Universitat de València

Facultat de Ciències Matemàtiques

Departament d'Estadística i Investigació Operativa

Programa de doctorat en Estadística i Optimització

Tesi Doctoral

Methods for the statistical analysis of spatio-temporal data

Adina Alexandra Iftimi

Directors:
Dr. Ottmar Cronie
Dr. Francisco Martínez Ruiz
Dr. Francisco Montes Suay

Maig 2017

Ottmar Cronie, Professor del Departament de Matemàtica i Estadística Matemàtica de la Universitat d'Umeå, Suècia, Francisco Martínez Ruiz, Professor Associat del Departament d'Estadística i Investigació Operativa de la Universitat de València, i Francisco Montes Suay, Catedràtic d'Universitat del Departament d'Estadística i Investigació Operativa de la Universitat de València,

CERTIFIQUEN que la present memòria d'investigació, titulada:

"**Methods for the statistical analysis of spatio-temporal data**"

ha estat realitzada sota la seua direcció per Adina Alexandra Iftimi i constitueix la seua tesi per optar al grau de Doctor per la Universitat de València Estudi General.

I perquè així conste, en compliment amb la normativa vigent, n'autoritzen la presentació davant la Facultat de Ciències Matemàtiques de la Universitat de València perquè en puga ser tramitada la lectura i defensa pública.

Burjassot, 5 de Maig de 2017.

Ottmar Cronie

Francisco Martínez Ruiz

Francisco Montes Suay

Acknowledgements

I would like to express my sincere appreciation and thanks to my advisor Paco Montes for his continuous help and guidance. His words of wise were of great aid whenever needed. Also, his patience, motivation and immense knowledge led me in all the time of research. Paco, you have been a tremendous mentor for me. I would like to thank you for encouraging my research and for allowing me to grow as a research scientist.

Also, I would like to thank my advisor Ottmar Cronie. You have taught me to seek and find the fun part in research. I am grateful for all your advice and humbled to call you my friend. Thank you for your continuous motivation and encouragement, but also for the hard conversations which stimulated me to widen my research. And to my advisor Paco Martínez, who provided me with valuable data and advice. Without your help, it would not have been possible to conduct this research.

I am also grateful to all the staff from the University of València (Spain), University Jaume I of Castellón (Spain), University of Umeå (Sweden), CWI and University of Twente (The Netherlands), in particular to Jorge Mateu for his undoubted support and advice and to Marie-Colette van Lieshout for her guidance in writing Chapter 4 of this thesis.

I would like to thank Francisco González from Dirección General de Salud Pública, Servicio de Vigilancia y Control Epidemiológico, Generalitat Valenciana for providing the varicella data. And also Ana Míguez for her very useful comments on epidemiology related issues.

I gratefully acknowledge financial support from the Ministry of Education, Culture, and Sports (Grants FPU12/04531 and EST15/00174), also Fundació Ferran Sunyer i Balaguer. This allowed me to visit the CWI, University of Twente and University of Umeå.

I thank all my colleagues and friends for stimulating discussions and for all the fun we had in the last four years.

Special thanks to my family, my parents and my sisters for supporting me throughout the writing of this thesis and my life in general. Last but by no means least, I would like to express appreciation to my beloved husband Antonio. You are always my support.

to my family

Contents

Acknowledgements **i**

1 Introduction **1**

2 Methods for statistical analysis of spatial data **7**
- 2.1 Introduction . . . 7
- 2.2 Methods . . . 9
  - 2.2.1 Descriptive analysis . . . 9
  - 2.2.2 Hybrid models for point process data . . . 11
  - 2.2.3 Diagnostics . . . 12
- 2.3 Spatial analysis for the varicella point pattern . . . 13
  - 2.3.1 The study area and the varicella dataset . . . 13
  - 2.3.2 Descriptive analysis . . . 17
  - 2.3.3 Inhomogeneous Baddeley–Geyer hybrid model . . . 23
- 2.4 Conclusions and discussion . . . 26

3 Methods for statistical analysis of space-time lattice data **29**
- 3.1 Introduction . . . 29
- 3.2 Methods . . . 30
  - 3.2.1 Standardisation of rates . . . 30
  - 3.2.2 Exploratory space-time data analysis . . . 31
    - 3.2.2.1 Moran index and Moran scatter plot . . . 32
    - 3.2.2.2 Directional Moran scatter plot . . . 33
    - 3.2.2.3 Spatial Markov chain models . . . 35
  - 3.2.3 Space-time models . . . 35
    - 3.2.3.1 Models with parametric time trend . . . 36
    - 3.2.3.2 Models with non-parametric time trend . . . 36
- 3.3 Space-time analysis for the varicella data . . . 38
  - 3.3.1 Exploratory analysis . . . 38
  - 3.3.2 Space-time models . . . 41
- 3.4 Conclusions and discussion . . . 43

4 A multi-scale area-interaction model for spatio-temporal point patterns 47
4.1 Introduction . . . 47
4.2 Preliminaries . . . 48
4.3 Space-time area-interaction processes . . . 49
  4.3.1 Markov properties . . . 50
4.4 Simulation . . . 53
  4.4.1 The Metropolis-Hastings algorithm . . . 53
  4.4.2 Birth-and-death-processes . . . 55
4.5 Inference . . . 56
  4.5.1 Pseudo-likelihood method . . . 56
  4.5.2 Simulation and parameter estimation of a spatio-temporal area interaction process . . . 58
4.6 Application. Varicella data . . . 61
4.7 Conclusions . . . 66

5 The second-order analysis of marked spatio-temporal point processes 67
5.1 Introduction . . . 67
5.2 Marked spatio-temporal point processes . . . 69
  5.2.1 Preliminaries . . . 69
  5.2.2 Marked spatio-temporal point processes . . . 71
  5.2.3 Intensity functions . . . 72
    5.2.3.1 Separability . . . 74
  5.2.4 Pair correlation functions . . . 75
  5.2.5 Specific marking structures . . . 75
    5.2.5.1 Common mark distribution . . . 75
    5.2.5.2 Independent marks and random labelling . . . 76
  5.2.6 Palm distributions . . . 76
    5.2.6.1 Reduced Palm distributions with respect to the mark sets 77
5.3 Examples of models . . . 77
5.4 Marked inhomogeneous second-order measures of spatio-temporal interaction . . . 81
  5.4.1 Second order intensity-reweighted stationarity . . . 81
  5.4.2 Marked spatio-temporal second-order reduced moment measures . 82
    5.4.2.1 Representation in terms of Palm distributions . . . 84
    5.4.2.2 Changing the order of the mark sets . . . 84
  5.4.3 Marked spatio-temporal inhomogeneous K-functions . . . 85
  5.4.4 Marked stationary spatio-temporal K-functions . . . 87
  5.4.5 Multivariate spatio-temporal K-functions . . . 88
    5.4.5.1 Multivariate stationary spatio-temporal K-functions . . . 89
  5.4.6 Further properties . . . 89
    5.4.6.1 Independent thinning . . . 89
    5.4.6.2 Scaling . . . 90

5.5 Statistical inference . . . 91
  5.5.1 Voronoi intensity estimation . . . 91
    5.5.1.1 Simplifying assumptions . . . 93
  5.5.2 Estimation of the second-order summary statistics . . . 94
    5.5.2.1 Smoothing . . . 98
    5.5.2.2 Multivariate versions . . . 98
    5.5.2.3 Common mark distribution versions . . . 98
    5.5.2.4 Homogeneous versions . . . 99
  5.5.3 Independence assumptions . . . 99
    5.5.3.1 Testing random labelling . . . 102
5.6 Second order analysis of the earthquake data . . . 105
  5.6.1 Earthquakes data set . . . 105
  5.6.2 Analysis . . . 107
5.7 Conclusion . . . 111

6 Future work **113**

Bibliography **115**

Resumen **125**

List of Tables

2.1 Population by age group in 2010. 15
2.2 Evolution of the Valencia population (districts 1 to 16) and the number of varicella cases. 17
2.3 Estimated parameters for the inhomogeneous Poisson model. 22
2.4 The estimated parameters for the Baddeley-Geyer hybrid model. 24

3.1 Specifications for the four types of space-time interaction. 38
3.2 Moran index for SIR of varicella values from 2008 to 2013. 39
3.3 Moran directional summary with non-parametric p-values. 40
3.4 DIC and logarithmic score for the model with parametric time trend. . . . 42
3.5 DIC and logarithmic score for all models with non-parametric time trend. 42

4.1 Parameter estimates for simulation 1 59
4.2 Parameter estimates for simulation 2 59
4.3 Parameter estimates for the varicella data. 65

List of Figures

2.1 Location of the city of Valencia. Universal Transverse Mercator (UTM) coordinate system (distance in meters). 15

2.2 Valencia city map. (Top) Territorial division by districts. (Bottom) Territorial division by boroughs. 16

2.3 (Left) Spatial point pattern of varicella cases registered during 2013 in Valencia (Spain), together with quadrat plot for the point pattern. (Right) Point pattern of independent uniform random points. 18

2.4 (Left) Empirical G function (solid line) with pointwise maximum and minimum envelopes obtained under CSR null hypothesis (grey area). (Right) K-function (solid line) with pointwise maximum and minimum envelopes obtained under CSR null hypothesis (grey area). 18

2.5 Kernel smoothing estimate of the varicella point pattern intensity. (Left) Bandwidth = 153 (Diggle, 1985). (Right) Bandwidth = 300. 19

2.6 (Left) Residual G function (solid line). (Right) Residual K-function (solid line); and the corresponding envelopes obtained under the inhomogeneity assumption previously described (grey area). 19

2.7 (Left) Kernel estimate of population density under 14 years, 2013. (Centre) Kernel estimate of the distance function to the nearest school in Valencia. (Right) Kernel estimate of the average number of persons per family register. 20

2.8 (Left) Coordinates and corresponding p-values for the six main school clusters detected using `SatScan` software. (Right) Pattern of all schools in Valencia (black dots), together with the spatial locations of the main clusters (dots encircled and labelled with an X). 21

2.9 Non-parametric smoothed estimation of the intensity as a function of spatial covariates. (Left) Population density under 14 years, 2013. (Centre) nearest distance to school in Valencia. (Right) Average number of persons per family register. 22

2.10 (Left) Residual G function for the inhomogeneous Poisson process. (Right) Residual K-function for the inhomogeneous Poisson process. 23

2.11 Cumulative raw residuals for the x and y axis. 23

2.12 (Left) Residual G function for the hybrid model. (Right) Residual K-function for the hybrid model. 25

2.13 (Top) Cumulative raw residuals for the x and y axis for the hybrid model. (Bottom) Qqplot for the raw residuals for the hybrid model. 25
2.14 (Left) Spatial distribution of varicella point pattern in Valencia, Spain, cases registered during 2013. (Right) Simulation from the hybrid model. . 26
2.15 Cumulative raw residuals for the x and y axis; qqplot for the raw residuals. Hybrid model without the focal points. 27
2.16 Cross K-function between the varicella and schools patterns. 27

3.1 Moran scatter plot for a variable X and its spatial lag Y. 33
3.2 Upper row: simple Moran scatter plot. Lower row: directional Moran scatter plot. 34
3.3 Right: spatial distribution of varicella SIR in Valencia for 2008. Left: spatial distribution of varicella SIR in Valencia for 2013. 38
3.4 Standardised Moran directional scatter plots for 2008-2013 and 2010-2011. 39
3.5 Evolution of the number of varicella cases (solid line) and evolution of the population (dotted line) during the time period 2008-2013. 41
3.6 (Top) PIT-histograms for models with type I to II interaction. (Bottom) PIT-histograms for models with type III to IV interaction. 43
3.7 (Top) Distribution of the estimated relative risk of varicella according to models 3 and 4 for 2008. (Bottom) Distribution of the estimated relative risk of varicella according to models 3 and 4 for 2013. 44
3.8 (Left) Linear time trend for borough 3 from district 9 corresponding to the model with type IV interaction. (Right) Linear time trend for borough 4 from district 12 corresponding to the model with type IV interaction . . . 45

4.1 An illustration of possible $\mathbf{x} \oplus G$ (cylinders around the points), where the black dots represent points of the process. 50
4.2 (Left) An illustration of $F_{r_j}^{t_j}$ where the blue annulus corresponds to $\{(y,s) \in W_S \times W_T : r_{j-1} < ||x-y|| \leq r_j, |t-s| \leq t_{j-1}\}$, the two green annuli represent $\{(y,s) \in W_S \times W_T : r_{j-1} < ||x-y|| \leq r_j, t_{j-1} < |t-s| \leq t_j\}$ and the two red cylinders are $\{(y,s) \in W_S \times W_T : ||x-y|| \leq r_{j-1}, t_{j-1} < |t-s| \leq t_j\}$. (Right) Multi-scale behavior. 52
4.3 (Top left) Model parameters for *simulation 1*. (Top right) A realization of the first model. (Bottom left) Model parameters for *simulation 2*. (Bottom right) A realization of the second model. 60
4.4 Spatio-temporal pattern of weekly varicella cases in Valencia during 2013, where the spatio-temporal study area is $W_S \times W_T = ([0,9] \times [0,9]) \times [0,52]$ (km$^2\times$ weeks). 61
4.5 (Left) Spatial projection of the spatio-temporal point pattern for the varicella data. After projection, locations were jittered using a maximum jitter distance of 20 metres. (Right) Estimated pair correlation function for the jittered spatial point pattern shown in the left panel. 63

4.6 (Left) Weekly reports of varicella cases (∘) and fitted regression curve (–). (Right) Estimated auto-correlation function for the data shown in the left panel. 63

4.7 Model parameters for the varicella data. 64

4.8 (Left) Realization from the model fitted to the varicella data. (Right) Spatio-temporal pattern of weekly varicella cases in Valencia during 2013, where the spatio-temporal study area is $W_S \times W_T = ([0,9]\times[0,9])\times[0,52]$ (km$^2\times$ weeks). 66

5.1 (Left) Spatio-temporal Poisson process with intensity function $\lambda(x_1,y_1,t_1) = 5t_1 e^{5+0.5x_1}$ on $[0,1]^2\times[0,1]$ with independent Bernoulli distributed marks (parameter $p = 0.4$). (Centre) Spatial projections for the time interval $[0,0.5]$. (Right) Spatial projections for the time interval $(0.5,1]$. Here $\mathcal{M} = \{0,1\}$ and 'type 1' refers to a point having mark 0. 78

5.2 (Left) Randomly labelled spatio-temporal log-Gaussian Cox process with iid $Bernoulli(0.4)$-distributed marks. (Centre) Projection of the random intensity field of the log-Gaussian Cox process Y_g, at time $t_1 = 0.5$. (Right) Projection of the random intensity field of the log-Gaussian Cox process Y_g, at spatial coordinate $x_1 = 0.5$. Here $\mathcal{M} = \{0,1\}$ and 'type 1' refers to a point having mark 0. 79

5.3 (Left) A realisation of a bivariate spatio-temporal process. (Centre) Projection of the random intensity field of Y_2, at time $t_1 = 0.5$. (Right) Projection of the random intensity field of Y_2, at spatial coordinate $x_1 = 0.5$. 80

5.4 (Left) A realisation of the above defined geostatistically marked spatio-temporal log-Gaussian Cox process; the size of a point is proportional to the value of its continuous mark. (Right) The Gaussian random field generating the marks. 81

5.5 The four Hamilton principle scenarios for the estimator (5.24); min-max envelopes based on 99 realisations of the randomly labelled Poisson process in Example 1. 97

5.6 (Top) Envelopes for the estimate $\widehat{K}^{CD}_{\text{inhom}}(r,t) - \widehat{K}^{Y_g}_{\text{inhom}}(r,t)$, where $C = \{0\}$ and $D = \{1\}$, based on 99 realisations of the the model given in Example 2, for fixed temporal lags $t = 0.05$ and $t = 0.10$. (Centre) Envelopes for the estimate $\widehat{K}^{CD}_{\text{inhom}}(r,t) - \widehat{K}^{Y_g}_{\text{inhom}}(r,t)$,$t = 0.15$ for fixed temporal lags $t = 0.15$ and $t = 0.30$. (Bottom) The estimate $\widehat{K}^{CD}_{\text{inhom}}(r,t) - \widehat{K}^{Y_g}_{\text{inhom}}(r,t)$, for all space-time lags. 101

5.7 (Top) Envelopes for the estimated $K^{CD}_{\text{inhom}}(r,t) - 2\pi r^2 t$, where $C = \{0\}$ and $D = \{1\}$, based on 99 realisations of the the model given in Example 3, for fixed temporal lags $t = 0.05$ and $t = 0.10$. (Centre) Envelopes for the estimated $K^{CD}_{\text{inhom}}(r,t) - 2\pi r^2 t$, for fixed temporal lags $t = 0.15$ and $t = 0.30$. (Bottom) The estimate $\widehat{K}^{CD}_{\text{inhom}}(r,t) - 2\pi r^2 t$ for all r and t. . . . 103

5.8 The estimated $\Delta(r,t) = K^{CD}_{\text{inhom}}(r,t) - K^{DC}_{\text{inhom}}(r,t)$, $Y_C = \{(x,t,m) \in [0,1] \times [0,1] \times [0,0.5]\}$, $Y_D = \{(x,t,m) \in [0,1] \times [0,1] \times (0.5,1]\}$, for the realisation in Figure 5.4, together with $\Delta(r,t)$-envelopes generated by 99 resamples/permutations of the marks, for fixed temporal lag $t = 0.20$. . . 104

5.9 (Left) Spatial locations and magnitudes of the 1248 earthquakes registered in the Sumatra area. The sizes of the dots are proportional to the magnitudes. The red X:s correspond to the four important earthquakes described above. (Right) Temporal evolution of magnitude of all earthquakes. 106

5.10 Spatial locations and magnitudes of the earthquake data, annually, from 2004 to 2008. 107

5.11 The marked spatio-temporal pattern of the earthquake dataset. The sizes of the circles are proportional to the magnitudes 109

5.12 (Left) Estimated K-function for the Phuket data, $K^{CD}_{\text{inhom}}(r,t) - 2\pi r^2 t$. (Right) Smoothed K-function estimate, $\widetilde{K}^{CD}_{\text{inhom}}(r,t) - 2\pi r^2 t$. 110

5.13 (Left) Smoothed K-function estimate, $\widetilde{K}^{CD}_{\text{inhom}}(r,t) - 2\pi r^2 t$, for the time frame of one day, (centre) one week and (right) 50 days. 111

5.14 Estimate of $\Delta(r,t) = K^{CD}_{\text{inhom}}(r,t) - K^{DC}_{\text{inhom}}(r,t)$, for the earthquake dataset, together with 95% two-sided point-wise Monte-Carlo confidence bands, for fixed temporal lag $t = 836$ days. 111

Chapter 1

Introduction

The first attempts to introduce and interpret spatial data using statistics appeared in the form of data maps. Halley (1686) superimposed on a map the locations and directions of trade winds and monsoons situated around the tropics, and assigned a physical explanation to them. Later on, Snow (1855) drew maps of London, showing clusters of cholera cases from the 1854 Broad Street cholera outbreak. This is considered to be the first use of maps for spatial analysis.

Spatial statistics methodology was not introduced until much later. Gosset (1907), better known by his pen name Student, was interested in the study of the distribution of particles in a liquid, and instead of analysing the spatial positions of the particles, he aggregated the data into small areas. In the 1920's and 1930's, R. A. Fisher established the principles of the replica method, random sampling and blocks control. In the agricultural field, nearest neighbour algorithms (Altman, 1992) were used to conduct and analyse different studies and spatial dependence was taken into account (Haapanene et al., 2004). Currently, modern statistical methods are found in a wide range of research fields such as cartography, geology, ecology, biology (botanical studies of plants distribution, studies of biogeography), epidemiology (disease mapping), spatial econometrics, remote sensing (computational geometry), astronomy (studies of placement of galaxies) and many others.

Before proceeding with a description of spatial data, some comments on the data quality and the impact it has on statistical analysis, are to be made. In order to carry out a thorough study, it is important to have a complete and high-quality database. Sometimes one can stumble upon problems like typographical, encoding errors, incomplete or duplicate data. Therefore, it is crucial that, before beginning any type of statistical analysis, to always carry out a process of data validation in order to identify potential errors and, preferably, correct them. Researchers also have to take into consideration the right to privacy when collecting and processing data of personal nature. Studies which show unfavourable results, for example, areas with high levels of air or noise pollution, clusters with a high incidence of a disease, can cause discontent among the population. Therefore, many institutions impede the access to data using alleged protective policies. For researchers, the availability of data is very important, so the trust and support of the population are required.

Spatial data can be of continuous or discrete nature. It can be spatially aggregated into small regions, or it can be represented as individual observations. Spatial locations can be distributed in a regular or irregular manner into the spatial region.

Cressie (1991) proposed a classification of spatial data into three main groups: *geostatistical data*, *lattice data* and *spatial point patterns*. When a phenomenon is observed as measurements at a finite set of predetermined fixed spatial locations, the result is geostatistical data. When the spatial region (regular or irregular) is partitioned into a finite number of bounded regions, with well-defined limits, and data is obtained as aggregated observations in each of the defined regions, we get lattice data. And finally, when the spatial pattern is obtained observing random occurrences of a phenomenon, the data is called a point pattern.

Geostatistics was initially developed to predict mineral reserve using a limited number of spatial locations in a region. Thus, mining engineering has been one of the first fields where geostatistics was used extensively. Here, the fundamental key in the modelling of spatial relationships is *the variogram* which is estimated and modelled to adequately describe the observed phenomenon. The main objective of the application of geostatistics is usually to predict at a point or set of points in the observed region, where no measurements have been made. *Kriging* is a geostatistical technique used to interpolate the value of a random field at an unobserved location using the information from nearby observations and is one of the most used techniques in spatial prediction. Geostatistical methods are currently used in diverse disciplines, including petroleum geology, hydro-geology, hydrology, meteorology, oceanography, geochemistry, metallurgy, geography, forestry, environmental control, landscape ecology, soil science, agriculture and many others.

Lattice data sets are observed when the spatial region is partitioned into small regions. Statistical models for this type of data usually incorporate the relationship between observations of neighbouring locations. One of the goals of the lattice data analysis is obtaining an estimate of some measure of interest, in each region. Also, it may be of interest to analyse possible spatial structure in the data (clustering, repulsion). An example of lattice data is data collected by remote sensing satellites, which provide a very useful data collecting tool. For example, weather information, weather patterns, distribution of minerals or soil types can be observed using satellites, without the need of field sampling. The surface of the Earth is divided into small rectangles called pixels that are assigned a value of a given characteristic so that the data is received in the form of a regular network and each pixel is identified using its centre.

Spatial point data refer to phenomena that occur continuously and randomly at different locations in a region. A point pattern is the collection of locations of such events. Point processes are the theoretical mechanisms that allow modelling the development of the phenomena. One of the objectives of point pattern studies is to see whether an observed spatial pattern has a certain type of structure or, on the contrary, if it is simply the result of a (homogeneous) point process that acts independently in space. If we consider the locations of longleaf pines in a forest, we can raise several questions.

What is the biological meaning of the clustering of these trees? Are the tall trees clustered in groups? Do tall trees interact with small trees? Point patterns analysis aims at estimating parameters associated with specific models. Formal analysis of a spatial point pattern requires the use of multiple statistical techniques. First and second order summary functions are practical and useful tools to effectively describe and analyse the spatial structure. Diggle (2014) (Chapter 7) illustrates model-fitting using summary descriptions for several datasets.

Any phenomenon can be located in a geographic reference. Diseases, deaths, births, risk exposure or other events, can be associated with the place where they occur, which may be coordinates, addresses or areas. Adding a temporal component provides important insights into the understanding of the statistical process. The first law of geography by Waldo Tobler, *everything is related to everything else, but near things are more related than distant*, gives us a motivation for including neighbouring relation into any spatial analysis. Adding the temporal component transforms the analysis into a dynamic and integrative task, where information represents more than just numbers or static characteristics.

Space-time statistics are being used more and more in an attempt to explain the behaviour of this wide range of events. The main reason for using space–time methods is to reduce the initial variability of the risk by using and exchanging information between neighbouring spatial locations and proximal temporal moments.

Spatio-temporal data can be presented using maps for each time period that help detect high-risk areas or areas with unusual activity. Space-time epidemiology, for example, enables the identification of individual and aggregate risk factors for health, becoming an indispensable tool for decision making in public health. It also allows to dynamically evaluate risk factors, analyse their impact on the population and estimate the potential benefits that preventive measures are to public health.

It should be noted that in the context of spatio-temporal statistical analysis $2 + 1$ is not equal to 3 because the temporal dimension is fundamentally different from either spatial dimensions. Most spatial processes in nature are merely snapshots of the evolution of spatio-temporal processes, but using solely spatio-temporal analysis methods would be a mistake. The use of these techniques should be done when our interest involves both components, spatial and temporal, and should not be carried out using only statistical analyses of the two components separately.

Many real-life scenarios, such as the occurrence of earthquakes, disease incidents or fires, give rise to data collections of events where each event, besides having a spatial location and an event time, also carries further helpful information about the event in question. In the language of point processes, this ‘extra’ information is referred to as the *mark* of the event. To exemplify, in the case of earthquakes a mark could correspond to the magnitude of an event. When a mark is attached to a space-time point in this fashion, the random element/mechanism assumed to have generated the total collection of data is referred to as a *marked spatio-temporal point process (MSTPP)*, with the corresponding data referred to as a *marked spatio-temporal point pattern* (Daley and Vere-Jones, 2003; Diggle, 2014; Vere-Jones, 2009). Other applications of MSTPPs include occurrences of disease incidents, crimes, fires and many others.

A general structure of the thesis and contributions are presented as follows. Chapter 2 gives a basic step-by-step analysis for spatial data. Starting with a descriptive analysis by testing complete spatial randomness and inhomogeneity, then introducing descriptive analysis with covariates. This chapter presents a technique for constructing new Gibbs models for spatial point patterns proposed by Baddeley et al. (2013). In this chapter, the varicella spatial point pattern of all registered cases during 2013 in Valencia, Spain, is analysed. Descriptive analysis is used to get a glimpse of the basic properties of the point pattern. Covariate information such as the density of population (children under 14 years old) living in the study region, the distance to the nearest school, and the composition of families (expressed as the average number of persons per family) is used to describe the intensity of the process. The `SatScan` software (Kulldorff, 2010) is used to identify main clusters of schools. This information is further fed into the model. A comparison between different models is presented here, and also diagnostic methods to choose the best model to fit the data. Therefore, the first contribution of this thesis is the application of statistical analysis for spatial point patterns to epidemiology. As previously mentioned, this chapter puts forward a spatial analysis for the varicella data set through the use of hybrids of Gibbs models. The models presented in this chapter manage to describe interactions at different scales and also the spatial inhomogeneity considering covariate information. This chapter represents a minor but interesting contribution to the literature of point processes.

The second contribution of this thesis is in the field of spatio-temporal lattice data. Chapter 3 presents a second analysis of the varicella data set, only now in a lattice data format. The first part of this chapter is dedicated to exploratory space-time data analyses using a proposal of Rey (2014). The second part of the chapter puts forward statistical methodology regarding space-time parametric and non-parametric models for lattice data using a Bayesian framework. All these methods are applied to yearly varicella data. This chapter represents an interesting compendium of space-time analysis techniques applied to solve a specific epidemiological problem.

Chapter 4 proposes an extension of the area-interaction point process model (Baddeley and van Lieshout, 1995) to allow multi-scale interaction in a spatio-temporal framework. This model allows for interaction between points at different spatio-temporal scales and the inclusion of covariates. This chapter represents an important methodological contribution where we define and give the Markov properties of our spatio-temporal multi-scale area-interaction model. We then propose a modified Metropolis-Hastings and birth-and-death simulation algorithms. For inference, we use the pseudo-likelihood method and adapt the Berman-Turner procedure (Baddeley and Turner, 2000) to our context. The final part of this chapter applies the new model to the varicella data set. This chapter proposes a new model for analysing spatio-temporal point patterns and gives all the theoretical details on how to fit, simulate and validate the spatio-temporal multi-scale area-interaction model.

The fourth methodological contribution is presented in Chapter 5. This chapter aims at developing point process tools which allow us to perform so-called second-order non-parametric analyses of marked spatio-temporal point patterns. To analyse interaction in marked spatio-temporal point processes (MSTPPs), Chapter 5 introduces marked (cross) second-order reduced moment measures and K-functions for general inhomogeneous second-order intensity reweighted stationary MSTPPs. These summary statistics, which allow us to quantify dependence between different mark-categories of the points, are depending on the specific mark space and mark reference measure chosen. A new test for independent marking is proposed and unbiased minus-sampling estimators are derived for all statistics considered. In addition, we treat Voronoi intensity estimators for MSTPPs. These new statistics are finally employed to analyse the Andaman sea earthquake dataset. We find that clustering takes place between main and fore-/aftershocks at virtually all space and time scales. In addition, we find evidence that conditionally on the space-time locations of the earthquakes, the magnitudes do not behave like an iid sequence. Finally, Chapter 6 presents some future lines of research.

Chapter 2

Methods for statistical analysis of spatial data

Formal analysis of a spatial point pattern requires the use of multiple statistical techniques. First and second order summary statistics are practical and useful tools to effectively describe and analyse the spatial structure of a point pattern. Spatial statistical models take into account spatial dependency in regression analysis, putting forward information on spatial relationships between the covariates included in the model.

In this chapter we discuss the use of hybrids of Gibbs processes to model interactions at different spatial scales and to provide the statistical modelling of a specific data set. We introduce a dataset of varicella cases registered in Valencia, Spain. We also introduce an overview of different Gibbs model applications available in the literature. We then present tools for descriptive analysis, including complete spatial randomness and descriptive analysis with covariates. Also, we give theoretical tools for defining hybrid models and diagnostics methods. The novelty of this study is the epidemiological application presented at the end of the chapter. Section 2.3 puts forward a step-by-step statistical analysis for the varicella data. An important issue discussed in this chapter is the importance of schools locations and their contribution to explaining the spatial distribution of the disease. Finally, Section 2.4 gives some conclusions and a discussion on different aspects of the analysis.

2.1 Introduction

One important class of models for spatial point pattern analysis is the class of (finite) Gibbs point processes. Many theoretical applications of these models can be found in the literature (Geyer, 1999; Baddeley and Turner, 2000; Møller and Waagepetersen, 2004; Diggle, 2014). Moreover, Renshaw et al. (2009), Comas et al. (2009) and Comas and Mateu (2011) present different applications in forestry. Funwi-Gabga and Mateu (2012) use the area-interaction model to analyse the behaviour of the *Gorilla horilla diehli* in the Kagwene Sanctuary in Cameroon. Uria-Diez et al. (2013) use area-interaction and shot-noise Cox processes to model the distribution of three *Carex remota* cohorts in the north of Spain. Model-fitting, prediction, and simulation of Gibbs models are implemented in the `spatstat` package (Baddeley et al., 2015) of `R` (R Core Team, 2014). The most common use of Gibbs models is for *single* spatial scale interaction, which might not be realistic in real-life situations. In practical scenarios, human interaction as well as other phenomena like the disease spreading, exhibit spatial dependence at *multiple* scales. As stated before, point process models are generally used when only one type of spatial interaction (including only one scale of interaction) governs the structure of the point pattern. When there are indications that the spatial structure varies with ranges of distance, a global model is not suitable for describing the complex pattern of such interactions. Baddeley et al. (2013) propose local behaviour models called *hybrid models*. In their paper, the authors analyse human social interaction, studying the spatial locations of people sitting on the grass in a park on a sunny afternoon. This pattern clearly shows interaction at different scales. Another example of the multi-scale use of Gibbs models is Picard et al. (2009). The authors propose a marked area-interaction multi-scale model and apply the model to three examples from forestry that present different types of structure at different scales: a pine pattern, a bivariate kimboto pattern, and a marked pattern in Gabon, where the marks are represented by the tree diameter.

Motivated by the possible local scale behaviour, we propose an epidemiological application of hybrid models for a highly infectious disease, *varicella*.

In 1992 a multidisciplinary expert committee studied the threat of infectious diseases and related the emergence and re-emergence of these diseases to several factors, such as human demography, technological progress, economic development and land use, international trade, microbial adaptation, and failure in implementing public health measures (Lederberg et al., 1992). Later on, in relation to life-threatening infectious diseases such as smallpox or HIV, Brachman (2003) discusses the fact that modelling has become an important tool in how resources are distributed for purposes of control and prevention. High spreading risk makes infectious diseases difficult to contain and inhibit, in particular, airborne diseases, which are spread via air by coughing, sneezing, or talking. Airborne diseases of concern to emergency responders include *meningitis*, *varicella* (also known as *chicken pox*), *tuberculosis* or *influenza*. Airborne transmission depends on various endemic variables. The efficacy of airborne disease transmission is influenced by environmental factors (climate and geographical location) and/or socio-economic and living conditions. All these factors influence airborne diseases and give rise to complex multi-scale point patterns. In this chapter, we analyse the spatial point pattern of varicella cases registered during 2013. A detailed description of varicella in Valencia is given in Section 2.3.1.

The spatial point pattern of varicella reflects the complex structure inherited by an infectious disease. This complexity makes it difficult to understand the real causes behind its behaviour. Researchers in epidemiology usually make informal statements about disease spread and its relation to covariate information. We make use of stochastic processes in space, in particular of inhomogeneous spatial point processes, to provide a formal statistical procedure to disentangle the local interactions of such complex structures and highlight the role of available covariate information.

Varicella can be associated with different socio-demographic variables. Alp et al. (2005) show that the educational level influences the prevalence of anti-varicella antibodies, this prevalence being lower for children that are not attending school, followed by those who have at least attended elementary school. Cooper Robbinsa et al. (2011) provide a review of different studies on school-based vaccination. The effect school holidays have on the transmission of varicella is studied in Jackson et al. (2014). The authors show that reductions in contact between children during the summer break lead to a lower transmission of varicella. Due to the nature of the disease, it is well known that a relation exists between school locations and disease spread. However, to the best of our knowledge, this has not been confirmed by formal procedures. We use techniques of spatial point processes to analyse, check, and underline this relation.

2.2 Methods

2.2.1 Descriptive analysis

The intensity of a point process is the average density of points per unit area. It can either be constant at each point of the process (*homogeneous* process), or it can vary from one location to another (*inhomogeneous* process). In most applications, the intensity of the underlying process is not constant. For a spatial location $x \in \mathbb{R}^2$ we denote by $\lambda(x)$ the intensity function.

Testing for *complete spatial randomness* (henceforth, CSR) is equivalent to examine if a point pattern is a uniform Poisson point process with constant intensity λ. The basic properties to follow are: (i) the number of points in any region has a Poisson distribution with mean λ times the area of the region; (ii) given n points in a region, the locations of these point are independent and uniformly distributed; (iii) the counts of two disjoint regions are independent. We want to rule out the hypothesis of homogeneous Poisson process (the 'null hypothesis'). If this is not discarded then the original point pattern has nothing 'readable' because the points are completely unpredictable and have no type of dependence.

There are various methods for testing CSR. A classical test is the χ^2 test based on quadrat counts (Cressie and Read, 1984), but this testing procedure is very restrictive because it requires an equal-area partition of the region. A more general method, using 'random quadrats' is discussed in Cressie (1991). A more powerful test is the Kolmogorov-Smirnov test (Berman, 1986), in which the observed and the expected distribution of the values of some function T are compared. In this chapter we test for CSR using *summary statistics*, the nearest neighbour function and the method of K-functions.

When rejecting the CSR null hypothesis there is a fundamental ambiguity between inhomogeneity and clustering (Bartlett, 1964). Thus, the process could be an inhomogeneous Poisson process, a clustered process, a regular process or even a mixed process. The next step we take is to test for *inhomogeneity.* An inhomogeneous Poisson process is obtained when the constant intensity λ is substituted with a spatially varying function, $\lambda(x)$. For a visual inspection of the spatial point pattern and a first glimpse at its properties, we compute and plot a nonparametric intensity estimation, using kernel smoothing (Diggle, 1985). To test for inhomogeneity, we also use summary statistics as the nearest neighbour function and the K-function.

For a stationary point process X, *the nearest neighbour function*, henceforth the G function, is the cumulative distribution function of the distance $dist(x, X \setminus x)$ from a typical point $x \in X$. The K-function calculates the expected number of points within a certain distance from an arbitrary point, except the chosen point, divided by the average number of points per unit area Ripley (1976). For a nonstationary point process X, *the inhomogeneous G function* (Van Lieshout, 2011), is the intensity-reweighted equivalent of the nearest-neighbour distance distribution function G for homogeneous point processes. An estimator for the G function using generating functionals is given in (Van Lieshout, 2011). For a nonstationary point process X, *the inhomogeneous K-function* (Baddeley et al., 2000), a generalisation of the Ripley's K-function, represents the expected value, given that x is a point of X, of the sum of all terms $1/\lambda(y)$ over all points y in the process separated from x by a distance less than a certain value, r. An estimator for the inhomogeneous K-function is given in Baddeley et al. (2000). More details are available in Chiu et al. (2013). The K-function works for pairwise interactions, while the G function handles interactions of all orders.

From a practical point of view, it is of interest to relate the spatial structure with covariate information. Thus we often need to determine if and how a point pattern spatial distribution is associated with different types of covariates. Many times we want to determine whether the intensity of a point process is higher in areas where a certain feature of the population prevails.

Denote by $Z(x)$ the set of covariates for a spatial location x, and consider the intensity as a function of these covariates, $\lambda(x) = \rho\{Z(x)\}$. A more general formulation could be $\lambda(x) = \rho(x, Z(x))$, that is, the inhomogeneity is partially explained by the covariates. The function ρ describes the dependence between the intensity and the values of the covariates. A non-parametric estimate for ρ is proposed by Baddeley et al. (2012), providing a smoothing estimate of the intensity as a function of (continuous) spatial covariates. This method estimates $\rho(z)$ by the ratio between the (rescaled) density estimate obtained by smoothing the values of the covariates Z at the data points and the density estimate of the reference distribution of Z. This makes it possible to associate the intensity of the process with each covariate separately.

The advantage of parametric modelling using inhomogeneous Poisson processes is that it provides an easy method for analysing spatial point patterns via an intensity function depending on two or more covariates. This can be done by comparing summary statistics computed for simulations of the corresponding inhomogeneous Poisson processes with the intensity function depending on covariates, with the functions estimated from the initial point pattern.

2.2.2 Hybrid models for point process data

A spatial point pattern $\mathbf{x}$, a realisation of a point process X, is an unordered set $\mathbf{x} = \{x_1, \ldots, x_n\}, n \geq 0, x_i \in W$ of points x_i in a spatial 'window' $W \subset \mathbb{R}^d, d \geq 1$ (Baddeley et al., 2013). Descriptive analysis is required to learn about the basic properties of a spatial point pattern. With a view to finding further characteristics of the spatial structure and association between points, models for point patterns can be considered.

An important class of models is the class of Gibbs models, with many applications available in the literature (Ogata and Tanemura, 1984; Geyer, 1999; Daley and Vere-Jones, 2003; Baddeley et al., 2006; Renshaw et al., 2009; Comas and Mateu, 2011; Funwi-Gabga and Mateu, 2012; Diggle, 2014). These models are specified in terms of their *probability density.* The probability density function for a Poisson process with intensity 1 is $f(\mathbf{x}) = 1$. The uniform Poisson process with constant intensity $\lambda > 0$ has probability density $f(\mathbf{x}) = \alpha\lambda^{n(\mathbf{x})}$, where $n(\mathbf{x})$ is the number of points in the configuration $\mathbf{x}$ and α is the normalising constant, $\alpha = e^{(1-\lambda)|W|}$. We denote by h the *unnormalised* probability density.

Defining new functional forms for h is not trivial. Baddeley et al. (2013) propose techniques for defining new forms for h combining probability densities of known Gibbs models. In order to define a new class of models, they specify a different unnormalised probability density by multiplying the unnormalised probability densities $h_1, \ldots, h_n$, $n \geq 2$ of n models. A *hybrid* density $h(\mathbf{x}) = h_1(\mathbf{x}) \ldots h_n(\mathbf{x})$, or equivalently $\log h(\mathbf{x}) = \log h_1(\mathbf{x}) + \cdots + \log h_n(\mathbf{x})$, is obtained. Note that the likelihood of a hybrid is a product of component likelihoods, and the Papangelou conditional intensity of a hybrid is the product of the conditional intensities of the components. However, these product forms do not imply any kind of stochastic independence, since the usual factorisation lemma does not apply.

The unnormalised hybrid density has to satisfy certain properties. The functional form of h has to be *integrable* (its integral is finite); has to be *locally stable* (there is a finite constant B such that $h(x \cup \{u\}) \leq Bh(x)$, for all $x \in \mathbf{x}$ and $u \in W$); has to be *Ruelle stable* (there are finite constants A and M such that $h(x) \leq AM^{n(x)}$), meaning that h is dominated by an unnormalised Poisson density; and h has to be *hereditary*, or has to have *hereditary positivity* (for any configuration $\mathbf{x}$, $h(\mathbf{x}) > 0$ implies $h(\mathbf{y}) > 0$ for all sub-configurations $\mathbf{y} \subset \mathbf{x}$). For more details, see Ruelle (1969) and Møller and Waagepetersen (2004).

Gibbs models are generally applied to point processes with inhibitory patterns. Exceptions that fit aggregated models and are locally stable are the *Widom-Rowlinson penetrable sphere model* (Widom and Rowlinson, 1970), the *area-interaction processes* (Baddeley and van Lieshout, 1995), or the *Geyer saturation process* (Geyer, 1999). In this section we present and discuss the details of the *inhomogeneous Baddeley-Geyer hybrid model* (Baddeley et al., 2013).

We first define the unnormalised probability density for the stationary Geyer saturation process with parameters β, γ, r and s, given by

$$h_G(\mathbf{x}) = \beta^{n(\mathbf{x})} \prod_{i=1}^{n(\mathbf{x})} \gamma^{\min(s, t(x_i, \mathbf{x} \backslash x_i, r))} \tag{2.1}$$

where β controls the intensity of the process, $n(\mathbf{x})$ is the number of points in the pattern and γ is the interaction parameter. Furthermore, $t(x_i, \mathbf{x} \backslash x_i, r)$ is the number of neighbours of x_i in $\mathbf{x}$ within a radius r, that is, the number of points x_j with $j \neq i$ such that $||x_i - x_j|| \leq r$. The parameter $s > 0$ is a saturation threshold which ensures that each term in the product is bounded by γ^s, so that the density is integrable and Ruelle stable for all values of $\gamma > 0$. The process is clustered if $\gamma > 1$ and is regular when $\gamma < 1$.

Considering the hybrid of several Geyer densities of form (2.1), the unnormalised density of an inhomogeneous Baddeley-Geyer hybrid model is obtained by

$$h(\mathbf{x}) = \beta^{n(\mathbf{x})} \prod_{i=1}^{n(\mathbf{x})} \prod_{j=1}^{m} \gamma_j^{\min(s_j, t(x_i, \mathbf{x} \backslash x_i; r_j))},$$

where $r_1, r_2, \ldots, r_m$ are interaction ranges, $s_1, s_2, \ldots, s_m$ are saturation parameters, and $\gamma_1, \gamma_2, \ldots, \gamma_m$ are the interaction parameters.

We can introduce the effect of covariates by considering a local covariate effect $\beta(x_i) = \rho\{x_i, Z(x_i)\}$, where $Z(x_i)$ is a set of covariates for a spatial location $x_i \in \mathbf{x}$. The resulting probability density is

$$h(\mathbf{x}) = \prod_{i=1}^{n(\mathbf{x})} \beta(x_i) \prod_{j=1}^{m} \gamma_j^{\min(s_j, t(x_i, \mathbf{x} \backslash x_i, r_j))}, \tag{2.2}$$

that defines inhomogeneity in an elegant and straightforward way.

Estimating irregular parameters $r_1, \ldots, r_m$ and $s_1, \ldots, s_m$ in the inhomogeneous Baddeley-Geyer hybrid model raises an important issue on which very little statistical theory is available. Baddeley and Turner (2000) propose using profile pseudolikelihood. We use their technique to estimate the irregular parameters for the hybrid model.

2.2.3 Diagnostics

The next step in the analysis is to check if the model fits well and if each assumption of the model is appropriate. For a Poisson model, homogeneous or inhomogeneous, a χ^2 goodness-of-fit test based on quadrat counts (Cressie and Read, 1984) or a Berman test (Berman, 1986) can be applied. For hybrid models, no theory is available to support these tests. As an alternative, goodness-of-fit for hybrid models relies on the summary statistics functions. Baddeley et al. (2011) propose new tools for model validation. They suggest the use of residual G and K functions to compare and decide the best fitted model. In order to obtain these residual functions, we compute both the nonparametric estimate of the G function, and the one based on the corresponding model. Then the residual G function is obtained as the difference between the two measures. Likewise, we obtain the residual K-function. The residual G and K functions should be approximately zero if the model fits well. These functions provide a suitable diagnostic for the goodness-of-fit of a point process model.

Diagnostic plots based on residuals are another tool to measure the goodness-of-fit of a model and also to identify outliers in the data. These plots display the residuals from the fitted model. This diagnostic is followed up by Q-Q plots based on residuals from the model.

2.3 Spatial analysis for the varicella point pattern

2.3.1 The study area and the varicella dataset

The *varicella-zoster virus* (VZV) is a highly contagious virus spread worldwide which causes two clinical syndromes: *varicella*, also known as chickenpox, and *herpes zoster*, otherwise known as shingles (WHO, 2008). A case of varicella is defined as a sudden onset with moderate fever, minimum general symptoms and maculopapular skin rash for a few hours that transforms into vesicles for 3 or 4 days leaving granular crusts (Servicio de Vigilancia y Control Epidemiológico, 2007). The various phases coexist and injuries are superficial.

Varicella is transmitted from person to person by direct contact with the rash or inhalation of aerosolized droplets from respiratory tract secretions of patients with varicella. Varicella occurs worldwide and is a highly contagious human disease, no animal or insect source is known to exist. It is highly communicable and endemic in all countries worldwide. In temperate climates, at least 90 % of the population develops varicella by the age of fifteen and 95 % by the time they reach adulthood. Varicella is characterised by fever and a generalised vesicular rash, consisting of 200 to 500 lesions (European Centre for Disease Prevention and Control, 2014). The clinical course of varicella is generally mild in children. Adults may suffer from more severe symptoms and also have a higher risk of complications. Children infected with human immunodeficiency virus also may have severe, prolonged illness (Centers for Disease Control and Prevention, 2012). The disease may be fatal, especially in neonates and immunocompromised individuals. Complications include VZV-induced pneumonitis or encephalitis and invasive group A streptococcal infections. Following infection, the virus remains latent in neural ganglia; upon subsequent reactivation, VZV may cause zoster (shingles), a disease affecting mainly immunocompromised individuals and elderly people.

After recovery from varicella, patients usually have lifetime immunity. The first symptoms of varicella generally appear after a 10-21 days incubation period. It is characterised by an itchy, vesicular rash, fever and malaise. Varicella is generally self-limited and vesicles gradually develop crusts. It usually takes about 7 to 10 days for all the vesicles to dry out and for the crusts to disappear. This gives us a time period, from infection to completely dried vesicles, between 17 and 31 days. Most people with varicella make full recoveries, only 2-6 % of varicella cases develop complications (European Centre for Disease Prevention and Control, 2014).

Reported infection after household exposure ranges from 61%-100% (WHO, 2008; Gershon et al., 2008) which indicates small range interaction. The epidemiology of the disease is different in temperate and tropical climates. The reasons behind this behaviour may be related to climate, population density and risk of exposure (WHO, 2014; Health Department, 2014).

The diagnosis of varicella is primarily clinical. Laboratory tests are requested for more complicated cases or for epidemiological purposes. Recent studies have shown that varicella vaccination has influenced the incidence of the disease, decreasing the number, size, and duration of the outbreaks.

Varicella is widely spread around the world, with an annual incidence estimated at 80-90 million cases a year. Since 1997, Spain has been one of the European countries that reports every case of varicella. In Spain, chickenpox decreased during the time period 2005-2007 (Peña Rey et al., 2009).

Throughout the thesis we will apply different statistical methodology to the data set of varicella cases registered in Valencia, Spain, from 2008 to 2013, data gathered by the General Direction of Public Health, Epidemiological Area, Vigilance Service and Epidemiological Control, Generalitat Valenciana (Health Department, 2014).

Valencia is the capital of Autonomous Region of Valencia and the third largest city in Spain after Madrid and Barcelona, with approximately $800,000$ inhabitants in the administrative centre and an area of approximately 134 km^2 (Statistics Office, 2013). Valencia is located on the Mediterranean coast of Spain (39 °28'0"N, 0 °22'30"W). Figure 2.1 shows the location of the province of Valencia within Spain (left of the small subplot), the location of the city of Valencia within the Province of Valencia (right of the small subplot), and the city of Valencia itself (large subplot). Figure 2.2 (upper row) shows all the districts of Valencia, the centre of the city consisting of districts 1, 2 and 3. The majority of varicella cases were registered in districts 1 to 16, the remaining districts, 17, 18 and 19, are sparsely populated and far from the urban core, thus have been eliminated from all the analyses. Consequently, the study region corresponds to districts 1 to 16 of Valencia (see Figure 2.1, large subplot). Each of the 19 districts is divided in smaller areas called boroughs. Valencia has a total of 87 boroughs, 70 of them corresponding to districts 1 to 16. Figure 2.2 (lower row) shows the spatial distribution of all the boroughs.

As stated before, Valencia has a total population of about $800,000$ peoples living in the urban core. There are some general patterns of age distribution that can be observed. The city centre tends to be populated by an elderly community, whereas areas situated around the city centre and on the city's outskirts encompass a younger population. As an example of the age distribution, consider the Valencia population from 2010. Table 2.1 shows the percentage of population by age group living in the central area of the city in comparison with the ones living in the periphery. We can see that for the first two age groups the percentage of the population is higher in the periphery than in the central urban area. Table 2.2 displays a temporal evolution of the population in districts 1 to 16 and the number of children under 15 years from 2008 to 2013. We observe a decrease in population, as well as in the number of children under 15.

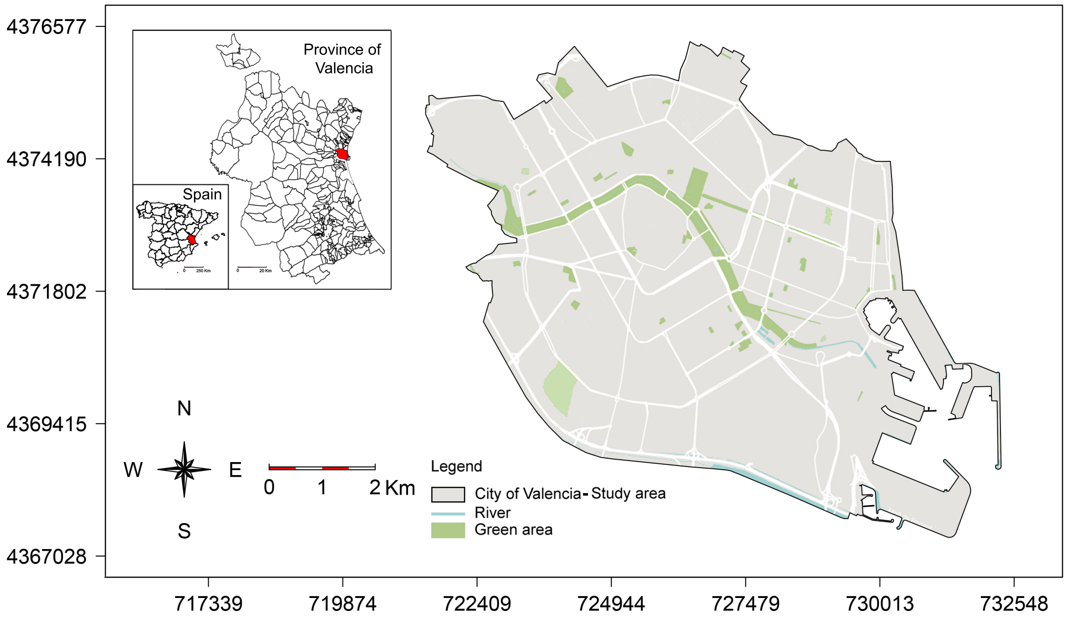

Figure 2.1. Location of the city of Valencia. Universal Transverse Mercator (UTM) coordinate system (distance in meters).

The dataset consists of varicella cases from 2008 to 2013, all residents in Valencia. There was a total of 11, 214 varicella cases registered during this time period of six years. Table 2.2 shows the temporal evolution of the total number of cases. We note the clear decrease in the number of cases from 2, 545 varicella cases observed in 2008, to 921 cases registered in 2013. The cases were recorded on a weekly basis.

	Age group									
	0-4	5-14	15-24	25-34	35-44	45-54	55-64	65-74	> 75	Total
Centre	5 571	9 504	10 252	18 069	19 996	16 560	13 374	11 519	15 049	119 894
	4.6%	*7.9%*	*8.6%*	*15.1%*	*16.7%*	*13.8%*	*11.2%*	*9.6%*	*12.6%*	
Periphery	32 580	57 859	66 509	106 944	109 093	91 593	73 720	56 386	54 433	649 117
	5.0%	*8.9%*	*10.2%*	*16.5%*	*16.8%*	*14.1%*	*11.4%*	*8.7%*	*8.4%*	
Valencia	38 151	67 363	76 761	125 013	129 089	108 153	87 094	67 905	69 482	769 011

Table 2.1. Population by age group in 2010.

Throughout the thesis, the varicella dataset will be used to apply different types of statistical analyses, from temporal to spatial and spatio-temporal analyses. Also, we will be studying the statistical properties of the varicella dataset as lattice data, aggregated into boroughs, and the underlying point process

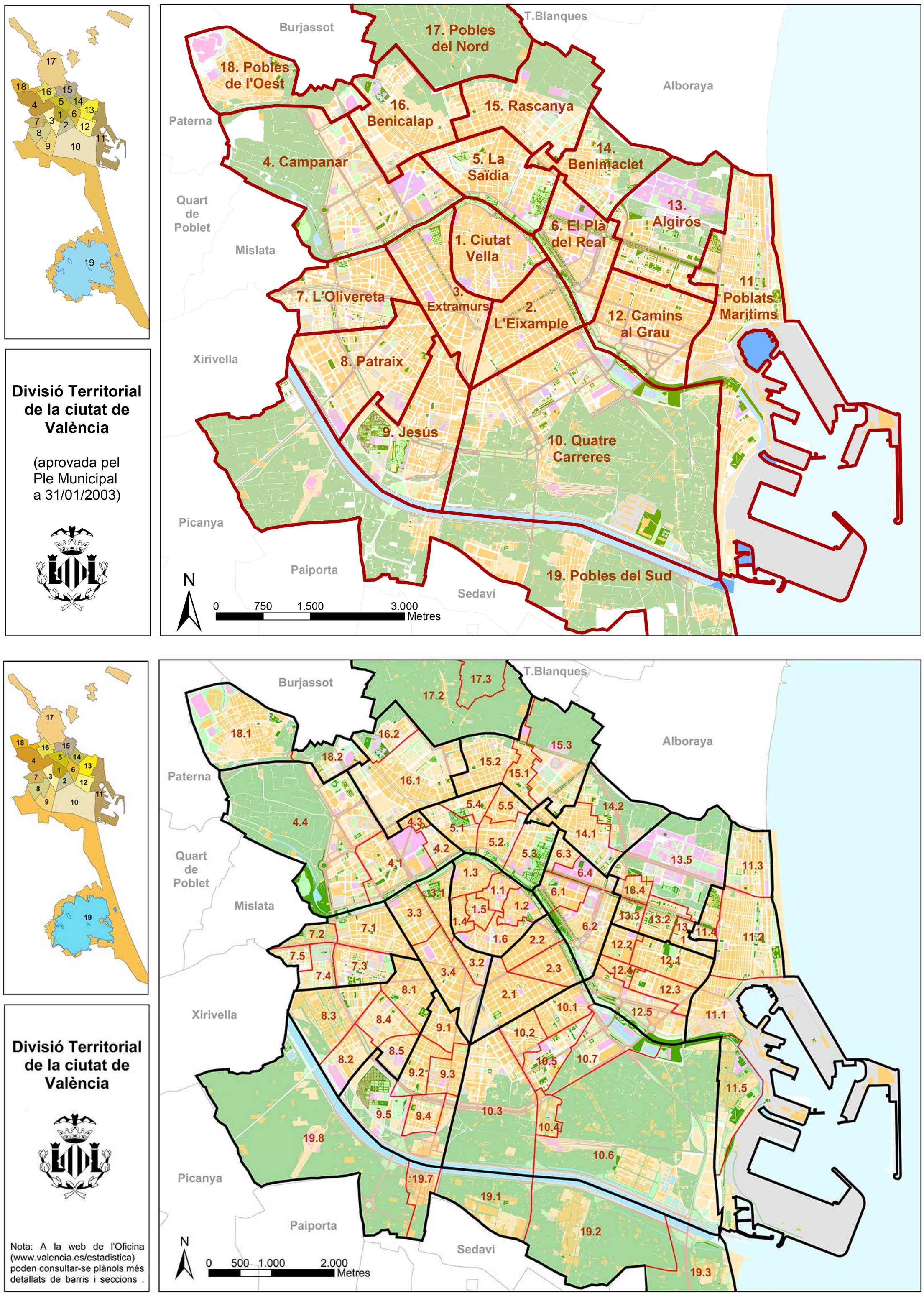

Figure 2.2. Valencia city map. (Top) Territorial division by districts. (Bottom) Territorial division by boroughs.

	2008	2009	2010	2011	2012	2013
Population	768,790	773,856	769,011	759,211	757,938	752,992
Population under 15	116,929	118,367	118,471	118,144	118,487	111,765
Varicella cases	2,545	2,152	2,438	1,898	1,260	921

Table 2.2. Evolution of the Valencia population (districts 1 to 16) and the number of varicella cases.

2.3.2 Descriptive analysis

With a view to apply spatial statistical methods for point patterns, we consider a subset of the varicella dataset we presented in Section 2.3.1. An important issue that this dataset emphasises is the presence of multiple points, which is caused due to an inherent discrete assignment in observations. The point pattern is constructed by considering the address of residence for each case, in particular, the building where a case resides. Therefore, one point (building) can have multiple cases assigned. There are different solutions for this problem. One can jitter all points by adding a small amount of noise to all coordinates. This way we remove all duplicated locations. Another alternative is to assign to each location only one varicella case, meaning, consider the locations without taking into account the number of cases observed at the specific location. The newly obtained point pattern is called the *ground* point pattern.

It is well known that varicella is primarily a disease of children. Therefore, it seems reasonable to analyse only children under the age of 14. Considering all the above, the subset we are analysing throughout this chapter consists of the ground point pattern of all varicella cases (children under the age of 14) registered during 2013, in Valencia. We need to mention that the analysis can be implemented similarly for any of the remaining years.

Figure 2.3 (left) shows the spatial point pattern of the varicella cases together with the quadrat counting of the points. The pattern of varicella is clearly not randomly scattered, with areas with a higher number of cases than the average. Figure 2.3 (right) shows the same amount of spatial points as the varicella cases uniformly distributed in the region. This figure is a visual confirmation that the varicella pattern is not completely spatially random.

Figure 2.4 (left) shows the empirical G function for the point pattern (solid line) and the point-wise minimum and maximum envelopes obtained from 99 simulations of a process under the CSR null hypothesis. Likewise, Figure 2.4 (right) shows the estimated K-function for the point pattern (solid line) and the point-wise minimum and maximum envelopes obtained by simulating a Poisson process 99 times. The hypothesis of complete spatial randomness is clearly rejected.

As stated before, a rejection of the CSR hypothesis is an indicator that our process can be an inhomogeneous Poisson process, a cluster process, or even a mixed process.

The intensity of the spatial point pattern has been estimated non-parametrically using a Gaussian kernel intensity estimator. This method is sensitive to bandwidth selection. Figure 2.5 (left) shows the kernel estimate for the bandwidth obtained using

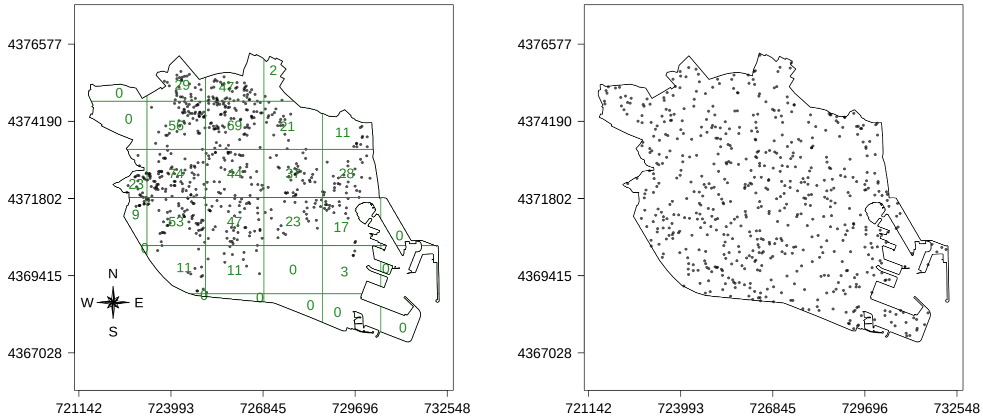

Figure 2.3. (Left) Spatial point pattern of varicella cases registered during 2013 in Valencia (Spain), together with quadrat plot for the point pattern. (Right) Point pattern of independent uniform random points.

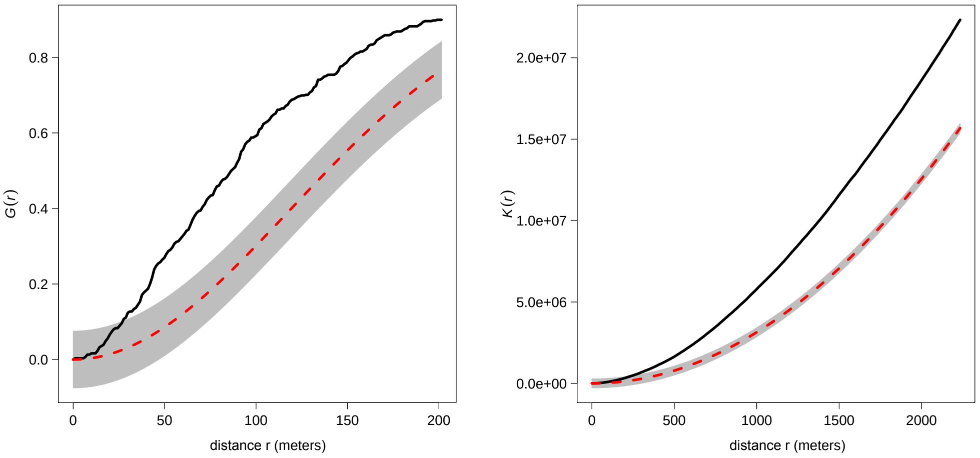

Figure 2.4. (Left) Empirical G function (solid line) with pointwise maximum and minimum envelopes obtained under CSR null hypothesis (grey area). (Right) K-function (solid line) with pointwise maximum and minimum envelopes obtained under CSR null hypothesis (grey area).

a cross-validation method, as proposed by Diggle (1985). Figure 2.5 (right) shows the intensity estimate for a bandwidth equal to 300. The latter provides a smoother estimate, whereas the former gives a more uneven estimate.

Both estimates in Figure 2.5 show several areas where varicella incidence is high: one in the north of Valencia, corresponding to the *Rascanya* district and a second one in the west in the *l'Olivereta* district. The population living in these areas generally tends to have a lower income level in comparison with the central areas of the city.

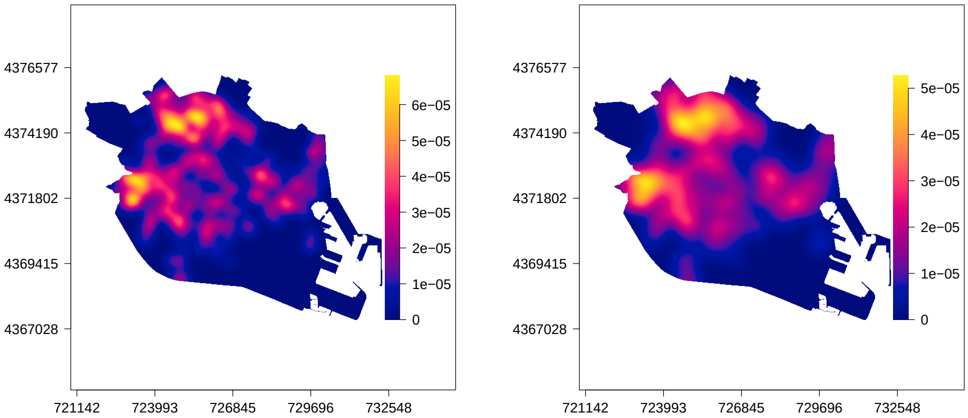

Figure 2.5. Kernel smoothing estimate of the varicella point pattern intensity. (Left) Bandwidth = 153 (Diggle, 1985). (Right) Bandwidth = 300.

Figure 2.6 (left) shows the residual G function with the corresponding estimate obtained under the hypothesis of inhomogeneity (the log-intensity depends on a linear combination of the coordinates of the point locations). Likewise, Figure 2.6 shows the residual K-function together with the corresponding envelopes. The irregular shape of the G function indicates some type of aggregation in the data.

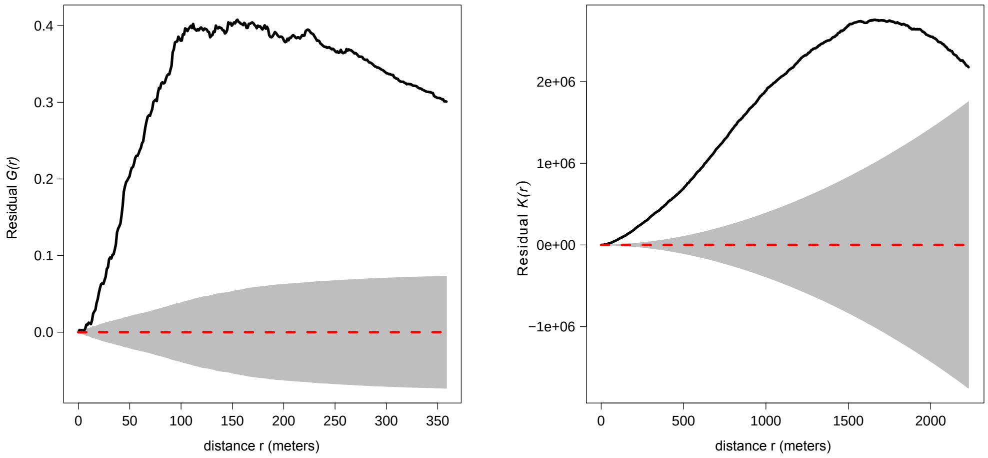

Figure 2.6. (Left) Residual G function (solid line). (Right) Residual K-function (solid line); and the corresponding envelopes obtained under the inhomogeneity assumption previously described (grey area).

The next step in the analysis is to assess the interaction between the varicella pattern and external factors. Almost all types of diseases, and varicella in particular, depend on the population. Whether it is a contagious disease or not, people living in a determined area are the most important factor in initiating and transmitting a disease. Regions with a high population, tend to register more cases, and scarcely populated areas tend to have lower disease incidence. In order to make this correction, we consider the density of population under 14 years as a covariate. Figure 2.7 (left) shows a kernel estimate of this population density in the city of Valencia for the year 2013. We can see that the spatial distribution of the population shows some sort of a pattern. The centre and the outskirts of the city are areas where the density of population is low, whereas the neighbourhoods surrounding the centre of the city are the most preferred to live in.

The location of the schools in Valencia can play a significant role in the analysis of the disease. Figure 2.8 (right) shows the pattern of all daycare centres, preschools and schools in Valencia. Varicella is primarily a children's disease, and almost all children either attend daycare centres, preschool or elementary school. Thus, the spatial point pattern in Figure 2.8 is considered a major factor in the clustering of varicella cases.

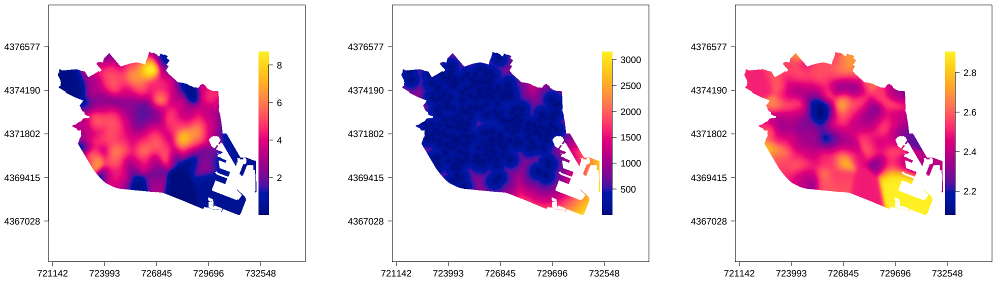

Figure 2.7. (Left) Kernel estimate of population density under 14 years, 2013. (Centre) Kernel estimate of the distance function to the nearest school in Valencia. (Right) Kernel estimate of the average number of persons per family register.

We also consider the distance to the nearest school as a covariate. The distance function of a set of locations (in our case the schools), $\mathbf{s} = \{s_1, \ldots, s_m\}, m > 0, s_i \in W$, where W is the study region, is a mathematical function such that, for any spatial location x the function value is the shortest euclidean distance from x to $\mathbf{s}$. More specifically,

$$dist(x) = \min \{||x - s_i||\}, \ i = 1, \ldots, m, \ x \in W. \tag{2.3}$$

Figure 2.7 (centre) shows a graphical representation of the distance function defined by (2.3). Areas, where the distance function has low values, match with neighbourhoods with an elevated number of schools. Similarly, areas with high values of the distance function are regions with fewer or no schools.

Estimating the school distance function can explain the relationship between the disease and the pattern of schools. Alternatively, the schools can be considered *focal* *points*. The point pattern of all schools in Figure 2.8 shows areas where points tend to cluster. The `SatScan` software (Kulldorff, 2010) is used to identify main clusters of schools. A summary of the main identified clusters is presented in the left-hand-side table of Figure 2.8. On the right-hand side of Figure 2.8 we see the spatial location of all schools as black dots, together with the main clusters detected by the software, encircled and labelled with an X. These six points can be interpreted as focal points (sources) around which varicella cases tend to group. For each source, we calculate the distance function in equation (2.3).

ID	Northing (m)	Easting (m)	p-value
1	4374194.64	726012.70	2.22E-06
2	4374257.90	723723.00	4.16E-06
3	4372123.33	728467.68	1.30E-02
4	4371909.12	724210.03	2.50E-02
5	4370990.51	726351.50	4.59E-01
6	4369998.62	724372.88	4.59E-01

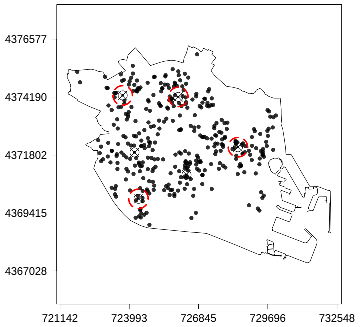

Figure 2.8. (Left) Coordinates and corresponding p-values for the six main school clusters detected using `SatScan` software. (Right) Pattern of all schools in Valencia (black dots), together with the spatial locations of the main clusters (dots encircled and labelled with an X).

Another factor taken into account is the composition of the families. Figure 2.7 (right) shows a kernel estimate of the average number of persons per family register, meaning the average number of persons a family consists of.

As stated before, the population density in Figure 2.7 (left) shows the centre and the peripheral areas of the city having less population than the areas surrounding the central neighbours of Valencia. In the north of Valencia, there is also an area where the population density is the highest. Figure 2.7 (right) shows a central area where the number of persons per family is low. This intensity increases as we move toward the peripheral neighbours of the city.

As explained before, the intensity can be considered as a function of the covariates, $\lambda(x) = \rho\{Z(x)\}$. Figure 2.9 shows the non-parametric estimate of $\lambda(x)$ for each individual covariate together with point-wise two-standard-deviation confidence limits (grey shading) (Baddeley et al., 2012). A clear relationship between the intensity of the process and the three covariates can be observed. For areas where the population density is low, the intensity of varicella cases is also low. Areas with high-density population are more likely to have a higher relative risk of varicella. For the distance function defined in Equation 2.3, Figure 2.9 (centre) shows a linear relation. The larger the distance from schools, the lower the incidence of varicella observed. Figure 2.9 (right) shows the high intensity of the relative risk of varicella in areas where small families live. As the number of family members increases, the risk of varicella decreases. This unusual and unexpected behaviour will be further discussed.

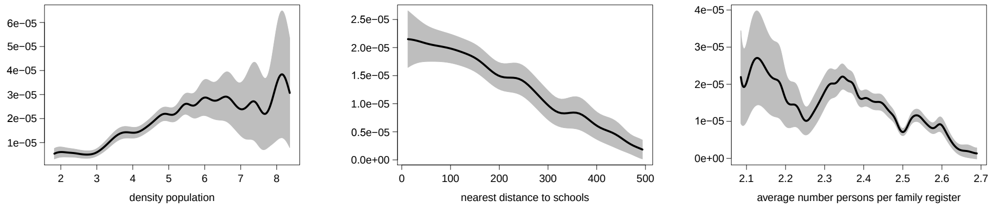

Figure 2.9. Non-parametric smoothed estimation of the intensity as a function of spatial covariates. (Left) Population density under 14 years, 2013. (Centre) nearest distance to school in Valencia. (Right) Average number of persons per family register.

The non-parametric estimates shown in Figure 2.9 could indicate that varicella risk is higher in neighbourhoods where families have a medium-low cultural and socio-economic status. We fit an inhomogeneous Poisson process where the log-intensity is a linear function of the three covariates shown in Figure 2.7, and use the diagnostics described in Section 2.2.3 to assess the fitted model. The coordinates of the spatial locations, x and y, were not considered in this model, as they do not improve the overall fitting.

	Estimate	S.E.	CI95.lo	CI95.hi	Ztest	Zval
(Intercept)	-4.11e+00	9.25e-01	-5.92e+00	-2.29e+00	***	-4.44e+00
population density	4.05e-01	3.61e-02	3.34e-01	4.76e-01	***	1.12e+01
distance function	-2.55e-03	3.68e-04	-3.27e-03	-1.83e-03	***	-6.94e+00
family register	-3.40e+00	4.04e-01	-4.19e+00	-2.61e+00	***	-8.41e+00

Table 2.3. Estimated parameters for the inhomogeneous Poisson model.

Table 2.3 shows the estimated parameters for the inhomogeneous Poisson process. All three covariates are significant and, as expected, population density has a positive effect on the overall risk. The sign for the second covariate shows an inverse relation between disease risk and the distance function. The parameter of the last covariate also shows an inverse association between varicella risk and the family register, as shown in Figure 2.9 (right). This would mean that for regions where large families live, varicella incidence is low. This seems to contrast with the opinion of epidemiologists and will be further discussed in Section 2.4.

Figure 2.10 shows the residual G and K functions for the inhomogeneous Poisson process with intensity depending on the covariates. The sharpness and the peak of the residual G function in Figure 2.10 (left) evidences 'unexplained' interaction at different spatial scales, between 75 and 125 meters. The residual K-function shows that the model does not fit well.

Figure 2.11 shows the cumulative raw errors for the inhomogeneous Poisson model. The size of the errors is considerable, and they have a particular shape.

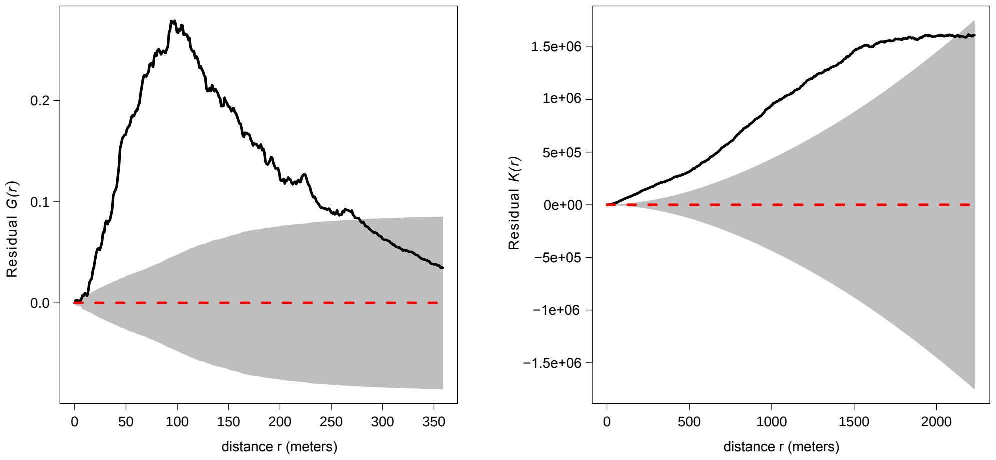

Figure 2.10. (Left) Residual G function for the inhomogeneous Poisson process. (Right) Residual K-function for the inhomogeneous Poisson process.

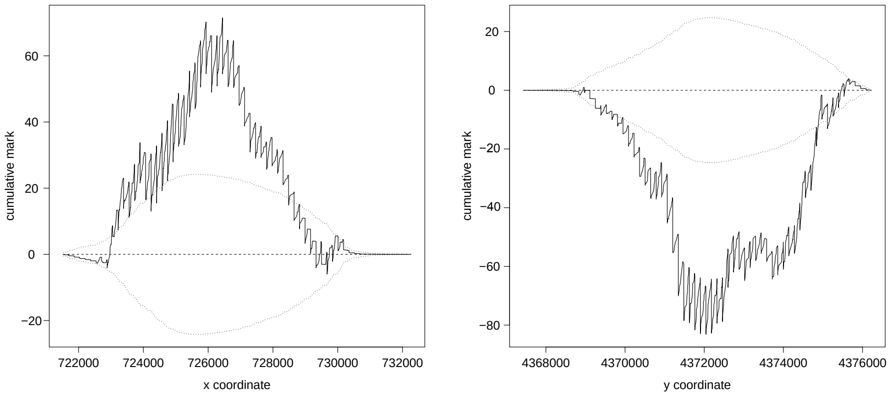

Figure 2.11. Cumulative raw residuals for the x and y axis.

2.3.3 Inhomogeneous Baddeley–Geyer hybrid model

To explain the multi-scale spatial interaction, we considered a hybrid Baddeley-Geyer process depending on the three covariates. We then added all subsets of focal points as covariates to this initial model and compared the results, using model diagnostics. The best model for the data includes all three covariates and four of the six focal points. These four focal points are circled in red in Figure 2.8.

We selected the irregular parameters, using maximum profile pseudolikelihood. We searched over a wide range of parameters $1 \leq r \leq 400$ and $s = 1, \ldots, 20$ and identified three significant spatial interaction components. The optimum combination is obtained for distances $r_0 = 47$, $r_1 = 104$, $r_2 = 286$ meters and saturation parameters $s_0 = 1$,

$s_1 = 3$, $s_2 = 2$.

	Estimate	S.E.	CI95.lo	CI95.hi	Ztest	Zval
(Intercept)	-7.55E+00	1.28E+00	-1.01E+01	-5.04E+00	***	-5.89E+00
population density	3.21E-01	4.71E-02	2.28E-01	4.13E-01	***	6.81E+00
distance function	-1.08E-03	5.05E-04	-2.07E-03	-9.20E-05	*	-2.14E+00
family register	-2.54E+00	5.85E-01	-3.69E+00	-1.39E+00	***	-4,34E+00
focal point 1	-8.12E-05	7.28E-05	-2.24E-04	6.14E-05		-1.12E+00
focal point 2	-4.82E-05	9.54E-05	-2.35E-04	1.39E-04		-5.05E-01
focal point 3	7.84E-05	9.34E-05	-1.05E-04	2.61E-04		8.40E-01
focal point 4	5.48E-05	5.74E-05	-5.78E-05	1.67E-04		9.54E-01
β_1	2.66E-01	1.01E-01	6.78E-02	4.65E-01	**	2.63E+00
β_2	2.42E-01	4.16E-02	1.60E-01	3.24E-01	***	5.81E+00
β_3	4.97E-01	2.03E-01	9.89E-02	8.94E-01	*	2.45E+00
γ_1	1.305					
γ_2	1.273					
γ_3	1.164					

Table 2.4. The estimated parameters for the Baddeley-Geyer hybrid model.

To determine the accuracy of the fitted model, we used three complementary methods. First, we looked at the residual G and K functions. Second, we used diagnostic plots to study the size of the errors. Finally, we simulated a point pattern from the fitted model and compared it with the empirical initial pattern.

The interaction parameters for the hybrid model are $\gamma_1 = 1.305$, $\gamma_2 = 1.273$, and $\gamma_3 = 1.643$, which confirm that the point pattern shows spatial aggregation at different spatial scales. Table 2.4 shows the estimates of the parameters for the fitted model. The positive coefficient of the population density shows a positive effect of population density on the incidence of varicella. The negative coefficient of the family register remains an unusual concern.

To assess the fit of the model, we used the residual G and K functions. Figure 2.12 shows a comparison between these two functions for the inhomogeneous Poisson model and the hybrid model. We see a definite improvement for the latter model. The values of the two functions oscillate around zero and are inside the envelopes, which indicates the model fits well. Using the same scales as in Figure 2.11, Figure 2.13 shows a diagnostic for the residuals of the fitted model. The range of the raw residuals in Figure 2.13 has substantially decreased compared to the residuals of the inhomogeneous Poisson process. The qqplot in Figure 2.13 shows that the errors are approximately normal.

Figure 2.14 shows a comparison between the initial point pattern of varicella and a simulation of the fitted hybrid model. We can see that the simulated pattern presents areas with more points than the initial point pattern, but overall, it does correspond quite well to the data pattern.

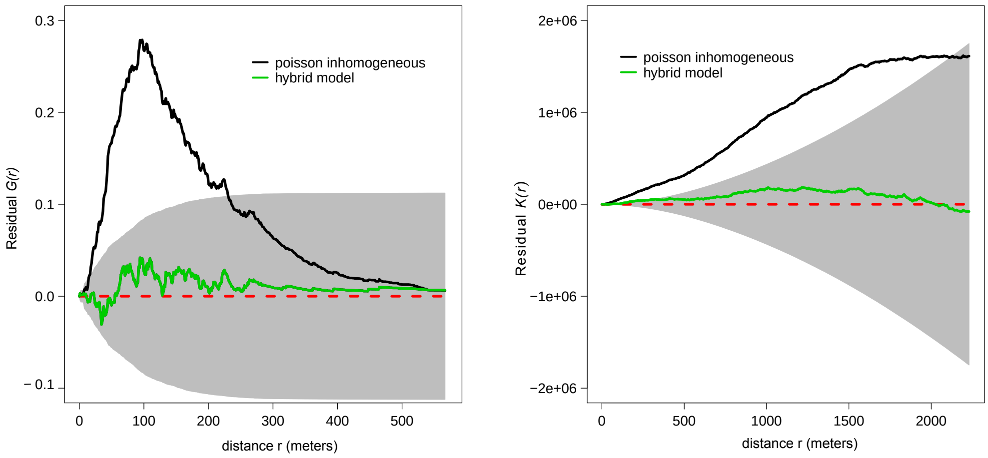

Figure 2.12. (Left) Residual G function for the hybrid model. (Right) Residual K-function for the hybrid model.

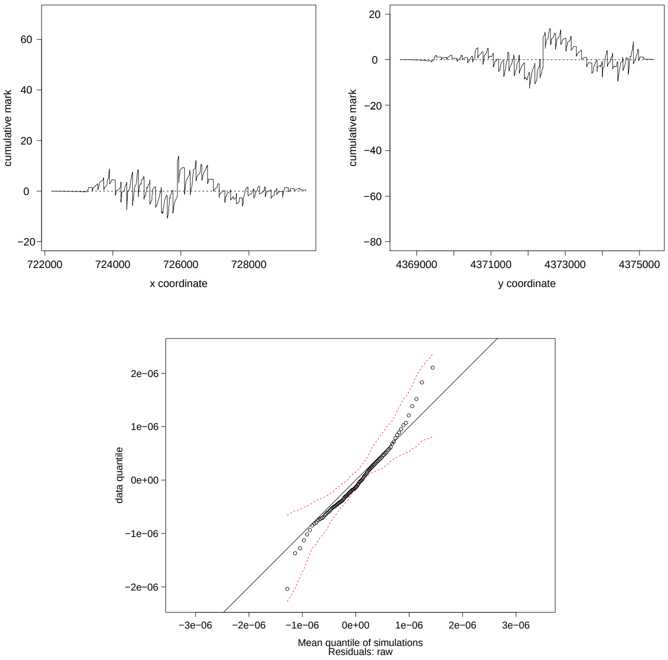

Figure 2.13. (Top) Cumulative raw residuals for the x and y axis for the hybrid model. (Bottom) Qqplot for the raw residuals for the hybrid model.

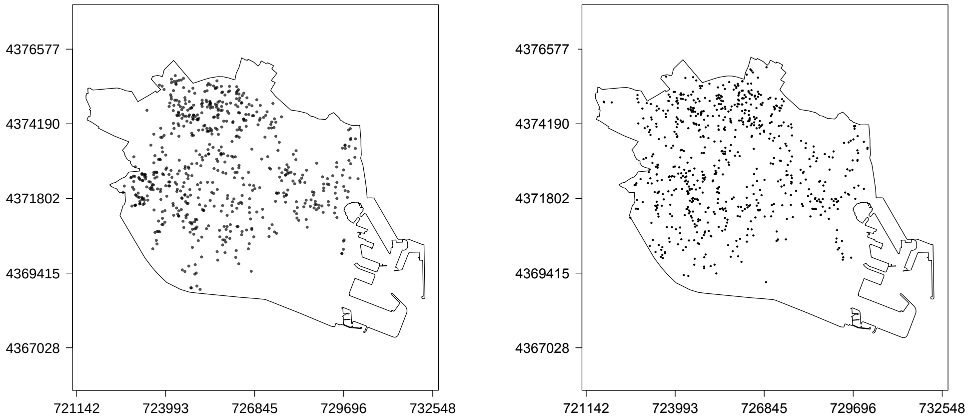

Figure 2.14. (Left) Spatial distribution of varicella point pattern in Valencia, Spain, cases registered during 2013. (Right) Simulation from the hybrid model.

2.4 Conclusions and discussion

This chapter presents an epidemiological application of hybrid of Gibbs models. The novelty of this study is the actual use of the statistical methods in the analysis of varicella data. The study considers intensity estimation including covariate information. This analysis was designed to find specific characteristics and relations between varicella incidence and the school point pattern. We conclude that there is a significant relation between the location of schools and possible varicella outbreaks. The final fitted model explains the behaviour of the disease and identifies aggregation at different scales: approximately 50, 100, and 290 meters. Equation (2.2) shows that every spatial scale r_j is bounded to its corresponding saturation parameter s_j, thus making the interpretation of these parameters more difficult.

We constructed the point pattern by considering the addresses of residence for each case. As a result, we obtained points with multiple cases which would lead us to a marked point process. We have studied the ground process, meaning the locations where the cases were observed, without the marks, as a first step toward discovering statistical properties of the pattern. The next natural step is to explore the properties of the marked point process. This is one of the main focuses of our future work.

The final fitted model includes, besides the distance to the nearest school, the focal points. We wanted to analyse the model without taking into consideration any focal point. The reason we did this is, partly, because we already used school information when we introduced the nearest-school distance function. We thought it was somehow redundant to use this information again, by considering the focal points. Figure 2.15 shows the representation of the errors, using the same scale as in Figure 2.11 and 2.13 and a qqplot for the hybrid model without the focal points. If we take a careful look, the shape of the errors changes, with certain areas of the study region exhibiting a clearly pronounced shape and high values. The qqplot shows that the errors are not normal. All this indicates that the inclusion of focal points helps to correct the unusual behaviour of the resulting residuals. Another way to verify this is to compute the cross K-function between the pattern of varicella cases and the pattern obtained considering the six focal points. Figure 2.16 shows the cross K-function estimate (solid line) and the envelopes simulated assuming a stationary (spatially homogeneous) random spatial point process (grey area). As this figure illustrates, the cases are not independently distributed with respect to the focal points.

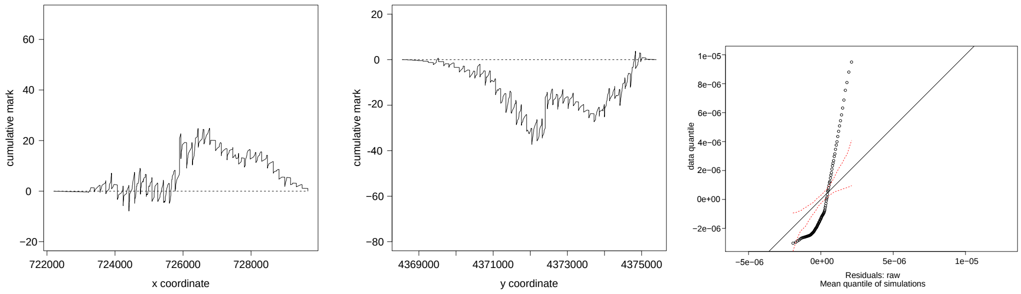

Figure 2.15. Cumulative raw residuals for the x and y axis; qqplot for the raw residuals. Hybrid model without the focal points.

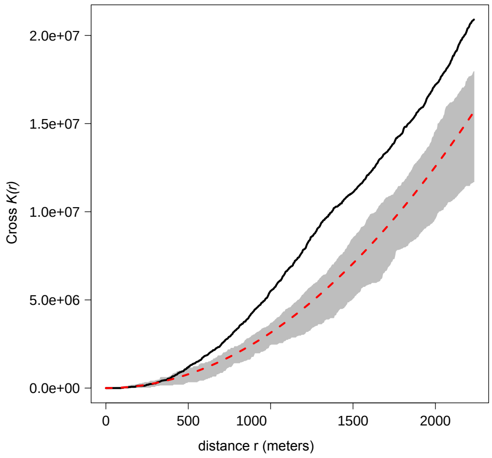

Figure 2.16. Cross K-function between the varicella and schools patterns.

Our fitted model explains one interesting feature that has been missed by other more epidemiological approaches and therefore adds value to our research. The intensity of incidence of varicella is low at peripheral areas even though families with a large number of children live there. Let us compare Figure 2.5 (right) and Figure 2.7 (right), the varicella point pattern estimate and the ‘average number of persons per family register’ estimate. On one hand, we observe that for centrally situated areas (areas with an older population) the average number of persons per family is low. As expected, the intensity of the process in these areas is also low. On the other hand, as stated before, the peripheral areas of the city correspond to low varicella incidence and large families. This seems odd, but a deeper and careful analysis indicates that these peripheral areas of the city are less populated, so less movement of people is expected, and the risk of varicella being contagious is reduced. In addition, children living in large families are quickly infected and become immune faster than those living in families with a reduced number of members.

We have been able to identify such interactions between the incidence of varicella cases and some available covariates. However, other covariates that have not been taken into account here could also play a role. For example, some economic and demographic variables could influence the risk of incidence.

Chapter 3

Methods for statistical analysis of space-time lattice data

This thesis combines two epidemiological applications and two methodological proposals. This chapter presents a second application, putting forward an analysis of aggregated spatio-temporal data. This chapter represents an analysis based on different statistical techniques used to analyse the varicella dataset introduced in Chapter 2, Section 2.3.1. In addition to the spatial analysis in Chapter 2 where the spatial component of the process in analysed, this chapter takes into account the temporal component in the dataset. We aggregate the data into small areas (boroughs) and consider the yearly evolution of the lattice data. The main objective in this chapter is to find a relation between the spatial and temporal components of the process and confirm it by means of models. This chapter proposes an exploratory space-time data analysis. Also, space-time parametric and non-parametric models are applied to the data to identify the behaviour of the underlying process. In Section 3.1 we take a look at an overview of some statistical models available in the literature. Then we introduce the tools for exploratory space-time data analysis, as proposed by Rey (2014) and space-time models. Exploratory space-time data analysis methods are applied to the varicella dataset, together with space-time parametric and non-parametric models, as proposed by Knorr-Held (2000).

3.1 Introduction

Parametric and non-parametric models with different types of interaction for analysing space-time data have been proposed in the literature. Bernardinelli et al. (1995) proposes a parametric space-time model assuming a linear time trend. The paper by Knorr-Held (2000) suggests a model that combines the spatial model by Besag et al. (1991) with dynamic models where the temporal trends are assumed to be non-linear and non-stationary. In particular, the time-changing effects can be seen as the temporal analogue of the spatially structured components in the Besag et al. (1991) model. This model allows for space-time interactions where four types of interactions arise naturally as the product of one of the two spatial effects with one of the two temporal main effects.

Markov Chain Monte Carlo (MCMC) methods have given the community of users a great tool to analyse these types of models for complex data. Recently, an approximation method for Bayesian inference, Integrated Nested Laplace Approximations (INLA) (Rue et al., 2009), has been developed. It represents a very useful tool for reducing computation time, providing Bayesian inference by means of the user-friendly library R-INLA.

A broad range of applications have implemented parametric and non-parametric models to detect particular space-time behaviour of patterns. Bernardinelli et al. (1995) illustrate the use of parametric models by analysing the cumulative prevalence of insulin-dependent diabetes mellitus in data gathered from all 18-year-old military recruits in Sardinia during 1936-1971. Ugarte et al. (2009) present an evaluation of the performance of the parametric and non-parametric space-time Bayesian models with a simulation study and an application for mortality data due to colorectal cancer in males from Navarra, Spain, corresponding to four 5-year time windows. Schrödle and Held (2011a) apply space-time models using INLA to cases of coxiellosis among Swiss cows from 2005 to 2008. Different extensions of the parametric and non-parametric models have been proposed by Schrödle and Held (2011b) and applied to counts of salmonellosis in cattle reported to the Swiss Federal Veterinary Office from 1991 to 2008. Schrödle et al. (2011) propose another application for these models to analyse reported cases of bovine viral diarrhoea in Swiss cattle from 2003 to 2007. Blangiardo et al. (2013) study the pattern of low birth weight for the 159 counties of US state of Georgia from 2000 to 2010, using the parametric formulation in Bernardinelli et al. (1995) and the non-parametric formulation in Knorr-Held (2000) with INLA.

This chapter presents a study of the spatial and temporal behaviour of varicella, divided into two parts, an exploratory space-time data analysis (ESTDA) and an application of space-time models.

3.2 Methods

3.2.1 Standardisation of rates

In epidemiology, the incidence rate of a disease is defined as the total number of observed cases (individuals who change their status, going from being healthy to sick patients) divided by the total number of people at risk (population at risk), both measured at the same moment in time. Most diseases are affected by a number of demographic factors such as age, sex, educational or socio-economic status. As a result, other variables have a direct or indirect impact on the calculation of incidence rates.

Standardisation (and other adjustment procedures) is used to compare two or more rates after eliminating the structural effects, leaving us with the residual effect, a good indicator of disease incidence among populations. Age and sex are two of the most common variables used for standardisation.

Indirect standardisation uses the age- and sex-specific rates for a standard population and applies them to the population distribution in the areas of study to calculate the expected number of cases.

Internal standardisation calculates the rates, using the information provided by the observed data.

Given the time component in the data, standardisation must be performed according to the population living in the study area at any given moment in time. For example, if we are handling daily, weekly or even monthly data, the standard population must correspond to the annual population (if there are no significant shifts in the population from one period of time to another).

Below, we exemplify internal standardisation for yearly observed data. We denote by O_l the number of observed cases for a reference year, i_0, and for the age and sex group $l = 1, \ldots, L$. We calculate the rate τ_l by dividing the number of cases O_l by the population at risk from the age and sex group l for the year i_0, denoted by P_l:

$$\tau_l = \frac{O_l}{P_l}.$$

To calculate the number of expected cases, we multiply τ_l by the population, P_{lk} in each area $k = 1, \ldots, K$ corresponding to the age and sex group l and the corresponding year:

$$E_k = \tau_l P_{lk}$$

In matrix form, we assume a $1 \times L$ vector for every year in the study of the rate. We multiply it with an $L \times K$ matrix corresponding to the population by age and sex group and area of study, obtaining a $1 \times K$ vector

$$\begin{bmatrix} \tau_{11} & \tau_{12} & \ldots & \tau_{1L} \end{bmatrix}_{1\times L} \begin{bmatrix} P_{11} & P_{12} & \ldots & P_{1K} \\ P_{21} & P_{22} & \ldots & P_{2K} \\ P_{31} & P_{32} & \ldots & P_{3K} \\ \ldots & \ldots & \ldots & \ldots \\ P_{L1} & P_{L2} & \ldots & P_{LK} \end{bmatrix}_{L\times K} = \begin{bmatrix} E_{11} & E_{12} & \ldots & E_{1K} \end{bmatrix}_{1\times K} \tag{3.1}$$

where $E_{1k} = \sum_{j=1}^{J} \tau_{1j} P_{jk}$.

Dividing the number of observed cases by the expected cases, we obtain the standardised ratio (SR). The areas with SR more than 1 are those where the number of observed cases is higher than would be expected if they behaved like the standard population; by contrast, in areas with SR less than 1, the number of observed cases is lower than expected.

3.2.2 Exploratory space-time data analysis

Spatial and spatio-temporal data exploration requires the setting of a neighbourhood structure between geographic units that represents the configuration of the analysed region. The units are usually administrative divisions established on a given territory.

The neighbourhood structure depends on the criteria used to define the concept of neighbour. If we define *neighbours* as the administrative units that share a border, an approach that seems best suited for a grid (lattice) as irregular as the one of the different boroughs of the city of Valencia, the contiguity matrix, W, has the following form:

$$w_{ij} = \begin{cases} 0, & i = j = 1, \ldots, n; \\ 1/n_i, & \text{if } j \in V(i), \text{ with } n_i = \#V(i); \\ 0, & \text{if } j \notin V(i), \end{cases}$$

where i and j represent whichever two of the n boroughs and $V(i)$ the set of neighbours of i. With this structure, no borough is its own neighbour, and the values in each row sum to unity because the weights w_{ij} are standardised. For other neighbouring structures, see Cliff and Ord (1973).

3.2.2.1 Moran index and Moran scatter plot

The first analysis of the geographical location data is to determine the existence of spatial autocorrelation. One of the most popular measures is Moran's index (Moran, 1950a,b), which allows for the measurement of spatial autocorrelation between observations of a variable X. Its most general expression for a non-standardised weight matrix W is

$$I = \frac{n \sum_i \sum_j w_{ij}(x_i - \bar{x})(x_j - \bar{x})}{S_0 \sum_i (x_i - \bar{x})^2}, \tag{3.2}$$

where $S_0 = \sum_{ij} w_{ij}$. If, as in our case, $w_{ij} = 1/n_i$, then $S_0 = n$ and (3.2) takes the following form:

$$I = \frac{\sum_i \sum_{j \in V(i)} \frac{1}{n_i}(x_i - \bar{x})(x_j - \bar{x})}{\sum_i (x_i - \bar{x})^2}. \tag{3.3}$$

Under the hypothesis of randomness in the spatial distribution of the observed values of X and asymptotic normality, its expectation $E(I) = -1/(n-1)$ and its variance can be obtained. This allows for the construction of a hypothesis test to check for spatial autocorrelation. However, there is an alternative Monte Carlo test based on random permutations which avoids the problem of asymptotic normality. Both tests are implemented in the `spdep` (Bivand, 2014) package in `R` (R Core Team, 2014).

Expression (3.3) gives us an interpretation of I in terms of a regression coefficient. It can be rewritten as

$$I = \frac{\sum_i \sum_{j \in V(i)} \frac{1}{n_i}(x_i - \bar{x})(x_j - \bar{x})}{\sum_i (x_i - \bar{x})^2} = \frac{\sum_i (x_i - \bar{x}) \left[\frac{1}{n_i} \sum_{j \in V(i)} (x_j - \bar{x})\right]}{\sum_i (x_i - \bar{x})^2} = \frac{cov(x, y)}{var(x)}, \tag{3.4}$$

where y represents the average of the values of x in the neighbourhood of i, also called first-order *spatial lag* because it refers to the neighbours sharing borders, which are at distance 1. Unequivocally, I is the regression coefficient of Y on X and represents the average spatial correlation of the data. Figure 3.1 shows a simulated example of this type of regression.

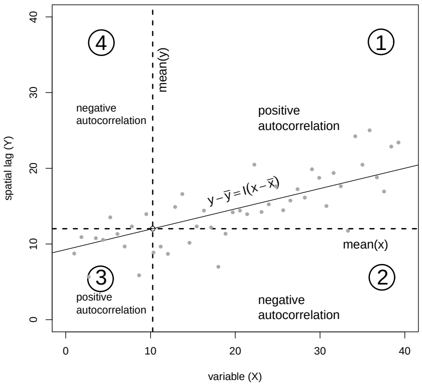

Figure 3.1. Moran scatter plot for a variable X and its spatial lag Y.

If there is no association between the values of x, the average of its neighbourhood does not vary systematically with it. However, if there is a positive association, high or low values of x surrounded by similar values, pairs of (x_i, y_i) would appear predominantly in two of the quadrants into which the lines $y = \bar{y}$ and $x = \bar{x}$ divide the plane, specifically in quadrants 1 and 3 as in the example shown in the graph. Each quadrant has an associated autocorrelation type as indicated in the figure. Such graphs are called *Moran scatter plots*. In classic notation, if L and H denote, respectively, lower or higher values than the mean, the four quadrants of Figure 3.1 are denoted by the following combinations HH, HL, LL and LH, where the first letter refers to the value of X and the second to the average of its neighbours.

3.2.2.2 Directional Moran scatter plot

Moran's index only provides spatial information on the behaviour of the variable. Since we have information over time, our interest lies not so much in the static exploration shown in Figure 3.1 but in the comparison of similar figures obtained for different timeframes. This comparison provides an overview of the spatial dynamics but can mask individual developments. The graphical part of this procedure is called *directional Moran scatter plot* and is shown in Figure 3.2.

The upper row of Figure 3.2 shows Moran plots corresponding to two different time instants of a certain variable X. In the bottom left-hand graph, both Moran plots are jointly represented, and the transition between the two periods for each zone is represented by a motion vector whose direction is $(\overrightarrow{t_1, t_2})$. The characteristics of these motion

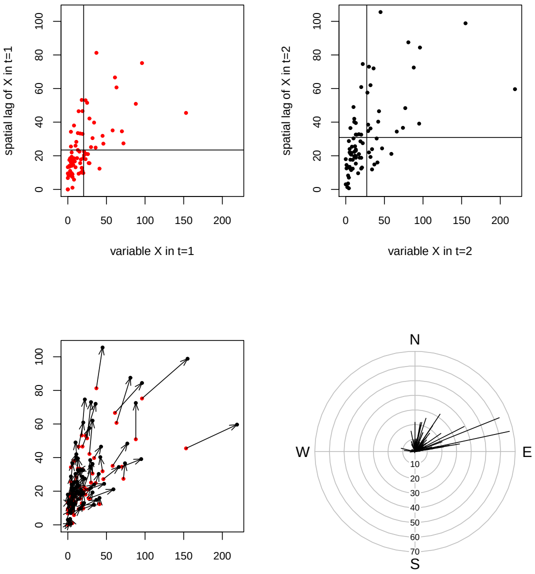

Figure 3.2. Upper row: simple Moran scatter plot. Lower row: directional Moran scatter plot.

vectors can be summarised graphically by the radial plot in the lower right-hand part of the figure, where the four quadrants now have a different meaning than their equivalents in the original Moran plot. For example, movements in the NE direction indicate an overall growth, both for the area and for its neighbours, as big as the magnitude of the vector.

The graphical analysis can be completed with a test (Rey et al., 2011) to contrast the spatial dependence of the observed motion against the null hypothesis that X and its spatial lag Y evolve independently. The originator of the modern Monte Carlo test is Barnard (1963). The test proposed here is based on M random permutations of the observed values over the neighbours of each area, whose value is fixed. For each permutation, the spatial lag and associated motion vectors are recalculated. Two tests and their corresponding p-values can be applied:

Normal approximation.- Based on,

$$z_i = \frac{n_i - \bar{n}_i^*}{s_{\bar{n}_i^*}},$$

where n_i is the number of vectors from the quadrant $i = 1, 2, 3, 4$ in the original radial plot, and $\bar{n}_i^*$ and $s_{\bar{n}_i^*}$ are the mean and the standard deviation of the number of observed vectors in i along the M random permutations. The associated p-value is calculated from the distribution function of the standard normal.

Non-parametric test.- Based on the range of n_i between the values obtained in the M random permutations. If m_i is the number of permutations equal to or greater than n_i, the associated p-value is obtained by

$$p = \frac{m_i + 1}{M + 1}. \tag{3.5}$$

3.2.2.3 Spatial Markov chain models

Markov chains were introduced by Quah (1993) for the study of regional income distribution as an alternative method to σ and β convergence (Barro and Sala-i-Martin, 2004). A transition probability matrix among the different classes of income is obtained, allowing for the representation of the dynamics of its evolution. In Rey (1993), the author extends the procedure, including the spatial structure.

If $i = 1, 2, \ldots, m$ are the different states of the variable X at a time k, verifying the condition

$$P(X_k = j | X_{k-1} = i, X_{k-2} = l, \ldots, X_0 = m) = P(X_k = j | X_{k-1} = i) = p_{ij},$$

the probabilities p_{ij} are the elements of the transition matrix P of dimension $m \times m$. Assuming temporal homogeneity, the maximum-likelihood estimation of the transition probabilities is

$$\hat{p}_{ij} = \frac{n_{ij}}{\sum_j n_{ij}},$$

where n_{ij} is the observed number of transitions from i to j.

Considering the two states L and H mentioned above, Rey (1993, 2014) proposed to separately estimate the transition matrices $P(X)$, corresponding to the variable X, and $P(Y)$ for the spatial lag of X. Both matrices have a dimension of 2×2. The matrix of joint transitions of each area and its spatial lag, $P(XY)$, is also estimated. Their dimension is 4×4, since they can jointly be in any of the four quadrants (states) in Figure 3.1. Under the null hypothesis of independence or absence of spatial dynamics

$$\widehat{P(XY)} = \widehat{P(X)} \otimes \widehat{P(Y)}, \tag{3.6}$$

where $\otimes$ is the Kronecker product.

A χ^2 test with $4 \times (4 - 1) = 12$ degrees of freedom (Anderson and Goodman, 1957) allows for testing (3.6). Its rejection means that the respective Markov chains are not separable. That is, the evolution of the variable in a given area is not independent of that of its neighbours.

3.2.3 Space-time models

The available literature offers a large range of space-time models for lattice data. Here, we focus on the parametric models defined by Bernardinelli et al. (1995), the non-parametric models described by Knorr-Held (2000) and some variations of them.

Let O_{ij} be the number of observed cases and E_{ij} the expected number of cases for the ith area and jth period of time. We assume the number of observed cases is modelled as

$$O_{ij} \sim \text{Poisson}(r_{ij} E_{ij}),$$

where r_{ij} is the underlying disease rate.

3.2.3.1 Models with parametric time trend

A parametric formulation for space-time modelling was introduced by Bernardinelli et al. (1995) and assumes the log-rate as a linear function of time

$$\log(r_{ij}) = \alpha + (\theta_i + \phi_i) + (\beta + \delta_i)t_j, \tag{3.7}$$

where α is the intercept, θ_i and ϕ_i two area effects adopting the standard Besag et al. (1991) model with structured and unstructured components, β is the mean linear time trend over all areas, and δ_i is the interaction between the time effect and the space effect.

With respect to the model identifiability, Bernardinelli et al. (1995) suggest a parametrisation in which $\sum \phi_i = 0$ and $\sum \delta_i = 0$. Thus, δ_i is the difference between the area-specific trend and the mean trend β. Therefore, a value of $\delta_i < 0$ implies that the area-specific trend is less steep than the mean linear time trend, whilst a value of $\delta_i > 0$ implies that the area-specific trend is steeper than the mean trend.

To allow close regions to have similar incidence rates, the spatially structured component $\boldsymbol{\theta}$ is modelled as an intrinsic Gaussian Markov random field (iGMRF) where the prior density of $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_n)^T$ can be written as

$$\pi(\boldsymbol{\theta}|\sigma^2_{\boldsymbol{\theta}}) \propto \exp\left(-\frac{1}{2\sigma^2_{\boldsymbol{\theta}}} \sum_{i\sim i'} (\theta_i - \theta_{i'})^2\right), \tag{3.8}$$

where $i \sim i'$ represents all pairs of adjacent regions.

The parameter ϕ_i represents the unstructured residual, modelled using an exchangeable prior: $\phi_i \sim \text{Normal}(0, \sigma^2_{\phi})$.

The parameter δ_i, the so-called differential trend, captures the interaction between the linear time trend and the two spatial effects $\boldsymbol{\theta}$ and $\boldsymbol{\phi}$. It can be interpreted as the difference between the time trend of region i and the overall time trend β. It can be specified as an independent and identically, mean-zero normally distributed effect or specified as in (3.8).

3.2.3.2 Models with non-parametric time trend

The models described above assume a linear time trend in each region. Knorr-Held (2000) proposes a model specification combining the spatial model by Besag et al. (1991) with dynamic models. This allows for a non-parametric estimation of temporal trends, which neither assumes linearity nor stationarity and can be seen as the temporal analogue of the spatially structured component in Besag et al. (1991). According to Knorr-Held (2000), a non-parametric, additive space-time model can be written as

$$\log(r_{ij}) = \alpha + \theta_i + \phi_i + \rho_j + \gamma_j + \delta_{ij}, \tag{3.9}$$

where θ_i and ϕ_i have the same parametrisation as in (3.7), and ρ_j and γ_j represent the structured and unstructured temporal effects. An independent mean-zero normal prior which assumes no temporal structure is used for the temporal effect $\boldsymbol{\gamma}$. In contrast, the parameter $\boldsymbol{\rho}$ displays temporal structure. Schrödle and Held (2011b) suggest the use of first- and second-order random walk. We consider a first-order random walk (RW1) for $\boldsymbol{\rho}$, with a prior in which effects for neighbouring time points tend to be alike.

$$\pi(\boldsymbol{\rho}|\sigma^2_{\boldsymbol{\rho}}) \propto \exp\left(-\frac{1}{2\sigma^2_{\boldsymbol{\rho}}}\sum_{j=2}^{T}(\rho_i - \rho_{i'})^2\right), \tag{3.10}$$

with precision matrix $k_{\boldsymbol{\rho}}K_{\boldsymbol{\rho}}$, where $k_{\boldsymbol{\rho}}$ is an unknown scalar and $K_{\boldsymbol{\rho}}$ is the structure matrix:

$$K_{\boldsymbol{\rho}} = \begin{pmatrix} 1 & -1 & & & & & \\ -1 & 2 & -1 & & & & \\ & -1 & 2 & -1 & & & \\ & & \vdots & \vdots & \vdots & & \\ & & & -1 & 2 & -1 & \\ & & & & -1 & 2 & -1 \\ & & & & & -1 & 1 \end{pmatrix}. \tag{3.11}$$

The parameter $\boldsymbol{\delta}$ represents the space-time interactions. The specification of the prior for $\boldsymbol{\delta}$ depends on which spatial and temporal effects are assumed to interact. Knorr-Held (2000) proposes four types of interaction as shown here in Table 3.1. Clayton (1996) and Knorr-Held (2000) suggest specifying $K_{\boldsymbol{\delta}}$ as the Kronecker product of the structure matrices of the corresponding effects which are assumed to interact. For example, type II interaction assumes that the unstructured spatial component and the structured temporal component of the model interact, meaning that δ follows a random walk (of first or second order, depending on the specification), independently of all other areas. Knorr-Held (2000) provides a symbolic representation that can help better understand the four possible types of interaction.

In these non-parametric models, the parameters are not identifiable. We must set a sum-to-zero constraint on all θ_i and ρ_j in order to ensure the identifiability of the intercept α. To assure identifiability for the interaction term δ, specific sum-to-zero constraints have to be used. Moreover, for type II, III and IV interaction more constraints must be considered, taking into account the rank deficiency of the structure matrix of δ.

Space-time interaction	K_{δ}
Type I	$K_{\phi} \otimes K_{\gamma}$
Type II	$K_{\phi} \otimes K_{\rho}$
Type III	$K_{\theta} \otimes K_{\gamma}$
Type IV	$K_{\theta} \otimes K_{\rho}$

Table 3.1. Specifications for the four types of space-time interaction.

These constraints can be defined using the results of Rue and Held (2005), and applied as in Schrödle and Held (2011b). The space-time models can be implemented using the R-INLA library (Rue et al., 2009).

3.3 Space-time analysis for the varicella data

3.3.1 Exploratory analysis

Figure 3.3 shows the spatial distribution of the standardised incidence ratio (SIR) of varicella in the boroughs of Valencia during the first year of the study, 2008, and the last year of the study, 2013. Population changes within a year are minimal and do not affect the ratio of calculated incidence. We can see an evolution of the spatial distribution of varicella incidence from one year to another. Areas in the south and east of Valencia are showing a decrease in incidence, and some of the boroughs in the centre of the city show an increase. Higher values of SIR are observed in the periphery of the city, which indicates varicella has a specific spatial behaviour. Table 3.2 shows the values for the Moran's index calculated for each year and reflects the existence of a spatial effect for all years. So far, Figure 3.3 and Table 3.2 suggest spatial dependence but are not able to show potential spatial dynamics and underlying movement patterns.

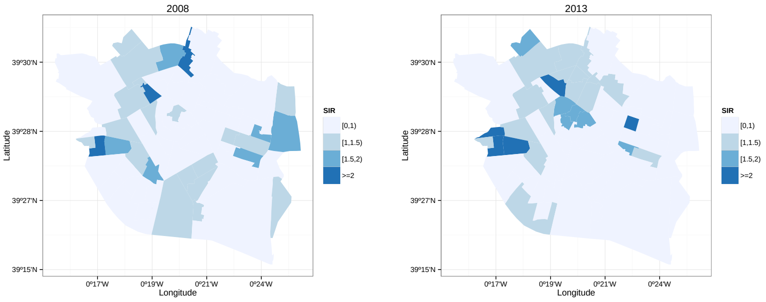

Figure 3.3. Right: spatial distribution of varicella SIR in Valencia for 2008. Left: spatial distribution of varicella SIR in Valencia for 2013.

The exploratory analysis using the directional Moran scatter plot summarised in Table 3.3 will help us identify dynamics in the distribution of our data. Table 3.3 shows the number of *observed* vectors in each of the four quadrants, the *expected*, which is the average obtained from the $M = 499$ random permutations, and the associated *p-value*

	Asymptotic normality test				Montecarlo test	
	Moran I	Expectation	Variance	p-value	observed rank	p-value
SIR-08	0.1791	-0.0145	0.0050	3.09e-03	995	0.005
SIR-09	0.2612	-0.0145	0.0047	2.87e-05	1000	0.001
SIR-10	0.2489	-0.0145	0.0051	1.06e-04	1000	0.001
SIR-11	0.3617	-0.0145	0.0052	8.59e-08	1000	0.001
SIR-12	0.2381	-0.0145	0.0047	1.20e-04	999	0.001
SIR-13	0.3216	-0.0145	0.0042	1.06e-07	999	0.001

Table 3.2. Moran index for SIR of varicella values from 2008 to 2013.

calculated by (3.5) for the five year transitions. We have highlighted in grey those p-values that lead to the rejection of the null hypothesis: *the values of a borough and the values of its neighbours evolve independently.* These p-values are significant in most of the transitions, except the last one, 2012-13, which does not reject independence in any of the quadrants.

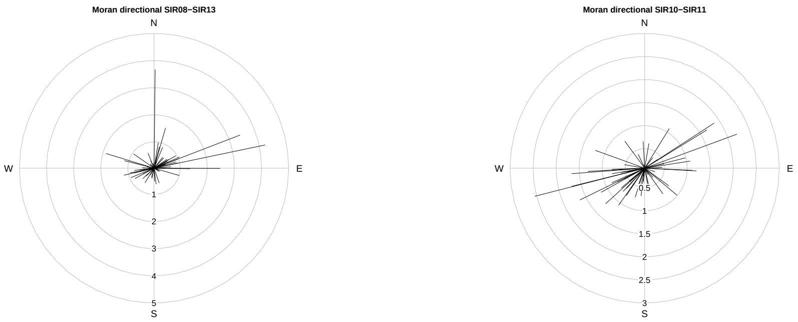

Figure 3.4. Standardised Moran directional scatter plots for 2008-2013 and 2010-2011.

Figure 3.4 shows two radial directional Moran plots corresponding to the periods 2008-2013 and 2010-2011. Movement vectors directed towards NE suggest an increase in the incidence of co-movements (movements of one borough together with its neighbours). Movements towards SW reflect decreasing co-movements of the observed incidence and its neighbours in the distribution. The length of the directional vectors stands for the amount of change in varicella incidence. We will focus on the directions of the vectors, capturing spatial dynamic distribution. Figure 3.4 seems to display more vectors in the NE and SW directions reflecting co-movements of boroughs and their neighbours. This could indicate spatial dependence in distribution dynamics in these two periods.

Table 3.3 shows that the most frequent type of movement along the study period is in the NE and SW directions. This means that boroughs and their neighbours move in the same direction within the distribution. Areas with high/low varicella incidence will move in the same direction in terms of growth/decline as their neighbouring areas, maintaining a high/low varicella rate. The last three rows of Table 3.3 summarise the movements between the initial year, 2008, and the final year, 2013. We continue to

		N-E	N-W	S-W	S-E
2008-09	observed	20	13	25	12
	expected	15.4609	19.4128	18.5872	16.5391
	p-value	0.0640	0.0080	0.0260	0.0420
2009-10	observed	27	12	19	12
	expected	21.6734	17.8918	13.1082	17.3267
	p-value	0.0520	0.0100	0.0280	0.0140
2010-11	observed	14	6	37	13
	expected	10.0701	16.0261	26.9740	16.9299
	p-value	0.0800	0.0000	0.0040	0.0360
2011-12	observed	26	12	27	5
	expected	15.6573	20.7535	18.2465	15.3427
	p-value	0.0020	0.0020	0.0060	0.0000
2012-13	observed	27	15	15	13
	expected	23.6072	18.4689	11.5311	16.3928
	p-value	0.1860	0.0620	0.1420	0.0980
2008-13	observed	23	11	23	13
	expected	23.7214	18.5431	11.4569	16.2786
	p-value	0.3380	0.0020	0.0020	0.1080

Table 3.3. Moran directional summary with non-parametric p-values.

see higher frequency in the NE and SW directions, but only the latter has a significant p-value with many more cases than expected.

The lack of independence has important implications for varicella distribution, as high or low aggregation is more likely to occur among neighbouring areas within a similar varicella incidence. Therefore, interesting behaviour is not so much to be seen in an individual borough, but rather in its interaction with neighbouring boroughs.

The analysis of the dynamics of the SIR values along the study period requires prior clarification. The threshold for determining whether the state of the SIR of varicella in a borough is L (low) or H (high) is not the average, but the value 1, so that $SIR \leq 1 \Rightarrow L$ and $SIR > 1 \Rightarrow H$. With this approach, the estimates of the transition matrices that appear on both sides of (3.6) are

$$\widehat{P(XY)} = \begin{bmatrix} 0.6145 & 0.1567 & 0.1084 & 0.1204 \\ 0.5000 & 0.1552 & 0.1034 & 0.2414 \\ 0.3636 & 0.0728 & 0.2727 & 0.2909 \\ 0.2958 & 0.1972 & 0.1690 & 0.3380 \end{bmatrix}$$

and

$$\widehat{P(X)} \otimes \widehat{P(Y)} = \begin{bmatrix} 0.5198 & 0.2213 & 0.1816 & 0.0773 \\ 0.3906 & 0.3504 & 0.1366 & 0.1224 \\ 0.3285 & 0.1398 & 0.3729 & 0.1588 \\ 0.2468 & 0.2214 & 0.2803 & 0.2515 \end{bmatrix}$$

The statistic to compare the matrices is $\chi_0 = 45.7818$, which for χ_{12} provides a p-value $p = 7.56504e - 06$. This means that the SiRs of varicella in a borough and in its surroundings evolve jointly, a result that will be taken into account in the following space-time modelling.

3.3.2 Space-time models

The models described in Section 3.2.3 are now applied to varicella data. All computation has been carried out using `INLA` (Rue et al., 2009). The data analysed in this section represents the space-time process observed in 70 boroughs during a six-year time period, from 2008 to 2013. Figure 3.5 shows the temporal evolution of the total number of varicella cases from 2008 to 2013 (solid line). Also, it shows the temporal evolution of the population (dotted line). A clear decrease in the number of cases can be seen during the last four years.

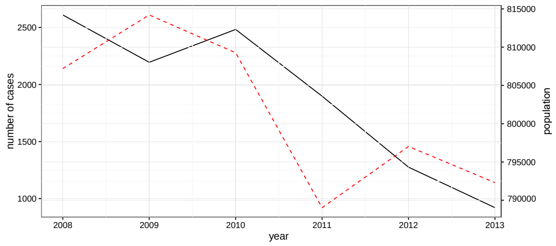

Figure 3.5. Evolution of the number of varicella cases (solid line) and evolution of the population (dotted line) during the time period 2008-2013.

The number of varicella cases has decreased from over $2,500$ cases in 2008 to less than $1,000$ cases in 2013. This decrease could be a direct consequence of Public Health preventive measures. Moreover, the population have participated in the vaccination programmes implemented by the Spanish Government. Unfortunately, information regarding vaccination is not available for the boroughs of Valencia thus can not be considered in the analysis.

Parametric and non-parametric models have been fitted to the varicella data. *Deviance information criterion* (DIC), a well-known Bayesian model choice criterion, has been used to select the best model fit. DIC is defined as the sum of the posterior mean of the deviance and the effective number of parameters. The smaller the DIC, the better the trade-off between model fit and complexity. The *logarithmic score* (Gneiting and Raftery, 2007) indicates the predictive quality of the model and can easily be obtained in `R`; the smaller the resulting score, the better the predictive quality of the model is.

Calibration refers to the statistical consistency between the probabilistic forecasts and the observations (Czado et al., 2009). If the observations are drawn from the predictive distribution, an ideal and desirable situation, the *probability integral transform* (PIT) has a standard uniform distribution. The calibration of the models can be checked by plotting the PIT-histograms as suggested by Czado et al. (2009). The closer the histograms are to being uniform, the better the calibration is.

	Bernardinelli model (model 0)
DIC	4,036.88
Deviance	3,997.30
Effective number of parameters	66.57
Logarithmic score	5.28

Table 3.4. DIC and logarithmic score for the model with parametric time trend.

	Interaction			
	Type I (model 1)	Type II (model 2)	Type III (model 3)	Type IV (model 4)
DIC	3,039.23	2,943.67	2,654.76	2,667.86
Deviance	2,600.96	2,543.04	2,356.02	2,373.55
Effective number of parameters	438.27	400.63	298.73	294.31
Logarithmic score	4.648	4.366	3.468	3.502

Table 3.5. DIC and logarithmic score for all models with non-parametric time trend.

Tables 3.4 and 3.5 show the previous criteria for the Bernardinelli model and the non-parametric models. In view of the results of the two tables, the non-parametric models represent a better fit for our data. Table 3.5 shows models 3 and 4 have the smallest DIC, $2,654.76$ and $2,667.86$, and also the smallest logarithmic score, 3.468 and 3.502. Models 3 and 4 are models with interaction type III and IV. The PIT-histograms in Figure 3.6 (lower row) are closest to being uniform. Both models can be chosen as valid for modelling our data.

Figure 3.7 shows the estimated mode for the varicella-fitted relative risk in all 70 boroughs of Valencia, for models 3 and 4, for 2008 and 2013. The spatial distribution of the estimated mode for model 3 is similar to the one estimated for model 4. For 2008, Figure 3.7 (left, upper row) and Figure 3.7 (left, lower row) show two small areas (area 1 and 2) where the relative risk has a smoother estimation for model 4. The same applies to 2013. Figure 3.7 (right, upper row) and Figure 3.7 (right, lower row) show a more uniform estimation for model 4, in the central-north part of the city and in the south.

Figure 3.7 also shows a spatial structure for model 3 where areas with low varicella risk and areas with high varicella risk tend to cluster together. We can identify areas changing from high to low relative risk in the south of the city. In the centre of the city, areas with low varicella risk do not experience changes during the period of the study.

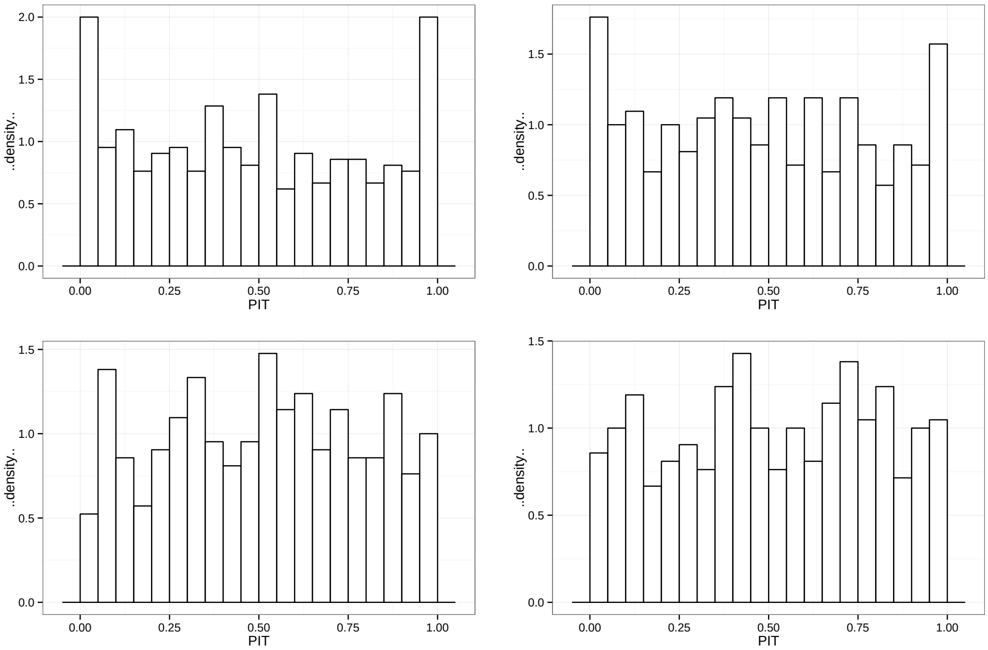

Figure 3.6. (Top) PIT-histograms for models with type I to II interaction. (Bottom) PIT-histograms for models with type III to IV interaction.

High varicella risk boroughs are clustered in the western and eastern areas of Valencia. A similar behaviour is observed for model 4. We can see that the varicella spatial pattern varies over the years.

Figure 3.8 shows the estimated linear time trend for model 4 corresponding to the *La Creu Coberta* borough in district 9 and the *Cami Fondo* borough in district 12. We can see that model 4 manages to describe the underlying decrease in the rate.

The spatial structure shows a higher relative risk in economically disadvantaged peripheral areas, or areas with a younger population. In 2008, Figure 3.7 shows an area in the centre of Valencia where the relative risk of varicella is less than 0.7, spreading as far as the north-west part of the city. In 2013, we identified low varicella risk areas in the centre and eastern parts of Valencia. All these areas are boroughs with an elderly population and a community more involved in the vaccination programmes.

3.4 Conclusions and discussion

We carried out an applied epidemiological and statistical study to see how an airborne disease like varicella behaves spatially and temporally. We applied a wide range of descriptive spatio-temporal tools and spatio-temporal models to analyse varicella. The exploratory space-time data analysis showed an important relationship between the spatial and the temporal component of the data. We used directional statistics in order to

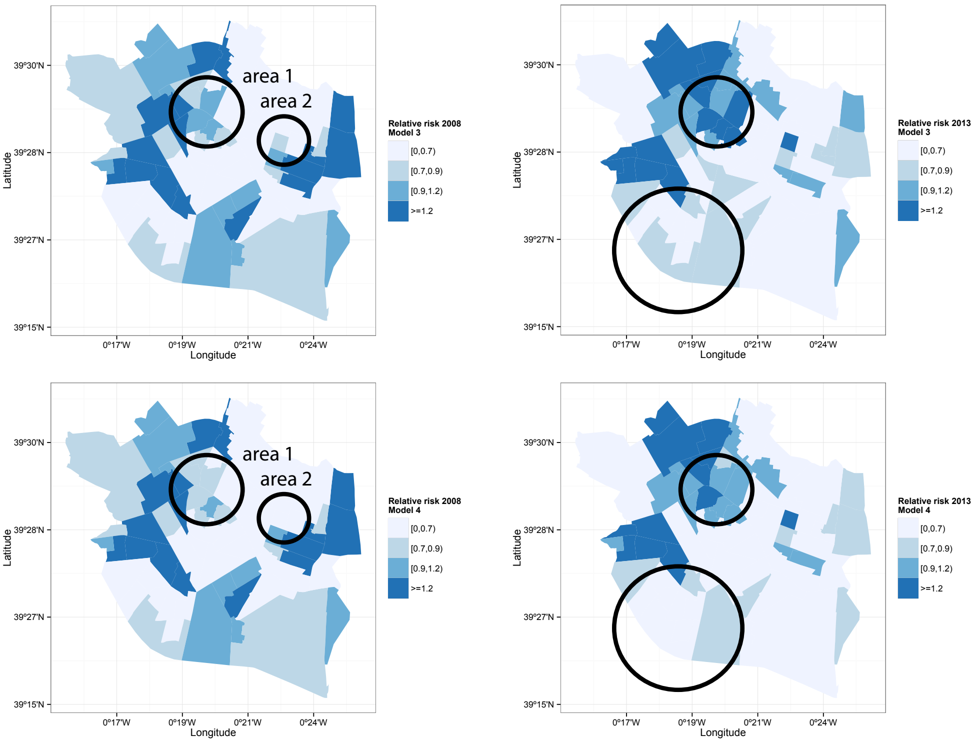

Figure 3.7. (Top) Distribution of the estimated relative risk of varicella according to models 3 and 4 for 2008. (Bottom) Distribution of the estimated relative risk of varicella according to models 3 and 4 for 2013.

obtain relevant information from movements within the space-time distribution. This dynamic approach relates the spatial dependence to the statistical relationship between an area and its neighbours across space and over time. Our results suggest boroughs do not act independently, and we have identified a range of movements, mostly in the NE and SW directions. This proves a strong spatio-temporal relationship between the small areas of Valencia.

The space-time models confirm the existence of an interaction between the spacial and the temporal component in our data. Both models with type III and type IV interaction are identified as a good fit for our data. We consider model 4 a more complex model capable of explaining all underlying space-time effects. The results from the exploratory analysis also allow us to support this decision.

Epidemiologists confirm that there has been a decrease in the incidence of varicella due to disease vaccination. Since 2006, the varicella vaccine has been included in the children vaccination schedule of the Community of Valencia (CV) for 11-year-old children with no previous history of vaccination or illness (two doses one month apart). The Spanish Association of Paediatrics (SAP) had previously recommended two varicella vaccine doses, one from 12 to 15 -months and a second one, from 4 to 6 years. Most paediatricians carried out the vaccination this way until last year, when the vaccine was

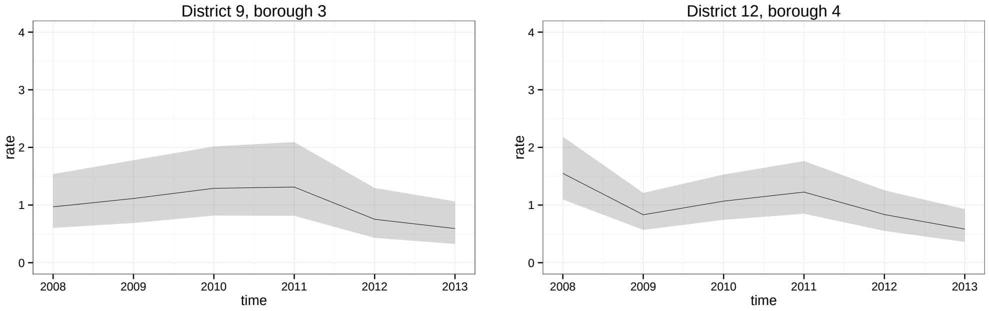

Figure 3.8. (Left) Linear time trend for borough 3 from district 9 corresponding to the model with type IV interaction. (Right) Linear time trend for borough 4 from district 12 corresponding to the model with type IV interaction

withdrawn from pharmacies and SAP recommended not administering it until the age of 12. Vaccination coverage is quite high: 39.12% for 2008, 39.29% for 2009, 42.83% for 2010, 49.54% for 2011, 53.41% for 2012 and 41.92% for 2013.

The coverage rates are quite parallel to the incidence rates of varicella; when varicella vaccination increases, we observe a decrease in the incidence of the disease. The vaccines given before the age of 11 in Valencia were not funded by Public Health; thus, parents bought the vaccines on recommendation of paediatricians and then administered them to their children in private clinics. The population living in the centre of the city of Valencia generally tends to have a fairly high income level, and in those neighbourhoods, more children were probably vaccinated than in peripheral areas. In mid-2013, the varicella vaccine was withdrawn from pharmacies and SAP banned its administration. In fact, data from 2014 confirmed that varicella incidence has once again risen. As we already stated, most chickenpox cases occur before the age of 10, so when the vaccination schedule starts at age 11, most children have already had chickenpox unless they have been vaccinated, which was the case in the years encompassed by our study.

The next interesting step in the analysis of the varicella data is to examine the spatio-temporal point pattern. This will be further discussed in Chapter 4.

Chapter 4

A multi-scale area-interaction model for spatio-temporal point patterns

This chapter presents an extension of the area-interaction model (Baddeley and van Lieshout, 1995) to a space-time framework. This extension is an important contribution to the field of statistical analysis of point patterns. The new model we introduce includes different types of interaction at different spatio-temporal scales. We first present some preliminaries in relation to notation and terminology. We follow by defining the probability density for the new model and the Markov properties. We then adapt simulation algorithms, such as the Metropolis-Hastings and the birth-and-death algorithms, to our context. Pseudo-likelihood method for inference and an extension of the Berman-Turner procedure for the new *spatio-temporal multi-scale area-interaction* model are proposed afterwards. All the proposed algorithms are illustrated using a simulated example. Varicella data is used to show an application of the newly introduced model.

4.1 Introduction

Spatio-temporal patterns are increasingly observed in many different fields, including ecology, epidemiology, seismology, astronomy and forestry. The common feature is that all observed events have two basic characteristics: the location and the time of the event. Here we are mainly concerned with epidemiology (Stallybrass, 1931), which studies the distribution, causes and control of diseases in a defined human population. The locations of the occurrence of cases give information on the spatial behaviour of the disease, whereas the times, measured on different scales (days, weeks, years, period of times), give insights on the temporal response of the overall process. An essential point to take into consideration is that people are not uniformly distributed in space, hence information on the spatial distribution of the population at risk is crucial when analysing spatio-temporal patterns of diseases.

Realistic models to fit epidemiological data should incorporate spatio-temporal inhomogeneity and allow for different types of dependence between points. One important class of such models is the family of Gibbs point processes, defined in terms of their probability density function (Ripley, 1988, 1989; Van Lieshout, 2000), and, in particular, the sub-class of pairwise interaction processes. Well-known examples of pairwise interaction processes are the Strauss model (Strauss, 1975; Kelly and Ripley, 1990) or the hard core process, a particular case of the Strauss model where no points ever come closer to each other than a given threshold. However, pairwise interaction models are not a suitable choice for fitting clustered patterns. A family of Markov point processes that can fit both clustered and ordered patterns is that of the area- or quermass-interaction models (Baddeley and van Lieshout, 1995; Kendall et al., 1999). These models are defined in terms of stochastic geometric functionals and display interactions of all orders. Methods for inference and perfect simulation are available in Kendall (1998); Häggström et al. (1999); Møller and Helisová (2010); Dereudre et al. (2014).

Most natural processes exhibit interaction at multiple scales. The classical Gibbs processes model spatial interaction at a single scale, nevertheless *multi-scale* generalizations have been proposed in the literature (Ambler and Silverman, 2010; Gregori et al., 2003; Picard et al., 2009).

4.2 Preliminaries

A realization of a spatio-temporal point process consists of a finite number $n \geq 0$ of distinct points (x_i, t_i), $i = 1, \ldots, n$, that are observed within a compact spatial domain $W_S \subseteq \mathbb{R}^2$ and time interval $W_T \subseteq \mathbb{R}$. The pattern formed by the points will be denoted by $\mathbf{x} = \{(x_i, t_i)\}_{i=1}^n$. For a mathematically rigorous account, the reader is referred to Daley and Vere-Jones (2003, 2008).

We define the Euclidean norm $||x|| = (x_1^2+x_2^2)^{1/2}$ and the Euclidean metric $d_{\mathbb{R}^2}(x, y) = ||x-y||$ for $x = (x_1, x_2) \in \mathbb{R}^2$ and $y = (y_1, y_2) \in \mathbb{R}^2$. We need to treat space and time differently, thus on $\mathbb{R}^2 \times \mathbb{R}$ we consider the supremum norm $||(x, t)||_\infty = \max\{||x||, |t|\}$ and the supremum metric $d((x, t),(y, s)) = ||(x, t) - (y, s)||_\infty = \max\{||x - y||, |t - s|\}$, where $(x, t),(y, s) \in \mathbb{R}^2 \times \mathbb{R}$. Note that $(\mathbb{R}^2 \times \mathbb{R}, d(\cdot, \cdot))$ as well as its restriction to $W_S \times W_T$ is a complete, separable metric space. We write $\mathcal{B}(\mathbb{R}^2 \times \mathbb{R}) = \mathcal{B}(\mathbb{R}^2) \otimes \mathcal{B}(\mathbb{R})$ for the Borel σ-algebra and ℓ for Lebesgue measure. We denote by $\oplus$ the Minkowski addition of two sets $A, B \subset \mathbb{R}^2$, defined as the set $A \oplus B = \{a + b : a \in A, b \in B\}$.

As stated in Section 4.1, Gibbs models form an important class of models able to fit epidemiological data exhibiting spatio-temporal inhomogeneity and interaction between points. In space, the Widom-Rowlinson *penetrable sphere model* (Widom and Rowlinson, 1970) produces clustered point patterns; the more general area-interaction model (Baddeley and van Lieshout, 1995) fits both clustered and inhibitory point patterns. In its most simple form, the area-interaction model is defined by its probability density

$$p(\mathbf{x}) = \alpha \lambda^{n(\mathbf{x})} \gamma^{-A(\mathbf{x})} \tag{4.1}$$

with respect to a unit rate Poisson process on W_S. Here α is the normalizing constant, $\mathbf{x}$

is a spatial point configuration in $W_S \subset \mathbb{R}^2$, $n(\mathbf{x})$ is the cardinality of $\mathbf{x}$ and $A(\mathbf{x})$ is the area of the union of discs of radius r centered at $x_i \in \mathbf{x}$ restricted to W_S. The positive scalars λ, γ and $r > 0$ are the parameters of the model. Note that, as emphasized in Van Lieshout (2000), Gibbsian interaction terms can be combined to yield more complex models. Doing so, Ambler and Silverman (2010); Gregori et al. (2003); Picard et al. (2009) develop an extension of the area-interaction process which incorporates both inhibition and attraction. We propose a further generalization of the area-interaction model to allow multi-scale interaction in a spatio-temporal framework.

4.3 Space-time area-interaction processes

Let $\mathbf{x}$ be a finite spatio-temporal point configuration on $W_S \times W_T \subset \mathbb{R}^2 \times \mathbb{R}$, that is, a finite set of points, including the empty set.

Definition 1. *The* spatio-temporal multi-scale area-interaction process *is the point process with density*

$$p(\mathbf{x}) = \alpha \prod_{(x,t)\in\mathbf{x}} \lambda(x,t) \prod_{j=1}^{m} \gamma_j^{-\ell(\mathbf{x}\oplus G_j)} \tag{4.2}$$

with respect to a unit rate Poisson process on $W_S \times W_T$*, where* $\alpha > 0$ *is a normalizing constant,* $\lambda \geq 0$ *is a measurable and bounded function,* ℓ *is Lebesgue measure restricted to* $W_S \times W_T$*,* $\gamma_j > 0$ *are the interaction parameters,* G_j *are some compact subsets of* $\mathbb{R}^2 \times \mathbb{R}$ *with size depending on* j*,* $j = 1, \dots, m$*,* $m \in \mathbb{N}$*, and* $\oplus$ *denotes Minkowski addition.*

Note that when $\mathbf{x}$ is the empty set, $p(\mathbf{x}) = \alpha$. The interaction parameters have the same interpretation as for the spatial area-interaction model (4.1). For fixed $j \in \{1, \dots, m\}$, when $0 < \gamma_j < 1$ we would expect to see *inhibition* between points at spatio-temporal scales determined by the definition of the compact set G_j. On the other hand, when $\gamma_j > 1$ we expect *clustering* between the points. We observe that (4.2) reduces to an inhomogeneous Poisson process when $\gamma_j = 1$ for all $j \in \{1, \dots, m\}$.

Covariates can be introduced in the model by letting the intensity function λ be a measurable and bounded function $\lambda(x,t) = \rho(Z(x,t))$ of the covariate vector $Z(x,t)$.

The new model proposed in (4.2) successfully extends the area-interaction model to multi-scale interaction for spatio-temporal point patterns.

Lemma 1. *The density* (4.2) *is measurable and integrable for all* γ_j*,* $j = 1, \dots, m$*,* $m \in \mathbb{N}$*.*

Proof. Consider a point configuration, $\mathbf{x}$. Since ℓ is σ-finite and G_j is compact, the map $\mathbf{x} \mapsto \ell(\mathbf{x} \oplus G_j)$ is measurable for any $j = 1, \dots, m$. It follows that the map $\mathbf{x} \mapsto \exp[-\ell(\mathbf{x} \oplus G_j) \log \gamma_j]$ is measurable for any $j = 1, \dots, m$. The map $\mathbf{x} \mapsto \prod_{x_i \in \mathbf{x}} \lambda(x_i, t_i)$ is also measurable by assumption, hence the density (4.2) is measurable.

To determine if (4.2) is integrable, we observe that $0 \leq \ell(\mathbf{x} \oplus G_j) \leq \ell(W_S \times W_T) < \infty$. The function λ is integrable by assumption, hence (4.2) is dominated by an integrable function, and therefore integrable. $\square$

As a further simplification, for fixed $j \in \{1, \dots, m\}$, consider the case where $\mathbf{x} \oplus G_j = \bigcup_{(x,t) \in \mathbf{x}} \mathcal{C}_{r_j}^{t_j}(x, t)$ is the union of all cylinders with radius (r_j, t_j) centered in (x, t) taken over all $(x, t) \in \mathbf{x}$. We define the cylinder with radius (r_j, t_j) by

$$\mathcal{C}_{r_j}^{t_j}(x, t) = \{(y, s) \in W_S \times W_T : ||x - y|| \leq r_j, |t - s| \leq t_j\}.$$

Then $\mathbf{x} \oplus G_j$ is the set of all points within the cylinders $\mathcal{C}_{r_j}^{t_j}(x, t)$ centered in points of $\mathbf{x}$ and the expression (4.2) reads

$$p(\mathbf{x}) = \alpha \prod_{(x,t) \in \mathbf{x}} \lambda(x, t) \prod_{j=1}^{m} \gamma_j^{-\ell(\cup_{(x,t) \in \mathbf{x}} \mathcal{C}_{r_j}^{t_j}(x,t))}, \tag{4.3}$$

where (r_j, t_j) are pairs of *irregular parameters* (Baddeley et al., 2015) of the model and γ_j are interaction parameters, $j = 1, \dots, m$. The function λ is here assumed known for simplicity, but could also depend on further parameters.

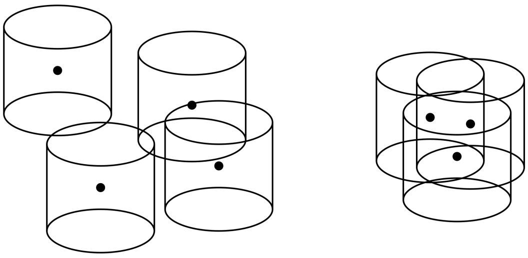

Figure 4.1. An illustration of possible $\mathbf{x} \oplus G$ (cylinders around the points), where the black dots represent points of the process.

Figure 4.1 shows an illustration of $\mathbf{x} \oplus G_j$. When $0 < \gamma_j < 1$, point configurations such as the one on the left are likely to be observed (inhibition between points), whereas for large $\gamma_j > 1$, point configurations such as the one on the right are more likely to be observed (attraction between points).

4.3.1 Markov properties

Let $\sim$ on $\mathbb{R}^2 \times \mathbb{R}$ be a symmetric and reflexive relation on $\mathbb{R}^2 \times \mathbb{R}$, i.e. for any $(x, t), (y, s) \in \mathbb{R}^2 \times \mathbb{R}$, $(x, t) \sim (y, s) \Leftrightarrow (y, s) \sim (x, t)$ and $(x, t) \sim (x, t)$. Two points (x, t) and (y, s) are said to be *neighbours* if $(x, t) \sim (y, s)$. An example of a fixed range relation on $\mathbb{R}^2 \times \mathbb{R}$ is

$$(x, t) \sim (y, s) \Leftrightarrow (x, t) \oplus G \cap (y, s) \oplus G \neq \emptyset, \tag{4.4}$$

where $G = \mathcal{C}_{r_1}^{t_1}$ is a cylinder of radius (r_1, t_1).

Definition 2. *A point process has the* Markov property *(Van Lieshout, 2000; Ripley and Kelly, 1977) with respect to the symmetric, reflexive relation* $\sim$*, if, for all point configurations* $\mathbf{x}$ *with* $p(\mathbf{x}) > 0$*, the following conditions are fulfilled:*

1. *$p(\mathbf{y}) > 0$ for all $\mathbf{y} \subseteq \mathbf{x}$;*
2. *the likelihood ratio $\dfrac{p(\mathbf{x} \cup \{(y,s)\})}{p(\mathbf{x})}$ for adding a new point (y,s) to a point configuration $\mathbf{x}$ depends only on points $(x,t) \in \mathbf{x}$ such that $(y,s) \sim (x,t)$, i.e. depends only on the neighbours of (y,s).*

Lemma 2. *The spatio-temporal multi-scale area-interaction process* (4.2) *is a Markov point process with respect to the relation* (4.4) *in the sense of Ripley and Kelly (1977).*

Proof. Note that if $p(\mathbf{x}) > 0$, since $\lambda(x,t) > 0$ for all $(x,t) \in \mathbf{x}$, then whenever $\mathbf{y} \subseteq \mathbf{x}$, also $p(\mathbf{y}) > 0$. The likelihood ratio

$$\begin{aligned}\frac{p(\mathbf{x} \cup \{(y,s)\})}{p(\mathbf{x})} &= \frac{\alpha \left(\prod\limits_{(x,t)\in\mathbf{x}} \lambda(x,t) \right) \lambda(y,s) \prod\limits_{j=1}^{m} \gamma_j^{-\ell((\mathbf{x}\cup\{(y,s)\})\oplus G_j)}}{\alpha \prod\limits_{(x,t)\in\mathbf{x}} \lambda(x,t) \prod\limits_{j=1}^{m} \gamma_j^{-\ell(\mathbf{x}\oplus G_j)}} \\ &= \lambda(y,s) \prod_{j=1}^{m} \gamma_j^{-\ell(((y,s)\oplus G_j)\setminus(\mathbf{x}\oplus G_j))}. \end{aligned} \tag{4.5}$$

Note that

$$\begin{aligned}((y,s)\oplus G_j) \setminus (\mathbf{x} \oplus G_j) &= ((y,s)\oplus G_j) \cap \left[\bigcup_{(x,t)\in\mathbf{x}} (x,t) \oplus G_j \right]^c \\ &= ((y,s)\oplus G_j) \cap \left[\bigcup_{(x,t)\sim(y,s)} (x,t) \oplus G_j \right]^c, \quad \forall j = 1,\dots,m.\end{aligned}$$

Thus (4.5) depends only on the newly added point (y,s) and its neighbors. Hence (4.2) defines a Markov point process with respect to $\sim$. □

It follows that the density $p(\cdot)$ in (4.2) is Markov at range $2\max\{(r_j,t_j)\}$, $j = 1,\dots,m$.

Define the Papangelou conditional intensity of a point process with density p defined by

$$\lambda((y,s);\mathbf{x}) = \frac{p(\mathbf{x} \cup \{(y,s)\})}{p(\mathbf{x})},$$

whenever $p(\mathbf{x}) > 0$ and $(y,s) \notin \mathbf{x}$. Then, for the spatio-temporal multi-scale area-interaction process, by the proof of Lemma 2 we obtain that

$$\lambda((y,s);\mathbf{x}) = \lambda(y,s) \prod_{j=1}^{m} \gamma_j^{-\ell(\mathcal{C}_{r_j}^{t_j}(y,s)\setminus\bigcup_{(x,t)\in\mathbf{x}} \mathcal{C}_{r_j}^{t_j}(x,t))}, \tag{4.6}$$

or, upon transformation to a logarithmic scale,

$$\log \lambda((y,s);\mathbf{x}) = \log \lambda(y,s) - \sum_{j=1}^{m} (\log \gamma_j)\, \ell \left(\mathcal{C}_{r_j}^{t_j}(y,s) \setminus \bigcup_{(x,t)\in \mathbf{x}} \mathcal{C}_{r_j}^{t_j}(x,t) \right).$$

Note that $\lambda(y,s)$ may be 0, thus making $\log \lambda(y,s)$ ill-defined.

Write $\eta_j = \log \gamma_j$. Then, whenever well-defined,

$$\begin{aligned} \log \lambda((y,s);\mathbf{x}) &= \log \lambda(y,s) - \sum_{j=1}^{m} \eta_j \int_{W_S \times W_T} \mathbf{1}\{(z,u) \in \mathcal{C}_{r_j}^{t_j}(y,s) \setminus \bigcup_{(x,t)\in \mathbf{x}} \mathcal{C}_{r_j}^{t_j}(x,t)\}\, dz\, du \\ &= \log \lambda(y,s) - \sum_{j=1}^{m} \int_{\mathrm{F}_{r_j}^{t_j}(y,s)} \sum_{i=j}^{m} \eta_i\, \mathbf{1}\{(z,u) \notin \bigcup_{(x,t)\in \mathbf{x}} \mathcal{C}_{r_i}^{t_i}(x,t)\}\, dz\, du, \end{aligned} \tag{4.7}$$

where $\mathrm{F}_{r_j}^{t_j}(x,t)$ is the difference between two concentric cylinders $\mathcal{C}_{r_j}^{t_j}(x,t)$ and $\mathcal{C}_{r_{j-1}}^{t_{j-1}}(x,t)$.

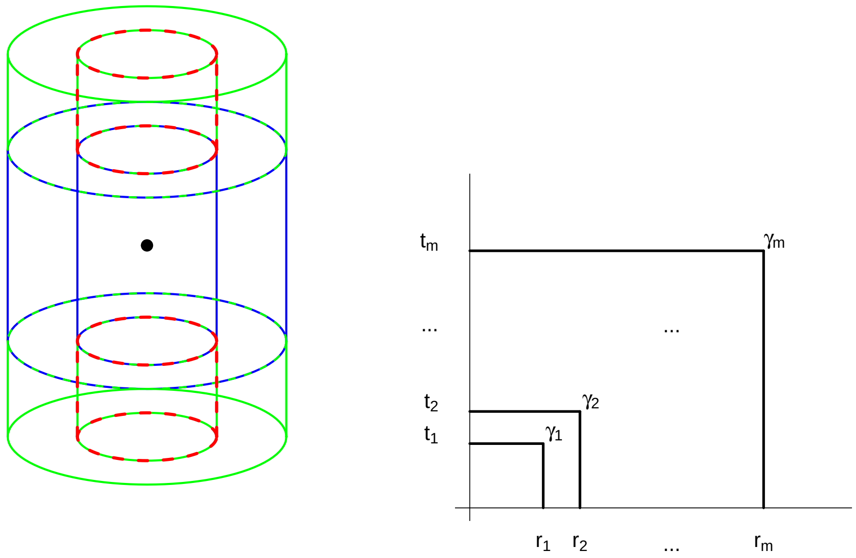

Figure 4.2. (Left) An illustration of $F_{r_j}^{t_j}$ where the blue annulus corresponds to $\{(y,s) \in W_S \times W_T : r_{j-1} < ||x-y|| \leq r_j, |t-s| \leq t_{j-1}\}$, the two green annuli represent $\{(y,s) \in W_S \times W_T : r_{j-1} < ||x-y|| \leq r_j, t_{j-1} < |t-s| \leq t_j\}$ and the two red cylinders are $\{(y,s) \in W_S \times W_T : ||x-y|| \leq r_{j-1}, t_{j-1} < |t-s| \leq t_j\}$. (Right) Multi-scale behavior.

Indeed,

$$\begin{aligned}
\mathrm{F}_{r_j}^{t_j}(x,t) &= \mathcal{C}_{r_j}^{t_j}(x,t) \setminus \mathcal{C}_{r_{j-1}}^{t_{j-1}}(x,t) \\
&= \left\{ (y,s) \in W_S \times W_T : \begin{array}{l} r_{j-1} < ||x-y|| \leq r_j, |t-s| \leq t_{j-1} \text{ or} \\ r_{j-1} < ||x-y|| \leq r_j, t_{j-1} < |t-s| \leq t_j \text{ or} \\ ||x-y|| \leq r_{j-1}, t_{j-1} < |t-s| \leq t_j \end{array} \right\},
\end{aligned}$$

with $0 = r_0 < r_1 < \cdots < r_m$ and $0 = t_0 < t_1 < \cdots < t_m$. The left-most panel of Figure 4.2 shows an illustration of $F_{r_j}^{t_j}$ for fixed r_j, t_j. The blue annulus corresponds to $\{(y,s) \in W_S \times W_T : r_{j-1} < ||x-y|| \leq r_j, |t-s| \leq t_{j-1}\}$, the two green annuli represent $\{(y,s) \in W_S \times W_T : r_{j-1} < ||x-y|| \leq r_j, t_{j-1} < |t-s| \leq t_j\}$ and the two red cylinders form $\{(y,s) \in W_S \times W_T : ||x-y|| \leq r_{j-1}, t_{j-1} < |t-s| \leq t_j\}$. If, for $(y,s), ||y-x|| > 2r_m, |s-t| > 2t_m, \forall (x,t) \in \mathbf{x}$, then

$$\begin{aligned}
\log \lambda((y,s);\mathbf{x}) &= \log \lambda(y,s) - \sum_{j=1}^{m} \left(\sum_{i=j}^{m} \eta_i \right) \ell(\mathrm{F}_{r_j}^{t_j}(y,s)) \\
&= \log \lambda(y,s) - \sum_{j=1}^{m} \eta_j \ell(C_{r_j}^{t_j}(y,s)).
\end{aligned}$$

To conclude this discussion, note that, in accordance with Gregori et al. (2003),

$$p(\mathbf{x}) = \alpha \prod_{(x,t)\in\mathbf{x}} \lambda(x,t) \exp\left[-\sum_{j=1}^{m} \alpha_j \, \ell(F_{r_j}^{t_j}(\mathbf{x}))\right],$$

where $\alpha_j = \sum_{i \geq j} \eta_i$ and $F_{r_j}^{t_j}(\mathbf{x})) = (\mathbf{x} \oplus G_j) \backslash (\mathbf{x} \oplus G_{j-1})$ As before, $G_j = C_{r_j}^{t_j}$.

The model in (4.2) with Papangelou conditional intensity defined by (4.7) allows for models whose interaction behaviour varies across spatio-temporal scales, for example, inhibition at small scales, attraction at larger scales and randomness beyond. The different spatio-temporal scales, (r_j, t_j), are defined according to $\mathrm{F}_{r_j}^{t_j}$. Indeed, a point (z,u) in $F_{r_j}^{t_j}(\mathbf{x})$ contributes a term α_j to the energy (the negative of the exponential term) in $p(\mathbf{x})$. The right-most panel of Figure 4.2 shows a visual representation of this multi-scale behavior.

An important property of Markov densities is the fact that the Papangelou conditional intensity, $\lambda((y,s);\mathbf{x})$, depends only on (y,s) and its neighbours in $\mathbf{x}$, and is computationally convenient. This property will be exploited in the next section to design simulation algorithms for generating realisations of the model.

4.4 Simulation

4.4.1 The Metropolis-Hastings algorithm

Consider a Markov point process on $W_S \times W_T \subseteq \mathbb{R}^2 \times \mathbb{R}$ defined by its density $p(\cdot)$. The *Metropolis-Hastings algorithm*, first introduced in statistical physics (Barker, 1965;

Metropolis et al., 1953), is a tool for constructing a Markov process with limit distribution defined by $p(\cdot)$.

Metropolis-Hastings algorithms are *discrete time Markov* processes where transitions are defined as the proposal of a new state that is accepted or rejected based on the likelihood of the proposed state compared with the old state. We consider two types of proposals: addition (*birth*) and deletion (*death*) of a point. The likelihood ratio of the new state in comparison with the old state, for these type of transitions, is the (reciprocal) conditional intensity.

More precisely, consider the point configuration $\mathbf{x}$. We can propose either a birth or a death with respective probabilities $q(\mathbf{x})$ and $1-q(\mathbf{x})$ that depend on $\mathbf{x}$. For a birth, a new point $u \in W_S \times W_T$ is sampled from a probability density $b(\mathbf{x},\cdot)$ and the new point configuration $\mathbf{x}\cup\{u\}$ is accepted with probability $A(\mathbf{x},\mathbf{x}\cup\{u\})$, otherwise the state remains unchanged, $\mathbf{x}$. For a death, the point $x \in \mathbf{x}$ chosen to be eliminated is selected according to a discrete probability distribution $d(\mathbf{x},\cdot)$ on $\mathbf{x}$, and the proposal $\mathbf{x}\setminus\{x\}$ is accepted with probability $A(\mathbf{x},\mathbf{x}\setminus\{x\})$, otherwise the state remains unchanged.

In general, we can choose $b(\cdot,\cdot), d(\cdot,\cdot)$ and $q(\cdot)$ as we prefer. However, an important condition to consider is that of *detailed balance*, and therefore time-reversibility of the Markov process,

$$\begin{aligned}&q(\mathbf{x})\,b(\mathbf{x},u)\,A(\mathbf{x},\mathbf{x}\cup\{u\})\,p(\mathbf{x}) =\\ &(1-q(\mathbf{x}\cup\{u\}))\,d(\mathbf{x}\cup\{u\},u)\,A(\mathbf{x}\cup\{u\},\mathbf{x})\,p(\mathbf{x}\cup\{u\}).\end{aligned} \tag{4.8}$$

For simplicity, consider the case that births and deaths are equally likely and sampled uniformly, that is, $q \equiv 1/2$, $b \equiv 1/\ell(W_S \times W_T)$ and $d(\mathbf{x},\cdot) = 1/n(\mathbf{x})$, where $n(\mathbf{x})$ is the number of points in the point configuration $\mathbf{x}$. Then (4.8) reduces to

$$\begin{aligned}\frac{1}{2}\times\frac{1}{\ell(W_S\times W_T)}A(\mathbf{x},\mathbf{x}\cup\{u\})\,p(\mathbf{x}) &= \left(1-\frac{1}{2}\right)\frac{1}{n(\mathbf{x})+1}A(\mathbf{x}\cup\{u\},\mathbf{x})\,p(\mathbf{x}\cup\{u\})\\ \frac{1}{\ell(W_S\times W_T)}A(\mathbf{x},\mathbf{x}\cup\{u\})\,p(\mathbf{x}) &= \frac{1}{n(\mathbf{x})+1}A(\mathbf{x}\cup\{u\},\mathbf{x})\,p(\mathbf{x}\cup\{u\})\\ \frac{A(\mathbf{x},\ x\cup\{u\})}{A(\mathbf{x}\cup\{u\},\mathbf{x})} &= \underbrace{\frac{\ell(W_S\times W_T)}{n(\mathbf{x})+1}\times\frac{p(\mathbf{x}\cup\{u\})}{p(\mathbf{x})}}_{=r(\mathbf{x},u)}.\end{aligned}$$

Thus, more likely configurations can be favored by setting $A(\mathbf{x},\mathbf{x}\cup\{u\}) = \min\{1, r(\mathbf{x},u)\}$, and $A(\mathbf{x}\cup\{u\},\mathbf{x}) = \min\{1, 1/r(\mathbf{x},u)\}$. Therefore, using equation (4.6), for the spatio-temporal multi-scale area-interaction process (4.2), the ratio $r(\mathbf{x},u)$ for $u = (y,s)$ reduces to

$$r(\mathbf{x},u) = \frac{\ell(W_S\times W_T)}{n(\mathbf{x})+1}\lambda(y,s)\prod_{j=1}^{m}\gamma_j^{-\ell(\mathcal{C}^{t_j}_{r_j}(y,s)\setminus\bigcup_{(x,t)\in\mathbf{x}}\mathcal{C}^{t_j}_{r_j}(x,t))}. \tag{4.9}$$

In practice, we will use the logarithmic form of the conditional intensity as given in equation (4.7). When the region $W_S \times W_T$ is irregular we use rejection sampling to generate a point uniformly at random from $W_S \times W_T$.

4.4.2 Birth-and-death-processes

In this section we discuss methods for simulating (4.2) using birth-and-death processes (Preston, 1977). The birth-and-death process is a *continuous time Markov* process where the transition from one state to another is given by either a *birth* or a *death*. A birth is the transition from a point configuration $\mathbf{x} \in W_S \times W_T \subseteq \mathbb{R}^2 \times \mathbb{R}$ to $\mathbf{x} \cup \{u\}$ by adding the point $u \in W_S \times W_T$. A death is the transition from a point configuration $\mathbf{x}$ to $\mathbf{x} \setminus \{x\}$ by eliminating a point $x \in \mathbf{x}$. We denote by $b(\mathbf{x}, u)\, du$ the transition rate for a birth and by $d(\mathbf{x}, x)$ the transition rate of a death. The total birth rate from $\mathbf{x}$ is the integral

$$B(\mathbf{x}) = \int_{W_S \times W_T} b(\mathbf{x}, u)\, du$$

and the total death rate is

$$D(\mathbf{x}) = \sum_{x \in \mathbf{x}} d(\mathbf{x}, x).$$

The process stays in state $X^{(n)} = \mathbf{x}$ for an exponentially distributed random sojourn time $T^{(n)}$ with mean $1/(B(\mathbf{x}) + D(\mathbf{x}))$. The detailed balance equations are given by

$$b(\mathbf{x}, u)\, p(\mathbf{x}) = d(\mathbf{x} \cup \{u\}, u)\, p(\mathbf{x} \cup \{u\}). \tag{4.10}$$

We consider the particular case when the death rate is constant (Ripley, 1977), $d(\mathbf{x}, x) = 1$. Hence, for the spatio-temporal multi-scale area-interaction process (4.2), the birth rate is given by the conditional intensity (cf. equation (4.6))

$$b(\mathbf{x}, (y, s)) = \frac{p(\mathbf{x} \cup \{(y, s)\})}{p(\mathbf{x})} = \lambda(y, s) \prod_{j=1}^{m} \gamma_j^{-\ell(\mathcal{C}^{t_j}_{r_j}(y,s) \setminus \bigcup_{(x,t)\in\mathbf{x}} \mathcal{C}^{t_j}_{r_j}(x,t))}. \tag{4.11}$$

For computation of the ratio in equation (4.11) we will use the logarithmic form of the conditional intensity as in equation (4.7).

Following Van Lieshout (1995, 2000) we define an algorithm for simulating a birth-and-death process and generate the successive states $X^{(n)}$ and the sojourn times $T^{(n)}$ as detailed in Algorithm 1 which incorporates a rejection sampling step for computational convenience. Define a threshold $w(\mathbf{x})$, and, for $u \notin \mathbf{x}$, set

$$g(\mathbf{x}, u) = \begin{cases} b(\mathbf{x}, u), & \text{if } b(\mathbf{x}, u) \geq w(\mathbf{x}) \\ w(\mathbf{x}), & \text{otherwise.} \end{cases}$$

A common choice is to take $w(\mathbf{x})$ equal to an upper bound to the conditional intensity.

Denote by $G(\mathbf{x})$ the integral of g. We generate the sequence of $(X^{(n)}, T^{(n)})$ as follows.

Algorithm 1. *Initialize* $X^{(0)} = \mathbf{x}_0$ *for some finite point configuration with density function* $p(\mathbf{x}_0) > 0$*. For* $n = 0, 1, \ldots,$ *if* $X^{(n)} = \mathbf{x}$*, compute* $D = D(\mathbf{x})$*,* $G = G(\mathbf{x})$ *and set* $T^{(n)} = 0$*.*

- *Add an exponentially distributed time to* $T^{(n)}$ *with mean* $1/(D+G)$*;*
- *with probability* $D/(D+G)$ *generate a death* $X^{(n+1)} = \mathbf{x} \setminus \{x\}$ *by eliminating one of the current points* $x \in \mathbf{x}$ *at random according to distribution* $d(\cdot,\cdot)$ *and stop;*
- *else sample a point* u *from* $g(\mathbf{x},u)/G$*; with probability* $b(\mathbf{x},u)/g(\mathbf{x},u)$ *accept the birth* $X^{(n+1)} = \mathbf{x} \cup \{u\}$ *and stop; otherwise repeat the whole algorithm.*

4.5 Inference

4.5.1 Pseudo-likelihood method

In this section, we assume that the function λ is known and denote by $\theta = (\gamma_1, \gamma_2, \ldots, \gamma_m)$ the interaction parameters in model (4.3). To estimate θ, we may use pseudo-likelihood which aims to optimize

$$PL(\mathbf{x},\theta) = \exp\left(-\int_{W_S}\int_{W_T} \lambda_\theta((u,v);\mathbf{x})\,du\,dv\right) \prod_{(x,t)\in\mathbf{x}} \lambda_\theta((x,t);\mathbf{x}\setminus\{(x,t)\}), \tag{4.12}$$

where $\lambda_\theta((u,v);\mathbf{x})$ is the conditional intensity that depends on θ (Besag, 1975).

For a Poisson process the conditional intensity is equal to the intensity function, hence pseudo-likelihood is equivalent to maximum likelihood. In general, the pseudo-likelihood $PL(\mathbf{x},\theta)$ is only an approximation of the true likelihood. However, no sampling is needed and the computational load will be considerably smaller than for the maximum likelihood method.

The maximum pseudo-likelihood normal equations are then given by

$$\frac{\partial}{\partial\theta} \log PL(\mathbf{x},\theta) = 0, \tag{4.13}$$

where

$$\log PL(\mathbf{x},\theta) = \sum_{(x,t)\in\mathbf{x}} \log \lambda_\theta((x,t);\mathbf{x}\setminus\{(x,t)\}) - \int_{W_S}\int_{W_T} \lambda_\theta((u,v);\mathbf{x})\,du\,dv. \tag{4.14}$$

As seen in Section 4.3.1, the Papangelou conditional intensity of the spatio-temporal multi-scale area-interaction model is

$$\lambda_\theta((y,s);\mathbf{x}) = \lambda(y,s) \prod_{j=1}^{m} \gamma_j^{-\ell(\mathcal{C}_{r_j}^{t_j}(y,s)\setminus\cup_{\mathbf{x}}^j)},$$

where $\cup_{\mathbf{x}}^j = \bigcup_{(x,t)\in\mathbf{x}} \mathcal{C}_{r_j}^{t_j}(x,t)$; its logarithm reads

$$\log \lambda_\theta((y,s);\mathbf{x}) = \log \lambda(y,s) - \sum_{j=1}^{m} (\log \gamma_j)\, \ell(\mathcal{C}_{r_j}^{t_j}(y,s) \setminus \cup_{\mathbf{x}}^j).$$

Following Baddeley and Turner (2000) we denote by $S_j(y,s) = \ell(\mathcal{C}^{t_j}_{r_j}(y,s) \setminus \cup^j_{\mathbf{x}})$ the sufficient statistics, hence $\log \lambda_\theta((y,s);\mathbf{x}) = \log \lambda(y,s) - \theta^T \begin{bmatrix} S_1(y,s) \\ \cdots \\ S_m(y,s) \end{bmatrix}$. This notation will be further used in Algorithm 2.

Thus, equation (4.13) gives us the pseudo-likelihood equations

$$\frac{\partial}{\partial\theta}\Bigg(\sum_{(x,t)\in\mathbf{x}} \left[\log \lambda(x,t) - \sum_{j=1}^{m} (\log \gamma_j)\, \ell(\mathcal{C}^{t_j}_{r_j}(x,t) \setminus \cup^j_{\mathbf{x}\setminus\{(x,t)\}}) \right] - \\ - \int_{W_S}\int_{W_T} \lambda(u,v) \prod_{j=1}^{m} \gamma_j^{-\ell(\mathcal{C}^{t_j}_{r_j}(u,v)\setminus\cup^j_{\mathbf{x}})}\, du\, dv \Bigg) = 0. \tag{4.15}$$

For every parameter γ_i, $i = 1, 2, \ldots, m$, the equations (4.15) read

$$\sum_{(x,t)\in\mathbf{x}} \frac{\ell(\mathcal{C}^{t_i}_{r_i}(x,t) \setminus \cup^i_{\mathbf{x}\setminus\{(x,t)\}}))}{\gamma_i} = \\ = \int_{W_S}\int_{W_T} \lambda(u,v) \frac{\ell(\mathcal{C}^{t_i}_{r_i}(u,v) \setminus \cup^i_{\mathbf{x}})}{\gamma_i} \prod_{j=1}^{m} \gamma_j^{-\ell(\mathcal{C}^{t_j}_{r_j}(u,v)\setminus\cup^j_{\mathbf{x}})}\, du\, dv. \tag{4.16}$$

The major difficulty is to estimate the integrals on the right hand side of equations (4.16). Baddeley and Turner (Baddeley and Turner, 2000) propose using the Berman-Turner method to approximate the integral in (4.14) by

$$\int_{W_S}\int_{W_T} \lambda_\theta((u,v);\mathbf{x})\, du\, dv \approx \sum_{j=1}^{n} \lambda_\theta((u_j,v_j);\mathbf{x})\, w_j,$$

where (u_j, v_j) are points in $W_S \times W_T$ and w_j are quadrature weights. This yields an approximation for the log pseudo-likelihood of the form

$$\log PL(\mathbf{x},\theta) \approx \sum_{(x,t)\in\mathbf{x}} \log \lambda_\theta((x,t);\mathbf{x}\setminus\{(x,t)\}) - \sum_{j=1}^{n} \lambda_\theta((u_j,v_j);\mathbf{x})\, w_j. \tag{4.17}$$

Note that if the set of points $\{(u_j,v_j), j = 1,\ldots,n\}$ includes all the points $(x,t) \in \mathbf{x}$, we can rewrite (4.17) as

$$\log PL(\mathbf{x},\theta) \approx \sum_{j=1}^{n} (y_j \log \lambda_j - \lambda_j)\, w_j, \tag{4.18}$$

where $\lambda_j = \lambda_\theta((u_j,v_j);\mathbf{x}\setminus\{(u_j,v_j)\})$, $y_j = z_j/w_j$ and

$$z_j = \begin{cases} 1, & \text{if } (u_j,v_j) \in \mathbf{x} \text{ (is a point)}, \\ 0, & \text{if } (u_j,v_j) \notin \mathbf{x} \text{ (is a dummy point)}. \end{cases} \tag{4.19}$$

The right hand side of (4.18), for fixed $\mathbf{x}$, is formally equivalent to the log-likelihood of independent Poisson variables $Y_j \sim \text{Poisson}(\lambda_j)$ taken with weights w_j. Therefore (4.18) can be maximized using software for fitting generalized linear models.

In summary, the method is as follows.

Algorithm 2.

- *Generate a set of dummy points and merge them with all the data points in* $\mathbf{x}$ *to construct the set of quadrature points* $(u_j, v_j) \in W_S \times W_T$*;*
- *compute the quadrature weights* w_j*;*
- *obtain the indicators* z_j *defined in* (4.19) *and calculate* $y_j = z_j/w_j$*;*
- *compute the values* $S_j(u_j, v_j)$ *of the sufficient statistics at each quadrature point;*
- *fit a generalized log-linear Poisson regression model with parameters* $\log \lambda_j$ *given by* $\log \lambda(u_j, v_j) - \theta^T S(u_j, v_j)$*, responses* y_j *and weights* w_j*.*

The coefficient estimates returned by Algorithm 2 give the maximum pseudo-likelihood estimator $\hat{\theta}$ for θ.

In order to estimate the parameters $\theta = (\gamma_1, \gamma_2, \ldots, \gamma_m)$ using the above method we need to have values for the *irregular* parameters r_j and t_j for $j = 1, \ldots, m$. Baddeley and Turner (2000) suggest fitting the model for a range of values of these parameters and choose the values which maximize the pseudo-likelihood. Additionally, we recommend first to compute some summary statistics, such as the pair correlation or auto-correlation function, to narrow down the search.

We construct the quadrature scheme as a partition of $W_S \times W_T$ dividing the spatio-temporal area into cubes C_k of equal volume. In the centre of each cube C_k we place exactly one dummy point. We then assign to each dummy or data point (u_j, v_j) a weight $w_j = v/n_j$ where v is the volume of each cube, and n_j is the number of points, dummy or data, in the same cube as (u_j, v_j). These weights are called the counting weights (Baddeley and Turner, 2000).

We conclude this section by mentioning briefly an alternative way to define the quadrature scheme (Algorithm 2). Indeed, Baddeley and Turner (2000) suggest the use of a Dirichlet tessellation to generate the quadrature weights. A quadrature scheme generated this way would mean that the weight of each point would be equal to the volume of the corresponding Dirichlet 3-dimensional cell. The computational cost of such a method is very high. Therefore we partition $W_S \times W_T$ into cubes of equal volume, as described above.

4.5.2 Simulation and parameter estimation of a spatio-temporal area interaction process

For illustration purposes, we simulate two multi-scale spatio-temporal area interaction processes as defined in (4.3), one which exhibits small scale inhibition and large scale clustering (*simulation 1*) and a second one which exhibits small scale clustering and large scale inhibition (*simulation 2*).

We consider the spatio-temporal domain $W_S \times W_T = ([0,1] \times [0,1]) \times [0,1]$ and in both cases take constant $\lambda \equiv 50$. For the irregular parameters we choose the same spatio-temporal scales $r_1 = 0.03$, $r_2 = 0.05$, $t_1 = 0.03$ and $t_2 = 0.05$ for both simulations. We use the Metropolis-Hastings algorithm described in Section 4.4.1 with $20,000$ iterations implemented in the `MPPLIB C++` library (Steenbeek et al., 2016). To estimate the parameters we follow the steps in Algorithm 2. We partition $W_S \times W_T$ into $10^3 = 1,000$ cubes of volume 10^{-3}. In the center of each cube we place a dummy point, obtaining a total of $1,000$ dummy points. We then compute the sufficient statistics for each data and dummy point using the `MPPLIB C++` library and apply Algorithm 2 to obtain the estimates for the parameters. For the implementation of the pseudo-likelihood method we use the statistical software `R` (R Core Team, 2014) together with the `spatstat` (Baddeley et al., 2015) package. The theoretical background for computing the 'envelopes', that is the confidence interval bounds given as 2.5% and 97.5% in Tables 4.1 and 4.2 for a Poisson process is exhaustively described in Kutoyants (1998).

Figure 4.3 (top left) shows the interaction parameters for *simulation 1*, $2\pi r_1^2 t_1 \log(\gamma_1) = -5$ and $2\pi r_2^2 t_2 \log(\gamma_2) = 5$. This setting of parameters gives us the spatio-temporal point configuration shown in the top right panel of Figure 4.3 which indeed shows small scale inhibition between points and large scale clustering. The parameter estimates are shown in Table 4.1.

	Estimate	2.5 %	97.5 %
$\log \lambda$	6.07	3.57	8.02
$2\pi r_1^2 t_1 \log(\gamma_1) = -5$	-2.45	-5.48	0.37
$2\pi r_2^2 t_2 \log(\gamma_2) = 5$	4.48	2.44	6.48

Table 4.1. Parameter estimates for simulation 1

For *simulation 2* we choose interaction parameters $2\pi r_1^2 t_1 \log(\gamma_1) = 5$ and $2\pi r_2^2 t_2 \log(\gamma_2) = -5$, as shown in the bottom left panel of Figure 4.3. The bottom right panel of Figure 4.3 shows a realization of the process with these parameters. We observe small scale clustering and large scale inhibition between points. The estimates of the parameters are given in Table 4.2.

Note that Figure 4.3 and Tables 4.1–4.2 correspond to a single realization of the multi-scale area-interaction model and one should be hesitant to draw any conclusions on the efficacy or otherwise of the pseudo-likelihood method from this illustration.

	Estimate	2.5 %	97.5 %
$\log \lambda$	8.25	4.56	10.50
$2\pi r_1^2 t_1 \log(\gamma_1) = 5$	7.17	2.69	12.00
$2\pi r_2^2 t_2 \log(\gamma_2) = -5$	-2.39	-6.40	1.19

Table 4.2. Parameter estimates for simulation 2

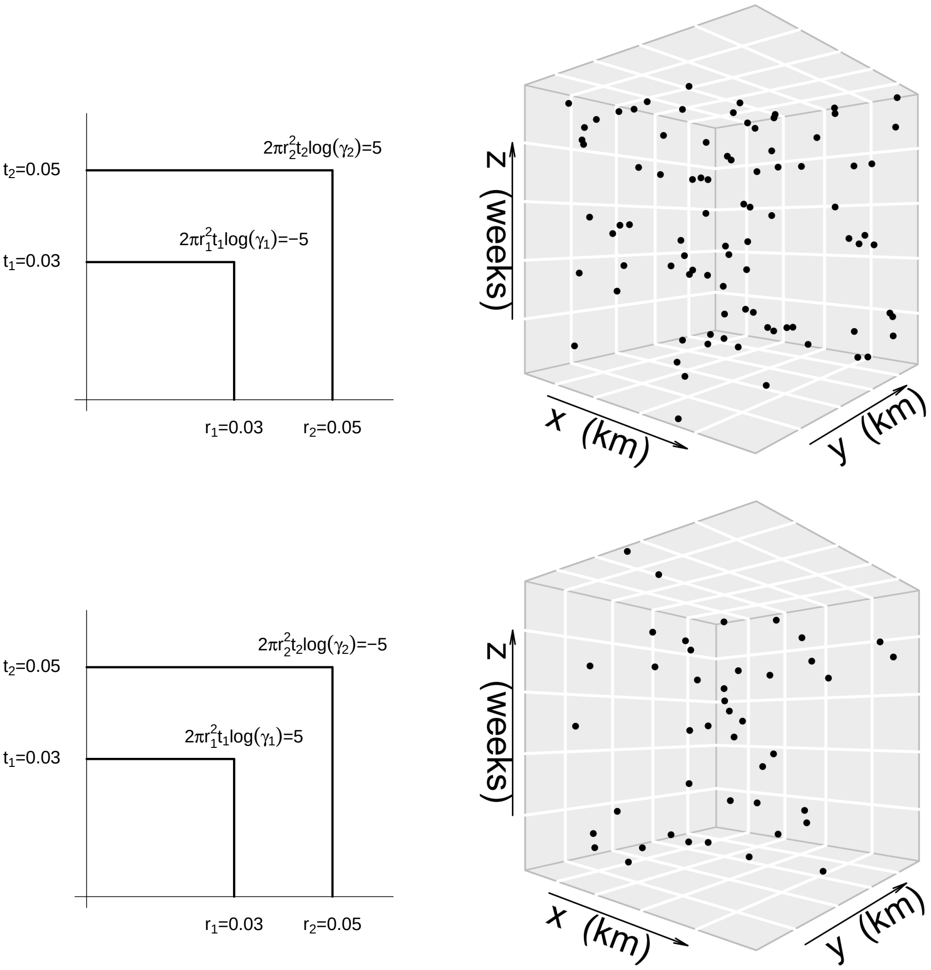

Figure 4.3. (Top left) Model parameters for *simulation 1.* (Top right) A realization of the first model. (Bottom left) Model parameters for *simulation 2.* (Bottom right) A realization of the second model.

4.6 Application. Varicella data

In this chapter we analyse varicella cases registered in Valencia, Spain, during 2013. The spatial coordinates of the varicella cases are expressed in latitude and longitude. First we transform them from longitude/latitude to UTM scale expressed in meters (Snyder, 1987). We then re-scale the spatial coordinates to kilometers such that the spatial study area reduces to $[0, 9] \times [0, 9]$. The temporal component of the process takes values from 0 to 51. For computational purposes to be explained later, we take the interval $[0, 52]$ as the time window. Therefore, we set the spatio-temporal study area to $W_S \times W_T = ([0, 9] \times [0, 9]) \times [0, 52]$ (km$^2\times$ weeks). The spatio-temporal pattern of all varicella cases thus obtained is shown in Figure 4.4. The x- and y-axis represent the spatial coordinates in kilometres and the z-axis represents the time component in weeks.

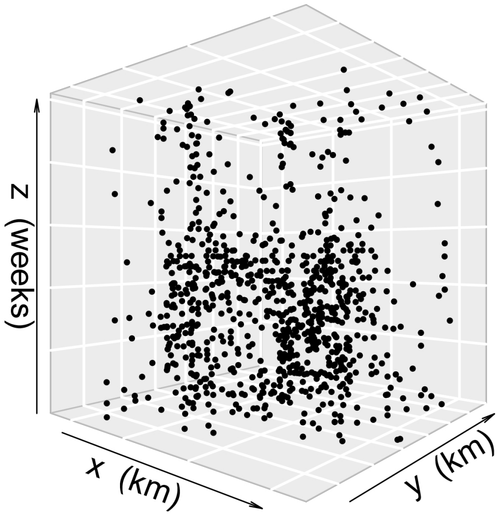

Figure 4.4. Spatio-temporal pattern of weekly varicella cases in Valencia during 2013, where the spatio-temporal study area is $W_S \times W_T = ([0, 9] \times [0, 9]) \times [0, 52]$ (km$^2\times$ weeks).

The main focus of our varicella data analysis is to quantify the interactions across a range of spatio-temporal scales. We do so by using the spatio-temporal multi-scale area-interaction model introduced in Section 4.3.

First we need to get some idea about a plausible upper bound to the values of the irregular parameters (r_j, t_j), $j = 1, \ldots, m$, in model (4.3). To this end, we use summary statistics for the spatial and temporal projections of the space-time point pattern shown in Figure 4.4.

The left panel in Figure 4.5 shows the projection of all points onto the spatial region. The sizes of the circles are proportional to time, the bigger the circle, the more recent the event. Due to the projection, duplicate locations are observed, so we jitter the coordinates uniformly on the spatial region around the duplicated points using a maximum jittering distance of 20 meters. To get a rough indication of the spatial interaction range, we pretend that the pattern is stationary and isotropic, and estimate the pair correlation function. The result is shown in the right panel of Figure 4.5. Recall that for a Poisson process the pair correlation function is equal to 1. Values of the pair correlation function lower than 1 indicate inhibition and values larger than 1 suggest clustering. Figure 4.5 suggests that the pair correlation function is approximately constant from 2 kilometers onward, which indicates a maximum value for the r_i of around 1 kilometer. On a cautionary note, we need to keep in mind that the estimator only takes into account the spatial pattern of points and assumes isotropy.

The left panel in Figure 4.6 shows the temporal evolution of varicella over the 52 weeks, where the small circles $\circ$ represent the number of registered cases. The right panel displays the estimated auto-correlation function which measures the correlation between the values of the series at different times as a function of the time lag between them. Figure 4.6 suggests possible correlation for time lags as big as 10 weeks, however, we take a larger time lag of 15 weeks. This gives us an estimate for the maximum value for the t_i of about 7.5 weeks. Note that caveats similar to the spatial case apply.

Now that we have estimated the maximum spatial and temporal range for the model, the following step in our analysis is to consider covariate information. The most important factor in the transmission of any kind of disease, and especially a highly contagious one such as varicella, is the population. In areas with very low population we will probably not register as many varicella cases as in highly populated areas. Thus, the pattern of varicella cases can drastically change from one area to another, depending on the spatial distribution of the population, and from one week to the next one. We express the spatio-temporal inhomogeneity term in equation (4.3) as a product $\lambda(x,t) = \lambda(x)\, Z(t)$, $x \in [0,9]^2$, $t \in \{0,\ldots,51\}$, between a non-parametric estimate of the population density $\lambda(x)$ and a re-scaled parametric estimate of the temporal component $Z(t)$.

First consider the spatial component $\lambda(x)$. The population data available to us consist of the number of people living in each census section of the city of Valencia, a total number of 559 sections (districts 1 to 16). We randomly generate within each section p points, where p is equal to the number of people living in that particular section. This way, we obtain a sample of the population for the city of Valencia. We estimate its intensity by a kernel estimator, keeping in mind that the bandwidth has to be chosen carefully, to get $\lambda(x)$, $x \in W_S$.

We fit a harmonic regression (Halberg et al., 1967) to the pattern of the weekly varicella counts

$$Z(t) = c_0 + \sum_{j=1}^{3} \left(c_j \cos(2\pi jt/52) + d_j \sin(2\pi jt/52)\right) + c(a+bt), \tag{4.20}$$

where $Z(t)$ denotes the number of varicella cases at time t, $t = 0,\ldots,51$, and c_0, a, b, c, c_j, d_j, $j = 1,2,3$, are the parameters of the model.

The left panel in Figure 4.6 shows the fitted regression curve. We observe a period at the beginning of the year, from winter until spring, with large numbers of varicella cases, and a second period starting around week 26, in which the number of cases decreases.

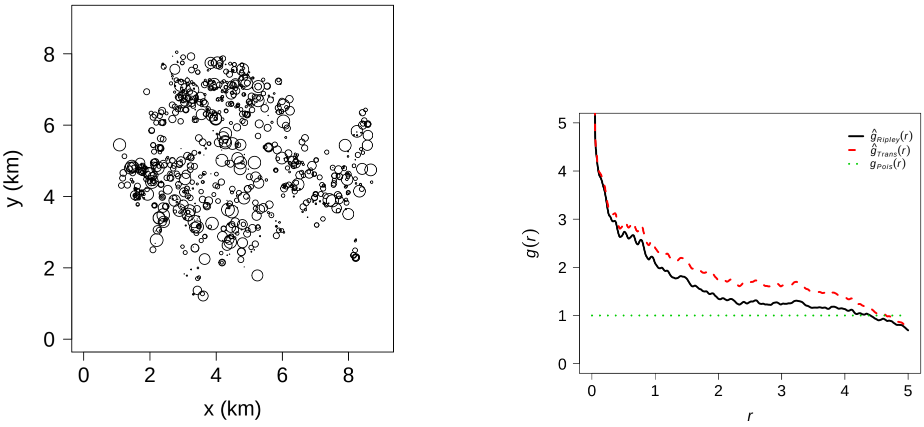

Figure 4.5. (Left) Spatial projection of the spatio-temporal point pattern for the varicella data. After projection, locations were jittered using a maximum jitter distance of 20 metres. (Right) Estimated pair correlation function for the jittered spatial point pattern shown in the left panel.

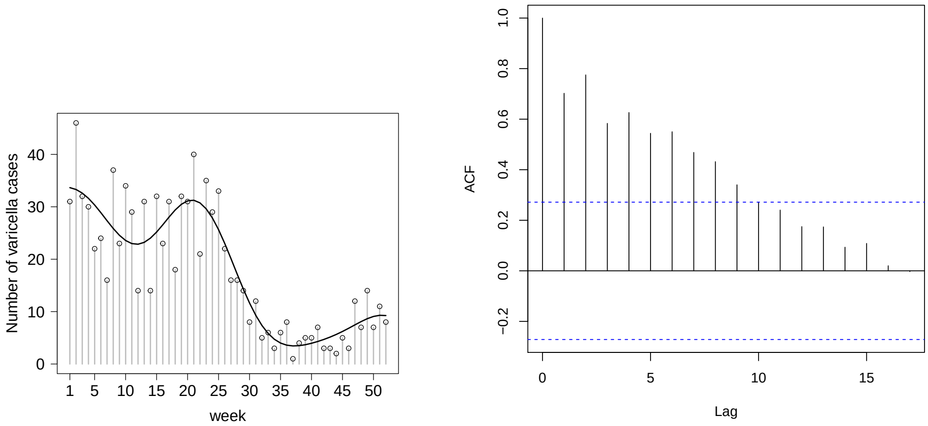

Figure 4.6. (Left) Weekly reports of varicella cases (○) and fitted regression curve (–). (Right) Estimated auto-correlation function for the data shown in the left panel.

These periods correspond roughly with the school term and the summer break. Also, in 2013, in Spain, there were several holidays besides the summer and winter holidays. On March 19, San Jose is celebrated and the period from the 24th to the 31st of March corresponds to the Easter holidays. As a consequence we can observe in Figure 4.6 a decrease during the 11th and 12th week. Towards the end of the year, the number of cases picks up again as the Michaelmas term begins.

Finally, we re-scale the parametric estimate of the temporal component $Z(t)$ by 100, in order to avoid obtaining extreme values for the spatio-temporal inhomogeneity term $\lambda(x,t)$.

Since realizations of (4.3) do not contain points with equal time stamps, we jitter in time as well as space. More precisely, the week index is replaced by a time stamp that is uniformly distributed in the indicated week so that the temporal component falls in $W_T = [0, 52]$. To estimate the parameters, we follow the steps described in Algorithm 2. For constructing the quadrature points we partition $W_S \times W_T$ into $9 \times 9 \times 52 = 4,212$ cubes of equal volume 1 and place one dummy point in the center of each cube. Doing so, we obtain a total of $5,133$ dummy and data points. We attribute to each point a weight equal to the volume of the cube divided by the number of dummy and data points inside the cube containing the point. We then compute the sufficient statistics $S_j(\cdot,\cdot)$ corresponding to each point using the `MPPLIB C++` library of Steenbeek et al. (2016). We follow Algorithm 2 and obtain estimates for the parameters γ. The analysis and visual representations have been carried out using the statistical software `R` (R Core Team, 2014) together with the `spatstat` (Baddeley et al., 2015), `plot3D` (Soetaert, 2016) and `rgdal` (Bivand et al., 2016) packages.

Recall that we found indications for the maximum spatial range to be about 2 kilometers, the maximum temporal range 15 weeks. As suggested in Baddeley and Turner (2000) we fitted the model for a range of values (r_j, t_j), $j = 1, \dots, m$, in the larger domain $[0, 2] \times [0, 15]$ and for different $m \in \{3, 4, 5, 6, 7, 8\}$ to choose the optimal combination.

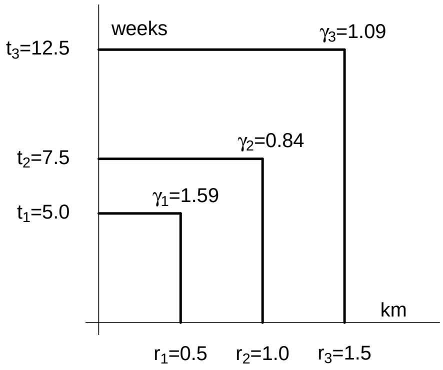

Figure 4.7. Model parameters for the varicella data.

We estimate $m = 3$, that is, three spatio-temporal scales and the corresponding parameters. For the spatial scales we selected $r_1 = 0.5, r_2 = 1$ and $r_3 = 1.5$ kilometers and for the temporal scales $t_1 = 5$, $t_2 = 7.5$ and $t_3 = 12.5$ weeks. Figure 4.7 shows the multi-scale interaction in the data together with the estimated values of the model parameters. Also, Table 4.3 shows the estimated parameters of the model together with a confidence interval and a significance level.

	Spatial scale (r_i)	Time scale (t_i)	Parameters (γ_i)	2.5%	97.5%
(Intercept)			1.20	1.09	1.31
	0.5	5.0	1.57	1.39	1.78
	1.0	7.5	0.84	0.74	0.95
	1.5	12.5	1.10	1.00	1.23

Table 4.3. Parameter estimates for the varicella data.

As stated before, the time period from infection to completely dried vesicles is between approximately 17 and 31 days. In the fitted model we observe that for a spatial lag of 0.5 kilometer and a temporal lag of 5 weeks there is clustering (significant $\gamma_1 = 1.57$). This means that for a period of five weeks and at rather small distance (as far as 0.5 kilometers), a phenomenon of aggregation is observed between cases of varicella. The time lag corresponds more or less with the period of 31 days indicated by the epidemiologists. This is caused by the main feature of chickenpox, being a contagious disease.

The fitted model also exhibits inhibition for spatial lags as far as 1 kilometer and temporal lags up to 7.5 weeks (significant $\gamma_2 = 0.84$). This might be a result of the fact that after recovery from varicella, patients usually have lifetime immunity. For higher spatial and temporal lags the model suggests no interaction ($\gamma_3 \approx 1$), which corresponds with the rough bounds we found before and is in accordance with the beliefs of the epidemiologists. If you are situated far away from a varicella case, both in space and in time, you are less susceptible to contract the disease due to the contagious factor. Also, the probability of contracting the disease would be the same as the incidence of varicella.

To validate our model, we simulated a number of space-time multi-scale area-interaction processes with the fitted parameters using the Metropolis-Hastings algorithm described in Section 4.4.1 for $20,000$ iterations, which seems enough for the algorithm to converge based on diagnostic plots. Figure 4.8 (left) shows one such simulation. Comparing Figure 4.8, left and right, we note that the simulated spatio-temporal point pattern is similar to the varicella point pattern.

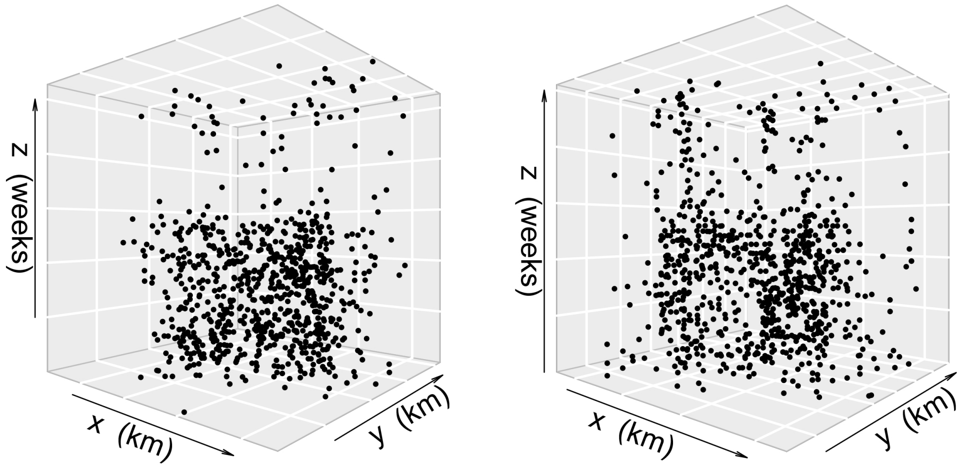

Figure 4.8. (Left) Realization from the model fitted to the varicella data. (Right) Spatio-temporal pattern of weekly varicella cases in Valencia during 2013, where the spatio-temporal study area is $W_S \times W_T = ([0,9] \times [0,9]) \times [0,52]$ (km$^2\times$ weeks).

4.7 Conclusions

In this chapter we developed an extension of the area-interaction model that is able to incorporate different types of interaction at different spatio-temporal scales and proposed methods to simulate this process. We discussed inference and demonstrated the pseudo-likelihood method on simulated data. Additionally, we analysed a spatio-temporal point pattern of varicella in the city of Valencia, Spain.

Chapter 5

The second-order analysis of marked spatio-temporal point processes

This chapter represents a significant methodological contribution to the field of point processes, and in particular to the field of marked spatio-temporal point processes. Our contribution introduces measures of second-order spatio-temporal interaction, which quantify the interaction between categories of marked points.

The earthquake dataset presented in Section 5.6.1 has largely motivated the development of the methodology in this chapter. We first introduce the notion of marked spatio-temporal point process (MSTPP), together with a summary on mark spaces, reference measures and intensity functions. Next we give examples of some MSTPP models which will be used for evaluation throughout the chapter. In Section 5.4, we introduce second-order intensity-reweighted stationarity for MSTPP and define the marked spatio-temporal second-order reduced moment measure together with the marked spatio-temporal inhomogeneous K-function, $K_{\text{inhom}}^{CD}(r,t)$. In Section 5.4 we also provide some representation results. We also propose estimators for intensity functions (a Voronoi tessellation based approach), as well as for the new second-order summary statistics. In addition, we consider ideas for testing independence assumptions of the marks. The final section of the chapter presents the second-order analysis of the earthquake dataset.

5.1 Introduction

Classically, when analysing (marked) STPPs, the analysis has been based on *conditional intensity functions (CIs)* (see e.g. (Ogata, 1998; Choi and Hall, 1999; Schoenberg et al., 2002; Daley and Vere-Jones, 2003; Marsan and Lengliné, 2008)). In principle, a conditional intensity function gives us the expected number of further events in a coming infinitesimal period, given the history of events up to that point. The beauty and appeal of CIs is that, when existing, they specify the whole distribution of the MSTPP.

As pointed out by e.g. Diggle (2014), however, not all MSTPP models have available/-tractable CIs. Furthermore, much of the CI-based analysis is carried out within the framework of a given class of models.

Recalling that we want to define a general fully non-parametric analysis, we will proceed with a non-CI based approach, thus following a random set/random measure formulation (see e.g. (Van Lieshout, 2000; Daley and Vere-Jones, 2003; Møller and Waagepetersen, 2004; Chiu et al., 2013; Diggle, 2014)). In this context, when analysing marked spatio-temporal point patterns, the first thing one starts with is to try to explain where and when events of a given mark category of the data tended to happen. Since where and when is a univariate property, in the sense that we are not dealing explicitly with possible dependencies between the points, we are dealing with analysing *intensity*. Before proceeding to propose specific models for the intensity structure, through the observed point pattern, one usually starts by obtaining a non-parametric estimate of the *intensity function* (see e.g. (Diggle, 2014)). The intensity function, in essence, reflects the infinitesimal probability of finding a point of the MSTPP at a given spatial location, at a given time, with a given mark. Note that it is different from the previously mentioned conditional intensity, which is defined as a conditional equivalence. In the simplest of worlds, we would simply assume homogeneity, i.e. that univariately it is equally likely to observe an event, with any mark, at any space-time position. This is, however, not the slightest realistic so we proceed by assuming inhomogeneity. Although the most natural candidate for this type of non-parametric estimation is kernel estimation (Silverman, 1986; Van Lieshout, 2012; Diggle, 2014), due to the abrupt changes in activity of the earthquakes, both spatially and temporally, we here make the choice to consider an adaptive approach, namely a *Voronoi intensity estimation* approach (see e.g. (Barr and Schoenberg, 2010)).

Having obtained a non-parametric estimate of the intensity function, so that we have a description of the univariate properties, we may proceed to studying the inherent dependence structure of the data-generating mechanism, i.e. the underlying MSTPP. We here focus on second-order summary statistics, thus ignoring higher-order summary statistics, e.g. the spatio-temporal J-function and its components (Cronie and van Lieshout, 2015) and the marked J-functions and their components (Van Lieshout, 2006; Cronie and van Lieshout, 2016). In the context of unmarked spatio-temporal point processes, Diggle et al. (1995) extended Ripley's K-function $K(r)$ (Ripley, 1976; Ripley and Kelly, 1977) to the stationary spatio-temporal context. Recall that this function, $K(r,t)$, gives us the expected number of further space-time points from an arbitrary space-time point of the process, given that the points in question have space and time separation $r \geq 0$ and $t \geq 0$, respectively. After the introduction of the spatial inhomogeneous K-function $K_{\text{inhom}}(r)$ (Baddeley et al., 2000), which is defined as an integral of the pair correlation function, Gabriel and Diggle (2009) extended its definition to the spatio-temporal context, resulting in the function $K_{\text{inhom}}(r,t)$. Note that under inhomogeneity, given only one realisation, we cannot e.g. visually distinguish between regions of high intensity and clustering/aggregation. For general marks in the purely spatial setting, Van Lieshout (2006) defined a marked version $K^{CD}(r)$ of Ripley's

K-function; loosely speaking it gives us Ripley's K-function under the condition that we restrict the interaction to take place between points with marks belonging to some mark set (category) C and points with marks in a mark set D. In addition, inspired by Van Lieshout (2006), Cronie and van Lieshout (2016) introduced a marked version of the inhomogeneous K-function, $K_{\text{inhom}}^{CD}(r)$, which reduces to the multivariate version introduced in Møller and Waagepetersen (2004), when we assume that the marks are integer-valued (a multivariate/multi-type inhomogeneous point process). It reduces to the one in Van Lieshout (2006) when we assume stationarity. Our contribution combines the ideas of Gabriel and Diggle (2009) with those of Cronie and van Lieshout (2016) to define a K-function $K_{\text{inhom}}^{CD}(r,t)$ for inhomogeneous MSTPPs, which reduces to a combination of $K(r,t)$ and $K^{CD}(r)$ when we assume stationarity. Loosely speaking, $K_{\text{inhom}}^{CD}(r,t)$ describes the interaction, in a (Gabriel and Diggle, 2009) sense, between points belonging to mark set C and points belonging to mark set D, for an inhomogeneous MSTPP. Note that for all summary statistics above, one of the main foci has been to consider their non-parametric estimation. Here, as well we will allocate a significant part of this chapter to the non-parametric estimation. Having developed $K_{\text{inhom}}^{CD}(r,t)$ and its estimation schemes, it turns out that we may also devise some statistical testing procedures which we will also look a bit closer at.

Once we have developed the statistical tools, we analyse the earthquake data with the aim of quantifying the interactions, so that we may asses the space-times propagations of the shocks.

5.2 Marked spatio-temporal point processes

In order to formally define a *marked spatio-temporal point process* Y, with locations x_i in $\mathbb{R}^d$, event times t_i in $\mathbb{R}$ and marks m_i in some suitable mark space $\mathcal{M}$, there are some technical details that need to be tended to.

5.2.1 Preliminaries

The most natural way of measuring distances in $\mathbb{R}^d$ is provided by the Euclidean metric $d_{\mathbb{R}^d}(x,y) = \|x-y\|_{\mathbb{R}^d}$, $\|x\|_{\mathbb{R}^d} = (\sum_{i=1}^d x_i^2)^{1/2}$, $x, y \in \mathbb{R}^d$. Hence, we measure distances between spatial locations by means of $d_{\mathbb{R}^d}(\cdot,\cdot)$ and between temporal locations by means of $d_{\mathbb{R}}(\cdot,\cdot)$, i.e. absolute values. To combine the spatial and the temporal distances in a good way, such that we treat space and time differently, we endow our space-time domain $\mathbb{R}^d \times \mathbb{R}$ with the supremum norm $\|(x,t)\|_\infty = \max\{\|x\|_{\mathbb{R}^d}, |t|\}$ and the supremum metric

$$\begin{aligned} d_\infty((x,t),(y,s)) = \|(x,t)-(y,s)\|_\infty &= \max\{d_{\mathbb{R}^d}(x,y), d_{\mathbb{R}}(t,s)\} \\ &= \max\{\|x-y\|_{\mathbb{R}^d}, |t-s|\}, \end{aligned}$$

where $(x,t),(y,s) \in \mathbb{R}^d \times \mathbb{R}$. Hereby, we have combined two complete separable metric (csm) spaces, into the spatio-temporal csm space $(\mathbb{R}^d \times \mathbb{R}, d_\infty(\cdot,\cdot))$ (Daley and Vere-Jones, 2003).

Note that the d_∞-induced Borel σ-algebra $\mathcal{B}(\mathbb{R}^d \times \mathbb{R}) = \mathcal{B}(\mathbb{R}^d) \otimes \mathcal{B}(\mathbb{R})$, the product σ-algebra, since the underlying space is csm. Following Cronie and van Lieshout (2015), we define a spatio-temporal point process as a simple point process in $(\mathbb{R}^d \times \mathbb{R}, \mathcal{B}(\mathbb{R}^d \times \mathbb{R}))$.

Definition 3. *A spatio-temporal point process with spatial locations in $\mathbb{R}^d$ and event times in $\mathbb{R}$ is a point process in $(\mathbb{R}^d \times \mathbb{R}, \mathcal{B}(\mathbb{R}^d \times \mathbb{R}))$.*

Remark 1. *If we would endow $\mathbb{R}^d \times \mathbb{R} = \mathbb{R}^{d+1}$ with the Euclidean distance*

$$d_{\mathbb{R}^{d+1}}((x,t,(y,s)) = \|(x,t)-(y,s)\|_{\mathbb{R}^{d+1}} = ((t-s)^2 + \sum_{i=1}^{d}(x_i^2 - y_i^2))^{1/2},$$

we would encounter the problem that space and time are not treated differently. Indeed, this space is topologically equivalent to $(\mathbb{R}^d \times \mathbb{R}, d_{\mathbb{R}^{d+1}}(\cdot,\cdot))$ and we note that there are other (less natural) ways of combining $\|\cdot\|_{\mathbb{R}^d}$ and $|\cdot|$ such that $\mathbb{R}^d \times \mathbb{R}$ becomes a csm space.

Depending on what kind of mark types we want to consider for the data, in the construction of the related MSTPP model we have to choose an appropriate mark space and for the purpose of integration also appropriate associated reference measure. E.g., having recorded earthquakes we may either partition the magnitude scale, so that we consider a *multivariate STPP* (see Section 5.4.5), or treat the marks as continuous. Depending on the choice, the statistical analyses differ so the choice made can be important.

Formally, regarding the mark space $\mathcal{M}$, we assume that it is a *complete separable metric* (csm) space with corresponding metric $d'(\cdot,\cdot)$ and Borel sets $\mathcal{B}(\mathcal{M})$. We equip $(\mathbb{R}^d \times \mathbb{R}) \times \mathcal{M}$ with the Borel sets $\mathcal{B}((\mathbb{R}^d \times \mathbb{R}) \times \mathcal{M})$, which become the product σ-algebra $\mathcal{B}(\mathbb{R}^d) \otimes \mathcal{B}(\mathbb{R}) \otimes \mathcal{B}(\mathcal{M})$. It is mostly natural to generate this structure through the metric

$$\begin{aligned} d((x_1,t_1,m_1),(x_2,t_2,m_2)) &= \max\{d_\infty((x_1,t_1),(x_2,t_2)), d'(m_1,m_2)\} \\ &= \max\{\|x_1 - x_2\|_{\mathbb{R}^d}, |t_1 - t_2|, d'(m_1,m_2)\}, \end{aligned} \tag{5.1}$$

where $(x_1,t_1,m_1),(x_2,t_2,m_2) \in (\mathbb{R}^d \times \mathbb{R}) \times \mathcal{M}$.

For a general discussion on mark spaces and their metric structures, see e.g. (Van Lieshout, 2000), (Møller and Waagepetersen, 2004, Appendix B.3) and (Chiu et al., 2013). When e.g. $\mathcal{M} \subseteq \mathbb{R}^l$, $l \geq 1$, we let $d'(m_1,m_2) = \|m_1 - m_2\|_{\mathbb{R}^l}$. In the case of our application, naturally we will consider $\mathcal{M} \subseteq \mathbb{R}$, i.e. $d'(m_1,m_2) = |m_1 - m_2|$. For the case where $\mathcal{M}$ is a finite collection of labels $\{1,\dots,k\}$, $k \geq 2$, (let $d'(m_1,m_2) = |m_1 - m_2|$), this is referred to as the multivariate/multi-type case.

For the purpose of e.g. integration over $(\mathbb{R}^d \times \mathbb{R}) \times \mathcal{M}$, we need to endow the underlying space $((\mathbb{R}^d \times \mathbb{R}) \times \mathcal{M}, \mathcal{B}((\mathbb{R}^d \times \mathbb{R}) \times \mathcal{M}))$ with a reference measure. The choice of reference measure may seem as a mathematical detail and of little practical importance at first, but it will become clear that it plays a significant role also in the statistical analysis. A fact that is often overlooked in statistical settings.

We will choose as reference measure the product reference measure

$$\ell \otimes \nu = \ell_{d+1} \otimes \nu = \ell_d \otimes \ell_1 \otimes \nu,$$

where ℓ_d is the Lebesgue measure on $\mathbb{R}^d$, $d \geq 1$, and ν is some suitable finite reference measure on the mark space. Throughout, $(\ell \otimes \nu)^n$ will represent the n-fold product measure of $\ell \otimes \nu$ with itself.

When well-defined, we write

$$\int f(x,t,m)[\ell \otimes \nu](d(x,t,m)) = \int \int \int f(x,t,m)\nu(dm)\mathrm{d}t\mathrm{d}x$$

for the integral of some $f : (\mathbb{R}^d \times \mathbb{R}) \times \mathcal{M} \to \mathbb{R}$. When $\mathcal{M} = \mathbb{R}^k$, $k \geq 1$, it is reasonable to choose $\nu(\cdot)$ as some suitable probability law and when $\mathcal{M} \subseteq \mathbb{R}^k$, $k \geq 1$, is bounded, we may simply let $\nu(\cdot) = \ell_k(\cdot)$ (or normalise to have a uniform distribution as reference measure). For other mark spaces, see e.g. (Chiu et al., 2013).

5.2.2 Marked spatio-temporal point processes

In analogy with Cronie and van Lieshout (2015), let the *unmarked/ground process* (Daley and Vere-Jones, 2003) Y_g of space-time events (x_i, t_i) be given by a spatio-temporal point process (STPP), as defined in Definition 3. Informally, we assign marks $m_i \in \mathcal{M}$ (random variables) to the points of Y_g to obtain the marked spatio-temporal point process Y.

More formally, consider first the collection N_{lf} of all simple non-negative integer valued measures $\varphi(\cdot) = \sum_{i=1}^{n} \delta_{(x_i,t_i,m_i)}(\cdot) = \sum_{i=1}^{n} \mathbf{1}\{(x_i,t_i,m_i) \in \cdot\}$, $0 \leq n \leq \infty$ ($n = 0$ corresponds to the null measure), on $\mathcal{B}((\mathbb{R}^d \times \mathbb{R}) \times \mathcal{M})$ which are locally finite, i.e. $\varphi(B \times C) < \infty$ for bounded $B \times C \in \mathcal{B}((\mathbb{R}^d \times \mathbb{R}) \times \mathcal{M})$, with the additional property that the (spatio-temporal) ground measure $\varphi_g(\cdot) = \varphi(\cdot \times \mathcal{M})$ is locally finite on $\mathcal{B}(\mathbb{R}^d \times \mathbb{R})$. Note that the term simple refers to $\varphi(\{(x,t,m)\}) \in \{0,1\}$ for any $(x,t,m) \in (\mathbb{R}^d \times \mathbb{R}) \times \mathcal{M}$. The support of such a measure $\varphi(\cdot) \in N_{lf}$ will also be denoted by φ, hence, $\varphi = \{(x_i,t_i,m_i)\}_{i=1}^{n} \subseteq (\mathbb{R}^d \times \mathbb{R}) \times \mathcal{M}$.

Let $\mathcal{N}$ be the smallest σ-algebra on N_{lf} to make the mappings $\varphi \mapsto \varphi(A) \in \{0,1,\ldots\}$ measurable for bounded $A \in \mathcal{B}((\mathbb{R}^d \times \mathbb{R}) \times \mathcal{M})$. Denoting also the collection of all supports by N_{lf}, we note that there analogously exists a σ-algebra $\mathcal{N}$ that is generated by the mappings $\varphi \mapsto |\varphi \cap A| \in \{0,1,\ldots\}$, for bounded $A \in \mathcal{B}((\mathbb{R}^d \times \mathbb{R}) \times \mathcal{M})$ and all supports φ.

Definition 4. *A marked spatio-temporal point process (MSTPP)* $Y(\cdot) = \sum_{i=1}^{N} \delta_{(x_i,t_i,m_i)}(\cdot)$, $0 \leq N \leq \infty$, *is a measurable mapping from some probability space* $(\Omega, \mathcal{F}, \mathbb{P})$ *into the measurable space* $(N_{lf}, \mathcal{N})$. *If* $N < \infty$ *almost surely (a.s.) then* Y *is called a* finite *MSTPP.*

By the above arguments we may treat a MSTPP Y as a random measure as well as a random subset $Y = \{(x_i,t_i,m_i)\}_{i=1}^{N}$ of $(\mathbb{R}^d \times \mathbb{R}) \times \mathcal{M}$ and thus conveniently jump between the two notions. By this duality, $Y(A)$ and $|Y \cap A|$ may both be used to denote the cardinality of the number of points of Y belonging to $A \in \mathcal{B}((\mathbb{R}^d \times \mathbb{R}) \times \mathcal{M})$.

Note that by definition the ground process $Y_g = \{(x_i, t_i)\}_{i=1}^N$ is a well defined (spatio-temporal) point process on $\mathbb{R}^d \times \mathbb{R}$. We also write $P(\cdot) = \mathbb{P}(Y \in \cdot)$ for the distribution of Y, i.e. the probability measure that Y induces on $(N_{lf}, \mathcal{N})$.

If $\{(x_i, t_i)\}_{i=1}^N = Y_g \overset{d}{=} Y_g + (a, b) = \{(x_i + a, t_i + b)\}_{i=1}^N$ for any $(a, b) \in \mathbb{R}^d \times \mathbb{R}$, where $\overset{d}{=}$ denotes equality in distribution, we say that Y is stationary (Chiu et al., 2013; Daley and Vere-Jones, 2003). In practise stationarity is rarely realistic.

5.2.3 Intensity functions

Let Y be a MSTPP with ground process Y_g. We will next consider the joint distributional properties of the points of Y, which we describe through the so-called product densities.

For any $n \geq 1$, assume that the nth *factorial moment measure* $\alpha^{(n)}(\cdot)$ of Y exists (as a locally finite measure on $\mathcal{B}((\mathbb{R}^d \times \mathbb{R}) \times \mathcal{M})^n)$ and assume that $\alpha^{(n)}$ is absolute continuous with respect to $(\ell \otimes \nu)^n$. Then its permutation invariant Radon-Nikodym derivative $\rho^{(n)}(\cdot) \geq 0$ (Chiu et al., 2013; Daley and Vere-Jones, 2003; Diggle, 2014), the so-called nth *intensity function/product density/factorial moment density*, may be defined through the so-called *Campbell formula*: For any measurable function $f \geq 0$,

$$\begin{aligned}
&\mathbb{E}\left[\sum\nolimits^{\neq}_{(x_1,t_1,m_1),\ldots,(x_n,t_n,m_n)\in Y} f((x_1,t_1,m_1),\ldots,(x_n,t_n,m_n))\right] = \qquad (5.2)\\
&= \int \cdots \int f((x_1,t_1,m_1),\ldots,(x_n,t_n,m_n)) \times \\
&\times \rho^{(n)}((x_1,t_1,m_1),\ldots,(x_n,t_n,m_n)) \prod_{i=1}^n \nu(dm_i)\mathrm{d}x_i\mathrm{d}t_i,
\end{aligned}$$

which includes the case where both sides are infinite. Here $\sum^{\neq}$ denotes summation over n-tuples $((x_1,t_1,m_1),\ldots,(x_n,t_n,m_n))$ of distinct points. Regarding the interpretation of $\rho^{(n)}(\cdot)$, by the simpleness of Y,

$$\begin{aligned}
&\mathbb{P}(Y(d(x_1,t_1,m_1)) = 1,\ldots,Y(d(x_n,t_n,m_n)) = 1) = \\
&= \rho^{(n)}((x_1,t_1,m_1),\ldots,(x_n,t_n,m_n)) \prod_{i=1}^n \nu(dm_i)\mathrm{d}x_i\mathrm{d}t_i.
\end{aligned}$$

This is the infinitesimal probability of observing points of Y_g in the space-time neighbourhoods $d(x_i,t_i) \subseteq \mathbb{R}^d \times \mathbb{R}$ of (x_i,t_i), with associated marks $m_i \in dm_i \subseteq \mathcal{M}$, where $[\ell \otimes \nu](d(x_i,t_i,m_i)) = \nu(dm_i)\ell(d(x_i,t_i)) = \nu(dm_i)\mathrm{d}x_i\mathrm{d}t_i$, $i = 1,\ldots,n$. Note that $\rho^{(n)}(\cdot)$ does not give us the joint density of *all* points of Y, unless we condition on the total number of points $Y((\mathbb{R}^d \times \mathbb{R}) \times \mathcal{M}) = N = n$ (Daley and Vere-Jones, 2003, Lemma 5.4.III).

To make the statistical analysis more practically feasible, we sometimes make the additional pragmatic assumption that Y_g may be treated as either of the point processes

$$\begin{aligned}
Y_S &= \{x : (x,t) \in Y_g\} \subseteq \mathbb{R}^d, \qquad (5.3)\\
Y_T &= \{t : (x,t) \in Y_g\} \subseteq \mathbb{R},
\end{aligned}$$

with marks in $\mathbb{R}$ and $\mathbb{R}^d$, respectively (c.f. Møller and Ghorbani (2012)). Note that e.g. in the former case this holds when we have $\mathbb{E}[Y_g(B \times \mathbb{R})] < \infty$ for any bounded $B \in \mathcal{B}(\mathbb{R}^d)$, which in turn holds e.g. when Y_g a.s. has no points outside $\mathbb{R}^d \times W_T$, for some bounded $W_T \in \mathcal{B}(\mathbb{R})$. The other case is analogous. Both are naturally permitted if Y_g (and thus Y) is a finite point process, i.e. if $N < \infty$ a.s.. Hence, from a practical point of view it is a very mild assumption.

Remark 2. *Such additional marking is facilitated by the imposed space-time metric* $d_\infty(\cdot,\cdot)$ *(Van Lieshout, 2000, p. 8).*

Since the ground process Y_g is well-defined by definition we may also define its nth factorial moment measure

$$\alpha_g^{(n)}(B_1 \times \cdots \times B_n) = \alpha^{(n)}((B_1 \times \mathcal{M}), \ldots, (B_n \times \mathcal{M})), \quad B_1, \ldots, B_n \in \mathcal{B}(\mathbb{R}^d \times \mathbb{R}),$$

assuming local finiteness. The next result, which is standard and a slight modification of e.g. (Heinrich, 2013, Section 4.1.2), shows that $\rho^{(n)}(\cdot)$ can be written as a product of the ground product density $\rho_g^{(n)}(\cdot)$ and a conditional density of the marks, given the spatio-temporal locations. Note that $\rho_g^{(n)}((x_1,t_1),\ldots,(x_n,t_n)) \prod_{i=1}^n \mathrm{d}x_i \mathrm{d}t_i$ gives us the probability of finding points of Y_g in infinitesimal neighbourhoods of $(x_i,t_i) \in \mathbb{R}^d \times \mathbb{R}$, $i = 1, \ldots, n$.

Lemma 3. *If* $\alpha_g^{(n)}(\cdot)$ *exists, then*

$$\begin{aligned}\rho^{(n)}((x_1,t_1,m_1),\ldots,(x_n,t_n,m_n)) =& f^{\mathcal{M}}_{(x_1,t_1),\ldots,(x_n,t_n)}(m_1,\ldots,m_n)\times \\ & \times \rho_g^{(n)}((x_1,t_1),\ldots,(x_n,t_n))\end{aligned} \tag{5.4}$$

almost everywhere (a.e.), where $\rho_g^{(n)}(\cdot)$ *is the* n*th product density of* Y_g *and* $f^{\mathcal{M}}_{(x_1,t_1),\ldots,(x_n,t_n)}(\cdot)$ *is the density of the conditional probability* $M^{(x_1,t_1),\ldots,(x_n,t_n)}(C)$, $C \in \mathcal{B}(\mathcal{M}^n)$*, of the marks of* n *points of* Y*, given that they have space-time locations* $(x_1,t_1),\ldots,(x_n,t_n) \in \mathbb{R}^d \times \mathbb{R}$*.*

When, in addition, Y_S *is well defined,*

$$\rho_g^{(n)}((x_1,t_1),\ldots,(x_n,t_n)) = f^T_{x_1,\ldots,x_n}(t_1,\ldots,t_n)\rho_S^{(n)}(x_1,\ldots,x_n),$$

and if Y_T *is well defined,*

$$\rho_g^{(n)}((x_1,t_1),\ldots,(x_n,t_n)) = f^S_{t_1,\ldots,t_n}(x_1,\ldots,x_n)\rho_T^{(n)}(t_1,\ldots,t_n),$$

where $\rho_S^{(n)}(\cdot)$ *and* $\rho_T^{(n)}(\cdot)$ *denote the respective* n*th product densities of* Y_S *and* Y_T*.*

Proof. This follows by disintegration (Daley and Vere-Jones, 2003): there exists a family of regular conditional probabilities $\{M^{(x_1,t_1),\ldots,(x_n,t_n)}(\cdot) : (x_1,t_1),\ldots,(x_n,t_n) \in \mathbb{R}^d \times \mathbb{R}\}$ on $\mathcal{B}(\mathcal{M})^n$, usually referred to as the *n-point mark distributions*, such that for $B \times C \in \mathcal{B}((\mathbb{R}^d \times \mathbb{R}) \times \mathcal{M})^n$ (Chiu et al., 2013, p. 118, p. 125)

$$\alpha^{(n)}(B \times C) = \int_B M^{(x_1,t_1),\ldots,(x_n,t_n)}(C)\alpha_g^{(n)}(d((x_1,t_1),\ldots,(x_n,t_n))).$$

The imposed absolute continuity of $\alpha^{(n)}$ with respect to $(\ell\otimes\nu)^n$ ensures that $M^{(x_1,t_1),\ldots,(x_n,t_n)}(\cdot)$ has Radon-Nikodym derivative $f^{\mathcal{M}}_{(x_1,t_1),\ldots,(x_n,t_n)}(\cdot)$ and $\alpha_g^{(n)}(\cdot)$ has Radon-Nikodym derivative $\rho_g^{(n)}(\cdot)$. $\square$

Turning to the explicit univariate properties of Y, setting $n = 1$ we obtain the *intensity measure* $\Lambda(B\times C) = \alpha^{(1)}(B\times C) = \mathbb{E}[Y(B\times C)] = \int_{B\times C}\lambda(x,t,m)\nu(dm)\mathrm{d}x\mathrm{d}t$, $B\times C \in \mathcal{B}((\mathbb{R}^d\times\mathbb{R})\times\mathcal{M})$, where, as indicated in (Vere-Jones, 2009), the *intensity function* is given by

$$\lambda(x,t,m) = \rho^{(1)}(x,t,m) = f^{\mathcal{M}}_{(x,t)}(m)\lambda_g(x,t).$$

Here $\lambda_g(x,t) = \rho_g^{(1)}(x,t)$ is the intensity function of the ground process. Also, when Y_S and Y_T are well defined, $\lambda_g(x,t) = f_x^T(t)\lambda_S(x)$ and $\lambda_g(x,t) = f_t^S(x)\lambda_T(t)$, respectively, where $\lambda_S(\cdot)$ and $\lambda_T(\cdot)$ are the respective intensity functions of Y_S and Y_T. Heuristically, in order to obtain $\lambda_g(x,t)$, we rescale $\mathbb{P}(Y_S\cap dx \neq \emptyset) = \lambda_S(x)\mathrm{d}x$ by the conditional infinitesimal probability that this event, with spatial location in $Y_S\cap dx$, occurs at time t.

At times one makes the assumption that the intensity is constant as a function of space, time or both. This is referred to as homogeneity.

Definition 5. *If $\lambda_g(x,t) = \lambda_T(t)$ only depends on $t\in\mathbb{R}$, we say that Y is* spatially homogeneous*, whereas if $\lambda_g(x,t) = \lambda_S(x)$ only depends on $x\in\mathbb{R}^d$, we say that Y is* temporally homogeneous*.*

We say that Y is (spatio-temporally) homogeneous *if its ground process is homogeneous, i.e. if $\lambda_g(x,t)\equiv\lambda>0$ and $\lambda(x,t,m) = f^{\mathcal{M}}_{(x,t)}(m)\lambda$, and we call it* inhomogeneous *otherwise.*

Some things should be noted here. Firstly, stationarity implies homogeneity. Secondly, the functions $\lambda_S(\cdot)$ and $\lambda_T(\cdot)$ are non-unique since e.g. $\lambda_g(x,t) = \lambda_T(t) = c\frac{\lambda_T(t)}{c} = c\widetilde{\lambda}_T(t)$ for any $c>0$. Also, statistically, homogeneity is a strongly simplifying assumption and it is seldom realistic nor advised to assume that the data under consideration is generated by a homogeneous process (unless one is very confident that the application in mind behaves accordingly).

5.2.3.1 Separability

We next consider the notion of separability (Møller and Ghorbani, 2012; Gabriel and Diggle, 2009), of which homogeneity is an example.

Definition 6. *If the ground intensity can be expressed as a (non-unique) product $\lambda_g(x,t) = \lambda_1(x)\lambda_2(t)$ of two non-negative measurable functions $\lambda_1(\cdot)$ and $\lambda_2(\cdot)$, we say that Y is* separable*.*

When $\lambda_S(\cdot)$ and $\lambda_T(\cdot)$ exist we may e.g. set $\lambda_1(x) = \lambda_S(x)$ and $\lambda_2(t) = f^T(t) = \lambda_T(t)/\int_{\mathbb{R}}\lambda_T(s)\mathrm{d}s$, or $\lambda_1(x) = f^S(x) = \lambda_S(x)/\int_{\mathbb{R}^d}\lambda_1(y)\mathrm{d}y$ and $\lambda_2(t) = \lambda_T(t)$ (note that $f^S(x)$ and $f^T(t)$ are probability densities).

It should be noted that separability mainly is a practical assumption, imposed to simplify the analysis, and it is not always justified. It is mainly suitable when Y_g has a repetitive behaviour in the sense that the intensity may be treated as a temporal/spatial rescaling of an overall temporal/spatial intensity, where the rescaling happens independently.

5.2.4 Pair correlation functions

Having defined the product densities, we may proceed to defining a further central summary statistic for point processes, the *pair correlation function (pcf)* (Chiu et al., 2013; Møller and Waagepetersen, 2004),

$$g((x_1,t_1,m_1),(x_2,t_2,m_2)) = \frac{\rho^{(2)}((x_1,t_1,m_1),(x_2,t_2,m_2))}{\lambda(x_1,t_1,m_1)\lambda(x_2,t_2,m_2)}.$$

By expression (5.4), the pcf satisfies

$$\begin{aligned} g((x_1,t_1,m_1),(x_2,t_2,m_2)) &= \frac{f^{\mathcal{M}}_{(x_1,t_1),(x_2,t_2)}(m_1,m_2)}{f^{\mathcal{M}}_{(x_1,t_1)}(m_1)f^{\mathcal{M}}_{(x_2,t_2)}(m_2)}\frac{\rho_g^{(2)}((x_1,t_1),(x_2,t_2))}{\lambda_g(x_1,t_1)\lambda_g(x_2,t_2)} \qquad (5.5) \\ &= \frac{f^{\mathcal{M}}_{(x_1,t_1),(x_2,t_2)}(m_1,m_2)}{f^{\mathcal{M}}_{(x_1,t_1)}(m_1)f^{\mathcal{M}}_{(x_2,t_2)}(m_2)} g_g((x_1,t_1),(x_2,t_2)), \end{aligned}$$

where $g_g(\cdot)$ is the pcf of the ground process Y_g. Due to expression (5.8) below, for a Poisson process on $(\mathbb{R}^d \times \mathbb{R}) \times \mathcal{M}$ the pcf satisfies $g(\cdot) = g_g(\cdot) \equiv 1$. Hence, for a MSTPP Y with intensity $\lambda(\cdot)$ and $g((x_1,t_1,m_1),(x_2,t_2,m_2)) > 1$, there is clustering between points of Y located around (x_1,t_1) and (x_2,t_2), with associated marks m_1 and m_2. Similarly, $g((x_1,t_1,m_1),(x_2,t_2,m_2)) < 1$ indicates inhibition. Non-parametric estimates of pcf:s are used extensively to analyse whether data exhibits interaction (Diggle, 2014).

5.2.5 Specific marking structures

Below follow some possible marking structures that may be imposed. We will consider these in more depth further on and, in particular, we will see how they influence summary statistics that we will derive. Hereby they also play a role in the statistical analysis.

5.2.5.1 Common mark distribution

Starting with the univariate properties, we next introduce the notion of a common (marginal) mark distribution.

Definition 7. *We say that a MSTPP Y has a* common (marginal) mark distribution *$M(C)$, $C \in \mathcal{B}(\mathcal{M})$, if all marks have the same marginal distributions; $M^{(x,t)}(\cdot) \equiv M(\cdot)$ for any $(x,t) \in \mathbb{R}^d \times \mathbb{R}$ and $f^{\mathcal{M}}_{(x,t)}(\cdot) \equiv f^{\mathcal{M}}(\cdot)$, $(x,t) \in \mathbb{R}^d \times \mathbb{R}$, for a* common mark density.

If, in addition, $M(\cdot)$ *and the reference measure* $\nu(\cdot)$ *coincide, so that* $f^{\mathcal{M}}_{(x,t)}(\cdot) \equiv 1$ *and* $\lambda(x,t,m) = \lambda_g(x,t)$*, we say that the reference measure is given by* the *mark distribution (Chiu et al., 2013, p. 119).*

It should be emphasised that Y having a common mark distribution means that all marks $m_1, \dots, m_N$ have the same marginal distribution $M(\cdot)$; they may, however, very well be mutually dependent. Note that Y being homogeneous with a common mark distribution results in $\lambda(x,t,m) = f^{\mathcal{M}}(m)\lambda$, so that $\lambda(x,t,m) = \lambda$ if the reference measure is given by the mark distribution.

5.2.5.2 Independent marks and random labelling

In order to provide a complete marking structure for Y we have to define all joint distributions of the marks m_i, $i = 1, \dots, N$ (conditionally on the ground process). This includes e.g. such elaborate structures as geostatistical marking (see e.g. (Illian et al., 2008)). However, one possible simplifying assumption is to let the marks be independent. Following e.g. (Daley and Vere-Jones, 2003, Def. 6.4.III), we consider the following two definitions:

1. Y has *independent marks* if, given the ground process Y_g, the marks are mutually independent random variables such that the distribution of a mark depends only on the spatio-temporal location of the corresponding event. Here we have $f^{\mathcal{M}}_{(x_1,t_1),\dots,(x_n,t_n)}(m_1, \dots, m_n) = \prod_{i=1}^{n} f^{\mathcal{M}}_{(x_i,t_i)}(m_i)$ for any $n \geq 1$.
2. If, in addition to independent marking, Y has a common mark distribution, i.e. if the marks are independent and identically distributed, then we say that Y has the *random labelling property.* Here $f^{\mathcal{M}}_{(x_1,t_1),\dots,(x_n,t_n)}(m_1, \dots, m_n) = \prod_{i=1}^{n} f^{\mathcal{M}}(m_i)$ for any $n \geq 1$, where we recall the common mark density $f^{\mathcal{M}}(\cdot)$.

5.2.6 Palm distributions

In order to consider conditioning on the event that Y has a point somewhere in $(\mathbb{R}^d \times \mathbb{R}) \times \mathcal{M}$ (this will needed for our summary statistics), we turn to Palm distributions (Daley and Vere-Jones, 2003; Van Lieshout, 2000; Chiu et al., 2013). The family of *reduced Palm distributions* of Y, $\{P^{!(x,t,m)}(\cdot) : (x,t,m) \in \mathbb{R}^d \times \mathbb{R} \times \mathcal{M}\}$, may formally be defined as the family of regular probabilities (Daley and Vere-Jones, 2003) satisfying the *reduced Campbell-Mecke formula* (see e.g. (Van Lieshout, 2000)): For any measurable function $f : (\mathbb{R}^d \times \mathbb{R} \times \mathcal{M}) \times N_{lf} \to [0, \infty)$,

$$\mathbb{E}\left[\sum_{(x,t,m)\in Y} f((x,t,m), Y \setminus \{(x,t,m)\})\right] = \tag{5.6}$$
$$= \int_{\mathbb{R}^d\times\mathbb{R}\times\mathcal{M}} \mathbb{E}^{!(x,t,m)}\left[f((x,t,m),Y)\right] \lambda(x,t,m)\nu(dm)\mathrm{d}x\mathrm{d}t.$$

Note that $\mathbb{E}^{!(x,t,m)}[\cdot]$ is the expectation corresponding to the probability measure $P^{!(x,t,m)}(\cdot) = \mathbb{P}^{!(x,t,m)}(Y \in \cdot)$. Concerning its interpretation, the MSTPP with distribution $P^{!(x,t,m)}(\cdot)$ on $(N_{lf}, \mathcal{N})$, the *reduced Palm process* at (x,t,m), may be interpreted as the conditional MSTPP $(Y|\{Y \cap \{(x,t,m)\} \neq \emptyset\}) \setminus \{(x,t,m)\}$. Under stationarity, $P^{!(x,t,m)}(\cdot)$ is constant as a function of (x,t,m), whereby one sets $P^{!(x,t,m)}(\cdot) \equiv P^{!(0,0,m)}(\cdot)$.

5.2.6.1 Reduced Palm distributions with respect to the mark sets

It will sometimes be convenient to consider conditioning with respect to a whole mark set $C \in \mathcal{B}(M)$, instead of just one specific mark value as in $P^{!(x,t,m)}(\cdot)$. To do so, following Cronie and van Lieshout (2016), we may define ν-averaged reduced Palm distributions.

Definition 8. *The* ν-averaged reduced Palm distribution (at $(x,t) \in \mathbb{R}^d \times \mathbb{R}$), with respect to $C \in \mathcal{B}(\mathcal{M})$*, is defined as*

$$P_C^{!(x,t)}(R) = \mathbb{P}_C^{!(x,t)}(Y \in R) = \frac{1}{\nu(C)} \int_C P^{!(x,t,m)}(R)\nu(dm), \qquad R \in \mathcal{N}. \tag{5.7}$$

Note that this is a probability measure since $0 \leq P^{!(x,t,m)}(\cdot) \leq 1$. *Expectation under* $P_C^{!(x,t)}(\cdot)$ *is given by* $\mathbb{E}_C^{!(x,t)}[\cdot] = \frac{1}{\nu(C)} \int_C \mathbb{E}^{!(x,t,m)}[\cdot]\nu(dm)$, *by Fubini's theorem.*

In the case that the reference measure is given by the mark distribution (recall Definition 7),

$$\mathbb{P}_C^{!(x,t)}(Y \in \cdot) = \frac{\int_C P^{!(x,t,m)}(\cdot)M(dm)}{M(C)}$$

may be interpreted as the conditional distribution

$$\mathbb{P}\left(Y \setminus (\{(x,t)\} \times \mathcal{M}) \in \cdot \,|\, Y \cap (\{(x,t)\} \times C) \neq \emptyset\right).$$

Under stationarity, where $\mathbb{P}_C^{!(0,0)}(Y \in \cdot) = \mathbb{P}_C^{!(x,t)}(Y + (x,t) \in \cdot)$ for almost any $(x,t) \in \mathbb{R}^d \times \mathbb{R}$, we refer to $P_C^{!(x,t)}(\cdot)$ as the *reduced Palm distribution with respect to the mark set* C (see (Van Lieshout, 2006) and (Chiu et al., 2013, p. 135)).

5.3 Examples of models

We next briefly recall and consider some properties of two particular models that we will consider in this chapter.

Poisson processes are the benchmark models for absence of (spatio-temporal) interaction (Chiu et al., 2013; Diggle, 2014; Van Lieshout, 2000). For a Poisson process Y on $(\mathbb{R}^d \times \mathbb{R}) \times \mathcal{M}$, due to the independence of its points, the product densities and the pcf satisfy

$$\begin{aligned} \rho^{(n)}((x_1,t_1,m_1),\ldots,(x_n,t_n,m_n)) &= \prod_{i=1}^n \lambda(x_i,t_i,m_i) = \prod_{i=1}^n f^{\mathcal{M}}_{(x_i,t_i)}(m_i)\lambda_g(x_i,t_i), \quad n \geq 1, \\ g((x_1,t_1,m_1),(x_2,t_2,m_2)) &= g_g((x_1,t_1),(x_2,t_2)) \equiv 1. \end{aligned} \tag{5.8}$$

Hence, it may be regarded as independently marked (see Section 5.2.5). We stress that this differs from a Poisson process Y_g on $\mathbb{R}^d \times \mathbb{R}$ to which we assign marks according to families $\{f^{\mathcal{M}}_{(x_1,t_1),\dots,(x_n,t_n)}(\cdot) : (x_1,t_1),\dots,(x_n,t_n) \in \mathbb{R}^d \times \mathbb{R}\}$, $n \geq 1$, of densities on $\mathcal{M}^n$; its pcf is given by $f^{\mathcal{M}}_{(x_1,t_1),(x_2,t_2)}(m_1,m_2)(f^{\mathcal{M}}_{(x_1,t_1)}(m_1) f^{\mathcal{M}}_{(x_2,t_2)}(m_2))^{-1}$. Indeed, the two concepts coincide when we have independent marking for the latter (see e.g. (Haenggi, 2012, Theorem 7.5)).

Example 1 (Poisson process)**.** *We consider a spatio-temporal (ground) Poisson process* $Y_g = \{(x_i, y_i, t_i)\}_{i=1}^N$ *on* $W_S \times W_T = [0,1]^2 \times [0,1]$*, with intensity function*

$$\lambda(x_1, y_1, t_1) = 5t_1 e^{5+0.5x_1}, \qquad (x_1, y_1, t_1) \in W_S \times W_T.$$

Conditionally on the number of points, N*, we further consider* N *independent Bernoulli distributed random variables* $m_1, \dots, m_N$*, with parameter* $p = 0.4$*, and assign these to* Y_g*, as marks. Hereby the mark space is* $\mathcal{M} = \{0,1\}$ *and* $Y = \{(x_1, y_1, t_1, m_1) : (x_1, y_1, t_1) \in Y_g\} \subseteq W_S \times W_T \times \mathcal{M}$ *is the resulting MSTPP. The reference measure considered is the counting measure (see the Appendix for details on multivariate STPPs).*

Figure 5.1 shows a realisation of such a process, together with spatial projections for two different time intervals, $[0, 0.5]$ *(middle) and* $(0.5, 1]$ *(right).*

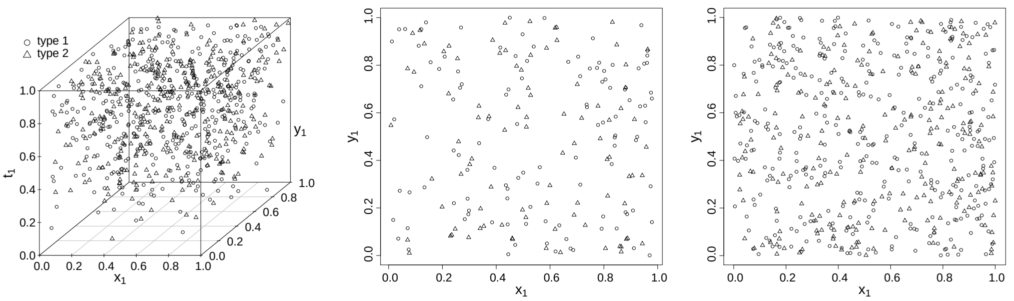

Figure 5.1. (Left) Spatio-temporal Poisson process with intensity function $\lambda(x_1, y_1, t_1) = 5t_1 e^{5+0.5x_1}$ on $[0,1]^2 \times [0,1]$ with independent Bernoulli distributed marks (parameter $p = 0.4$). (Centre) Spatial projections for the time interval $[0, 0.5]$. (Right) Spatial projections for the time interval $(0.5, 1]$. Here $\mathcal{M} = \{0,1\}$ and 'type 1' refers to a point having mark 0.

Recall that a *spatio-temporal log-Gaussian Cox process (LGCP)* Y_g (Møller et al., 1998; Cronie and van Lieshout, 2015; Diggle, 2014) is a spatio-temporal Poisson process for which the intensity function is given by the realisation of some (a.s. locally integrable non-negative) random field $X(x,t) = \mathrm{e}^{\mu(x,t)+Z(x,t)}$, where $Z(x,t)$ is a zero-mean Gaussian random field on $\mathbb{R}^d \times \mathbb{R}$. Such a random field Z is characterised by its expectation function $\mathbb{E}[Z(x,t)]$ and its covariance function $\mathrm{Cov}(Z(x_1,t_1), Z(x_2,t_2))$. The simplest class of space-time covariance models are separable models, which are given by

$$\mathrm{Cov}(Z(x_1,t_1), Z(x_2,t_2)) = \mathrm{Cov}((x_1,t_1),(x_2,t_2)) = \mathrm{Cov}_S(x_1,x_2)\,\mathrm{Cov}_T(t_1,t_2),$$

where Cov_S is a covariance function on $\mathbb{R}^d$ and Cov_T is a covariance function on $\mathbb{R}$. If, in addition, we assume stationarity in space and time, the covariance function depends only on the space-time lag between the points, whereby

$$\mathrm{Cov}(Z(x_1,t_1),Z(x_1+h,t_1+u)) = C(h,u) = C_S(h)C_T(u), \tag{5.9}$$

where $(h,u) \in \mathbb{R}^d \times \mathbb{R}$ is the space-time lag between the points (Gneiting, 2002). To simulate random fields, we employ the `R` package `RandomFields` (Schlather et al., 2015).

Example 2 (Independently marked LGCP)**.** *We consider a univariate spatio-temporal LGCP,* $Y_g = \{(x_i,y_i,t_i)\}_{i=1}^N$*, on the spatio-temporal domain* $W_S \times W_T = [0,1]^2 \times [0,1]$*, with mean function given by* $\mu(x_1,y_1,t_1) = \log(750) - 0.5(y_1+t_1) - \sigma^2/2$*, where* $\sigma^2 = (1/4)^2 = 1/16$*. We further consider a separable space-time covariance function for* Z*, where the spatial covariance function is given by the stationary and isotropic Whittle-Matérn covariance model:*

$$C_S(h) = \sigma^2 \frac{2^{1-\nu}}{\Gamma(\nu)}(c\|h\|)^\nu K_\nu(c\|h\|),$$

where $\nu > 0$ *is a smoothness parameter,* c *is a nonnegative scaling parameter and* K_ν *denotes the modified Bessel function of the second kind of order* ν*. The temporal covariance function is constant and given by* $C_T(u) = 1$*.*

As in Example 1, we consider N *independent random Bernoulli distributed marks, with parameter* $p = 0.4$*, and obtain the MSTPP* $Y = \{(x_1,y_1,t_1,m_1) : (x_1,y_1,t_1) \in Y_g\} \subseteq W_S \times W_T \times \mathcal{M}$*, where we again consider the counting measure as mark reference measure* $\nu(\cdot)$*.*

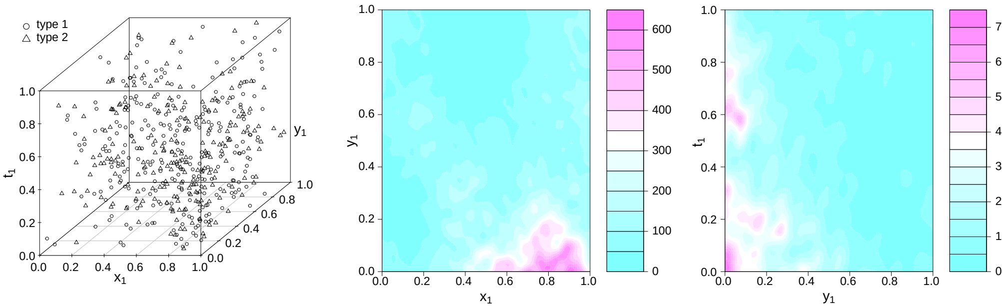

Figure 5.2. (Left) Randomly labelled spatio-temporal log-Gaussian Cox process with iid $Bernoulli(0.4)$-distributed marks. (Centre) Projection of the random intensity field of the log-Gaussian Cox process Y_g, at time $t_1 = 0.5$. (Right) Projection of the random intensity field of the log-Gaussian Cox process Y_g, at spatial coordinate $x_1 = 0.5$. Here $\mathcal{M} = \{0,1\}$ and 'type 1' refers to a point having mark 0.

Figure 5.2 shows a realisation of such an independently marked spatio-temporal log-Gaussian Cox process (left), together with a temporal projection ($t_1 = 0.5$*) and spatial projection (*$x_1 = 0.5$*) of the Gaussian random field (middle and right).*

Example 3 (Bivariate spatio-temporal process)**.** *We consider a spatio-temporal Poisson process, Y_1, with the same intensity function as in Example 1, on the spatio-temporal domain $W_S \times W_T = [0,1]^2 \times [0,1]$. In the same spatio-temporal observation window we consider a spatio-temporal log-Gaussian Cox process, Y_2, with mean function given by $\mu(x_1, y_1, t_1) = \log(750) - 1.5(y_1 + t_1) - \sigma^2/2$, where $\sigma^2 = (1/4)^2 = 1/16$. We consider the same spatio-temporal covariance function as in Example 2.*

We assign the numerical mark 1 to all points coming from Y_1 and the numerical mark 2 to the second component process, Y_2. Hence, the mark space is $\mathcal{M} = \{1, 2\}$ and the bivariate STPP Y is obtained by combining Y_1 and Y_2 into $Y = Y_1 \cup Y_2$. Note that this is a multivariate STPP and as usual it is natural to let $\nu(\cdot)$ be given by the counting measure.

Figure 5.3 shows a realisation of such a bivariate STPP. Figure 5.3 also shows projections of a realisation of the random intensity field of Y_1, at time $t_1 = 0.5$ (centre) and at spatial coordinate $x_1 = 0.5$ (right).

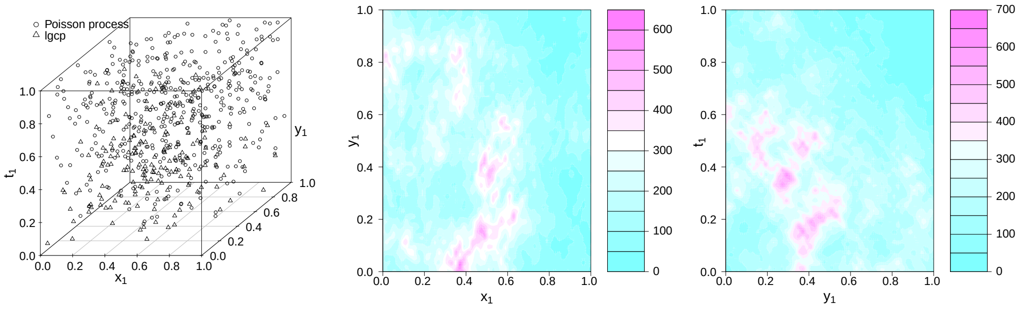

Figure 5.3. (Left) A realisation of a bivariate spatio-temporal process. (Centre) Projection of the random intensity field of Y_2, at time $t_1 = 0.5$. (Right) Projection of the random intensity field of Y_2, at spatial coordinate $x_1 = 0.5$.

Example 4 (Geostatistically marked LGCP)**.** *We consider a spatio-temporal log-Gaussian Cox process, Y_g, on the spatio-temporal domain $W_S \times W_T = [0,1]^2 \times [0,1]$, with underlying mean function $\mu(x_1, y_1, t_1) = \log(750) - 0.5(y_1 + t_1) + \sigma^2/2$, where $\sigma^2 = (1/4)^2 = 1/16$. As covariance function, we consider the separable spatio-temporal covariance function described in Example 2. We then simulate a spatio-temporal Gaussian random field, $\{R(x, y, t) : (x, y, t) \in [0,1]^2 \times [0,1]\}$, with covariance function given by the stationary isotropic exponential model, $C(h, u) = C_S(h)C_T(u) = exp(-h)$; here $h \geq 0$ is the spatial Euclidean distance between two points (a separable model). In order to assign marks to Y_g, we let $m_i = R(x_i, t_i)$ for all $(x_i, t_i) \in Y_g$, whereby $\mathcal{M} = \mathbb{R}$ and the mark reference measure $\nu(\cdot)$ is assumed to be the Lebesgue measure on $\mathbb{R}$.*

Figure 5.4 (left) shows a realisation of such a geostatistically marked spatio-temporal log-Gaussian Cox process, where the size of a circle around a point is proportional to the value of its continuous mark, together with the Gaussian random field of the marks (right).

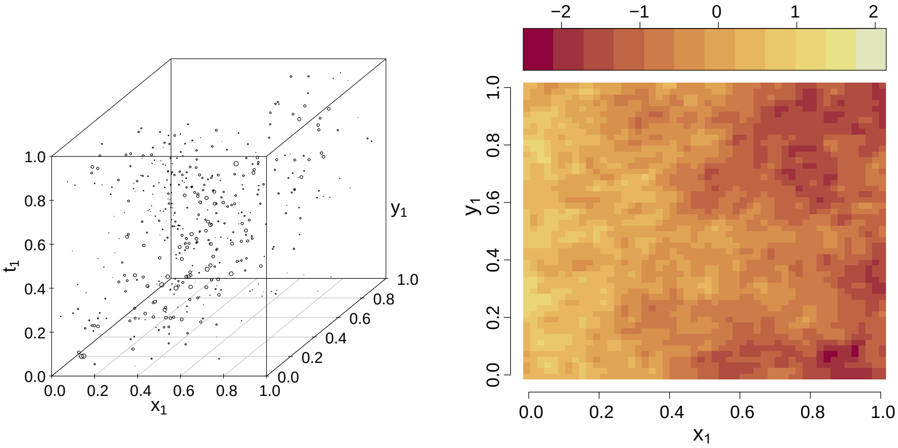

Figure 5.4. (Left) A realisation of the above defined geostatistically marked spatio-temporal log-Gaussian Cox process; the size of a point is proportional to the value of its continuous mark. (Right) The Gaussian random field generating the marks.

5.4 Marked inhomogeneous second-order measures of spatio-temporal interaction

As seen above, the intensity function of a MSTPP governs its univariate properties and the pcf governs second-order interactions. We now proceed by defining cumulative summary statistics/measures of spatio-temporal interaction for MSTPPs. The key idea is the extension of the marked inhomogeneous K-function of Cronie and van Lieshout (2016) to the spatio-temporal context.

5.4.1 Second order intensity-reweighted stationarity

A weaker form of stationarity that we impose when we consider the inhomogeneous MSTPPs below is *second-order intensity-reweighted stationarity (SOIRS)* (see e.g. (Baddeley et al., 2000)).

Definition 9. *We say that a MSTPP Y is* second-order intensity-reweighted stationary (SOIRS) *if the pcf exists and satisfies*

$$g((x_1+a,t_1+b,m_1),(x_2+a,t_2+b,m_2)) = g((x_1,t_1,m_1),(x_2,t_2,m_2))$$

a.e., for any $(a,b)\in\mathbb{R}^d\times\mathbb{R}$.

Avoiding the degenerate case where $\lambda(x,t,m)=\lambda_g(x,t)\equiv 0$ a.e., we must require that $\lambda_g(x,t)>0$ a.e. (Baddeley et al., 2000). Furthermore, a homogeneous SOIRS point process is second-order stationary. If in addition to SOIRS we have that

$$g((x_1,t_1,m_1),(x_2,t_2,m_2)) = \overline{g}(\|x_1-x_2\|_{\mathbb{R}^d},|t_1-t_2|,m_1,m_2), \tag{5.10}$$

i.e. the pcf is given by some function $\bar{g}(\cdot)$ that spatio-temporally depends only on the spatial distances and the temporal distances, we refer to Y as *SOIRS with isotropy (SOIRSI)* (Gabriel and Diggle, 2009; Diggle, 2014). C.f. the isotropy part of (Møller and Waagepetersen, 2004, p.34).

5.4.2 Marked spatio-temporal second-order reduced moment measures

As an alternative to the pcf as a marked measure of second-order spatio-temporal interaction, we may instead consider cumulative versions of it. Throughout we will assume that Y is SOIRS (see Definition 9).

We start by defining the *marked spatio-temporal second-order reduced moment measure* (c.f. (Møller and Waagepetersen, 2004, Definition 4.5)), which is our main building block. It describes how points of Y, with marks in some Borel set $C \subseteq \mathcal{M}$, interact with points of Y with marks in some Borel set $D \subseteq \mathcal{M}$, when their spatio-temporal separation vectors lie in some $E \subseteq \mathbb{R}^d \times \mathbb{R}$.

Definition 10. *Let $B \subseteq \mathbb{R}^d \times \mathbb{R}$, $\ell(B) > 0$, be arbitrary and let $C, D \subseteq \mathcal{M}$ be fixed Borel sets with $\nu(C), \nu(D) > 0$. The* marked spatio-temporal second-order reduced moment measure *of a SOIRS MSTPP Y is defined as*

$$\mathcal{K}^{CD}(E) = \frac{1}{\ell(B)\nu(C)\nu(D)} \times \tag{5.11}$$

$$\times\, \mathbb{E}\left[\sum_{(x_1,t_1,m_1),(x_2,t_2,m_2)\in Y}^{\neq} \frac{\mathbf{1}\{(x_1,t_1,m_1) \in B \times C\}\mathbf{1}\{(x_2,t_2) \in (x_1,t_1) \oplus E\}\mathbf{1}\{m_2 \in D\}}{\lambda(x_1,t_1,m_1)\lambda(x_2,t_2,m_2)} \right],$$

for $E \in \mathcal{B}(\mathbb{R}^d \times \mathbb{R})$ (through measure extension of locally finite measures on the ring of bounded Borel sets (see e.g. Halmos (1974))).

The marked spatio-temporal second-order reduced moment measure may also be expressed using the pair correlation function.

Lemma 4. *Given a SOIRS MSTPP Y and Borel mark sets $C, D \subseteq \mathcal{M}$, $\nu(C), \nu(D) > 0$, the measure $\mathcal{K}^{CD}(\cdot)$ defined in* (5.11) *satisfies*

$$\begin{aligned}\mathcal{K}^{CD}(E) &= \frac{1}{\ell(B)\nu(C)\nu(D)} \int_B \int_C \int_{(x_1,t_1)+E} \int_D g((x_1,t_1,m_1),(x_2,t_2,m_2))\mathrm{d}x_2\mathrm{d}t_2\mathrm{d}x_1\mathrm{d}t_1 \\ &= \frac{1}{\nu(C)\nu(D)} \int_C \int_D \left[\int\!\!\int_E g((0,0,m_1),(x,s,m_2))\mathrm{d}x\mathrm{d}s \right] \nu(dm_2)\nu(dm_1)\end{aligned} \tag{5.12}$$

for any $E \in \mathcal{B}(\mathbb{R}^d \times \mathbb{R})$. Hence, expression (5.11) *does not depend on the choice of B.*

Proof. Define the non-negative measurable function $f : ((\mathbb{R}^d \times \mathbb{R}) \times \mathcal{M})^2 \to \mathbb{R}$,

$$f((x_1,t_1,m_1),(x_2,t_2,m_2)) = \frac{\mathbf{1}\{(x_1,t_1,m_1) \in B \times C\}\mathbf{1}\{(x_2,t_2,m_2) \in E \times D\}}{\lambda(x_1,t_1,m_1)\lambda(x_2,t_2,m_2)}$$

where $B, E \in \mathcal{B}(\mathbb{R}^d \times \mathbb{R})$ and $C, D \in \mathcal{B}(\mathcal{M})$. By the Campbell formula, Fubini's theorem and the translation invariance obtained under SOIRS,

$$
\begin{aligned}
\nu(C)\nu(D)\mathcal{K}^{CD}(E) &= \mathbb{E}\left[\sum_{(x_1,t_1,m_1),(x_2,t_2,m_2)\in Y}^{\neq} f((x_1,t_1,m_1),(x_2,t_2,m_2))\right] \\
&= \frac{1}{\ell(B)\nu(C)\nu(D)} \int_B \int_C \int_E \int_D \frac{\rho^{(2)}((x_1,t_1,m_1),(x_2,t_2,m_2))}{\lambda(x_1,t_1,m_1)\lambda(x_2,t_2,m_2)} dx_1 dt_1 \nu(dm_1) dx_2 dt_2 \nu(dm_2) \\
&= \frac{1}{\ell(B)} \int_B \int_C \int_{(x_1,t_1)+E} \int_D g((x_1,t_1,m_1),(x_2,t_2,m_2))\nu(dm_2)\mathrm{d}x_2\mathrm{d}t_2\nu(dm_1)\mathrm{d}x_1\mathrm{d}t_1 \\
&= \frac{1}{\ell(B)} \int_B \int_C \int_D \int_{(x_1,t_1)+E} g((0,0,m_1),(x_2-x_1,t_2-t_1,m_2))\mathrm{d}x_2\mathrm{d}t_2\nu(dm_2)\nu(dm_1)\mathrm{d}x_1\mathrm{d}t_1 \\
&= \frac{1}{\ell(B)}\ell(B) \int_C \int_D \int_E g((0,0,m_1),(z,s,m_2))\mathrm{d}z\mathrm{d}s\nu(dm_2)\nu(dm_1) \\
&= \int_C \int_D \left[\int_E g((0,0,m_1),(u,v,m_2))dudv\right] \nu(dm_1)\nu(dm_2).
\end{aligned}
$$

This proves Equation (5.12) and that $\mathcal{K}^{CD}(E)$ does not depend on the choice of B. □

Since $\rho^{(2)}((x_1,t_1,m_1),(x_2,t_2,m_2)) = \lambda(x_1,t_1,m_1)\lambda(x_2,t_2,m_2)$ for a Poisson process on $(\mathbb{R}^d \times \mathbb{R}) \times \mathcal{M}$, by the Campbell formula we have that

$$\mathcal{K}^{CD}(E) = \ell(E).$$

Writing

$$Y_C = \{(x,t) : (x,t,m) \in Y, m \in C\} \subseteq Y_g \tag{5.13}$$

for the collection of points of Y_g that have marks belonging to $C \in \mathcal{B}(\mathcal{M})$ (i.e. the projection of Y on $\mathbb{R}^d \times \mathbb{R}$), note that we do not necessarily have that $Y_C \cap Y_D = \emptyset$, since we have allowed that $C \cap D \neq \emptyset$. However, it may be highly unnatural to consider C and D such that $C \cap D \neq \emptyset$.

5.4.2.1 Representation in terms of Palm distributions

Recalling $\mathbb{E}_C^{!(x,t)}[\cdot]$ from Definition 8, we may obtain a further representation and interpretation of $\mathcal{K}^{CD}(E)$. By the reduced Campbell-Mecke formula and (5.7),

$$\begin{aligned}\mathcal{K}^{CD}(E) &= \frac{1}{\ell(B)\nu(C)\nu(D)}\times \qquad (5.14)\\ &\times \mathbb{E}\left[\sum\nolimits^{\neq}_{(x_1,t_1,m_1),(x_2,t_2,m_2)\in Y} \frac{\mathbf{1}\{(x_1,t_1,m_1)\in B\times C\}\mathbf{1}\{(x_2,t_2)\in(x_1,t_1)+E\}\mathbf{1}\{m_2\in D\}}{\lambda(x_1,t_1,m_1)\lambda(x_2,t_2,m_2)}\right]\\ &= \frac{1}{\ell(B)\nu(C)\nu(D)}\int_{B\times C}\mathbb{E}^{!(x_1,t_1,m_1)}\left[\sum_{(x_2,t_2,m_2)\in Y\cap((x_1,t_1)+E)\times D}\frac{1}{\lambda(x_2,t_2,m_2)}\right]\mathrm{d}x_1\mathrm{d}t_1\nu(dm_1)\\ &= \frac{1}{\ell(B)\nu(D)}\int_B \mathbb{E}_C^{!(x_1,t_1)}\left[\sum_{(x_2,t_2,m_2)\in Y\cap((x_1,t_1)+E)\times D}\frac{1}{\lambda(x_2,t_2,m_2)}\right]\mathrm{d}x_1\mathrm{d}t_1.\end{aligned}$$

In other words, $\mathcal{K}^{CD}(E)$ may be obtained either through averaging over the mark space, as in (5.12), or through averaging over the spatio-temporal domain, as in (5.14).

Let Y be stationary with ground intensity $\lambda_g(x,t)\equiv\lambda>0$ and mark density $f^{\mathcal{M}}_{(x,t)}(m)=f^{\mathcal{M}}(m)$, with respect to the mark reference measure $\nu(\cdot)$. From (5.14), we have that (see e.g. (Møller and Waagepetersen, 2004, Theorem C.1))

$$\begin{aligned}\mathcal{K}^{CD}(E) &= \qquad (5.15)\\ &= \frac{1}{\ell(B)\nu(C)\nu(D)}\int_{B\times C}\mathbb{E}^{!(0,0,m_1)}\left[\sum_{(x_2,t_2,m_2)\in Y\cap E\times D}\frac{1}{f^{\mathcal{M}}(m_2)\lambda}\right]\mathrm{d}x_1\mathrm{d}t_1\nu(dm_1)\\ &= \frac{1}{\lambda\nu(C)\nu(D)}\int_C \mathbb{E}^{!(0,0,m_1)}\left[\sum_{(x_2,t_2,m_2)\in Y\cap E\times D}\frac{1}{f^{\mathcal{M}}(m_2)}\right]\nu(dm_1)\\ &= \frac{\mathbb{E}_C^{!(0,0)}\left[\sum_{(x,t,m)\in Y\cap E\times D} f^{\mathcal{M}}(m)^{-1}\right]}{\lambda\nu(D)},\end{aligned}$$

whereby $\mathcal{K}^{CD}(E)=\frac{1}{\lambda\nu(D)}\mathbb{E}_C^{!(0,0)}\left[Y(E\times D)\right]$ if $\nu(\cdot)$ and $M(\cdot)$ are equal (or, equivalently, $f^{\mathcal{M}}(\cdot)\equiv 1$).

5.4.2.2 Changing the order of the mark sets

It may be noted that $\mathcal{K}^{CD}(\cdot)$ is not necessarily symmetric in C and D, i.e. it is not certain that $\mathcal{K}^{CD}(\cdot)=\mathcal{K}^{DC}(\cdot)$ in general. The next result provides some conditions under which this is satisfied. The main function of the result is to indicate that estimators of marked spatial dependence between points with marks in C and D, which are based on Definition 10, may look a bit different depending on the order chosen for C and D. In addition, it may be used to test hypotheses for the marking of Y (see Section 5.5.3.1).

Theorem 1. *Let* Y *be a SOIRS MSTPP and consider any Borel mark sets* $C, D \subseteq \mathcal{M}$, $C \neq D$, *with* $\nu(C), \nu(D) > 0$. *Either of*

1. $f^{\mathcal{M}}_{(x_1,t_1),(x_2,t_2)}(m_1, m_2) = f^{\mathcal{M}}_{(x_1,t_1)}(m_1) f^{\mathcal{M}}_{(x_2,t_2)}(m_2)$, *which includes* Y *being independently marked (and thus randomly labelled),*
2. Y *has a common mark distribution* $M(\cdot)$ *and, conditional on the associated locations in* $\mathbb{R}^d \times \mathbb{R}$, *any two marks* m_i, m_j, $i \neq j$, *are exchangeable random variables (this includes them being pairwise independent),*

implies that the measures $\mathcal{K}^{CD}(\cdot)$ *and* $\mathcal{K}^{DC}(\cdot)$ *coincide.*

Note that the conditional exchangeability in Theorem 1 refers to that, for almost every $(x_1, t_1) \neq (x_2, t_2)$,

$$\begin{aligned} M^{(x_1,t_1),(x_2,t_2)}(C_1 \times C_2) &= \int_{C_1 \times C_2} f^{\mathcal{M}}_{(x_1,t_1),(x_2,t_2)}(m_1, m_2)\nu(dm_1)\nu(dm_2) \\ &= \int_{C_1 \times C_2} f^{\mathcal{M}}_{(x_1,t_1),(x_2,t_2)}(m_2, m_1)\nu(dm_1)\nu(dm_2) \\ &= M^{(x_1,t_1),(x_2,t_2)}(C_2 \times C_1), \quad C_1, C_2 \in \mathcal{B}(\mathcal{M}). \end{aligned}$$

Proof. Through (5.12) we see that $\mathcal{K}^{CD}(\cdot) = \mathcal{K}^{DC}(\cdot)$ requires that

$$\frac{f^{\mathcal{M}}_{(x_1,t_1),(x_2,t_2)}(m_1, m_2)}{f^{\mathcal{M}}_{(x_1,t_1)}(m_1) f^{\mathcal{M}}_{(x_2,t_2)}(m_2)} \overset{a.e.}{=} \frac{f^{\mathcal{M}}_{(x_1,t_1),(x_2,t_2)}(m_2, m_1)}{f^{\mathcal{M}}_{(x_1,t_1)}(m_2) f^{\mathcal{M}}_{(x_2,t_2)}(m_1)}.$$

If Y is independently marked this is clearly satisfied. Turning to the second option, the common mark distribution translates the above statement into

$$f^{\mathcal{M}}_{(x_1,t_1),(x_2,t_2)}(m_1, m_2) \overset{a.e.}{=} f^{\mathcal{M}}_{(x_1,t_1),(x_2,t_2)}(m_2, m_1),$$

which holds by the exchangeability.

□

By de Finetti's theorem, this is equivalent to saying that, conditionally on the ground locations, pairwisely the marks can be expressed as mixtures of iid random variables.

Remark 3. *As an alternative one could proceed by considering a symmetrised version* $\overline{\mathcal{K}}^{DC}(E) = (\mathcal{K}^{CD}(E) + \mathcal{K}^{DC}(E))/2$, $E \in \mathcal{B}(\mathbb{R}^d \times \mathbb{R})$.

5.4.3 Marked spatio-temporal inhomogeneous K-functions

We have defined a marked inhomogeneous spatio-temporal measure, $\mathcal{K}$, to quantify second-order interactions. By specifying the set E we may obtain different measures of spatio-temporal interaction between points with different mark classifications C and D. In what follows we will look closer at such choices.

Assume that Y is SOIRS and consider two mark sets $C, D \in \mathcal{B}(\mathcal{M})$, with $\nu(C), \nu(D) > 0$. A first natural candidate for E would be the closed, origin centred ball $B[(0,0), r] = \{(x,s) : d_\infty((0,0),(x,s)) \leq r\} = \{(x,s) : \|x\| \leq r, |s| \leq r\}$ of radius $r \geq 0$, where $\|\cdot\|$ is an abbreviation of $\|\cdot\|_{\mathbb{R}^d}$. Hereby we would obtain a direct extension of the marked inhomogeneous K-function of Cronie and van Lieshout (2016) to the spatio-temporal setting:

$$\begin{aligned}
K^{CD}_{\text{inhom}}(r) = \mathcal{K}^{CD}(B[(0,0),r]) = \frac{1}{\ell(B)\nu(C)\nu(D)} \times \\
\times \mathbb{E}\left[\sum_{(x_1,t_1,m_1)\in Y\cap B\times C}\ \sum_{(x_2,t_2,m_2)\in Y\setminus\{(x_1,t_1,m_1)\}} \frac{\mathbf{1}\{(x_2,t_2,m_2)\in B[(x_1,t_1),r]\times D\}}{\lambda(x_1,t_1,m_1)\lambda(x_2,t_2,m_2)}\right] \\
= \frac{1}{\nu(C)\nu(D)} \int_C \int_D \int_{B[(0,0),r]} g((0,0,m_1),(x,s,m_2))\ell(d(x,s))\nu(dm_2)\nu(dm_1).
\end{aligned}$$

However, since the spatial scale is different than the temporal scale, it is more natural to treat space and time lags separately. Hence, we instead choose $(x_1, t_1) + E$ to be the closed cylinder $\mathcal{C}^t_r(x_1, t_1)$, with centre $(x_1, t_1) \in \mathbb{R}^d \times \mathbb{R}$, radius $r \geq 0$ and height $t \geq 0$, i.e.

$$\mathcal{C}^t_r(x_1,t_1) = (x_1,t_1) + \mathcal{C}^t_r(0,0) = \{(x_2,t_2) \in \mathbb{R}^d \times \mathbb{R} : \|x_1 - x_2\| \leq r, |t_1 - t_2| \leq t\}.$$

Note that when $d = 2$, $\mathcal{C}^t_r(0,0)$ is obtained by taking a disk (2-dimensional Euclidean ball) with radius r and stretching it in the t-dimension until it becomes the cylinder of height $2t$. Furthermore, $B[(0,0), r] = \mathcal{C}^r_r(0,0)$, whereby $K^{CD}_{\text{inhom}}(r) = \mathcal{K}^{CD}(\mathcal{C}^r_r(0,0))$.

Definition 11. *For any SOIRS MSTPP Y and mark sets $C, D \in \mathcal{B}(\mathcal{M})$, $\nu(C), \nu(D) > 0$, the* marked inhomogeneous spatio-temporal K-function *is defined as*

$$\begin{aligned}
K^{CD}_{\text{inhom}}(r,t) &= \mathcal{K}^{CD}(\mathcal{C}^t_r(0,0)) \qquad (5.16)\\
&= \frac{1}{\ell(B)\nu(C)\nu(D)} \times \\
&\quad \times\mathbb{E}\left[\sum_{(x_1,t_1,m_1)\in Y\cap B\times C}\ \sum_{(x_2,t_2,m_2)\in Y\setminus\{(x_1,t_1,m_1)\}} \frac{\mathbf{1}\{(x_2,t_2,m_2)\in \mathcal{C}^t_r(x_1,t_1)\times D\}}{\lambda(x_1,t_1,m_1)\lambda(x_2,t_2,m_2)}\right] \\
&= \frac{1}{\nu(C)\nu(D)} \int_C \int_D \int_{\|x\|\leq r} \int_{-t}^{t} g((0,0,m_1),(x,s,m_2))\mathrm{d}x\mathrm{d}s\nu(dm_2)\nu(dm_1)
\end{aligned}$$

for $r, t \geq 0$ and any $B \in \mathbb{R}^d \times \mathbb{R}$, $\ell(B) > 0$, by expression (5.12)*. Note that $K^{CD}_{\text{inhom}}(r,r) = K^{CD}_{\text{inhom}}(r)$.*

To connect Definition 11 with $K^{Y_g}_{\text{inhom}}(r,t) = K_{\text{inhom}}(r,t)$, i.e. the inhomogeneous spatio-temporal (ground) K-function in (Gabriel and Diggle, 2009; Møller and Ghorbani, 2012), we note that $K^{\mathcal{M}\mathcal{M}}_{\text{inhom}}(r,t)$ reduces to $K^{Y_g}_{\text{inhom}}(r,t)$ if the reference measure is given by the mark distribution (recall Definition 7). Furthermore, when Y is SOIRSI (recall (5.10)) with $g((x_1,t_1,m_1),(x_2,t_2,m_2)) = \bar{g}(\|x_1-x_2\|,|t_1-t_2|,m_1,m_2)$, by a transformation (to hyper-spherical coordinates),

$$\begin{aligned}K^{CD}_{\text{inhom}}(r,t) &= \frac{1}{\nu(C)\nu(D)}\int_C\int_D\int_{\|x\|\le r}\int_{-t}^{t}\bar{g}(\|x\|,|s|,m_1,m_2)\mathrm{d}x\mathrm{d}s\nu(dm_2)\nu(dm_1)\\ &= \frac{1}{\nu(C)\nu(D)}\int_C\int_D\int_{-t}^{t}\int_0^r \omega_d\bar{g}(u,v,m_1,m_2)u^{d-1}\mathrm{d}u\mathrm{d}v\nu(dm_2)\nu(dm_1)\end{aligned}$$

and we note the resemblance with $K_{\text{inhom}}(r,t)$.

To give the motivation behind $K^{CD}_{\text{inhom}}(r,t)$, recall that for a Poisson process on $(\mathbb{R}^2\times\mathbb{R})\times\mathcal{M}$ we have that

$$K^{CD}_{\text{inhom}}(r,t) = \ell(\mathcal{C}^t_r(0,0)) = 2tr^d\omega_d = 2tr^d\pi^{d/2}/(\Gamma(d/2+1)),$$

where $\omega_d = \pi^{d/2}/(\Gamma(d/2+1))$ is the (Lebesgue) volume of the d-dimensional Euclidean unit ball and $\Gamma(\cdot)$ is the Gamma function. The Lebesgue measure for the ball of radius r centered in the origin is $\ell(B[0,r]) = \omega_d r^d$, where $\omega_d = (\pi^{d/2})/(\Gamma(\frac{d}{2}+1))$ is the volume of the unit ball in $\mathbb{R}^d$. Thus, the volume of the cylinder $\mathcal{C}^t_r(0,0) = \{(y,s)\in\mathbb{R}^d\times\mathbb{R} : ||y||\le r \text{ and } |s|\le t\}$ is $2\omega_d r^d t$. Then is natural to think that

- If $K^{CD}_{\text{inhom}}(r,t) > 2\omega_d r^d t$ we have an indication that points of Y with marks in D have a tendency to cluster around the points with marks in C (in a pairwise sense), having compensated for the inhomogeneity. This is referred to as *clustering/aggregation.*

- When $K^{CD}_{\text{inhom}}(r,t) < 2\omega_d r^d t$, points with marks in D tend to avoid being close to the points with marks in C (in a pairwise sense), taking the inhomogeneity into account. This is called *regularity/inhibition.*

Here closeness is understood in terms of one of the points being inside the cylinder neighbourhood $\mathcal{C}^t_r$ of the other. In other words, we have defined a way of measuring spatio-temporal interaction between points belonging to two mark sets C and D, in terms of spatial lags $r\ge 0$ and temporal lags $t\ge 0$, in the presence of inhomogeneity.

5.4.4 Marked stationary spatio-temporal K-functions

Equation (5.15) leads us to the definition of the marked stationary K-function.

Definition 12. *Given a stationary MSTPP Y with intensity $\lambda > 0$, under the assumption that $\nu(\cdot) = M(\cdot)$, its* marked stationary spatio-temporal K-function *is given by*

$$\begin{aligned}K^{CD}(r,t) &= \frac{1}{\lambda\nu(C)\nu(D)}\int_C \mathbb{E}^{!(0,0,m)}\left[Y(\mathcal{C}^t_r(0,0)\times D)\right]\nu(dm)\\ &= \frac{\mathbb{E}^{!(0,0)}_C\left[Y(\mathcal{C}^t_r(0,0)\times D)\right]}{\lambda\nu(D)},\end{aligned}$$

for any $C, D \in \mathcal{B}(\mathcal{M})$ *with* $\nu(C), \nu(D) > 0$*. This is a spatio-temporal version of the form proposed by Van Lieshout (2006).*

In the multivariate case, where $\lambda_i(x,t) \equiv \lambda_i > 0$*,* $i \in \mathcal{M}$*, and* $C = \{i\}$ *and* $D = \{j\}$*, we obtain*

$$K^{ij}(r,t) = \frac{1}{\ell(B)\lambda_j}\mathbb{E}^{!(0,0,i)}\left[Y_j(\mathcal{C}_r^t(0,0)\right]\int_B dx_1dt_1 = \frac{\mathbb{E}^{!(0,0,i)}\left[Y_j(\mathcal{C}_r^t(0,0))\right]}{\lambda_j} = \frac{\mathbb{E}_i^{!(0,0)}\left[Y_j(\mathcal{C}_r^t(0,0))\right]}{\lambda_j},$$

a spatio-temporal version of the classical multivariate stationary K*-function (Diggle, 2014, p. 60). In particular,* $i = j$ *results in the* K*-function of Diggle et al. (1995) for* Y_i*.*

In other words, given that there is a typical point of Y, located at the origin, with mark belonging to C, $K^{CD}(r,t)$ asks what the expected number of further points is, which are located within the cylinder $\mathcal{C}_r^t(0,0)$ and have marks belonging to D. In the multivariate case, assuming that $i \neq j$, $\lambda_j K^{ij}_{\text{inhom}}(r,t) = \mathbb{E}^{!(0,0,i)}\left[Y_j(\mathcal{C}_r^t(0,0)\right]$ gives us the expected number of points of Y_j that fall within spatial distance r and temporal distance time t of a typical point of Y_i. Note that by the Slivniyak-Mecke theorem (Chiu et al., 2013), $\mathbb{E}_C^{!(0,0)}[Y(\mathcal{C}_r^t(0,0) \times D)] = \mathbb{E}[Y(\mathcal{C}_r^t(0,0) \times D)] = \lambda\nu(D)\ell(\mathcal{C}_r^t(0,0))$ for a Poisson process Y, as has already been established in the more general SOIRS case. Hence, $K^{CD}(r,t) - 2tr^d\omega_d > 0$ indicates clustering between points with marks in C and D and $K^{CD}(r,t) - 2tr^d\omega_d < 0$ indicates regularity.

5.4.5 Multivariate spatio-temporal K-functions

As stated in Section 5.2, when the mark space is given by a finite set, say $\mathcal{M} = \{1, \dots, k\}$, $k > 1$, a marked point process Y is referred to as a multivariate STPP with components Y_i, $i = 1, \dots, k$, where Y_i contains all the points of Y_g with i as associated mark.

Following Cronie and van Lieshout (2016), we further find that for a multivariate STPP, the ν-averaged reduced Palm distribution at $(x,t) \in \mathbb{R}^d \times \mathbb{R}$, with respect to $C = \{i\}$, is given by

$$P_i^{!(x,t)}(R) = P_{\{i\}}^{!(x,t)}(R) = \frac{P^{!(x,t,i)}(R)\nu(\{i\})}{\nu(\{i\})} = P^{!(x,t,i)}(R), \quad i \in \mathcal{M} = \{1, \dots, k\}, R \in \mathcal{N},$$

and is thus independent of the specific choice of $\nu(\cdot)$. By expression (5.14) it now follows that

$$\begin{aligned}
\mathcal{K}^{CD}(E) &= \frac{1}{\ell(B)\nu(D)}\int_B \sum_{i\in C}\mathbb{E}_i^{!(x_1,t_1)}\left[\sum_{j\in D}\sum_{(x_2,t_2)\in Y_j\cap((x_1,t_1)+E)}\frac{\nu(j)}{\lambda_j(x_2,t_2)}\right]\mathrm{d}x_1\mathrm{d}t_1 \\
&= \sum_{i\in C}\sum_{j\in D}\frac{\nu(j)}{\nu(D)}\frac{1}{\ell(B)}\int_B \mathbb{E}_i^{!(x_1,t_1)}\left[\sum_{(x_2,t_2)\in Y_j\cap((x_1,t_1)+E)}\frac{1}{\lambda_j(x_2,t_2)}\right]\mathrm{d}x_1\mathrm{d}t_1 \\
\mathcal{K}^{\{i\}\{j\}}(E) &= \frac{1}{\ell(B)\nu(j)}\int_B \mathbb{E}_i^{!(x_1,t_1)}\left[\sum_{(x_2,t_2)\in Y_j\cap((x_1,t_1)+E)}\frac{\nu(j)}{\lambda_j(x_2,t_2)}\right]\mathrm{d}x_1\mathrm{d}t_1,
\end{aligned}$$

for any $C, D \subseteq \mathcal{M}$, since $\lambda(x,t,i) = \lambda_i(x,t)/\nu(i)$, $(x,t,i) \in \mathbb{R}^d \times \mathbb{R} \times \mathcal{M}$ for a multivariate MSTPP. Here $\mathbb{E}_i^{!(x_1,t_1)}[\cdot] = \mathbb{E}^{!(x_1,t_1,i)}[\cdot]$ is the expectation under the reduced Palm distribution of Y_i. Note that we, in essence, scale each j-contribution by the probability $\nu(j)/\nu(D)$.

We may now define the *i-to-j inhomogeneous spatio-temporal cross K-function*:

$$\begin{aligned} K_{\mathrm{inhom}}^{ij}(r,t) &= \mathcal{K}^{\{i\}\{j\}}(\mathcal{C}_r^t(x_1,t_1)) \\ &= \frac{1}{\ell(B)\nu(j)} \int_B \mathbb{E}_i^{!(x_1,t_1)} \left[\sum_{(x_2,t_2)\in Y_j \cap \mathcal{C}_r^t(x_1,t_1)} \frac{\nu(j)}{\lambda_j(x_2,t_2)} \right] \mathrm{d}x_1 \mathrm{d}t_1 \qquad (5.17) \\ &= \frac{1}{\ell(B)} \mathbb{E} \left[\sum_{(x_1,t_1)\in Y_i \cap B} \sum_{(x_2,t_2)\in Y_j \cap \mathcal{C}_r^t(x_1,t_1) \setminus \{(x_1,t_1)\}} \frac{1}{\lambda_i(x_1,t_1)\lambda_j(x_2,t_2)} \right] \\ &= \frac{1}{\ell(B)} \int_B \mathbb{E}_i^{!(x_1,t_1)} \left[\sum_{(x_2,t_2)\in Y_j \cap \mathcal{C}_r^t(x_1,t_1)} \frac{1}{\lambda_j(x_2,t_2)} \right] \mathrm{d}x_1 \mathrm{d}t_1, \qquad (5.18) \end{aligned}$$

c.f. (Møller and Waagepetersen, 2004, Definition 4.8). Note that when $i = j$, $K_{\mathrm{inhom}}^{ij}(r,t)$ reduces to the inhomogeneous spatio-temporal K-function of Y_i, i.e. $K_{\mathrm{inhom}}^{i}(r,t)$. Also, the *$i$-to-any inhomogeneous spatio-temporal cross K-function* is given by

$$K_{\mathrm{inhom}}^{i\bullet}(r,t) = K_{\mathrm{inhom}}^{\{i\}\mathcal{M}}(r,t) = \sum_{j\in\mathcal{M}} \frac{\nu(j)}{\nu(\mathcal{M})} \frac{1}{\ell(B)} \int_B \mathbb{E}_i^{!(x_1,t_1)} \left[\sum_{(x_2,t_2)\in Y_j \cap \mathcal{C}_r^t(x_1,t_1)} \frac{\nu(j)}{\lambda_j(x_2,t_2)} \right] \mathrm{d}x_1 \mathrm{d}t_1,$$

where each $\nu(j) = 1$, $j \in \mathcal{M}$, if $\nu(\cdot)$ is the counting measure on $\mathcal{M}$.

5.4.5.1 Multivariate stationary spatio-temporal K-functions

Considering the stationary case, where $\lambda_i(x,t) \equiv \lambda_i > 0$, $i \in \mathcal{M}$, we have that $K_{\mathrm{inhom}}^{ij}(r,t)$ reduces to

$$\begin{aligned} K^{ij}(r,t) &= \frac{1}{\ell(B)\lambda_j} \mathbb{E}^{!(0,0,i)} \left[Y_j(\mathcal{C}_r^t(0,0)\right] \int_B dx_1 dt_1 \\ &= \frac{\mathbb{E}^{!(0,0,i)} \left[Y_j(\mathcal{C}_r^t(0,0)\right]}{\lambda_j} = \frac{\mathbb{E}_i^{!(0,0)} \left[Y_j(\mathcal{C}_r^t(0,0)\right]}{\lambda_j}, \qquad (5.19) \end{aligned}$$

a spatio-temporal version of the classical multivariate stationary K-function (Diggle, 2014, p. 60). Assuming that $i \neq j$, note that $\lambda_j K_{\mathrm{inhom}}^{ij}(r,t) = \mathbb{E}^{!(0,0,i)} \left[Y_j(\mathcal{C}_r^t(0,0)\right]$ gives us the expected number of points of Y_j that fall within spatial distance r and temporal distance time t of a typical point of Y_i.

5.4.6 Further properties

5.4.6.1 Independent thinning

The first thing that may be pointed out is that when applying independent thinning to Y, i.e. when we retain each $(x,t,m) \in X$ according to some probability function

$0 \leq p(x,t,m) \leq 1$, $(x,t,m) \in (\mathbb{R}^d \times \mathbb{R}) \times \mathcal{M}$, the pcf of the thinned process coincides with the original one (Baddeley et al., 2000). This implies that $\mathcal{K}^{CD}(\cdot)$ and $K^{CD}_{\text{inhom}}(r,t)$ are not affected by this type of thinning.

5.4.6.2 Scaling

We next give a scaling result, which indicates the relationship between the two definitions of K-functions.

Theorem 2. *Consider $C, D \in \mathcal{B}(\mathcal{M})$ with $\nu(C)$ and $\nu(D)$ positive. For any $\beta = (\beta_S, \beta_T) \in (0,\infty)^2$ and a SOIRS MSTPP $Y = \{(x_i, t_i, m_i)\}_{i=1}^N$, define the rescaling*

$$\beta Y = \{(\beta_S x_i, \beta_T t_i, m_i)\}_{i=1}^N.$$

The marked inhomogeneous spatio-temporal K-function $K^{CD}_{\text{inhom}}(r,t;\beta)$ of βY satisfies

$$K^{CD}_{\text{inhom}}(r,t;(\beta_S,\beta_T)) = K^{CD}_{\text{inhom}}(r/\beta_S, t/\beta_T; (1,1)), \quad r,t \geq 0,$$

where $K^{CD}_{\text{inhom}}(r,t;(1,1)) = K^{CD}_{\text{inhom}}(r,t)$ is the marked inhomogeneous spatio-temporal K-function of Y.

Proof. Denote the pcf of Y by $g_Y(\cdot)$. As in (Cronie and van Lieshout, 2015, Section 4.3.), through a change of variables and the Campbell formula we find that the pcf of βY is given by

$$\begin{aligned} g_{\beta Y}((x_1,t_1,m_1),(x_2,t_2,m_2)) &= \frac{(\beta_S^d \beta_T)^{-2} \rho^{(2)}((x_1/\beta_S, t_1/\beta_T, m_1),(x_2/\beta_S, t_2/\beta_T, m_2))}{(\beta_S^d \beta_T)^{-1} \lambda(x_1/\beta_S, t_1/\beta_T, m_1)(\beta_S^d \beta_T)^{-1} \lambda(x_2/\beta_S, t_2/\beta_T, m_2)} \\ &= g_Y((x_1/\beta_S, t_1/\beta_T, m_1),(x_2/\beta_S, t_2/\beta_T, m_2)). \end{aligned}$$

Hence, by a change of variables,

$$\begin{aligned} &K^{CD}_{\text{inhom}}(r,t;\beta) = \\ &= \frac{1}{\nu(C)\nu(D)} \int_C \int_D \int_{\|x\| \leq r} \int_{|s| \leq t} g_Y((0,0,m_1),(x/\beta_S, s/\beta_T, m_2)) \mathrm{d}x \mathrm{d}s \nu(dm_2)\nu(dm_1) \\ &= \frac{1}{\nu(C)\nu(D)} \int_C \int_D \int_{\|\beta_S x\| \leq r} \int_{|\beta_T s| \leq t} g((0,0,m_1),(x,s,m_2)) \mathrm{d}x \mathrm{d}s \nu(dm_2)\nu(dm_1) \\ &= K^{CD}_{\text{inhom}}(r/\beta_S, t/\beta_T). \end{aligned}$$

□

Theorem 2 essentially tells us two things. To begin with, if we rescale the spatial and/or the temporal domain, and thereby the space-time locations of Y, then $K^{CD}_{\text{inhom}}(\cdot)$ changes in a natural way. Secondly, we note that, equivalently,

$$K^{CD}_{\text{inhom}}(r,t) = K^{CD}_{\text{inhom}}(r\beta_S, t\beta_T; (\beta_S,\beta_T)), \quad r,t \geq 0,$$

whereby $K^{CD}_{\text{inhom}}(r,t) = K^{CD}_{\text{inhom}}(r,r;(1,r/t)) = K^{CD}_{\text{inhom}}(r;(1,r/t))$, $r \geq 0$. In other words, $K^{CD}_{\text{inhom}}(r,t)$ may always be obtained through $K^{CD}_{\text{inhom}}(r)$ by applying proper scaling of Y_g, i.e. considering $\beta Y = \{(x_i, \beta_T t_i, m_i)\}_{i=1}^{N}$, where $\beta_T = r/t$. There are practical implications of this results; it is sufficient to define an estimator for $K^{CD}_{\text{inhom}}(r)$, $r \geq 0$ (however, this is not the choice that we will make when we define our estimators).

It may be noted from the proof of Theorem 2 that we may obtain a more general result, pertaining to $\mathcal{K}^{CD}(E)$. More specifically, we have that the marked spatio-temporal second-order reduced moment measure $\mathcal{K}^{CD}(\cdot;\beta)$ of βY satisfies $\mathcal{K}^{CD}(E;\beta) = \mathcal{K}^{CD}(\{(\beta_S x, \beta_T s) : (x,s) \in E\})$.

5.5 Statistical inference

The intensity function as well as our second-order summary statistics are probabilistic entities used to quantify first and second-order properties of a given point process. Turning to the real world, where we are given a marked spatio-temporal point pattern $\{(x_i, t_i, m_i)\}_{i=1}^{n}$, such as the earthquake data set, we are naturally interested in how we can statistically estimate these quantities, to better understand the data-generating mechanism in question. We do this by assuming that we have observed a realisation of a SOIRS MSTPP Y. Its ground process, Y_g, is formally defined on $\mathbb{R}^d \times \mathbb{R}$ but in practice we treat it as only observed within some bounded spatio-temporal region $W_S \times W_T \subseteq \mathbb{R}^d \times \mathbb{R}$, which is often referred to as the *study region*. We also restrict ourselves to the case where only one single point pattern is observed but we point out that most arguments below can be averaged over if one would have repetitions.

Being able to estimate the relevant quantities, we then proceed to considering different specific marking structures (recall Section 5.2.5). In particular we will consider some related hypothesis testing.

5.5.1 Voronoi intensity estimation

Writing $N = Y(W_S \times W_T \times \mathcal{M})$, if we can assume homogeneity in space-time, with $\nu(\cdot) = M(\cdot)$, so that $\lambda(x,t,m) \equiv \lambda > 0$, we simply estimate λ by means of $N/[\ell \otimes \nu](W_S \times W_T \times \mathcal{M})$. This is, however, a scenario that is rarely or never seen in practise, in particular not in the case of earthquakes.

As pointed out in (Vere-Jones, 2009, Section 3.2), when estimating the intensity function of a MSTPP, unless one can assume homogeneity, one should use a local/adapted/-variable approach, as opposed to global smoothing techniques, such as single bandwidth kernel estimators (Diggle, 2014; Van Lieshout, 2012; Silverman, 1986). Motivated by Barr and Schoenberg (2010), and in particular their study of earthquakes (in a purely spatial setting), we choose to consider a marked spatio-temporal version of the *Voronoi intensity estimator*.

We start by defining the Voronoi estimators for $\lambda(x,t,m)$ and $\lambda_g(x,t)$. They are constructed through Voronoi tessellations (see e.g. (Chiu et al., 2013)) generated by the metrics $d_\infty(\cdot,\cdot)$ and $d(\cdot,\cdot)$ in expression (5.1).

Definition 13. *The* spatio-temporal Voronoi intensity estimator *is defined by*

$$\widehat{\lambda}_g(x,t) \quad = \sum_{(y,v)\in Y_g\cap W_S\times W_T} \frac{\mathbf{1}\{(x,t)\in \mathcal{V}^g_{(y,v)}\cap W_S\times W_T\}}{\ell(\mathcal{V}^g_{(y,v)}\cap W_S\times W_T)}, \tag{5.20}$$

where $(x,t)\in W_S\times W_T$ *and the* spatio-temporal Voronoi tessellation *is given by*

$$\begin{aligned}\mathcal{V}_g = \{\mathcal{V}^g_{(x,t)}\}_{(x,t)\in Y_g} = \big\{(u,v)\in\mathbb{R}^d\times\mathbb{R} \quad &: \quad d_\infty((u,v),(x,t))\le d_\infty((u,v),(y,s)) \\ &\quad \textit{for any } (y,s)\in Y_g\setminus\{(x,t)\}\big\}_{(x,t)\in Y_g}.\end{aligned}$$

Recalling the metric $d'(\cdot,\cdot)$ *in* (5.1)*, the* marked spatio-temporal Voronoi tessellation *generated by* Y *is defined as* $\mathcal{V}=\{\mathcal{V}_{(x,t,m)}\}_{(x,t,m)\in Y}$*, where*

$$\begin{aligned}\mathcal{V}_{(x,t,m)} &= \big\{(u,v,z)\in(\mathbb{R}^d\times\mathbb{R})\times\mathcal{M} \quad : \quad d((x,t,m),(u,v,z))\le d((y,s,k),(u,v,z)) \\ &\qquad\qquad \textit{for any } (y,s,k)\in Y\setminus\{(x,t,m)\}\big\}_{(x,t,m)\in Y} \\ &= \big\{(u,v,z)\in(\mathbb{R}^d\times\mathbb{R})\times\mathcal{M} \quad : \quad \max\{\|x-u\|_{\mathbb{R}^d},|t-v|,d'(m,z)\}\le \\ &\qquad\qquad \le\max\{\|y-u\|_{\mathbb{R}^d},|s-v|,d'(k,z)\} \\ &\qquad\qquad \textit{for any } (y,s,k)\in Y\setminus\{(x,t,m)\}\big\}_{(x,t,m)\in Y}.\end{aligned}$$

Furthermore, the marked spatio-temporal Voronoi intensity estimator *is defined as*

$$\widehat{\lambda}(x,t,m) = \sum_{(x_i,t_i,m_i)\in Y\cap W_S\times W_T\times\mathcal{M}} \frac{\mathbf{1}\{(x,t,m)\in\mathcal{V}_{(x_i,t_i,m_i)}\}}{[\ell\otimes\nu](\mathcal{V}_{(x_i,t_i,m_i)}\cap W_S\times W_T\times\mathcal{M})}, \tag{5.21}$$

for $(x,t,m)\in W_S\times W_T\times\mathcal{M}$*.*

Note the explicit dependence on the choice of space-time-mark metric and reference measure above.

We next give the mass preservation and the unbiasedness of the estimators above.

Theorem 3. *The estimators* (5.21) *and* (5.23) *are mass-preserving, i.e. they integrate to the total number of points* N*, and unbiased as estimators of the expected total number of points.*

Proof. We only consider (5.21) since (5.23) is analogous. Starting with the mass-preservation,

$$\begin{aligned}&\int_{W_S\times W_T\times\mathcal{M}}\widehat{\lambda}(x,t,m)\nu(dm)\mathrm{d}x\mathrm{d}t = \\ &= \sum_{(x_i,t_i,m_i)\in Y\cap W_S\times W_T\times\mathcal{M}} \frac{\int_{W_S\times W_T\times\mathcal{M}}\mathbf{1}\{(x,t,m)\in\mathcal{V}_{(x_i,t_i,m_i)}\}\nu(dm)\mathrm{d}x\mathrm{d}t}{[\ell\otimes\nu](\mathcal{V}_{(x_i,t_i,m_i)}\cap W_S\times W_T\times\mathcal{M})} = N.\end{aligned}$$

Taking expectations on both sides yields

$$\mathbb{E}\left[\int_{W_S\times W_T\times\mathcal{M}}\widehat{\lambda}(x,t,m)\nu(dm)\mathrm{d}x\mathrm{d}t\right] = \mathbb{E}[N] = \int_{W_S\times W_T\times\mathcal{M}}\lambda(x,t,m)\nu(dm)\mathrm{d}x\mathrm{d}t.$$

□

5.5.1.1 Simplifying assumptions

Ideally, one does not impose too many conditions when finding the intensity estimate, unless convinced that specific conditions such as separability hold. We will next look at a few scenarios where we impose simplifying assumptions and we note that the need for them often is related to computational expenses.

We here need to introduce the Voronoi cells of the projections of Y (assuming that they are well defined). Recalling the projections Y_S and Y_T from (5.3) and defining the projection Y_M of Y on $\mathcal{M}$ in an identical fashion, let

$$\begin{aligned}
\mathcal{V}_S &= \{\mathcal{V}^S_x\}_{x\in Y_S} = \{u\in\mathbb{R}^d : \|u-x\|_{\mathbb{R}^d} \le \|u-y\|_{\mathbb{R}^d} \text{ for any } y\in Y_S\setminus\{x\}\}_{x\in Y_S},\\
\mathcal{V}_T &= \{\mathcal{V}^T_t\}_{t\in Y_T} = \{v\in\mathbb{R} : |v-t| \le |v-s| \text{ for any } s\in Y_T\setminus\{t\}\}_{t\in Y_T},\\
\mathcal{V}_M &= \{\mathcal{V}^M_m\}_{m\in Y_M} = \{z\in\mathcal{M} : d'(m,z) \le d'(k,m) \text{ for any } k\in Y_M\setminus\{m\}\}_{m\in Y_M},\\
\mathcal{V}_{T\times M} &= \{\mathcal{V}^{T\times M}_{(t,m)}\}_{(t,m)\in\mathbb{R}\times\mathcal{M}}\\
&= \big\{(v,z)\in\mathbb{R}\times\mathcal{M} : \max\{|t-v|, d'(m,z)\} \le \max\{|s-v|, d'(k,z)\}\\
&\qquad \text{for any } (s,k)\in Y_T\times Y_M\setminus\{(t,m)\}\big\}_{(t,m)\in Y_T\times Y_M}.
\end{aligned} \tag{5.22}$$

Some simplified setups are given by:

1. Separability and a common mark distribution:

$$\begin{aligned}
&\widehat{\lambda}(x,t,m) = \frac{1}{N^2}\widehat{\lambda}_S(x)\widehat{\lambda}_T(t)\widehat{\lambda}_M(m)\\
&= \frac{1}{N^2}\sum_{y\in Y_S\cap W_S}\frac{\mathbf{1}\{x\in\mathcal{V}^S_y\cap W_S\}}{\ell_d(\mathcal{V}^S_y\cap W_S)}\sum_{v\in Y_T\cap W_T}\frac{\mathbf{1}\{t\in\mathcal{V}^T_v\cap W_T\}}{\ell_1(\mathcal{V}^T_v\cap W_T)}\sum_{z\in Y_M}\frac{\mathbf{1}\{m\in\mathcal{V}^M_z\}}{\nu(\mathcal{V}^M_z)}.
\end{aligned}$$

   If we assume that the common mark distribution is given by $\nu(\cdot)$, we set $\lambda_M(m)/N \equiv 1$ above.

2. Non-separability and a common mark distribution:

$$\widehat{\lambda}(x,t,m) = \widehat{f}^{\mathcal{M}}(m)\widehat{\lambda}_g(x,t) = \frac{\widehat{\lambda}_M(m)}{N}\widehat{\lambda}_g(x,t).$$

   If the mark distribution and the reference measure coincide, we set $\widehat{f}^{\mathcal{M}}(m) \equiv 1$ above.

3. Separability and time-mark dependence:

$$\widehat{\lambda}(x,t,m) = \frac{\widehat{\lambda}_S(x)}{N}\sum_{(v,z)\in Y_T\times Y_M\cap W_T\times\mathcal{M}}\frac{\mathbf{1}\{(t,m)\in\mathcal{V}^{T\times M}_{(v,z)}\cap W_T\times\mathcal{M}\}}{[\ell_1\otimes\nu](\mathcal{V}^{T\times M}_{(v,z)}\cap W_T\times\mathcal{M})}. \tag{5.23}$$

   The case of separability and space-mark dependence is analogous.

As a corollary to Theorem 3 (the proof is identical), we obtain mass preservation and unbiasedness for the estimators above.

Corollary 1. *All the estimators above are mass preserving and unbiased.*

5.5.2 Estimation of the second-order summary statistics

We next give the definitions of the estimators of our previously defined second-order statistics. In order to account for edge effects (Cronie and Särkkä, 2011; Chiu et al., 2013; Gabriel, 2014) when defining the estimators below, we apply a *minus sampling/border correction scheme*. Denoting the boundaries of W_S and W_T by ∂W_S and ∂W_T, respectively, we write $W_S^{\ominus r} = \{x \in W_S : d_{\mathbb{R}^d}(x, \partial W_S) \geq r\} = \{x \in W_S : B_{\mathbb{R}^d}[x, r] \subseteq W_S\}$ and $W_T^{\ominus t} = \{s \in W_T : d_{\mathbb{R}}(x, \partial W_T) \geq t\}$ for the eroded spatial and temporal domains, respectively. Here $B_{\mathbb{R}^d}[x, r]$ is the closed ball in $\mathbb{R}^d$ with centre x and radius r.

Throughout we consider a SOIRS MSTPP Y, and assume that $\ell_d(W_S^{\ominus r}) > 0$, $\ell_1(W_T^{\ominus t}) > 0$ and $C, D \in \mathcal{B}(\mathcal{M})$, with $\nu(C), \nu(D) > 0$.

Definition 14. *The estimator $\widehat{K}_{\text{inhom}}^{CD}(r,t)$ of the marked inhomogeneous spatio-temporal K-function $K_{\text{inhom}}^{CD}(r,t)$, $r, t \geq 0$, based on $Y \cap W_S \times W_T \times \mathcal{M}$, is defined by*

$$\begin{aligned} &\ell_d(W_S^{\ominus r})\ell_1(W_T^{\ominus t})\nu(C)\nu(D)\widehat{K}_{\text{inhom}}^{CD}(r,t) = \\ &= \sum_{(x_1,t_1,m_1)\in Y\cap W_S^{\ominus r}\times W_T^{\ominus t}\times C} \; \sum_{(x_2,t_2,m_2)\in Y\cap \mathcal{C}_r^t(x_1,t_1)\times D\setminus\{(x_1,t_1,m_1)\}} \frac{1}{\lambda(x_1,t_1,m_1)\lambda(x_2,t_2,m_2)}. \end{aligned} \tag{5.24}$$

By replacing $\mathcal{C}_r^t(x_1,t_1)$ by $(x_1,t_1) + E$ in (5.24)*, $E \in \mathcal{B}(\mathbb{R}^d \times \mathbb{R})$, we obtain an estimator $\widehat{\mathcal{K}}^{CD}(E)$ of the marked spatio-temporal second-order reduced moment measure $\mathcal{K}^{CD}(E)$.*

Next, in Lemma 5, we turn to the unbiasedness of the estimators above

Lemma 5. *The estimators in Definition 14 are unbiased. The variance of $\widehat{\mathcal{K}}^{CD}(E)$ is given in expression* (5.25)*.*

Proof. By the Campbell formula and expression (5.12),

$$\begin{aligned} \mathbb{E}[\widehat{\mathcal{K}}^{CD}(E)] &= \frac{\int_{W_S^{\ominus r}\times W_T^{\ominus t}\times C}\int_{E\times D} g((x_1,t_1,m_1),(x_2,t_2,m_2))\mathrm{d}x_1\mathrm{d}t_1\nu(dm_2)\mathrm{d}x_2\mathrm{d}t_2\nu(dm_1)}{\ell(W_S^{\ominus r})\ell(W_T^{\ominus t})\nu(C)\nu(D)} \\ &= \frac{\ell\left(W_S^{\ominus r}\right)\ell\left(W_T^{\ominus t}\right)\int_C\int_{E\times D} g((0,0,m_1),(u,v,m_2))\mathrm{d}u\mathrm{d}v\nu(dm_2)\nu(dm_1)}{\ell(W_S^{\ominus r})\ell(W_T^{\ominus t})\nu(C)\nu(D)} \\ &= \mathcal{K}^{CD}(E), \end{aligned}$$

which implies that (5.24) is unbiased. We next turn to the variance and for simplicity we write $A = W_S^{\ominus r} \times W_T^{\ominus t}$. It follows that

$$
\begin{aligned}
&[\ell(A)\nu(C)\nu(D)]^2 \widehat{\mathcal{K}}^{CD}(E)^2 = \\
&= \sum_{(x_1,t_1,m_1),(x_2,t_2,m_2),(x_3,t_3,m_3),(x_4,t_4,m_4)\in Y} \frac{\mathbf{1}\{(x_1,t_1,m_1)\in A\times C\}\mathbf{1}\{(x_3,t_3,m_3)\in A\times C\}}{\lambda(x_1,t_1,m_1)\lambda(x_2,t_2,m_2)\lambda(x_3,t_3,m_3)\lambda(x_4,t_4,m_4)} \times \\
&\times \mathbf{1}\{(x_2,t_2,m_2)\in E\times D\setminus\{(x_1,t_1,m_1)\}\}\mathbf{1}\{(x_4,t_4,m_4)\in E\times D\setminus\{(x_3,t_3,m_3)\}\} \\
&= \sum_{(x_1,t_1,m_1)\in Y\cap A\times C} \; \sum_{(x_2,t_2,m_2)\in Y\cap E\times D\setminus\{(x_1,t_1,m_1)\}} \frac{1}{\lambda(x_1,t_1,m_1)^2\lambda(x_2,t_2,m_2)^2} \\
&+ \sum_{(x_1,t_1,m_1)\in Y\cap A\times C} \; \sum^{\neq}_{(x_2,t_2,m_2),(x_4,t_4,m_4)\in Y\cap E\times D\setminus\{(x_1,t_1,m_1)\}} \frac{\lambda(x_1,t_1,m_1)^{-2}}{\lambda(x_2,t_2,m_2)\lambda(x_4,t_4,m_4)} \\
&+ \sum^{\neq}_{(x_1,t_1,m_1),(x_2,t_2,m_2)\in Y\cap A\times C} \; \sum_{(x_3,t_3,m_3)\in Y\cap E\times D\setminus\{(x_1,t_1,m_1),(x_2,t_2,m_2)\}} \frac{\lambda(x_3,t_3,m_3)^{-2}}{\lambda(x_1,t_1,m_1)\lambda(x_2,t_2,m_2)} \\
&+ \sum^{\neq}_{(x_1,t_1,m_1),(x_2,t_2,m_2)\in Y\cap A\times C} \; \sum^{\neq}_{(x_3,t_3,m_3),(x_4,t_4,m_4)\in Y\cap E\times D\setminus\{(x_1,t_1,m_1),(x_2,t_2,m_2)\}} \frac{1}{\prod_{i=1}^4 \lambda(x_i,t_i,m_i)} \\
&= S_1 + S_2 + S_3 + S_4.
\end{aligned}
$$

By the Campbell formula,

$$
\mathbb{E}[S_4] = \int_{A\times C}\int_{A\times C}\int_{E\times D}\int_{E\times D} \frac{\rho^{(4)}((x_1,t_1,m_1),\ldots,(x_4,t_4,m_4))}{\lambda(x_1,t_1,m_1)\cdots\lambda(x_4,t_4,m_4)} \prod_{i=1}^4 \mathrm{d}x_i \mathrm{d}t_i \nu(dm_i),
$$

$$
\mathbb{E}[S_3] = \int_{A\times C}\int_{A\times C}\int_{E\times D} \frac{1}{\lambda(x_3,t_3,m_3)} \frac{\rho^{(3)}((x_1,t_1,m_1),\ldots,(x_3,t_3,m_3))}{\lambda(x_1,t_1,m_1)\cdots\lambda(x_3,t_3,m_3)} \prod_{i=1}^3 \mathrm{d}x_i \mathrm{d}t_i \nu(dm_i),
$$

$$
\mathbb{E}[S_2] = \int_{A\times C}\int_{E\times D}\int_{E\times D} \frac{1}{\lambda(x_1,t_1,m_1)} \frac{\rho^{(3)}((x_1,t_1,m_1),\ldots,(x_3,t_3,m_3))}{\lambda(x_1,t_1,m_1)\cdots\lambda(x_3,t_3,m_3)} \prod_{i=1}^3 \mathrm{d}x_i \mathrm{d}t_i \nu(dm_i),
$$

$$
\mathbb{E}[S_1] = \int_{A\times C}\int_{E\times D} \frac{1}{\lambda(x_1,t_1,m_1)\lambda(x_2,t_2,m_2)} g((x_1,t_1,m_1),(x_2,t_2,m_2)) \prod_{i=1}^2 \mathrm{d}x_i \mathrm{d}t_i \nu(dm_i),
$$

whereby

$$
\mathrm{Var}(\widehat{\mathcal{K}}^{CD}(E)) = \frac{\sum_{i=1}^4 \mathbb{E}[S_i]}{[\ell(A)\nu(C)\nu(D)]^2} - \mathcal{K}^{CD}(E)^2. \tag{5.25}
$$

□

Clearly, in practise $\lambda(\cdot)$ is not known so each $\lambda(x_i, t_i, m_i)$ must be replaced by an estimate $\widehat{\lambda}(x_i, t_i, m_i)$, which may obtained by e.g. the Voronoi estimation approach presented previously. In estimators such as $\widehat{\mathcal{K}}^{CD}(E)$ and (5.24), Stoyan and Stoyan (2000) advocate the Hamilton principle, which suggests replacing $\ell_d(W_S^{\ominus r})\ell_1(W_T^{\ominus t})\nu(C)$ by

$$\sum_{(x,s,m)\in Y\cap W_S^{\ominus r}\times W_T^{\ominus t}\times C} \frac{1}{\lambda(x,s,m)};$$

the latter is an unbiased estimator of the former, due to the Campbell formula. In essence, we may have one of the following scenarios:

1. All of $\ell_d(W_S^{\ominus r})$, $\ell_1(W_T^{\ominus t})$, $\nu(C)$ and $\nu(D)$ are (assumed) known: employ (5.24) for the estimation of $K_{\text{inhom}}^{CD}(r,t)$.
2. $\nu(C)$ and/or $\nu(D)$ is unknown but $\ell_d(W_S^{\ominus r})\ell_1(W_T^{\ominus t})$ is known: use the estimator

   $$\widehat{\nu(C)} = \frac{1}{\ell_d(W_S^{\ominus r})\ell_1(W_T^{\ominus t})} \sum_{(x,s,m)\in Y\cap W_S^{\ominus r}\times W_T^{\ominus t}\times C} \frac{1}{\lambda(x,s,m)}$$

   in (5.24). This is all analogous for $\nu(D)$.
3. $\nu(C)$ and $\nu(D)$ are known explicitly but $\ell_d(W_S^{\ominus r})\ell_1(W_T^{\ominus t})$ is unknown, with the ground intensity $\lambda_g(\cdot)$ (assumed) known explicitly: use the estimator

   $$\widehat{\ell_d(W_S^{\ominus r})\ell_1}(W_T^{\ominus t}) = \sum_{(x,s)\in Y_g\cap W_S^{\ominus r}\times W_T^{\ominus t}} \frac{1}{\lambda_g(x,s)}$$

   in (5.24).
4. Neither of $\ell_d(W_S^{\ominus r})$, $\ell_1(W_T^{\ominus t})$, $\nu(C)$ and $\nu(D)$ are (assumed) known but the ground intensity $\lambda_g(\cdot)$ is (assumed) known explicitly: estimate $\ell_d(W_S^{\ominus r})\ell_1(W_T^{\ominus t})\nu(C)\nu(D)$ by means of

   $$\frac{\sum_{(x,s,m)\in Y\cap W_S^{\ominus r}\times W_T^{\ominus t}\times C} \lambda(x,s,m)^{-1} \sum_{(x,s,m)\in Y\cap W_S^{\ominus r}\times W_T^{\ominus t}\times D} \lambda(x,s,m)^{-1}}{\sum_{(x,s)\in Y_g\cap W_S^{\ominus r}\times W_T^{\ominus t}} \lambda_g(x,s)^{-1}}$$

   and plug this into (5.24).

Note that this, in fact, means that when we are given the intensity functions $\lambda(x,t,m)$ and $\lambda_g(x,t)$, we do not need to explicitly know/provide $\nu(C)$ and $\nu(D)$. This setup provides (ratio) unbiased estimators when the intensity is known. To evaluate the performance of the four scenarios above, we employ each one to 99 realisations of the model in Example 1 and generate min/max envelopes (Diggle, 2014). The results can be found in Figure 5.5 and it seems that knowing the mark set measures is the most crucial part. Note, however, that the most realistic practical scenario is number 2.

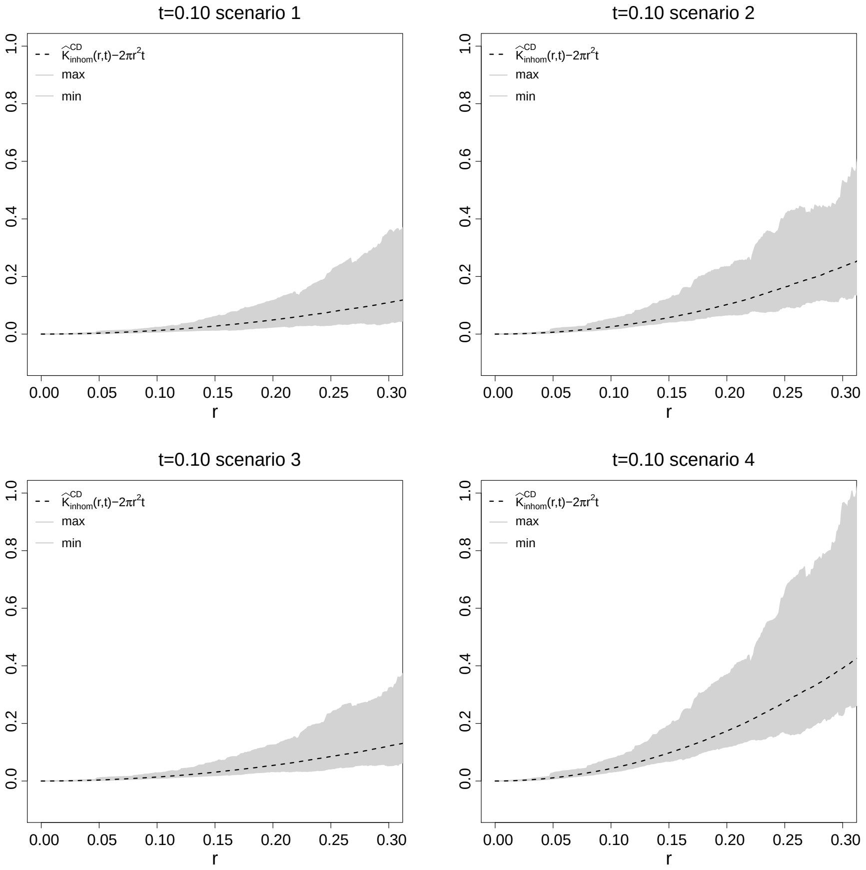

Figure 5.5. The four Hamilton principle scenarios for the estimator (5.24); min-max envelopes based on 99 realisations of the randomly labelled Poisson process in Example 1.

5.5.2.1 Smoothing

Recall from Section 5.4.6.1 that $K_{\text{inhom}}^{CD}(r,t)$ is invariant under independent thinning. This may be exploited to obtain a smoothing/thinning/bootstrapping scheme for the estimation of $K_{\text{inhom}}^{CD}(r,t)$. More specifically, let $\widehat{K}_{\text{inhom}}^{CD}(r,t;Y_i^p)$, $i=1,\dots,n$, be the estimators generated by n independent thinnings Y_i^p, $i=1,\dots,n$, of Y, using retention probability function $p(x,t,m)\equiv p\in(0,1)$. The resulting smoothed estimator is given by

$$\widetilde{K}_{\text{inhom}}^{CD}(r,t)=\frac{1}{n}\sum_{i=1}^{n}\widehat{K}_{\text{inhom}}^{CD}(r,t;Y_i^p).$$

In essence, we are averaging over n different unbiased estimators of $K_{\text{inhom}}^{CD}(r,t)$; hereby also $\widetilde{K}_{\text{inhom}}^{CD}(r,t)$ is unbiased. A clear gain with this approach is that we even out the negative effects of using only one misspecified plug-in intensity estimate, which has been generated by only one sample, as is the case of $\widehat{K}_{\text{inhom}}^{CD}(r,t)$. The drawback is that we get an increased variance. Regarding the choice of $p\in(0,1)$, we generally consider $p=0.5$ to be a decent choice (unless the dataset is small, which requires additional caution).

Remark 4. *In principle, one could consider bootstrap-type regions/envelopes for $K_{\text{inhom}}^{CD}(r,t)$, based on $\widehat{K}_{\text{inhom}}^{CD}(r,t;Y_i^p)$, $i=1,\dots,n$, provided that we choose some suitable function space metric (c.f. e.g. (Myllymäki et al., 2016)).*

5.5.2.2 Multivariate versions

We next turn to the estimation of a multivariate SOIRS STPP Y. From the general estimator in Definition 14, where $C=\{i\}$ and $D=\{j\}$, $i\neq j$, we obtain

$$\begin{aligned}&\widehat{K}_{\text{inhom}}^{ij}(r,t)=\\&=\frac{1}{\ell_d(W_S^{\ominus r})\ell_1(W_T^{\ominus t})}\sum_{(x_1,t_1)\in Y_i\cap W_S^{\ominus r}\times W_T^{\ominus t}}\frac{1}{\lambda_i(x_1,t_1)}\sum_{(x_2,t_2)\in Y_j\cap\mathcal{C}_r^t(x_1,t_1)}\frac{1}{\lambda_j(x_2,t_2)},\end{aligned}$$

and we see that this does not require explicit knowledge of $\nu(\cdot)$. Although not necessary here, it is common to assume that $\nu(\cdot)$ is the counting measure on $\mathcal{M}$.

Since $\lambda(x,t,i)=\lambda_i(x,t)/\nu(i)$, in practise, for each $i\in\mathcal{M}=\{1,\dots,k\}$ we obtain an estimate $\widehat{\lambda}_i(x,t)$ based on Y_i, which we plug into the estimators above. One may e.g. use either of the separable or non-separable ground process Voronoi intensity estimators proposed previously.

5.5.2.3 Common mark distribution versions

When there is a common mark distribution $M(\cdot)$, which coincides with the reference measure $\nu(\cdot)$ (recall Definition 7), we may estimate $\nu(C)\nu(D)=M(C)M(D)$, $C,D\in$

$\mathcal{B}(\mathcal{M})$, by $\widehat{\nu(C)\nu(D)} = Y(W_S \times W_T \times C)Y(W_S \times W_T \times D)/Y_g(W_S \times W_T)^2$ to obtain

$$\begin{aligned}
&\ell_d(W_S^{\ominus r})\ell_1(W_T^{\ominus t})\widehat{\nu(C)\nu(D)}\widehat{K}^{CD}_{\text{inhom}}(r,t) = \\
&= \sum_{(x_1,t_1)\in Y_C \cap W_S^{\ominus r}\times W_T^{\ominus t}} \; \sum_{(x_2,t_2)\in Y_D \cap \mathcal{C}_r^t(x_1,t_1)\setminus\{(x_1,t_1)\}} \frac{1}{\lambda_g(x_1,t_1)\lambda_g(x_2,t_2)},
\end{aligned}$$

where we plug in an estimate of $\lambda_g(\cdot)$ in practise. As before, we may perform Hamilton principle correction by replacing $\ell_d(W_S^{\ominus r})\ell_1(W_T^{\ominus t})$ by $\sum_{(x,s)\in Y_g \cap W_S^{\ominus r}\times W_T^{\ominus t}} \lambda_g(x,s)^{-1}$.

5.5.2.4 Homogeneous versions

In the stationary case, when the reference measure is given by the mark distribution (see Definition 7), given the ground intensity $\lambda > 0$, from the general estimator in Definition 14 we obtain

$$\widehat{K}^{CD}(r,t) = \frac{\sum_{(x_1,t_1)\in Y_C \cap W_S^{\ominus r}\times W_T^{\ominus t}} Y_D(\mathcal{C}_r^t(x_1,t_1)\setminus\{(x_1,t_1)\})}{\lambda^2 \ell_d(W_S^{\ominus r})\ell_1(W_T^{\ominus t})\nu(C)\nu(D)},$$

where we in practise replace λ by the estimate $\widehat{\lambda} = Y_g(W_S \times W_T)/(\ell_d(W_S)\ell_1(W_T))$ and $\nu(C)\nu(D)$ by $\widehat{\nu(C)\nu(D)} = Y(W_S \times W_T \times C)Y(W_S \times W_T \times D)/Y_g(W_S \times W_T)^2$. In the stationary and multivariate case we obtain

$$\widehat{K}^{ij}_{\text{inhom}}(r,t) = \frac{1}{\lambda_i\lambda_j\ell_d(W_S^{\ominus r})\ell_1(W_T^{\ominus t})} \sum_{(x_1,t_1)\in Y_i \cap W_S^{\ominus r}\times W_T^{\ominus t}} Y_j(\mathcal{C}_r^t(x_1,t_1)),$$

where λ_i is estimated by $\widehat{\lambda}_i = Y_i(W_S \times W_T)/(\ell_d(W_S)\ell_1(W_T))$, $i \in \mathcal{M} = \{1,\dots,k\}$.

5.5.3 Independence assumptions

We next look closer at how $K^{CD}_{\text{inhom}}(r,t)$ is affected by making different independence assumptions that are related to the marking structure. Recalling the definitions from Section 5.2.5, we start by looking at independent marking, which includes random labelling, to see how $K^{CD}_{\text{inhom}}(r,t)$ is affected. We then proceed to considering the scenario where points of Y with marks that belong to different mark sets C and D are independent. It should be noted that the main part of the results below, in essence, are translated versions of the results in (Cronie and van Lieshout, 2016).

Lemma 6 below suggests finding evidence of independent marking by comparing $K^{CD}_{\text{inhom}}(r,t)$ with its unmarked counterpart, i.e. considering $K^{CD}_{\text{inhom}}(r,t) - K^{Y_g}_{\text{inhom}}(r,t)$, where we recall the inhomogeneous K-function of the ground process, $K^{Y_g}_{\text{inhom}}(r,t) = \int_{\mathcal{C}_r^t(0,0)} g_g((0,0),(x,s))\mathrm{d}x\mathrm{d}s$.

Lemma 6. *Let $C, D \subseteq \mathcal{M}$ be Borel sets with $\nu(C)$, $\nu(D) > 0$ and assume that Y has independent marks. Then, Y and Y_g have the same pcf's (note the equivalence in SOIRS) and $K^{CD}_{\text{inhom}}(r,t) = K^{Y_g}_{\text{inhom}}(r,t)$.*

Proof. Recall that under the assumption of independent marks we have that $f^{\mathcal{M}}_{(x_1,t_1),\dots,(x_n,t_n)}(m_1,\dots,m_n) = \prod_{i=1}^n f^{\mathcal{M}}_{(x_i,t_i)}(m_i)$. Using Equation (5.5), we obtain that

$$\begin{aligned} g((x_1,t_1,m_1),(x_2,t_2,m_2)) &= \frac{f^{\mathcal{M}}_{(x_1,t_1),(x_2,t_2)}(m_1,m_2)}{f^{\mathcal{M}}_{(x_1,t_1)}(m_1) f^{\mathcal{M}}_{(x_2,t_2)}(m_2)} g_g((x_1,t_1),(x_2,t_2)) \\ &= g_g((x_1,t_1),(x_2,t_2)), \end{aligned}$$

whereby Y_g is SOIRS whenever Y is and

$$\begin{aligned} K^{CD}_{\text{inhom}}(r,t) &= \frac{1}{\nu(C)\nu(D)} \int_C \int_D \int_{\mathcal{C}^t_r(0,0)} g((0,0,m_1),(x,s,m_2)) \mathrm{d}x\mathrm{d}s\nu(dm_2)\nu(dm_1) \\ &= \int_{\mathcal{C}^t_r(0,0)} g_g((0,0),(x,s)) \mathrm{d}x\mathrm{d}s. \end{aligned}$$

□

We next evaluate Lemma 6 numerically, to ensure that our estimator is behaving properly. In order to do so, we simulate 99 realisations of the model given in Example 2 and for the fixed temporal lags $t \in \{0.05, 0.10, 0.15, 0.30\}$ we construct min/max-envelopes (see e.g. Diggle (2014)) for $\widehat{K}^{CD}_{\text{inhom}}(r,t) - \widehat{K}^{Y_g}_{\text{inhom}}(r,t)$, where $C = \{0\}$ and $D = \{1\}$, based on these 99 realisations. Figure 5.6 shows the envelopes obtained for the different values of t and we see that our estimator is behaving properly since the envelopes centre around 0. Also, in Figure 5.6 we find the estimates of $K^{CD}_{\text{inhom}}(r,t) - K^{Y_g}_{\text{inhom}}(r,t)$ for space lags $r \in [0, 0.3]$ and time lags $t \in [0, 0.3]$. One can see that the values of the estimated $\widehat{K}^{CD}_{\text{inhom}}(r,t) - \widehat{K}^{Y_g}_{\text{inhom}}(r,t)$ are close to 0.

Consider next the concept of independent components, which is the scenario where the restrictions $Y|_C = Y \cap (\mathbb{R}^d \times \mathbb{R} \times C)$ and $Y|_D = Y \cap (\mathbb{R}^d \times \mathbb{R} \times D)$, with ground processes Y_C and Y_D, are independent. This can be exemplified by considering a marked bivariate process $Y = (Y_1, Y_2)$, where each component $Y_j = \{(x_{ij}, t_{ij}, m_{ij})\}_{i=1}^{N_j}$, $j = 1, 2$, is a (dependently) marked process, but where Y_1 and Y_2 are mutually independent. In essence, this is the merging of two mutually independent populations, which have dependent marking structures within. Assessing possible dependence between $Y|_C$ and $Y|_D$, Lemma 7 below suggests comparing $K^{CD}_{\text{inhom}}(r,t)$ with $2\omega_d r^d t$; when $D = \mathcal{M} \setminus C$, it further suggests comparing $K^{C\mathcal{M}}_{\text{inhom}}(r,t)$ with $\frac{\nu(\mathcal{M}\setminus C)}{\nu(\mathcal{M})} 2\omega_d r^d t + \frac{\nu(C)}{\nu(\mathcal{M})} K^{CC}_{\text{inhom}}(r,t)$.

Lemma 7. *Let $C, D \in \mathcal{B}(\mathcal{M})$, with $\nu(C)$ and $\nu(D) > 0$, and let be Y is SOIRS, with $Y|_C$ and $Y|_D$ mutually independent. It follows that $K^{CD}_{\text{inhom}}(r,t) = 2\omega_d r^d t$ and when $D = \mathcal{M} \setminus C$, we have that $K^{C\mathcal{M}}_{\text{inhom}}(r,t) = \frac{\nu(\mathcal{M}\setminus C)}{\nu(\mathcal{M})} 2\omega_d r^d t + \frac{\nu(C)}{\nu(\mathcal{M})} K^{CC}_{\text{inhom}}(r,t)$.*

Proof. Under the assumption of independence between $Y|_C$ and $Y|_D$,

$$\begin{aligned} &\rho^{(2)}((x_1,t_1,m_1),(x_2,t_2,m_2)) = \\ &= \begin{cases} \rho^{(2)}((x_1,t_1,m_1),(x_2,t_2,m_2)) & \text{if } (m_1,m_2) \in C \times C \text{ or } (m_1,m_2) \in D \times D, \\ \lambda(x_1,t_1,m_1)\lambda(x_2,t_2,m_2) & \text{if } (m_1,m_2) \in C \times D \text{ or } (m_2,m_1) \in C \times D. \end{cases} \end{aligned}$$

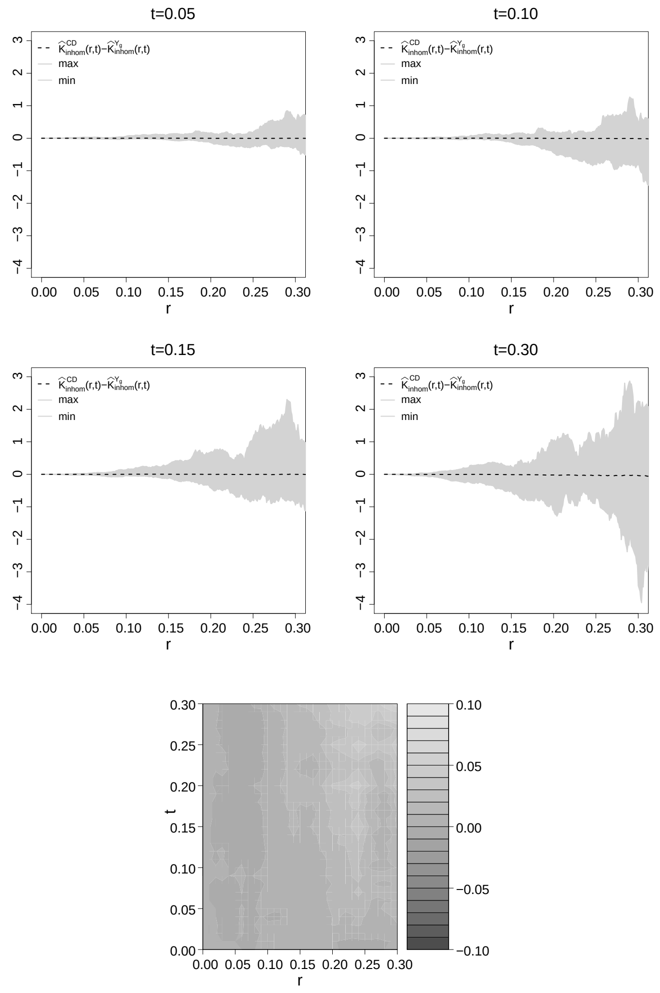

Figure 5.6. (Top) Envelopes for the estimate $\widehat{K}^{CD}_{\text{inhom}}(r,t) - \widehat{K}^{Y_g}_{\text{inhom}}(r,t)$, where $C = \{0\}$ and $D = \{1\}$, based on 99 realisations of the the model given in Example 2, for fixed temporal lags $t = 0.05$ and $t = 0.10$. (Centre) Envelopes for the estimate $\widehat{K}^{CD}_{\text{inhom}}(r,t) - \widehat{K}^{Y_g}_{\text{inhom}}(r,t)$,$t = 0.15$ for fixed temporal lags $t = 0.15$ and $t = 0.30$. (Bottom) The estimate $\widehat{K}^{CD}_{\text{inhom}}(r,t) - \widehat{K}^{Y_g}_{\text{inhom}}(r,t)$, for all space-time lags.

Hence, in the former case,

$$K^{CD}_{\text{inhom}}(r,t) = \frac{1}{\nu(C)\nu(D)} \int_C \int_D \int_{\mathcal{C}^t_r(0,0)} \mathrm{d}u\mathrm{d}v\nu(dm_1)\nu(dm_2) = \int_{\mathcal{C}^t_r(0,0)} \mathrm{d}u\mathrm{d}v = 2\omega_d r^d t$$

and in the latter case,

$$\begin{aligned}
K^{C\mathcal{M}}_{\text{inhom}}(r,t) &= \frac{1}{\nu(C)\nu(\mathcal{M})} \int_C \int_{\mathcal{M}\setminus C} \int_{\mathcal{C}^t_r(0,0)} \mathrm{d}u\mathrm{d}v\nu(dm_1)\nu(dm_2) + \\
&\quad + \frac{1}{\nu(C)\nu(\mathcal{M})} \int_C \int_C \int_{\mathcal{C}^t_r(0,0)} g((0,0,m_1),(u,v,m_2))\mathrm{d}u\mathrm{d}v\nu(dm_1)\nu(dm_2) \\
&= \frac{\nu(C)\nu(\mathcal{M}\setminus C)}{\nu(C)\nu(\mathcal{M})} \ell(\mathcal{C}^t_r(0,0)) + \\
&\quad + \frac{\nu(C)}{\nu(\mathcal{M})} \frac{1}{\nu(C)\nu(C)} \int_C \int_C \int_{\mathcal{C}^t_r(0,0)} g((0,0,m_1),(u,v,m_2))\mathrm{d}u\mathrm{d}v\nu(dm_1)\nu(dm_2) \\
&= \frac{\nu(\mathcal{M}\setminus C)}{\nu(\mathcal{M})} 2\omega_d r^d t + \frac{\nu(C)}{\nu(\mathcal{M})} K^{CC}_{\text{inhom}}(r,t).
\end{aligned}$$

□

To evaluate the above results numerically, we simulate 99 realisations of the model in Example 3 and consider $K^{CD}_{\text{inhom}}(r,t) - 2\omega_d r^d t$, where $C = \{0\}$ and $D = \{1\}$, for each one. The corresponding envelopes, which cover 0, are illustrated in Figure 5.7.

5.5.3.1 Testing random labelling

We here offer an alternative to testing the hypothesis of random labelling in the context of general MSTPPs, which does not require a particular shape of the study region, as is the case in e.g. the (inhomogeneous) Lotwick-Silverman test (Lotwick and Silverman, 1982; Cronie and van Lieshout, 2016). Note that we merely indicate how such a test may be constructed and that we do not formally test hypotheses here. For Monte-Carlo tests (see e.g. (Diggle, 2014)) such as the one described here we note that there are issues related to the choice of the number of simulations used to construct envelopes (see e.g. (Myllymäki et al., 2016)); unless executed properly, it is wise not to draw too strong conclusions and instead use them more loosely, as mere indicators of some hypothesis. Although $\widehat{K}^{CD}_{\text{inhom}}(r,t) - \widehat{K}^{Y_g}_{\text{inhom}}(r,t)$ gives us an indication on whether we have independent marking/random labelling, we cannot say exactly how large it has to be for us to infer anything. Hence, we need some formal way of testing such a hypothesis.

To construct a test, with the hypotheses H_0 : the marks are randomly labelled, and H_1 : the marks are not randomly labelled, we recall from Theorem 1 that a necessary condition for H_0 to hold is that $\mathcal{K}^{CD}(E) = \mathcal{K}^{DC}(E)$ for any $E \in \mathcal{B}(\mathbb{R}^d \times \mathbb{R})$ and any mark Borel sets C, D, with non-null ν-content. Hence, as test statistic we will use

$$\Delta(r,t) = K^{CD}_{\text{inhom}}(r,t) - K^{DC}_{\text{inhom}}(r,t), \quad r,t \geq 0.$$

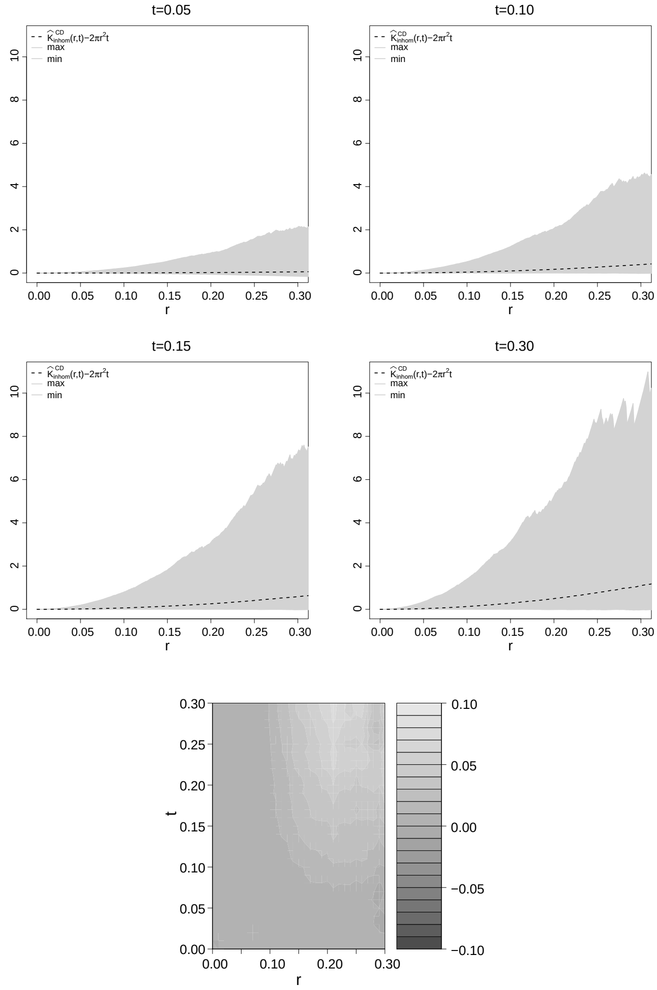

Figure 5.7. (Top) Envelopes for the estimated $K^{CD}_{\text{inhom}}(r,t) - 2\pi r^2 t$, where $C = \{0\}$ and $D = \{1\}$, based on 99 realisations of the the model given in Example 3, for fixed temporal lags $t = 0.05$ and $t = 0.10$. (Centre) Envelopes for the estimated $K^{CD}_{\text{inhom}}(r,t) - 2\pi r^2 t$, for fixed temporal lags $t = 0.15$ and $t = 0.30$. (Bottom) The estimate $\widehat{K}^{CD}_{\text{inhom}}(r,t) - 2\pi r^2 t$ for all r and t.

This may be exploited to construct a Monte-Carlo test, where the envelopes are generated by resampling the marks of Y, *without replacement*, and for each such mark-permuted version of Y estimate $\Delta(r,t)$. In essence, rejection of H_0 is based on whether the estimate of the original $\Delta(r,t)$, based on Y, sticks out of the envelopes for any $r, t \geq 0$ and any C, D.

Remark 5. *Note that the resampling of the marks requires that we assume that there is a common mark distribution, i.e. that we have random labelling. If one would have repeated observations of Y, on the other hand, one would also be able to test for independent marking.*

Furthermore, an alternative which we will not mention any further here is to consider, instead, resampling the marks with *replacement.*

We next evaluate the test above for a realisation of Example 4. More explicitly, we estimate $\Delta(r,t)$ for the realisation found in Figure 5.4 and then permute the marks in order to generate estimates $\Delta_i(r,t)$, $i = 1, \ldots, 99$, which in turn give rise to the envelopes. As we can see in Figure 5.8, the estimate of $\Delta(r,t)$ for $t = 0.20$ moves outside the envelopes, for certain values of r, which indeed would indicate that we do not have random labelling.

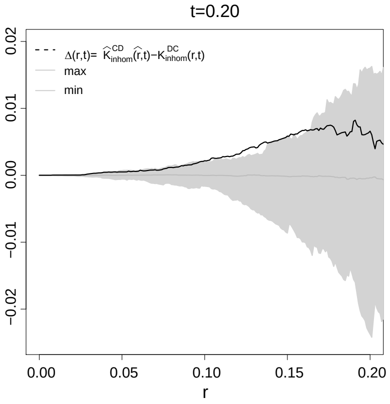

Figure 5.8. The estimated $\Delta(r,t) = K^{CD}_{\text{inhom}}(r,t) - K^{DC}_{\text{inhom}}(r,t)$, $Y_C = \{(x,t,m) \in [0,1] \times [0,1] \times [0,0.5]\}$, $Y_D = \{(x,t,m) \in [0,1] \times [0,1] \times (0.5,1]\}$, for the realisation in Figure 5.4, together with $\Delta(r,t)$-envelopes generated by 99 resamples/permutations of the marks, for fixed temporal lag $t = 0.20$.

Through Theorem 1 and its proof we note that the stronger the (spatio-temporal) dependence between the marks, the more clear the deviation of $\Delta(r,t)$ from the envelopes. Note further that the larger the size of the sample, the better the performance of the test, as one would expect.

5.6 Second order analysis of the earthquake data

5.6.1 Earthquakes data set

An earthquake is characterised by the shaking of the surface of the Earth and can range from being imperceptible to being devastating, with enormous damage and thousands of people killed. Historical data of earthquakes have shown that, on a year to year basis, there are some general patterns to be found. There are mainly three large areas of the earth with significant activity: i) the world's greatest earthquake belt, the circum-Pacific seismic belt, also known as *the Ring of Fire*, ii) *the Alpide*, which extends from Java to Sumatra through the Himalayas, to the Mediterranean, towards the Atlantic, and iii) the submerged *mid-Atlantic Ridge* (USGS, 2012).

On the 26th of December 2004 a huge earthquake, the *Sumatra-Andaman event* hit the Andaman sea with a magnitude of 8.8. As expected, and as most are aware of, the consequences were terrible, resulting in both tremendous material damage as well as a massive number of human lives ended. The epicentre of the earthquake was located offshore, thus creating a huge tsunami which led to the tragedy where a large number of people died. As stated in (Vigny et al., 2005), after the Sumatra-Andaman earthquake there were further small co-seismic jumps detected up to over $3,000$ kilometres (km) from the earthquake epicentre, within 10 minutes from the earthquake. Also, Vigny et al. (2005) state that post-seismic motion continued for a long period, 50 days after the earthquake in 2004 the island of Phuket moved 34 cm. Hence, the high magnitude earthquakes tend to produce a sort of domino effect, with small aftershocks triggering each other. Following this event, on the 28th of March 2005, another earthquake of 8.4 magnitude hit Nias, an area close to the Sumatra-Andaman region. This process started slowly and spread in two directions, first toward the north for approximately 100 km and then, after 40 seconds of delay, towards the south for about 200 km (Walker et al., 2005). Later, on the 12th of September 2007, two more earthquakes occurred in the Mentawai area, with magnitudes 8.5 and 8.1. According to Konca et al. (2008), the potential for a large event in this area remains high.

Earthquakes are registered using a seismographic network and the most common measure is the magnitude, which is a measure of the size of the earthquake source (USGS, 2012). Earthquakes of magnitude 3 or lower are almost undetectable and rarely felt. Earthquakes of magnitudes higher than 3 can cause landslides, which in turn can have fatal outcomes. Shocks of magnitude 7 and higher can cause severe landscape and building damage, and consequently human fatalities. When the epicentre of the earthquake is located offshore, there is also the possibility of tsunami development. Furthermore, often very large earthquakes are followed by a sequence of aftershocks, where the magnitude of the aftershocks can vary and some large aftershocks can have their own associated aftershock sequences (Harte, 2010).

Earthquake records often come in the format where an event (shock), in addition to having a spatial location and an event time, also carries further helpful information, such as magnitude.

In this thesis, and in particular, in Chapter 5, we analyse earthquake data from the

Sumatra region, registered from 2004 to 2008. The data in question can be downloaded freely from the R package `PtProcess` (Harte, 2010). It was originally extracted from the *preliminary determination of epicentres catalogue*, provided by the *US Geology Survey* (`ftp://hazards.cr.usgs.gov/pde/`). More specifically, it includes earthquakes registered in the area of Sumatra, Indonesia (part of the Alpide), with magnitudes (rounded to one decimal) larger than or equal to 5. The spatial region considered has boundaries 89° E, 105° E, 16° N and 5° S. We will consider a change in the spatial coordinates from longitude/latitude to UTM scale (Snyder, 1987). The time frame stretches from the midnight of the 1st of January 2004 until the 30th of December 2008, the day of the last registered shock. The first registered shock took place on the 16th of February 2004. A total of $1,248$ earthquakes were recorded during this period. Figure 5.9 (left) shows the spatial distribution of the point pattern of all $1,248$ earthquakes registered in the Sumatra area from the 16 February 2004 to 30 December 2008. The sizes of the black dots are proportional to the magnitudes of the events. The red X:s represent the four important earthquakes described previously. Figure 5.9 (right) shows the temporal development of the magnitudes. Furthermore, Figure 5.10 shows all earthquakes annually.

Note that an *aftershock* is an earthquake following a previous large shock, the *main shock*. The magnitude of an aftershock is smaller than the main shock. If the aftershock is larger than the main shock, the aftershock is labelled as main shock and the original main quake is labelled *foreshock* (USGS, 2012).

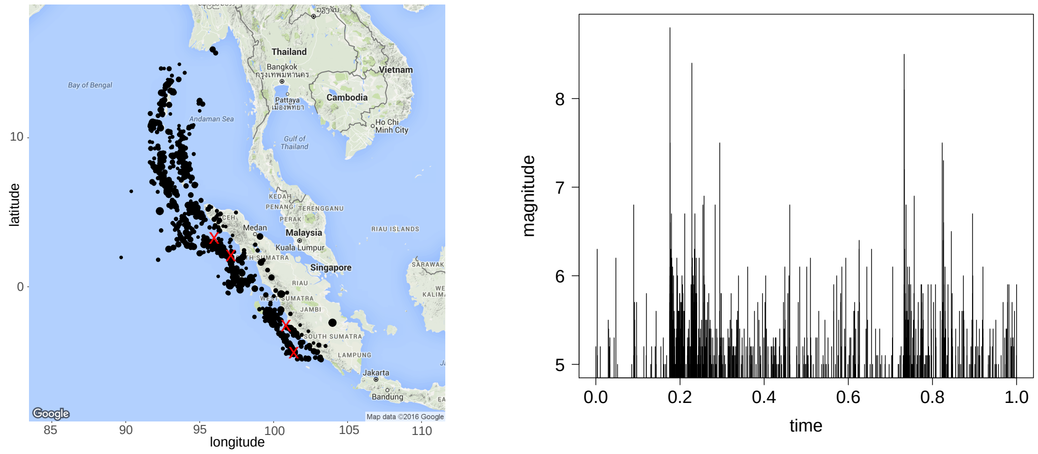

Figure 5.9. (Left) Spatial locations and magnitudes of the 1248 earthquakes registered in the Sumatra area. The sizes of the dots are proportional to the magnitudes. The red X:s correspond to the four important earthquakes described above. (Right) Temporal evolution of magnitude of all earthquakes.

Looking at Figure 5.9 (left) and Figure 5.10, we note that the earthquakes tend to appear in the same spatial region each year, the region being a reflection of the seismic belt. Hence, the spatial distribution of points in a given time period (Figure 5.10) may, essentially, be treated as a rescaling of an overall spatial distribution (Figure 5.9 left).

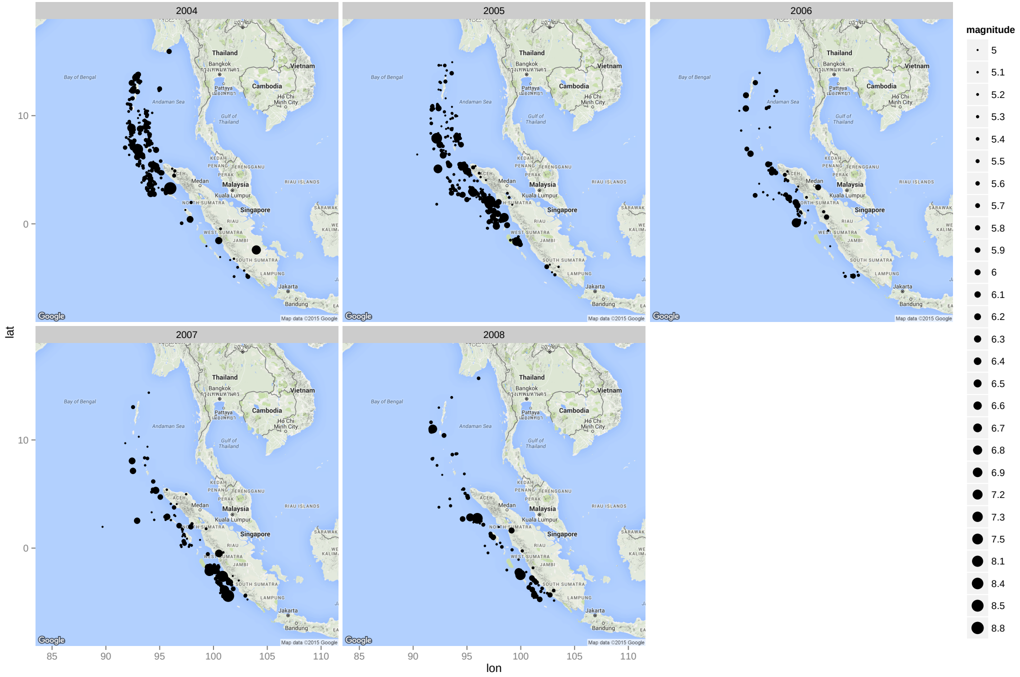

Figure 5.10. Spatial locations and magnitudes of the earthquake data, annually, from 2004 to 2008.

This observation makes reference to *space-time separability*, concept which will be further exploited in the analysis. Considering Figure 5.9 (right), we cannot exclude that there is dependence between the event times and the magnitudes, which may be an effect of earthquakes having fore-/aftershocks. Hence, we will not explicitly assume (first order) independence between the temporal component and the mark component of the data.

5.6.2 Analysis

As stated in Section 5.6.1, earthquakes are a huge threat to mankind's safety. Large magnitude earthquakes have produced serious landscape damage, but also human casualties; recall the effects of the huge Sumatra-Andaman event from 2004 with magnitude 8.8. The epicentre of the earthquake was located offshore, thus creating a huge tsunami which led to the tragedy where a large number of people died. In their paper, Vigny et al. (2005) state that, after the Sumatra-Andaman earthquake, post-seismic motion was detected at more than $3,000$ km away, and as late as 50 days after. This is an indication of a domino effect triggered by a big earthquake.

An *aftershock* is an earthquake following a previous large shock, the *main shock*. A major event tends to displace the crust of a tectonic plate, thus giving rise to the formation of aftershocks. The magnitude of an aftershock is smaller than the main shock. If the aftershock is larger than the main shock, the aftershock is labelled main shock and the original main quake is labelled *foreshock* (USGS, 2012). We want to study how far in space and time one may find aftershocks or foreshocks of different sizes.

The data analysed in this section, which consists of a total of $n = 1248$ earthquakes registered from 2004 to 2008, includes all earthquakes with magnitude larger than or equal to 5. The modified Mercalli intensity scale (USGS, 2012) classifies earthquakes into twelve classes, where shocks with magnitude larger than 6 can cause severe building and landscape damage, and human fatalities. Approximately 94.8% of all earthquakes registered in the Sumatra area have magnitude ≤ 6. These events cause minor wreckage, with limited damage to buildings and other structures. We want to study how far in space and time aftershocks and foreshocks (earthquakes with magnitude ≤ 6) appear after a big shock (magnitude > 6).

As previously mentioned, we have focused our analysis on developing point process tools which allow us to carry out second-order non-parametric analyses; recall that we consider the magnitudes as marks. Our objective is to analyse the interaction between different types of earthquakes, classified according to their magnitudes. More precisely, the K-function will give us information about the spatial and temporal scales at which points with marks (magnitude) in a certain category C, e.g. $C = \{\text{magnitude larger than } 6\}$, tend to cluster or tend to separate from points with marks in some other category D, e.g. $D = \{\text{magnitude less than or equal to } 6\}$, in the presence of inhomogeneity.

Formally, we consider a marked spatio-temporal point pattern $Y = \{(x_i, y_i, t_i, m_i)\}_{i=1}^{n} \subseteq (W_S \times W_T) \times \mathcal{M}$, $n = 1248$. Here $(x_i, y_i) \in \mathbb{R}^2$ is the spatial location of the ith event, $t_i \in \mathbb{R}$ is the number of days passed since the midnight of 1 January 2004 until the occurrence of the ith event, and m_i is the associated magnitude. As explained in Section 5.6.1, we transform the spatial latitude/longitude coordinates to UTM coordinates expressed in metres and rescale them. We use the following rescaling. Define $a = \min(y_i)$ and $b = \max(y_i)$. The new rescaled coordinates are $x_i' = (x_i - a)/(b - a)$ and $y_i' = (y_i - a)/(b - a)$, respectively, and $|b - a| = 2295032$ metres. The spatial study region becomes $W_S = [0, 0.7] \times [0, 1]$. We also rescale time. If $c = \min(t_i)$ and $d = \max(t_i)$, then the rescaled temporal component is $t_i' = (t_i - c)/(d - c)$, where $d - c = 1779.242$ days. The temporal window hereby becomes $W_T = [0, 1]$. Theorem 2 in Section 5.4.6.2 tells us that if we rescale the spatial and/or the temporal domain, the actual K-function estimates are obtained by simply scaling back the spatial and temporal lags. The largest earthquake ever to be recorded was in 1960, in Chile (Kanamori, 1977), with a magnitude of 9.5. Therefore, we set the magnitude scale to $[0, 10]$. Hence, we consider earthquakes with magnitude greater than 6 as belonging to mark set C, and shocks with magnitude less than or equal to 6 to D. Furthermore, as reference measure for the mark space we use the Lebesgue measure on the mark space $\mathcal{M} = [0, 10]$.

Figure 5.11 shows the marked spatio-temporal point pattern of all 1248 earthquakes registered in the Sumatra area from 16 of February 2004 until 30 of December 2008. Here the marks are represented as circles, with the size being proportional to the magnitude of the event. We suspect that the pattern is not regular since there are points that tend to be close to other points at all scales (in other words, not just inhomogeneity), so there seems to be clustering. We can identify small shocks (small circles) gathering around big earthquakes (large circles), but we cannot visually conclude anything significant. There are some areas of the study region where isolated small events are observed. Figure 5.10 (last plot) shows the temporal evolution of the earthquakes' magnitudes. This figure hints that, temporally, big earthquakes are preceded and followed by smaller foreshocks and aftershocks.

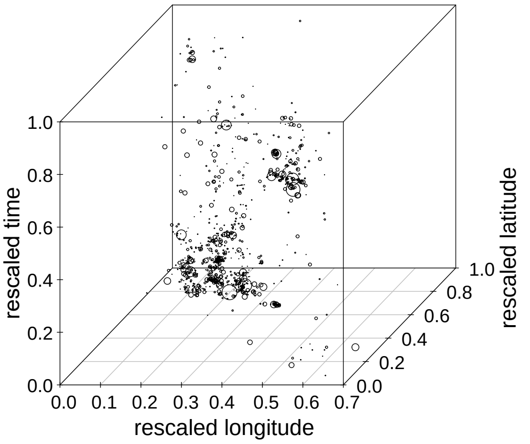

Figure 5.11. The marked spatio-temporal pattern of the earthquake dataset. The sizes of the circles are proportional to the magnitudes

Recall from Section 5.6.1 that we do not assume that there is first-order dependence between the spatial and the temporal components, i.e. we think it is justified to assume separability. In addition, as already mentioned above and indicated in Section 5.6.1, we will assume that there is first-order dependence between the temporal component and the marks. This leads us to the intensity estimator $\widehat{\lambda}(x, t, m)$ given in expression (5.23). This estimator requires that we use the Voronoi tessellation $\mathcal{V}_{T\times M}$ in (5.22). Our numerical implementation of this max-metric tessellation turned out to be too slow for the analysis of this data set. As an approximation, we chose to replace $\mathcal{V}_{T\times M}$ in (5.23) by the Euclidean Voronoi tessellation

$$\begin{aligned}\widetilde{\mathcal{V}}_{T\times M} &= \{\widetilde{\mathcal{V}}^{T\times M}_{(t,m)}\}_{(t,m)\in\mathbb{R}\times\mathcal{M}} \\ &= \big\{(v,z)\in\mathbb{R}\times\mathcal{M} : \|(v,z)-(t,m)\|_{\mathbb{R}^2} \le \|(v,z)-(s,k)\|_{\mathbb{R}^2} \\ &\qquad \text{for any } (s,k)\in Y_T\times Y_M\setminus\{(t,m)\}\big\}_{(t,m)\in Y_T\times Y_M}\end{aligned}$$

and evaluated it numerically by means of the implementation found in the `R` package `spatstat` (Baddeley and Turner, 2005). We believe that this approximation generates intensity estimates of a similar kind (the difference will be particularly small when employing the smoothed $K^{CD}_{\text{inhom}}(r, t)$ estimate).

Figure 5.12 (left) shows the estimate of $K^{CD}_{\text{inhom}}(r,t) - 2\pi r^2 t$ for spatial lags r range between 0 and 575 km, and temporal lags t range between 0 and 445 days. Figure 5.12 (right) shows the smoothed K-function estimate (retention probability $p = 0.5$ and 100 bootstrap samples), for the same spatial and temporal lags r and t. The behaviour does not change significantly for different choices of p. Figure 5.13 shows the smoothed K-function for three different smaller temporal scales, chosen as day, week and 50 days.

Figure 5.12 indicates clustering, since the K-functions are larger than $2\pi r^2 t$, at all spatio-temporal scales. This indicates that events in category D, meaning foreshocks or aftershocks, tend to cluster around events in category C. The strongest clustering between main shocks and foreshocks/aftershocks seems to occur at a temporal lag of approximately $200 - 300$ days, at all spatial scales. There seems to be an almost linear build-up of interaction and afterwards there seems to be a rapid decay in clustering. The majority of the fore-/aftershocks seem to occur at spatial distances larger than 200 km from a main shock. We emphasise that aftershocks are observed at distances quite far from the main shock. Looking at Figure 5.13, there seem to be predominant inter-event distances at which most fore-/aftershocks tend to occur; note the peaks around 300 and 500 km. Figure 5.13 (left) shows that within a day, aftershocks tend to travel far, even as far as 500 km. Looking at the temporal lags in all three representations in Figure 5.13 we can see that there are fore-/aftershocks occurring in direct connection to the main shock. We note that close in space and time there seem to be few fore-/aftershocks in connection to a main earthquake.

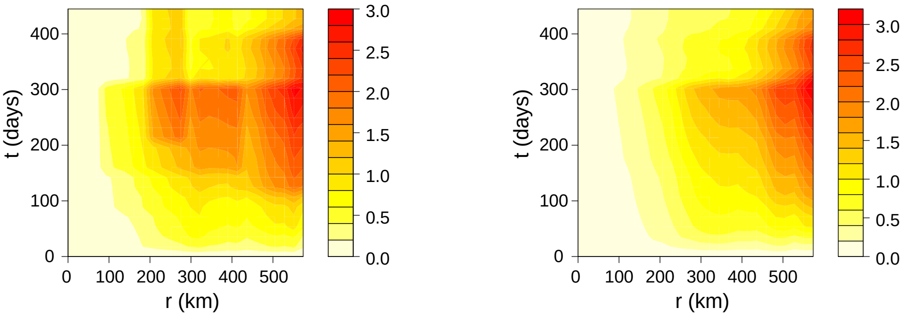

Figure 5.12. (Left) Estimated K-function for the Phuket data, $K^{CD}_{\text{inhom}}(r,t) - 2\pi r^2 t$. (Right) Smoothed K-function estimate, $\widetilde{K}^{CD}_{\text{inhom}}(r,t) - 2\pi r^2 t$.

In the literature it is sometimes considered that magnitude does not depend on the spatio-temporal location of the event (USGS, 2012). We next briefly look for indications of this belief by means of executing our random labelling test in Section 5.5.3.1, based on 99 permutations of the marks, where we have used 95% two-sided point-wise confidence bands. We found that for small and medium t the estimate of $\Delta(r,t)$ stays within the envelopes for all considered spatial lags r. For very large t, as indicated in Figure 5.14, we see that the estimate of $\Delta(r,t)$ sticks out of the envelope, thus indicating the possibility of the marks not being randomly labelled. It is advised not to draw too strong

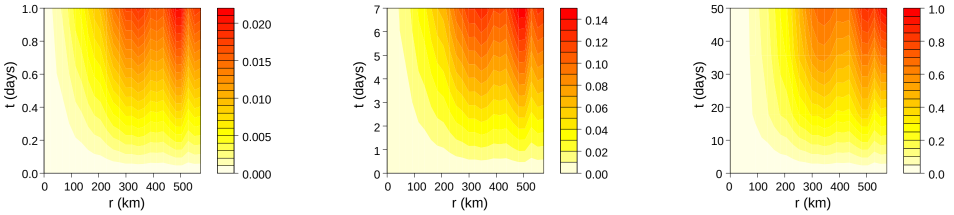

Figure 5.13. (Left) Smoothed K-function estimate, $\widetilde{K}^{CD}_{\text{inhom}}(r,t) - 2\pi r^2 t$, for the time frame of one day, (centre) one week and (right) 50 days.

conclusions, however, as indicated in Section 5.5.3.1.

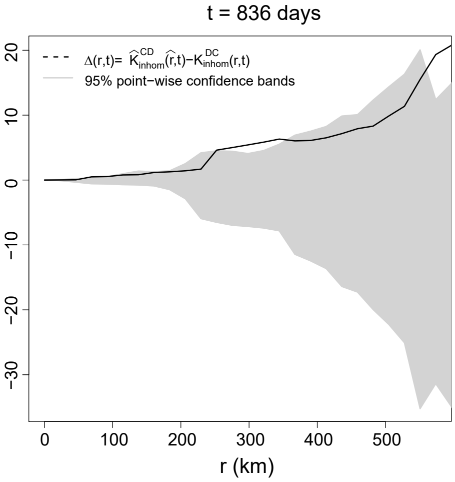

Figure 5.14. Estimate of $\Delta(r,t) = K^{CD}_{\text{inhom}}(r,t) - K^{DC}_{\text{inhom}}(r,t)$, for the earthquake dataset, together with 95% two-sided point-wise Monte-Carlo confidence bands, for fixed temporal lag $t = 836$ days.

5.7 Conclusion

In this chapter we have treated the second-order analysis of marked spatio-temporal point processes. In particular, we have defined measures of second-order spatio-temporal interaction, which allow us to quantify interactions between categories of marked points. For all statistics defined we derive unbiased estimators. In addition, we have considered an unbiased marked spatio-temporal Voronoi intensity estimation scheme, which allows us to estimate the underlying intensity function in an adaptive fashion. The set-up is quite general in the sense that the mark space as well as the corresponding mark reference measure are allowed to be arbitrary. We also exploit our newly defined tools to devise tests for particular marking structures.

The motivation behind this work comes from the necessity to analyse the interaction between main earthquakes and their fore-/aftershocks. We apply our methods to a well studied earthquake dataset (Harte, 2010) and conclude that there are strong and far-reaching interactions between main shocks and other shocks. Also, we see some evidence that, given the spatio-temporal locations, the magnitudes are not behaving like an iid sequence of random variables (random labelling).

Chapter 6

Future work

In this thesis we have studied different types of statistical analysis depending on the problem we wanted to work on. We have analysed spatial, spatio-temporal and marked spatio-temporal data. We have worked with aggregated data and point patterns. We want to emphasize that the thesis is composed by two applications, Chapters 2 and 3 and two methodological contributions, Chapters 4 and 5. This thesis represents a starting point. Some of the current and future lines of research are presented down below.

As we have already mentioned, varicella is a highly infectious disease. We are certain that, in addition to the effect of the population, there are other covariates that may influence the spatio-temporal behaviour of the disease. Therefore, an important goal for future work is to consider adding covariates that can improve our analyses. For example, WHO (2014) suggests that there are some climatic factors that can influence the epidemiology of varicella. Thus, covariates such as the monthly average temperatures, weekly average levels of rainfall, average hours of sunshine, or other climate related covariates, may provide useful information to the analysis of varicella. Also, additional information on the income per capita or other socio-economical covariates might improve our analysis.

Another interesting factor that could influence the spread of the disease is the location of the river Turia. Epidemiologists believe any natural geographic barrier somehow marks a difference, as children who live on one side of the river mostly go to schools on that side of the river and those living on the opposite side attend schools on that side. The role of schools and kindergartens in the transmission of chickenpox is crucial. Considering that today most couples only have one or two children, transmission within the family probably has less influence than school transmission (including childcare).

For future work, it would be interesting to apply the model presented in Chapter 4 to other diseases that may exhibit interaction at several scales in space and time. It would also be very interesting to apply the model to data that are not necessarily related to epidemiology. Earthquake patterns, for instance, tend to show aggregation but also inhibition at different scales. Indeed, we believe that the proposed model may find applications in a wide range of research fields, such as forestry, geology and sociology.

Other direct applications of the methodology presented in Chapter 5 can be found in e.g. epidemiology and criminology. We are currently looking at datasets related to these fields. In particular, our dataset of varicella cases. Furthermore, we are analysing crime data in Valencia, Spain. Note that for both applications it may be more relevant to consider multivariate versions of the summary statistics.

We are currently also working on extending the ideas presented in Chapter 5 to marked spatio-temporal point processes on linear networks (Ang et al., 2012). In addition, we are looking at combining the results of Cronie and van Lieshout (2015, 2016) in the same way as done in Chapter 5, in order to obtain inhomogeneous marked spatio-temporal versions of the nearest-neighbour distance distribution function, the empty-space function and the J-function. We believe that these may be more powerful tools for the purpose of non-parametric statistical analyses.

Bibliography

Alp, H., Altınkaynak, S., Ertekin, V., Kılıcaslan, B., and Giiraksin, A. (2005). Seroepidemiology of varicella-zoster virus infection in a cosmopolitan city (erzurum) in the eastern turkey. *Health Policy*, 72:119–124.

Altman, N. (1992). An introduction to kernel and nearest-neighbor nonparametric regression. *The American Statistician*, 46(3):175–185.

Ambler, G. K. and Silverman, B. W. (2010). Perfect simulation using dominated coupling from the past with application to area-interaction point processes and wavelet thresholding. In Bingham, N. H. and Goldie, C. M., editors, *Probability and Mathematical Genetics*, pages 64–90. The London Mathematical Society, Cambridge.

Anderson, T. W. and Goodman, L. A. (1957). Statistical inference about markov chains. *Ann. Math. Statist.*, 28(1):89–110.

Ang, Q. W., Baddeley, A., and Nair, G. (2012). Geometrically corrected second order analysis of events on a linear network, with applications to ecology and criminology. *Scandinavian Journal of Statistics*, 39(4):591–617.

Baddeley, A., Chang, Y., Song, Y., and Turner, R. (2012). Nonparametric estimation of the dependence of a point process on spatial covariates. *Statistics and Its Interface*, 5(2):221–236.

Baddeley, A., Gregori, P., Mateu, J., Stoica, R., and Stoyan, D. (2006). *Case studies in Spatial Point Pattern Modelling.* Springer-Verlag, New York.

Baddeley, A., Møller, J., and Waagepetersen, R. (2000). Non- and semi-parametric estimation of interaction in inhomogeneous point patterns. *Statistica Neerlandica*, 54:329–350.

Baddeley, A., Rubak, E., and Møller, J. (2011). Score, pseudo-score and residual diagnostics for spatial point process models. *Statistical Science*, 26(4):613–646.

Baddeley, A., Rubak, E., and Turner, R. (2015). *Spatial Point Patterns: Methodology and Applications with R.* Chapman and Hall/CRC Press, London.

Baddeley, A. and Turner, R. (2005). spatstat: An R package for analyzing spatial point patterns. *Journal of Statistical Software*, 12(6):1–42.

Baddeley, A., Turner, R., Mateu, J., and Bevan, A. (2013). Hybrids of Gibbs Point Process Models and Their Implementation. *Journal of Statistical Software*, 55(11):1–43.

Baddeley, A. J. and Turner, R. (2000). Practical maximum pseudolikelihood for spatial point patterns. *Australian and New Zealand Journal of Statistics*, 42(3):283–322.

Baddeley, A. J. and van Lieshout, M. N. M. (1995). Area-interaction point processes. *Annals of the Institute of Statistical Mathematics*, 47:601–619.

Barker, A. A. (1965). Monte carlo calculation of radial distribution functions for a proton-electron plasma. *Australian Journal of Physics*, 18:119–133.

Barnard, G. A. (1963). Discussion of The spectral analysis of point processes by M. S. Bartlett. *Journal of the Royal Statistical Society. Series B (Methodological)*, 25(2):294–296.

Barr, C. D. and Schoenberg, F. P. (2010). On the Voronoi estimator for the intensity of an inhomogeneous planar poisson process. *Biometrika*, 97(4):977–984.

Barro, R. and Sala-i-Martin, X. (2004). *Economic Growth.* MIT Press, Cambridge, Massachusetts, second edition.

Bartlett, M. S. (1964). The spectral analysis of two-dimensional point processes. *Biometrika*, 51(3/4):299–311.

Berman, M. (1986). Testing for spatial association between a point process and another stochastic process. *Applied Statistics*, 35:54–62.

Bernardinelli, L., Clayton, D., Pascutto, C., Montomoli, C., Ghislandi, M., and Songini, M. (1995). Bayesian analysis of space-time variation in disease risk. *Statistics in medicine*, 14(21-22):2433–2443.

Besag, J. (1975). Statistical analysis of non-lattice data. *The Statistician*, 24(3):179–195.

Besag, J., York, J., and Mollié, A. (1991). Bayesian image restoration, with two applications in spatial statistics. *Annals of the Institute of Statistical Mathematics*, 43(1):1–20. Cited By (since 1996):989.

Bivand, R. (2014). spdep: Spatial dependence: weighting schemes, statistics and models.

Bivand, R., Keitt, T., and Rowlingson, B. (2016). *rgdal: Bindings for the Geospatial Data Abstraction Library.* R package version 1.1-10.

Blangiardo, M., Cameletti, M., Baio, G., and Rue, H. (2013). Spatial and spatio-temporal models with R-INLA. *Spatial and Spatio-temporal Epidemiology*, 7:39–55.

Brachman, P. S. (2003). Infectious diseases - past, present, and future. *International Journal of Epidemiology*, 32:684–686.

Centers for Disease Control and Prevention (2012). Epidemiology and Prevention of Vaccine-Preventable Disease. Varicella. *The Pink Book: Course Textbook*, 12.

Chiu, S. N., Stoyan, D., Kendall, W. S., and Mecke, J. (2013). *Stochastic Geometry and Its Applications*. Wiley, 3nd ed. edition.

Choi, E. and Hall, P. (1999). Nonparametric Approach to Analysis of Space-Time Data on Earthquake Occurrences. *Journal of Computational and Graphical Statistics*, 8(4):733–748.

Clayton, D. (1996). Generalized linear mixed models. *Markov Chain Monte Carlo in Practice Gilks W, et al (eds)*, pages 275–301.

Cliff, A. and Ord, J. (1973). *Spatial Autocorrelation*. Pion, London.

Comas, C. and Mateu, J. (2011). Statistical inference for Gibbs point processes based on field observations. *Stochastic Environmental Research and Risk Assessment*, 25(2):287–300.

Comas, C., Palahi, M., Pukkala, T., and Mateu, J. (2009). Characterising forest spatial structure through inhomogeneous second order characteristics. *Stochastic Environmental Research and Risk Assessment*, 23(3):387–397.

Cooper Robbinsa, S., Wardb, K., and Skinnera, S. R. (2011). School-based vaccination: A systematic review of process evaluations. *Vaccine*, 29:9588–9599.

Cressie, N. and Read, T. (1984). Multinomial goodness-of-fit tests. *Journal of the Royal Statistical Society, Series B*, 46:440–464.

Cressie, N. A. C. (1991). *Statistics for spatial data*. Wiley, New York.

Cronie, O. and Särkkä, A. (2011). Some edge correction methods for marked spatio-temporal point process models. *Computational Statistics and Data Analysis*, 55(7):2209 – 2220.

Cronie, O. and van Lieshout, M. N. M. (2015). A J-function for inhomogeneous spatio-temporal point processes. *Scandinavian Journal of Statistics*, 42:562–579.

Cronie, O. and van Lieshout, M. N. M. (2016). Summary statistics for inhomogeneous marked point processes. *Annals of the Institute of Statistical Mathematics*, 68(4):905–928.

Czado, C., Gneiting, T., and Held, L. (2009). Predictive model assessment for count data. *Biometrics*, 65(4):1254–1261.

Daley, D. J. and Vere-Jones, D. (2003). *An Introduction to the Theory of Point Processes: Volume I: Elementary Theory and Methods*. Springer, New York.

Daley, D. J. and Vere-Jones, D. (2008). *An Introduction to the Theory of Point Processes: Volume II: General Theory and Structure*. Springer, New York.

Dereudre, D., Lavancier, F., and Helisová, K. (2014). Estimation of the intensity parameter of the germ-grain quermass-interaction model when the number of germs is not observed. *Scandinavian Journal of Statistics*, 41(3):809–829.

Diggle, P. (1985). A kernel method for smoothing point process data. *Applied Statistics (Journal of the Royal Statistical Society, Series C)*, 34:138–147.

Diggle, P. (2014). *Statistical Analysis of spatial and spatio-temporal point patterns*. CRC Press, Taylor Francis Group.

Diggle, P. J., Chetwynd, A., Häggkvist, R., and Morris, S. (1995). Second-order analysis of space-time clustering. *Statistical Methods in Medical Research*, 4:124–136.

European Centre for Disease Prevention and Control (2014). Varicella vaccine in the European Union. *Stockholm: ECDC*.

Funwi-Gabga, N. and Mateu, J. (2012). Understanding the nesting spatial behaviour of gorillas in the kagwene sanctuary, cameroon. *Stochastic Environmental Research and Risk Assessment*, 26:793–811.

Gabriel, E. (2014). 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.

Gabriel, E. and Diggle, P. J. (2009). Second-order analysis of inhomogeneous spatio-temporal point process data. *Statistica Neerlandica*, 63(1):43–51.

Gershon, A., Takahashi, M., and Seward, J. (2008). *Vaccines*, chapter Varicella Vaccine, pages 916–958. Philadelphia, WB Saunders, 5th edition.

Geyer, C. J. (1999). Likelihood inference for spatial point processes. In Barndorff-Nielsen, O. E., Kendall, W. S. and van Lieshout, M. N. M., editor, *Stochastic Geometry: Likelihood and Computation*. Chapman and Hall/CRC.

Gneiting, T. (2002). Nonseparable, stationary covariance functions for space-time data. *Journal of the American Statistical Association*, 97(458):590–600.

Gneiting, T. and Raftery, A. E. (2007). Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association*, 102(477):359–378.

Gosset, W. S. (1907). The probable error of a mean. *Biometrika*, 6(1).

Gregori, P., van Lieshout, M. N. M., and Mateu, J. (2003). Modelización de procesos area-interacción generalizados. In *27 Congreso Nacional de Estadística e Invesigación Operativa*.

Haapanene, R., Ek, A. R., Bauer, M. E., and Finley, A. O. (2004). Delineation of forest/nonforest land use classes using nearest neighbor methods. *Remote Sensing of Environment*, 89:265–271.

Haenggi, M. (2012). *Stochastic Geometry for Wireless Networks*. Cambridge University Press, New York, NY, USA.

Häggström, O., van Lieshout, M. N. M., and Møller, J. (1999). Characterization results and markov chain monte carlo algorithms including exact simulation for some spatial point processes. *Bernoulli*, 5(41):641–658.

Halberg, F., Tong, Y. L., and Johnson, E. A. (1967). *The Cellular Aspects of Biorhythms*, chapter Circadian System Phase – An Aspect of Temporal Methodology; Procedures and Illustrative Examples. Springer.

Halley, E. (1686). *An historical account of the Trade Winds, and Monffons, observable in the seas between and near the Tropicksm with an attempt to assign the Pacifical cause of the Sand Winds*.

Halmos, P. R. (1974). *Measure Theory*. Springer.

Harte, D. (2010). Ptprocess: An R package for modelling marked point processes indexed by time. *Journal of Statistical Software*, 35(8):1–32.

Health Department (2014). Varicella Report. Epidemiological Surveillance 2013. Public Health. Epidemiology and Health Surveillance. Epidemiological Surveillance and Control.

Heinrich, L. (2013). *Asymptotic Methods in Statistics of Random Point Processes*, pages 115–150. Springer, Berlin, Heidelberg.

Illian, J., Penttinen, A., Stoyan, H., and Stoyan, D. (2008). *Statistical analysis and modelling of spatial point patterns*. John Wiley & Sons.

Jackson, C., Mangtani, P., Fine, P., and Vynnycky, E. (2014). The effects of school holidays on transmission of varicella zoster virus, england and wales, 1967–2008. *PLOS ONE*, 9(6):1–9.

Kanamori, H. (1977). The energy release of great earthquakes. *J. Geophys. Res.*, 82:2981–2987.

Kelly, F. P. and Ripley, B. D. (1990). On strausss model for clustering. *Biometrika*, 63:357–360.

Kendall, W. S. (1998). *Proceedings of the Symposium on Probability towards the year 2000*, chapter Perfect simulation for the area-interaction point process. Springer-Verlag, Berlin.

Kendall, W. S., Van Lieshout, M. N. M., and Baddeley, A. J. (1999). Quermass-interaction processes: conditions for stability. *Advances in Applied Probability*, 31(2):315–342.

Knorr-Held, L. (2000). Bayesian modelling of inseparable space-time variation in disease risk. *Statistics in medicine*, 19(17-18):2555–2567.

Konca, A. O., Avouac, J., Sladen, A., Meltzner, A. J., Sieh, K., Fang, P., Li, Z., Galetzka, J., Genrich, J., Chlieh, M., Natawidjaja, D. H., Bock, Y., Fielding, E. J., Ji, C., and Helmberger, D. V. (2008). Partial rupture of a locked patch of the sumatra megathrust during the 2007 earthquake sequence. *Nature*, 456(7222):631–635.

Kulldorff, M. (2010). *SatScan User Guide for version 9.0.*

Kutoyants, Y. A. (1998). *Statistical Inference for Spatial Poisson Processes.* Springer.

Lederberg, J., Shope, R. E., and Stanley, C. O. (1992). Emerging infections: Microbial threats to health in the united states. *National Academy Press, Washington DC.*

Van Lieshout, M. N. M. (1995). *Stochastic geometry models in image analysis and spatial statistics.* CWI Tract.

Van Lieshout, M. N. M. (2000). *Markov Point processes and their applications.* Imperial College Press.

Van Lieshout, M. N. M. (2006). A J-function for marked point patterns. *Annals of the Institute of Statistical Mathematics*, 58:235–259.

Van Lieshout, M. N. M. (2011). A J-function for inhomogeneous point processes. *Statistica Neerlandica*, 65(2):183–201.

Van Lieshout, M. N. M. (2012). On estimation of the intensity function of a point process. *Methodology and Computing in Applied Probability*, 14(3):567–578.

Lotwick, H. and Silverman, B. (1982). Methods for analysing spatial processes of several types of points. *Journal of the Royal Statistical Society. Series B (Methodological)*, 44(3):406–413.

Marsan, D. and Lengliné, O. (2008). Extending Earthquakes' Reach Through Cascading. *Science*, 319(5866):1076–1079.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., and Teller, A. H. (1953). Equation of state calculations by fast computing machines. *Journal of Chemical Physics*, 21:1087–1092.

Møller, J. and Ghorbani, M. (2012). Aspects of second-order analysis of structured inhomogeneous spatio-temporal point processes. *Statist. Neerlandica*, 66:472–491.

Møller, J. and Helisová, K. (2010). Likelihood inference for unions of interacting discs. *Scandinavian Journal of Statistics*, 3:365–381.

Møller, J., Syversveen, A. R., and Waagepetersen, R. P. (1998). Log Gaussian Cox processes. *Scandinavian Journal of Statistics*, 25(3):451–482.

Møller, J. and Waagepetersen, R. (2004). *Statistical Inference and simulation for spatial point processes*. Boca Raton: CRC Press.

Moran, P. A. P. (1950a). Notes on continuous stochastic phenomena. *Biometrika*, 37:17–23.

Moran, P. A. P. (1950b). A test for the serial dependence of residuals. *Biometrika*, 37:178–181.

Myllymäki, M., Mrkvička, T., Grabarnik, P., Seijo, H., and Hahn, U. (2016). Global envelope tests for spatial processes. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*.

Ogata, Y. (1998). Space-time point-process models for earthquake occurrences. *Annals of the Institute of Statistical Mathematics*, 50:379–402.

Ogata, Y. and Tanemura, M. (1984). Likelihood analysis of spatial point patterns. *Journal of the Royal Statistical Society. Series B (Methodological)*, 46(3):496–518.

Peña Rey, I., De Aragón, M. V. M., Hueso, A. V., Arellano, M. T., Cabero, E. A., and Rodríguez, B. S. (2009). Epidemiology of varicella in Spain pre-and post-vaccination periods. *Revista española de salud pública*, 83(5):711–724.

Picard, N., Bar-Hen, A., Mortier, F., and Chadœuf, J. (2009). The multi-scale marked area-interaction point process: a model for the spatial pattern of trees. *Scandinavian Journal of Statistics*, 36(1):23–41.

Preston, C. J. (1977). Spatial birth-and-death processes. *Bulletin of the International Statistical Institute*, 46:371–391.

Quah, D. (1993). Galtons fallacy and tests of the convergence hypothesis. *The Scandinavian Journal of Economics*, 95(4):427–443.

R Core Team (2014). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Renshaw, E., Comas, C., and Mateu, J. (2009). Analysis of forest thinning strategies through the development of space-time growth-interaction simulation models. *Stochastic Environmental Research and Risk Assessment*, 23(3):275–288.

Rey, S. J. (1993). Spatial empirics for economic growth and convergence. *Geographical Analysis*, 33(3):195–214.

Rey, S. J. (2014). Spatial dynamics and space-time data analysis. In Fischer, M. M. and Nijkamp, P., editors, *Handbook of Regional Science*, pages 1365–1383. Springer Berlin Heidelberg.

Rey, S. J., Murray, A. T., and Anselin, L. (2011). Visualizing regional income distribution dynamics. *Letters in Spatial and Resource Sciences*, 4(1):81–90.

Ripley, B. (1976). The second-order analysis of stationary point processes. *Journal of Applied Probability*, (13):255–266.

Ripley, B. D. (1977). Modelling spatial patterns (with discussion). *Journal of the Royal Statistical Society. Series B (Methodological)*, 39(2):172–212.

Ripley, B. D. (1988). *Statistical inference for spatial processes.* Cambridge University Press, U.K.

Ripley, B. D. (1989). *Spatial Statistics: Past, Present and Future*, chapter Gibbsian interaction models, pages 1–19. Image, New York.

Ripley, B. D. and Kelly, F. P. (1977). Markov point processes. *Journal London Mathematical Society*, 15:188–192.

Rue, H. and Held, L. (2005). *Gaussian Markov Random Fields.* Chapman and Hall/CRC.

Rue, H., Martino, S., and Chopin, N. (2009). Approximate bayesian inference for latent gaussian models using inte- grated nested laplace approximations (with discussion). *Journal of the Royal Statistical Society, Series B*, 71(2):319–392.

Ruelle, D. (1969). *Statistical Mechanics: Rigorous Results.* W.A. Benjamin, Reading, Mass.

Schlather, M., Malinowski, A., Menck, P. J., Oesting, M., and Strokorb, K. (2015). Analysis, simulation and prediction of multivariate random fields with package RandomFields. *Journal of Statistical Software*, 63(8):1–25.

Schoenberg, F. P., Brillinger, D. R., and Guttorp, P. (2002). *Point Processes, Spatial-Temporal*, volume 3, pages 1573–1577. John Wiley & Sons, Ltd, Chichester, UK.

Schrödle, B. and Held, L. (2011a). A primer on disease mapping and ecological regression using INLA. *Computational Statistics*, 26(2):241–258.

Schrödle, B. and Held, L. (2011b). Spatio-temporal disease mapping using INLA. *Environmetrics*, 22(6):725–734.

Schrödle, B., Held, L., Riebler, A., and Danuser, J. (2011). Using integrated nested Laplace approximations for the evaluation of veterinary surveillance data from Switzerland: A case-study. *Journal of the Royal Statistical Society. Series C: Applied Statistics*, 60(2):261–279.

Servicio de Vigilancia y Control Epidemiológico (2007). *Protocolos EDO.* Generalitat Valenciana. Conselleria de Sanitat. Dirección General de Salud Pública.

Silverman, B. W. (1986). *Density Estimation for Statistics and Data Analysis.* Chapman & Hall, London.

Snow, J. (1855). *On the Mode of Communication of Cholera.* London: John Churchill, New Burlington Street, England.

Snyder, J. P. (1987). *Map Projections – A Working Manual.* United States Government Printing Office, Washington, D.C.

Soetaert, K. (2016). *plot3D: Plotting Multi-Dimensional Data.* R package version 1.1.

Stallybrass, C. O. (1931). *The principles of epidemiology and the process of infection.* G Routledge and Son Ltd, London, United Kingdom.

Statistics Office (2013). Municipal register of inhabitants at 1st of January. Valencia Local Council.

Steenbeek, A. G., Lieshout, M. N. M., Stoica, R. S., with contributions from Gregori, P., Berthelsen, K. K., and Iftimi, A. (2016). *MPPBLIB, a C++ library for marked point processes.*

Stoyan, D. and Stoyan, H. (2000). Improving ratio estimators of second order point process characteristics. *Scandinavian Journal of Statistics*, 27:641–656.

Strauss, D. J. (1975). A model for clustering. *Biometrika*, 63:467–475.

Ugarte, M., Goicoa, T., Ibáñez, B., and Militino, A. (2009). Evaluating the performance of spatio-temporal bayesian models in disease mapping. *Environmetrics*, 20:647–665.

Uria-Diez, J., Ibáñez, R., and Mateu, J. (2013). Importance of habitat heterogeneity and biotic processes in the spatial distribution of a riparian herb (Carex remota L.): a point process approach. *Stochastic Environmental Research and Risk Assessment*, 27(1):59–76.

USGS (2012). Earthquake hazards program. earthquake glossary.

Vere-Jones, D. (2009). Some models and procedures for space-time point processes. *Environmental and Ecological Statistics*, 16(2):173–195.

Vigny, C., Simons, W. J. F., Abu, S., Bamphenyu, R., Satirapod, C., Choosakul, N., Subarya, C., Socquet, A., Omar, K., Abidin, H. Z., and Ambrosius, B. A. C. (2005). Insight into the 2004 Sumatra-Andaman earthquake from GPS measurements in southeast Asia. *Nature*, 436(7048):201–206.

Walker, K. T., Ishii, M., and Shearer, M. (2005). Rupture details of the 28 March 2005 Sumatra Mw 8.6 earthquake imaged with teleseismic P waves. *Geophysical Research Letters*, 32(L24303).

WHO (2008). The Immunological Basis for Immunization Series. Module 10: Varicella-zoster virus. WHO Press.

WHO (2014). Weekly epidemiological record. Varicella and herpes zoster vaccines: WHO position paper, June 2014.

Widom, B. and Rowlinson, J. S. (1970). New model for the study of liquid-vapor phase transitions. *The Journal of Chemical Physics*, 52:1670–1684.

Resumen

Los primeros intentos de introducir e interpretar datos espaciales aparecieron en forma de mapas de datos. Halley (1686) superpuso en un mapa las ubicaciones y direcciones de los alisios y monzones situados alrededor de los trópicos, y les asignó una explicación física. Más adelante, Snow (1855) dibujó un mapa de Londres, mostrando los casos del brote de cólera de la calle Broad de 1854. Se considera que este es el primer uso de mapas que se hizo con fines de una análisis espacial de datos.

La metodología de la estadística espacial no fue introducida hasta mucho más tarde. Gosset (1907), conocido por su seudónimo Student, estaba interesado en el estudio de la distribución de las partículas en un líquido y, en lugar de analizar las posiciones de las partículas, agregó los datos en pequeñas áreas. En los años 1920 y 1930, R. A. Fisher estableció los principios del método de réplica, muestreo aleatorio y control de bloques. En el campo agrícola se utilizaron algoritmos de vecinos más cercanos para analizar diferentes estudios donde se tuvo en cuenta la dependencia espacial. Actualmente, los métodos estadísticos modernos se aplican en una amplia gama de campos de investigación como cartografía, geología, ecología, biología (estudios botánicos de distribución de plantas, estudios de biogeografía), epidemiología (mapas de enfermedades), econometría espacial, teledetección (geometría computacional), astronomía (estudios de localización de galaxias) y muchos otras.

Antes de proceder a una descripción de los datos espaciales, deben hacerse algunos comentarios sobre la calidad de los datos y el impacto que tiene en el análisis estadístico. Para llevar a cabo un estudio, es importante tener una base de datos completa y de alta calidad. A veces podemos encontrarnos con errores tipográficos, de codificación, datos incompletos o duplicados. Por lo tanto, es crucial que, antes de comenzar cualquier tipo de análisis estadístico, se lleve a cabo siempre un proceso de validación de datos para identificar posibles errores y, preferentemente, corregirlos. Los investigadores también tienen que tener en cuenta el derecho a la privacidad cuando se recogen y procesan datos de carácter personal. Los estudios que muestran resultados desfavorables, por ejemplo, áreas con altos niveles de contaminación atmosférica o ruidosa, grupos con alta incidencia de una enfermedad, pueden causar descontento entre la población, razón por la cual muchas instituciones no permiten el acceso a los datos argumentando supuestas políticas de protección. Para los investigadores la disponibilidad de datos es muy importante, por lo que se requiere la confianza y el apoyo de la población.

Los *datos espaciales* pueden ser de naturaleza continua o discreta. Puede agruparse en regiones pequeñas, o pueden representarse como observaciones individuales. Las ubicaciones espaciales pueden estar distribuidas de manera regular o irregular en una región. De acuerdo con ello, Cressie (1991) propuso una clasificación de los datos espaciales en tres principales grupos: *datos geoestadísticos*, *datos agregados* y *patrones puntuales*. Cuando un fenómeno se observa como mediciones en un conjunto finito de localizaciones fijas predeterminadas, el resultado son datos geoestadísticos. Cuando la región espacial (regular o irregular) se divide en un número finito de áreas, con límites bien definidos, y los datos se obtienen como observaciones agregadas en cada una de las regiones definidas, obtenemos datos agregados. Y finalmente, cuando el patrón espacial se obtiene observando las localizaciones aleatorias de un fenómeno, estamos en presencia de un patrón puntual.

La geoestadística fue desarrollada inicialmente para predecir la reserva mineral utilizando un número limitado de ubicaciones en una región. Así, la ingeniería minera ha sido uno de los primeros campos en los que se utilizó ampliamente. La clave fundamental en el modelado de las relaciones espaciales en este tipo de fenómenos es *el variograma*, que se estima y modeliza para describir adecuadamente el fenómeno observado. El objetivo principal en la aplicación de la geoestadística es generalmente predecir en un punto o conjunto de puntos de la región observada en la que no se han hecho mediciones. *Kriging* es una técnica de la geoestadística utilizada para interpolar el valor de un campo aleatorio en una localización en donde no hay observaciones, utilizando la información de observaciones cercanas, y es una de las técnicas más utilizadas en la predicción espacial. Los métodos geoestadísticos se utilizan actualmente en diversas disciplinas, incluyendo geología del petróleo, hidrogeología, hidrología, meteorología, oceanografía, geoquímica, geometalurgia, geografía, silvicultura, control ambiental, ecología del paisaje, ciencias del suelo, agricultura y muchas otras.

Los datos agregados se observan cuando la región espacial se divide en áreas pequeñas. Los modelos estadísticos para este tipo de datos usualmente incorporan la relación entre observaciones vecinas. Uno de los objetivos del análisis de los datos agregados es obtener una estimación de alguna medida de interés en cada área. También puede ser de interés para analizar la posible estructura espacial en los datos (agrupación, repulsión). Un ejemplo de datos agregados son los datos recopilados por satélites de teledetección, que proporcionan una herramienta muy útil de recolección de datos. Por ejemplo, la información meteorológica, los patrones climáticos, la distribución de minerales o tipos de suelos pueden ser observados usando satélites, sin necesidad de muestreo de campo. La superficie de la Tierra se divide en pequeños rectángulos llamados píxeles a los que se les asignan un valor de una característica dada, de modo que los datos se reciben en forma de una red regular y cada píxel se identifica utilizando su centro.

Los patrones puntuales se refieren a fenómenos que se producen de forma continua y aleatoria en diferentes lugares de una región. Un patrón de puntos es la colección de ubicaciones de tales eventos. Los procesos puntuales son los mecanismos teóricos que permiten modelar el desarrollo de los fenómenos. Uno de los objetivos de los estudios de patrones puntuales es ver si un patrón espacial observado tiene un cierto tipo de estructura o, por el contrario, si es simplemente el resultado de un proceso puntual (homogéneo) que actúa independientemente en el espacio. Si consideramos las localizaciones de los pinos de hoja larga en un bosque, podemos plantear varias preguntas. ¿Cuál es el significado biológico de la agrupación de estos árboles? ¿Se observa una agrupación de los árboles altos? ¿Los árboles altos interactúan con árboles pequeños? El análisis de patrones puntuales tiene como objetivo estimar parámetros asociados con modelos específicos. El análisis formal de un patrón puntual requiere el uso de múltiples técnicas estadísticas. Las funciones de resumen de primer y segundo orden son herramientas prácticas y útiles para describir y analizar eficazmente la estructura espacial. Diggle (2014) (Capítulo 7) introduce una gran variedad de conjuntos de datos y los posibles métodos para analizarlos.

Cualquier fenómeno puede localizarse utilizando una referencia geográfica. Las enfermedades, las muertes, los nacimientos, la exposición al riesgo u otros eventos, pueden asociarse al lugar donde ocurren, que puede ser coordenadas, direcciones o áreas. Añadir un componente temporal proporciona una mejor visión de la comprensión del proceso estadístico. La primera ley de geografía de Waldo Tobler, *todas está relacionado con todo lo demás, pero las cosas cercanas están más relacionadas entre si*, nos da una motivación para incluir la relación de vecindad en cualquier análisis espacial. La adición del componente temporal transforma el análisis en una tarea dinámica e integradora, donde la información representa más que sólo números o características estáticas.

Los métodos estadísticos espacio-temporales se están utilizando cada vez más en un intento de explicar el comportamiento de esta amplia gama de eventos. La principal razón para usarlos es reducir la variabilidad inicial del riesgo mediante el uso e intercambio de información entre localizaciones cercanas y momentos temporales próximos.

Los *datos espacio-temporales* se pueden presentar usando mapas para cada período de tiempo que ayudan a detectar áreas de alto riesgo o áreas con actividad inusual. La epidemiología espacio-temporal, por ejemplo, permite identificar los factores de riesgo individuales y agregados para la salud, convirtiéndose en una herramienta indispensable para la toma de decisiones en salud pública. También permite evaluar dinámicamente los factores de riesgo, analizar su impacto en la población y estimar los potenciales beneficios de las medidas preventivas para la salud pública.

Cabe señalar que en el contexto del análisis estadístico espacio-temporal, $2 + 1$ no es igual a 3, debido a que la dimensión temporal es fundamentalmente diferente de las dimensiones espaciales. La mayoría de los procesos espaciales en la naturaleza son sólo instantáneas de la evolución de los procesos espacio-temporales, pero usar únicamente métodos de análisis espacio-temporales sería un error. El uso de estas técnicas debe hacerse cuando nuestro interés involucra a ambos componentes, espacial y temporal, y no debe llevarse a cabo utilizando únicamente análisis estadísticos de los dos componentes por separado.

Muchos escenarios de la vida real como los terremotos, incidentes de enfermedades o incendios, dan lugar a recopilaciones de datos en donde cada dato, además de tener una localización espacial y una ocurrencia en el tiempo, también lleva otra información útil sobre el evento en cuestión. En el lenguaje de los procesos puntuales, esta información 'extra' se conoce como la *marca* del suceso. Para ejemplificar, en el caso de los terremotos una marca podría corresponder a la magnitud del mismo. Cuando se asocia una marca a un dato espacio-temporal de esta manera, el mecanismo aleatorio que supuestamente generó la colección total de datos se denomina como un *proceso puntual marcado espacio-temporal (MSTPP)*, y los datos correspondientes se denominan patrón puntual marcado espacio-temporal ((Daley and Vere-Jones, 2003; Diggle, 2014; Vere-Jones, 2009). Otras aplicaciones de MSTPP incluyen, entre muchos otros, incidentes de ciertas enfermedades, crímenes, incendios.

Una estructura general de la tesis y sus contribuciones es la siguiente:

El capítulo 2 proporciona un análisis secuencial de los datos espaciales. Comienza con un análisis descriptivo contrastando la aleatoriedad espacial completa y la inhomogeneidad, y continúa con un análisis descriptivos con covariables. Este capítulo presenta una técnica para construir nuevos modelos de Gibbs para patrones puntuales espaciales, propuesta por Baddeley et al. (2013). En este capítulo se analiza el patrón puntual de todos los casos de varicela registrados durante 2013 en Valencia, España. El análisis descriptivo se utiliza para obtener una visión de las propiedades básicas del patrón de puntos. Se hace también uso de la información proporcionada por covariables, como la densidad de la población (niños menores de 14 años) que vive en el área de estudio, la distancia a la escuela más cercana y la composición de las familias (expresada como el promedio de personas por familia), para describir la intensidad del proceso. El software `SatScan` (Kulldorff, 2010) se utiliza para identificar los principales clusters de escuelas. Esta información se introduce posteriormente en el modelo. Se presenta una comparación entre diferentes modelos, así como métodos de diagnóstico para elegir el mejor modelo para ajustar los datos. Esta primera contribución de la tesis es la aplicación del análisis estadístico para los patrones puntuales espaciales a la epidemiología. Como se mencionó anteriormente, este capítulo presenta un análisis espacial para el conjunto de datos de la varicela a través del uso de híbridos de los modelos de Gibbs. Los modelos presentados en este capítulo consiguen describir interacciones a diferentes escalas y también la inhomogeneidad espacial considerando la información de las covariables mencionadas. Se trata de una contribución menor pero interesante a la literatura de los procesos puntuales.

Una segunda contribución atañe al campo de los datos espacio-temporales en retículo. El capítulo 3 presenta un segundo análisis del conjunto de datos de varicela en un formato de datos agregados. En la literatura se han propuesto modelos paramétricos y no paramétricos con diferentes tipos de interacción para analizar datos espacio-temporales. Bernardinelli et al. (1995) propone un modelo paramétrico que asume una tendencia lineal. El artículo de Knorr-Held (2000) sugiere un modelo que combina el modelo espacial definido por Besag et al. (1991) con modelos dinámicos en los que se asume que las tendencias temporales son no lineales y tampoco estacionarias. En particular, los efectos temporales pueden ser vistos como el análogo temporal de los componentes espaciales estructurados en el modelo de Besag et al. (1991). Este modelo permite interacciones espacio-temporales donde cuatro tipos de interacciones surgen naturalmente como el producto de uno de los dos efectos espaciales con uno de los dos efectos temporales. Los métodos de Markov Chain Monte Carlo (MCMC) han dado a la comunidad de usuarios una gran herramienta para analizar estos tipos de modelos para datos complejos. Recientemente se ha desarrollado un método de aproximación para la inferencia bayesiana (INLA) (Rue et al., 2009), que representa una herramienta muy útil al reducir el tiempo de cómputo y presentar una interfaz muy fácil de usar por medio del paquete R-INLA.

Una amplia gama de aplicaciones han implementado modelos paramétricos y no paramétricos para detectar el comportamiento espacio-temporal de los patrones. Bernardinelli et al. (1995) ilustran el uso de modelos paramétricos mediante el análisis de la prevalencia acumulada de la diabetes mellitus insulino-dependiente en los datos recogidos a todos los reclutas militares de 18 años de edad en Cerdeña durante 1936-1971. Ugarte et al. (2009) presenta una evaluación de los modelos bayesianos espacio-temporales paramétricos y no paramétricos con un estudio de simulación y una aplicación para datos de mortalidad por cáncer colorrectal en varones de Navarra, correspondientes a cuatro ciclos de 5 años. Schrödle and Held (2011a) aplican modelos espacio-temporales utilizando INLA a casos de coxiellosis en vacas suizas de 2005 a 2008. Diferentes extensiones de los modelos paramétricos y no paramétricos han sido propuestas por Schrödle and Held (2011b) y aplicados a los recuentos de salmonelosis en el ganado bovino de Suiza, desde 1991 hasta 2008. Schrödle et al. (2011) proponen otra aplicación de estos modelos para analizar los casos de diarrea viral bovina en ganado suizo de 2003 a 2007. Blangiardo et al. (2013) estudian el patrón de nacimientos con bajo peso para los 159 condados del estado de Georgia de los EE.UU. de 2000 a 2010, utilizando formulación paramétrica (Bernardinelli et al., 1995) y formulación no paramétrica en (Knorr-Held, 2000) con INLA.

Las técnicas descritas en el capítulo se aplican al estudio del comportamiento espacial y temporal de los datos de varicela. El estudio de ha dividido en dos partes, un análisis exploratorio de datos espacio-tiempo (ESTDA) y una aplicación de modelos espacio-temporales. Para la primera parte del estudio, el análisis exploratorio espacio-temporal, seguimos la propuesta de Rey (2014). En la segunda parte se propone una metodología estadística con modelos paramétricos y no paramétricos espacio-temporales para datos agregados utilizando un marco bayesiano. Este capítulo representa un compendio interesante de técnicas de análisis espacio-temporales aplicadas para resolver un problema epidemiológico específico.

Los patrones espacio-temporales se observan cada vez más en muchos campos diferentes, como por ejemplo la ecología, la epidemiología, la sismología, la astronomía y la silvicultura. El aspecto común es que todos los sucesos observados tienen dos características básicas: la localización y el momento en el que se registra el evento. En el capítulo 4 nos interesa principalmente la epidemiología, que estudia la distribución, las causas y el control de las enfermedades en una población humana definida. Las localizaciones de los casos dan información sobre el comportamiento espacial de la enfermedad, mientras que los tiempos, medidos en diferentes escalas (días, semanas, años, período de tiempo), dan una idea de la respuesta temporal del proceso global. Un punto esencial a tener en cuenta es que las personas no están uniformemente distribuidas en el espacio. La distribución espacial de la población en riesgo es crucial cuando se analizan patrones espacio-temporales de enfermedades. Modelos realistas que se ajusten a los datos epidemiológicos deben incorporar la inhomogeneidad espacio-temporal y permitir diferentes tipos de dependencia entre los puntos. Una clase importante de modelos es la familia de procesos puntuales de Gibbs, definida por su función de densidad de probabilidad (Van Lieshout, 2000; Ripley, 1988, 1989), y en particular, la subclase de procesos de interacción por pares. Ejemplos bien conocidos de procesos de interacción por pares son el modelo de Strauss (Kelly and Ripley, 1990; Strauss, 1975) o el proceso hard-core, un caso particular del modelo de Strauss en el que ningún punto se aproxima a otro más allá de un umbral dado. Sin embargo, los modelos de interacción por pares no siempre son una opción adecuada para ajustar los patrones agrupados. Una familia de procesos puntuales de Markov que pueden analisar patrones puntuales es la de los modelos área-interacción (Baddeley and van Lieshout, 1995; Kendall et al., 1999). Los métodos para la inferencia y la simulación perfecta están disponibles en Dereudre et al. (2014); Häggström et al. (1999); Kendall (1998); Møller and Helisová (2010). La mayoría de los procesos naturales exhiben interacción a múltiples escalas. Los procesos clásicos de Gibbs modelan la interacción espacial en una sola escala, sin embargo las generalizaciones *multi-escala* han sido propuestas en la literatura (Ambler and Silverman, 2010; Gregori et al., 2003; Picard et al., 2009).

El capítulo 4 propone una nueva extensión del modelo área-interacción para procesos puntuales (Baddeley and van Lieshout, 1995) que permite la interacción a multi-escalas en un marco espacio-temporal, permitiendo también la inclusión de covariables. El contenido del capítulo representa una contribución metodológica importante en la que definimos y damos las propiedades de Markov de nuestro modelo. A continuación, proponemos una modificación de los algoritmos de simulación de Metropolis-Hastings y birth-and-death. Para la inferencia utilizamos el método de la pseudo-verosimilitud y adaptamos el procedimiento de Berman-Turner (Baddeley and Turner, 2000) a nuestro contexto. La última parte del capítulo aplica el nuevo modelo propuesto al conjunto de datos de varicela.

Una segunda contribución metodológica se presenta en el capítulo 5. Este capítulo tiene como objetivo desarrollar herramientas de proceso puntuales que nos permitan realizar los llamados análisis no paramétricos de segundo orden de patrones puntuales espacio-temporales marcados. Para analizar la interacción en este tipo de procesos se introduce medidas de momentos reducidos de segundo orden marcados y K- funciones.

Clásicamente, al analizar los STPPs (marcados), el análisis se ha basado en *funciones de intensidad condicional (CIs)* (véase, por ejemplo, (Choi and Hall, 1999; Daley and Vere-Jones, 2003; Marsan and Lengliné, 2008; Ogata, 1998; Schoenberg et al., 2002)). En principio, una función de intensidad condicional nos da el número esperado de sucesos adicionales en el período infinitesimal próximo, dada la historia de los eventos hasta ese momento. El atractivo de las ICs es que, cuando existen, especifican toda la distribución del MSTPP. Sin embargo, como se ha señalado por ejemplo en Diggle (2014), no todos los modelos MSTPP tienen CIs disponibles/manipulables. Además, gran parte del análisis basado en CI se lleva a cabo dentro del marco de una clase dada de modelos.

Recordando que queremos definir un análisis general no paramétrico, procederemos con un enfoque no basado en CI, siguiendo así una formulación conjunto aleatorio/medida aleatoria (ver, por ejemplo, (Chiu et al., 2013; Daley and Vere-Jones, 2003; Diggle, 2014; Van Lieshout, 2000; Møller and Waagepetersen, 2004)). En este contexto, cuando se analizan patrones de puntos espacio-temporales marcados, lo primero que se hace es intentar explicar dónde y cuándo ocurrieron sucesos de una categoría de marca dada. Puesto que donde y cuándo es una propiedad univariante, en el sentido de que no estamos tratando explícitamente con las posibles dependencias entre los puntos, estamos tratando de analizar la *intensidad*. Antes de proceder a proponer modelos específicos para la estructura de intensidad, a través del patrón de puntos observados, se comienza habitualmente obteniendo una estimación no paramétrica de la función de intensidad (véase, por ejemplo, (Diggle, 2014)). La función de intensidad, en esencia, refleja la probabilidad infinitesimal de encontrar un punto del MSTPP en una ubicación espacial dada, en un momento dado, con una marca dada. Obsérvese que es diferente de la intensidad condicional antes mencionada, que se define como una equivalencia condicional. En el caso más sencillo simplemente asumiríamos la homogeneidad, es decir, que es igualmente probable que observe un evento, con cualquier marca, en cualquier posición espacio-temporal. Esto no es, sin embargo, un supuesto realista, así que procederemos asumiendo inhomogeneidad.

Aunque el candidato más natural para este tipo de estimación no paramétrica es la estimación kernel, como nuestro objetivo es aplicar el método al estudio y análisis de los terremotos ocurridos en del mar de Andamán (región de Sumatra) en el periodo 2004-2008, y debido a los abruptos cambios en la actividad que los terremotos muestran tanto espacial como temporalmente, hemos optado por de considerar un enfoque adaptativo, es decir, un enfoque de estimación de la intensidad de Voronoi (véase, por ejemplo, (Barr and Schoenberg, 2010)). Habiendo obtenido una estimación no paramétrica de la función de intensidad, de modo que tenemos una descripción de las propiedades univariantes, podemos proceder al estudio de la estructura de dependencia inherente del mecanismo generador de datos, es decir, el MSTPP subyacente. En el capítulo 5 nos centramos en los estadísticos resumen de segundo orden, ignorando así los estadísticos resumen de orden superior, como por ejemplo la J- función espacio-temporal y sus componentes (Cronie and van Lieshout, 2015) y la J- funcion marcada y sus componentes (Cronie and van Lieshout, 2016; Van Lieshout, 2006).

Nuestra contribución combina las ideas de Gabriel and Diggle (2009) con las de Cronie and van Lieshout (2016) para definir una K- función, $K^{CD}_{rminhom}(r,t)$ para los MSTPPs no homogéneos. En términos generales, la $K^{CD}_{\mathrm{inhom}}(r,t)$ describe la interacción, en el sentido de (Gabriel and Diggle, 2009), entre puntos pertenecientes al conjunto de marcas C y los puntos pertenecientes al conjunto de marcas D, para un MSTPP no homogéneo. Obsérvese que para todos los estadísticos de resumen anteriores, uno de los principales objetivos ha sido considerar su estimación no paramétrica. Una parte significativa del capítulo 5 está dedicado a la estimación no paramétrica. Estos estadísticos de resumen, que nos permiten cuantificar la dependencia entre diferentes marcas-categorías de los puntos, dependen del espacio de las marcas y de la medida marcada de referencia elegida. Se propone un nuevo test para el marcado independiente y se derivan los estimadores de minus-sampling para todas las estadísticas consideradas. Además, estudiamos los estimadores Voronoi de la intensidad de los MSTPPs. Estos nuevos estadísticos se emplean finalmente para analizar el conjunto de datos de los terremotos antes mencionados. Encontramos que el fenómeno de agrupamiento se observa entre los terremotos principales y terremotos anteriores/posteriores en prácticamente todas las escalas de espacio y tiempo. Además, encontramos evidencia de que, condicionalmente en las localizaciones espacio-temporales de los terremotos, las magnitudes no se comportan como una secuencia independiente e idénticamente distribuida.

El trabajo presentado representa un punto de partida dé lo que podrían ser futuras líneas de investigación, algunas de las cuales describimos a continuación.

Como ya hemos mencionado, la varicela es una enfermedad altamente infecciosa. Estamos seguros de que, además del efecto de la población, existen otras covariables que pueden influir en el comportamiento espacio-temporal de la enfermedad. Por lo tanto, una meta importante para un trabajo futuro es considerar la inclusión de covariables que pueden mejorar nuestros análisis. Por ejemplo, WHO (2014) sugiere que existen algunos factores climáticos que pueden influir en la epidemiología de la varicela. Así, covariables tales como las temperaturas medias mensuales, los niveles medios semanales de lluvia, las horas medias de sol u otras relacionadas con el clima, pueden proporcionar información útil para el análisis de la varicela. Además, información adicional sobre los ingresos per cápita u otras covariables socioeconómicas podrían mejorar nuestro análisis.

Otro factor interesante que podría influir en la propagación de la enfermedad en la ciudad de Valencia es la ubicación del río Turia, que la cruza de Oeste a Este y la divide en dos partes de extensión parecida. Los epidemiólogos creen que cualquier barrera geográfica natural marca algunas diferencias, ya que los niños que viven en un lado del río van a escuelas en ese lado del río y los que viven en el lado opuesto asisten a las escuelas de ese lado. El papel de las escuelas y guarderías en la transmisión de la varicela es crucial. Teniendo en cuenta que hoy en día la mayoría de las parejas sólo tienen uno o dos hijos, la transmisión dentro de la familia probablemente tiene menos influencia que la transmisión escolar.

Para un trabajo futuro, sería interesante aplicar el modelo presentado en el capítulo 4 a otras enfermedades que pueden exhibir interacción en varias escalas en el espacio y el tiempo. También sería muy interesante aplicar el modelo a datos que no están necesariamente relacionados con la epidemiología. Los patrones de terremotos, por ejemplo, tienden a mostrar agregación, pero también inhibición a diferentes escalas. De hecho, creemos que el modelo propuesto puede ser aplicado en una amplia gama de campos de investigación, tales como la silvicultura, la geología y la sociología.

Otras aplicaciones directas de la metodología presentada en el capítulo 5 se pueden encontrar en, por ejemplo, epidemiología y criminología. Actualmente estamos estudiando conjuntos de datos relacionados con estos campos. En particular, nuestro conjunto de datos de casos de varicela. Además estamos analizando los datos sobre delitos en Valencia, España. Téngase en cuenta que para ambas aplicaciones puede ser más relevante considerar las versiones multivariantes de los estadísticos de resumen.

Actualmente estamos trabajando también en extender las ideas presentadas en el capítulo 5 a procesos puntuales espacio-temporales marcados en redes lineales (Ang et al., 2012). Además, buscamos combinar los resultados de Cronie and van Lieshout

(2015, 2016) de la misma manera que se hizo en el Capítulo 5, para obtener versiones espaciales y temporales no homogéneas de la función distribución de la distancia al vecino más próximo, la función empty-space y la J- función. Creemos que estas pueden ser herramientas más poderosas para el propósito de análisis estadístico no paramétrico.