



JOHNS HOPKINS
BLOOMBERG
SCHOOL of PUBLIC HEALTH

Johns Hopkins University, Dept. of Biostatistics Working Papers

6-27-2005

Spatio-temporal Point Processes: Methods and Applications

Peter J. Diggle

Medical Statistics Unit, Lancaster University, UK & Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health, p.diggle@lancaster.ac.uk

Suggested Citation

Diggle, Peter J., "Spatio-temporal Point Processes: Methods and Applications" (June 2005). *Johns Hopkins University, Dept. of Biostatistics Working Papers*. Working Paper 78.
<http://biostats.bepress.com/jhubiostat/paper78>

This working paper is hosted by The Berkeley Electronic Press (bepress) and may not be commercially reproduced without the permission of the copyright holder.

Copyright © 2011 by the authors

Spatio-temporal Point Processes: Methods and Applications

Peter J Diggle

(Department of Mathematics and Statistics, Lancaster University
and

Department of Biostatistics, Johns Hopkins University School of Public Health)

June 27, 2005

1 Introduction

This chapter is concerned with the analysis of data whose basic format is $(x_i, t_i) : i = 1, \dots, n$ where each x_i denotes the location and t_i the corresponding time of occurrence of an *event* of interest. We shall assume that the data form a complete record of all events which occur within a pre-specified spatial region A and a pre-specified time-interval, $(0, T)$. We call a data-set of this kind a *spatio-temporal point pattern*, and the underlying stochastic model for the data a *spatio-temporal point process*.

1.1 Motivating examples

1.1.1 Amacrine cells in the retina of a rabbit

One general approach to analysing spatio-temporal point process data is to extend existing methods for purely spatial data by considering the time of occurrence as a distinguishing feature, or *mark*, attached to each event. Before giving an example of this, we give an even simpler example of a marked spatial point pattern, in which the events are of just two qualitatively different types. Each event in Figure 1 represents the location of an amacrine cell in the retina of a rabbit. These cells play a fundamental role in mammalian vision. One type transmits information when a light goes *on*, the other type similarly transmits information when a light goes *off*. The data consist of the locations of 152 on cells and 142 off cells in a rectangular region of dimension 1060 by 662 μm .

The primary goal for the analysis of these data is to discriminate between two competing developmental hypotheses. The first hypothesis is that the pattern forms initially

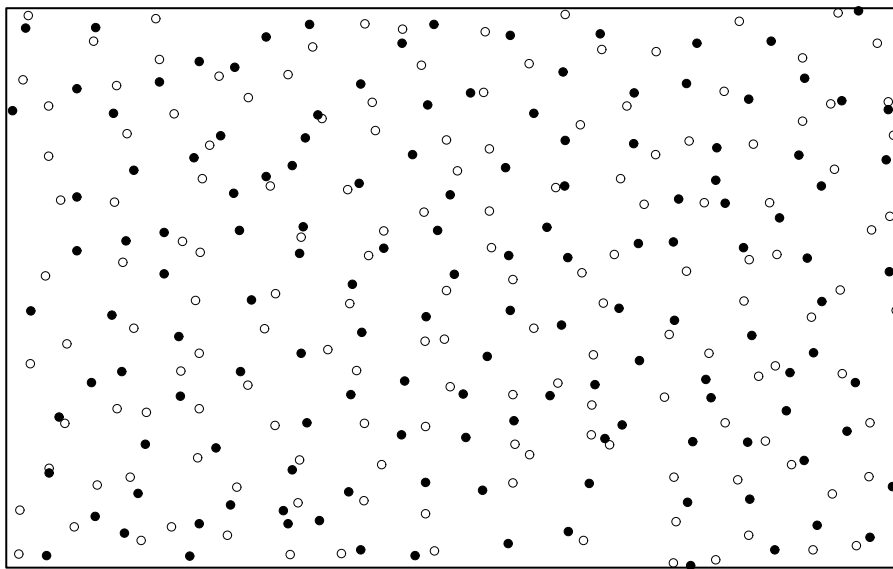


Figure 1: Amacrine cells in the retina of a rabbit. On and off cells are shown as open and closed circles, respectively. The rectangular region on which the cells are observed has dimension 1060 by 662 μm .

in two separate layers, corresponding to their pre-determined functionality, the second that the pattern forms initially in a single, undifferentiated layer with function determined at a later developmental stage. One way to formalise this in statistical terms is to ask whether the two component patterns are statistically independent. Approximate independence would favour the first hypothesis. As we shall discuss in Section 2, this statement is a slight over-simplification but it provides a sensible starting point for an analysis of the data.

Our description and later analysis of these data is based on material in Diggle, Eglen and Troy (2005). For a general discussion of the biological background, see Hughes (1985)

1.1.2 Bovine tuberculosis in Cornwall, UK

Our second example concerns the spatio-temporal distribution of reported cases of bovine tuberculosis (BTB) in the county of Cornwall, UK, over the years 1991 to 2002. Individual cases are identified from annual inspections of farm-herds, hence the effective time-resolution of the data is one year.

The prevalence of BTB has been increasing during the twelve-year period covered by the data, but the observed annual counts exaggerate this effect because the scale of the annual inspection programme has also increased. Each recorded case is classified genetically, using the method of spoligotyping (Durr, Hewinson and Clifton-Hadley, 2000). The main scientific interest in these data lies not so much in the overall spatio-temporal distribution of the disease, but rather in the degree of spatial segregation

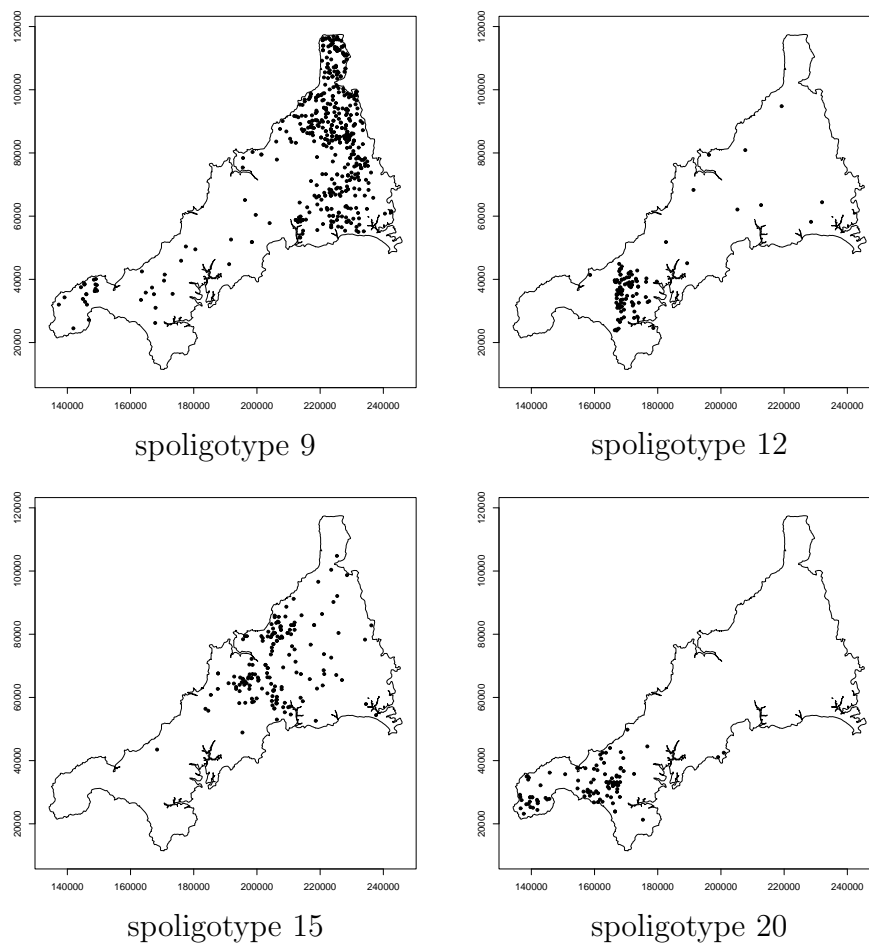


Figure 2: Spatial distributions of the four most common spoligotype data over the fourteen years.

amongst the different spoligotypes, and whether this spatial segregation is or is not stable over time. If the predominant mode of transmission is through local cross-infection, we might expect to find a stable pattern of spatial segregation, in which locally predominant spoligotypes persist over time, whereas if the disease is spread primarily by the introduction of animals from remote locations which are bought and sold at market, the resulting pattern of spatial segregation should be less stable over time (Diggle, Zheng and Durr, 2005).

Figure 2 shows the spatial distributions of cases corresponding to each of the four most common spoligotypes. The visual impression is one of strong spatial segregation, with each of the four types predominating in particular sub-regions.

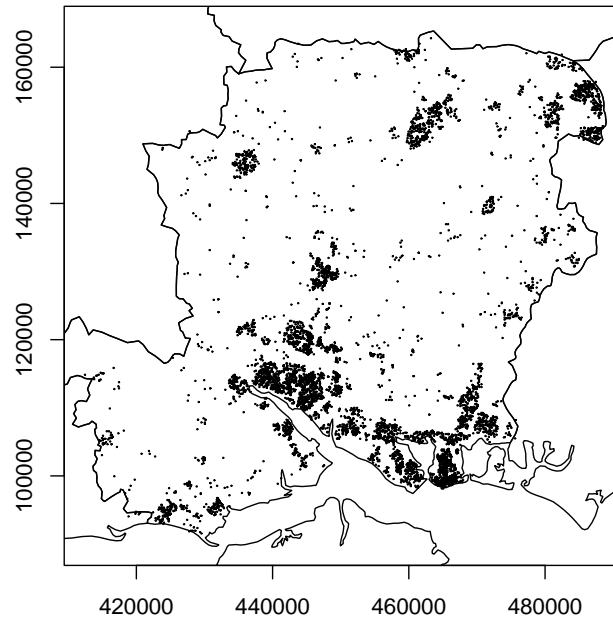


Figure 3: Locations of 7167 incident cases of non-specific gastroenteric disease in Hampshire, 1 January 2001 to 31 December 2002.

1.1.3 Gastroenteric disease in Hampshire, UK

Our third example concerns the spatio-temporal distribution of gastroenteric disease in the county of Hampshire, UK, over the years 2001 and 2002. The data are derived from calls to NHS Direct, a 24-hour, 7-day phone-in service operating within the UK's National Health Service. Each call to NHS Direct generates a data-record which includes the caller's post-code, the date of the call and a symptom code (Cooper, Smith, O'Brien, Hollyoak and Baker, 2003). Figure 3 shows the locations of the 7167 calls from patients resident in Hampshire whose assigned symptom code corresponded to acute, non-specific gastroenteric disease. The spatial distribution of cases largely reflects that of the population of Hampshire, with strong concentrations in the large cities of Southampton and Portsmouth, and smaller concentrations in other towns and villages. Inspection of a dynamic display of the space-time coordinates of the cases suggests the kind of pattern typical of an endemic disease, in which cases can occur at any point in the study region at any time during the two-year period. Occasional outbreaks of gastroenteric disease, which arise as a result of multiple infections from a common source, should result in anomalous, spatially and temporally localised concentrations of cases.

The data were collected as part of the AEGISS project (Diggle, Knorr-Held, Rowling-

son, Su, Hawtin and Bryant, 2003), whose overall aim was to improve the timeliness of the disease surveillance systems currently used in the UK. The specific statistical aims for the analysis of the data are to establish the normal pattern of spatial and temporal variation in the distribution of reported cases, and hence to develop a method of real-time surveillance to identify as quickly as possible any anomalous incidence patterns which might signal the onset of an outbreak requiring some form of public health intervention.

1.1.4 The UK 2001 epidemic of foot-and-mouth disease

Foot-and-mouth disease (FMD) is a highly infectious viral disease of farm livestock. The virus can be spread directly between animals over short distances in contaminated airborne droplets, and indirectly over longer distances, for example via the movement of contaminated material. The UK experienced a major FMD epidemic in 2001, which resulted in the slaughter of more than 6 million animals. Its estimated total cost to the UK economy was around £8 billion (UK National Audit Office, 2002). The epidemic affected 44 counties, and was particularly severe in the counties of Cumbria, in the north-west of England, and Devon, in the south-west. Figure 4 shows the spatial distributions of all farms in Cumbria and Devon which were at risk at the start of the epidemic, and of the farms which experienced the disease. In sharp contrast to the data on gastroenteric disease in Hampshire, the case-farms are strongly concentrated in sub-regions within each of the two counties. Dynamic plotting of the space-time locations of case-farms confirms the typical pattern of a highly infectious, epidemic disease. The predominant pattern is of transmission between near-neighbouring farms, but there are also a few, apparently spontaneous outbreaks of the disease far from any previously infected farms.

The main control strategies used during the epidemic involved the pre-emptive slaughter of animal-holdings at farms thought to be at high risk of acquiring, and subsequently spreading, the disease. Factors which could affect whether a farm is at high risk include, most obviously, its proximity to infected farms, but also recorded characteristics such as the size and species composition of its holding. One objective in analysing these data is to formulate and fit a model for the dynamics of the disease which incorporates these effects. A model of this kind could then provide information on what forms of control strategy would be likely to prove effective in any future epidemic.

1.2 Chapter outline

In Section 2, we give a brief review of statistical methods for spatial point patterns, illustrated by an analysis of the amacrine cell data shown in Figure 1. We refer the reader to Diggle (2003) or Moller and Waagepetersen (2004) for more detailed accounts of the methodology, and to Diggle, Eglen and Troy (2005) for a full account of the data-analysis.

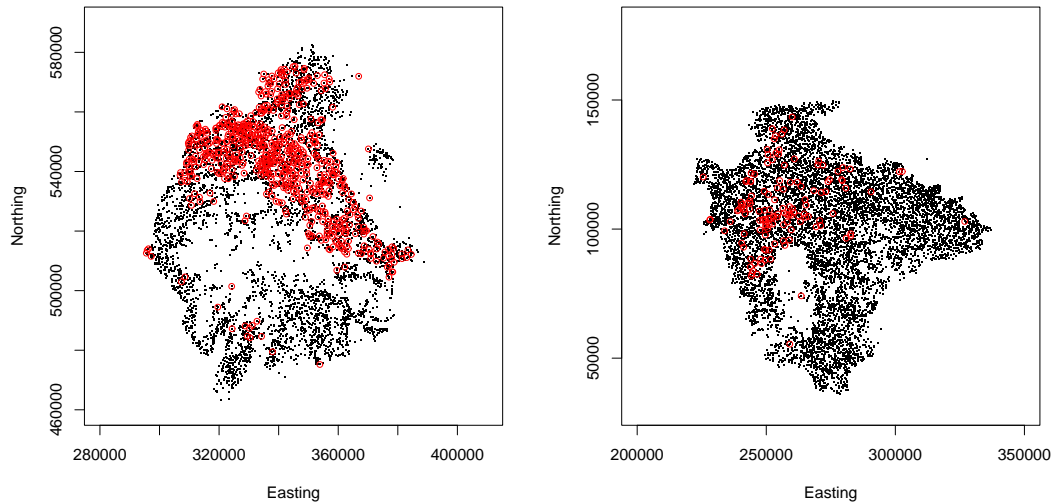


Figure 4: Locations of at-risk farms (black) and FMD-case farms (red) in Cumbria (left-hand panel) and in Devon (right-hand panel).

In Section 3, we discuss strategies for analysing spatio-temporal point process data. We argue that an important distinction in practice is between data for which the individual events (x_i, t_i) occur in a space-time continuum, and data for which the time-scale is either naturally discrete, or is made so by recording only the aggregate spatial pattern of events over a sequence of discrete time-periods. Our motivating examples include instances of each of these scenarios. Other scenarios which we do not consider further are when the locations are coarsely discretised by assigning each event to one of a number of sub-regions which form a partition of A . Methods for the analysis of spatially discrete data are typically based on Markov random field models. An early, classic reference is Besag (1974). Book-length treatments include Cressie (1991), Banerjee, Carlin and Gelfand (2003) and Rue and Held (2005).

In later sections, we describe some of the available models and methods through their application to our motivating examples. This emphasis on specific examples is to some extent a reflection of the author's opinion that generic methods for analysing spatio-temporal data-sets have not yet become well-established; certainly, they are less well established than is the case for purely spatial data. Nevertheless, in the final section of the chapter we will attempt to draw some general conclusions which go beyond the specific examples considered, and can in that sense be regarded as pointers towards an emerging general methodology.

2 Statistical methods for spatial point processes

2.1 Descriptors of pattern: spatial regularity, complete spatial randomness and spatial aggregation

A convenient, and conventional, starting point for the analysis of a spatial point pattern is to apply one or more tests of the hypothesis of *complete spatial randomness* (CSR), under which the data are a realisation of a homogeneous Poisson process. A homogeneous Poisson process is a point process which satisfies two conditions: the number of events in any planar region A follows a Poisson distribution with mean $\lambda|A|$, where $|\cdot|$ denotes area and the constant λ is the *intensity*, or mean number of events per unit area; and the numbers of events in disjoint regions are independent. It follows that, conditional on the number of events in any region A , the locations of the events form an independent random sample from the uniform distribution on A (see, for example, Diggle, 2003, Section 4.4). Hence, CSR embraces two quite different properties: a uniform marginal distribution of events over the region A ; and independence of events. We emphasise that this is only a starting point, and that the hypothesis of CSR is rarely of any scientific interest. Rather, CSR is a dividing hypothesis (Cox, 1977), a test of which leads to a qualitative classification of an observed pattern as regular, approximately random or aggregated.

We do not attempt a precise mathematical definition of the descriptions “regular” and “aggregated.” Roughly speaking, a regular pattern is one in which events are more evenly spaced throughout A than would be expected under CSR, and typically arises through some form of inhibitory dependence between events. Conversely, an aggregated pattern is one in which events tend to occur in closely spaced groups. Patterns of this type can arise as a consequence of marginal non-uniformity or a form of attractive dependence, or both. In general, as shown by Bartlett (1964), it is not possible to distinguish empirically between underlying hypotheses of non-uniformity and dependence, using the information presented by a single observed pattern. Figure 5 shows an example of a regular, a completely random and an aggregated spatial point pattern. The contrasts amongst the three are clear.

2.2 Functional summary statistics

Tests of CSR which are constructed from functional summary statistics of an observed pattern are useful for two reasons: when CSR is conclusively rejected, their behaviour gives clues as to the kind of model which might provide a reasonable fit to the data; and they may suggest preliminary estimates of model parameters. Two widely used ways of constructing functional summaries are through nearest neighbour and second-moment properties. Third and higher-order moment summaries are easily defined, but appear to be rarely (possibly too rarely) used in data-analysis; an exception is Peebles and Groth (1975). They do feature more prominently in the theoretical

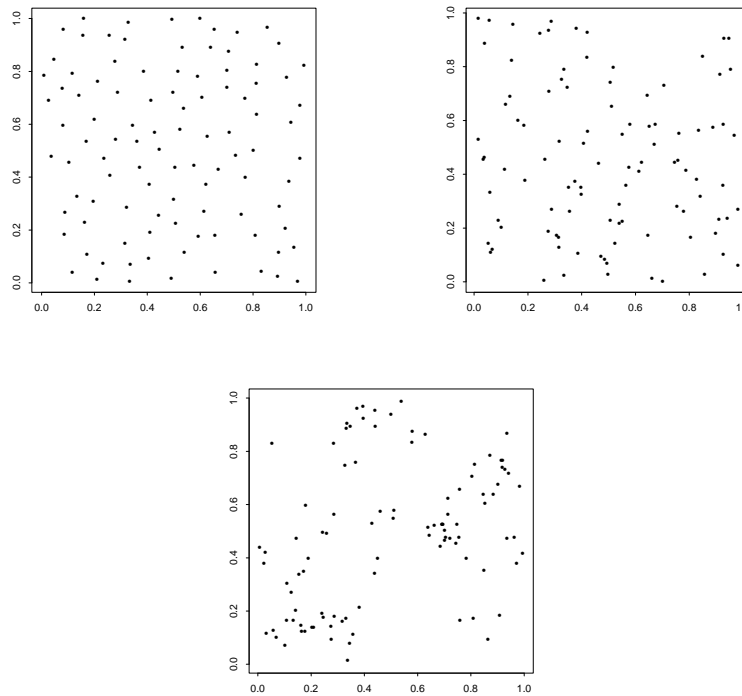


Figure 5: Examples of a regular (upper-left panel), a completely random (upper-right panel) and an aggregated (lower panel) spatial point pattern.

analysis of ecological models, as discussed elsewhere in this volume, and undoubtedly offer potential insights which are not captured by second-moment properties.

Two nearest neighbour summaries are the distribution functions of X , the distance from an arbitrary origin of measurement to the nearest event of the process, and of Y , the distance from an arbitrary event of the process to the nearest other event. We denote these by $F(x)$ and $G(y)$, respectively. The empirical counterpart of $F(x)$ typically uses the distances, d_i say, from each of m points in a regular lattice arrangement to the nearest event, leading to the estimate $\tilde{F}(x) = m^{-1} \sum I(d_i \leq x)$ where $I(\cdot)$ is the indicator function. Similarly, if e_i is the distance from each of n events to its nearest neighbour, then $\tilde{G}(y) = n^{-1} \sum I(e_i \leq y)$. Edge-corrected versions of these simple estimators are sometimes preferred, and are necessary if we wish to compare empirical estimates with the corresponding theoretical properties of a stationary point process.

Derivations, and further discussion, of results in the remainder of this Section can be found, for example, in Diggle (2003, Chapter 4).

Under CSR, $F(x) = G(x) = 1 - \exp(-\lambda\pi x^2)$, where λ is the *intensity*, or mean number of events per unit area. Typically, in a regular pattern $G(x) < F(x)$, whereas in an aggregated pattern $G(x) > F(x)$.

To describe the second-moment properties of a spatial point process, we need some

additional notation. Let dx denote an infinitesimal neighbourhood of the point x , and $N(dx)$ the number of events in dx . Then, the *intensity function* of the process is

$$\lambda(x) = \lim_{|dx| \rightarrow 0} \left\{ \frac{E[N(dx)]}{|dx|} \right\}.$$

Similarly, the *second-moment intensity function* is

$$\lambda_2(x, y) = \lim_{\substack{|dx| \rightarrow 0 \\ |dy| \rightarrow 0}} \left\{ \frac{E[N(dx)N(dy)]}{|dx||dy|} \right\}$$

and the *covariance density* is

$$\gamma(x, y) = \lambda_2(x, y) - \lambda(x)\lambda(y).$$

The process is stationary and isotropic if its statistical properties do not change under translation and rotation, respectively. If we now assume that the process is stationary and isotropic, the intensity function reduces to a constant, λ , equal to the expected number of events per unit area. Also, the second-moment intensity reduces to a function of distance, $\lambda_2(x, y) = \lambda_2(r)$ where $r = \|x - y\|$ is the distance between x and y , and the covariance density is $\gamma(r) = \lambda_2(r) - \lambda^2$. In this case, the scaled quantity $\rho(r) = \lambda_2(r)/\lambda^2$ is called, somewhat misleadingly, the *pair correlation function*. For a homogeneous Poisson process, $g(r) = 1$ for all r .

A more tangible interpretation of the pair correlation function is obtained if we integrate over a disc of radius s . This gives the reduced second-moment measure, or *K-function*,

$$K(s) = 2\pi \int_0^s \rho(r)rdr. \tag{1}$$

Ripley (1976, 1977) introduced the *K-function* as a tool for data-analysis. One of its advantages over the pair correlation function is that it can be interpreted as a scaled expectation of an observable quantity. Specifically, let $E(s)$ denote the expected number of further events within distance s of an arbitrary event. Then,

$$K(s) = \lambda^{-1}E(s). \tag{2}$$

The result (2) leads to several useful insights. Firstly, it suggests a method of estimating $K(s)$ directly by the method of moments, without the need for any smoothing; this is especially useful for relatively small data-sets. Secondly, it explains why $K(s)$ is a good descriptor of spatial pattern. For a completely random pattern, events are positioned independently, hence $E(s) = \lambda\pi s^2$ and $K(s) = \pi s^2$. This gives a benchmark against which to assess departures from CSR. For aggregated patterns, $K(s)$ is relatively large at small distances s because each event typically forms part of a “cluster” of mutually close events. Conversely, for regular patterns $K(s)$ is relatively small at small distances s because each event tends to be surrounded by empty space. Another useful property is that $K(s)$ is invariant to random thinning, i.e. retention or deletion of events according to a series of independent Bernoulli trials. This follows

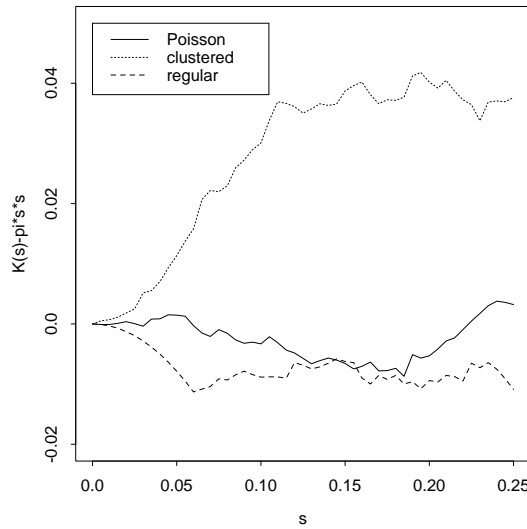


Figure 6: Estimates $\hat{K}(s) - \pi s^2$ for a regular (dashed line), a completely random (Poisson process, solid line) and an aggregated or clustered (dotted line) point pattern.

immediately from (2), which expresses $K(s)$ as the ratio of two quantities, both of which vary by the same constant of proportionality under random thinning.

We use the following edge-corrected method of moments estimator proposed originally by Ripley (1976, 1977). For data $x_i \in A : i = 1, \dots, n$, a natural estimator for $E(s)$ is

$$\tilde{E}(s) = n^{-1} \sum_{i=1}^n \sum_{j \neq i} I(r_{ij} \leq s), \quad (3)$$

where $r_{ij} = \|x_i - x_j\|$. Except for very small values of s , this estimator suffers from substantial negative bias because events outside A are not recorded in the data. A remedy is to replace the simple count in (3) by a sum of weights w_{ij} , where w_{ij}^{-1} is the proportion of the circumference of the circle with centre x_i and radius r_{ij} which lies within A . Finally, we estimate λ by $(n-1)/|A|$ where $|A|$ denotes the area of A , to give

$$\hat{K}(s) = |A| \{n(n-1)\}^{-1} \sum_{i=1}^n \sum_{j \neq i} w_{ij} I(r_{ij} \leq s). \quad (4)$$

Ripley used $n/|A|$ to estimate λ . Our preference for $(n-1)/|A|$ has a slightly arcane theoretical justification which is discussed in Chetwynd and Diggle (1998), but is clearly of no great consequence when n is large.

Figure 6 shows estimates $\hat{K}(s) - \pi s^2$ for each of the three point patterns shown in Figure 5. Subtraction of the CSR benchmark, $K(s) = \pi s^2$, emphasises departures from CSR, in effect acting as a magnifying glass applied to the estimate $\hat{K}(s)$.

Multivariate extensions of the K -function and its estimator were proposed by Lotwick and Silverman (1982). For a stationary, isotropic process let $\lambda_j : j = 1, \dots, m$ denote

the intensity of type j events. Define functions $K_{ij}(s) = \lambda_j^{-1} E_{ij}(s)$, where $E_{ij}(s)$ is the expected number of further type j events within distance s of an arbitrary type i event. Note that $K_{ij}(s) = K_{ji}(s)$. Although this equality is not obvious from the above definitions, it follows immediately from the multivariate analogue of our earlier definition (1) of $K(s)$ as an integrated version of the pair correlation function. However, direct extension of (4) to the multivariate case leads to two different estimates $\tilde{K}_{ij}(s)$ and $\tilde{K}_{ji}(s)$ which, following Lotwick and Silverman (1982), we can combine to give the single estimate

$$\hat{K}_{ij}(s) = \{n_i \tilde{K}_{ij}(s) + n_j \tilde{K}_{ji}(s)\} / (n_i + n_j). \quad (5)$$

Two useful benchmark results for multivariate K -functions are:

- (i) if type i and type j events form independent processes, then $K_{ij}(s) = \pi s^2$;
- (ii) if type i and type j events form a random labelling of a univariate process with K -function $K(s)$, then $K_{ii}(s) = K_{jj}(s) = K_{ij}(s) = K(s)$.

2.3 Functional summary statistics for the amacrine data

Figure 7 shows estimates $\hat{K}_{ij}(s) - \pi s^2$ for the amacrine cell data. Our interpretation of the three estimates is as follows. Firstly, the near-equality of $\hat{K}_{11}(s)$ and $\hat{K}_{22}(s)$ suggests that the underlying biological process may be the same for both types of cell. Informally, the difference between $\hat{K}_{11}(s)$ and $\hat{K}_{22}(s)$ gives an upper bound to the size of the sampling fluctuations in the estimates. Secondly, both estimates show a strong inhibitory effect, with no two cells of the same type occurring within a distance of around $30\mu\text{m}$. Thirdly, the estimate $\hat{K}_{12}(s)$ fluctuates around a value close to zero at small distances s , suggesting that the two component patterns are approximately independent. More specifically, $\hat{K}_{12}(s)$ does not show the strong inhibitory effect exhibited by both $\hat{K}_{11}(s)$ and $\hat{K}_{22}(s)$.

Collectively, these results are consistent with the first of the two developmental hypotheses for these data, namely that the component patterns of on and off cells form initially in two separate layers which later fuse to form the mature retina. Specifically, the separate layer hypothesis would imply statistical independence between the two component patterns, hence $K_{12}(s) = \pi s^2$. In fact, as we discuss below, the component patterns cannot strictly be independent because of the physical space required by each cell body. The data are clearly not compatible with random labelling of an initially undifferentiated pattern, as this would require all three estimated K -functions to be equal to within sampling variation. Furthermore, it is difficult to imagine how any biologically plausible labelling process could preserve strict inhibition between any two cells of the same type without imposing a similar constraint on two cells of opposite type. Hence, the analysis summarised in Figure 7 strongly favours the separate layer hypothesis.

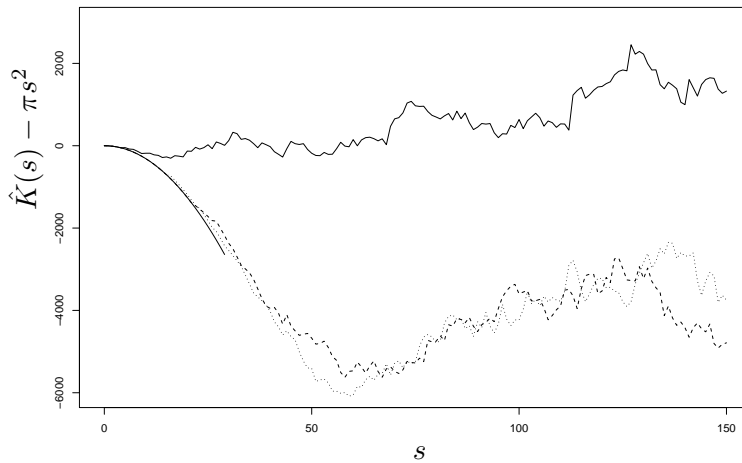


Figure 7: Estimates of the K -functions for the amacrine cell data. Each plotted function is $\hat{K}(s) - \pi s^2$. The dashed line corresponds to $\hat{K}_{11}(s)$ (on cells), the dotted line to $\hat{K}_{22}(s)$ (off cells) and the solid line to $\hat{K}_{12}(s)$. The parabola $-\pi s^2$ is also shown as a solid line.

2.4 Likelihood-based methods

Classical maximum likelihood estimation is straightforward for Poisson processes, but notoriously intractable for other point process models. Two more tractable alternatives are maximum pseudo-likelihood and Monte Carlo maximum likelihood. Both are particularly well-suited to estimation in a class of models known as pairwise interaction point processes, and it is this context that we discuss them here.

A third variant of likelihood-based estimation uses a *partial likelihood*. This method is best known in the context of survival analysis (Cox, 1972, 1975). We describe its adaptation to spatio-temporal point processes in Section 3.2.2.

2.4.1 Pairwise interaction point processes

Pairwise interaction processes form a sub-class of Markov point processes (Ripley and Kelly, 1977). They are defined by their likelihood ratio, $f(\cdot)$, with respect to a Poisson process of unit intensity. Hence, if $\chi = \{x_1, \dots, x_n\}$ denotes a configuration of n points in a spatial region A , then $f(\chi)$ measures in an intuitive sense how much more likely is the configuration χ than it would be as a realisation of a Poisson process of unit intensity. For a pairwise interaction process, we need to specify a parameter β which governs the mean number of events per unit area and an *interaction function* $h(r)$, where r denotes distance. Intuitively, $h(r)$ is related to the likelihood that the model will generate pairs of events separated by a distance r , in the sense that the likelihood for a particular configuration of events depends on the product of $h(\|x_i - x_j\|)$ over all distinct pairs of events. Hence, for example, a value $h(r) = 0$ for all $r < \delta$ would

imply that no two events can be separated by a distance less than δ . The likelihood ratio $f(\chi)$ for the resulting pairwise interaction point process is

$$f(\chi) = c(\beta, h)\beta^n \prod_{j < i} h(\|x_i - x_j\|), \quad (6)$$

where $c(\beta, h)$ is a normalising constant which is generally intractable. Note that in a homogeneous Poisson process, the number of points in A follows a Poisson distribution with mean proportional to $|A|$ and, conditional on the number of points in A , their locations form an independent random sample from the uniform distribution on A . It follows that a homogeneous Poisson process is a special case of a pairwise interaction process in which $h(u) = 1$ for all u , and β is the intensity. More generally, in (6) the parameter β is related to, but not necessarily equal to, the intensity. Provided that the specified form of $h(\cdot)$ is legitimate, values of $h(r)$ less than or greater than 1 correspond to processes which generate regular or aggregated patterns, respectively.

A sufficient condition for legitimacy is that $h(r) \leq 1$ for all r , as this guarantees a finite intensity for the resulting point process. It also leads to point patterns whose character is inhibitory, meaning that close pairs of events are relatively unlikely by comparison with a Poisson process of the same intensity. Pairwise interaction point process of this kind are widely used for modelling regular spatial point patterns.

Specifications in which $h(u) > 1$ are more problematic. An intuitive explanation for this is that if, over a range of distances r , the interaction function takes values $h(r) \geq h_0$, where $h_0 > 1$, then the product term on the right hand side of (6) can be as large as $h_0^{n(n-1)/2}$, and this cannot be balanced by adjusting the value of β in (6). Hence, the likelihood increases without limit as $n \rightarrow \infty$. This perhaps explains why, even if we are prepared to consider n as fixed, pairwise interaction processes with $h(r) > 1$ tend to generate unrealistically strong spatial aggregation, with large clusters of near-coincident events. For a rigorous discussion of the properties of pairwise interaction processes with $h(r) > 1$, see Gates and Westcott (1986).

2.4.2 Maximum pseudo-likelihood

The method of maximum pseudo-likelihood was originally proposed by Besag (1975, 1978) as a method for real-valued, spatially discrete processes. Besag et al (1982) derived a point process version by considering a limit of binary-valued processes on a lattice, as the lattice spacing tends to zero. For a finite-dimensional probability distribution, the pseudo-likelihood is the product of the full conditional distributions, i.e. the conditional distributions of each Y_i given the values of all other Y_j . Hence, if $Y = (Y_1, \dots, Y_n)$ has joint probability density $f(y)$, then the pseudo-likelihood is, in an obvious notation, $\prod_{i=1}^n f(y_i|y_j, j \neq i)$.

For a point process, the pseudo-likelihood uses conditional intensities in place of the full conditional distributions. In particular, for a Markov point process with likelihood ratio $f(\cdot)$, the conditional intensity for an arbitrary point u given the

observed configuration X on $A - \{u\}$ is

$$\lambda(u; X) = \begin{cases} f(X \cup \{u\})/f(X) & : u \notin X \\ f(X)/f(X - \{u\}) & : u \in X \end{cases}$$

and the log-pseudo-likelihood is

$$\sum_{i=1}^n \log \lambda(x_i; X) - \int_A \lambda(u; X) du.$$

2.4.3 Monte Carlo maximum likelihood

Monte Carlo maximum likelihood estimation, as described here, was proposed by Geyer and Thompson (1992). Geyer (1999) and Moller and Waagepetersen (2004) discuss the method in the context of point process models including, but not restricted to, pairwise interaction point processes.

Conditional on the number of events in a specified region A , the likelihood for a pairwise interaction point process can be written in principle as

$$\ell(\theta) = c(\theta)f(X; \theta), \tag{7}$$

where $X = \{x_1, \dots, x_n\}$ is the observed configuration of the n events in A . For most models of interest, the normalising constant $c(\theta)$ in (7) is intractable. However, note that

$$\begin{aligned} c(\theta)^{-1} &= \int_X f(X; \theta) dX \\ &= \int_X f(X; \theta) \times \frac{c(\theta_0)}{c(\theta_0)} \times \frac{f(X; \theta_0)}{f(X; \theta_0)}, \end{aligned}$$

for any value of θ_0 . If we now define $r(X; \theta, \theta_0) = f(X; \theta)/f(X; \theta_0)$ then we can write

$$\begin{aligned} c(\theta)^{-1} &= c(\theta_0)^{-1} \int_X r(X; \theta, \theta_0) c(\theta_0) f(X; \theta) dX \\ &= c(\theta_0)^{-1} E_{\theta_0}[r(X; \theta, \theta_0)] dX, \end{aligned}$$

where $E_{\theta_0}[\cdot]$ denotes expectation with respect to the distribution of X when $\theta = \theta_0$. This in turn allows us to re-express the likelihood (7) as

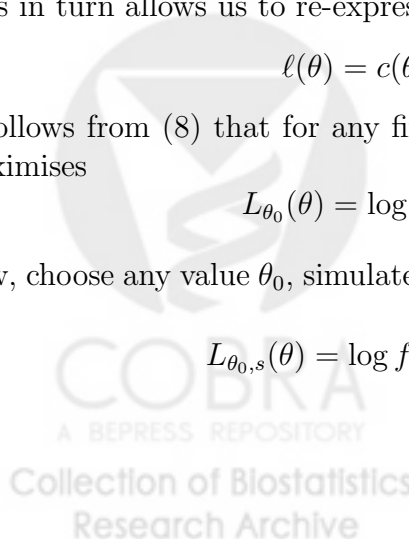
$$\ell(\theta) = c(\theta_0) f(X; \theta) / E_{\theta_0}[r(X; \theta, \theta_0)]. \tag{8}$$

It follows from (8) that for any fixed value θ_0 , the maximum likelihood estimator $\hat{\theta}$ maximises

$$L_{\theta_0}(\theta) = \log f(X; \theta) - \log E_{\theta_0}[r(X; \theta, \theta_0)]. \tag{9}$$

Now, choose any value θ_0 , simulate realisations $X_j : j = 1, \dots, s$ with $\theta = \theta_0$ and define

$$L_{\theta_0, s}(\theta) = \log f(X; \theta) - \log s^{-1} \sum_{j=1}^s [r(X_j; \theta, \theta_0)]. \tag{10}$$



Then, the value $\hat{\theta}_{MC}$ which maximises (10) is a *Monte Carlo maximum likelihood estimator* (MCMLE) for θ . Note the indefinite article. A Monte Carlo log-likelihood $L_{\theta_0, s}(\theta)$ is typically a smooth function of θ and can easily be maximised numerically, but it is also a function of θ_0 , s and the simulated realisations X_j . In practice, we would want to choose s sufficiently large that the Monte Carlo variation introduced by using a sample average in place of the expectation on the right-hand side of (9) is negligible. However, for given s the behaviour of the MCMLE is critically dependent on the choice of θ_0 , the ideal being to choose θ_0 equal to $\hat{\theta}$ in which case the Monte Carlo variation in (10) is zero at $\theta = \hat{\theta}$. More generally, obtaining a sufficiently accurate Monte Carlo approximation to the intractable expectation in (9) raises a number of practical issues which, in the author's experience, make it difficult to automate the procedure.

2.5 Bivariate pairwise interaction point processes

The family of pairwise interaction point processes can readily be extended to the multivariate case by specifying a set of interaction functions, one for each possible pair of types of event. In the bivariate case, and again treating the numbers of events of each type as fixed, we denote the data by a pair of configurations, $X_1 = \{x_{1i} : i = 1, \dots, n_1\}$ and $X_2 = \{x_{2i} : i = 1, \dots, n_2\}$.

In a bivariate pairwise interaction process, the joint density for X_1 and X_2 is $f(X_1, X_2) \propto P_{11}P_{22}P_{12}$, where

$$P_{11} = \prod_{i=2}^{n_1} \prod_{j=1}^{i-1} h_{11}(\|x_{1i} - x_{1j}\|),$$

$$P_{22} = \prod_{i=2}^{n_2} \prod_{j=1}^{i-1} h_{22}(\|x_{2i} - x_{2j}\|)$$

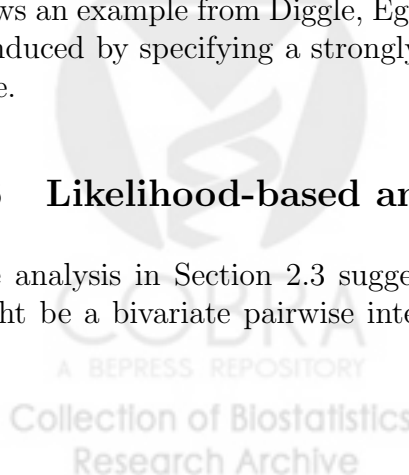
and

$$P_{12} = \prod_{i=1}^{n_1} \prod_{j=1}^{n_2} h_{12}(\|x_{1i} - x_{2j}\|).$$

As in the univariate case, a sufficient condition for the legitimacy of the model is that $0 \leq h_{ij}(r) \leq 1$ for all r . However, and in contrast to the univariate case, a model of this kind can easily generate spatially aggregated component patterns. Figure 8 shows an example from Diggle, Eglen and Troy (2005), in which marginal aggregation is induced by specifying a strongly inhibitory interaction between events of opposite type.

2.6 Likelihood-based analysis of the amacrine cell data

The analysis in Section 2.3 suggested that a suitable model for the amacrine data might be a bivariate pairwise interaction process with strongly inhibitory marginal



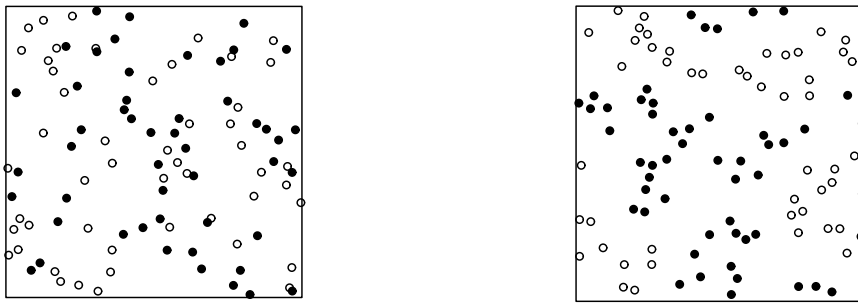


Figure 8: Simulated realisations of bivariate pairwise interaction point processes each with 50 events of either type on the unit square and simple inhibitory interaction functions. In both panels, the minimum permissible distance between any two events of the same type is 0.025. In the left-hand panel, the two component patterns are independent. In the right-hand panel, the minimum permissible distance between any two events of opposite types is 0.1.

properties and approximate independence between the two components.

Our first stage in fitting a model of this kind is to use maximum pseudo-likelihood estimation in conjunction with a piece-wise constant specification of $h(u)$ to identify a candidate model for the interaction within each component pattern. We then use Monte Carlo maximum likelihood to fit a suitable parametric model to each component. Figure 9 shows the result, together with a Monte Carlo maximum likelihood estimate using the parametric model

$$h(u; \theta) = \begin{cases} 0 & : u \leq \delta \\ 1 - \exp[-\{(u - \delta)/\phi\}^\alpha] & : u > \delta \end{cases} . \quad (11)$$

The fit adopts common parameter values for the two types of cell, on the basis of a Monte Carlo likelihood ratio test under the assumption that the two component processes are independent; for details, see Diggle, Eglen and Troy (2005). In fitting the parametric model, we used a fixed value $\delta = 10\mu\text{m}$, representing the approximate physical size of each cell body, and estimated the remaining parameters as $\hat{\phi} = 49.08$ and $\hat{\alpha} = 2.92$.

For the bivariate analysis, we use the same parametric form (11) for the three interaction functions $h_{11}(r)$, $h_{22}(r)$ and $h_{12}(r)$. For ϕ tending to zero, the model for $h_{12}(r)$ reduces to a simple inhibitory form, $h_{12}(r) = 0$ for $r < \delta_{12}$ and $h_{12}(r) = 1$ otherwise, with independence of the two components as the special case $\delta_{12} = 0$. Independence is strictly impossible because no two cells can occupy the same location, but it is reasonable to treat δ_{12} as a parameter to be estimated because the two types of cell are located at slightly different depths within the retina. Diggle, Eglen and Troy (2005) conclude that a bivariate model with a simple inhibitory $h_{12}(r)$ and $\hat{\delta}_{12} = 4.9$ gives a reasonable fit to the data.

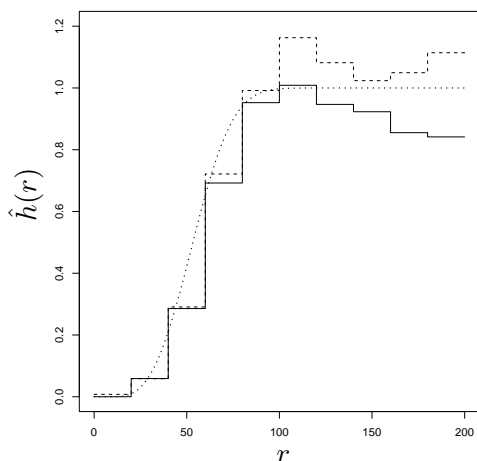


Figure 9: Non-parametric maximum pseudo-likelihood estimates of the pairwise interaction functions for on cells (solid line) and for off cells (dashed line), together with parametric fit assuming common parameter values for both types of cell (dotted line).

3 Strategies for the analysis of spatio-temporal point patterns

Many of the tools used to analyse spatial point process data can be extended to the spatio-temporal setting. Functional summaries based on low-order moments can be extended in the obvious way by considering configurations of events at specified spatial and temporal positions. For example, Diggle, Chetwynd, Haggkvist and Morris (1995) defined a spatio-temporal K -function $K(s, t)$ such that $\lambda K(s, t)$ is the expected number of further events within distance s and time t of an arbitrary event of the process. Bhopal, Diggle and Rowlingson (1992) used an estimate of $K(s, t)$ to analyse the spatio-temporal distribution of apparently sporadic cases of legionnaire's disease. However, the spatio-temporal setting opens up other modelling and analysis strategies which take more explicit account of the directional character of time, and the consequently richer opportunities for scientific inference.

3.1 Strategies for discrete-time data

As noted earlier, discrete-time spatio-temporal point process data can arise in two ways; either the underlying process genuinely operates in discrete-time, or an underlying continuous-time process is observed at a discrete sequence of time-points. A hypothetical example of the former would be the yearly sequence of spatial point distributions formed by the natural regeneration of an annual plant community. The Cornwall BTB data described in Section 1.1.2 are an example of the latter.

3.1.1 Transition models

For genuinely discrete-time processes, a natural strategy is to build a transition model to describe the changes between successive times. In symbolic notation, if \mathcal{P}_t denotes the spatial point process at time t a transition model for the joint distribution of $\mathcal{P} = \{\mathcal{P}_1, \mathcal{P}_2, \dots, \mathcal{P}_t\}$ takes the form

$$[\mathcal{P}] = [\mathcal{P}_1][\mathcal{P}_2|\mathcal{P}_1]\dots[\mathcal{P}_t|\mathcal{P}_{t-1}, \dots, \mathcal{P}_1] \quad (12)$$

A convenient working assumption would be that the process is Markov in time, and this may have some mechanistic justification when times correspond to successive generations, as in our hypothetical example.

3.1.2 A transition model for spatial aggregation

A standard example of a spatial point process model which leads to spatially aggregated patterns is a Neyman-Scott clustering process (Neyman and Scott, 1958), defined as follows. *Parent* events form a homogeneous Poisson process with intensity ρ events per unit area. The parents then generate numbers of *offspring* as an independent random sample from the Poisson distribution with mean μ . The positions of the offspring relative to their parents are an independent random sample from the bivariate Normal distribution with mean zero and variance matrix $\sigma^2 I$, where I denotes the identity matrix. The observed point pattern is then taken to be the superposition of the offspring from all parents.

An obvious way to turn the Neyman-Scott process into a transition model is to let the offspring of one generation become the parents for the next generation. Kingman (1977) discusses a model of this kind in questioning whether the basic Neyman-Scott formulation can arise as the equilibrium distribution of a spatio-temporal process. Note in particular that the spatio-temporal process defined in this way may die out after a finite number of generations. Figure 10 shows the result of a simulation on the unit square with periodic boundary conditions, i.e. events are generated on a torus, which is then unwrapped to form the unit square region A . The model parameters are $\rho = 100$, $\sigma = 0.025$ and $\mu = 1$. Setting $\mu = 1$ implies that the mean number of events in each generation is ρ , but with a variance which increases from one generation to the next; if N_t denotes the number of events in the t th generation, then $E[N_t] = \rho$ for all t and $\text{Var}(N_t) = t\rho$.

The simulation shows how the spatial aggregation in the resulting patterns tends to increase with successive generations, although it is not clear that this informal description can be expressed rigorously, not least because on any finite region the process is certain eventually to become extinct.

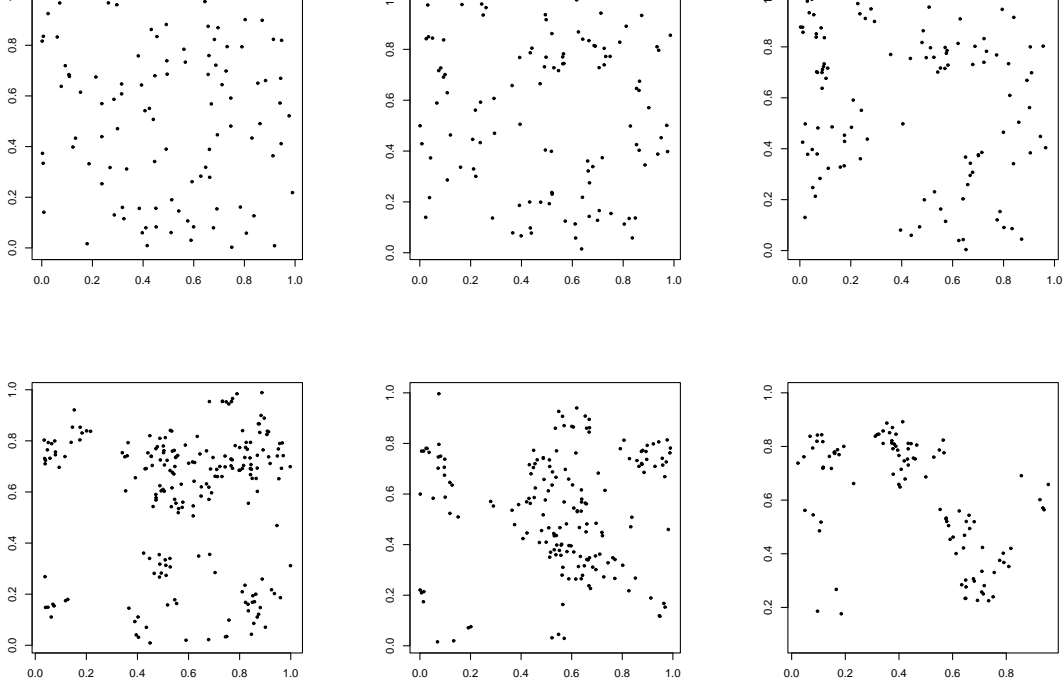


Figure 10: Simulated realisation of a transition model for spatial aggregation. The first row shows the first three generations of the process, starting from a completely random spatial distribution. The second row shows the 50th, 70th and 90th generations.

3.1.3 Marked point process models

When time is artificially discrete as a consequence of the data-recording process, a principled approach would be to formulate a continuous-time model and to deduce from the model the statistical properties of the observed, discrete-time data. A more pragmatic strategy is to analyse the data as a marked spatial point process, treating time as an ordered categorical mark attached to the spatial location of each event. From this point of view, methods for the analysis of multivariate spatial point processes, which are marked point processes with categorical marks, can be applied directly and the discussion of multivariate point processes in Section 2 is relevant. However, the way in which the results of any such analysis are interpreted should, so far as is possible, take account of the natural ordering of time.

This pragmatic strategy is unlikely to deliver a model with a direct, mechanistic interpretation, but is a useful approach for descriptive analysis. In Section 4 we illustrate the approach with an analysis of the Cornwall BTB data, as reported in Diggle, Zheng and Durr (2005).

3.2 Analysis strategies for continuous-time data

Even when continuous-time data are available, we shall preserve a distinction between empirical and mechanistic modelling. An empirical model aspires to provide a good

descriptive fit to the data, but does not necessarily admit a context-specific scientific interpretation. A mechanistic model is more ambitious, embodying features which relate directly to the underlying science. To some extent, this is a false dichotomy. On the one hand, a good empirical model will include parameters which are interpretable in ways relevant to the scientific context and, minimally, should furnish an answer to a scientifically interesting question. On the other hand, even a mechanistic model will be at best an idealised, and quite possibly a crude, approximation to the truth. From a statistical perspective, a simple but well-identified model may be more valuable than an over-complicated model incorporating more parameters than can reasonably be estimated from the available data.

3.2.1 Empirical modelling: log-Gaussian spatio-temporal Cox processes

A Cox process, introduced in one time-dimension by Cox (1955), is a Poisson process with a varying intensity which is itself a stochastic process. For our purposes, we need a model for a non-negative valued spatio-temporal stochastic process $\Lambda(x, t)$. Then, conditionally on $\Lambda(x, t)$ our point process is a Poisson process with intensity $\Lambda(x, t)$. This implies that, again conditionally on $\Lambda(x, t)$, the number of events in any spatio-temporal region, say $A \times (0, T)$, is Poisson-distributed with mean

$$\mu_{A,T} = \int_0^T \int_A \Lambda(x, t) dx dt$$

and the locations and times of the events are an independent random sample from the distribution on $A \times (0, T)$ with probability density proportional to $\Lambda(x, t)$.

By far the most tractable class of real-valued spatio-temporal stochastic processes is the Gaussian process, $S(x, t)$ say, for which the joint distribution of $S(x_i, t_i)$ for any set of points (x_i, t_i) is multivariate Normal. A log-Gaussian Cox process is Cox process whose intensity is of the form $\Lambda(x, t) = \exp\{S(x, t)\}$, where $S(x, t)$ is a Gaussian process (Moller, Syversveen and Waagepetersen, 1998).

The properties of a log-Gaussian Cox process are determined by the mean and covariance structure of $S(x, t)$. Note firstly that any spatial and/or temporal variation in the mean of $S(x, t)$ translates into a multiplicative, deterministic component to $\Lambda(x, t)$, hence we can always re-express our model as $\Lambda(x, t) = \lambda(x, t) \exp\{S(x, t)\}$ where the mean of $S(x, t)$ is a constant. In the stationary case, a convenient parameterisation is to set $E[S(x, t)] = -0.5\sigma^2$, where $\sigma^2 = \text{Var}\{S(x, t)\}$. This gives $E[\exp\{S(x, t)\}] = 1$, hence $\lambda(x, t)$ is the unconditional space-time intensity, or mean number of events per unit time per unit area in an infinitesimal neighbourhood of the point (x, t) . In the remainder of this Section, we assume that $\lambda(x, t) = 1$ and focus on the specification of the stochastic component $\exp\{S(x, t)\}$.

In general, if $S(x, t)$ has mean $-0.5\sigma^2$ and covariance function

$$\text{Cov}\{S(x, t), S(x', t')\} = \sigma^2 \rho(x, x', t, t')$$

then the covariance function of $\exp\{S(x, t)\}$ is

$$\gamma(x, x', t, t') = \exp\{\sigma^2 \rho(x, x', t, t')\} - 1,$$

and $\gamma(\cdot)$ is also the covariance density of the Cox process (see Brix and Diggle, 2001, but note also the correction in Brix and Diggle, 2003). In the stationary case, $\gamma(x, x', t, t') = \gamma(u, v)$, where $u = \|x - x'\|$ and $v = |t - t'|$.

Brix and Moller (2001) consider a sub-class of spatio-temporal log-Gaussian Cox processes in which $S(x, t) = S(x) + g(t)$ where $g(t)$ is a deterministic function. One interpretation of this sub-class is that $S(x)$ represents spatial environmental variation which does not vary over time, and $\exp\{g(t)\}$ represents a time-varying birth-rate for new events. A natural extension would be to replace $g(t)$ by a stationary stochastic process $G(t)$, which would give an additive decomposition of the correlation function of $S(x, t)$ as

$$\rho(u, v) = \{\sigma_S^2 \rho_S(u) + \sigma_G^2 \rho_G(v)\} / (\sigma_S^2 + \sigma_G^2).$$

Brix and Diggle (2001) develop an approach to spatio-temporal prediction using a separable spatio-temporal correlation function

$$\rho(u, v) = r(u) \exp(-v/\beta). \quad (13)$$

In (13), $r(u)$ is any valid spatial correlation function, whilst the exponential term reflects the underlying Markov-in-time structure of the model for $S(x, t)$, which they derive as follows.

First, consider a discretisation of continuous two-dimensional space into a fine grid, say of size M by N , and write S_t for the MN -element vector of values of $S(x, t)$ at the grid-points. Now, assume that S_t evolves over time according to the stochastic differential equation

$$dS_t = (A - BS_t)dt + dU_t \quad (14)$$

where A is an MN -element vector, B is a non-singular $MN \times MN$ matrix and U_t is a discrete-space approximation of spatial Brownian motion. In the stationary case, (14) corresponds to a Gaussian process $S(x, t)$ with spatio-temporal covariance structure

$$\text{Cov}\{S(x, t), S(x - u, t - v)\} = \sigma^2 r(u) \exp(-vB).$$

Brix and Diggle focus on the special case in which $B = \beta^{-1}I$ for some scalar $\beta > 0$, in which case $S(x, t)$ has variance σ^2 and separable spatio-temporal correlation function given by (13).

Separability of the spatial and temporal correlation properties is a reasonable working assumption for an empirical, descriptive model, but may be too inflexible for some applications. For example, it implies that for any single location, x_0 say, the conditional distribution of $S(x_0, t)$ given the *whole* of the process $S(x, t')$ for some $t' < t$ depends only on $S(x_0, t')$. Gneiting (2002) reviews the relevant literature and proposes a general class of stationary, non-separable spatio-temporal covariance functions.

Within the log-Gaussian Cox process framework, model-specification corresponds exactly to the problem of specifying a model for a spatio-temporal Gaussian process. See, for example, the chapters by Higdon, and by Gneiting, Genton and Guttorp, in this volume.

3.2.2 Mechanistic modelling: conditional intensity and a partial likelihood

Accepting that the distinction between what we have chosen to call empirical and mechanistic models is not sharp, a mechanistic model is one which seeks to explain how the evolution of the process depends on its past history, in a way which can be interpreted in terms of underlying scientific mechanisms. A natural way to specify a model of this kind is through its conditional intensity function. Let \mathcal{H}_t denote the accumulated history of the process, i.e. the complete set of locations and times of events occurring up to time t . Then, the conditional intensity function, $\lambda(x, t | \mathcal{H}_t)$, represents the conditional intensity for an event at location x and time t , given \mathcal{H}_t . This assumes amongst other things, that multiple, coincident events cannot occur; for a rigorous discussion, see for example Daley and Vere-Jones (1988, Chapter 2).

A defining property of a Poisson process is that its conditional intensity function is equal to its unconditional intensity, in other words the future of the process is stochastically independent of its past. A more interesting example of a conditional intensity function is the following, which bears some resemblance to the discrete-time transition model illustrated in Section 3.1.2. Each event of the process at time zero subsequently produces offspring according to an inhomogeneous temporal Poisson process with intensity $\alpha(u)$, realised independently for different events. As in the earlier, discrete-time example, the positions of the offspring relative to their parents are an independent random sample from the bivariate Normal distribution with mean zero and variance matrix $\sigma^2 I$. Each offspring, independently, then follows the same rules as their parent: it produces offspring according to an inhomogeneous Poisson process with intensity $\alpha(u - t)$, where t denotes its birth-time, and each offspring is spatially dispersed relative to its parent according to a bivariate Normal distribution with mean zero and variance $\sigma^2 I$. The events of the process are the resulting collection of locations x and birth-times t . If we order the events of the process so that $t_i < t_{i+1}$ for all i , then the history at time t is $\mathcal{H}_t = \{(x_i, t_i) : i = 1, \dots, N_t\}$, where N_t is the number of events to have occurred by time t . Writing $f(x) = (2\pi\sigma^2)^{-1} \exp\{-x'x/(2\sigma^2)\}$, the conditional intensity function is

$$\lambda(x, t | \mathcal{H}_t) = \sum_{i=1}^{N_t} \alpha(t - t_i) f(x - x_i).$$

The number of offspring produced by any event of this process is Poisson-distributed, with mean

$$\mu = \int_0^\infty \alpha(u) du,$$

which we therefore assume to be finite. The number of events as a function of time, N_t , forms a simple branching process, and eventual extinction is certain if $\mu \leq 1$. Otherwise, the probability of extinction depends both on μ and on the initial conditions at time $t = 0$. The spatial character of the process varies considerably, according to both the detailed model specification and the initial conditions. However, the cumulative spatial distribution of events occurring up to time t tends to become progressively more strongly aggregated as t increases, because of the combined effects of the successive clustering of groups of offspring around their respective parents together with the extinction of some lines of descent.

For data $(x_i, t_i) \in A \times (0, T) : i = 1, \dots, n$, with $t_1 < t_2 < \dots < t_n$, the log-likelihood associated with any point process specified through its conditional intensity function can be written as

$$L(\theta) = \sum_{i=1}^n \log \lambda(x_i, t_i | \mathcal{H}_{t_i}) - \int_0^T \int_A \lambda(x, t | \mathcal{H}_t) dx dt. \quad (15)$$

See, for example, Daley and Vere-Jones (1988, Chapter 13). Two major obstacles to the use of (15) in practice are that the form of the conditional intensity may itself be intractable, and that even when the conditional intensity is available, as in our example above, direct evaluation of the integral term in (15) is seldom straightforward. Monte Carlo methods are becoming more widely available for problems of this kind (Geyer, 1999; Moller and Waagepetersen, 2004). However, in practice these methods often need careful tuning to each application and the associated cost of developing and running reliable code can be an obstacle to their routine use.

As an alternative, computationally simpler approach to inference for models which are defined through their conditional intensity, Diggle (2005) proposes a partial likelihood, which is obtained by conditioning on the locations x_i and times t_i and considering the resulting log-likelihood for the observed time-ordering of the events $1, \dots, n$. Now let

$$p_i = \lambda(x_i, t_i | \mathcal{H}_{t_i}) / \sum_{j=i+1}^n \lambda(x_j, t_j | \mathcal{H}_{t_i}). \quad (16)$$

Then, the partial log-likelihood is

$$L_p(\theta) = \sum_{i=1}^n \log p_i. \quad (17)$$

This method is a direct adaptation to the space-time setting of the seminal proposal in Cox (1972) for proportional hazards modelling of survival data; see also Moller and Sorensen (1994). As discussed in Cox (1975), estimates obtained by maximising the partial likelihood inherit the general asymptotic properties of maximum likelihood estimators, although their use may entail some loss of efficiency by comparison with full maximum likelihood estimation. Also, some parameters of the original model may be unidentifiable from the partial likelihood. The loss of identifiability can be advantageous if the non-identified parameters are nuisance parameters. This often

applies, for example, in the proportional hazards model for survival data (Cox, 1972), where it is helpful that the baseline hazard function can be left unspecified. Otherwise, and again as exemplified by the proportional hazards model for survival data, other methods of estimation are needed to recover the unidentified parameters; see, for example, Andersen, Borgan, Gill and Keiding (1992).

4 Bovine tuberculosis: non-parametric smoothing methods for estimating spatial segregation

Recall that the data for this application are the locations, genotypes and year of detection of all known cases of bovine tuberculosis (BTB) amongst farm animals in the county of Cornwall, UK. Figure 2 showed maps of the locations for each of the four most common genotypes, collapsed over time. These maps give a clear impression of spatial segregation, by which we mean that the county can be partitioned approximately into sub-regions where one or other of the four genotypes predominates. More formally, suppose that the pattern of cases in Figure 2 is generated by a multivariate Poisson process with intensities $\lambda_k(x)$ corresponding to cases of genotype $k = 1, \dots, 4$. Then, the probability that a case at location x will be of type j , conditional on there being a case of one of the four types at x , is

$$p_j(x) = \lambda_j(x) / \sum_{k=1}^4 \lambda_k(x).$$

We say that the pattern is *unsegregated* if $p_j(x) = p_j$ for all j and all x . At the opposite extreme, the pattern is *completely segregated* if, at each x , $p_j(x) = 1$ for one of $j = 1, \dots, 4$. Estimating and mapping the type-specific probabilities $p_j(x)$ allows an assessment of intermediate degrees of spatial segregation.

A specific question posed by these data is whether the pattern of spatial segregation is stable over time, as this would point towards constant re-infection by the locally predominant genotype, rather than to introduced infections transmitted through cattle bought and sold at market. We present here an analysis taken from Diggle, Zheng and Durr (2005) which is intended to answer this question. Our aim is to estimate possibly time-varying type-specific probability surfaces $p_{jt}(x)$, the conditional probabilities that a case at location x in time-period t will be of type j .

In this spatio-temporal setting, we assume that cases form a discrete-time sequence of multivariate Poisson processes with intensity functions $\lambda_{jt}(x)$. Although this is not strictly consistent with the infectious nature of the disease, it provides a reasonable working model within which we can answer the specific question of interest. The Poisson assumption lends itself more naturally to the general approach described in Sections 2 and 3.1.3, where we consider the time-dimension as a qualitative mark attached to the spatial location of each event, rather than to a transition modelling approach as described in Sections 3.1.1 and 3.1.2. The latter would be more natural if

the goal of the analysis was to model the transmission of the disease between farms, but this is ruled out because of the very coarse time-resolution of the annual inspection regime. We therefore approach the analysis of the data as a problem in non-parametric intensity estimation for a multivariate Poisson process.

The spatial distribution of farms at risk is not uniform over the county although it is, to a good approximation, constant over the twelve-year time-period covered by the data. Also, as noted earlier, variation in the temporal intensity of the disease is confounded with variation in the extent of the annual testing programme. We therefore factorise $\lambda_{jt}(x)$ as

$$\lambda_{jt}(x) = \lambda_0(x)\mu_0(x, t)\rho_{jt}(x), \quad (18)$$

where $\lambda_0(x)$ represents the spatial intensity of farms, $\mu_0(x, t)$ the spatio-temporal intensity of recorded cases of unspecified type and the functions $\rho_{jt}(x)$ represent the spatially, temporally and genotypically varying risks which are the quantities of interest. The type-specific probabilities associated with (18) are

$$p_{jt}(x) = \lambda_{jt}(x) / \sum_{k=1}^4 \lambda_{kt}(x) = \rho_{jt}(x) / \sum_{k=1}^4 \rho_{kt}(x), \quad (19)$$

which implies that spatio-temporal variations in the pattern of segregation can be estimated without our knowing either the spatial distribution of farms (although this is available if required) or, more importantly (because it is not available), the spatio-temporal variation in the extent of the inspection programme. Note that to estimate non-specific spatial variation in risk, we would need to make the additional assumption that $\mu_0(x, t) = \mu_0(t)$.

We estimate the type-specific probability surfaces using a simple kernel regression method. Let n_t denote the number of recorded cases in time-period t . Write the data for time-period t in the form of a set of non-specific case-locations $x_{it} : i = 1, \dots, n_t$ and associated labels Y_{it} giving the genotype of each case. Then, our kernel regression estimator of $p_{jt}(x)$ is

$$\hat{p}_{jt}(x) = \sum_{i=1}^{n_t} w_{ij}(x) I(Y_{it} = j), \quad (20)$$

where

$$w_{ij}(x) = w_j(x - x_{it}) / \sum_{k=1}^{n_t} w_j(x - x_{kt}),$$

$$w_j(x) = w(x/h_j)/h_j^2 \text{ and} \quad w(x) = \exp(-\|x\|^2/2). \quad (21)$$

The Gaussian kernel function (21) could be replaced by any other non-negative-valued weighting function. The choice of kernel function is usually less important than the choice of the band-width constants, h_j , which determine the amount of smoothing applied to the data in estimating the $p_{jt}(x)$. Using a common set of band-widths across time-periods makes for ease of interpretation. We shall also use

a common band-width, h , for all 4 components of the p -surface, as this ensures that $\sum_{j=1}^m \hat{p}_j(x) = 1$ for every location x .

To choose the value of h , we proceed as follows. Conditioning on the case-locations, the log-likelihood for the labels Y_{it} under the Poisson process model has the multinomial form,

$$L(h) = \sum_{t=1}^m \sum_{i=1}^{n_t} \sum_{j=1}^m I(Y_{it} = j) \log p_{jt}(x_{it}; h).$$

Maximising $L(h)$ would give the degenerate solution $\hat{h} = 0$. To avoid this, we use a cross-validated form of the log-likelihood,

$$L_c(h_1, \dots, h_4) = \sum_{t=1}^r \sum_{i=1}^n \sum_{j=1}^m I(Y_{it} = j) \log \hat{p}_{jt}^{(it)}(x_{it}),$$

where $\hat{p}_{jt}^{(it)}(x_{it})$ denotes the kernel estimator (20), but based on all of the data except (x_{it}, Y_{it}) .

The data from the years prior to 1997 are too sparse to allow non-parametric estimation of the $p_{jt}(x)$. We therefore applied the kernel regression method to data from the years 1997 to 2002. To maintain sufficient numbers of cases per genotype per time-period, we also combined data from successive years to give $r = 3$ discrete two-year time-periods.

The null hypothesis of interest is that $p_{jt}(x) = p_j(x)$ for $j = 1, \dots, 4$ and $t = 1, \dots, r$. An *ad hoc* statistic for a Monte Carlo test of this hypothesis is

$$T^P = \sum_{j=1}^4 \sum_{x \in X} \sum_{t=1}^r (\hat{p}_j(x, t) - \bar{p}_j(x))^2, \quad (22)$$

where $X = \{x_i^t : i = 1, 2, \dots, n_t; t = 1, 2, \dots, r\}$, $\hat{p}_j(x, t)$ is the estimated type-specific probability surface for type j in time-period t and $\bar{p}_j(x) = r^{-1} \sum_{t=1}^r \hat{p}_j(x, t)$.

The band-width which maximises the cross-validated log-likelihood is $h = 9647$ metres. A Monte Carlo test based on the statistic (22) using 999 simulations gave an attained significance level of 0.015, hence the null hypothesis is rejected at the conventional 5% level. However, the changes in the segregation pattern over time appear to be somewhat subtle. Figures 11, 12 and 13 show the estimated type-specific probability surfaces over the three consecutive time-periods. The general pattern is of an increase in the extent of spatial segregation over time. Thus, genotype 15 becomes progressively more dominant in north central Cornwall, genotype 20 has established near-dominance in the far west by 2001/02, and splogotype 9 is dominant in the east of Cornwall, but within a territory which becomes more confined to an area close to the eastern boundary as time proceeds. Finally, genotype 15 shows an apparently stable spatial distribution over the three time-periods.

In summary, our findings from our analysis of the BTB data are that there is very strong spatial segregation amongst the four most common genotypes, and that the

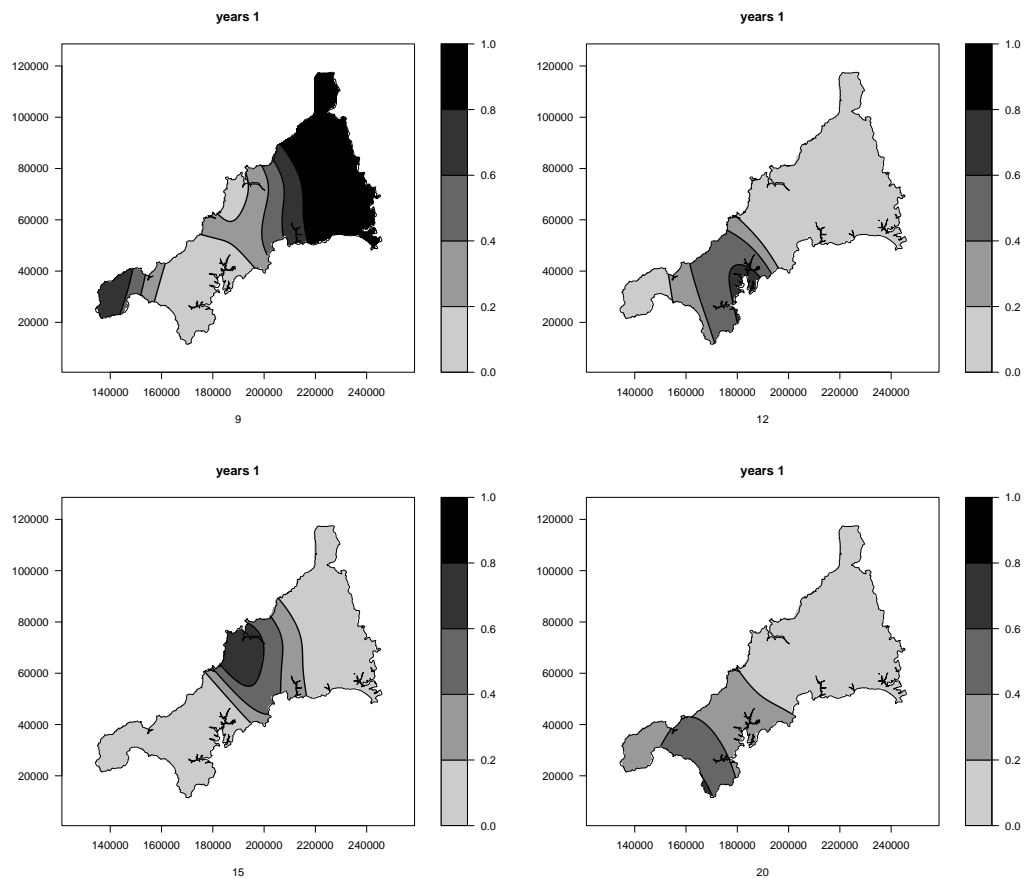


Figure 11: Kernel regression estimates of type-specific probability surfaces for cases in years 1997-98.

pattern of spatial segregation is broadly consistent over time, albeit with subtle but statistically significant differences between successive two-year periods.

Further methodological developments in progress include replacing the non-parametric kernel regression methodology by a hierarchical stochastic model, in which the spatio-temporal intensities are determined by a multivariate, discrete-time log-Gaussian process, hence $\lambda_j(x, t) = \exp\{S_j(x, t)\}$ where $\{S_1(x, t), \dots, S_4(x, t)\}$ is a quadri-variate Gaussian process. The hierarchical stochastic modelling framework is better able to deal with the relatively small numbers of cases of each genotype in each time-period by exploiting the assumed statistical dependencies between genotypes and over time.

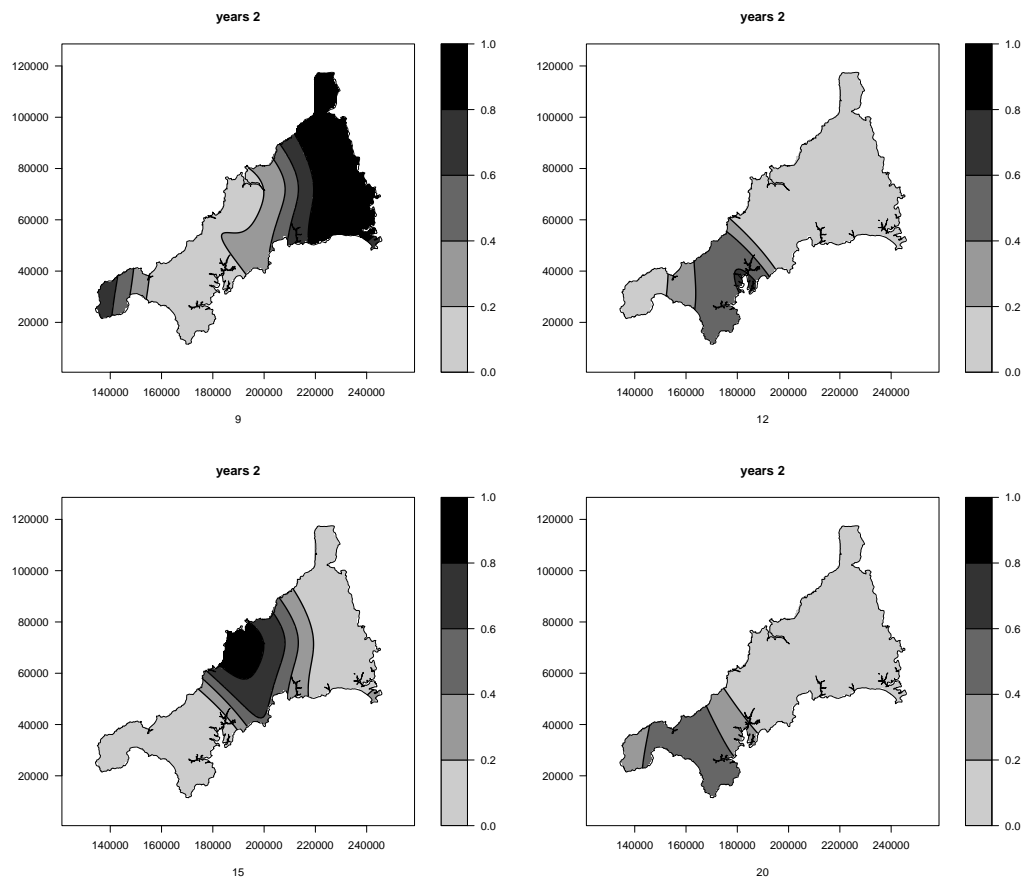


Figure 12: Kernel regression estimates of type-specific probability surfaces for cases in years 1999-2000.

5 Real-time surveillance for gastroenteric disease: log-Gaussian Cox process modelling

Our second application has previously been reported in greater detail by Diggle, Knorr-Held, Rowlingson, Su, Hawtin and Bryant (2003) and by Diggle, Rowlingson and Su (2005). It concerns the development of a real-time surveillance system for non-specific gastroenteric disease in the county of Hampshire, UK, using the data described in Section 1.1.3.

The methodological problem is to build a descriptive model of the normal pattern of spatio-temporal variation in the distribution of incident cases, and to use the model to identify unusual, spatially and temporally localised, departures from this pattern, which we call “anomalies.” The wider aim is that statistical evidence of current anomalies in the spatio-temporal distribution of incident cases can then be combined with other forms of evidence, for example reports from pathology laboratory analyses of faecal samples, to trigger an earlier response to an emerging problem than

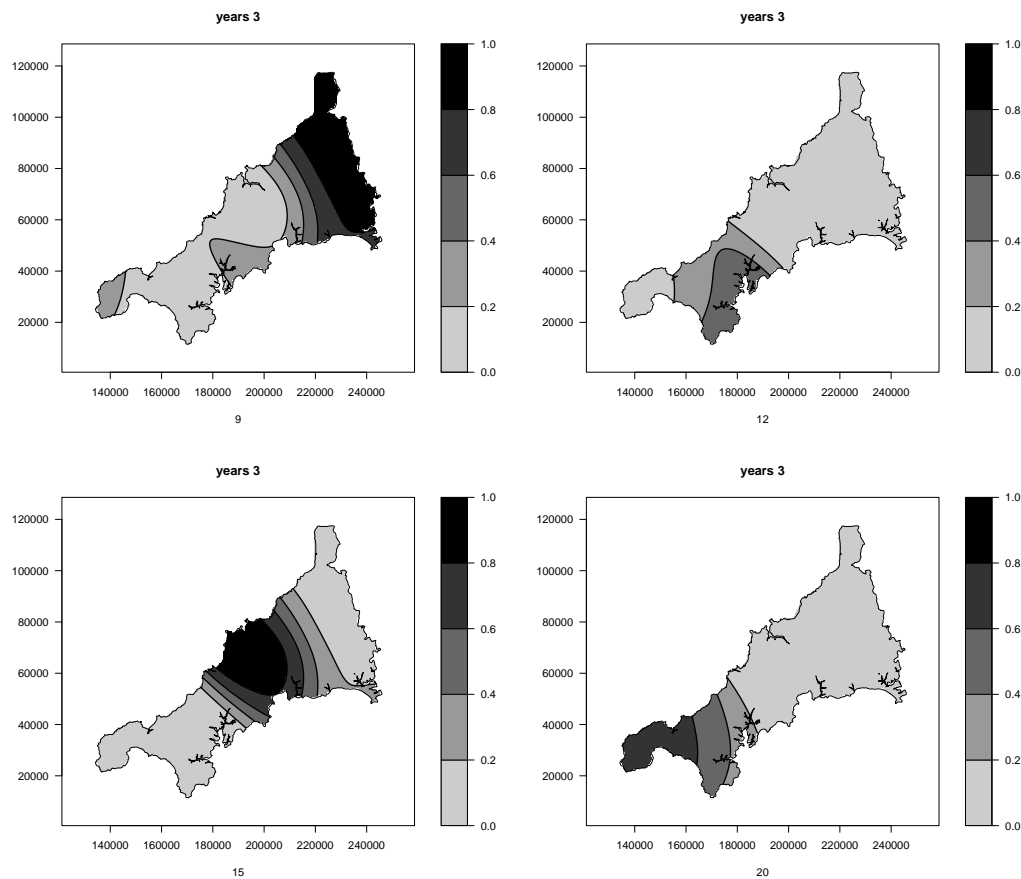


Figure 13: Kernel regression estimates of type-specific probability surfaces for cases in years 2001-02.

is typically achieved by current surveillance systems. For further discussion, see Diggle *et al* (2003).

Our descriptive model is a log-Gaussian Cox process of the kind proposed by Brix and Diggle (2001) and discussed in Section 3.2.1. Our model needs to allow for both spatial and temporal heterogeneity in the rate of calls to NHS Direct. The heterogeneity arises through a combination of factors including spatial variation in the density of the population at risk and aspects of the pattern of usage of NHS Direct by different sectors of the community; for example, there is a clear day-of-the-week effect due to the relative inaccessibility of other medical service at weekends, whilst anecdotal evidence suggests that NHS Direct is used disproportionately more often by young families than by the elderly. We assume that these spatial and temporal effects operate independently, whereas the spatially and temporally localised anomalies which we wish to detect are governed by a spatio-temporally correlated stochastic process. Hence, within the framework of log-Gaussian Cox processes, we postulate a spatio-

temporal intensity function for incident cases of the form

$$\lambda(x, t) = \lambda_0(x)\mu_0(t) \exp\{S(x, t)\}, \quad (23)$$

where $S(x, t)$ is a stationary, spatio-temporal Gaussian process with expectation $E[S(x, t)] = -0.5\sigma^2$, variance $\text{Var}\{S(x, t)\} = \sigma^2$ and separable correlation function $\text{Corr}\{S(x, t), S(x - u, t - v)\} = \rho(u, v) = r(u) \exp(-v/\beta)$. As noted earlier, this specification guarantees that $E[\exp\{S(x, t)\}] = 1$ for all x and t . We now add the constraint that $\lambda_0(x)$ integrates to 1 over the study-region. The function $\mu_0(t)$ then represents the time-varying total intensity, or mean number per unit time, of incident cases over the whole county, whilst $\lambda_0(x)$ becomes a probability density function for the spatial distribution of incident cases, averaged over time. The spatial correlation function $r(u)$ is in principle arbitrary, but we have found that a simple exponential, $r(u) = \exp(-u/\phi)$, gives a reasonable fit to the Hampshire data.

To fit the model, we need to estimate the functions $\lambda_0(x)$ and $\mu_0(t)$, and the additional parameters σ^2 , ϕ and β which specify the Gaussian process $S(x, t)$. We consider each of these estimation problems in turn.

For the spatial density $\lambda_0(x)$, it is hard to envisage a suitable parametric model. Also, we cannot assume that the spatial distribution of the relevant population, namely users of NHS Direct, matches that of the overall population distribution over the county of Hampshire, hence we cannot use census information to estimate $\lambda_0(x)$. We therefore use a non-parametric kernel density estimation method. This is very similar to the kernel regression method discussed in Section 4 but adapted to the density estimation setting. Using the Gaussian kernel (21), and band-width h , the kernel estimate of $\lambda_0(x)$ is

$$\hat{\lambda}_0(x) = n^{-1} \sum_{i=1}^n h^{-2} w\{h^{-1}(x - x_i)\}, \quad (24)$$

where $x_i : i = 1, \dots, n$ are the case-locations. Because of the very severe variations in population density across the county, any choice of a fixed band-width h is liable to be an unsatisfactory compromise between the relatively large and small band-widths which would be appropriate in the more rural and urban areas, respectively. We therefore follow a suggestion in Silverman (1985), in which we first construct a fixed band-width pilot estimator $\tilde{\lambda}_0(x_i)$ using (24) with a subjectively chosen band-width h_0 , then calculate \tilde{g} , the geometric mean of $\tilde{\lambda}_0(x_i) : i = 1, \dots, n$, and use a locally adaptive kernel estimator

$$\hat{\lambda}_0(x) = n^{-1} \sum h_i^{-2} \phi\{(x - x_i)/h_i\}$$

with $h_i = h_0\{\tilde{\lambda}_0(x_i)/\tilde{g}\}^{-0.5}$. This has the required effect of applying a larger band-width in the more rural areas, where $\tilde{\lambda}_0(x)$ is relatively small. Figure 14 shows the resulting estimate $\hat{\lambda}_0(x)$.

To estimate the mean number of cases per day, $\mu_0(t)$, a parametric approach is more reasonable. We use a Poisson log-linear regression model incorporating day-of-the-week effects as a 7-level factor, time-of-year effects as sine-cosine wave at frequency

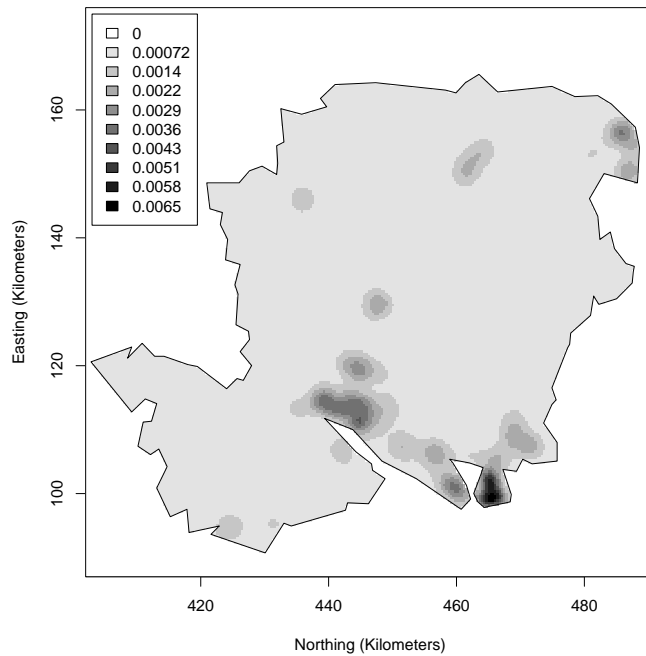


Figure 14: Locally adaptive kernel estimate of $\lambda_0(x)$ for the Hampshire gastroenteric disease data.

$\omega = 2\pi/365$, plus its first harmonic, and a linear trend to reflect progressive up-take in the usage of NHS Direct during the period covered by the data. Hence,

$$\log \mu_0(t) = \delta_{d(t)} + \alpha_1 \cos(\omega t) + \eta_1 \sin(\omega t) + \alpha_2 \cos(2\omega t) + \eta_2 \sin(2\omega t) + \gamma t, \quad (25)$$

where $d(t)$ codes the day of the week. The Poisson formulation does not account for the extra-Poisson variation which, as anticipated, the data exhibit, but nevertheless produces consistent estimates on the assumption that (25) is a correct specification for $\mu_0(t)$. Figure 15 compares the resulting estimate, centred on the average of the seven daily intercepts $\delta_{d(t)}$, with the observed numbers of calls per day, averaged over each week. The spring peak in incidence is a well-known feature of this group of diseases.

To estimate the parameters of the stochastic component of the model, $S(x, t)$, we have used a simple method-of-moments approach, based on matching empirical and theoretical second-moment properties of the data and model, respectively. We are currently developing an implementation of a Monte Carlo maximum likelihood method, based on material in Moller and Waagepetersen (2004). A partial justification for the method-of-moments approach is that the main goal of the analysis is real-time spatial prediction, whose precision is limited by the relatively low daily incidence of cases, whereas parameter estimation draws on the complete set of data; hence, efficiency of parameter estimation is not crucial.

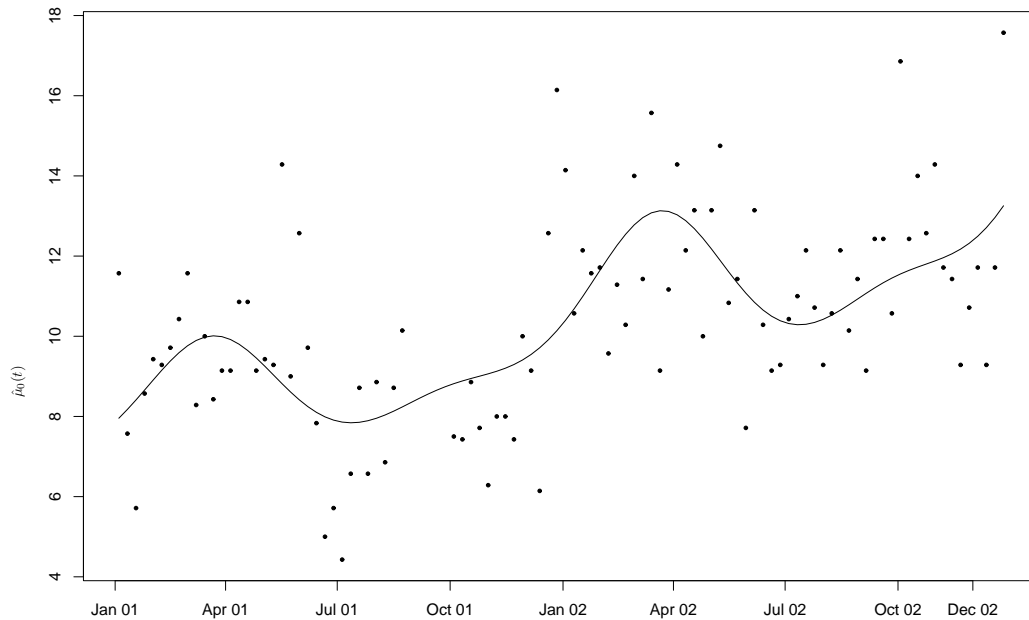


Figure 15: Fitted log-linear model for the mean number of gastroenteric disease cases per day adjusted for day-of-the-week effects (solid line), and observed numbers of cases per day averaged over successive weeks (solid dots).

Consider first the parameters of the spatial covariance structure of $S(x, t)$, namely

$$\text{Cov}\{S(x, t), S(x - u, t)\} = \sigma^2 \exp(-|u|/\phi).$$

The corresponding spatial pair correlation function is $g(u) = \exp\{\sigma^2 \exp(-|u|/\phi)\}$, and the estimation method consists of minimising the criterion

$$\int_0^{u_0} [\{\log \hat{g}(u)\} - \{\log g(u)\}]^2 du, \quad (26)$$

where $u_0 = 2\text{km}$ is chosen subjectively as the apparent range of the spatial correlation, and $\hat{g}(u)$ is a non-parametric estimate of the pair correlation function. We use a time-averaged kernel estimator,

$$\hat{g}(u) = \frac{1}{2\pi u T |A|} \sum_{t=1}^T \sum_{i=1}^n \sum_{i \neq j} \frac{K_h(u - \|x_i - x_j\|) w_{ij}}{\lambda_t(x_i) \lambda_t(x_j)}. \quad (27)$$

In (27), each of the summations over $i \neq j$ refers to pairs of events occurring on the same day, T is the length of the study-period, A the study area, w_{ij} is Ripley's (1977) edge-correction as used in (4), $\lambda_t(x) = \hat{\lambda}_0(x) \hat{\mu}_0(t)$ and

$$K_h(u) = \begin{cases} 0.75h^{-1}(1 - u^2/h^2) & : -h \leq u \leq h \\ & : \text{otherwise.} \end{cases}$$

To estimate the temporal correlation parameter β , we again match empirical and theoretical second-moment properties, as follows. Let N_t denote the numbers of incident cases on day t . For our model, the time-variation in $\mu_0(t)$ makes the covariance structure of N_t non-stationary. We obtain

$$\begin{aligned} \text{Cov}(N_t, N_{t-v}) &= \mu_0(t)\mathbf{1}(v=0) + \{\mu_0(t)\mu_0(t-v)\} \times \\ &\quad \left\{ \int_W \int_W \lambda_0(x_1)\lambda_0(x_2) \exp[\sigma^2 \exp(-v/\beta) \exp(-u/\phi)] dx_1 dx_2 - 1 \right\} \end{aligned} \quad (28)$$

Note that the expression for $\text{Cov}(N_t, N_{t-v})$ given by Brix and Diggle (2001) is incorrect. To estimate β we minimise

$$\sum_{v=1}^{v_0} \sum_{t=v+1}^n \{\hat{C}(t, v) - C(t, v; \beta)\}^2,$$

where $v_0 = 14$ days, $\hat{C}(t, v) = N_t N_{t-v} - \hat{\mu}_0(t)\hat{\mu}_0(t-v)$ and $C(t, v; \theta) = \text{Cov}(N_t, N_{t-v})$ as defined in (28) but plugging in the previously estimated values for σ^2 and ϕ .

The resulting estimates are shown in Figure 16. Note in particular that the scales on which the spatial and temporal dependence decay are broadly consistent with the known character of the diseases in question, which are spread by direct contact between infected individuals and generally have latent periods of the order of a few days.

We now use the fitted model for spatial prediction, as follows. Firstly, and with the same justification as given above for the use of potentially inefficient methods of parameter estimation, we use plug-in values of the estimated model parameters, thereby ignoring the effects of parameter uncertainty on the predictive distributions of interest. We then use a Metropolis-adjusted Langevin algorithm, as described in Brix and Diggle (2001), to generate samples from the conditional distribution of $S(x, t)$ given data up to time t . For display purposes, we choose a critical threshold value $c > 1$ and map predictive exceedance probabilities,

$$p_t(x) = \text{P}(\exp\{S(x, t)\} > c | \text{data}) \quad (29)$$

We would argue that mapping $p_t(x)$ is more relevant than mapping predictive estimates, $\hat{S}(x, t)$. The latter are highly variable because of the relatively low daily incidence and would be liable to over-interpretation. We suggest that the value of c should be chosen by the public health practitioner to represent a multiplicative increase in the local daily incidence which, if verified, would be regarded as practically significant. A high predictive probability that this threshold has been exceeded at a particular time and place would then suggest that some kind of follow-up action may be required, for example to ascertain whether recent cases in the immediate vicinity might have a common cause, or are more likely to be unrelated chance occurrences.

Figure 17 shows an example of the resulting map of $p_t(x)$ for one day in March 2003, using the threshold value $c = 2$. Note that the colour-scale for $p_t(x)$ is continuous,

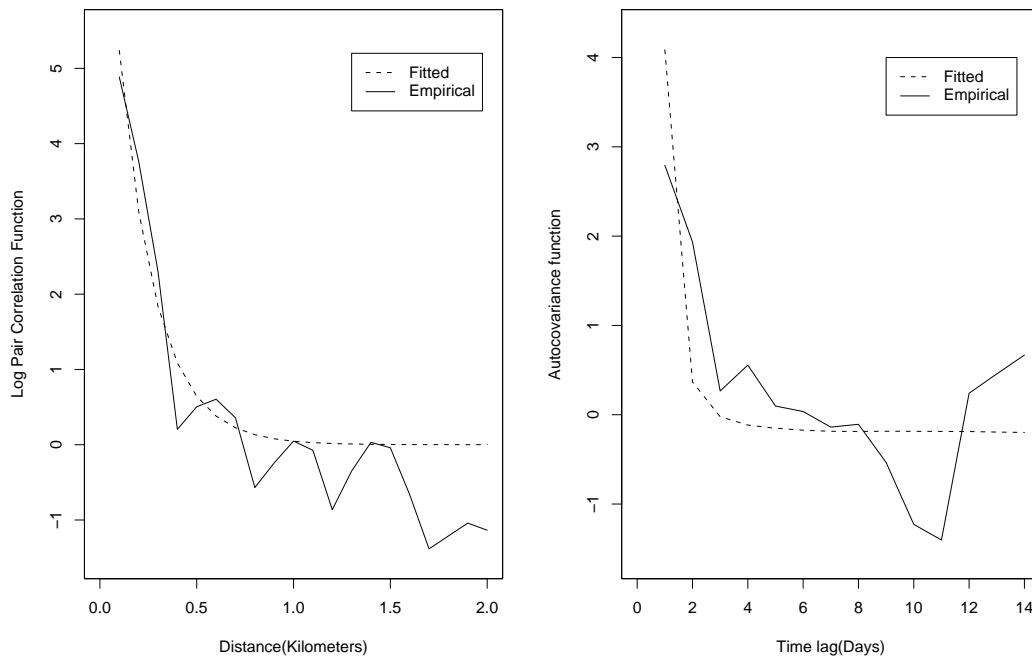


Figure 16: Empirical (solid line) and fitted (dashed line) second-moment properties for the Hampshire gastroenteric disease data. The left-hand panel shows the logarithm of the pair correlation function, the right-hand panel the temporal autocovariance function.

but non-linear so that only predictive probabilities close to 1 show up as orange or red. We also emphasise that the value $c = 2$ was chosen for illustrative purposes and is much smaller than would be associated, in this application, with a genuine outbreak of public health concern.

As noted above, one way to refine the methodology associated with this application would be to use Monte Carlo maximum likelihood estimation for the covariance parameters of $S(x, t)$. A brief outline of this method follows. We approximate the continuous spatio-temporal domain $A \times (0, T)$ by a fine lattice, and denote by X and S the point process data and the latent Gaussian process in this discretised domain. The likelihood for the model, $L(\theta; X)$ say, is the marginal distribution of X , which is intractable, whereas the joint distribution of X and S can be written explicitly as the product of the conditional distribution of X given S and the marginal distribution of S , both of which have known standard forms – Poisson and Gaussian, respectively. It follows that for any fixed value θ_0 ,

$$\begin{aligned}
 L(\theta; X) &= \int f(S, X; \theta) dS \\
 &= \int f(S, X; \theta) \times \frac{L(\theta_0; X)}{L(\theta_0; X)} \times \frac{f(S, X; \theta_0)}{f(S, X; \theta_0)} dS
 \end{aligned}$$

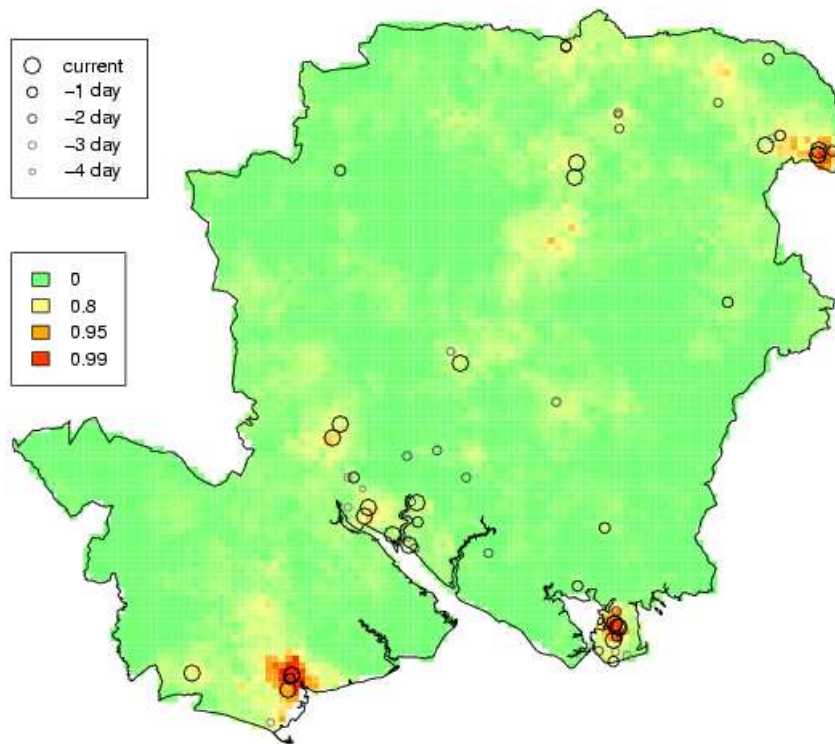


Figure 17: Predictive exceedance probabilities $p_t(x)$ for the Hampshire gastroenteric disease data on a day in March 2003, using the threshold value $c = 2$. Incident cases over five consecutive days are shown as circles of diminishing size to correspond to the progressive discounting of past data in constructing the predictive distribution of $S(x, t)$.

$$= L(\theta_0; X) \int \frac{f(S, X; \theta)}{f(S, X; \theta_0)} f(S|X; \theta_0) dS.$$

Hence, the likelihood ratio between θ and θ_0 is

$$L(\theta; X)/L(\theta_0; X) = E_{\theta_0} \left[\frac{f(S, X; \theta)}{f(S, X; \theta_0)} \right], \quad (30)$$

where $E_{\theta_0}[\cdot]$ denotes expectation with respect to the conditional distribution of S given X at $\theta = \theta_0$. In (30), we now replace the theoretical expectation by a sample mean over Monte Carlo simulations, using the same Metropolis-adjusted Langevin algorithm as used earlier to simulate samples from the predictive distribution of S given X . Note also that writing $S = a(\theta) + B(\theta)Z$, where Z is a vector of independent standard Gaussian variates, gives $f(S, X; \theta) = f(X|Z; \theta)f(Z)$, in which case the ratio within the expectation term on the right hand side of (30) reduces to $f(X|Z; \theta)/f(X|Z; \theta_0)$.

Another possible refinement of the methodology would be to use a stochastic model of daily incidence in place of the deterministic $\mu_0(t)$. In particular, the rising trend

evident during the two-year period covered by the data which were used to develop the model was not sustained thereafter. More generally, patterns in the usage of NHS Direct are liable to fluctuate in response to changes in policy-related factors within the NHS system whose combined effects are hard to capture in a deterministic model. A pragmatic strategy would be the following. Postulate a temporal log-Gaussian Cox process for the daily incident counts, N_t , as a series of conditionally independent Poisson counts with conditional expectations $\exp(\alpha_t + M_t)$, where α_t encodes day-of-the-week effects and M_t is a latent Gaussian process. Use this model to generate real-time predictions $\hat{\alpha}_t + \hat{M}_t$, which can then be plugged into the spatial prediction algorithm in place of the log-linear regression estimates $\hat{\mu}_0(t)$.

6 Foot-and-mouth disease: mechanistic modelling and partial likelihood analysis

To illustrate the partial likelihood method described in Section 3.2.2, we analyse the foot-and-mouth data using a mechanistic model proposed by Keeling et al (2001). The analysis reported here is taken from Diggle (2005).

The model assumes that the rate of transmission of infection between an infected farm i and a susceptible farm j is given by

$$\lambda_{ij}(t) = \lambda_0(t)A_iB_jf(\|x_i - x_j\|)I_{ij}(t), \quad (31)$$

where $\lambda_0(t)$ is a baseline rate of infection, A_i and B_j encode characteristics of the farms in question which affect their infectivity and susceptibility respectively, $f(\cdot)$ is a *transmission kernel* which models the spread of infection as a function of distance, and $I_{ij}(t)$ is an indicator that farm i is infective and farm j susceptible at time t . The data include, as applicable for each animal-holding farm, the dates on which the disease was reported and on which the stock was culled. To allow for reporting delays, we assume that a farm is reported as a case τ days after it becomes infective. A farm remains infective unless and until its stock is culled. A farm is susceptible if it is not infective, and has not had its stock culled.

In the analysis reported here, we assume that $\tau = 5$ days and model the infectivity and susceptibility factors as

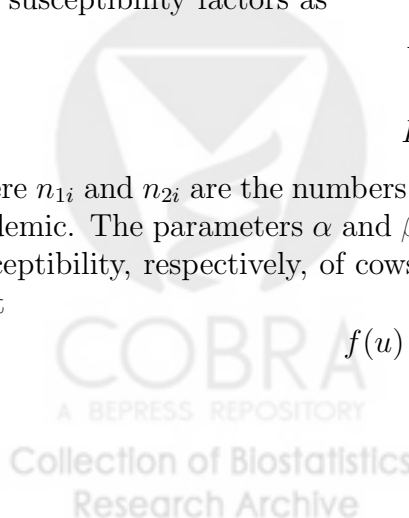
$$A_i = (\alpha n_{1i} + n_{2i}) \quad (32)$$

and

$$B_j = (\beta n_{1j} + n_{2j}). \quad (33)$$

where n_{1i} and n_{2i} are the numbers of cows and sheep held by farm i at the start of the epidemic. The parameters α and β therefore represent the relative infectiousness and susceptibility, respectively, of cows to sheep. For the transmission kernel, we assume that

$$f(u) = \exp\{-(u/\phi)^\kappa\} + \rho. \quad (34)$$



Keeling et al (2001), did not specify a functional form for $f(\cdot)$, but (34) captures the qualitative features of the results which they reported in graphical form. In particular, they reported a sharper-than-exponential decay with distance, which in (34) would correspond to $\kappa < 1$, whilst ρ in (34) represents the contribution to the epidemic from apparently spontaneous cases which occur remotely from any previous case.

For any farm k , we define $\lambda_k(t) = \sum_j \lambda_{jk}(t)$, from which we obtain the conditional intensities

$$\lambda(x_i, t_i | \mathcal{H}_{t_i}) = \lambda_i(t_i) / \sum_k \lambda_k(t_i).$$

The partial log-likelihood follows by substitution of the conditional intensities into (16) and (17). To maximise the partial log-likelihood we use the Nelder-Mead simplex algorithm (Nelder and Mead, 1965) as implemented in the R function `optim()`, which provides a numerical estimate of the Hessian matrix.

In the model for the transmission kernel, the parameters κ and ρ are poorly identified because the cases which appear to correspond to long-range transmission are few in number, and can be explained empirically either by including a small, positive value of ρ or by adjusting the value of κ . Because ρ corresponds formally to what is known to be a real effect, namely the indirect spread of infection via the movement of farm equipment and staff, we retain ρ as a positive-valued parameter to be estimated, but fix $\kappa = 0.5$ to correspond to the observation in Keeling et al (2001) that the transmission kernel is more sharply peaked than exponential.

We first investigated whether the data in Cumbria and Devon support the assumption of a common set of parameters in the two counties. The likelihood ratio test statistic for common versus separate parameters is 2.98 on 4 degrees of freedom, hence $p = 0.56$ and we therefore accepted the hypothesis of common parameter values. We then obtained common parameter estimates $(\hat{\alpha}, \hat{\beta}, \hat{\phi}, \hat{\rho}) = (4.92, 30.68, 0.39, 9.9 \times 10^{-5})$. For all practical purposes, $\hat{\rho} \approx 0$, although a likelihood ratio test formally rejects $\rho = 0$ because the likelihood is sensitive to the precise probabilities which the model assigns to rare events.

One question of specific interest is whether the infectivities and susceptibilities for individual farms, A_i and B_j , are linear or sub-linear in the numbers of animals. To investigate this, we extend (32) and (33) to $A_i = (\alpha n_{1i}^\gamma + n_{2i}^\gamma)$ and $B_j = (\beta n_{1i}^\gamma + n_{2i}^\gamma)$, respectively, where γ is an additional parameter to be estimated. Fitting this five-parameter model results in a large increase in the maximised log-likelihood, from -6196.3 to -5861.4 .

Another possible extension of the model would be to include farm-level covariates by defining $A_i = (\alpha n_{1i}^\gamma + n_{2i}^\gamma) \exp(z_i' \delta)$, where z_i is a vector of covariates for farm i , with a similar expression for the susceptibilities B_j . The z_i might, for example, codify management practices or other measured characteristics of individual farms which could affect their propensity to transmit, or succumb to, the disease. By way of illustration, we consider adding a log-linear effect of farm area to the model. The likelihood ratio statistic for the covariate effect is 3.26 on 1 degree of freedom,

Parameter	Estimate	95% confidence interval	
α	1.42	1.13	1.78
β	36.17	0.19	692.92
ϕ	0.41	0.36	0.48
ρ	1.3×10^{-4}	8.5×10^{-5}	2.1×10^{-4}
γ	0.13	0.09	0.21

Table 1: Parameter estimation for the five-parameter model fitted to combined data from Cumbria and Devon

corresponding to $p = 0.07$. However, we can expect this test to be rather weak, because the observed distribution of farm area is extremely skewed, and the few farms with large areas will therefore have high leverage.

Estimates for the five-parameter model are shown in Table 1, together with approximate 95% confidence limits deduced from the numerical estimate of the Hessian matrix. Optimisation was conducted on the log-scale for all parameters, which is why the confidence limits are not symmetric about the point estimates. Estimated correlations amongst the parameter estimates are all small, the largest being 0.25 between $\log \phi$ and $\log \rho$. The results in Table 1 indicate a strongly sub-linear dependence of infectivity and susceptibility on the numbers of animals. Note that under the weak form of dependence implied by the estimate of γ , the estimate of β is very imprecise.

These results are qualitatively similar to those reported in Keeling et al (2001), although they only considered the case $\gamma = 1$. They reported point estimates $\tilde{\alpha} = 1.61$ and $\tilde{\beta} = 15.2$. They did not specify a parametric model for the transmission kernel but their Figure 1B shows similar behaviour to our fitted model, decaying from 1 at $u = 0$ to approximately 0.1 at $u = 1\text{km}$, compared with our $\hat{f}(1) = 0.21$.

Finally, we use a simple adaptation of the Nelson-Aalen estimator (Andersen, Borgan, Gill and Keiding, 1992, Chapter 4) to obtain a non-parametric estimate of the cumulative base-line hazard,

$$\hat{\Lambda}_0(t) = \int_0^t \hat{\lambda}_0(u) du.$$

We re-write (31) as $\lambda_{ij}(t) = \lambda_0(t)\rho_{ij}(t)$ and define $\rho(t) = \sum_i \sum_j \rho_{ij}(t)$. The Nelson-Aalen estimator is now given by

$$\begin{aligned} \hat{\Lambda}_0(t) &= \int_0^t \hat{\rho}(u)^{-1} dN(u) \\ &= \sum_{i:t_i \leq t} \hat{\rho}(t_i)^{-1}, \end{aligned} \tag{35}$$

where $\hat{\rho}(t)$ is the parametric estimate of $\rho(t)$ implied by the fitted model. Figure 18 shows the Nelson-Aalen estimates obtained from the Cumbria and Devon data. The

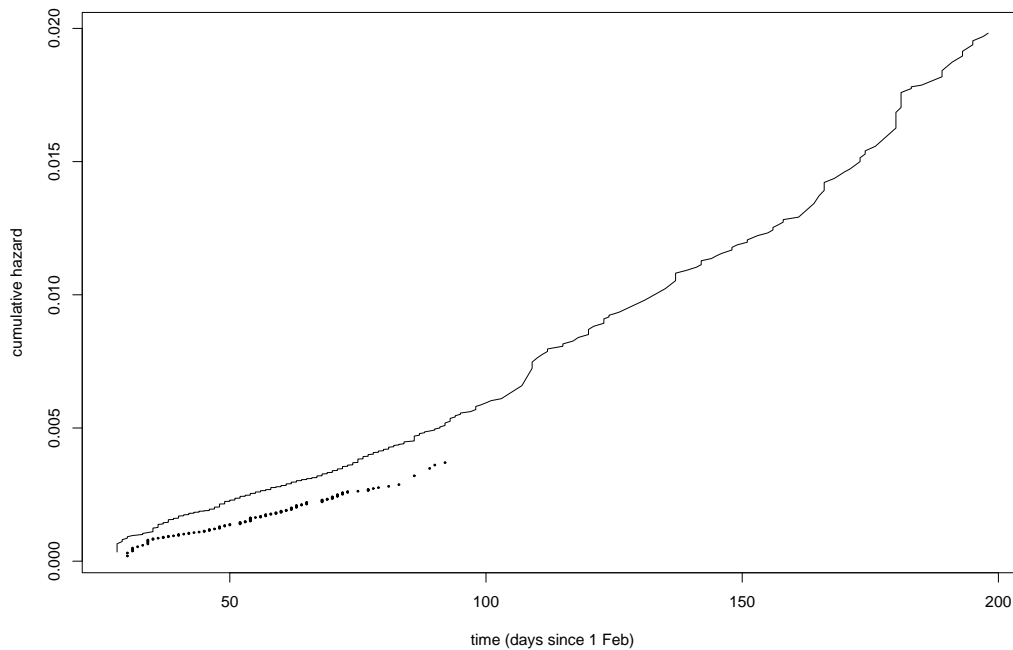


Figure 18: Estimated cumulative baseline hazards in Cumbria (solid line) and in Devon (dotted line) for the five-parameter model.

generally lower estimates for Devon are consistent with the lower overall prevalence of the disease (137 cases out of 8182 at-risk farms in Devon, 657 cases out of 5090 at-risk farms in Cumbria). Both estimates are approximately linear over the first two to three months, by which time the epidemic in Devon has almost run its course. The slope of the Cumbria estimate increases thereafter. This does not necessarily imply a failure of the culling strategies being applied, since the model already takes account of their effects, but rather suggests that external environmental effects, for example the increase in animal movements outdoors in spring and summer, may have promoted an increase in the virulence of the disease process.

This analysis of the foot-and-mouth data demonstrates the feasibility of using the partial likelihood approach to answer a variety of questions relevant to an understanding of the underlying disease process. The method of fitting is quite flexible, and it would be straightforward to extend the model in various ways, for example by including additional farm-level covariates. A fuller analysis of the data will be reported separately. More generally, the partial likelihood provides a useful method for analysing any spatio-temporal point process model which is specified via its conditional intensity function. The method is based on a generally accepted principle of inference with known asymptotic properties, whilst being computationally straightforward and therefore well-suited to routine use.

7 Discussion

In this chapter, we have tried to indicate the scope for simple adaptation of existing methods, and for development of new methods, to analyse spatio-temporal point process data. Data-sets of this kind are becoming more widely available, perhaps reflecting the increased ability in many fields of science to capture such data routinely. More fundamentally, interesting scientific questions concerning the dynamic spatio-temporal behaviour of natural systems cannot easily be addressed using purely spatial or temporal information.

In the author's opinion, simple adaptations of existing spatial statistical methods, as illustrated by our analysis of the bovine tuberculosis data, will continue to be valuable for descriptive analysis, and are more or less forced on us when the data are coarsely discretised in time. But when data with a fine resolution in both the spatial and temporal dimensions are available, models which explicitly recognise the directional nature of time by conditioning future behaviour on past outcomes are likely to be more insightful. Our analyses of the gastroenteric disease data and of the foot-and-mouth data illustrate this general philosophy. For the gastroenteric disease application our model for the latent stochastic process $S(x, t)$ is, essentially, a multivariate time series model incorporating a qualitatively sensible, if admittedly also computationally convenient, Markov dependence structure in time. The cross-correlation structure of the model is then chosen so as to have a parsimonious spatial interpretation whilst giving a reasonable empirical fit to the data. In contrast, for the foot-and-mouth data we use a previously proposed mechanistic model in which the conditioning on past events is used explicitly to quantify the current risk of transmission of the disease from an infective to a susceptible farm.

As in other areas of statistics, the development of computationally intensive, Monte Carlo methods of inference has greatly enhanced our ability to fit relatively complex and realistic models. However, it is still all too easy to find combinations of data and model for which fitting by Monte Carlo methods is computationally infeasible. Also, the requirement to tune Monte Carlo algorithms to each non-standard application imposes very real constraints on the statistician's ability to compare a range of candidate models within a reasonable time-scale whilst ensuring that Markov chain Monte Carlo algorithms have converged to their equilibrium distributions and that inferences about parameter combinations of interest are appropriately insensitive to pragmatic choices of multivariate priors. For these reasons, we see a place for methods of inference such as the partial likelihood method proposed in Diggle (2005) and applied here to the foot-and-mouth data, which are capable of routine implementation whilst still being based on generally accepted statistical principles.

Acknowledgements

This work was supported by the UK Engineering and Physical Sciences Research Council through the award of a Senior Fellowship to Peter Diggle (Grant number GR/S48059/01) and by the USA National Institute of Environmental Health Science through Grant number 1 R01 ES012054, Statistical Methods for Environmental Epidemiology.

I thank Prof. Abbie Hughes and Dr. Elzbieta Wieniawa Narkiewicz for providing the amacrine cell data, and Laura Green, Matt Keeling, Graham Medley, Sir David Cox and Hugues Lassalle for very helpful discussions concerning the foot-and-mouth data.

The application to the bovine tuberculosis data was supported by the UK Department for Environment, Food and Rural Affairs (Grant number SE3020). Roger Sainsbury of the State Veterinary Service helped collect the data. Jackie Inwald of the Department of Bacterial Diseases, VLA Weybridge, carried out the spoligotyping.

The application to gastroenteric disease surveillance was supported by a grant from the Food Standards Agency, UK, and from the National Health Service Executive Research and Knowledge Management Directorate.

References

- Andersen, P.K., Borgan, O., Gill, R.D. and Keiding, N. (1992). *Statistical Models Based on Counting Processes*. New York : Springer.
- Banerjee, S., Carlin, B.P and Gelfand, A.E. (2003). *Hierarchical Modelling and Analysis for Spatial Data*. London: CRC Press.
- Besag, J. (1974). Spatial interaction and the statistical analysis of lattice systems (with Discussion). *Journal of the Royal Statistical Society B* **36**, 192–225.
- Bhopal, R.S., Diggle, P.J. and Rowlingson, B.S. (1992). Pinpointing clusters of apparently sporadic cases of legionnaires' disease. *British Medical Journal*, **304**, 1022-7.
- Brix, A. and Diggle, P.J. (2001). Spatio-temporal prediction for log-Gaussian Cox processes. *Journal of the Royal Statistical Society, B* **63**, 823–41.
- Brix, A. and Diggle, P.J. (2003). Corrigendum: Spatio-temporal prediction for log-Gaussian Cox processes. *Journal of the Royal Statistical Society, B* **65**, 946.
- Brown, P.E., Karsen, K., Roberts, G.O. and Tonellato, S. (2000). Blur-generated non-separable space-time models. *Journal of the Royal Statistical Society, B* **62**, 847–60.
- Cooper, D.L., Smith, G.E., O'Brien, S.J., Hollyoak, V.A. and Baker, M. (2003). What can analysis of calls to NHS Direct tell us about the epidemiology of gastrointestinal

infections in the community? *Journal of Infection*, **46**, 101–5.

Cox, D.R. (1972). Regression models and life tables (with Discussion). *Journal of the Royal Statistical Society B*, **34**, 187–220.

Cox, D.R. (1975). Partial likelihood. *Biometrika*, **62**, 269–75.

Cox, D.R. (1977). The role of significance tests. *Scandinavian Journal of Statistics*, **4**, 49–71.

Cressie, N.A.C. (1991). *Statistics for Spatial Data*. New York : Wiley.

Daley D J and Vere-Jones D (1988). *An Introduction to the Theory of Point Processes*. New York : Springer.

Diggle, P.J. (2003). *Statistical Analysis of Spatial Point Patterns (second edition)*. London : Arnold.

Diggle, P.J. (2005). A partial likelihood for spatio-temporal point processes. *Johns Hopkins University, Department of Biostatistics Working Papers*, Working Paper 75. <http://www.bepress.com/jhubiostat/paper75>

Diggle, P.J., Chetwynd, A.G., Haggkvist, R. and Morris, S. (1995). Second-order analysis of space-time clustering. *Statistical Methods in Medical Research*, **4**, 124–36.

Diggle, P.J., Eglen, S.J. and Troy, J.B. (2005). Modelling the bivariate spatial distribution of amacrine cells. In *Case Studies in Spatial point Processes*, ed A. Baddeley, J. Mateu, D. Stoyan. New York : Springer (to appear).

Diggle, P.J., Knorr-Held, L., Rowlingson, B., Su, T., Hawtin, P. and Bryant, T. (2003). Towards On-line Spatial Surveillance. In *Monitoring the Health of Populations: Statistical Methods for Public Health Surveillance.*, ed R. Brookmeyer and D. Stroup. Oxford : Oxford University Press.

Diggle, P.J., Rowlingson, B. and Su, T-L. (2005). Point process methodology for on-line spatio-temporal disease surveillance. *Environmetrics*, **16**, 423–34.

Diggle, P.J., Zheng, P. and Durr, P. (2005). Non-parametric estimation of spatial segregation in a multivariate point process. *Applied Statistics*, **54**, 645–58.

Durr, P.A., Hewinson, R.G. and Clifton-Hadley, R.S. (2000). Molecular epidemiology of bovine tuberculosis: I. *Mycobacterium bovis* genotyping. *Revue Scientifique et Technique de l'office Internationale des Epizooties*, **19**, 675–688.

Geyer, C.(1999). Likelihood inference for spatial point processes. In *Stochastic Geometry: likelihood and computation*, ed O.E. Barndorff-Nielsen, W.S. Kendall and M.N.M. van Lieshout, 79–140.

Geyer, C.J. and Thompson, E.A. (1992). Constrained Monte Carlo maximum likelihood for dependent data (with Discussion). *Journal of the Royal Statistical Society, B* **54**, 657–99.

Gneiting, T. (2002). Nonseparable, stationary covariance functions for space-time

data. *Journal of the American Statistical Association*, **97**, 590–600.

Hughes, A. (1985) New perspectives in retinal organisation. *Progress in Retinal Research* **4** 243–314.

Keeling, M.J., Woolhouse, M.E.J., Shaw, D.J., Matthews, L., Chase-Topping, M., Haydon, D.T., Cornell, S.J., Kappey, J., Wilesmith, J. and Grenfell, B.T. (2001). Dynamics of the 2001 UK foot and mouth epidemic: stochastic dispersal in a heterogeneous landscape. *Science*, **294**, 813–7.

Moller, J. and Sorensen, M. (1994). Statistical analysis of a spatial birth-and-death process model with a view to modelling linear dune fields. *Scandinavian Journal of Statistics*, **21**, 1–19.

Moller, J. and Waagepetersen, R.P. (2004). *Statistical Inference and Simulation for Spatial Point Processes*. London : Chapman and Hall.

Moller, J., Syversveen, A. and Waagepetersen, R. (1998). Log Gaussian Cox processes. *Scandinavian Journal of Statistics*, **25**, 451–82.

Nelder, J.A. and Mead, R. (1965). A simplex algorithm for function minimisation. *Computer Journal*, **7**, 308–13.

Neyman, J. and Scott, E.L. (1958). Statistical approach to problems of cosmology. *Journal of the Royal Statistical Society, Series B* **20**, 1–43.

Peebles, P.J.E. and Groth, E.J. (1975). Statistical analysis of catalogs of extragalactic objects, V. Three point correlation function for the galaxy distribution in the Zwicky catalog. *Astrophysical Journal*, **196**, 1–11.

Ripley, B.D. (1976). The second-order analysis of stationary point processes. *Journal of Applied Probability*, **13**, 255–66.

Ripley, B.D. (1977). Modelling spatial patterns (with discussion). *Journal of the Royal Statistical Society B* **39**, 172–212.

Rue, H. and Held, L. (2005). *Gaussian Markov Random Fields: Theory and Applications*. London: CRC Press.

UK National Audit Office (2002). *The 2001 Outbreak of Foot and Mouth Disease*. Report by the Comptroller and Auditor General, HC 939, Session 2001–2002. London: The Stationery Office.

