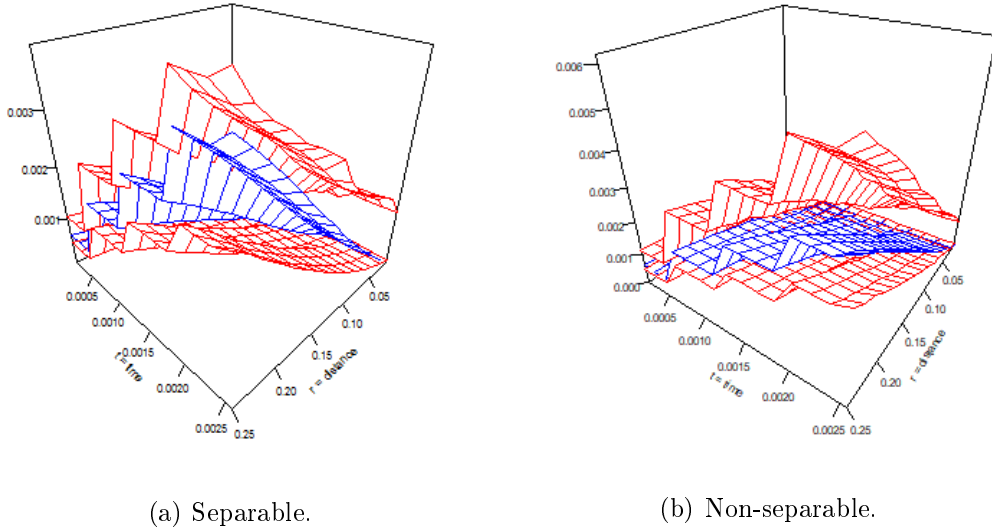


Figure 4.10: 95% envelope of the spatio-temporal K -function for separable (left) and non-separable (right) covariance functions, with the blue surface representing the estimated K -function for the observed data, and the red surfaces being the 95% envelopes obtained through simulations under the fitted model.

This diagnostic tool assesses the departure of the estimated K -function, weighted by the inverse of the estimated intensity of the process, from the theoretical one. Smaller discrepancies indicate that the estimated intensity is closer to the actual intensity of the process.

According to the results in Adelfio *et al.* (2020)⁴⁰, the expected value of the weighted K -function is equal to the expectation of the Equation (2.31) if the estimated $\hat{\lambda}$ is the driving intensity of the process. Specifically, for a Poisson process, then $E[K^w(r, h)] = \pi r^2 h$. Similarly, if the estimated intensity is the real intensity of the process, then $E[\hat{K}_i(r, h)] = E[\hat{K}_i^w(r, h)] = \pi r^2 h$. We measure this distance by calculating the χ^2 statistic of the difference between the observed and expected values of both the

local and global K -weighted statistics, deriving its distribution over 100 simulations (for further details, we refer the reader to Adelfio *et al.*, 2020).

4.4.1 Diagnosis results

In this section, the results obtained from assessing the goodness of fit of the model fitted under the three scenarios considered are presented. This analysis is conducted over 100 simulations of the fitted point process for each scenario. Additionally, for comparative purposes, we include a case with a constant intensity.

The diagnosis based on K -function reveals a better performance for the non-separable case. The global space-time diagnostic exhibits larger differences for the separable and mixed scenarios (Figures 4.11(a), 4.11(b), 4.11(c), and 4.11(d)). Specifically, the mean sum of residuals is 123.89, 124.02 and 93.85 for the separable, mixed and non-separable scenarios, respectively. Conversely, when considering a constant intensity as the generator of the process, we obtain a sum of residuals of 222.61. The χ^2 statistic of the differences yields lower values for the non-separable case, while the separable and mixed scenarios present similar performance (see Figure 4.12). Moreover, all scenarios considered outperform the use of a constant intensity, resulting in a statistic $\chi^2 = 45.43$.

Regarding the local space-time diagnostic, which reflects the behavior of the intensity when fitting each individual point of the process, we observe that the non-separable model consistently improves the results compared to the other models considered (see Figure 4.13(a)) with a 95% quantile lower than 0.0026. Contrary to the global diagnosis, the use of a non-separable covariance model enhances the local structure of the predicted data across all models considered. Moreover, each model provides better results compared to those obtained by using a constant intensity (see Figure 4.13(b)), for which the quantile is equal to 0.47.

4.5 Risk mapping

In this section, we present the results obtained through inference by using the estimated model from the previous section for each scenario.

The intensity function of the deterministic model is predicted for sixteen future time

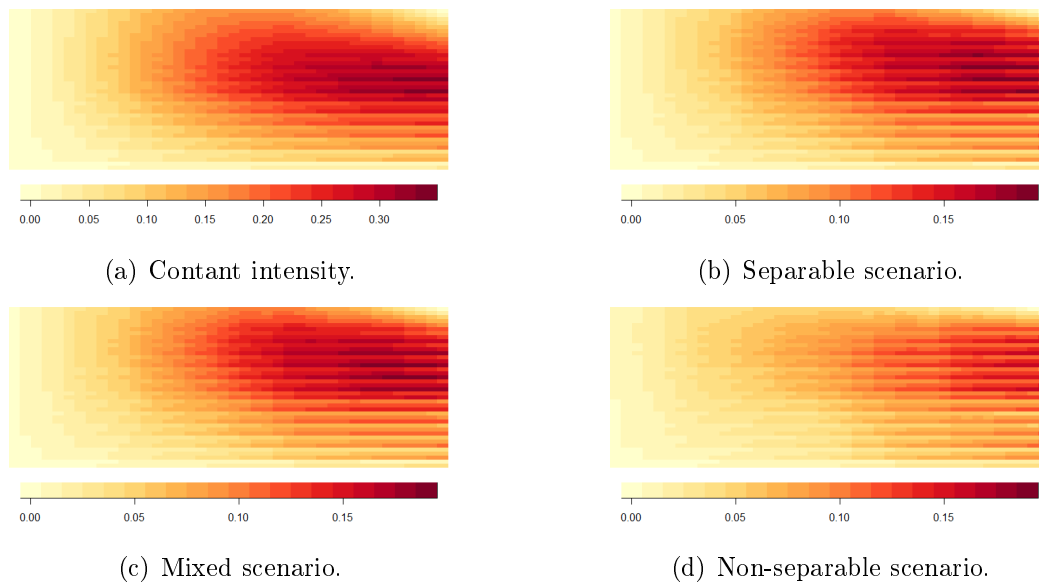


Figure 4.11: Surfaces of the mean differences between the estimated weighted K -function and the theoretical expected value for the separable (upper right), mixed (bottom left) and non-separable scenario (bottom right), and surface of the differences considering a constant intensity (upper left).

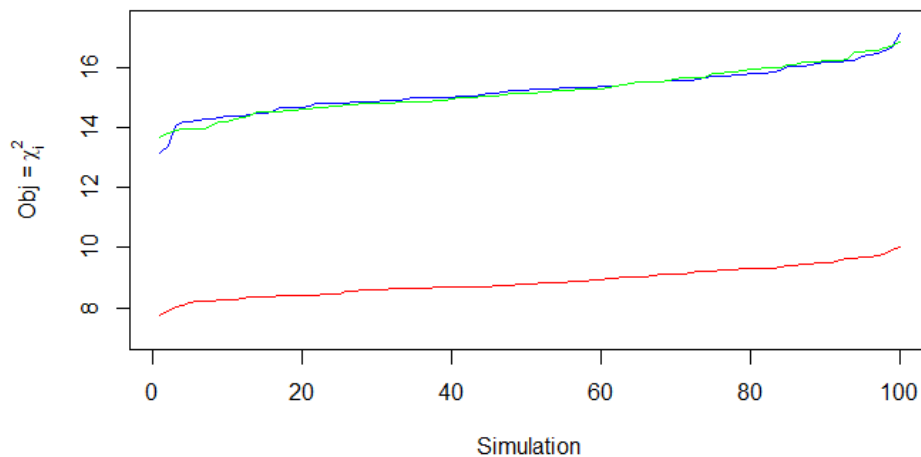


Figure 4.12: Values of the χ^2 statistic of the differences between the estimated weighted K -function and the theoretical expected value of each simulated field for the non-separable (red line), separable (green line) and mixed (blue line) scenarios.

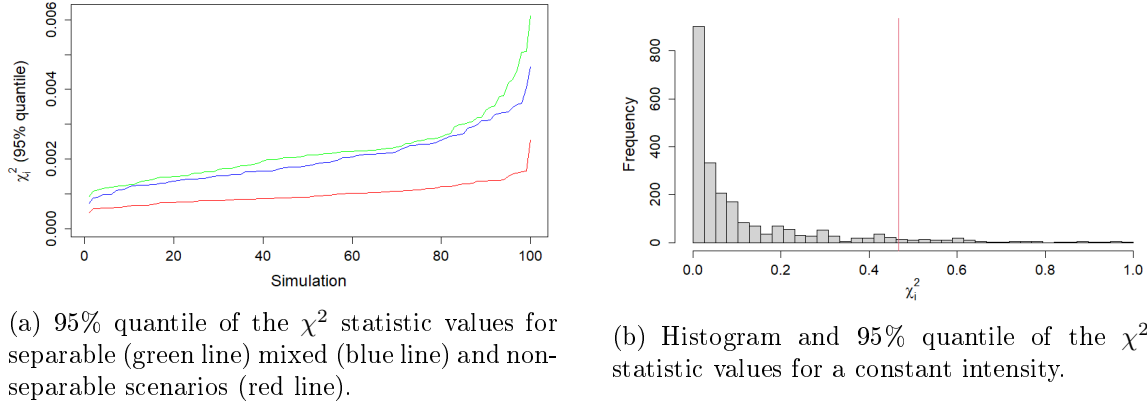


Figure 4.13: 95% quantile of the χ^2 statistic values of the difference between the estimated local weighted K -function and the theoretical expected value for each simulated field considering separable, mixed and non-separable scenarios (Figure 4.13(a)). Figure 4.13(b) displays the histogram of the χ^2 statistic values considering a constant intensity as the driving intensity field of the process. The vertical red line displays the 95% quantile of the χ^2 statistic values.

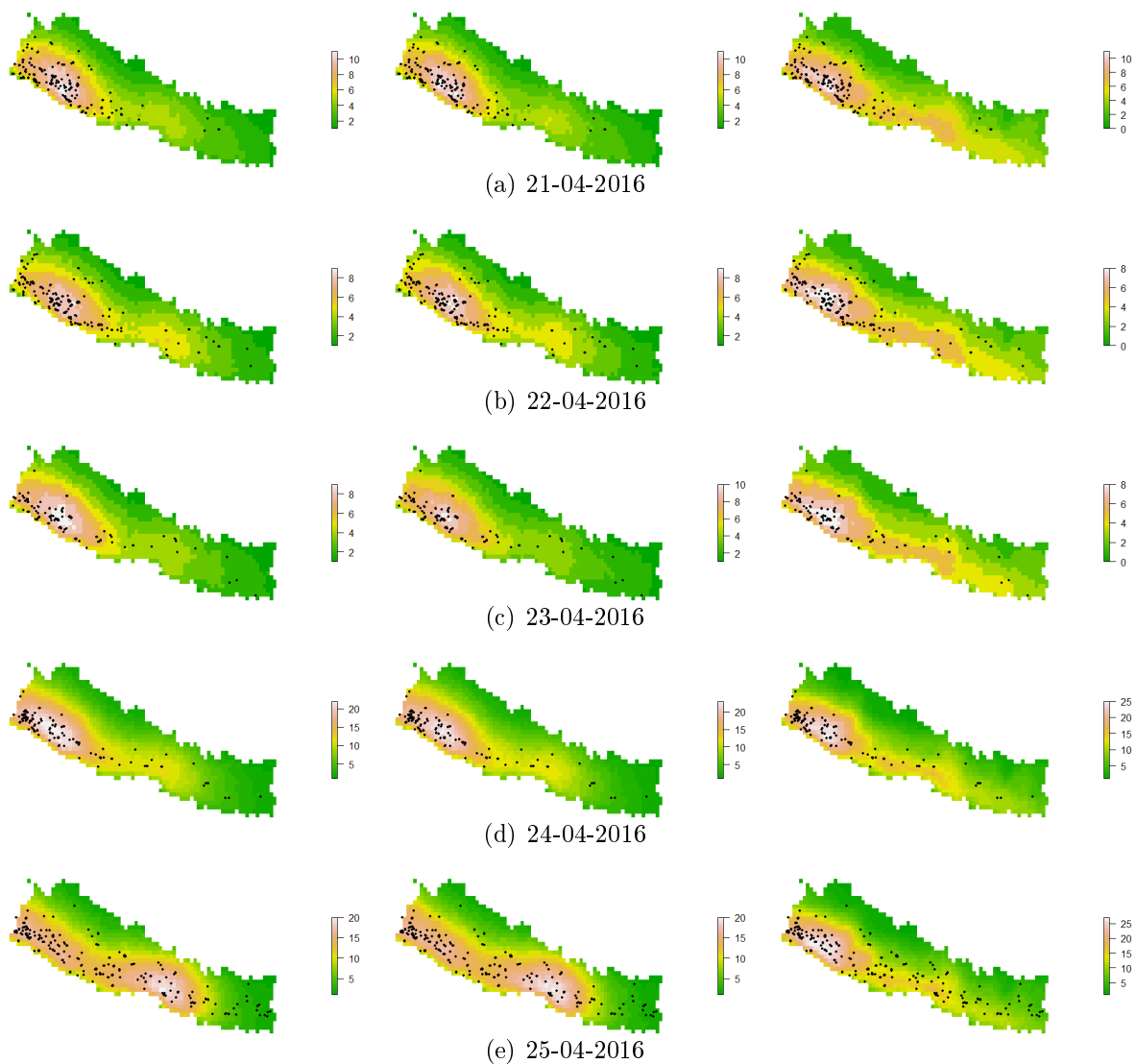
periods across a spatial grid of size 50×50 . Simultaneously, the stochastic component is obtained through the conditional simulation of 500 realizations $G(\mathbf{u}, t)$ from a spatio-temporal Gaussian random field using the DeIaco-Cesare covariance model with parameters $\hat{\theta}$ for each scenario on a grid of size $50 \times 50 \times 16$. Subsequently, we simulate 500 Poisson point patterns for each period under consideration using the resulting predicted intensities.

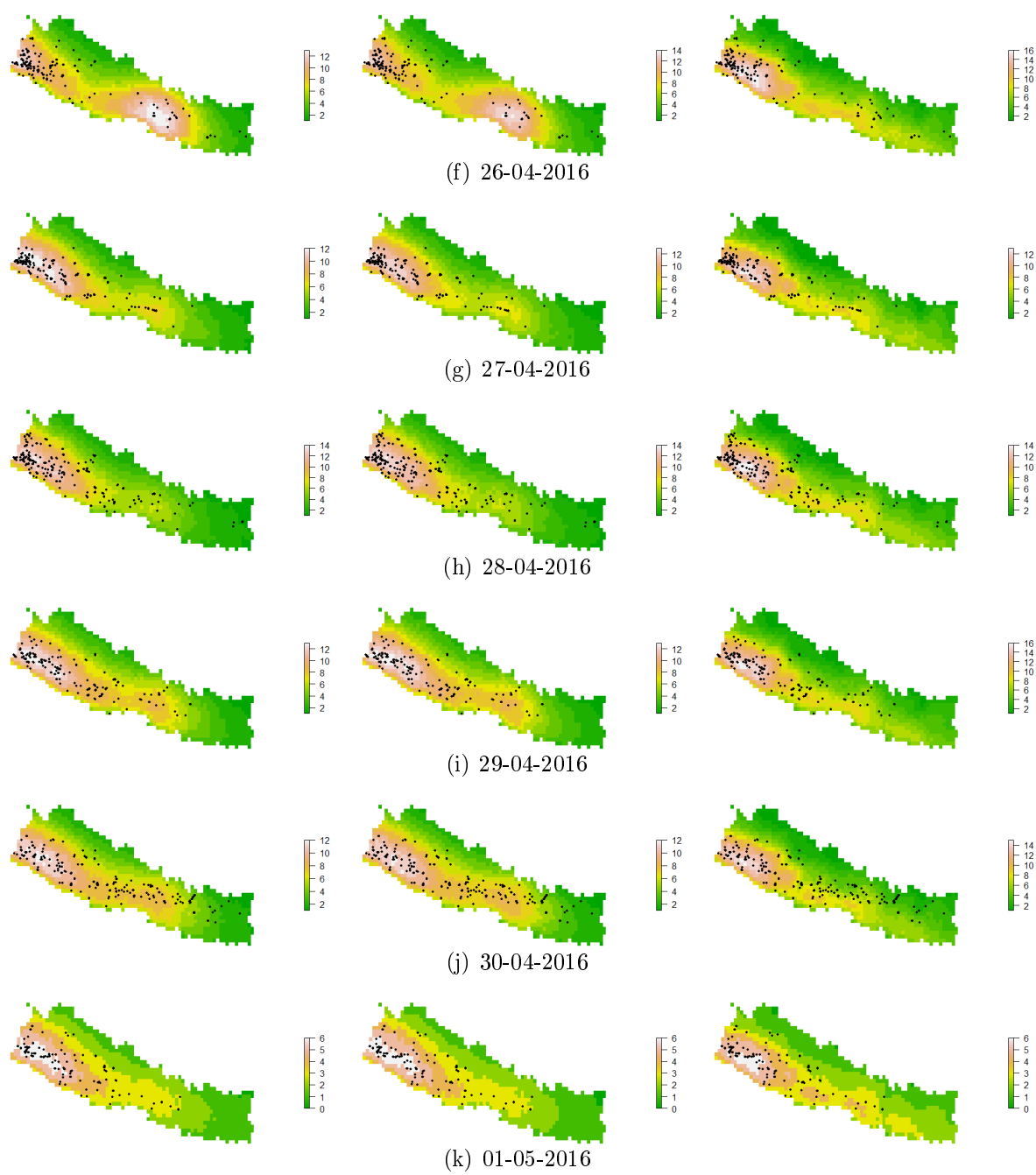
To visualize the risk of forest fires in W for each future period $t + k$ with $k \in [1, 16]$, we compute risk maps (Figures 4.14(a)-4.14(p)) through a quadrat counting procedure within the bounding box of W . This bounding box has dimensions $W = [80.05858, 88.55858] \times [26.34784, 26.84784]$ and we use a sliding window of size 0.5 with a step of 0.1. Each of these subregions W_i is referred to as quadrat.

Among the various families of risk measures, one widely employed in the field of geostatistics and point processes is the quantile-based risk measure family (see Romero *et al.*, 2018⁵⁹). Specifically, Value-at-Risk (VaR) is commonly used due to its straightforward interpretation and ease of computation. The VaR measure evaluates, for a given confidence level α , the threshold beyond which the probability distribution function of losses exceeds α . This threshold is determined by identifying the quantile that leaves to its right the worst $(1 - \alpha)$ % results.

Let $\mathbf{X}_j$ denote the j -th realization, then, for each quadrat S_i and each realization $\mathbf{X}_j = 1, \dots, 500$ we count the number of points $n_{ij} = n(\mathbf{X}_j \cup S_i)$. The value at risk for the region S_i with $\alpha = 0.95$ is defined as the number of forest fires in the region S_i occurring in the 5% of the worst scenarios, i.e.,

$$VaR_\alpha(\mathbf{n}_i) = \inf \{m \in \mathbb{R} | P(\mathbf{n}_i + m < 0) \leq \alpha\} \quad (4.11)$$





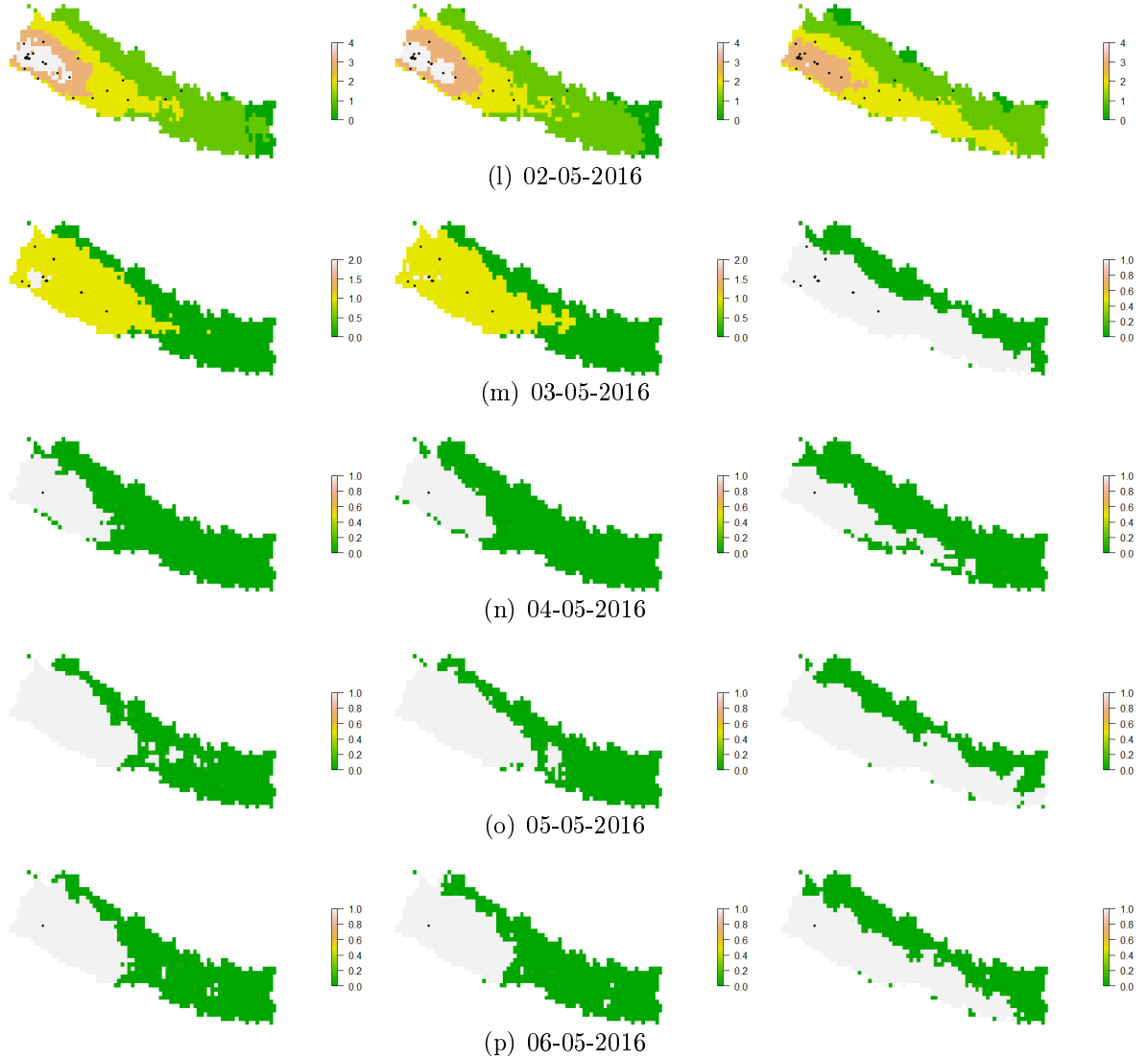


Figure 4.14: Risk maps of the predicted point patterns for the separable (left), mixed (center) and non-separable (right) scenarios from 21-04-2016 to 06-05-2016.

4.6 Synthesis

In this chapter, we have proposed a doubly non-separable model within the class of log-Gaussian Cox processes, where the deterministic and stochastic terms reflect the spatio-temporal dependence of a point pattern, applied to a dataset of forest fires that occurred in Nepal from 2012 to 2016. To assess its predictive accuracy, we compare

this non-separable model with a separable model and a mixed model that combines a separable deterministic intensity with a non-separable covariance structure, conducting this comparison by using local and global weighted second-order statistics.

The exploratory analysis presented in Section 4.3 shows clear seasonal patterns within the temporal structure of the data. In addition, there appears to be a potential interaction between the temporal and spatial dimensions of the process, suggesting that it may be appropriate to adopt a non-separable model for the fitting. In this sense, for the temporal deterministic component of the model we use a generalized linear model with meteorological covariates, while the spatio-temporal deterministic component is estimated using a two-step procedure that combines two intensities obtained through a weighted kernel density estimator and a generalized linear model using the meteorological covariates by district and time.

The results from the deterministic component of the model, which is responsible for the spatio-temporal structure of the underlying intensity, show a better performance in capturing the spatio-temporal dynamics within the non-separable model. This approach allows us to observe an increase in forest fires from April to June, particularly in areas and time intervals characterized by strong winds and mild temperatures. Conversely, a decrease in forest fires is observed in areas with higher precipitation and extreme temperatures.

For the stochastic component, we used an exponential family of covariances for both spatial and temporal terms in the separable case and the DeIaco-Cesare family for non-separable and mixed scenarios. Parameter estimation is performed by minimizing the quadratic distance between the empirical and parametric correlation functions. We use an empirical approach to analyze the results obtained by plotting the envelopes of the non-parametric spatio-temporal second-order summary statistics $\hat{K}(r, h)$, obtaining that the exponential covariance does not accurately reflect the spatio-temporal structure of the data, since it lies outside the envelope.

In Section 4.4 we perform inference over the considered scenarios. By examining the predictive accuracy using the weighted spatio-temporal K -function, we observe that the use of the non-separable scenario improves the fitting, as we obtain lower values of the χ^2 statistics of the differences between the estimated and theoretical weighted K -function for both global and local diagnostics.

Regarding the separable and mixed scenarios, they exhibit similar global performance.

However, it was observed that considering a non-separable covariance structure improves the local fitting.

The analysis carried out motivates the extension of this study through the use of analytical methods to evaluate the goodness-of-fit of the model and its predictive performance, such as entropy-based contrasts to quantify the amount of information transferred from the fitted intensity to the observed pattern. Another idea being developed by the authors is to consider other spatial supports such as linear networks.

Chapter 5

A Comparative Tool for Assessing the Goodness of Fit of Spatial Point Process Models

In this chapter, we propose the use of information and complexity measures to compare the goodness of fit for spatial point process models, focusing on log-Gaussian Cox processes. As discussed in Chapter 3, the structural properties of these processes can be characterized by using information and complexity measures. Specifically, we were able to characterize the information transfer between the two steps involved in the generation of a log-Gaussian Cox process. The aim of this chapter is to extend those findings by analyzing the performance of information and complexity measures as a goodness-of-fit tool and their accuracy in representing the spatial structure of the data.

In that sense, we conduct a simulation study designing nine different scenarios by varying the covariance structure through different configurations of σ^2 and ϕ parameters. For each configuration, we estimate an intensity function by considering various degrees of smoothing in addition to a constant intensity scenario.

We employ information and complexity measures, including Shannon entropy, exponential LMC complexity and Kullback-Leibler divergence, to compare the reference and simulated point patterns. Sensitivity with respect to varying values of the model parameters and the cell size of the quadrat scheme are explored and discussed.

Furthermore, we apply this procedure to a real dataset of forest fires in Nepal. By

comparing the observed spatial patterns to those predicted by different models, we explore the practical usefulness of the information and complexity measures involved in assessing model fitting and guiding the selection of appropriate models for spatial data analysis.

5.1 Introduction

Information and complexity measures are widely employed in many sciences to assess the degree of randomness and heterogeneity of observations. The earliest attempts to introduce these concepts applied on point processes are due to Rudemo (1964)⁹⁴, McFadden (1965)⁹⁵ and Fritz (1973)⁹⁶. Rudemo and McFadden defined Shannon entropy for Poisson processes, while Fritz extended the concept of Rényi entropy to point processes in $\mathbb{R}$. Developing further from these contributions, Papangelou (1978)⁷¹ studied the entropy rate of stationary point processes on the real line.

Subsequently, Daley and Vere-Jones (2004)⁷³ and Baccelli and Oh Woo (2016)⁹⁷, extended and generalized these concepts for homogeneous point processes on a d -dimensional support. More recently, Angulo *et al.*, (2021)⁵³ developed methods for analyzing point patterns using information and complexity measures within a multifractal framework. In the context of log-Gaussian Cox processes, Chapter 3 evaluates the global and local structural information transfer from the intensity field to the subsequently generated point patterns through entropy, divergence, and complexity measures (see Medialdea *et al.*, 2021).

Clark (2020)⁹⁸ introduced a collection of information statistics for point processes based on entropy and divergences measures, including the entropy-population covariance, which quantifies the correlation between local and global entropy. These measures have been widely used in physics, where Kittel and Kroemer (1980)⁹⁹ introduced the concept of configurational entropy, describing the number of configurations of the particles of a system using Boltzmann equation, first introduced by Planck (1914)¹⁰⁰. Cushman (2021)¹⁰¹ extended the concept of configurational entropy to other domains such as surfaces, point patterns and landscape mosaics.

Recently, these measures have gained popularity in environmental studies (Altieri and Cocchi, 2024¹⁰²). Shannon entropy, in particular, has proven useful as a diversity index for assessing the degree of heterogeneity among species and landscape patterns (see, for

example, Magurran, 2004¹⁰³ and Frosini, 2006¹⁰⁴). This application has been explored across various types of spatial data such as areal, geostatistical and point processes. Altieri *et al.* (2021)¹⁰⁵ further contributed to this field by developing the **SpatEntropy** R package, which contains functions to describe the heterogeneity of multicategorical lattice and point data through the computation of various entropy and spatial entropy measures.

Despite the broad applicability and relevance in characterizing spatial pattern configurations and system dynamics, information and complexity measures have not yet been fully explored as diagnostic tools for assessing the goodness of fit in point process models.

5.2 Simulation study

We assume a LGCP with an intensity $\Lambda(\mathbf{x}) = \exp\{G(\mathbf{x})\}$, where $G(\mathbf{x})$ is a Gaussian random field with $E[G(\mathbf{x})] = \mu$ and $C(\mathbf{u}, \mathbf{v}) = C(r) = \sigma^2 \exp(-r/\phi)$, displayed in a fine grid with $\mathbf{x} \in W = [0, 1] \times [0, 1]$, the unit square window. We design nine scenarios (see Table 5.1) by varying the covariance structure through different configurations of σ^2 and ϕ parameters.

	σ^2	ϕ
Configuration 1	1	0.05
Configuration 2	1	0.1
Configuration 3	1	0.25
Configuration 4	2.5	0.05
Configuration 5	2.5	0.1
Configuration 6	2.5	0.25
Configuration 7	5	0.05
Configuration 8	5	0.1
Configuration 9	5	0.25

Table 5.1: Parameter configurations for the scenarios considered in the analysis.

For each configuration C_i , $i = 1, \dots, 9$, an intensity function Λ_i is generated, each of which subsequently generates $X_{i,1}, \dots, X_{i,1000}$ point patterns with a mean value of one thousand points. Figure 5.1 shows an example of the simulated patterns for each configuration.

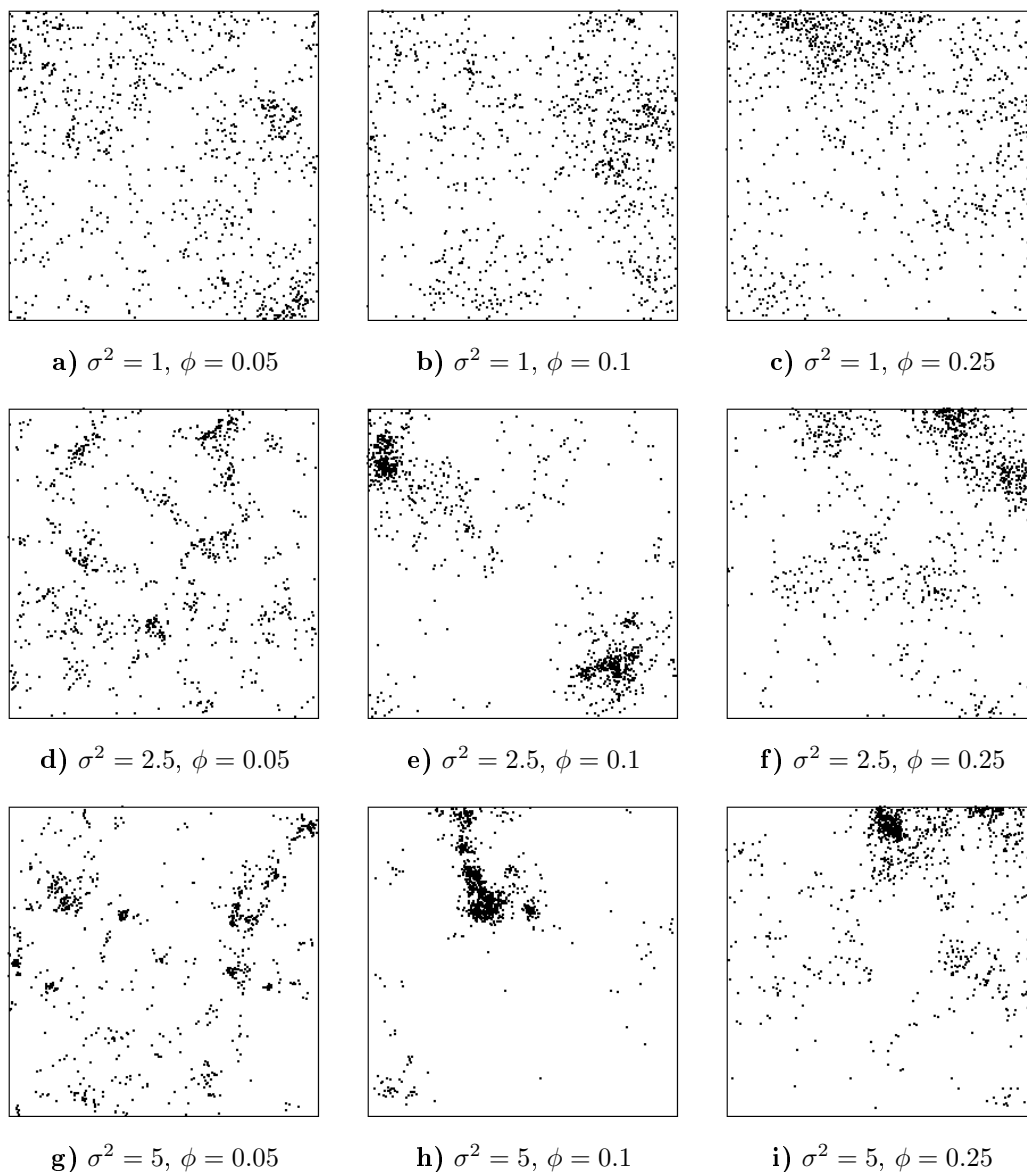


Figure 5.1: Examples of realizations from a log-Gaussian Cox process (LGCP) for different configurations of the covariance parameters σ^2 and ϕ .

The intensity function for each realization is estimated through a kernel density estimation procedure. The bandwidth τ of the kernel is selected by using Diggle's mean-square error criterion given in Equation 2.25, configuring different degrees of accuracy in the estimated intensity by setting the bandwidth parameter as τ , 2τ and 3τ . Additionally, we consider a constant intensity equal to $n/|W|$. Conditional on these

intensity estimates we perform a realization of a Poisson process.

Simulation and estimation procedure

Step 1:

For C_i $i=1, \dots, 9$: generate Λ_i
 for $j=1, \dots, 1000$: generate $\Lambda_i \rightsquigarrow X_{i,j}$

Step 2:

Estimate intensities: $\hat{\lambda}_{i,j}|X_{i,j}$ through KDE considering τ (case 1), 2τ (case 2), 3τ (case 3) and $\hat{\lambda}_{i,j} = \frac{n(X_{i,j})}{|W|}$ (case 4)
 Perform a realization: $\hat{\lambda}_{i,j} \rightsquigarrow \hat{X}_{i,j}$

We compute information and complexity measures, specifically focusing on Shannon entropy (Equation 2.55) and exponential LMC complexity (Equation 2.63), for the probability distributions of both the reference and simulated point patterns, as well as for each configuration of the covariance parameters considered. Additionally, we assess the Kullback-Leibler divergence (see Equation 2.65) between each pair $(X_{i,j}, X_{i,k})$ of the reference patterns and each pair $(X_{i,j}, \hat{X}_{i,j})$. We also compute indicators based on Shannon entropy and exponential LMC complexity expressed respectively as

$$I_H(\mathbf{p}||\mathbf{q}) = \frac{H(\mathbf{p})}{H(\mathbf{q})}, \quad (5.1)$$

$$I_{C_{LMC}^{\text{exp}}}(\mathbf{p}||\mathbf{q}) = \frac{C_{LMC}^{\text{exp}}(\mathbf{p})}{C_{LMC}^{\text{exp}}(\mathbf{q})}, \quad (5.2)$$

with $\mathbf{p}$ being the probability distribution of the simulated pattern (or observed pattern) and $\mathbf{q}$ the probability distribution of the pattern simulated from their respectively estimated intensities. Both indices range from $[0, \infty]$.

For I_H , values within $[0, 1]$ indicate an structural loss in the estimated pattern compared to the observed pattern. Conversely, values within $[1, \infty]$ suggest an structural gain, *i.e.* the estimated pattern is less homogeneous than the observed pattern. When $I_H = 1$, the estimated pattern perfectly reflects the structural degree of heterogeneity the observed pattern.

For I_{LMC}^{exp} , values between $[0, 1]$ suggest that the estimated pattern has greater structural complexity than the reference pattern. Values in $[1, \infty]$ indicates an structural complexity loss relative to the observed pattern, while $I_{LMC}^{exp} = 1$ suggests that the estimated pattern and reference pattern have the same degree of structural complexity.

In Figures 5.2, 5.3, 5.4, we show the mean values and 95% confidence intervals for Shannon entropy, Kullback-Leibler divergence and exponential LMC complexity for both the reference and simulated patterns.

The Shannon entropy of the simulated patterns matches the entropy of the reference patterns almost perfectly for the most accurate estimate of λ , $\hat{\lambda}_\tau$. As we increase the smoothing bandwidth, $\hat{\lambda}$ becomes smoother, leading to poorer estimations with entropy values above the confidence intervals. When we consider a constant intensity, we obtain more homogeneous realizations, yielding Shannon entropy values closer to $H_{\max} = \ln(N)$.

When we compare the Kullback-Leibler divergence between the reference patterns within each scenario, we obtain larger values than those obtained when comparing each of them with its corresponding simulated pattern, considering the optimal smoothing bandwidth τ . This divergence value increases as we increase the smoothing bandwidth. The exponential LMC complexity shows a loss of structure in the estimated pattern compared to the reference pattern. This loss becomes more pronounced as the accuracy of the estimation decreases, while the use of a constant intensity leads to complexity values near zero due to the trivial structure of the resulting pattern.

The selected quadrat scheme plays a crucial role in the computation of information and complexity measures, according to the p-values obtained from simulation by comparing the discrepancy between the observed and estimated patterns with respect to the internal variability of the observed patterns (see Tables 5.2, 5.3 and 5.4). For patterns exhibiting lower variance (*i.e.*, less clustered structure), Shannon entropy performs well regardless of the chosen quadrat size. However, as variance increases, it is required the use of a smaller quadrat size, with the exception of cases involving high values of variance combined with low scale parameter values (*i.e.*, highly concentrated patterns). In such cases, Shannon entropy shows a poor performance, and it is recommended the use of exponential LMC complexity and Kullback-Leibler divergence for global and local assessment, respectively.

Kullback-Leibler divergence exhibits good performance with medium quadrat sizes ($q = 20$) for low structured patterns. As patterns become more clustered, increasing the

number of cells in the quadrat scheme becomes beneficial for capturing local variability, especially in scenarios with low scale parameter values.

Regarding the exponential LMC complexity measure, optimal performance for sparsely clustered patterns is achieved by using larger quadrat sizes. Conversely, when the spatial dependence range is large, this measure has a good performance across all quadrat schemes. In highly clustered patterns, the choice of the quadrat size depends on the scale parameter value; as the pattern becomes more diffuse, an increase in quadrat size is recommended.

The mean values of the indices (5.1) and (5.2) are presented in Table 5.5 and Table 5.6, respectively. According to the entropy indices, λ_τ yields the closest values to 1 in each scenario and across all quadrat sizes, confirming it as the most accurate estimate. As we increase the bandwidth and the estimates are consequently smoothed, the entropy indices decrease. This decrease is more pronounced for patterns generated with larger variance parameters, *i.e.*, highly clustered patterns, which results in poorer representations and the lowest values of entropy indices when we consider a constant intensity. In terms of quadrat size, the entropy index becomes more useful for discriminating between models for lower to medium quadrat size ($q = 10$ and $q = 20$), as they better capture the global structure of the patterns.

We obtain similar results regarding complexity indices. However, they generally perform better than entropy indices, as their magnitudes more clearly highlights the differences in goodness of fit across the models considered. As with entropy indices, complexity indices are particularly useful for highly clustered patterns (large variance parameter) and strong spatial dependence (large scale parameter). Regarding quadrat size, for small quadrat size ($q = 50$), the indices show less variation across estimates, suggesting a limited ability to differentiate between estimates, as they are poorly capture the local variability.

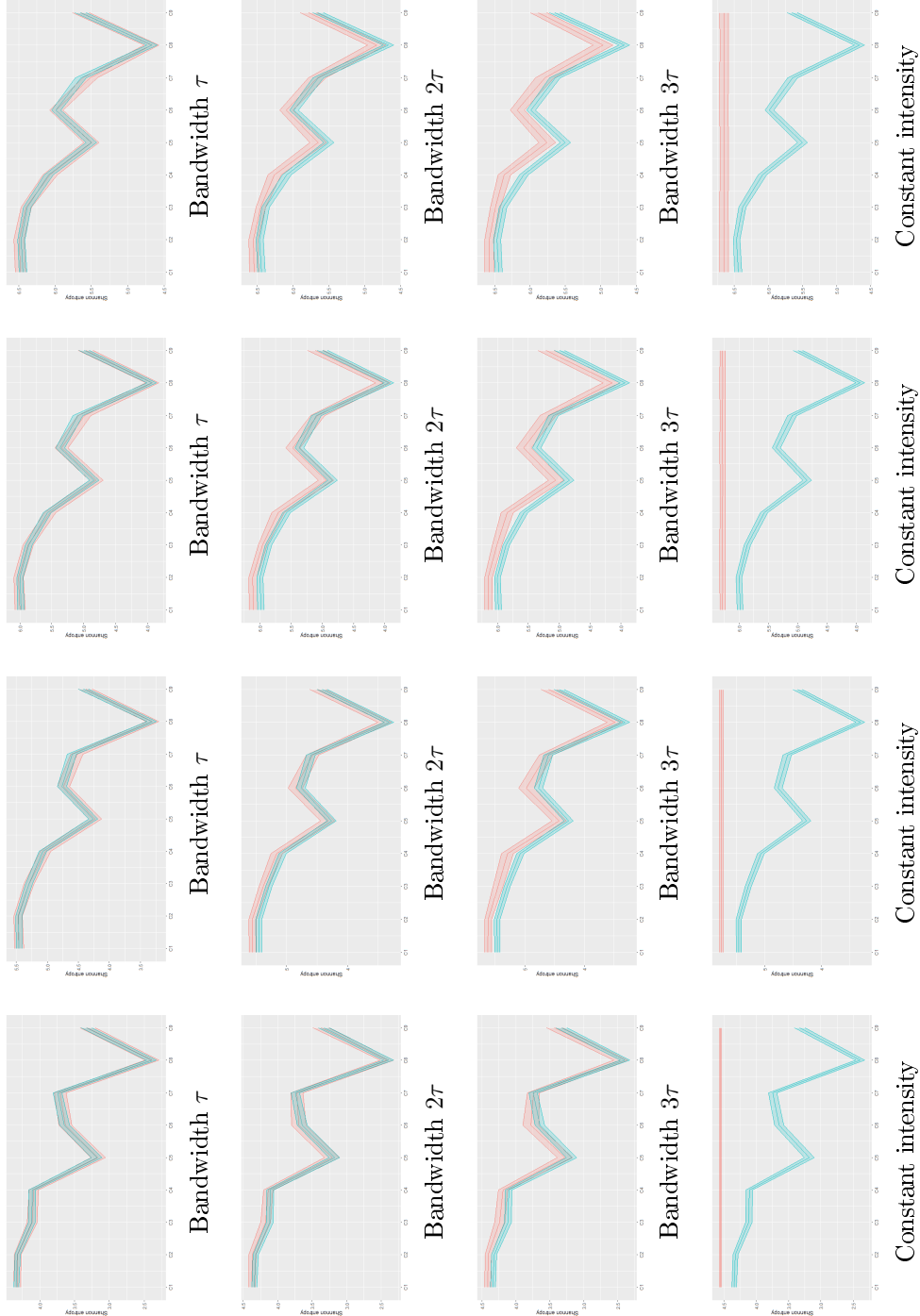


Figure 5.2: Shannon entropy and 95% confidence interval for observed (blue) and estimated patterns (red) for $[10 \times 10]$ (left), $[20 \times 20]$ (center left), $[30 \times 30]$ (center right) and $[50 \times 50]$ (right) quadrat scheme configurations.

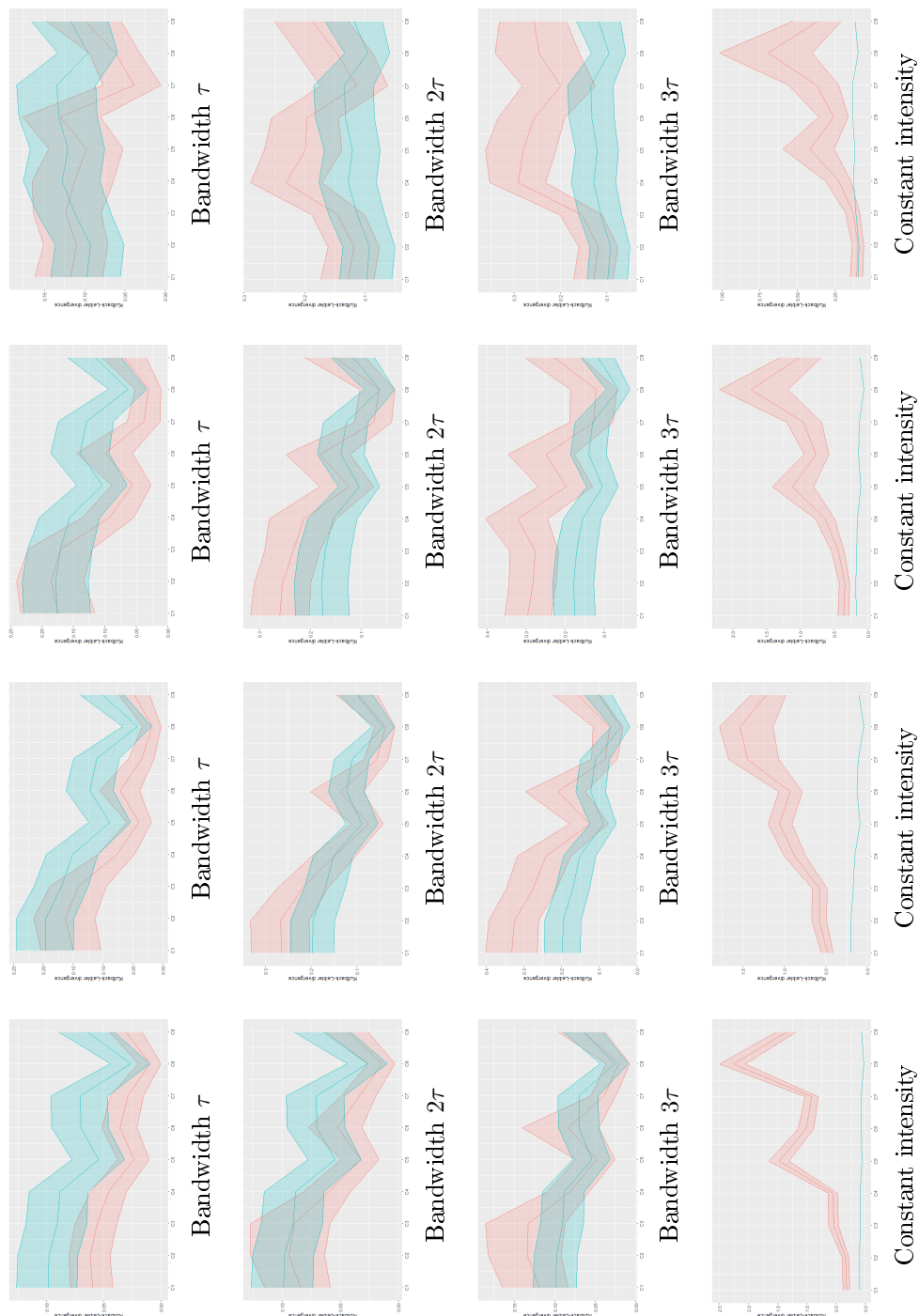


Figure 5.3: Kullback-Leibler divergence and 95% confidence interval for each pair of observed (blue) and estimated patterns (red), for $[10 \times 10]$ (left), $[20 \times 20]$ (center left), $[30 \times 30]$ (center right) and $[50 \times 50]$ (right) quadrat scheme configurations.

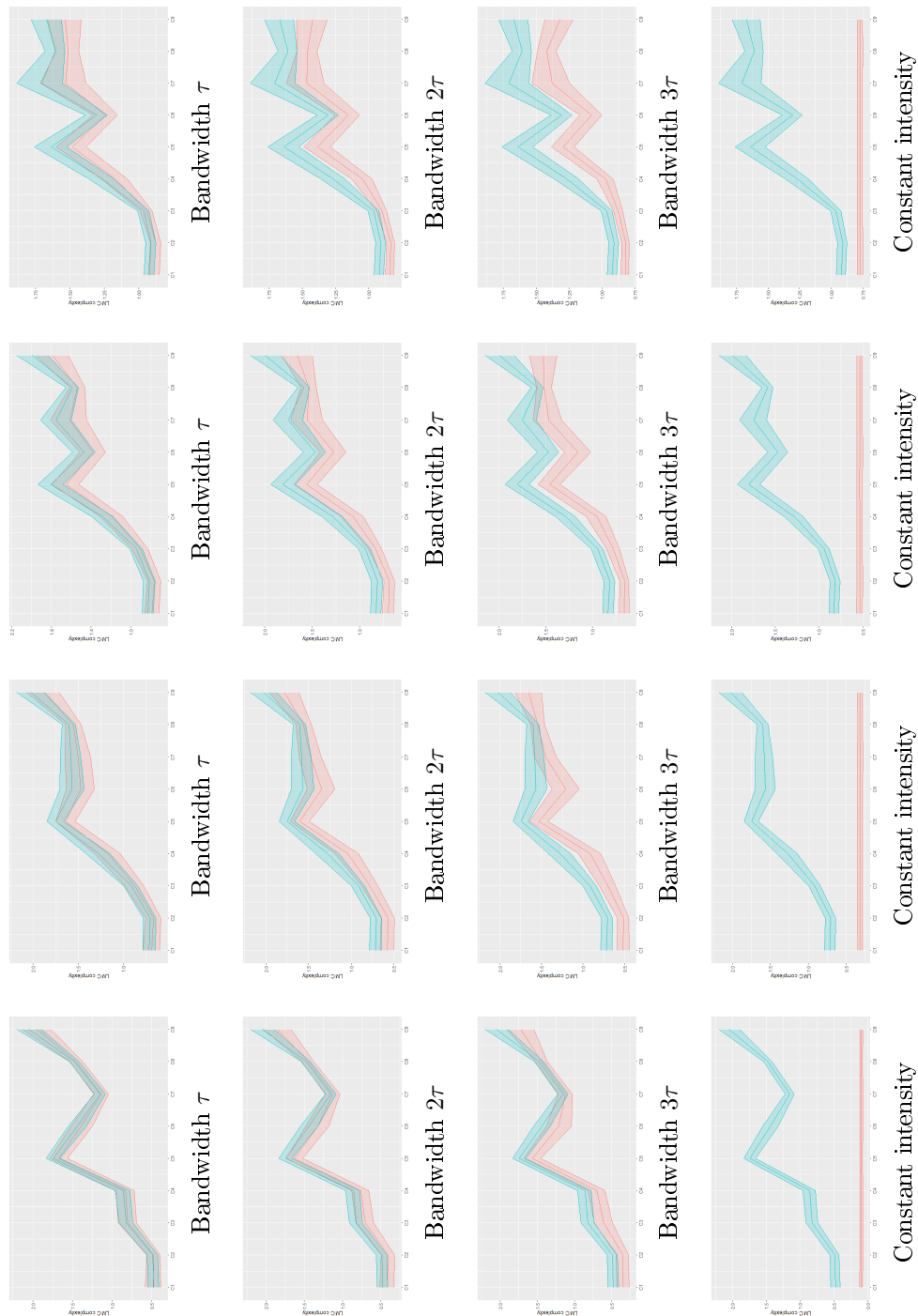


Figure 5.4: Exponential LMC complexity and 95% confidence interval for observed (blue) and estimated patterns (red) for $[10 \times 10]$ (left), $[20 \times 20]$ (center left), $[30 \times 30]$ (center right) and $[50 \times 50]$ (right) quadrat scheme configurations.

σ^2	ϕ	$q = 10$				$q = 20$				$q = 30$				$q = 50$			
		λ_τ	$\lambda_{2\tau}$	$\lambda_{3\tau}$	λ_c	λ_τ	$\lambda_{2\tau}$	$\lambda_{3\tau}$	λ_c	λ_τ	$\lambda_{2\tau}$	$\lambda_{3\tau}$	λ_c	λ_τ	$\lambda_{2\tau}$	$\lambda_{3\tau}$	λ_c
1	0.05	0.021	0.080	0.571	1	0.023	0.627	0.974	1	0.020	0.740	0.980	1	0.034	0.498	0.806	0.997
	0.1	$< 10^{-4}$	0.161	0.661	1	0.003	0.563	0.940	1	0.015	0.640	0.930	1	0.020	0.354	0.633	0.987
	0.25	0.003	0.150	0.691	1	0.006	0.312	0.832	1	0.008	0.342	0.826	1	0.008	0.170	0.531	1
2.5	0.05	0.004	0.016	0.332	1	0.023	0.269	0.995	1	0.035	0.593	0.995	1	0.013	0.844	1	1
	0.1	0.001	0.009	0.184	1	0.006	0.056	0.546	1	0.011	0.205	0.834	1	0.007	0.508	0.970	1
	0.25	0.001	0.032	0.428	1	0.006	0.227	0.791	1	0.007	0.277	0.852	1	0.006	0.349	0.889	1
5	0.05	0.003	0.002	0.002	1	0.040	0.004	0.023	1	0.138	0.007	0.113	1	0.301	0.014	0.478	1
	0.1	$< 10^{-4}$	0.003	0.048	1	$< 10^{-4}$	0.071	0.586	1	$< 10^{-4}$	0.174	0.794	1	0.003	0.598	0.985	1
	0.25	0.001	0.011	0.149	1	0.003	0.071	0.552	1	0.003	0.222	0.810	1	0.002	0.364	0.903	1

Table 5.2: Significance level of Shannon entropy measure for each quadrat scheme q , intensity function $\hat{\lambda}_\tau, \hat{\lambda}_{2\tau}, \hat{\lambda}_{3\tau}, \hat{\lambda}_c$, and configuration of covariance parameters considered.

σ^2	ϕ	$q = 10$				$q = 20$				$q = 30$				$q = 50$			
		λ_τ	$\lambda_{2\tau}$	$\lambda_{3\tau}$	λ_c	λ_τ	$\lambda_{2\tau}$	$\lambda_{3\tau}$	λ_c	λ_τ	$\lambda_{2\tau}$	$\lambda_{3\tau}$	λ_c	λ_τ	$\lambda_{2\tau}$	$\lambda_{3\tau}$	λ_c
1	0.05	$< 10^{-4}$	0.018	0.352	1	0.002	0.668	0.990	1	0.061	0.788	0.972	1	0.168	0.301	0.292	0.078
	0.1	$< 10^{-4}$	0.064	0.610	1	0.007	0.711	0.982	1	0.095	0.742	0.928	1	0.140	0.299	0.232	0.058
	0.25	$< 10^{-4}$	0.099	0.685	1	0.005	0.545	0.936	1	0.047	0.626	0.924	1	0.077	0.292	0.406	0.140
2.5	0.05	$< 10^{-4}$	0.001	0.022	1	$< 10^{-4}$	0.059	0.866	1	$< 10^{-4}$	0.558	0.999	1	0.024	0.933	1	0.825
	0.1	$< 10^{-4}$	$< 10^{-4}$	0.036	1	$< 10^{-4}$	0.014	0.459	1	$< 10^{-4}$	0.235	0.890	1	0.014	0.788	0.994	0.998
	0.25	$< 10^{-4}$	0.009	0.270	1	$< 10^{-4}$	0.225	0.824	1	$< 10^{-4}$	0.420	0.917	1	0.032	0.654	0.968	0.910
5	0.05	$< 10^{-4}$	$< 10^{-4}$	$< 10^{-4}$	1	$< 10^{-4}$	$< 10^{-4}$	0.003	1	$< 10^{-4}$	$< 10^{-4}$	0.083	1	$< 10^{-4}$	0.024	0.609	0.988
	0.1	$< 10^{-4}$	$< 10^{-4}$	0.013	1	$< 10^{-4}$	0.027	0.523	1	$< 10^{-4}$	0.096	0.732	1	$< 10^{-4}$	0.608	0.990	1
	0.25	$< 10^{-4}$	$< 10^{-4}$	0.103	1	$< 10^{-4}$	0.074	0.674	1	$< 10^{-4}$	0.411	0.945	1	0.015	0.708	0.986	0.987

Table 5.3: Significance level of Kullback-Leibler divergence measure for each quadrat scheme q , intensity function $\hat{\lambda}_\tau, \hat{\lambda}_{2\tau}, \hat{\lambda}_{3\tau}, \hat{\lambda}_c$, and configuration of covariance parameters considered.

σ^2	ϕ	$q = 10$				$q = 20$				$q = 30$				$q = 50$			
		λ_τ	$\lambda_{2\tau}$	$\lambda_{3\tau}$	λ_c	λ_τ	$\lambda_{2\tau}$	$\lambda_{3\tau}$	λ_c	λ_τ	$\lambda_{2\tau}$	$\lambda_{3\tau}$	λ_c	λ_τ	$\lambda_{2\tau}$	$\lambda_{3\tau}$	λ_c
1	0.05	0.005	0.051	0.411	1	0.018	0.513	0.919	1	0.049	0.735	0.960	1	0.078	0.618	0.886	1
	0.1	$< 10^{-4}$	0.101	0.181	1	0.013	0.538	0.910	1	0.023	0.556	0.869	1	0.063	0.453	0.706	0.995
	0.25	0.001	0.188	0.730	1	0.001	0.329	0.847	1	0.007	0.326	0.836	1	0.020	0.210	0.581	1
2.5	0.05	0.001	0.063	0.342	1	0.013	0.330	0.855	1	0.034	0.503	0.960	1	0.051	0.759	0.985	1
	0.1	0.035	0.157	0.532	1	0.096	0.278	0.672	1	0.089	0.429	0.855	1	0.112	0.578	0.942	1
	0.25	0.002	0.089	0.461	1	0.011	0.314	0.809	1	0.016	0.342	0.848	1	0.024	0.448	0.904	1
5	0.05	0.001	0.003	0.005	1	0.009	0.023	0.067	1	0.016	0.087	0.379	1	0.029	0.140	0.460	1
	0.1	0.002	0.003	0.008	1	0.007	0.016	0.034	1	0.029	0.052	0.134	1	0.133	0.531	0.823	1
	0.25	0.016	0.139	0.549	1	0.044	0.295	0.739	1	0.123	0.616	0.942	1	0.097	0.655	0.964	1

Table 5.4: Significance level of exponential LMC complexity measure for each quadrat scheme q , intensity function $\hat{\lambda}_\tau, \hat{\lambda}_{2\tau}, \hat{\lambda}_{3\tau}, \hat{\lambda}_c$, and configuration of covariance parameters considered.

σ^2	ϕ	$q = 10$				$q = 20$				$q = 30$				$q = 50$			
		λ_τ	$\lambda_{2\tau}$	$\lambda_{3\tau}$	λ_c	λ_τ	$\lambda_{2\tau}$	$\lambda_{3\tau}$	λ_c	λ_τ	$\lambda_{2\tau}$	$\lambda_{3\tau}$	λ_c	λ_τ	$\lambda_{2\tau}$	$\lambda_{3\tau}$	λ_c
1	0.05	1.0034	0.9926	0.9831	0.9530	1.0005	0.9816	0.9719	0.9463	0.9974	0.9801	0.9724	0.9522	0.9949	0.9842	0.9797	0.9685
	0.1	1.0006	0.9890	0.9804	0.9487	0.9984	0.9817	0.9735	0.9473	0.9965	0.9819	0.97558	0.9553	0.9955	0.9868	0.9831	0.9722
	0.25	1.0013	0.9851	0.9715	0.9038	0.9999	0.9836	0.9723	0.9179	0.9983	0.9847	0.9756	0.9330	0.9976	0.9888	0.9865	0.9594
2.5	0.05	1.0048	0.9926	0.9801	0.9017	1.0080	0.9820	0.9599	0.8791	1.0082	0.9762	0.9539	0.8867	1.0030	0.9728	0.9568	0.9154
	0.1	1.0045	0.9898	0.9702	0.6978	1.0070	0.9864	0.9644	0.7387	1.0063	0.9798	0.9571	0.7699	1.0008	0.9740	0.9553	0.8272
	0.25	1.0035	0.9868	0.9653	0.8006	1.0051	0.9809	0.9595	0.8263	1.0042	0.9813	0.9631	0.8542	1.0001	0.9825	0.9691	0.9001
5	0.05	1.0101	1.0063	1.0006	0.8221	1.0176	1.0059	0.9920	0.7976	1.0215	1.0037	0.9847	0.8106	1.0230	0.9976	0.9752	0.8485
	0.1	1.0025	0.9895	0.9749	0.5259	1.0015	0.9783	0.9518	0.5744	1.0025	0.9765	0.9494	0.6261	0.9989	0.9662	0.9392	0.7013
	0.25	1.0044	0.9873	0.9692	0.7302	1.0082	0.9839	0.9616	0.7653	1.0052	0.9775	0.9574	0.7944	1.0012	0.9771	0.9612	0.8494

Table 5.5: Mean values of Shannon entropy indices for each quadrat scheme q , intensity function $\hat{\lambda}_\tau, \hat{\lambda}_{2\tau}, \hat{\lambda}_{3\tau}, \hat{\lambda}_c$, and configuration of covariance parameters considered.

σ^2	ϕ	$q = 10$				$q = 20$				$q = 30$				$q = 50$			
		λ_τ	$\lambda_{2\tau}$	$\lambda_{3\tau}$	λ_c	λ_τ	$\lambda_{2\tau}$	$\lambda_{3\tau}$	λ_c	λ_τ	$\lambda_{2\tau}$	$\lambda_{3\tau}$	λ_c	λ_τ	$\lambda_{2\tau}$	$\lambda_{3\tau}$	λ_c
1	0.05	0.9993	1.1724	1.4114	5.1410	1.0479	1.2462	1.4000	2.2486	1.0588	1.1888	1.2669	1.5574	1.0404	1.0949	1.1209	1.1934
	0.1	1.0231	1.2053	1.4036	5.2879	1.0660	1.2533	1.3823	2.2151	1.0658	1.1830	1.2463	1.5243	1.0407	1.0869	1.1086	1.1761
	0.25	1.0304	1.1744	1.3705	8.9096	1.0476	1.1756	1.3132	2.8748	1.0420	1.1333	1.2158	1.7669	1.0234	1.0677	1.1002	1.2448
2.5	0.05	1.0367	1.1268	1.2401	9.3148	1.0930	1.2297	1.3995	4.0056	1.0807	1.2037	1.3499	2.4039	1.0836	1.2105	1.3033	1.6357
	0.1	1.0407	1.0699	1.1159	18.8739	1.0661	1.0925	1.1414	5.4850	1.0953	1.1580	1.2417	3.3786	1.1051	1.1823	1.2654	2.1153
	0.25	1.0379	1.0902	1.1939	15.6294	1.0679	1.1681	1.3064	4.9297	1.0657	1.1579	1.2778	2.7594	1.0548	1.1326	1.2120	1.6890
5	0.05	1.0152	1.0214	1.0295	12.5820	1.0575	1.0746	1.0931	4.9855	1.0871	1.1222	1.1959	3.2833	1.1152	1.1670	1.2396	2.2059
	0.1	1.0158	1.0189	1.0231	15.9560	1.0312	1.0293	1.0369	5.0392	1.0340	1.0352	1.0484	2.9753	1.0636	1.1026	1.1344	2.0778
	0.25	1.0551	1.1137	1.1931	22.3911	1.0785	1.1416	1.2301	6.3565	1.1178	1.2060	1.3102	3.7386	1.0884	1.1756	1.2681	2.1586

Table 5.6: Mean values of exponential LMC complexity indices for each quadrat scheme q , intensity function $\hat{\lambda}_\tau, \hat{\lambda}_{2\tau}, \hat{\lambda}_{3\tau}, \hat{\lambda}_c$, and configuration of covariance parameters considered.

5.3 Entropy-based Monte Carlo goodness-of-fit test

In this section, we propose a goodness-of-fit test statistic based on Shannon entropy. The testing framework follows the Monte Carlo procedure described in Diggle (1983)¹²⁰, as outlined in Section 2.2.5.

As a summary description, we use the normalized Shannon entropy (see Equation (2.56)), evaluated across multiple spatial resolutions by partitioning the study region A into grids of varying quadrat sizes. For each partition k , we apply a quadrat counting method to compute the empirical distribution of event count, $\mathbf{p}_k$, with $k = 1, \dots, n$.

The normalized entropy $H_{\text{norm}}(\mathbf{p}_k)$, for simplicity hereinafter also denoted as $H(\mathbf{p}_k)$, is then calculated for each $\mathbf{p}_k$. The choice of the normalized Shannon entropy removes the influence of quadrat size. While Shannon entropy is bounded between $[0, \ln(N)]$, where N is the number of quadrats in the region A , the normalized entropy rescales the values to the fixed interval $[0, 1]$. This makes it possible to compare entropy values across different grid resolutions.

To construct the statistic, we simulate $m-1$ realizations under the null hypothesis and compute, for each simulation i , the discrepancy between the estimated and theoretical normalized entropy as

$$u_i = \sup_k |\hat{H}_i(\mathbf{p}_k) - \bar{H}_i(\mathbf{p}_k)|, \quad (5.3)$$

with

$$\bar{H}_i(\mathbf{p}_k) = (m-2)^{-1} \sum_{\substack{j=2 \\ j \neq i}}^m \hat{H}_j(\mathbf{p}_k) \quad (5.4)$$

While the test is intended to assess how well a fitted model captures the observed spatial structure, we evaluate its power, following the approach in Diggle (1979)¹²¹, by testing the ability of the entropy-based summary to distinguish between clustered and regular patterns from complete spatial random patterns.

Three alternative spatial point processes are considered. The first is a log-Gaussian Cox process, defined in Equation (2.40), which produces clustered patterns. The second process is the simple sequential inhibition process, as described in Diggle, Besag and Gleaves (1976)¹²⁷. In this process, n events are sequentially placed within the region A , such that each new event lies at least a fixed distance δ from all previously placed events. A key parameter of this process is the packing density,

$$\tau = n\pi\delta^2/4|A| \quad (5.5)$$

representing the proportion of the area A covered by n non-overlapping discs of diameter δ , controlling the degree of regularity in the resulting point pattern.

The third process considered is a homogeneous Poisson process, to test the significance level under the null hypothesis of complete spatial randomness.

For these alternative processes, and a range of parameters settings for both inhibition and inhomogeneous processes, we generate 1000 realizations, with a mean of 1000 event in the unit square, to assess the performance of the test under different types of spatial structures. We consider $k = 29$ partitions of the study region from 2×2 to 30×30 quadrats, selecting the later to maintain an average of approximately three events per cell.

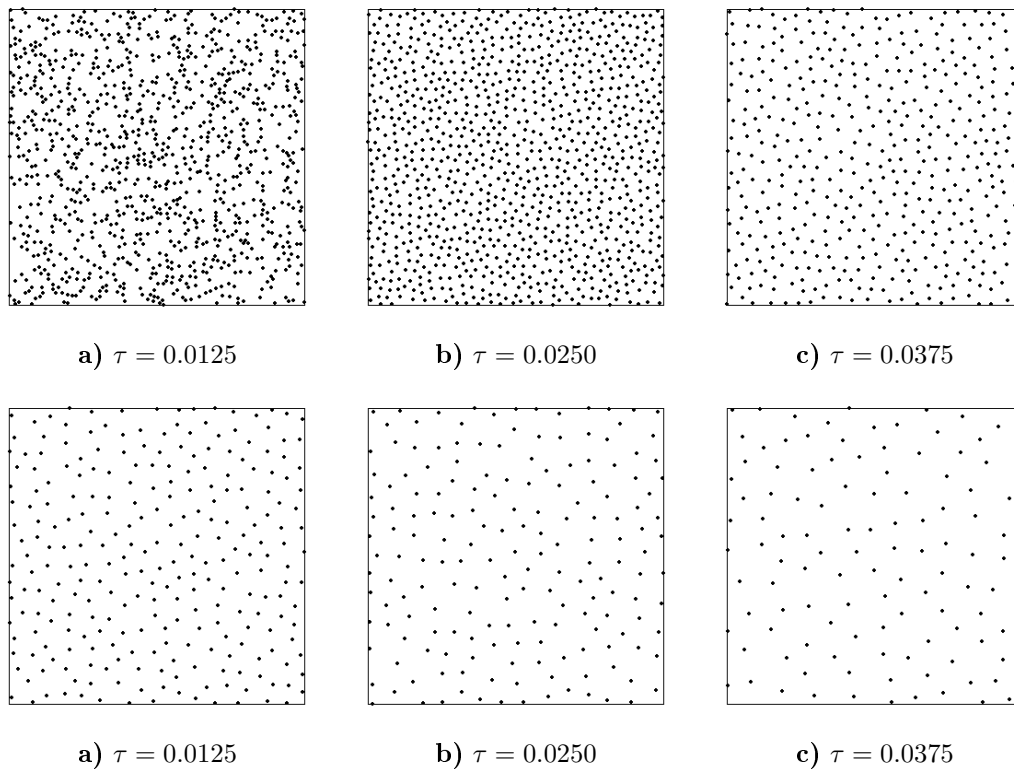


Figure 5.5: Examples of realizations from simple sequential inhibition processes for different configurations of the packing density τ .

The discrepancy statistic (5.3) is tested following a Monte Carlo procedure based on 100 simulations under the complete spatial randomness hypothesis. Tables 5.7 and 5.8 record the number of cases in which complete spatial randomness was rejected by a 4% test for the LGCP and inhibition models, respectively. Under the homogeneous Poisson process, the test rejected the null hypothesis in only four out of 100 cases.

	$\sigma^2 = 1$	$\sigma^2 = 2.5$	$\sigma^2 = 5$
$\phi = 0.05$	100	100	100
$\phi = 0.1$	100	100	100
$\phi = 0.25$	100	100	100

Table 5.7: Number of rejections of null hypothesis of complete spatial randomness against an LGCP with parameters σ^2 and ϕ by a 4% test.

The results show that the entropy-based statistic is highly effective in identifying de-

$\tau = 0.0125$	$\tau = 0.0250$	$\tau = 0.0375$	$\tau = 0.0500$	$\tau = 0.0625$	$\tau = 0.0750$
100	100	100	100	100	100

Table 5.8: Number of rejections of null hypothesis of complete spatial randomness against a simple sequential inhibition process with packing density τ by a 4% test.

viations from complete spatial randomness. In all evaluated configurations of the LGCP and inhibition processes, the test rejected the null hypothesis in every replication. This indicates that the test is sensitive to both clustering and regularity, effectively distinguishing between different spatial structures. Moreover, the test obtains a high significance level under the null hypothesis, showing appropriate calibration.

5.4 Data application: forest fires in Nepal

In this section, we apply the proposed entropy-based goodness-of-fit test to a dataset of forest fires recorded in Nepal from 2012 to 2016. The aim is to evaluate, for different models, their coherence with the observed spatial pattern by using the normalized Shannon entropy as a summary statistic. The spatial distribution of fire events is depicted in Figure 5.6.

To describe the spatial behavior of the process we consider two strategies: a kernel density estimation and a more sophisticated approach by using a spatial log-Gaussian Cox model. The latter allows us to account for the second-order characteristics of the process.

We use Diggle’s kernel density estimator with an Epanechnikov smoothing kernel. The bandwidth, τ , is selected by using the mean-square error criterion of Diggle, Equation 2.25, and edge correction is applied using the Jones-Diggle’s method (Equation 2.29). The resulting estimated intensity is shown in Figure 5.7(a).

For the second strategy, we use the modelling framework of a log-Gaussian Cox process model with intensity function,

$$\Lambda(\mathbf{s}) = \lambda(\mathbf{s}) \exp \{G(\mathbf{s})\} \quad (5.6)$$

where $\lambda(\mathbf{s})$ is the previously fitted intensity from the kernel density estimation, and $G(\mathbf{s})$ is a Gaussian process that captures the residual variation, adding stochastic structure to

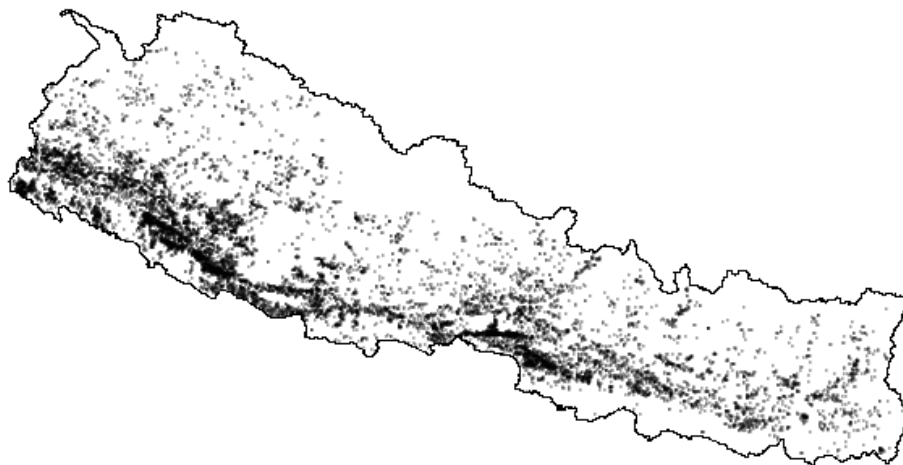


Figure 5.6: Forest fires occurred in Nepal from 2012 to 2016.

the model. The estimation of G is performed by using the Metropolis-Adjusted Langevin algorithm, as described in Section 2.2.4. We consider an exponential covariance model to describe the correlation structure of the process, with parameters estimated via a minimum contrast procedure, resulting in $\hat{\sigma}^2 = 1.6997$ and $\hat{\phi} = 0.3213$. The estimated intensity is shown in Figure 5.7(b).

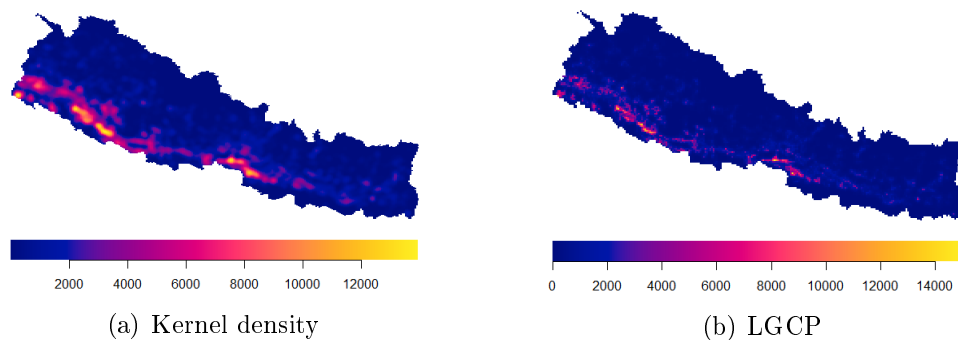


Figure 5.7: Estimated intensities by using a kernel density estimator (left) and by considering a log-Gaussian Cox process model (right).

For each model considered, we simulate 1000 realizations conditioned on each fitted intensity surface.

We set $k = 29$ partitions of the study region, ranging from $[4 \times 2]$ to $[60 \times 30]$ quadrats. These configurations are based on the bounding box containing Nepal, which has a width twice its height, $W = [80.0586, 88.2017] \times [26.3478, 30.4730]$. We perform a

quadrat counting procedure over each defined grid, on both the observed forest fire pattern and the 1000 realizations under each model considered. Afterwards, we compute, for each probability distribution derived from the quadrat counting, the normalized Shannon entropy.

Let $\hat{H}_1(\mathbf{q}_k)$ denote the normalized entropy of the observed pattern, and $\hat{H}_i(\mathbf{q}_k)$ the entropy for the i -th simulation under a given model. We define the upper and lower simulation envelopes as:

$$U(k) = \max_{i=1,\dots,100} \hat{H}_i(\mathbf{q}_k); \quad L(k) = \min_{i=1,\dots,100} \hat{H}_i(\mathbf{q}_k) \quad (5.7)$$

for each partition k .

Since the theoretical distribution of the Shannon entropy under the null hypothesis is unknown, we use the unbiased estimate $\bar{H}_i$, defined in Equation (5.4).

We use u_i , defined in Equation (5.3), as a measure of discrepancy between $\hat{H}_i(\mathbf{p}_k)$ and $\bar{H}_i(\mathbf{p}_k)$ over the whole range of k quadrat configurations.

We perform a test based on the rank of u_1 , yielding a p-value of 0.13 for the model based on kernel density estimation and a p-value of 0.45 for the log-Gaussian Cox process model. The null hypothesis in each case is that the observed spatial pattern of forest fires is a realization of from the corresponding fitted model. Figure 5.8 shows the function $\hat{H}_1$, computed for the observed pattern, and the envelope from the 1000 realizations of both fitted models, together with $\bar{H}_1$. Both models provide a satisfactory fit, with the observed curve largely lying within the simulation bounds. However, the model fitted by using a kernel density shows a greater deviation at finer spatial resolutions (smaller quadrat size), indicating a weaker local fit compared to the LGCP model, which remains within the envelopes across all scales considered.

Figure 5.9 shows the values of the statistics u_i against u_1 , obtained from the discrepancy between the summary statistic of the observed pattern and the unbiased estimator of $H(\mathbf{q}_k)$. The lower discrepancy observed for the LGCP model indicates a greater accuracy in capturing the spatial structure of the forest fire pattern.

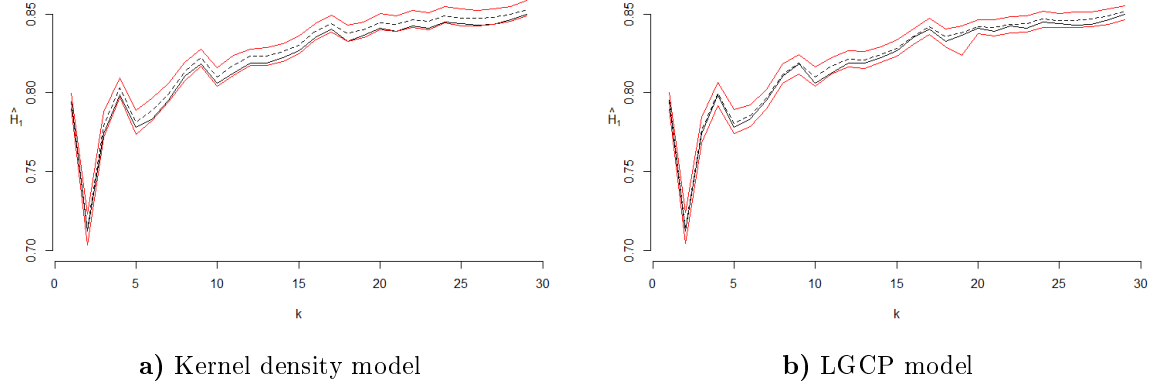


Figure 5.8: $\hat{H}_1(\mathbf{q}_k)$ (black line), $\bar{H}_1(\mathbf{q}_k)$ (dotted line), and upper and lower envelopes from 100 simulations (red dotted lines).

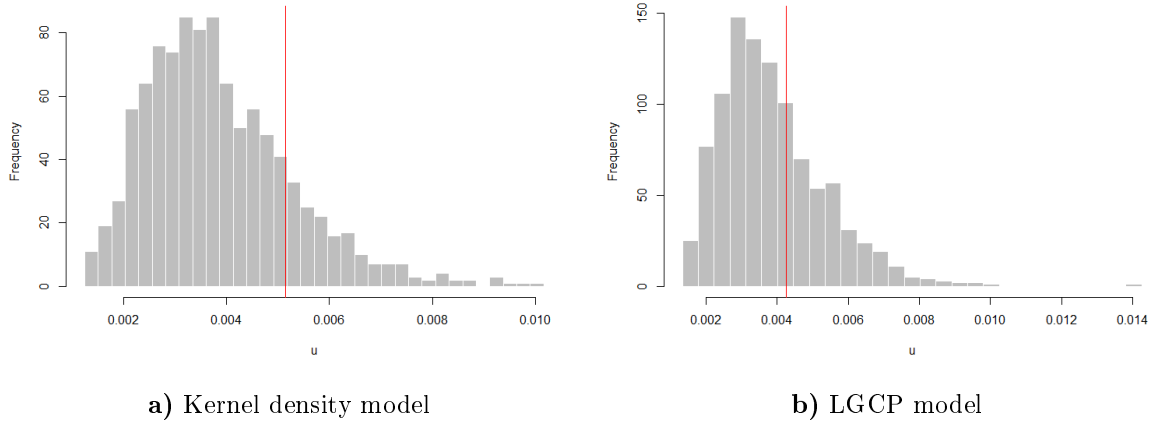


Figure 5.9: Histogram of discrepancy values u_i , and u_1 (red line)

5.5 Synthesis

In this chapter, we present a preliminary study aimed at developing a comparative tool for assessing the goodness of fit of spatial point process models, with a focus on log-Gaussian Cox processes (LGCP). Through a simulation study, we evaluated different configurations of the covariance parameters, σ^2 and ϕ , across nine scenarios, using a kernel density estimation with varying bandwidth to estimate the intensity functions of the observed patterns.

The results obtained show an almost perfect matching in terms of the Shannon entropy when comparing the original pattern with the patterns obtained through the most

accurate estimate considered, regardless of quadrat size. Performance declines as the accuracy of the estimates was reduced. Constant intensity models yielded more homogeneous realizations, resulting in entropy values closer to $H_{\max} = \ln(N)$.

Regarding the Kullback-Leibler measure, we observed that the divergence within the reference patterns was higher than that between reference and estimated patterns when we consider the most accurate estimate, with this indicating a strong source of variability in the two-steps pattern generation process of the LGCP. Larger bandwidths in the estimation led to poorer performance of the measure, as expected.

The exponential LMC complexity measures revealed a loss of structure in the estimated patterns, which became more pronounced with less accurate intensity estimates. For the constant intensity scenario, we obtain complexity values near zero, indicating a trivial structure.

The selection of the quadrat scheme has a crucial impact in the computation of these measures. For less clustered patterns, Shannon entropy performs well across all quadrat sizes. However, for processes with higher variance, smaller quadrats become necessary, except in cases of high variance combined with low scale parameters, where exponential LMC complexity and Kullback-Leibler divergence are more effective.

Extending these results, we proposed an entropy-based Monte Carlo goodness-of-fit test, which allows comparison across multiple spatial resolutions. This statistic was evaluated through simulations under three different spatial point process models; log-Gaussian Cox processes, simple sequential inhibition processes, and homogeneous Poisson processes. The test exhibited strong power in detecting both clustering and regularity, consistently rejecting the null hypothesis of complete spatial randomness in LGCP and inhibition processes, while maintaining appropriate calibration under the homogeneous Poisson process.

This framework was further applied to a real dataset of forest fires in Nepal. We assessed the fit of two models: a non-parametric kernel density estimator and an LGCP model. Simulated envelopes and discrepancy statistics were computed for each case. The results indicate that while kernel density-based model offers a good fits, the LGCP model describes more accurately the spatial structure, particularly at finer spatial resolutions. The observed entropy values for the kernel density estimates fall outside the simulation envelopes at smaller scales, suggesting an underestimation of local clustering. In contrast, the LGCP model maintained consistency across all quadrat configurations, highlighting

its ability to capture both first- and second-order properties of the data.

Overall, this approach offers a comprehensive framework for evaluating the goodness of fit of spatial point process models, providing valuable insights for both simulated and real data. However, further research is required to resolve some issues that arise from this assessment, such as the appropriate selection of the quadrat scheme, which has been observed to be highly dependent on the covariance parameters and can significantly influence the sensitivity and stability of the entropy-based statistics across different spatial structures. Additionally, extensions to spatio-temporal point processes and other spatial supports, such as networks, should be considered.

Chapter 6

Results

In this chapter we present the key results derived from this thesis work.

The characterization of both global and local structural information transfer between the two phases involved in the generation of a LGCP, as detailed in Chapter 3, has revealed that the amount of information transferred, as well as the structural characteristics of the process, are determined primarily by the values of the mean, variance and scale parameters. In contrast, changes in the smoothness parameter have a minimal impact. Specifically, increases in the variance lead to greater system complexity and reduced diversity in the marginal structure of both phases. Conversely, increases in the scale parameter lead to higher stochasticity and reduced structural coherence between phases. This study has led to the publication of the paper: Medialdea A., Angulo J.M., Mateu J. (2021) *Structural complexity and information transfer in spatial log-Gaussian Cox processes*, Entropy, **29**, 1135.

In Chapter 4, we have proposed a fully non-separable model within the class of LGCPs. This model incorporates a non-separable deterministic intensity and a non-separable spatio-temporal correlation structure, enabling it to capture the spatio-temporal dependence of point patterns. To demonstrate its practical usefulness, the model has been applied to a real dataset of forest fires in Nepal. Its predictive accuracy has been assessed using the weighted spatio-temporal K -function and compared against a separable model and a mixed model that combines a separable first-order intensity with a non-separable covariance structure. The results indicate that the fully non-separable model provides improved overall fitting. While the separable and mixed

scenarios perform similarly in global terms, the non-separable covariance structure of the mixed model improves the local fitting. This contribution has resulted in the paper: Medialdea A., Angulo J.M., Mateu J., Adelfio G. *A fully non-separable log-Gaussian Cox process to model forest fires* (pending submission).

A comparative analysis based on information and complexity measures has been presented in Chapter 5 to assess the goodness of fit of spatial point process models, with a particular focus on LGCPs. In a simulation study across nine scenarios with varying covariance parameters has been observed that the structural properties of these processes can be effectively captured by using Shannon entropy, exponential LMC complexity and Kullback-Leibler divergence. This enables us to discriminate between models in terms of goodness of fit and support model selection. However, the performance of the measures is sensitive to the choice of the quadrat scheme, which can affect their usefulness under certain parameter settings. Based on these results, an entropy-based Monte Carlo goodness-of-fit test has been proposed, allowing comparisons across multiple spatial resolution and thereby mitigating the influence of the quadrat configuration. Simulation results have confirmed the ability of the test to discriminate both clustering and regularity, with appropriate calibration, against complete spatial randomness. Finally, this methodology has been applied to the spatial dataset of forest fires in Nepal, demonstrating its practical applicability. A paper based on this work is currently in progress.

Overall, the results of this thesis work have provided significant insights into the structural properties and modeling of LGCPs. The analysis of information transfer has shown that the covariance parameters play a crucial role in determining the heterogeneity and global structural complexity of the process, as well as the local coherence between phases. The development of a fully non-separable LGCP model has led to improved predictive performance, globally through the inclusion of a non-separable deterministic intensity, and locally due to the non-separable correlation structure, as demonstrated by its application to forest fire data. Furthermore, the introduction of information and complexity measures for assessing the goodness-of-fit has proven to be an effective tool for model discrimination. The proposed entropy-based Monte Carlo goodness-of-fit test has shown to be capable in detecting deviations from complete spatial randomness across varying spatial resolutions, thereby serving as a useful tool for model selection. Together, these contributions result in a methodological framework for the analysis and modeling of spatial and spatio-temporal point processes, with a practical applicability to real data.

Capítulo 6

Resultados

En este capítulo se presentan los resultados clave derivados de este trabajo de tesis.

La caracterización de la transferencia de información estructural tanto global como local entre las dos fases involucradas en la generación de un LGCP, tal como se detalla en el Capítulo 3, ha revelado que la cantidad de información transferida, así como las características estructurales del proceso, están determinadas principalmente por los valores de los parámetros de media, varianza y escala. En contraste, cambios en el parámetro de suavizamiento producen un impacto mínimo. Específicamente, aumentos en la varianza conducen a una mayor complejidad del sistema y a una reducción en la diversidad de la estructura marginal de ambas fases. Por otro lado, incrementos en el parámetro de escala resultan en una mayor estocasticidad y en una disminución de la coherencia estructural entre las fases. Este estudio ha dado lugar a la publicación del artículo: Medialdea A. Angulo J.M., Mateu J. (2021) *Structural complexity and information transfer in spatial log-Gaussian Cox processes*, Entropy, **29**, 1135.

En el Capítulo 4, se ha propuesto un modelo completamente no separable dentro de la clase de los LGCPs. Este modelo incorpora una intensidad determinista no separable y una estructura de correlación espacio-temporal no separable, lo que permite capturar la dependencia espacio-temporal de los patrones puntuales. Para demostrar su utilidad práctica, el modelo ha sido aplicado a un conjunto de datos reales sobre incendios forestales en Nepal. Su precisión en términos de predicción ha sido evaluada mediante la versión ponderada de la función K espacio-temporal y comparada con un modelo separable y un modelo mixto que combina una intensidad de primer orden separable con

una estructura de covarianza no separable. Los resultados indican que el modelo completamente no separable proporciona un mejor ajuste general. Mientras que los escenarios separable y mixto presentan desempeños similares a nivel global, la estructura de covarianza no separable del modelo mixto mejora el ajuste local. Esta contribución ha dado lugar al artículo: Medialdea A., Angulo J.M., Mateu J., Adelfio, G. *A fully non-separable log-Gaussian Cox process to model forest fires* (pendiente de someter).

En el Capítulo 5 se presenta un análisis comparativo basado en medidas de información y complejidad para evaluar el ajuste de modelos de procesos puntuales espaciales, con un enfoque particular en los LGCPs. Tras un estudio de simulación para nueve escenarios con distintos parámetros de covarianza, se ha mostrado que las propiedades estructurales de estos procesos pueden ser descritas mediante el uso de la entropía de Shannon, la complejidad LMC exponencial y la divergencia de Kullback-Leibler. Estas medidas permiten discriminar entre modelos en términos de bondad de ajuste, facilitando la selección de un modelo apropiado. Sin embargo, el desempeño de estas medidas es sensible a la elección del esquema de la cuadrícula de conteo, lo que puede afectar a su utilidad bajo ciertas configuraciones de parámetros. A partir de estos resultados, se ha propuesto un test Monte Carlo de bondad de ajuste basado en entropía que permite comparaciones entre múltiples resoluciones espaciales y, por lo tanto, reduce la influencia de la configuración de la cuadrícula de conteo. Los resultados de simulación confirman la capacidad del test para discriminar tanto agrupamiento como regularidad, manteniendo una calibración adecuada, frente a aleatoriedad espacial completa. Finalmente, esta metodología ha sido aplicada al conjunto de datos espaciales de incendios forestales en Nepal, demostrando su aplicabilidad práctica. Actualmente, hay un artículo en desarrollo basado en este trabajo.

En conjunto, los resultados de esta tesis han contribuido al análisis de las propiedades estructurales y a la modelización de los LGCPs. Mediante el análisis de la transferencia de información se ha mostrado que los parámetros de covarianza determinan gran parte de la heterogeneidad y la complejidad estructural global del proceso, así como la coherencia local entre sus fases. El desarrollo de un modelo completamente no separable para LGCPs se ha traducido en una mejora del desempeño en términos de predicción, globalmente mediante la inclusión de una intensidad determinista no separable, y localmente debido a la estructura de correlación no separable, como se demuestra en su aplicación a datos de incendios forestales. Además, la introducción de medidas de información y complejidad

como herramientas para evaluar la bondad de ajuste ha demostrado ser efectivo para discriminar entre modelos. El test Monte Carlo de bondad de ajuste basado en entropía se ha mostrado eficaz para detectar desviaciones respecto a la aleatoriedad espacial completa a través de distintas resoluciones espaciales, siendo una herramienta útil para la selección de modelos. En conjunto, estas contribuciones conforman un marco metodológico para el análisis y modelización de procesos puntuales espaciales y espacio-temporales, con aplicabilidad práctica a datos reales.

Capítulo 7

Conclusions and Future Research Lines

In this thesis work, we study various aspects related to the analysis of log-Gaussian Cox processes and the description of their structural characteristics based on Information and Complexity measures.

Specifically, in Chapter 3 we present an assessment focused on the analysis of the global and local structural information transfer from the intensity fields to the point patterns of log-Gaussian Cox processes. We illustrate that such transfer of information is primarily influenced by the mean, variance and scale parameters. The smoothness parameter, however, does not significantly affect the information transferred between phases. In particular, higher variance leads to increased system complexity and decreased diversity of the resulting patterns. The study also shows that an increase in mean and variance enhances the structural information transfer, while an increase in the scale parameter diminishes the structure of the intensity field and increases the stochastic nature of pattern generation. The study suggests the potential for further research into the analytical aspects of structural complexity in these processes, including multifractal Cox processes and spatio-temporal predictive risk assessment.

In Chapter 4 we propose a doubly non-separable model for log-Gaussian Cox processes, with the aim of capturing the spatio-temporal dependencies of these processes. Applied to a dataset of forest fires in Nepal, we obtain that the model outperforms separable and mixed models in terms of fitting and predictive accuracy, particularly in capturing seasonal patterns and interactions between spatial and temporal dimensions. The non-separable covariance structure enhances the local fitting, suggesting the ap-

appropriateness of this model. The results suggest the extension of the study to include analytical methods for evaluating model fitting and to consider alternative spatial supports like linear networks.

In Chapter 5 we develop a comparative analysis for assessing the goodness of fit in spatial point process models, with a focus on log-Gaussian Cox processes. Through simulation, we find that Shannon entropy performs well under all scenarios considered, regardless of covariance parameter values and quadrat size, while Kullback-Leibler divergence and exponential LMC complexity measures are more effective for patterns with high variance or low scale parameters. The study also highlights the importance of the quadrat scheme selection, which significantly impacts the performance of these measures. Based on the outcomes of this analysis, we proposed an entropy-based goodness-of-fit test within a Monte Carlo framework. The application of this approach to a dataset of forest fires illustrates its practical usefulness in detecting deviations from assumed models. Nonetheless, further research is needed to refine the methodology, particularly in terms of quadrat scheme selection and extending the analysis to spatio-temporal processes and other spatial supports.

The development of this thesis work has opened various lines for further research, including different aspects:

- **Multifractality:** Multifractal Cox processes (Baile *et al.*, 2021¹¹⁷) constitute an important extension for analysis based on the connection between Information and Complexity measures and multifractal dimensions (see, for example, Angulo *et al.*, 2021⁵³).
- **Spatio-temporal assessment based on Information and Complexity measures:** Implications with reference to predictive risk assessment and mapping (Romero *et al.*, 2018⁵⁹), regarding the information-complexity characterization of different scenarios.
- **Goodness-of-fit test based on other information and complexity measures:** Extending the use of complexity measures such as LMC complexity and Kullback-Leibler divergence, as well as, generalizations based on a deformation parameter (*e.g.* Rényi and Tsallis), as alternatives for goodness-of-fit testing, allowing a more complete characterization of the spatial dynamics of point processes.

- Alternative spatial supports: Investigating other spatial supports, such as linear networks, and extending the proposed model and methodologies to different domains.
- Quadrat scheme selection: As the selection of the quadrat scheme significantly impacts the performance of information and complexity measures, further research is needed to develop guidelines or automated methods for optimal scheme selection based on model parameters.

Capítulo 7

Conclusiones y Futuras Líneas de Investigación

En este trabajo de tesis, se estudian diversos aspectos relacionados con el análisis de procesos de Cox log-gaussianos y la descripción de sus características estructurales basadas en medidas de información y complejidad.

Específicamente, en el Capítulo 3 presentamos una evaluación centrada en el análisis de la transferencia de información estructural global y local desde los campos de intensidad hacia los patrones de puntos de procesos de Cox log-gaussianos. Ilustramos que dicha transferencia de información está influenciada principalmente por los parámetros de media, varianza y escala. El parámetro de suavizamiento, sin embargo, no afecta significativamente a la información transferida entre las fases de generación del proceso. En particular, una mayor varianza produce un incremento en la complejidad del sistema y una disminución de la diversidad de los patrones resultantes. El estudio también muestra que un aumento en los valores de la media y la varianza mejora la transferencia de información estructural, mientras que un aumento en el parámetro de escala disminuye la estructura del campo de intensidad e incrementa la naturaleza estocástica de la generación de patrones. El estudio sugiere la necesidad de investigación adicional sobre aspectos analíticos de la complejidad estructural de estos procesos, incluyendo los procesos de Cox multifractales y la evaluación predictiva del riesgo en contextos espacio-temporales.

En el Capítulo 4 se propone un modelo doblemente no separable para procesos de

Cox log-gaussianos, con el objetivo de capturar la dependencia espacio-temporal de estos procesos. Aplicado a un conjunto de datos de incendios forestales en Nepal, el modelo proporciona mejores resultados que los modelos separables y mixtos en términos de ajuste y precisión en la predicción, particularmente en la descripción de patrones estacionales e interacciones entre la dimensión espacial y temporal. La estructura de covarianza no separable mejora el ajuste local, lo que implica que este modelo es más adecuado. Los resultados sugieren la extensión del estudio mediante la inclusión de métodos analíticos para evaluar la bondad de ajuste del modelo, así como considerar otras alternativas de soporte espacial como redes lineales.

En el Capítulo 5 se desarrolla un análisis comparativo para evaluar la bondad de ajuste en modelos de procesos puntuales espaciales, con un enfoque en los procesos de Cox log-gaussianos. A través de un estudio de simulación, encontramos que la entropía de Shannon funciona bien en todos los escenarios considerados, independientemente de los valores de los parámetros de covarianza y del tamaño de la cuadrícula de conteo, mientras que la divergencia de Kullback-Leibler y la complejidad LMC son más efectivas para patrones con valores altos de varianza o valores reducidos del parámetro de escala. El estudio también resalta la importancia de la selección del tamaño de la cuadrícula de conteo al impactar significativamente en el rendimiento de estas medidas. En base a los resultados de este análisis, se propone un test de bondad de ajuste basado en entropía dentro del marco de un algoritmo de Monte Carlo. La aplicación de este enfoque a un conjunto de datos de incendios forestales nos permite ilustrar su utilidad práctica para detectar desviaciones respecto a los modelos considerados. No obstante, se requiere investigación adicional para refinar la metodología, en particular, en cuanto a la selección de la cuadrícula de conteo y la posible extensión del análisis a procesos espacio-temporales y otros soportes espaciales.

El desarrollo de este trabajo de tesis ha abierto varias líneas de investigación adicional, que incluyen diferentes aspectos:

- Multifractalidad: Los procesos de Cox multifractales (Baile *et al.*, 2021¹¹⁷) constituyen una extensión importante para el análisis basado en la conexión entre las medidas de información y complejidad y las dimensiones multifractales (ver, por ejemplo, Angulo *et al.*, 2021⁵³).
- Evaluación espacio-temporal basada en medidas de información y complejidad: Im-

plicaciones referentes a la evaluación predictiva y la creación de mapas de riesgo (Romero *et al.*, 2018⁵⁹), en relación a la caracterización de la información y complejidad en diferentes escenarios.

- Test de bondad de ajuste basado en otras medidas de información y complejidad: Extensión del uso de medidas de complejidad y divergencia, como la complejidad LMC y la divergencia de Kulback-Leibler; así como generalizaciones basadas en un parámetro de deformación (*e.g.* Rényi and Tsallis), como alternativas para la construcción del test de bondad de ajuste, permitiendo una captura más completa de la dinámica espacial de los procesos puntuales.
- Soportes espaciales alternativos: Investigación de otros soportes espaciales, como redes lineales, y extensión del modelo y las metodologías propuestas a diferentes dominios.
- Selección de la cuadrícula de conteo: Dado que la selección del tamaño de la cuadrícula de conteo impacta significativamente en el desempeño de las medidas de información y complejidad, se requiere investigación adicional para el desarrollo de reglas o métodos automatizados basados en los parámetros del modelo para la selección óptima de la cuadrícula.

Bibliography

- [1] Feller W. (1950) *An Introduction to Probability Theory and its Applications*. Wiley, New York
- [2] Smith W. (1958) Renewal theory and its ramifications. *Journal of the Royal Statistical Society Series B: Statistical Methodology* **20**, 243-284
- [3] Ellis R.L. (1844) On the foundation of theory of probabilities. *Transactions of the Cambridge Philosophical Society*, Part I (1842), 1-6. Read February 14, 1842. Reprinted in Ellis [1863, pp. 12-37]
- [4] Palm C. (1943) Intensity variations in telephone traffic. *Ericsson Technics* **44**, 1-189
- [5] Artzner P., Delbaen F., Eber J.M., Heath D. (1999) Coherent measures of risk. *Mathematical Finance* **9**, 203-228
- [6] Föllmer H., Schied A. (2002) Convex measures of risk and trading constraints. *Finance and Stochastics* **6**, 429-447
- [7] Frittelli M., Rosazza Gianin E. (2002) Putting order in risk measures. *Journal of Banking and Finance* **26**, 1473-1486
- [8] Acerbi C. (2002) Spectral measures of risk: A coherent representation of subjective risk aversion. *Journal of Banking and Finance* **26**, 1505-1518
- [9] Shannon C.E. (1948) A mathematical theory of communication. *Bell System Technical Journal* **27**, 379-423
- [10] Rényi A. (1961) On measures of entropy and information. *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability* **1**, 547-561
- [11] Tsallis C. (1988) Possible generalization of Boltzmann-Gibbs statistics. *Journal of Statistical Physics* **52**, 479-487
- [12] Havrda J., Charvát F. (1967) Quantification method of classification processes. Concept of structural α -entropy. *Kybernetika* **3**, 30-35
- [13] Chilès J.P., Delfiner P. (2012) *Geostatistics: Modeling Spatial Uncertainty*. Wiley, New York

-
- [14] Adler R.J., Taylor J.E. (2007) *Random Fields and Geometry*. Springer, New York
 - [15] Stein M.L. (1999) *Interpolation of Spatial Data*. Springer, New York
 - [16] Christakos G. (1992) *Random Field Models in Earth Sciences*. Academic Press, San Diego
 - [17] Hristopulos D.T. (2020) *Random Fields for Spatial Data Modeling*. Springer Netherlands, Dordrecht
 - [18] Matérn B (1986) *Spatial Variation*. Springer, New York
 - [19] DeIaco S., Myers D., Posa D. (2002) Nonseparable space-time covariance models: some parametric families. *Mathematical Geology* **34**, 23-42
 - [20] Gneiting T. (2002) Nonseparable, stationary covariance function for space-time data. *Journal of the American Statistical Association* **97**, 590-600
 - [21] Rodríguez-Iturbe I., Mejía J.M. (1974) The design of rainfall networks in time and space. *Water Resources Research* **10**, 713-729
 - [22] Diggle P., Tawn J.A., Moyeed R.A. (1998) Model-based geostatistics. *Applied Statistics* **47**, 299-326
 - [23] Whittle P. (1954) On stationary processes in the plane. *Biometrika* **41**, 434-449
 - [24] Myers D.E., Journel A.G. (1990) Variograms with zonal anisotropies and noninvertible kriging systems. *Mathematical Geology* **22**, 779-785
 - [25] DeCesare L., Myers D., Posa D. (2001) Estimating and modeling space-time correlation structures. *Statistics and Probability Letters* **51**, 9-14
 - [26] Cressie N., Huang H. (1999) Classes of nonseparable, spatio-temporal stationary covariance functions. *Journal of the American Statistical Association* **448**, 1330-1340
 - [27] Ripley B. (1977) Modelling spatial patterns. *Journal of the Royal Statistical Society: Series B (Methodological)* **39**, 172-192
 - [28] Diggle P. (1985) A kernel method for smoothing point process data. *Applied Statistics* **34**, 138-147
 - [29] Baddeley A., Rubak E., Turner R. (2015) *Spatial Point Patterns: Methodology and Applications with R*. Chapman and Hall/CRC, New York
 - [30] Scott D. (1992) *Multivariate Density Estimation*. Wiley, New Jersey
 - [31] Cronie O., Van Lieshout M. (2018) A non-model-based approach to bandwidth selection for kernel estimators of spatial intensity functions. *Biometrika* **105**, 455-462

- [32] Gasser T., Müller H.G. (1979) Kernel estimation of regression functions. In: *Smoothing Techniques for Curve Estimation. Lecture Notes in Mathematics*, Gasser T., Rosenblatt M. (eds.) 23-58. Springer, Berlin
- [33] Jones M.C. (1993) Simple boundary correction for kernel density estimation. *Statistics and Computing* **3**, 135-146
- [34] Diggle P. (2010) Nonparametric methods. In: *Handbook of Spatial Statistics*, Gelfand A.E., Diggle P., Guttorp P, Fuentes M. (eds.) 299-316. CRC Press, Boca Raton
- [35] Loader C. (1999) *Local Regression and Likelihood*. Springer, New York
- [36] Getis A. (1984) Interaction modeling using second-order analysis. *Environment and Planning A: Economy and Space* **16**, 173-183
- [37] Stoyan D., Stoyan H. (1994) *Fractals, Random Shapes and Point Fields: Method of Geometrical Statistics*. Wiley, New York
- [38] Baddeley A.J., Møller J., Waagepetersen R. (2000) Non and semi-parametric estimation of interaction in inhomogeneous point patterns. *Statistica Neerlandica* **54**, 329-350
- [39] Gabriel E., Diggle P.J. (2009) Second-order analysis of inhomogeneous spatio-temporal point process data. *Statistica Neerlandica* **63**, 43-51
- [40] Adelfio G., Siino M., Mateu J., 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**, 149-168
- [41] Cox D.R. (1955) Some statistical methods connected with series of events. *Journal of the Royal Statistical Society. Series B (Methodological)* **17**, 129-157
- [42] Neyman J., Scott E.L. (1958) A statistical approach to problems of cosmology. *Journal of the Royal Statistical Society. Series B (Methodological)* **20**, 1-43
- [43] Brix A., Diggle P. (2001) Spatiotemporal prediction for log-Gaussian Cox processes. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)* **63**, 823-841
- [44] Møller J., Syversveen A., Waagepetersen R. (1998) Log Gaussian Cox processes. *Scandinavian Journal of Statistics* **25**, 451-482
- [45] Brix A., Møller J. (2001) Space-time multi type log Gaussian Cox processes with a view to modelling weeds. *Scandinavian Journal of Statistics* **28**, 471-448
- [46] Taylor B., Davies T., Rowlingson B. (2013) lgcp: An R package for inference with spatial and spatio-temporal log-Gaussian Cox processes. *Journal of Statistical Software* **52**, 1-40
- [47] Wood A., Chan G. (1994) Simulation of stationary Gaussian processes in $[0, 1]^d$. *Journal of Computational and Graphical Statistics* **3**, 409-432

-
- [48] Taylor B., Diggle P. (2014) INLA or MCMC? A tutorial and comparative evaluation for spatial prediction in log-Gaussian Cox processes. *Journal of Statistical Computation and Simulation* **84**, 2266-2284
 - [49] Roberts G., Tweedie R. (1996) Exponential convergence of Langevin distributions and their discrete approximations. *Bernoulli* **2**, 341-363
 - [50] Kullback S., Leibler R.A. (1951) On information and sufficiency. *The Annals of Mathematical Statistics* **22**, 79-86
 - [51] Catalán R.G., Garay J., López-Ruiz R. (2002) Features of the extension of a statistical measure of complexity to continuous systems. *Physical Review E* **66**, 011102
 - [52] López-Ruiz R., Mancini H., Calbet X. (1995) A statistical measure of complexity. *Physics Letters A* **209**, 321-326
 - [53] Angulo J.M., Esquivel F.J., Madrid A.E., Alonso F.J. (2021) Information and complexity analysis of spatial data. *Spatial Statistics* **42**, 100462
 - [54] Campbell L.L. (1966) Exponential entropy as a measure of extent of a distribution. *Z. Wahrscheinlichkeitstheorie verw Gebiete* **5**, 217-225
 - [55] López-Ruiz R., Nagy Á., Romera E., Sañudo J. (2009) A generalized statistical complexity measure: Applications to quantum systems. *Journal of Mathematical Physics* **50**, 123528
 - [56] Romera E., Sen K.D., Nagy Á. (2011) A generalized relative complexity measure. *Journal of Statistical Mechanics: Theory and Experiment* **2011**, 9-16
 - [57] Hartley R.V.L. (1928) Transmission of information. *Bell System Technical Journal* **7**, 535-563
 - [58] Koch E. (2017) Spatial risk measures and applications to max-stable processes. *Extremes* **20**, 635-670
 - [59] Romero J.L., Madrid A.E., Angulo J.M. (2018) Quantile-based spatiotemporal risk assessment of exceedances. *Stochastic Environmental Research and Risk Assessment* **32**, 2275-2291
 - [60] Koch E. (2019) Spatial risk measures and rate of spatial diversification. *Risks* **7**, 52
 - [61] Fölmer H., Schied A. (2004) *Stochastic Finance: An Introduction in Discrete Time*. De Gruyter, Berlin
 - [62] Fölmer H., Knispel T. (2011) Entropic risk measures: coherence vs. convexity, model ambiguity and robust large deviations. *Stochastics and Dynamics* **11**, 333-351
 - [63] Coles P., Jones B. (1991) A lognormal model for the cosmological mass distribution. *Monthly Notices of the Royal Astronomical Society* **248**, 1-13

-
- [64] Rathbun S. (1996) Estimation of Poisson intensity using partially observed concomitant variables. *Biometrics* **52**, 226-242
- [65] Illian J., Martino S., Sørbye S., Gallego-Fernández J., Zunzunegui M., Esquivias M., Travis J. (2013) Fitting complex ecological point process models with integrated nested Laplace approximation. *Methods in Ecology and Evolution* **4**, 305-315
- [66] Diggle P., Zheng P., Durr P. (2005) Nonparametric estimation of spatial segregation in a multivariate point process: bovine tuberculosis in Cornwall, UK. *Journal of the Royal Statistical Society. Series C (Applied Statistics)* **54**, 645-658
- [67] Illian J., Møller J., Waagepetersen R. (2009) Hierarchical spatial point process analysis for a plant community with high biodiversity. *Environmental and Ecological Statistics* **16**, 389-405
- [68] Waagepetersen R., Schweder T. (2006) Likelihood-based inference for clustered line transect data. *Journal of Agricultural, Biological, and Environmental Statistics* **11**, 264-279
- [69] Rodrigues A., Diggle P. (2012) Bayesian estimation and prediction for inhomogeneous spatiotemporal log-Gaussian Cox processes using low-rank models, with applications to criminal surveillance. *Journal of the American Association* **107**, 93-101
- [70] Siino M., Adelfio G., Mateu J., Chiodi M., D'Alessandro A. (2017) Spatial pattern analysis using hybrid models: an application to the Hellenic seismicity. *Stochastic Environmental Research and Risk Assessment* **31**, 1633-1648
- [71] Papangelou F. (1978) On the entropy rate of stationary point processes and its discrete approximation. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete* **44**, 191-211
- [72] Baratpour S., Ahmadi J., Arghami N. (2007) Some characterizations based on entropy of order statistics and record values. *Communications in Statistics - Theory and Methods* **36**, 47-57
- [73] Daley D., Vere-Jones D. (2004) Scoring probability forecast for point processes: the entropy score and information gain. *Journal of Applied Probability* **41**, 297-312
- [74] Flannigan M.D., Wotton B.M. (2001) Climate, weather and area burned. In: *Forest Fires: Behavior and Ecological Effects* Johnson E.A., Miyanishi K. (eds.) 351-373. Academic Press, San Diego
- [75] Parajuli A., Chand D.B., Rayamajhi B., Khanal R., Baral S., Malla Y., Poudel S. (2015) Spatial and temporal distribution of forest fires in Nepal. In: *XIV World Forestry Congress*
- [76] Bhujel K., Maskey-Byanju R., Gautam A. (2017) Wildfire dynamics in Nepal from 2000-2016. *Nepal Journal of Environmental Science* **5**, 1-8

-
- [77] Pechony O., Shindell D. (2010) Driving forces of global wildfires over the past millennium and the forthcoming century. *Proceedings of the National Academy of Sciences* **107**, 19167-19170
- [78] Juan P., Mateu J., Saez M. (2012) Pinpointing spatio-temporal interactions in wildfire patterns. *Stochastic Environmental Research and Risk Assessment* **26**, 1131-1150
- [79] Pereira P., Turkman K., Amaral Turkman M., Sá A., Pereira J. (2013) Quantification of annual wildfire risk; A spatio-temporal point process approach. *Statistica* **73**, 55-68
- [80] Møller J., Díaz-Ávalos C. (2010) Structured spatio-temporal shot-noise Cox point process models, with a view to modelling forest fires. *Scandinavian Journal of Statistics* **37**, 2-25
- [81] Serra L., Juan P., Varga D., Mateu J., Saez M. (2013) Spatial pattern modelling of wildfires in Catalonia, Spain 2004-2008. *Environmental Modelling and Software* **40**, 235-244
- [82] Serra L., Saez M., Mateu J., Varga D., Juan P., Díaz-Ávalos C., Rue H. (2014) Spatio-temporal log-Gaussian Cox processes for modelling wildfire occurrence: the case of Catalonia 1994-2008. *Environmental and Ecological Statistics* **21**, 531-563
- [83] Gabriel E., Opitz T., Bonneu F. (2017) Detecting and modeling multi-scale space-time structures: the case of wildfire occurrences. *Journal of the French Statistical Society* **158**, 86-105
- [84] Opitz T., Bonneu F., Gabriel E. (2020) Point-process based Bayesian modeling of space-time structures of forest fire occurrences in Mediterranean France. *Spatial Statistics* **40**, 100429
- [85] Kelling C., Haran M. (2020) A two-stage Cox process model with spatial and nonspatial covariates. *Spatial Statistics* **51**, 100685
- [86] Gilardi A., Borgoni R., Mateu J. (2021) A non-separable first-order spatio-temporal intensity for events on linear network: An application to ambulance interventions. *Annals of Applied Statistics* **18**, 529-554
- [87] Siino M., Adelfo G., Mateu J. (2018) Joint second-order parameter estimation for spatio-temporal log-Gaussian Cox processes. *Stochastic Environmental Research and Risk Assessment* **32**, 3525-3539
- [88] Kolovos A., Christakos G., Hristopulos D.T., Serre M.L. (2004) Methods for generating non-separable spatiotemporal covariance models with potential environmental applications. *Advances in Water Resources* **27**, 815-830
- [89] Diggle P.J., Moraga P., Rowlingson B.M., Taylor B. (2013) Spatial and spatio-temporal log-Gaussian Cox processes: extending the geostatistical paradigm. *Statistical Science* **28**, 542-563

-
- [90] Gneiting T., Genton M., Guttorp P. (2006) *Geostatistical Space-Time Models, Stationarity, Separability, and Full Symmetry*. Chapman and Hall/CRC, New York
- [91] Zhou Z., Matteson D.S., Woodard D.B., Henderson S.G., Micheas A.C. (2015) A spatio-temporal point process model for ambulance demand. *Journal of the American Statistical Association* **110**, 6-15
- [92] Gabriel E. (2012) Estimating second-order characteristics of inhomogeneous spatio-temporal point processes: influence of edge correction methods and intensity estimates. *Methodology and Computing in Applied Probability* **16**, 411-431
- [93] Sheather S.J., 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**, 683-690
- [94] Rudemo M. (1964) Dimension and entropy for a class of stochastic processes. *Publication of the Mathematical Institute of the Hungarian Academy of Sciences* **9**, 73-87
- [95] McFadden J. (1965) The entropy of a point process. *Journal of the Society for Industrial and Applied Mathematics* **13**, 998-994
- [96] Fritz J. (1973) An approach to the entropy of point processes. *Periodica Mathematica Hungarica* **3**, 73-83
- [97] Baccelli F., Woo J. (2016) On the entropy and mutual information of point processes. *2016 IEEE International Symposium on Information Theory (ISIT)*, 695-699
- [98] Clark D. (2020) Local entropy statistics for point processes. *IEEE Transactions on Information Theory* **66**, 1155-1163
- [99] Kittel C., Kroemer H. (1980) *Thermal Physics*. W.H Freeman and Company, New York
- [100] Planck M. (1914) *The Theory of Heat Radiation*. P. Blakiston's Son and Co., Philadelphia
- [101] Cushman S. (2021) Generalizing Boltzmann configurational entropy to surfaces, point patterns and landscape mosaics. *Entropy* **23**, 1616
- [102] Altieri L., Cocchi D. (2024) *Entropy Measures for Environmental Data*. Springer Nature, Singapore
- [103] Magurran A.E. (2004) *Measuring Biological Diversity*. Blackwell Publishing, Oxford
- [104] Frosini B.V. (2006) Descriptive measures of ecological diversity. In: *Environmetrics. Encyclopedia of life support systems* Jureckova J., El-Shaarawi A.H. (eds.), Eolss Publishers, Oxford.
- [105] Altieri L., Cocchi D., Roli G. (2021) Spatial entropy for biodiversity and environmental data: The R-package SpatEntropy. *Environmental Modelling and Software* **144**, 105149

-
- [106] Ma C. (2003) Families of spatio-temporal stationary covariance models. *Journal of Statistical Planning and Inference* **116**, 489-501
- [107] Guttorp P., Gneiting T. (2006) Studies in the history of probability and statistics XLIX On the Matérn correlation family. *Biometrika* **93**, 989-995
- [108] Chen W., Genton M.G., Sun Y. (2021) Space-time covariance structures and models. *Annual Review of Statistics and Its Application* **8**, 191-215
- [109] Schlather M., Malinowski A., Menck P.J., Oesting M., Strokorb K. (2015) Analysis, simulation and prediction of multivariate random fields with package RandomFields. *Journal of Statistical Software* **63**, 1548-7660
- [110] Wackernagel H. (2003) *Multivariate Geostatistics*. Springer, Berlin
- [111] Arbia G. (2006) *Spatial Econometrics: Statistical Foundations and Applications to Regional Convergence*. Springer-Verlag, Berlin
- [112] Holt C. (2004) Forecasting seasonals and trends by exponentially weighted moving averages. *International Journal of Forecasting* **20**, 5-10
- [113] Winters P. (1960) Forecasting sales by exponentially weighted moving averages. *Management Science* **6**, 324-342
- [114] Raskshit S., Davies T., Moradi M., McSwiggan G., Nair G., Mateu J., Baddeley A. (2019) Fast kernel smoothing of point patterns on a large network using two-dimensional convolution. *International Statistical Review* **87**, 531-556
- [115] Handcock M., Stein M. (1993) A Bayesian analysis of kriging. *Technometrics* **35**, 403-410
- [116] Guttorp P., Gneiting T. (2005) *On the Whittle-Matérn correlation family*. Technical Report 80, NRCSE
- [117] Baïle R., Muzy J., Silvani X. (2021) Multifractal point processes and the spatial distribution of wildfires in French Mediterranean regions. *Physica A: Statistical Mechanics and its Applications* **568**, 125697
- [118] Angulo J.M. (2016) *La Teoría Moderna de Medidas de Riesgo en el Contexto de la Complejidad*. Academia de las Ciencias Matemáticas, Físico-Químicas y Naturales de Granada. Godel Impresiones Digitales SL, ISBN 978-84-16478-80-4.
- [119] Veen A., Schoenberg F.P. (2005) Assessing spatial point process models using weighted K-functions: Analysis of California earthquakes. *UC's eScholarship Repository*
- [120] Diggle P.J. (1983) *Statistical Analysis of Spatial Point Patterns*. Mathematics in Biology. Academic-Press, London.

-
- [121] Diggle P.J. (1979) On parameter estimation and goodness-of-fit testing for spatial point patterns. *Biometrics* **35**, 87-101
 - [122] Besag J., Diggle P.J. (1977) Simple Monte Carlo tests for spatial patterns. *Applied Statistics* **26**, 327-333
 - [123] Hope A.C.A. (1968) A simplified Monte Carlo significance test procedure. *Journal of the Royal Statistical Society. Series B (Methodological)* **30**, 582-598
 - [124] Clark P.J., Evans F.C. (1954) Distance to nearest neighbor as a measure of spatial relationships in populations. *Ecology* **35**, 445-453
 - [125] Marriott F.H.C. (1954) Barnard's Monte Carlo tests: how many simulations? *Journal of the Royal Statistical Society. Series C (Applied Statistics)* **28**, 75-77
 - [126] Fisher R.A., Thornton H.G., Mackenzie W.A. (1922) The accuracy of the plating method of estimating the density of bacterial populations, with particular reference to the use of Thornton's agar medium with soil samples *Annals of Applied Botany* **9**, 325-359
 - [127] Diggle P.J., Besag J., Gleaves J.T. (1976) Statistical analysis of spatial point patterns by means of distance methods. *Biometrics* **32**, 659-667

Appendix

This appendix collects a selection of the most relevant **R** codes developed for the elaboration of the thesis.

```
#####
## Simulation of LGCP with varying parameter values *
## Value of fixed parameters: *
## sigma = 1 *
## mu = 0 *
## nu = 1 *
## scale = 1 *
#####

library(spatstat)
spatstat.options(npixel = c(180, 180))

# LGCP with varying sigma = seq (0, 5, 0.25)
var <- rep(seq(0, 0.5, 0.25), each = 100)
sim_var <- lapply(var, FUN = function(x){rLGCP(model = "matern", nu = 1,
                                              mu = 0, scale = 1, var = x,
                                              win = owin(c(0, 10), c(0, 10)),
                                              nsim = 100, saveLambda = TRUE)})

# LGCP with varying mu = seq (0, 4, 0.2)
mu <- rep(seq(0, 4, 0.2), each = 100)
```

```

sim_mu <- lapply(mu, FUN = function(x){rLGCP(model = "matern", nu = 1,
                                             mu = x, scale = 1, var = 1,
                                             win = owin(c(0, 10), c(0, 10)),
                                             nsim = 100, saveLambda = TRUE)})

# LGCP with varying nu = seq (0.25, 2.75, 0.125)
nu <- rep(seq(0.25, 2.75, 0.125), each = 100)
sim_nu <- lapply(nu, FUN = function(x){rLGCP(model = "matern", nu = x,
                                             mu = 0, scale = 1, var = 1,
                                             win = owin(c(0, 10), c(0, 10)),
                                             nsim = 100, saveLambda = TRUE)})

# LGCP with varying scale = seq (1, 11, 0.5)
scale <- rep(seq(1, 11, 0.5), each = 100)
sim_scale <- lapply(scale, FUN = function(x){rLGCP(model = "matern", nu = 1,
                                                    mu = 0, scale = x, var = 1,
                                                    win = owin(c(0, 10), c(0, 10)),
                                                    nsim = 100, saveLambda = TRUE)})

#####
## Simulation of reference processes                                     ##
#####
# Homogeneous Poisson process with intensity lambda = 10
upp <- rpoispp(10, win = owin(c(0, 10), c(0, 10)), nsim = 100)

# Poisson process with white noise intensity
wnp <- rpoispp(exp(rnoise(rnorm, owin(c(0, 10), c(0, 10))))), nsim = 100)

#####
## Preparing data, distributions and quadrat counting                 ##
#####
# Split generated processes on intensity fields and point patterns

```

```

extract_elements <- function(process){
pattern_list <- list()
intensity_list <- list()
for(i in 1:length(process)){
  pattern_list[[i]] <- list()
  intensity_list[[i]] <- attr(process[[i]][[1]], "Lambda")
  for(j in 1:length(process[[i]])){
    pattern_list[[i]][[j]] <- process[[i]][[j]]
    attr(pattern_list[[i]][[j]], "Lambda") <- NULL
  }}
  return(list(patterns = pattern_list, intensities = intensity_list))
}

# Distribution of counts for point patterns and intensity fields
pp_dist <- function(pp, numq = 10){
  qc <- lapply(pattern_list, function(x) lapply(x, function(y){
    as.vector(t(quadratcount(y, numq))))))
  dist_pattern <- lapply(qc, function(x) lapply(x, function(y){y/sum(y)}))
  return(dist_pattern)
}

rf_dist <- function(lambda){
  rast <- lapply(intensity_list, function(x){as.vector(aggregate(raster(x),
    fact = 18, FUN = mean))})
  dist_intensity <- lapply(rast, function(x){x/sum(x)})
  return(dist_intensity)
}

#####
##                                DEFINITION OF INFORMATION AND COMPLEXITY MEASURES                                ##
#####

#####

```

```

** SHANNON ENTROPY                                     *#
#####
Shannon.E <- function(x){
  aux <- vector()
  for(i in 1:length(x)){
    ifelse(x[[i]]==0, aux[[i]]<-0, aux[[i]]<-(x[[i]]*log(x[[i]])))
  }
  entropy <- -sum(aux)
  return(entropy)
}

# Maximum entropy
max.entropy <- function(x){
  max.entropy <- log(length(x))
  return(max.entropy)
}

# Normalized Shannon entropy
Shannon.norm <- function(x){
  ifelse(length(x)==1, norm.entropy<-0,
          norm.entropy<-Shannon.E(x)/max.entropy(x))
  return(norm.entropy)
}

#####
** RENYI ENTROPY                                       *#
#####
Renyi.E <- function(x,order){
  aux <- vector()
  for(i in 1:length(x)){
    ifelse(x[[i]]==0, aux[[i]]<-0, aux[[i]]<-x[[i]]^(order))
  }
  entropy<-ifelse(order==1, Shannon.E(x), (1/(1-order)*log(sum(aux))))
}

```

```

    return(entropy)
}

Renyi.norm <- function(x,order){
  ifelse(length(x)==1, norm.entropy<-0, ifelse(order==1,
    norm.entropy <-Shannon.E(x)/max.entropy(x),
    norm.entropy<-Renyi.E(x,order)/max.entropy(x)))
  return(norm.entropy)
}

#####
## TSALLIS ENTROPY                                     *#
#####

Tsallis.E <- function(x,order){
  aux <- vector()
  for(i in 1:length(x)){
    ifelse(x[[i]]==0, aux[[i]]<-0, aux[[i]]<-x[[i]]^order)
  }
  entropy <- ifelse(order==1, Shannon.E(x), (1/(order-1))*(1-(sum(aux))))
  return(entropy)
}

# Tsallis maximum entropy
Tsallis.max <- function(x,order){
  ifelse(order==1,max.ent<-max.entropy(x),
    max.ent<-(1-length(x)^(1-order))/(order-1))
  return(max.ent)
}

# Normalized Tsallis entropy
Tsallis.norm <- function(x,order){
  ifelse(length(x)==1,norm.entropy<-0,
    ifelse(order==1,norm.entropy<-Shannon.E(x)/max.entropy(x),

```

```

        norm.entropy<-Tsallis.E(x,order)/Tsallis.max(x,order)))
    return(norm.entropy)
}

#####
** DIVERGENCE MEASURES                                     *#
#####

# Kullback-Leibler divergence
KL.divergence <- function(x,y){
  aux<-vector()
  for(i in 1:length(x)){
    ifelse(x[[i]]==0|y[[i]]==0, aux[[i]]<-0,
           aux[[i]]<-x[[i]]*log(x[[i]]/y[[i]]))
  }
  kl.div<-sum(aux)
  ifelse(length(x)==length(y), return(kl.div),
         return(c("Error. Both probability densities must have the same length")))
}

# Renyi divergence
R.divergence <- function(x,y,order){
  aux<-vector()
  for(i in 1:length(x)){
    ifelse(x[[i]]==0|y[[i]]==0, aux[[i]]<-0,
           ifelse(order==1, aux[[i]]<-x[[i]]*log(x[[i]]/y[[i]]),
                  aux[[i]]<-(x[[i]]^order)/(y[[i]]^(order-1))))
  }
  ifelse(order==1, r.div<-sum(aux), r.div<-(1/(order-1))*log(sum(aux)))
  ifelse(length(x)==length(y), return(r.div),
         return(c("Error. Both probability densities must have the same length")))
}

```

```

# Tsallis divergence
T.divergence <- function(x,y,order){
  aux <- vector()
  for(i in 1:length(x)){
    ifelse(x[[i]]==0|y[[i]]==0, aux[[i]]<-0,
           ifelse(order==1, aux[[i]]<-x[[i]]*log(x[[i]]/y[[i]]),
           aux[[i]]<-(x[[i]]^order)/(y[[i]]^(order-1))))
  }
  ifelse(order==1, t.div<-sum(aux), t.div<-(1/(1-order))*(1-sum(aux)))
  ifelse(length(x)==length(y), return(t.div),
  return(c("Error. Both probability densities must have the same length")))
}

#####
## COMPLEXITY MEASURES                                     *#
#####
# Exponential LMC complexity
exp.LMC <- function(x){
  aux<-vector()
  for(i in 1:length(x)){
    aux[[i]]<-(x[[i]]-(1/length(x)))^2
  }
  diseq<-sum(aux)
  complex<-exp(Shannon.E(x))*diseq
  return(c(exp.LMC=complex,exp.Entropy=exp(Shannon.E(x)),diseq=diseq))
}

# Generalized complexity
gen.complex<-function(x,alpha,beta){
  complexity<-exp(Renyi.E(x,alpha)-Renyi.E(x,beta))
  return(complexity)
}

```

```

# Relative complexity
rel.complex <- function(x,y,alpha,beta){
  return(rel.complex=exp(R.divergence(x, y, alpha)- R.divergence(x, y, beta))
}

#####
** Code for inter-intra analysis of LGCP                                **
#####

# Shannon entropy, means and standard deviations
HS_ent <- function(dist, type){
  mean <- vector(); sd <- vector();
  inf <- vector(); sup <- vector()
  if(type == "pattern"){
    entropia <- list()
    for(i in 1:length(dist)){
      entropia[[i]] <- vector()
      for(j in 1:length(dist[[i]])){
        entropia[[i]][j] <- Shannon.E(dist[[i]][[j]])
      }
      mean[i] <- mean(entropia[[i]])
    }
    sd <- apply(matrix(mean, nrow = 100, ncol = 21), 2, FUN = sd)
    mean <- colSums(matrix(mean/100, nrow = 100, ncol = 21))
  }else if(type == "field"){
    entropia <- vector()
    for(i in 1:length(dist)){
      entropia[i] <- Shannon.E(dist[[i]])
    }
    mean <- apply(matrix(entropia, nrow = 100, ncol = 21), 2, FUN = mean)
  }
}

```

```

sd <- apply(matrix(entropia, nrow = 100, ncol = 21), 2, FUN = sd)
}

for(i in 1:length(mean)){
  inf[i] <- mean[i] - 1.96 * (sd[i]/sqrt(100))
  sup[i] <- mean[i] + 1.96 * (sd[i]/sqrt(100))
}
return(list("entropy" = mean, "inf" = inf, "sup" = sup))
}

# LMC complexity, means and standard deviations
LMC <- function(dist, type){
  mean <- vector(); sd <- vector();
  inf <- vector(); sup <- vector()
  if(type == "pattern"){
    complex <- list()
    for(i in 1:length(dist)){
      complex[[i]] <- vector()
      for(j in 1:length(dist[[i]])){
        complex[[i]][j] <- Shannon.E(dist[[i]][[j]])*(sum(dist[[i]][[j]]^2)
          -1/length(dist[[i]][[j]]))
      }
    }
    mean[i] <- mean(complex[[i]])
  }
  sd <- apply(matrix(mean, nrow = 100, ncol = 21), 2, FUN = sd)
  mean <- colSums(matrix(mean/100, nrow = 100, ncol = 21))
} else if(type == "field"){
  complex <- vector()
  for(i in 1:length(dist)){
    complex[i] <- Shannon.E(dist[[i]])*(sum(dist[[i]]^2)
      -1/length(dist[[i]]))
  }
  mean <- apply(matrix(complex, nrow = 100, ncol = 21), 2, FUN = mean)
  sd <- apply(matrix(complex, nrow = 100, ncol = 21), 2, FUN = sd)
}

```

```

    }
    for(i in 1:length(mean)){
      inf[i] <- mean[i] - 1.96 * (sd[i]/sqrt(100))
      sup[i] <- mean[i] + 1.96 * (sd[i]/sqrt(100))
    }
    return(list("complexity" = mean, "inf" = inf, "sup" = sup))
  }

# Kullback-Leibler divergence, means and standard deviations
KL <- function(dist1, dist2){
  divergence <- list()
  mean <- vector(); sd <- vector()
  inf <- vector(); sup <- vector()
  for(i in 1:length(dist2)){
    divergence[[i]] <- vector()
    for(j in 1:length(dist1[[i]])){
      divergence[[i]][[j]] <- KL.divergence(dist1[[i]][[j]], dist2[[i]])
    }
    mean[i] <- mean(divergence[[i]])
  }
  sd <- apply(matrix(mean, nrow = 100, ncol = 21), 2, FUN = sd)
  mean <- colSums(matrix(mean/100, nrow = 100, ncol = 21))
  for(i in 1:length(mean)){
    inf[i] <- mean[i] - 1.96 * (sd[i]/sqrt(100))
    sup[i] <- mean[i] + 1.96 * (sd[i]/sqrt(100))
  }
  return(list("divergence" = mean, "inf" = inf, "sup" = sup))
}

#####
## Preparing Nepal forest fires dataset                                     *#
#####

```

```
library(rgeos)
library(magrittr)
library(dplyr)

# Open boundary file
boundary <- readShapeSpatial("BBDD/Data/Data/Boundary/NationalBoundary")
y <- as(boundary, "SpatialPolygons")
bwin <- as.owin(y)

# Open fire data
fire <- readShapeSpatial("BBDD/Data/Data/ForestFire/ForestFire_Nepal")
fire$days <- as.Date(fire$ACQ_DATE, format = "%Y-%m-%d")
fire_pp <- as(fire, "ppp")
fire_pp$window <- bwin

# Define train and test sets
train_fire <- subset(fire_pp, fire_pp$marks$days >= "2012-02-01"
                     & fire_pp$marks$days < "2016-04-21")
df <- data.frame(Date = as.Date(train_fire$marks$days, "%Y/%m/%d"))
train_fire$marks$tempo <- c(1, 1 + cumsum(as.vector(diff(df$Date))))
max_train <- max(train_fire$marks$tempo)

x <- train_fire$marks$LONGITUDE
y <- train_fire$marks$LATITUDE
t <- train_fire$marks$tempo
trainset <- data.frame(x, y, t)
trainset <- as.3dpoints(trainset %>% distinct())

test_fire <- subset(fire_pp, fire_pp$marks$days >= "2016-04-21"
                   & fire_pp$marks$days < "2016-06-19")
df <- data.frame(Date = as.Date(test_fire$marks$days, "%Y/%m/%d"))
test_fire$marks$tempo <- c(max_train + 1,
                           max_train + 1 + cumsum(as.vector(diff(df$Date))))
```

```

x <- test_fire$marks$LONGITUDE
y <- test_fire$marks$LATITUDE
t <- test_fire$marks$tempo
testset <- data.frame(x, y, t)
testset <- as.3dpoints(testset %>% distinct())

#####
## Preparing covariates                                     *#
#####

climate_p1 <- read.csv(file = "BBDD/Climate_data/climate_pi.csv")
climate_p2 <- read.csv(file = "BBDD/Climate_data/climate_pii.csv")
climate <- as.data.frame(rbind(climate_p1, climate_p2))
climate$date <- mdy(climate[, 1])

# Train and test sets
train_clim <- subset(climate, climate$date >= "2012-02-01"
                    & climate$date < "2016-04-21")
test_clim <- subset(climate, climate$date >= "2016-04-21"
                   & climate$date < "2016-06-19")
df <- data.frame(Date = as.Date(test_clim$date, "%Y/%m/%d"))
test_clim$tempo <- c(max_train+1, max_train+1+
                    cumsum(as.vector(diff(df$Date))))

# Polygonal districts of Nepal
regions <- readShapeSpatial("BBDD/Data/Data/Regions/local_unit")

# Aggregate climate covariates by district
wind <- aggregate(train_clim$ws10m, by = list(train_clim$district),
                  FUN = median)
prectot <- aggregate(train_clim$prectot, by = list(train_clim$district),
                    FUN = median)$x

```

```

wb_temp <- aggregate(train_clim$t2mwet, by = list(train_clim$district),
                     FUN = median)$x
clim_df <- data.frame(Group.1 = wind[, 1], wind = wind[, 2], prectot,
                     wb_temp)

regions$wind <- rep(NA, nrow(regions))
regions$prectot <- rep(NA, nrow(regions))
regions$wb_temp <- rep(NA, nrow(regions))
for(i in 1:nrow(regions)){
regions$wind <- as.numeric(replace(regions$wind,
                                regions$DISTRICT == clim_df[i, 1], clim_df[i, 2]))
regions$prectot <- as.numeric(replace(regions$prectot,
                                regions$DISTRICT == clim_df[i, 1], clim_df[i, 3]))
regions$wb_temp <- as.numeric(replace(regions$wb_temp,
                                regions$DISTRICT == clim_df[i, 1], clim_df[i, 4]))
}

poly_district <- gUnaryUnion(regions, id = regions$DISTRICT)

# Transform polygons to tess object
list_owin <- lapply(poly_district, FUN = as.owin)
tess_district <- tess(tiles = list_owin)

# Plot covariate tessellations
plot(tess_district, do.col = TRUE, values = as.numeric(clim_df[, 2]),
     main = "")
plot(tess_district, do.col = TRUE, values = as.numeric(clim_df[, 3]),
     main = "")
plot(tess_district, do.col = TRUE, values = as.numeric(clim_df[, 4]),
     main = "")

# Covariate functions
function_wind <- as.function.tess(tess_district,

```

```

                                values = as.numeric(clim_df[, 2]))
funcion_prectot <- as.function.tess(tess_district,
                                values = as.numeric(clim_df[, 3]))
funcion_wb_temp <- as.function.tess(tess_district,
                                values = as.numeric(clim_df[, 4]))

#####
## DETERMINISTIC INTENSITY                                     *#
#####
## Temporal component                                         *#
#####
## Daily number of forest fires
t <- train_fire$marks$days
days <- data.frame(x, y , t)
days <- days %>% distinct()

AA <- days$t
BB <- seq(range(AA)[1], range(AA)[2], by = "day")
CC <- c(AA, BB)
A1 <- data.frame(table(CC))
counts1 <- data.frame(table(strftime(CC)))
counts2 <- data.frame(table(strftime(days$t)))

counts <- merge.data.frame(counts1, counts2, by = "Var1", all.x = T)
counts$Freq.y[is.na(counts$Freq.y)] <- 0
zm <- ts(counts$Freq.y, frequency = 365, start = c(2012,5))

## Model covariates
# Aggregate climate covariates by date
wind <- aggregate(train_clim$ws10m, by = list(strftime(train_clim$date)),
                  FUN = median)$x
prectot <- aggregate(train_clim$prectot, by = list(strftime(train_clim$date)),

```

```

      FUN = median)$x
wb_temp <- aggregate(train_clim$t2mwet, by = list(strftime(train_clim$date)),
      FUN = median)$x

omegat <- 1:1541
month <- lubridate::month(unique(train_clim$date))
month <- factor(month, levels = 1:12, labels = month.name)

df_covariates <- data.frame(zm, prectot, wb_temp, wind)

# Generalized lineal model
Mu <- glm(zm ~ month + wind + wb_temp + prectot +
      sin(2 * pi * omegat / 365) + sin(4 * pi * omegat / 365) +
      cos(2 * pi * omegat / 365) + cos(4 * pi * omegat / 365),
      family = "poisson")

#####
## Spatial component - Separable                                     *#
#####
planar_trainset <- unique(as.ppp(as.data.frame(cbind(trainset[, 1],
                                                    trainset[, 2]))))

A <- ppm(planar_trainset ~ wind + temp + prectot,
      covariates = list(wind = function_wind,
                        temp = function_wb_temp,
                        prectot = function_prectot),
      interaction = Poisson(), use.gam = TRUE)

# Separable deterministic intensity with spatio-temporal weights for 16 days
# Preparing covariates for testset
clim_subset <- subset(test_clim, test_clim$tempo > 1540 &
      test_clim$tempo <= 1556)
wind <- aggregate(clim_subset$ws10m, by = list(test_clim$district),

```

```

      FUN = median)
prectot <- aggregate(test_clim$prectot, by = list(test_clim$district),
      FUN = median)$x
wb_temp <- aggregate(test_clim$t2mwet, by = list(test_clim$district),
      FUN = median)$x
clim_df <- data.frame(Group.1 = wind[, 1], wind = wind[, 2], prectot,
      wb_temp)

regions$wind <- rep(NA, nrow(regions))
regions$prectot <- rep(NA, nrow(regions))
regions$wb_temp <- rep(NA, nrow(regions))
for(j in 1:nrow(regions)){
  regions$wind <- as.numeric(replace(regions$wind,
      regions$DISTRICT == clim_df[j, 1],
      clim_df[j, 2]))
  regions$prectot <- as.numeric(replace(regions$prectot,
      regions$DISTRICT == clim_df[j, 1],
      clim_df[j, 3]))
  regions$wb_temp <- as.numeric(replace(regions$wb_temp,
      regions$DISTRICT == clim_df[j, 1],
      clim_df[j, 4]))
}

poly_district <- gUnaryUnion(regions, id = regions$DISTRICT)
list_owin <- lapply(poly_district, FUN = as.owin)
tess_district <- tess(tiles = list_owin)

function_wind <- as.function.tess(tess_district,
      values = as.numeric(clim_df[, 2]))
function_prectot <- as.function.tess(tess_district,
      values = as.numeric(clim_df[, 3]))
function_wb_temp <- as.function.tess(tess_district,
      values = as.numeric(clim_df[, 4]))

```

```

# Predict intensity function
lambda_sep <- predict.ppm(A, covariates = list(wind = function_wind,
                                                temp = function_wb_temp,
                                                prectot = function_prectot))

#####
## Spatial component - Non-separable                                     *#
#####

library(optimx)
# Weight function
my_weight <- function(lag, p1, p2, p3, p4) {
  p1 ^ lags + p2 ^ lags*(p3 ^ ((2*sin(pi * lags / 365)) **2))*
    (p4 ^ ((sin(pi * lags / 730)) **2))
}

# Estimate MSE (i.e. squared error between positive part of ACF and
# weight function)
my_MSE <- function(acf, lags, par) {
  mean((acf-par[1]*my_weight(lags, par[2], par[3], par[4], par[5]))^2)
}

# Estimate positive part of ACF
my_acf <- acf(zm, plot = FALSE, lag.max = 1541)
my_positive_acf <- pmax(0, my_acf$acf[, , 1][-1])

min_value <- vector()
try(my_rhos <- optimx(
  par = rep(0.5, 5),
  fn = my_MSE,
  method = "nlm",

```

```

    lower = rep(0, 5),
    upper = rep(1, 5),
    acf = my_positive_acf,
    lags = seq_len(1079)
  )
)

# Non-separable deterministic intensity with spatio-temporal weights
# for 16 future time periods (lambda_1)
fire_data <- superimpose(train_fire, test_fire)
marks(fire_data) <- fire_data$marks$tempo

lambda_1 <- list()
for(i in 1:16){
  fire_subset <- subset(fire_data, fire_data$marks <= i+1540 &
                        fire_data$marks >= i+1540-est_lag)
  weight_st <- my_weight(abs(fire_subset$marks-(i+1541)), my_rhos$p2,
                        my_rhos$p3, my_rhos$p4, my_rhos$p5)
  lambda_1[[i]] <- density(unmark(fire_subset), weights = weight_st,
                        edge = TRUE, at = "pixels", leaveoneout = TRUE,
                        adjust = 1, diggle = FALSE, se = FALSE,
                        kernel = "gaussian",
                        scalekernel = is.character(kernel),
                        positive = TRUE)
}

lambda_1.norm <- lapply(lambda_1, function(x){x/sum(x, na.rm = TRUE)})

# Non-separable deterministic intensity through glm model
# coefficients (lambda_2)
for(i in 1:16){
  # Preparing covariates for testset
  clim_subset <- subset(test_clim, test_clim$tempo == i+1540)

```



```

# Predict intensity function
lambda_2[[i]] <- predict.ppm(A, covariates = list(wind = function_wind,
                                                  temp = function_wb_temp,
                                                  prectot = function_prectot))

}

lambda_2.norm <- lapply(lambda_2, function(x){x/sum(x, na.rm = TRUE)})

# Non-separable predicted intensity for 16 future time periods
lambda_nonsep <- mapply(function(x, y){sqrt(x*y)}, lambda_1.norm,
                        lambda_2.norm)

#####
## STOCHASTIC COMPONENT                                     *#
#####

library(raster)
library(geostatsp)
library(lgcp)
library(RandomFields)
RFOptions(force=TRUE,strategy=1,spConform=FALSE,cPrintlevel=0,Ttriple=TRUE)

# Function with input a point pattern for redistribute the duplicates whitin
# a circumference of a given radius
split_dup<-function(X,radius=0.04,set.seed=2){
  m <- multiplicity(X)
  first <- !duplicated(X)
  table(multiplicity(X))
  Xbis <- X[first] %mark% m[first]
  marks(Xbis)
  id<-which(marks(Xbis)>1)
  Xter<-Xbis[id]

```

```

    totpun=npoints(Xter)
    for(j in 1:totpun){
      id_rep<-which(X$x==Xter$x[j]&X$y==Xter$y[j])
      Y<-X[id_rep]
      aa<-rjitter(Y, radius=radius, retry=TRUE, giveup = 10000)
      X[id_rep]<-aa
    }
    return(list(X=X,multX=table(multiplicity(X))))
  }

fire_coords <- unmark(train_fire)
# Replacing coordinates in [0, 1]x[0,1] box
window <- owin(xrange = range((fire_pp$x-min(fire_pp$x))/(max(fire_pp$x)
  -min(fire_pp$x))+0.1),
  yrange = range((fire_pp$y-min(fire_pp$y))/(max(fire_pp$y)
  -min(fire_pp$y))+0.1))
fire_coords$x <- (fire_coords$x-min(fire_pp$x))/(max(fire_pp$x)-
  min(fire_pp$x))+0.1
fire_coords$y <- (fire_coords$y-min(fire_pp$y))/(max(fire_pp$y)-
  min(fire_pp$y))+0.1
fire_coords$window <- window

bb <- split_dup(fire_coords, radius = 1)
pun <- bb$X
punst <- as.3dpoints(x = pun$x, y = pun$y,
  t = as.numeric((train_fire$marks$tempo - 1)/(1557)+0.1))
t.region <- range((train_fire$marks$tempo - 1)/(1557)+0.1)
s.region <- rbind(c(range(fire_coords$x)[1], range(fire_coords$y)[1]),
  c(range(fire_coords$x)[2], range(fire_coords$y)[1]),
  c(range(fire_coords$x)[2], range(fire_coords$y)[2]),
  c(range(fire_coords$x)[1], range(fire_coords$y)[2]))

uu <- punst

```

```
sim <- data.frame(x = uu[, 1], y = uu[, 2], t = uu[, 3])
tlim = max(t.region)
max_dist <- max(dist(s.region))
max_dist_t <- max(diff(t.region))
lgrid_s <- 50
rsup_t <- max_dist_t/4
xytimes <- sim$t
dt <- dist(xytimes)
bw_vector_t <- dpik(dt, kernel = "box", range.x = c(min(dt), max(dt)))
rmin_vector_t2 <- (bw_vector_t*((1+lgrid_s)/lgrid_s))
data_2_t <- cbind(rmin_vector_t2, bw_vector_t)

rsup_s <- max_dist/4
pun_s <- ppp(sim$x, sim$y, window = owin(xrange = range(s.region[, 1]),
                                         yrange = range(s.region[, 2])))

xy <- cbind(sim$x, sim$y)
ds <- dist(xy)
bw2_vector <- dpik(ds, kernel = "epanech", range.x = c(min(ds), max(ds)))
rmin3_vector <- bw2_vector*((1+lgrid_s)/lgrid_s)
data_2_s <- cbind(rmin3_vector, bw2_vector)

start_par <- c(log(dim(sim)[1])/2, max_dist/100, max_dist_t/100)
itnmax <- 100
l = 2
vals_s <- data_2_s
vals_t <- data_2_t

u <- seq(vals_s[1], rsup_s, len = lgrid_s)
v <- seq(vals_t[1], rsup_t, len=lgrid_s)
us0 <- vals_s[2]
vt0 <- vals_t[2]
print(c(us0, vt0))
```

```

# Estimator of the pair correlation function for the observed point
# pattern
g0 <- PCFhat(xyt = uu, s.region = s.region, t.region = t.region,
             dist = u, times = v, ks = "epanech", hs = us0, kt = "box",
             ht = vt0)
nonpar_g_st <- g0

# Estimate covariate parameters by minimizing the quadratic distance
# between the empirical and parametric form of the pair correlation
# function
# --- Separable assumption (exp-exp covariance)
g_st_exp_exp <-function (param, useq, vseq,ghat, transform, power){
uv<-expand.grid(useq,vseq)
g.parametric <- exp(param[1]*exp(-uv[,2]/param[3])* exp(-uv[,1]/param[2]))
ghat2<-as.vector(ghat)
return(sum((((ghat2))^power - (g.parametric)^power)^2))
}
min_vals<-c(0.01, 0.001,0.001)
max_vals<-c(70,70,70)
min_vals2<-min_vals; max_vals2<-max_vals
start_par<-c(log(dim(sim)[1])/2,max_dist/100,max_dist_t/100)
MINCON_EXP_EXP<-c(NA,NA,NA)
try({MINCON_EXP_EXP<- optimx(par = start_par,
                             fn = g_st_exp_exp,
                             useq=nonpar_g_st$dist,
                             vseq=nonpar_g_st$times,
                             ghat = nonpar_g_st$pcf,
                             transform = NULL,
                             power = 1,
                             method = "nlm",
                             lower=min_vals2,
                             upper=max_vals2)
})

```

```

res_EXP_EXP <- as.numeric(as.vector(MINCON_EXP_EXP[1:3]))

# Simulation of residual component
sim_stpp<- list()
dgrid2=dgrid=50
M=100
set.seed(20005)

t.region <- range((test_fire$marks$tempo - 1)/(1557)+0.1)
nx=dgrid2
ny=dgrid2
nt=dgrid
xlim<-c(min(s.region[,1]),max(s.region[,1]))
ylim<-c(min(s.region[,2]),max(s.region[,2]))
zlim<-c(min(t.region),max(t.region))

gridwidth<-diff(xlim)/nx
gridwidth2<-diff(ylim)/ny
gridwidth3<-diff(zlim)/nt
x.seq<-seq(xlim[1]+0.5*gridwidth,xlim[2]-0.5*gridwidth,by=gridwidth)
y.seq<-seq(ylim[1]+0.5*gridwidth2,ylim[2]-0.5*gridwidth2,by=gridwidth2)
z.seq<-c(zlim[1], (zlim[2]-zlim[1])/16 , 16)

sigma2 <- res_EXP_EXP[1]
scale_s <- res_EXP_EXP[2]
scale_t <- res_EXP_EXP[3]

mod <- "mo <- RMexp(var=sigma2,scale=scale_s,proj = 'space') *
RMexp(proj = 'time',scale=scale_t) + RMtrend(mean=-0.5*sigma2)"
eval(parse(text=mod))

Z_exp_exp<-RandomFields::RFsimulate(model=mo, x=x.seq, y=y.seq, T=z.seq,
```

```

                                n=500)

# --- Non-separable assumption (iaco covariance)
g_st_iaco <- function(param, useq, vseq, ghat, transform, power){
  uv <- expand.grid(useq, vseq)
  g.parametric <- exp((as.numeric(param[1])*(1+(uv[,1]/as.numeric(param[2]))^
as.numeric(param[4])+(uv[,2]/as.numeric(param[3]))^as.numeric(param[5]))^
(-as.numeric(param[6]))))
  ghat2<-as.vector(ghat)
  return(sum((((ghat2)) ^ power - (g.parametric) ^ power) ^ 2))
}

min_vals<-c(0.01,0.01,0.001,0.01,0.01,1.1)
max_vals<-c(log(dim(sim)[1])/2,Inf,Inf,2,2,15)
min_vals2<-min_vals; max_vals2<-max_vals
start_par<-c(log(dim(sim)[1])/2,max_dist/100,max_dist_t/100,1,1,8)
MINCON_IACO<-c(NA,NA,NA,NA,NA,NA)
try({MINCON_IACO<- optimx(par = start_par,
  fn = g_st_iaco,
  useq=nonpar_g_st$dist,
  vseq=nonpar_g_st$times,
  ghat = nonpar_g_st$pcf,
  transform = NULL,
  power = 1,
  method = "nml",
  lower = min_vals2,
  upper=max_vals2)
})

res_IACO<- as.numeric(as.vector(MINCON_IACO[1:6]))

# Simulation of residual component
param=list(mean=-0.5*res_IACO[1],

```

```

var=res_IACO[1],
scale_s=res_IACO[2],
scale_t=res_IACO[3],
power_s=res_IACO[4],
power_t=res_IACO[5],
power2 = res_IACO[6])

mod <- "mo<- RMIaco(var=as.numeric(param$var),
nu=as.numeric(param$power_s),
lambda=as.numeric(param$power_t),
delta=as.numeric(param$power2),
scale=as.numeric(param$scale_s))+
RMtrend(mean=param$mean)"
eval(parse(text=mod))

Z_IACO<-RandomFields::RFsimulate(model=mo, x=x.seq, y=y.seq, T=z.seq,
                                n=500)

#####
## PREDICTION                                     *#
#####
# Separable assumption
L_sep <- list()
uu_sep <- list()
for(i in 1:16){
  L_sep[[i]] <- list()
  uu_sep[[i]] <- list()
  for(h in 1:500){
    L_sep[[i]][[h]] <- exp(log(lambda_sep[[i]])+Z_EXP_EXP[, , i, h])
    uu_sep[[i]][[h]] <- rpoispp(L_sep[[i]][[h]])
  }
}

```

```

# Mixed assumption
L_mixed <- list()
uu_mixed <- list()
for(i in 1:16){
  L_mixed[[i]] <- list()
  uu_mixed[[i]] <- list()
  for(h in 1:500){
    L_mixed[[i]][[h]] <- exp(log(lambda_sep[[i]])+Z_IACO[, ,i,h])
    uu_mixed[[i]][[h]] <- rpoispp(L_mixed[[i]][[h]])
  }
}

# Non-separable assumption
L_nonsep <- list()
uu_nonsep <- list()
for(i in 1:16){
  L_nonsep[[i]] <- list()
  uu_nonsep[[i]] <- list()
  for(h in 1:500){
    L_nonsep[[i]][[h]] <- exp(log(lambda_nonsep[[i]])+Z_IACO[, ,i,h])
    uu_nonsep[[i]][[h]] <- rpoispp(L_nonsep[[i]][[h]])
  }
}

#*****
#* GOODNESS OF FIT ANALYSIS                                     *#
#*****

x <- test_fire$marks$LONGITUDE
y <- test_fire$marks$LATITUDE
t <- test_fire$marks$tempo-max(train_fire$marks$tempo)
testset <- data.frame(x, y, t)
testset <- as.3dpoints(testset %>% distinct())

```

```

x <- testset[, 1]
y <- testset[, 2]
t <- testset[, 3]

# Extract values from intensities
extract.int <- function(x, y, t, L, ntimes, nsims){
  df_int <- lapply(1:nsims, \(i) data.table::copy(data.frame(x, y, t,
                                                            int = NA)))

  L_raster <- lapply(L, function(z){lapply(z, raster)})
  int <- list()
  for(i in 1:nsims){
    int[[i]] <- 0
    for(j in 1:ntimes){
      aux <- subset(df_int[[i]], select = x:y, df_int[[i]]$t == j)
      int[[i]] <- c(int[[i]], raster::extract(L_raster[[j]][[i]], aux))
    }
    df_int[[i]]$int <- int[[i]][-1]
    df_int[[i]] <- subset(df_int[[i]], !is.na(df_int[[i]]$int))
  }
  return(df = df_int)
}

extract_sep <- extract.int(x, y, t, L_sep, 16, 500)
extract_mixed <- extract.int(x, y, t, L_mixed, 16, 500)
extract_nonsep <- extract.int(x, y, t, L_nonsep, 16, 500)
pp <- as.3dpoints(x, y, t)
class(pp) <- "stpp"

# Global weighted K-function
diag_global <- function(pattern, intensity, u = seq(0, 0.25, length = 41),
  v = seq(0, 4, length = 41), correction_g = "isotropic"){
  hat.norm <- list() ;x2_global <- list();
  diff <- list(); diff_surface <- list()

```

```

    for(i in 1:length(intensity)){
      global_k <- STIKhat(xyt = pp, lambda = intensity[[i]]$int,
                        dist = u, times = v,
                        correction = correction_g)
      normalize <- function(x){(x-min(x))/((max(x)-min(x))+0.0001)}
      theo.norm <- normalize(global_k$Ktheo)
      hat.norm[[i]] <- normalize(global_k$Khat)
      chi_global[[i]] <- sum((theo.norm-hat.norm[[i]])^2/
                          (hat.norm[[i]] + 0.5))
      diff[[i]] <- sum(abs(theo.norm - hat.norm[[i]]))
      diff_surface[[i]] <- abs((theo.norm - hat.norm[[i]]))
    }
  return(list(surface = Reduce('+', diff_surface)/length(diff_surface),
             chi = unlist(chi_global)))
}

diag_g_sep <- diag_global(pp, extract_sep,)
diag_g_mixed <- diag_global(pp, extract_mixed)
diag_g_nonsep <- diag_global(pp, extract_nonsep)

# Local weighted K-function -- Compute a list of statistics for each
# simulation
diag_local <- function(pattern, intensity, u = seq(0, 0.25, length = 41),
                      v = seq(0, 4, length = 41), correction_g = "isotropic",
                      infectious = FALSE, p = 0.95){
  nn <- nrow(pattern)
  out <- list()
  for(i in 1:100){
    print(i)
    local_k_1 <- KLISTAhat(xyt = pattern,
                        lambda = intensity[[i]], dist = u, times = v,
                        correction = correction_g)
    theo_local_k1 <- local_k_1$klistattheo
  }
}

```

```

        diff_1_local <- array(NA, c(length(u) - 1, length(v) - 1,
                                   nn))
        x2_1 <- vector(l = nn)
        for(j in 1:nn){
            diff_1_local[, , j] <- local_k_1$list.KLISTA[, , j]
            -theo_local_k1
            x2_1[j] <- sum(diff_1_local[, , j]^2/
                          (theo_local_k1 + 0.1))
        }
        normalize <- function(x){(x-min(x))/((max(x)-min(x))+0.0001)}
        x21_1 <- normalize(x2_1)
        minx <- min(x21_1)
        maxx <- max(x21_1)
        breaksx <- seq(minx, maxx, length = 40)
        hist(x21_1, breaks = breaksx,
             main = latex2exp::TeX(r"(Obj = $\chi_i^2$)"), xlab = " ") #
        q <- quantile(x21_1, p)
        abline(v = q, col = 2)
        id_points <- which(x21_1 > quantile(x21_1, p = p))
        out[[i]] <- list(pattern = pattern,
                        listas = local_k_1$list.KLISTA,
                        ids = id_points, chi = q)
    }
    return(out)
}

diag_l_sep <- diag_local(pp, extract_sep,)
diag_l_mixed <- diag_local(pp, extract_mixed)
diag_l_nonsep <- diag_local(pp, extract_nonsep)

#####
** RISK MAPPING                                     **#

```

```

#####
# Sliding window -- Windows of size 0.5 overlap with a step of 0.1.
# Beginning in bottom left (80.05858, 88.55858)x(26.3478, 26.84784),
# we move upwards along y-axis, when the maximum is reached, we move
# 0.5 units along the x-axis.

cont_i <- seq(bwin$ xrange[1], bwin$ xrange[2], 0.1)
cont_j <- seq(bwin$ yrange[1], bwin$ yrange[2], 0.1)
win <- list()
for(i in 1:(length(cont_i)-5)){
  win[[i]] <- list()
  for(j in 1:(length(cont_j)-5)){
    win[[i]][[j]] <- owin(xrange = c(cont_i[i], cont_i[i+4]),
                           yrange = c(cont_j[j], cont_j[j+4]))
  }
}

win <- unlist(win, recursive = FALSE)

# Quadrat counting using sliding window for each time unit
qq_sw <- function(pp, win_list){
  cont_sw <- list()
  for(h in 1:length(uu)){
    cont_sw[[h]] <- matrix(NA, nrow = length(uu[[i]]), ncol = length(win))
    for(i in 1:length(uu[[i]])){
      for(j in 1:length(win)){
        cont_sw[[h]][i, j] <- uu[[h]][[i]][win[[j]]]$n
      }
    }
  }
  return(cont_sw)
}

```

```

cont_sep <- qq_sw(uu_sep, win)
cont_mixed <- qq_sw(uu_mixed, win)
cont_nonsep <- qq_sw(uu_nonsep, win)

# Computing VaR for each window and creating risk maps
VaR_maps <- function(cont_sw, alpha){
  VaR[[h]] <- vector()
  for(i in 1:ncol(cont_sw[[h]])){
    VaR[[h]][[i]] <- quantile(cont_sw[[h]][, i], alpha)
  }
  VaR_matrix[[h]] <- matrix(VaR[[h]], ncol = sqrt(VaR[[h]]))
  VaR_map[[h]] <- as(VaR_matrix[[h]][nrow(VaR_matrix[[h]]):1, ], "RasterLayer")
  extent(VaR_map[[h]]) <- c(bwin$xrange[1], bwin$xrange[2], bwin$yrange[1],
                           bwin$yrange[2])

  return(VaR_map)
}

#####
## Simulation study of information and complexity measure as a      ##
## goodness of fit tool                                           ##
#####
# Compute power and significance for 1000 simulations             ##
#####
par_grid <- data.frame(var = rep(c(1, 2.5, 5), each = 3),
                      scale = rep(c(0.05, 0.1, 0.25), times = 3))

qc_dim <- 10 # Change to 20, 30, 40 for each quadrat size
max_entropy <- log(10^2, 20^2, 30^2, 50^2)

# Simulate real intensity with 1000 realizations each
s.region <- rbind(c(0,0), c(1,0), c(1,1), c(0,1))
xlim <- ylim <- c(0, 1)

```

```

ngrid <- 500
gridwidth <- diff(xlim)/ngrid
x.seq <- y.seq <- seq(xlim[1]+0.5*gridwidth, xlim[2]-0.5*gridwidth,
                      by = gridwidth)

Z <- list()
pat_sim <- list()
for(i in 1:nrow(par_grid)){
  Z[[i]] <- RandomFields::RFsimulate(model = RMexp(var = par_grid[i, 1],
                                                  scale = par_grid[i, 2]),
                                     x = x.seq, y = y.seq)
  pat_sim[[i]] <- rpoispp(as.im(exp(log(1000/mean(exp(Z[[i]])))+Z[[i]]),
                           W = owin(xlim, y lim)), nsim = 1000)
}

#####
# Estimate intensity                                     *#
#####
# Case 1: Kernel density sigma
sim_dens <- list()
for(i in 1:nrow(par_grid)){
  dens <- lapply(pat_sim[[i]], FUN = function(x) density(x, bw.diggle(x),
    kernel = "epanechnikov",
    diggle = TRUE,
    scalekernel = TRUE,
    edge = TRUE,
    dimyx = c(500, 500),
    positive = TRUE))
  sim_dens[[i]] <- lapply(dens, FUN = rpoispp)
}

# Case 2: Kernel density 2*sigma
sim_dens2 <- list()

```

```

for(i in 1:nrow(par_grid)){
  dens <- lapply(pat_sim[[i]], FUN = function(x) density(x, 2*bw.diggle(x),
    kernel = "epanechnikov",
    diggle = TRUE,
    scalekernel = TRUE,
    edge = TRUE,
    dimyx = c(500, 500),
    positive = TRUE))
  sim_dens2[[i]] <- lapply(dens, FUN = rpoispp)
}

# Case 3: Kernel density 3*sigma
sim_dens3 <- list()
for(i in 1:nrow(par_grid)){
  dens <- lapply(pat_sim[[i]], FUN = function(x) density(x, 3*bw.diggle(x),
    kernel = "epanechnikov",
    diggle = TRUE,
    scalekernel = TRUE,
    edge = TRUE,
    dimyx = c(500, 500),
    positive = TRUE))
  sim_dens3[[i]] <- lapply(dens, FUN = rpoispp)
}

# Case 4: Constant intensity
sim_cons <- list()
for(i in 1:nrow(par_grid)){
  dens <- lapply(pat_sim[[i]], npoints)
  sim_cons[[i]] <- lapply(dens, FUN = rpoispp)
}

```

```

#*****

```

```

## Perform quadrat counting, compute measure and confidence bands      ##
#####
qc <- list()
prob <- list()
shannon <- list()
divergence <- list()
LMC <- list()
for(i in 1:nrow(par_grid)){
  qc[[i]] <- lapply(pat_sim[[i]], nx = qc_dim, FUN = quadratcount)
  prob[[i]] <- lapply(qc[[i]], function(x) x/sum(x))
  shannon[[i]] <- unlist(lapply(prob[[i]], FUN = Shannon.E), recursive = FALSE)
  divergence[[i]] <- unlist(mapply(expand.grid(prob[[i]], prob[[i]])[[1]],
                                   expand.grid(prob[[i]], prob[[i]])[[2]],
                                   FUN = KL.divergence))
  LMC[[i]] <- unlist(lapply(prob[[i]], FUN = exp.LMC), recursive = FALSE)
}

shannon_mean <- unlist(lapply(shannon, FUN = mean))
shannon_upper <- unlist(lapply(shannon, FUN = quantile, prob = 0.95))
shannon_lower <- unlist(lapply(shannon, FUN = quantile, prob = 0.05))
divergence_mean <- unlist(lapply(divergence, FUN = mean))
divergence_upper <- unlist(lapply(divergence, FUN = quantile, prob = 0.95))
divergence_lower <- unlist(lapply(divergence, FUN = quantile, prob = 0.05))
LMC_mean <- unlist(lapply(LMC, FUN = mean))
LMC_upper <- unlist(lapply(LMC, FUN = quantile, prob = 0.95))
LMC_lower <- unlist(lapply(LMC, FUN = quantile, prob = 0.05))

#####
## Compute significance values      ##
#####
x <- shannon_dens # shannon_dens2, shannon_dens3, shannon_cons
ind <- mapply(x = x, y = shannon, z = shannon_lower,
t = shannon_upper,

```

```

function(x, y, z, t){ifelse(abs(x - y) > t-z, 1, 0)}}
sig_shdens <- colSums(ind)/nrow(ind)

x <- LMC_dens # LMC_dens2, LMC_dens3, LMC_cons
ind <- mapply(x = LMC_dens, y = LMC, z = LMC_lower,
t = LMC_upper,
function(x, y, z, t){ifelse(abs(x - y) > t-z, 1, 0)}})
sig_LMCdens <- colSums(ind)/nrow(ind)

x <- divergence_dens # divergence_dens2, divergence_dens3, divergence_cons
ind <- mapply(x = divergence_dens, y = divergence_upper,
function(x, y){ifelse(x > y, 1, 0)}})
sig_divdens <- colSums(ind)/nrow(ind)

#####
## Goodness-of-fit test                                     *#
#####
# Power - CSR null hypothesis vs. inhomogeneous Poisson process
malla <- data.frame(var = rep(c(1, 2.5, 5), each = 3),
scale = rep(c(0.05, 0.1, 0.25), times = 3))
qc_dim <- seq(2,30,1) # quadrat size
nsim <- 100

s.region <- rbind(c(0, 0), c(1, 0), c(1, 1), c(0, 1))
xlim <- ylim <- c(0, 1)
ngrid <- 500
gridwidth <- diff(xlim)/ngrid
x.seq <- y.seq <- seq(xlim[1]+0.5*gridwidth,xlim[2]-0.5*gridwidth,by=gridwidth)

sig <- list()
for(k in 1:nrow(malla)){
sig[[k]] <- vector()
for(h in 1:50){

```

```

Z <- RandomFields::RFsimulate(model = RMexp(var = malla[k, 1],
scale = malla[k, 2]),
x = x.seq, y = y.seq)
obs_pattern <- rpoispp(as.im(exp(log(1000/mean(exp(Z))) + Z),
W = owin(xlim, ylim)))
qc <- list()
for(i in 1:length(qc_dim)){
qc[[i]] <- quadratcount(obs_pattern, nx = qc_dim[i])
}

prob <- lapply(qc, function(x) x/sum(x))
hat_H1 <- unlist(lapply(prob, FUN = Shannon.E_norm), recursive = FALSE)

sim_pat <- rpoispp(obs_pattern$n, W = owin(xlim, ylim), nsim = 100)
qc_sim <- list()
prob_sim <- list()
hat_Hi <- list()
for(i in 1:length(qc_dim)){
qc_sim <- lapply(sim_pat, nx = qc_dim[[i]], FUN = quadratcount)
prob_sim <- lapply(qc_sim, function(x) x/sum(x))
hat_Hi[[i]] <- unlist(lapply(prob_sim, FUN = Shannon.E_norm), recursive = FALSE)
}

U_env <- unlist(lapply(hat_Hi, FUN = max))
L_env <- unlist(lapply(hat_Hi, FUN = min))

bar_H1 <- unlist(lapply(hat_Hi, FUN = function(x){(1/nsim)*sum(x)}))
bar_Hi <- list()
hat_Hi_t <- list()
for(i in 1:nsim){
bar_Hi[[i]] <- vector()
hat_Hi_t[[i]] <- vector()
for(j in 1:length(qc_dim)){

```

```

bar_Hi[[i]][j] <- (1/(nsim-1))*sum(hat_Hi[[j]][-i])
hat_Hi_t[[i]][j] <- hat_Hi[[j]][i]
}
}

ui <- mapply(function(x,y){max(abs(x-y))}, x = hat_Hi_t, y = bar_Hi)
u1 <- max(abs(hat_H1 - bar_H1))
ind <- lapply(ui, FUN = function(x){ifelse(x > u1, 1, 0)})
sig[[k]][h] <- sum(unlist(ind))/length(ind)
}
}

# Power - CSR null hypothesis vs. simple sequential inhibition process
malla <- c(0.0125, 0.0250, 0.0375, 0.0500, 0.0625, 0.0750)

qc_dim <- seq(2,30,1) # quadrat size
nsim <- 100

s.region <- rbind(c(0, 0), c(1, 0), c(1, 1), c(0, 1))
xlim <- ylim <- c(0, 1)
ngrid <- 500
gridwidth <- diff(xlim)/ngrid
x.seq <- y.seq <- seq(xlim[1]+0.5*gridwidth,xlim[2]-0.5*gridwidth,by=gridwidth)

sig <- list()
for(k in 1:length(malla)){
  sig[[k]] <- vector()
  for(h in 1:50){
    obs_pattern <- rSSI(r = malla[k], n = 1000, win = owin(xlim, ylim))
    qc <- list()
    for(i in 1:length(qc_dim)){
      qc[[i]] <- quadratcount(obs_pattern, nx = qc_dim[i])
    }
  }
}

```

```

prob <- lapply(qc, function(x) x/sum(x))
hat_H1 <- unlist(lapply(prob, FUN = Shannon.E_norm), recursive = FALSE)
sim_pat <- rpoispp(obs_pattern$n, W = owin(xlim, ylim), nsim = 100)
qc_sim <- list()
prob_sim <- list()
hat_Hi <- list()
for(i in 1:length(qc_dim)){
  qc_sim <- lapply(sim_pat, nx = qc_dim[[i]], FUN = quadratcount)
  prob_sim <- lapply(qc_sim, function(x) x/sum(x))
  hat_Hi[[i]] <- unlist(lapply(prob_sim, FUN = Shannon.E_norm), recursive = FALSE)
}

U_env <- unlist(lapply(hat_Hi, FUN = max))
L_env <- unlist(lapply(hat_Hi, FUN = min))

bar_H1 <- unlist(lapply(hat_Hi, FUN = function(x){(1/nsim)*sum(x)}))
bar_Hi <- list()
hat_Hi_t <- list()
for(i in 1:nsim){
  bar_Hi[[i]] <- vector()
  hat_Hi_t[[i]] <- vector()
  for(j in 1:length(qc_dim)){
    bar_Hi[[i]][j] <- (1/(nsim-1))*sum(hat_Hi[[j]][-i])
    hat_Hi_t[[i]][j] <- hat_Hi[[j]][i]
  }
}

ui <- mapply(function(x,y){max(abs(x-y))}, x = hat_Hi_t, y = bar_Hi)

u1 <- max(abs(hat_H1 - bar_H1))

ind <- lapply(ui, FUN = function(x){ifelse(x > u1, 1, 0)})
sig[[k]][h] <- sum(unlist(ind))/length(ind)

```

```

}
}

#Calibration - CSR null hypothesis vs. homogeneous Poisson process
qc_dim <- seq(2,30,1) # quadrat size
nsim <- 100

s.region <- rbind(c(0, 0), c(1, 0), c(1, 1), c(0, 1))
xlim <- ylim <- c(0, 1)
ngrid <- 500
gridwidth <- diff(xlim)/ngrid
x.seq <- y.seq <- seq(xlim[1]+0.5*gridwidth,xlim[2]-0.5*gridwidth,by=gridwidth)

sig <- list()
for(h in 1:50){
  print(h)
  obs_pattern <- rpoispp(1000, W = owin(xlim, ylim))
  qc <- list()
  for(i in 1:length(qc_dim)){
    qc[[i]] <- quadratcount(obs_pattern, nx = qc_dim[i])
  }
  prob <- lapply(qc, function(x) x/sum(x))
  hat_H1 <- unlist(lapply(prob, FUN = Shannon.E_norm), recursive = FALSE)

  sim_pat <- rpoispp(obs_pattern$n, W = owin(xlim, ylim), nsim = 100)
  qc_sim <- list()
  prob_sim <- list()
  hat_Hi <- list()
  for(i in 1:length(qc_dim)){
    qc_sim <- lapply(sim_pat, nx = qc_dim[[i]], FUN = quadratcount)
    prob_sim <- lapply(qc_sim, function(x) x/sum(x))
    hat_Hi[[i]] <- unlist(lapply(prob_sim, FUN = Shannon.E_norm), recursive = FALSE)
  }
}

```

```

U_env <- unlist(lapply(hat_Hi, FUN = max))
L_env <- unlist(lapply(hat_Hi, FUN = min))

bar_H1 <- unlist(lapply(hat_Hi, FUN = function(x){(1/nsim)*sum(x)}))
bar_Hi <- list()
hat_Hi_t <- list()
for(i in 1:nsim){
  bar_Hi[[i]] <- vector()
  hat_Hi_t[[i]] <- vector()
  for(j in 1:length(qc_dim)){
    bar_Hi[[i]][j] <- (1/(nsim-1))*sum(hat_Hi[[j]][-i])
    hat_Hi_t[[i]][j] <- hat_Hi[[j]][i]
  }
}

ui <- mapply(function(x,y){max(abs(x-y))}, x = hat_Hi_t, y = bar_Hi)
u1 <- max(abs(hat_H1 - bar_H1))

ind <- lapply(ui, FUN = function(x){ifelse(x > u1, 1, 0)})
sig[h] <- sum(unlist(ind))/length(ind)
}

#####
## Application shannon-based goodness-of-fit test on forest fires      *#
## dataset                                                            *#
#####

# Read data
fire <- readShapeSpatial(".././../BBDD/Data/Data/ForestFire/ForestFire_Nepal")
fire_pp <- unmark(as(fire, "ppp"))
fire_pp$window <- bwin

```

```
# Model A
A <- kernel_dens <- density(fire_pp, bw.diggle(fire_pp), kernel = "epanechnikov",
                           diggle = TRUE, scalekernel = TRUE, edge = TRUE,
                           positive = TRUE, dimyx = c(128, 128))
pat_A <- rpoispp(A, nsim = 100)

# Model B
parest <- minimum.contrast(fire_pp, model = "exponential", method = "g")
sar <- spatialAtRisk(A)
mcmc.ctrlr <- mcmcpars(mala.length = 120000, burnin = 20000, retain = 100,
                      adaptivescheme = andrieuthomsh(inith = 1,
                                                    alpha = 0.5,
                                                    C = 1,
                                                    targetacceptance = 0.574))

pred_B <- lgcpPredictSpatial(sd = fire_pp,
                             model.parameters = lgcppars(
                               sigma = parest$estimates[[2]],
                               phi = parest$estimates[[1]]),
                             spatial.covmodel = "exponential", gridsize = 256,
                             spatial.intensity = sar,
                             mcmc.control = mcmc.ctrlr, ext = 4)

B <- intens(pred_B)
pat_B <- rpoispp(B, nsim = 100)

# Compute measures for real pattern, model A and model B
qc_dimy <- seq(2, 30, 1)
qc_dimx <- 2*qc_dimy

# Quadrat-counting and summary description for real pattern

qc <- list()
```

```

for(i in 1:length(qc_dimx)){
qc[[i]] <- quadratcount(fire_pp, nx = qc_dimx[i], ny = qc_dimy[i])
}
prob <- lapply(qc, function(x) x/sum(x))
shannon <- lapply(prob, FUN = Shannon.E_norm)
hat_H1 <- unlist(shannon)

# Model A (Kernel)
qcA <- list()
probA <- list()
shannonA <- list()
for(i in 1:length(qc_dim)){
print(i)
qcA[[i]] <- lapply(pat_A, nx = qc_dimx[i], ny = qc_dimy[i], FUN = quadratcount)
probA[[i]] <- lapply(qcA[[i]], function(x) x/sum(x))
shannonA[[i]] <- unlist(lapply(probA[[i]], FUN = Shannon.E_norm), recursive = FALSE)
}

hat_HiA <- shannonA

U_env <- unlist(lapply(hat_HiA, FUN = max)) #Upper envelope
L_env <- unlist(lapply(hat_HiA, FUN = min)) #Lower envelope

bar_H1 <- unlist(lapply(hat_HiA, FUN = function(x){(1/100)*sum(x)}))

bar_HiA <- list()
hat_HiA_t <- list()
for(i in 1:100){
bar_HiA[[i]] <- vector()
hat_HiA_t[[i]] <- vector()
for(j in 1:29){
bar_HiA[[i]][j] <- (1/99)*sum(hat_HiA[[j]][-i])
hat_HiA_t[[i]][j] <- hat_HiA[[j]][i]
}
}

```

```
}
}
```

```
plot(x = hat_H1, ylim = c(0.705,0.86), type = "l",
     ylab = "", xlab = TeX(r'($k$)'), frame = FALSE)
mtext(TeX(r'($\hat{H}_1$)'), side=2, line=2, las = 1)
lines(bar_H1, lty = "dashed")
lines(L_env, col = "red")
lines(U_env, col = "red")
```

```
ui <- mapply(function(x,y){max(abs(x-y))}, x = hat_HiA_t, y = bar_HiA)
u1 <- max(abs(hat_H1[1:29] - bar_H1))
```

```
ind <- lapply(ui, FUN = function(x){ifelse(x > u1, 1, 0)})
sigA <- sum(unlist(ind))/length(ind)
```

```
#Histogram
minx <- min(ui)
maxx <- max(ui)
breaksx <- seq(minx, maxx, length = 15)
par(mfrow=c(1,1))
hist(ui,breaks=breaksx,main=" ",xlab=TeX(r'($u$)'),col="grey",border=F,
     freq = FALSE)#true
abline(v = u1, col = "red")
```

```
# Model B (LGCP)
```

```
qcB <- list()
probB <- list()
shannonB <- list()
for(i in 1:length(qc_dimx)){
  qcB[[i]] <- lapply(pat_B, nx = qc_dimx[i], ny = qc_dimy[i], FUN = quadratcount)
```

```

probB[[i]] <- lapply(qcB[[i]], function(x) x/sum(x))
shannonB[[i]] <- unlist(lapply(probB[[i]], FUN = Shannon.E_norm), recursive = FALSE)
}

hat_HiB <- shannonB

U_env <- unlist(lapply(hat_HiB, FUN = max))
L_env <- unlist(lapply(hat_HiB, FUN = min))

bar_H1 <- unlist(lapply(hat_HiB, FUN = function(x){(1/100)*sum(x)}))

bar_HiB <- list()
hat_HiB_t <- list()
for(i in 1:100){
  bar_HiB[[i]] <- vector()
  hat_HiB_t[[i]] <- vector()
  for(j in 1:29){
    bar_HiB[[i]][j] <- (1/99)*sum(hat_HiB[[j]][-i])
    hat_HiB_t[[i]][j] <- hat_HiB[[j]][i]
  }
}

plot(x = hat_H1[1:29], ylim = c(0.705,0.86), type = "l",
     ylab = "", xlab = TeX(r'($k$)'), frame = FALSE)
mtext(TeX(r'($\hat{H}_1$)'), side=2, line=2, las = 1)
lines(bar_H1, lty = "dashed")
lines(L_env, col = "red")
lines(U_env, col = "red")

ui <- mapply(function(x,y){max(abs(x-y))}, x = hat_HiB_t, y = bar_HiB)
u1 <- max(abs(hat_H1[1:29] - bar_H1))

```

```
ind <- lapply(ui, FUN = function(x){ifelse(x > u1, 1, 0)})
sigB <- sum(unlist(ind))/length(ind)
sigB

minx <- min(ui)
maxx <- max(ui)
breaksx <- seq(minx, maxx, length = 13)
par(mfrow=c(1,1))
hist(ui,breaks=breaksx,main=" ",xlab=TeX(r'($u$)'),col="grey",border=F)#true
abline(v = u1, col = "red")
```