Spatial and Spatiotemporal Point Process Modelling in Epidemiology

A thesis presented in partial fulfilment of the requirements for the degree of

Doctor of Philosophy

in

Statistics

at Massey University, Manawatu
New Zealand

Tilman Marcus Davies

December, 2011

Abstract

Geographical epidemiology encapsulates those problems that aim to understand space and/or space time trends in the disease(s) of interest, a goal clearly important in both public health and economic contexts. The increasing availability of point-location, coordinate data for studies in geographical epidemiology calls for a greater scrutiny of statistical point process theory with respect to these applications. Though the statistical analysis of planar point patterns is now theoretically well-established, there remain many aspects which warrant further research, especially from a practical perspective. The need becomes even greater when we consider the relative youth of methods for the analysis of spatiotemporal observations, where non-trivial variation and even dependence throughout the space-time continuum can exist.

This work aims to address these issues by careful review, theoretical refinement, empirical testing, and real-world analyses of certain statistical tools used in point process problems in geographical epidemiology. We scrutinise the kernel-smoothed density-ratio estimator of the so called *relative risk function*, a particularly flexible approach given the anticipated spatial heterogeneity of the observations over a given geographical region. This discussion introduces the adaptive (i.e. variable bandwidth) risk function, as well as novel asymptotic methods for computation of tolerance contours designed to identify sub-regions of statistically significant fluctuations in risk.

More sophisticated statistical methodology is warranted in certain situations, where it may be assumed that both 'global' heterogeneity and 'local' correlation drives the space and/or space-time disease dispersion. A comprehensive review of the stochastic *log-Gaussian Cox process*, in both purely spatial and spatiotemporal contexts, is conducted. A suite of novel numerical experiments investigate the performance of convenient, yet *ad hoc*, minimum contrast parameter estimation techniques for the dependence structure of the latent Gaussian process.

The computer code arising from the review and refinement of the above methodologies was instrumental in the release of two separate software packages. These are available in the R environment, and also showcased here. A number of additional collaborations with applied researchers around the world serve to further highlight the contributions made throughout the course of this research project and the importance of sound statistical methods in geographical epidemiology.

Acknowledgements

To my amazing family, both across the Tasman and up in Germany, for their unwavering and continuous support;

To my wonderful cat, Sigma, for always being happy to see me and for maintaining a consistent supply of fluff and semi-deceased birds;

To the statistics academics across Europe for their interest in my work and generous hospitality during my trip in early 2011;

To my external collaborators of various peer-reviewed works that took an interest in the contents of this thesis while still in production; this academic experience at such an early stage of my career was indeed invaluable, and of course extremely rewarding. Special thanks are reserved for M. Stevenson, EpiCentre, Massey University (New Zealand), B. Taylor, B. Rowlingson and P. Diggle, Division of Medicine, Lancaster University (U.K.), and Z. Zhang, Department of Epidemiology, Fudan University (People's Republic of China);

To my local co-supervisors, Associate Professor Geoff Jones and Doctor Jonathan C. Marshall, a massive thanks for never failing to provide that extra support and additional points of view absolutely invaluable to understanding and implementing novel and/or complex material;

And most importantly to my primary supervisor, Professor Martin L. Hazelton, without whom this work quite simply would not exist. Responsible not just for this project's supervision, but for inspiring me in undergraduate years to further pursue statistics in the first place, he ensured that the positive times far outweighed the negative. True intelligence is not just knowledge, but the ability to communicate, and teach, said knowledge. Martin is, without a doubt, the most intelligent academic I've ever met.

Contents

List of Figures

List of Tables

List of Publications

Accredited publications arising during the course of this research.

1. Taylor BM, **Davies TM**, Rowlingson BS, Diggle PJ (2012) **lgcp** - An R package for inference with spatial and spatiotemporal log-Gaussian Cox processes, *Journal of Statistical Software* [in press].

2. Zhang ZJ, Chen DM, Chen Y, **Davies TM**, Vaillancourt JP, Liu WB (2012) Risk signals of an influenza pandemic caused by highly pathogenic avian influenza subtype H5N1: Spatio-temporal perspectives, *The Veterinary Journal* **192** 3 417-421.

3. Zhang ZJ, **Davies TM**, Jiang QW (2012) Identification of high-risk regions for schistosomiasis in the Guichi region of China: an adaptive kernel density estimation-based approach, *Submitted for publication – Geospatial Health*.

4. **Davies TM**, Hazelton ML, Marshall JC (2011) **sparr**: Analyzing spatial relative risk using fixed and adaptive kernel density estimation in R, *Journal of Statistical Software* **39** 1 1-14.

5. Sanson RL, Harvey N, Garner MG, Stevenson MA, **Davies TM**, Hazelton ML, O'Connor J, Dubé C, Forde-Folle KN, Owen K (2011) Foot-and-mouth disease model verification and 'relative validation' through a formal model comparison, *Revue Scientifique et Technique-Office International des Epizooties* **30** 2 527-540.

6. **Davies TM**, Hazelton ML (2010) Adaptive kernel estimation of spatial relative risk, *Statistics in Medicine* **29** 23 2423-2437.

7. Hazelton ML, **Davies TM**, (2009) Inference based on kernel estimates of the relative risk function in geographical epidemiology, *Biometrical Journal* **51** 1 98-109 (**Pre-Ph.D.**).

List of Acronyms

Definitions of acronyms and abbreviations, and the section of their first occurrence in the thesis.

2D	Two-dimensional	1.2
3D	Three-dimensional	4.1.2
ADAPT	Adaptive-bandwidth relative risk function	2.5
AEGISS	Ascertainment and Enhancement of Gastrointestinal Infection Surveillance and Statistics	3.4.2
AMISE	Asymptotic mean integrated squared error	2.3.1
ARE	Absolute relative error	3.5.1
ASY	Asymptotic	2.4
BCV	Biased cross-validation	2.3.1
CRAN	Comprehensive R Archive Network	1.3
CSR	Complete spatial randomness	1.2.1
CV	Cross-validation	2.3.1
DFT	Discrete Fourier transform	3.3
DMCS	Deterministic Minimum Contrast Scenario	3.5.2
ETAS	Epidemic type aftershock sequence	5.4
EU	European Union	5.3.2
EW	Epidemic wave	5.2.2
FAO	Food and Agriculture Organisation of the United Nations	5.2.2
FFT	Fast-Fourier transform	3.3
FIXCV	Fixed-bandwidth relative risk function (CV bandwidth)	2.5
FIXOS	Fixed-bandwidth relative risk function (OS bandwidth)	2.5
FMD	Foot-and-mouth disease	5.3.1
GCMT	Global Centroid Moment Tensor	5.4
GPS	Global positioning system	5.1.2
iid	Independent and identically distributed	2.1.1
IQR	Interquartile range	2.3.1
ISE	Integrated squared error	2.3.1
KDE	Kernel density estimation	2.1
LGCP	Log-Gaussian Cox process	3.1
LOWESS	Locally weighted scatterplot smoothing	3.4.2
LSCV	Least-squares cross-validation	2.3.1
MALA	Metropolis-adjusted Langevin algorithm	3.3
MC	Monte-Carlo	2.4
MCMC	Markov-chain Monte Carlo	3.3.2
MCS	Minimum Contrast Scenario	3.5.1
MH	Metropolis-Hastings	3.3.2
MISE	Mean integrated squared error	2.3.1
NAADSM	North American Animal Disease Spread Model	5.3.1
netCDF	Network common data form	4.2.2
OIE	World Organisation for Animal Health	5.2.2
OS	Oversmoothing (maximal smoothing principle)	2.3.1
PBC	Primary biliary cirrhosis	2.6.2
PCF	Pair correlation function	1.2.3
PDF	Probability density function	2.1.1
RED	Relative Euclidian distance	3.5.2
SMR	Standardised mortality ratio	1.1
SNCP	Shot-noise Cox process	3.1.4
STS	SpatioTemporal Scenario	3.5.3
WISE	Weighted integrated squared error	2.5
WMISE	Weighted mean integrated squared error	2.3.1

Chapter 1

Introduction

The field of epidemiology is concerned with the study of disease incidence and prevalence within a given human or animal population. It is not only intuitively sensible, but often essential, to examine the spatial and temporal variation present in cases of a given disease in order to best understand its aetiology. Statisticians and epidemiologists are often close collaborators, as the complexities involved in the modelling of space-time data call for a reliance on sound statistical methods, knowledge of how to best construct and fit the models, and an understanding of the allowances and limitations associated with the methodology in use.

This chapter aims to give an overview of the area shared by statistics and geographical epidemiology. Section 1.1 presents some motivating examples and discusses the two data formats most commonly encountered in the applications of interest and issues associated with both. A general background on the statistical theory of point processes as required for the core chapters is provided in Section 1.2. The objectives and structure of the thesis itself are outlined in Section 1.3.

1.1 Data in geographical epidemiology

Geographical epidemiology involves the presence of data collected within finite geographical regions, with information on the time of observation contributing to a spatiotemporal (as opposed to purely spatial) analysis where available. Additional covariate information associated with individual records or specific subsets of the region or observation dates, may also be collected.

1.1.1 Illustrative examples

To motivate the subsequent thesis, we present here some interesting historical examples of epidemiological analyses where the statistical emphasis has been on the spatial or spatiotemporal aspects of the disease(s) of interest. Some hypothetical questions are also posed to further encourage this statistical viewpoint, and additional remarks are provided in Section 1.1.2.

Figure 1.1: Snow's recorded locations of cholera deaths (·) and water pumps (×) during the 1854 outbreak in London.

Snow's cholera

One of the first detailed academic examples in geographical epidemiology is that of John Snow's famous discovery of the source of an outbreak of cholera in mid-nineteenth century London (subsequently published in Snow, 1855). Snow, a physician skeptical of the then-held belief that poor air quality was to blame for such illnesses, proceeded to record the spatial locations of the infected individuals' homes. A version of his original map is shown in Figure 1.1.

With the benefit of hindsight, we may challenge any appearance of 'unusually large' groupings of cholera cases in this image unless we assume the London population is uniformly distributed over the area of interest. Would comparing a map of a sample of uninfected individuals shed any light on why the observations have assumed this spatial pattern?

In any case, Snow concluded that the source of the outbreak was a contaminated drinking water pump on Broad Street due to the concentration of cases in its immediate vicinity, which is indeed clearly visible on the map. Though conflicting with the previous notions concerning cholera transmission, the pump was decommissioned, effectively stopping the outbreak. John Snow is widely considered to be one of the founding fathers of epidemiology,

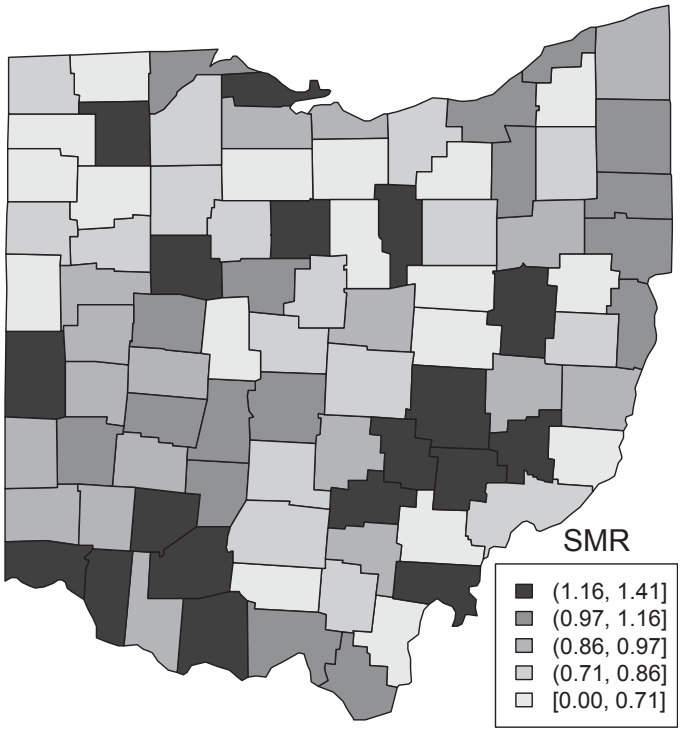


Figure 1.2: Standardised mortality ratio (SMR) of respiratory cancer by county in Ohio, USA, for the year 1985.

and a memorial to his work in the form of a Victorian pump with its handle removed sits at the same site today.

Ohio cancer mortality

This example comprises recorded deaths due to respiratory cancer in each of the 88 counties of the U.S. state of Ohio in 1985. Part of a far larger spatiotemporal data set spanning the 21 years beginning 1968 to end 1988, the mortality counts can also be broken down into groups specified by age, sex and race (a good description of these data as well as references to earlier analyses may be found in Lawson, 2006). Figure 1.2 shows the standardised mortality ratio (SMR) by county, given in each case by the quotient of the observed and expected number of mortalities.

Naturally, the SMR and resulting image (hence our interpretation of the variation in spatial mortality rates) is dependent on the determination of the expected counts. In our case, this has been straightforwardly determined as the overall mortality rate (i.e. total Ohio deaths divided by total Ohio population), subsequently multiplied by the at-risk population

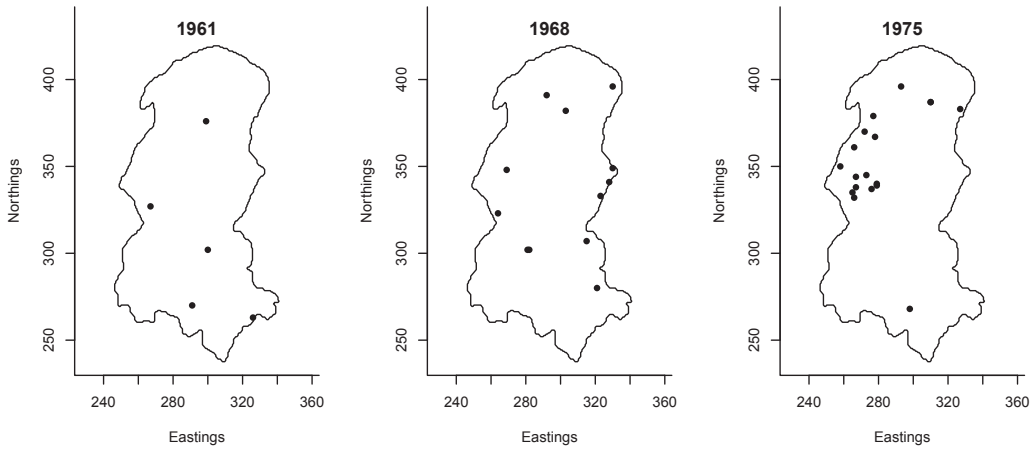


Figure 1.3: Spatial locations of Burkitt's lymphoma occurrences in the Western Nile district of Uganda, Africa, for the years 1961, 1968 and 1975.

counts within each county. Would accounting for risk factors associated with respiratory cancer or other covariates in estimation of the expected number of mortalities dramatically change our interpretation of the spatial risk of this disease over Ohio?

Burkitt's lymphoma in Uganda

Burkitt's lymphoma, so named after the researcher who first described the disease during the 1950s, is a particular type of cancer affecting the human lymphatic system. This data set consists of the recorded locations of cases of Burkitt's lymphoma in the Western Nile district of Uganda, Africa, over the years 1961-1975. It is a popular example in geographical epidemiology; see Williams et al. (1978) and references therein for the earliest analyses.

Figure 1.3 displays three particular years of the data set. There are a number of cautionary notes that should be made about the data (in its current form) prior to concluding any space-time trends. For example, though the total number of cases has increased over the three illustrated years from 5 (1961) to 12 (1968) to 15 (1975), does this indicate a true increase in prevalence or has reporting simply become more thorough? Do spatial groupings suggest actual heightened risk of contraction, or do they simply reflect the natural heterogeneity in the underlying population over the study region?

1.1.2 Preliminary discussion

As we have just seen, we may classify the types of data commonly encountered in geographical epidemiology into two broad categories. *Aggregated* or *areal* count data deals with information restricted to well-defined sub-regions of the geographical area under scrutiny, such as states in a country or postal zones/census tracts in a city. The (spatial) resolution of the data does not increase further than the level of these sub-regions. The second type of data frequently observed is typically labelled *point location* or *case-event* data. In contrast

to aggregated counts, point location data provides the specific coordinates of the disease cases within the overall study region.

There are general goals in the analysis of spatial/spatiotemporal disease data that are encountered in both situations. The foundation for many studies is an expression of *relative risk* of disease contraction over the study region of interest. This summary measure is derived based on the observed cases with respect to some distribution of the at-risk, uninfected population, often referred to as *controls*. As encountered in the Ohio cancer example, relative risk is most commonly computed as a morbidity or mortality ratio, whereby values greater than 1 would indicate an increase or excess in the estimated risk of infection within a given sub-region (aggregated counts) or specific location (case-events). Indeed, the purpose of Chapter 2 is to examine in detail and improve nonparametric estimation of relative risk surfaces for case-event data.

Understandably, there are both advantages and disadvantages associated with these two data formats. In human epidemiology, specific event locations are typically recorded as street addresses which have the potential to be misleading with respect to studying disease aetiology. For example, it is quite possible that infection occurred not at home, but at the workplace. Alternatively, the location at which the event was recorded could be different to the location of infection following a move. These difficulties notwithstanding, case-event data (if verified/validated by some means to a satisfactory extent) allows for a detailed mapping of the space/space-time disease trends and can be of great use in epidemiological problems. Statistical methods used to analyse this data format fall into the category of *point process modelling*.

Aggregated count data can avoid some of the problems associated with the uncertainty in 'exact' recorded locations, at the expense of other potential drawbacks. This kind of count data is usually far easier to obtain than specific case coordinates, and can indeed be the only feasible option for data acquisition if we are dealing with, for example, occurrence of a disease over an entire country. However, the discontinuity of estimated statistics, such as the relative risk, between the disjoint sub-regions is unnatural. How can we expect the rate of disease occurrence to change 'instantly' upon crossing from one sub-region to another? Nevertheless, statistical techniques that can create and are equipped to deal with aggregate data (an important part, for example, of *small area estimation* methodology – a good review can be found by Ghosh and Rao, 1994) can yield informative results about diseases spread over large regions, time spans and populations. Of course, counts in defined sub-regions can be obtained from case event data, another part of small area estimation. It is generally not recommended, however, to perform this style of 'binning' when we do have coordinate-specific data available due to the obvious loss of spatial information.

For up-to-date notes on these issues, as well as discussions on some existing approaches to spatial data collection, small area estimation, and analysis from an applied perspective, the reader is referred to texts such as Waller and Gotway (2004) and Lawson (2006).

This work focuses purely on case-event situations where the data are observed as a point pattern within a geographical region. The benefits associated with the increase in detail have in turn led to an increased availability of these kinds of datasets as more resources are geared towards this type of collection. Though now theoretically well-established, applied

statistical modelling of spatial and in particular spatiotemporal point patterns remains a relatively young field, previously hindered by the lack of suitable methodology as well as the often high computational requirements. It is therefore important that the methodology and available practical tools continue to evolve and improve, so that epidemiologists and other researchers may gain the most from their data.

1.2 Point processes

The term *point process* is used in reference to the underlying stochastic mechanism, function, or phenomenon that has generated the finite (i.e. countable) point pattern we observe. The assumptions we make about the process have an important impact on how we model the problem at hand as well as interpretation of the results. It is therefore necessary that we cover some fundamental aspects of spatial point process theory in order to be best equipped to tackle the kind of case-event applications we are likely to encounter in geographical epidemiology.

This section offers an overview of terms and methods for spatial point processes, to the extent necessary for the digestion of the remainder of the thesis. Though we focus on a 'purely spatial' discussion here, the ideas extend naturally to space-time data and this will be clarified as required in Chapter 3. The vast majority of spatial and spatiotemporal point process problems in epidemiology deal only with 'x-y' planar coordinates and where necessary we restrict attention to two-dimensional (2D) data. That is, although we may theoretically work in a general dimension d, we will concentrate on the case where $d = 2$. It is also worth noting that some authors prefer to use the term 'planar' for $d = 2$ data and reserve the label 'spatial' for the situation $d = 3$ in order to clearly distinguish between the two (e.g. Illian et al., 2008). We do not make this distinction, however, and will continue to refer to our 2D point patterns as 'spatial' as the specific case of $d = 3$ is not considered in this work.

Finally, we mention that much of the historical literature on point processes is heavily technical. Here, we are concerned more with the applied aspects with respect to geographical epidemiology, and focus on refinement of issues associated with the implementation of the methodology. The following discussions represent a review of point process theory most relevant for our purposes. Where necessary, the reader is directed to some useful and accessible textbook references for additional theoretical rigor.

1.2.1 Complete spatial randomness

A natural starting place in the analysis of many applications involving point process data is to ask the question "are we observing a completely random phenomenon?" In other words, it is of interest to determine whether or not there exists structure in the observed points that is unlikely to occur under *complete spatial randomness* (CSR).

Suppose we observe n points within the finite region $\mathcal{R} \subset \mathbb{R}^d$ with Lebesgue measure $|\mathcal{R}|$ (when $d = 2$, this simply refers to the area of the study region). CSR would imply (a) that the observed points are an independent random sample from the uniform distribution on $\mathcal{R}$,

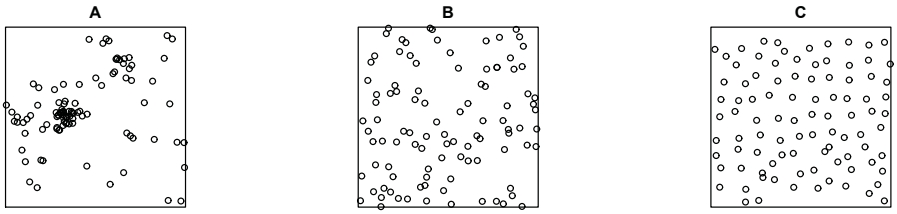


Figure 1.4: Three hypothetical examples of 2D point patterns.

and (b) that n follows a Poisson distribution with mean $\lambda|\mathcal{R}|$ (Diggle, 2003). The parameter λ is referred to the *intensity* of the point process and is interpreted as *the expected number of points per unit area*. The intensity (or *intensity function* $\lambda(\boldsymbol{x})$; $\boldsymbol{x} \in \mathcal{R}$) is frequently the core focus of point process analyses and will be discussed in greater detail in the following section.

Consider as illustrative examples the three artificially generated point patterns in Figure 1.4, labelled A, B and C. Patterns A and C were specifically generated to violate the condition of CSR. Pattern B was generated under the condition of CSR.

- Suppose these patterns were observed in practice. In pattern A, there appears to be at least one obvious large grouping of observations. Is any grouping perhaps more prominent than what we would expect for a consignment of points randomly scattered throughout the region? In terms of the point intensity, we seem to be observing *heterogeneity*.

- It could reasonably be argued that pattern B exhibits a closer resemblance to complete randomness, with groupings and separations of points not appearing to be 'too extreme'. In terms of the intensity, we seem to be observing *uniformity* or *homogeneity*.

- Point pattern C does not contain any obvious point groupings. But on the other hand, there seems to be an unusual systematic structure in the pattern unlikely to arise for uniformly distributed points. In terms of the intensity, we seem to be observing *regularity*.

A visual interpretation of a given pattern is of course by itself a rather naïve judgement of CSR. Much of the early literature on point process analysis deals with the development of statistical tests under a null hypothesis of complete randomness. Tests based on *interpoint distances*, *nearest neighbour distances* and *point-to-nearest-event distances*, for example, all compare what we expect to see of these types of distances under CSR with what we actually have observed. A significant departure from this expectation, usually achieved with exceedance of Monte-Carlo CSR significance 'envelopes' by our observed distance measures, would hence constitute sufficient evidence to reject our randomness assumption. Alternative tests that do not require evaluation of location-to-location distances investigate the distribution of event counts within $\mathcal{S}$ disjoint sub-regions of equal area. Use of *quadrat counts*

offers a simple and fast alternative to distance-based methods at the cost of the requirement to specify an arbitrary grid resolution and a loss of detail.

Good descriptions of these and some other simple preliminary tests of a given point pattern for CSR can be found in e.g. Diggle (2003) and Illian et al. (2008). Generally speaking, examples of point processes in epidemiology exhibit the type of inhomogeneity we noted in pattern A; patterns B and especially C are rarely, if ever, encountered in this field. Our focus remains on the more complicated aspect of modelling, and prediction based on, these patterns. Any testing for CSR is automatically absorbed into estimation of the intensity function and examination of second-order properties (this will become clearer in Section 1.2.3).

1.2.2 The intensity function

The fundamental first order property describing the mean structure of a point process is known as the *intensity function*. The intensity of a particular point process is key in dictating the structure and appearance of the observed points. For this reason, a large part of inference based on spatial point processes deals with (sensible) estimation of the intensity. Let $N(\mathcal{B})$ denote the number of points or events within the window $\mathcal{B}$, and $\delta\boldsymbol{x}$ a small circular region centered at location $\boldsymbol{x}$. Formally, the intensity function $\lambda(\cdot)$ is then expressed as

$$\lambda(\boldsymbol{x}) = \lim_{|\delta\boldsymbol{x}|\to 0} \left\{ \frac{\mathbb{E}\left[N(\delta\boldsymbol{x})\right]}{|\delta\boldsymbol{x}|} \right\}; \qquad \boldsymbol{x} \in \mathcal{R}. \tag{1.1}$$

We interpret $\lambda(\boldsymbol{x})$ (as mentioned in the previous section) as the *expected number of points per unit area* when situated at $\boldsymbol{x}$. If $|\delta\boldsymbol{x}|$ represents an infinitely small area then $\lambda(\boldsymbol{x})|\delta\boldsymbol{x}|$ denotes the probability that there is a point within this subset of the space. Also note that

$$\mathbb{E}\left[N(\mathcal{R})\right] = \int_{\mathcal{R}} \lambda(\boldsymbol{x})\, \mathrm{d}\boldsymbol{x}. \tag{1.2}$$

From (1.2) we see that scaling of (1.1) by $N(\mathcal{R})^{-1}$ provides the continuous probability *density* function with support on $\mathcal{R}$. This highlights an assumption underpinning much of the conventional theory on point process modelling: *simplicity* (the term used by Illian et al., 2008). Also referred to as an *orderly* point process by Diggle (2003), such a process is one where there is zero probability of observing more than one point at a specific location $\boldsymbol{x}$. In practice, this can be violated due to, for example, too coarse a coordinate grid resolution when recording events.

Though the intensity function is the standard for point process analyses as it incorporates information on the size of the data set, the density functions are useful when, for example, we wish to compare two different point patterns 'on equal playing fields'. Densities are used in Chapter 2.

Poisson processes

The *homogeneous Poisson process* is generally considered the foundation for the theory of point process modelling. It is the ideal theoretical representation of CSR, and the simplest

stochastic mechanism for generation of the observations (Diggle, 2003). Repeated from Section 1.2.1, this implies for $\boldsymbol{x} \in \mathcal{R}$ we observe

$$\text{a constant intensity } \lambda(\boldsymbol{x}) \equiv \lambda; \qquad N(\mathcal{R}) \sim \text{POI}\left(\int_{\mathcal{R}} \lambda(\boldsymbol{x})\, \mathrm{d}\boldsymbol{x}\right); \qquad \text{and}$$

$$\text{the } N(\mathcal{R}) \text{ events are independent,}$$

where we use the term 'independent' in the sense that the placement of a given point is not affected by the specific placement of any others. With no change in the latter two features, we may straightforwardly generalise these conditions to the *inhomogeneous* or *heterogeneous Poisson process* whereby a non-constant intensity $\lambda(\boldsymbol{x})$ exists.

It is worth noting that a *stationary* spatial point process is one which possesses a constant mean and covariance structure across space. The homogeneous Poisson process, with an unchanging intensity and consistent independence amongst points, is therefore considered stationary. The inhomogeneous Poisson process is clearly *nonstationary* – the constant independence remains but the mean (i.e. the intensity) now varies according to location $\boldsymbol{x}$. These terms become particularly important when we consider the more complicated processes below.

A natural estimator of the constant or 'overall' *intensity measure* is $\rho = N(\mathcal{R})/|\mathcal{R}|$. This quantity is useful under both stationarity (where our estimate of the true intensity is simply $\hat{\lambda} = \rho$) and nonstationarity, though we obviously lose local information on the distribution if the point process is indeed inhomogeneous. For estimation of a non-constant intensity $\hat{\lambda}(\boldsymbol{x})$ there exist both parametric (e.g. regression) and nonparametric (e.g. smoothing) approaches. The restrictive nature of fully parametric methods, coupled with the high heterogeneity often observed in epidemiological problems make the flexibility afforded by nonparametric approaches very appealing. One particularly useful method is known as *kernel density/intensity estimation*. There exist a number of difficulties with this estimation technique in practice however, and we therefore defer the detailed discussion on two-dimensional kernel density estimation until Chapters 2 and 3, where it will be used extensively.

Cox processes

Through an inhomogeneous Poisson process we may observe point patterns like A, where some degree of point aggregation is present. However, the fundamental aspect of Poisson processes is the presence of a fixed, unknown true intensity function $\lambda(\boldsymbol{x})$ – should subsequent, additional patterns be generated, they would follow the precisely defined intensity structure.

In some applications, the inherent *independence* between the individual observations in a Poisson process is not necessarily an intuitively sensible treatment of the process(es) at hand. It may well be more appropriate to consider the observations to be correlated depending on some measure of distance to other points, in the sense that positive correlation would produce *clustered* patterns; negative correlation would induce *repulsion*. As such, this implies we are no longer constrained by a single, fixed planar structure in terms of point generation.

This quite different modelling ideology can be obtained if we assume the underlying in-

tensity is itself stochastic. The models which take this form are occasionally referred to as *doubly stochastic*, as the occurrence of events now takes two steps: only after the intensity function has been generated can the points themselves be generated. The idea of a stochastically occurring intensity function is intuitively sensible in geographical epidemiology. As alluded to above, without special conditions on the population of interest it may even be unreasonable to think of the dispersion of a disease as following a fixed, pre-defined, underlying structure – even if that structure is highly heterogeneous.

One particular class of doubly stochastic models of great use in geographical epidemiology is that of the *Cox process* (Cox, 1955). Cox processes provide the user with a powerful framework with which to model a variety of patterns, in particular those which exhibit the type of clustering as seen in example pattern A.

For $\{\boldsymbol{x}, \boldsymbol{y}\} \in \mathbb{R}^d$ let $\Lambda(\boldsymbol{x})$ denote a non-negative d-dimensional random field with mean and covariance functions

$$\mu(\boldsymbol{x}) = \mathbb{E}[\Lambda(\boldsymbol{x})] \qquad \text{and} \qquad \mathcal{C}(\boldsymbol{x}, \boldsymbol{y}) \equiv \mathrm{Cov}[\Lambda(\boldsymbol{x}), \Lambda(\boldsymbol{y})]$$

respectively. Then *conditional upon* $\Lambda(\boldsymbol{x}) = \lambda(\boldsymbol{x})$, the observed events are an inhomogeneous Poisson process with intensity function $\lambda(\boldsymbol{x})$ as defined above. For our purposes in $d = 2$, this *infinite* process becomes *finite* when we restrict attention to the measurable geographical study region such that $\{\boldsymbol{x}, \boldsymbol{y}\} \in \mathcal{R} \subset \mathbb{R}^2$.

If the underlying stochastic process is second-order stationary, i.e. $\mu \equiv \mu(\boldsymbol{x})$ and $\mathcal{C}(\boldsymbol{x}, \boldsymbol{y}) \equiv \mathcal{C}(||\boldsymbol{x} - \boldsymbol{y}||)$ where $||\cdot||$ denotes Euclidian distance, then the resulting point process will also be stationary. This does not mean that any subsequently generated point patterns must exhibit CSR. Suppose we have a stationary Cox process with a certain non-trivial correlation structure generating our intensity functions. *Given* a realisation of the stochastic process, the subsequent point pattern will thus be inhomogeneous Poisson, hence (conditionally) nonstationary. Cox processes relinquish the specification of the locations of point clusters (i.e. they are now random), instead using dependence between generated points specified by $\mathcal{C}$ to govern the appearance of the pattern.

This clarifies the important distinction between (unconditional) Poisson processes and doubly stochastic processes. If we are able to uncover *any* statistical evidence of dependence (i.e. correlation) among the observations in a particular point pattern, then an *a priori* assumption that the data have arisen from a Poisson process could well be violated. However, it is understandably extremely difficult to formally distinguish between (finite) Poisson and Cox processes. In fact, it is impossible to do so given only a single realisation, unless rather restrictive assumptions are made e.g. concerning the rate of change of the mean intensity function. A good discussion on disentangling heterogeneity (i.e. variation in intensity) and clustering (arising from correlation) can be found in Section 9.6 of Diggle (2003). This will also be highlighted when we discuss second-order properties in Section 1.2.3.

Inhibitive processes

Cox and Poisson processes are arguably those which most sensibly and conveniently model problems such as examples A and B. Though it is perfectly possible for an inhomogeneous

Poisson process to, for example, produce patterns like C, it may be unreasonable given the application to assume unrestricted independence. We may make use of a more appropriate category of models to deal with observed *regularity*.

The most basic class of these processes are labelled *simple inhibition processes*. The only departure from CSR for these processes is the imposition of some minimal allowable distance between two points. The intensity of such processes is often expressed in terms of a *packing intensity*, explained as the expected proportion of $\mathcal{R}$ covered by non-overlapping balls with a diameter equal to the specified minimum distance. A simple inhibition rule can, however, be rather restrictive. We may introduce more flexibility into this setup by considering types of *Markov point processes*, characterised generally by a likelihood ratio function compared with a Poisson process with a constant unit intensity. Models such as *Gibbs* and *Strauss point processes* have been shown to be particularly useful in practice.

In general, processes that exhibit such inhibition among the observed events are not particularly common in geographical epidemiology, where we would expect events to have a proclivity to cluster. Repulsion in point patterns is more often observed in fields such as ecology, where for example we might be observing two rival species of plant attempting to co-exist in the same area. It is for this reason Cox and Poisson processes are the focus of this thesis; here we are not directly concerned with inference based on patterns behaving as in example C. For more on regular point patterns and associated modelling strategies the reader is directed to the Diggle (2003) and Illian et al. (2008) textbooks as a good introduction.

1.2.3 Second-order characteristics

The intensity function describes a point process at one location $\boldsymbol{x} \in \mathcal{R}$. *Second-order* properties by definition study the relationship between distinct *pairs* of points, and are hence essential in helping to gauge the nature of covariance structures in observed patterns. For this reason, many generally regard examination of estimated second-order summary statistics as one of the most important activities carried out when investigating point processes.

A practical examination of second-order properties of a general point process is very difficult. This pursuit is greatly eased by assuming, in addition to simplicity, the important aspects of *stationarity* and *isotropy*. Stationarity and isotropy imply that the properties of the process at hand are invariant under translation and rotation of the points respectively. Though upon first glance this may seem somewhat restrictive, it does not in any way eliminate the possibility of generating highly heterogeneous point patterns based on e.g. a Cox process, and renders a mathematical examination (hence a statistical summary) of second-order properties far more accessible than would otherwise be the case.

Here, we will briefly discuss the second-order characteristics most relevant for the purposes of this thesis. Fortunately, the theoretical properties and the corresponding estimators of interest are intuitively sensible, and are therefore relatively straightforward to conceptualise from an applied perspective.

Second-order product-density

At the heart of much of the theory is the idea of the *second-order density* (or *intensity*). With $\delta\boldsymbol{x}$ and $\delta\boldsymbol{y}$ representing infinitesimal regions containing locations $\boldsymbol{x}$ and $\boldsymbol{y}$ respectively, then

$$\lambda_2(\boldsymbol{x},\boldsymbol{y}) = \lim_{|\delta\boldsymbol{x}|,|\delta\boldsymbol{y}|\to 0}\left\{\frac{\mathbb{E}\left[N(\delta\boldsymbol{x})N(\delta\boldsymbol{y})\right]}{|\delta\boldsymbol{x}||\delta\boldsymbol{y}|}\right\}; \qquad \boldsymbol{x},\boldsymbol{y}\in\mathcal{R}; \qquad \boldsymbol{x}\neq\boldsymbol{y}, \tag{1.3}$$

whereby $\lambda_2(\boldsymbol{x},\boldsymbol{y})|\delta\boldsymbol{x}||\delta\boldsymbol{y}|$ is interpreted as the probability that there contains a point in each of $\delta\boldsymbol{x}$ and $\delta\boldsymbol{y}$. For a stationary, isotropic process $\lambda_2(\boldsymbol{x},\boldsymbol{y})$ reduces to $\lambda_2(||\boldsymbol{x}-\boldsymbol{y}||)$ where $||\cdot||$ represents the Euclidian vector norm. That is, the second-order intensity depends only on the distance between the two locations, as opposed to their specific positions. For a Poisson process, independence dictates that $\lambda_2(\boldsymbol{x},\boldsymbol{y})\equiv\lambda(\boldsymbol{x})\lambda(\boldsymbol{y})$.

The second-order product-density is the fundamental building block for the K- and pair correlation functions below and as such, all three quantities are closely related. Coverage of $\hat{\lambda}_2$, the estimator of λ_2, is therefore absorbed into the following discussions.

K-function and variants

A particularly useful second-order statistic for describing the nature of a given point pattern is that of the K-function, credited to Ripley (1977). Ripley's K-function characterises the pattern at hand in terms of the count of points within an area a specified distance r from the position of an arbitrary observation (in practice, 'arbitrary' refers to any event in the pattern selected at random). Theoretically, for a simplistic, stationary and isotropic process,

$$K(r) = 2\pi\rho^{-2}\int_0^r t\lambda_2(t)\,\mathrm{d}t, \tag{1.4}$$

which clarifies the relationship between K and the second-order product-density (recall that ρ is the overall intensity measure defined in Section 1.2.2). We may also note from (1.4) that the K-function is closely related to the cumulative distribution function of the interpoint distances.

The K-function is popular due to the ease with which it may be estimated and the intuitive fashion with which we may interpret the results. Let $\boldsymbol{X}$ be a point pattern of size n observed within $\mathcal{R}$, with overall estimated intensity measure $\hat{\lambda}=n/|\mathcal{R}|$. Then the K-function can be estimated with

$$\hat{K}(r) = \hat{\lambda}^{-2}|\mathcal{R}|^{-1}\sum_{\boldsymbol{u},\boldsymbol{v}\in\boldsymbol{X}}^{\boldsymbol{u}\neq\boldsymbol{v}}\mathbf{1}(||\boldsymbol{u}-\boldsymbol{v}||\leq r)w_{\mathcal{R}}(\boldsymbol{u},\boldsymbol{v}); \qquad 0\leq r\leq r_{\max}, \tag{1.5}$$

where we may straightforwardly interpret the quantity $\hat{\lambda}\hat{K}(r)$ as 'the expected number of further observations of $\boldsymbol{X}$ falling within a disc of radius r randomly positioned at an event in $\mathcal{R}$'. In (1.5), $\mathbf{1}$ denotes the indicator function, $r_{\max}$ is some upper bound on the distances able to be evaluated by the function (discussed further in a moment), and $w_{\mathcal{R}}(\boldsymbol{u},\boldsymbol{v})$ is an edge-correction factor necessary due to the finiteness of the study region. Edge-correction is a common problem in spatial statistics due to the fact that working in a restricted subset of space, there may well exist unobserved points outside of the region that

would otherwise impact the estimation of desired quantities. Edge-correcting is generally an appropriate reweighting step designed to compensate for this loss of information. Though there exist several different approaches to defining w in (1.5), we will make use of the correction specifically geared towards stationary and isotropic processes and also introduced in Ripley (1977). Working in $d = 2$, let $c(\boldsymbol{a}, b)$ denote the perimeter of the disc centered at $\boldsymbol{a}$ and with radius b. Then

$$w_{\mathcal{R}}(\boldsymbol{u}, \boldsymbol{v})^{-1} = \frac{\{c(\boldsymbol{u}, ||\boldsymbol{u} - \boldsymbol{v}||) \cap \mathcal{R}\}}{2\pi||\boldsymbol{u} - \boldsymbol{v}||} \tag{1.6}$$

i.e. $w_{\mathcal{R}}$ is the the reciprocal of the proportion of the circumference of the circle centered at $\boldsymbol{u}$ and intersecting $\boldsymbol{v}$ that falls within $\mathcal{R}$. As such, it is an intuitively sensible technique: interpoint distances resulting from observations near the boundary of $\mathcal{R}$ will be given a weight > 1 in (1.5), while pairs of points a suitable distance from the edges will have $w_{\mathcal{R}}(\boldsymbol{u}, \boldsymbol{v}) = 1$.

Ripley (1977) shows that (1.5) is approximately unbiased for suitably small r. The issue of edge-correction due to observations falling within a restricted interval implies, however, we must specify a maximum allowable distance $r_{\max}$ for $\hat{K}$. Obviously, we cannot count the number of points within some distance t of the location $\boldsymbol{s} \in \mathcal{R}$ if this distance is excessively large in the sense that the intersection of $\mathcal{R}$ with the perimeter of the disc at $\boldsymbol{s}$ possessing a radius of t constitutes the empty set. This leads naturally to defining $r_{\max}$ to be the maximum distance possible such that every coordinate within the study region is still able to locate another interior position $r_{\max}$ units away. In practice though, the evaluation of the K-function estimator at such large distances is rarely of use and is ill-advised due to an obvious increase in sampling variability as $r \to \infty$.

We may use plots of $\hat{K}$ to examine a given point pattern as a diagnostic of sorts for indications of attraction or repulsion at specific distances given by r. Consider a realised pattern from the homogeneous Poisson process with intensity λ. Clearly, the expected number of points within a disc of radius r from an arbitrary point will be equal to $\lambda\pi r^2$. Plots of $\hat{K}$ for any pattern at suitable distances r can therefore be compared to πr^2, the 'theoretical' curve; sufficiently extreme departures from which would suggest the presence of residual attraction ($\hat{K}(r) > \pi r^2$) or repulsion ($\hat{K}(r) < \pi r^2$). Determining what would constitute a significant departure from the theoretical trend can be aided by empirical tolerance envelopes. These are typically made up of e.g. the 0.025th and 0.975th quantiles of $\hat{K}(r)$ values computed from N iterated point patterns generated under the assumption of a homogeneous Poisson process within the same window $\mathcal{R}$ with constant intensity equal to the overall intensity measure of $\boldsymbol{X}$, namely $\hat{\lambda}$. To highlight this functionality, we calculate $\hat{K}$ for the point patterns shown in Figure 1.4, and display these in Figure 1.5 along with the theoretical line. Also included are empirical tolerance envelopes computed as described here with $N = 100$.

- As we would expect, there is clear evidence of clustering in pattern A, with positive exceedances of the envelopes at almost all evaluated distances. This suggests that the assumption that the points in A were generated from a homogeneous Poisson process is invalid due to the detection of positive dependence.

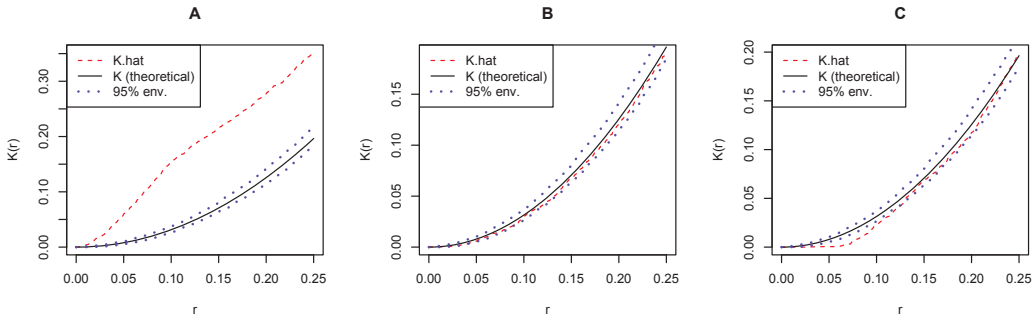


Figure 1.5: Estimates of Ripley's K-function for the three point patterns A, B and C.

- Pattern B, on the other hand, has shown no evidence against the homogeneous Poisson process assumption. There are no visible breaches of the tolerance envelopes.

- The apparent repulsion evident from a visual inspection of pattern C is supported by the exceedance of the lower limit of the empirical envelopes based on the theoretical Poisson process curve. Note that unlike the attractive process, detection of repulsion at larger distances is understandably more difficult.

The fact that the $\hat{K}$ results for patterns A and C do not agree with those of the theoretical appearance of a homogeneous Poisson process is unsurprising. It is important to remember, however, that this suggests correlation between the points only in the case we assume (perhaps naaïvely) that the intensities are constant. This highlights the aforementioned fact that from a single realised point pattern, it is not possible to determine if we are dealing with, say, a (finite) Cox process or an (unconditional) inhomogeneous Poisson process on $\mathcal{R}$.

In an effort to aid in the detection of additional dependence if we assume (and account for) an underlying inhomogeneous intensity function in two dimensions, Ripley's K-function was extended to the *inhomogeneous* K-function by Baddeley et al. (2000). The validity of K_{inhom} is contingent upon the observed point pattern exhibiting *second-order intensity reweighted stationarity*, a definition coined by the same authors which requires that the quantity $g_{\text{inhom}}(\boldsymbol{x}, \boldsymbol{y}) = \frac{\lambda_2(\boldsymbol{x},\boldsymbol{y})}{\lambda(\boldsymbol{x})\lambda(\boldsymbol{y})} \equiv g_{\text{inhom}}(r)$ where $r = ||\boldsymbol{x} - \boldsymbol{y}||$. That is, that g_{inhom} depends only on the distance between $\boldsymbol{x}$ and $\boldsymbol{y}$, and that the intensity function λ is bounded away from zero. This is important as it implies that together, λ and g completely define the second-order (co-varying) nature of the process with an inhomogeneous intensity and allows for the definition of (1.4) to be naturally extended to

$$K_{\text{inhom}}(r) = 2\pi \int_0^r t g_{\text{inhom}}(t)\, \mathrm{d}t. \tag{1.7}$$

The corresponding estimator is given by

$$\hat{K}_{\text{inhom}}(r) = |\mathcal{R}|^{-1} \sum_{\boldsymbol{u},\boldsymbol{v}\in\boldsymbol{X}}^{\boldsymbol{u}\neq\boldsymbol{v}} \frac{\mathbf{1}(||\boldsymbol{u} - \boldsymbol{v}|| \leq r) w_{\mathcal{R}}(\boldsymbol{u}, \boldsymbol{v})}{\hat{\lambda}(\boldsymbol{u})\hat{\lambda}(\boldsymbol{v})}; \qquad 0 \leq r \leq r_{\max}, \tag{1.8}$$

where now $\hat{\lambda}(\cdot)$ is a suitable estimate of the non-constant intensity. Note that both (1.7) and (1.8) reduce to their stationary counterparts in equations (1.4) and (1.5) if we experience a constant intensity. Once the intensity has been estimated, the interpretation of plots of $\hat{K}_{\text{inhom}}$ remain the same as for $\hat{K}$, in that significant departures from the theoretical πr^2 would suggest dependence in excess of the assumed $\hat{\lambda}(\cdot)$.

A transformation of Ripley's K-function and also commonly used in practice is the L-function suggested by Besag (1977), whereby $\hat{L}(r) = \sqrt{\hat{K}(r)/\pi}$ in the case of $d = 2$. Though obviously representing the same information as the K-function, we compare the results of $\hat{L}$ with a straight line; under a Poisson assumption $\hat{L}(r) = r$. It is argued that deviations from this straight line are easier to detect and interpret in the sense of the second-order behaviour of the process than deviations from the parabolic curve $\hat{K}(r) = \pi r^2$. Analogously, $\hat{L}_{\text{inhom}}(r) = \sqrt{\hat{K}_{\text{inhom}}(r)/\pi}$.

Pair correlation function

Where the K-function is related to the cumulative distribution function of the interpoint distances, the pair correlation function (PCF) is related to the *density* function of the interpoint distances. As the name would suggest, the PCF g defines the strength of the correlation (if any) between pairs of locations on $\mathbb{R}^d$ for a given point process. Theoretically, for a simplistic, stationary and isotropic process with a homogeneous intensity,

$$g(\boldsymbol{x}, \boldsymbol{y}) = g(r) = \frac{\lambda_2(r)}{\lambda^2}. \tag{1.9}$$

Assume the presence of the 2D point pattern $\boldsymbol{X}$ defined above. Then (1.9) may be estimated with

$$\hat{g}(r) = \frac{\hat{\lambda}_2(r)}{\hat{\lambda}^2} = \frac{1}{2\pi\hat{\lambda}^2 r|\mathcal{R}|} \sum_{\boldsymbol{u},\boldsymbol{v}\in\boldsymbol{X}}^{\boldsymbol{u}\neq\boldsymbol{v}} \kappa_h(||\boldsymbol{u}-\boldsymbol{v}|| - r) w_{\mathcal{R}}(\boldsymbol{u}, \boldsymbol{v}); \qquad 0 < r \leq r_{\max}, \tag{1.10}$$

where κ is a univariate smoothing kernel. This is typically elected to be the Epanechnikov density where $\kappa_h(u) = \frac{3}{4}h^{-1}(1 - u^2/h^2)\mathbf{1}(-h \leq u \leq h)$. As such, the smoothing parameter h is critical to the quality of the resulting estimate.

In fact, we have already met the PCF by introducing the inhomogeneous K-function (1.7). As second-order intensity reweighted stationarity implies that $g_{\text{inhom}}(\boldsymbol{x}, \boldsymbol{y})$ need only depend upon $r = ||\boldsymbol{x}-\boldsymbol{y}||$ even if we assume an underlying inhomogeneous intensity function, we may estimate g_{inhom} with

$$\hat{g}_{\text{inhom}}(r) = \frac{1}{2\pi r|\mathcal{R}|} \sum_{\boldsymbol{u},\boldsymbol{v}\in\boldsymbol{X}}^{\boldsymbol{u}\neq\boldsymbol{v}} \frac{\kappa_h(||\boldsymbol{u}-\boldsymbol{v}|| - r) w_{\mathcal{R}}(\boldsymbol{u}, \boldsymbol{v})}{\hat{\lambda}(\boldsymbol{u})\hat{\lambda}(\boldsymbol{v})}; \qquad 0 < r \leq r_{\max}, \tag{1.11}$$

where, once again, $\hat{\lambda}(\cdot)$ is now an estimate of the non-constant intensity.

The interpretation of the PCF is fairly straightforward, and is aided by thinking about its direct relationship with the second-order product-density (1.9). Poisson processes by definition exhibit independence among the observed points and thus with other Poisson processes.

Doubly stochastic processes such as the Cox process do not meet this independence assumption when the specified covariance structure $\mathcal{C}$ is non-trivial in the sense that the position of any given observation impacts the placement of any other point(s). Formally speaking, the probability $p_2(\boldsymbol{x}, \boldsymbol{y})$ that we observe a point in each of the infinitesimal sub-regions $\delta\boldsymbol{x}$ and $\delta\boldsymbol{y}$ in $\mathcal{R}$ is

$$p_2(\boldsymbol{x}, \boldsymbol{y}) = \lambda^2 g(r)|\delta\boldsymbol{x}||\delta\boldsymbol{y}| \qquad \text{and} \qquad p_2(\boldsymbol{x}, \boldsymbol{y}) = \lambda(\boldsymbol{x})\lambda(\boldsymbol{y}) g_{\text{inhom}}(r)|\delta\boldsymbol{x}||\delta\boldsymbol{y}|$$

for simplistic, stationary and isotropic processes assuming homogeneous and inhomogeneous intensities over $\mathcal{R}$ respectively. In order for the latter specification to hold, the process must also be second-order intensity reweighted stationary.

As such, the PCF can be viewed as a 'correlation factor' required to correctly calculate p_2. The independence assumption of Poisson processes means that $g(r) \equiv 1$ for all distances, coinciding with the multiplicative theorem of probability theory. Clearly, this will not be the case if the process does exhibit dependence i.e. we will observe $0 < g(r) < 1$ at small distances if there exists repulsion; $g(r) > 1$ for attraction. Some applications have a *hard-core* distance u_0 such that $g(u) = 0$ for $u \leq u_0$, indicating a zero probability of another observation u units away; this however is of limited interest in our applications. Finally, we note that is generally assumed in practice that $g(r) \to 1$ as $r \to \infty$, and that this collection of comments are clearly also valid for $g_{\text{inhom}}(r)$ when we impose second-order intensity reweighted stationarity.

In the same way that plots of $\hat{K}$ and $\hat{K}_{\text{inhom}}$ can be compared to the theoretical πr^2, we may therefore compare plots of $\hat{g}$ and $\hat{g}_{\text{inhom}}$ to the horizontal line at 1; the theoretical value of the PCF in the absence of any dependence. Assuming a homogeneous underlying intensity in the same way as the K-function example, Figure 1.6 shows plots of $\hat{g}$ for patterns A, B and C. Also included are 95% empirical tolerance envelopes computed in the same way as earlier.

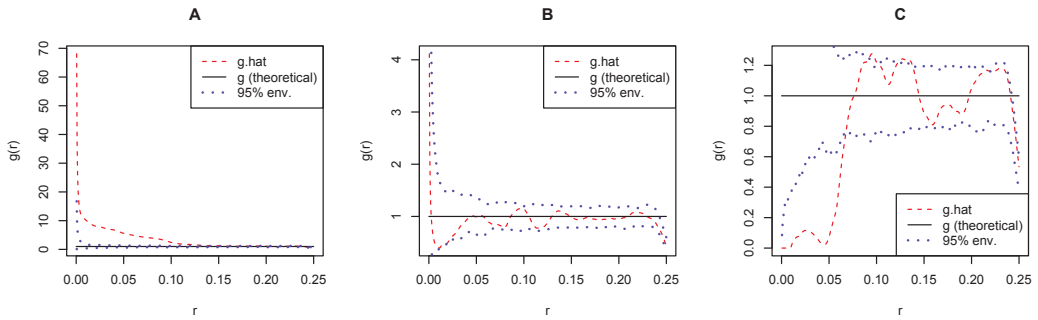


Figure 1.6: Estimates of the PCF based on a homogeneous underlying intensity for the three point patterns A, B and C.

Naturally, the same information is gleaned from Figure 1.6 as from Figure 1.5 in terms of indications of dependence. Where pattern B does not suggest any significant deviation under the assumption of a homogeneous Poisson process, the plots corresponding to patterns A and C do. We note an obvious positive breach of the envelopes for the former, indicating

attraction. Negative breaches are visible for the latter, supporting the notion of repulsion. Note also that the extremity of the exceedances reduces to nil as the distances become large.

We have spent the time to discuss in detail these important second-order characteristics and their estimators, in particular the ability to search for dependence in a given point pattern, due to the important role they play in quantifying said dependence for more intensive spatial and spatiotemporal modelling. This is investigated further in Chapter 3.

1.2.4 Higher-order properties

The second-order product-density naturally extends to the kth order, whereby

$$\lambda_k(\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_k) = \lim_{|\delta\boldsymbol{x}_1|,|\delta\boldsymbol{x}_2|,\ldots,|\delta\boldsymbol{x}_k|\to 0} \left\{ \frac{\mathbb{E}\left[N(\delta\boldsymbol{x}_1) \times N(\delta\boldsymbol{x}_2) \times \ldots \times N(\delta\boldsymbol{x}_k)\right]}{|\delta\boldsymbol{x}_1| \times |\delta\boldsymbol{x}_2| \times \ldots \times |\delta\boldsymbol{x}_k|} \right\} \quad (1.12)$$

for infinitesimal disjoint areas $\{\delta\boldsymbol{x}_1, \delta\boldsymbol{x}_2, \ldots, \delta\boldsymbol{x}_k\} \in \mathcal{R}$. This describes the distribution of the configuration of k points within the study region, such that $\lambda_k(\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_k) \times |\delta\boldsymbol{x}_1| \times |\delta\boldsymbol{x}_2| \times \ldots \times |\delta\boldsymbol{x}_k|$ equals the probability of observing a point in each of the k areas.

The value of exploring second-order characteristics when analysing point patterns is clear. However, first- and second-order properties do not by themselves uniquely define a particular process; examples have been provided in the literature in which different processes have yielded identical intensities and/or K- and g (PCF) functions (Baddeley and Silverman, 1984; Diggle, 2003; Tscheschel and Stoyan, 2006). Finer details of point processes can be explored by examining third- and higher-order characteristics. Understandably, the available techniques and interpretation of estimated quantities become less straightforward.

For most practical intents and purposes however, the exploration of first- and second-order properties is sufficient in order to gain insight into the structure and behaviour of the point process(es) at hand, and this is where our applied focus will remain. For more on higher-order statistical analysis of point patterns, as well as some additional references, see e.g. Møller and Waagepetersen (2004); Illian et al. (2008).

1.3 Thesis aims and structure

This work seeks to investigate the role of statistical methods for investigating two-dimensional point process problems in epidemiology. The relative youth of this type of continuous space and space-time modelling (particularly the latter) means there are a number of issues requiring attention in order to maximise the benefits associated with these types of detailed analyses. The complexities of the methodology also suggest an effort be made to streamline and condense the most relevant and advantageous theoretical advancements for accessibility by epidemiologists and other applied researchers. To these ends, the work in this thesis revolves around three core **Aims**:

1. **To identify, collate, summarise, test, and refine or improve where possible, areas of statistical methodology currently used to analyse spatial and spatiotemporal point patterns.**

2. **To aid in bridging the gap between powerful yet theoretically complex models and applied researchers.**

3. **To trial advancements made on real-world data.**

The current chapter, *Introduction*, has provided a historical review of the kinds of applications and methodologies most frequently encountered in geographical epidemiology, thereby motivating the the remainder of the dissertation. This is structured as follows. Chapter 2, *Spatial Relative Risk*, begins with an investigation into one of the most common initial exploratory analyses: the dispersion of spatially recorded disease cases in relation to the uninfected, at-risk population. In so doing, the first detailed appraisal of a particular variable smoothing approach for the kernel-estimated density-ratio is conducted.

Chapter 3, *Spatial and Spatiotemporal Modelling with the Log-Gaussian Cox Process*, delves deeper into the *a priori* assumptions made about the genesis of the observed data. This allows access to a sophisticated and highly flexible modelling framework for both spatial and spatiotemporal point patterns. Only recently detailed in the literature, there is much which remains unknown about its practical performance. Of particular interest is the gauging of the quality of estimation techniques for critically important parameters of these models. Both Chapters 2 and 3 seek to address **Aim 1**.

The assessment and development of this material requires a considerable programming effort. The associated computational implementations were the first step in the development of two separate software packages, detailed in Chapter 4: *Software Releases*. The sizeable collection of functions reflect the methodology examined in Chapters 2 and 3; the package corresponding to the latter was developed in close collaboration with researchers at Lancaster University, United Kingdom. Addressing **Aim 2**, the R-language releases are available in the public domain at `http://cran.r-project.org/` – the online Comprehensive R Archive Network (CRAN).

It is important to not only examine the methodology, but to ensure to a certain extent that any identified theoretical benefits can be realised in practice when dealing with real-world data. **Aim 3** is addressed throughout the aforementioned chapters by the inclusion and interpretation of analyses on non-synthetic data sets. In addition to these included examples, researchers from several other institutions have expressed interest in the subject material at hand. Chapter 5, *Collaborative Applications and Case Studies*, describes and carries out some analyses (both collaborative and independent) drawn from the non-statistical literature which have made use of work detailed in this thesis. Finally, a review and summary are provided in the *Conclusion* (Chapter 6).

Chapter 2

Spatial Relative Risk

It is sensible to consider the spatial element in epidemiological applications. Intuitively, we can expect the incidence and prevalence of disease to vary over a given geographical region, due to an inhomogeneous population dispersion. The observed complexity of these dispersions in terms of their heterogeneity can leave purely parametric approaches inadequate for dealing with this type of variation. Many studies in geographical epidemiology, particularly those concerning point process data, therefore make use of kernel density estimation (KDE) for analysing spatial disease trends (see for example Kelsall and Diggle, 1995*b*; Sabel et al., 2000; Prince et al., 2001; Danese et al., 2008; Clough et al., 2009, among many others). This methodology affords the user an unmatched degree of flexibility, though not without cost. Several issues arise that must be addressed should we wish to implement KDE to model spatial disease risk, and if improperly dealt with they can have adverse effects on both the theoretical and practical results. Also worthy of consideration is the bias/variance trade-off problem ubiquitous in smoothing methods. These issues, combined with the increasing popularity of this approach in the epidemiological literature, mean an examination (and refinement) of the specific methods in an effort to provide a better understanding of their performance with respect to these types of applications is warranted. In addition, it is important to motivate a more data-driven or unified approach to these analyses, in contrast to *ad hoc* decisions being made on a problem-by-problem basis.

The goal in this chapter is to therefore provide a comprehensive discussion on the use of KDE with a view to estimating the variation in disease risk for point process data. Section 2.1 describes the bivariate (2-dimensional) kernel density estimator, for both a fixed and a spatially adaptive smoothing parameter. Section 2.2 defines the density-ratio estimator of relative risk and discusses. Practical issues such as bandwidth selection are examined in Section 2.3. A novel method for highlighting sub-regions of statistically significant risk based on asymptotic theory is outlined in Section 2.4, and we study the empirical performance of the relative risk function on several synthetic scenarios in Section 2.5. Section 2.6 illustrates the methodology with two real-world examples, and closing comments are given in Section 2.7. Advancements made in this chapter, namely the use of adaptive smoothing for the density-ratio, its theoretical properties, corresponding asymptotic tolerance contours, and some additional empirical results, have been published in Davies and Hazelton (2010)

(*Statistics in Medicine*). This naturally followed on from a similar investigation utilising only a fixed bandwidth conducted during this author's Honours year (pre-Ph.D.), published in the *Biometrical Journal* (see Hazelton and Davies, 2009).

2.1 Bivariate kernel density estimation

Kernel density estimation (KDE) is a nonparametric method that allows each observation to contribute to the final density estimate following specification of a *kernel function* and a *smoothing parameter* (also known as a *bandwidth* or *window width*). In contrast to, for example, a relatively crude histogram, a kernel estimate provides the researcher with a smooth density for continuous data. The fundamental ideas of KDE have changed little since the seminal papers of Rosenblatt (1956), Parzen (1962) and Cacoullos (1966). One important extension to the methodology was made by Breiman et al. (1977), where a variable smoothing parameter is used (in place of an otherwise constant amount of smoothing specified by the bandwidth) for each observation. Defining a particular version of the observation-wise variable bandwidth approach, Abramson (1982) demonstrated that under certain conditions, his *adaptive* kernel density estimator yielded a significantly lower level of bias than the *fixed*-bandwidth estimator; a theoretical benefit echoed in e.g. Hall and Marron (1988). More recently, Diggle, Rowlingson and Su (2005) showed an improvement in integrated square error (ISE) by using this adaptive method over the fixed when estimating intensity functions in a simulation study.

There exist several alternative options for variable bandwidth or adaptive KDE. Of note is the idea of *local* kernel estimators, where, in contrast to the methods in Breiman et al. (1977) and Abramson (1982), the smoothing parameter is varied not at each observation, but for each *location* at which the density itself is estimated. It can be shown (see for example the relevant summary in Section 2.10 of Wand and Jones, 1995) that there are theoretical benefits to local variable kernel methods over the fixed if the bandwidths are chosen 'optimally' according to some appropriate criterion. One particularly well-known local smoothing technique which attempts to achieve this is that of the *nearest neighbour* method (Loftsgaarden and Quesenberry, 1965), where information on the distances from the point of estimation to the kth closest observation is used in order to compute the final result. However, aside from the need to find an appropriate value for k, a consequence of the fact that the bandwidths are now a function of estimation location means that the final result does not necessarily integrate to 1, in which case a bona-fide density would not be achieved. It has also been noted that there can be sudden changes in the nearest neighbour distances when altering the point of estimation, resulting in a potential lack of smoothness (see again Wand and Jones, 1995).

In light of these issues, and the potentially significant bias reductions mentioned earlier, we will consider only Abramson's (1982) method in this chapter. Hereinafter, all mentions of the 'adaptive' approach thus refer to Abramson's variable bandwidth calculator. Further reading on KDE and smoothing methods in general can be found in comprehensive texts such as Silverman (1986), Scott (1992), Wand and Jones (1995), Simonoff (1996) and Bowman and Azzalini (1997). Also of note is the paper by Jones (1990), which provides a detailed

discussion on the differences between observation-wise variable bandwidth estimators and specific location-wise estimators in the univariate setting.

2.1.1 Estimator

Consider $n > 0$ independent, identically distributed (iid) bivariate observations $\boldsymbol{X}_1$, $\boldsymbol{X}_2$, $\ldots$, $\boldsymbol{X}_n$. The true density thereof, f, is estimated by $\hat{f}$ with

$$\hat{f}(\boldsymbol{z}) = \frac{1}{n}\sum_{i=1}^{n} \frac{1}{h(\boldsymbol{X}_i)^2} K\left(\frac{\boldsymbol{z}-\boldsymbol{X}_i}{h(\boldsymbol{X}_i)}\right), \tag{2.1}$$

where K, the kernel function, is ideally chosen to be a radially symmetric probability density function (PDF). The degree of smoothing is controlled by the bandwidth, indicated in (2.1) by the function $h(\cdot)$.

A fixed-bandwidth kernel estimate of a density, that is, an estimate utilising a constant amount of smoothing for all observations, is equivalent to setting $h(\boldsymbol{u}) = h_{\mathcal{F}}$, some fixed scalar value, for all $\boldsymbol{u}$. For adaptive smoothing under Abramson (1982), $h(\boldsymbol{u})$ is set to

$$h(\boldsymbol{u}) = \frac{h_0 \lambda_f(\boldsymbol{u})}{\gamma_f}, \tag{2.2}$$

where $\lambda_f(\boldsymbol{u}) = f(\boldsymbol{u})^{-\frac{1}{2}}$, h_0 is a global smoothing parameter we will call the *global bandwidth* and $\gamma_f = \exp\left\{\frac{1}{n}\sum_i \log\left[f(\boldsymbol{X}_i)^{-\frac{1}{2}}\right]\right\}$. The presence of the geometric mean term γ_f in (2.2) frees the bandwidth factors from dependence on the scale of the recorded data, and allows h_0 to be considered on the same scale as $h_{\mathcal{F}}$ in a corresponding fixed-bandwidth density estimate. Noteworthy is the fact that the components that make up (2.2) are directly related to the density we are attempting to estimate.

In real-world applications, however, the true density f is unknown. It is therefore standard practice to replace all instances of f in (2.2) with a *pilot estimate* thereof which we will represent as $\tilde{f}_{\tilde{h}}$. In our case, $\tilde{f}_{\tilde{h}}$ is a fixed kernel density estimate with the bandwidth $h_{\mathcal{F}} = \tilde{h}$ (referred to as the *pilot bandwidth*).

To be thorough, we re-express (2.2) to reflect this pilot estimation:

$$h(\boldsymbol{u}) = \frac{h_0 \tilde{\lambda}_{\tilde{f}}(\boldsymbol{u})}{\gamma_{\tilde{f}}}, \tag{2.3}$$

where $\tilde{\lambda}_{\tilde{f}}(\boldsymbol{u}) = \tilde{f}_{\tilde{h}}(\boldsymbol{u})^{-\frac{1}{2}}$ and $\gamma_{\tilde{f}} = \exp\left\{\frac{1}{n}\sum_i \log\left[\tilde{f}_{\tilde{h}}(\boldsymbol{X}_i)^{-\frac{1}{2}}\right]\right\}$.

In the interests of clarity, we introduce new notation to distinguish fixed-bandwidth bivariate kernel density estimates from adaptive. Fixed estimates of f will be denoted by $\hat{f}_{\text{fix}}$, adaptive by $\hat{f}_{\text{adv}}$.

2.1.2 Theoretical properties

We may gain insight into the behaviour of the kernel density estimator, for both fixed and adaptive smoothing regimens, via theoretical properties obtained from large-sample approximations. Prior to deriving these properties let us clarify some notation and assume the

satisfaction of certain conditions. As in the previous section, boldface characters represent 2-dimensional vectors i.e. $\boldsymbol{x} = [x_1, x_2]^\top$. We use $\mathcal{D}_\nu(\boldsymbol{x})$ to denote the 2×1 vector of first-order partial derivatives at location $\boldsymbol{x}$ for an appropriate function ν and $\mathcal{H}_\nu(\boldsymbol{x})$ the corresponding 2×2 Hessian matrix. The order notation $o(\cdot)$ for the real-valued sequences s_1 and s_2 where $s_1 = o(s_2)$ indicates that $\lim_{n\to\infty}|s_1/s_2| = 0$, and $s_1 = O(s_2)$ that $\lim_{n\to\infty}|s_1/s_2| < \infty$. Furthermore, we assume that the 'target' density f is continuously differentiable up to some required order p, and that K satisfies

$$\int K(\boldsymbol{u})\,\mathrm{d}\boldsymbol{u} = 1, \qquad \int \boldsymbol{u}K(\boldsymbol{u})\,\mathrm{d}\boldsymbol{u} = 0, \qquad \text{and} \qquad \int u_1^r K(\boldsymbol{u})\,\mathrm{d}\boldsymbol{u} = k_r \neq 0, \tag{2.4}$$

with the final condition assuming $r > 1$ and also holding for u_2. Note that if K is a radially symmetric PDF with compact support, all of the conditions in (2.4) are satisfied. Unless otherwise stated, the definitions provided in this section hold throughout the remainder of the chapter.

We begin with the fixed-bandwidth estimator, with smoothing parameter $h_{\mathcal{F}}$. This approach has received more attention in the literature than the adaptive, and good theoretical summaries are available in texts such as Silverman (1986) and Wand and Jones (1995). For the bias of $\hat{f}_{\text{fix}}$ at a given coordinate $\boldsymbol{z}$, we write

$$\begin{aligned} \text{Bias}[\hat{f}_{\text{fix}}(\boldsymbol{z})] &= \mathbb{E}[\hat{f}_{\text{fix}}(\boldsymbol{z})] - f(\boldsymbol{z}) \\ &= \int \frac{1}{h_{\mathcal{F}}^2} K\left(\frac{\boldsymbol{z}-\boldsymbol{y}}{h_{\mathcal{F}}}\right) f(\boldsymbol{y})\,\mathrm{d}\boldsymbol{y} - f(\boldsymbol{z}). \end{aligned} \tag{2.5}$$

To find an approximate expression for the bias of the fixed estimator using (2.5), we perform the change of variable $\boldsymbol{y} = \boldsymbol{z} - h_{\mathcal{F}}\boldsymbol{t}$ and invoke the multivariate form Taylor's theorem:

$$\begin{aligned} \text{Bias}[\hat{f}_{\text{fix}}(\boldsymbol{z})] &= \int K(\boldsymbol{t}) f(\boldsymbol{z} - h_{\mathcal{F}}\boldsymbol{t})\,\mathrm{d}\boldsymbol{t} - f(\boldsymbol{z}) \\ &= \int K(\boldsymbol{t})[f(\boldsymbol{z}) - h_{\mathcal{F}}\mathcal{D}_f(\boldsymbol{z})^\top \boldsymbol{t} + \frac{1}{2}h_{\mathcal{F}}^2 \boldsymbol{t}^\top \mathcal{H}_f(\boldsymbol{z})\boldsymbol{t} + o(h_{\mathcal{F}}^2)]\,\mathrm{d}\boldsymbol{t} - f(\boldsymbol{z}) \\ &= f(\boldsymbol{z}) \int K(\boldsymbol{t})\,\mathrm{d}\boldsymbol{t} - h_{\mathcal{F}}\mathcal{D}_f^\top \int \boldsymbol{t}K(\boldsymbol{t})\,\mathrm{d}\boldsymbol{t} + \frac{1}{2}h_{\mathcal{F}}^2 \text{trace}\left[\mathcal{H}_f(\boldsymbol{z}) \int \boldsymbol{t}\boldsymbol{t}^\top K(\boldsymbol{t})\,\mathrm{d}\boldsymbol{t}\right] \\ &\quad - f(\boldsymbol{z}) + o(h_{\mathcal{F}}^2) \\ &\approx \frac{1}{2}h_{\mathcal{F}}^2 k_2 \text{trace}[\mathcal{H}_f(\boldsymbol{z})]. \end{aligned} \tag{2.6}$$

Much the same approach is used to find an asymptotic approximation to the variance:

$$\begin{aligned}
\mathrm{Var}[\hat{f}_{\mathrm{fix}}(\boldsymbol{z})] =& \frac{1}{nh_{\mathcal{F}}^4} \int K\left(\frac{\boldsymbol{z}-\boldsymbol{y}}{h_{\mathcal{F}}}\right)^2 f(\boldsymbol{y})\,\mathrm{d}\boldsymbol{y} - \frac{1}{n}\mathbb{E}[\hat{f}_{\mathrm{fix}}(\boldsymbol{z})]^2 \\
=& \frac{1}{nh_{\mathcal{F}}^2} \int K(\boldsymbol{t})^2 f(\boldsymbol{z}-h_{\mathcal{F}}\boldsymbol{t})\,\mathrm{d}\boldsymbol{t} - \frac{1}{n}\left(f(\boldsymbol{z}) + \mathrm{Bias}[\hat{f}_{\mathrm{fix}}(\boldsymbol{z})]\right)^2 \\
=& \frac{1}{nh_{\mathcal{F}}^2} \int K(\boldsymbol{t})^2 f(\boldsymbol{z})\,\mathrm{d}\boldsymbol{t} + o(n^{-1}h_{\mathcal{F}}^{-2}) \\
\approx& \frac{1}{nh_{\mathcal{F}}^2} R(K) f(\boldsymbol{z}),
\end{aligned} \tag{2.7}$$

where $R(K) = \int K(\boldsymbol{t})^2\,\mathrm{d}\boldsymbol{t}$.

The same properties for the adaptive kernel estimator are more difficult to find due to the pilot density estimation stage in calculating the variable bandwidths at each observation. Through what Silverman (1986) describes as "extremely tedious algebra", following use of an algebraic manipulation software package he gives

$$\mathrm{Bias}[\hat{f}_{\mathrm{adv}}(\boldsymbol{z})] = \frac{h_0^4}{24\gamma_f^4 f(\boldsymbol{z})} A_f(\boldsymbol{z}) k_4 + o(h_0^4) \tag{2.8}$$

where $A_f(\boldsymbol{z})$ is a bounded, well-defined, yet complicated function of f and its derivatives. Hall and Marron (1988) do provide some commentary on evaluation of this quantity by hand, paying particular attention to the univariate setting. They note however that this would require laborious repeated integration-by-parts and, given the limited practical interest of the explicit result, we do not pursue it further here.

We note that the bias of $\hat{f}_{\mathrm{adv}}$ is reduced from $O(h_{\mathcal{F}}^2)$ in the fixed setting to $O(h_0^4)$ in the adaptive. See Abramson (1982) and Hall and Marron (1988) for further theoretical details on this interesting fact. This is a primary motivator for exploring the adaptive kernel density-ratios in the following section.

An approximation to the variance of $\hat{f}_{\mathrm{adv}}$ can be found by drawing on results in Hall and Marron (1988). Firstly, we define the quantities

$$D_1 = \breve{f} - \mathbb{E}[\breve{f}] \qquad \text{and} \qquad D_2 = \hat{f}_{\mathrm{adv}} - \breve{f}$$

where $\breve{f}$ is the 'ideal' adaptive estimator; that is, implementation of (2.1) using (2.2) directly (no pilot estimation). Theorem 3.2 of Hall and Marron (1988) then states that the variables $(\zeta_1, \zeta_2) \leftarrow n^{1/2} h_0 \gamma_f^{-1} (D_1, D_2)$ follow a bivariate normal distribution with zero mean vector and

$$\mathrm{Var}[\zeta_1] \approx f^2 \int K(\boldsymbol{u})^2\,\mathrm{d}\boldsymbol{u}; \qquad \mathrm{Var}[\zeta_2] \approx \frac{1}{4} f^2 \int L(\boldsymbol{u})^2\,\mathrm{d}\boldsymbol{u}; \qquad \text{and}$$

$$\mathrm{Cov}[\zeta_1, \zeta_2] \approx \frac{1}{2} f^2 \int K(\boldsymbol{u}) L(\boldsymbol{u})\,\mathrm{d}\boldsymbol{u};$$

where $L(\boldsymbol{u}) = 2K(\boldsymbol{u}) + \boldsymbol{u}^\top \mathcal{D}_K(\boldsymbol{u})$. Assuming certain conditions on the kernel function, such as those in (2.4), symmetry, and that K is continuously differentiable (at least twice) throughout its support, Hall and Marron (1988) also note that in the 2-dimensional setting,

$\int KL = \int K^2$. Now:

$$\begin{aligned}
\hat{f}_{\text{adv}} =& \hat{f}_{\text{adv}} - \breve{f} + \breve{f} - \mathbb{E}[\breve{f}] + \mathbb{E}[\breve{f}] \\
=& D_1 + D_2 + \mathbb{E}[\breve{f}] \\
\Rightarrow \text{Var}[\hat{f}_{\text{adv}}] =& \text{Var}[D_1] + \text{Var}[D_2] + 2\text{Cov}[D_1, D_2] \\
\approx& \frac{\gamma_f^2}{nh_0^2} \left(\text{Var}[\zeta_1] + \text{Var}[\zeta_2] + 2\text{Cov}[\zeta_1, \zeta_2]\right) \\
\approx& \frac{\gamma_f^2}{nh_0^2} \left[f^2 \int K^2 + \frac{1}{4} f^2 \int L^2 + 2\left(\frac{1}{2} f^2 \int KL\right)\right].
\end{aligned}$$

So finally,

$$\text{Var}[\hat{f}_{\text{adv}}(\boldsymbol{z})] \approx \frac{\gamma_f^2}{nh_0^2} f(\boldsymbol{z})^2 \left[2 \int K(\boldsymbol{y})^2 \, \mathrm{d}\boldsymbol{y} + \frac{1}{4} \int L(\boldsymbol{y})^2 \, \mathrm{d}\boldsymbol{y} \right]. \tag{2.9}$$

We note in closing that these statistics for both the fixed and adaptive kernel estimators reflect the bias-variance trade-off ubiquitous in smoothing problems: the greater the amount of smoothing (the bandwidth) the lower the variance but higher the bias, and vice-versa.

2.2 The relative risk function

In this section we introduce the density-ratio estimator of so-called *relative risk*. This technique has become popular, particularly in epidemiological applications, owing to ease of implementation and interpretation. Due to the high spatial heterogeneity often observed in the target populations in these problems, kernel smoothing provides a high degree of flexibility in appropriately capturing this spatial variation. Specifically, the aim here is to extend the methodology to use adaptive kernel smoothing and study any advantages and/or disadvantages over the fixed-smoothing estimator; the only option considered in the external literature to date.

To avoid confusion, it is worth mentioning that in epidemiological circles the definition of 'relative risk' often refers to a ratio of infected entity values between groups according to their exposure to a certain risk factor (see for example the definitions in Dawson and Trapp, 2004). As we shall see below, however, the kernel-smoothed density-ratio is defined as a ratio of infected to *non-infected* density estimates; technically resulting in what epidemiologists would refer to as a location-wise *odds ratio*. The term 'relative risk' with respect to the kernel density-ratio was coined by the original authors of the relevant statistical papers (references follow) and has been used since; we will remain with the statistical terminology in the following discussions.

2.2.1 Definition

Consider now the presence of two point patterns, $\boldsymbol{X} = \{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_{n_1}\}$ and $\boldsymbol{Y} = \{\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_{n_2}\}$, both observed in some finite observation window $W \subset \mathbb{R}^2$. Assuming that $\boldsymbol{X}$ and $\boldsymbol{Y}$ have arisen from independent Poisson point processes with density functions f and

g respectively, the *relative risk function* (Bithell, 1990, 1991) is defined as

$$r(\boldsymbol{z}) = \frac{f(\boldsymbol{z})}{g(\boldsymbol{z})}, \qquad \boldsymbol{z} \in W. \tag{2.10}$$

Kelsall and Diggle (1995*a*) replace the numerator and denominator in (2.10) with kernel density estimates $\hat{f}_{\text{fix}}$ and $\hat{g}_{\text{fix}}$ respectively. In addition, the authors advocate the use of the log-risk function $\rho = \log[\text{r}]$ to symmetrise treatment of the two densities. We seek to investigate the use of $\hat{f}_{\text{adv}}$ and $\hat{g}_{\text{adv}}$ instead, in light of the potential bias reductions observed in the individual density estimates as per (2.8). Thus, we study and compare the relative risk function estimators

$$\hat{\rho}_{\text{fix}}(\boldsymbol{z}) = \log\left[\frac{\hat{f}_{\text{fix}}(\boldsymbol{z})}{\hat{g}_{\text{fix}}(\boldsymbol{z})}\right] \qquad \text{and} \qquad \hat{\rho}_{\text{adv}}(\boldsymbol{z}) = \log\left[\frac{\hat{f}_{\text{adv}}(\boldsymbol{z})}{\hat{g}_{\text{adv}}(\boldsymbol{z})}\right]. \tag{2.11}$$

These constructions are particularly useful in geographical epidemiology, as they allow the researchers to compare the spatial distributions of observed disease *cases* as $\boldsymbol{X}$ with the at-risk population $\boldsymbol{Y}$ (dubbed *controls*), both scaled by their respective sample sizes. Significant departures from uniformity can therefore point to an excess disease contraction 'risk', or lack thereof, in specific areas of W after taking into account the underlying distribution.

Generally regarded as an exploratory data analysis tool, the interpretation of these kernel-smoothed risk surfaces has important consequences in practice. Should significant sub-regions of risk be identified, this has the potential to drive important health policy decisions. For example, Kelsall and Diggle (1995*a*) investigated the relative risk of laryngeal cancer in relation to proximity to a disused industrial incinerator in an area of Lancashire, England. Prince et al. (2001) studied the variation in risk of primary biliary cirrhosis in northeastern England in an effort to better target health resources. Sanson et al. (2011) used the kernel-smoothed relative risk methodology to compare the performance of different models of foot-and-mouth disease prediction in Irish farms. Many other examples exist in the literature. It is therefore critical that we aim to achieve the best possible estimates of spatial relative risk in these applications such that any conclusions drawn from the analyses reflect as closely as possible the true variation across the study region.

2.2.2 Properties

As always, an understanding of the fundamental properties of any estimator is important. Using the asymptotic approximations discussed for the kernel estimator in Section 2.1.2, we may approximate the same properties for the log-relative risk functions in (2.11). For the fixed estimator, the bias and variance terms are already available (see Kelsall and Diggle, 1995*a*), and can be shown to be

$$\text{Bias}\left[\hat{\rho}_{\text{fix}}(\boldsymbol{z})\right] \approx \frac{k_2}{2}\left\{\frac{\text{trace}[\mathcal{H}_f(\boldsymbol{z})]h^2_{\mathcal{F}(1)}}{f(\boldsymbol{z})} - \frac{\text{trace}[\mathcal{H}_g(\boldsymbol{z})]h^2_{\mathcal{F}(2)}}{g(\boldsymbol{z})}\right\}, \tag{2.12}$$

and

$$\mathrm{Var}\left[\hat{\rho}_{\mathrm{fix}}(\boldsymbol{z})\right] \approx R(K)\left\{\frac{1}{n_1 h_{\mathcal{F}(1)}^2 f(\boldsymbol{z})} + \frac{1}{n_2 h_{\mathcal{F}(2)}^2 g(\boldsymbol{z})}\right\}, \tag{2.13}$$

where $h_{\mathcal{F}(1)}$ and $h_{\mathcal{F}(2)}$ denote the fixed bandwidths used for the case and control density estimates respectively.

We must now derive the same properties for the adaptive risk function, and do so following the same approach as that used to find (2.12) and (2.13) by first-order Taylor expansion of the log function. Let $\epsilon_\nu(\boldsymbol{x}) = [\hat{\nu}_{\mathrm{adv}}(\boldsymbol{x}) - \nu(\boldsymbol{x})]/\nu(\boldsymbol{x})$ for some density ν, and let $h_{0(1)}$ and $h_{0(2)}$ be the global bandwidths for the $\hat{f}_{\mathrm{adv}}$ and $\hat{g}_{\mathrm{adv}}$ respectively. Assuming that n_1 and n_2 are large (as required for the validity of the asymptotic expansions), these error terms will be small. Therefore,

$$\begin{aligned}
\hat{\rho}_{\mathrm{adv}}(\boldsymbol{z}) =& \log[\hat{f}(\boldsymbol{z})] - \log\left[\hat{g}(\boldsymbol{z})\right] \\
=& \log\left[f(\boldsymbol{z})\right] + \log\left[1 + \epsilon_f(\boldsymbol{z})\right] - \log\left[g(\boldsymbol{z})\right] - \log\left[1 + \epsilon_g(\boldsymbol{z})\right] \\
=& \rho(\boldsymbol{z}) + \epsilon_f(\boldsymbol{z}) - \epsilon_g(\boldsymbol{z}) + O_p(\epsilon^2)
\end{aligned}$$

where ϵ^2 represents squared epsilon terms regardless of density. Now, using (2.8) and ignoring the higher-order terms of $h_{0(1)}$ and $h_{0(2)}$,

$$\begin{aligned}
\mathrm{Bias}\left[\hat{\rho}_{\mathrm{adv}}(\boldsymbol{z})\right] = \mathbb{E}\left[\hat{\rho}_{\mathrm{adv}}(\boldsymbol{z})\right] - \rho(\boldsymbol{z}) \approx& \mathbb{E}[\epsilon_f(\boldsymbol{z})] - \mathbb{E}[\epsilon_g(\boldsymbol{z})] \\
\approx& \frac{\mathbb{E}[\hat{f}_{\mathrm{adv}}(\boldsymbol{z})] - f(\boldsymbol{z})}{f(\boldsymbol{z})} - \frac{\mathbb{E}[\hat{g}_{\mathrm{adv}}(\boldsymbol{z})] - g(\boldsymbol{z})}{g(\boldsymbol{z})} \\
\approx& \frac{k_4}{24}\left\{\frac{A_f(\boldsymbol{z}) h_{0(1)}^4}{f(\boldsymbol{z})^2 \gamma_f^4} - \frac{A_g(\boldsymbol{z}) h_{0(2)}^4}{g(\boldsymbol{z})^2 \gamma_g^4}\right\};
\end{aligned} \tag{2.14}$$

and similarly, using (2.9),

$$\begin{aligned}
\mathrm{Var}\left[\hat{\rho}_{\mathrm{adv}}(\boldsymbol{z})\right] \approx& \mathrm{Var}[\epsilon_f(\boldsymbol{z})] + \mathrm{Var}[\epsilon_g(\boldsymbol{z})] \\
\approx& \frac{\mathrm{Var}[\hat{f}_{\mathrm{adv}}(\boldsymbol{z})]}{f(\boldsymbol{z})^2} + \frac{\mathrm{Var}[\hat{g}_{\mathrm{adv}}(\boldsymbol{z})]}{g(\boldsymbol{z})^2} \\
\approx& S(K)\left[\frac{\gamma_f^2}{n_1 h_{0(1)}^2} + \frac{\gamma_g^2}{n_2 h_{0(2)}^2}\right]
\end{aligned} \tag{2.15}$$

where

$$S(K) = 2\int K(\boldsymbol{z})^2\,\mathrm{d}\boldsymbol{z} + \frac{1}{4}\int L(\boldsymbol{z})^2\,\mathrm{d}\boldsymbol{z}$$

and L is as defined for (2.9).

Two important points are immediately apparent. Firstly, consider the situation where we have either a common fixed bandwidth for case and control density estimates in $\hat{\rho}_{\mathrm{fix}}$; i.e. $h_{\mathcal{F}(1)} = h_{\mathcal{F}(2)} = h_{\mathcal{F}}$, or a common global bandwidth for the separate estimates in $\hat{\rho}_{\mathrm{adv}}$; i.e. $h_{0(1)} = h_{0(2)} = h_0$. Assume also that for the adaptive version we use a common scaling parameter $\gamma_f = \gamma_g = \gamma_\omega$. Then, in areas of W where $f \approxeq g$, we can expect the bias to diminish in both fixed and adaptive relative risk functions; see (2.12) and (2.14). While this was mentioned by Kelsall and Diggle (1995*a*) for the fixed version, it is useful to observe

much the same result for the adaptive estimator. The common bandwidth multiplier γ_ω is computed as a 'pooled' geometric mean term in the sense that case and control marks are ignored. Letting $\boldsymbol{P} = \boldsymbol{X} \cup \boldsymbol{Y}$ be the pooled data set, one easy option is to set $\gamma_\omega = \exp\{(n_1+n_2)^{-1} \sum_{\boldsymbol{u}\in\boldsymbol{P}} \log[\tilde{\omega}_{\tilde{h}_\omega}(\boldsymbol{u})^{-1/2}]\}$, where $\tilde{\omega}$ is the pooled pilot kernel density estimate with fixed bandwidth $\tilde{h}_\omega$. The γ_ω term would then replace γ_f and γ_g in the individual case/control kernel estimates, with (2.14) and (2.15) affected in an obvious fashion.

The second point arises upon inspection of the asymptotic variances in (2.13) and (2.15). By a convenient cancellation of squared density terms in the second line of (2.15), we note approximate variance stabilisation resulting for the adaptive risk function. The same cannot be said for the fixed version, where volatility exists due to explicit $\boldsymbol{z}$ terms in the right-hand side of (2.13). The fact that $\mathrm{Var}\,[\hat{\rho}_{\mathrm{fix}}]$ is made up of the terms $1/f(\boldsymbol{z})$ and $1/g(\boldsymbol{z})$ is particularly serious, as it implies potentially extreme variance in areas of sparse data.

2.3 Practical considerations

There are a number of key issues that must be addressed when implementing kernel-smoothed relative risk functions in practice. Here we aim to discuss sensible approaches to dealing with them, but it is important to realise that there will always be a certain limit to generalising techniques for approaching smoothing problems. Indeed, *ad hoc* solutions will to a certain extent remain in real-world applications; though it is helpful to be able to characterise and identify, where possible, known behaviour of the methodology in question.

2.3.1 Bandwidths

Though there exist several feasible approaches for calculating bandwidths (termed *bandwidth selectors*) based on some optimisation criterion for kernel density estimation, specific selectors for kernel relative risk functions *per se* are scarce. A thorough review of bandwidth selectors for density estimation is unsuitable for this chapter; though fundamental issues, particularly with respect to the risk function, are discussed here. The interested reader is directed to texts such as Wand and Jones (1995) for further investigation.

Assignment

Before we are able to consider the method of smoothing parameter calculation, *bandwidth assignment* must be decided upon. That is, in the fixed context we can elect to calculate two distinct values (one for the case density, the other for the control), or a single bandwidth, common to both. For implementation of $\hat{\rho}_{\mathrm{adv}}$, up to four distinct values can be chosen, as specification of pilot and global bandwidths for both case and control density estimates is required.

Outside of the established large-sample approximations to the bias and variance terms of the risk function, there is little to guide the user in deciding upon sensible bandwidth assignment. What is clear from Section 2.2.2 is that the potential bias cancellations arising from assigning a common bandwidth in the fixed setting, or a common global bandwidth in the adaptive, is too valuable to ignore.

We therefore proceed by imposing these conditions upon the bandwidth assignment. This leaves the issue of assigning pilot bandwidths when utilising adaptive smoothing. Being abstracted away from the final density estimate, it is natural to expect varying the pilots to have less of an impact than the globals. To date, there exists no evidence for or against choosing a common pilot bandwidth. Considering the global bandwidths are elected to be the same for cases and controls, it may be sensible to calculate the pilots separately for each group. It could be argued that these different pilot bandwidths can help preserve detail specific to each density estimate for subsequent calculation of the variable bandwidths. This intuitive reasoning is the basis for choosing, in later sections, separate pilots for the adaptive relative risk function.

Selection methods

The aim of bandwidth selectors for the kernel estimator is to use the observed data to yield a bandwidth which produces a kernel estimate as 'close' to the true density as possible. A useful and hence typically employed criterion for density estimation is to therefore minimise the *mean integrated squared error* (MISE), which for the risk function can be expressed as

$$\mathbb{E}\int \left[\hat{\rho}_{\text{fix}}(\boldsymbol{x};h_{\mathcal{F}})-\rho(\boldsymbol{x})\right]^2\,\mathrm{d}\boldsymbol{x} \qquad \text{and} \qquad \mathbb{E}\int \left[\hat{\rho}_{\text{adv}}(\boldsymbol{x};h_0)-\rho(\boldsymbol{x})\right]^2\,\mathrm{d}\boldsymbol{x}. \tag{2.16}$$

Note that under our bandwidth assignment conditions, we ideally would require an approach that will minimise these quantities for a common $h_{\mathcal{F}}$ and a common h_0 between case and control densities in the fixed and adaptive settings respectively.

One available method that can achieve this based on leave-one-out least-squares cross-validation (LSCV) was introduced by Kelsall and Diggle (1995*a*), and involves minimisation with respect to some bandwidth h of

$$\text{LSCV}_{\rho}(h) = -\int \left[\hat{\rho}(\boldsymbol{x};h)\right]^2\,\mathrm{d}\boldsymbol{x} - \frac{2}{n_1}\sum_{i=1}^{n_1}\left[\frac{\hat{\rho}^{-i}(\boldsymbol{x}_i;h)}{\hat{f}^{-i}(\boldsymbol{x}_i;h)}\right] + \frac{2}{n_2}\sum_{j=1}^{n_2}\left[\frac{\hat{\rho}^{-j}(\boldsymbol{y}_j;h)}{\hat{g}^{-j}(\boldsymbol{y}_j;h)}\right], \tag{2.17}$$

where $\hat{\chi}^{-i}$ denotes the kernel estimate of χ (where χ is either a density or risk function) constructed without the ith observation, $\hat{\rho}$ (and $\hat{f},\hat{g}$) can indicate either the fixed or adaptive estimators, and h either $h_{\mathcal{F}}$ or h_0 depending on which estimator we are employing. The authors also note that if the $\hat{\rho}$ and ρ terms are replaced by the density $\hat{\nu}$ and ν in (2.16), then (2.17) reduces to the MISE-minimising least-squares cross-validation method for a standalone kernel density estimate for corresponding data $\boldsymbol{m}_1,\boldsymbol{m}_2,\ldots,\boldsymbol{m}_s$, which we will similarly denote as LSCV_{ν}:

$$\text{LSCV}_{\nu}(h) = \int \hat{\nu}(\boldsymbol{x};h)^2\,\mathrm{d}\boldsymbol{x} - 2s^{-1}\sum_{i=1}^{s}\hat{\nu}^{-i}(\boldsymbol{m}_i;h). \tag{2.18}$$

A related approach specific to choosing a common bandwidth for the risk function estimator was discussed in Hazelton (2008). We do not consider this method in the current work, for several reasons. Firstly, it was designed in terms of the fixed-bandwidth risk function estimator only, cancelling identical edge-correction factors (see the following section)

in the ratio. This cannot be achieved in the adaptive setting, which is the primary interest in this chapter. Secondly, the selector was derived for use with the raw density ratio, not the log-risk function estimator as we will use here. Finally, Hazelton's (2008) selector is designed to minimise a MISE weighted by the control density, making direct comparison of it with approaches based on (2.16) less straightforward.

Least-squares cross-validation can be a somewhat troubling approach in the context of density estimation, however. Studies on the asymptotics of LSCV (2.18) showed a high level of variability in the estimated smoothing parameters (assuming certain regularity conditions – see Hall and Marron, 1987; Scott and Terrell, 1987). This variability can be mitigated by introducing additional bias in the objective function, leading to approaches such as *biased cross-validation* (BCV). However, both LSCV and BCV indicate rather poor rates of convergence, in the order of only $o(n^{-1/10})$, to the theoretically optimal bandwidth (i.e. that which minimises the theoretical expression for the MISE of a kernel density estimate). In addition, BCV is well-defined for individual densities, but appropriate analogues for $\hat{r}$ or $\hat{\rho}$ are not readily available. These findings are summarised clearly in Section 3.8 of Wand and Jones (1995).

Though based on sound theoretical measures of discrepancy of the estimated relative risk function with its true counterpart, Kelsall & Diggle's (and Hazelton's) 'jointly optimal' LSCV selectors appear to be extremely unreliable in practice. The problems arise due to the simple fact we are dealing with a ratio of two kernel estimates, and in the case of (2.17), this ratio is repeated in the last two terms on the right-hand side of the equation. Instability occurs most often when there exist 'sufficiently isolated' (in light of the bandwidth) observations in either or both of the case and control datasets. When the estimate is evaluated at these locations, small denominator terms can lead to explosions in the cross-validation objective function.

This, coupled with the natural variability in leave-one-out LSCV, means employing (2.17) in practice is rather difficult. Indeed, Kelsall and Diggle (1995*a*) themselves note that during their simulation study, the bandwidth chosen by LSCV_ρ would occasionally be so inappropriate as to "...give rise to an extremely high ISE". They go on to label these instances as "...a breakdown of the methodology". This author's own (unpublished) experiences (cf. Davies, 2007) agree with these statements, and note that even when a clear minimum in the CV objective function was identified, there would be a tendency for the corresponding bandwidth to produce severely undersmoothed estimates.

An illustrative example of these deficiencies is warranted. Consider the 62 cases of childhood leukaemia and lymphoma collected in the North Humberside region of the U.K., between 1974 and 1986. Along with 141 controls randomly selected from the birth register, these data were first presented and analysed by Cuzick and Edwards (1990). Figure 2.1 shows the cases and controls inside the geographical study region, as well as the LSCV_ρ objective function values for a fine sequence of bandwidths between 10 and 50, using fixed smoothing. These values are displayed on both the raw and log-scales to assist in visualisation due to the extreme variability of the results.

Immediately we note there is concerning instability for the smaller bandwidth values. This is due to the leave-one-out operations resulting in the aforementioned minuscule den-

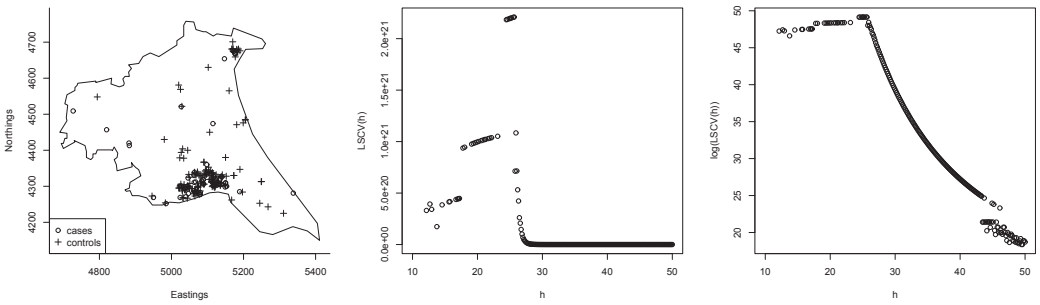


Figure 2.1: Humberside case-control observations (left), along with the LSCV_ρ function values for a sequence of bandwidths (center), and the same LSCV_ρ function values on the log-scale (right).

sity magnitude at isolated observations. Some of these values are missing, owing to the internal computations being unable to cope with the scale of the returned results. As the bandwidth is increased, stability improves. However, the LSCV_ρ results remain monotonic decreasing as the amount of smoothing becomes larger. An inspection of the function for smoothing parameters $>>$ 50 (not shown) further reveals no clear minimum, and as the bandwidth approaches excessively large values (i.e. completely unsuitable given the scale of the observations and area of the study region), issues with edge-correction (see the following section) re-introduce further instability and incomputable quantities.

While there are indeed examples where (2.17) will indicate a clear local minimum and thus 'successfully' select a bandwidth, the behaviour observed with the Humberside data is unfortunately not uncommon. Even when synthetically generated data are considered (as later in Section 2.5), the same problems often occur, making implementation of LSCV_ρ in a comprehensive simulation study rather difficult.

The difficulties discussed in terms of practical implementation of LSCV_ρ mean we do not consider this specific method further. As an alternative solution to the bandwidth selection problem, we may use one of the many selectors defined for single density estimation and simply apply it to the pooled case-control observations $\boldsymbol{P} = \boldsymbol{X} \cup \boldsymbol{Y}$. This opens up a wider range of options for data-driven bandwidth calculation, but the appropriateness of this 'shortcut' is questionable. For instance, if $n_1 << n_2$, then this pooling approach could result in a bandwidth based too heavily on the features of the control data. However, if care is taken, pooling can be effective and convenient. In spite of the drawbacks associated with LSCV methodology, note that LSCV_ν is more stable (in terms of identification of a clear local minimum) than LSCV_ρ, primarily due to the lack of division by leave-one-out density values. We therefore consider using the value of h_{CV} given by

$$h_{\text{CV},\omega} = \arg\min \text{LSCV}_\omega(h)$$

where

$$\text{LSCV}_\omega(h) = \int \hat{\omega}_{\text{fix}}(\boldsymbol{x}; h)^2 \, \mathrm{d}\boldsymbol{x} - 2(n_1 + n_2)^{-1} \sum_{\boldsymbol{x} \in \boldsymbol{P}} \hat{\omega}_{\text{fix}}^{-\boldsymbol{x}}(\boldsymbol{x}; h), \tag{2.19}$$

$\hat{\omega}$ is used to denote the pooled-data density estimate, and $\hat{\chi}^{-\boldsymbol{u}}$ represents the estimate $\hat{\chi}$ constructed without the observation $\boldsymbol{u}$. The selected h is then used as the common bandwidth in the estimates $\hat{f}$ and $\hat{g}$.

The susceptibility of leave-one-out CV to undersmoothing is still concerning. Let us consider once more the potential bias cancellations that drive the common bandwidth assignment decision – meaning we may be inclined to choose slightly larger bandwidths for risk function estimation in an effort to reduce variance. This makes it feasible to consider a bandwidth selector based on the *maximal smoothing principle* (Terrell, 1990). This bandwidth (also known as the *oversmoothing* bandwidth) is given by

$$h = U \times \left[\frac{(d+8)^{(d+6)/2} \pi^{d/2} R(K)}{16n\Gamma\left[(d+8)/2\right] d(d+2)} \right]^{\frac{1}{d+4}} \tag{2.20}$$

where d is the dimension of our data ($d = 2$ for our purposes), $R(K)$ is as defined for (2.7), and U is, in the formal definition, the standard deviation of our data (as we are smoothing isotropically, any estimate of U must be scalar). Though it is acknowledged that this approach can produce smoothing parameter estimates that are too large for standalone density estimation, the aforementioned variance control for the risk function estimator could be beneficial. By again using the pooled approach to bandwidth calculation, it is possible to implement this method for choosing the common $h_{\mathcal{F}}$ or h_0 values for $\hat{\rho}_{\text{fix}}$ or $\hat{\rho}_{\text{adv}}$ respectively. Therefore, we will also consider using (2.20) based on $\boldsymbol{P}$ as follows:

$$h_{\text{OS}} = U_{\text{IQR}} \left[\frac{625\pi R(K)}{192n^{\star}} \right]^{\frac{1}{6}}$$

which, when setting $R(K) = (4\pi)^{-1}$ for the Gaussian kernel, is

$$h_{\text{OS}} = U_{\text{IQR}} \left[\frac{625}{768n^{\star}} \right]^{\frac{1}{6}}, \tag{2.21}$$

where $n^{\star} = \sqrt{n_1 n_2}$ and the scaling value $U_{\text{IQR}} = [\text{IQR}(\boldsymbol{x}_{(1)}) + \text{IQR}(\boldsymbol{x}_{(2)})]/2.68$. Here, with $\boldsymbol{x}_{(1)}$ and $\boldsymbol{x}_{(2)}$ we denote the vectors of coordinates on the horizontal and vertical axes respectively, where $\boldsymbol{x} \in \boldsymbol{P}$. Thus, U_{IQR} is interpreted as the mean of the interquartile ranges on the two axes, divided by 1.34 (the factor which converts the interquartile range to a robust estimator of the standard deviation for normally distributed data). Finally we note that the use of $n^{\star}$ rather than $(n_1 + n_2)$ in (2.21) owes to the earlier comments of caution in the event of noticeable differences between the case and control sample sizes when using $\boldsymbol{P}$ for bandwidth calculation. To prevent the oversmoothing (OS) bandwidths from being too heavily influenced by the larger sample size (typically the controls in epidemiological applications) the geometric mean of the two sample sizes, $n^{\star}$, is used in place of the pooled total. This gears the 'effective' sample size to lie in between n_1 and n_2 when $n_1 \neq n_2$, with $n^{\star} - \min(n_1, n_2) < \max(n_1, n_2) - n^{\star}$.

These bandwidth selectors are considered in choosing a common smoothing parameter for estimation of the relative risk function later in this chapter, where assignment specifics using OS and the various forms of single-density LSCV are mentioned. It remains a fact

that more sophisticated methods for selection of a jointly optimal bandwidth for the density-ratio are extremely difficult to identify. One possible option is to pursue Bayesian ideas, though historically these methods do not perform very well in KDE problems (Brewer, 1998, 2000). Alternative pursuits may involve modifications to the minimisation criterion, such as a consideration of different weighted mean integrated squared errors (WMISE) as in Hazelton (2008), or the possibility of investigating asymptotic mean integrated squared error (AMISE) expressions. However, and as we will see in Section 2.5, it seems the OS bandwidth selector does perform competitively when compared to the more formal LSCV methods in practice.

2.3.2 Edge-correction

Up until this point, we have conveniently ignored the fact that $\boldsymbol{X}$ and $\boldsymbol{Y}$ have both been observed within the finite window W. This fact, however, has important consequences for kernel estimation of the densities on W, as it is possible there exist points just outside the region that were simply not collected. Consider the presence of points near the boundary of the region. Then, dependent on the bandwidth(s), it is likely that a portion of the kernel weight assigned to these observations will fall outside W, resulting in potentially severe negative bias around the edges. To remedy this issue, we employ an *edge-correction* technique which reduces this bias to an asymptotically negligible level.

A relatively simple fix, this methodology was examined for the fixed bandwidth setting in Diggle (1985) and Kelsall and Diggle (1995*a*), and in the adaptive setting by Marshall and Hazelton (2010). These methods rescale the density estimates by division at each coordinate $\boldsymbol{z} \in W$ by

$$q_{h(\boldsymbol{z})}(\boldsymbol{z}) = \int_W \frac{1}{h(\boldsymbol{z})^2} K\left(\frac{\boldsymbol{x}-\boldsymbol{z}}{h(\boldsymbol{z})}\right) \mathrm{d}\boldsymbol{x}. \tag{2.22}$$

It is easily recognised that (2.22) is simply the proportion of the appropriate bandwidth-scaled kernel at location $\boldsymbol{z}$ that falls within W. Strictly speaking, this means the correction need only be used for locations close to the boundaries, as $q_{h(\boldsymbol{z})} \approxeq 1$ for interior locations. Recall also that $h(\boldsymbol{z}) \equiv h_{\mathcal{F}}$ in the fixed version.

To demonstrate how this correction factor achieves the desired result, consider the third line of equation (2.6) where we derive the asymptotic bias for $\hat{f}_{\text{fix}}$ using a multivariate Taylor expansion. With respect to W, this line would now read

$$\begin{aligned}\text{Bias}[\hat{f}_{\text{fix}}(\boldsymbol{z})] =& f(\boldsymbol{z}) \int_W K(\boldsymbol{t})\,\mathrm{d}\boldsymbol{t} - h_{\mathcal{F}} \mathcal{D}_f^{\top} \int_W \boldsymbol{t} K(\boldsymbol{t})\,\mathrm{d}\boldsymbol{t} + \frac{1}{2} h_{\mathcal{F}}^2 \text{trace}\left[\mathcal{H}_f(\boldsymbol{z}) \int_W \boldsymbol{t}\boldsymbol{t}^{\top} K(\boldsymbol{t})\,\mathrm{d}\boldsymbol{t}\right] \\ & - f(\boldsymbol{z}) + o(h_{\mathcal{F}}^2).\end{aligned} \tag{2.23}$$

Due to the change of variable $\boldsymbol{y} = \boldsymbol{z} - h_{\mathcal{F}}\boldsymbol{t}$ and the finiteness of W, it is now not possible to use the conditions in (2.4) for all $\boldsymbol{z}$, the first of which sets the leading integral in the right-hand side of (2.23) to one, allowing the cancellation of the $f(\boldsymbol{z})$ terms. Division by (2.22) normalises this integral, rendering the bias once again asymptotically manageable. Much the same operations occur in the adaptive setting. Of course, the division of the other terms in these Taylor expansions by $q_{h(\boldsymbol{z})}$ means the bias is still larger (that is, of a

lower order) at boundary locations than in the interior. Further bias elimination is possible with more complicated edge-correction techniques based on implementing special classes of *boundary kernels*. However, numerical results in Marshall and Hazelton (2010) showed this methodology to have a minimal impact in practice at best.

Understandably, the variance terms are also affected by edge-correction. This is important, as it will in turn affect the derivation of the theoretical risk function tolerance contours in Section 2.4. Recall again that as per our bandwidth assignment we assume common fixed $h_{\mathcal{F}}$ and global h_0 bandwidths for the fixed and adaptive settings respectively. For the variance of $\hat{\rho}_{\text{fix}}$ in (2.13), we must replace $R(K)$ with

$$R_{W,\boldsymbol{z}}(K) = \frac{1}{h_{\mathcal{F}}^2 q_{h_{\mathcal{F}}}(\boldsymbol{z})^2} \int_W K\left(\frac{\boldsymbol{x}-\boldsymbol{z}}{h_{\mathcal{F}}}\right)^2 \mathrm{d}\boldsymbol{x}, \tag{2.24}$$

such that

$$\text{Var}\left[\hat{\rho}_{\text{fix}}(\boldsymbol{z})\right] \approx R_{W,\boldsymbol{z}}(K) h_{\mathcal{F}}^{-2} \left\{\frac{1}{n_1 f(\boldsymbol{z})} + \frac{1}{n_2 g(\boldsymbol{z})}\right\}. \tag{2.25}$$

Correcting the variance of $\hat{\rho}_{\text{adv}}$ is slightly more complicated as we must now account for the different sets of variable bandwidths resulting from the separate case and control datasets. To that end, the uncorrected variance in equation (2.15) changes to

$$\text{Var}\left[\hat{\rho}_{\text{adv},q}(\boldsymbol{z})\right] \approx \frac{\gamma_\omega^2}{h_0^2} \left\{\frac{S^{\tilde{f}}_{W,\boldsymbol{z}}(K)}{n_1} + \frac{S^{\tilde{g}}_{W,\boldsymbol{z}}(K)}{n_2}\right\} \tag{2.26}$$

where we assume the presence of the pooled geometric mean term γ_ω, and

$$S^{\tilde{\nu}}_{W,\boldsymbol{z}}(K) = \left[\frac{1}{h(\boldsymbol{z};\tilde{\nu}) q_{h(\boldsymbol{z};\tilde{\nu})}(\boldsymbol{z})}\right]^2 \times \left[2\int_W K\left(\frac{\boldsymbol{x}-\boldsymbol{z}}{h(\boldsymbol{z};\tilde{\nu})}\right)^2 \mathrm{d}\boldsymbol{x} + \frac{1}{4}\int_W L\left(\frac{\boldsymbol{x}-\boldsymbol{z}}{h(\boldsymbol{z};\tilde{\nu})}\right)^2 \mathrm{d}\boldsymbol{x}\right] \tag{2.27}$$

for a given pilot density $\tilde{\nu}$. It is also important to note that edge-correction must be employed at all stages of estimation, including the pilot density when smoothing adaptively. This is to prevent the variable bandwidths resulting from equation (2.3) being too heavily influenced by boundary bias in the pilot estimate.

Indeed, any operations involving a kernel density estimate where the observations are restricted to a finite region should be edge-corrected. This includes the LSCV bandwidth selection methods detailed in Section 2.3.1. Numerical experience with these selectors indicates that not correcting for boundary bias can often lead to very different suggestions of the optimal bandwidth.

Henceforth it is assumed without further mention that all fixed and adaptive density and therefore risk function estimates, including those appearing in bandwidth selectors, have been edge-corrected as per (2.22).

2.3.3 Kernel function

It is widely acknowledged that the choice of kernel function K is very much a secondary matter in these problems, with a relatively minimal impact upon the performance of the

kernel estimator itself (e.g. Silverman, 1986; Wand and Jones, 1995; Bowman and Azzalini, 1997). Of particular interest are the investigations in Wand and Jones (1995) which clearly demonstrate the small differences in estimation efficiency of substantially different kernel shapes.

Due to satisfaction of the conditions (2.4) for convenient expression of the asymptotic properties, it is generally recommended to let K be a radially symmetric probability density function with the mode being centered at the origin. This also ensures the resulting estimate will itself be a bona-fide density. Several options exist in this case. In the analysis of spatial point patterns (i.e. with $d = 2$), the most common choices in practice tend to be the Gaussian K_{G}, Epanechnikov K_{E}, and biweight K_{B} functions, with

$$K_{\mathrm{G}}(\boldsymbol{u}) = (2\pi)^{-1}\exp\left[-\frac{1}{2}\boldsymbol{u}^{\top}\boldsymbol{u}\right],$$

$$K_{\mathrm{E}}(\boldsymbol{u}) = \begin{cases} 2\pi^{-1}\left(1-\boldsymbol{u}^{\top}\boldsymbol{u}\right) & \text{if } \boldsymbol{u}^{\top}\boldsymbol{u} < 1 \\ 0 & \text{otherwise} \end{cases}, \quad \text{and}$$

$$K_{\mathrm{B}}(\boldsymbol{u}) = \begin{cases} 3\pi^{-1}\left(1-\boldsymbol{u}^{\top}\boldsymbol{u}\right)^2 & \text{if } \boldsymbol{u}^{\top}\boldsymbol{u} < 1 \\ 0 & \text{otherwise} \end{cases}.$$

There do exist certain theoretical benefits, such as exact calculation of edge-correction factors, for kernels with compact support like K_{E} and K_{B}. Nevertheless, the aforementioned texts advocate choosing the kernel function based on reasons other than mathematical performance of the estimator, such as computational convenience or even aesthetics, due to the almost indistinguishable differences in estimate quality between different Ks.

For this reason, many researchers have opted for K_{G} in applications, particularly in geographical epidemiology. In addition to satisfying the conditions on the kernel function necessary for validation of the theoretical properties of the various estimators examined in Sections 2.1 and 2.2, the Gaussian kernel has the appealing property of unbounded support which is useful in areas of W where data are sparse. This ensures that the final density estimate will be defined everywhere in W, where a bounded kernel can easily result in an estimate that is discontinuous within the study region. As well as the perhaps minor issue of being decidedly unaesthetic, a discontinuous estimate can more importantly result in obvious problems involving subsequent inference to do with the density. Indeed, mere estimation of a relative risk function where the case and control densities are undefined (and which could well be undefined in *different* subregions of W if smoothing is adaptive) would prove troublesome. We therefore consider only the Gaussian kernel in this chapter (namely the implementations in Sections 2.5 and 2.6) as well as the collaborative applications in Sections 5.1-5.3.

2.3.4 Trimming

The final issue that must be considered applies only to the adaptive setting, in that the bandwidths given by equation (2.3) require some minor, yet important, adjustments post-calculation. Hall et al. (1995) show that the significant bias reductions possible with this smoothing regimen can be almost entirely eliminated should the density estimate at any given location be influenced too heavily by isolated data points. At these locations, the pilot density can be very small and our λ values (hence our $h(\cdot)$ values) become unacceptably large, creating a problem we will refer to as *bandwidth inflation*. In addition to the major concern of losing any advantageous drops in estimator bias (when compared to the fixed setting), bandwidth inflation can have obvious adverse effects on the estimated density if excessively large bandwidths lead to individual kernels spanning the entire finite study region and beyond. These detrimental impacts will flow directly through to any subsequent relative risk function estimate.

Hall et al. (1995) demonstrate that the attractive asymptotics which the adaptive estimator is capable of possessing can still be achieved with a relatively simple fix. This involves specifying some maximum threshold with respect to the magnitude of the calculated variable bandwidths, also known as *bandwidth trimming*. Note that the nature of this trimming criterion can be defined in a number of different ways; some alternative suggestions can be found in the work referenced above and appropriate references therein. Importantly, Hall et al. (1995) indicate that, provided any 'sensible' method of trimming is employed (i.e. one that does not too greatly restrict the maximum allowable bandwidth), the beneficial asymptotic bias levels of the adaptive estimator will still be realised.

To eliminate any of these unreasonably large bandwidths, we define $h_M = \text{median}(\boldsymbol{h})$ and use the following criterion when adaptively estimating some density ν from a sample of size n:

$$h(\cdot;\tilde{\nu}) \leftarrow \begin{cases} h(\cdot;\tilde{\nu}) & \text{if } h(\cdot;\tilde{\nu}) < \beta h_M \\ \beta h_M & \text{otherwise} \end{cases} \tag{2.28}$$

where $\boldsymbol{h}$ represents the $n \times 1$ vector of variable bandwidths originally resulting from (2.3). Specifically, we use some constant β times the median of the untrimmed adaptive bandwidths (at the locations of each observation) to disallow values greater than that quantity. Use of a robust statistic in such a trimming technique is intuitively sensible given it is extreme values that we are attempting to avoid. As β is increased, we move toward the completely untrimmed estimator, and the adverse effects of bandwidth inflation return. If β is too small, the theoretical benefits of the adaptive smoothing scheme cannot be realised due to excessive restriction of the larger bandwidths required for isolated observations. As a general rule-of-thumb, this author's experience leads to the recommendation of setting $4 < \beta < 8$.

In practice, we have found working with $\beta = 5$ still allows for generously sized values in our set of adaptive bandwidths but limits any adverse effects due to bandwidth inflation. Therefore, as well as assuming edge-correction, we will also assume without further mention that this trimming criterion (2.28) with $\beta = 5$ has been employed for all subsequent adaptive density and relative risk function estimates.

2.4 Asymptotic tolerance contours

It is often of interest to be able to assess the statistical significance of peaks and/or troughs in an estimated risk function. In other words, we wish to highlight subregions of W where we have observed abnormally large (or small) risk estimates after adjusting for the underlying population dispersion. Formally, the appropriate hypotheses in testing for statistically significant risk elevation would be expressed as

$$\begin{array}{lrcll} \mathrm{H}_0\colon & \rho(\boldsymbol{z}) & = & 0 & \\ \mathrm{H}_\mathrm{A}\colon & \rho(\boldsymbol{z}) & > & 0; & \boldsymbol{z} \in W, \end{array} \tag{2.29}$$

where tests for unusually low risk or double-sided tests would be expressed with '$<$' and '$\neq$' in H_A respectively.

Naturally, these *anomalies* can only be identified if we understand the variation possible in the fixed or adaptive risk function estimate $\hat{\rho}$. In turn, a corresponding map of p-values with respect to the hypotheses can be calculated over W, which the researcher may then use to superimpose upon an image of $\hat{\rho}$ so-called *tolerance contours* at pre-defined significance levels α.

Until now, the only existing approach to calculating the p-value surface was via Monte-Carlo (MC) methods (Kelsall and Diggle, 1995*a*). First, the case and control location data are pooled, then n_1 points are resampled without replacement to represent the simulated cases and the remaining n_2 points used as the simulated controls. The fixed and adaptive kernel estimators are then applied on these simulated datasets for each of N iterations providing simulated log-risk values $\{\hat{\rho}^*_1(\boldsymbol{z}), \hat{\rho}^*_2(\boldsymbol{z}), \ldots, \hat{\rho}^*_N(\boldsymbol{z})\}$. For each point $\boldsymbol{z}$, we therefore have the originally observed value and N simulated, so the p-value based on significantly elevated risk at location $\boldsymbol{z}$ is obtained by the expression $[1 + \sum_i \mathbf{1}(\hat{\rho}(\boldsymbol{z}) \leq \hat{\rho}^*_i(\boldsymbol{z}))]/(N+1)$, where $\mathbf{1}(\cdot)$ is the indicator function. The MC approach is easy to understand and implement. However there is an obvious computational cost that can be excessive, prohibitively so for the adaptive smoothing methodology. This motivates the need for an approach to p-value surface (and hence tolerance contour) calculation that is more accessible in both fixed and adaptive settings.

Fundamental theory of kernel density estimation (e.g. Parzen, 1962) states that given $\boldsymbol{z}$ the density estimate $\hat{\nu}(\boldsymbol{z})$ (fixed or adaptive) of target ν will be asymptotically normally distributed, provided certain natural assumptions hold about the elected bandwidth(s) (i.e. positivity, finiteness, tending to zero as the sample size increases). Given a sufficiently large sample size, this means we are able to make use of the approximations to the variances of the fixed and adaptive (log) relative risk functions to define a normally distributed test statistic at location $\boldsymbol{z}$, leading naturally to a p-value surface over W (a move supported in the adaptive setting by findings in Hall and Marron, 1988). Not requiring multiple evaluations of the kernel-smoothed density-ratio, this asymptotic approach (abbreviated to ASY) to constructing the desired tolerance contours can be expected to require only a fraction of the computation time of the MC contours. Thus, a feasible option for highlighting statistically significant fluctuations in the adaptive risk function is now available.

Straightforwardly, we standardise the estimated risk functions in order to obtain the field

of test statistics. Specifically, for both fixed and adaptive estimators, the test statistic $\mathcal{Z}$ is found by

$$\mathcal{Z}(\boldsymbol{z}) = \frac{\hat{\rho}(\boldsymbol{z}) - 0}{\sqrt{\mathrm{Var}[\hat{\rho}(\boldsymbol{z})]}}.$$

In the fixed-bandwidth setting this means

$$\mathcal{Z}_{\mathrm{fix}}(\boldsymbol{z}) = \frac{\hat{\rho}_{\mathrm{fix}}(\boldsymbol{z})}{R_{W,\boldsymbol{z}}(K)^{1/2} h_{\mathcal{F}}^{-1} \left\{ [n_1 f(\boldsymbol{z})]^{-1} + [n_2 g(\boldsymbol{z})]^{-1} \right\}^{1/2}}. \tag{2.30}$$

Under the null hypothesis in (2.29), the constant risk function indicates identical case and control densities, leading to a natural cancellation of the leading bias terms. By approximating the unknown terms in the denominator of (2.30) with a kernel density estimate $\hat{\omega}_{\mathrm{fix}}$ based on the pooled case-control data $\boldsymbol{P}$, the test statistics can be evaluated in practice with

$$\mathcal{Z}_{\mathrm{fix}}(\boldsymbol{z}) = \frac{\hat{\rho}_{\mathrm{fix}}(\boldsymbol{z}) h_{\mathcal{F}} \hat{\omega}_{\mathrm{fix}}^{1/2}(\boldsymbol{z})}{R_{W,\boldsymbol{z}}(K)^{1/2} \left\{ n_1^{-1} + n_2^{-1} \right\}^{1/2}}. \tag{2.31}$$

Here, we assume for simplicity that the pooled kernel estimate $\hat{\omega}_{\mathrm{fix}}$ is constructed with the same fixed bandwidth as for the individual case and control density estimates making up $\hat{\rho}_{\mathrm{fix}}$, namely $h_{\mathcal{F}}$. There exists however no strict reason why this must be so; indeed one may even use an adaptive kernel estimate for this component.

This is extended for the adaptive estimator in the same way as above. The corresponding test statistics, assuming common global bandwidths and geometric mean scalers (h_0 and γ_ω respectively), avoid the problem of having to compute an additional, pooled density estimate due to the variance stabilisation noted in (2.15). The same standardisation yields

$$\mathcal{Z}_{\mathrm{adv}}(\boldsymbol{z}) = \frac{\hat{\rho}_{\mathrm{adv}}(\boldsymbol{z}) h_0}{\gamma_\omega \left\{ S_{W,\boldsymbol{z}}^{\tilde{f}}(K) n_1^{-1} + S_{W,\boldsymbol{z}}^{\tilde{g}}(K) n_2^{-1} \right\}^{1/2}}. \tag{2.32}$$

Assuming sufficiently large sample sizes, $\mathcal{Z}_{\mathrm{fix}}(\boldsymbol{z}) \dot{\sim} \mathrm{N}(0,1)$ and $\mathcal{Z}_{\mathrm{adv}}(\boldsymbol{z}) \dot{\sim} \mathrm{N}(0,1)$. We therefore obtain our p-value surface map on $\boldsymbol{z} \in W$ by the standard normal quantiles.

This novel method of p-value surface and tolerance contour calculation is examined in practice in Section 2.6.

2.5 Simulation study

However theoretically appealing the adaptive approach to estimation of a risk function may be, it is possible that there are practical drawbacks to the method, such as excessive noise introduced at the pilot estimation stage. For this reason, it is of interest to compare the widely employed fixed bandwidth technique with the adaptive in settings where the true risk functions are known, and measure discrepancies between these and kernel estimates thereof based on correspondingly generated data.

We compare the performance of the fixed and adaptive kernel relative risk estimators in ten synthetic problem scenarios, all defined on the unit square such that $W = [0,1] \times$

$[0, 1]$. For each scenario, the 'true' control density g is defined in terms of normal mixtures, and scaled to integrate to unity over W. A corresponding 'true' relative risk surface r is then defined using a particular formulation involving exponential functions. Additional scaling takes place for r to ensure $\int_W g(\boldsymbol{u})r(\boldsymbol{u})\,\mathrm{d}\boldsymbol{u} = 1$. At each iteration, n_2 controls are sampled randomly from g. Case 'candidates' are sampled from g and accepted/rejected with probabilities proportional to r, until the desired number n_1 is reached. Figures 2.2 and 2.3 display filled contour plots of the control densities and corresponding log-risk functions for Problems 1 to 5 and Problems 6 to 10 respectively; the specific formulations are provided in Table 2.1. In this table, $\phi_W(a, b\boldsymbol{I})$ represents the bivariate Gaussian density edge-corrected to W, centered at a and with standard deviation b ($\boldsymbol{I}$ refers to the 2×2 identity matrix).

The problem scenarios have been designed in an effort to loosely reflect what we may be likely to observe in epidemiological applications in practice and are ordered in terms of the maximum level of relative risk, least to most extreme. Problem 1 is considered a 'baseline' situation: both the control density and risk function are perfectly uniform; raw risk range is [1, 1]. Problem 2 indicates a single risk mode on the northern border, arising from a wider dispersion of controls; raw risk ranging from 0.95 to 1.50. In Problem 3 we observe a mildly bimodal yet relatively diffuse dispersion of raw risk ranging from 0.42 to 1.67. From a multimodal control density a smaller, sharper increase in risk (raw interval of [0.86, 2.22]) contained within a diffuse risk 'hotspot' is visible in Problem 4. Problem 5 has an underlying control population characterised by an elongated 'bump' and a smaller circular density on the upper border. There exists a small, sharp peak in risk within the latter area, with the raw risk ranging from a low 0.35 to 2.88. The control density of Problem 6 is a single mode in the center of W, with the risk surface exhibiting three smaller modes of equal size located in the tail regions of g; risk interval of [0.61, 3.66]. Problem 7 exhibits several areas of point groupings surrounding a larger off-central mode and a corresponding risk interval of [0.94, 3.92]. A wide, dispersed mode along the upper border contains three concentrated risk peaks in close proximity for Problem 8 with a raw risk range of [0.92, 4.23]. Three even higher, yet more separated hotspots of risk are observed in Problem 9; raw risk interval [0.93, 5.99]. The final problem, Problem 10, is designed in contrast to Problem 1 to have an extreme maximum risk value. The control density remains uniform, but the risk surface now includes a tight peak in the center of W; risk interval is [0.94, 9.45]. Though less likely to occur in practice, the results of this scenario will nevertheless be interesting in comparison to the others.

Based on each of 100 iterations for each problem, the methodology is examined using two fixed-bandwidth and one adaptive-bandwidth kernel estimator of the relative risk function. Two sample sizes are used: $n_1 = 100, n_2 = 100$ and $n_1 = 500, n_2 = 1000$. The first estimator we label FIXCV, with the common fixed bandwidth computed using the LSCV single-density criterion on the pooled data set. That is, $h_{\mathcal{F}} = h_{\mathrm{CV},\omega}$ based on the criterion (2.19). The second, FIXOS, has the bandwidth fixed at the particular oversmoothing criterion discussed in Section 2.3.1, again based on the pooled case-control observations: $h_{\mathcal{F}} = h_{\mathrm{OS}}$ as detailed by equation (2.21). The adaptive estimates (labelled ADAPT) arise from distinct pilot bandwidths based on the case and control data separately, using the standalone-density LSCV approach. Specifically, $\tilde{h}_f = h_{\mathrm{CV},f}$ and $\tilde{h}_g = h_{\mathrm{CV},g}$, using (2.18). The common global bandwidth is set equal to the pooled-data calculated OS bandwidth i.e. $h_0 = h_{\mathrm{OS}}$.

Problem	Target definition
1	$g(\boldsymbol{u}) = 1$ $r(\boldsymbol{u}) = 1$
2	$g(\boldsymbol{u}) = 0.1 + 0.3\phi_W(\boldsymbol{u} - [0.3, 0.9]^\top, 0.2\boldsymbol{I}) + 0.3\phi_W(\boldsymbol{u} - [0.5, 0.8]^\top, 0.2\boldsymbol{I})$ $+0.3\phi_W(\boldsymbol{u} - [0.7, 0.6]^\top, 0.2\boldsymbol{I})$ $r(\boldsymbol{u}) \propto 1 + 0.2\exp\left(-100\|\boldsymbol{u} - [0.4, 0.9]^\top\|^2\right) + 0.5\exp\left(-100\|\boldsymbol{u} - [0.3, 0.9]^\top\|^2\right)$
3	$g(\boldsymbol{u}) = 0.1 + 0.3\phi_W(\boldsymbol{u} - [0.5, 0.5]^\top, 0.1\boldsymbol{I}) + 0.3\phi_W(\boldsymbol{u} - [0.7, 0.6]^\top, 0.1\boldsymbol{I})$ $+0.3\phi_W(\boldsymbol{u} - [0.8, 0.9]^\top, 0.2\boldsymbol{I})$ $r(\boldsymbol{u}) \propto 0.1 + 0.2\exp\left(-50\|\boldsymbol{u} - [0.4, 0.5]^\top\|^2\right) + 0.3\exp\left(-25\|\boldsymbol{u} - [0.7, 0.6]^\top\|^2\right)$
4	$g(\boldsymbol{u}) = 0.1 + 0.2\phi_W(\boldsymbol{u} - [0.5, 0.5]^\top, 0.1\boldsymbol{I}) + 0.2\phi_W(\boldsymbol{u} - [0.5, 0.7]^\top, 0.1\boldsymbol{I})$ $+0.2\phi_W(\boldsymbol{u} - [0.2, 0.7]^\top, 0.1\boldsymbol{I}) + 0.3\phi_W(\boldsymbol{u} - [0.8, 0.2]^\top, 0.2\boldsymbol{I})$ $r(\boldsymbol{u}) \propto 1 + \exp\left(-50\|\boldsymbol{u} - [0.4, 0.5]^\top\|^2\right) + \exp\left(-500\|\boldsymbol{u} - [0.5, 0.5]^\top\|^2\right)$
5	$g(\boldsymbol{u}) = 0.1 + 0.1\phi_W(\boldsymbol{u} - [0.1, 0.1]^\top, 0.1\boldsymbol{I}) + 0.2\phi_W(\boldsymbol{u} - [0.2, 0.2]^\top, 0.1\boldsymbol{I})$ $+0.2\phi_W(\boldsymbol{u} - [0.3, 0.3]^\top, 0.1\boldsymbol{I}) + 0.2\phi_W(\boldsymbol{u} - [0.4, 0.4]^\top, 0.1\boldsymbol{I})$ $+0.2\phi_W(\boldsymbol{u} - [0.5, 0.9]^\top, 0.1\boldsymbol{I})$ $r(\boldsymbol{u}) \propto 0.2 + \exp\left(-50\|\boldsymbol{u} - [0.1, 0.1]^\top\|^2\right) + \exp\left(-50\|\boldsymbol{u} - [0.2, 0.3]^\top\|^2\right)$ $+1.75\exp\left(-1000\|\boldsymbol{u} - [0.6, 0.9]^\top\|^2\right)$
6	$g(\boldsymbol{u}) = 0.1 + 0.9\phi_W(\boldsymbol{u} - [0.5, 0.5]^\top, 0.2\boldsymbol{I})$ $r(\boldsymbol{u}) \propto 0.2 + \exp\left(-62.5\|\boldsymbol{u} - [0.2, 0.7]^\top\|^2\right) + \exp\left(-50\|\boldsymbol{u} - [0.8, 0.2]^\top\|^2\right)$ $+\exp\left(-62.5\|\boldsymbol{u} - [0.7, 0.8]^\top\|^2\right)$
7	$g(\boldsymbol{u}) = 0.1 + 0.5\phi_W(\boldsymbol{u} - [0.6, 0.7]^\top, 0.2\boldsymbol{I}) + 0.1\phi_W(\boldsymbol{u} - [0.2, 0.9]^\top, 0.1\boldsymbol{I})$ $+0.1\phi_W(\boldsymbol{u} - [0.1, 0.6]^\top, 0.1\boldsymbol{I}) + 0.1\phi_W(\boldsymbol{u} - [0.3, 0.3]^\top, 0.1\boldsymbol{I})$ $+0.1\phi_W(\boldsymbol{u} - [0.8, 0.2]^\top, 0.1\boldsymbol{I})$ $r(\boldsymbol{u}) \propto 0.2 + 0.1\exp\left(-500\|\boldsymbol{u} - [0.2, 0.9]^\top\|^2\right) + 0.7\exp\left(-500\|\boldsymbol{u} - [0.1, 0.6]^\top\|^2\right)$ $+0.2\exp\left(-500\|\boldsymbol{u} - [0.3, 0.3]^\top\|^2\right) + 0.1\exp\left(-500\|\boldsymbol{u} - [0.8, 0.2]^\top\|^2\right)$
8	$g(\boldsymbol{u}) = 0.1 + 0.1\phi_W(\boldsymbol{u} - [0.1, 0.9]^\top, 0.2\boldsymbol{I}) + 0.1\phi_W(\boldsymbol{u} - [0.3, 0.9]^\top, 0.2\boldsymbol{I})$ $+0.2\phi_W(\boldsymbol{u} - [0.5, 0.8]^\top, 0.2\boldsymbol{I}) + 0.1\phi_W(\boldsymbol{u} - [0.7, 0.9]^\top, 0.2\boldsymbol{I})$ $+0.1\phi_W(\boldsymbol{u} - [0.9, 0.9]^\top, 0.2\boldsymbol{I}) + 0.3\phi_W(\boldsymbol{u} - [0.5, 0.1]^\top, 0.1\boldsymbol{I})$ $r(\boldsymbol{u}) \propto 1 + 2\exp\left(-500\|\boldsymbol{u} - [0.2, 0.9]^\top\|^2\right) + 3\exp\left(-500\|\boldsymbol{u} - [0.1, 0.6]^\top\|^2\right)$ $+4\exp\left(-500\|\boldsymbol{u} - [0.3, 0.3]^\top\|^2\right)$
9	$g(\boldsymbol{u}) = 0.1 + 0.2\phi_W(\boldsymbol{u} - [0.1, 0.5]^\top, 0.1\boldsymbol{I}) + 0.2\phi_W(\boldsymbol{u} - [0.3, 0.5]^\top, 0.1\boldsymbol{I})$ $+0.5\phi_W(\boldsymbol{u} - [0.7, 0.6]^\top, 0.2\boldsymbol{I})$ $r(\boldsymbol{u}) \propto 1 + 8\exp\left(-5000\|\boldsymbol{u} - [0.8, 0.8]^\top\|^2\right) + 5\exp\left(-500\|\boldsymbol{u} - [0.9, 0.5]^\top\|^2\right)$ $+6\exp\left(-500\|\boldsymbol{u} - [0.8, 0.3]^\top\|^2\right)$
10	$g(\boldsymbol{u}) = 1$ $r(\boldsymbol{u}) \propto 0.1 + \exp\left(-500\|\boldsymbol{u} - [0.5, 0.5]^\top\|^2\right)$

Table 2.1: Definitions of control densities g and risk functions r for the synthetic problem scenarios.

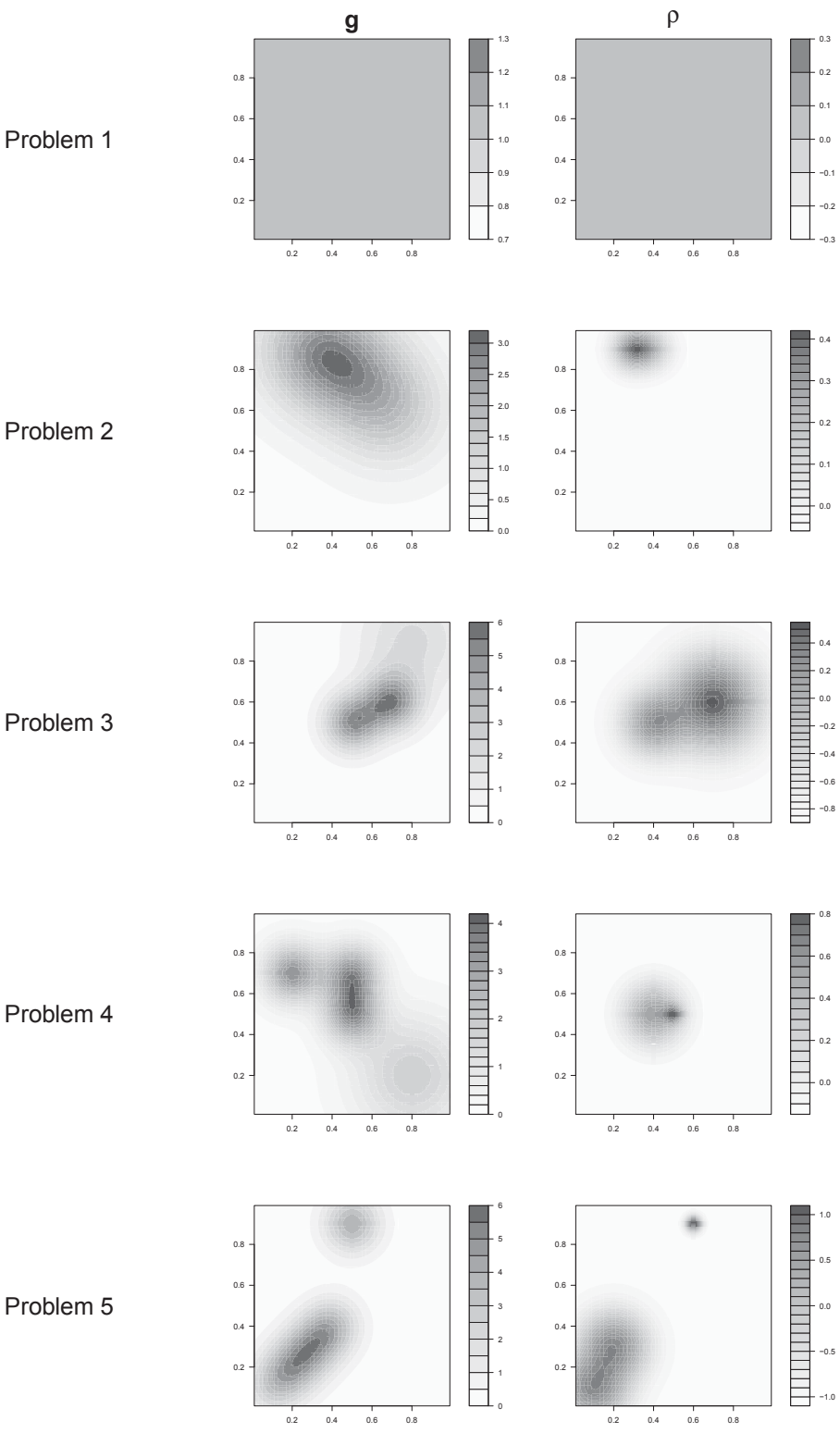


Figure 2.2: Filled contour plots of the synthetic control densities and log-risk functions for Problems 1 to 5.

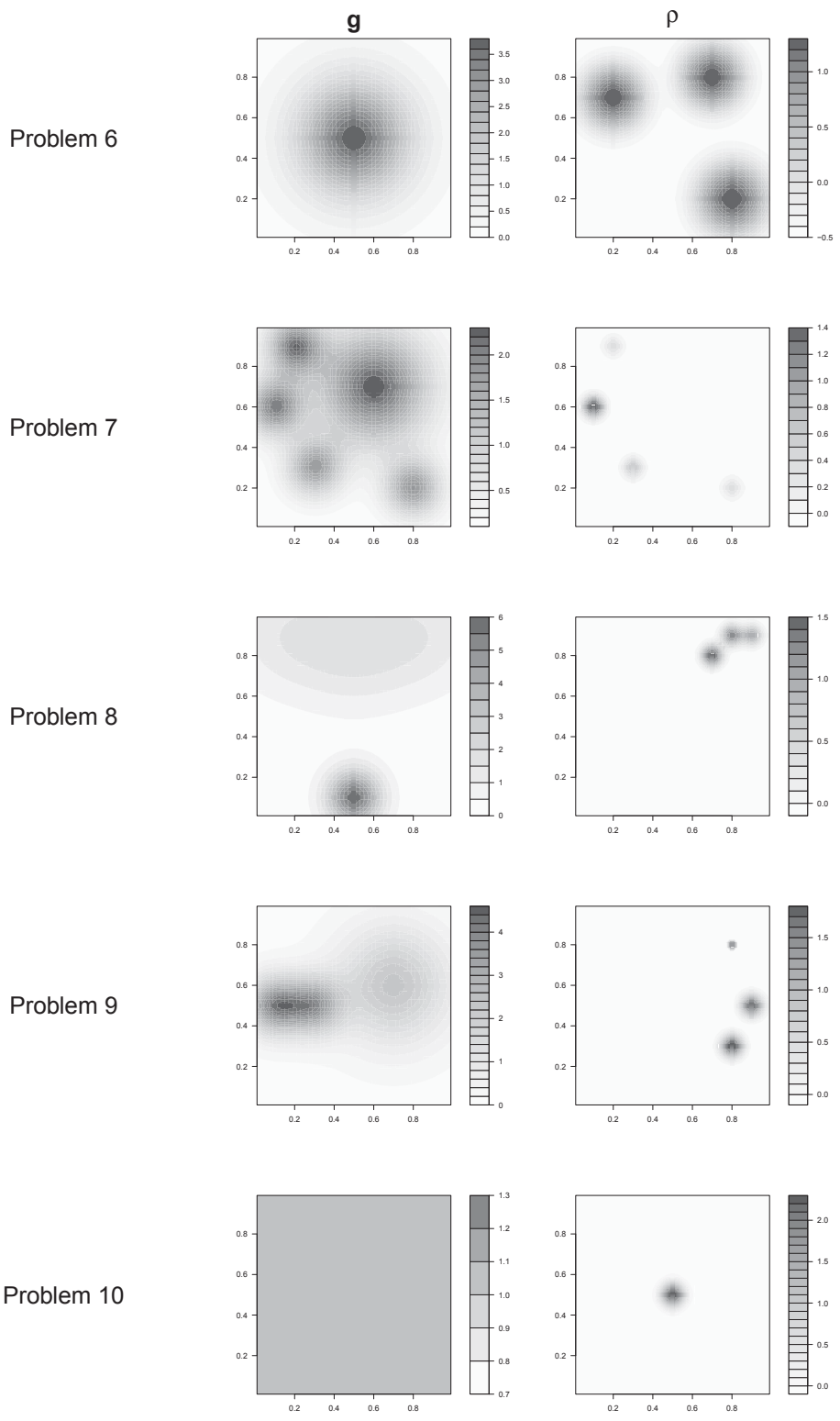


Figure 2.3: Filled contour plots of the synthetic control densities and log-risk functions for Problems 6 to 10.

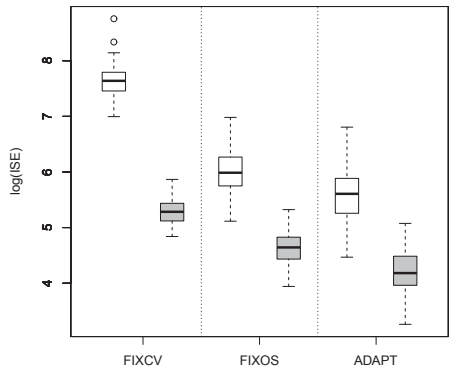


Figure 2.4: Problem 1 log-scaled ISE measure for sample sizes $n_1 = 100$, $n_2 = 100$ (white) and $n_1 = 500$, $n_2 = 1000$ (grey).

The reader is referred back to the comments on bandwidth assignment for reasoning behind these decisions.

The 'quality' or 'performance' of each of FIXCV, FIXOS and ADAPT in terms of the proximity to the true risk surface for each iterated data set is measured in two ways. We compute the integrated squared error (ISE) $\int_W [\hat{\rho}(\boldsymbol{x}) - \rho(\boldsymbol{x})]^2 \, \mathrm{d}\boldsymbol{x}$ as well as a *weighted* integrated squared error (WISE) given by $\int_W \rho(\boldsymbol{x})^2 [\hat{\rho}(\boldsymbol{x}) - \rho(\boldsymbol{x})]^2 \, \mathrm{d}\boldsymbol{x}$. Though the ISE measures overall proximity to the actual surface with equal importance given to all areas of W, the purpose of the WISE is to amplify the error in areas where the risk is more extreme. In other words, WISE prevents the measure of quality being too heavily influenced by areas where there is no activity in the surface. This is particularly important in examining the scenarios where the risk is mostly uniform, as in these cases the preference in terms of overall equal error would simply be to 'flatten' the estimated surface, perhaps smoothing over and ignoring a small yet clear risk increase. The ISE and WISE statistics are displayed on the log-scale as boxplots in Figures 2.4-2.13 for Problems 1 to 10 respectively (for obvious reasons, only the ISE is available for Problem 1).

Inspection of the errors for Problem 1 (Figure 2.4) shows a clear disadvantage for FIXCV, which is to be expected. The perfectly uniform risk surface calls for large bandwidths, punishing undersmoothing occurring from least-squares cross-validation. FIXOS and ADAPT are closer in terms if ISE, but ADAPT seems to return lower errors on average for both sample sizes. It appears the adaptive approach, which as we recall is in fact using the FIXOS bandwidth as the common h_0, is still managing to provide the greater amount of smoothing required here.

Much the same message is conveyed in the results for Problem 2 in Figure 2.5. The relatively mild increase in risk is again inadequately expressed in FIXCV, with FIXOS and ADAPT again the best performers. ADAPT consistently provides the lowest median errors for both ISE and WISE. The WISE results are particularly encouraging, as they indicate the extra smoothing hypothesised to be beneficial for ADAPT in Problem 1 is not so excessive

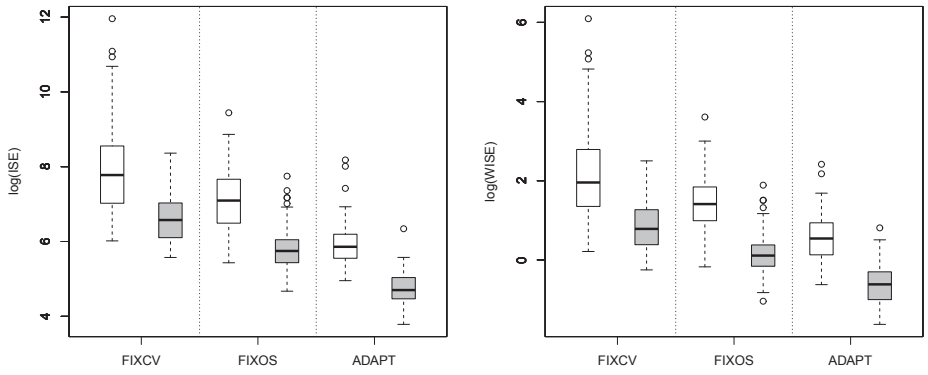


Figure 2.5: Problem 2 log-scaled ISE (left) and WISE measures (right) for sample sizes $n_1 = 100$, $n_2 = 100$ (white) and $n_1 = 500$, $n_2 = 1000$ (grey).

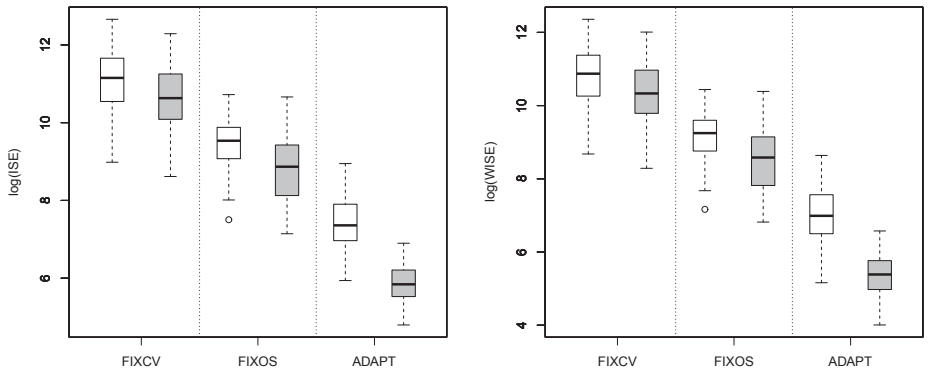


Figure 2.6: Problem 3 log-scaled ISE (left) and WISE measures (right) for sample sizes $n_1 = 100$, $n_2 = 100$ (white) and $n_1 = 500$, $n_2 = 1000$ (grey).

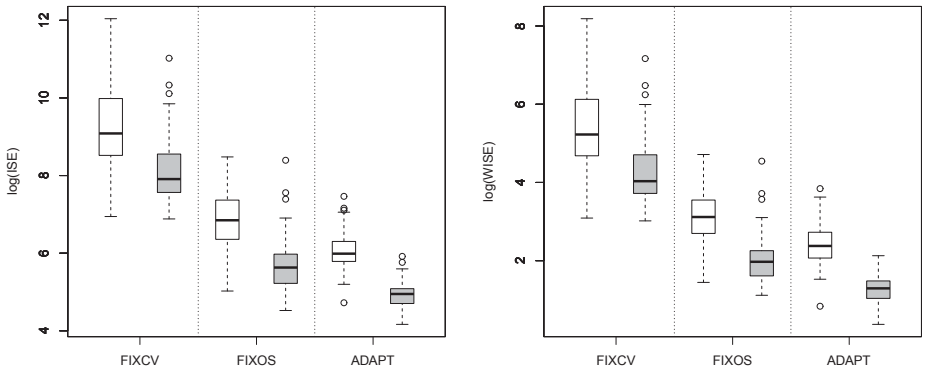


Figure 2.7: Problem 4 log-scaled ISE (left) and WISE measures (right) for sample sizes $n_1 = 100$, $n_2 = 100$ (white) and $n_1 = 500$, $n_2 = 1000$ (grey).

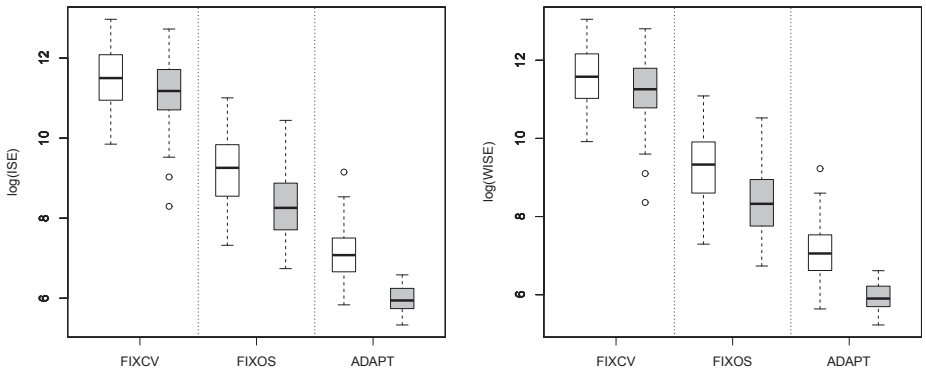


Figure 2.8: Problem 5 log-scaled ISE (left) and WISE (right) measures for sample sizes $n_1 = 100$, $n_2 = 100$ (white) and $n_1 = 500$, $n_2 = 1000$ (grey).

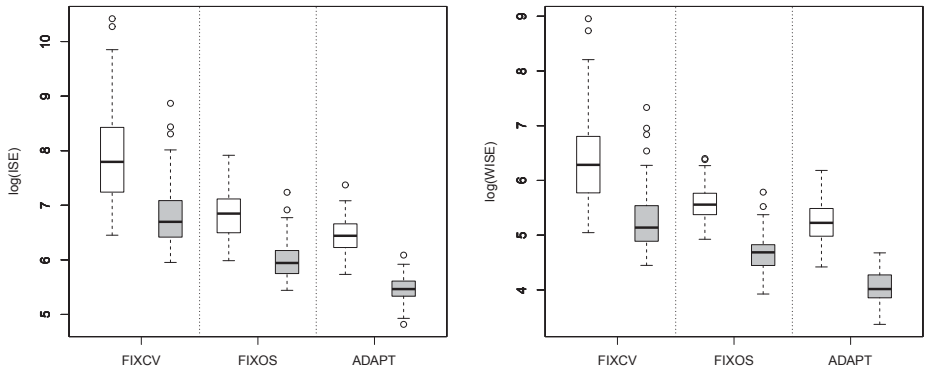


Figure 2.9: Problem 6 log-scaled ISE (left) and WISE (right) measures for sample sizes $n_1 = 100$, $n_2 = 100$ (white) and $n_1 = 500$, $n_2 = 1000$ (grey).

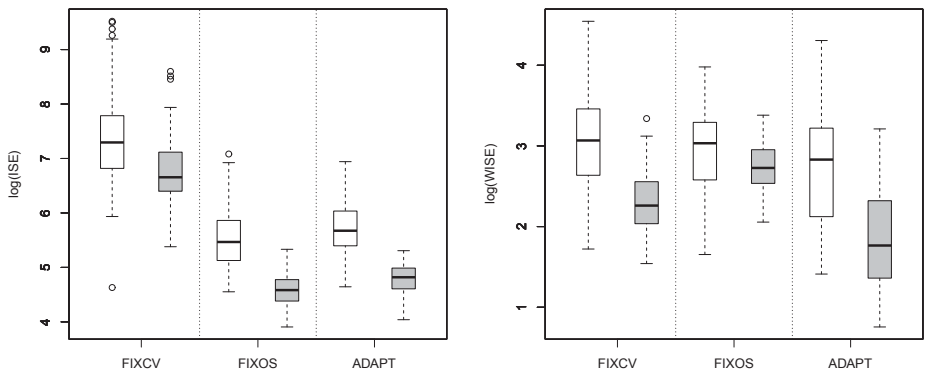


Figure 2.10: Problem 7 log-scaled ISE (left) and WISE (right) measures for sample sizes $n_1 = 100$, $n_2 = 100$ (white) and $n_1 = 500$, $n_2 = 1000$ (grey).

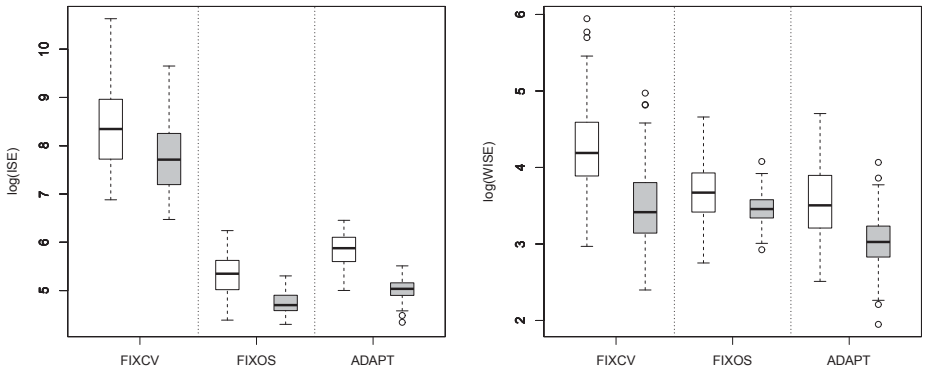


Figure 2.11: Problem 8 log-scaled ISE (left) and WISE (right) measures for sample sizes $n_1 = 100$, $n_2 = 100$ (white) and $n_1 = 500$, $n_2 = 1000$ (grey).

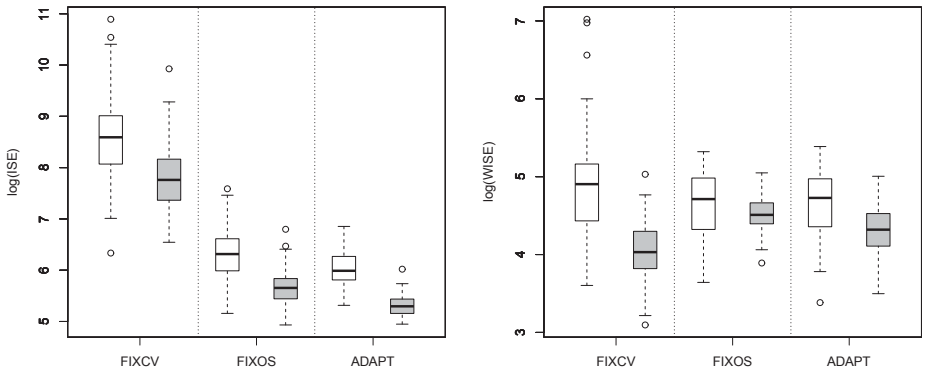


Figure 2.12: Problem 9 log-scaled ISE (left) and WISE (right) measures for sample sizes $n_1 = 100$, $n_2 = 100$ (white) and $n_1 = 500$, $n_2 = 1000$ (grey).

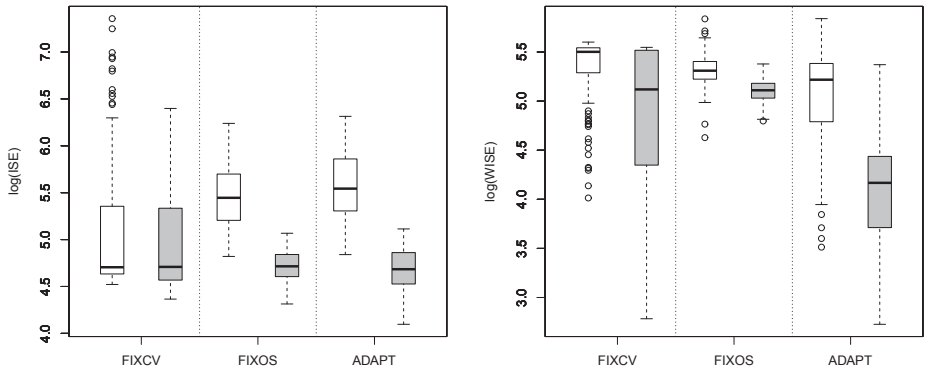


Figure 2.13: Problem 10 log-scaled ISE (left) and WISE (right) measures for sample sizes $n_1 = 100$, $n_2 = 100$ (white) and $n_1 = 500$, $n_2 = 1000$ (grey).

as to miss the actual fluctuations altogether.

Even more distinct are the results for Problem 3 (Figure 2.6). The ISE and WISE measures are very similar, with FIXCV performing badly in comparison to the other two considered risk function estimators. ADAPT is clearly the best performer.

Problem 4 in Figure 2.7 reveals ADAPT once again providing superior error measures to the fixed-bandwidth estimates. FIXOS comes close in second place with undersmoothing continuing to plague FIXCV. It is difficult to discern any clear differences between the ISE and WISE boxplots for this scenario, though as the extremity of maximum risk continues to rise as we proceed through the different problems, we can expect WISE to begin to play a bigger role in interpreting the estimator quality.

Problem 5 (Figure 2.8) shows plot similar to those for Problem 3 in terms of the quality of ADAPT in comparison to FIXOS and FIXCV. This particular scenario could be rather difficult for fixed-bandwidth approaches to handle, due to the presence of both a diffuse, elogated risk increase as well as a concentrated, circular risk hotspot. The adaptive estimator is naturally more flexible than the fixed, and it appears to work well in this example in terms of best describing the appearance of the surface.

Boxplots corresponding to Problem 6, shown in Figure 2.9, once more point to strongest performance of ADAPT, followed by FIXOS then FIXCV. The improvement of ADAPT is not as noticeable as in Problem 5, most likely owing to the fact that the three modes of risk are of almost identical appearance. Unfortunately for the fixed bandwidth estimator, it could be argued that this is less likely to occur in real-world applications.

Problem 7 with a maximum risk of 3.92, is where we begin to notice a change in the pattern of ISE measures (Figure 2.10). FIXCV remains the worst performer, but FIXOS now seems marginally better than ADAPT. This is most likely due to the 'flatness' of risk over most of W, where the adaptive estimator may still be attempting to capture detail. This highlights the importance of the WISE statistic where ADAPT once more shows the smallest median error, especially for the larger sample size. This indicates although FIXOS

has, to a certain extent, stabilised the uniform areas of the region, it has not performed as well as ADAPT in reaching the true height of the peaks in risk that are indeed present.

We note precisely this behaviour once more in Figure 2.11, where the results for Problem 8 are displayed. Like Problem 7, this scenario shows a mostly uninteresting risk surface, albeit with three tight, tall peaks in the upper right quadrant. FIXOS appears superior in terms of ISE, but the WISE results clearly show lack of ability of FIXOS to capture the actual level of risk for which ADAPT is better equipped to handle. FIXCV stays behind the other two estimates in quality.

Problem 9 shows some interesting plots in Figure 2.12. There seems to be a mild improvement of ADAPT over FIXOS, with FIXCV trailing behind in ISE. For the WISE, however, FIXCV shows some extremely competitive results, providing the best median error for the larger sample size. It seems the very small bandwidths issued by LSCV_ω which punished FIXCV in ISE have done well in reaching the tall peaks in this scenario (maximum raw risk of 5.99). ADAPT has still outperformed FIXOS in WISE, but it would seem there has been a minor trade-off in both attempting to reach the height of the hotspots *and* smooth over the rest of W as required.

Recall the final scenario, Problem 10, provides a contrasting example to Problem 1 with a uniform control density and a corresponding risk surface with one obvious peak reaching 9.45 on the raw risk scale. The results in Figure 2.13 are somewhat difficult to interpret. The small sample sizes show ADAPT struggling to perform well in ISE when compared to the fixed-bandwidth approaches. However, this trend is reversed when we use the larger n_1 and n_2. It is possible that in this particular scenario we require more data in order for the asymptotic benefits of ADAPT to be observed. That being noted, the WISE results do show ADAPT as superior, mildly so for the small sample size and stronger for the large.

Broadly speaking, the simulation study has shown favourable performance of the adaptive kernel estimator of the relative risk function for point process data over the fixed-bandwidth version used to date. Intuitively, this makes sense. The high level of heterogeneity commonly observed in epidemiological applications means it is to our benefit in order to be able to observe more detail in heavily populated areas and not lend too much density height to isolated observations (i.e. where there is less information). Nevertheless, it is encouraging to see the theoretical benefits of Abramson's (1982) smoothing scheme do appear to be realised in practice.

2.6 Examples

This section implements the relative risk methodology described in this chapter using real-world examples. It also provides an opportunity to examine the behaviour of the tolerance contour calculation approaches.

2.6.1 Chorley-Ribble cancer data

The first example comprises the locations of 58 'cases' of laryngeal cancer encountered in Section 1.1 coupled with 971 'controls' of lung cancer in the Chorley-Ribble health regions

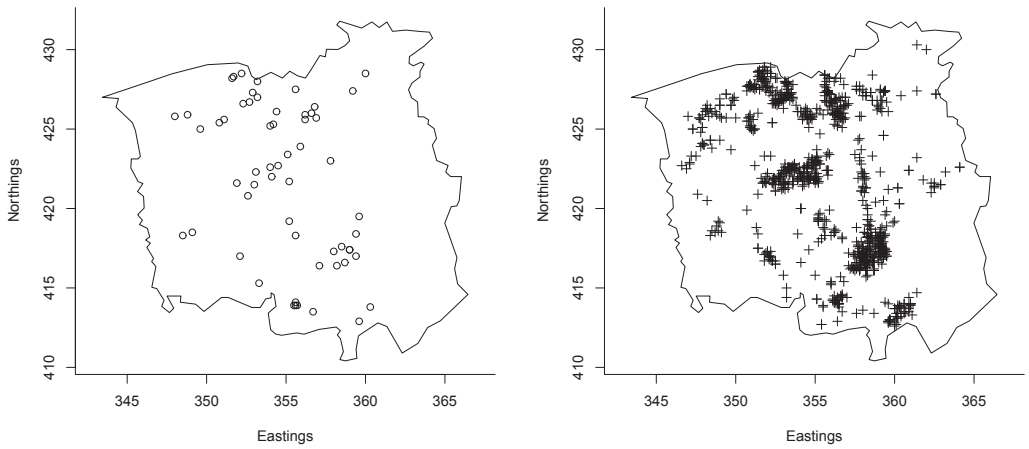


Figure 2.14: Laryngeal cancer cases (left) and lung cancer controls (right) for the Chorley-Ribble dataset.

of southeastern England. Cases and controls are displayed side-by-side in Figure 2.14. As mentioned earlier, a larger version of these data was first presented and analysed by Diggle (1990).

Due to the poor performance of FIXCV in the simulation study we compute only the ADAPT and FIXOS edge-corrected log-risk function estimates (the different case/control pilot bandwidths for ADAPT are now set to half of the OS-computed figures for the case and control data separately). Shown in Figure 2.15 are two copies of these plots, one with MC (Monte-Carlo) tolerance contours, the other with ASY (asymptotic) contours. These contours were computed with respect to (2.29) i.e. an 'upper-tailed' test searching for significant risk elevations.

Some comments are warranted. Although the range of ADAPT log-risk appears approximately the same as the FIXOS surface, we would expect taller peaks to be generated in ADAPT, due to the reduced smoothing applied to dense groupings of observations. FIXOS has helped to stabilise the variability resulting from a surface based on a smaller bandwidth in the fixed setting, at the expense of potentially smoothing too greatly over any 'hotspots' of interest.

This can be seen in the included tolerance contours. It seems that the ADAPT risk surface has had more 'success' in flagging significant subregions than FIXOS. This has most likely been a result of both taller peaks in the case density estimate, as well as the asymptotic variance stabilisation. The lack of responsiveness in terms of flagging hotspots by FIXOS is what we would expect with a greater amount of smoothing limiting the magnitude of risk fluctuations. Indeed, there were no significant ASY contours for the FIXOS surface.

A comparison of ASY vs. MC methods for tolerance contour calculation shows the relative stability of the asymptotic approach. In both ADAPT and FIXOS MC contours, sub-regions on the western border were suggested as being significant. However, examining Figure 2.14, we note there are no observed cases in these areas. Over all four images, the most 'agreed

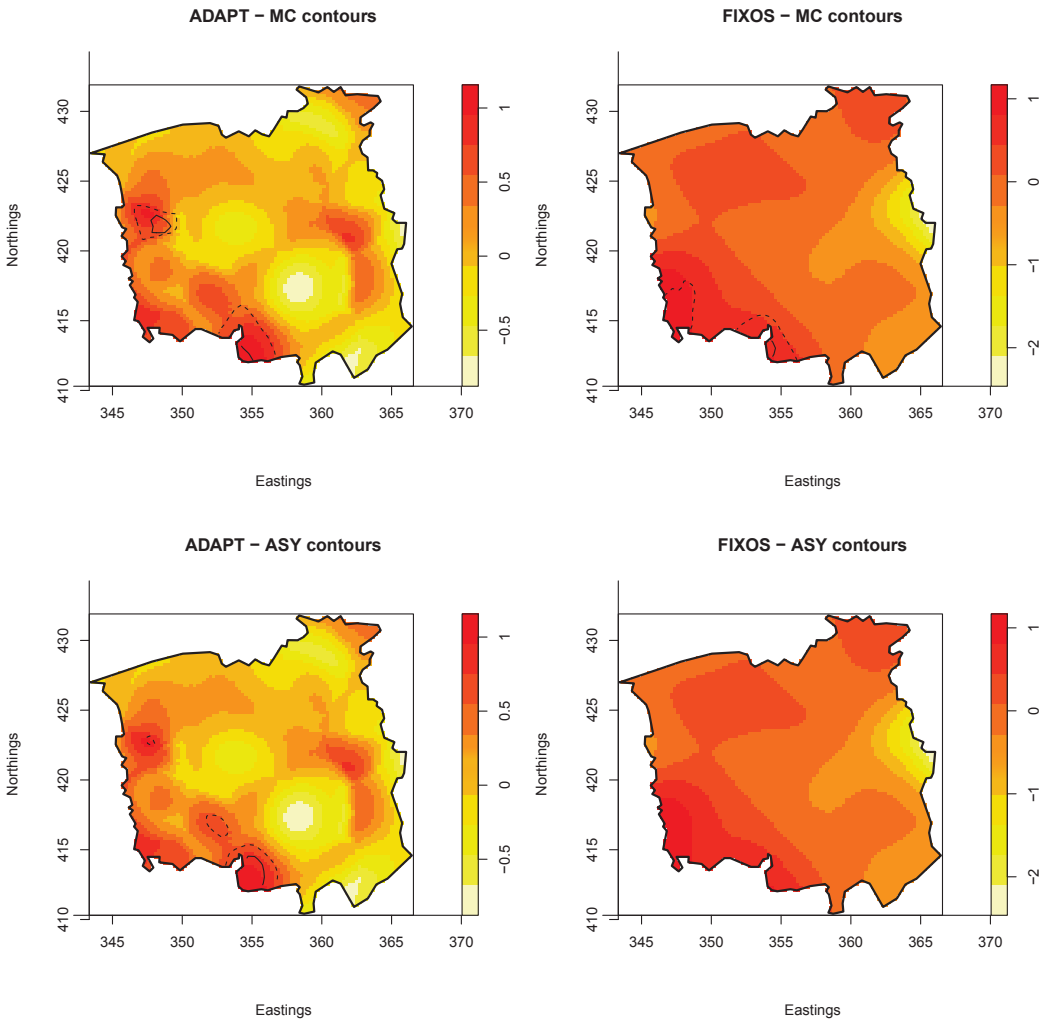


Figure 2.15: Chorley-Ribble log-risk function estimates utilising the ADAPT and FIXOS smoothing regimens, with MC tolerance contours (top row) and ASY tolerance contours (bottom row) superimposed at significance levels of 0.05 (solid line) and 0.1 (dashed line).

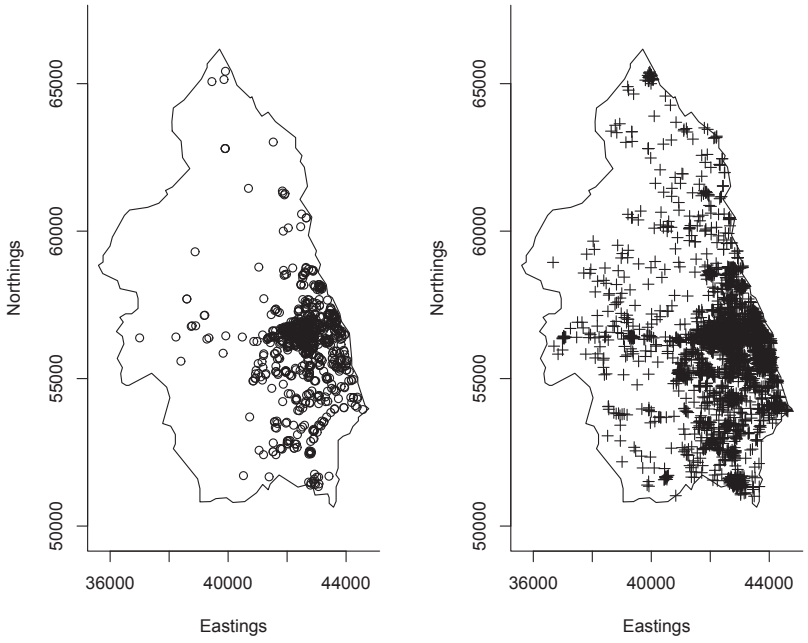


Figure 2.16: PBC cases (left) and population controls (right) for the PBC dataset.

upon' sub-region of potential significance is located on the center of the southern border. At any rate, conclusions based on ASY vs. MC must be reached with care in this example, given $n_1 = 58$ only. Finally, we note that the vast improvement in computational expense of the ASY p-value surfaces is undeniable; there being no need for repeated evaluations of the risk function leads to total execution time in a matter of seconds, as opposed to hours.

2.6.2 Primary biliary cirrhosis in northeastern England

These data were first presented and analyzed in Prince et al. (2001). We have the spatial coordinates of 761 cases of primary biliary cirrhosis (PBC) along with 3020 controls randomly selected from the underlying population using weighted postal zones, collected within several adjacent health regions in northeastern England (Figure 2.16). As in the previous example, the ADAPT and FIXOS risk surfaces are computed with ASY and MC contours for the PBC data; these are displayed in Figure 2.17.

Immediately we note the differences in appearance between the fixed and adaptive surfaces over the study region. As would be expected from the methodology, ADAPT has provided more detail. Again, due to a lack of spatial adaptability, FIXOS has been unable to capture the same level of detail in the heavily populated areas. In addition to this, note the scale of the estimates of log-risk. In contrast to ADAPT's range of about -4 to 1, the FIXOS lower estimates reach below -30 on the log-scale, indicating (entirely implausibly) almost complete protection from PBC in those western areas. ADAPT, with its larger bandwidths in sparsely populated sub-regions, has avoided these extremities.

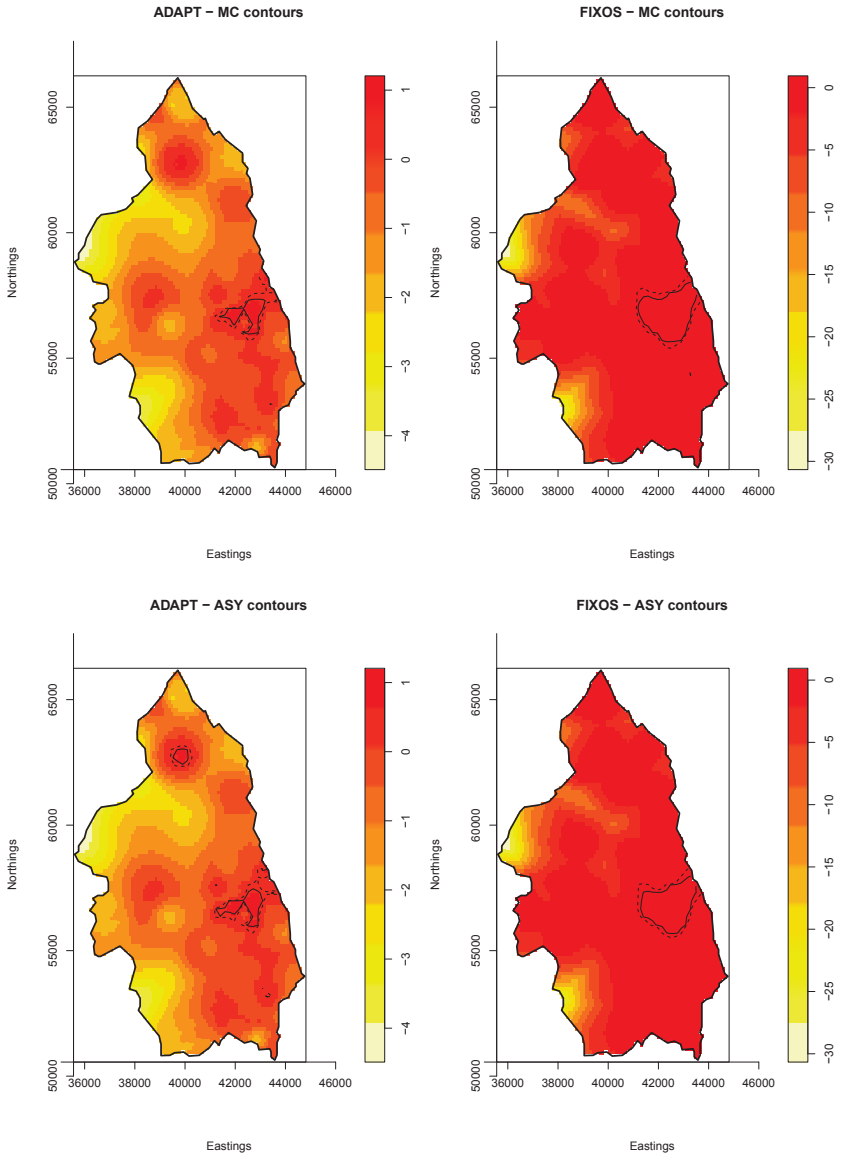


Figure 2.17: PBC log-risk function estimates utilising the ADAPT and FIXOS smoothing regimens, with MC tolerance contours (top row) and ASY tolerance contours (bottom row) superimposed at significance levels of 0.05 (solid line) and 0.1 (dashed line).

Examination of the tolerance contours shows that all are in agreement over a general area of increased risk on the eastern border (this supports the original findings in Prince et al., 2001). In line with the difference between the surface estimates, the ASY and MC contours for ADAPT have been a little more detailed in the specific appearance of the significant area(s). Interestingly, a small group of cases in the north of the study region has been flagged by the ASY contours as a significant hotspot. This may warrant further investigation in terms of the application itself, though the exploratory nature of this methodology means contours must be interpreted as only an indication, rather than definitive proof. Once more, the reduction in computational requirements for ASY as opposed to MC was impressive.

2.7 Closing remarks

The investigation of spatial relative risk for both aggregated count and case-event (point process) data is a useful, straightforward and therefore common first step in the analysis of problems in geographical epidemiology. When we observe point process data it is often convenient to employ the flexible method of kernel density estimation to provide a smooth and continuous estimate of said risk (or log-risk) through a density-ratio function.

The use of a variable bandwidth in this approach has been very scarce in the literature, especially in comparison to the fixed bandwidth technique. This is in spite of known theoretical benefits in using e.g. Abramson's (1982) adaptive calculation prescription, and the fact that several works have noted spatial adaptability in the amount of smoothing is intuitively sensible for modelling geographical dispersions of human and animal populations (e.g. Benschop et al., 2008). It is reasonable to conjecture that this gap in the literature is due, at least in part, to the fact that the fixed estimator is relatively straightforward to implement, with comparatively little guidance until now as to how one should go about variable smoothing in the density-ratio.

This chapter has therefore sought to study in detail and improve the kernel-smoothed density-ratio estimator of relative risk when we have observed point locations. Through detailing some asymptotic properties of the adaptive risk function, we found that in addition to the original discovery of bias reduction, we are likely to experience approximate variance stabilisation not available in the fixed bandwidth setting. This is a particularly pleasant feature, as it helps provide numerical stability in subsequent analyses involving the surface estimate.

Though the adaptive estimator is somewhat more computationally expensive than the fixed, this is only likely to become a tangible problem if repeated evaluations of the estimator are required. One such example would be the calculation of Monte-Carlo tolerance contours, used in an effort to identify areas of statistically significant fluctuations in risk. We therefore exploited the asymptotic normality of kernel density estimates, coupled with the derived expressions for the approximate variance of the estimator, to develop a computationally inexpensive approach to calculation of the corresponding p-value surfaces. As well as a huge improvement in computational expense, the asymptotic tolerance contours also appeared to be slightly more resistant to artefactual oddities (such as signalling significance in areas with little or no data) than the MC approach.

These interesting discoveries, however, still depend on the all-important assignment and selection of initial smoothing parameters (fixed, global and pilot). Though it is acknowledged that the choices made in the preceding work are somewhat *ad hoc*, there is at present little to guide the researcher in the 'best' selection methods. Indeed, aside from having high computational requirements (especially for the adaptive estimator) the 'jointly optimal' selectors that do exist perform rather poorly in provision of a sensible smoothing parameter in practice. Furthermore, the fact that the relatively 'cheap and dirty' oversmoothing technique we employed seemed to do quite well in terms of integrated squared error indicates there is still much to be learned concerning adequate methodology for common case/control bandwidth optimisation.

Overall, however, we may conclude from the work here that use of Abramson's adaptive smoothing regimen for kernel estimation of the case and control densities in the relative risk function shows considerable promise, due to certain resultant theoretical and practical benefits. This is particularly true when we expect to see the type of spatial heterogeneity often encountered in epidemiological applications.

Chapter 3

Spatial and Spatiotemporal Modelling with the Log-Gaussian Cox Process

Modelling spatial point patterns, or a sequence of spatially and temporally dependent point patterns, is a complicated endeavour. In the previous chapter, we considered a common exploratory tool often used in epidemiology for the comparison of two spatial point patterns. This focused on refining implementation of the relative risk function, with less emphasis on how the points came to be 'generated'. The essence of interpreting such a risk surface is contingent upon the individual intensity functions being fixed, i.e. unconditionally inhomogeneous Poisson.

As we observed in Chapter 1, the notion of an unchanging, 'true' intensity function may not necessarily be intuitively sensible in epidemiological applications. This especially applies to spatiotemporal data sets, where we would expect the spatial intensity to vary dynamically over time.

In this chapter, we turn our attention to one particularly useful Cox process for investigating spatial and spatiotemporal point patterns, where we expect to experience some form of clustering. Though technically a little more challenging, this class of models is both powerful and flexible in terms of the scope of problems to which it can be applied and the modifications that can be made to the general framework to suit the researcher's data and goals. Nevertheless, there are a number of unresolved issues pertaining to implementation of such processes for epidemiological modelling.

Section 3.1 introduces the *log-Gaussian Cox process* and specifies its fundamental properties for both purely spatial and spatiotemporal applications, as well as discussing the inclusion of multiplicative deterministic components; useful for epidemiological applications. Perhaps the most pressing issue in the practical implementation of this methodology, that of appropriate parameter estimation for the nature of the assumed dependence, is discussed in Section 3.2. Section 3.3 details how the *fast-Fourier transform* is exploited in order to gain a significant computational improvement in practice, as well as the Markov-Chain Monte-

Carlo techniques required for conditional simulation of a log-Gaussian Cox process. Section 3.4 provides spatial and spatiotemporal analyses of two real-world problems using the reviewed methods. A comprehensive collection of simulations in Section 3.5 provide valuable insight into how well the most widely employed parameter estimation technique performs in practice, and also examine the sensitivity of anomaly-detection methods to variation in the estimated parameters. A selection of potential theoretical refinements are discussed in Section 3.6, and we summarise findings and conclude in Section 3.7.

This thorough review of the LGCP serves not only to motivate the novel numerical experiments in Section 3.5, but also as an invaluable step in the first R-language implementation of the methodology. The corresponding software package is discussed later on in Chapter 4.

3.1 Definitions and properties

The key aspect of the Cox process, from Chapter 1, is the fact that the underlying intensity function which we assume to have generated the observed points is itself driven by some stochastic mechanism. In the case of the log-Gaussian Cox process (LGCP), the basis of this mechanism is a Gaussian random field with a given covariance structure. Here we discuss the model in both spatial and spatiotemporal contexts, and present some of the most important theoretical properties thereof.

3.1.1 Spatial

The introduction of the spatial LGCP in a statistical setting is accredited to Møller et al. (1998). The intensity we observe, $\lambda(\boldsymbol{x})$, is a realisation of the intensity-generating mechanism $\Lambda(\boldsymbol{x})$ (denoted with '$\leftarrow$'), such that

$$\lambda(\boldsymbol{x}) \leftarrow \Lambda(\boldsymbol{x}) = \exp[\mathcal{Y}(\boldsymbol{x})]; \qquad \boldsymbol{x} \in \mathbb{R}^2, \tag{3.1}$$

where $\mathcal{Y}$ is a real-valued Gaussian random field. Naturally, properties of the LGCP are found with respect to properties of $\mathcal{Y}$. To distinguish between the intensity function conditional upon a realisation of the random field, $\lambda(\boldsymbol{x})$, and the properties of the stochastic mechanism itself, $\Lambda(\boldsymbol{x})$, we will use $\rho_\lambda^{(k)}$ to denote its kth-order product density. Using the moment generating function of the normal distribution, we have

$$\begin{aligned}
\rho_\lambda^{(k)}(\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_k) &= \mathbb{E}\left[\prod_{i=1}^{k} \Lambda(\boldsymbol{x}_i)\right] \\
&= \mathbb{E}\left[\exp\left\{\sum_{i=1}^{k} \mathcal{Y}(\boldsymbol{x}_i)\right\}\right] \\
&= \exp\Bigg\{\sum_{i=1}^{k}\left[\mu_\lambda(\boldsymbol{x}_i) + \frac{\sigma_\lambda^2(\boldsymbol{x}_i)}{2}\right] \\
&\quad + \mathbf{1}(k>1)\sum_{i=1}^{k-1}\sum_{j=i+1}^{k} \sigma_\lambda(\boldsymbol{x}_i)\sigma_\lambda(\boldsymbol{x}_j) r_\lambda(\boldsymbol{x}_i, \boldsymbol{x}_j; \phi_\lambda)\Bigg\},
\end{aligned} \tag{3.2}$$

where $\mu_\lambda(\boldsymbol{x}) = \mathbb{E}[\mathcal{Y}(\boldsymbol{x})]$ (the mean of the Gaussian field at $\boldsymbol{x}$), $\sigma^2_\lambda(\boldsymbol{x}) = \text{Var}[\mathcal{Y}(\boldsymbol{x})]$, and r is the correlation function of $\mathcal{Y}$ controlled by a parameter ϕ_λ such that $\text{Cov}(\boldsymbol{x}_1, \boldsymbol{x}_2) = \sigma_\lambda(\boldsymbol{x}_1)\sigma_\lambda(\boldsymbol{x}_2)r(\boldsymbol{x}_1, \boldsymbol{x}_2; \phi_\lambda)$. The mean structure of the LGCP is therefore found with

$$\rho^{(1)}_\lambda(\boldsymbol{x}) = \exp\left\{\mu_\lambda(\boldsymbol{x}) + \frac{\sigma^2_\lambda(\boldsymbol{x})}{2}\right\}, \tag{3.3}$$

and the second-order characteristics with

$$\begin{aligned}\rho^{(2)}_\lambda(\boldsymbol{x}_1, \boldsymbol{x}_2) &= \exp\left\{\mu_\lambda(\boldsymbol{x}_1) + \mu_\lambda(\boldsymbol{x}_2) + \frac{\sigma^2_\lambda(\boldsymbol{x}_1) + \sigma^2_\lambda(\boldsymbol{x}_2)}{2} + \sigma_\lambda(\boldsymbol{x}_1)\sigma_\lambda(\boldsymbol{x}_2)r(\boldsymbol{x}_1, \boldsymbol{x}_2; \phi_\lambda)\right\}, \\ &= \rho^{(1)}_\lambda(\boldsymbol{x}_1)\rho^{(1)}_\lambda(\boldsymbol{x}_2)\exp\left\{\sigma_\lambda(\boldsymbol{x}_1)\sigma_\lambda(\boldsymbol{x}_2)r(\boldsymbol{x}_1, \boldsymbol{x}_2; \phi_\lambda)\right\}. \end{aligned} \tag{3.4}$$

Recall definition (1.9) of the pair correlation function in Section 1.2.3. The above result (3.4) therefore leads to a pleasantly simple form of the PCF for the LGCP, namely

$$g_\lambda(\boldsymbol{x}_1, \boldsymbol{x}_2) = \exp\left\{\sigma_\lambda(\boldsymbol{x}_1)\sigma_\lambda(\boldsymbol{x}_2)r(\boldsymbol{x}_1, \boldsymbol{x}_2; \phi_\lambda)\right\}. \tag{3.5}$$

Imposing the conditions of stationarity and isotropy upon the Gaussian field $\mathcal{Y}$ can serve to greatly simplify theoretical investigations into the LGCP, make available certain methods for fitting the models, and offer considerable computational benefits. For most practical intents and purposes (especially the kinds of epidemiological applications we will consider momentarily), these conditions are not overly restrictive. We will therefore assume stationarity and isotropy throughout the remainder of this chapter.

Stationarity and isotropy of $\mathcal{Y}$ implies $\mu_\lambda(\boldsymbol{x}) \equiv \mu_\lambda$, $\sigma^2_\lambda(\boldsymbol{x}) \equiv \sigma^2_\lambda$, and $r(\boldsymbol{x}_1, \boldsymbol{x}_2; \phi_\lambda) \equiv r(u; \phi_\lambda)$ where $u = ||\boldsymbol{x}_1 - \boldsymbol{x}_2||$. This simplifies (3.2), thereby implying $\rho_\lambda = \rho^{(1)}_\lambda(\boldsymbol{x}) = \exp\left\{\mu_\lambda + \frac{1}{2}\sigma^2_\lambda\right\}$, $\rho^{(2)}_\lambda(u) = \rho^2_\lambda \exp\left\{\sigma^2_\lambda r(u; \phi_\lambda)\right\}$ and $g_\lambda(u) = \exp\left\{\sigma^2_\lambda r(u; \phi_\lambda)\right\}$. From these statements it is clear that a particular LGCP will be stationary and isotropic if and only if $\mathcal{Y}$ is stationary and isotropic. Estimators for these properties are naturally based on data $\boldsymbol{X}_W$ collected within a finite spatial region so that our process is defined on $\boldsymbol{x} \in W \subset \mathbb{R}^2$. The estimates $\hat{\rho}_\lambda$, $\hat{\rho}^{(2)}_\lambda$ and therefore $\hat{g}_\lambda$ can be computed precisely as outlined in Section 1.2.3. In addition, we may straightforwardly compute $\hat{K}_\lambda$, the estimate of Ripley's K-function.

The LGCP is of particular interest from the point of view of epidemiological applications for several important reasons. Firstly, the behaviour/appearance of any generated intensity is essentially only constrained in terms of the functional form of r. This means we are able to generate, model and capture a wide variety of point patterns without, for example, restricting point clusters to be of a certain shape; flexibility which is appealing in the current setting for obvious reasons. Secondly, note that the LGCP is completely specified by relatively simple formulæ for the first- and second-order characteristics, limiting the number of parameters that will require estimation in practice. Thirdly, because the distribution of the Gaussian field and hence the LGCP is known given these specifications, the need for edge-correction (due to the fact that in practice we are restricted to W) is conveniently obviated when we wish to simulate therefrom. Finally, and as we shall see in the following sections, the LGCP can be extended in an intuitive fashion to deal with space-time point processes, as well as modified to take into account some underlying, yet fixed, heterogeneity.

3.1.2 Spatiotemporal

The spatially varying intensity conditional upon a realisation of the Gaussian process yields the expected number of points per unit area over the plane. This becomes a spatiotemporal intensity ψ when the observations are time-dependent; the expected number of points at location $\boldsymbol{x}$ and time t such that

$$\psi(\boldsymbol{x}, t) \leftarrow \Psi(\boldsymbol{x}, t) = \exp[\mathcal{Z}(\boldsymbol{x}, t)]; \qquad \boldsymbol{x} \in \mathbb{R}^2, \quad t \in \mathbb{R}^+, \tag{3.6}$$

for a spatiotemporal Gaussian field $\mathcal{Z}$. This model, stochastic in both space and time (for nonnegative reals denoted with $\mathbb{R}^+$), was suggested by Brix and Diggle (2001) and builds directly upon the earlier spatial work by Møller et al. (1998). As such, the same appealing features remain: a powerful and flexible modelling framework with relatively tractable first- and second-order properties.

Before we delve further into this extension, it is necessary to clarify the behaviour of $\mathcal{Z}$. Where in the purely spatial case we need only be concerned with spatial correlation (specified by r), it is clear that in the spatiotemporal version we must be able to handle covariance between distinct points in the space-time continuum i.e. $\mathbb{R}^2 \times \mathbb{R}^+$. One particularly important feature that we will assume is present is that of a *separable* correlation structure. This dictates that the correlation between two points in space-time can be straightforwardly factorised into 'purely spatial' and 'purely temporal' components, the result of which is to greatly increase mathematical accessibility to moment-based methods for parameter estimation (Section 3.2), and to render computationally feasible the only current methods available to perform conditional simulation (Section 3.3). We assume hereinafter that

$$\text{Corr}\left[\mathcal{Z}(\boldsymbol{x}_1, t_1), \mathcal{Z}(\boldsymbol{x}_2, t_2)\right] = r(\boldsymbol{x}_1, \boldsymbol{x}_2; \phi_\psi)\tau(t_1, t_2; \theta),$$

where r denotes the spatial correlation function, controlled by the parameter ϕ_ψ, and τ is the temporal correlation function controlled by θ. Stationarity and isotropy imply this expression can be further simplified to $r(u; \phi_\psi)\tau(\nu; \theta)$, where $u = ||\boldsymbol{x}_1, \boldsymbol{x}_2||$ and $\nu = t_2 - t_1$; $t_2 \geq t_1$.

Though the goal is to consider a continuous space-time model, in many applications we only observe events in discrete time. Formally, the intensity at time t is therefore expressed as

$$\begin{aligned} \psi_t(\boldsymbol{x}) &= \int_{t_0}^{t} \exp\left\{\mathcal{Z}(\boldsymbol{x}, u)\right\} \, \mathrm{d}u \\ &\approx (t - t_0)\exp\left\{\mathcal{Z}(\boldsymbol{x}, t)\right\}, \end{aligned} \tag{3.7}$$

where t_0 is some previous timestamp $t_0 < t$. Note that the approximation (3.7) is only valid if we are able to assume the time between successive observations is appropriate such that the spatial intensity may be considered approximately constant between timestamps. In turn, it is this approximation which allows inference for the spatiotemporal LGCP as per Brix and Diggle (2001).

Assuming spatiotemporal stationarity and isotropy, and the initial specification of sepa-

rability, we have for the spatial margin

$$\rho_{\psi,t} = \Delta t \times \exp\left\{\mu_\psi + \frac{\sigma_\psi^2}{2}\right\}, \tag{3.8}$$

where $\mu_\psi = \mathbb{E}[\mathcal{Z}(\cdot,\cdot)]$, $\sigma_\psi^2 = \mathrm{Var}[\mathcal{Z}(\cdot,\cdot)]$, and Δt refers to the amount of time elapsed since the previous time point where data have been observed in the same way as $(t - t_0)$ in (3.7). Also, for pairwise distances $u = ||\boldsymbol{x}_1 - \boldsymbol{x}_2||$,

$$\rho_{\psi,t}^{(2)}(u) = \rho_{\psi,t}^2 \exp\left\{\sigma_\psi^2 r(u;\phi_\psi)\right\}, \tag{3.9}$$

implying the PCF $g(u) = \rho_{\psi,t}^{(2)}(u)/\rho_{\psi,t}^2 = \exp\left\{\sigma_\psi^2 r(u;\phi_\psi)\right\}$ yielding the same result as earlier.

In contrast to the spatial-only setting, estimators for (3.8) and (3.9) which lead to the useful intensity, K-function and PCF estimates, must now however be able to take into account our temporally marked data. Suppose we have observed data $\boldsymbol{X}_W = \{\boldsymbol{X}_{W,1} \cup \boldsymbol{X}_{W,2} \cup \ldots \cup \boldsymbol{X}_{W,T}\}$ where W is the defined spatial region and $\boldsymbol{X}_{W,t}$ represents only those data occurring at (or corresponding to) time t for some measurable maximum T. We therefore assume the observation of data at equal, unit time intervals $\Delta t = 1$ for all t in the sense that now $t \in \mathbb{Z}^+$ where $\mathbb{Z}^+$ denotes the set of positive integers. A natural estimator for $\rho_{\psi,t}$ is then

$$\hat{\rho}_\psi \equiv \hat{\rho}_{\psi,t} = \frac{\mathcal{N}(\boldsymbol{X}_W)}{|W|T}, \tag{3.10}$$

where $\mathcal{N}(\boldsymbol{X})$ represents the total number of observations in the pattern $\boldsymbol{X}$. Our stationarity assumption is thus highlighted by the above expression which naturally yields a constant per-unit-space, per-unit-time intensity estimate.

Second-order property estimators of the spatial margin are averaged over time; using (1.10),

$$\hat{g}_\psi(u) = \left[2\pi T u |W| \hat{\rho}_\psi^2\right]^{-1} \sum_{t=1}^{T} \sum_{\boldsymbol{x}_1,\boldsymbol{x}_2 \in \boldsymbol{X}_{W,t}}^{\boldsymbol{x}_1 \neq \boldsymbol{x}_2} \kappa_h(||\boldsymbol{x}_1 - \boldsymbol{x}_2|| - u) w_W(\boldsymbol{x}_1, \boldsymbol{x}_2); \qquad 0 < u \leq u_{\max}, \tag{3.11}$$

where $u_{\max} = r_{\max}$, with $r_{\max}$, w, and κ_h defined as in Section 1.2.3. Similarly, we suggest

$$\hat{K}_\psi(u) = (T|W|)^{-1} \hat{\rho}_\psi^{-2} \sum_{t=1}^{T} \sum_{\boldsymbol{x}_1,\boldsymbol{x}_2 \in \boldsymbol{X}_{W,t}}^{\boldsymbol{x}_1 \neq \boldsymbol{x}_2} \mathbf{1}(||\boldsymbol{x}_1 - \boldsymbol{x}_2|| \leq u) w_W(\boldsymbol{x}_1, \boldsymbol{x}_2); \qquad 0 \leq u \leq u_{\max} \tag{3.12}$$

as an alternative second-order characterisation. It is also possible to use the union of all observations regardless of time, $\boldsymbol{X}_W$, and use the spatial-only estimators in order to obtain these properties. This was pursued in Møller and Díaz-Avalos (2010) who used a different type of Cox process to model the spatiotemporal spread of fires (see Section 3.1.4). In comparison to temporal averaging we do not recommend the union-technique, however, due to the potential loss of finer spatial information specific to data observed at distinct, individual times.

Where the spatial behaviour can be described by examining the aforementioned second-order characteristics, we must also consider the temporal margin. This aspect concerns the covariance between the expected total number of cases of disease at two points in time. Recall that the functions r and τ describe the spatial and temporal correlation structure of the Gaussian field $\mathcal{Z}$ respectively, and let $\nu = |t_2 - t_1|$ such that $0 \leq \nu \leq (T-1)$. Our equal unit time interval and stationarity assumptions allow us to consider the temporal covariance in terms of the integer lag by first considering some arbitrary time point t. Following Brix and Diggle (2001) we have

$$\begin{aligned}
\mathcal{C}(\nu;\theta) &= \mathcal{C}(t, t-\nu;\theta) \\
&= \mathrm{Cov}\left[\mathcal{N}(\boldsymbol{X}_{W,t}), \mathcal{N}(\boldsymbol{X}_{W,t-\nu})\right] \\
&= \mathbb{E}[\mathcal{N}(\boldsymbol{X}_{W,t})\mathcal{N}(\boldsymbol{X}_{W,t-\nu})] - \mathbb{E}[\mathcal{N}(\boldsymbol{X}_{W,t})]\mathbb{E}[\mathcal{N}(\boldsymbol{X}_{W,t-\nu})] \\
&= \mathbb{E}[\mathbb{E}[\mathcal{N}(\boldsymbol{X}_{W,t})\mathcal{N}(\boldsymbol{X}_{W,t-\nu})|\Psi(\cdot,t), \Psi(\cdot,t-\nu)]] \\
&\quad - \mathbb{E}[\mathbb{E}[\mathcal{N}(\boldsymbol{X}_{W,t})|\Psi(\cdot,t)]]\mathbb{E}[\mathbb{E}[\mathcal{N}(\boldsymbol{X}_{W,t-\nu})|\Psi(\cdot,t-\nu)]] \\
&= \int_W \int_W \mathbb{E}[\psi_t(\boldsymbol{x}_1)\psi_{t-\nu}(\boldsymbol{x}_2)]\,\mathrm{d}\boldsymbol{x}_1\,\mathrm{d}\boldsymbol{x}_2 - \int_W \mathbb{E}[\psi_t(\boldsymbol{x})]\,\mathrm{d}\boldsymbol{x} \int_W \mathbb{E}[\psi_{t-\nu}(\boldsymbol{x})]\,\mathrm{d}\boldsymbol{x} \\
&= \int_W \int_W \mathbb{E}[\exp\{\mathcal{Z}(\boldsymbol{x}_1,t) + \mathcal{Z}(\boldsymbol{x}_2,t-\nu)\}]\,\mathrm{d}\boldsymbol{x}_1\,\mathrm{d}\boldsymbol{x}_2 - \int_W \rho_{\psi,t}(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} \int_W \rho_{\psi,t-\nu}(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} \\
&= \int_W \int_W \mathbb{E}[\exp\{\mathcal{Z}(0,0) + \mathcal{Z}(u,\nu)\}]\,\mathrm{d}\boldsymbol{x}_1\,\mathrm{d}\boldsymbol{x}_2 - \left[\rho_\psi |W|\right]^2 \\
&= \int_W \int_W \exp\{2\mu_\psi + \sigma^2_\psi[1 + r(u;\phi_\psi)\tau(\nu;\theta)]\}\,\mathrm{d}\boldsymbol{x}_1\,\mathrm{d}\boldsymbol{x}_2 - \exp\left[2\mu_\psi + \sigma^2_\psi\right]|W|^2.
\end{aligned} \tag{3.13}$$

The final result is obtained through the use of the joint moment-generating function for the spatiotemporal Gaussian field at two time points separated by a lag of ν. A natural estimator for $\mathcal{C}$ is through the empirical autocovariance function, defined simply as

$$\hat{\mathcal{C}}(\nu) = n(\nu)^{-1} \sum_{t_1=1}^{T} \sum_{t_2=t_1}^{T} \{\mathbf{1}[(t_2 - t_1) = \nu]\mathcal{N}(\boldsymbol{X}_{W,t_1})\mathcal{N}(\boldsymbol{X}_{W,t_2})\} - \left[\hat{\rho}_\psi |W|\right]^2, \tag{3.14}$$

where $n(\nu) = \sum_{t_1=1}^{T} \sum_{t_2=t_1}^{T} \mathbf{1}[t_2 - t_1 = \nu]$ denotes the number of pairs contributing to the summation.

3.1.3 The LGCP in epidemiology: deterministic components

The previous discussion has defined the spatial and spatiotemporal intensities purely in terms of the random fields $\mathcal{Y}$ and $\mathcal{Z}$ respectively. This implies that any/all heterogeneity observed in the functions is due purely to the dependence structure of the process at hand. In other words, it is only the correlation between points, arising from spatial and temporal distances, which dictates the appearance of the patterns observed. Patterns which, in the absence of *any* correlation, would exhibit CSR due to the assumption of stationarity.

As mentioned in both Chapters 1 and 2, the idea that the underlying populations are uniformly distributed over the study region is rather unlikely in epidemiological applications,

if not entirely unreasonable. Thus, thinking about an observed disease dispersion specific to a certain population as being entirely down to pointwise correlation may well be inappropriate. Spatial disease variation will, to a certain extent, reflect the geographical distribution of the at-risk population. A disease could also, for example, be more prevalent in winter, leading to overall seasonal changes in the expected number of cases observed throughout a given year.

A very natural approach to tackling this issue is to include within the LGCP framework deterministic components which specify the assumed underlying trends. This changes the role the Gaussian fields play in modelling our problems. Now, 'global' variation in space and/or space-time distributions is captured by the deterministic component(s), leaving the stochastic component to describe the residual 'local' or 'small-scale' correlation (obviously far more difficult to model deterministically). By straightforwardly combining knowledge of the overall variation with a flexible random quantity for finer details the global descriptors fail to capture, we are left with an intuitively sensible model for applications where we expect natural inhomogeneity, while still being sensitive to natural local fluctuations *from* these global trends.

Spatial

We modify the spatial LGCP from Section 3.1.1 to include a multiplicative deterministic term, such that

$$\lambda(\boldsymbol{x}) = \zeta_\lambda(\boldsymbol{x})\exp\{\mathcal{Y}(\boldsymbol{x})\}; \qquad \boldsymbol{x} \in W \subset \mathbb{R}^2, \tag{3.15}$$

where $\zeta_\lambda(\cdot)$ represents the global spatial intensity. For identifiability, we specify that the mean of the LGCP must be equal to one, implying

$$\mathbb{E}[\exp\{\mathcal{Y}(\cdot)\}] = \exp\left\{\mu_\lambda + \frac{\sigma_\lambda^2}{2}\right\} = 1; \qquad \therefore \mu_\lambda = -\frac{\sigma_\lambda^2}{2}. \tag{3.16}$$

We retain the notation $r(||\boldsymbol{x}_1 - \boldsymbol{x}_2||;\phi_\lambda) \equiv r(u;\phi_\lambda)$ to describe the spatial stationary and isotropic correlation structure of the residual stochastic process.

With the mean of $\mathcal{Y}$ specified, the first- and second-order properties (3.3) and (3.4) of the LGCP are altered in an obvious fashion. From (3.16) we therefore have that $\rho_\lambda(\boldsymbol{x}) = \mathbb{E}[\lambda(\boldsymbol{x})] = \zeta_\lambda(\boldsymbol{x})$; the estimator for which depends solely on the specification of ζ_λ. Furthermore,

$$\rho_\lambda^{(2)}(\boldsymbol{x}_1, \boldsymbol{x}_2) = \zeta_\lambda(\boldsymbol{x}_1)\zeta_\lambda(\boldsymbol{x}_2)\exp\{\sigma_\lambda^2 r(u;\phi_\lambda)\}. \tag{3.17}$$

Clearly, the estimators for the pair correlation and K-functions must now take ζ_λ into account and to do this we utilise the inhomogeneous versions of these functions described in Section 1.2.3. Specifically, where without a deterministic intensity component we would use (1.5) and (1.10) to estimate the K- and pair correlation functions respectively; in the current setting we should employ (1.8) and (1.11), with ζ_λ positioned in the denominator.

Spatiotemporal

The spatiotemporal LGCP follows in a similar fashion:

$$\psi(\boldsymbol{x},t)=\zeta_\psi(\boldsymbol{x})\eta(t)\exp\{\mathcal{Z}(\boldsymbol{x},t)\};\qquad \boldsymbol{x}\in W\subset\mathbb{R}^2,\quad t\in\mathbb{Z}^+, \tag{3.18}$$

where ζ_ψ denotes the global spatial intensity, η describes global temporal trends in disease counts, and $\mathcal{Z}$ is the Gaussian field describing residual space-time correlation. Spatiotemporal correlation in $\mathcal{Z}$ is again specified to be separable; $\text{Corr}[\mathcal{Z}(\boldsymbol{x}_1,t_1),\mathcal{Z}(\boldsymbol{x}_2,t_2)]=r(\boldsymbol{x}_1,\boldsymbol{x}_2;\phi_\psi)\tau(t_1,t_2;\theta)\equiv r(u;\phi_\psi)\tau(\nu;\theta)$; $u=||\boldsymbol{x}_1-\boldsymbol{x}_2||$, $0\leq\nu\leq(T-1)$.

Identifiability is achieved following Diggle, Rowlingson and Su (2005). We set $\mu_\psi=-0.5\sigma^2_\psi$ and ζ_ψ is scaled to integrate to one, so that η may now provide the expected number of cases at each timestamp. For a set point in time $t\in\{1,\dots,T\}$ we have that $\rho_{\psi,t}(\boldsymbol{x})=\mathbb{E}[\psi(\boldsymbol{x},t)]=\zeta_\psi(\boldsymbol{x})\eta(t)$ and

$$\rho^{(2)}_{\psi,t}(\boldsymbol{x}_1,\boldsymbol{x}_2)=\zeta_\psi(\boldsymbol{x}_1)\zeta_\psi(\boldsymbol{x}_2)\eta(t)^2\exp\{\sigma^2_\psi r(u;\phi_\psi)\} \tag{3.19}$$

This leads to the pair correlation and K-functions being estimated through the inhomogeneous versions of the time-averaged estimators (3.11) and (3.12), namely

$$\hat{g}_\psi(u)=[2\pi Tu|W|]^{-1}\sum_{t=1}^{T}\eta(t)^{-2}\sum_{\boldsymbol{x}_1,\boldsymbol{x}_2\in\boldsymbol{X}_{W,t}}^{\boldsymbol{x}_1\neq\boldsymbol{x}_2}\frac{\kappa_h(||\boldsymbol{x}_1-\boldsymbol{x}_2||-u)w_W(\boldsymbol{x}_1,\boldsymbol{x}_2)}{\zeta_\psi(\boldsymbol{x}_1)\zeta_\psi(\boldsymbol{x}_2)};\quad 0<u\leq u_{\max} \tag{3.20}$$

and

$$\hat{K}_\psi(u)=(T|W|)^{-1}\sum_{t=1}^{T}\eta(t)^{-2}\sum_{\boldsymbol{x}_1,\boldsymbol{x}_2\in\boldsymbol{X}_{W,t}}^{\boldsymbol{x}_1\neq\boldsymbol{x}_2}\frac{\mathbf{1}(||\boldsymbol{x}_1-\boldsymbol{x}_2||\leq u)w_W(\boldsymbol{x}_1,\boldsymbol{x}_2)}{\zeta_\psi(\boldsymbol{x}_1)\zeta_\psi(\boldsymbol{x}_2)};\quad 0\leq u\leq u_{\max}. \tag{3.21}$$

Note that with the heterogeneous deterministic spatial trend, provided we assume stationary and isotropic (and separated) spatial correlation, (3.17) and (3.19) indicate the validity of using (1.8), (1.11), (3.20) and (3.21) due to the satisfaction of second-order intensity-reweighted stationarity as per Baddeley et al. (2000).

In terms of temporal covariance, equation (3.13) becomes

$$\mathcal{C}(t,t-\nu;\theta)=\eta(t)\eta(t-\nu)\left[\int_W\int_W\zeta_\psi(\boldsymbol{x}_1)\zeta_\psi(\boldsymbol{x}_2)\exp\{\sigma^2_\psi r(u;\phi_\psi)\tau(\nu;\theta)\}\,\mathrm{d}\boldsymbol{x}_1\,\mathrm{d}\boldsymbol{x}_2-1\right] \tag{3.22}$$

for lags $\nu>0$.

Our empirical estimator is straightforwardly adapted from (3.14) to give

$$\hat{\mathcal{C}}(t,t-\nu)=\mathcal{N}(\boldsymbol{X}_{W,t})\mathcal{N}(\boldsymbol{X}_{W,t-\nu})-\eta(t)\eta(t-\nu);\qquad \nu>0. \tag{3.23}$$

Note that although we have stationarity in the temporal covariance of the Gaussian random field, the inclusion of the deterministic component η renders the temporal covariance structure of Ψ nonstationary, unless η is itself constant.

3.1.4 An alternative model

We focus specifically on the spatial and spatiotemporal log-Gaussian Cox process in this epidemiologically motivated work due to its impressive flexibility in terms of the wide variety of intensities and point patterns which it is capable of naturally describing. There exists an alternative Cox process worth briefly mentioning here, due to its modelling potential with respect to the applications of interest.

The spatial and spatiotemporal *shot-noise Cox process* (SNCP), credited to Møller (2003) and Møller and Díaz-Avalos (2010) respectively, represents a general class of cluster-oriented processes which assume that observed 'children' points are born (i.e. randomly scattered) from 'parent' points (the distribution of which is also stochastic). Restricting attention to two dimensions, realisations of this process give intensities $\xi(\boldsymbol{x})$ (spatial) and $\xi(\boldsymbol{x}, t)$ (spatiotemporal), where

$$\xi(\boldsymbol{x}) \leftarrow \Xi(\boldsymbol{x}) = \sum_{\boldsymbol{y} \in \Phi} \gamma_{\boldsymbol{y}} R(\boldsymbol{x}, \boldsymbol{y}) \tag{3.24}$$

and, assuming a discrete time measure/indexer,

$$\xi(\boldsymbol{x}, t) \leftarrow \Xi(\boldsymbol{x}, t) = \sum_{m=-\infty}^{\infty} \sum_{\boldsymbol{y} \in \Phi_m} \gamma_{\boldsymbol{y},m} S(\boldsymbol{x}, \boldsymbol{y}, t, m); \qquad \{\boldsymbol{x}, \boldsymbol{y}\} \in \mathbb{R}^2, \quad \{t, m\} \in \mathbb{Z}. \tag{3.25}$$

The spatial processes Φ and Φ_m, possessing in practice uniform intensities ω and ω_m, determine the locations of the parent nodes, or *mothers* $\boldsymbol{y}$. With *cluster intensity* $\gamma_{\boldsymbol{y}}$ or $\gamma_{\boldsymbol{y},m}$, the *offspring* arise according to the *dispersion density* R or S. The domain $\mathbb{Z}$ denotes the full set of integers.

The SNCP is a useful and intuitive approach to problems where we expect this kind of mother-offspring behaviour and, like the LGCP, has relatively tractable first- and second-order properties. Similar benefits in terms of theoretical accessibility are noticed if we assume stationarity, isotropy, and, in the spatiotemporal case, separability, of the dispersion kernels. We may also include deterministic components in much the same way as in Section 3.1.3.

Possible examples where a SNCP could be considered appropriate include observing the distribution of saplings or seeds of a certain species around adult plants, or modelling forest fire occurrences with respect to ignition locations (as studied in Møller and Díaz-Avalos, 2010). For applications in geographical epidemiology the idea of disease cases arising in the proximity of specific outbreak sources may well be sensible, though there exist problems with the SNCP revolving around the need to specify a particular dispersion kernel. Though clusters of varying forms may be generated via 'overlapping' kernels, the notion of constraining these individual 'infection areas' around each mother point to be a certain shape (e.g. Gaussian, rectangular) may be rather unnatural. We must also expect the choice of dispersion density to affect the local properties of the model. This constitutes a lack of flexibility for diseases in which the range of infection is small.

Another noteworthy issue is the fact that edge-correction, not a problem for a LGCP restricted to a finite interval, cannot be ignored for the SNCP. This can be observed from the definitions in (3.24) and (3.25) as well as being intuitively obvious. If we have observed offspring within the finite interval, it is quite possible that some of these observations have

been generated by mother points *outside* the interval. A simulation of such a process must therefore attempt to account for this potential influence by exterior mothers and we do so by defining an appropriately extended (yet still finite and measurable) space or space-time window in which to allow parent generation. The extended interval is generally computed with respect to the effective range of R or S.

Though the SNCP constitutes an interesting alternative to modelling these types of heterogeneous point patterns, we remain focused on the LGCP for the above reasons.

3.2 Minimum contrast parameter estimation

In practice, the parameters controlling the dependence structure of an assumed Cox process must be estimated. By our definitions, we have in the purely spatial case ϕ_λ and σ^2_λ, with ϕ_ψ, σ^2_ψ and θ for the spatiotemporal model. The convenience of generally requiring only a relatively small number of correlation parameters is one of the appealing features of the LGCP, though this can to a certain extent be viewed as a double-edged sword. The fact that this small collection of parameters fully defines rather complex correlation structures means we must expect each of these parameters to exert considerable influence over the behaviour of resulting patterns. In turn, even slightly misspecified parameter estimates (in reference to the 'true' process values) could have the potential to adversely affect conclusions based on a given LGCP analysis.

The extent of this influence, if any, is difficult to gauge, and will be examined through novel empirical results in Section 3.5. Here we focus on defining the most common approach used to date for parameter estimation in the case of the log-Gaussian Cox process in order to further motivate the need to understand the scope of variation in and subsequent quality of the estimates themselves.

3.2.1 Correlation structures

So far, we have only encountered the separate spatial and temporal correlation structures as generalised functions dependent on the parameters of interest as r and τ respectively. Stationarity and isotropy dictate that both are able to be defined in terms of a scalar lag in either space or time, and we assume that the strength of either correlation diminishes as the lag increases. This means that the same style of functions are appropriate for either r or τ.

Defining a universal correlation function which may represent r or τ as $b(s;\alpha)$, where s represents the scalar lag and α the parameter (ϕ_λ or ϕ_ψ for r, and θ for τ), Table 3.1 gives a few possible forms which b may take. Figure 3.1 provides a visual of these functions for varying values of α.

The images give a good idea of how we can expect the correlation to behave should we choose one of these functional forms for r or τ. The parameter α can clearly be interpreted as a 'scale' value; increasing α increases the effective range of correlation.

The 'raw' spatial correlation r is multiplied by the variance of $\mathcal{Y}$ (or $\mathcal{Z}$), namely σ^2_λ (or σ^2_ψ), and then exponentiated to obtain the PCF for the LGCP as per (3.5). Since this value

Function	$b(s;\alpha) = \ldots$
Linear	$\mathbf{1}(s < \alpha)(1 - s/\alpha)$
Exponential	$\exp\{-s/\alpha\}$
Gaussian	$\exp\{-(s/\alpha)^2\}$
Stable	$\exp\{-(s/\alpha)^{1/2}\}$
Cardinal sine	$\sin(s/\alpha)/(s/\alpha)$
Hyperbolic	$(1 + s/\alpha)^{-1}$

Table 3.1: Some examples of stationary, isotropic correlation functions.

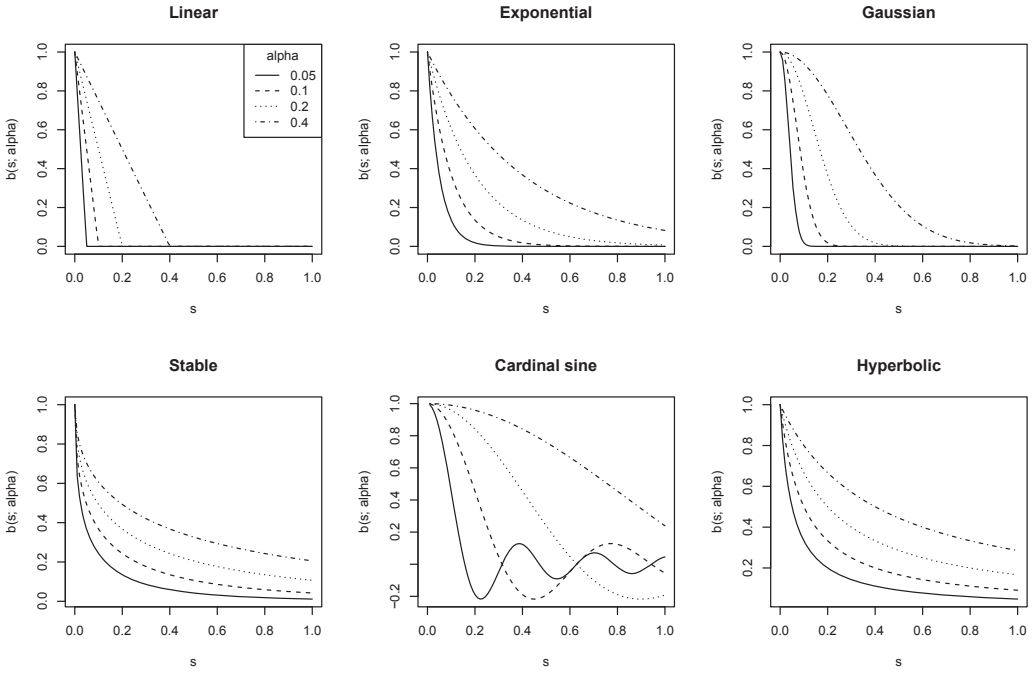


Figure 3.1: Correlation functions from Table 3.1 plotted with $\alpha \in \{0.05, 0.1, 0.2, 0.4\}$; a legend is placed in the 'Linear' function plot.

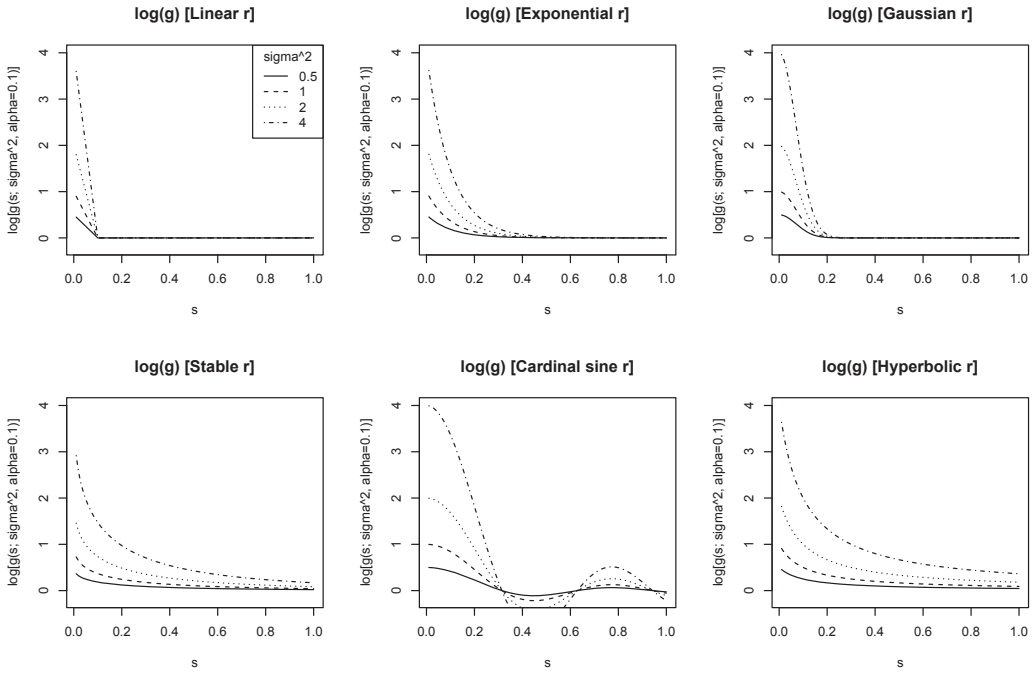


Figure 3.2: Log-pair correlation functions g using as r the functional forms in Table 3.1, plotted with $\sigma^2 \in \{0.5, 1, 2, 4\}$ and a fixed scale $\alpha = 0.1$.

must also be estimated, it is of interest to visually examine the impact of altering the variance of the Gaussian field on the behaviour of the PCF (assuming a fixed scale parameter). Recall that $g(u; \sigma^2, \alpha) = \exp\{\sigma^2 r(s; \alpha)\}$, where σ^2 may now represent σ^2_λ or σ^2_ϕ and α either ϕ_λ or ϕ_ψ depending on whether we are working in the spatial or spatiotemporal settings respectively. Thus, the functions plotted in Figure 3.1 are equivalent to $\log[g(u; \sigma^2, \alpha)]$ with $\sigma^2 = 1$.

Figure 3.2 displays the log-PCFs corresponding to the different correlation functions in Table 3.1, this time for a varying σ^2 and a fixed α. As before, the range of s is kept constant to aid in a visual comparison. It is clear that increasing σ^2 results in 'amplification' of the correlation over the range controlled by α.

It is important to note that these six examples are by no means the only options available to us in terms of defining the correlation structure(s) of the LGCP. Commentary on these and other options, including examples of subsequently generated Gaussian random fields, and what may be considered an 'appropriate' parametric PCF, can be found in Møller et al. (1998).

3.2.2 Concept of minimum contrast

The *minimum contrast* approach is advocated in the seminal works of Møller et al. (1998) and Brix and Diggle (2001) as the best available option for estimation of the correlation parameters of the LGCP (as well as the SNCP in Møller, 2003; Møller and Díaz-Avalos, 2010). As such, it has been employed in the applied modelling of real-world data such

as in Møller et al. (1998), Diggle, Rowlingson and Su (2005) and Møller and Díaz-Avalos (2010). It is therefore of interest to define and examine this technique in the present work; a systematic empirical examination of its performance will be carried out in Section 3.5.

Minimum contrast estimation is used due to the fact that direct likelihood-based inference for the parameters of interest is generally not possible. Such an approach requires direct derivation of the density function for the observed point configuration; a function not analytically tractable in most cases (see for example the comments in Section 7 of Møller et al., 1998). However, and as we shall discuss further in the following two sections, the intuitiveness of the minimum contrast procedure is offset by the numerous subjective decisions that must be made in order to implement the criterion in practice.

In Section 3.2.1 we viewed several possible parametric forms of correlation structures that could be considered appropriate for our purposes. From Section 3.1, recall the various nonparametric definitions of the second-order properties of the LGCP and its modifications. Thus, a natural approach to estimation of the parameters of interest is to minimise some discrepancy between a particular nonparametric function and the elected parametric form thereof.

Let ξ denote a vector of parameters upon which the function J is dependent and allow $\hat{J}$ to represent a nonparametric estimate of J based on the observed data. The minimum contrast 'optimal' estimates of ξ, $\hat{\xi}$, are found by the combination of values in ξ which minimise

$$\mathcal{M}_J(\xi) = \int_{s_0}^{s_{\max}} \omega(s) \left\{ v[\hat{J}(s)] - v[J(s;\xi)] \right\}^2 \mathrm{d}s \approx \mathcal{S}_{\mathrm{diff}} \sum_{s \in \mathcal{S}} \omega(s) \left\{ v[\hat{J}(s)] - v[J(s;\xi)] \right\}^2, \tag{3.26}$$

where s_0 and $s_{\max}$ are the lower and upper lag limits of the contrast criterion, $\omega(\cdot)$ denotes some scalar weight associated with each lag s, $v[\cdot]$ represents some transformation (if any) of its argument, and the approximation to $\mathcal{M}_J$ is obtained simply by summing over a (fine) sequence of lags $\mathcal{S} = \{s_0, s_1, \ldots, s_{\max}\}$ equally spaced so that $\mathcal{S}_{\mathrm{diff}} = s_{a+1} - s_a$; $s_a, s_{a+1} \in \mathcal{S}$.

Minimum contrast is popular due to its ease of computation and intuitively sensible fashion in which we may interpret the criterion. Consider as an example the intensity surface $\lambda(\cdot)$ on the $\mathbb{R}^2$ unit square, which is a single realisation of a stationary and isotropic spatial LGCP with $\rho_\lambda = 1000$, $\sigma_\lambda^2 = 2$, $\phi_\lambda = 0.1$ and a Gaussian correlation structure so $g_\lambda(u) = \exp\{2\exp[-(10u)^2]\}$. A spatial point pattern according to this intensity of size $n \sim \mathrm{POI}(\rho_\lambda)$ observations is generated. The intensity surface and associated data set are displayed in Figure 3.3.

We now use (3.26) in an attempt to estimate the parameters of the process based on the observed point pattern. No deterministic global trend is assumed for this illustrative example; any heterogeneity in the 'observed data' is entirely due to dependence between individual points. We use (1.5) and (1.10) to nonparametrically estimate the K- and pair correlation functions respectively. For the latter, we set $v[\cdot] = \log[\cdot]$ and $(s_0 = 0) < s \leq (s_{\max} = 0.25)$ in the minimum contrast function, and choose h for the kernel κ_h according to the guidelines in Stoyan and Stoyan (1994). The theoretical K-function is obtained using the relationship (1.4), and $v[\cdot] = \sqrt[4]{\cdot}$ with $(s_0 = 0) \leq s \leq (s_{\max} = 0.25)$. There exists no additional information to support using weights, so in both cases we set $\omega(\cdot) = 1$. These

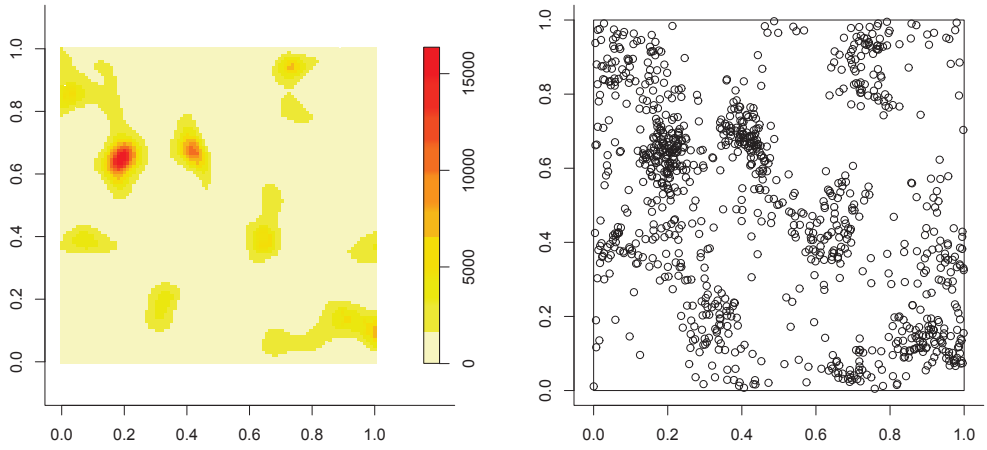


Figure 3.3: A realisation of a LGCP with a Gaussian correlation structure (left) and a subsequently generated point pattern (right).

choices are explained further in the following section.

Figure 3.4 shows the nonparametric estimates of $\hat{K}(s)^{1/4}$ and $\log[\hat{g}(s)]$, along with the identically transformed theoretical versions of these functions defined using the minimum contrast optimal parameters $\hat{\xi} = \{\hat{\sigma}^2_\lambda, \hat{\phi}_\lambda\}$. We note a visually pleasing fit, with the parametric functions closely tracking the trends described by the data. However, the estimated parameters seem slightly disappointing: use of the K-function has provided $\{1.13, 0.079\}$; the PCF has yielded $\{1.15, 0.081\}$. We will investigate the issues of minimum contrast estimate proximity to the true values, as well as whether use of either K or g provides superior estimates over the other, in Section 3.5.

Estimation of θ, for spatiotemporal data, is performed in much the same way by minimising the squared discrepancy between (3.13) and (3.14) for a finite sequence of temporal lags. When including deterministic components as defined at (3.15), the comparison is between the quantities (3.22) and (3.23) which are assumed to be nonstationary in time, and the contrast criterion is

$$\mathcal{M}_{\mathcal{C}}(\theta) = \sum_{\nu=1}^{\nu_{\max}} \sum_{t=\nu+1}^{T} \left[\hat{\mathcal{C}}(t, t-\nu) - \mathcal{C}(t, t-\nu; \theta)\right]^2, \tag{3.27}$$

for some user-specified value of $\nu_{\max} < T$ which may be viewed in the same way as $s_{\max}$ above. Inspection of the theoretical version of the temporal covariance shows it is dependent upon the spatial correlation parameters; these must therefore be estimated first and are subsequently plugged in to the temporal minimum contrast procedure.

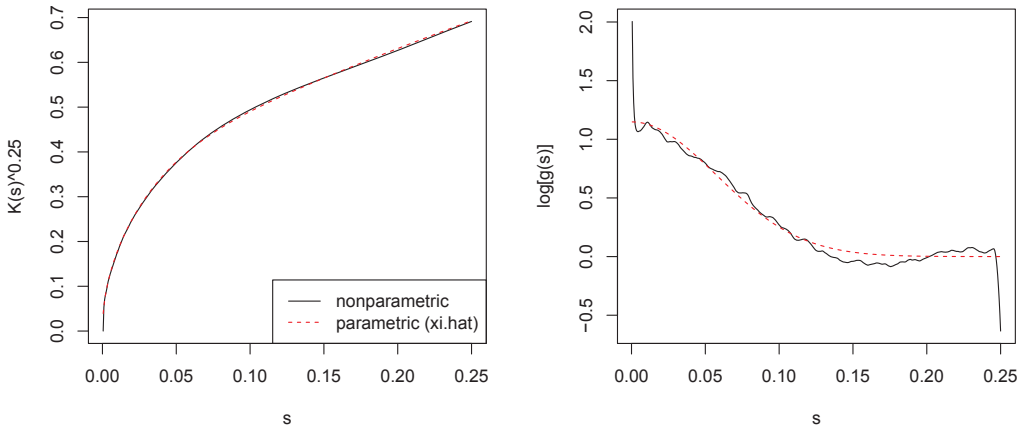


Figure 3.4: Nonparametric estimates of the transformed K- and pair correlation functions for the synthetic data, along with the parametric curves corresponding to the minimum contrast parameters $\hat{\xi}$.

3.2.3 Practical issues

Focusing for the moment on (3.26), there clearly exist undeniable criticisms concerning the arbitrariness of minimum contrast. How do we choose appropriate contrast limits s_0 and $s_{\max}$? How should we weight using $\omega(\cdot)$, if at all? Are the transformations via $v[\cdot]$ important? If performing a contrast procedure using the PCF, how do we choose the bandwidth h?

These are extremely difficult questions to answer. Indeed, universally preferable options may not exist. The convenience and relative computational simplicity of minimum contrast has driven the popularity of this method for estimation of Cox process parameters, and it could well be the case that any deficiencies in terms of accuracy and precision of the resulting estimates can safely be considered negligible in practice. This 'allowable scope' in quality, however, is not well understood, and the importance of sound parameter estimates cannot be ignored.

Some theoretical efforts have been made in the literature to better understand the minimum contrast methods in the stationary Cox process setting. Guan and Sherman (2007) study this approach with respect to the K-function for the LGCP, providing both theoretical and empirical indicators of performance. Ignoring edge-correction, they show that, assuming certain regularity conditions, the corresponding minimum contrast estimates are consistent and asymptotically normally distributed. These results are echoed in a related article by Waagepetersen and Guan (2009) who studied mainly Neyman-Scott processes (a special type of SNCP or 'Poisson cluster process'; see for example Diggle, 2003; Møller and Waagepetersen, 2004). Expressing the minimum contrast criterion as $\int_0^{s_{\max}} \omega(s)[\hat{K}(s)^c - K(s)^c]^2 \, \mathrm{d}s$, Guan and Sherman (2007) argued that a sensible way to choose $s_{\max}$ and c when $\omega(\cdot) = 1$ is to find the combination which minimises their derived asymptotic variance of the corresponding parameter estimates. This approach is of limited use in practice, however. Firstly, the expression for the variance depends upon the value of

the true parameters, meaning implementation with respect to a real-world problem would require pilot estimates of the unknown parameters. Secondly, the authors remark that the performance of the minimum contrast criterion in terms of proximity of the parameter estimates to the true values appeared to be quite sensitive to the choice of these pilot estimates. Thirdly, a simulation study indicated any improvement given by this method of choosing $s_{\max}$ and c in the 'ideal' scenario where the true parameter values were used as the pilots was not substantial when compared to subjective choices of $s_{\max}$ and c based on past empirical evidence of various authors (discussed in a moment). Guan and Sherman (2007) also suggested an alternative sub-sampling approach which can aid in choosing the form of the weight function ω and $s_{\max}$, setting $c = 1$. Again, any practical improvements over the vanilla *ad hoc* minimum contrast rules were difficult to place. Finally, it is not clear if these findings could be applied in a similar fashion if using the pair correlation function rather than K, or if edge-correction would have a significant impact on their results. Although Guan (2009) investigated use of the PCF against K and concluded that use of the PCF provided in general more stable parameter estimates, this study was carried out with respect to Neyman-Scott processes, not the LGCP.

For these reasons, and our practical interest in the quality of minimum contrast estimates, we will not pursue the above techniques further. Our applications involving minimum contrast will calibrate as needed the various components of the criterion based largely on the past numerical experiences of applied researchers, as well as our own experience. Specifically, Diggle (2003) cites empirical evidence to advocate setting $v[\cdot] = \sqrt[4]{\cdot}$ and $\omega(\cdot) = 1$ when using the K-function to estimate spatial correlation parameters for Cox processes where one has observed point aggregation. This is due to the natural alleviation of weight passed to evaluations at large s where we experience greater variance fluctuations. Diggle (2003) also suggests alternatively setting $v[\cdot] = \cdot$ and using ω to weight the discrepancy measure by the inverse (approximate) variance of the K-function (which Guan and Sherman, 2007, used their sub-sampling method to achieve). This author's preference is to use the former approach in order to avoid additional problems such as numerical instability associated with the variance approximation.

Use of the log-transform for the PCF i.e. $v[\cdot] = \log[\cdot]$ was first used in Møller et al. (1998) and subsequently in Diggle, Rowlingson and Su (2005) owing to its stabilisation of extreme values (e.g. those observed for small s in the nonparametric estimator). There exists little if any evidence to suggest using anything other than $\omega(\cdot) = 1$ when performing minimum contrast with the PCF, and as with the K-function, we prefer to reserve use of ω for only those specialised situations in which we possess additional metadata which may assist steering parameter estimation. Implementation of (3.26) with the PCF also requires specification of the bandwidth h for the Epanechnikov smoothing kernel κ in $\hat{g}$. Much of the literature which has dealt with estimation of the PCF has cited the empirical guidelines given by Stoyan and Stoyan (1994) who suggest setting $h = a/\sqrt{N(W)/|W|}$ i.e. some constant a divided by the square root of the overall intensity measure (as per the definitions in Section 1.2.2), where $0.1 \leq a \leq 0.2$. This author's early empirical experimentations have presented only small differences in the arrived parameter estimates for a varying a within this range, and we therefore elect to set $a = 0.15$ in all subsequent implementations. Guan

(2007) did suggest a least-squares cross-validation method for selection of h in the PCF, further developed in Guan (2009). Again, however, these were evaluated with respect to Poisson cluster processes, not LGCPs. Empirical results in Guan (2007) showed similar performance to bandwidths selected using Stoyan & Stoyan's (1994) guidelines when dense point clustering was observed. This, coupled with the additional computational expense associated with the LSCV selector, means we do not consider it further in our studies.

The contrast limits are the final consideration. It is natural to set $s_0 = 0$ but there is an important distinction between the K and PCF versions of (3.26) in that the K-function can be evaluated at zero but the PCF cannot. As a result, for the PCF minimum contrast procedures we follow Møller et al. (1998) and set s_0 as the smallest interpoint distance in the observed data so that $s_0 = \min(||\boldsymbol{x}_i - \boldsymbol{x}_j||); \quad i \neq j$. Note that an inappropriate choice of s_0 (e.g. a value too close to zero) can have adverse affects on the performance of $\mathcal{M}_g$, even with the log-transformation present. The upper limit $s_{\max}$ can be considered less important provided the transformation function is employed sensibly such as above. Of course, we are restricted by the limits imposed by the edge-correction in the nonparametric estimators, but we are generally not interested in the second-order properties at such large distances anyway, particularly for clustered point patterns. A good rule of thumb, which is supported by comments in Diggle (2003), is to set $s_{\max} = \min(x_W, y_W)/4$ where x_W and y_W represent the maximum width and height of the observation region W respectively. Note also that this is the default approach employed in the comprehensive spatial point pattern software package **spatstat** in R (Baddeley and Turner, 2005). Alternatively this can be determined by trial-and-error, by inspecting plots of the nonparametric trend of the PCF for different $s_{\max}$ and setting this upper contrast limit at a lag where the dependence appears to drop to a negligible level (i.e. zero on the log-scale).

More emphasis has been placed in the preceding discussion of minimum contrast estimation on the spatial margin, for several reasons. Firstly, we have two equally natural options to estimate the second-order behaviour of the spatial aspect, namely g and K, though the single most intuitively sensible option for gauging dependence in time is arguably the temporal covariance function. Secondly, there tend to be more issues such as the PCF bandwidth h, and edge-correction over W, which require consideration. Thirdly, and perhaps most importantly, estimation of the temporal correlation parameter θ depends itself on the estimated values of σ^2_ψ and ϕ_ψ. If these are poorly estimated, the adverse effect on $\hat{\theta}$ could be severe. We revisit minimum parameter estimation in Sections 3.4 and 3.5 for both real-world and synthetic data.

3.2.4 Numerical likelihood-based alternatives

Møller et al. (1998) also note that it is possible to use a Monte-Carlo-estimated likelihood function to obtain estimates of the required quantities, but favour minimum contrast due to potentially restrictive multimodality in the latter. Brix and Diggle (2001) support this decision, and add that any theoretical shortcomings of minimum contrast are mitigated when we expect to deal with large data sets.

We pay particular attention to the minimum contrast technique for estimation of the Cox

process parameters here due to its prevalence in the applied literature. Further technicalities concerning likelihood-based methods can be found in Møller and Waagepetersen (2004); for a short summary including useful references the reader is also directed to Section 7.2 of Illian et al. (2008).

3.3 Computational considerations

The following discussion focuses on the computational aspects most relevant to working with the LGCP. In particular, we follow Appendix E in Møller and Waagepetersen (2004) and detail how the *discrete Fourier transform* (DFT) is considered in order to exploit massive reductions in computational costs via the *fast-Fourier transform* (FFT) when simulation of the specified Gaussian fields is required. The aim is to augment Møller and Waagepetersen's (2004) impressive yet theory-dominated instructions with the provision of a more conceptual description appropriate for the nature of this thesis. We also review the *Metropolis-adjusted Langevin algorithm* (MALA) in the context of the LGCP; the Markov-chain Monte-Carlo technique necessary for simulation of our models conditional upon observed data.

3.3.1 General simulation: the fast-Fourier transform

As we encountered in Section 3.1.2, it is necessary to discretise the temporal margin in order to theoretically analyse the LGCP. In practice, the spatial margin must also be discretised in order to implement the methodology. This is achieved by partitioning the finite spatial study region W into a suitably fine set of disjoint cells. Note therefore that the following discussion holds for both fields $\mathcal{Y}$ (purely spatial) and $\mathcal{Z}$ (spatiotemporal), when the latter is evaluated at a single time point t. We will work only with the former symbol here.

Though we would generally expect W to be an irregular polygon on $\mathbb{R}^2$, the simplest discretisation strategy remains to construct a $M \times N$ grid over the encapsulating rectangle $\mathcal{A}_W$; $M, N > 2$. Represent the maximum ranges spanning W (and hence $\mathcal{A}_W$) along the x- and y-axes with $R_x(\mathcal{A}_W)$, and $R_y(\mathcal{A}_W)$ respectively. Defining $\Delta_x = R_x(\mathcal{A}_W)/(M-1)$ and $\Delta_y = R_y(\mathcal{A}_W)/(N-1)$ to be the horizontal and vertical grid spacings respectively, we re-define $\mathcal{A}_W$ in terms of a lattice; the discretised interval I. This is given as

$$I = \big[(x_{\min} + \Delta_x m, y_{\min} + \Delta_y n) : m \in \{0, \ldots, M-1\}, n \in \{0, \ldots, N-1\}\big], \tag{3.28}$$

where obviously $R_x(I) \equiv R_x(\mathcal{A}_W)$ and $R_y(I) \equiv R_y(\mathcal{A}_W)$, and $x_{\min}$, $y_{\min}$ are the lowest (bottom left) coordinates of $\mathcal{A}_W$.

Considering the fact that important calculations such as finding inter-cell distances and determining region inclusion are defined in terms of the cell centroids, it is useful to also explicitly state these as

$$\begin{aligned} C^{(m,n)} = &\big[[x_{\min} + (m-0.5)\Delta_x, y_{\min} + (n-0.5)\Delta_y] : \\ &m \in \{1, \ldots, M-1\}, n \in \{1, \ldots, N-1\}\big]. \end{aligned} \tag{3.29}$$

In fact, it is this grid design coupled with the absence of edge-effect concerns in the LGCP,

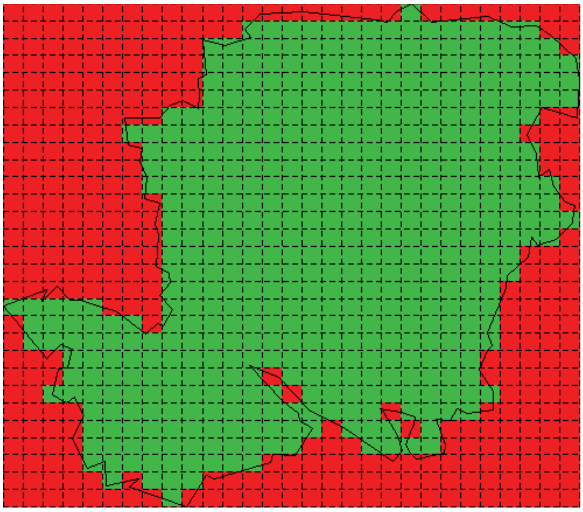


Figure 3.5: A rectangular grid approximating an irregular spatial region.

which enables us to access the FFT methods described in a moment.

From (3.28), simulation of the spatially continuous Gaussian field $\mathcal{Y}$ will therefore approximately translate to the simulation of a high-dimensional Gaussian variable with a non-trivial correlation structure $Y(\boldsymbol{u})$; $\boldsymbol{u} \in C$, such that

$$Y \sim \mathrm{N}(\boldsymbol{\mu}, \Sigma_Y), \tag{3.30}$$

where, according to our assumptions about the process, $\boldsymbol{\mu}$ is the $[(M-1)\times(N-1)]\times 1$ vector of the stationary process mean μ, and Σ_Y is the $[(M-1)(N-1)]\times[(M-1)(N-1)]$ covariance matrix. The covariance matrix is computed from the $[(M-1)(N-1)] \times [(M-1)(N-1)]$ cell-centroid-wise matrix of Euclidian distances D, whereby

$$D[i,j] = ||C^{(\boldsymbol{i})} - C^{(\boldsymbol{j})}||;$$

$$i \to \boldsymbol{i}, j \to \boldsymbol{j} = \big[(m,n) : m \in \{1,\ldots,M-1\}, n \in \{1,\ldots,N-1\}\big],$$

so $i, j \in \{1,\ldots,(M-1)(N-1)\}$, thus allowing

$$\Sigma_Y[i,j] = \sigma^2 r(D[i,j];\phi), \tag{3.31}$$

where r is the stationary, isotropic correlation function (possibly one of those described in Section 3.2.1). The stationary scalar mean, variance and correlation parameters, μ, σ^2 and ϕ respectively, can represent either μ_λ, σ^2_λ, ϕ_λ or μ_ψ, σ^2_ψ, ϕ_ψ depending on whether we are examining a spatial or spatiotemporal data set as per our earlier expositions in Sections 3.1.1 and 3.1.2. Note also that $i \to \boldsymbol{i}$ signifies the fact that a unique scalar cell indexer i translates to a unique combination of the cell centroid coordinates $\boldsymbol{i}$.

Consider Figure 3.5, which shows some irregular spatial region upon which we have superimposed I constructed as an evenly-spaced rectangular grid with $M = N = 30$ (coarse for illustrative purposes). We therefore have an overall grid approximation of 29×29 cells;

of interest are obviously those which fall inside the geographical region (green). These are the cells whose centroids C are inside the irregular polygon i.e. $C^{(m,n)} \in W$ for interior cell indices m and n from the sets in (3.29). Nevertheless, we repeat that it is easiest (and necessary for the FFT) to consider all cells over the encapsulating rectangle.

Immediately the computational ramifications of modelling our problems as a function of a correlated Gaussian field become apparent. Direct simulation from such a process for even inappropriately coarse grids carries massive computational requirements; for example in the case of the illustrative cell assignment of 29×29 we must already consider eigendecomposition of an 841×841 covariance matrix. Using standard methods (e.g. Cholesky decompositions) the associated costs are so high that many computer languages capable of this level of mathematical operation will struggle to run implemented programs based on (3.30) directly. On the latest desktop machines R, for example, will cite inadequate memory allocation and refuse to even begin evaluation for grids little finer than the above size.

Møller et al. (1998) suggested a clever solution to this problem from work in Wood and Chan (1994) using the two-dimensional discrete Fourier transform and the subsequent advantages of associated quick-evaluation algorithms termed *fast-Fourier transforms* (FFT). Briefly, the FFT makes use of specialised complex number transformations which in our context allows the derivation of eigenvalues of *block circulant* matrices at a dramatically reduced cost of computational resources (for simulation involving component-dependent multivariate Gaussian variables, the eigendecomposition of the covariance matrix is required).

Albeit symmetrical, the covariance matrix Σ_Y is however in its current form not block circulant. To satisfy this requirement of the FFT, *circulant embedding* is performed using an extended rectangular grid which must be at least twice the size in both axes as the original lattice. This is achieved as follows. First define the extended grid resolution as $M_{\text{ext}} \times N_{\text{ext}}$, with $M_{\text{ext}} \geq (2M-1)$, $N_{\text{ext}} \geq (2N-1)$ so we have

$$I_{\text{ext}} = \big[(x_{\min} + \Delta_x m, y_{\min} + \Delta_y n) : m \in \{0, \ldots, M_{\text{ext}} - 1\}, n \in \{0, \ldots, N_{\text{ext}} - 1\}\big] \quad (3.32)$$

and

$$\begin{aligned} C_{\text{ext}}^{(m,n)} = &\big[[x_{\min} + (m - 0.5)\Delta_x, y_{\min} + (n - 0.5)\Delta_y] : \\ &m \in \{1, \ldots, M_{\text{ext}} - 1\}, n \in \{1, \ldots, N_{\text{ext}} - 1\}\big]. \end{aligned} \quad (3.33)$$

The approximation is now given as $Y_{\text{ext}}(\boldsymbol{u})$, $\boldsymbol{u} \in C_{\text{ext}}$;

$$Y_{\text{ext}} \sim \text{N}(\boldsymbol{\mu}_{\text{ext}}, \Sigma_{Y_{\text{ext}}}), \quad (3.34)$$

where μ_{ext} and $\Sigma_{Y_{\text{ext}}}$ are defined as in (3.30), except now increased in size to correspond to the extended interval I_{ext}. Our example lattice in Figure 3.5 is extended the minimum allowable lengths such that I_{ext} is a 59×59 grid with 58×58 cells. This is re-plotted in Figure 3.6.

Secondly, we must define the extended cell-centroid-wise Euclidian distance matrix. This is not, however, computed with 'raw' Euclidian distances between distinct centroids of the extended lattice. To satisfy the circulant requirement of the covariance matrix, the distances

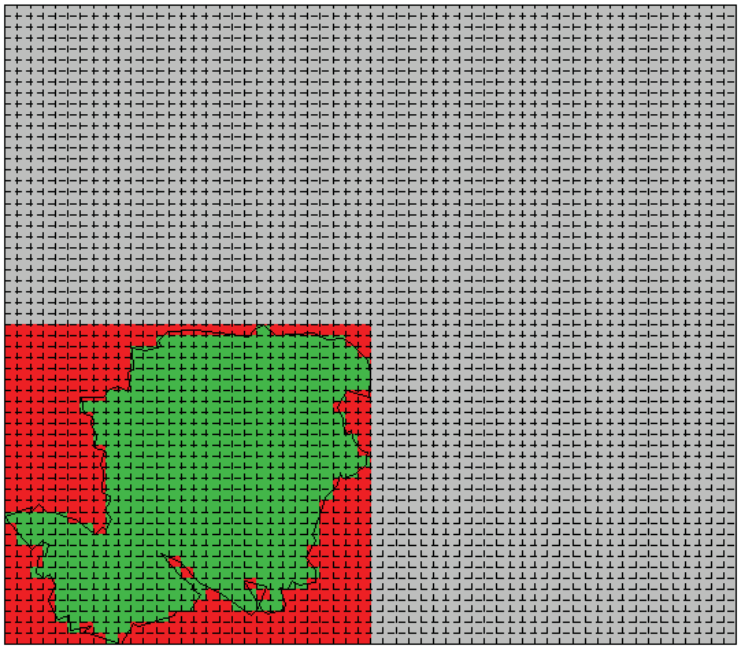


Figure 3.6: The extended rectangular grid in preparation for circulant embedding of cell-wise distances.

upon which this is defined must correspond to wrapping the extended lattice on a torus and only thereafter computing the shortest Euclidian distances. Consequently, the definition of D_{ext} differs from the definition of D for (3.31) in the sense that

$$D_{\text{ext}}[i,j] = \Big\{ \min\left[\text{abs}_m(\boldsymbol{i},\boldsymbol{j}), R_x(I_{\text{ext}}) - \text{abs}_m(\boldsymbol{i},\boldsymbol{j})\right]^2 + \min\left[\text{abs}_n(\boldsymbol{i},\boldsymbol{j}), R_y(I_{\text{ext}}) - \text{abs}_n(\boldsymbol{i},\boldsymbol{j})\right]^2 \Big\}^{\frac{1}{2}}, \tag{3.35}$$

where $\text{abs}_m(\boldsymbol{i},\boldsymbol{j}) = |C^{(\boldsymbol{i})}_{\text{ext},m} - C^{(\boldsymbol{j})}_{\text{ext},m}|$ and $C^{(\boldsymbol{i})}_{\text{ext},m}$ denotes the horizonal centroid coordinate of the cell in I_{ext} indexed by $i \to \boldsymbol{i}$; subscript n the vertical, so finally,

$$\Sigma_{Y_{\text{ext}}}[i,j] = \sigma^2 r(D_{\text{ext}}[i,j];\phi); \qquad i,j \in \{1,\ldots,(M_{\text{ext}}-1)(N_{\text{ext}}-1)\}. \tag{3.36}$$

Figure 3.7 displays the difference between simply evaluating the 'raw' Euclidian distances on I_{ext} (left panel) and smallest Euclidian distance around the torus (right) for two arbitrarily chosen centroids. Note therefore that D_{ext} is correctly defined in terms of the minimum Euclidian distance on the torus-wrapped extended lattice by selecting the shortest of the two for each centroid pair. Figure 3.8 shows the extended lattice in Figure 3.6 wrapped on a ring torus in abstract space. The distinction between 'abstract space' and $\mathbb{R}^2$ is made here to clarify the fact that it is not the physical surface area on this 'inflated' torus that is used to compute inter-cell differences; to obtain the distances highlighted in Figure 3.7 we are essentially working on a uniformly 'deflated' torus.

Thirdly, though it may seem somewhat counter-intuitive to *increase* the dimension of our computationally costly problem, by definition (3.36) $\Sigma_{Y_{\text{ext}}}$ is now block circulant, and

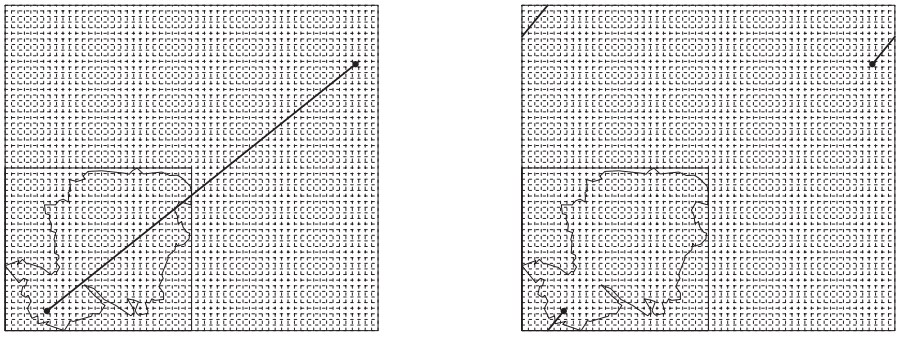


Figure 3.7: Comparing raw (left) and torus (right) Euclidian distance calculation between two distinct cell centroids on the extended lattice.

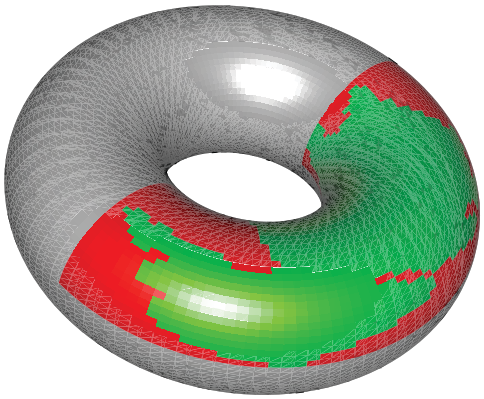


Figure 3.8: The extended lattice with irregular polygon wrapped on a torus in abstract space.

its spectral decomposition can therefore be computed efficiently using the FFT. It turns out that the required eigenvalues can be derived using only the first row of $\Sigma_{Y_{\text{ext}}}$, meaning we must only initially compute and consider a $(M_{\text{ext}}-1)\times(N_{\text{ext}}-1)$ matrix rather than, at the very least, the original $(M-1)(N-1)\times(M-1)(N-1)$ covariance matrix Σ_Y.

To elaborate a little further, let F_L (L>0) denote the normalised $L\times L$ DFT matrix, such that

$$F_L[m,n]=\frac{\exp\{-\hat{\imath}2\pi mn/L\}}{\sqrt{L}};\qquad m,n\in\{0,\ldots,L-1\},\tag{3.37}$$

where $\hat{\imath}$ is the complex imaginary unit. Now, let $\tilde{\Sigma}_{Y_{\text{ext}}}$ represent the first row of $\Sigma_{Y_{\text{ext}}}$ arranged, as mentioned above, as a $(M_{\text{ext}}-1)\times(N_{\text{ext}}-1)$ matrix. Then the $(M_{\text{ext}}-1)\times(N_{\text{ext}}-1)$ matrix of eigenvalues, $\tilde{\Lambda}$, which corresponds to the full, extended, torus-wrapped covariance matrix $\Sigma_{Y_{\text{ext}}}$, can be computed quickly by

$$\tilde{\Lambda}=\sqrt{(M_{\text{ext}}-1)(N_{\text{ext}}-1)}\bar{F}_{M_{\text{ext}}-1}\tilde{\Sigma}_{Y_{\text{ext}}}\bar{F}_{N_{\text{ext}}-1},\tag{3.38}$$

where $\bar{F}_L$ denotes the complex conjugate of (3.37). To then produce a realisation of the Gaussian field on the extended lattice, Y_{ext} as per (3.34), the steps are as follows. First, compute the 'dummy' matrix $\tilde{H}$, such that

$$\tilde{H}=F_{M_{\text{ext}}-1}UF_{N_{\text{ext}}-1},\tag{3.39}$$

where U is a trivial $(M_{\text{ext}}-1)\times(N_{\text{ext}}-1)$ matrix containing independent standard normal variables. Transform $\tilde{H}$ into H by the one-to-one operation

$$H=\tilde{H}\circ\tilde{\Lambda}^{\odot 1/2},\tag{3.40}$$

where we use $A^{\odot a}$ to denote the element-wise power of the matrix A to the scalar a, and $\circ$ the Hadamard product. Finally, the realisation on the extended lattice is found with

$$Y_{\text{ext}}=\bar{F}_{M_{\text{ext}}-1}H\bar{F}_{N_{\text{ext}}-1}+\boldsymbol{\mu}_{\text{ext}},\tag{3.41}$$

and then any cells in Y_{ext} which correspond possess centroids falling outside of W are discarded to in order to obtain the realised field within the study region of interest. Note that use of FFT methods occurs implicitly in the computer implementation of (3.38), (3.39) and (3.41).

Some additional notes are warranted. The validity of these DFT steps is contingent upon the covariance matrix $\Sigma_{Y_{\text{ext}}}$ being positive semidefinite i.e. that its eigenvalues are non-negative, due to the step in (3.40). Møller et al. (1998) note that positive semidefiniteness is not guaranteed in practice, though this has rarely, if ever, been a problem provided suitably-sized values of M_{ext} and N_{ext} are used. Indeed, the issue of negative eigenvalues has not arisen in this author's experience. It is also a well-known fact (we defer to Wood and Chan, 1994, for details) that the FFT quick-evaluation algorithms perform at their best in terms of computational speed and efficiency when the grid cell resolutions are powers of two. In most applications, this author has found that setting $(M-1)=(N-1)=2^6=64$ or

$(M-1) = (N-1) = 2^7 = 128$ provides an acceptable balance between aesthetically pleasing images and computational expense. Of course, the grid resolution must be fine enough to adequately capture the scale of dependence controlled predominantly by ϕ; it could well be the case that even higher values of M and N are required for a given application.

From a programming perspective, the need for computationally efficient functions is critical with respect to practical implementation of the spatial and spatiotemporal LGCP. This is especially true when we consider the iterative nature of corresponding methods for conditional simulation (discussed in the following section). Where multiple generations of high-dimensional Gaussian random fields are required, the circulant embedding/FFT methodology is, quite simply, essential.

3.3.2 Conditional simulation: the Metropolis-adjusted Langevin algorithm

The preceding discussion focused on how circulant embedding of the covariance matrix for a discretised approximation to a spatially continuous Gaussian field can result in a significant decrease in computational requirements for simulation of said field via the FFT. By "simulation of [the] field", we refer to the fact that, given a correlation function r and values or estimates of the parameters μ, σ^2 and ϕ, we may subsequently generate a realisation of the Gaussian variable with this specific covariance structure.

Alone, this is termed *unconditional simulation*, owing to the fact that the realisations of the process have been generated entirely at random i.e. without any need to take into account observed data. Clearly, if we assume the spatial or spatiotemporal intensities of a given point pattern (or at least a component thereof) is defined in terms of a LGCP, then the observed data must play a crucial role in quantifying the assumed process. This *conditional simulation* of LGCPs, hence Gaussian fields, is of course not at all straightforward due to their stochastic nature.

Conditional simulation of the LGCP is therefore also referred to as *prediction*: based on data we aim to predict the specific form of the stochastic intensity process instrumental in their generation. Though the conditional distribution of the stochastic process given the data is analytically intractable (Møller et al., 1998), using Bayes' theorem it is possible to express this function up to a constant of proportionality. This motivates a Bayesian approach whereby characteristics of the posterior can be estimated through long-run averages of a suitably designed Markov chain. It is assumed the reader is familiar with the basic premise of Bayesian analysis and Markov chain Monte Carlo (MCMC); a brief introduction of the methods and terminology relevant for our purposes is given here.

At the heart of MCMC is the *random walk Metropolis-Hastings* (MH) *algorithm* (Metropolis et al., 1953; Hastings, 1970). Suppose we have some objective variable of interest, V, which we wish to simulate based on the *target density* thereof, f. Given that the current position, or *state*, of the chain is $V^{(t)}$, we

1. Generate a *candidate* $V' \sim q[\cdot|V^{(t)}; h]$.

2. Take

$$V^{(t+1)} = \begin{cases} V' & \text{with probability} \quad p(V', V^{(t)}) \\ V^{(t)} & \text{otherwise.} \end{cases}$$

where

$$p(x, y) = 1 \wedge \frac{f(x)q[y|x; h]}{f(y)q[x|y; h]} \tag{3.42}$$

3. Repeat 1. and 2. until the desired length of the chain is reached.

Here, p is called the *acceptance probability* and q is referred to as the *candidate* or *proposal* distribution. A candidate for the next transition is generated according to this distribution, which is conditional upon the current state of the chain and controlled by a *tuning* parameter $h > 0$. Alternatively we may write the algorithm as having $q(\cdot\,|V^{(t)}; h) = V^{(t)} + \epsilon_t$, where $\epsilon \sim q_\epsilon(h)$ and $q_\epsilon(h)$ is some standardised error distribution independent of $V^{(t)}$.

Thus, the MH algorithm is useful when we are able to evaluate the target density f up to a constant of proportionality but not directly, due to the cancellation which naturally occurs in computation of p. By subsequently obtaining multiple instances of variables $V \sim f$, it is possible to therefore draw conclusions about the general posterior distribution of f.

The *Metropolis-adjusted Langevin algorithm* (MALA), also known as the *Langevin-Hastings algorithm*, was first suggested in statistics in a discussion paper by Besag (1994); a detailed study by Roberts and Tweedie (1997) followed. The MALA is essentially a more sophisticated random walk MH algorithm, where gradient information on the target variable is exploited in order to assist steering the suggested candidates toward higher areas of the target density. To denote this dependence of q on the gradient function, we write $q[\cdot\,|V^{(t)}; h, \nabla(V^{(t)})]$, where $\nabla(x) = \partial \log[f(x)]/\partial x$. This, for the MALA, takes the specific form $q[\cdot\,|V^{(t)}; h, \nabla(V^{(t)})] = V^{(t)} + 0.5h\nabla(V^{(t)}) + \sqrt{h}\epsilon_t$, where $\epsilon \sim q_\epsilon(\cdot)$. Note that the standardised error distribution q_ϵ for the MALA in this form does not need to account for the tuning parameter h as it already scales the deterministic part of the candidate generating function. Following re-specification of q, progression of the MALA then follows the same steps as above.

Møller et al. (1998) advocate use of the MALA over a vanilla random walk for conditional simulation of spatial LGCPs due to superior convergence rates of the chains as a result of including the gradient information. This was echoed in Brix and Diggle (2001) for the spatiotemporal LGCP. Superior behaviour in terms of mixing characteristics of the MALA chain for the LGCP was also apparent in Christensen et al. (2001) and Christensen and Waagepetersen (2002). Below we summarise the quantities necessary to implement the MALA in both spatial and spatiotemporal settings. Further details are can be found in the above papers, as well as in Chapter 10 of Møller and Waagepetersen (2004).

Spatial

The variable of interest is the Gaussian random field $Y(m, n)$ for cell centroids $(m, n) \in C$ as per (3.29). To take advantage of the FFT, we extend our field of interest to $Y_{\text{ext}}(m, n)$; $(m, n) \in C_{\text{ext}}$ (3.33). Take Q to be the matrix which diagonalises $\Sigma_{Y_{\text{ext}}}$ and for ease of notation set $d = (M_{\text{ext}} - 1)(N_{\text{ext}} - 1)$. Then, it is possible to express Y_{ext} as a $d \times 1$ vector

$\tilde{Y}_{\text{ext}}$ such that

$$\tilde{Y}_{\text{ext}} = Q\Gamma + \mu_\lambda \tag{3.43}$$

where $\Gamma \sim N_d(0, \boldsymbol{I})$, N_d denotes the d-dimensional standard normal distribution, and $\boldsymbol{I}$ the d-dimensional identity matrix. In addition to a slight reduction in computational complexity, Møller et al. (1998) note that working with Γ rather than $\tilde{Y}_{\text{ext}}$ directly when constructing the Markov chain appeared to improve its mixing, leading to an accelerated drop in autocorrelation between successively accepted draws from the target density. It is also important to note that obtaining $\tilde{Y}_{\text{ext}}$ from Γ via (3.43) is a relatively trivial operation when employing the Fourier transforms; indeed Q need not be computed explicitly. Once $\tilde{Y}_{\text{ext}}$ has been recovered, this is easily mapped back to the original lattice I_{ext}, and stored in a relevant $(M_{\text{ext}} - 1) \times (N_{\text{ext}} - 1)$ matrix. The lack of concern over edge effects means we simply discard all cells whose centroids $C_{\text{ext}} \notin W$, and we are left with an appropriate realisation of the Gaussian field, conditional upon the data, within the study region of interest.

Note that the posterior of concern is conditional upon the observed point pattern, in the sense we are interested in $f(\gamma|\boldsymbol{X})$ (the conditional density of $\Gamma|\boldsymbol{X}$). Møller et al. (1998) provide the relevant formula.

$$\begin{aligned}\log\left[f(\gamma|\boldsymbol{X})\right] = \text{constant} + \sum_{(m,n)\in C_{\text{ext}}} \Big\{ & \tilde{y}_{\text{ext}}(m,n)\mathcal{N}(m,n) \\ & - \exp[\tilde{y}_{\text{ext}}(m,n)A(m,n)]\Big\} - 0.5||\gamma||^2,\end{aligned} \tag{3.44}$$

where $\tilde{y}_{\text{ext}}$ is obtained from the particular γ via (3.43), $\mathcal{N}(m,n)$ denotes the total number of observations falling inside the cell referenced by its centroid at (m,n), and A represents the area of each cell i.e. $A = \Delta_x \times \Delta_y$ as per (3.32).

Denote the $d \times 1$ vector of cell counts corresponding to the order of $\tilde{y}_{\text{ext}}$ by $\tilde{\mathcal{N}}_{\text{ext}}$, and the cell area vector by $\tilde{A}_{\text{ext}}$. The gradient of (3.44) is then the $d \times 1$ vector given with

$$\nabla(\gamma) = \frac{\partial \log[f(\gamma|\boldsymbol{X})]}{\partial\gamma} = Q\left[\tilde{\mathcal{N}}_{\text{ext}} - \exp\{\tilde{y}_{\text{ext}}\} \circ \tilde{A}_{\text{ext}}\right] - \gamma; \tag{3.45}$$

recall $\circ$ represents the Hadamard product. It is important to note that in (3.44) we set $\mathcal{N}(m,n) = A(m,n) = 0$ if $(m,n) \notin W$; resulting in the vectors $\tilde{\mathcal{N}}_{\text{ext}}$ and $\tilde{A}_{\text{ext}}$ in (3.45) being altered accordingly.

Now suppose $\gamma^{(k)}$ is the current state of the chain. For the LGCP, $q_\epsilon = N_d(0, \boldsymbol{I})$, which translates to the proposed candidate, γ', being generated as

$$\gamma' = \Gamma' \sim q[\Gamma'|\gamma^{(k)}; h, \nabla(\gamma^{(k)})] = N_d(\gamma^{(k)} + 0.5h\nabla(\gamma^{(k)}), h\boldsymbol{I}). \tag{3.46}$$

Note that the h therefore represents the (stationary) variance of the mutually independent

multivariate Gaussian distribution. The move from $\gamma^{(k)}$ to γ' is accepted with probability

$$\begin{aligned} p(\gamma', \gamma^{(k)}) &= \frac{f(\gamma'|\boldsymbol{X})q[\gamma^{(k)}|\gamma'; h, \nabla(\gamma')]}{f(\gamma^{(k)}|\boldsymbol{X})q[\gamma'|\gamma^{(k)}; h, \nabla(\gamma^{(k)})]} \\ &= \frac{f(\gamma'|\boldsymbol{X})\text{exp}\{-||\gamma^{(k)} - \gamma' - 0.5h\nabla(\gamma')||^2/(2h)\}}{f(\gamma^{(k)}|\boldsymbol{X})\text{exp}\{-||\gamma' - \gamma^{(k)} - 0.5h\nabla(\gamma^{(k)})||^2/(2h)\}}. \end{aligned} \tag{3.47}$$

using the definitions in (3.44) and (3.45).

To begin the chain, we must specify an initial state, $\gamma^{(0)}$. In spite of better performance of the MALA over the random walk in terms of accelerated convergence, Christensen et al. (2003) demonstrated the apparent increased sensitivity of the MALA chain to the choice of the initial state. They concluded that, for the LGCP, electing $\gamma^{(0)} = \Gamma^{(0)} \sim N_d(0, \boldsymbol{I})$ provided the most stable performance; this author's experience supports their findings.

Setting h is a little more difficult. However, there exists evidence in the literature to support choosing this variance of the proposal distribution to achieve an overall acceptance rate of around 0.574 to provide the fastest rates of convergence (Roberts and Rosenthal, 1998). The optimal value of h will differ according to various aspects of the problem such as d and chain length, and is typically set following examination of some shorter preliminary runs of the algorithm.

Should we include a multiplicative deterministic intensity ζ_λ in our model as described in Section 3.1.3, this changes the specific form of the target density and gradient functions. The target (3.44) becomes

$$\begin{aligned} \log[f(\gamma|\boldsymbol{X})] = \text{constant} + \sum_{(m,n)\in C_{\text{ext}}} \Big\{ &[\log\{\zeta_\lambda(m,n)\} + \tilde{y}(m,n)]\mathcal{N}(m,n) \\ &- \zeta_\lambda(m,n)\text{exp}[\tilde{y}(m,n)]A(m,n)\Big\} - 0.5||\gamma||^2, \end{aligned} \tag{3.48}$$

and (3.45) becomes

$$\nabla(\gamma) = \frac{\partial \log[f(\gamma|\boldsymbol{X})]}{\partial \gamma} = Q\left[\tilde{\mathcal{N}}_{\text{ext}} - \tilde{\zeta}_{\text{ext},\lambda} \circ \text{exp}\{\tilde{y}_{\text{ext}}\} \circ \tilde{A}_{\text{ext}}\right] - \gamma, \tag{3.49}$$

where $\tilde{\zeta}_{\text{ext},\lambda}$, is the vectorised version of $\zeta_\lambda(m,n)$ with order corresponding to $\tilde{y}_{\text{ext}}$. Like $\mathcal{N}(m,n)$ and $A(m,n)$, $\zeta_\lambda(m,n)$ and the relevant entries of $\tilde{\zeta}_{\text{ext},\lambda}$ are set to zero for $m, n \notin W$. The acceptance probability retains the same form as in (3.47).

Spatiotemporal

Dealing with a spatiotemporal LGCP MALA means we must consider multiple spatial intensities *at each iteration*, rather than just single realisations as above, due to the effect of the temporal correlation on the spatial variation. Under the modelling assumptions made in Section 3.1.2, Brix and Diggle (2001) simulate $\Gamma_j \sim N_d(0, \boldsymbol{I})$ such that

$$\tilde{Z}_{\text{ext},1} = Q\Gamma_1 + \mu_\psi; \quad \tilde{Z}_{\text{ext},j} = Q\Gamma_j + \mu_\psi[1 - \tau(1;\theta)] + \tau(1;\theta)\tilde{Z}_{\text{ext},j-1}; \quad j \in \{2, 3, \ldots, T\}; \tag{3.50}$$

recall that τ is the temporal correlation function and unit time intervals dictate the form of the subsequent γs at times $j > 1$. We use Z to denote the spatially discretised $\mathcal{Z}$; $\tilde{Z}_{\text{ext},t}$ is the $d \times 1$ vector giving the torus-wrapped Gaussian field on the extended spatial lattice at time t.

In disease surveillance, the goal is typically to simulate the intensity at the latest time point for which data have been observed, given data up until this time i.e. $\tilde{Z}_{\text{ext},T}$ given $\boldsymbol{X}_{1:T}$. However, and as Brix and Diggle (2001) are quick to point out, the nature of the model dictates it is easier to simulate $\tilde{Z}_{\text{ext},1:T}$ given $\boldsymbol{X}_{1:T}$; owing to the fact that we only know the likelihood function up to a constant of proportionality for this situation, and not $\tilde{Z}_{\text{ext},T}|\boldsymbol{X}_{1:T}$.

In practice, for large data sets, the simulation of T spatial intensities at each iteration of the chain presents another computational efficiency problem. This is avoided by considering only the most recent s times with data; a sensible solution given that we would expect the distant past to have an ever-diminishing impact on the state of the intensity at the present. Let $U = T - s + 1$. We are only interested in, and hence only simulate, $\tilde{Z}_{\text{ext},U:T}$ given $\boldsymbol{X}_{U:T}$. The value of s is chosen according to the strength of temporal correlation for the application at hand, which we recall is controlled by θ. We will think along these lines for the remainder of the definitions, which means the simulation design in (3.50) changes with $\Gamma_1 = \Gamma_U$ and $j \in \{U+1, U+2, \ldots, T\}$.

The log-target in the spatiotemporal setting is given by Brix and Diggle (2001) as

$$\begin{aligned}\log[f(\gamma_{U:T}|\boldsymbol{X}_{U:T})] = \text{constant} + \sum_{t=U}^{T} \sum_{(m,n)\in C_{\text{ext}}} \Big\{ & \tilde{y}_{\text{ext},t}(m,n)\mathcal{N}_t(m,n) \\ & - \exp[\tilde{y}_{\text{ext},t}(m,n)]A(m,n)\Big\} - \sum_{t=U+1}^{T} \left\{ \frac{||\gamma_t||^2}{2[1-\tau(1;\theta)^2]} \right\} \\ & - 0.5||\gamma_U||^2, \end{aligned} \tag{3.51}$$

where $\mathcal{N}_t(m,n)$ denotes the number of observations falling in the cell with centroid (m,n) at time t.

As earlier, let $\tilde{\mathcal{N}}_{\text{ext},t}$ represent the $d \times 1$ vector of cell counts at time t with order corresponding to $\tilde{y}_{\text{ext}}$. We express the gradient as the $d \times s$ array with

$$\nabla(\gamma_{U:T}) = \left\{ \left[\frac{\partial}{\partial \gamma_U} \log f(\gamma_{U:T}|\boldsymbol{X}_{U:T}) \right], \ldots, \left[\frac{\partial}{\partial \gamma_T} \log f(\gamma_{U:T}|\boldsymbol{X}_{U:T}) \right] \right\},$$

where

$$\begin{aligned}\frac{\partial}{\partial \gamma_t} \log f(\gamma_{U:T}|\boldsymbol{X}_{U:T}) = & \mathbf{1}[t < T] \sum_{i=t}^{T-1} \left\{ Q\left[\tau(1;\theta)(\tilde{\mathcal{N}}_{\text{ext},T+t-i} - \exp[\tilde{y}_{\text{ext},T+t-i}] \circ \tilde{A}_{\text{ext}})\right] \right\}^{\odot[T-i]} \\ & - [1-\tau(1;\theta)^2]^{-1}\gamma_t, \end{aligned} \tag{3.52}$$

and using similar notation as for (3.40), $a^{\odot b}$ here refers to the elementwise power of the vector a to the scalar b. In both (3.51) and (3.52), the $\tilde{y}_{\text{ext},t}$s are obtained with the corresponding γ_ts via (3.50), and the terms involving (m,n) are again set to zero for those $(m,n) \notin W$.

Conceptually, it is easiest to think of $\nabla(\gamma_{U:T})$ as some list object with s components, each a vector of length d describing the gradient at each spatial cell for each of the timestamps $\{U, \ldots, T\}$.

Candidates for each transition in the MALA are generated in much the same way as in the purely spatial case. They are now treated as 'blocks' of s consecutive spatial intensities such that the candidate $\gamma'_{U:T}$ can be viewed as a $d \times s$ matrix, with each column corresponding to a single spatial intensity on I_{ext}. Assuming $\gamma^{(k)}_{U:T}$ is the current state of the chain, and that the gradient term is an appropriately constructed $d \times s$ matrix, we have

$$\gamma'_{U:T} = \Gamma'_{U:T} \sim N_d(\gamma^{(k)}_{U:T} + 0.5h\nabla(\gamma^{(k)}_{U:T}), h\boldsymbol{I}). \tag{3.53}$$

The move $\gamma^{(k)}_{U:T} \to \gamma'_{U:T}$ is accepted with the now familiar probability

$$p(\gamma'_{U:T}, \gamma^{(k)}_{U:T}) = \frac{f(\gamma'_{U:T}|\boldsymbol{X}_{U:T})\exp\{-||\gamma^{(k)}_{U:T} - \gamma'_{U:T} - 0.5h\nabla(\gamma'_{U:T})||^2/(2h)\}}{f(\gamma^{(k)}_{U:T}|\boldsymbol{X}_{U:T})\exp\{-||\gamma'_{U:T} - \gamma^{(k)}_{U:T} - 0.5h\nabla(\gamma^{(k)}_{U:T})||^2/(2h)\}}. \tag{3.54}$$

An initial state follows the same guidelines mentioned earlier, in the sense that $\gamma^{(0)}_{U:T}$ is given with $\gamma^{(0)}_i \sim N_d(0, \boldsymbol{I})$ for $i \in \{U, \ldots, T\}$. The tuning parameter h in subsequent candidate generation is again chosen to achieve an overall approximate acceptance rate of 0.574.

The inclusion of deterministic components describing 'global' spatial and/or temporal trends, ζ_ψ and η as in Section 3.1.3, alters the target density and gradient as we would expect, with (3.51) becoming

$$\begin{aligned}\log[f(\gamma_{U:T}|\boldsymbol{X}_{U:T})] = \text{constant} + \sum_{t=U}^{T} \sum_{(m,n)\in C_{\text{ext}}} \Big\{ &\{\tilde{y}_{\text{ext},t}(m,n) + \log[\zeta_\psi(m,n)\eta(t)]\}\mathcal{N}_t(m,n) \\ &- \zeta_\psi(m,n)\eta(t)\exp[\tilde{y}_{\text{ext},t}(m,n)]A(m,n)\Big\} \\ &- \sum_{j=U+1}^{T} \left\{\frac{||\gamma_j||^2}{2[1-\tau(1;\theta)^2]}\right\} - 0.5||\gamma_U||^2,\end{aligned} \tag{3.55}$$

and the entries for time t in (3.52) becoming

$$\begin{aligned}\frac{\partial}{\partial\gamma_t}\log f(\gamma_{U:T}|\boldsymbol{X}_{U:T}) = \mathbf{1}[t<T] \sum_{i=t}^{T-1} \Big\{ &Q\left[\tau(1;\theta)\left(\tilde{\mathcal{N}}_{\text{ext},T+t-i}\right.\right. \\ &\left.\left. -\eta(T+t-i)\tilde{\zeta}_{\text{ext},\psi} \circ \exp[\tilde{y}_{\text{ext},T+t-i}] \circ \tilde{A}_{\text{ext}}\right)\right]\Big\}^{\odot[T-i]} \\ &- [1-\tau(1;\theta)^2]^{-1}\gamma_t,\end{aligned} \tag{3.56}$$

where, as in the purely spatial setting, $\tilde{\zeta}_{\text{ext},\psi}$ is the corresponding vector of deterministic spatial values over I_{ext}.

A noteworthy comment, pertaining to both purely spatial and spatiotemporal models, concerns *gradient truncation*. The MALA can, in some cases, spend lengthy periods away from the modes of the target density which can in turn adversely affect the mixing

and convergence of the chain, and therefore our final results (Roberts and Tweedie, 1997; Møller et al., 1998). This can be avoided, without disturbing the ergodic properties of the chain, by imposing a constraint on the magnitude of the cell-wise gradient values. The *truncated* MALA is obtained by replacing $\exp[\tilde{y}_{\text{ext},t}]$ in (3.45), (3.49), (3.52) and (3.56) by $\{\boldsymbol{a} \overset{\circ}{\wedge} \exp[\tilde{y}_{\text{ext},t}]\}$, where $\boldsymbol{a}$ represents the $d \times 1$ vector with all entries equal to the scalar constant a, and $\overset{\circ}{\wedge}$ denotes the component-wise minimum. At the time of writing there appears to be no research concerning 'optimal' calculation of a, though this is not necessarily a major concern in practice. As mentioned in Møller et al. (1998), truncation is generally not needed provided sensible specification of the tuning parameter h; typically achieved if we search for the optimal acceptance rate of 0.574.

3.4 Real-world examples

We now take a break from the theory and demonstrate epidemiological modelling with the LGCP using two real-world data sets; one purely spatial, the other spatiotemporal. The examples are here mainly for illustrative purposes, though we do provide conjecture on possible conclusions. Relative to the existing literature there are novel aspects to both analyses in terms of the data used and methods employed.

3.4.1 Humberside leukaemia and lymphoma

In the first example, we aim to estimate the intensity of the spatial dispersion of childhood leukaemia and lymphoma in the North Humberside region of the United Kingdom, based on 62 observations collected between 1974 and 1986; data first presented and analysed by Cuzick and Edwards (1990). The reader might recall encountering this data set in our discussion of the difficulties associated with leave-one-out least-squares cross-validation for selection of jointly optimal bandwidths pertaining to kernel-smoothed density-ratio estimates in Section 2.3.1.

Figure 3.9 displays the study region W and data $\boldsymbol{X}$, normalised to the unit square. The relatively small number of observations, coupled with vast, empty areas of the study region and several clearly isolated points, are what have contributed to the problems associated with some forms of bandwidth selection in kernel intensity estimation. For this reason, we will use kernel methods in tandem with the LGCP to help produce a sensible estimate of the spatially varying intensity function. To the best of this author's knowledge, this analysis constitutes the first instance of modelling the Humberside data set using the LGCP.

The intensity $\lambda \leftarrow \Lambda$ will be modelled as the product of a deterministic 'global' component and a stochastic 'local' component as per (3.15). Our goal is to therefore obtain a sensible prediction, given the observed data, of $\Lambda(\boldsymbol{x}) = \zeta_\lambda(\boldsymbol{x})\exp[Y(\boldsymbol{x})]$; $\boldsymbol{x} \in W$, where ζ_λ is the global trend describing the underlying spatial heterogeneity, and $Y \approx \mathcal{Y}$ (as per Section 3.3.1) is a stationary Gaussian field on W with mean μ_λ, variance σ^2_λ, and controlled by the stationary, isotropic correlation function $r(\cdot;\phi_\lambda)$. As per the the identifiability constraints discussed in Section 3.1.3, ζ_λ must integrate to $\mathcal{N}(\boldsymbol{X})$ over W, and $\mu_\lambda = \mathbb{E}[\exp\{Y(\boldsymbol{x})\}] = -\sigma^2_\lambda/2$. The estimate will be denoted as $\hat{\Lambda}(\boldsymbol{x}) = \hat{\zeta}_\lambda(\boldsymbol{x})\exp[\hat{Y}(\boldsymbol{x})]$ with the Gaussian field $\hat{Y}$ controlled by

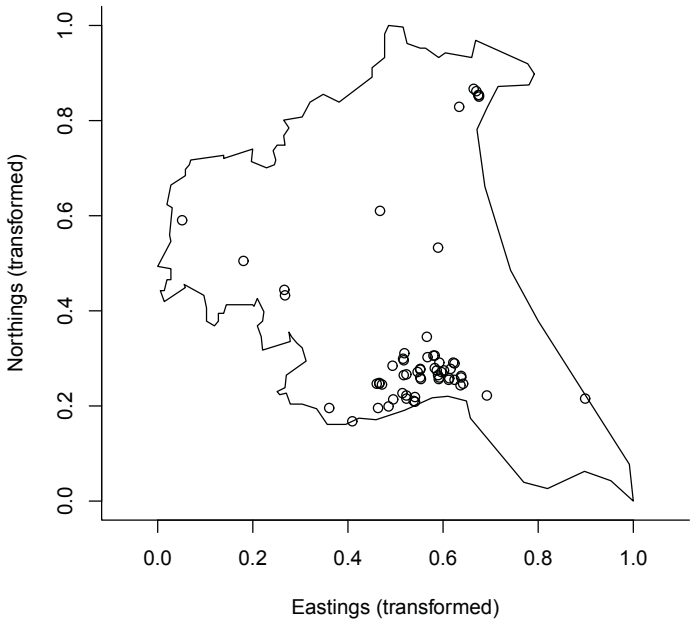


Figure 3.9: 62 recorded spatial locations of childhood leukaemia and lymphoma in North Humberside, 1974–1986.

the corresponding minimum contrast parameter estimates $\hat{\mu}_\lambda$, $\hat{\sigma}^2_\lambda$, and $\hat{\phi}_\lambda$.

Though some may view it as unusual to model cancer occurrences using dependence between individual observations, this component can be interpreted as serving as a surrogate for unobserved covariates which impact the final form of the intensity (an argument used in a similar multiplicative decomposition of deterministic and stochastic components in Møller and Díaz-Avalos, 2010). Also, there exists evidence to suggest some forms of leukaemia can be virally contracted (see for example the report by Leonard, 1998), providing an additional rationale for stochastic modelling of the clusters.

The first step is to specify the form of the deterministic component. As mentioned above, this spatial trend will be specified nonparametrically using a fixed-bandwidth kernel intensity estimate, where we elect the fixed version over the adaptive for simplicity. More importantly, however, this choice is made as too little is known about the relationship between the global and pilot bandwidths in the adaptive setting and the subsequent impact on minimum contrast parameter estimation for the LGCP. The smoothing bandwidth $h_{\mathcal{F}}$ is chosen subjectively to be 0.15, a large bandwidth given the scaling of the data. This oversmoothing is employed owing to the simple fact that ζ_λ is intended to capture only the 'global' heterogeneity, leaving small-scale effects to the partnering LGCP. In fact, through the novel empirical experiments conducted in Section 3.5, we demonstrate that oversmoothing the multiplicative deterministic spatial trend in this way is likely to produce minimum contrast estimates of the correlation parameters controlling the Gaussian field closer to the true values (assuming the specified model) than smoothing parameters based on standard

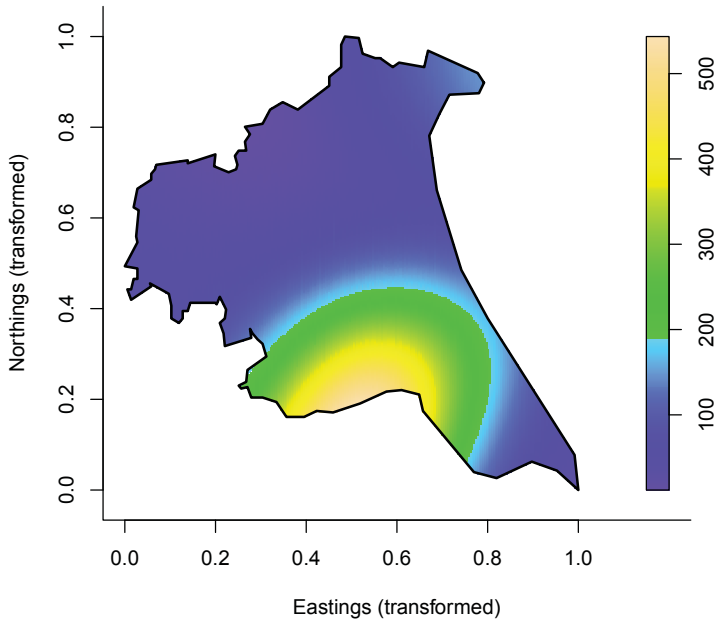


Figure 3.10: Fixed-bandwidth kernel intensity estimate of the global spatial variation of the Humberside data.

bandwidth selectors (e.g. LSCV). Figure 3.10 shows the estimated global trend, $\hat{\zeta}_\lambda$, for the Humberside data set.

The next order of business is to decide upon a functional form of r. This could, hypothetically, be chosen according to some *a priori* knowledge concerning the type of disease we are considering, as well as the kind of overall structure our model will take. For this small data set, we opt for the Stable correlation function as described in Section 3.2.1; $r(u;\phi_\lambda) = \exp[-\sqrt{u/\phi_\lambda}]$. This indicates the parametric PCF will be expressed by $g(u) = \exp\{\sigma^2_\lambda \exp[-\sqrt{u/\phi_\lambda}]\}$, and the theoretical K may be found by the relationship (1.4).

Having elected a correlation function, the next step is to estimate the unknown process parameters μ_λ, σ^2_λ and ϕ_λ. The spatial dependence, described by the latter two terms, will be estimated first via the minimum contrast approach specified by (3.26). Recall in the relevant discussions in Section 3.2.3 the various decisions made in calibrating $\mathcal{M}_J$ and the supporting arguments. For $J = K$, we choose $v[\cdot] = \sqrt[4]{\cdot}$, and when $J = g$, $v[\cdot] = \log[\cdot]$ and we set $h = 0.15/\sqrt{\mathcal{N}(\boldsymbol{X})/|W|}$. In both cases, $\omega(\cdot) = 1$, $s_0 = 0$ and $s_{\max} = 0.05$ (owing to the fact that the correlation parameters only describe a *residual* dependence structure, we shorten the interval of interest considerably).

Figure 3.11 shows the result of estimating σ^2_λ and ϕ_λ using minimum contrast. Both versions of the procedure show reasonable agreement of the parametric and nonparametric functions. Use of the K-function provides $\hat{\sigma}^2_\lambda = 5.514$ and $\hat{\phi}_\lambda = 0.0086$; use of g yields $\hat{\sigma}^2_\lambda = 5.826$ and $\hat{\phi}_\lambda = 0.0069$. It seems that although the K-function version has indicated a slightly larger range of dependence through ϕ_λ, the PCF estimates have leant towards

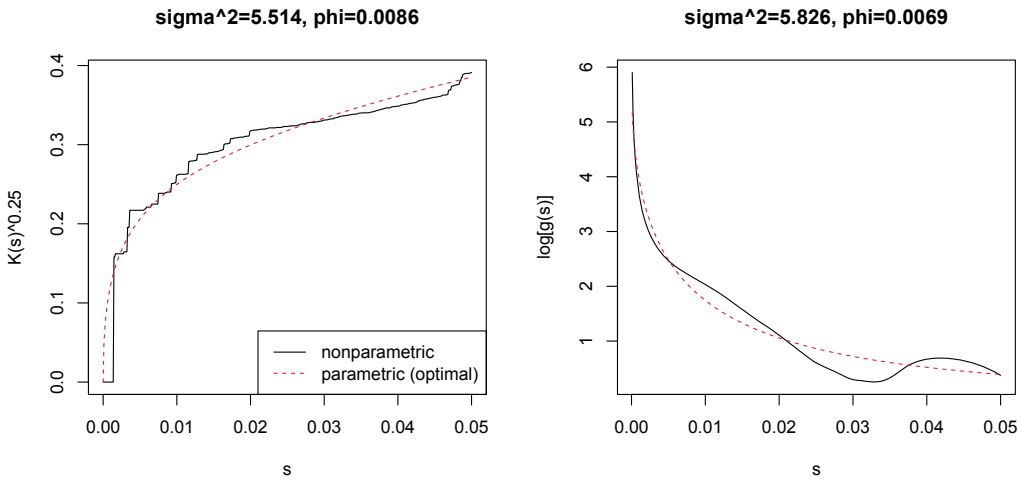


Figure 3.11: Minimum contrast estimation of the spatial correlation parameters for the Humberside data based on the K-function (left) and PCF (right).

slightly higher concentrations of intensity at cluster locations though σ^2_λ. It is unclear which of these combinations is 'better' for the Humberside data in terms of proximity to the true values under our current modelling assumptions. In light of the satisfactory nonparametric versus parametric trend matches for both versions we will therefore work with their means, setting $\hat{\sigma}^2_\lambda = 5.670$ and $\hat{\phi}_\lambda = 0.0078$ (though it should be noted that this cannot be considered a universally appropriate step). Our estimate of μ_λ, the stationary mean of the Gaussian field Y, is therefore given by $\hat{\mu}_\lambda = -\hat{\sigma}^2_\lambda/2 = -5.670/2 = -2.835$.

We now have the ability to simulate as many Gaussian fields, corresponding to the estimated dependency structure and point intensity of the Humberside data, as we like. Of interest will be simulation of the Gaussian random processes *conditional upon the observed data*, which in turn will allow us to actually estimate the 'true' appearance of λ over W.

A MALA run as discussed in Section 3.3.2 is thus required. The progression follows (3.42); the structure of our Humberside LGCP dictates use of (3.48) and (3.49) as the target density and gradient functions. To take full advantage of the computational concessions via the FFT, the cell-dimension of the rectangular lattice $(M-1) \times (N-1)$ over W is $2^6 \times 2^6$ (64×64); the overlaying extended lattice is given cell-dimension $2^7 \times 2^7$ (128×128). The MALA chain is instructed to complete 100000 iterations, with a discarded burn-in of 10000 to allow convergence to properly begin, and retaining every 90^{th} result thereafter to curb the adverse effects of autocorrelation arising in consecutively realised fields. This provides us with a thinned chain of length 1000 with which to estimate the intensity (for the reader less familiar with Bayesian methods, note that burn-in and thinning is considered standard practice in MCMC execution for the above reasons). After some preliminary runs, the MALA tuning parameter h was set to 0.058.

Figure 3.12 displays some trace plots and diagnostics associated with our run of the MALA; a satisfactory acceptance rate of around 60% was reported. Horizontal lines indicate the posterior mean of each chain. The trace plots show convergence appears to have been

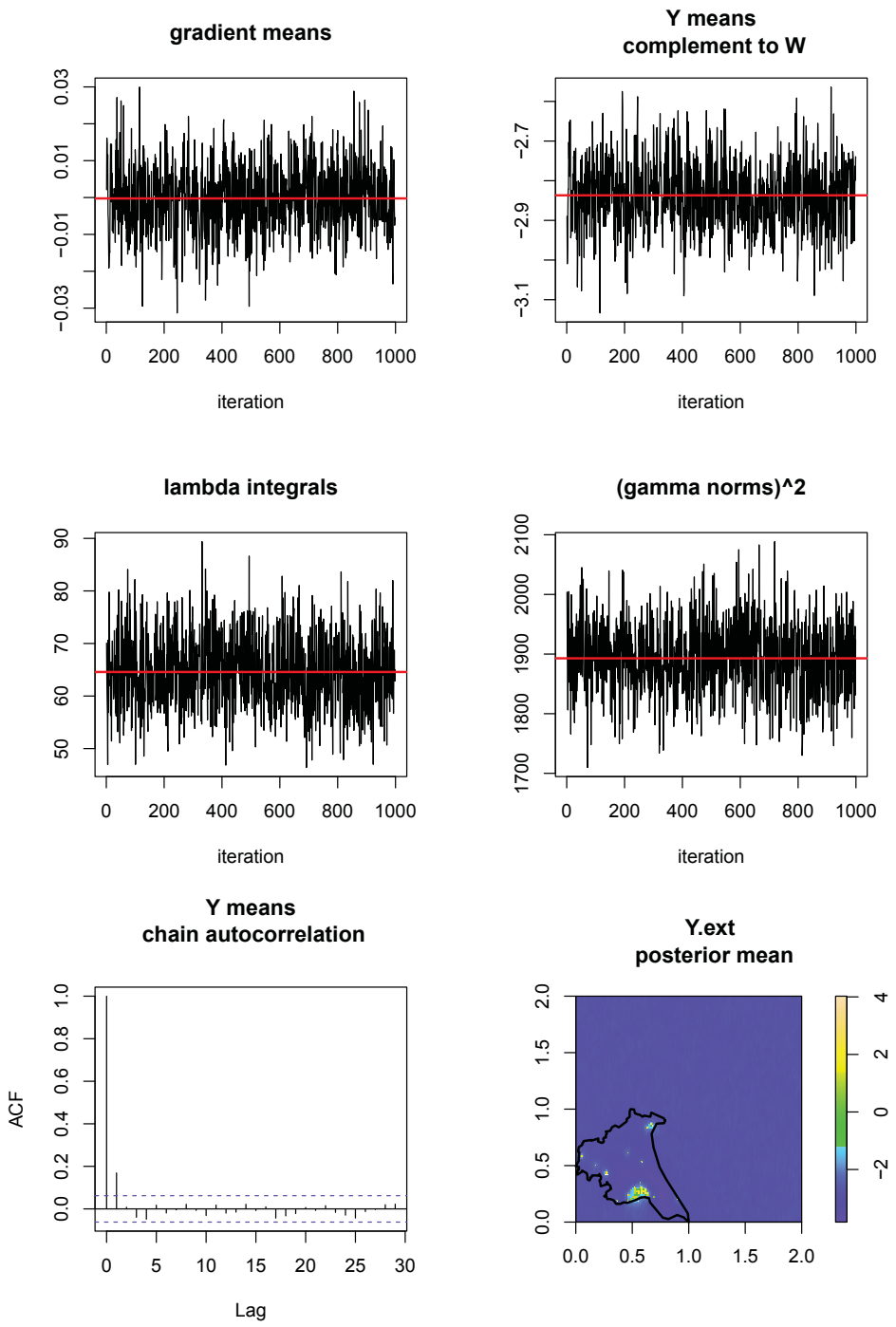


Figure 3.12: Trace plots and diagnostics for the Humberside MALA run.

successfully achieved with the chains mixing nicely. Top left we observe 'gradient means'; the mean value of the gradient function on I_{ext} associated with the candidate generation for each retained field. Appropriately centered at (approximately) zero, this is the expected appearance of the mean gradient value when convergence has been achieved. Top right displays the mean cell values for each retained field for all locations $\boldsymbol{x} \notin W$, which should, and does, center on roughly $\hat{\mu}_\lambda = -2.835$ (note that the field outside W and away from the boundary corresponds to unconditional simulation associated with the supplied parameter estimates; cf. Møller et al., 1998). Perhaps the most important trace plot is given center left, and illustrates the integral of each retained, realised intensity function $\hat{\lambda}(\boldsymbol{x}) \leftarrow \hat{\Lambda}(\boldsymbol{x}) = \hat{\zeta}_\lambda(\boldsymbol{x})\exp[\hat{Y}(\boldsymbol{x})]$; $\boldsymbol{x} \in W$. Centered appropriately close to $\mathcal{N}(\boldsymbol{X}) = 62$, this indicates that the conditionally realised intensities are of the correct scaling, and the algorithm is generating and accepting appropriate candidates (posterior mean is in fact 64.56). Referring back to Section 3.3.2 for notation, center right provides the squared vector norms associated with the generated γs prior to backtransformation to y, roughly centered on the number of cells falling inside W (1915). Recall the generated γ values are in fact $N_d(0, \boldsymbol{I})$ variables; the collective sum of the squared components will therefore be Chi-squared with expectation equal to the number of interior cells. Bottom left we observe that, by computing an autocorrelation plot of the means of each retained $\tilde{y}_{\text{ext}}$, the level of thinning we have employed has all but eliminated any concerns we may have had arising from consecutively realised fields being too heavily correlated. Finally, we observe the posterior mean Gaussian field on I_{ext} i.e. on the torus, bottom right. Modes in the field inside W correspond to where data have been observed. The aforementioned unconditional simulation occurring outside W is clearly visible in comparison to the activity inside W.

Figure 3.13 displays our final estimate of the intensity function, $\bar{\hat{\lambda}}$, over W, given by the posterior mean of the conditionally realised and retained exponentiated Gaussian fields from the run of the MALA. The superimposition upon this plot of the original observations indicates a plausible prediction.

3.4.2 AEGISS gastrointestinal infections

Our second example concerns the spatiotemporal modelling of the AEGISS (Ascertainment and Enhancement of Gastrointestinal Infection Surveillance and Statistics) data set. We perform a similar analysis to that reported in Diggle, Rowlingson and Su (2005). Our work here differs in that we inspect an updated (i.e. larger) version of the data set and alter some approaches to estimation of the included deterministic components, naturally affecting computation of the correlation parameters.

The analysis revolves around sensible estimation of the spatiotemporal intensity, specified by the model $\psi(\boldsymbol{x}, t) \leftarrow \Psi(\boldsymbol{x}, t) = \zeta_\psi(\boldsymbol{x})\eta(t)\exp[Z(\boldsymbol{x}, t)]$ as per the definitions in Section 3.1.3; $Z \approx \mathcal{Z}$. The goals in this example extend a little further than the Humberside analysis, however, in that we are not just interested in obtaining an estimate of the intensity. The specification of our model, as per (3.18), will as mentioned include the deterministic components ζ_ψ (global spatial trend) and η (global temporal trend). With the spatiotemporal LGCP $\exp[Z(\boldsymbol{x}, t)]$ thus representing residual departures from the underlying heterogeneity,

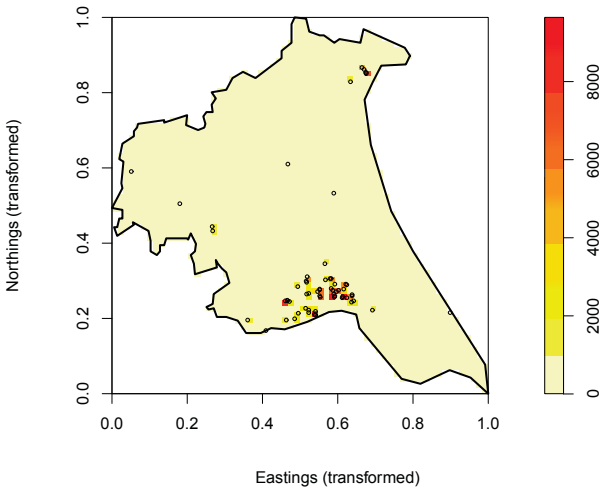


Figure 3.13: Posterior mean of the intensity surface $\bar{\hat{\lambda}}(\boldsymbol{x})$ from the Humberside MALA run with original observations.

the additional objective is to empirically compute *exceedance probabilities*, which are able to describe, based on the estimated extremity of said residual departures, any 'anomalous' behaviour in the space-time pattern of gastroenteritis. Given the nature of this application is *surveillance* of the disease, it is clearly of interest to monitor the disease occurrences and locate perceived 'hotspots' of cases. In other words, we wish to somehow 'flag' clusters of observations in a local spatiotemporal window which are sufficiently large enough to *exceed* what we would deem 'normal' background variation. Again, $\hat{\psi}(\boldsymbol{x}, t)$ will represent a realisation of the model $\hat{\Psi}(\boldsymbol{x}, t) = \hat{\zeta}_{\psi}(\boldsymbol{x})\hat{\eta}(t)\exp[\hat{Z}(\boldsymbol{x}, t)]$ implemented with the relevant deterministic and correlation parameter estimates.

Let us first inspect the data. We have in total 10554 spatial observations in Southampton, U.K., recorded daily over the three-year period 1 January 2001 to 31 December 2003. Again normalising the spatial observation window to the unit square, Figure 3.14 displays the margins of spatial disease dispersion and temporal variation in the number of cases (for visual simplicity the latter plot contains counts averaged over successive weekly periods). We note clear spatial heterogeneity; presumably a direct reflection of the natural population dispersion over W. An underlying trend in the counts of cases over time is less obvious, though some structure does appear present.

To account for these global trends, thereby allowing the stochastic component to describe the aforementioned 'local', residual correlations, we fit deterministic models to represent ζ_{ψ} and η. The spatial trend is estimated following Diggle, Rowlingson and Su (2005) using an adaptive (via the method of Abramson, 1982) kernel density estimate. The bandwidths h_0 and $\tilde{h}$ (as per notation in Chapter 2), are computed using the oversmoothing method discussed in Section 2.3.1, with $\tilde{h} = 0.5h_0$. For the temporal component we again choose a nonparameteric approach, fitting a LOWESS smoother (locally weighted polynomial regres-

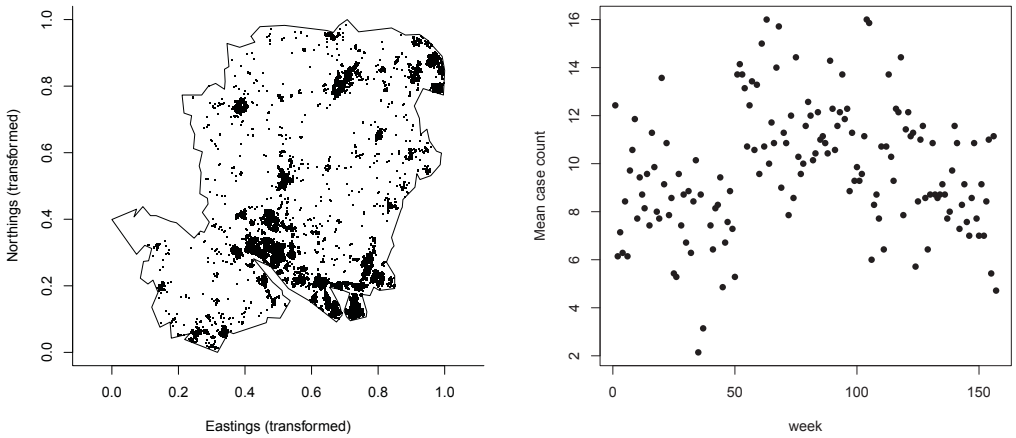


Figure 3.14: Spatial (left) and temporal (right) margins of the AEGISS data set.

sion – see Cleveland, 1979, 1981) to the daily count data; the proportion of points affecting the curve is set to 0.2. Diggle, Rowlingson and Su (2005) originally used harmonic regression in the form of a generalised linear model for their fit of η, but we found this approach to perform poorly on the updated data set; hence the choice of LOWESS. Figure 3.15 displays our estimates $\hat{\zeta}_\psi$ and $\hat{\eta}$ (again averaged over each week to correspond to the earlier plot) for our model.

Having specified the global trends, we must now estimate the parameters controlling the residual spatiotemporal correlation. Recall the assumption of separability from Section 3.1.2, which states that for our Gaussian process $\mathcal{Z}$, hence its finely discretised approximation Z, $\text{Cov}[Z(\boldsymbol{x}_1, t_1), Z(\boldsymbol{x}_2, t_2)] = \sigma^2 r(u; \phi_\psi)\tau(\nu; \theta)$; $u = ||\boldsymbol{x}_1 - \boldsymbol{x}_2||$, $0 \leq \nu \leq (T-1)$. That is, for two distinct points in the space-time continuum, the covariance of Z can be expressed in separate 'spatial only' and 'temporal only' correlation functions r and τ respectively. This allows straightforward access to minimum contrast estimation procedures for σ^2_ψ, ϕ_ψ and θ.

For our AEGISS model we select the Exponential correlation function for both margins such that $r(u; \phi_\psi) = \exp[-u/\phi_\psi]$ and $\tau(\nu; \theta) = \exp[-\nu/\theta]$. For estimation of σ^2_ψ and ϕ_ψ, we make use of the inhomogeneous time-averaged estimators of the PCF and K-function given by (3.20) and (3.21) respectively, and compare them to their theoretical counterparts which in this case are $g(u) = \exp\{\sigma^2_\psi \exp[-|u|/\phi_\psi]\}$ and $K(u) = 2\pi \int_0^u tg(t)\,\mathrm{d}t$.

Figure 3.16 displays the results of the minimum contrast procedure for estimation of σ^2_ψ and ϕ_ψ, again employing the relevant transforms mentioned in Section 3.2.3. Like the Humberside analysis, the residual nature of the dependence structure means we use a small upper contrast limit; $s_{\max} = 0.025$ spatial units. Note that use of an unreasonably large value for $s_{\max}$ when the LGCP merely describes local features has the potential to produce misleading parameter estimates.

In this case, both versions of the minimum contrast method for estimating σ^2_ψ and ϕ_ψ differ somewhat in terms of their respective ideas of the two parameters. Indeed there is a mildly disappointing match, in both trend and discrepancy, between the optimal parametric

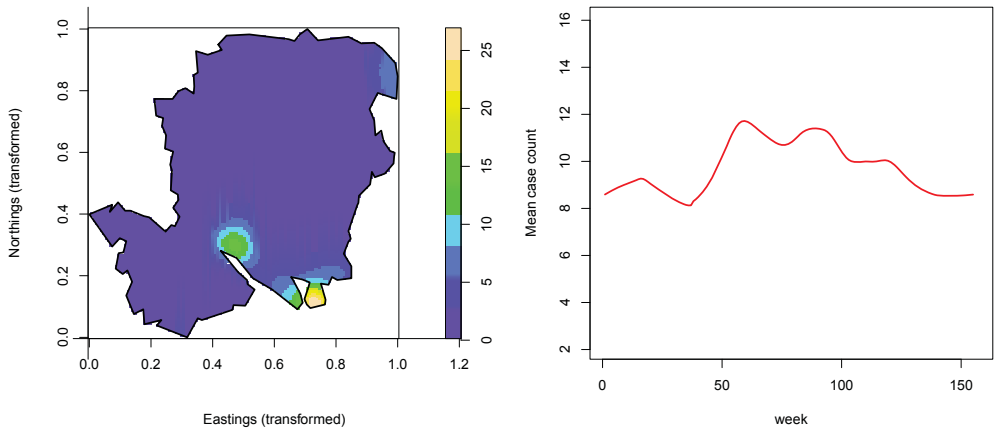


Figure 3.15: Implemented deterministic global spatial trend $\hat{\zeta}_\psi$ (left) and global temporal trend $\hat{\eta}$ (right) for our AEGISS model.

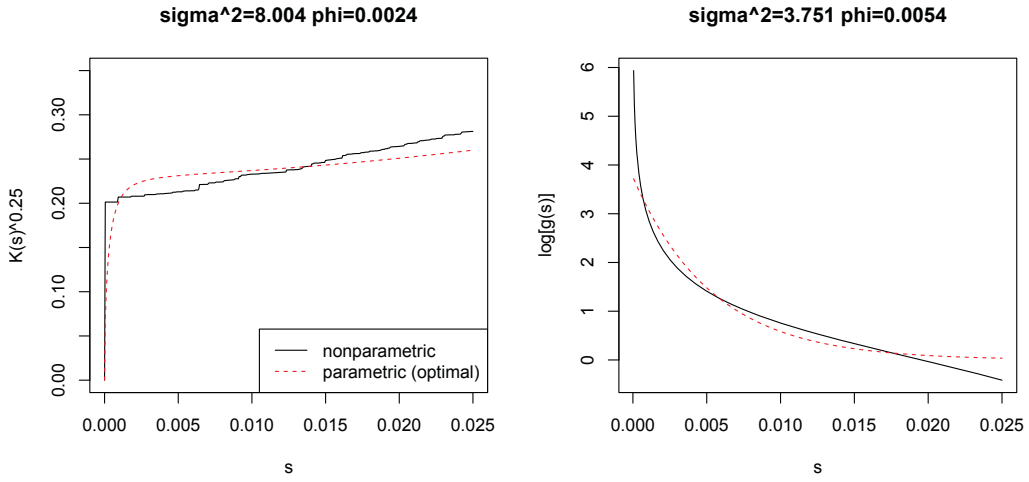


Figure 3.16: Minimum contrast parameter estimation for the spatial dependence structure of the AEGISS data, via the time-averaged K-function (left) and PCF (right) for the selected Exponential correlation function.

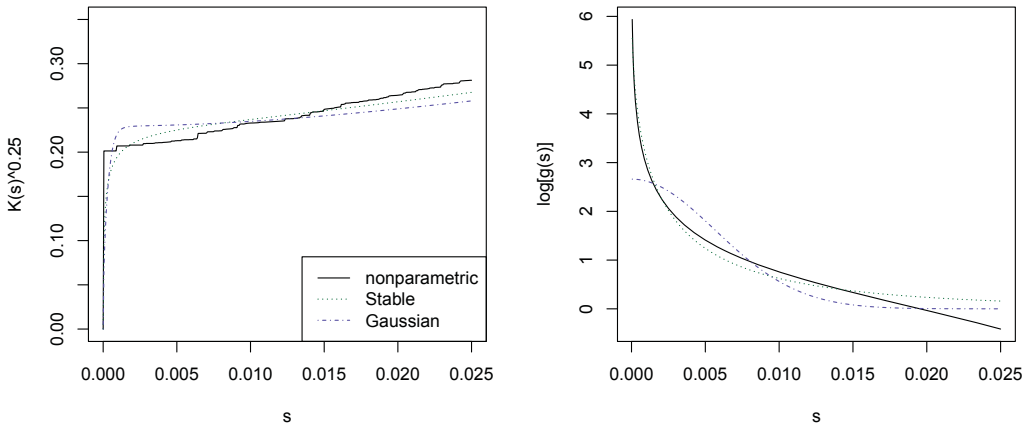


Figure 3.17: Minimum contrast procedures for the spatial dependence structure of the AEGISS data, via the time-averaged K-function (left) and PCF (right), and 'optimal' curves based on the Gaussian and Stable correlation functions.

and nonparametric versions of $K(s)^{0.25}$. The plots of $\log[g(s)]$, however, show a more suitable parametric versus nonparametric agreement.

One possible option to increase our confidence in the selected parameters (especially with respect to K) would be to visually explore the fits arising from the use of different correlation functions for r, but there is no guarantee any 'better' matches will be found; indeed the very definition of a 'better match' in this general situation is also extremely difficult. This is highlighted with additional experimentation using the Gaussian and Stable correlation functions (cf. Section 3.2.1). Figure 3.17 shows the minimum-contrast-optimal parametric curves arising from this modification to the spatial dependence structure.

The trend mismatch in the K-function plot remains, suggesting no real practical improvement in model specification should we select the Gaussian or Stable correlation functions over the Exponential. It does appear in both cases that the Stable curves are more visually appealing than the Gaussian. It is conceivable that there are additional factors out of our control which are contributing to the almost linear appearance of the nonparametric K-function plots for $s > 0$, in turn resulting in the slightly disappointing parametric matches for any of the candidate correlation functions. One reasonable conjecture concerns the resolution of the recorded data; the number of duplicated coordinates which arise when the temporal marking of the events is ignored may be adversely affecting the validity of the estimators of the second-order spatial characteristics at certain lags.

At any rate, it is clear that altering the specification of the correlation function for the AEGISS data has had little if any success in producing nonparametric versus parametric trend matches that are superior to those originally assumed. For this reason, the findings discussed above, the illustrative nature of this example, and the fact that we aim to follow the same modelling strategies used by Diggle, Rowlingson and Su (2005), we will remain with the Exponential form of the correlation function in this analysis, and incorporate only

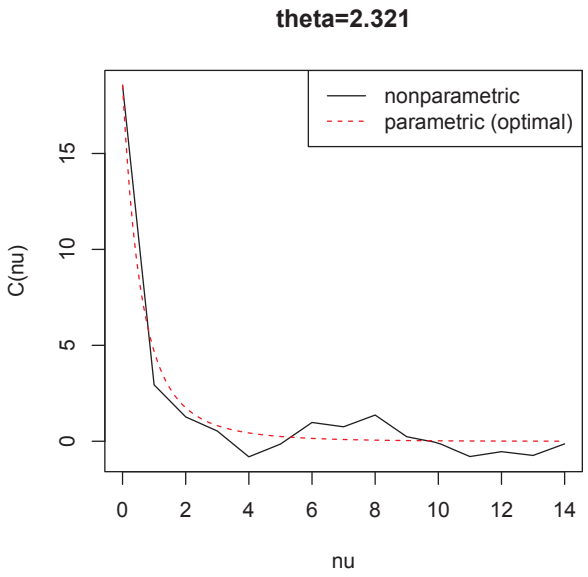


Figure 3.18: Minimum contrast parameter estimation (Exponential correlation function) for the temporal dependence structure of the AEGISS data, based on comparing the parametric and nonparametric functions for the temporal covariance in the total number of cases per time point.

the results from use of g. That is, we set $\hat{\sigma}^2_\psi = 3.751$ and $\hat{\phi}_\psi = 0.0054$.

Estimation of θ is performed using (3.27), and our choices of $\hat{\sigma}^2_\psi$ and $\hat{\phi}_\psi$. Figure 3.18 provides a visual of the $\mathcal{C}$ (3.22) using the optimal estimate of θ versus $\hat{\mathcal{C}}$ (3.23). Time is measured in days, with each time interval equal to a single day. Minimum contrast is performed over lags up to $\nu_{\max} = 14$.

Inspection of Figure 3.18 indicates a satisfactory match between the nonparametric $\hat{\mathcal{C}}$ and optimally selected theoretical $\mathcal{C}$. Our estimate of the temporal correlation parameter is $\hat{\theta} = 2.321$. Experimentation with varying $\nu_{\max}$ (not shown) had minimal impact on the resulting parameter estimate.

Finding a method of computing the aforementioned exceedance probabilities requires us once again to consider the interpretation of the various components of our model. Note that the 'natural' or 'expected' background variation is captured in this case for the most part by the deterministic trends ζ_ψ and η. We would expect this deterministic spatiotemporal behaviour of disease occurrences to be affected by small to moderate levels of residual spatiotemporal dependence; accounted for by the stochastic component for which we have just obtained the requisite parameter estimates. Following Diggle, Rowlingson and Su (2005), we define an *exceedance* of the expected levels of variation and correlation in the observed spatiotemporal intensity of cases to occur when the residual process at location $\boldsymbol{x}$ and time t positively breaches some pre-defined threshold ϵ, based on the observed data. This is empirically estimated using the conditional field realisations from the MALA, such that the *probability of exceedance* $P_\epsilon(\boldsymbol{x}, t) = P\{\exp[Z(\boldsymbol{x}, t)] > \epsilon | \text{data up to } t\}$ is calculated via the proportion of (retained) fields that have recorded $\exp[Z(\boldsymbol{x}, t)] > \epsilon$. Thus, the higher $P_\epsilon(\boldsymbol{x}, t)$,

the greater the chance that we have in fact observed 'abnormally large' intensity levels (given historical data), and we may, for example, visually 'flag' these localisations on subsequent plots of either the posterior mean spatial intensity functions over W at each time point as an animation, or on similar maps of the exceedance probabilities themselves.

Definitions of appropriate thresholds can be subjectively chosen based on past experience, or, perhaps more sensibly, by targeting sufficiently large exceedance probabilities as the P_ϵ^{th}-quantiles. Using the properties of the log-Gaussian distribution these are easily derived. For a specific probability p, $\epsilon = \exp[\mu_\psi + \sigma_\psi \Phi^{-1}(p)]$ where $\Phi^{-1}(\cdot)$ denotes the quantile function of the standard normal distribution. Under our assumed model, $\mu_\psi = -0.5\sigma_\psi^2$, and we replace σ_ψ^2 by its minimum contrast estimate, giving $\hat{\epsilon} = \exp[1.94\Phi^{-1}(p) - 1.88]$. The interest is typically in high exceedance probabilities; $P_{\hat{\epsilon}} \in \{0.99, 0.95, 0.9\}$. Via the above formula, the corresponding thresholds for the AEGISS analysis are $\hat{\epsilon} \in \{13.88, 3.71, 1.83\}$.

We are now ready to execute the spatiotemporal LGCP MALA using (3.55) and (3.56). Predictive distributions are generated for the last 5 days of data; based on $\hat{\theta}$ and Figure 3.18 we assume in addition that only the previous 5 days impact the intensity at any given time (as for larger lags it seems we have temporal independence). Thus, we run 5 individual MALAs, with the target time of interest being $T - 4$ up to $T = 1095$; for predictions at $T - 4$, for example, we take into account the data from times $T - 9$ to $T - 5$ (as well as the $T - 4$ data themselves). Note that were we to specify a large enough 'historical' window only one MALA would need to be run; the choice of separate MALAs was made here for computational reasons.

The spatial discretisation is set with $(M - 1) \times (N - 1) = 64 \times 64$ and $(M_{\text{ext}} - 1) \times (N_{\text{ext}} - 1) = 128 \times 128$ (finer resolutions are of course possible with an obvious impact on computational expense). Each MALA is run for 120000 iterations with a burn-in length of 20000, and thinned by retaining every 100^{th} spatiotemporal field, yielding a set of 1000 realisations from which to calculate posterior means of the intensities and exceedance probabilities on each of the 5 days of interest. Preliminary runs lead us to select $h = 0.004$ as an appropriate tuning parameter for the algorithms.

Inspection of diagnostics in the same way as the Humberside example indicates convergence was successfully achieved. Figure 3.19 shows one such example. These trace plots from the MALA of the value of the Gaussian field Y at each retained iteration for three randomly selected cells corresponding to the final day of data T also display good mixing of the chain.

The product of $\hat{\zeta}_\psi$, $\hat{\eta}$ and the posterior means of the exponentiated spatiotemporal fields obtained from the MALAs for the 'final' day of each of the days of interest constitute our prediction of the intensity $\bar{\hat{\psi}}$. Figure 3.20 shows the predicted intensities for the days $T-4$ to $T - 1$, upon which we have superimposed the observations occurring on that day, as well as the previous days' observations (restricted to just the single previous day for visual clarity). The most recent day of observations, clearly of the most interest from a disease surveillance perspective, is displayed separately in Figure 3.21.

Immediately noticeable from the posterior intensity plots, like the Humberside example, is the fact that the highest peaks in the surfaces naturally correspond to the cells which contain the most recent observations. By including the previous days' observations on each

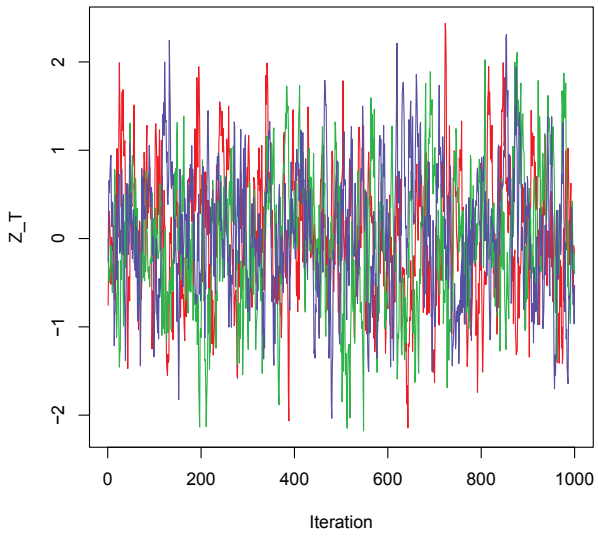


Figure 3.19: Trace plots of the realised, retained values of $\hat{Z}$ for three distinct, randomly selected cells (identified simply as 'blue', 'red', and 'green') in the defined AEGISS space-time window for the MALA at $T = 1095$.

image, we can see the effect on the 'current' intensity – there remain higher intensity levels where points may not be currently observed, but where they have very recently occurred. Sensibly, the intensity at $T - 2$ is the 'flattest': there happened to be no observations on that day.

Recall that the exceedance probabilities, the means by which we will attempt to detect anomalous behaviour in the spatiotemporal occurrence of disease cases, are defined according to the positive extremity of the predicted residual process conditional upon the observed data. If, at a given space-time location, the residual process breaches a predefined threshold, this is classified as an exceedance; the collective proportions of exceedances from the retained exponentiated fields after a MALA run (or runs) then constitute the spatiotemporal maps of empirical exceedance probabilities. Another interpretation of this definition is recognition of the fact that when a 'sufficiently large' number of cases are observed in a 'sufficiently contained' spatiotemporal localisation with respect to the assumed global trends, the residual process attempts to capture this with exceedingly high values.

Figure 3.22 displays the 95^{th} percentile spatial exceedance probabilities (corresponding to a threshold of $\hat{\epsilon} = 3.73$) for the AEGISS data for days $T - 4$ to $T - 1$; the most recent day T is given in Figure 3.23.

As we would expect, the magnitude of the probabilities reflects the observed data. The final time point of interest, T, provides little reason for concern. The highest calculated exceedance probability is less than around 0.7; in practice we would most likely be interested in higher probabilities (over, say, 0.9), since this would bolster our confidence in interpreting such an occurrence as a true 'anomaly'. However, the choice of an appropriate maximum limit for the interpretation of larger exceedance probabilities as 'anomalies' does remain at

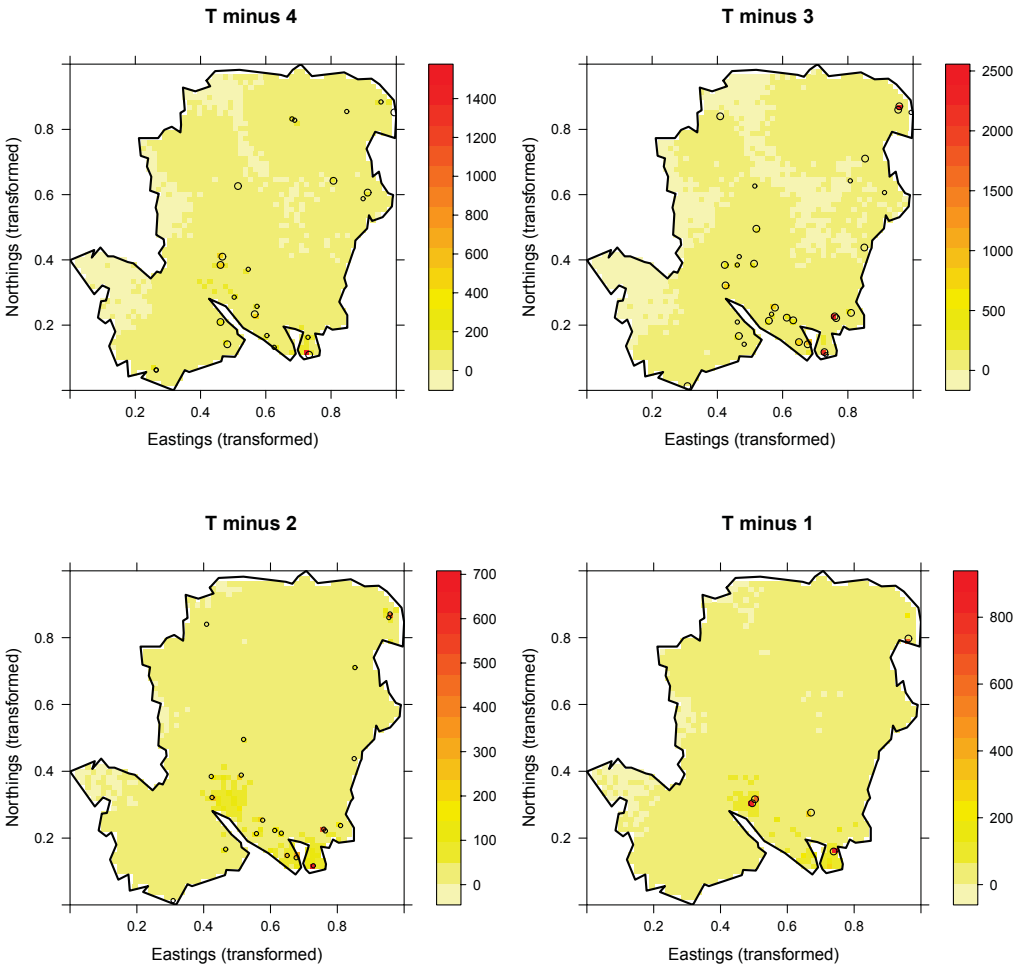


Figure 3.20: Posterior mean intensity surfaces for each of $T-4$ to $T-1$ for the AEGISS analysis. Large circles are the observations on that day, small represent the locations of those from the previous day.

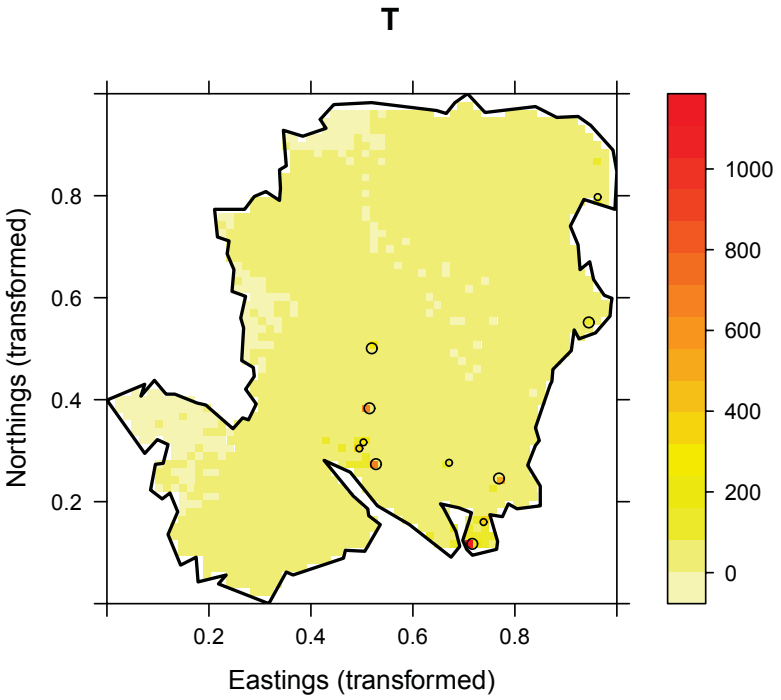


Figure 3.21: Posterior mean intensity surface for the most recent time T. Large circles are the observations on that day, small represent the locations of those from the previous day $T-1$.

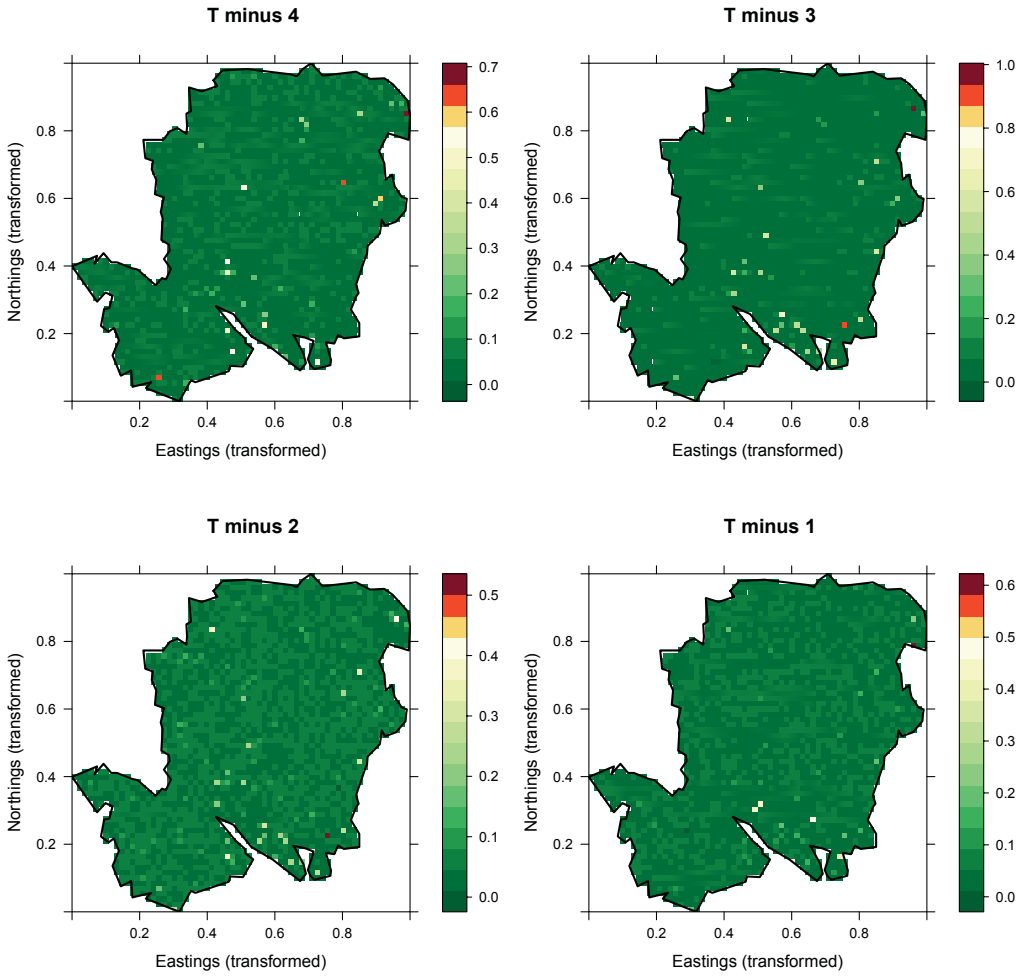


Figure 3.22: 95$^{\text{th}}$ percentile empirical exceedance probabilities based on the realised exponentiated Gaussian fields for each of $T-4$ to $T-1$ for the AEGISS analysis.

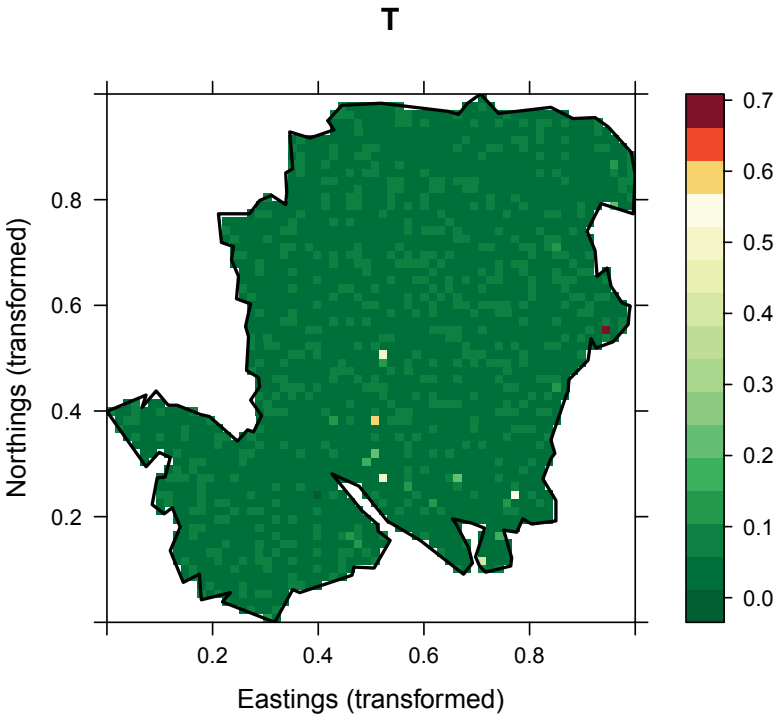


Figure 3.23: 95^{th} percentile empirical exceedance probabilities based on the realised exponentiated Gaussian fields for the most recent time point T.

the discretion of the researcher, and can be application-specific. Inspection of the preceding time points in Figure 3.22 suggests little else, with the exception of the plot corresponding to time $T-3$. We note two particular cells which have returned high proportions of exceedances. Interpretation of such a result, however, must be made with care. Diggle, Rowlingson and Su (2005) sensibly mention that identification of genuine outbreaks or anomalies based on such a model must of course take into account the neighbouring cells in both space and time. Only the persistent presence of high exceedance probabilities over multiple time points in the same spatial area should warrant flagging, and possible intervention, by the relevant health officials.

3.5 Simulation studies

Here, we begin the novel numerical explorations into key issues concerning modelling and prediction based on the LGCP. The importance of sensible estimation of the parameters controlling the dependence structure of the latent Gaussian process is crucial, and it is this issue which dominates the following three simulation studies.

These studies aim to contribute to the knowledge concerning the performance of the minimum contrast parameter estimation methods described in Section 3.2, as well as how variation of the correlation parameters affects subsequent analyses. We begin with Study 1 (Section 3.5.1), shedding light on the difference between using the K-function or the PCF when estimating the relevant parameters in spatial analyses. Section 3.5.2 expands on this issue by introducing Study 2, demonstrating how variation of the level of smoothing applied to a kernel estimate of a 'global' spatial trend affects the estimation of the parameters of the residual stochastic process. Finally, we explore the impact of parameter variation on the detection of anomalies via exceedance probabilities (as defined and utilised in the AEGISS example above) in spatiotemporal surveillance in Study 3 (Section 3.5.3). In Studies 1 and 2, we suppress the subscript $_\lambda$ for ease of notation.

3.5.1 Study 1: K or g?

We have available two sensible choices for estimation of the parameters concerning spatial dependence in an assumed Cox process via either the K-function or pair correlation function (g). Some of the simplest and most intuitive forms of the covariance structure are concerned with two parameters: the variance of the Gaussian field σ^2 and the scale of dependence, ϕ. We will restrict attention to this setting.

Though representing much the same information, it is unclear whether one performs better than the other in terms of producing estimates of the required parameters close to the true, unknown values. This is particularly important in spatiotemporal models; estimation of the parameter(s) controlling the temporal correlation depend directly on the estimates arrived at for the spatial component. Illian et al. (2008) note in Section 4.3.2:

> "The question which of the two functions, g or K, yields the better minimum contrast estimates remains unanswered and requires further research..."

The authors go on to say, without further exposition, that their experience tends to favour the PCF. We therefore aim to establish some initial numerical results with respect to this issue.

We define nine stationary, isotropic log-Gaussian Cox processes $\Lambda(\boldsymbol{x}) = \exp[Y(\boldsymbol{x})]$, with $Y \approx \mathcal{Y}$ and $\boldsymbol{x} \in W$. All are controlled by the ubiquitous Exponential correlation structure such that $\text{Cov}[Y(\boldsymbol{x}_1), Y(\boldsymbol{x}_2)] = \sigma^2 \exp[-||\boldsymbol{x}_1 - \boldsymbol{x}_2||/\phi]$. Working on the unit square $W = [0,1] \times [0,1]$, we design the Minimum Contrast Scenario (MCS1 through MCS9) covariance structures by providing three possible values for each of σ^2 and ϕ; small, medium, large in the sense that $\sigma^2 \in \{1, 2.5, 5\}$ and $\phi \in \{0.05, 0.1, 0.25\}$.

Figure 3.24 shows a single example realisation for each MCS, best displayed using a correspondingly generated data set (in the Figure, each point pattern has $N(W) = 1000$). The impact of varying σ^2 and ϕ is clear, with larger variance leading to clusters of points that are more concentrated, and a larger scale leading to patterns that vary in appearance more slowly due to the longer range of dependence. According to the ordering of the definitions in this image, it is therefore easy to see why MCS3 has the smallest, most tightly concentrated clusters (smallest ϕ, largest σ^2) and MCS7 produces the most diffuse patterns (largest ϕ, smallest σ^2).

For each scenario, 100 intensity functions are generated, each of which subsequently generates a point pattern of size $n \in \{300, 900\}$ (although technically these figures should represent the expected value of the respective Poisson distributions to determine the actual sample sizes, we force them to be fixed to remove this source of variation). Minimum contrast estimation is then performed on the realised data set, using either K or g. We build on the previous experiences of other authors (e.g. Stoyan and Stoyan, 1994; Diggle, 2003) and set the upper contrast limit, function transformations and bandwidth for g as outlined in Section 3.2.3 and as used in the examples in Section 3.4. Minimisation was performed using the Nelder-Mead optimisation algorithm (Nelder and Mead, 1965).

Results by way of the mean variance and scale parameters $\bar{\sigma}^2$ and $\bar{\phi}$, along with their empirical standard errors, are reported in Table 3.2. There are several comments that can be made. Almost all estimates appear to *underestimate* the true values to a certain extent. These discrepancies become more pronounced as we increase the range of dependence through ϕ (e.g. MCS8, MCS9), which is understandable – we would expect the quality of the parameter estimation to degrade as the window in which we are working becomes inadequately small in terms of measuring the actual, relatively large, scale. Indeed, this is supported by conclusions drawn in some earlier works e.g. Heinrich (1993) deals with some specialised cases; recall also the results in Guan and Sherman (2007): under reasonable conditions we should expect the minimum contrast procedures to converge to the true values as the area captured by W and the sample size increases. However, even the small-scale, low variability estimates (e.g. MCS1, MCS2) exhibit this negative bias, for both K and g. Though well within what we could expect 'reasonable variation' given the standard errors for the estimates in these scenarios, it is the persistent presence of the underestimation for both σ^2 and ϕ which is concerning.

Whether K or g is better at estimation of the required parameters is difficult to gauge using just Table 3.2. To aid in this, we produce plots of the *absolute relative error* (ARE)

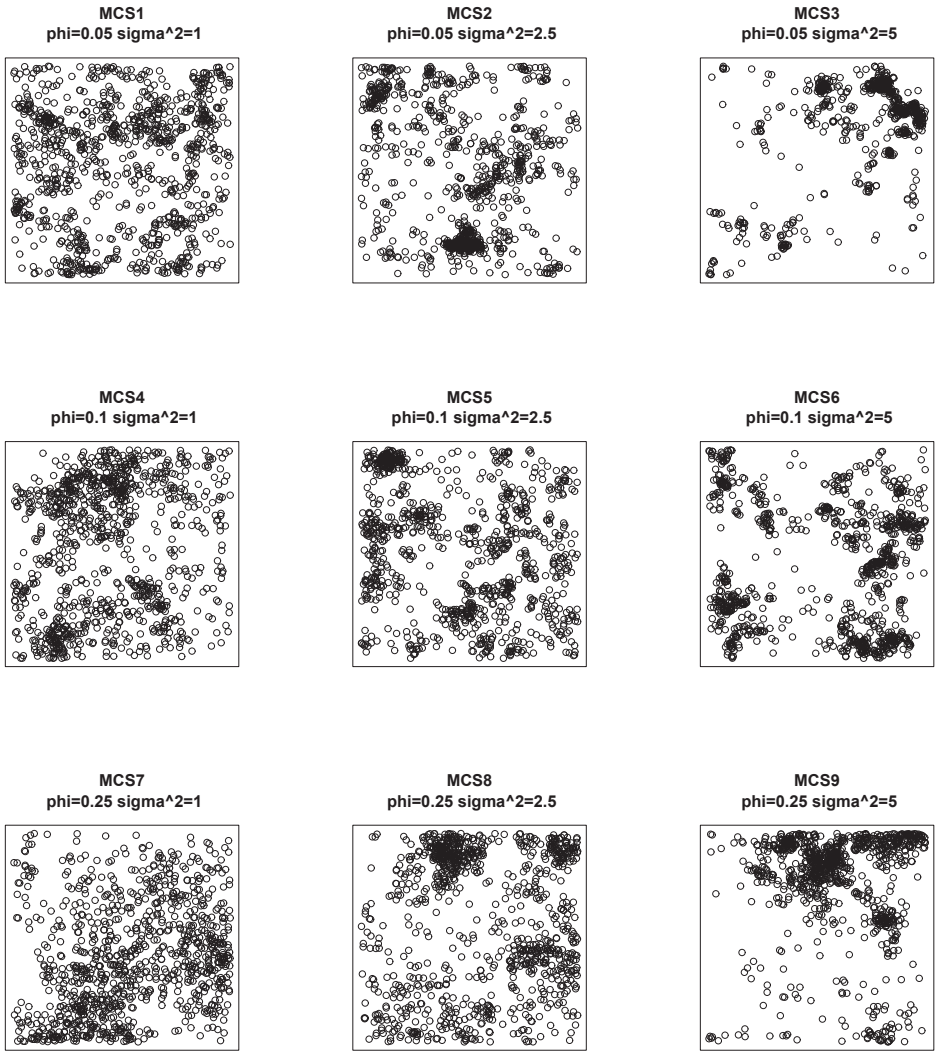


Figure 3.24: Examples of the kinds of point patterns produced under MCS1 to MCS9.

MCS		n	K		g	
			$\bar{\phi}$ (s.e.[$\bar{\phi}$])	$\bar{\sigma}^2$ (s.e.[$\bar{\sigma}^2$])	$\bar{\phi}$ (s.e.[$\bar{\phi}$])	$\bar{\sigma}^2$ (s.e.[$\bar{\sigma}^2$])
1	$\{\phi = 0.05,$ $\sigma^2 = 1\}$	300	0.047 (0.018)	0.964 (0.223)	0.043 (0.015)	1.091 (0.226)
		900	0.042 (0.011)	1.004 (0.189)	0.043 (0.010)	1.034 (0.161)
2	$\{\phi = 0.05,$ $\sigma^2 = 2.5\}$	300	0.044 (0.013)	2.362 (0.601)	0.043 (0.012)	2.421 (0.575)
		900	0.041 (0.013)	2.305 (0.578)	0.042 (0.012)	2.311 (0.554)
3	$\{\phi = 0.05,$ $\sigma^2 = 5\}$	300	0.043 (0.014)	3.844 (0.876)	0.042 (0.013)	3.906 (0.943)
		900	0.044 (0.014)	3.934 (0.953)	0.044 (0.013)	3.973 (1.029)
4	$\{\phi = 0.1,$ $\sigma^2 = 1\}$	300	0.082 (0.034)	0.925 (0.285)	0.074 (0.028)	1.013 (0.272)
		900	0.076 (0.022)	0.947 (0.233)	0.075 (0.020)	0.985 (0.224)
5	$\{\phi = 0.1,$ $\sigma^2 = 2.5\}$	300	0.079 (0.027)	1.951 (0.485)	0.074 (0.024)	2.040 (0.495)
		900	0.074 (0.023)	2.044 (0.446)	0.072 (0.021)	2.087 (0.433)
6	$\{\phi = 0.1,$ $\sigma^2 = 5\}$	300	0.074 (0.030)	3.195 (0.862)	0.071 (0.027)	3.277 (0.866)
		900	0.073 (0.023)	3.481 (0.892)	0.068 (0.021)	3.607 (1.028)
7	$\{\phi = 0.25,$ $\sigma^2 = 1\}$	300	0.165 (0.075)	0.725 (0.263)	0.137 (0.062)	0.828 (0.262)
		900	0.162 (0.062)	0.775 (0.280)	0.150 (0.054)	0.827 (0.283)
8	$\{\phi = 0.25,$ $\sigma^2 = 2.5\}$	300	0.153 (0.054)	1.625 (0.579)	0.138 (0.049)	1.720 (0.598)
		900	0.148 (0.052)	1.571 (0.454)	0.138 (0.046)	1.636 (0.477)
9	$\{\phi = 0.25,$ $\sigma^2 = 5\}$	300	0.134 (0.050)	2.412 (0.726)	0.124 (0.048)	2.531 (0.797)
		900	0.128 (0.046)	2.583 (0.656)	0.119 (0.044)	2.685 (0.719)

Table 3.2: Study 1: MCS results, based on the K-function and PCF.

using the mean parameter estimates in the table. Specifically, for each MCS, we compute $\text{ARE}_V(\bar{\phi}) = |\bar{\phi} - \phi|/\phi$ and $\text{ARE}_V(\bar{\sigma}^2) = |\bar{\sigma}^2 - \sigma^2|/\sigma^2$ for both $V = K$ and $V = g$. These quantities are plotted separately for the two sample sizes in Figure 3.25.

As we observed in the table, the discrepancies between the true and mean estimated parameter values worsen as the scale and variability of the dependence structure increases alongside the MCS number. Generally speaking, it appears that the K-function produces better estimates than g of the scale parameter ϕ, at the cost of poorer estimates of the variance parameter σ^2, and vice-versa. It is therefore extremely difficult to conclude if, overall, one function produces more accurate results than the other. Increasing the sample size has had little impact on the pattern of AREs, though it could be argued that the 'poorer' σ^2 estimates from K are closer to those obtained from g than before.

These results have demonstrated a tendency for the minimum contrast methods applied to spatial LGCPs using the K- and pair correlation functions to underestimate the true values. Now, recall once more the findings in Guan and Sherman (2007), discussed in Section 3.2.3, where the authors theoretically demonstrated consistency of minimum contrast parameter estimates using the K-function. It would seem that at the sample sizes employed here, the discrepancies of the estimates with respect to the true values is still noticeable. Particularly interesting is Figure 1 in Guan and Sherman (2007). Visually inspecting the histograms of their estimated ϕ and σ^2 values (1000 iterations, sample sizes of 400), it does appear that underestimation (of the mean result) is also present in their work (though they do not comment on this feature explicitly). This is especially visible in the plots corresponding to their estimates of σ^2. Our study mirrors these tendencies, and it would seem similar behaviour is occurring in the PCF results.

It could be the case that the exhibited discrepancies are of negligible practical concern;

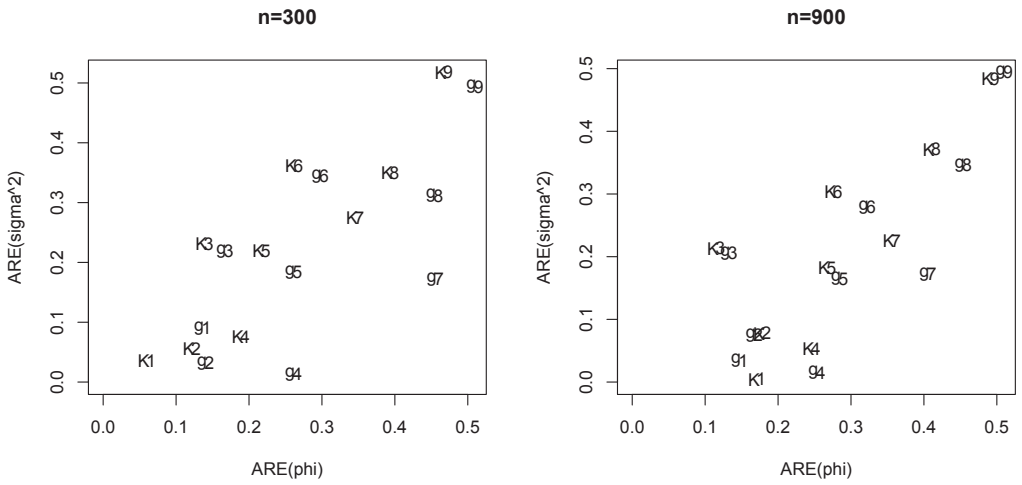


Figure 3.25: Absolute relative errors (ARE) for the mean parameter estimates produced from the simulations of MCS1 to MCS9, with respective labelling 'K1' to 'K9' for the K-function results, and 'g1' to 'g9' for the PCF results.

this will be the focus of Study 3 below. The issue of K versus g is unclear, though our results have indicated conflicting performances depending on which parameter we are focused on. One possible 'rule-of-thumb' strategy, based on Figure 3.25, would be for the researcher to decide which parameter is of the most interest, choosing K if it is ϕ and g if it is σ^2. Alternatively, we might choose to use K to estimate ϕ, ignoring the K-chosen σ^2, but then use g to estimate σ^2, ignoring the g-chosen ϕ. As always in numerical experiments, however, it is stressed that appropriate caution should be observed with respect to the drawn conclusions.

3.5.2 Study 2: Deterministic bandwidth impact on spatial parameter estimation

The previous study specifically examined the performance of the popular minimum contrast approach to parameter estimation for the stationary and isotropic LGCP. However we argued in Section 3.1.3 that it is more natural to consider the LGCP as a stochastic term describing 'residual' or local correlation, and involve a multiplicative deterministic quantity to actually define the global structure of any heterogeneity. We may subsequently take advantage of the assumption of second-order intensity-reweighted stationarity, and estimate the spatial residual process parameters σ^2 and ϕ through straightforward extensions of K and g; namely their 'inhomogeneous' counterparts (Baddeley et al., 2000).

As we have also seen, particularly in Chapter 2, kernel smoothing provides a convenient and flexible method of estimating the underlying spatial intensity functions deterministically. Pivotal, however, is the choice of bandwidth(s) used to smooth the global trend; this having an obvious impact on the resulting minimum contrast estimates of the stochastic process parameters. Too great, and we will overestimate the true level of dependence. Too small,

and we may miss detecting any correlation whatsoever.

It is unknown just how sensitive the resulting estimates of σ^2 and ϕ are to the initial choice of smoothing should we assume our intensity is modelled this way, namely $\Lambda(\boldsymbol{x}) = \zeta(\boldsymbol{x})\exp[Y(\boldsymbol{x})]$, where $\zeta(\boldsymbol{x})$, the global spatial trend, is replaced by $\hat{\zeta}(\boldsymbol{x})$: a kernel-smoothed intensity estimate; $\boldsymbol{x} \in W$, $W = [0,1] \times [0,1]$. Study 2 therefore involves specification of several synthetic Deterministic Minimum Contrast Scenarios (DMCS) to address this issue.

We begin by designing three different fixed global intensities – with what we deem to roughly be described as mild, moderate, and strong levels of 'natural' inhomogeneity; labelled 'A', 'B' and 'C' respectively. These, defined using appropriately scaled normal mixtures, are displayed in Figure 3.26 and given algebraically in Table 3.3, where $\phi_W(\boldsymbol{a}, b\boldsymbol{I})$ denotes the two-dimensional normal density edge-corrected to W and centered at $\boldsymbol{a}$ with scalar standard deviation b; $\boldsymbol{I}$ is the 2×2 identity matrix.

Label	DMCS	$\zeta(\boldsymbol{x}) \propto \ldots$
A	1,2	$0.1 + 0.3\phi_W(\boldsymbol{x} - [0.5, 0.5]^\top, 0.1\boldsymbol{I}) + 0.3\phi_W(\boldsymbol{x} - [0.7, 0.6]^\top, 0.1\boldsymbol{I})$ $+0.3\phi_W(\boldsymbol{x} - [0.8, 0.9]^\top, 0.2\boldsymbol{I})$
B	3,4	$0.1 + 0.1\phi_W(\boldsymbol{x} - [0.1, 0.9]^\top, 0.2\boldsymbol{I}) + 0.1\phi_W(\boldsymbol{x} - [0.3, 0.9]^\top, 0.2\boldsymbol{I})$ $+0.2\phi_W(\boldsymbol{x} - [0.5, 0.8]^\top, 0.2\boldsymbol{I}) + 0.1\phi_W(\boldsymbol{x} - [0.7, 0.9]^\top, 0.2\boldsymbol{I})$ $+0.1\phi_W(\boldsymbol{x} - [0.9, 0.9]^\top, 0.2\boldsymbol{I}) + 0.3\phi_W(\boldsymbol{x} - [0.5, 0.1]^\top, 0.1\boldsymbol{I})$
C	5,6	$0.1 + 0.3\phi_W(\boldsymbol{x} - [0.1, 0.5]^\top, 0.1\boldsymbol{I}) + 0.2\phi_W(\boldsymbol{x} - [0.7, 0.9]^\top, 0.1\boldsymbol{I})$ $+0.4\phi_W(\boldsymbol{x} - [0.9, 0.2]^\top, 0.05\boldsymbol{I})$

Table 3.3: Definitions of global spatial trends used in Study 2.

It is the product of these global intensity functions and the stochastic component which will define the intensity for each DMCS, from which each data set will subsequently be generated. For each global trend, the magnitude of the residual correlation is assigned one of two possible levels: 'concentrated', with $\sigma^2 = 9$ and $\phi = 0.009$, and 'diffused', with $\sigma^2 = 2$ and $\phi = 0.02$ (the Exponential correlation function is again employed as r). We therefore have six scenarios, with DMCS1 through DMCS6 representing 'A/Concentrated', 'A/Diffused', 'B/Concentrated', 'B/Diffused', 'C/Concentrated' and 'C/Diffused' global/local inhomogeneities respectively. It is important to remember that the parameters controlling dependence in the presence of a deterministic intensity as we have defined in Section 3.1.3 and use in this study must clearly be considered on far smaller, local scales in comparison to

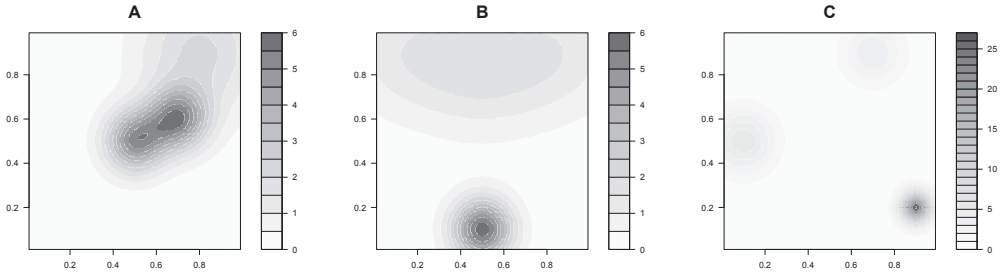


Figure 3.26: Deterministic spatial trends for Study 2.

intensities whose heterogeneity is due to correlation alone (as originally defined in Section 3.1.1 and used in Study 1). Also, note that we use the terms 'concentrated' and 'diffused' rather loosely here and that all additional dependence is of course attractive, not repulsive. 'Concentrated' refers to the fact we observe short-range, more intense clusters, and 'diffused' refers to a longer range yet less intense level of residual correlation.

To give the reader a better idea of the behaviour of these slightly more complicated scenarios, Figure 3.27 gives single point pattern examples of size $N(W) = 1000$ for each of DMCS1 through DMCS6, alongside an example of a point pattern generated using the fixed heterogeneous spatial trend alone ('no correlation').

The study proceeds as follows. For each scenario, 100 realisations of the intensity function are generated, followed by the generation of corresponding point patterns of size $n \in \{300, 900\}$. Estimation of the global trends is then performed using a fixed-bandwidth kernel density estimator with the Gaussian kernel function (for computational reasons and the complexities of bandwidth assignment the adaptive estimator is not considered here). This estimation is conducted using a sequence of bandwidths between 0.03 and 0.3, followed by minimum contrast estimation of σ^2 and ϕ using K and g with the same calibrations as earlier; though $s_{\max}$ is reduced to 0.05 in recognition of the fact the LGCP is now describing only local features.

Considering the two parameters as coordinates, Table 3.4 gives the pair of mean estimates $\bar{\sigma}^2$ and $\bar{\phi}$ which are nearest to the coordinates of the true values via what we will term the *relative Euclidian distance* (RED). This is given as $\text{RED}_V(\bar{\phi}, \bar{\sigma}^2) = \sqrt{\text{ARE}_V(\bar{\sigma}^2)^2 + \text{ARE}_V(\bar{\phi})^2}$, where $\text{ARE}_V(\cdot)$ is as defined in Study 1 ($V = K$ or $V = g$); thus we identify $(\bar{\phi}_\star, \bar{\sigma}^2_\star) = \text{argmin } \text{RED}_V(\bar{\phi}, \bar{\sigma}^2)$. Also given is the bandwidth $h_\star$ used in kernel estimation of the global trend ζ which achieved the result. Visually, the results for $n = 900$ are given in Figures 3.28 (DMCS1 to DMCS3) and 3.29 (DMCS4 to DMCS6).

All scenarios seem to perform relatively well when it comes to estimation of the scale parameter ϕ, though there is considerable underestimation of σ^2 in the 'concentrated' problems when $\sigma^2 = 9$, where it seems the minimum contrast technique, coupled with the kernel estimation of the global trend, is simply unable to reach such high levels on our given observation window. There appears to be little difference in the values of $\bar{\phi}_\star$, $\bar{\sigma}^2_\star$ and $h_\star$ between g and K, though there is a tendency for g to provide slightly smaller values of $\bar{\phi}_\star$ with slightly larger values of $\bar{\sigma}^2_\star$ when compared with K. Mirroring Study 1, altering the sample size as we have done has had minimal impact on the estimation of the parameters in these scenarios.

Perhaps most interesting, however, is the fact that the 'optimal' pair of parameter means in Table 3.4 is consistently achieved with rather large bandwidths. Smoothing parameters for the Gaussian kernel intensities in the order of 0.2 on the unit square correspond to what many would describe as oversmoothed estimates; an 'off-the-shelf' bandwidth selector such as least-squares cross-validation (LSCV) almost always produces bandwidths less than 0.1 for the data sets we considered.

Intuitively, this does make sense. Many bandwidth selectors for kernel intensity estimation, like LSCV, are derived with some form of the assumption that we aim to describe *all* features of the density from which the points have been generated with the kernel estimate

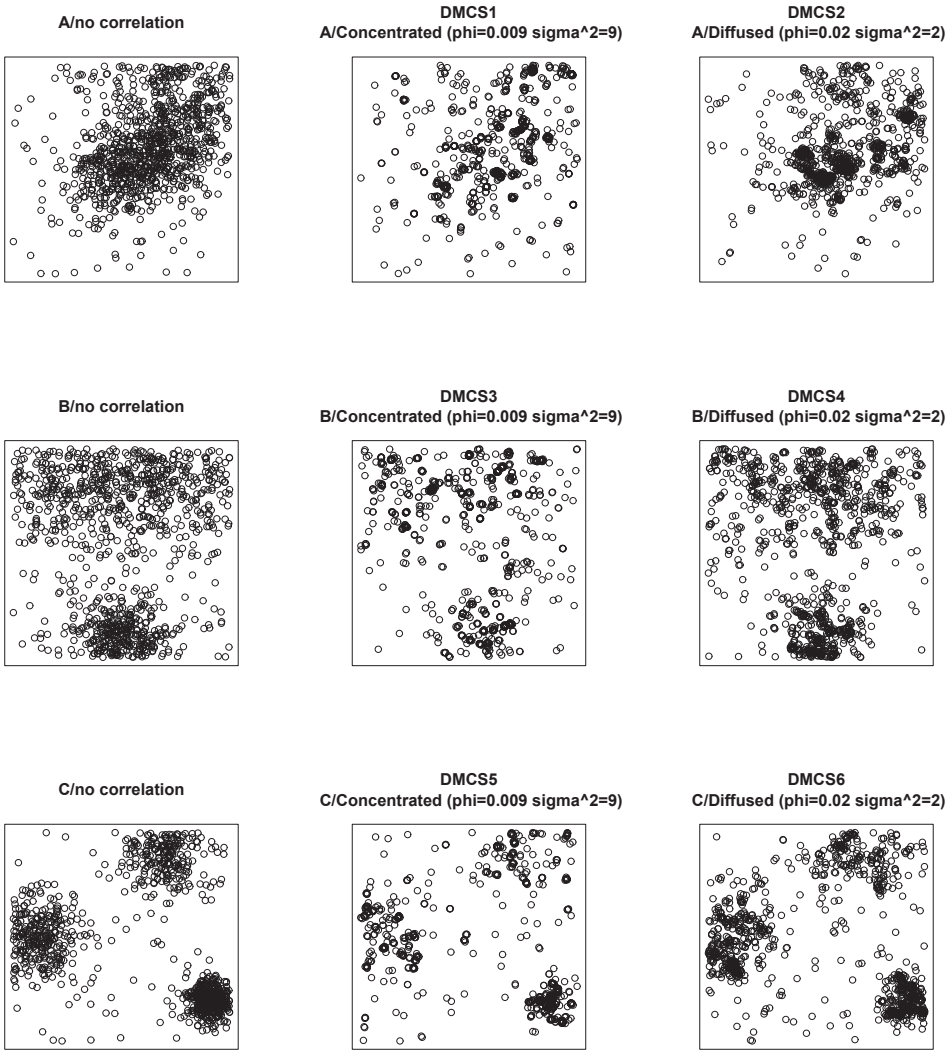


Figure 3.27: Examples of the kinds of point patterns produced under DMCS1 to DMCS6, alongside uncorrelated realisations based on the global trends 'A', 'B' and 'C' alone.

DMCS	V	n	$\bar{\phi}_\star$ (s.e.$[\bar{\phi}_\star]$)	$\bar{\sigma}^2_\star$ (s.e.$[\bar{\sigma}^2_\star]$)	$h_\star$
1	g	300	0.0094 (0.0002)	5.7853 (0.0971)	0.19
		900	0.0091 (0.0002)	6.4541 (0.0998)	0.21
	K	300	0.0092 (0.0002)	6.3661 (0.0976)	0.21
		900	0.0092 (0.0002)	6.2515 (0.0762)	0.20
2	g	300	0.0197 (0.0005)	2.1285 (0.0362)	0.21
		900	0.0199 (0.0004)	2.098 (0.0339)	0.20
	K	300	0.0198 (0.0005)	1.9169 (0.0469)	0.18
		900	0.0204 (0.0004)	1.9388 (0.0345)	0.19
3	g	300	0.0095 (0.0001)	5.6898 (0.0762)	0.18
		900	0.0092 (0.0002)	6.3714 (0.1037)	0.20
	K	300	0.0095 (0.0002)	6.3532 (0.0763)	0.21
		900	0.0094 (0.0001)	6.3083 (0.0740)	0.20
4	g	300	0.0198 (0.0004)	2.1347 (0.0318)	0.20
		900	0.0197 (0.0003)	2.1070 (0.0270)	0.19
	K	300	0.0201 (0.0004)	1.8537 (0.0397)	0.17
		900	0.0206 (0.0003)	1.8992 (0.0250)	0.18
5	g	300	0.0093 (0.0002)	5.4666 (0.0710)	0.14
		900	0.0092 (0.0002)	5.8938 (0.0775)	0.15
	K	300	0.0094 (0.0002)	6.1425 (0.0686)	0.16
		900	0.0092 (0.0001)	6.0686 (0.0742)	0.15
6	g	300	0.0188 (0.0004)	2.1596 (0.0372)	0.14
		900	0.0199 (0.0003)	2.1043 (0.0307)	0.14
	K	300	0.0205 (0.0005)	1.9351 (0.0492)	0.13
		900	0.0202 (0.0004)	1.8398 (0.0276)	0.13

Table 3.4: Study 2: optimal pair of mean parameter estimates which have minimised the RED to the true values, with parenthesised standard errors, and corresponding deterministic bandwidth.

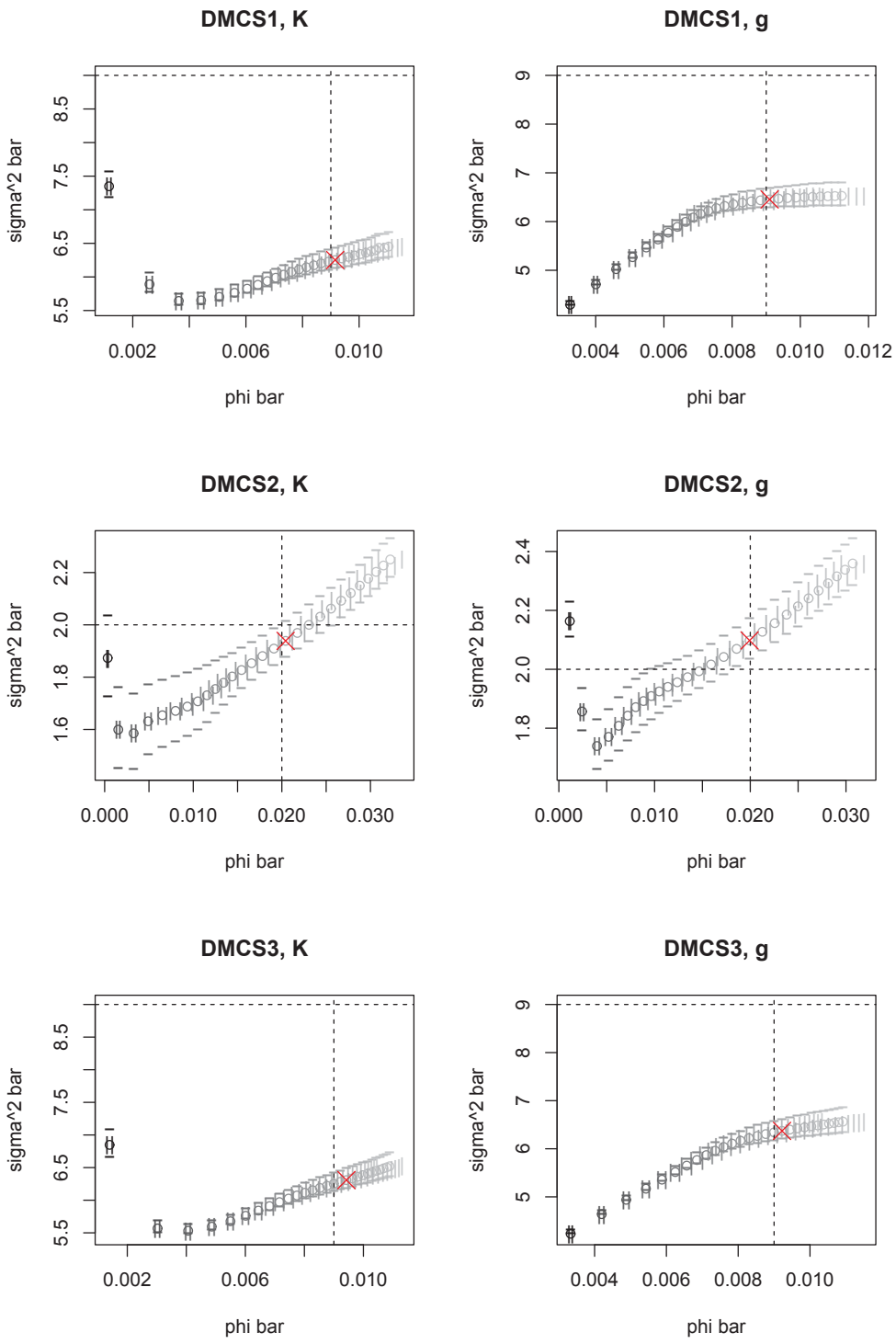


Figure 3.28: Plotting the results of Study 2 for $n = 900$: DMCS1 to DMCS3. Variability of each mean is given as $\pm$2s.e.; '|' for $\bar{\phi}$, '-' for $\bar{\sigma}^2$. The pair minimising the RED is marked with a '$\times$'. Points black$\rightarrow$grey indicate increasing deterministic bandwidth. The location of the true parameter values is given with dashed lines.

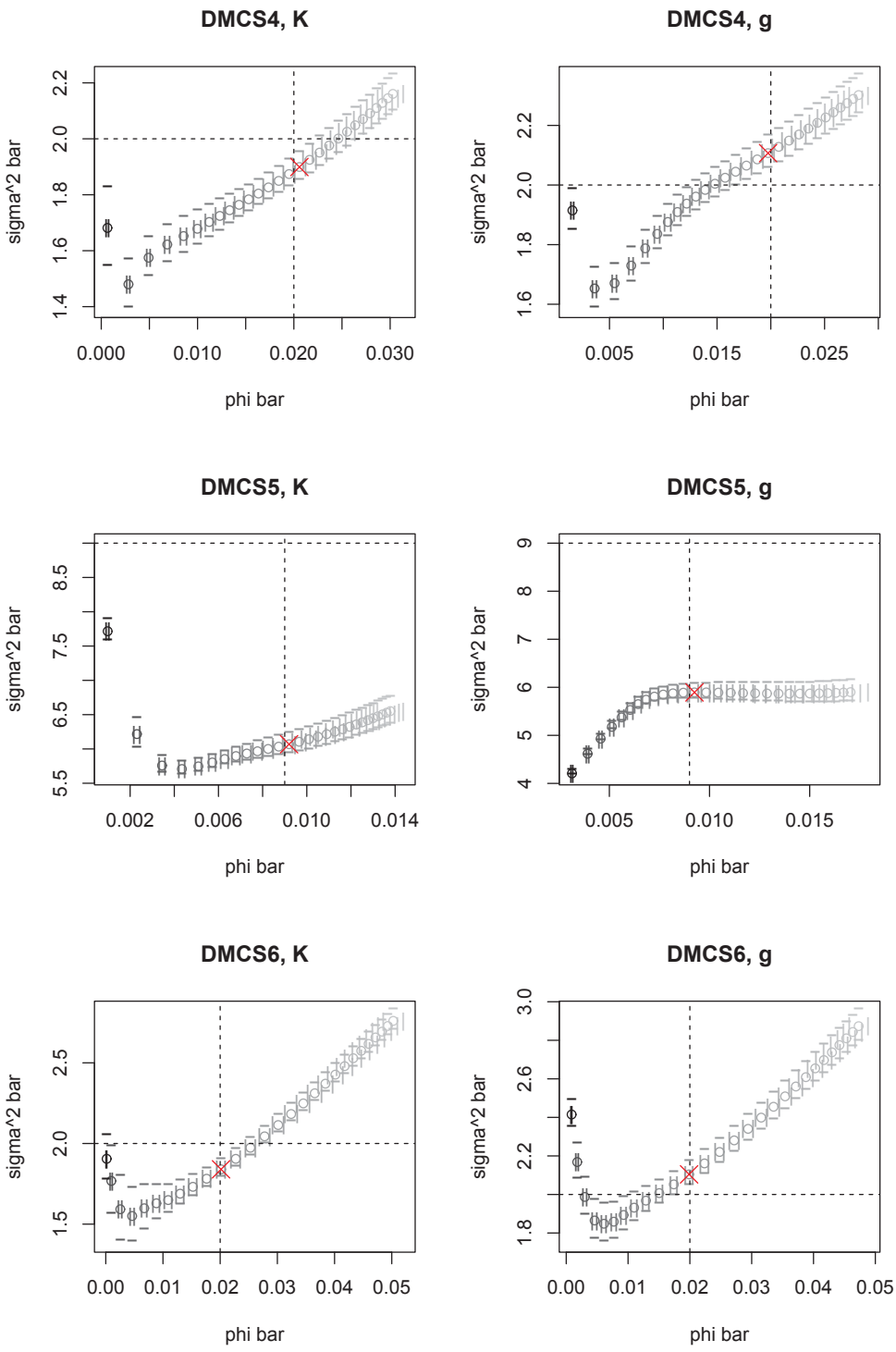


Figure 3.29: Plotting the results of Study 2 for $n = 900$: DMCS4 to DMCS6. Variability of each mean is given as $\pm$2s.e.; '|' for $\bar{\phi}$, '-' for $\bar{\sigma}^2$. The pair minimising the RED is marked with a '$\times$'. Points black$\rightarrow$grey indicate increasing deterministic bandwidth. The location of the true parameter values is given with dashed lines.

itself. However, the global trend ζ clearly does not solely describe the behaviour of the point process under scrutiny in this situation. In an effort to capture both the global heterogeneity *and* the dependency imparted by the residual LGCP, stock-standard bandwidth selectors are therefore very likely to yield inappropriate bandwidths *for estimating the global trend alone*, which in turn will adversely affect our minimum contrast estimates describing the actual levels of dependency present. The notion that we should, to a certain extent, oversmooth the nonparametric estimate of the deterministic trend for our attractive processes is reasonable, and nicely illustrated by these results – a 'general' idea of the overall, global trend requires a large bandwidth, leaving the residual LGCP to be responsible for describing 'finer' details. As we move from the mild global heterogeneity described by 'A' (DMCS1 and DMCS2) to the stronger natural heterogeneity specified by 'C' (DMCS5 and DMCS6) note that the bandwidth yielding the parameter estimates nearest to the truth does drop off somewhat from around 0.2 to roughly 0.15. Presumably, the smaller bandwidth is required to adequately capture the modes of the 'more heterogeneous' global intensity. At the end of the day however, 0.15 is still considered to be a very large level of smoothing for our unit square W.

Inspection of Figures 3.28 and 3.29 shows similar trends between g and K for each DMCS. The inability of both minimum contrast using both of these functions to capture the true variance for the 'concentrated' scenarios is clear in the plots of DMCS1, DMCS3 and DMCS5. This is more difficult to explain; it is likely affected by both the corresponding true ϕ and the size of W. The images also highlight the dangers associated with 'reckless' oversmoothing. This will effectively force the global trend to be uniform, and in contrast to undersmoothing, will therefore leave the residual LGCP responsible for describing all heterogeneity in the points. In turn, we might expect the resulting correlation parameters to be overestimated.

This study has demonstrated the impact of varying the bandwidth used to smooth a kernel estimate of the intensity of a global spatial trend, when the spatial intensity can be multiplicatively decomposed using this deterministic global trend and a unit-mean LGCP describing residual correlation. The simulations have revealed little difference between use of g and K for subsequent minimum contrast estimation of the residual process parameters. In particular, we conclude that there appears to be a preference towards restrained oversmoothing of the global trend estimate in order to adequately capture the true levels of dependence when we are dealing with the kind of attractive processes more likely to be observed in epidemiological applications.

Nevertheless, it should be stressed that this conclusion merely reflects the results from a set of artificial scenarios. In practice, it is possible that expert epidemiological knowledge can also justify the choice of bandwidth and in some settings lead to the same conclusion. For example, the researchers may well have an educated idea of the effective infective range of the disease under scrutiny. If this is on a small scale relative to the size of W, then they may wish to use the stochastic component to capture this behaviour as residual correlation, leaving an oversmoothed deterministic component to appropriately model the fixed heterogeneity in overall disease prevalence throughout the study region.

3.5.3 Study 3: Spatiotemporal exceedance probability sensitivity

The previous two simulation studies have examined the performance of the commonly employed, albeit quite *ad hoc*, minimum contrast approach to parameter estimation of the terms defining the correlation structure in the spatial LGCP. They illustrated the degree to which we can expect the estimates of the spatial parameters to vary, and the rates at which they do so when additional aspects of the modelling framework, such as the deterministic kernel-smoothed heterogeneous 'global' intensity, are present. Though these results are very useful on their own, a key question remains: does this degree of parameter estimate variation *actually matter in practice*? We pose the question in the context of spatiotemporal disease surveillance where, as in the AEGISS analysis, the goal is to produce real-time maps of exceedance probabilities which describe the chances of having observed 'anomalies' (i.e. unusually extreme behaviour) under the assumed model at each space-time location.

To examine this issue, we design three distinct Spatiotemporal Scenarios (STS) each of which have a 'true' form of the space-time stationary, isotropic Gaussian field which, following the exponentiation that provides the unit-mean 'residual' LGCP, constitutes the sole ingredient for computing the exceedance probabilities; see Section 3.4.2. The true Gaussian field is set as a single, unconditionally generated realisation with a pre-defined σ^2_ψ, ϕ_ψ and θ; using the Exponential correlation function throughout. Spatially constrained to the unit square W, each STS is split into two versions according to a pre-defined temporal interval with a maximum number of integer time points T and constant mean number of observations η in the sense that $\eta = \eta(t)$; $t \in \{1, \ldots, T\}$. In all cases, the global spatial trend is ignored (i.e. made uniform such that $\zeta_\psi(\boldsymbol{x}) = 1$ for all $\boldsymbol{x} \in W$).

The two versions of each STS have either $T = 10, \eta = 100$ or $T = 100, \eta = 10$, i.e. either a spatiotemporal field defined over 10 distinct time points with a mean number of 100 observations (Poisson distributed) at each time, or a longer-term field over 100 time points but with only a mean of 10 observations per time. STS1 has $\sigma^2_\psi = 4$, $\phi_\psi = 0.04$, $\theta = 4$; STS2 has $\sigma^2_\psi = 8$, $\phi_\psi = 0.1$, $\theta = 6$; and STS3 has $\sigma^2_\psi = 2$, $\phi_\psi = 0.02$, $\theta = 2$. The first scenario, STS1, represents data sets exhibiting a moderate level of dependence and cluster concentration relative to the others: STS2 has the 'longest-reaching' dependencies in terms of the spatial and temporal lags; as well as the most extreme variance. STS3 denotes a considerably 'weaker' or 'shorter' dependence structure, as well as the smallest field variance. For each of the six problems (three scenarios times two versions), 10 distinct data sets were generated and stored; each one of these representing a single instance of a hypothetically observed data set under each set of conditions for a single realised sample size determined by the original unconditional spatiotemporal field realisations defining the STSs themselves.

Phase 1

This simulation study takes the form of two phases, the first of which involves minimum contrast estimation of the three parameters controlling the stochastic covariance. Incidentally, this phase in itself can be considered interesting; a (limited) opportunity to examine the accuracy and precision of minimum contrast estimation of the temporal dependence

parameter θ being provided. For a given STS and version, and using the corresponding 10 pre-generated data sets, σ^2_ψ, ϕ_ψ and θ are estimated for each; the former two based on the time-averaged version of either K or g, the latter on $\mathcal{C}$ using the results arrived at for σ^2_ψ and ϕ_ψ. Calibration of the contrast integrals (e.g. function transformation) follows the decisions made in Section 3.2.3 and also followed in the previous two simulation studies. Table 3.5 gives the mean of the estimated parameters ($\bar{\sigma}^2_\psi$, $\bar{\phi}_\psi$ and $\bar{\theta}$) as well as the corresponding standard errors for each scenario and version.

STS	$T;\eta$	V	$\bar{\phi}_\psi$ (s.e.[$\bar{\phi}_\psi$])	$\bar{\sigma}^2_\psi$ (s.e.[$\bar{\sigma}^2_\psi$])	$\bar{\theta}$ (s.e.[$\bar{\theta}$])
1 $\{\phi = 0.04, \sigma^2 = 4, \theta = 4\}$	10; 100	K	0.0361 (0.0002)	3.9810 (0.0196)	1.5673 (0.0094)
		g	0.0376 (0.0001)	3.9039 (0.0183)	1.5294 (0.0087)
	100; 10	K	0.0263 (0.0003)	3.6088 (0.0335)	4.5564 (0.0598)
		g	0.0287 (0.0002)	3.1826 (0.0152)	3.9776 (0.0202)
2 $\{\phi = 0.1, \sigma^2 = 8, \theta = 6\}$	10; 100	K	0.1043 (0.0002)	5.2463 (0.0065)	4.4921 (0.0196)
		g	0.0956 (0.0002)	5.4259 (0.0069)	4.9436 (0.0254)
	100; 10	K	0.0522 (0.0003)	5.0739 (0.0189)	6.1134 (0.0405)
		g	0.0493 (0.0002)	5.0622 (0.0162)	6.1699 (0.0318)
3 $\{\phi = 0.02, \sigma^2 = 2, \theta = 2\}$	10; 100	K	0.0302 (0.0005)	1.8243 (0.0250)	0.8185 (0.0067)
		g	0.0226 (0.0003)	2.1881 (0.0199)	0.9411 (0.0062)
	100; 10	K	0.0136 (0.0004)	2.0677 (0.0732)	1.6500 (0.0363)
		g	0.0099 (0.0002)	2.6932 (0.0094)	1.9772 (0.0055)

Table 3.5: Study 3, Phase 1: Means and associated standard errors of the minimum contrast-estimated correlation parameters for both versions of each STS, based on the hypothetically observed data sets.

An examination of these figures shows generally what we would expect to see given our design specifications for each STS, though there are some exceptions. Where $T = 10$ but $\eta = 100$, note the estimates of the spatial correlation parameters, σ^2_ψ and ϕ_ψ, tend to be closer to their true counterparts compared to when $T = 100$, $\eta = 10$. Naturally, the fact that we have far more spatial data at each possible time dictates these values, based on the second-order spatial characteristics, should be more reliable than the version where $T = 100$ but η is only 10. Unfortunately, this means the estimates of θ for the $T = 10$, $\eta = 100$ versions suffer for much the same reasons! With far fewer time points in the data sets, the estimation procedure struggles to accurately capture the the temporal correlation. For example, note in STS1 for $T = 10$, $\eta = 100$, the mean estimated spatial parameters are notably superior than the $T = 100$, $\eta = 10$ version (for use of both K and g), but θ is drastically underestimated, sitting at around 1.5 (true value is 4). Conversely, inspection of $T = 100$, $\eta = 10$ gives far superior estimates of θ at the cost of misspecifying σ^2_ψ and ϕ_ψ to a greater extent. This story is mirrored in the results for STS2 and once more in STS3, albeit with some overestimation of the spatial parameters in the latter. Note that STS3 is the scenario which has the weakest range of spatiotemporal correlation, with $\phi_\psi = 0.02$ and $\theta = 2$. Estimation of these 'finer' parameters to a similar level of satisfaction as the earlier scenarios could conceivably require a far larger data set, in both space and time. This is where the issue of spatial and temporal discretisation levels also comes into play, with smaller dependency ranges requiring finer discrete approximations in order to be detectable

in practice. We leave STS3 as-is, and consider this a scenario where the researcher has small-scale dependence in their space-time data and may therefore be more susceptible to utilising too coarse a discretisation in conditional simulation, which we perform in the following phase of this simulation study.

As observed in Study 1 and Study 2, the differences between results based on using either the K-function or the pair correlation function for estimation of the spatial parameters are relatively small. It is nevertheless comforting to note that the time-averaged versions of the relevant nonparametric estimators appear to behave similarly to the 'spatial only' estimators. As a consequence, the differences in estimation of θ between K and g is also relatively minor. One noteworthy point is the fact that in almost all cases, the level of variation associated with use of g was lower than that of the estimates obtained with K, also noticed in the previous simulation studies. This is of course of little value, however, if there exists any considerable bias in the estimates themselves.

Prior to entering Phase 2 of the study, we remark that Phase 1 is essentially a 'by-product' of the procedure necessary for the conditional prediction which follows. To draw more precise conclusions on the performance of minimum contrast parameter estimation for the spatiotemporal LGCP, it is of course desirable to use a larger number of data sets. However, the results observed with the 10 generated sets for each version per STS are already useful: they are intuitively sensible, and reflect earlier discoveries in the purely spatial setting. The reason for such a relatively small number of data sets in this simulation study was made primarily for computational reasons in the operations which follow; these details will be clarified momentarily.

Phase 2

By now, we have gained a good understanding of the behaviour of minimum contrast parameter estimation in terms of the proximity of the estimates to the true values. As mentioned at the introduction of this study, however, it remains to be seen if these types of discrepancies impact upon our practical interpretations of the process when we use our observed data to predict the intensity itself. On its own, this is unfortunately a rather vague challenge. The computational expense associated with conditional simulation of the spatiotemporal LGCP (which will in a numerical exercise require multiple MALA runs) dictates that all aspects of the design of Phase 2 must be carefully considered in order to attain meaningful results with respect to clearly defined objectives in a computationally feasible way.

Disease surveillance, as described for the AEGISS data in Diggle, Rowlingson and Su (2005) and Section 3.4.2, is amongst the primary objectives of interest for epidemiological applications utilising the spatiotemporal LGCP. Flagging of 'unusual' space-time clusters of observations in view of the fitted model is achieved with exceedance probabilities, computed empirically as the proportion of retained, exponentiated Gaussian fields (from a corresponding MALA run) which exceed a certain user-defined threshold (chosen to be a suitably high limit such that the researcher is confident in interpreting higher values as a true 'anomaly') at each space-time location. Naturally we should be concerned if, at a given time, the spatial exceedance probability surfaces obtained in practice fail to detect a true anomaly, or conversely, flag an anomaly where there exists none. When these are interpreted as poten-

tial disease outbreaks, the consequences of non-detection or false-flagging in terms of public health as well as economic cost could well be substantial. Do the minimum contrast parameter estimates, and the discrepancies they exhibit with reference to the true values, diminish our ability to arrive at a 'suitably correct' representation of exceedances?

Brix and Diggle (2001) and Diggle, Rowlingson and Su (2005) do provide some sensible conjecture on this issue, albeit without further formal attention. They argue that, in situations where we have a suitably large data set, uncertainty in the minimum contrast estimates can safely be ignored when it comes to conditional simulation of a spatiotemporal LGCP. This statement is made owing to the fact that we use *all* historical data to estimate the parameters, yet only the *last few* time points (user-specified, depending on the perceived strength of the temporal dependence), with relatively few observations, for prediction of the 'current' spatial intensity. We should therefore expect variation in the prediction procedure to dominate (compared to parameter estimation variation); the conditional simulation algorithm subsequently being *relatively* insensitive to the choice of initial parameters.

This is a perfectly reasonable notion, but does depend heavily on just how much historical data we have, as well as the (underlying) expected number of observations per time point. The fact remains that in both Studies 1 and 2, as well as Phase 1 of the current study, we have in some cases noticed rather extreme departures of the minimum contrast estimates from the 'truth'. Are we to expect that even these kinds of discrepancies will have little impact on anomaly detection?

This motivates Phase 2, where we use each of the 10 data sets generated from each STS (and version) to predict the exponentiated spatial Gaussian field at the final time point (as would be of the most interest in a real-time surveillance setting) using the MALA. MALAs of length 44000 are run based on (a) use of the true parameter values; (b) use of the K-function estimated parameters; and (c) use of the parameters estimated with g. In all cases, there is a burn-in period of 4000 iterations, followed by retention of every 100^{th} iteration.

To compare the estimated intensities arising from the conditional simulations with the true form of the latent LGCP we first determine, for each scenario and version, which cells of the $(M-1) \times (N-1) = 64 \times 64$ grid at the last time point T do in fact exceed the 95^{th} percentile threshold. This results in a binary matrix; 1 denoting an exceedance, 0 no exceedance. Label this the 'true anomaly matrix'. For a given STS and version, we then obtain for each of the 10 data sets a collection of 400 retained posterior exponentiated spatial Gaussian fields on the same 64×64 cells from a MALA of type (a), (b) or (c). Subsequent exceedance probability surfaces are computed based on the 95^{th} percentile of the corresponding lognormal distribution (this is dependent in each case on the variance, or estimated variance, of the Gaussian field itself). The performance of the MALAs is judged based on those cells with empirical probability greater than 0.5, that is, the cells for which more than half of the realisations have flagged an exceedance. This results in another binary structure which is labelled the 'predicted anomaly matrix'.

The Jaccard similarity coefficient is used to evaluate the agreement between the true and predicted anomaly matrices. Briefly, for two sets A and B, the Jaccard index is calculated as the size of their intersection divided by the size of their union; $\mathcal{J} = |A \cap B|/|A \cup B|$; $0 \leq \mathcal{J} \leq 1$. In our case, the sets represent the two anomaly matrices; elements are the

cell-wise 'exceedance' (1) or 'no-exceedance' (0) flags. By summing the two binary matrices, $\mathcal{J}$ can quickly be computed as the ratio of the number of cells whose value is 2 to the number of cells whose value is greater than or equal to 1. Thus, a 'perfect' match will be found with $\mathcal{J} = 1$; absolutely no match with $\mathcal{J} = 0$. It is important to note that we do not expect to see any perfect matches, regardless of how well the prediction algorithm executes. Precisely *where* cells are flagged in the predicted anomaly matrix depends on a number of variable components, some of which are extremely difficult to control. The number of observations per time point, cell resolution with respect to the magnitude of the LGCP parameters, as well as the convergence and mixing performance of the MALA chain, will all impact the final exceedance results and magnitudes of the recorded $\mathcal{J}$ statistics. As such, we stress that use of the Jaccard coefficient in this setting should be viewed as a very simple within-scenario, between-MALA type of quality measure.

Prior to execution, it is worth highlighting the variability in the 95% exceedance thresholds themselves. Dependent upon the variance of the Gaussian field, σ^2_ψ, discrepancies between different estimates thereof will flow through to these cutoff values, thereby potentially having an effect on the location and prevalence of flagged exceedances (and therefore our resulting Jaccard indices). Figure 3.30 illustrates this with respect to the 10 data sets specific to each STS and version, using the true and mean parameters estimates of σ^2_ψ in Table 3.5. All scenarios using the minimum contrast-estimated values of the variance exhibit on average an increased 95% threshold when compared to the value associated with the true variance. Furthermore, recall that the LGCP is a 'residual process' with unit mean: the impact on the 'location' (stationary mean) parameter $\mu_\psi = -0.5\sigma^2_\psi$ is also present. These issues, in addition to the obvious use of σ^2_ψ in defining the discretised covariance matrix of the Gaussian field, therefore means that there exists considerable potential for the precise magnitude of the supplied σ^2_ψ to affect a given MALA.

The computational demands of Phase 2 cannot be emphasised enough. The decisions regarding the size of the prediction grid (64×64); the number of distinct data sets (10); the number of scenarios (3); the number of versions of each scenario (2); the number of different model parameterisations or 'MALA types' (3); the MALA length and burn-in (44000 and 4000 respectively); the retention rate (1/100); and the number of previous time points to consider for prediction at the latest time ('lag length' – 5); were all made after careful examination of the performance of the required operations. The above figures were decided upon as a fine balance between relatively manageable computational costs and meaningfulness of the resulting empirics (in terms of variety of scenarios and accuracy and precision of results, as well as validity of the MCMC methods based on an assessment of mixing and convergence). Finally, the $3 \times 2 \times 3 \times 10 = 180$ individual 44000-iteration, minus-5-lagged spatiotemporal MALAs were run, for which several further weeks of execution time using multiple machines was required.

Results are given in Table 3.6 as the mean Jaccard scores $\bar{\mathcal{J}}_{\text{TRUE}}$, $\bar{\mathcal{J}}_K$ and $\bar{\mathcal{J}}_g$, obtained for each of the 10 data sets specific to each STS and version by comparing the corresponding predicted anomaly matrix with the true anomaly matrix. Note the subscripts TRUE, K and g refer to MALAs (a), (b) and (c) respectively, that is, using the true parameter values, the individually-derived K-function parameter values and the individually-derived

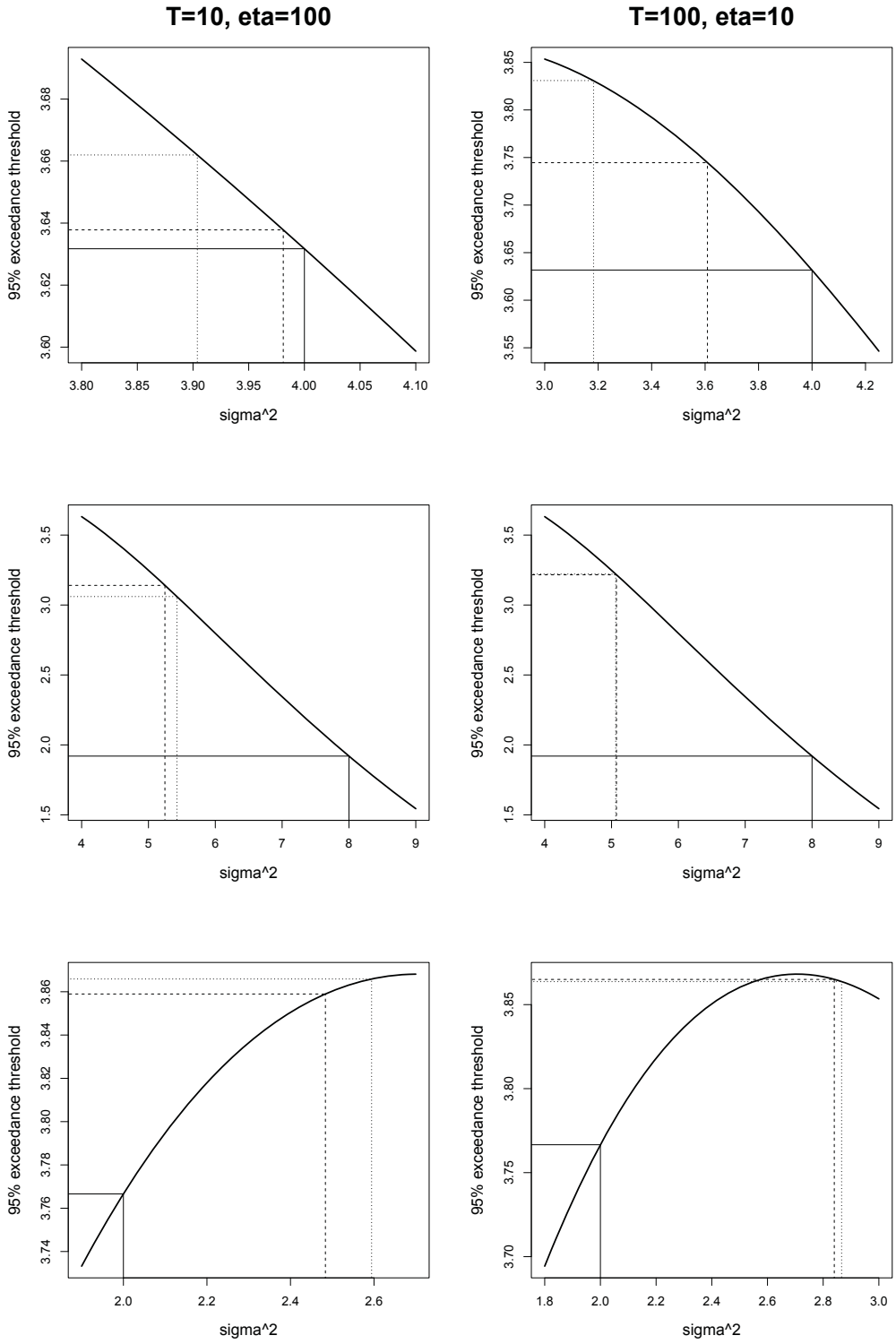


Figure 3.30: Cutoff threshold value (95^{th} percentile) by variance σ^2_ψ (solid, bold line). Values for STS1 (top), STS2 (middle) and STS3 (bottom), are given for the true variance (solid line), mean minimum contrast estimate by K (dashed line), and mean minimum contrast estimate by g (dotted line). Versions $T = 10; \eta = 100$ and $T = 100; \eta = 10$ are given in the left and right columns respectively.

PCF parameter values.

STS	$T; \eta$	$\bar{\mathcal{J}}_{\text{TRUE}}$ (s.d.$[\mathcal{J}_{\text{TRUE}}]$)	$\bar{\mathcal{J}}_K$ (s.d.$[\mathcal{J}_K]$)	$\bar{\mathcal{J}}_g$ (s.d.$[\mathcal{J}_g]$)
1 $\{\phi = 0.04$ $\sigma^2 = 4$ $\theta = 4\}$	10; 100	0.1631 (0.1004)	0.2166 (0.0367)	0.2094 (0.0429)
	100; 10	0.0566 (0.0312)	0.0521 (0.0141)	0.0513 (0.0248)
2 $\{\phi = 0.1$ $\sigma^2 = 8$ $\theta = 6\}$	10; 100	0.5212 (0.0606)	0.5361 (0.0800)	0.5825 (0.0663)
	100; 10	0.3905 (0.0752)	0.2763 (0.0405)	0.2622 (0.0365)
3 $\{\phi = 0.02$ $\sigma^2 = 2$ $\theta = 2\}$	10; 100	0.1036 (0.0226)	0.0873 (0.0295)	0.0976 (0.0267)
	100; 10	0.0073 (0.0071)	0.0078 (0.0105)	0.0108 (0.0083)

Table 3.6: Study 3, Phase 2: Means and standard deviations of the Jaccard indices for both versions of each STS, based on MALA prediction of the final time point exceedances.

In general, 'within scenarios', there appears to be little effect on the Jaccard scores of using the minimum contrast-estimated parameters when performing prediction of the intensity on the final day, regardless of STS or version, in comparison to using the true parameters. Indeed, some cases even show higher scores for the MALAs corresponding to the estimated parameters when compared to the MALAs executed using the true values themselves (for example STS1, version $T = 10, \eta = 100$; STS3, version $T = 100, \eta = 10$). The fact that we sometimes get better results with estimated parameters may not just be due to simulation noise. In each case, the generated datasets and associated results are based on a single realisation of the spatiotemporal Gaussian process – it could be that the characteristics of these processes (in particular, for anomaly detection) are better described using the estimated rather than the true parameters. Between the two versions within each scenario, there are noticeably lower Jaccard scores in the $T = 100, \eta = 10$ setting. This is understandable: with far fewer observations at the time point of prediction, there are clearly fewer cells which will be capable of producing intensity levels of a magnitude sufficient to flag an exceedance.

Between scenarios, the highest Jaccard indices are reserved for those problems with longer ranges of dependence (e.g. STS2): this is intuitively sensible when we consider the fact that the larger the range of space-time dependence, the larger the exceedance sub-regions are likely to be in terms of the number of adjacent cells making up a given hotspot. This in turn means any generated data will be less 'sporadically clustered' in terms of where point aggregation occurs, there therefore being less chance of observing 0-1, 1-0 'disagreements' between the true and predicted anomaly matrices when computing the Jaccard indices.

The take-home message, however, remains the same: comfortingly, these results have indicated no substantial drop in overall agreement quality of detected exceedances when comparing the mean Jaccard scores across the different parameterisations for any of the scenarios. Thus, in spite of the discrepancies observed in minimum contrast estimation (cf. Table 3.5), this method of obtaining parameter values appears to be adequate for sensible

prediction of exceedance probabilities in a disease surveillance setting.

The importance of sample sizes at each time point, as well as the strength of the spatiotemporal correlation in the specific cells flagged as exceedances, is illustrated for STS1, STS2, and STS3 in Figures 3.31, 3.32, and 3.33 respectively. For both versions of each scenario, these images provide the true anomaly matrices, as well as the 'best' predicted anomaly matrix defined as the exceedance surface with the highest Jaccard index (generically labelled as $\mathcal{J}_{\max}$), for each MALA.

Plots corresponding to STS1 (Figure 3.31) show that although $\bar{\mathcal{J}}_{\text{TRUE}}$ was less than $\bar{\mathcal{J}}_K$ and $\bar{\mathcal{J}}_g$ for version $T = 10, \eta = 100$ in Table 3.6, the highest individual Jaccard score from use of the true parameter values was greater than the highest individual score for either of the MALAs using the estimated parameter values. This highlights the fact that the mean Jaccard indices do not by themselves determine the quality of agreement, and that a modest degree of variation should be expected due to the restriction on the number of distinct spatiotemporal data sets which had to be made for computational reasons. In all MALA predictions for both versions, it is clear from comparison with the appropriate true anomaly matrix that they fail to locate absolutely all anomalous cells. In reference to an earlier comment, however, we simply cannot expect near-perfect matches: the specific positions of cells detected as exceedances will depend on the (stochastically) realised point patterns themselves, the sample sizes at each timestamp, and the quality of the executed MALA. However, a visual interpretation of the anomalous sub-regions that *are* detected from the MALA runs (how a surveilled process would typically be scrutinised in practice), provides much the same message about their actual locations, with little (if any) real loss of information if we compare the best MALA (a) (true parameter values) results with those for MALAs (b) and (c) (K and g minimum contrasts respectively). Cell flagging, as we would expect, is more likely as the size of the anomalous sub-region increases, and so it is the locations of the 'larger' clusters visible in the true anomaly matrices that tend to be detected.

Images for STS2 (Figure 3.32) show clearly the effect of longer-range dependence in the spatiotemporal data. The true anomaly matrix for the $T = 10, \eta = 100$ version shows three clear subregions made up of exceeding cells. The largest contiguous sub-region in the southwestern quadrant is satisfactorily detected in all three of the different MALA runs, and the smaller anomalous area in the northwestern quadrant is also identified. The MALAs seem to have struggled in these particular examples in detecting the 'weakest' anomalous region on the eastern border, though there is some minor identification in MALA (c). This could change on a case-by-case basis however: one of the other data sets could well have generated more data in that particular area resulting in a clearer image. The same comments can be made for the $T = 100, \eta = 10$ version of STS2: satisfactory detection of the largest contiguous anomaly, with sporadic anomalies a little more difficult to identify.

STS3 (Figure 3.33), with the shortest range of spatiotemporal dependence, shows the reader the widespread cellular anomalies, many of which are isolated from any others. Again, those few anomalous sub-regions made up of multiple adjacent cells are those which the algorithm is more likely to predict; version $T = 10, \eta = 100$ shows some visual level of agreement between the different MALAs. Results for the $T = 100, \eta = 10$ version are particularly poor,

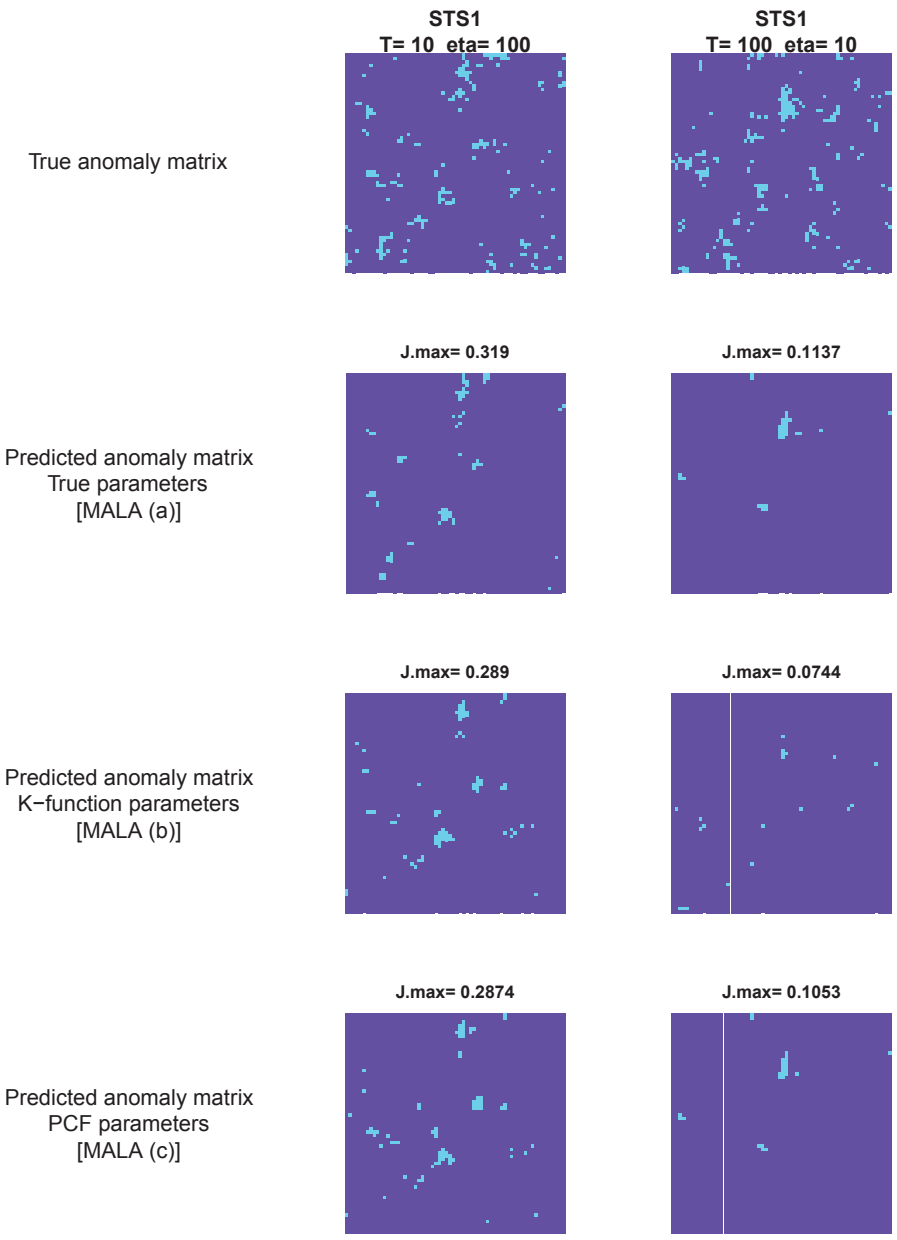


Figure 3.31: True and 'best' predicted anomaly matrices (based on the surface with the highest Jaccard score, $\mathcal{J}_{\max}$, in each case) for both versions of STS1. Cells deemed anomalous as per the discussion are light, dark cells are those which are not.

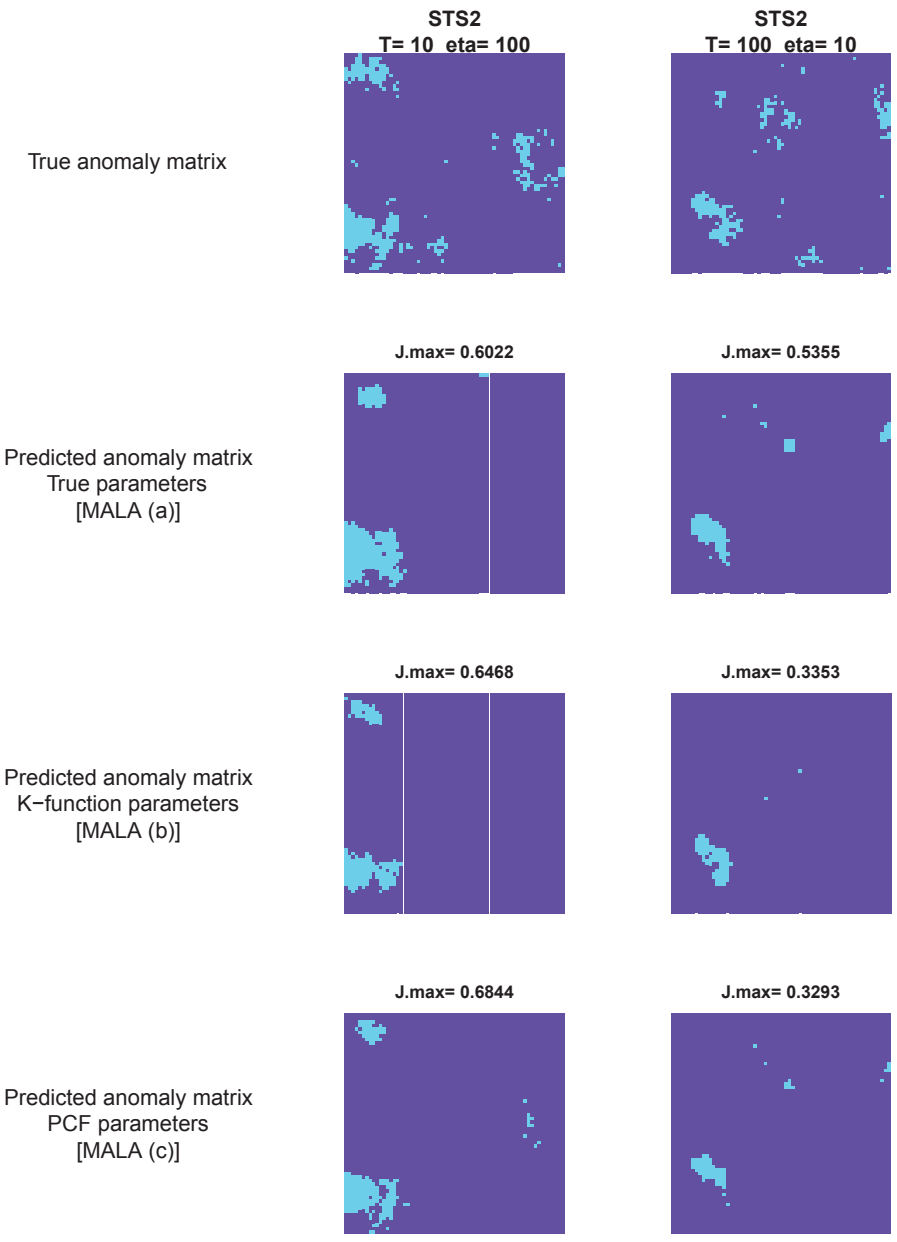


Figure 3.32: True and 'best' predicted anomaly matrices (based on the surface with the highest Jaccard score, $\mathcal{J}_{\max}$, in each case) for both versions of STS2. Cells deemed anomalous as per the discussion are light, dark cells are those which are not.

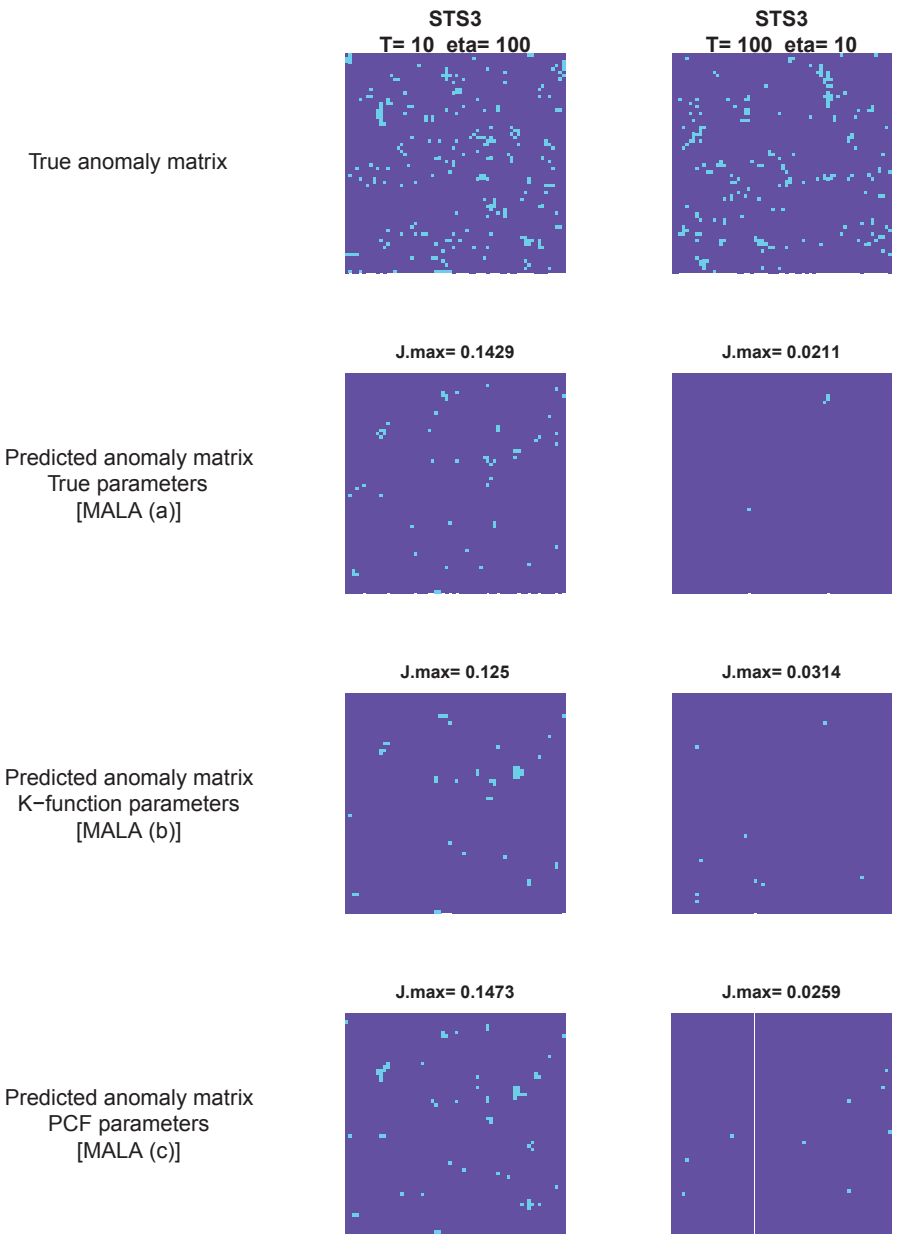


Figure 3.33: True and 'best' predicted anomaly matrices (based on the surface with the highest Jaccard score, $\mathcal{J}_{\max}$, in each case) for both versions of STS3. Cells deemed anomalous as per the discussion are light, dark cells are those which are not.

however. The low timestamp-wise counts, coupled with the weak correlation, have yielded only a handful of detections, thereby making it extremely difficult to compare the results of the various MALAs (even with reference to the true anomaly matrix). Naturally, it is unfair to expect the model to flag exceedances where there exist no data. The relatively low Jaccard scores in this scenario may also be related to the level of discretisation which has been employed for conditional simulation (64×64). With our study region as the unit square, this equates to square cells each of approximate size 0.0156×0.0156. With $\phi_\psi = 0.02$ in STS3, not to mention the occasionally even smaller minimum contrast estimates of this parameter, the given grid resolution may be considered quite coarse, resulting in an inability to evaluate the correlation levels at influential distances. This in turn seems to have affected the capability of adequate exceedance detection, highlighting the caution that researchers must take with respect to definition of the lattice upon which prediction takes place.

From this phased simulation study, we may therefore draw some unique and useful conclusions. Phase 1 indicated performance of the minimum contrast approach, on this occasion making use of the time-averaged versions of K and g, to be consistent with what we had come to expect from Simulation Studies 1 and 2. Intuitively, better estimates of the spatial parameters in terms of proximity to the true values tended to occur when we had more spatial data, and better estimates of the strength of the temporal correlation was apparent when we utilised a greater 'historical' window. Regardless of the observed discrepancies between true and estimated parameters, however, Phase 2 indicated predictive inference to be fairly insensitive to this type of variation. Though understandably unable to provide 100% detection rates when compared with the true form of the exceedances for the final time point, the various MALAs employed showed similar performance to one another in terms of overall agreement via the Jaccard similarity coefficient. A visual inspection of the detected anomalies (based on the prediction with the highest Jaccard score out of the 10 data sets for which the MALA was run) supported these results. Though it is obvious that a number of complex factors influence the exact nature of a predicted anomaly surface, it is clear that more success is achieved in detection of exceedances where there are larger sub-regions of anomalies comprised of several adjacent exceeding cells. Identification of anomalies in problems with short-range dependence, coupled with small event counts, is more difficult; this is also where the issue of too coarse an evaluation grid can easily begin to adversely affect our results. Nevertheless, the fact that within each scenario and version, use of minimum contrast-estimated parameters did not appear to significantly hinder identification of anomalies via empirical exceedance probabilities (when compared to use of the true values), is encouraging: in spite of the many justifiable criticisms associated with this method of parameter estimation, the present results indicate little if any loss of information in an epidemiological surveillance setting.

3.6 Potential theoretical refinements

The relative youth of the LGCP, particularly in the spatiotemporal case, leaves open several avenues for future theoretical advancements. In this section, we briefly describe and discuss three such avenues which build on existing ideas in the more general arena of spatial and

spatiotemporal point process modelling.

3.6.1 Improved 'ratio estimators' of second-order characteristics

The first future research direction is the investigation of so called *adapted ratio estimators* of the second-order properties of a planar point process based on an observed pattern $\boldsymbol{X}$. These ideas, first discussed at length by Stoyan and Stoyan (2000), deal with a modification to the standard $\hat{\lambda} = \mathcal{N}(\boldsymbol{X})/|W|$ estimator of the stationary intensity over the compact region W. The modifications are designed specifically to improve our subsequent estimates of the homogeneous versions of the K-function and the PCF, which in turn means minimum contrast parameter estimation for the LGCP might also be improved in these situations. The following discussion will for simplicity focus on K and g in the purely spatial setting, though the ideas communicate naturally to their spatiotemporal counterparts.

Consider again the nonparametric estimators of the K-function and PCF given as (1.5) and (1.10) in Section 1.2.3 and rewritten here.

$$\hat{K}(u) = \hat{\lambda}^{-2}|W|^{-1} \sum_{\boldsymbol{x},\boldsymbol{y}\in\boldsymbol{X}}^{\boldsymbol{x}\neq\boldsymbol{y}} \mathbf{1}(||\boldsymbol{x}-\boldsymbol{y}|| \leq u) w_W(\boldsymbol{x},\boldsymbol{y}); \qquad 0 \leq u \leq u_{\max},$$

and

$$\hat{g}(u) = \frac{1}{2\pi\hat{\lambda}^2 u|W|} \sum_{\boldsymbol{x},\boldsymbol{y}\in\boldsymbol{X}}^{\boldsymbol{x}\neq\boldsymbol{y}} \kappa_h(||\boldsymbol{x}-\boldsymbol{y}|| - u) w_W(\boldsymbol{x},\boldsymbol{y}); \qquad 0 < u \leq u_{\max},$$

with $u_{\max}$, w_W and κ_h being the upper evaluation limit, Ripley's (1977) isotropic edge-correction factor, and the univariate kernel respectively; all as defined earlier. The supposed problem with both of these estimators lies in the leading $\hat{\lambda}^{-2}$ term, the use of which introduces bias and considerable variance when this quantity is simply the negative square of the $\hat{\lambda} = \mathcal{N}(\boldsymbol{X})/|W|$ estimate.

Stoyan and Stoyan (2000) study some alternative estimators to $\hat{\lambda}$ for use in $\hat{K}$ and $\hat{g}$, which they assess via simulation. Their results do appear to indicate a reduction in the variance of the resulting estimates when compared to using just $\hat{\lambda}$, though any reduction in the bias alone is minimal. These alternatives are referred to as 'adapted' intensity estimates, as they depend upon the distance u at which we are evaluating $\hat{K}$ or $\hat{g}$. Illian et al. (2008) summarise two of these estimators in Section 4.3.3. The first, used in estimation of the K-function, is referred to as the *volume-weighted* estimator, and is given as

$$\hat{\lambda}_V(u) = \frac{\sum_{\boldsymbol{x}\in\boldsymbol{X}} |W \cap D(\boldsymbol{x},u)|}{2\pi \int_0^u y\bar{\gamma}_W(y)\,\mathrm{d}y}, \tag{3.57}$$

where $D(\boldsymbol{x},u)$ represents the disc centered at $\boldsymbol{x}$ with radius u, and $\bar{\gamma}_W(u)$ is the *isotropised set covariance* of the window W (see e.g. Stoyan and Stoyan, 1994; Illian et al., 2008). If W is a rectangle or disc there exist simple, explicit expressions for $\bar{\gamma}_W$; more complicated windows require sophisticated methods such as Fourier transforms in order to be practically feasible (Illian et al., 2008, Appendix B).

The estimator recommended for use in nonparametric computation of the PCF is the

surface-weighted estimator, with

$$\hat{\lambda}_S(u) = \frac{\sum_{\boldsymbol{x}\in\boldsymbol{X}} l[W \cap c(\boldsymbol{x}, u)]}{2\pi u \bar{\gamma}_W(u)}, \tag{3.58}$$

where $l[W \cap c(\boldsymbol{x}, u)]$ denotes the length of the circumference of the disc centered at $\boldsymbol{x}$ with radius u that falls within W.

We therefore achieve the adapted estimators for K and g by replacing $\hat{\lambda}$ above by $\hat{\lambda}_V$ and $\hat{\lambda}_S$ respectively. These techniques hold potential for further investigation with respect to parameter estimation for the LGCP, though the computational burden of (3.57) and (3.58) may prove problematic. Furthermore, it is not necessarily clear how these adapted ratio estimators could be extended to the situation in which our LGCP modelling framework includes deterministic components and the inhomogeneous estimators for K and g must be used.

3.6.2 Multivariate LGCP modelling

In this chapter we have been concerned with the idea of *univariate* LGCPs. These models can be generalised to the situation in which we have a *marked* point process i.e. that the points which make up our generated data have qualitative or even quantitative measures associated with them. Møller et al. (1998) briefly define the dichotomously-marked LGCP in the spatial setting in Section 5, and further technical definitions related to the multivariate LGCP can be found in Møller and Waagepetersen (2004). However, there is much which remains unclear about the multivariate LGCP, including issues surrounding the specification of appropriate correlation and *cross-correlation* (i.e. between-mark dependence) functions, parameter estimation given observed data, and simulation. Furthermore, the extension to spatiotemporal data has not yet been considered.

This relative lack of appropriate methodology means a corresponding lack of applications of this potentially very powerful model in the literature, though one epidemiological example can be found in Liang et al. (2009). The authors consider the spatial variation in occurrences of colon and rectal cancer in Minnesota, USA, and use a rather *ad hoc* Bayesian framework to aid in their modelling. A similar approach to tackling more complicated problems, however, would be extremely difficult to implement. One example is the multivariate point pattern examined in Diggle, Zheng and Durr (2005) who analysed four different types of bovine tuberculosis in Cornwall, UK. The authors used kernel methods to investigate the spatial variation present by comparing, in a pairwise fashion, the estimated distributions of the marginal type-specific data sets. It would be extremely interesting to investigate a more sophisticated approach to this analysis, and the multivariate LGCP is a viable candidate.

A final interesting note can be made with reference to the kernel-smoothed density-ratio studied at length in Chapter 2. In dealing with case-control data, we may consider this as a bivariate point process with marks 'case' and 'control'. It stands to reason that the multivariate LGCP may provide an interesting alternative to kernel smoothing for estimation of the density-ratio, an issue this author has raised in preliminary correspondence with Dr. B. M. Taylor and Prof. P. J. Diggle at Lancaster University, UK. Suppose we have a bivariate

point process in W such that $\boldsymbol{X} = \{\boldsymbol{X}_f, \boldsymbol{X}_g\}$. In its simplest form, we would assume the mark-specific point processes $\boldsymbol{X}_f$ and $\boldsymbol{X}_g$ and their respective intensity functions f and g are independent of one another. We could then treat as separate LGCPs $f(\boldsymbol{x}) = \exp\{\mathcal{F}(\boldsymbol{x})\}$ and $g(\boldsymbol{x}) = \exp\{\mathcal{G}(\boldsymbol{x})\}$ given realised Gaussian fields $\mathcal{F}$ and $\mathcal{G}$, $\boldsymbol{x} \in W$. The subsequent (log-transformed) LGCP density-ratio ρ_L is given with

$$\begin{aligned}
\rho_L(\boldsymbol{x}) &= \log\left[\frac{f(\boldsymbol{x})/\int_W f(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}}{g(\boldsymbol{x})/\int_W g(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}}\right] \\
&= \log\left[\frac{\exp\{\mathcal{F}(\boldsymbol{x})\}\int_W g(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}}{\exp\{\mathcal{G}(\boldsymbol{x})\}\int_W f(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}}\right] \\
&= \mathcal{F}(\boldsymbol{x}) - \mathcal{G}(\boldsymbol{x}) + \log\left[\int_W \exp\{\mathcal{G}(\boldsymbol{x})\}\,\mathrm{d}\boldsymbol{x}\right] - \log\left[\int_W \exp\{\mathcal{F}(\boldsymbol{x})\}\,\mathrm{d}\boldsymbol{x}\right]. \qquad (3.59)
\end{aligned}$$

leaving the attractively simple difference between the two realised Gaussian fields as the fundamental quantity. However, problems lie in sensible estimation of the quantities given in the right-hand side of (3.59), an issue exacerbated by the complexities of conditional simulation of the LGCP given observed data. Furthermore, it may not be sensible to assume complete independence between $\boldsymbol{X}_f$ and $\boldsymbol{X}_g$ (and f and g themselves). All this points to the fact that a deeper investigation into the theory and methods associated with the multivariate LGCP is indeed warranted.

3.6.3 Nonseparable spatiotemporal correlation structures

A third future research direction is the relaxation of the assumption of a separable correlation structure made in Section 3.1.2. Recall for the spatiotemporal Gaussian field $\mathcal{Z}$, in addition to stationarity and isotropy we imposed the condition that

$$\mathrm{Corr}\left[\mathcal{Z}(\boldsymbol{x}_1, t_1), \mathcal{Z}(\boldsymbol{x}_2, t_2)\right] = r(||\boldsymbol{x}_1 - \boldsymbol{x}_2||; \phi_\psi)\tau(|t_2 - t_1|; \theta),$$

i.e. that the dependence could be factorised into purely spatial and temporal components. The result of this assumption is to greatly ease model-based inference, such as the derivation of the second-order properties in Section 3.1.2, and accessibility of the methods associated with spatiotemporal minimum contrast (Section 3.2.2) and conditional simulation (Section 3.3.2).

However, this comes at a cost of flexibility, and without restricting attention to the LGCP several authors have proposed general classes of *nonseparable* correlation functions for spatiotemporal modelling. One particularly interesting piece of work is the article by Rodrigues and Diggle (2010), who coin the terms *positive* and *negative nonseparability*. Let $\pi(u, \nu)$ represent the correlation between two points in space-time separated by spatial and temporal lags $u \geq 0$ and $\nu \geq 0$ respectively. A positively nonseparable correlation structure occurs when $\pi(u, \nu)/\pi(u, 0)\pi(0, \nu) > 1$ for all $u > 0$ and $\nu > 0$, negative nonseparability when $\pi(u, \nu)/\pi(u, 0)\pi(0, \nu) < 1$. We have separability when $\pi(u, \nu)/\pi(u, 0)\pi(0, \nu) = 1$.

This characterisation is important as it allows the authors to compare the flexibility of methods proposed in earlier works. For example, they identify that a class of functions de-

fined in Gneiting (2002) cannot accommodate negative nonseparability, whereas the models discussed in De Cesare et al. (2001) cannot accommodate positive nonseparability. This leads Rodrigues and Diggle (2010) to go on and define a class of covariance/correlation models that can accommodate both.

To date, this author is unaware of any efforts to implement the spatiotemporal LGCP with a nonseparable correlation structure. To successfully do so would require a number of issues to be addressed. For example, an investigation into how one might approach the problem of parameter estimation under such dependence models is paramount. Also, computational aspects would need to be considered – is use of the FFT still possible to aid in inversion of the covariance matrix? How would conditional simulation via the MALA be affected? Finally, one would need to attempt to answer the question of whether or not the additional complexity of including a nonseparable correlation structure in the spatiotemporal LGCP is in fact 'worth it' in the sense of providing noticeable benefits (for example, significantly increased accuracy and precision of predictions) in practice.

3.7 Summary and concluding remarks

In many applications dealing with two-dimensional point processes, it can be more intuitively sensible to view the underlying intensity function as stochastic (Cox process), rather than fixed (Poisson process). The corresponding 'doubly stochastic' models can be particularly well-suited to problems in spatial and spatiotemporal disease mapping. In this chapter we have examined one such model, the log-Gaussian Cox process, from an applied epidemiological perspective.

Through the LGCP, we are afforded a powerful framework which is capable of modelling a wide variety of point patterns; a detailed review of the spatial and spatiotemporal versions occurring in Section 3.1. In its most simple form, the LGCP dictates that the intensity function is viewed as an exponentiated Gaussian random field; heterogeneity in which is controlled by defined correlation structure; the appeal of the model lying primarily in its obvious flexibility. Firstly, the model is mathematically tractable (Møller et al., 1998); theoretical properties found with relative ease using moment generating functions of the lognormal distribution. Secondly, the researcher is able to choose from a wide selection of functional forms of the assumed correlation structure; assumptions of stationarity and isotropy (and in the spatiotemporal setting, separability) made for theoretical simplicity have few drawbacks for most practical intents and purposes. Thirdly, the LGCP can be fully defined using parameters gleaned from its first- and second-order properties. Lastly, the LGCP is easily extended and modified to cope with more complicated intensity specifications, such as the inclusion of deterministic components to model 'global' heterogeneity.

Despite its justified appeal, further investigation into issues related to the fitting and simulation of such models is warranted, and is ongoing. One of the most pressing problems is that of sensible estimation of the parameters defining the correlation structure. Section 3.2 discussed the only option that has demonstrated feasibility in practice – minimum contrast – whereby empirical and theoretical descriptors of second-order characteristics are matched to one another in order to obtain 'optimal' parameter values. Though drawing criticism

for its *ad hoc* nature, the few alternative suggestions such as numerical likelihood methods are unable to match its applied performance. Computational specifics in Section 3.3 began with a visual walkthrough of circulant embedding and torus-wrapping of the spatially discretised covariance matrix, intended to be viewed alongside more theoretical discussions on the procedure. Most importantly, it was noted that the subsequent use of the fast-Fourier transform is absolutely essential in practice, the high-dimensionality associated with implementation of the LGCP equating to massive computation and storage requirements. The discussion ended with a contained review of the Metropolis-adjusted Langevin algorithm for both spatial and spatiotemporal versions, the MCMC technique designed to provide superior convergence rates over a vanilla random walk when performing conditional simulation (prediction) of the Gaussian field. The many intricate details and occasionally complicated nature of these computational necessities highlight the importance of functional and accessible code, should we wish to see further applied analyses in epidemiology using the LGCP. Real-world examples were demonstrated in Section 3.4, illustrating the model specification, fitting, and conditional simulation stages and their performance in both the purely spatial and spatiotemporal settings.

With the noted concerns over sensible parameter estimation, it was deemed necessary to contribute to the relatively limited numerical literature focusing on performance of the minimum contrast procedure. Three novel simulation studies were detailed in Section 3.5, designed with epidemiological applications in mind. The first aimed to assess the impact of using either the K-function or the pair correlation function (g) for estimation of spatial process parameters. It was concluded that there was little discernable difference between the two methods over the artificial scenarios, though there seemed to be a mild tendency of the K-function to produce estimates of the spatial dependence scale parameter closer to the true values, with g providing better estimates of the field variance. These differences were very small, however; an inspection of the absolute relative errors (true parameter value versus estimated) showing extremely similar trends when plotted for the different scenarios for both K and g. The numerical investigations continued by turning our attention to the importance of the (fixed) bandwidth when a multiplicative spatial deterministic component as a kernel density estimate is included in the definition of the intensity function with a view to modelling the global trends, leaving the LGCP to capture local dependence. It was found that oversmoothing this component was beneficial in terms of finding minimum contrast parameter estimates closest to the true values describing the residual correlation structure. Along the same lines, it was important to note that many 'off-the-shelf' bandwidth selectors used in kernel density estimation would be likely to undersmooth the estimate in its role as a purely global, and not a *total* (global-plus-local), descriptor. Finally, the spatiotemporal LGCP was scrutinised through the design of phased simulation study. As we had noted, some discrepancy between the true and estimated parameter values is to be expected from minimum contrast methods – but it was unclear if the magnitude of these discrepancies actually matters when we aim to produce exceedance probability surfaces from conditional predictions of the intensity function. Motivated by real-time disease surveillance, this was addressed by quantifying the agreement between the true and predicted exceedance surfaces, each straightforwardly expressed as a binary matrix. Phase 1 of this study performed min-

imum contrast estimation of the spatiotemporal correlation parameters for the pre-defined scenarios; our findings reflected what we had come to expect based on results of the previous two simulation studies. Phase 2 performed a large collection of MALA runs providing the predicted exceedance surfaces, and it was found, using the Jaccard similarity coefficient, that there encouragingly did not appear to be any disadvantage in using the minimum contrast-estimated parameters as opposed to the true values when attempting to detect anomalous sub-regions.

Much of this chapter has reviewed existing methodology, the aim being to provide a comprehensive, coherent synthesis of the theoretical and practical aspects of the LGCP most relevant for epidemiological applications. It is therefore hoped that the preceding content will serve to motivate further epidemiological analyses using this capable modelling framework, particularly for spatiotemporal data. An equally important outcome is the collaboratively developed R package, capable of the discussed operations, which has recently been made available in the public domain and is explored in the following chapter.

Chapter 4

Software Releases

Writing self-contained software is a powerful way to distribute statistical innovations and allow users of the methodology (either novices or veterans of the methods in question) easier access to these advancements for their own applications. Producing programs for general consumption presents its own unique challenges. Careful structuring and standardisation of functions and other quantities such as data objects is required to streamline the package. Computational efficiency, which might be sacrificed in personal code to a certain extent for improved readability and editing, must now take highest priority. Comprehensive error handling, from basic argument checks to dealing with the specifics of internal computations, requires implementation. Thorough documentation is essential in order to best guide correct usage and interpretation of results.

In this chapter two generalised software packages are introduced. Written in and for the statistical programming environment R (R Development Core Team, 2011), these packages make available to fellow researchers material and advancements in Chapters 2 and 3. In addition to the manuscripts directly describing them, the development of these programs has spawned several interesting collaborations and additional research directions, some of which are covered in Chapter 5. The package discussed in Section 4.2 is developed in collaboration with researchers at Lancaster University.

4.1 R package sparr

There already exist a number of packages that can perform bivariate kernel density/intensity estimation in R; see for example **spatstat** (Baddeley and Turner, 2005), **KernSmooth** (Wand and Ripley, 2011), **ks** (Duong, 2007), and **spatialkernel** (Zheng and Diggle, 2011). Following release of **sparr**, the latest versions of **spatstat** and **spatialkernel** have also included functions to handle the computation of fixed-bandwidth density-ratios. Neither of these, however, provide the explicit capability to estimate the edge-corrected adaptive relative risk functions discussed in detail in Chapter 2. Package **sparr** also possesses the unique functionality needed for p-value surface computation, via both Monte-Carlo (MC) and the new asymptotic (ASY) methods.

The R package **sparr** (spatial relative risk), freely available on the Comprehensive R

Archive Network (CRAN) at `http://cran.r-project.org/web/packages/sparr/index.html`, was developed in an effort to aid in the growing demand for the aforementioned functionality. This package is specifically detailed in the *Journal of Statistical Software* article Davies, Hazelton and Marshall (2011).

4.1.1 Aims and scope

The objective of **sparr** is to allow researchers to perform fixed and adaptive kernel estimation of spatial relative risk. The package is designed with applied users (e.g. epidemiologists) in mind. Below we describe briefly the main components of **sparr**. The reader is directed to the package documentation (Appendix A), for specific details, and to refer back to Chapter 2 for theoretical specifics.

- `PBC` (data set): previously unavailable in R, this object contains the 761 cases along with 3020 controls for the primary biliary cirrhosis data analysed in Section 2.6.2. Also provided is the 115-vertex polygon describing the study region (Peter Diggle is thanked for permission to publish these data in the package).

- `bivariate.density`: A flexible function to perform kernel density/intensity estimation on bounded regions in $\mathbb{R}^2$. Providing edge-corrected fixed and adaptive smoothing regimens, the user can also control a number of important features such as evaluation grid resolution, additional coordinates at which to evaluate the estimate, and specification of duplicate observations. The data themselves can be passed to this function in a variety of ways. Detailed information is contained in the returned '`bivden`' object, allowing careful inspection and follow-up analysis of the estimate.

- `risk`: This function performs the quotient operation based on two kernel density estimates as required for a relative risk surface. The user may opt either to supply two pre-calculated `bivden` objects (in which case the operations performed by `risk` are trivial), or simply pass in the raw data directly (in which case the kernel estimation is performed internally). The former option allows more control over the estimation of the individual densities, whereas the latter sacrifices this control (i.e. imposes a particular set of default conditions) in favour of ease of access. This function returns a detailed object of class '`rrs`'. The user can fully retrieve both '`bivden`' objects used as case and control density estimates, regardless of the initial choices in running `risk`.

- `tolerance`: Compute the asymptotically derived (ASY) or Monte-Carlo (MC) p-value surfaces for a given risk function, potentially leading to tolerance contours. Though provided for the sake of completeness, the MC approach is significantly more computationally expensive than ASY. Indeed, it is not recommended that this method be used for adaptive risk functions. Large sample sizes, coupled with fine grid resolutions, will also impact on computational expense. To curb this problem, the option is available to reduce the resolution of the evaluation grid used for the p-value surface to coarser levels if necessary. The returned component is a matrix of p-values corresponding to the elected test (i.e. lower-tailed, upper-tailed, or two-tailed).

- `plot.bivden` and `plot.rrs`: These functions allow the plotting of bivariate density and risk function estimates i.e. `bivden` and `rrs` objects. The exploratory nature of relative risk functions mean visualisation of the surface(s) is generally always an objective in a given application. Specific methods for viewing two-dimensional kernel estimates (let alone risk surfaces) are not a priority in other contributions, where the user is expected to make use of the generic options available to construct their images. These functions remove the need for this careful *ad hoc* tailoring, allowing flexible visualisation of the aforementioned objects in a number of different ways. Some of these are showcased in the following section. However, through the detailed return information of the estimated densities and surfaces, the user retains the choice to manually plot their objects in any desired fashion.

4.1.2 Example walkthrough and code snippets

The functionality of **sparr**, as with many software packages, is best illustrated by example. In this section we re-create the example in Section 2.6.2 by estimating and displaying an adaptive relative risk function for the PBC data, along with asymptotically estimated tolerance contours.

Data

Initially, it is sensible to simply view the raw data. The following code loads the data set into the current workspace, re-formats the data structures, and yields Figure 4.1.

```
R> library(sparr)
R> data(PBC)
R> pbc.cases <- split(PBC)[[1]]
R> pbc.controls <- split(PBC)[[2]]
R> plot(pbc.controls,pch=3,main="")
R> box(bty="L");axis(1);axis(2);title(ylab="Northings",xlab="Eastings")
R> points(pbc.cases,col=2)
```

Inspection reveals, as we would expect, similar case-control dispersions based on the natural inhomogeneity in the population. Comparing the distribution of cases versus the distribution of controls, and attempting to identify any apparent differences, is our key objective.

Bandwidths

Firstly, we must sensibly calculate our isotropic bandwidth(s) for the required kernel smoothing. Recall the decision in Section 2.3.1 to select a common global bandwidth and separate pilot bandwidths for the case and control density estimates when implementing variable soothing. Furthermore, recall the poor performance of the single-density leave-one-out least-squares cross-validation bandwidth selector when compared to Terrell's (1990) oversmoothing principle in the numerical experiments in Section 2.5. With this in mind, observe the following code:

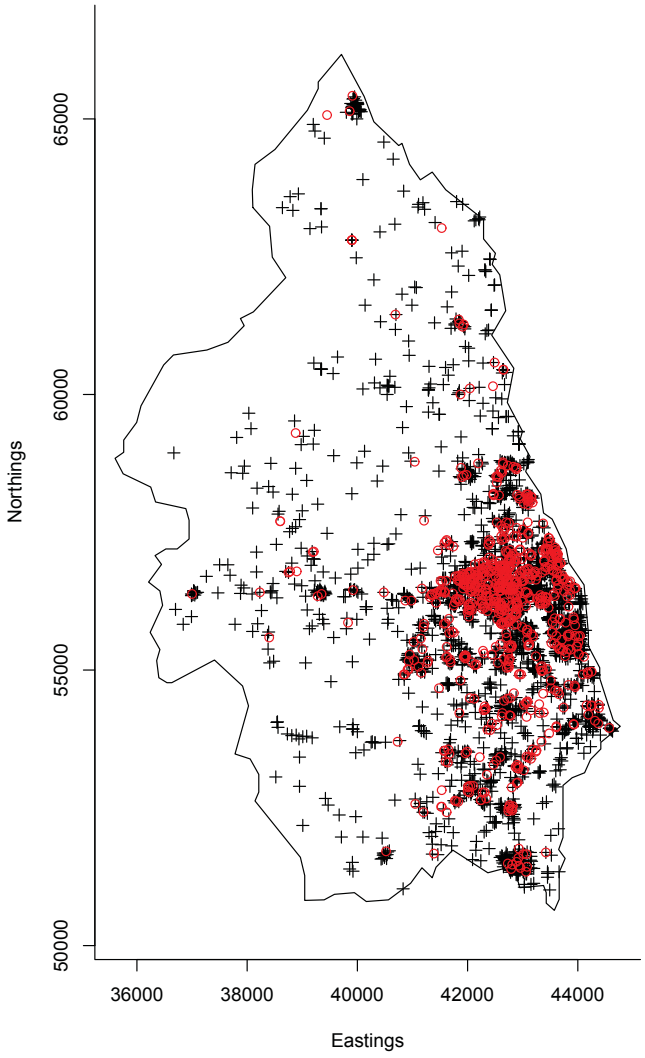


Figure 4.1: Cases (red '○') and controls (black '+') for the PBC data.

```
R> OS.f <- OS(pbc.cases)
R> OS.f

[1] 317.4248

R> OS.g <- OS(pbc.controls)
R> OS.g

[1] 330.2433

R> OS.pool <- OS(PBC, nstar=sqrt(pbc.cases$n*pbc.controls$n))
R> OS.pool

[1] 349.8445
```

These commands calculate the single-density oversmoothing bandwidths for the case, control and pooled data via the **sparr** function `OS`, which uses the mean of the two axis-specific interquartile ranges scaled by 1/1.34 to represent an *ad hoc* robust scalar measure of standard deviation as required for the computation of the oversmoothing bandwidths. The argument `nstar` for `OS.pool` corresponds to our discussion in Section 2.3.1, where we set the 'effective' sample size to the geometric mean of the separate case and control sample sizes to prevent too much influence by the larger of the two.

Density estimation

Estimation of the densities may now take place. We execute

```
R> pbc.pooled.density <- bivariate.density(PBC,pilotH=0.5*OS.pool,
R> + globalH=OS.pool,adaptive=T,res=128,WIN=PBC$window,comment=F)
R> pbc.case.density <- bivariate.density(pbc.cases,pilotH=0.5*OS.f,
R> + globalH=OS.pool,adaptive=T,res=128,WIN=PBC$window,
R> + gamma=pbc.pooled.density$gamma,comment=F)
R> pbc.control.density <- bivariate.density(pbc.controls,pilotH=0.5*OS.g,
R> + globalH=OS.pool,adaptive=T,res=128,WIN=PBC$window,
R> + gamma=pbc.pooled.density$gamma,comment=F)
```

which provides us with the required objects. The object `pbc.pooled.density` is simply the density estimate of the pooled data set, i.e. ignoring the case/control marking of the observations; this is needed for computation of the subsequent p-value surfaces. Global bandwidths are ordered to be equal to the `OS` values computed from the pooled data. Note that the pilot bandwidths are set to half of the `OS`-computed values; the case and control estimates utilise the separately computed oversmoothing bandwidths for the pilots in alignment with earlier comments (see Section 2.6.2). We also employ a common geometric mean scaling term computed from the pooled data, γ_ω, for the case/control estimates as the argument `gamma` (the reader is referred back to Section 2.2.2). Finally note that the grid resolution of our kernel estimates is set using the argument `res=128`, this indicates we wish to evaluate the densities on a 128×128 grid, automatically discarding any cells lying outside the irregular

polygon which defines our study region, specified by `WIN`. Function progress is commented upon during execution, we have disabled this through use of the argument `comment=F`. The option `comment=T` is particularly useful when we are dealing with large data sets and/or very fine grid resolutions.

The objects returned by `bivariate.density` are given their own class, `bivden`. We can inspect basic information concerning our kernel estimates using class-specific versions of the generic R `print` and `summary` commands. For example, `print`ing a `bivden` object yields

```
R> pbc.case.density

Bivariate kernel density/intensity estimate

Adaptive isotropic smoothing with (pilot) h = 158.7124
    global h = 349.8445 unit(s)
No. of observations: 761
```

Risk surface and tolerance contour computation

The relative risk function may now be estimated. This is achieved through use of the function `risk` which by default computes the log-transformed density ratio as advocated by e.g. Kelsall and Diggle (1995*a*). We run this command using our pre-computed case and control density estimate objects `pbc.case.density` and `pbc.control.density`, again suppressing output with `comment=F`. By default, `risk` also produces a pixel image plot of the resulting risk surface; this has been disabled for the moment (the various plotting options will be explored once we have computed the asymptotic tolerance contours).

```
R> pbc.risk <- risk(f=pbc.case.density,g=pbc.control.density,
R> + comment=F,plotit=F)
R> summary(pbc.risk)

Log-Relative risk function.

Estimated log-risk range -4.485437 to 1.207464.
9095 grid cells out of 16384 fall inside study region.

--Numerator (case) density--
Bivariate kernel density/intensity estimate

Adaptive isotropic smoothing with (pilot) h = 158.7124
    global h = 349.8445 unit(s)
No. of observations: 761
Evaluated over 128 by 128 rectangular grid.
Defined study region is a polygon with 115 vertices.

Estimated density range 4.509172e-12 to 5.134165e-07.
```

```
9095 grid cells out of 16384 fall inside study region.

--Denominator (control) density--
Bivariate kernel density/intensity estimate

Adaptive isotropic smoothing with (pilot) h = 165.1216
    global h = 349.8445 unit(s)
No. of observations: 3020
Evaluated over 128 by 128 rectangular grid.
Defined study region is a polygon with 115 vertices.

Estimated density range 3.544127e-10 to 3.027201e-07.
9095 grid cells out of 16384 fall inside study region.
```

The `risk` function produces an object of class `rrs` for which `print` and `summary` commands are also available. Use of `summary.rrs`, which subsequently calls `summary.bivden` on both objects supplied as `f` and `g`, is shown in the above output. A numerical summary of `pbc.risk` shows a maximum density-ratio value of $\exp(1.207) = 3.343$. Can we consider risk magnitudes at this level 'unusually extreme' for our current data set?

Natural sampling variability will clearly produce benign fluctuations in the resulting spatial risk surface. It is of interest to highlight those sub-regions, if indeed there exist any, that are sufficiently extreme such that we would label them 'statistically significant' in the typical frequentist interpretation of the phrase. In spatial disease risk mapping, the search for significant positive fluctuations is the more common goal; embellishment of any 'hotspots' in the risk of disease contraction is clearly of interest to epidemiologists and public health officials.

Upon definition of appropriate hypotheses, we discussed in Section 2.4 the derivation of a p-value surface corresponding to a kernel-smoothed relative risk function. Denoting our study region with $\mathcal{R}$, suppose we denote our (log) risk function of interest as $\rho(\boldsymbol{x})$, $\boldsymbol{x} \in \mathcal{R}$. As also covered in these earlier discussions, hotspot detection would correspond to the hypotheses $\mathrm{H}_0 : \rho(\boldsymbol{x}) = 0$, $\mathrm{H}_\mathrm{A} : \rho(\boldsymbol{x}) > 0$; diminished risk would imply testing for $\mathrm{H}_\mathrm{A} : \rho(\boldsymbol{x}) < 0$; a double-sided test dictates $\mathrm{H}_\mathrm{A} : \rho(\boldsymbol{x}) \neq 0$. Once we have specified hypotheses and estimated the corresponding p-value surface, any significant sub-regions are highlighted on a plot of the estimated risk function via tolerance contours at pre-defined significance levels.

Historically, p-value surfaces were computed via Monte-Carlo methods, namely random permutations of case/control marks and repeated evaluation of the kernel-smoothed risk function against which the actual estimated risk function values for each spatial coordinate would be compared. In Chapter 2, we proposed a theoretical alternative, based on the asymptotic normality of kernel density estimates (e.g. Parzen, 1962) coupled with asymptotic expressions for the variances of the fixed and adaptive kernel-estimated density-ratios. Numerically stable and computationally cheaper than MC methods, the asymptotic approach to tolerance contour calculation offers competitive performance, particularly for large data sets and when employing adaptive smoothing.

The p-value surfaces are computed in **sparr** with the function `tolerance`. To illustrate the difference in computational expense between the two approaches, 'MC' and 'ASY', we evaluate both, using 100 iterations for the former.

```
R> pbc.tol.mc <- tolerance(pbc.risk,pbc.pooled.density,method="MC",ITER=100)

[1] "Tue Sep 20 22:01:47 2011"

Monte-Carlo iteration no.
10  20  30  40  50  60  70  80  90

[1] "Wed Sep 21 01:54:21 2011"

R> pbc.tol.asy <- tolerance(pbc.risk,pbc.pooled.density,method="ASY")

[1] "Wed Sep 21 09:39:12 2011"

--Adaptive-bandwidth asymptotics--
calculating integrals K2...
--f--
--g--
calculating integrals L2...
--f--
--g--

[1] "Wed Sep 21 09:40:16 2011"
```

Supply of the pre-calculated `rrs` object, as well as the relevant pooled kernel density estimate, is required. Function progress is printed to the console on-the-fly; as for earlier functions this can be disabled by setting `comment=F`. Note that for only 100 iterations, the MC p-value surface required almost four hours to complete within the current workspace for our adpative PBC example. By comparison, the asymptotic method took a little over a minute. The comments referring to `integrals K2` and `L2` correspond to estimation of the square-bracketed terms in equation (2.9) for the case and control densities `f` and `g` respectively. Returned is a named list including the component `P`, a matrix of equal dimensions to the original rectangular evaluation grid, containing the p-values of interest over $\mathcal{R}$.

Visualisation

The final step is to produce plots of our estimated risk function and any statistically significant sub-regions therein through the use of tolerance contours. To this end, **sparr** provides several flexible forms of risk surface visualisation. Arguably, one of the most intuitive is a pixel image or 'heatplot'. The following code produces Figure 4.2.

```
R> par(mfrow=c(1,2))
R> plot(pbc.risk,display="heat",col=heat.colors(10)[10:1],
```

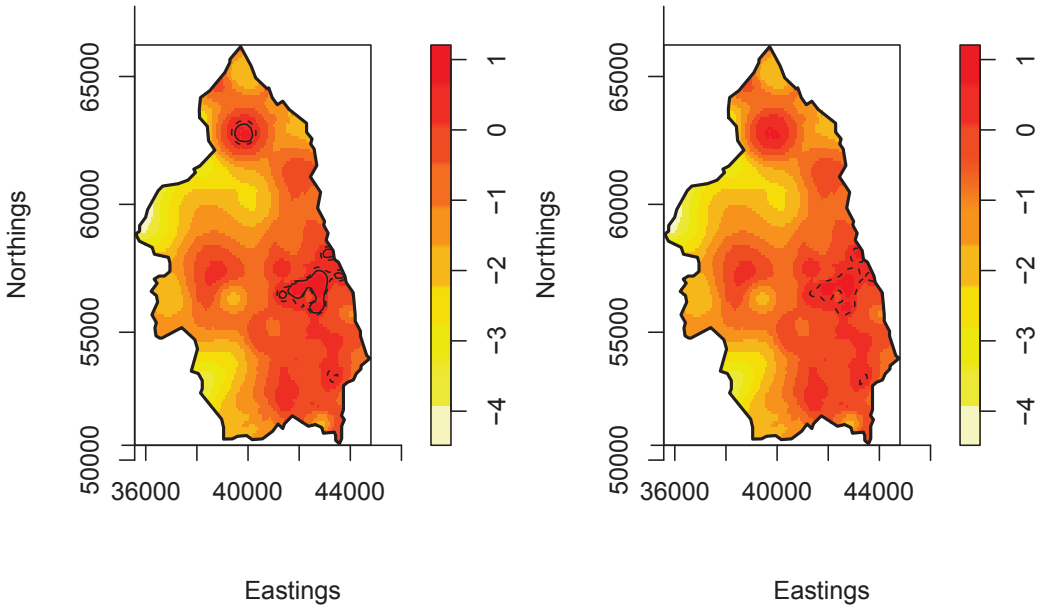


Figure 4.2: Heatplots of the adaptive log-relative risk surfaces for the PBC data generated using **sparr**, displaying asymptotic and Monte-Carlo tolerance contours (left and right respectively), superimposed at significance levels of 0.01 (solid line) and 0.05 (dashed).

```
R> + tolerance.matrix=pbc.tol.asy$P,
R> + tol.opt=list(levels=c(0.01,0.05),lty=c(1,2)),
R> + main="",xlab="Eastings",ylab="Northings")
R> plot(pbc.risk,display="heat",col=heat.colors(10)[10:1],
R> + tolerance.matrix=pbc.tol.mc$P,
R> + tol.opt=list(levels=c(0.01,0.05),lty=c(1,2)),
R> + main="",xlab="Eastings",ylab="Northings")
```

Inspection of the plots shows some distinctions between the theoretically and empirically computed contours. Note that the `plot.rrs` function allows the user to directly include the calculated tolerance contours through the use of the argument `tolerance.matrix`. The appearance of the contours can be minimally controlled via a named list argument `tol.opt`. Should the user wish more control over their inclusion, they may set `tolerance.matrix` to `NULL`, in which case `tol.opt` is ignored, and subsequently add to a pre-plotted risk surface the desired contours using the standard R command `contour` (package **graphics**) with argument `add=T`.

In this example, the asymptotic method has appeared to amplify the significance of individual peaks somewhat, flagging a sub-region on the eastern border as well as a smaller northern area at the 1% level. The Monte-Carlo contours agree with the apparent significance of the eastern border sub-region, albeit only at the 5% level, but do not indicate much

interest in any other areas (although we have only made use of 100 iterations for this method to spare excessive computation time). In any case, it should be stressed that superimposition upon a risk surface image of tolerance contours of this sort are intended to provide a general idea of possible sub-regions of interest as an exploratory tool only, as opposed to definitive proof of anomalous behaviour. The agreement of the significant eastern-border area between the two approaches would lead us in this case to conclude that this is an area that most warrants further investigation by the relevant health officials.

In addition to the standard heatplot, the user can view the risk function as a contour or perspective plot, setting `display="contour"` or `display="persp"`. One particularly unique feature to **sparr** is the fact that we may also produce an interactive three-dimensional perspective plot based on the powerful features of the **rgl** package. Allowing full rotation and zooming via left and right mouse button holds respectively, it produces images from which it can be easier to assess the relative magnitudes of peaks and troughs in a particular surface compared to e.g. from a 2D heat plot.

The following code opens an interactive graphics device, plots the coloured 3D surface, and adds the asymptotically derived tolerance contours as in the left-hand panel of Figure 4.2. Figure 4.3 shows two screenshots of the interactive plot.

```
R> asp <- diff(PBC$window$yrange)/diff(PBC$window$xrange)
R> plot(pbc.risk,display="3d",aspect=c(1,asp,1),col=heat.colors(10)[10:1],
R> + tolerance.matrix=pbc.tol.asy$P,
R> + tol.opt=list(levels=c(0.01,0.05),raise=0.025,col="blue"))

approximating 3D boundary...
approximating 3D tolerance contours...
```

To preserve the correct x:y axis aspect ratio, we must specify our desired relative scale via the argument `aspect=c(x,y,z)`. This is achieved by firstly finding the correct ratio which we store as the value `asp`. Also noteworthy: the `tol.opt` option `raise` is only relevant to `display="3d"` plots when `!is.null(tolerance.matrix)`. This is a scalar constant which artificially translates the superimposed 3D contours `raise` units along the z-axis, used to prevent segments of the included contours falling 'below' the plotted surface (an artefact of the approximations made during the 3D plotting).

Much the same plotting choices exist for standalone density estimates (i.e. objects of class `bivden` as opposed to `rrs` *per se*), and greater flexibility than can be fully covered here is of course available. In fact, in addition to the arguments for `plot.bivden` and `plot.rrs` we have encountered and discussed thus far, there is an additional `...` field reserved for arbitrary arguments depending on which `display` type we have elected. In the accompanying package documentation, the user is directed to the R help files for functions in external packages for proper use of `...` in fine-tuning density and risk function plots. For the four `display`s `"heat"`, `"contour"`, `"persp"` and `"3d"`, consultation should be made to `?plot.im`, `?persp`, `?contour` and `?persp3d` respectively.

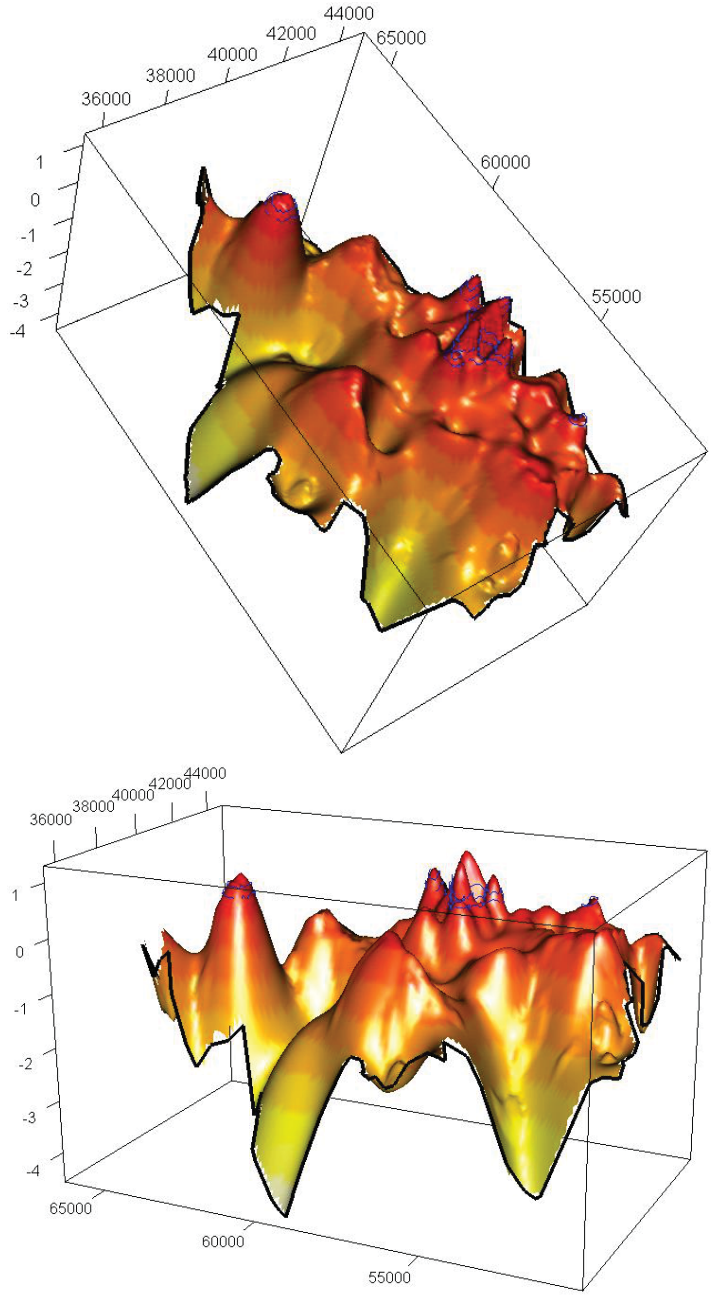


Figure 4.3: Two snapshots of the interactive 3D adaptive bandwidth PBC risk function, with superimposed asymptotic tolerance contours.

4.1.3 Commentary regarding fast and efficient use

The operations **sparr** can perform would be computationally expensive if one were to evaluate objects such as density estimates and p-value surfaces directly. This is especially true when using an adaptive bandwidth. For this reason, various coding techniques and some approximations have replaced what would otherwise be prohibitively slow code. Encountered in Section 3.3.1 for the LGCP, the fast-Fourier transform (FFT) can also be applied to kernel density estimation. A good introductory discussion on how this is performed can be found in Appendix D of Wand and Jones (1995). This is therefore a key tool in accelerating execution time for the relevant operations in **sparr** and is accessed using the fixed-bandwidth kernel intensity estimate function `density.ppp` from one of **sparr**'s dependencies, **spatstat**. Though a trivial 'piggyback' translation for `bivariate.density` to make in the fixed bandwidth case, more thought is required to utilise the advantageous speed of the **spatstat** code for adaptive estimation, where we require bandwidths for all possible grid coordinates to edge-correct (cf. Marshall and Hazelton, 2010). In this situation, `bivariate.density` calculates these bandwidths and finds the unique values of the 100 integer quantiles thereof. Each of these unique values is then treated as a fixed bandwidth, and the corresponding fixed-bandwidth edge-correction factors are extracted from the FFT-calculated objects. A given coordinate within the study region is then matched to the nearest appropriate integer quantile value by its bandwidth, and the edge-correction factor corresponding to that bandwidth and coordinate is assigned. This approximation is significantly faster than directly evaluating factors at each distinct coordinate and associated bandwidth, at a negligible cost of accuracy. Nevertheless, the user can turn off the option to use this fast-Fourier transform assistance by setting the `bivariate.density` argument `use.ppp.methods=F`.

Other approximations of note are made for the integral components of the asymptotic p-value surfaces in `tolerance`, again using the fast-Fourier methods of **spatstat** as well as approaches constituting only single evaluations of a function subsequently translated to all possible grid coordinates. See the documentation pages for `bivariate.density` and `tolerance` for further information.

In addition, it is worth noting that a reduction in computation time may be had by reducing the resolution of the grid over which we estimate the density or p-value surface. The default resolution is to use a 50×50 grid, which is considered to be a suitable minimum resolution that provides serviceable estimates and images, while keeping computation costs at a relatively low level. The user may find that a modest increase in resolution, as in the preceding PBC example, may be warranted for aesthetic reasons. In the interests of flexibility it is also possible, via the optional argument `reduce` for the function `tolerance`, to evaluate a p-value surface over a coarser grid than the corresponding risk function density estimates. This allows tolerance contours to be computed relatively quickly, and then superimposed over a more aesthetically pleasing image of the risk surface.

Also mentioned is the fact that the complexity of the polygon describing the study region can, in extreme cases, result in a prohibitive computational cost for density and p-value surface estimation. A polygon with many thousands of vertices (e.g. one obtained from a detailed geographical map) will cause delays, particularly when the complex region needs

to be checked by the function multiple times for calculation of edge-correction factors. A simple workaround is to reduce order of the polygon by, for example, retaining only every kth vertex. Analysis can then be performed at a generally negligible cost of accuracy (depending on the magnitude of k) using the reduced polygon. For imaging, the researchers can easily display the efficiently computed results using the original geographical map.

It is worth clarifying the aspects of kernel density estimation which directly impact computational expense. Producing adaptive over fixed estimates will increase this cost. The size of the data set, opting to edge-correct, as well as grid resolution, also impact on execution time. Though **sparr** has sought to minimise this cost as much as possible, it remains up to the user to find an acceptable balance between the aforementioned issues and computing time for their projects.

4.1.4 Limitations and future extensions

Package **sparr** is a freely available collection of R functions which include the most recent advancements in the applied estimation of relative risk functions based on planar kernel smoothing. First becoming available the public domain in late 2010, the popularity of the methodology and an accessible software package is clearly apparent. Since initial release, this author has corresponded with researchers from a wide variety of backgrounds, from epidemiologists to archaeologists, seeking advice on sensible usage of the approaches and interpretation of returned results. Some of these communications have resulted in collaborative projects and are summarised in Chapter 5.

Nevertheless, there remains scope for further extensions to **sparr** (at the time of writing, version 0.3-0). The issue of bandwidth selection is a difficult problem with respect to risk functions, and the somewhat *ad hoc* approach taken by most authors (present included) in tackling practical applications of the methodology has shifted the focus of **sparr** to the actual analysis and interpretation of a risk function. Though, as mentioned, the package does provide the functionality of some bandwidth selectors for density estimation (e.g. `OS`), **sparr** aims to emphasise the importance of user-consideration of the bandwidth choice(s) for risk functions on a case-by-case basis. This is not to say that the available suite of bandwidth selectors cannot be expanded should user demand exist. Also, as almost all practical applications of the methodology use the Gaussian kernel function, this is the only kernel currently supported for the density estimation in **sparr**. The infinite tails of this kernel are useful in areas of a given region with sparse observations, where a bounded kernel can result in unbounded risk function estimates if we experience zero case and/or control density levels within the study region. Future versions of this package will, however, endeavour to expand the available choices for kernel functions.

Finally, we acknowledge that the advancements in this package would not be possible without several other important contributions to CRAN; these are reflected as **sparr**'s package dependencies. Package **rgl** by Adler and Murdoch (2009) provides the ability to plot in 3D and interact with the device. The comprehensive spatial point pattern package **spatstat** (Baddeley and Turner, 2005) provides functions to enable efficient region handling, as well as the aforementioned FFT implementation. Lastly, package **MASS** (Venables and

Ripley, 2002) provides utility support for internal functions.

4.2 R package lgcp

The synthesis of the theory and applications of the log-Gaussian Cox process in Chapter 3 for planar point patterns required a substantial volume of code. These efforts constituted the first functional, generally accessible platform for spatiotemporal modelling and FFT MALA prediction with the LGCP in the R language, piquing the interest of the researchers Benjamin M. Taylor, Barry S. Rowlingson and Peter J. Diggle at the Division of Medicine, Lancaster University, during this author's visit to various universities in Europe; January 2011. Arguably among the world leaders in applied epidemiological spatial and spatiotemporal point process modelling, Markov-Chain Monte-Carlo expert Taylor with Rowlingson and Diggle were instrumental in the acceleration and final quality of the package construction.

The R package **lgcp** is motivated primarily by problems in the real-time disease surveillance as in Diggle, Rowlingson and Su (2005) and Sections 3.4.2 and 3.5.3 of this work. As such, it is the spatiotemporal LGCP where the focus of **lgcp** lies; a corresponding paper Taylor, Davies, Rowlingson and Diggle (2012) is at the time of writing in press with the *Journal of Statistical Software*. The development of the package in terms of re-coding and refinement of this author's raw code, as well as authorship of additionally required functions, package 'beautification' and paper fundamentals were primarily the responsibility of Taylor, hence his lead position in the article author list itself. Davies appears second owing to the original coding efforts forming the core functionality of the package and ongoing testing and developmental suggestions. Specifically, this author's coding contributions featuring in **lgcp** focused on the torus-wrapping FFT methodology, MALA simulation, and computation of exceedance probabilities. Both Taylor and Rowlingson were responsible for the interactive features of **lgcp**, data-to-disk dumping functions and post-processing options. Diggle, behind a considerable amount of the relevant literature, provided input (both code and direction), invaluable to the quality of the released product. Version 0.9-0 is available on CRAN, at `http://cran.r-project.org/web/packages/lgcp/index.html`.

4.2.1 Aims and scope

The key objective of **lgcp** is to allow users to model spatiotemporal planar point patterns using the stationary, isotropic LGCP, and conditionally simulate from (i.e. predict using) the fits. Considerably larger than **sparr**, this package provides unique tools for deterministic component and stochastic process parameter estimation, as well as prediction using the MALA and exceedance probability calculation. Functionality for data-to-disk dumping and retrieval is included for large files resulting from MALA runs. Tools for post-processing and visualisation of results are also available. Below, the most interesting features of **lgcp** are summarised.

- `stppp`: the generic function for converting R data structures (e.g. `ppp` or `list`) into an object of a class of the same name (space-time planar point pattern). All functions in **lgcp** requiring observed data input make use of the `stppp` class. The object contains

information on the spatial locations, the spatial observation window, the times at which each observation occurred, and the temporal interval limits.

- `spatialAtRisk` and `temporalAtRisk`: generic conversion functions, and resulting object class names, for R objects corresponding to the deterministic 'global' space (ζ_ψ) and time (η) trends respectively. The reader is referred back to Section 3.1.3 for theoretical details. These items are able to be specified in a number of different ways, including, for example, a `spatialAtRisk` object from a **sparr** `bivden` object.

- `lambdaEst`: function to interactively estimate the deterministic global spatial component via fixed-bandwidth estimation. As we shall see in the following section, this opens a specialised **rpanel** device, allowing the user to vary the bandwidth and immediately visualise the impact on the resulting density estimate.

- `muEst`: function to estimate the deterministic global temporal trend using a LOWESS smoother.

- `ginhomAverage` and `KinhomAverage`: functions to compute the time-averaged inhomogeneous pair correlation and K-functions respectively, defined in Section 3.1.3 as equations (3.20) and (3.21). These are required for e.g. minimum contrast estimation of the spatial correlation parameters.

- `spatialparsEst` and `thetaEst`: interactive, visual **rpanel** estimation of spatial and temporal correlation parameters respectively. Devices are opened containing plots of the appropriate nonparametric and theoretical curves; use of a slider handle displays the impact of varying the parameters affecting the theoretical curve. The object is to match, as closely as possible, the two lines. This will also be demonstrated in the following section.

- `lgcpPredict`: execution of a FFT MALA to obtain samples and summary statistics from the predictive distribution of interest. The user is afforded control over almost all aspects of the MCMC chain. This function returns an object with class of the same name, for which there exist a suite of post-processing functions.

- `lgcpargs` and `mcmcargs`: functions governing control over `lgcpPredict`. Correlation parameters are specified by the former, chain characteristics (e.g. burn-in, thinning) by the latter.

- `setoutput`: a function controlling result computation and storage. For example, if dumping to disk where, and what, do we wish to return?

- `meanfield`, `varfield`, `rr`, `serr` and `intens`: examples of post-processing functions on the returned `lgcpPredict` object. These return the cell-wise Monte-Carlo mean and variance of the retained Gaussian fields, the cell-wise mean and standard error of the exponentiated Gaussian fields, and the estimated mean cell-wise Poisson intensity respectively.

- `print.lgcpPredict` and `plot.lgcpPredict`: support for the generic R functions `print` and `plot` for objects of class `lgcpPredict`. The former displays detailed information on the run, the latter plots the Monte-Carlo mean of the exponentiated spatial Gaussian fields for each stored timestamp.

- `plotExceed`: a specialised plotting function for exceedance probability surfaces. This requires specific instruction to compute the exceedance probabilities at the commencement of `lgcpPredict`, and will be illustrated in our example walkthrough.

Note that the first version of the full package documentation (written by B.M. Taylor) can be obtained at `http://cran.r-project.org/web/packages/lgcp/lgcp.pdf`. There exist a host of smaller functions in **lgcp**; for simple reference purposes we therefore only provide the help files for the functions featuring the most in the following section in Appendix B.

4.2.2 Example walkthrough and code snippets

We will now demonstrate the functionality of **lgcp**. To do so, the AEGISS data from the example in Section 3.4.2 will be used, with the aim of obtaining a spatial exceedance probability surface for the 'current' (i.e. most recent) time point as would be the objective in real-time surveillance. Rather than replicate the precise steps and decisions made earlier, to best showcase **lgcp** we will make full use of the package functions for estimation of the deterministic components ζ_ψ and η, as well as the correlation parameters σ^2_ψ, ϕ_ψ and θ.

Data specification

In R, call `library(lgcp)`. Suppose we have present the `data.frame` object `AEGISS` and `W`, a polygonal object of class `owin` (**spatstat**) describing the spatial study region. The columns of `AEGISS` named `x`, `y` and `t` provide the spatial x-y coordinates of each recorded case of gastroenteritis, as well as the time (in this example each integer represents a 'day' starting on January 1, 2001) at which it was recorded, respectively. We have

```
R> nrow(AEGISS)

[1] 10554

R> range(AEGISS$t)

[1]    1 1095
```

indicating a total of 10554 recorded cases, over the time range [1, 1095].

To view the spatial and temporal margins, thereby resulting in Figure 4.4, call:

```
R> par(mfrow=c(1,2))
R> AEGISS.ppp <- ppp(x=AEGISS$x,y=AEGISS$y,window=W)
R> plot(AEGISS.ppp,main="",xlim=c(0,1),ylim=c(0,1),pch='.')
R> box(bty="L");axis(1,at=pretty(c(0,1)));axis(2,at=pretty(c(0,1)))
R> title(xlab="Eastings",ylab="Northings")
R> plot(unique(AEGISS$t),as.numeric(table(AEGISS$t)),
```

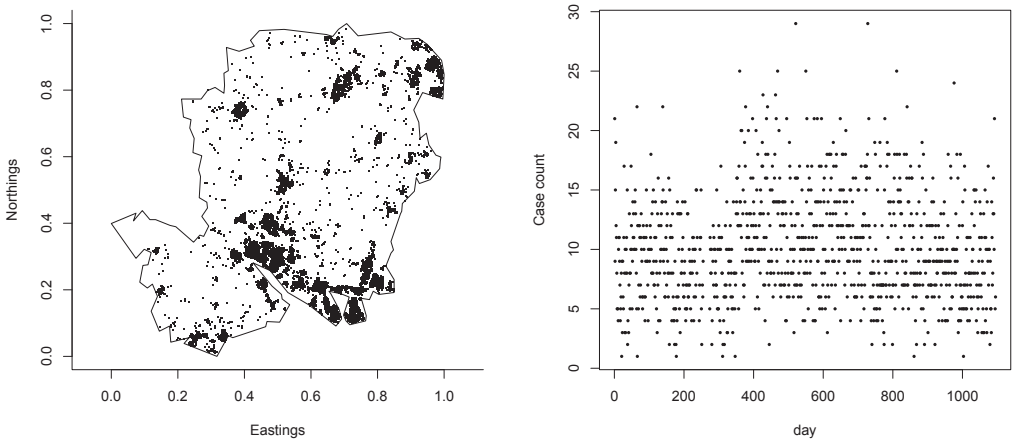


Figure 4.4: Plotted spatial coordinates and temporal margins (raw daily counts) of the AEGISS data from the corresponding R commands.

```
R> + xlab="day",ylab="Case count",pch=16,cex=0.5)
R> par(mfrow=c(1,1))
```

Note that unlike Figure 3.14, we have not averaged over successive weekly periods for the temporal margin here.

The first step is to now convert our data structures into the class accepted by **lgcp**.

```
R> xyt <- stppp(list(data=cbind(AEGISS$x,AEGISS$y,AEGISS$t),
R> + tlim=range(AEGISS$t),window=W))
R> xyt

Space-time point pattern
 planar point pattern: 10554 points
window: polygonal boundary
enclosing rectangle: [0, 1] x [0, 1] units
   Time Window : [ 1 , 1095 ]
```

Deterministic components

Next, we use the interactive fixed-bandwidth edge-corrected kernel density estimation tool to estimate the deterministic global spatial trend ζ_ψ. We call

```
R> zeta <- lambdaEst(xyt,axes=T)
```

which opens or accesses the active graphics device and produces an initial plot of the spatial density estimate. Simultaneously opened is a device from the **rpanel** package, linked to the graphical display. This panel contains a field in which the user may manually enter a value for the bandwidth; the resulting density estimate is seen in the corresponding plot. Figure 4.5 shows a screenshot of this functionality for our AEGISS example. The field

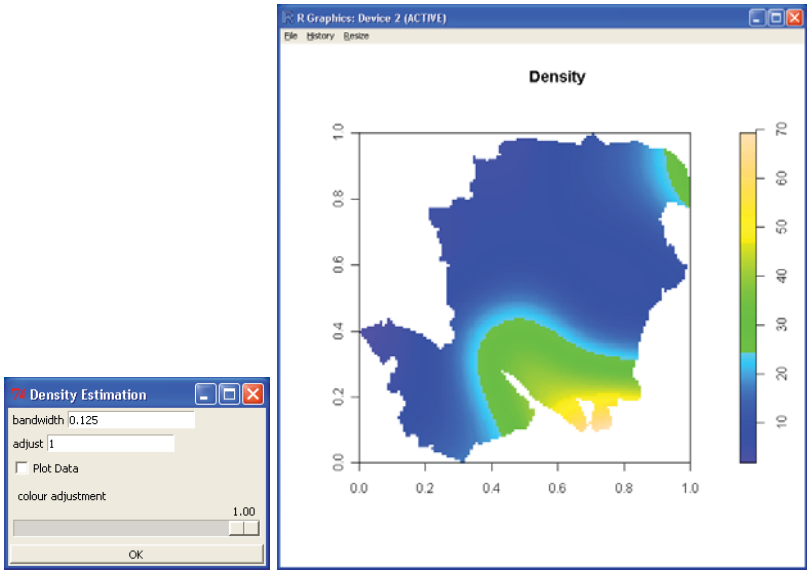


Figure 4.5: Interactive fixed-bandwidth density estimation of the AEGISS spatial data in **lgcp**.

`bandwidth` has the obvious purpose, and additional fields in the panel assist with calculation and visualisation. The user may scale the selected bandwidth by a factor of `adjust`, and include the spatial observations on the plot by checking `Plot Data`. The slider handle labelled `colour adjustment` alters the binning of colour ribbon, a feature that can be useful when the observed data are highly heterogeneous.

Drawing on the results from the second simulation study in Section 3.5.2, we opt for a relatively large bandwidth (given the data have been rescaled to fall inside the unit square) of 0.125; this is the image displayed in Figure 4.5. Clicking `OK` closes the panel device and stores the resulting density estimate, `zeta`, as a pixel image object (class `im`). This we subsequently convert to an object of class `spatialAtRisk` with

```
R> zeta.sar <- spatialAtRisk(zeta)
```

A function for estimating the global temporal trend in terms of the expected number of observed cases per time point via a LOWESS smoother is included. We execute

```
R> eta.tar <- muEst(xyt,f=0.2)
```

where `f` has the same interpretation as the argument of the same name for the R function `lowess`; part of R's built-in, automatically loaded package **stats**. Setting smoothing span to 20%, the function `muEst` automatically returns an appropriately scaled `temporalAtRisk` object; there is no need for explicit conversion here. The fitted global daily trend is superimposed upon the raw daily counts in Figure 4.6; this plot generated by the following code:

```
R> plot(unique(xyt$t),as.numeric(table(xyt$t)),
R> + xlab="day",ylab="Case count",pch=16,cex=0.5)
R> times <- xyt$tlim[1]:xyt$tlim[2]
R> lines(times,sapply(times,eta.tar),col=2,lwd=2)
```

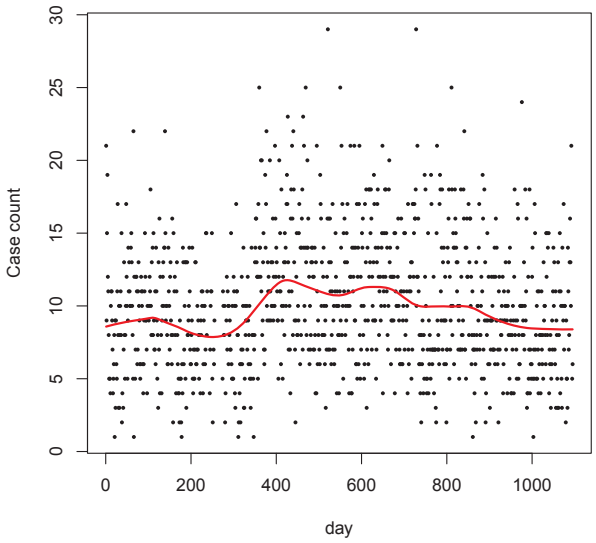


Figure 4.6: Showing the estimated global temporal trend $\eta(t)$.

In situations where the researcher wishes to assume a constant global temporal trend (i.e. any heterogeneity in the temporal counts is to be explained by the stochastic component), **lgcp** provides the function `constantInTime` which also returns a correctly scaled `temporalAtRisk` object.

Specification of `zeta.sar` and `eta.tar` does not necessarily have to be performed using the above tools; they are provided for basic self-contained functionality. As mentioned in Section 4.2.1, we can supply a variety of different data structures to the functions `spatialAtRisk` and `temporalAtRisk` to obtain the required trends in the correct format. This provides the user with an extremely flexible framework for modelling the global variation in the observed spatiotemporal data.

Stochastic component parameter estimation

Once satisfied with the deterministic components, we turn to estimation of the parameters controlling the stochastic component, namely the residual spatiotemporal LGCP. The stationary, isotropic exponential function will describe the theoretical correlation structures r and τ; we require estimates of σ^2_ψ, ϕ_ψ and θ and begin with the spatial parameters.

To use minimum contrast methods (or, equivalently, the interactive features of **lgcp**) the time-averaged nonparametric estimates of either g (pair correlation function) or K are needed. This is easily achieved by calling

```
R> rvals <- seq(0,0.025,length=513)
R> g.mean <- ginhomAverage(xyt,spatial.intensity=zeta.sar,
R> + temporal.intensity=eta.tar,rvals=rvals)

  |================================================| 100%
```

```
Returning an average of 1047 curves
```

```
R> K.mean <- KinhomAverage(xyt,spatial.intensity=zeta.sar,
R> + temporal.intensity=eta.tar,rvals=rvals)

  |================================================| 100%
Returning an average of 1047 curves
```

The raw data as well as the estimated global trend objects must be passed to the functions `ginhomAverage` and `KinhomAverage`. The sequence of spatial distances at which we are interested in evaluating these functions is pre-specified by `rvals`; note the short-range upper limit owing to the residual, 'local' nature of the stochastic component. Some computation time must be expected, and progress bars are incremented during execution (on this author's machine the total execution time of the above code block was roughly 2.5 minutes). We also note that depending on the raw data `xyt`, some benign warnings can be issued by R concerning duplicated spatial coordinates; we have suppressed these in the above output. Provided the level of duplication is not too extreme, the point process theory on which our models are based can still be considered (approximately) valid and it is relatively safe in practice to ignore such warnings. Now, the interactive tools for minimum contrast estimation of the parameters can be demonstrated.

```
R> g.sigmaphi <- spatialparsEst(g.mean,sigma.range=c(0.5,3),
R> + phi.range=c(0.0001,0.01))
R> K.sigmaphi <- spatialparsEst(K.mean,sigma.range=c(0.5,3),
R> + phi.range=c(0.0001,0.01))
```

Like `lambdaEst`, a call to `spatialparsEst` opens a control window linked to a graphical display. Shown are the nonparametric and corresponding theoretical curves of either g or K, depending on which of the outputs from either `ginhomAverage` or `KinhomAverage` have been supplied to `spatialparsEst`. Both are displayed in Figure 4.7.

The control panel provides three options for the functional form of the spatial correlation: Exponential, Matérn, or Whittle. The latter two require specification of an additional 'shape' parameter, named `nu` here. Our choice of `exponential` means we can ignore this (the reader is directed to the help file `?CovarianceFct` of package **RandomFields** for further details on these forms). Also provided is the option to view the curves on their respective 'raw' scales, or to transform them according to the preferences indicated for minimum contrast procedures in Section 3.2.3. In both cases, we choose the transformed version of the curves.

The objective when running these functions is to move the slider handles for `sigma` and `phi` to visually find the theoretical curve which most appropriately matches, to the satisfaction of the user, the nonparametric counterpart. Inspection of Figure 4.7 shows the values this author arrived at for both g and K versions of the tool. The selections are $\sigma_\psi^2 = 2^2 = 4$, $\phi_\psi = 0.00599$ when using g and $\sigma_\psi^2 = 2.88^2 = 8.294$, $\phi_\psi = 0.0021$ when using K. As we would expect, these parameter values are very close to those found when numerically minimising the discrepancy between these curves in Section 3.4.2; Figure 3.16. Remaining unimpressed with the K-function match as before, we discard these results and work with those corresponding to g. Refer back to Section 3.4.2 for further discussion.

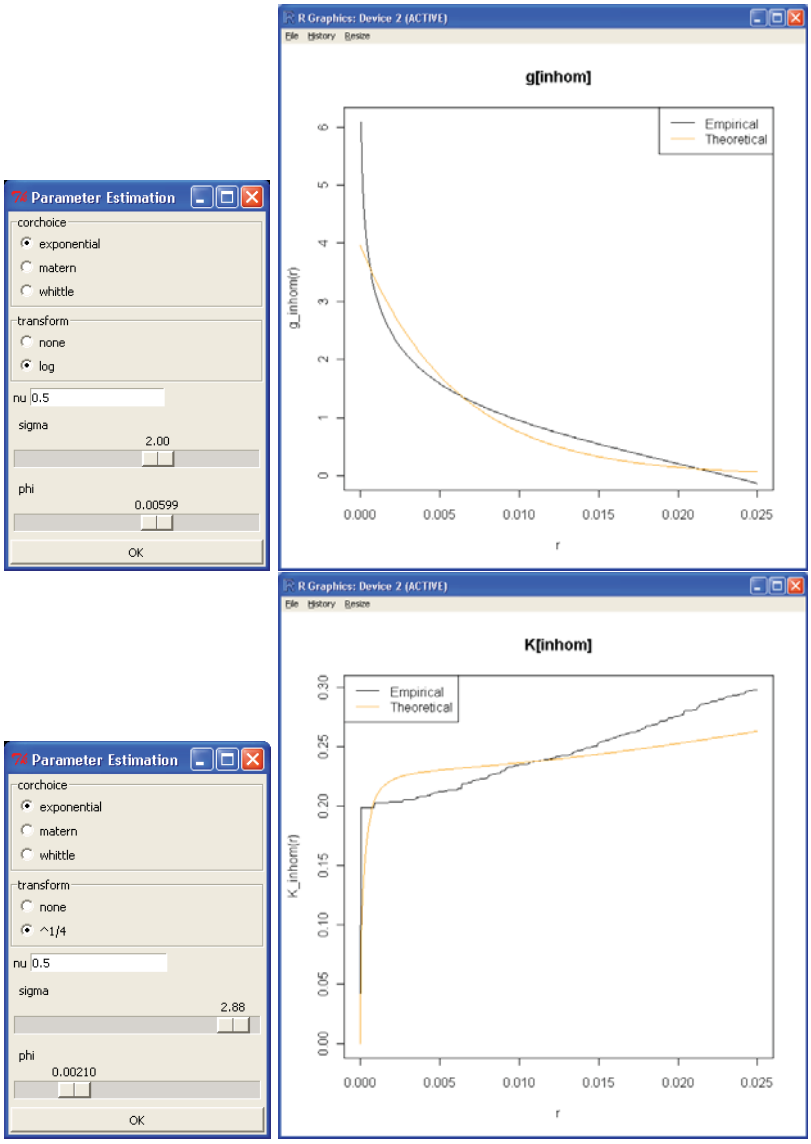


Figure 4.7: Interactive spatial parameter estimation in **lgcp**, based either on the PCF (top) or K-function (bottom).

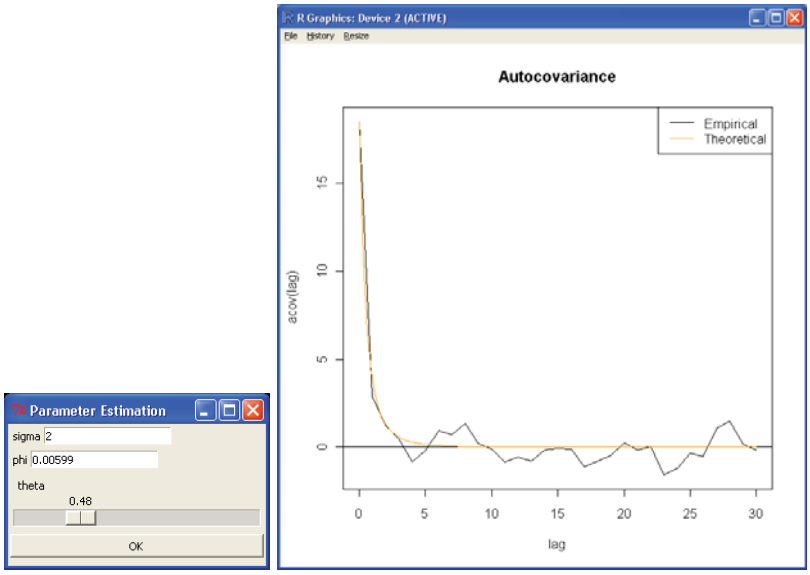


Figure 4.8: Interactive temporal parameter estimation for θ^{-1} in **lgcp**, based on the the temporal covariance function.

It remains to interactively estimate θ. Currently, the Exponential correlation function is the only form coded for the temporal correlation τ in **lgcp**. Furthermore, recall that our discussions thus far have defined, for an integer lag ν, $\tau(\nu) = \exp[-\nu/\theta]$ i.e. θ is a *divisive* scaling parameter. Package **lgcp** defines θ as a *multiplicative* scaler in all relevant functions. Therefore if viewed in the context of our earlier theoretical expositions we should consider the following estimation steps as actually finding $1/\theta$.

```
R> C.theta <- thetaEst(xyt,spatial.intensity=zeta.sar,
R> + temporal.intensity=eta.tar,sigma=g.sigmaphi$sigma,phi=g.sigmaphi$phi,
R> + theta.range=c(0,2),spatial.covmodel="exponential")
```

The above code snippet, calling `thetaEst`, requires the `stppp` object of our data and the spatial and temporal global trends. In accordance with the earlier decision to use only the PCF-selected parameters, these are also passed to the function. A viewable range of interest for $1/\theta$ is specified, as well as the selected functional form for the spatial correlation. Opened for the final time is another **rpanel** controlling the active graphical display, where the theoretical and nonparametric temporal covariance functions $\mathcal{C}$ (3.22) and $\hat{\mathcal{C}}$ (3.23) are plotted. Figure 4.8 shows the results after a satisfactory visual match using the slider handle for `theta`$\rightarrow \theta^{-1}$.

Note that the values of `sigma` and `phi` passed to `thetaEst` are automatically set as the values affecting the theoretical curve. The user may enter different values to observe the impact they have on the behaviour of $\mathcal{C}$, but in searching for a suitable θ, both σ and ϕ should be kept fixed at the values which we intend to use for conditional simulation. The chosen value for θ^{-1} is stored as `C.theta` upon clicking `OK`. We have $\theta = 1/0.48 = 2.083$ which again is close to the numerically minimised value in Section 3.4.2; Figure 3.18.

Finally, note that as with computation of the deterministic components, these supplied

interactive approaches for estimation of σ_ψ, ϕ_ψ and θ do not have to be used, particularly in the event the researcher has additional information which affects their interpretation of the behaviour of the stochastic dependence. These tools are provided in **lgcp** primarily for exploratory purposes; their presence highlights the *ad hoc* nature of minimum contrast methods and associated drawbacks. Indeed, their use restricts the option of spatial correlation functions to those implemented and, being visually based, there is obviously no guarantee that the results arrived at will be equal to the numerical minimum discrepancy.

Conditional simulation: running the MALA

We wish to predict the true form of the spatiotemporal intensity at the most recent timestamp based on the observed data and, additionally, to obtain exceedance probabilities based on the extremity of the predicted residual correlation. We consider these goals fundamental in real-time disease surveillance; they are achieved through a run of the spatiotemporal Metropolis-adjusted Langevin algorithm.

Package **lgcp** provides the first publicly accessible R implementation of the MALA for this objective. The core function of the package, `lgcpPredict`, performs the required operations and returns an object containing detailed result and chain diagnostic information.

```
R> args(lgcpPredict)

function (xyt, T, laglength, model.parameters = lgcppars(),
    spatial.covmodel = "exponential", covpars = c(), cellwidth = NULL,
    gridsize = NULL, spatial.intensity, temporal.intensity,
    mcmc.control, output.control = setoutput(),
    autorotate = FALSE, gradtrunc = NULL)
```

Naturally, the observed data must be supplied as an object of class `stppp` for the argument `xyt`. The argument `T` specifies the time point at which we wish to predict the spatial intensity, and `laglength` gives the number of previous time points to consider during execution of the chain. As we assume the temporal correlation between two time points diminishes as the lag between them increases, it makes little sense to incorporate absolutely all historical intensities in our prediction of the 'current' intensity. Though perfectly possible in theory, in practice this would result in ever-increasing amounts of storage space and execution time, to a negligible impact on the final results (see our earlier discussion in Section 3.3.2 and also Brix and Diggle, 2001). Our minimally required spatial and temporal correlation parameters (σ^2_ψ, ϕ_ψ and θ) are specified through the `lgcppars` function to `model.parameters`. Defaulting to `"exponential"`, the functional form of the spatial correlation can take any of those specified for `CovarianceFct` in **RandomFields**, though the reader is reminded in this particular example we are restricted to those used to interactively estimate the parameters in `spatialparsEst`. The argument `covpars` must specify parameters additional to those in `model.parameters`, if there exist any for the elected `spatial.covmodel`. The arguments `cellwidth` and `gridsize` are used to determine the resolution of the rectangular evaluation grid; only one need be provided. Global deterministic trends, the `spatialAtRisk` and `temporalAtRisk` objects, must be passed as `spatial.intensity` and `temporal.intensity`

respectively. Desired characteristics of the Markov chain are given to `mcmc.control`, the required form of the which is best demonstrated when we execute `lgcpPredict` below. Similarly, control over output including disk storage location and exceedance probability calculation is specified in `output.control`. The argument `autorotate=TRUE` determines if any efficiency is yielded by rotating the irregular study region, subsequently requiring storage of fewer cells of the rectangular evaluation grid. Defaulting to `FALSE`, this can be useful if we have, for example, a long, thin study region that is diagonally oriented. Finally, the user can specify the level of gradient truncation through `gradtrunc`. By default, `gradtrunc=NULL` will automatically compute the required quantity by simulating a (relatively) small number of variables from the target density, evaluating the gradient norm at each, and using the largest of these values as the truncation.

Prior to commencing the MALA, we must decide upon appropriate thresholds for the exponentiated Gaussian field. For a given retained field, any spatial locations at the time point of interest breaching these thresholds will be classified as an 'exceedance'. These are found, as in Section 3.4.2, using appropriately high quantiles of the corresponding lognormal distribution based on the estimated value of σ_ψ^2. In this particular example, the commands

```
R> quans <- c(0.9,0.95,0.99)
R> thresholds <- exp(-0.5*g.sigmaphi$sigma^2+g.sigmaphi$sigma*qnorm(quans))
R> thresholds

[1]  1.756114  3.631723 14.192246
```

provide the thresholds of interest for the 90^{th}, 95^{th} and 99^{th} percentiles. The command

```
R> exceed <- exceedProbs(thresholds)
```

must then be run; here `exceedProbs` takes a vector of desired thresholds and returns a new function (named `exceed` here) which can in turn be provided with a matrix. For each of the supplied thresholds, the function `exceed` then returns a logical matrix of the same dimensions as its argument after checking the entries thereof against the supplied thresholds; `TRUE` is returned for any entry which positively breaches them.

These steps are necessary owing to the fashion in which `lgcpPredict` constructs and analyses its Markov chain. We are now ready to execute.

```
R> aegissMALA <- lgcpPredict(xyt=xyt,
R> + T=xyt$tlim[2],
R> + laglength=5,
R> + model.parameters=lgcppars(sigma=g.sigmaphi$sigma,phi=g.sigmaphi$phi,
R> +                           theta=C.theta),
R> + gridsize=64,
R> + spatial.intensity=zeta.sar,
R> + temporal.intensity=eta.tar,
R> + mcmc.control=mcmcpars(mala.length=120000,burnin=20000,retain=100,
R> +                       MCMCdiag=3,adaptivescheme=constanth(0.004)),
R> + output.control=setoutput(gridfunction=dump2dir(dirname=
```

```
R> +                          "H:/PhD/Thesis/Chap4/aegissMALA",
R> +                          forceSave=TRUE,lastonly=T),
R> +                          gridmeans=MonteCarloAverage("exceed",
R> +                          lastonly=T))
R> + )

FFT Grid size: [128 , 128]
WARNING: disk space required for saving is approximately 187.61 Mb
Computing gradient truncation ...
|======================================================================| 100%
Using gradient truncation of 1232535
Netcdf file: H:/PhD/Thesis/Chap4/aegissMALA/simout.nc created
Initialising MALA ...
MALA successfully initialised ... please wait
  |=============================================| 100% : Run [119501/120000]

Files written to H:/PhD/Thesis/Chap4/aegissMALA/
```

The arguments to `lgcpPredict` are set to our desired values. The time point of interest, `T`, is the most recent, given by the maximum limit of the temporal window. The model parameters σ^2_ψ, ϕ_ψ and θ are given as the results from use of the PCF for estimation. Setting `gridsize` to `64` instructs the function to use a 64×64 evaluation grid. After passing in the deterministic global trends, `lgcpPredict` must now be told about the chain itself.

Values for `mcmc.control` are specified through the secondary function `mcmcpars`, and we provide the `mala.length=120000` iterations, a `burnin` of `20000` iterations and instruct to thin by `retain`ing only every 100^{th} non-burn-in iteration. The argument `MCMCdiag` is for diagnostic purposes post-execution; this is given a positive integer value instructing `lgcpPredict` to store the individual results from `3` randomly selected cells on the torus. The tuning parameter controlling candidate generation, h, is specified with the value of `adaptivescheme`. Use of `constanth(0.004)` sets $h = 0.004$.

As an aside, note that some methodology exists for sensible definitions of variable tuning parameters; code included for this objective variable in **lgcp** by B.M. Taylor. The approach uses the methods detailed in Andrieu and Thoms (2008) and, rather than using `constanth(...)`, is specified using `andrieuthomsh(...)`. This work has utilised a fixed h as in the original LGCP publications by Møller et al. (1998) and Brix and Diggle (2001); this is where we will remain for the current discussion. The reader is directed to Andrieu and Thoms (2008), the package documentation for `?andrieuthomsh`, and our **lgcp** article for further details on the alternative.

The function `lgcpPredict` is told what to compute, and how to store results, through `output.control=setoutput(...)`. Not performed by default, we elect to dump results to disk at the folder location given by `gridfunction=dump2dir(...)`. Additional arguments to this function tell `lgcpPredict` to `forceSave` (i.e. write) to the hard disk drive regardless of estimated final file size, rather than prompt the user at the beginning of execution (default), and to store the results of the `lastonly` time `T`, rather than also saving the previous

`laglength` timestamp results.

Finally, additional posterior results are instructed to be computed via the `setoutput gridmeans=MonteCarloAverage(...)` argument. Standard results such as cell-wise posterior mean and variances of the raw and exponentiated Gaussian field are returned automatically. Our desire for exceedance probabilities must be given here, however. The function `MonteCarloAverage` computes the results for a vector of appropriately defined functions; in this example we require only the results from the earlier-defined function `exceed`. The additional argument `lastonly` has the same effect as in `dump2dir`.

Commentary and progress bars show the completion status during execution. Note that `lgcpPredict` uses the minimum required grid extension for the FFT; an evaluation grid of $(M-1)\times(N-1) = 64\times 64$ means an extended grid of $(M_{\text{ext}}-1)\times(N_{\text{ext}}-1) = 128\times 128$. If used, the gradient truncation stage completes relatively quickly, with total execution time dominated by the MALA burn-in and run itself. Should execution be performed on a remotely accessed machine, it is also prudent to store the returned `lgcpPredict` object; this can be achieved with standard R functions such as `dput` (R package **base**).

Post-processing and visualisation

By `print`ing the returned object, we gain a useful summary of the run:

```
R> aegissMALA

lgcpPredict object.

  General Information
  -------------------
      FFT Gridsize: [ 128 , 128 ]

              Data:
       Time |     1090     1091     1092     1093     1094     1095
     Counts |       13       10       21        0        5        6

        Parameters: sigma=2, phi=0.006, theta=0.48
    Dump Directory: H:/PhD/Thesis/Chap4/aegissMALA/

     Grid Averages:
         Function Output Class
           exceed        array

        Time taken: 15.158 hours

  MCMC Information
  ----------------
    Number Iterations: 120000
              Burn-in: 20000
```

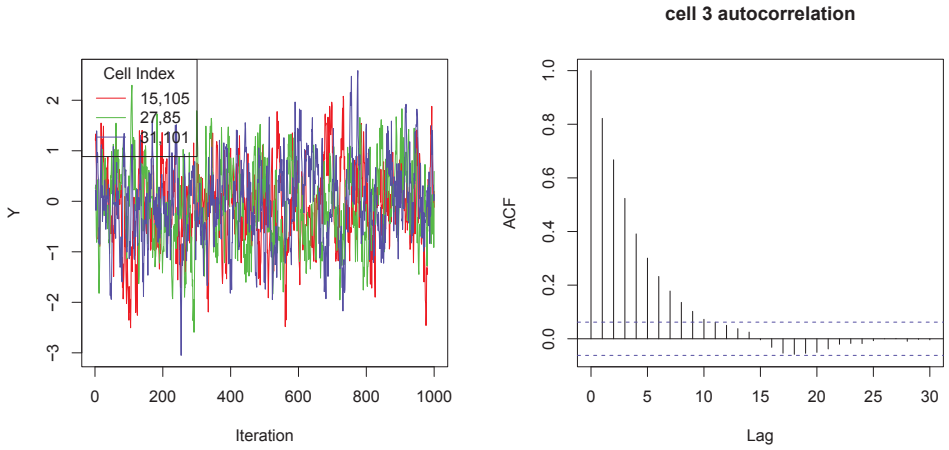


Figure 4.9: Trace plots of the values of the retained Gaussian field for the three randomly selected torus cells (left) and an autocorrelation plot of the retained chain values for the third cell (right).

```
        Thinning: 100
Mean Acceptance: 0.594
Adaptive Scheme: constanth
         Last h: 0.004
```

New information provided here includes the total execution time, in this case around 15 hours, and the mean acceptance rate, 0.594.

Mixing and convergence can be assessed quickly using the command `mcmctrace`. This retrieves the values of the Gaussian field from `aegissMALA` at each iteration for each of the `MCMCdiag=3` randomly selected cells. The result of the following instructions is given in Figure 4.9.

```
R> aegiss.diag <- mcmctrace(aegissMALA)
R> plot(aegiss.diag)
R> acf(aegiss.diag$trace[,3],main="cell 3 autocorrelation")
```

The trace plots show satisfactory mixing and appear to have achieved convergence. To a certain extent, correlation is exhibited in the resulting values; this is typical of MCMC and of less concern than if we had not observed the relatively well-behaved trace plots. The level of correlation has the potential to be far more severe if we had used the (discretised) Gaussian field Z directly in the iterations rather than the linear transformation Γ – refer back to Section 3.3.2 and Section 8 of Møller et al. (1998). Dumping the results to disk, rather than just working with the returned `lgcpPredict` object, allows for particularly detailed diagnostic summaries: with the entire contents of the retained chain available, the user can perform any number of operations on the posterior realisations in order to gain a better understanding of the behaviour of the MALA.

The dumped data is given as a 'netCDF' (Network Common Data Form) file named `simout.nc` in the folder specified in `dump2dir`. These file types are specially designed to allow

relatively fast access to large volumes of information and can be accessed via the **ncdf** package (Pierce, 2011), automatically loaded as a dependency when calling `library(lgcp)`. Briefly, to load the entire retained chain of the raw Gaussian fields into the current workspace, we call

```
R> aegiss.ncdf <- open.ncdf("H:/PhD/Thesis/Chap4/aegissMALA/simout.nc")
R> aegiss.chain <- get.var.ncdf(aegiss.ncdf)
R> close.ncdf(aegiss.ncdf)
R> dim(aegiss.chain)

[1]   64   64 1000
```

which indicates we have an `array` containing 1000 stored 64×64 matrices: the spatial Gaussian fields for the final time point only (recall `lastonly=TRUE`) for each of the retained iterations (of which there are clearly 1000 in total). Had we specified `lastonly=FALSE` in `dump2dir` the array given by `aegiss.chain` would have an extra dimension of size `laglength+1` to provide the spatial intensities corresponding to the historical window in addition to the `lastonly` timestamp. Should we not desire surface-wise inspection of the whole chain, **lgcp** provides the function `extract.lgcpPredict`, which allows rapid access to sub-samples without requiring the importation of the entire data set. Storage of the chain is also useful should the user decide post-execution that they require summaries in addition to those specified by `MonteCarloAverage` as `gridmeans` in `setoutput` for `output.control` in `lgcpPredict`. To this end, the function `expectation.lgcpPredict` is supplied.

Important summary statistics are returned automatically and do not require dumping of the chain to disk. Use of the functions `meanfield(aegissMALA)` and `varfield(aegissMALA)` return the matrices representing the cell-wise posterior mean and variance of the Gaussian field on W respectively for each `laglength+1` time points. The functions `rr(aegissMALA)` and `serr(aegissMALA)` return the cell-wise posterior mean and associated standard errors for the exponentiated Gaussian field (the reason for their names is clarified below) for each relevant time point. The posterior mean intensity, again at each influential time point `T-laglength` to `T` inclusive and obtained by the product of the spatial and temporal deterministic components with the posterior mean of the residual LGCP, is extracted via `intens(aegissMALA)`. The spatial x- and y-axis coordinates which were used to approximate the spatially continuous field are retrieved with `xvals(aegissMALA)` and `yvals(aegissMALA)` respectively.

Plotting the posterior may be done through use of `plot.lgcpPredict`. We have two automatic options available: view the cell-wise posterior mean of the exponentiated Gaussian field $\overline{\exp[Z_T]}$, or view the final intensity estimate $\bar{\psi}_T = \zeta_\psi \eta \overline{\exp[Z_T]}$ (we assume for simplicity the only time of interest is the final time point T=`T`). Note that the former option can in the current setting be referred to as the 'relative risk' plot in the sense that positive extremity in the residual process indicates heightened local 'risk' of observing a disease case. This should not be confused with the kernel-smoothed density-ratio estimator of relative risk discussed in Chapter 2 and dealt with in **sparr**. Also, note that the user is not restricted to the plotting methods demonstrated below; with access to each raw, retained Gaussian field, the user can retrieve and plot their results with any suitable functions in the R language.

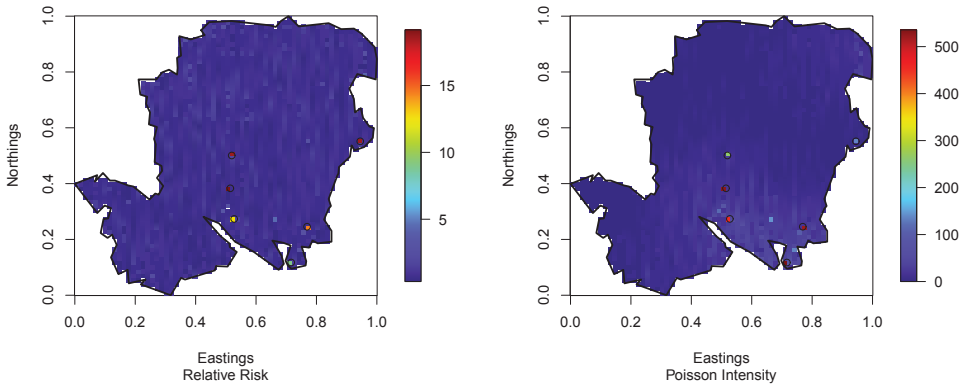


Figure 4.10: Posterior mean plots for the AEGISS data. Left: mean relative risk $\overline{\exp[Z_T]}$. Right: mean Poisson intensity $\zeta_\psi \eta \overline{\exp[Z_T]}$.

The following code snippet produces Figure 4.10, with the argument `type` distinguishing between the two options of 'relative risk' and 'Poisson intensity'. By default, this class-specific version of the generic `plot` will plot all of the `laglength+1` surfaces; a feature controlled by argument `sel`. With our earlier specification of `laglength=5`, setting `sel=6` instructs the function to display the surface at the latest time point `T`. When multiple time points are plotted, `plot.lgcpPredict` will also by default prompt the user to update the graphics device via `ask`. Viewing only one timestamp, we set `ask=FALSE`. Denying the function the ability to provide the observations automatically by setting `plotdata=FALSE`, here we add them manually with the standard R function `points`.

```
R> par(mfrow=c(1,2))
R> lastday <- xyt[xyt$t==xyt$tlim[2]]
R> plot(aegissMALA,type="relrisk",sel=6,plotdata=FALSE,ask=FALSE,
R> + xlab="Eastings",ylab="Northings")
R> points(lastday,pch=8)
R> plot(aegissMALA,type="intensity",sel=6,plotdata=FALSE,ask=FALSE,
R> + xlab="Eastings",ylab="Northings")
R> points(lastday,pch=8)
R> par(mfrow=c(1,1))
```

Naturally, peaks in both the the predicted relative risk and overall Poisson intensity surfaces correspond to where data have currently and/or have very recently, been observed. The effect of including the multiplicative deterministic components is clear in that the spatial variation now reflects the underlying global trend and the integral of the function now represents the scale of the number of daily observations.

Finally, from a disease surveillance perspective, we wish to determine if the peaks in relative risk we have currently observed constitute any anomalous behaviour in terms of consistent exceedance by the retained realisations of the pre-computed percentile thresholds. The proportion of $\exp[Z_T]$ realisations which breach a given threshold for each spatial

location inside the study region constitute the exceedance probabilities, and can be automatically plotted in **lgcp**. The following code produces Figure 4.11.

```
R> par(mfrow=c(length(thresholds),1))
R> plotExceed(aegissMALA,"exceed",ask=FALSE,
R> + xlab="Eastings",ylab="Northings")
R> par(mfrow=c(1,1))
```

In much the same way as Figures 3.22 and 3.23, we would typically insist upon high exceedance probabilities (i.e. 0.8 at the very least) before flagging a cell or localisation of cells as an anomaly. Furthermore, it is reasonable to suggest that a 'true' anomaly is far more likely to persist over time, rather than appearing at one time point and disappearing the next (Diggle, Rowlingson and Su, 2005), so regular monitoring of such a system over time would be recommended.

In any case, the exceedance maps in Figure 4.11 do not indicate any real cause for alarm. Use of the smallest threshold fails to indicate any cells providing a sufficiently high probability, dictating that the larger thresholds will yield even smaller exceedance proportions. In other words, the data observed at time `T=1095` as well as the recent history of the observations at the `laglength=5` previous time points have fallen within what the model estimates is 'normal' variability.

4.2.3 Some interesting notes on the FFT

As discussed in detail in Section 3.3.1, critical to manageable execution times in **lgcp** is the fast-Fourier transform. The technique of circulant embedding applied to the covariance matrix corresponding to the discretised spatial region allows appropriate use of the FFT for quick diagonalisation. This spectral decomposition of the covariance matrix is necessary so that we may repeatedly simulate from the high-dimensional Gaussian fields as required for the MALA in a timely fashion (and work with the linear transformation Γ rather than the field directly – refer back to Section 3.3.2). The command `fft`, part of the standard R package **stats**, computes the unnormalised discrete Fourier transform for a specified array and allows us to recover (via the *inverse* transform) the eigenvalues of a positive semidefinite block circulant matrix in the same way as more computationally expensive decomposition methods would.

Suppose, for illustrative purposes, we are working on a ridiculously coarse 2×2 square grid. The minimum extended grid size will be 4×4. The extended-grid covariance matrix will therefore have dimension 16×16, and have 16 eigenvalues. If the matrix of interpoint distances have been computed as per the minimum torus distances in Section 3.3.1, then this extended matrix will be block circulant. As per Appendix E in Møller and Waagepetersen (2004), if this matrix is also positive semidefinite, then it represents the covariance matrix of the circulant Gaussian random field on the extended grid. Suppose this object is stored as `extended.covariance.matrix` in the current R workspace. We may find the eigenvalues using the **base** R function `eigen`, which below makes use of the QR decomposition.

```
R> ext.dim <- dim(extended.covariance.matrix)
R> ext.dim
```

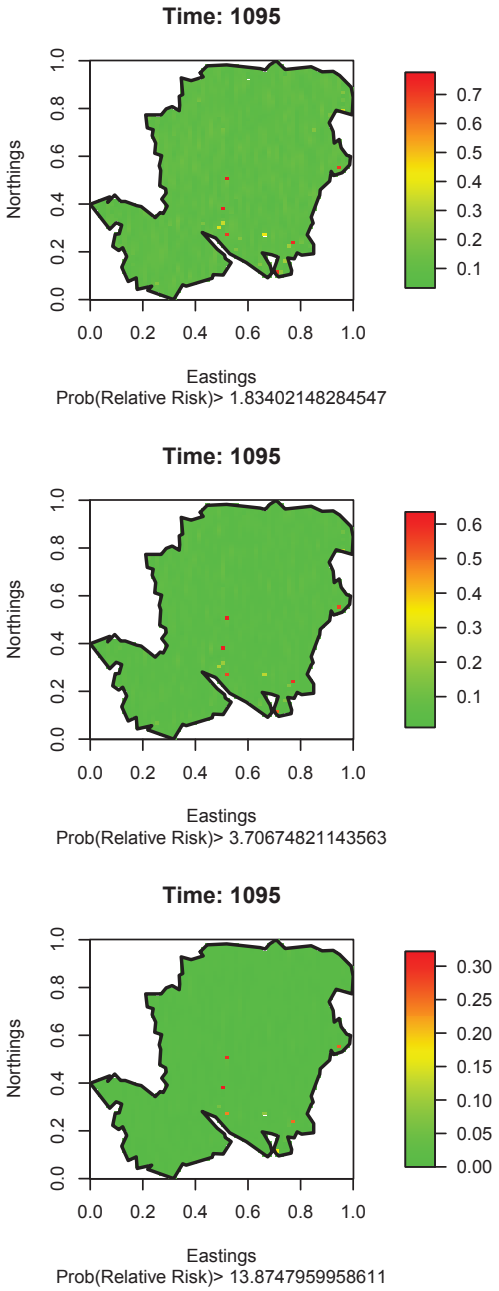


Figure 4.11: Exceedance probabilities for the final time point for the AEGISS analysis in **lgcp**.

```
[1] 16 16
```

```
R> eigs <- eigen(extended.covariance.matrix,symmetric=T)$values
R> eigs
```

```
[1] 16.805091  5.851589  5.851589  5.851589  5.851589  3.153741  3.153741
     3.153741  3.153741  1.675565  1.675565  1.675565  1.675565  1.612855
     1.429238  1.429238
```

Note that this method has required the entire covariance matrix (strictly speaking, knowledge of one diagonal half thereof due to symmetry), which has itself required the construction and covariance function evaluation of the extended torus-wrapped distance matrix.

The block circulant nature of the extended matrix means that all the information required for equivalent results using the FFT are contained in a 4×4 matrix formed *from the first row of* `extended.covariance.matrix` *only*.

```
R> fast.matrix <- matrix(extended.covariance.matrix[1,],
R> + nrow=sqrt(ext.dim),ncol=sqrt(ext.dim))
R> eigs.fft <- rev(sort(Re(fft(fast.matrix,inverse=T))))
R> eigs.fft
```

```
[1] 16.805091  5.851589  5.851589  5.851589  5.851589  3.153741  3.153741
     3.153741  3.153741  1.675565  1.675565  1.675565  1.675565  1.612855
     1.429238  1.429238
```

A quick inspection reveals that using just a fraction of the size of the entire extended covariance matrix yields identical eigenvalues via this alternative. As an aside, note that this example was run simply to demonstrate the equivalence of the two functions in the special case of the positive semidefinite block circulant matrix. In practice for the LGCP, the `eigen` method would technically not be evaluated on the extended covariance matrix. In this situation, however, subsequent inversion of the matrix of eigenvectors (the derivation of which constitutes the major contributing factor to computational costs), provided by the call to `eigen` with respect to the covariance matrix on the original lattice, would have to occur in order to find the diagonalisation matrix Q (refer back to Section 3.3.2 for details on this quantity).

For such small dimensions, the difference in computational expense between the two approaches is of course not noticeable. As the grid resolution becomes finer though, use of functions such as `eigen` becomes all but impossible. As a quick example, consider (a still relatively coarse) 40×40 grid; covariance matrix 1600×1600. The corresponding extension would be (at least) of size 80×80, with covariance matrix 6400×6400, but as we require only the first row, this dramatically reduces to computation of a certain 80×80 array only. Without needing the grid extension if working with `eigen`, we still however seek full eigendecomposition of the original 1600×1600 covariance matrix. Consider the following 'dummy' example, where matrices (symmetric for use in `eigen`) of the specified dimensions are simply filled with randomly generated independent standard normal values to demonstrate the impact of excessively high dimensionality on the functions of interest.

```
R> slow.mat <- toeplitz(rnorm(1600))
R> start <- Sys.time()
R> slow <- eigen(slow.mat,symmetric=T)
R> end <- Sys.time()
R> end-start

Time difference of 7.906 secs

R> fast.mat <- matrix(rnorm(80^2),80,80)
R> start <- Sys.time()
R> fast <- fft(fast.mat,inverse=T)
R> end <- Sys.time()
R> end-start

Time difference of 0.03200006 secs
```

The improvement in computational expense is clear; this dummy example roughly exhibiting a 250-fold improvement of `fft` over `eigen`. Considering the MALA must repeat these kinds of operations tens of thousands of times, not to mention at much finer resolutions, this speed and efficiency is an invaluable feature of the FFT methodology. Indeed, working with evaluation grids larger than around 40×40, `eigen` in fact fails to complete, citing insufficient memory allocations in R. It is also important to remember, as mentioned in Section 3.3.1, that use of the FFT requires the extended grid in order to be a valid approach due to certain periodicity conditions that must be met. Without this, the spectral decomposition would be incorrect; results being contaminated by 'wrap-around effects'.

The high-dimensionality associated with the discrete approximations to planar Gaussian fields means the performance of **lgcp** is heavily dependent on the benefits afforded by FFT. Indeed, much the same code as briefly illustrated here was used by this author in initial programming efforts concerning the relevant operations for implementation of the MALA and remains a key component of the package.

4.2.4 Limitations and future extensions

Package **lgcp** provides functionality for model specification, fitting, and prediction using the sophisticated space-time log-Gaussian Cox process of Brix and Diggle (2001). Primarily motivated by problems in disease surveillance, emphasis is placed on conditional simulation based on the observed data which involves implementation of the fast-Fourier-transformed Metropolis-adjusted Langevin algorithm for high-dimensional, spatiotemporally correlated Gaussian fields. By collaborating with researchers at Lancaster University, the end result is a functional, powerful tool with a number of additional features enhancing this author's original suite of executable R methods.

This R package depends upon a modest collection of other available packages in order to operate, some of which have been mentioned already. The important spatial data analysis packages **spatstat** (Baddeley and Turner, 2005), **sp** (Pebesma and Bivand, 2005; Bivand et al., 2008), and to a slightly lesser extent **fields** (Furrer et al., 2011), **maptools** (Lewin-Koh

and Bivand, 2011) and **rgdal** (Keitt et al., 2011) provide critical support for spatial region handling, summary statistic calculations, data set class specifications and other internal operations. The useful package **RandomFields** (Schlather, 2011) provides access to a suite of different spatial covariance functions, and **ncdf** (Pierce, 2011) yields the necessary tools for handling the aforementioned netCDF file format. Packages **rpanel** (Bowman et al., 2007), **rgl** (Adler and Murdoch, 2009) and **tcltk** (a standard part of the R installation) provide graphical assistance such as the interactive visuals observed in the previous section. Further support for internal functions is provided by **iterators** (Revolution Analytics, 2011) and the included-in-installation R package **methods**.

At the time of writing the version currently released is 0.9-0, which happens to be the first version submitted to CRAN. As such, there are many sensible directions in which **lgcp** could (and at some point, will) be expanded. The explicit capability to carry out analyses for 'purely spatial' data sets with the LGCP, as initially introduced into statistics by Møller et al. (1998), is seen as one of the most useful extensions to carry out. Theoretically, it is also relatively straightforward to include the capability to handle covariates in the structure of the LGCP, analogous in many ways to random effects models in regression, achieved by exponentiating not just the Gaussian field itself, but the *sum* of the Gaussian field and the product of the covariate value vector with a transposed coefficient vector at each spatial location (see e.g. Møller and Waagepetersen, 2007). This is another interesting feature under consideration for **lgcp**.

The problem of parameter estimation for the dependency structure remains a tricky one. At present, the interactive nature of the tools included in **lgcp** highlights the *ad hoc* nature of minimum contrast. In many ways, the decision to leave the user to themselves decide how to specify these parameters is similar to the decision in **sparr** to only provide a basic level of bandwidth selectors for the kernel-smoothed density-ratio: a certain level of 'user-discretion' will be warranted from one application to the next. Nevertheless, automated functions to e.g. return the minimum contrast parameters estimates would be useful, and this is planned for near-future versions.

The relative youth of the LGCP, particularly with respect to spatiotemporal data sets, is clear. Indeed, the only known application of the spatiotemporal framework to date is of the AEGISS data in Diggle, Rowlingson and Su (2005); we add to this practical experience in Section 5.4 with a preliminary exploration of some earthquake data. The rich variety of spatial and spatiotemporal intensity functions that the LGCP is capable of capturing means that this model and its modifications deserve greater attention in the applied literature. Relevant theoretical improvements at this point also require further investigation from an applied perspective; examples involve the nonseparable correlation structures explored by Rodrigues and Diggle (2010) and a principled Bayesian approach to parameter estimation (preliminary correspondence with Lancaster group). This knowledge will continue to evolve, and by monitoring and contributing to the literature, we aim to enable **lgcp** to provide users a unique portal to the latest developments in this area of spatial and spatiotemporal point process modelling.

Chapter 5

Collaborative Applications and Case Studies

In this chapter, collaborative applied work with academics and other scientists around the globe is summarised. These collaborations have made use of the novel statistical research developed during the course of this project. The first three examples discussed here have been either published or submitted for publication with appropriate co-authorship in relevant peer-reviewed epidemiological journals; the projects are not replicated in their entirety and the discussions vary in length. Rather, emphasis of the summaries is placed on this author's contributions, namely all or a portion of the statistical strategies implemented through programming, interpretation of the results, and write-up in the relevant sections of corresponding articles. The statistical research carried out in the preceding chapters chronologically began with investigations into kernel density estimation and the associated relative risk function: the three published examples have thus all exploited this methodology.

Section 5.1 covers a study that aims to identify high-risk areas of schistosomiasis in a particular region of China. Global trends in outbreaks of avian influenza are explored in Section 5.2, and Section 5.3 details the efforts made to verify the agreement of different models used to simulate the spread of foot-and-mouth disease throughout Irish farms in the event of an epidemic. Finally, for something different, Section 5.4 seeks to apply the spatiotemporal log-Gaussian Cox process to a Japanese earthquake data set. Obviously not an epidemiological example, it nevertheless allows further experimentation of this model in an applied setting and clearly demonstrates the flexibility of the LGCP.

Theoretical specifics are not repeated here. The reader is instructed to consult the relevant sections in Chapter 2 for Sections 5.1-5.3, and Chapter 3 for Section 5.4.

5.1 Schistosomiasis in Guichi, China

Soon after the work on adaptive kernel density estimation of relative risk functions became publicly available (Davies and Hazelton, 2010), this author was contacted by epidemiologists based in China investigating the spatial variation of schistosomiasis in the Guichi region of

the Anhui province. They wished to apply the variable smoothing method to analyse the risk of the disease within this region, and compare it to their previously obtained results using the fixed-bandwidth approach.

Assistance was provided in implementation of the methods (using, in fact, our R package **sparr** – see Section 4.1) and discussion of the results. The corresponding article, Zhang, Davies and Jiang (2011), is currently under review with *Geospatial Health*.

5.1.1 Background

Schistosomiasis, known colloquially as 'snail fever', is a parasitic disease which is estimated to affect over 200 million people worldwide (World Health Organization, 2010). Mainly prevalent in Africa and Asia, widespread infection can have significant economic and public health consequences. Carried by certain snail species and released into waterways when mature larvae, hosts become infected through penetration of the skin by the parasite when in contact with the affected water. The larvae subsequently develop into adult worms residing in the blood vessels; females releasing eggs. Eggs that remain trapped in the body trigger an immune reaction, leading to the visible symptoms. These can range from mild to severe (e.g. fever, abdominal discomfort, diarrhoea, blood in faeces and urine), and depend upon the species of parasite, but the illness is generally treatable. Long-term, unchecked infection, however, can lead to major damage to vital organs and death. It is therefore important that the prevalence of the disease in affected areas continues to be monitored so that control measures can be targeted effectively.

A species of the parasite particularly prevalent in China is *Schistosomiasis japonicum*, which affects the intestinal tract. China has strategies in place to control and limit the incidence rates, though this is extremely difficult due to the extensive habitat of the sole intermediary host (*Oncomelania hupensis*) coupled with extreme weather events such as floods. It remains a significant public health problem in this country; see for example the works by McManus et al. (2005) and Zhou et al. (2007). In Zhang et al. (2009), the authors used fixed-bandwidth kernel density estimation to estimate the relative risk function of schistosomiasis based on data collected in Guichi, Anhui province. The current work analyses the same data with adaptive kernel estimation of spatial relative risk, and draws comparisons with the previous approach.

5.1.2 Data and methods

The data comprise 83 spatial point locations of *S. japonica* cases in the Guichi region, which were recorded using a hand-held Global Positioning System (GPS), alongside 83 controls randomly selected from the at risk population (see Zhang et al., 2008, 2009, for further details on these data and their collection).

Figure 5.1 shows the data set and study region, rescaled to the unit square. Also included are local river systems, namely the Yangtze River bordering the north, and the Qiupu River which feeds into the Yangtze within the Guichi region itself. Particularly noticeable is the tendency of the observed cases to fall near the waterways. Is this simply a reflection of the at-risk population dispersion? A visual inspection may suggest not, though a more

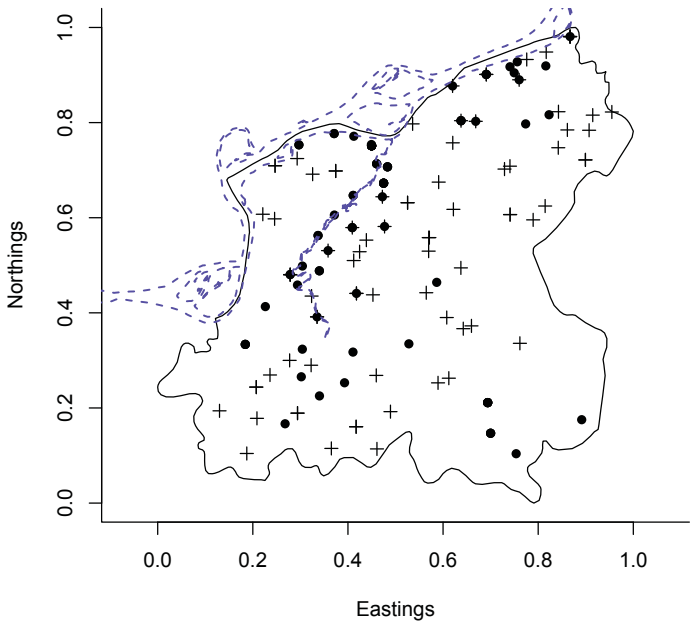


Figure 5.1: Cases (•) and controls (+) for the Guichi schistosomiasis data, including study region (solid border line) and local waterways (dashed blue lines).

sophisticated investigation is of course desirable.

Log-transformed kernel-smoothed density-ratio relative risk surfaces were constructed for these data, using both fixed and adaptive bandwidth versions of the estimator; refer back to Section 2.2. Bandwidths were assigned as per the guidelines discussed in Section 2.3.1, namely a common case-control fixed bandwidth, or a common case-control global bandwidth with separately computed pilots in the adaptive version. The single-density oversmoothing (OS) and least-squares cross-validation (LSCV) selectors were employed, the latter method only used for pilot bandwidth calculation in the adaptive estimates. Tests for significantly positive extremities in risk were performed using both Monte-Carlo (MC) and asymptotic (ASY) methods; refer back to Section 2.4. The former was calculated based on 1000 iterations.

In addition to the risk surfaces computed with the specific choices of bandwidths listed above and described in Chapter 2, the standard error for the estimated risk functions was assessed. The issue of estimated risk function variability frequently goes unmentioned in the literature, despite the critical role it plays in p-value surface and tolerance contour calculation for a given kernel risk surface. Via the ASY method, the standard error surfaces are given as the square root of the asymptotic expressions for the (edge-corrected) log-risk function variance provided in equations (2.25) and (2.26) for the fixed and adaptive functions respectively. The standard errors corresponding to the MC p-value surfaces are simply empirically computed as the standard deviation of the collection of re-sampled risk surface values for each cell.

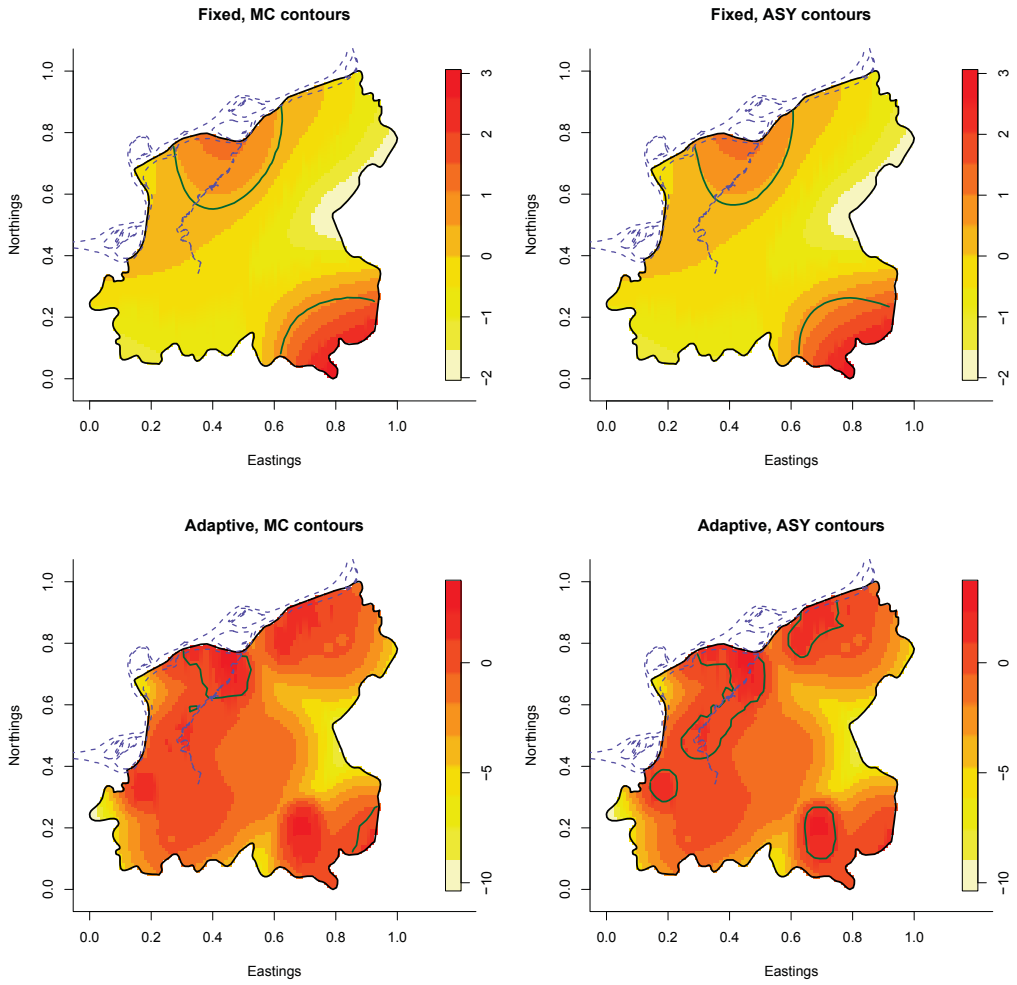


Figure 5.2: Initial risk function results for the schistosomiasis data, showing both fixed and adaptive versions with either MC or ASY upper-test tolerance contours plotted at a 5% level of significance.

5.1.3 Summary of results

Use $\boldsymbol{X}$ and $\boldsymbol{Y}$ to denote the case and control data sets, and let '$\leftarrow$ OS' read as "computed via the OS method on the pooled data set" and '$\leftarrow$ LSCV$_{\boldsymbol{Z}}$' read as "computed via the LSCV method on the data set $\boldsymbol{Z}$". For the interpreted results we therefore have $h_{\mathcal{F}} \leftarrow$OS, and $h_0 \leftarrow$ OS; $\tilde{h}_f \leftarrow$ LSCV$_{\boldsymbol{X}}$; $\tilde{h}_g \leftarrow$ LSCV$_{\boldsymbol{Y}}$, where as per the notation in Chapter 2, $h_{\mathcal{F}}$ denotes the common bandwidth in the fixed setting and h_0 the common global bandwidth in the adaptive setting, with case and control pilot bandwidths $\tilde{h}_f$ and $\tilde{h}_g$ respectively.

These risk surfaces are displayed in Figure 5.2. There appears to be little difference between the ASY and MC contours for the fixed-bandwidth risk surface. Two clear areas of interest are highlighted, one on the northern border where the Qiupu feeds into the Yangtze, and the other in the southeastern quadrant. The latter sub-region has occurred where isolated case observations have occurred with no controls present, though the size of

this significant area is questionable.

The results for the adaptive surface are particularly interesting. The ASY tolerance contours clearly track the Qiupu for most of its length, reflecting our visual examination of the observations in Figure 5.1. There are also three much smaller, more distinct sub-regions flagged on this plot. The appearance of the MC contours for the adaptive surface differ considerably, with only one clear sub-region at the intersection of the Qiupu and Yangtze being identified. The 'enthusiasm' of the ASY contours to flag significant areas, particularly for adaptive risk function estimates, can also be observed at some level in our earlier analysis of the Chorley-Ribble data set in Section 2.6.1. However, it is important to note that the reliability of the ASY approach for both fixed and adaptive smoothing regimens will likely suffer when the sample sizes are small: note that a number of large-sample approximations are made in order to derive valid expressions for the properties of the risk surfaces and subsequent formulæ for the ASY p-value surfaces.

Thus, drawing conclusions based on this analysis alone should be performed with caution. With only 83 cases and 83 controls, we can conceivably expect relatively high levels of variability in the numerically derived MC p-value surfaces. Also, as mentioned above, the collection of theoretical approximations that are made in order to derive the asymptotic properties on which the ASY p-value surfaces depend require by definition the presence of 'large' sample sizes; whether or not our schistosomiasis data set is 'suitably large' is difficult to discern. This can however be aided by examining the standard errors of the computed risk surfaces. Figure 5.3 shows plots of estimated cell-wise standard error for the risk surfaces over the Guichi region. For computational reasons, these were evaluated on a coarser grid than that used in Figure 5.2.

Between ASY and MC standard errors we note less variability, both in terms of size and spanning range, for ASY. Edge-correction, and the associated approximations, lead to higher standard errors near the boundary. Indeed, this should (approximately) be the only contributing factor for changes in the adaptive ASY s.e. surface; recall the asymptotic variance stabilisation uncovered in Section 2.2.2. This is reflected in the corresponding image and as such, the adaptive ASY variability is the best performer with the lowest maximum standard error and smallest span, followed by the fixed ASY s.e. surface. The empirical standard errors are far more unstable with sub-regions where little data have been observed understandably exhibiting higher values. Interestingly, the fixed MC s.e. surface does not reach the high maximum of the adaptive. However, it would appear that even for the relatively small sample sizes observed in this analysis, reduced variability when testing for significant risk fluctuations is still present when working with theoretical over empirical p-value surfaces.

In the corresponding article, we also visually explored the effects of combinations of halving and doubling the computed values for $h_{\mathcal{F}}$, h_0, $\tilde{h}_f$ and $\tilde{h}_g$. Generally speaking, the observed results from this informal 'sensitivity analysis' were similar to those for the unscaled interpretations discussed above. Competitive performance of ASY over MC in terms of variability was evident and, in adaptive smoothing, varying the global bandwidth had a greater impact on the resulting surfaces and tolerance contours than altering the pilots. Typically, we would expect heavier oversmoothing as opposed to heavier undersmoothing of

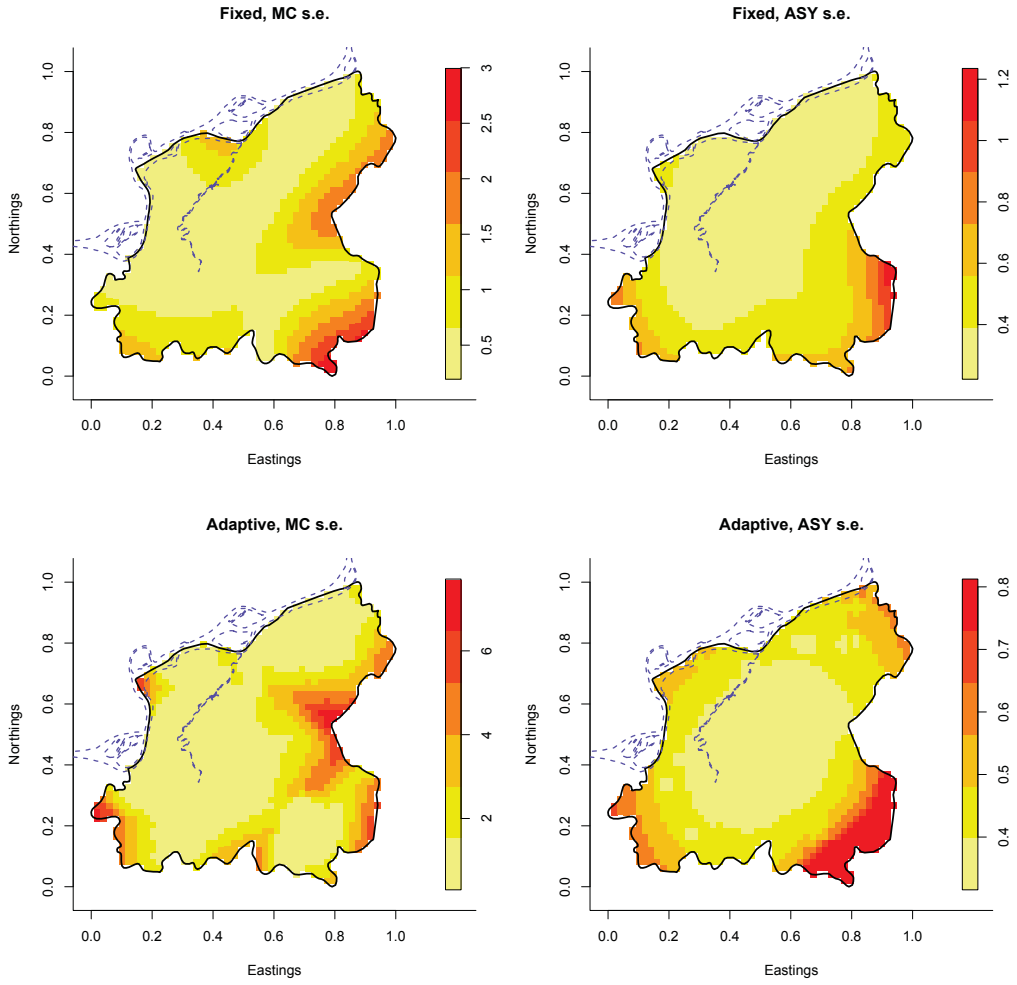


Figure 5.3: Estimated cell-wise standard error (s.e.) surfaces based on the estimated fixed and adaptive risk functions, computed via both ASY and MC methods.

the densities to be 'the lesser of two evils', and this was also observed; the excessive instability in estimates associated with the latter situation making for noisy and unpredictable behaviour in the resulting p-value surfaces.

5.1.4 Conclusions

A particularly pleasing example of Abramson's (1982) adaptive smoothing approach in the context of the kernel-estimated risk function, the asymptotically derived p-value surface and correspondingly plotted tolerance contours in Figure 5.2 clearly 'track' the Qiupu River within the Guichi region itself, where use of a fixed bandwidth leaves us unable to capture this detail in the surface. Were we (purely for the sake of argument) unaware of the importance of the proximity to natural waterways, this kind of result could be very useful. However, it seems the ASY method is more sensitive in terms of fluctuations in risk estimates and flagging significant sub-regions when implementing adaptive smoothing for small sample sizes. If the adaptive method is to be employed for applications where the sample sizes are considered small, it seems in future it would be prudent to consider using more stringent significance levels when implementing the asymptotic approach.

Nevertheless, the standard error surfaces indicate the ASY p-value surfaces to be based on a noticeably lower level of variability for both fixed and adaptive density-ratios when compared to MC. By definition, the MC method conditions upon the locations of cases and controls, which can (and in this example, has) allowed a high variance in sparsely populated areas. Coupled with the massive additional computing requirements of MC p-value surfaces, ASY is concluded as the preferred method for flagging significant sub-regions.

It is a known fact that contraction by humans of *S. japonica* in the marshland regions of China occurs when individuals are exposed to the parasite in affected water systems. This is reflected nicely in the analysis of the Guichi schistosomiasis data, with all four versions of the calculated 5% tolerance contours roughly agreeing on a northern area where the Qiupu meets the Yangtze of statistically significant increases in the risk of infection based on the kernel-smoothed density-ratio methodology. In addition, a smaller area in the southeastern quadrant may warrant further investigation.

5.2 Avian influenza – global trends

The geographical epidemiologist who initiated the schistosomiasis work described in the previous section, Z.J. Zhang, also voiced an interest in the adaptive kernel smoothing methodology to estimate the density function of globally collected avian influenza data, with a view to describing the recent spatial variation of a particular highly pathogenic subtype of the disease over several recent time periods.

This author contributed to actions and interpretations involving the kernel estimate, as well as a write-up of this method from a statistical perspective, included as supplementary material to the corresponding article. Other statistical evaluations, namely the implementation and interpretation of a spatiotemporal scan statistic, was performed by the collaborating authors. A relatively short, yet interesting paper – Zhang, Chen, Chen, Davies, Vaillancourt

and Liu (2012) – is published with *The Veterinary Journal*.

5.2.1 Background

Avian influenza, 'bird flu', is a disease affecting mainly poultry stocks, though certain subtypes have succeeded in crossing the human-animal species barrier. Subtype H5N1, a highly pathogenic strain of the influenza A virus, is one such example. In 1997, an outbreak in China infected the first humans; it was quickly established the handling of or close exposure to infected birds and their raw innards was a primary risk factor (World Health Organization, 2011). Since then, the disease has spread rapidly throughout Asia, Africa, and some parts of Europe, infecting many avian populations (commercial and wild) as well as several hundred humans.

The mortality rate in humans is high compared to other forms of influenza; early symptoms include high fever, diarrhoea, vomiting, abdominal pain, bleeding from the nose and gums, and bloody sputum production. Though H5N1 has only affected a relatively small number of people to date, major public health concerns arise due to the ever-present potential for a worldwide pandemic and mutation of the virus enabling human-to-human transmission.

This potential for severe detrimental effects on global economies and public heath dictates it is important that the space-time trends of infection in avian populations continue to be monitored alongside various control strategies employed to limit the impact of its presence; a statement particularly true when we note sporadic outbreaks continue to be reported around the world (Capua and Alexander, 2010). Previous examinations of this type have dealt mainly with short-term data up to around 2006; see for example the China-specific study by Oyana et al. (2006) and the worldwide study discussed in Si et al. (2009). Inspection of more recently recorded data is therefore clearly of interest, and one such data set is the focus of this analysis.

5.2.2 Data, methods, results

Data concerning the spatial locations of H5N1 outbreaks reported in domestic poultry over the period December 2003 to December 2009 were collected from the World Organisation for Animal Health (OIE) and the Food and Agriculture Organization of the United Nations (FAO). These were compiled into a single data set according to region sub-districts; see Zhang et al. (2010) for specifics. In this analysis, an 'outbreak' is simply defined following Toma et al. (1999) as "*the confirmed presence of disease, clinically expressed or not, in at least one individual in a defined location and during a specified period of time*".

The temporal period over which the data were observed was partitioned into six easily identifiable sub-periods, labelled *epidemic waves* (EW); EW1 covered the period December 2003 to June 2004; EW2 July 2004 to May 2005; EW3 June 2005 to October 2006; EW4 November 2006 to September 2007; EW5 October 2007 to October 2008; EW6 November 2008 to December 2009. The EWs are marked in plots (Figure 5.4 replicated directly from the article) showing the total counts of observed H5N1 outbreaks by day, week and month over the total temporal period. Clearly visible is the overall downturn in outbreaks, quite possibly a result of implemented disease containment and control procedures such as animal

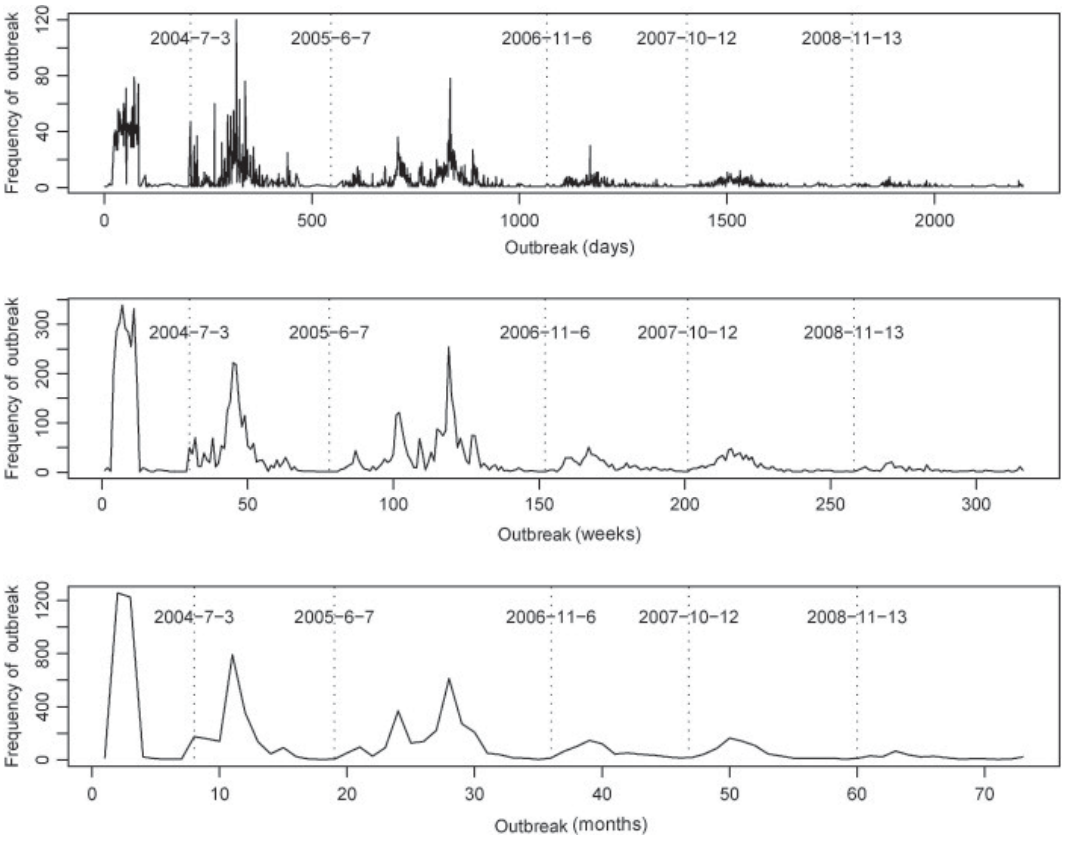


Figure 5.4: Daily, weekly, and monthly H5N1 outbreak counts, showing temporal partitioning of the six epidemic waves. Image reproduced with kind permission from Z.J. Zhang.

movement restrictions, quarantining, preemptive culling, and vaccination.

The 'overall' spatial distribution of H5N1 outbreaks within each EW was described with an adaptive kernel density estimate; supplementary details on the theoretical specifics were included with the online version of the corresponding article. This is a direct reflection of the methodology described in Chapter 2, most importantly equations (2.1), (2.3), and (2.22). In addition, likely clusters of observations were flagged using a space-time scan statistic, a straightforward extension of the well-known spatial scan statistic introduced by Martin Kulldorff; see Kulldorff (1997). Briefly, the spatiotemporal version (Kulldorff et al., 1998, 2005) as employed here uses a large number of overlapping, variably sized spatiotemporal windows – cylinders (base width controlling spatial window size, height controlling temporal window span) – and evaluates the observed and expected (i.e. were no clustering to exist) number of outbreak points falling within. From this information, a likelihood ratio score may be computed for each cylinder: the higher the score, the more likely we have observed a true cluster not occurring due to random variation. The spatiotemporal window which has maximum evaluated likelihood ratio statistic is elected as the 'most likely' ('primary') cluster; non-overlapping windows with smaller, yet still high, scores are identified as 'secondary' clusters. To evaluate the statistical significance of a given score, a large number of uniformly random permutations of the data are generated to coincide with a null hypothesis of 'no

clustering', conditional upon spatial location and temporal totals of the original observations. The scan statistic is employed on the synthetic data sets and the likelihood scores are recomputed, against which the ranks of the original scores are compared, yielding empirical p-values. Though, naturally, spatiotemporal inhomogeneity is expected for the H5N1 data, the scan statistic provides a useful embellishment on plots of the data of the location of the 'heaviest' clusters.

The evaluation of Kulldorff's space-time scan statistic was performed using the data within each EW; results superimposed upon maps of the kernel density estimates. There was little visible difference between use of the daily, weekly, and monthly versions of the data set; Figure 5.5 gives the EW maps for the weekly data (reproduced from the article). Outbreak locations are given as green dots. A greyscale pixel image provides the density estimate for each EW, the higher the density, the darker the shading (in some cases this was artificially enhanced for better visibility). Cylindrical windows deemed to contain the most likely clusters are denoted by discs annotated with the corresponding time span; red-solid for the primary; blue-dashed for secondaries.

We note in EW1 and EW2 the 'historic' density peak and associated clustering occurring in eastern/southeastern China; behaviour which persists throughout all EWs. It is understood that H5N1 is now endemic in native wild bird species, leaving the door open for sporadic infection of domestic poultry populations (Paul et al., 2010).

By EW3 however, the disease had progressed west, producing a second main density mode at the junction of Asia, Africa and eastern Europe; specific countries include those surrounding the Black and Mediterranean Seas. This 'novel mode' has previously gone unreported; likely a consequence of other analyses using older data sets. The space-time scan statistic continues to flag the primary cluster in the proximity of the historic mode, though secondary clusters are flagged and persist in the proximity of the novel mode from EW3. In terms of long-term trends, the specific sizes and locations of the highlighted scan statistic cylinders is less important than their detection shifting from the historic east to occur in the western areas of the map.

5.2.3 Concluding remarks

Monitoring of space-time trends of highly pathogenic avian influenza subtype H5N1 infections in poultry is critical to the appropriate fine-tuning of control and eradication strategies. Previous analyses have identified areas of interest in east and southeast Asia, and while these affected regions remain important, our investigation has indicated that the western junction of Africa and eastern Europe also warrants ongoing attention. This is additionally important when we consider the elements of the new environments in which the disease has become prevalent (e.g. new breeding grounds, altered climate, new avian species), increasing the opportunity for further mutations that could potentially spark a new pandemic.

Although the overall number of outbreaks has dropped significantly over time, it is the persistence of the virus in two major regions of the large study area which is the main concern. We conclude that the appropriate health officials in the affected and/or adjacent countries should continue to monitor the most recent information available and, where

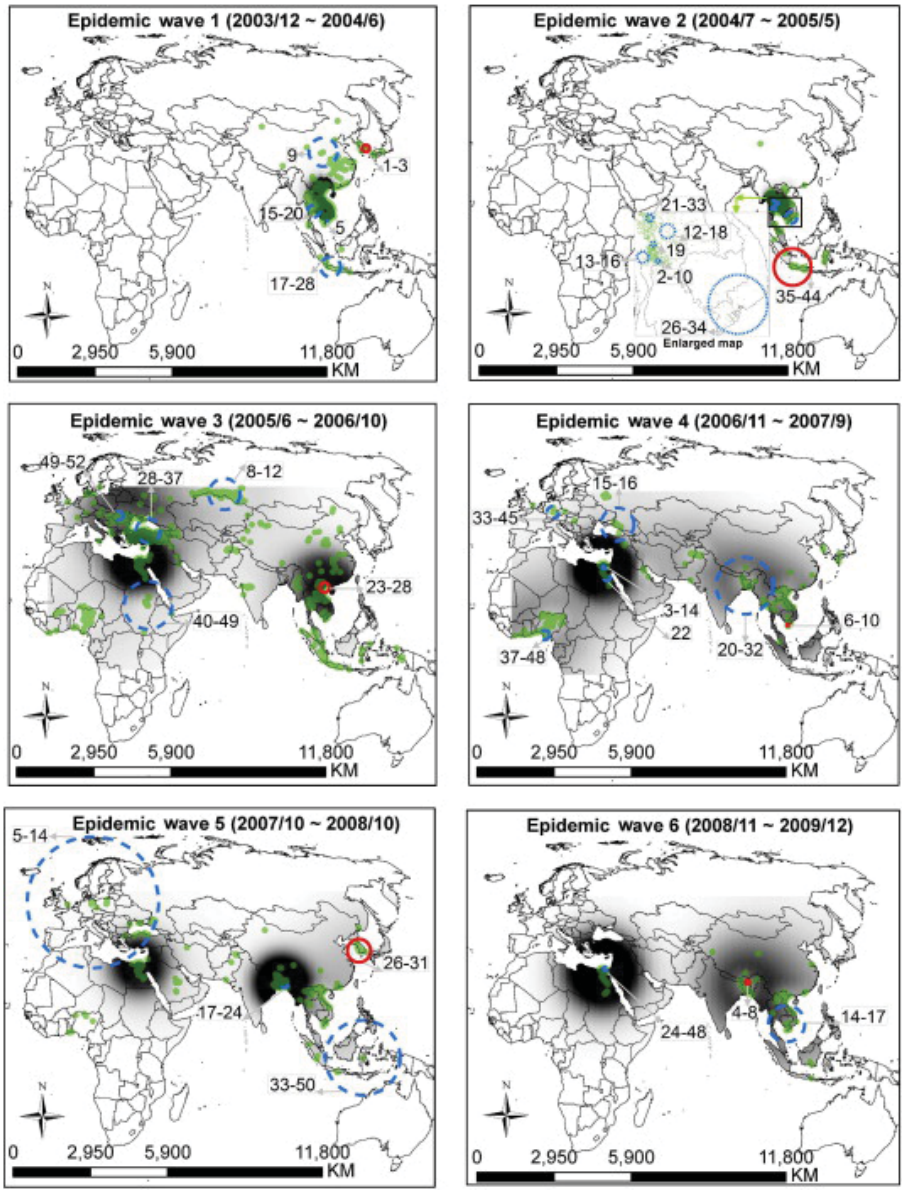


Figure 5.5: Spatial adaptive kernel density estimates of each epidemic wave for the week-based H5N1 outbreak data set, also indicating significant clusters as per Kulldorff's space-time permutation scan statistic. Image reproduced with kind permission from Z.J. Zhang.

possible, strengthen or refine existing control/containment systems in place should further outbreaks continue. Appropriate and thorough preparation for a 'worst-case scenario' remains key to effective minimisation of the detrimental impacts on public health and local economies resulting from a potential pandemic.

5.3 Foot-and-mouth disease model verification, Ireland

This section summarises the work in Sanson, Harvey, Garner, Stevenson, Davies, Hazelton, O'Connor, Dubé, Forde-Folle and Owen (2011), published in the *Revue Scientifique et Technique-Office International des Epizooties* (*OIE Scientific and Technical Review*) where three different disease control strategy evaluation models (what is meant by this is defined below) were used to examine the effectiveness of various control strategies for a hypothetical outbreak of foot-and-mouth disease and corresponding epidemic across a number of different types of farms in Ireland. This author, along with primary thesis advisor M.L. Hazelton, contributed to the statistical analysis concerning identification of the spatial areas within the study region which represented a heightened risk of infection for each evaluation model and control strategy. Like the case studies in Sections 5.1 and 5.2, these activities drew on advancements in Chapter 2, and made use of **sparr** (Section 4.1). The contribution also included a write-up of the statistical methods included in the article itself.

5.3.1 Project summary

In both previous case studies, references have been made to various types of 'control strategies' used by intervening health authorities of the affected regions to limit the spread of a given disease post-detection. The design of such systems typically involves expertise from various disciplines and is often based upon knowledge of previous epidemics. The quality of design, however, is just one of several important aspects concerning these strategies – it is critical that they can perform well (i.e. significantly hinder, halt, or eradicate the disease and its spread) should they in fact be needed in practice.

Determining this is a rather difficult goal, however. The best way to assess the performance of control strategies in this way is their implementation with respect to a real epidemic! In the (thankful) absence of such an event, researchers turn to *disease control strategy evaluation models* (shortened to *disease models*), which are used to evaluate and compare different kinds of control strategies via realistic depictions of disease spread, with the ultimate objective of increasing end-user confidence as to their effectiveness in real-world scenarios (as well as confidence in the evaluation models as the corresponding assessment tools). This is particularly useful when certain control strategies have not been used in practice before, or are expensive to implement or controversial in some other way. Frustratingly, however, this leads to another level of uncertainty – where different control strategy evaluation models exist, we would naturally expect some variation in their results concerning the effectiveness of the control strategies themselves. As a simple hypothetical, evaluation model 'A' might suggest control strategy '1' as the most effective in limiting disease spread in a given scenario, but evaluation model 'B' might prefer, for the same scenario, control

strategy '2'.

This particular work made up the final step of a three-stage project of evaluation model verification and validation with respect to epidemics of veterinary foot-and-mouth disease (FMD). In the epidemiological terminology, *verification* is the term used in reference to ensuring the explicit behaviour of the disease transmission has been accurately and succinctly represented via the creation of relevant computer programs; *validation* is the procedure of ensuring the disease model caters for, reflects and/or depicts realistic disease process conditions (see Taylor, 2003).

Three different FMD models in use by the Quadrilateral Group of countries (QUADs: Australia, New Zealand, Canada and the United States) were examined: AusSpread (Garner and Beckett, 2005, Australia), InterSpread Plus (Sanson, 1993, New Zealand) and the North American Animal Disease Spread Model (NAADSM) (Harvey et al., 2007, Canada; USA). The first two stages of the project, detailed in Dubé et al. (2007), covered initial verification efforts of the three disease models and compared their predictions based on some relatively simple synthetic scenarios (termed *relative validation*). The third stage, detailed here, deals with relative validation of the models using carefully designed, and subsequently controlled, hypothesised epidemics of FMD in an approximate circular region in central Ireland encapsulating 50,125 farms. The design of these sophisticated outbreak scenarios, in contrast to the earlier study, used real world data provided by the Irish collaborators concerning the farms at (potential) risk, including their actual spatial locations, animal types and counts, livestock movements, marketing specifics, and even local atmospheric conditions to describe chances of airborne infection.

5.3.2 Study design specifics

AusSpread, InterSpread Plus, and NAADSM were employed to gauge the spread of an outbreak of FMD based on a 'primary' infection at a farm in roughly the middle of the study region, followed by a 16-day 'silent spread' period during which 41 farms became infected without detection. The first farm detected fell on day 17; at which point four different control strategies were implemented.

Replicated from the accompanying article, Figure 5.6 shows this initial behaviour within Ireland, upon which the approximate study region has been superimposed post-copy.

Briefly, the four control strategies employed were: (a) Standard European Union (EU) stamping-out policy (this involves depopulation and livestock movement limitations); (b) EU stamping-out policy plus a contiguous cull (0.5km radius) around infected premises; (c) EU stamping-out policy plus 3km suppressive ring vaccination (8-day delay); and (d) EU stamping-out policy plus 3km suppressive ring vaccination (21-day delay; delays in vaccination are common in practice due to logistical and planning-related issues). In large-scale epidemic control, a 'contiguous cull' refers to the somewhat controversial preemptive measure of the killing of apparently healthy livestock; 'ring vaccination' is typically far more expensive than a contiguous cull and refers to total vaccination of uninfected livestock within a specified distance from a given infected premises.

In total, five outbreak scenarios were defined. Scenario 1 involved no control strategies

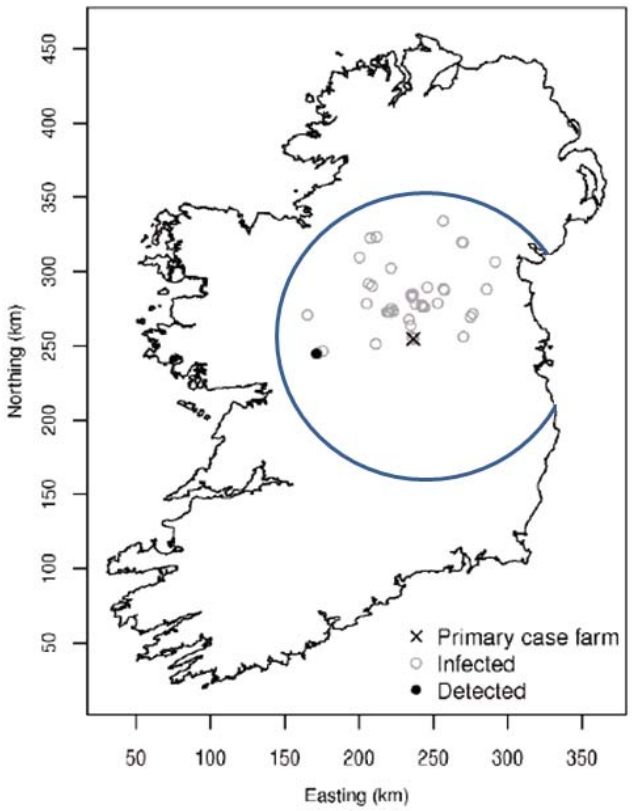


Figure 5.6: Map of Ireland, giving the primary infection location (×), silently infected premises (grey ∘), and point of first detection (•), as well as the approximate boundary of the circular study region. Image ©2011 OIE publications; reproduced with permission.

and the epidemic continued unabated. Scenarios 2-5 initially followed the infection-detection behaviour illustrated in Figure 5.6, at which point the four control strategies (a) to (d) were implemented respectively.

To obtain results for these scenarios by using the detailed real-world information mentioned above, the Australian, New Zealand, and North American disease control strategy evaluation models were instructed to generate and analyse each of 40 iterations of hypothetical outbreaks of FMD amongst the 50,125 farms in the defined study region. Monitoring/prediction took place from the date of implementation to the 60-day mark (Scenario 1), or eradication (Scenarios 2-5). 'Eradication' refers to the point at which the model suggested no new infections would/could subsequently occur.

5.3.3 Analysis and results

Relative validation of the three models was performed with respect to their prediction of various quantities associated with the generated outbreaks and their corresponding spatial spread as time progressed. For a given generated FMD epidemic, these quantities included the total predicted number of infected premises, the total predicted duration of the epidemic in days, and the area of the spatial sub-region affected (computed via construction of a convex hull), at eradication.

For each Scenario 2-5, the median value for each of these quantities was compared between the three disease models using the Kruskal-Wallis one-way analysis of variance (by ranks) test. Statistically significant results were encountered for the median predicted infected premises counts and epidemic duration, but not for the predicted median area values. Figure 5.7 replicates the boxplots corresponding to the counts of infected premises at eradication for each scenario and disease model over the 40 iterations.

We note it is the North American model which is primarily influencing the significant within-scenario, between-model difference result in this case. Variations in the specifically programmed models are clearly responsible for such differences. For example, NAADSM only selected non-detected farms to receive livestock movements from infected but not yet detected farms, thereby increasing the chances of pass-on infection. This is in contrast to the Australian and New Zealand models, in which both infected *and* detected farms could still 'legally' receive these movements. More importantly for our validation exercise, however, the relative within-model between-scenario differences remained similar; trends repeated for the other recorded quantities of epidemic duration and area value. Scenario 4 (use of control strategy (c)), appeared to provide the most promising absolute results according to all three disease models.

This author's specific contributions proceeded as follows. In addition to the above numerical analyses, the novel asymptotic p-value surfaces and corresponding tolerance contours for kernel-smoothed relative risk functions described in Chapter 2 (Section 2.4) were employed to discern the actual spatial positions of the statistically significant areas of increased risk of FMD infection for the Australian, New Zealand, and North American models. Via this methodology, the 'risk' is defined as the (log-transformed) ratio of 'case' to 'non-case' farms; the former defined as those in a given iteration for a given scenario and disease model that

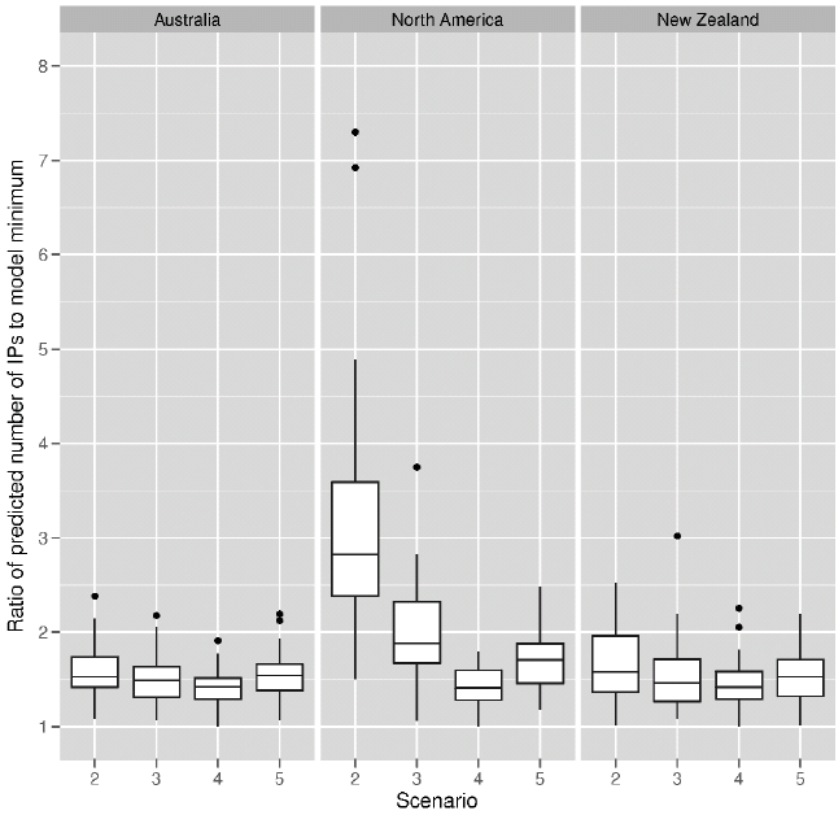


Figure 5.7: Boxplots displaying the predicted counts of infected premises at eradication relative to the minimum count for each scenario and disease model. Image ©2011 OIE publications; reproduced with permission.

were predicted to become infected, the latter those that were not. The relatively high sample size counts dictated use of the adaptive smoothing regimen was not practically feasible; the fixed version was therefore employed in all situations (Gaussian kernel function, common bandwidth of 9.3km). The p-value surfaces were averaged over each of the iterations; corresponding tolerance contours were plotted upon the study region at the 5% significance level.

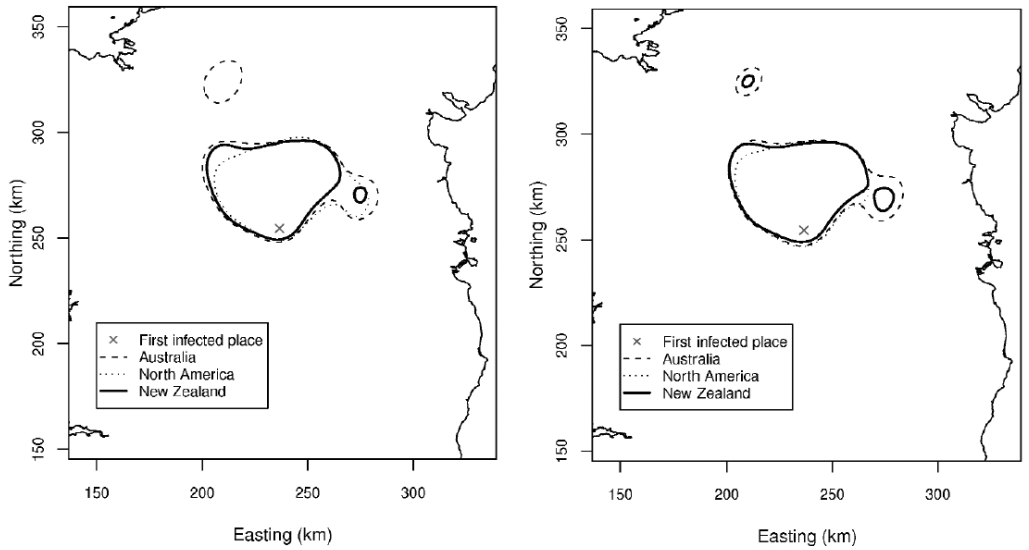
Figure 5.8 shows the averaged tolerance contours for each scenario and disease model. We note that the three models are fairly consistent in their predicted 'high-risk' sub-regions, with all models in agreement in the approximate size, shape and positioning of a central hotspot encapsulating the location of the primary infection. In Scenarios 3 and 5 (Figures 5.8(b) and 5.8(d) respectively) the Australian and New Zealand models flagged an additional, smaller hotspot to the north; only detected in Scenario 2 (Figure 5.8(a)) by the Australian model. All three models detected this additional hotspot in Scenario 4 (Figure 5.8(c)).

5.3.4 Discussion

The efforts involved in the design, implementation and testing of disease control strategy evaluation models are considerable. A multidisciplinary task, teams involved typically include epidemiologists, statisticians, physicians, and policy-makers. Detailed data is required on various aspects of the disease, population(s) at risk, and local geographical features. The current examination had to overcome several challenges at both the programming and simulation stages prior to the collection and dissemination of results (see the corresponding article for more on this issue).

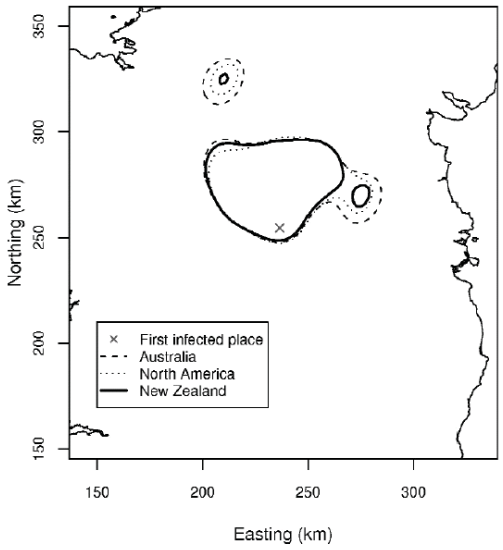
This work designed complex, realistic scenarios and simulated outbreaks of FMD amongst a large number of farms of several different types in central Ireland. Various disease control strategies intervened and, using detailed real-world data, the three disease models AusSpread, InterSpread Plus and NAADSM, were employed to predict the spread of the epidemic through space as time progressed. Although some differences were detected when comparing measures such as the number of infected premises between models, within scenarios, the relative within-model, between-scenario differences were consistent. This relative validation exercise has therefore highlighted the fact that policy and decision-makers would be likely to arrive at the same conclusions as to the efficacy of the various control strategies tested, regardless of the choice of disease model implemented to predict the spread of FMD in the event of a real epidemic in the region. High-risk regions of infection as determined through use of asymptotic tolerance contours for the fixed-bandwidth kernel-smoothed risk function were also identified as almost identical for each of AusSpread, InterSpread Plus and NAADSM.

We therefore conclude that appropriate modelling assumptions appear to have been made at the programming stage of all three models. Perhaps more importantly, the similar patterns observed in the results of various features of the iterated epidemics, obtained from the different disease models when employed as predictive tools, serve to substantially increase end-user confidence in their practical performance.

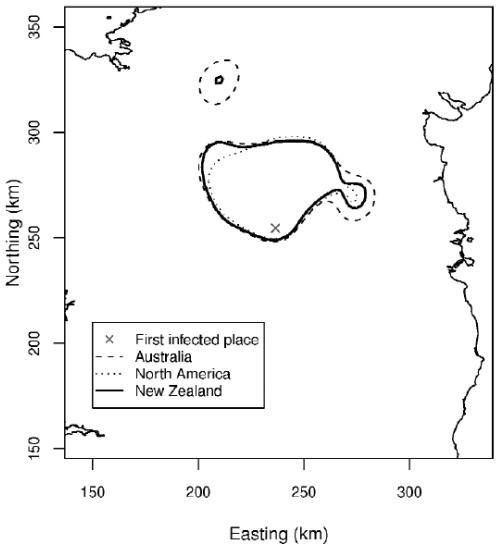


(a) Scenario 2

(b) Scenario 3



(c) Scenario 4



(d) Scenario 5

Figure 5.8: Averaged 5% asymptotic tolerance contours delineating heightened risk of infection of farms by FMD in Ireland, based on fixed-bandwidth kernel-smoothed risk surfaces, for each scenario and disease model. Images ©2011 OIE publications; reproduced with permission.

5.4 Spatiotemporal mapping of Japanese earthquakes, 1926–2005: a preliminary exploration

To conclude this set of case-studies applying the methods and novel material described in the previous chapters, we present an implementation of the type of spatiotemporal log-Gaussian Cox process introduced by Brix and Diggle (2001) for estimation and prediction of the space-time intensity of earthquakes over central Japan. While not an epidemiological application, we nevertheless proceed in an effort to showcase the flexibility of the LGCP and motivate further practical examinations.

Earthquakes can easily be described as one of the most devastating natural events to affect human populations. Far from just an immediate threat to safety at the time of occurrence, subsequent impacts on urban infrastructure, for example, can continue to long-lasting economic consequences. Recent examples include the earthquakes centered near Christchurch, New Zealand (2010, 2011), and off the northeastern coast of Japan (2011), which have resulted in widespread destruction and loss of life. Where possible, spatiotemporal monitoring is therefore clearly an important exercise in an effort to best understand the areas of a particular region which are most prone to experiencing these events.

A selection of global earthquake catalogues exist, which log comprehensive data on the events. In spatial and spatiotemporal earthquake applications, event locations are typically provided as the epicenter coordinates, and associated data such as depth and magnitude are often fundamental components of relevant statistical models. Recording these data in a reliable fashion is itself a complex task, see for example the discussion concerning the Global Centroid Moment Tensor (GCMT) catalogue in Ekström et al. (2005).

Like applications in geographical epidemiology, statistical methods for point processes and the spatial (and spatiotemporal) modelling of earthquakes are closely linked. Special attention is typically given to the effect of a given earthquake precipitating further events; we will briefly describe two important approaches to this kind of analysis. A general discussion of statistical techniques used in earthquake modelling can be found in e.g. Vere-Jones et al. (2005).

Particularly well-known is the *epidemic type aftershock sequence* (ETAS) model of Ogata (1988, 1998). This is an adaptation of a so-called space-time *Hawkes process* (Hawkes, 1971), which the author describes as a 'self-exciting' process where the occurrence of events increases the probability of further events. Parallels can be drawn with the shot-noise Cox process discussed briefly in Section 3.1.4, as an event in the space-time continuum can become a 'mother' point associated with 'offspring' densities. Aiming to produce event-intensity surfaces, Ogata's model takes into account earthquake magnitudes as both an impacting variable on potential offspring events, as well as specification of a 'magnitude distribution' providing the capability to model the magnitude of the offspring earthquakes themselves. A good summary of the ETAS model can be found in Vere-Jones (2009), who also notes that statistically significant departures from an ETAS fit are often used as diagnostic tools in identification of anomalous spatiotemporal localisations of earthquakes.

Another important class of statistical methods for earthquake trend analyses in space and time are known as *stress-release models*. In their simplest form, these models aim to

produce earthquake intensity estimates within a given seismic region which are a function of time and some unobserved (or partially observed) stress variable. When an earthquake occurs it induces a 'stress drop' in the model, heightening the chance of further events. A *linked* stress-release model is obtained when we consider the simple version over a finite number of contiguous, disjoint spatial regions; the models incorporate the possibility of 'stress transfers' between the regions. Like the ETAS model, linked stress-release models can be used to estimate spatial earthquake intensity functions (within a certain geographical area) as time progresses. In one sense, the consideration of the spatial intensity over a larger geographical region as a finite collection of smaller, disjoint spatial areas is similar to the discretisation that must be carried out in order to practically implement the LGCP. Further details on these models can be found in e.g. Bebbington and Harte (2001) and Daley and Vere-Jones (2003).

Attempting to model space-time earthquake intensities via the LGCP nevertheless remains a rather different approach to this kind of problem. As such, it is emphasised that what follows is labelled a short 'preliminary exploration'; our perspective in the following exposé is taken from the point process theory covered in the preceding chapters which has had an epidemiological slant. A more careful and detailed investigation, preferably in collaboration with appropriate geoscientists, is called for in order to draw proper and contextually valid conclusions based on any analysis of these data. Our focus remains on gauging the suitability of the LGCP for such an application in terms of the performance of fitting procedures such as parameter estimation, as well the appraisal of Markov-chain Monte-Carlo diagnostics obtained from experimentation with conditional simulation.

5.4.1 Data and analysis proposals

The data for this examination comprise the spatial locations of the epicenters of 98 earthquakes with a detected magnitude of greater than or equal to 6 on the Richter scale in a rectangular window over central Japan, and associated timestamps over the 80-year period 1926-2005 inclusive. Briefly summarised in Section 6.10.3 of Illian et al. (2008), the Japan Meteorological Agency is acknowledged for the availability of these data at `http://users.jyu.fi/~penttine/ppstatistics/data/Earthquakes_Fig_6_21.txt` (site hosted by Professor Antti Penttinen, University of Jyväskylä, Finland).

Figure 5.9 displays the spatial and temporal margins of the data, the former following rescaling to latitude-longitude coordinates in degrees, the latter giving annual total counts.

Obvious spatial clustering is noted in the observations, a natural feature of earthquake data. We would expect certain subregions to receive more events than others according to e.g. the underlying locations of fault lines, and so the underlying 'global' spatial variation of events in the study region is likely to be non-uniform. In addition, any given earthquake may provoke others of varying magnitude in the immediate vicinity (a key component of e.g. the ETAS and stress-release models discussed above). It could therefore be reasonably expected that smaller-scale 'local' positive correlation may be present, in addition to the aforementioned natural global inhomogeneity.

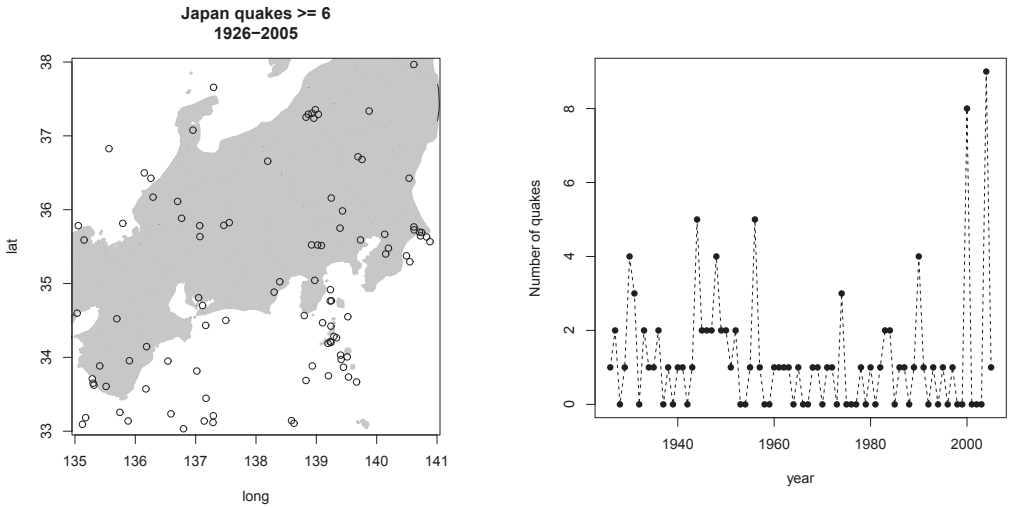


Figure 5.9: Spatial (left) and temporal (right) margins of the Japanese earthquake data. Spatial positioning of Japan's central land mass (grey shading) is visible in the former plot.

Temporally, it is known that the underlying global trend in terms of the frequency of events takes many hundred or even thousands of years to alter, and so this trend over the time span for which we have data may be considered constant (viewing, as per the spatial point process theory discussed in Chapter 1, the actual realised counts as being Poisson-distributed – see also the relevant comments in Section 7 of Vere-Jones, 2009). This view appears to be supported by the plot on the right-hand side of Figure 5.9, in turn suggesting that the heterogeneous global *spatial* trend can be considered fixed over the given time period. Nevertheless, relatively short-term local positive correlation may easily be present in the raw temporal frequencies.

From these comments on the nature of our observed data, it would seem that the log-Gaussian Cox process could be an appropriate framework with which to estimate and interpret the spatiotemporal intensity function $\psi_t(\boldsymbol{x})$ (yielding the expected number of events at time t and location $\boldsymbol{x}$) of this relatively recent period of earthquakes over central Japan. As per the model formulation and notation in Chapter 3, equation (3.18), and the above conjectures, our goal here is to therefore estimate and perform subsequent inference on $\psi_t(\boldsymbol{x}) = \zeta_\psi(\boldsymbol{x})\eta \exp[Z_t(\boldsymbol{x})]$, where ζ_ψ is the fixed heterogeneous spatial trend and η is the constant-in-time expected number of events associated with each timestamp (year). Recall also that the quantity Z is a stationary, isotropic spatiotemporal Gaussian field with variance σ^2_ψ, mean $\mu_\psi = -0.5\sigma^2_\psi$, and separable spatial and temporal correlation functions r and τ respectively. This component is used in the attempt to capture the natural 'residual' space-time correlation between earthquake occurrences. Appropriate scaling for each of ζ_ψ, η and $\exp[Z_t(\boldsymbol{x})]$ is imposed so as to achieve the identifiability constraints imposed in Section 3.1.3.

The steps in the following section are similar to those taken in the analysis of the AEGISS gastrointestinal infection data in Sections 3.4.2 and 4.2.2.

5.4.2 Model specification, parameter estimation

The fixed-through-time, heterogeneous global spatial trend ζ_ψ is modelled with a two-dimensional fixed-bandwidth kernel density estimate; the smoothing parameter set using the oversmoothing rule (refer back to Section 2.3.1). Note that, based on the results from simulation Study 2 in Section 3.5.2, a generous amount of smoothing for kernel estimation of the deterministic spatial component can aid in producing better estimates (i.e. estimates closer to the true values under the assumed model) of the parameters controlling the stochastic correlation (performed below). The constant expected number of events per time point, η, is set to $98/80 = 1.225$.

Figure 5.10(a) plots our deterministic spatial component ζ_ψ; Figure 5.10(b) superimposes the assumed value of η upon a plot of the temporal frequencies for the study period 1926-2005. We note a satisfactory representation for both margins in terms of their respective underlying global trends under the assumed modelling structure.

Next, we define the nature of the correlation structure. For simplicity, and in the absence of additional information, exponential functional forms are selected (just as in the earlier analysis of the AEGISS data). Recapping, this dictates that the spatial *covariance* function is expressed as $\sigma^2_\psi \exp[-u/\phi_\psi]$, and that the temporal correlation is simply $\exp[-\nu/\theta]$, for scalar spatial and temporal lags u and ν respectively. The parameters controlling the strength of this positive correlation, ϕ_ψ and θ, as well as the constant variance of the stationary and isotropic Gaussian field Z, σ^2_ψ, thus require estimation.

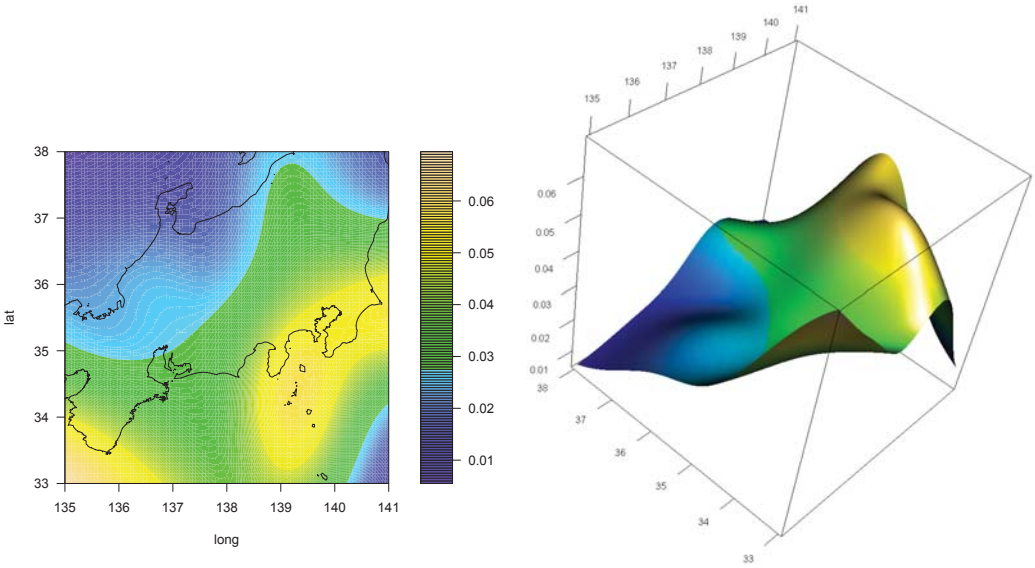
This is performed using the minimum contrast techniques used throughout Chapter 3, by matching as closely as possible the appropriate theoretical and empirical descriptors of the second-order properties of the process. The 'spatial' parameters σ^2_ψ and ϕ_ψ must be estimated first; this is achieved through use of either the time-averaged K-function or the time-averaged pair correlation function (g). Dependent upon the estimated values of σ^2_ψ and ϕ_ψ, estimation of the temporal correlation parameter subsequently takes place using the temporal covariance function $\mathcal{C}$. Refer back to Section 3.1 for details on the explicit parametric and non-parametric forms of these functions. Minimum contrast proceeds using the same calibration decisions (such as function transformation and upper contrast limits) made and justified in Section 3.2.3.

Figure 5.11 gives the transformed non-parametric K, g, as well as $\mathcal{C}$, along with the theoretical curve corresponding to the optimised parameter combination of σ^2_ψ and ϕ_ψ for the former two, and θ for the latter.

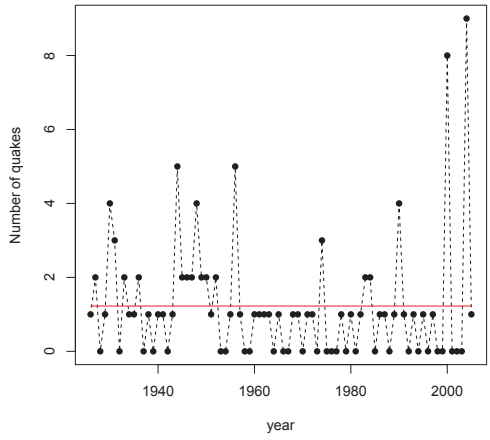
Both K and g provide similar estimates of the spatial parameters, with the parametric versions of the curves closely following the nonparametric trends. For this example, we simply set the corresponding parameter values to the mean of their K- and pair correlation function estimates (as in the Humberside analysis; cf. Section 3.4.1). That is, $\sigma^2_\psi = 4.57$ and $\phi_\psi = 0.31$, leading to the optimal $\theta = 2.16$.

5.4.3 Concurrent prediction and future-forecasting

To estimate the form of the spatial intensity at any given time point, we must execute a spatiotemporal Metropolis-adjusted Langevin algorithm to sample from the target distribution



(a) A heatplot (left) and 3D perspective plot (right) of the assumed global spatial trend (fixed-bandwidth kernel density).



(b) Constant-in-time value (red line) giving the expected number of events per year, against the raw observed counts.

Figure 5.10: Assumed spatial and temporal global trends for the Japanese earthquake LGCP during the period 1926-2005.

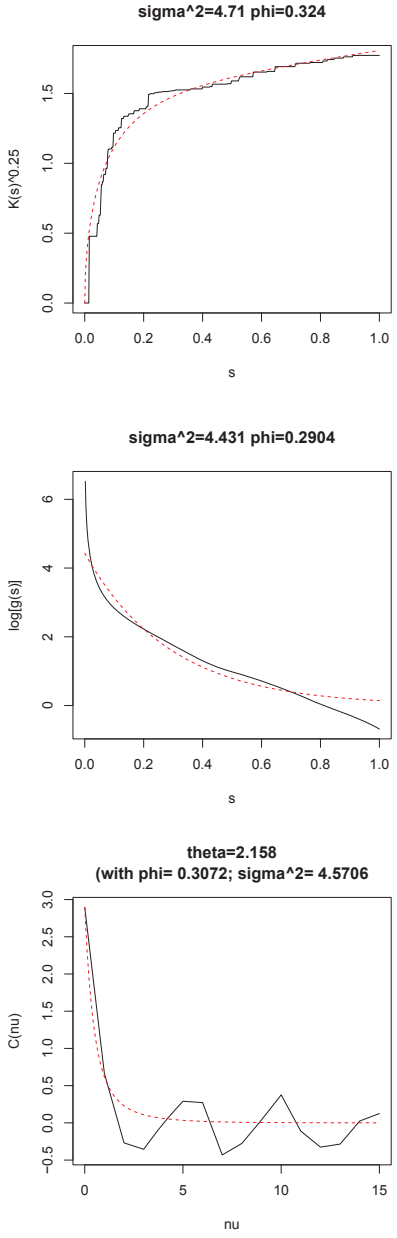


Figure 5.11: Minimum contrast estimation of correlation parameters for the Japanese earthquake data over the period 1926-1965. Top: transformed K-function; center: transformed pair correlation function; bottom: temporal covariance function. Nonparametric versions are given a solid line, optimal parametric curves a dashed line.

of interest, thereby predicting the residual LGCP. Here we must be careful to distinguish between 'concurrent' prediction and 'future-forecasting'; the former producing estimates of the intensity based on observed data *up to and including* the time point of interest, the latter based only on data *up to* the time point of interest (for 'one-step-ahead' prediction). Forecasting potentially high-risk areas of future earthquakes is clearly an important, yet extremely difficult, aspect of earthquake modelling.

We will consider both concurrent prediction and a simple form of one-step-ahead future-forecasting for our data; interest focused on final year ($T = 2005$). A MALA is run (details follow), from which we can straightforwardly extract the concurrent prediction of the spatial intensity of earthquakes greater than or equal to magnitude 6 for the year 2005. Then, assuming 2005 has not yet occurred i.e. it is not possible to yet have observations for this hypothetical 'future' time-point, we extract the (retained) realisations from the MALA run corresponding to the previous time point, 2004. For each of the retained realisations, we randomly generate a corresponding one-step-ahead Gaussian field based on the recursive definition of the spatiotemporal LGCP (see equation (3.50) in Section 3.3.2). The same summary statistics as can be computed for the concurrent version can thus be evaluated for the 'future' version of predictions, and the entire forecasting operation can be repeated any number of times.

The MALA is run on a 64×64-celled spatial grid for 120000 iterations with a burn-in period of 20000 iterations, retaining every 100^{th} realisation thereafter. From the specified strength of the temporal correlation, the algorithm is instructed to only make use of data from a maximum of 5 previous time points to minimise computation and storage costs. The tuning parameter is set at $h = 0.007$, achieving a satisfactory acceptance rate in view of the recommendations in Roberts and Rosenthal (1998). Total execution time was around 10 hours.

A trace plot of the value of the Gaussian field Z_T (concurrent prediction) at three randomly selected cells is given in Figure 5.12, along with an autocorrelation plot of the values from the first cell. This indicates convergence has been achieved, albeit with some mild correlation between successive realisations, with a stabilised chain and good mixing. This is particularly encouraging given the relatively small sample sizes over time.

The concurrent prediction of the spatial intensity for the final time point is displayed in Figure 5.13(a). This is given as the posterior mean of the retained exponentiated Gaussian fields (i.e. the estimated residual LGCP), as well as the Poisson intensity itself (found by the product of the deterministic global spatial and temporal components with the estimated residual LGCP). The hypothetical future-forecast version of these quantities for the same time point is given in Figure 5.13(b). These results are based 100 repetitions of the empirical forecasting procedure producing posterior field means for 2005; the grand average of these posterior means is subsequently taken.

Both the concurrent and future predictions have highlighted sharp peaks in the intensity surface in the northern and southern areas of the study region. These are the locations which have 'most recently' (and in the case of the concurrent prediction, 'presently') experienced events and, using the assumed form of the spatiotemporal correlation, are where the model deems the highest chances of observing further events exist. Note that the level of this 'risk'

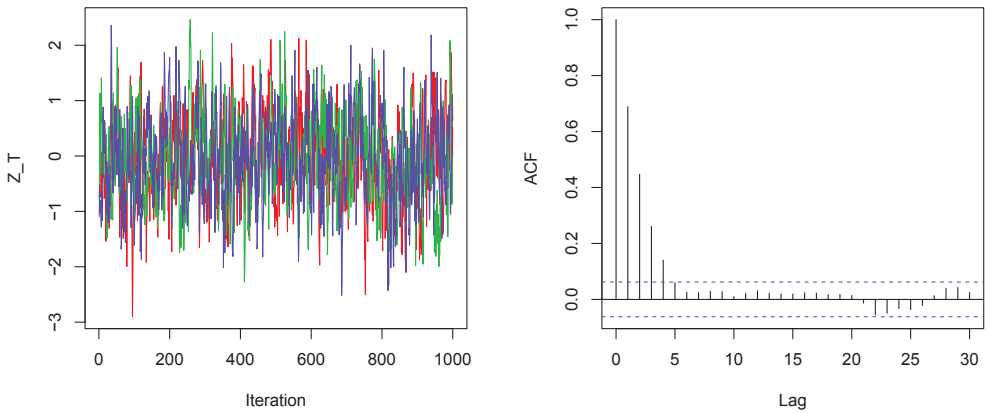


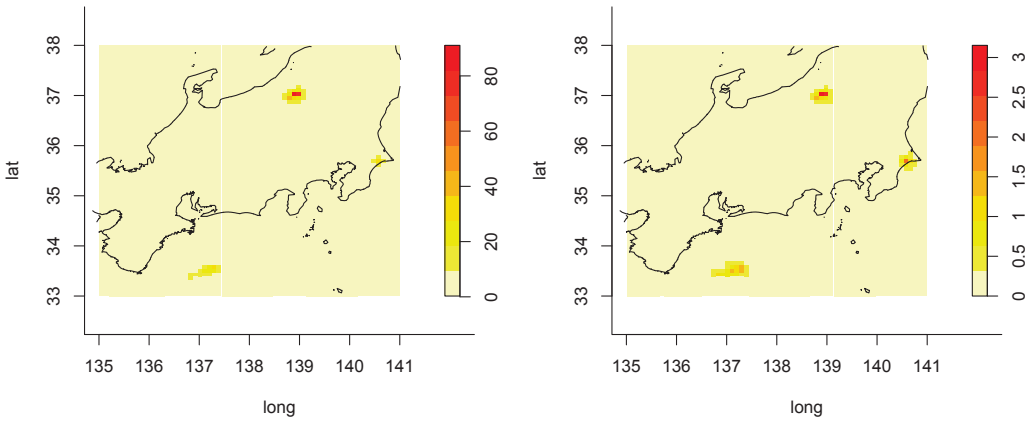
Figure 5.12: Trace plot of the value of the Gaussian field at three random cell-locations: 'red', 'green' and 'blue' (left), and an autocorrelation plot for the 'blue' cell (right).

is far higher in the future-forecast version of 2005 – this can be explained by examining the temporal variation in counts in Figure 5.10(b). Though there were a relatively high number of earthquakes in 2004 (9), in 2005 there occurred only 1. The concurrent prediction has taken this into account, but the future-forecast intensity has no such luxury: its prediction for 2005 has the strongest temporal dependence on the most recent time point (2004), and has therefore assumed a higher number of events were likely in the following year. The spatial location of the single event that did occur in 2005 is visible in the plots of the concurrently predicted intensity; this did not occur in either of the two 'historical' hotspots and a mild peak can be seen on the eastern edge of the region. Without the observed data, the future-forecast intensity was naturally unable to properly capture the magnitude of this third peak, an issue which clearly highlights the dangers of l-step-ahead forecasting for this model, particularly as l increases beyond 1.

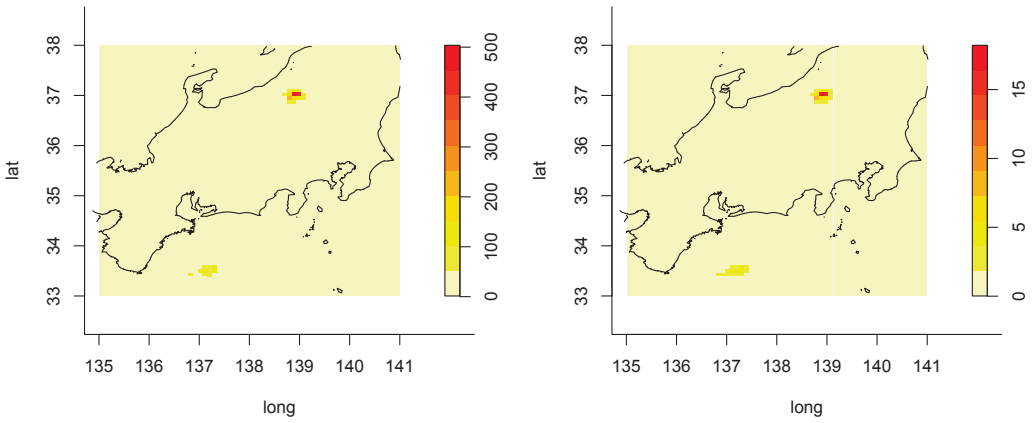
5.4.4 Closing comments, critiques

Originally motivated by applications in disease surveillance, we have investigated the feasibility of applying the spatiotemporal log-Gaussian Cox process to a data set comprised of earthquakes greater than or equal to a magnitude of 6 over central Japan. The notions of 'global' variation and 'local' correlation can to a certain extent be considered sensible for earthquake data, and are straightforwardly incorporated into the aforementioned framework using deterministic and stochastic components.

From a purely methodological standpoint, fitting the model posed little reason for concern with the deterministic global components easily specified. Perhaps more importantly, the minimum contrast methods used to estimate the strength of the assumed dependence structure performed well. Both K and g produced similar estimates of the spatial correlation parameters, producing theoretical curves which satisfactorily matched the trends exhibited by their nonparametric counterparts. The estimated strength of the temporal correlation,



(a) Concurrent prediction (i.e. using 2005 data): posterior mean residual LGCP (left) and Poisson intensity (right).



(b) Future-forecasting (i.e. assuming no data yet available): empirical mean residual LGCP (left) and Poisson intensity (right).

Figure 5.13: Estimating the spatial intensity for the Japanese earthquake data (2005).

derived in part using the estimates of the spatial correlation parameters, also reflected a sensible parametric versus nonparametric fit.

In terms of conditional simulation, the MALA appeared to behave well for these data. Diagnostic tracers indicated successful convergence and good mixing; any chain autocorrelation can be easily controlled via appropriate thinning. The concurrent prediction for 2005 reflected the most recent and 'current' data observed, the residual LGCP driving a large part of the final appearance of the intensity function. Treating 2005 as a future time point, the forecasted intensity heavily reflected the spatiotemporal variation and correlation inherent in the most recent time points. As such, it both overestimated the magnitude of the concurrently predicted intensity and failed to produce a (smaller) peak where an earthquake did in fact occur. Nevertheless, these results indicated the described forecasting procedure functioned as one would expect for the LGCP; an assessment of its practical performance in other spatiotemporal modelling problems may indeed prove worthwhile.

The described approach could easily be refined if need be. One possible extension would be to consider using covariate information, such as depth and precise magnitude of the events, to help describe the dependence structure. A collaborative effort with experienced geoscientists is warranted for the most effective implementation of this modelling framework with respect to the current problem, and vital to drawing valid contextual conclusions. For example, to compare the LGCP framework with an ETAS or stress-release modelling strategy would likely prove a challenging yet interesting research project.

Chapter 6

Discussion

An examination of the spatial and spatiotemporal trends inherent in disease prevalence and incidence within a given population is the key focus of geographical epidemiologists. This information can yield important insight into the location of the most heavily affected geographical regions, subsequently enabling the most economically viable and effective targeting of public-health-related control strategies. The additional detail associated with case-event or point-location data, as opposed to aggregated sub-region data, is of obvious value in the statistical modelling of these kinds of problems. This in turn has seen an increase in recent years of the availability of point-location data with respect to epidemiological applications, as researchers attempt to fully understand the space-time trends of the disease(s) of interest.

It is therefore the responsibility of biostatisticians to ensure the available (2-dimensional) point process theory on which most, if not all, of the associated statistical methods (employed in the data analysis stage of a point-location epidemiological study) are based, continues to evolve with aforementioned goals in mind. It is additionally important that the more technically challenging techniques are practically applicable. Their complexity often warrants collaborative efforts between applied researchers and statisticians, so the methods employed can interpreted correctly in view of the problem at hand. This inherent need for a literal translation from statistical theory to valid practical implementation is critical, and so the biostatistician's responsibility is not just the *development* or *refinement* of new/existing methodology, but also the *translation* of said methodologies to applied, real-world problems and the *appraisal* of their performance in practice.

It is these italicised issues which have been the focus of this thesis. This chapter concludes the efforts made to address them by way of a review of the completed work and discussion of the most interesting findings, as well as problems that have future research potential. Section 6.1 reiterates the three core thesis **Aims** and details the progress made; novel contributions best illustrated by the published and submitted peer-reviewed papers associated with the preceding chapters. The dissertation closes in Section 6.2, leaving the reader with a taste of things to come.

6.1 Summary and accomplishments

In general terms, the primary goal of this thesis was to review, appraise, and improve where possible, statistical methods for analysing 2-dimensional point processes from the point of view of epidemiological applications. It was envisaged that the work to be undertaken in pursuit of this primary objective would require working within all three fundamental facets of statistics: methodological, computational, and applied. Specifically, this author identified three core **Aims** in the *Introduction* (Chapter 1, Section 1.3), which would constitute the focus of his efforts. Repeated here, these were

1. **To identify, collate, summarise, test, and refine or improve where possible, areas of statistical methodology currently used to analyse spatial and spatiotemporal point patterns (1 associated paper, plus 1 based on work from Honours year);**
2. **To aid in bridging the gap between powerful yet theoretically complex models and applied researchers (2 associated papers); and**
3. **To trial advancements made on real-world data (3 associated papers);**

with the number of credited accepted/submitted publications at the time of writing providing the reader with an early indication of achievement of the three objectives.

6.1.1 Chapter 1

Although theoretical specifics of the considered models are contained within the required chapters (2 and 3), several more general aspects of planar point processes had to be identified and summarised in order for adequate theoretical digestion of the remaining work.

Different types of planar point patterns were illustrated alongside the idea of *complete spatial randomness*, and the *Poisson* and *Cox processes* were introduced. The fundamental *intensity function*, from which the observed points have arisen, was defined. *Second-order properties* and some associated summary statistics, used to study the relationship between distinct pairs of points, were explained. Intricate theoretical details were left to referenced works; an applied, intuitive emphasis serving to best set up the remainder of the thesis.

6.1.2 Chapter 2

Addressed: ***Aims 1*** *and* ***3***

An extremely useful and thus commonly employed strategy to summarise the spatial variation in disease prevalence over a certain geographical region is the expression of some measure of the 'risk' of infection. This is typically derived as an estimated ratio of the distributions of infected (case) to non-infected (control) entities (Bithell, 1990, 1991). The flexibility of kernel density estimation renders it a particularly attractive solution for computation of the desired quantities. Historically explored in the context of a fixed (constant) scalar smoothing parameter for the density-ratio (Kelsall and Diggle, 1995*a*), this chapter sought to further build upon work in Hazelton and Davies (2009) and investigate the worthiness of using a particular variable-bandwidth calculator (Abramson, 1982) for the same purpose.

Chapter 2 described in detail the *2-dimensional kernel density estimator*, and associated *density-ratio estimator of relative risk*, for both fixed and adaptive isotropic smoothing. With findings published in Davies and Hazelton (2010), **Aim 1** was addressed as a theoretical exploration took place. Expressing fundamental properties of the adaptive density-ratio for the first time, approximate *variance stabilisation* of this estimator was uncovered. A novel, computationally inexpensive method based on this asymptotic theory for computation of *tolerance contours*, which describe statistically significant sub-regions of risk, was also introduced. This offered an alternative to computationally costly Monte-Carlo approaches. Finally, **Aim 3** was addressed in the context of these methods by way of the analysis of two separate real-world examples. The adaptive density-ratio and asymptotic tolerance contour techniques indicated benefits over the older methods, and formed a unique portion of a correspondingly developed and released R package.

6.1.3 Chapter 3

Addressed: ***Aims 1*** *and* ***3***

This chapter discussed in detail the *log-Gaussian Cox process*, where a stochastic Gaussian field is used to describe attractive dependence between points. By collating, summarising, and interpreting the theory associated with this model from an applied, epidemiological perspective, we have addressed **Aim 1** for the LGCP. This involved scrutiny of the material in Møller et al. (1998), Brix and Diggle (2001), Møller and Waagepetersen (2004) and Diggle, Rowlingson and Su (2005). Discussions were split into spatial and spatiotemporal sections, with the assumptions of stationarity, isotropy and separability imposed upon the model for theoretical accessibility. *Minimum contrast parameter estimation* was defined and examined. Special attention was also given to *circulant embedding* and the *fast-Fourier transform* for simulation of the process. The *Metropolis-adjusted Langevin algorithm* was covered in detail with respect to conditional simulation of both the spatial and spatiotemporal versions of the LGCP. **Aim 3** was addressed through the analysis of two real-world examples, one of which also detailed the useful notion of *exceedance probabilities* used to detect anomalous behaviour within the ever-changing intensity in real-time disease surveillance.

Of particular note were the suite of novel numerical experiments carried out in order to gain a unique insight into the performance of the LGCP and associated fitting and simulation techniques in practice. *Study 1* revealed little difference between use of the K-function or pair correlation function in terms of performance of the minimum contrast methods. When a deterministic spatial component was included in the modelling framework as a kernel density estimate in *Study 2*, it was found a clear preference is given to 'restrained oversmoothing' of the 2-dimensional density – this would result in the stochastic process parameter estimates closest to the true quantities. Finally, a carefully designed set of spatiotemporal LGCP scenarios were utilised in order to judge the appearance of exceedance probability surfaces following conditional simulation based on variations of the supplied dependence parameters in *Study 3*. It was concluded that although variations of the quality statistic employed did seem to exist between parameterisations, they did not differ greatly in a broad interpretation of the locations of anomalous sub-regions: a comforting result given the critical eye with

which minimum contrast is often (and correctly so) viewed.

This review coupled with the novel simulation studies revealed the model to hold considerable potential for a variety of applications. The ease with which the LGCP can be extended or modified, e.g. to account for global space-time trends, is a particularly attractive feature. By shedding new light on the relative robustness of currently employed LGCP disease surveillance techniques, it was concluded that the LGCP certainly warrants further attention in epidemiological applications, and efforts should be made to improve accessibility to the methodology.

6.1.4 Chapter 4

Addressed: ***Aim 2***

This chapter described the software implementation (by way of two separate R packages) of the methods introduced in Chapters 2 and 3. Both are currently available on the Comprehensive R Archive Network.

The first package constituted the efforts surrounding kernel estimation of spatial relative risk, and was named **sparr**. Quite simply, it is a direct, coded version of the contents of Chapter 2 and the associated articles. The flexibility and accessibility of **sparr**, showcased in Davies, Hazelton and Marshall (2011), were the main catalysts in initiating the published collaborations summarised in Chapter 5.

Following a visit to Lancaster University in January, 2011, several researchers became interested in this author's implementation of the LGCP and associated functionality in the R language. These included academics who had played an instrumental role in the development of the relevant methodology. This precipitated a collaborative effort to polish the existing code and release the package **lgcp**. A corresponding article is at the time of writing in press as Taylor, Davies, Rowlingson and Diggle (2012). The primary motivation for this package is the use of the LGCP for real-time disease surveillance, and the included functions therefore focus on fitting and conditional simulation of the spatiotemporal version of the model. This powerful tool, the first of its kind in R, affords the user control over all aspects of the initiation, execution, summary statistic calculation, and storage of the Markov chain. Flexible options for retrieval and post-processing of the results, including visualisation of diagnostics, the intensity itself, and exceedance probability surfaces, are supplied.

6.1.5 Chapter 5

Addressed: ***Aim 3***

The benefits yielded by having available the software capable of certain statistical methods are illustrated by the speed with which **sparr** (released online in late 2010 – earlier than **lgcp** which was released in late 2011) attracted the attention of researchers working in human and veterinary epidemiology. Chapter 5 summarised the collaborative works initiated by these epidemiologists which this author contributed to in terms of statistical support.

The first example dealt with analysing the spatial variation in *relative risk of schistosomiasis* in a particular region of China. Currently submitted for publication as Zhang, Davies

and Jiang (2011), the aim was to examine the performance of the newly-described adaptive bandwidth kernel density-ratio estimator and asymptotic p-value surfaces of Chapter 2 in comparison to the previously employed fixed-bandwidth versions. The second example, available as Zhang, Chen, Chen, Davies, Vaillancourt and Liu (2012), studied *global trends in a highly pathogenic subtype of avian influenza*. Use of Abramson's (1982) adaptive kernel density estimator, as well as Kulldorff's space-time scan statistic (Kulldorff et al., 1998, 2005), served to identify an emerging 'hotspot' of infected poultry cases. The third example detailed international efforts to 'verify' and 'validate' three different *disease control strategy evaluation models* designed to simulate the spread (and subsequent control) of veterinary foot-and-mouth, and is published as Sanson, Harvey, Garner, Stevenson, Davies, Hazelton, O'Connor, Dubé, Forde-Folle and Owen (2011). The final section of Chapter 5 simply aimed, via a non-epidemiological example, to highlight the flexibility of the LGCP. We provided an informal *preliminary exploration of some Japanese earthquake data*, which also provided an opportunity to experiment with some straightforward 'future-forecast' intensity prediction ideas. The spatiotemporal LGCP was shown to fit well, and conditional simulation functioned satisfactorily.

6.2 Future research, thesis closure

This research has clearly demonstrated the role statistics plays in geographical epidemiology. In particular, we focused on those applications where we have observed two-dimensional spatial and/or spatiotemporal point patterns, and examined in detail two distinct approaches geared toward gaining a deeper insight into the dispersion of disease observations within a given study region.

Consider the discussions surrounding the kernel-estimated relative risk function. An issue which crops up fairly often in smoothing problems is selection of some 'optimal' degree of smoothing based on the observed data. Although the adaptive bandwidth density-ratio was shown to perform competitively with the fixed-bandwidth version in Chapter 2, it was demonstrated that the instability of leave-one-out least-squares cross-validation rendered the corresponding method of optimal bandwidth calculation rather unreliable in practice. One future goal could be the search for feasible, and more numerically stable, jointly optimal (i.e. in light of the two density estimates – case and control) bandwidth selectors. This may well involve investigating the performance of any new techniques on the fixed and adaptive versions of the relative risk function separately, owing to the additional complexity of requiring both 'global' and 'pilot' specifications in the latter.

Branching out from Poisson point processes, the discussion of the spatial and relatively young spatiotemporal log-Gaussian Cox process in Chapter 3 showed the model has the ability to be able to cope well with epidemiological data. Discussed in detail was the issue of minimum contrast estimation of the parameters controlling the correlation structure of the high-dimensional Gaussian field. It would be extremely useful to continue to expand the methods available for this type of parameter estimation, particularly when we consider the many different kinds of point patterns that the LGCP is capable of simulating/capturing. One possible avenue would be to explore Bayesian methods (ideas arising from preliminary

discussions with the Lancaster group), though any such approach would clearly involve high computational expense to the point where it may be considered infeasible for that very reason.

By successfully addressing clearly defined objectives directly related to the needs of researchers working with space and space-time modelling of point processes, particularly in epidemiology, this thesis and its contents have resulted in a number of refinements to existing methodology; a better understanding of the models' numerical performance; newly developed and generally accessible computer programs; a host of real-world applications; and a collection of associated publications. In addition, the work completed has presented a number of interesting avenues that have strong potential as future research pursuits.

Appendices

Appendix A

Package documentation: sparr

Full package documentation for **sparr**, written solely by this author. The details of cited works are located together with those in the main body of this thesis in 'References'.

A.1 'Welcome' help file: `sparr-package`

`sparr-package` *The sparr Package: SPAtial Relative Risk*

Description

Provides functions to estimate fixed and adaptive kernel-smoothed relative risk surfaces via the density-ratio method and perform subsequent inference.

Details

Package:	sparr
Version:	0.3-1
Date:	2011-12-13
License:	GPL (>= 2)

Kernel smoothing, and the flexibility afforded by this methodology, provides an attractive approach to estimating complex probability density functions. This is particularly of interest when exploring problems in geographical epidemiology, the study of disease dispersion throughout some spatial region, given a population. The so-called 'relative risk surface', constructed as a ratio of estimated case to control densities (Bithell, 1990, 1991), describes the variation in the 'risk' of the disease, given the underlying at-risk population. This is a technique that has been applied successfully for mainly exploratory purposes in a number of different examples (see for example Sabel et al., 2000; Prince

et al., 2001; Wheeler, 2007).

This package provides functions for bivariate kernel density estimation (KDE), implementing both fixed and 'variable' or 'adaptive' (Abramson, 1982) smoothing parameter options (see the function documentation for more information). A selection of bandwidth calculators for bivariate KDE and the relative risk function are provided, including one based on the maximal smoothing principle (Terrell, 1990), and others involving a leave-one-out least-squares cross-validation (see below). In addition, the ability to construct asymptotically derived p-value surfaces ('tolerance' contours of which signal statistically significant sub-regions of 'extremity' in a risk surface - Hazelton and Davies, 2009; Davies and Hazelton, 2010), as well as some flexible visualisation tools, are provided.

The content of `sparr` can be broken up as follows:

Datasets
`PBC` a case/control planar point pattern (`ppp`) concerning liver disease in northern England. Also available is the case/control dataset `chorley` of the `spatstat` package, which concerns the distribution of laryngeal cancer in an area of Lancashire, England.

Bandwidth calculators
`OS` estimation of an isotropic smoothing parameter for bivariate KDE, based on the oversmoothing principle introduced by Terrell (1990).
`NS` estimation of an isotropic smoothing parameter for bivariate KDE, based on the optimal value for a normal density (bivariate normal scale rule - see e.g. Wand and Jones, 1995).
`LSCV.density` a least-squares cross-validated (LSCV) estimate of an isotropic bandwidth for bivariate KDE (see e.g. Bowman and Azzalini, 1997).
`LSCV.risk` a least-squares cross-validated (LSCV) estimate of a jointly optimal, common isotropic case-control bandwidth for the kernel-smoothed risk function (see Kelsall and Diggle, 1995*a*,*b*; Hazelton, 2008).

Bivariate functions
`KBivN` bivariate normal (Gaussian) kernel
`KBivQ` bivariate quartic (biweight) kernel
`bivariate.density` kernel density estimate of bivariate data; fixed or adaptive smoothing

Relative risk and p-value surfaces
`risk` estimation of a (log) relative risk function
`tolerance` calculation of asymptotic p-value surface

Printing and summarising objects
`S3` methods (`print.bivden`, `print.rrs`, `summary.bivden` and `summary.rrs`) are available for the bivariate density and risk function objects.

Visualisation

Most applications of the relative risk function in practice require plotting the relative risk within the study region (especially for an inspection of tolerance contours). To this end, `sparr` provides a number of different ways to achieve attractive and flexible visualisation. The user may produce a heat plot, a perspective plot, a contour plot, or an interactive 3D perspective plot (that the user can pan around and zoom - courtesy of the powerful `rgl` package; see below) for either an estimated relative risk function or a bivariate density estimate. These capabilities are available through `S3` support of the `plot` function; see

`plot.bivden` for visualising a single bivariate density estimate from `bivariate.density`, and

`plot.rrs` for visualisation of an estimated relative risk function from `risk`.

Dependencies

The `sparr` package depends upon some other important contributions to CRAN in order to operate; their uses here are indicated:

`spatstat` - Fast-fourier transform assistance with fixed and adaptive density estimation, as well as region handling; see Baddeley and Turner (2005).

`rgl` - Interactive 3D plotting of densities and surfaces; see Adler and Murdoch (2009).
`MASS` - Utility support for internal functions; see Venables and Ripley (2002).

Citation

To cite use of `sparr` in publications, the user may refer to the following work:
Davies, T.M., Hazelton, M.L. and Marshall, J.C. (2011), `sparr`: Analyzing spatial relative risk using fixed and adaptive kernel density estimation in `R`, *Journal of Statistical Software* **39**(1), 1-14.

Author(s)

T.M. Davies, M.L. Hazelton and J.C. Marshall
Institute of Fundamental Sciences - Statistics, Massey University, Palmerston North, New Zealand.

Maintainer: T.M.D. `<t.davies@massey.ac.nz>`
Feedback welcomed.

A.2 Conversion to pixel image class `im`: `as.im.bivden`, `as.im.rrs`

as.im.bivden	*Converting a* sparr *bivariate kernel density estimate or relative risk surface object into a* spatstat *pixel image.*

Description

as.im methods for classes "bivden" and "rrs"

Usage

```
## S3 method for class 'bivden'
as.im(X, ...)
## S3 method for class 'rrs'
as.im(X, ...)
```

Arguments

X	An object of class "bivden" resulting from a call to bivariate.density, or an object of class "rss" resulting from a call to risk.
...	Ignored.

Value

An object of class im corresponding to the supplied argument. Additional return information originally part of X is lost.

Author(s)

T.M. Davies

See Also

im, as.im

Examples

```
data(chorley)

ch.bivden <- bivariate.density(chorley, ID = "lung", adaptive = FALSE,
 + pilotH = 1.5)
ch.im <- as.im(ch.bivden)
summary(ch.im)
```

A.3 Bivariate (2D) kernel density estimation: `bivariate.density`

`bivariate.density` *Bivariate kernel density estimates*

Description

Provides an adaptive or fixed bandwidth kernel density estimate of bivariate data.

Usage

```
bivariate.density(data, ID = NULL, pilotH, globalH = pilotH,
  adaptive = TRUE, edgeCorrect = TRUE, res = 50, WIN = NULL,
  counts = NULL, intensity = FALSE, xrange = NULL,
  yrange = NULL, trim = 5, gamma = NULL, atExtraCoords = NULL,
  use.ppp.methods = TRUE, comment = TRUE)
```

Arguments

`data`	An object of type `data.frame`, `list`, `matrix`, or `ppp` giving the observed data from which we wish to calculate the density estimate. Optional ID information (e.g. a dichotomous indicator for cases and controls) may also be provided in these four data structures. See 'Details' for further information on how to properly specify each one.
`ID`	If `data` is a data structure with a third component/column indicating case (1) or control (0) status, `ID` must specify which of these groups we wish to estimate a density for. If `ID` is `NULL` (default), a density is estimated for all present observations, regardless of any status information.
`pilotH`	A single numeric, positive 'smoothing parameter' or 'bandwidth'. When `adaptive` is `TRUE` (default), this value is taken to be the pilot bandwidth, used to construct the bivariate pilot density required for adaptive smoothing (see 'Details'). For a fixed bandwidth kernel density estimate, `pilotH` simply represents the fixed amount of smoothing. Currently, all smoothing is isotropic in nature.
`globalH`	A single numeric, positive smoothing multiplier referred to as the global bandwidth, used to calculate the adaptive bandwidths (see 'Details'). When `adaptive` is `TRUE`, this defaults to be the same as the pilot bandwidth. Ignored for a fixed density estimate.
`adaptive`	Boolean. Whether or not to produce an adaptive (variable bandwidth) density estimate, with the alternative being a fixed bandwith density estimate. Defaults to `TRUE`.

`edgeCorrect` Boolean. Whether or not to perform edge-correction on the density estimate according to the methods demonstrated by Diggle (1985) (fixed bandwidth) and Marshall and Hazelton (2010) (adaptive). This can have a noticeable effect on computation time in some cases. Defaults to `TRUE`. When `adaptive = TRUE`, the fixed-bandwidth pilot density is also edge-corrected according to `edgeCorrect`.

`res` A single, numeric, positive integer indicating the square root of the desired resolution of the evaluation grid. That is, each of the evaluation grid axes will have length `res`. Currently, only `res*res` grids are supported. Defaults to 50 for computational reasons.

`WIN` A polygonal object of class `owin` from the package `spatstat` giving the study region or 'window'. All functions in the package `sparr` that require knowledge of the specific study region make use of this class; no other method of defining the study region is currently supported. If no window is supplied (default), the function defines (and returns) it's own rectangular `owin` based on `xrange` and `yrange`. Ignored if `data` is an object of type `ppp`.

`counts` To perform binned kernel estimation, a numeric, positive, integer vector of giving counts associated with each observed coordinate in `data`, if `data` contains unique observations. If `NULL` (default), the function assumes each coordinate in `data` corresponds to one observation at that point. Should the `data` being supplied to `bivariate.density` contain duplicated coordinates, the function computes the counts vector internally (overriding any supplied value for `counts`), issues a warning, and continues with binned estimation. Non-integer values are rounded to the nearest integer.

`intensity` A boolean value indicating whether or not to return an intensity (interpreted as the the expected number of observations per unit area and integrating to the number of observations in the study region) function, rather than a density (integrating to one). Defaults to `FALSE`.

`xrange` Required only when no study region is supplied (`WIN = NULL`) and `data` is not an object of class `ppp`, and ignored otherwise. A vector of length 2 giving the upper and lower limits of the estimation interval for the x axis, in which case an evenly spaced set of values of length `res` is generated.

`yrange` As above, but for the y axis.

`trim` A numeric value (defaulting to `5`) that prevents excessively large bandwidths in adaptive smoothing by trimming the originally computed bandwidths **h** by `trim` times `median(h)`. A value of `NA` or a negative numeric value requests no trimming. Ignored when `adaptive` is `FALSE`.

`gamma` An optional positive numeric value to use in place of `gamma` for adaptive bandwidth calculation (see 'Details'). For adaptive relative risk estimation, this value can sensibly be chosen as common for both case and control densities (such as the `gamma` value from the adaptive density estimate of the 'pooled' (full) dataset): see Davies and Hazelton (2010). If nothing is supplied (default), this value is computed from the data being used to estimate the density in the defined fashion (again, see 'Details'). Ignored for fixed bandwidth estimation.

`atExtraCoords` It can occasionally be useful to retrieve the values of the estimated density at specific coordinates that are not the specific observations or the exact grid coordinates, for further analysis or plotting. `atExtraCoords` allows the user to specify an additional object of type `data.frame` with 2 colums giving the x `atExtraCoords[,1]` and y `atExtraCoords[,2]` coordinates at which to calculate and return the estimated density and other statistics (see 'Value').

`use.ppp.methods`
Boolean. Whether or not to switch to using methods defined for objects of class `ppp.object` from the package `spatstat` to estimate the density. This approach is much, much faster than forcing `bivariate.density` to do the explicit calculations (due to implementation of a Fast Fourier Transform; see `density.ppp`) and is highly recommended for large datasets. To further reduce computation time in the adaptive case when `use.ppp.methods = TRUE`, the variable edge-correction factors are calculated using the integer percentiles of the varying bandwidths. Defaults to `TRUE`.

`comment` Boolean. Whether or not to print function progress (including starting and ending times) during execution. Defaults to `TRUE`.

Details

This function calculates an adaptive or fixed bandwidth bivariate kernel density estimate, using the bivariate Gaussian kernel. Abramson's method is used for adaptive smoothing (Abramson, 1982). Suppose our `data` argument is a `data.frame` or `matrix`. Then for each observation `data[i,1:2]` ($i = 1, 2, \ldots n$), the bandwidth `h[i]` is given by

`h[i]`=`globalH` / (w(`data[i,1:2]`; `pilotH`)^(1/2)*`gamma`)

where w is the fixed bandwidth pilot density constructed with bandwidth `pilotH` and the scaling parameter `gamma` is the geometric mean of the w^(-1/2) values. A detailed discussion on this construction is given in Silverman (1986).

If the `data` argument is a `data.frame` or a `matrix`, this must have exactly two columns containing the x (`[,1]`) and y (`[,2]`) data values, or exactly three columns with the third (rightmost) column giving ID information by way of a numeric, dichotomous indicator.

Should `data` be a `list`, this must have two vector components of equal length named `x` and `y`. The user may specify a third component with the name `ID` giving the vector of corresponding ID information (must be of equal length to `x` and `y`). Alternatively, `data` may be an object of class `ppp` (see `ppp.object`). ID information can be stored in such an object through the argument `marks`. If `data` is a `ppp` object, the value of `window` of this object overrides the value of the argument `WIN` above.

Value

An object of class `"bivden"`. This is effectively a list with the following components:

`Zm`	a numeric matrix giving the value of the estimated (edge-corrected if elected) density at each of the coordinates of the grid. Values corresponding to points on the grid that fall outside the study region `WIN` are set to `NA`
`X`	a the sequence of values that were used as x grid coordinates. Will have length `res`
`Y`	a the sequence of values that were used as y grid coordinates. Will have length `res`
`kType`	the kernel function used in estimation. Currently fixed at `"gaus"`
`h`	a numeric vector with length equal to the number of observations, giving the bandwidths assigned to each observation in the order they appeared in `data`. For a fixed bandwidth estimate, this will simply be the identical value passed to and returned as `pilotH`
`pilotH`	the pilot or fixed bandwidth depending on whether adaptive smoothing is employed or not, respectively
`globalH`	the global bandwidth `globalH` if adaptive smoothing is employed, `NA` for fixed smoothing
`hypoH`	the matrix of 'hypothetical' bandwidths (with element placement corresponding to `Zm`) for each coordinate of the evaluation grid. That is, these values are the bandwidths at that grid coordinate if, hypothetically, there was an observation there (along with the original data). These are used for edge-correction in adaptive densities (Marshall and Hazelton, 2010). Will be `NA` for fixed bandwidth estimates
`zSpec`	a numeric vector with length equal to the number of observations used, giving the values of the density at the specific coordinates of the observations. Order corresponds to the order of the observations in `data`
`zExtra`	as `zSpec` for the observations in `atExtraCoords`, `NA` if `atExtraCoords` is not supplied
`WIN`	the object of class `owin` used as the study region

`qhz` a numeric matrix of the edge-correction factors for the entire evaluation grid (with placement corresponding to `Zm`. If `edgeCorrect = FALSE`, all edge correction factors are set to and returned as `1`

`qhzSpec` edge-correction factors for the individual observations; order corresponding to `data`

`qhzExtra` as `qhzSpec` for the observations in `atExtraCoords`; `NA` if `atExtraCoords` is not supplied

`pilotvals` the values of the pilot density used to compute the adaptive bandwidths. Order corresponds to the order of the observations in `data`. `NULL` when `adaptive = FALSE`

`gamma` the value of `gamma` that was passed to the function, or the geometric mean term of the reciprocal of the square root of the pilot density values used to scale the adaptive bandwidths if `gamma` is not supplied. `NULL` when `adaptive = FALSE`

`counts` the counts vector used in estimation of the density/intensity. If all values in `data` were unique and `counts = NULL`, the returned `counts` will be a vector of ones equal to the number of coordinates in `data`

`data` a two-column numeric data frame giving the observations in the originally supplied `data` that were used for the density estimation. If `data` originally contained duplicated coordinates, the returned `data` will contain only the unique coordinates, and should be viewed with respect to the returned value of `counts`

Warning

Explicit calculation of bivariate kernel density estimates is computationally expensive. The decision to produce adaptive over fixed bandwidth estimates, the size of the dataset, the evaluation grid resolution specified by `res`, the complexity of the study region and electing to edge-correct all have a direct impact upon the time it will take to estimate the density. Keeping `use.ppp.methods = TRUE` can drastically reduce this computational cost at the expense a degree of accuracy that is generally considered negligible for most practical purposes.

Author(s)

T.M. Davies

Examples

```
##Chorley-Ribble laryngeal cancer data ('spatstat' library)
data(chorley)

ch.lar.density <- bivariate.density(data = chorley, ID = "larynx",
 pilotH = 1.5, adaptive = FALSE)
```

```
plot(ch.lar.density, col = "lightblue", phi = 30, theta = -30,
 ticktype = "detailed", main = "chorley.larynx", display = "persp")

## Not run:

##PBC liver disease data
data(PBC)
pbc.adaptive.density <- bivariate.density(data = PBC$data, ID = 1,
 pilotH = 350, WIN = PBC$owin)

#3D plot - may need to adjust size of RGL device. Hold left click
# to pan, hold right to zoom.
plot(pbc.adaptive.density, display = "3d", col = heat.colors(20),
 main = "Density of PBC in north-east England", aspect = 1:2)

## End(Not run)
```

A.4 Single-density LSCV bandwidth: `LSCV.density`

`LSCV.density` — *Leave-one-out least-squares cross-validation (LSCV) for bivariate KDE bandwidths*

Description

Provides an isotropic LSCV bandwidth estimate for use in 2-dimensional kernel density estimation (see e.g. Bowman and Azzalini, 1997).

Usage

```
LSCV.density(data, hlim = NULL, res = 128, edge = TRUE,
 WIN = NULL, quick = TRUE, comment = TRUE)
```

Arguments

`data`	An object of type `data.frame`, `list`, `matrix`, or `ppp` describing the observed data from which we wish to calculate the LSCV bandwidth. See 'Details' for further information.
`hlim`	A numeric vector of length 2 giving the interval over which to search for the bandwidth that minimises the selection criterion. If `NULL` (default), the function attempts to automatically select an appropriate range based on multiples of Stoyan and Stoyan's (1994) rule-of-thumb. The user is strongly recommended to supply their own `hlim`.

`res`	Single integer giving the square grid resolution over which evaluation of the selection criterion takes place. Defaults to a 128 by 128 grid.
`edge`	Boolean. Whether or not to employ edge-correction in the calculations. Defaults to `TRUE`.
`WIN`	A polygonal `owin` object giving the study region. Ignored if `data` is already a `ppp.object`.
`quick`	Intended for advanced use; users are recommended not to change the default `TRUE`. Setting `quick = FALSE` forces the function to individually evaluate the CV objective function at each of `seq(hlim[1], hlim[2], length = 50)` bandwidths, returning the corresponding values. Can be useful for diagnostic purposes.
`comment`	Boolean. Whether or not to print function progress during execution. Defaults to `TRUE`.

Details

This function calculates a LSCV smoothing bandwidth for kernel density estimates of 2-dimensional (bivariate) data. If the `data` argument is a `data.frame` or a `matrix`, this must have exactly two columns containing the x (`[,1]`) and y (`[,2]`) data values. Should `data` be a `list`, this must have two vector components of equal length named `x` and `y`. Alternatively, `data` may be an object of class `ppp` (see `ppp.object`).

Value

A single numeric value of the estimated bandwidth (if `quick = FALSE`, this value is named `hopt`; additionally returned are the objective function values (`lscv`) and the index of the minimum value (`ind`)). The user may need to experiment with adjusting `hlim` to find a suitable minimum.

Warning

Leave-one-out LSCV for bandwidth selection in kernel density estimation is notoriously unstable in practice and has a tendency to produce rather small bandwidths. Satisfactory bandwidths are not guaranteed for every application. This method can also be computationally expensive for large data sets and fine evaluation grid resolutions.

Author(s)

T.M. Davies

See Also

`spatstat`'s function `bw.relrisk`

Examples

```
## Not run:
data(PBC)

##PBC cases
LSCV.density(split(PBC)[[1]],hlim=c(10,400))

##PBC controls
LSCV.density(split(PBC)[[2]],hlim=c(10,400))

## End(Not run)
```

A.5 Density-ratio LSCV bandwidth: `LSCV.risk`

`LSCV.risk` — *Leave-one-out least-squares cross-validation (LSCV) bandwidths for the relative risk function*

Description

Attempts to estimate a jointly optimal, common case-control fixed bandwidth for use in the kernel-smoothed relative risk function via leave-one-out least-squares cross-validation (LSCV). The user can choose between two methods described in Kelsall and Diggle (1995*a*,*b*) and Hazelton (2008).

Usage

```
LSCV.risk(cases, controls, hlim = NULL,
 method = c("kelsall-diggle", "hazelton"), res = 128,
 WIN = NULL, edge = TRUE, comment = TRUE)
```

Arguments

`cases`	An object of type `data.frame`, `list`, `matrix`, or `ppp` describing the observed case data from which we wish to calculate the LSCV bandwidth. See 'Details' for further information.
`controls`	As for `cases`, but for the control observations. Both `cases` and `controls` must be of the same object class.
`hlim`	A numeric vector of length 2 giving the interval over which to search for the common bandwidth which minimises the selection criterion. If `NULL` (default), the function attempts to automatically select an appropriate range based on multiples of Stoyan and Stoyan's (1994) rule-of-thumb. The user is strongly recommended to supply their own `hlim`.

`method`	A character vector giving the specific selection criterion to minimise; see either Kelsall and Diggle (1995*a*,*b*) or Hazelton (2008). See also 'Details'. Defaults to `"kelsall-diggle"`.
`res`	Single integer giving the square grid resolution over which evaluation of the selection criterion takes place. Defaults to a 128 by 128 grid.
`WIN`	A polygonal `owin` object giving the study region. Ignored if `data` is already a `ppp.object`.
`edge`	Boolean. Whether or not to employ edge-correction in the calculations. Defaults to `TRUE`.
`comment`	Boolean. Whether or not to print function progress during execution. Defaults to `TRUE`.

Details

This function calculates a 'jointly optimal', common isotropic LSCV bandwidth for the (Gaussian) kernel-smoothed relative risk function (case-control density-ratio). If the `cases`, `controls` arguments are `data.frame` or `matrix` objects, these must each have exactly two columns containing the x (`[,1]`) and y (`[,2]`) data values. Should they be `list`s, these must have two vector components of equal length named `x` and `y`. Alternatively, `cases` and `controls` may be objects of class `ppp` (see `ppp.object`), and the argument `WIN` can be ignored.

It can be shown that choosing a bandwidth that is equal for both case and control density estimates is preferable to computing 'separately optimal' bandwidths (Kelsall and Diggle, 1995*a*). Setting `method = "kelsall-diggle"`, `LSCV.risk` computes the common bandwidth which minimises the approximate mean integrated squared error of the log-transformed risk surface (see specifically Kelsall and Diggle, 1995*b*).

Alternatively, the user has the option of computing the common case-control bandwidth which minimises a *weighted* mean integrated squared error of the (raw) relative risk function (see Hazelton, 2008). Generally, this author has found the Kelsall-Diggle method to provide more stable performance.

Value

A single numeric value of the estimated bandwidth. The user may need to experiment with adjusting `hlim` to find a suitable minimum.

Warning

Leave-one-out LSCV for jointly optimal, common bandwidth selection in the kernel-smoothed risk function is even more unstable (in terms of high variability) than the standalone density version. Caution is advised; not all applications will yield a successful

result (this is termed "a breakdown of the methodology" by Kelsall and Diggle, 1995*a*). Undersmoothing has been noted in this author's personal experience. This method can also be computationally expensive for large data sets and fine evaluation grid resolutions.

Author(s)

T.M. Davies

See Also

`spatstat`'s function `bw.relrisk`

Examples

```
## Not run:
data(chorley)

LSCV.risk(cases = split(chorley)[[1]], controls = split(chorley)[[2]],
 hlim = c(0.1,2))

## End(Not run)
```

A.6 Bivariate (2D) Gaussian kernel: `KBivN`

`KBivN` *Standard bivariate normal kernel*

Description

Evaluates the standard bivariate normal (Gaussian) kernel function at specified values.

Usage

```
KBivN(X)
```

Arguments

`X`	A numeric vector of length 2 or a data frame with 2 columns.

Details

If `X` is a vector of length 2, then the two components `X[1]` and `X[2]` are taken to be the x and y coordinates respectively. For multiple evaluations at differing coordinates, `X` must be a data frame with `X[,1]` and `X[,2]` as the corresponding pairs of x and y coordinates respectively.

Value

A single numeric value if `X` is a vector, or `nrow(X)` values if `X` is a data frame, giving the result of the standard bivariate normal kernel at the specified coordinate(s).

Author(s)

T.M. Davies

Examples

```
KBivN(c(0.1,0.4))

x <- y <- seq(-4,4,length=50)
z <- KBivN(data.frame(cbind(sort(rep(x,50)),rep(y,50))))
persp(x,y,matrix(z,50,50,byrow=TRUE),main="bivariate Gaussian kernel",
 phi=30,theta=-30)
```

A.7 Bivariate (2D) quartic (biweight) kernel: `KBivQ`

`KBivQ` *Standard bivariate quartic (biweight) kernel*

Description

Evaluates the standard bivariate quartic (biweight) kernel function at specified values, for either the spherical or product derivation of the function.

Usage

```
KBivQ(X,type="spher")
```

Arguments

`X`	A numeric vector of length 2 or a data frame with 2 columns.
`type`	A character string.

`"spher"`	(default) spherical calculation of the kernel
`"prod"`	uses the product approach to kernel calculation

Details

If `X` is a vector of length 2, then the two components `X[1]` and `X[2]` are taken to be the x and y coordinates respectively. For multiple evaluations at differing coordinates, `X` must be a data frame with `X[,1]` and `X[,2]` as the corresponding pairs of x and y coordinates respectively.

Unlike the bivariate Gaussian kernel, it is necessary to specify the method of extending the univariate quartic kernel to the bivariate case; this can be done in two different ways, one way resulting in a slightly different kernel to the other. An explanation of these 'spherical' and 'product' approaches is given in Wand and Jones (1995).

Value

A single numeric value if `X` is a vector, or `nrow(X)` values if `X` is a data frame, giving the result of the standard bivariate quartic kernel at the specified coordinate(s) for the elected function derivation type.

Author(s)

T.M. Davies

Examples

```
KBivQ(c(0.1,0.4))

x <- y <- seq(-0.9,0.9,length=50)
z.spher <- KBivQ(data.frame(cbind(sort(rep(x,50)), rep(y,50))))
z.prod <- KBivQ(data.frame(cbind(sort(rep(x,50)), rep(y,50))), "prod")

par(mfrow=c(1,2))
contour(x, y, matrix(z.spher, 50, 50, byrow = TRUE),
 main = "bivariate quartic kernel (spherical)")
contour(x, y, matrix(z.prod, 50, 50, byrow=TRUE),
 main = "bivariate quartic kernel (product)")
```

A.8 Single-density normal scale rule bandwidth: `NS`

`NS` *Normal scale rule for bivariate KDE bandwidths*

Description

Provides the (isotropic) optimal bandwidth for a bivariate normal density based on a simple expression.

Usage

```
NS(data, nstar = NULL, scaler = NA)
```

Arguments

`data`	An object of type `data.frame`, `list`, `matrix`, or `ppp` giving the observed data from which we wish to calculate the NS bandwidth. See 'Details' for further information.
`nstar`	A single numeric, positive value to use in place of the number of observations n in the NS formula. If `NULL` (default), n will simply be the number of observations in `data`.
`scaler`	A single numeric, positive value to use for transforming the result with respect to the scale of the recorded data (i.e. a scalar representation of the standard deviation of the data). If `NA` (default), the scaling value is set as the mean of the interquartile ranges (IQR) of the x and y data values divided by 1.34 (Gaussian IQR).

Details

This function calculates a smoothing bandwidth for kernel density estimates of 2-dimensional data: the optimal value which would minimise the asymptotic mean integrated squared error of the bivariate normal density function, assuming the standard Gaussian kernel function. See Wand and Jones (1995) for example. If the `data` argument is a `data.frame` or a `matrix`, this must have exactly two columns containing the x (`[,1]`) and y (`[,2]`) data values. Should `data` be a `list`, this must have two vector components of equal length named `x` and `y`. Alternatively, `data` may be an object of class `ppp` (see `ppp.object`).

Value

A single numeric value of the estimated bandwidth.

Warning

The NS bandwidth is an approximation, and assumes *that the target density is bivariate normal*. This is considered rare in e.g. epidemiological applications. Nevertheless, it remains a quick and easy 'rule-of-thumb' method with which one may obtain a smoothing parameter in general applications.

Author(s)

T.M. Davies

Examples

```
data(PBC)
PBC.casedata <- split(PBC)[[1]]
PBC.controldata <- split(PBC)[[2]]

pbc.h <- NS(PBC, nstar = sqrt(PBC.casedata$n * PBC.controldata$n))
```

```
##Scaling bandwidth for case data using standard deviations
sd.x <- sd(PBC.casedata$x)
sd.y <- sd(PBC.casedata$y)

NS(data = PBC.casedata, scaler = mean(sd.x, sd.y))
```

A.9 Terrell's single-density oversmoothing bandwidth: OS

`OS` *Maximal smoothing principle (oversmoothing) for bivariate KDE bandwidths*

Description

Provides an (isotropic) bandwidth estimate for use in bivariate KDE based on the oversmoothing factor introduced by Terrell (1990).

Usage

```
OS(data, nstar = NULL, scaler = NA)
```

Arguments

`data`	An object of type `data.frame`, `list`, `matrix`, or `ppp` giving the observed data from which we wish to calculate the OS bandwidth. See 'Details' for further information.
`nstar`	A single numeric, positive value to use in place of the number of observations n in the OS formula. If `NULL` (default), n will simply be the number of observations in `data`.
`scaler`	A single numeric, positive value to use for transforming the result with respect to the scale of the recorded data. If `NA` (default), the scaling value is set as the mean of the interquartile ranges (IQR) of the x and y data values divided by 1.34 (Gaussian IQR). This approach was used in Davies and Hazelton (2010).

Details

This function calculates a smoothing bandwidth for kernel density estimates of bivariate data, following the maximal smoothing priciple of Terrell (1990). If the `data` argument is a `data.frame` or a `matrix`, this must have exactly two columns containing the x

([,1]) and y ([,2]) data values. Should data be a list, this must have two vector components of equal length named x and y. Alternatively, data may be an object of class ppp (see ppp.object).

Value

A single numeric value of the estimated bandwidth.

Author(s)

T.M. Davies

Examples

```
data(PBC)

##Using geometric mean of case and control counts as n
n1 <- sum(PBC$data[,3])
n2 <- nrow(PBC$data) - n1

OS(data = PBC$data[,1:2], nstar = sqrt(n1 * n2))

##Scaling bandwidth for case data using standard deviations
sd.x <- sd(PBC$data[PBC$data[,3]==1,1])
sd.y <- sd(PBC$data[PBC$data[,3]==1,2])

OS(data = PBC$data[PBC$data[,3]==1,1:2], scaler = mean(sd.x, sd.y))
```

A.10 Primary biliary cirrhosis (PBC) data: PBC

PBC *Primary biliary cirrhosis data*

Description

Data of the locations of 761 cases of primary biliary cirrhosis in several adjacent health regions of north-eastern England, along with 3020 controls representing the at-risk population, collected between 1987 and 1994. These data were first presented and analysed by Prince et al. (2001); subsequent analysis of these data in the spirit of sparr was performed in Davies and Hazelton (2010). Also included is the polygonal study region.

Usage

```
data(PBC)
```

Format

PBC is a dichotomously marked ppp.object.

Acknowledgements

The authors thank Prof. Peter Diggle at Lancaster University (http://www.lancs.ac.uk/staff/diggle/) for providing access to these data.

Examples

```
data(PBC)
summary(PBC)
plot(PBC)
```

A.11 Plotting density and risk function estimates: plot.bivden, plot.rrs

---

plot.bivden *Plotting a bivariate kernel density estimate object*

Description

plot methods for classes "bivden" and "rrs"

Usage

```
## S3 method for class 'bivden'
plot(x, ..., display = c("heat", "contour", "persp", "3d"),
 show.WIN = TRUE)
## S3 method for class 'rrs'
plot(x, ..., display = c("heat", "contour", "persp", "3d"),
 show.WIN = TRUE, tolerance.matrix = NULL,
 tol.opt = list(raise = 0.01, col = "black", levels = 0.05, lty = 1,
 + lwd = 1))
```

Arguments

x	An object of class "bivden" resulting from a call to bivariate.density, or an object of class "rss" resulting from a call to risk.
...	Additional graphical parameters to be passed to the relevant plot command depending on the value of display.
display	One of four possible character strings indicating the kind of plot desired (see 'Details'). Defaults to "heat".

`show.WIN` Boolean. Whether or not to draw the study region as an aesthetic enhancement to the plot of the density/risk surface. Defaults to `TRUE`.

`tolerance.matrix`
The matrix of p-values resulting from a call to `tolerance` and used to draw the asymptotic tolerance contours. If this argument is supplied, tolerance contours are automatically superimposed upon a `display = "heat"` or `display = "3d"` plot. Ignored for `display = "persp"` or `display = "contour"` plots. Defaults to `NULL`.

`tol.opt` A named list of components that control plotting of the tolerance contours given by `tolerance.matrix`. Components `col`, `levels`, `lty` and `lwd` are vectors of equal length controlling the colour, significance levels, line type (ignored for `display = "3d"`) and line width of the plotted contours respectively. The element `raise` is a single numeric value and is used only when `display = "3d"`. This vertically (i.e. with respect to the z axis) translates the contours upon the 3-D surface (see 'Details'). A value of `0` requests no translation. Defaults to `0.01`.

Details

There are currently four implemented plot types to visualise the estimated density or risk function. `"heat"` selects a heatplot, `"contour"` is simply a contour plot and `"persp"` creates a perspective plot. Selection of `"3d"` uses functions from the `rgl` package to open an RGL graphics device and creates a 3-dimensional surface which the user can interact with using the mouse. To use `...` to improve the appearance of the four possible plot types `"heat"`, `"contour"`, `"persp"` and `"3d"`, the reader is highly recommended to consult the relevant documentation in the help pages `plot.im`, `contour`, `persp` and `persp3d` respectively.

Adding tolerance contours to a `"3d"` relative risk plot requires the function to make some approximations to the vertical positioning of the contours at each corresponding coordinate. This can lead to some parts of normally visible contours falling 'underneath' the plotted surface, resulting in partially obscured contours. The element `raise` in `tol.opt` overcomes this issue by artificially raising the visible contours by a fixed amount. Care should be taken to find an appropriate value for `raise` for each analysis.

Value

Plots to the relevant graphics device.

Author(s)

T.M. Davies

See Also

`bivariate.density`, `risk`, `plot.default`, `plot.im`, `contour`, `persp`, `persp3d`, `par`, `par3d`

Examples

```
## see Examples in documentation for functions 'bivariate.density',
##  'risk' and 'tolerance'.
```

A.12 Relative risk function estimation: `risk`

`risk` *Bivariate relative risk function*

Description

Estimates a relative risk function based on the ratio of two bivariate kernel density estimates over identical grids and regions. In geographical epidemiology, the two densities would represent a set of disease cases (numerator) and a sample of controls illustrating the at-risk population (denominator).

Usage

```
risk(f, g, delta = 0, log = TRUE, h = NULL, adaptive = FALSE, res = 50,
 WIN = NULL, tolerate = FALSE, plotit = TRUE, comment = TRUE)
```

Arguments

`f`	Either a pre-calculated object of class `"bivden"` representing the 'case' density estimate, or an object of type `data.frame`, `list`, `matrix`, or `ppp` giving the observed case data. If this raw data is provided, a kernel density estimate is computed internally, with certain options available to the user in `bivariate.density` chosen/calculated automatically. See 'Details' for further information.
`g`	As for argument `f`, but for the controls. Whatever the type, the class of `g` must match that of `f`.
`delta`	A single numeric scaling parameter used for an optional additive constant to the densities; occasionally used for risk surface construction (see 'Details'). A negative or zero value for `delta` requests no additive constant (default).
`log`	Boolean. Whether or not to return the (natural) log-transformed relative risk function as recommended by Kelsall and Diggle (1995*a*). Defaults to `TRUE` with the alternative being the raw density ratio.
`h`	Ignored if `f` and `g` are already `"bivden"` objects. An optional numeric vector of length 1 OR 2, giving the global bandwidth(s) for internal estimation of the case and control densities if `adaptive = TRUE`, or

the fixed bandwidth(s) if `adaptive = FALSE`. When `h` is a single numeric value, this is elected as the common global/fixed bandwidth for case and control densities. When `h` has length 2, the values `h[1]` and `h[2]` are assigned as the case and control global/fixed bandwidths respectively. By default, a value of `h = NULL` tells the function to use the global/fixed smoothing parameters as outlined in 'Details' below. Note that for adaptive estimation, this argument does not affect calculation of the pilot bandwidths.

`adaptive` Ignored if `f` and `g` are already `"bivden"` objects. A boolean value specifying whether or not to employ adaptive smoothing for internally estimating the densities. A value of `FALSE` (default) elects use of fixed-bandwidth estimates.

`res` Ignored if `f` and `g` are already `"bivden"` objects. A numeric value giving the desired resolution (of one side) of the evaluation grid. Higher values increase resolution at the expense of computational efficiency. Defaults to a 50 by 50 grid.

`WIN` Ignored if `f` and `g` are already `"bivden"` objects OR objects of class `ppp` (in which case the study region is set to the value of the resident `window` component). A polygonal object of class `owin` giving the relevant study region in which the `f` and `g` data was collected.

`tolerate` Ignored if `f` and `g` are already `"bivden"` objects. A boolean value specifying whether or not to calculate a corresponding asymptotic p-value surface (for tolerance contours) for the estimated relative risk function. If `TRUE`, the p-value surface tests for elevated risk only (equivalent to setting `test = "greater"` in `tolerance`) and is evaluated over a maximum grid resolution of 50 by 50. Defaults to `FALSE` for computational reasons.

`plotit` Boolean. If `TRUE` (default), a heatplot of the estimated relative risk function is produced. If `tolerate = TRUE`, asymptotic tolerance contours are automatically added to the plot at a significance level of 5%.

`comment` Ignored if `f` and `g` are already `"bivden"` objects. Boolean. Whether or not to print function progress (including starting and ending datetimes) during execution. Defaults to `TRUE`.

Details

This function estimates a relative risk function via the density ratio method using fixed or adaptive bandwidth bivariate kernel density estimates. Both densities must be estimated using the same evaluation grid (and the same study window) in `bivariate.density`. In geographical epidemiology, the argument `f` represents the spatial distribution of the disease cases, and `g` the at-risk (control) population.

The option to supply the raw case and control data is available. If this is done, the function runs `bivariate.density` internally, abstracting certain decisions about the density estimation away from the user. If the user sets `adaptive = TRUE` (and `h` remains at `NULL`), the smoothing parameters are calculated as per the approach taken in Davies and Hazelton (2010): a common global bandwidth using the pooled data from `OS`, and differing pilot bandwidths using `CV.sm` on the case and control data separately. The scaling parameter `gamma` is common for the case and control density estimates, set as the `gamma` component of the pooled estimate. If a fixed relative risk is desired (`adaptive = FALSE`) and no specific bandwidths are given via the argument `h`, the case and control densities share a common bandwidth computed from the pooled data using `OS`. In supplying raw data to `risk`, the user must also specify an evaluation grid resolution (defaulting to 50 by 50) and the study region `WIN` (unless `f` and `g` are objects of class `ppp`, in which case the resident `window` component overrides `WIN`). All other arguments are set to their defaults as in `bivariate.density`.

If more flexibility is required for estimation of the case and control densities, the user must supply 'pre-calculated' objects of class `"bivden"` (from `bivariate.density`) as the `f` and `g` arguments. This drastically reduces the running time of a call to `risk` (as the density estimation step is already complete). However, the option of internally computing the asymptotic p-value surfaces (via the argument `tolerate`) is unavailable in this case; the user must run the `tolerance` function separately if tolerance contours are desired.

The relative risk function is defined here as the ratio of the 'case' density to the 'control' (Bithell, 1990, 1991). Using kernel density estimation to model these densities (Diggle, 1985), we obtain a workable estimate thereof. This function defines the risk function r in the following fashion:

r `= (f + delta*max(g))/(g + delta*max(g))`

Note the (optional) additive constants defined by `delta` times the maximum of each of the densities in the numerator and denominator respectively (see Bowman and Azzalini, 1997).

The log-risk function *rho*, given by $rho = \log[r]$, is argued to be preferable in practice as it imparts a sense of symmetry in the way the case and control densities are treated (Kelsall and Diggle, 1995*a*,*b*). The option of log-transforming the returned risk function is therefore selected by default.

Value

An object of class `"rrs"`. This is a marked list with the following components:

`rsM` a numeric `res*res` matrix (where `res` is the grid resolution as specified in the calls to `bivariate.density` for calculation of `f` and `g`) giving the values of the risk surface over the evaluation grid. Values cor-

	responding to grid coordinates outside the study region are assigned `NA`
`f`	the object of class `"bivden"` used as the numerator or 'case' density estimate
`g`	the object of class `"bivden"` used as the denominator or 'control' density estimate
`log`	whether or not the returned risk function is on the log-scale
`pooled`	the object of class `"bivden"` (based on the pooled data) calculated internally if `f` and `g` were raw data arguments, `NA` otherwise
`P`	a numeric 50 by 50 matrix of the asymptotic p-value surface if `tolerate = TRUE` and `f` and `g` were raw data arguments, `NA` otherwise

Warning

If raw data is supplied to `risk`, as opposed to previously computed objects of class `"bivden"`, the running time of this function will be greater. This is particularly the case if the user has also selected `tolerate = TRUE`. In the same fashion as `bivariate.density` and `tolerance`, setting `comment = TRUE` can keep the user appraised of the function progress during run-time.

Author(s)

T.M. Davies, M.L. Hazelton and J.C. Marshall

Examples

```
data(PBC)
PBC.casedata <- PBC$data[PBC$data[,3]==1,1:2]
PBC.controldata <- PBC$data[PBC$data[,3]==0,1:2]

pbc.h <- OS(PBC$data[,1:2],
 nstar = sqrt(nrow(PBC.casedata) * nrow(PBC.controldata)))

pbc.pool <- bivariate.density(data = PBC$data[,1:2], pilotH = pbc.h,
 adaptive = FALSE, WIN = PBC$owin)
pbc.case <- bivariate.density(data = PBC.casedata, pilotH = pbc.h,
 adaptive = FALSE, WIN = PBC$owin)
pbc.con <- bivariate.density(data = PBC.controldata, pilotH = pbc.h,
 adaptive = FALSE, WIN = PBC$owin)

pbc.rrs <- risk(f = pbc.case, g = pbc.con, plotit = FALSE)
summary(pbc.rrs)
```

A.13 Print summary of estimated density: `summary.bivden`

`summary.bivden` *Summarising a bivariate kernel density estimate object*

Description

`print` and `summary` methods for class `"bivden"`

Usage

```
## S3 method for class 'bivden'
print(x, ...)
## S3 method for class 'bivden'
summary(object, ...)
```

Arguments

`x, object`	An object of class `"bivden"`.
`...`	Ignored.

Author(s)

T.M. Davies

A.14 Print summary of estimated risk function: `summary.rrs`

`summary.rrs` *Summarising an estimated relative risk function object*

Description

`print` and `summary` methods for class `"rrs"`

Usage

```
## S3 method for class 'rrs'
print(x, ...)
## S3 method for class 'rrs'
summary(object, ...)
```

Arguments

`x, object`	An object of class `"rrs"` resulting from a call to `risk`.
`...`	Ignored.

Author(s)

T.M. Davies

A.15 Asymptotic and numerical tolerance contours (p-value surfaces): `tolerance`

`tolerance` *Asymptotic p-value surfaces*

Description

Calculates pointwise p-values based on asymptotic theory or Monte-Carlo (MC) permutations describing the extremity of risk over a given fixed or adaptive kernel-smoothed relative risk function.

Usage

```
tolerance(rs, pooled, test = "upper",
method = "ASY", reduce = 1, ITER = 1000,
exactL2 = TRUE, comment = TRUE)
```

Arguments

`rs`	An object of class `"rrs"` resulting from a call to `risk`, giving the fixed or adaptive kernel-smoothed risk function.
`pooled`	An object of class `"bivden"` resulting from a call to `bivariate.density` (or the component `pooled` from `rs` if it was created using raw data arguments) representing a density estimate based on the 'pooled' dataset of both 'case' and 'control' points. If separate from `rs`, this pooled density estimate must follow the same smoothing approach, evaluation grid and study region window as the densities used to create `rs`.
`test`	A character string indicating the kind of test desired to yield the p-values. Must be one of `"upper"` (default - performs upper tailed tests examining heightened risk 'hotspots'), `"lower"` (lower tailed tests examining 'troughs') or `"double"` (double-sided tests). See 'Details' for further information.

`method` A character string, either "ASY" (default) or "MC" indicating which method to use for calculating the p-value surface (asymptotic and Monte-Carlo approaches respectively). The MC approach is far more computationally expensive than the asymptotic method (see 'Warnings').

`reduce` A numeric value greater than zero and less than or equal to one giving the user the option to reduce the resolution of the evaluation grid for the pointwise p-values by specifying a proportion of the size of the evaluation grid for the original density estimates. For example, if the case and control `"bivden"` objects were calculated using `res = 100` and `tolerance` was called with `reduce = 0.5`, the p-value surface will be evaluated over a 50 by 50 grid. A non-integer value resulting from use of `reduce` will be `ceiling`ed.

`ITER` An integer value specifying the number of iterations to be used if `method = "MC"` (defaulting to 1000). Non-integer numeric values are rounded. Ignored when `method = "ASY"`.

`exactL2` Ignored if `rs` (and `pooled`) are fixed-bandwidth density estimates, or if `method = "MC"`. A boolean value indicating whether or not to separately calculate the 'L2' integral components for adaptive tolerance contours. A value of `FALSE` will approximate these components based on the 'K2' integrals for faster execution (depending on the size of the evaluation grid, this improvement may be small) at the expense of a small degree of accuracy. Defaults to `TRUE`. See the reference for adaptive p-value surfaces in 'Details' for definitions of these integral components.

`comment` Boolean. Whether or not to print function progress (including starting and ending times) during execution. Defaults to `TRUE`.

Details

This function implements developments in Hazelton and Davies (2009) (fixed) and Davies and Hazelton (2010) (adaptive) to compute pointwise p-value surfaces based on asymptotic theory of kernel-smoothed relative risk surfaces. Alternatively, the user may elect to calculate the p-value surfaces using Monte-Carlo methods (see Kelsall and Diggle, 1995*a*). Superimposing upon a plot of the risk surface contours of these p-values at given significance levels (i.e. 'tolerance contours') can be an informative way of exploring the statistical significance of the extremity of risk across the defined study region. The asymptotic approach to the p-value calculation is advantageous over a Monte-Carlo method, which can lead to excessive computation time for adaptive risk surfaces and large datasets. See the aforementioned references for further comments.

Choosing different options for the argument `test` simply manipulates the 'direction' of the p-values. That is, plotting tolerance contours at a significance level of 0.05 for

a p-value surface calculated with `test = "double"` is equivalent to plotting tolerance contours at significance levels of 0.025 and 0.975 for `test = "upper"`.

Value

A list with four components:

`X`	the equally spaced sequence of length `ceiling(reduce*res)` giving the evaluation locations on the x axis (where `res` is the grid resolution as specified in the calls to `bivariate.density` for calculation of the densities for `rs` and `pooled`)
`Y`	as above, for the y axis
`Z`	a numeric `ceiling(reduce*res)*ceiling(reduce*res)` matrix giving the values of the risk surface over the evaluation grid. Values corresponding to grid coordinates outside the study region are assigned `NA`. If `method = "MC"`, this will be a single value of `NA`
`P`	a `ceiling(reduce*res)*ceiling(reduce*res)` matrix giving the p-values corresponding to the evaluation grid in light of the elected `test`

Warning

Though far less expensive computationally than calculation of Monte-Carlo p-value surfaces, the asymptotic p-value surfaces (particularly for adaptive relative risk surfaces) can still take some time to complete. The argument of `reduce` provides an option to reduce this computation time by decreasing the resolution of the evaluation grid. However, the accuracy and appearance of the resulting tolerance contours can be severely degraded if `reduce` is assigned too small a value. Care must therfore be taken and consideration given to the resolution of the original evaluation grid when altering `reduce` from its default value. For most practical purposes, we have found a value of `reduce` resulting in evaluation of a p-value surface of size 50 by 50 is adequate.

The MC approach is provided as an option here for the sake of completeness only, and is coded exclusively in `R`. The computational cost of this approach for the adaptive risk function is enough to recommend against its use in this case, though it is faster for the fixed-bandwidth case if just comparing MC execution times between the two smoothing regimens. Comments on the issue of MC vs ASY are given in Section 3 of Hazelton and Davies (2009).

Author(s)

T.M. Davies and M.L. Hazelton

Examples

```
data(chorley)
ch.h <- CV.sm(chorley)
```

```
ch.pool <- bivariate.density(data = chorley, pilotH = ch.h,
 adaptive = FALSE)
ch.case <- bivariate.density(data = chorley, ID = "larynx", pilotH = ch.h,
 adaptive = FALSE)
ch.con <- bivariate.density(data = chorley, ID = "lung", pilotH = ch.h,
 adaptive = FALSE)

##Compute log-risk surface and asymptotic p-value surfaces
ch.rrs <- risk(f = ch.case, g = ch.con)
ch.tol <- tolerance(rs = ch.rrs, pooled = ch.pool)
contour(ch.tol$X, ch.tol$Y, ch.tol$P, levels = 0.05, add = TRUE)


## Not run:
data(PBC)
PBC.casedata <- PBC$data[PBC$data[,3]==1,1:2]
PBC.controldata <- PBC$data[PBC$data[,3]==0,1:2]

pbc.rrs.rawdata <- risk(f = PBC.casedata, g = PBC.controldata,
 WIN = PBC$owin, adaptive = TRUE, tolerate = TRUE)

plot(pbc.rrs.rawdata, display = "3d", aspect = 1:2, col = heat.colors(12)[12:1],
 tolerance.matrix = pbc.rrs.rawdata$P, tol.opt = list(col = "white", raise = 0.03))


## End(Not run)
```

A.16 Internal functions: `sparr-internal`

`sparr-internal` *Internal sparr functions*

Description

Internal sparr functions, only meant to be accessed by the user directly in a life-threatening situation. Email me.

Usage

```
AREA_BASIC(ndim = 2, lower, upper, minpts = 100, functn, ...)
areaLsq(xh,WIN,iter=1000)
assignColors(cols,MAT)
compute.coord(coord,data,h,n,WIN,type,q.opt=1,forAdapt=F,counts)
```

```
dupli.data.frame(data,w,comment)
getNearest(obs,gridx,gridy,WIN,anypoint = FALSE)
getQhz_Adaptive(ARG,kType,WIN,both=T,onlyk2=F)
getQhz_Fixed(Xseq,Yseq,kType,WIN,h,both=T,onlyk2=F)
identical_windows(w1,w2)
idQuan(h,q)
KBivN_OPT(x,WIN=NULL,mu=c(0,0),Sigma=matrix(c(1,0,0,1),2,2),sq=F,
+ forSimul=F,y=NULL)
KERNEL(X,type,dim=2,Qcomp="spher")
KSPEC(location,data,h,type,counts)
LSCV.density.single(h,data.ppp,res,edge,inside,ERRV=1000000)
LSCV.risk.single(h,cases,controls,method,res,edge,inside,ERRV=1000000)
Lsq(x,uh,WIN)
nearestBC_Vals(framevals,WIN)
rsmc(object,pooled,ITER,corrgrid,comment)
run_ppp(data,xy,h,WIN,counts)
tol3d(X,Y,TOL,RHO,levels=c(0.05),raise=0.1,col="white",lwd=2,lty=1)
tol3d.component(locations,gridx,gridy,gridz,raise)
translateCoord(gridCoord,mvmat,WIN)
```

Author(s)

T.M. Davies

Appendix B

Package documentation (excerpt): lgcp

An excerpt of the current version of the package documentation including the help files of the most interesting functions and those focused on in Section 4.2. The reader is reminded that at this point in time, the documentation has been primarily authored by B.M. Taylor and are included here for reference purposes. Some differences in style and notation compared to this author's relevant discussions in Chapter 3 should be expected. The details of cited works are located together with those in the main body of this thesis in 'References'.

B.1 Space-time data set object (generic): `stppp`

`stppp` *stppp function...*

Description

stppp function

Usage

```
stppp(P, ...)
```

Arguments

`P`	an object
`...`	additional arguments

Details

Generic function used in the construction of space-time planar point patterns. An stppp object is like a ppp object, but with extra 'slots' for (1) a vector giving the time at whcih the event occurred and (2) a time-window over which observations occurred. Observations are assumed to occur in the plane and the observation window is assumed not to change over time.

Value

method stppp

See Also

stppp, stppp.ppp, stppp.list

B.2 Global spatial variation object (generic): `spatialAtRisk`

`spatialAtRisk` *spatialAtRisk function...*

Description

spatialAtRisk function

Usage

```
spatialAtRisk(X, ...)
```

Arguments

`X`	an object
`...`	additional arguments

Details

Generic function used in the construction of spatialAtRisk objects. The class of spatialAtRisk objects provide a framework for describing the spatial inhomogeneity of the at-risk population, lambda(s). This is in contrast to the class of temporalAtRisk objects, which describe the global levels of the population at risk, mu(t).

Unless the user has specified lambda(s) directly by an R function (a mapping the from the real plane onto the non-negative real numbers, see ?spatialAtRisk.function), then it is only necessary to describe the population at risk up to a constant of proportionality, as the routines automatically normalise the lambda provided to integrate to 1.

For reference purposes, the following is a mathematical description of a log-Gaussian Cox Process, it is best viewed in the pdf version of the manual.

Let $\mathcal{Y}(s,t)$ be a spatiotemporal Gaussian process, $W \subset R^2$ be an observation window in space and $T \subset R_{\geq 0}$ be an interval of time of interest. Cases occur at spatio-temporal positions $(x,t) \in W \times T$ according to an inhomogeneous spatio-temporal Cox process, i.e. a Poisson process with a stochastic intensity $R(x,t)$, The number of cases, $X_{S,[t_1,t_2]}$, arising in any $S \subseteq W$ during the interval $[t_1,t_2] \subseteq T$ is then Poisson distributed conditional on $R(\cdot)$,

$$X_{S,[t_1,t_2]} \sim \text{Poisson}\left\{\int_S \int_{t_1}^{t_2} R(s,t)dsdt\right\}$$

Following Brix and Diggle (2001) and Diggle, Rowlingson and Su (2005), the intensity is decomposed multiplicatively as

$$R(s,t) = \lambda(s)\mu(t)\exp\{\mathcal{Y}(s,t)\}.$$

In the above, the fixed spatial component, $\lambda : R^2 \mapsto R_{\geq 0}$, is a known function, proportional to the population at risk at each point in space and scaled so that

$$\int_W \lambda(s)ds = 1,$$

whilst the fixed temporal component, $\mu : R_{\geq 0} \mapsto R_{\geq 0}$, is also a known function with

$$\mu(t)\delta t = E[X_{W,\delta t}],$$

for t in a small interval of time, δt, over which the rate of the process over W can be considered constant.

Value

method spatialAtRisk

See Also

lgcpPredict, linklgcpSim, spatialAtRisk.default, spatialAtRisk.fromXYZ, spatialAtRisk.im, spatialAtRisk.function, spatialAtRisk.SpatialGridDataFrame, spatialAtRisk.SpatialPolygonsDataFrame, spatialAtRisk.bivden

B.3 Global temporal variation object (generic): `temporalAtRisk`

`temporalAtRisk`	*temporalAtRisk function...*

Description

temporalAtRisk function

Usage

```
temporalAtRisk(obj, ...)
```

Arguments

`obj`	an object
`...`	additional arguments

Details

Generic function used in the construction of temporalAtRisk objects. A temporalAtRisk object describes the at risk population globally in an observation time window [t_1,t_2]. Therefore, for any t in [t_1,t_2], a temporalAtRisk object should be able to return the global at risk population, mu(t) = E(number of cases in the unit time interval containing t). This is in contrast to the class of spatialAtRisk objects, which describe the spatial inhomogeneity in the poopulation at risk, lambda(s).

Note that in the prediction routine, lgcpPredict, and the simulation routine, lgcpSim, time discretisation is achieved using `as.integer` on both observation times and time limits t_1 and t_2 (which may be stored as non-integer values). The functions that create temporalAtRisk objects therefore return piecewise cconstant step-functions. that can be evaluated for any real t in [t_1,t_2], but with the restriction that mu(t_i) = mu(t_j) whenever `as.integer(t_i)==as.integer(t_j)`.

A temporalAtRisk object may be (1) 'assumed known', or (2) scaled to a particular dataset. In the latter case, in the routines available (temporalAtRisk.numeric and temporalAtRisk.function), the stppp dataset of interest should be referenced, in which case the scaling of mu(t) will be done automatically. Otherwise, for example for simulation purposes, no scaling of mu(t) occurs, and it is assumed that the mu(t) corresponds to the expected number of cases during the unit time interval containing t. For reference purposes, the following is a mathematical description of a log-Gaussian Cox Process, it is best viewed in the pdf version of the manual.

Let $\mathcal{Y}(s,t)$ be a spatiotemporal Gaussian process, $W \subset R^2$ be an observation window in space and $T \subset R_{\geq 0}$ be an interval of time of interest. Cases occur at spatio-temporal positions $(x,t) \in W \times T$ according to an inhomogeneous spatio-temporal Cox process, i.e. a Poisson process with a stochastic intensity $R(x,t)$, The number of cases, $X_{S,[t_1,t_2]}$, arising in any $S \subseteq W$ during the interval $[t_1,t_2] \subseteq T$ is then Poisson distributed conditional on $R(\cdot)$,

$$X_{S,[t_1,t_2]} \sim \text{Poisson}\left\{\int_S \int_{t_1}^{t_2} R(s,t)dsdt\right\}$$

Following Brix and Diggle (2001) and Diggle, Rowlingson and Su (2005), the intensity is decomposed multiplicatively as

$$R(s,t) = \lambda(s)\mu(t)\exp\{\mathcal{Y}(s,t)\}.$$

In the above, the fixed spatial component, $\lambda : R^2 \mapsto R_{\geq 0}$, is a known function, proportional to the population at risk at each point in space and scaled so that

$$\int_W \lambda(s)ds = 1,$$

whilst the fixed temporal component, $\mu : R_{\geq 0} \mapsto R_{\geq 0}$, is also a known function with

$$\mu(t)\delta t = E[X_{W,\delta t}],$$

for t in a small interval of time, δt, over which the rate of the process over W can be considered constant.

Value

method temporalAtRisk

See Also

spatialAtRisk, lgcpPredict, lgcpSim, temporalAtRisk.numeric, temporalAtRisk.function, constantInTime, constantInTime.numeric, constantInTime.stppp, print.temporalAtRisk, plot.temporalAtRisk

B.4 Interactive kernel estimation of global spatial trend: `lambdaEst`

`lambdaEst` *lambdaEst function...*

Description

lambdaEst function

Usage

```
lambdaEst(xyt, weights=c(), edge=TRUE, ...)
```

Arguments

`xyt`	object of class stppp
`weights`	Optional vector of weights to be attached to the points. May include negative values. See ?density.ppp.
`edge`	Logical flag: if TRUE, apply edge correction. See ?density.ppp.
`...`	arguments to be passed to plot

Details

A tool for the visual estimation of lambda(s) via a 2 dimensional smoothing of the case locations. For parameter estimation, the alternative is to estimate lambda(s) by some other means, convert it into a spatialAtRisk object and then into a pixel image object using the build in coercion methods, this `im` object can then be fed to ginhomAverage, KinhomAverage or thetaEst for instance.

Value

This is an rpanel function for visual choice of lambda(s), the output is a variable, varname, with the density *per unit time* the variable varname can be fed to the function ginhomAverage or KinhomAverage as the argument density (see for example ?ginhomAverage), or into the function thetaEst as the argument spatial.intensity.

See Also

spatialAtRisk, ginhomAverage, KinhomAverage, spatialparsEst, thetaEst, muEst

B.5 Nonparametric estimation of global temporal trend: `muEst`

`muEst` *muEst function...*

Description

muEst function

Usage

```
muEst(xyt, ...)
```

Arguments

`xyt`	an stppp object
`...`	additional arguments to be passed to lowess

Details

Computes a non-parametric estimate of mu(t). For the purposes of performing prediction, the alternatives are: (1) use a parameteric model as in Diggle, Rowlingson and Su (2005), or (2) a constantInTime model.

Value

object of class temporalAtRisk giving the smoothed mut using the lowess function

See Also

temporalAtRisk, constantInTime, ginhomAverage, KinhomAverage, spatialparsEst, thetaEst, lambdaEst

B.6 Time-averaged inhomogeneous PCF: `ginhomAverage`

`ginhomAverage` *ginhomAverage function...*

Description

ginhomAverage function

Usage

```
ginhomAverage(xyt, spatial.intensity, temporal.intensity,
    time.window=xyt$tlim, rvals, correction="iso",
    suppresswarnings=FALSE)
```

Arguments

`xyt`	an object of class stppp
`spatial.intensity`	A spatialAtRisk object
`temporal.intensity`	A temporalAtRisk object
`time.window`	time interval contained in the interval xyt$tlim over which to compute average. Useful if there is a lot of data over a lot of time points.
`rvals`	Vector of values for the argument r at which g(r) should be evaluated (see ?pcfinhom). There is a sensible default.
`correction`	choice of edge correction to use, see ?pcfinhom, default is Ripley isotropic correction
`suppresswarnings`	Whether or not to suppress warnings generated by pcfinhom

Details

A function to estimate the inhomogeneous pair correlation function for a spatiotemporal point process. See equation (8) of Diggle, Rowlingson and Su (2005).

Value

time average of inhomogenous pcf, equation (13) of Brix and Diggle (2001).

See Also

KinhomAverage, spatialparsEst, thetaEst, lambdaEst, muEst

B.7 Time-averaged inhomogeneous K-function: `KinhomAverage`

`KinhomAverage` *KinhomAverage function...*

Description

KinhomAverage function

Usage

```
KinhomAverage(xyt, spatial.intensity, temporal.intensity,
    time.window=xyt$tlim, rvals, correction="iso",
    suppresswarnings=FALSE)
```

Arguments

`xyt`	an object of class stppp
`spatial.intensity`	A spatialAtRisk object
`temporal.intensity`	A temporalAtRisk object
`time.window`	time interval contained in the interval xyt$tlim over which to compute average. Useful if there is a lot of data over a lot of time points.
`rvals`	Vector of values for the argument r at which the inhmogeneous K function should be evaluated (see ?Kinhom). There is a sensible default.
`correction`	choice of edge correction to use, see ?Kinhom, default is Ripley's isotropic correction
`suppresswarnings`	Whether or not to suppress warnings generated by Kinhom

Details

A function to estimate the inhomogeneous K function for a spatiotemporal point process. The method of computation is similar to ginhomAverage, see eq (8) of Diggle, Rowlingson and Su (2005) to see how this is computed.

Value

time average of inhomogenous K function.

See Also

ginhomAverage, spatialparsEst, thetaEst, lambdaEst, muEst

B.8 Interactive spatial minimum contrast estimation: `spatialparsEst`

`spatialparsEst` *spatialparsEst function...*

Description

spatialparsEst function

Usage

```
spatialparsEst(gk, sigma.range, phi.range)
```

Arguments

`gk`	an R object; output from the function KinhomAverage or ginhomAverage
`sigma.range`	range of sigma values to consider
`phi.range`	range of phi values to consider

Details

Having estimated either the pair correlation or K functions using respectively ginhomAverage or KinhomAverage, the spatial parameters sigma and phi can be estimated. This function provides a visual tool for this estimation procedure.

Value

rpanel function to help choose sigma and phi by eye

See Also

ginhomAverage, KinhomAverage, thetaEst, lambdaEst, muEst

B.9 Interactive temporal contrast estimation: `thetaEst`

`thetaEst` *thetaEst function...*

Description

thetaEst function

Usage

```
thetaEst(xyt, spatial.intensity, temporal.intensity, sigma, phi,
    theta.range=c(0, 10), N=100, spatial.covmodel="exponential",
    covpars=c())
```

Arguments

`xyt`	object of class stppp
`spatial.intensity`	A spatial at risk object OR a bivariate density estimate of lambda, an object of class im (produced from density.ppp for example),
`temporal.intensity`	either an object of class temporalAtRisk, or one that can be coerced into that form. If NULL (default), this is estimated from the data, seee ?muEst
`sigma`	estimate of parameter sigma
`phi`	estimate of parameter phi
`theta.range`	range of theta values to consider
`N`	number of integration points in computation of C(v,beta) (see Brix and Diggle (2003), corrigendum to Brix and Diggle (2001))
`spatial.covmodel`	spatial covariance model
`covpars`	additional covariance parameters

Details

A tool to visually estimate the temporal correlation parameter theta; note that sigma and phi must have first been estimated.

Value

An r panel tool for visual estimation of temporal parameter theta NOTE if lambdaEst has been invoked to estimate lambda, then the returned density should be passed to thetaEst as the argument spatial.intensity

See Also

ginhomAverage, KinhomAverage, spatialparsEst, lambdaEst, muEst

B.10 Conditional simulation of a space-time LGCP via the MALA: `lgcpPredict`

`lgcpPredict` *lgcpPredict function...*

Description

lgcpPredict function

Usage

```
lgcpPredict(xyt, T, laglength, model.parameters=lgcppars(),
    spatial.covmodel="exponential", covpars=c(), cellwidth, gridsize,
    spatial.intensity, temporal.intensity, mcmc.control,
    output.control=setoutput(), autorotate=FALSE, gradtrunc)
```

Arguments

`xyt`	a spatio-temporal point pattern object, see ?stppp
`T`	time point of interest
`laglength`	specifies lag window, so that data from and including time (T-laglength) to time T is used in the MALA algorithm
`model.parameters`	values for parameters, see ?lgcppars
`spatial.covmodel`	correlation type see ?CovarianceFct
`covpars`	vector of additional parameters for certain classes of covariance function (eg Matern), these must be supplied in the order given in ?CovarianceFct
`cellwidth`	width of grid cells on which to do MALA (grid cells are square). Note EITHER gridsize OR cellwidthe must be specified.

`gridsize` size of output grid required. Note EITHER gridsize OR cellwidthe must be specified.

`spatial.intensity`
the fixed spatial component: an object of that can be coerced to one of class spatialAtRisk

`temporal.intensity`
the fixed temporal component: either a numeric vector, or a function that can be coerced into an object of class temporalAtRisk

`mcmc.control` MCMC paramters, see ?mcmcpars

`output.control`
output choice, see ?setoutput

`autorotate` logical: whether or not to automatically do MCMC on optimised, rotated grid.

`gradtrunc` truncation for gradient vector equal to H parameter Møller et al. (1998) pp 473. Set to NULL to estimate this automatically (default). Set to zero for no gradient truncation. further notes on autorotate argument: If set to TRUE, and the argument spatial is not NULL, then the argument spatial must be computed in the original frame of reference (ie NOT in the rotated frame). Autorotate performs bilinear interpolation (via interp.im) on an inverse transformed grid; if there is no computational advantage in doing this, a warning message will be issued. Note that best accuracy is achieved by manually rotating xyt and then computing spatial on the transformed xyt and finally feeding these in as arguments to the function lgcpPredict. By default autorotate is set to FALSE.

Details

The function `lgcpPredict` performs spatiotemporal prediction for log-Gaussian Cox Processes

The following is a mathematical description of a log-Gaussian Cox Process, it is best viewed in the pdf version of the manual.

Let $\mathcal{Y}(s,t)$ be a spatiotemporal Gaussian process, $W \subset R^2$ be an observation window in space and $T \subset R_{\geq 0}$ be an interval of time of interest. Cases occur at spatio-temporal positions $(x,t) \in W \times T$ according to an inhomogeneous spatio-temporal Cox process, i.e. a Poisson process with a stochastic intensity $R(x,t)$, The number of cases, $X_{S,[t_1,t_2]}$, arising in any $S \subseteq W$ during the interval $[t_1,t_2] \subseteq T$ is then Poisson distributed conditional on $R(\cdot)$,

$$X_{S,[t_1,t_2]} \sim \text{Poisson}\left\{\int_S \int_{t_1}^{t_2} R(s,t) ds dt\right\}$$

Following Brix and Diggle (2001) and Diggle, Rowlingson and Su (2005), the intensity is decomposed multiplicatively as

$$R(s,t) = \lambda(s)\mu(t)\exp\{\mathcal{Y}(s,t)\}.$$

In the above, the fixed spatial component, $\lambda : R^2 \mapsto R_{\geq 0}$, is a known function, proportional to the population at risk at each point in space and scaled so that

$$\int_W \lambda(s)ds = 1,$$

whilst the fixed temporal component, $\mu : R_{\geq 0} \mapsto R_{\geq 0}$, is also a known function with

$$\mu(t)\delta t = E[X_{W,\delta t}],$$

for t in a small interval of time, δt, over which the rate of the process over W can be considered constant.

NOTE: the xyt stppp object can be recorded in continuous time, but for the purposes of prediciton, discretisation must take place. For the time dimension, this is achieved invisibly by `as.integer(xyt$t)` and `as.integer(xyt$tlim)`. Therefore, before running an analysis please make sure that this is commensurate with the physical inerpretation and requirements of your output. The spatial discretisation is chosen with the argument cellwidth (or gridsize). If the chosen discretisation in time and space is too coarse for a given set of parameters (sigma, phi and theta) then the proper correlation structures implied by the model will not be captured in the output.

Before calling this function, the user must decide on the time point of interest, the number of intervals of data to use, the parameters, spatial covariance model, spatial discretisation, fixed spatial ($\lambda(s)$) and temporal ($\mu(t)$) components, mcmc parameters, and whether or not any output is required.

Value

the results of fitting the model in an object of class `lgcpPredict`

See Also

KinhomAverage, ginhomAverage, lambdaEst, muEst, spatialparsEst, thetaEst, spatialAtRisk, temporalAtRisk, lgcppars, CovarianceFct, mcmcpars, setoutput print.lgcpPredict, xvals.lgcpPredict, yvals.lgcpPredict, plot.lgcpPredict, meanfield.lgcpPredict, rr.lgcpPredict, serr.lgcpPredict, intens.lgcpPredict, varfield.lgcpPredict, gridfun.lgcpPredict, gridav.lgcpPredict, hvals.lgcpPredict, window.lgcpPredict, mcmctrace.lgcpPredict, plotExceed.lgcpPredict, quantile.lgcpPredict, identify.lgcpPredict, expectation.lgcpPredict, extract.lgcpPredict, showGrid.lgcpPredict, computeGradtrunc

B.11 Plotting posterior means from a MALA run: `plot.lgcpPredict`

`plot.lgcpPredict` *plot...*

Description

plot.lgcpPredict function

Usage

```
## S3 method for class 'lgcpPredict'
 plot(x, type="relrisk", sel=1:x$EY.mean$len,
    plotdata=TRUE, ask=TRUE, clipWindow=TRUE, ...)
```

Arguments

`x`	an object of class lgcpPredict
`type`	Character string: what type of plot to produce. Choices are "relrisk" (=exp(Y)); "serr" (standard error of relative risk); or "intensity" (=lambda*mu*exp(Y)).
`sel`	vector of integers between 1 and grid$len: which grids to plot. Default NULL, in which case all grids are plotted.
`plotdata`	whether or not to overlay the data
`ask`	logical; if TRUE the user is asked before each plot
`clipWindow`	whether to plot grid cells outside the observation window
`...`	additional arguments passed to image.plot

Details

Simple plotting function for objects of class `lgcpPredict`.

Value

plots the Monte Carlo mean of Y obtained via simulation

See Also

lgcpPredict

Bibliography

Abramson, I. S. (1982), 'On bandwidth estimation in kernel estimates – a square root law', *The Annals of Statistics* **10**(4), 1217–1223.

Adler, D. and Murdoch, D. (2009), **rgl***: 3D visualization device system (OpenGL)*. R package version 0.91.
URL: *http://CRAN.R-project.org/package=rgl*

Andrieu, C. and Thoms, J. (2008), 'A tutorial on adaptive MCMC', *Statistics & Computing* **18**, 343–373.

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

Baddeley, A. J. and Silverman, B. (1984), 'A cautionary example for the use of second-order methods for analysing point patterns', *Biometrics* **40**, 1089–1094.

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

Bebbington, M. and Harte, D. (2001), 'On the statistics of the linked stress-release model', *Journal of Applied Probability* **38A**, 176–187.

Benschop, J., Hazelton, M. L., Stevenson, M. A., Dahl, J., Morris, R. S. and French, N. (2008), 'Descriptive spatial epidemiology of subclinical Salmonella infection in Danish finisher pig herds: application of a novel method of spatially adaptive smoothing', *Veterinary Research* **39**(1), 02:1–11.

Besag, J. E. (1977), 'Contribution to the discussion of Dr Ripley's paper', *Journal of the Royal Statistical Society Series B* **39**, 193–195.

Besag, J. E. (1994), 'Discussion of the paper by Grenander and Miller', *Journal of the Royal Statistical Society Series B* **56**, 591–592.

Bithell, J. F. (1990), 'An application of density estimation to geographical epidemiology', *Statistics in Medicine* **9**, 691–701.

Bithell, J. F. (1991), 'Estimation of relative risk functions', *Statistics in Medicine* **10**, 1745–1751.

Bivand, R. S., Pebesma, E. J. and Gómez-Rubio, V. (2008), *Applied Spatial Data Analysis with R*, Springer, New York.

Bowman, A. W. and Azzalini, A. (1997), *Applied Smoothing Techniques for Data Analysis: The Kernel Approach with S-Plus Illustrations*, Oxford University Press Inc., New York.

Bowman, A. W., Crawford, E., Alexander, G. and Bowman, R. W. (2007), '**rpanel**: Simple interactive controls for R functions using the **tcltk** package', *Journal of Statistical Software* **17**(9), 1–19.

Breiman, L., Meisel, W. and Purcell, E. (1977), 'Variable kernel estimates of probability density estimates', *Technometrics* **19**, 135–144.

Brewer, M. J. (1998), A modelling approach to bandwidth selection in kernel density estimation, *in* 'Proceedings in Computational Statistics (COMPSTAT)', Bristol, Great Britain.

Brewer, M. J. (2000), 'A Bayesian model for local smoothing in kernel density estimation', *Statistics and Computing* **10**, 299–309.

Brix, A. and Diggle, P. J. (2001), 'Spatiotemporal prediction for log-Gaussian Cox processes', *Journal of the Royal Statistical Society Series B* **63**(4), 823–841.

Brix, A. and Diggle, P. J. (2003), 'Corrigendum: Spatiotemporal prediction for log-Gaussian Cox processes', *Journal of the Royal Statistical Society Series B* **65**(4), 946.

Cacoullos, T. (1966), 'Estimation of a multivariate density', *Annals of the Institute of Statistical Mathematics* **18**, 179–189.

Capua, I. and Alexander, D. (2010), 'Perpectives on the global threat: The challenge of avian influenza viruses for the world's veterinary community', *Avian Diseases* **54**, 176–178.

Christensen, O. F., Møller, J. and Waagepetersen, R. P. (2001), 'Geometric ergodicity of Metropolis-Hastings algorithms for conditional simulation in generalized linear mixed models', *Methodology and Computing in Applied Probability* **3**, 309–327.

Christensen, O. F., Roberts, G. O. and Rosenthal, J. S. (2003), 'Scaling limits for the transient phase of local Metropolis-Hastings algorithms', *Journal of the Royal Statistical Society Series B* **67**, 253–268.

Christensen, O. F. and Waagepetersen, R. P. (2002), 'Bayesian prediction of spatial count data using generalized linear mixed models', *Biometrics* **58**, 280–286.

Cleveland, W. S. (1979), 'Robust locally weighted regression and smoothing scatterplots', *Journal of the American Statistical Association* **74**, 829–836.

Cleveland, W. S. (1981), 'LOWESS: A program for smoothing scatterplots by robust locally weighted regression', *The American Statistician* **35**, 54.

Clough, H. E., Fenton, S. E., French, N. P., Miller, A. J. and Cook, A. J. C. (2009), 'Evidence from the UK Zoonoses Action Plan in favour of localised anomalies of salmonella infection on United Kingdom pig farms', *Preventive Veterinary Medicine* **89**(1-2), 67–74.

Cox, D. R. (1955), 'Some statistical methods related with series of events (with discussion)', *Journal of the Royal Statistical Society Series B* **17**, 129–164.

Cuzick, J. and Edwards, R. (1990), 'Spatial clustering for inhomogeneous populations', *Journal of the Royal Statistical Society Series B* **52**, 73–104.

Daley, D. J. and Vere-Jones, D. (2003), *Introduction to the theory of point processes*, 2nd edn, Springer, New York.

Danese, M., Lazzari, M. and Murgante, B. (2008), Kernel density estimation methods for a geostatistical approach in seismic risk analysis: the case study of potenza hilltop town (southern italy), *in* O. Gervasi, B. Murgante, A. Laganà, D. Taniar, Y. Mun and M. Gavrilova, eds, 'Lecture Notes in Computer Science vol. 5072', Springer-Verlag, Berlin, pp. 415–429.

Davies, T. M. (2007), Kernel density estimation for the analysis of relative risk in geographical epidemiology. Honours research dissertation (unpublished).

Davies, T. M. and Hazelton, M. L. (2010), 'Adaptive kernel estimation of spatial relative risk', *Statistics in Medicine* **29**(23), 2423–2437.

Davies, T. M., Hazelton, M. L. and Marshall, J. C. (2011), '**sparr**: Analyzing spatial relative risk using fixed and adaptive kernel density esimation in R', *Journal of Statistical Software* **39**(1), 1–14.

Dawson, B. and Trapp, R. G. (2004), *Basic & Clinical Biostatistics*, McGraw-Hill, U.S.A.

De Cesare, L., Myers, D. and Posa, D. (2001), 'Estimating and modeling space-time correlation structures', *Statistics and Probability Letters* **51**, 9–14.

Diggle, P. J. (1985), 'A kernel method for smoothing point process data', *Applied Statistics* **34**(2), 138–147.

Diggle, P. J. (1990), 'A point process modelling approach to raised incidence of a rare phenomenon in the vicinity of a prespecified point', *Journal of the Royal Statistical Society Series A* **153**, 349–362.

Diggle, P. J. (2003), *Statistical Analysis of Spatial Point Patterns*, Hodder Arnold, Great Britain.

Diggle, P. J., Rowlingson, B. and Su, T. L. (2005), 'Point process methodology for on-line spatio-temporal disease surveillance', *Environmetrics* **16**(5), 423–434.

Diggle, P. J., Zheng, P. and Durr, P. (2005), 'Nonparametric estimation of spatial segregation in a multivariate point process: bovide tuberculosis in Cornwall, UK', *Journal of the Royal Statistical Society Series C: Applied Statistics* **54**, 645–658.

Dubé, C., Stevenson, M. A., Garner, M. G., Sanson, R. L., Corso, B. A., Harvey, N., Griffin, J., Wilesmith, J. W. and Estrada, C. (2007), 'A comparison of predictions made by three simulation models of foot-and-mouth disease', *New Zealand Veterinary Journal* **55**, 280–288.

Duong, T. (2007), '**ks**: Kernel density estimation and kernel discriminant analysis for multivariate data in R', *Journal of Statistical Software* **21**(7), 1–16.

Ekström, G., Dziewonski, A. M., Maternovskaya, N. N. and Nettles, M. (2005), 'Global seismicity of 2003: centroid-moment-tensor solutions for 1087 earthquakes', *Physics of the Earth and Planetary Interiors* **148**, 327–351.

Furrer, R., Nychka, D. and Sain, S. (2011), **fields***: Tools for spatial data*. R package version 6.6.
URL: *http://CRAN.R-project.org/package=fields*

Garner, M. G. and Beckett, S. (2005), 'Modelling the spread of foot-and-mouth disease in Australia', *Australian Veterinary Journal* **83**, 30–38.

Ghosh, M. and Rao, J. N. K. (1994), 'Small area estimation: an appraisal', *Statistical Science* **9**, 55–76.

Gneiting, T. (2002), 'Nonseparable, stationary covariance functions for space-time data', *Journal of the American Statistical Association* **97**, 590–600.

Guan, Y. (2007), 'A least-squares cross-validation bandwidth selection approach in pair correlation function estimations', *Statistics and Probability Letters* **77**, 1722–1729.

Guan, Y. (2009), 'A minimum contrast estimation procedure for estimating the second-order parameters of inhomogeneous spatial point processes', *Statistics and Its Interface* **2**, 91–99.

Guan, Y. and Sherman, M. (2007), 'On least squares fitting for stationary spatial point processes', *Journal of the Royal Statistical Society Series B* **69**, 31–49.

Hall, P., Hu, T. C. and Marron, J. S. (1995), 'Improved variable window kernel estimates of probability densities', *The Annals of Statistics* **23**, 1–10.

Hall, P. and Marron, J. S. (1987), 'Extent to which least-squares cross-validation minimises integrated squared error in nonparametric density estimation', *Probability Theory and Related Fields* **74**, 567–581.

Hall, P. and Marron, J. S. (1988), 'Variable window width kernel density estimates of probability densities', *Probability Theory and Related Fields* **80**, 37–49.

Harvey, N., Reeves, A., Schoenbaum, M. A., Zagmutt-Vergara, F. J., Dubé, C., Hill, A. E., Corso, B. A., McNab, B., Cartwright, C. I. and Salman, M. D. (2007), 'The North American Animal Disease Spread Model: a simulation model to assist decision making in evaluating animal disease incursions', *Preventative Veterinary Medicine* **82**, 176–197.

Hastings, W. K. (1970), 'Monte Carlo sampling methods using Markov chains and their applications', *Biometrika* **57**, 97–109.

Hawkes, A. G. (1971), 'Spectra of some self-exciting and mutually exciting point processes', *Biometrika* **58**, 83–90.

Hazelton, M. L. (2008), 'Letter to the editor: Kernel estimation of risk surfaces without the need for edge correction', *Statistics in Medicine* **27**, 2269–2272.

Hazelton, M. L. and Davies, T. M. (2009), 'Inference based on kernel estimates of the relative risk function in geographical epidemiology', *Biometrical Journal* **51**, 98–109.

Heinrich, L. (1993), 'Asymptotic properties of minimum contrast estimators for parameters of Boolean models', *Metrika* **40**, 67–94.

Illian, J., Penttinen, A., Stoyan, H. and Stoyan, D. (2008), *Statistical Analysis and Modelling of Spatial Point Patterns*, Wiley, Great Britain.

Jones, M. C. (1990), 'Variable kernel density estimates and variable kernel density estimates', *Australian and New Zealand Journal of Statistics* **32**, 361–371.

Keitt, T. H., Bivand, R. S., Pebesma, E. J. and Rowlingson, B. S. (2011), **rgdal***: Bindings for the Geospatial Data Abstraction Library*. R package version 0.7-1.
URL: *http://CRAN.R-project.org/package=rgdal*

Kelsall, J. E. and Diggle, P. J. (1995*a*), 'Kernel estimation of relative risk', *Bernoulli* **1**, 3–16.

Kelsall, J. E. and Diggle, P. J. (1995*b*), 'Non-parametric estimation of spatial variation in relative risk', *Statistics in Medicine* **14**, 2335–2342.

Kulldorff, M. (1997), 'A spatial scan statistic', *Communications in Statistics: Theory and Methods* **26**, 1487–1496.

Kulldorff, M., Athas, W. F., Feuer, E. J., Miller, B. A. and Key, C. R. (1998), 'Evaluating cluster alarms: A space-time statistic and brain cancer in Los Alamos, New Mexico', *American Journal of Public Health* **88**, 1377–1380.

Kulldorff, M., Heffernan, R., Hartman, J., Assunção, R. and Mostashari, F. (2005), 'A space-time permutation scan statistic for disease outbreak detection', *PLoS Medicine* **2**, 216–224.

Lawson, A. B. (2006), *Statistical Methods in Spatial Epidemiology*, 2nd edn, Wiley, Great Britain.

Leonard, B. (1998), *Leukemia: A Research Report*, DIANE Publishing, U.S.A. National Institutes of Health (National Cancer Institute).

Lewin-Koh, N. J. and Bivand, R. S. (2011), **maptools***: Tools for reading and handling spatial objects*. R package version 0.8-10.
URL: *http://CRAN.R-project.org/package=maptools*

Liang, S., Carlin, B. P. and Gelfand, A. E. (2009), 'Analysis of Minnesota colon and rectum cancer point patterns with spatial and nonspatial covariate information', *Annals of Applied Statistics* **3**, 943–962.

Loftsgaarden, D. O. and Quesenberry, C. P. (1965), 'A nonparametric estimate of a multivariate density function', *Annals of Mathematical Statistics* **36**, 1049–1051.

Marshall, J. C. and Hazelton, M. L. (2010), 'Boundary kernels for adaptive density estimators on regions with irregular boundaries', *Journal of Multivariate Analysis* **101**, 949–963.

McManus, D. P., Feng, Z., Guo, J., Li, Y., Bartley, P. B., Loukas, A. and Williams, G. M. (2005), Pathways to improved, sustainable morbidity control and prevention of schistosomiasis in the People's Republic of China, *in* W. E. Secor and D. G. Colley, eds, 'Schistosomiasis', Springer, U.S.A., pp. 159–175.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H. and Teller, E. (1953), 'Equation of state calculations by fast computing machines', *Journal of Chemical Physics* **21**, 1087–1092.

Møller, J. (2003), 'Shot-noise Cox processes', *Advances in Applied Probability* **35**, 614–640.

Møller, J. and Díaz-Avalos, C. (2010), 'Structured spatio-temporal shot-noise Cox point process models, with a view to modelling forest fires', *Scandinavian Journal of Statistics* **37**, 2–25.

Møller, J., Syversveen, A. R. and Waagepetersen, R. P. (1998), 'Log-Gaussian Cox processes', *Scandianvian Journal of Statistics* **25**, 451–482.

Møller, J. and Waagepetersen, R. (2004), *Statistical Inference and Simulation for Spatial Point Processes*, Chapman and Hall, Boca Raton, USA.

Møller, J. and Waagepetersen, R. (2007), 'Modern statistics for spatial point processes', *Scandinavian Journal of Statistics* **34**, 643–684.

Nelder, J. A. and Mead, R. (1965), 'A simplex algorithm for function minimization', *Computer Journal* **7**, 308–313.

Ogata, Y. (1988), 'Statistical models for earthquake occurrence and residual analysis for point processes', *Journal of the American Statistical Association* **83**, 9–27.

Ogata, Y. (1998), 'Space-time point-process models for earthquake occurrences', *Annals of the Institute of Statistical Mathematics* **50**, 379–402.

Oyana, T. J., Dai, D. and Scott, K. E. (2006), 'Spatiotemporal distributions of reported cases of the avian influenza H5N1 (bird flu) in Southern China in early 2004', *Avian Diseases* **50**, 508–515.

Parzen, E. (1962), 'On estimation of a probability density function and mode', *Annals of Mathematical Statistics* **33**, 1065–1076.

Paul, M., Tavornpanich, S., Abrial, D., Gasqui, P., Charras-Garrido, M., Thanapongtharm, W., Xiao, X., Gilbert, M., Roger, F. and Ducrot, C. (2010), 'Anthropogenic factors and the risk of highly pathogenic avian influenza H5N1: Prospects from a spatial-based model', *Veterinary Research* **41**, 28.

Pebesma, E. J. and Bivand, R. S. (2005), 'Classes and methods for spatial data in R', *R News* **5**(2).
URL: *http://cran.r-project.org/doc/Rnews/*

Pierce, D. (2011), **ncdf***: Interface to Unidata netCDF data files*. R package version 1.6.6.
URL: *http://CRAN.R-project.org/package=ncdf*

Prince, M. I., Chetwynd, A., Diggle, P. J., Jarner, M., Metcalf, J. V. and James, O. F. W. (2001), 'The geographical distribution of primary biliary cirrhosis in a well-defined cohort', *Hepatology* **34**, 1083–1088.

R Development Core Team (2011), *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.
URL: *http://www.R-project.org*

Revolution Analytics (2011), **iterators***: Iterator construct for R*. R package version 1.0.5.
URL: *http://CRAN.R-project.org/package=iterators*

Ripley, B. D. (1977), 'Modelling spatial patterns (with discussion)', *Journal of the Royal Statistical Society Series B* **39**, 172–212.

Roberts, G. O. and Rosenthal, J. S. (1998), 'Optimal scaling of discrete approximations to langevin diffusions', *Journal of the Royal Statistical Society Series B* **60**, 255–268.

Roberts, G. O. and Tweedie, R. L. (1997), 'Exponential convergence of Langevin diffusions and their discrete approximations', *Bernoulli* **2**, 341–363.

Rodrigues, A. and Diggle, P. J. (2010), 'A class of convolution-based models for spatio-temporal processes with non-separable covariance structure', *Scandinavian Journal of Statistics* **37**, 553–567.

Rosenblatt, M. (1956), 'Remarks on some nonparametric estimates of a density function', *Annals of Mathematical Statistics* **27**, 832–837.

Sabel, C. E., Gatrell, A. C., Löytönen, M., Maasilta, P. and Jokelainen, M. (2000), 'Modelling exposure opportunitites: Estimating relative risk for motor disease in Finland', *Social Science & Medicine* **50**, 1121–1137.

Sanson, R. L. (1993), The development of a decision support system for an animal disease emergency, PhD thesis, Massey University, Palmerston North, New Zealand. 264 pp.

Sanson, R. L., Harvey, N., Garner, M. G., Stevenson, M. A., Davies, T. M., Hazelton, M. L., O'Connor, J., Dubé, C., Forde-Folle, K. N. and Owen, K. (2011), 'Foot-and-mouth disease model verification and 'relative validation' through a formal model comparison', *Revue Scientifique et Technique-Office International des Epizooties* **30**, 527–540.

Schlather, M. (2011), **RandomFields***: Simulation and Analysis of Random Fields*. R package version 2.0.45.
URL: *http://CRAN.R-project.org/package=RandomFields*

Scott, D. W. (1992), *Multivariate Density Estimation: Theory, Practice, and Visualization*, Wiley, New York.

Scott, D. W. and Terrell, G. R. (1987), 'Biased and unbiased cross-validation in density estimation', *Journal of the American Statistical Association* **82**, 1131–1146.

Si, Y., Skidmore, A. K., Wang, T., de Boer, W. F., Debba, P., Toxopeus, A. G., Li, L. and Prins, H. H. (2009), 'Spatio-temporal dynamics of global H5N1 outbreaks match bird migration patterns', *Geospatial Health* **4**, 65–78.

Silverman, B. W. (1986), *Density Estimation for Statistics and Data Analysis*, Chapman and Hall, New York.

Simonoff, J. S. (1996), *Smoothing Methods in Statistics*, Springer-Verlag, New York.

Snow, J. (1855), *On the Mode of Communication of Cholera*, 2nd edn, Churchill-Livingstone, London.

Stoyan, D. and Stoyan, H. (1994), *Fractals, Random Shapes and Point Fields*, Wiley, Chichester, U.K.

Stoyan, D. and Stoyan, H. (2000), 'Improving ratio estimators of second order point process characteristics', *Scandinavian Journal of Statistics* **27**, 641–656.

Taylor, B. M., Davies, T. M., Rowlingson, B. S. and Diggle, P. J. (2012), '**lgcp** – An R package for inference with spatial and spatiotemporal log-Gaussian Cox processes', *Journal of Statistical Software* [**in press**].

Taylor, N. (2003), *Review of the use of models in informing disease control policy development and adjustment*, Department for Environment, Food and Rural Affairs, London. Accessed 10/10/2011.
URL: *http://www.defra.gov.uk/science/documents/publications/2003/UseofModelsinDiseaseControlPolicy.pdf*

Terrell, G. R. (1990), 'The maximal smoothing principle in density estimation', *Journal of the American Statistical Association* **85**, 470–477.

Toma, B., Vaillancourt, J. P., Dufour, B., Eloit, M., Moutou, F., Marsh, W., Benet, J. J., Sanaa, M. and Michel, P., eds (1999), *Dictionary of Veterinary Epidemiology*, Iowa State University Press, U.S.A.

Tscheschel, A. and Stoyan, D. (2006), 'Statistical reconstruction of random point patterns', *Computational Statistics & Data Analysis* **51**, 859–871.

Venables, W. N. and Ripley, B. D. (2002), *Modern Applied Statistics with S*, 4th edn, Springer, New York. ISBN 0-387-95457-0.
URL: *http://www.stats.ox.ac.uk/pub/MASS4*

Vere-Jones, D. (2009), 'Some models and procedures for space-time point processes', *Environmental and Ecological Statistics* **16**, 173–195.

Vere-Jones, D., Ben-Zion, Y. and Zuniga, R., eds (2005), *Statistical Seismology*, Vol. 162, Pure and Applied Geophysics.

Waagepetersen, R. P. and Guan, Y. (2009), 'Two-step estimation for inhomogeneous spatial point processes', *Journal of the Royal Statistical Society Series B* **71**, 685–702.

Waller, L. A. and Gotway, C. A. (2004), *Applied Spatial Statisics for Public Health Data*, Wiley, New Jersey, U.S.A.

Wand, M. P. and Jones, M. C. (1995), *Kernel Smoothing*, Chapman and Hall, Boca Raton, USA.

Wand, M. P. and Ripley, B. D. (2011), **KernSmooth***: Functions for kernel smoothing for Wand & Jones (1995).* R package version 2.23-6.
URL: *http://CRAN.R-project.org/package=KernSmooth*

Wheeler, D. C. (2007), 'A comparison of spatial clustering and cluster detection techniques for childhood leukemia incidence in Ohio, 1996–2003', *International Journal of Health Geographics* **6**(13).

Williams, E. H., Smith, P. G., Day, N. E., Geser, A., Ellice, J. and Tukei, P. (1978), 'Space-time clustering of Burkitt's lymphoma in the West Nile district of Uganda 1961-1975', *British Journal of Cancer* **37**, 109–122.

Wood, A. T. A. and Chan, G. (1994), 'Simulation of stationary Gaussian processes in $[0, 1]^d$', *Journal of Computational and Graphical Statistics* **3**, 409–432.

World Health Organization (2010), *WHO Factsheet: Schistosomiasis.* Accessed 3/10/2011.
URL: *http://www.who.int/mediacentre/factsheets/fs115/en/index.html*

World Health Organization (2011), *WHO Factsheet: Avian influenza.* Accessed 7/10/2011.
URL: *http://www.who.int/mediacentre/factsheets/avian\textunderscoreinfluenza/en/index.html*

Zhang, Z. J., Carpenter, T. E., Chen, Y., Clark, A. B., Lynn, H. S., Peng, W., Zhou, Y., Zhao, G. and Jiang, Q. (2008), 'Identifying high-risk regions for schistosomiasis in Guichi, China: a spatial analysis', *Acta Tropica* **107**, 217–223.

Zhang, Z. J., Chen, D., Chen, Y., Davies, T. M., Vaillancourt, J. P. and Liu, W. (2012), 'Risk signals of an influenza pandemic caused by highly pathogenic avian influenza subtype H5N1: Spatio-temporal perspectives', *The Veterinary Journal* **192**, 417–421.

Zhang, Z. J., Chen, D. M., Chen, Y., Liu, W. B., Wang, W., Zhao, F. and Yao, B. (2010), 'Spatiotemporal data comparisons for global highly pathogenic avian influenza (HPAI) H5N1 outbreaks', *PLoS ONE* **5**, e15314.

Zhang, Z. J., Clark, A. B., Bivand, R., Chen, Y., Carpenter, T. E., Peng, W., Zhou, Y., Zhao, G. and Jiang, Q. (2009), 'Nonparametric spatial analysis to detect high-risk regions for schistosomiasis in Guichi, China', *Transactions of the Royal Society of Tropical Medicine and Hygiene* **103**, 1045–1052.

Zhang, Z. J., Davies, T. M. and Jiang, Q. W. (2011), Identification of high-risk regions for schistosomiasis in the Guichi region of China: an adaptive kernel density estimation-based approach. *Submitted for publication*.

Zheng, P. and Diggle, P. J. (2011), **spatialkernel***: Nonparameteric estimation of spatial segregation in a multivariate point process*. R package version 0.4-10.
URL: *http://CRAN.R-project.org/package=spatialkernel*

Zhou, X. N., Guo, J. G., Wu, X. H., Jiang, Q. W., Zheng, J., Dang, H., Wang, X. H., Xu, J., Zhu, H. Q., Wu, G. L., Li, Y. S., Xu, X. J., Chen, H. G., Wang, T. P., Zhu, Y. C., Qiu, D. C., Dong, X. Q., Zhao, G. M., Zhang, S. J., Zhao, N. Q., Xia, G., Wang, L. Y., Zhang, S. Q., Lin, D. D., Chen, M. G. and Hao, Y. (2007), 'Epidemiology of schistosomiasis in the People's Republic of China, 2004', *Emerging Infectious Diseases* **13**, 1470–1476.