CHAPTER 4

Extrapolating and Interpolating Spatial Patterns

M.N.M. van Lieshout *A.J. Baddeley*

4.1 Introduction

Observations of a spatial pattern are typically confined to a bounded region of space, while the original pattern of interest can often be imagined to extend outside. Much attention has been paid to statistical inference for models of the pattern given only the partial observations in the sampling window. Less attention has been given to *prediction* or *extrapolation* of the process (i.e. of the same realisation of the process) beyond the window of observation, conditional on the partially observed realisation. A motivating example is the charting of geological faults encountered during coal mining (Baddeley et al. 1993, Chilès 1989). It is of interest to predict the likely presence of geological faults outside the region mined so far, and thereby to choose between various mining strategies. Other examples may be found in image processing, for instance the problem of replicating a texture beyond the region where it has been observed as in the editing of a video image so that a foreground object is removed and replaced seamlessly by the background texture (De Bonet 1997). Partial observation of a spatial pattern may also include effects such as aggregation by administrative regions, deletion of part of the pattern, and the unobservability of a related pattern. Recovery of full information in this context might be called *interpolation*; it resembles a missing data problem. In the mining problem discussed above, mapped charts represent only those parts of geological faults which were physically encountered. Gaps may arise because the mined region is not convex both at its outer boundary and within this boundary, because pillars of unmined material remain. Hence it is of interest to join observed line segments together and to interpret them as part of the same continuous fault zone, a process that is known as 'interpretation' by geologists. As another example, geostatistics deals with predicting values of a spatial random process (e.g. precipitation or pollution measurements) from observations at known locations (e.g. Journel and Huijbregts 1978, Cressie 1993, Stein 1999) and interpolation techniques

have been developed under the name of conditional simulation for Gaussian and other second-order random fields, as well as for discrete Markov random field models. Relatively few conditional simulation techniques have been developed for spatial processes of geometric features such as points, line segments and filled shapes. Those that exist are based largely on Poisson processes and the associated Boolean models (Lantuéjoul 1997, Kendall and Thönnes 1999, van Lieshout and van Zwet 2001). A major obstacle is the scarcity of spatial models that are both realistic and tractable for simulation. Some exceptions are the following. There has been much interest in the conditional simulation of oil-bearing reservoirs given data obtained from one or more exploration wells (Haldorsen 1983, Chessa 1995). The wells are essentially linear transects of the spatial pattern of reservoir sand bodies. Typically the sand bodies are idealised as rectangles with horizontal and vertical sides of independent random lengths, placed at random locations following a Poisson point process. For line segment processes, Chilès (1989) presents some stochastic models with particular application to modelling geological faults (based largely on Poisson processes), geostatistical inference, and possibilities for conditional simulation; Hjort and Omre (1994) describe a pairwise interaction point process model for swarms of small faults in a fault zone, and Stoica et al. (2000,2001) study a line segment process for extracting linear networks from remote sensing images. Some of these authors have correctly noted the sampling bias effect attendant on observing a spatial pattern of geometric features within a bounded window (analogous to the 'bus paradox'). Techniques from stochastic geometry need to be enlisted to check the validity of simulation algorithms. Extrapolation or interpolation of a spatial pattern entails fitting a stochastic model to the observed data, and computing properties of the conditional distribution of this model given the observed data. We will discuss a variety of stochastic models for patterns of geometric objects, and treat typical issues such as edge effects, occlusion and prediction in some generality. Subsequently, we shall focus on the problem of identifying clusters in a spatial point pattern, which can be regarded as interpolation of a two-type point pattern from observations of points of one type only, the points of the other type being the cluster centres (Baddeley and van Lieshout 1993, Lawson 1993b, van Lieshout 1995, van Lieshout and Baddeley 1995). Applications may be found in epidemiology, forestry, archaeology, coal mining, animal territory research, and the removal of land mines.

4.2 Formulation and Notation

In this section we describe the general framework considered throughout. The spatial pattern is a random closed set (Matheron 1975, Stoyan et al. 1995) U in $\mathbb{R}^d$, typically $d = 2$ or 3. The distribution of U is governed by a parameter θ in some space Θ.

4.2.1 Germ-grain Models

All models considered in this chapter are *germ-grain models* (Stoyan et al. 1995) constructed as follows. There is an underlying process $X = \{X_i,\ i = 1, 2, \ldots\}$ of germs in $\mathbb{R}^d$, each associated with a random compact set Z_i (the 'grain') in $\mathbb{R}^d$ specified by a parameter in some space $\mathcal{Z}$. The 'complete data' process $W = \{(X_i, Z_i)\}$ consists of pairs of germs with their associated grains and hence can be seen as a marked point process. The union of the translated grains, $U = \bigcup_i (X_i + Z_i)$, forms the germ-grain model. We shall be concerned mostly with spatial cluster processes, which can be formulated as germ-grain models where the X_i are the cluster centres, Z_i is the cluster of points or objects associated with centre X_i translated back to the origin (i.e. Z_i is a random finite set of geometric objects), and U is the union pattern. We will sometimes refer to the X_i as the *parents* and to $X_i + Z_i$ as the *daughters* or *offspring* of X_i. If both the cluster centres and their offspring are points, $\mathcal{Z}$ is the space $\mathcal{N}$ consisting of all finite point patterns in $\mathbb{R}^d$. The complete data W then consists of the patterns X and U together with information mapping each member of U to its cluster centre in X. Note that if $X = \{X_i,\ i = 1, 2, \ldots\}$ is a homogeneous Poisson point process, and the Z_i are i.i.d. the random closed set U is a *Boolean model* (see pp. 484–502 Serra 1982). The common distribution of Z_i is called the distribution of the *typical grain*; the germs X_i play only an indirect role. In practice, one observes the intersection $Y = U \cap A$ of U with a compact window $A \subseteq \mathbb{R}^d$. Mostly the window A is fixed and known. More generally, one may assume that A is an observable random set and condition on it, effectively implying A should be ancillary for θ and independent of U. The requirement that A be observable excludes, for example, random thinning models (Cressie 1993, Stoyan et al. 1995). These are unidentifiable in the sense that one cannot distinguish between a point process of low intensity and a heavily thinned point process of higher intensity, without imposing further assumptions.

4.2.2 Problem Statement

The goal is, given data $\mathbf{y} = U \cap A$, to obtain estimates of the conditional expectations of random variables associated with U or W. Note that in the latter case, W will contain grains Z_i^* such that $X_i^* + Z_i^*$ hits the boundary of A. Hence, any extrapolation technique will have to extend Z_i^* as well as locate germ-grain pairs not hitting A. It is important to realise that the individual objects $X_i + Z_i$ in the germ-grain model are not assumed to be observable separately. They are merely an intermediate stage in the construction of the model for the random set U. For example, any object $X_i + Z_i$ which is completely occluded, i.e. contained in the union of other objects, is not observable and may as well be absent. Consequently our

analysis must depend only on the union set U and not on the representation of U as a union of objects $X_i + Z_i$. In other words, if the data image $\mathbf{y}$ can be represented in two different ways

$$\mathbf{y} = \bigcup_{i=1}^{n} (X_i + Z_i) \cap A = \bigcup_{j=1}^{m} (X'_j + Z'_j) \cap A$$

then inference based on either representation must yield identical results. This rules out mark-correlation techniques (Penttinen and Stoyan 1989). Specialising to spatial cluster analysis, inference focuses on the conditional expected number of clusters, the conditional mean number of points per cluster and the posterior distribution of centre locations as well as the strength of evidence for clustering. As for occlusion effects, the whole essence of the problem is that we do not know which data points belong to the same cluster. Below, we adopt a Bayesian strategy and base inference on the posterior distribution of W given $\mathbf{y}$. The parameter vector θ will be estimated by Monte Carlo maximum likelihood (Gelfand and Carlin 1993, Geyer 1999).

4.2.3 Edge Effects and Sampling Bias

Edge effects and sampling bias are bound to arise when a spatial pattern of unbounded extent is observed in a bounded frame (Baddeley 1999). In this section, we illustrate these problems for partial realisations of a Poisson process of geometric objects. Although the Poisson assumption allows for explicit computations, the essential complexities of the general problem are already present. Thus, assume that the germ process $X = \{X_i\}$ is a homogeneous Poisson point process in $\mathbb{R}^d$ with intensity $\lambda > 0$, that the grains Z_i are i.i.d. random compact sets, and that $A \subseteq \mathbb{R}^d$ is a fixed, compact window. We wish to generate a realisation of $U \cap A$. The approach taken will be to sample those objects which wholly or partly intersect A, and to clip the resulting pattern to the window A. First, note that a translated grain $X_i + Z_i$ intersects A if and only if $X_i \in A \oplus \check{Z}_i$, where $A \oplus B = \{a + b : a \in A, b \in B\}$ is the Minkowski sum of two sets $A, B \subseteq \mathbb{R}^d$ and $\check{A} = \{-a : a \in A\}$ is the reflection of A about the origin (Matheron 1975, Serra 1982, Stoyan et al. 1995). Hence, the germ–grain pairs (X_i, Z_i) for which $X_i + Z_i$ hits A form an inhomogeneous Poisson process whose intensity measure has density $\lambda \mathbf{1}\{x \in A \oplus \check{Z}\}$ with respect to the product of Lebesgue measure and the probability distribution of the grains. Write $|\cdot|$ for d-dimensional volume. Then the number of objects intersecting A is Poisson distributed with mean

$$\lambda E(|A \oplus \check{Z}|) \tag{4.1}$$

where the expectation is with respect to the distribution of the typical grain Z, provided (4.1) is finite. Given n objects are present, they are i.i.d.

with density $\mathbf{1}\{x \in A \oplus \check{Z}\}/E(|A \oplus \check{Z}|)$. Turning to the marginal grain distribution, it should be noted that the grains Z_i corresponding to objects which intersect A are not a random sample from the distribution of the typical grain Z. Instead, their distribution is weighted in the sense that Z_i are i.i.d. with distribution

$$P_A(Z \in \cdot) = \frac{E\left(\mathbf{1}\{Z \in \cdot\}\,|A \oplus \check{Z}|\right)}{E(|A \oplus \check{Z}|)} \tag{4.2}$$

where E denotes the expectation with respect to the distribution of the typical grain Z. Thus the sampling bias favours larger grains: a larger object is more likely than a smaller object to intersect A. The sampling bias also depends on the geometry and relative orientations of A and Z. For further information, see Serra (1982) or Stoyan et al. (1995). To simulate $U \cap A$, the properties just described can be used if the function $f(Z) = |A \oplus \check{Z}|$ and the distribution (4.2) can be evaluated analytically. In two dimensions, if A is a disc of radius r and Z is almost surely convex with nonempty interior, then by Steiner's formula (page 200 Santaló 1976)

$$|A \oplus \check{Z}| = |Z| + r\,\mathrm{length}(\partial Z) + \pi r^2 \tag{4.3}$$

almost surely, where $\mathrm{length}(\partial Z)$ denotes the length of the boundary of Z. Hence (4.2) is a mixture of the area-weighted, the length-weighted and the unweighted typical grain distribution. Of course, if Z is a cluster of points, it is not convex. However, if the diameter of Z is almost surely bounded by D (say), we can generate centres $X_i \in A \oplus B_D$ where B_D is the disc of diameter D, form the associated Z_i and clip $X_i + Z_i$ to A. Similarly, one can reduce to the case where A is convex, or even a disc, by simply enclosing A in a larger, convex region A^+ such as the convex hull or circumcircle of A, generating a simulated realisation of U in A^+, and clipping it to A.

4.2.4 Extrapolation

When extending a germ-grain model beyond the observation window, two cases may be distinguished, namely

(i) extending grains Z_i^* such that $X_i^* + Z_i^*$ hits the boundary of A based on $U \cap A$;

(ii) extending the pattern U beyond the window A based on $U \cap A$.

Below we discuss several geometric aspects in some generality. Specific aspects related to spatial cluster processes will be treated in subsequent sections. For Poisson germ-grain models, the conditional distribution of $\{X_i^* + Z_i^* : i = 1, 2, \ldots\}$ given $\{(X_i^* + Z_i^*) \cap A : i = 1, 2, \ldots\}$ is such that the $X_i^* + Z_i^*$ are conditionally independent, and the conditional distribution of $X_i^* + Z_i^*$ depends only on $(X_i^* + Z_i^*) \cap A$ (Daley and Vere-Jones 1988, Last 1990, Kingman 1993, Reiss 1993). Note that this conditional distribution as well as the law of $(X_i^* + Z_i^*) \cap A$ may have atoms,

as for example if there is a non-zero probability that a single object $X_i^* + Z_i^*$ covers A completely, or, in the conditional case, if a grain is specified fully by its restriction to A. Atoms need to be treated separately using integral-geometric factorisation techniques (Santaló 1976). If the distribution governing X is not that of a Poisson point process, as for spatial clustering problems, the grains can no longer be extended independently of each other. Other obstacles arise from the unobservability of the individual objects in the pattern (cf. section 4.2.2), and we need to extend grains based on the union set $U \cap A$. Sometimes, $U \cap A$ suffices to determine the individual sets $(X_i^* + Z_i^*) \cap A$; more often it will not be possible to determine the components uniquely from U especially if the window A is not convex or if objects may occlude one another. Indeed, the identification of the offspring partitioning is the whole point of spatial clustering. To conclude this section, note that alternative classes of models include the various Poisson-based constructions described in Chilès (1989), chapter XIII in Serra (1982), and Arak-Surgailis-Clifford mosaics and random graphs in Arak et al. (1993). We use germ–grain models mainly because they are quite flexible while remaining relatively simple from a computational point of view: Markov chain Monte Carlo simulation methods are available by combining existing methods for point processes and for Poisson processes of geometric objects, and parametric and nonparametric inferential methods can be carried over from existing methods for spatial point processes. Moreover, in the alternative models listed above, the geometric features may be connected (e.g. several line segments may have a common endpoint) in a fashion which is inappropriate to most of the applications considered here, although positively desirable for other applications such as random tessellations.

4.3 Spatial Cluster Processes

The identification of centres of clustering is of interest in many areas of applications, including archeology (Hodder and Orton 1976), mining (Chilès 1989, Baddeley et al. 1993) and animal territory research (Blackwell 1998). In disease mapping the identification of cluster centres is of interest (Marshall 1991; see also Chapter 14 of this volume) and mine field detection relies on separating clusters of land mines from clutter of other kinds (Dasgupta and Raftery 1998, Cressie and Lawson 2000). Most traditional clustering approaches build a tree based on some similarity measure, for example, Mardia et al. (1979), Chatfield and Collins (1980), Kaufman and Rousseeuw (1990) or other textbooks on multivariate statistics (see also Chapter 1). From this tree, the number of clusters and the corresponding partition are decided in an ad hoc (and mostly subjective) manner. More recently, model based clustering techniques (Dasgupta and Raftery 1998, Deibolt and Robert 1994) consider finite mixture models. The number of groups is determined by a Bayes factors or AIC criterion, and given the number of

mixture components, model parameters are estimated by maximum likelihood, often using a variant of the EM algorithm. Most applications also allow a 'do not know' class for outliers or noise. The cluster centres only play an implicit role – approximated by the centre of gravity, principal axis or other 'mean' of the detected clusters – if they appear at all. Notable exceptions are Lund et al. (1999) and Lund and Thönnes (2000) who model uncertainty in point locations by means of a cluster process consisting of at most a single point, and van Lieshout et al. (2001) who employ variational analysis in the space of intensity measures of the parent point process. In contrast, following up on earlier work (Baddeley et al. 1993, van Lieshout 1995, van Lieshout and Baddeley 1995), this chapter advocates the use of point process and germ–grain models (see Section 4.2.1). A virtue of this approach is that the number of clusters, the locations of their centres, and the grouping or labelling of observed points into clusters, are intrinsic aspects of the underlying process (rather than additional parameters) and are all treated simultaneously. The most general model we consider is the independent cluster process introduced in Section 4.3.1, but most attention will be focussed on the computationally convenient Cox cluster processes (Section 4.3.2). The cluster formation densities are derived in Section 4.3.3 below.

4.3.1 Independent Cluster Processes

Let X be a point process on $\mathbb{R}^d$ and associate with each X_i a finite cluster Z_i of points 'centred' at the origin of $\mathbb{R}^d$. Throughout we will assume that the grains Z_i are conditionally independent. The union of offspring $U = \cup_i (X_i + Z_i)$ is an *independent cluster process* (pp. 236–238 in Daley and Vere–Jones 1988, pp. 75–81 and 148 ff. in Cox and Isham 1980). Technical conditions of finiteness and measurability must be satisfied for such a process to exist, see p. 236 in Daley and Vere-Jones (1988). The data consist of a realisation of $Y = U \cap A$ in a compact window $A \subseteq \mathbb{R}^d$ of positive volume. Thus,

$$\mathbf{y} = \{ y_1, \ldots, y_m \}, \qquad m > 0, \quad \mathbf{y} \subseteq A$$

is a configuration of daughters in A. The above formulation is quite flexible, in that it retains the possibility of locating putative cluster parents outside the window A to counteract sampling bias effects (see the discussion in Section 4.2.3) and of grain characteristics such as the daughter intensity or the spread of the cluster to be randomly and spatially varying. In order to be able to base inference on penalised likelihoods, we shall restrict the germ process to lie inside some compact set $\mathcal{X} \subseteq \mathbb{R}^d$ of positive volume, and assume that its distribution is absolutely continuous with respect to a unit rate Poisson point process on $\mathcal{X}$. For each $\xi \in \mathcal{X}$ we are given the distribution Q_ξ of a finite point process Z_ξ on a compact subset $\tilde{\mathcal{X}}$

of $\mathbb{R}^d$; Z_ξ represents the offspring of a parent ξ translated back to the origin to fit in the general germ-grain model of Section 4.2.1. We assume that Q_ξ is absolutely continuous with a density $g(\cdot|\xi)$ with respect to the distribution of a unit rate Poisson process on $\tilde{\mathcal{X}}$. Thus $\mathcal{Z} = \mathcal{N} = \mathcal{N}_{\tilde{\mathcal{X}}}$, the family of finite point configurations in $\tilde{\mathcal{X}}$. To ensure existence of U, we shall assume that the family of densities is jointly measurable seen as a function $g : \mathcal{X} \times \mathcal{N} \to \mathbb{R}^+$. More generally, we could have set $\tilde{\mathcal{X}} = \mathbb{R}^d$ equipped with some finite diffuse intensity measure $\nu(\cdot)$, with the assumption that Q_ξ is absolutely continuous with respect to the distribution of a Poisson process with intensity measure $\nu(\cdot)$. It is of interest to note that when X is a Poisson process and we extend the process onto the whole of $\mathbb{R}^d$, then Q may be almost surely reconstructed from a single joint observation of parents and daughters (Milne 1970, Bendrath 1974). Table 4.1 summarises standard nomenclature for special cases of the independent cluster model (Stoyan et al. 1995, Daley and Vere-Jones 1988).

Parents X	Clusters Z	Name of process U
general	general	Independent cluster process
Poisson	general	Poisson cluster process
general	Poisson	Cox cluster process
Poisson	Poisson	Neyman-Scott process
Poisson (homogeneous)	Poisson (uniform in disc)	Matérn cluster process
Poisson (homogeneous)	Poisson (Gaussian)	Modified Thomas process

Table 4.1 *Standard nomenclature for independent cluster processes.*

4.3.2 Cox Cluster Processes

For simplicity, most attention will be focussed on the Cox cluster process model, where each grain Z_ξ, $\xi \in \mathcal{X}$, is a realisation of an *inhomogeneous Poisson point process* on $\tilde{\mathcal{X}}$ with *intensity function* $h(\cdot + \xi|\xi) : \tilde{\mathcal{X}} \to [0, \infty)$. In other words, a parent point ξ is replaced by a Poisson number of offspring with mean $H(\xi) = \int_{\tilde{\mathcal{X}}} h(t + \xi|\xi)\, dt \in (0, \infty)$, and given the number of offspring their locations are independently and identically distributed with probability density $f(\cdot) = h(\cdot|\xi)/H(\xi)$ on $\xi + \tilde{\mathcal{X}}$ (with respect to Lebesgue measure). We shall assume the intensity function $h(\cdot|\cdot)$ to be jointly measurable in its arguments, as well as integrable so that $H(\xi) < \infty$ for all $\xi \in \mathcal{X}$. As in Dasgupta and Raftery (1998), van Lieshout (1995) and van Lieshout and Baddeley (1995), scatter noise and outliers – also known as *orphans* – are modelled by a Poisson point process of constant intensity $\epsilon > 0$ independently of all Z_ξ. This fits into the germ–grain framework of Section 4.2.1 by introducing an extra dummy or 'ghost' parent x_0. We shall write $h(\cdot|x_0) \equiv \epsilon$, and denote its integral over $\mathcal{X} \oplus \tilde{\mathcal{X}}$ by

$H(x_0)$. By the superposition property of Poisson processes, conditional on $X = \mathbf{x} = \{x_1, \ldots, x_n\}$, the combined offspring form a Poisson point process on $\mathcal{X} \oplus \tilde{\mathcal{X}}$ with intensity function

$$\lambda(\cdot \mid \mathbf{x}) = \epsilon + \sum_{i=1}^{n} h(\cdot | x_i) \tag{4.4}$$

with the convention that $h(t|x_i) = 0$ if $t \notin x_i + \tilde{\mathcal{X}}$, $i = 1, \ldots, n$. The marginal distribution of U is that of a Cox point process Stoyan et al. (1995). Often, the intensity function $h(t|\xi)$ will depend only on the distance $d(\xi, t)$ between ξ and t. An example is

$$h(t|\xi) = \begin{cases} \mu & \text{if } d(\xi, t) \leq R_h \\ 0 & \text{otherwise} \end{cases} \tag{4.5}$$

which, if X is also a Poisson process, is known as the *Matérn cluster process* (Matérn 1986). Another interesting special case is (for $d = 2$, say)

$$h(t|\xi) = \frac{\mu}{2\pi\sigma^2} e^{-d(\xi,t)^2/2\sigma^2}. \tag{4.6}$$

According to (4.6), the daughters follow an isotropic Gaussian distribution with centre ξ. Again if X is a Poisson process, the distribution of U is called the *modified Thomas process*. More generally, the spread σ may depend on ξ. For further details, consult Diggle (1983), Daley and Vere-Jones (1988), or Stoyan et al. (1995). For a Cox cluster process, conditional on $X = \mathbf{x} = \{x_1, \ldots, x_n\}$ and the number m of offspring, the points are drawn independently from a finite mixture distribution (Hand 1981, Titterington et al. 1985) with $n + 1$ component distributions determined by the x_i and weights

$$p_i = \frac{H(x_i)}{\sum_{i=0}^{n} H(x_i)}, \qquad i = 0, \ldots, n.$$

If the intensity function h is translation invariant in the sense that $h(t + \xi|\xi) = h(t|0)$ for all $\xi \in \mathcal{X}$ – a common assumption in our spatial context – the weights are identical for all parents except the ghost, a rather unnatural restriction in the finite mixture context. Furthermore, the connection with mixture distributions is lost when the clusters are not Poisson processes. To conclude this section, note that some parents may be childless. In particular, if the clusters Z_ξ are Poisson processes, they have a positive probability of being empty. If in a particular application there is no interest in such parents, one could condition each Z_ξ on $\{Z_\xi \neq \emptyset\}$, or consider only those parents having at least one daughter.

4.3.3 Cluster Formation Densities

In order to be able to draw inference about parents and cluster membership, we need the (posterior) distribution of $W = \{(x_0, Z_0), \ldots, (X_n, Z_n)\}$, i.e.

of parents X_i marked by their associated grain Z_i, $i = 0, \ldots, n$. We will take a Bayesian approach based on

$$\begin{aligned} p_{W|U}(\{(x_i, \mathbf{z}_i)\}_{i \le n} \mid \mathbf{u}) &\propto P(\mathbf{z}_0, \ldots, \mathbf{z}_n \mid x_0, \ldots, x_n, \mathbf{u})\, p_{X|U}(\mathbf{x} \mid \mathbf{u}) \\ &= c(\mathbf{u})\, P(\mathbf{z}_0, \ldots, \mathbf{z}_n \mid x_0, \ldots, x_n, \mathbf{u})\, p_{U|X}(\mathbf{u} \mid \mathbf{x})\, p_X(\mathbf{x}), \end{aligned} \tag{4.7}$$

the posterior density of W with respect to a unit rate Poisson process on $\mathcal{X}$ marked at $\xi \in \mathcal{X}$ by a label in $\mathcal{P}(\mathbf{u} - \xi)$ according to the uniform distribution on the power set of $\mathbf{u}$ translated back to the origin. The term $p_X(\mathbf{x})$ is the prior density for X with respect to the distribution of a unit rate Poisson process on $\mathcal{X}$, and $c(\mathbf{u})$ a normalising constant depending on the 'data' $\mathbf{u}$. If only the cluster centres are of interest, the posterior density of X (with respect to the distribution of a unit rate Poisson process on $\mathcal{X}$) may be used instead:

$$p_{X|U}(\mathbf{x} \mid \mathbf{u}) = c'(\mathbf{u})\, p_{U|X}(\mathbf{u} \mid \mathbf{x})\, p_X(\mathbf{x}). \tag{4.8}$$

We will discuss the choice of prior later on and here describe only the 'forward terms' of cluster formation. Firstly, recall from Section 4.3.1 that conditional on $X = \mathbf{x} = \{x_1, \ldots, x_n\}$, the grains $Z_1, \ldots, Z_n$ associated with $x_1, \ldots, x_n$ respectively are independent with distributions that are absolutely continuous with respect to a unit rate Poisson process on $\tilde{\mathcal{X}}$. Thus, the conditional joint density of $(Z_1, \ldots, Z_n)$ equals

$$\prod_{i=1}^{n} g(\mathbf{z}_i | x_i)$$

with respect to the n-fold product measure of unit rate Poisson processes on $\tilde{\mathcal{X}}$. The orphans Z_0 are modelled as a Poisson process of rate $\epsilon > 0$. Again conditioning on $X = \mathbf{x}$, the superposition U is absolutely continuous with respect to the distribution of a unit rate Poisson process on $\mathcal{X} \oplus \tilde{\mathcal{X}}$. Its density at $\mathbf{u} = \{u_1, \ldots, u_m\}$ can be found by summing over all partitions in sibling clusters

$$p_{U|X}(\mathbf{u} \mid \mathbf{x}) = e^{(1-\epsilon)|\mathcal{X} \oplus \tilde{\mathcal{X}}| - n|\tilde{\mathcal{X}}|} \times \tag{4.9}$$

$$\sum_{\varphi} \epsilon^{n(\mathbf{u}_{\varphi^{-1}(\{0\})})} \prod_{i=1}^{n} g(\mathbf{u}_{\varphi^{-1}(\{i\})} - x_i | x_i) \mathbf{1}\{\mathbf{u}_{\varphi^{-1}(\{i\})} - x_i \subseteq \tilde{\mathcal{X}}\}$$

where the sum ranges over all possible offspring-to-parent assignment functions $\varphi : \{1, \ldots, m\} \to \{0, \ldots, n\}$, $\mathbf{u}_{\varphi^{-1}(\{i\})} = \{u_j : \varphi(j) = i\}$ consists of those u_j ascribed to parent x_i by φ, and $n(\cdot)$ denotes cardinality. Equation (4.9) is most readily derived using Janossy densities (page 122 Daley and Vere-Jones 1988). The details can be found in lemma 23 of van Lieshout (1995). Note that (4.9) can be expressed as

$$e^{(1-n-\epsilon)|\mathcal{X} \oplus \tilde{\mathcal{X}}|} \sum_{\varphi} \epsilon^{n(\mathbf{u}_{\varphi^{-1}(\{0\})})} \prod_{i=1}^{n} g'(\mathbf{u}_{\varphi^{-1}(\{i\})} - x_i | x_i)$$

where $g'(\cdot - x_i|x_i) = e^{|\mathcal{X}\oplus\tilde{\mathcal{X}}|-|\tilde{\mathcal{X}}|}g(\cdot - x_i|x_i)\mathbf{1}\{\cdot - x_i \subseteq \tilde{\mathcal{X}}\}$ is a density of the translated typical grain with respect to a unit rate Poisson process on $\mathcal{X} \oplus \tilde{\mathcal{X}}$. Next, consider the conditional distribution of the complete model given the cluster centres $x_1, \ldots, x_n$. Since we already derived the conditional joint density of $(Z_1, \ldots, Z_n)$, an identification

$$(\mathcal{Z}^n, \pi_{\mathcal{Z}}^n, \mathcal{A}) \leftrightarrow (\mathcal{N}_{\mathcal{X}\times\mathcal{Z}}, \xi_{\mathbf{x}}, \mathcal{B})$$

of grain vectors $(\mathbf{z}_1, \ldots, \mathbf{z}_n) \in \mathcal{Z}^n$ with the marked point configuration $\{(x_1, \mathbf{z}_1), \ldots, (x_n, \mathbf{z}_n)\} \in \mathcal{N}_{\mathcal{X}\times\mathcal{Z}}$ is needed. Here $\mathcal{Z} = \mathcal{N}_{\tilde{\mathcal{X}}}$ is the grain space (cf. Section 4.2.1) consisting of all finite point configurations, $\pi_{\mathcal{Z}}^n$ is the n-fold product measure of unit rate Poisson processes on $\tilde{\mathcal{X}}$, $\mathcal{A}$ is the usual Borel product σ-algebra of the weak topology (Daley and Vere-Jones 1988), and $\mathcal{B}$ the Borel σ-algebra of the weak topology on marked point patterns. To do so, define a measurable bijection $i_{\mathbf{x}}$ (in the sense that the complement of the range of $i_{\mathbf{x}}$ has measure zero under $\xi_{\mathbf{x}}$) depending on the parent pattern $\mathbf{x} = \{x_1, \ldots, x_n\}$ by

$$i_{\mathbf{x}} : (\mathbf{z}_1, \ldots, \mathbf{z}_n) \mapsto \{(x_1, \mathbf{z}_1), \ldots (x_n, \mathbf{z}_n)\}.$$

Using the identification thus defined, the measure $\xi_{\mathbf{x}}$ is given by $\xi_{\mathbf{x}}(B) = \pi_{\mathcal{Z}}^n(i_{\mathbf{x}}^{-1}(B))$ for all $B \in \mathcal{B}$. Finally, the conditional distribution of W or equivalently the marks Z_i given (X, U) is discrete, with probabilities

$$P(\mathbf{z}_0, \ldots \mathbf{z}_n \mid x_0, \ldots, x_n, \mathbf{u}) = \frac{\epsilon^{n(\mathbf{z}_0)} \prod_{i=1}^n g(\mathbf{z}_i|x_i)}{\sum_\varphi \epsilon^{n(\mathbf{u}_{\varphi^{-1}(\{0\})})} \prod_{i=1}^n g(\mathbf{u}_{\varphi^{-1}(\{i\})} - x_i|x_i)\mathbf{1}\{\mathbf{u}_{\varphi^{-1}(\{i\})} - x_i \subseteq \tilde{\mathcal{X}}\}} \tag{4.10}$$

provided the union $\cup_i(x_i + \mathbf{z}_i)$ equals $\mathbf{u}$. If $g(\cdot - \xi|\xi)$ is hereditary (cf. Section 4.4.1) for each $\xi \in \mathcal{X}$, the sum in the denominator of (4.10) over all functions φ ascribing parents to each offspring, is non-zero. Otherwise, we have to impose the condition that the grain partition and (X, U) are compatible, in the sense that there exists at least one φ for which the term $\epsilon^{n(\mathbf{u}_{\varphi^{-1}(\{0\})})} \prod_{i=1}^n g(\mathbf{u}_{\varphi^{-1}(\{i\})} - x_i|x_i)\mathbf{1}\{\mathbf{u}_{\varphi^{-1}(\{i\})} - x_i \subseteq \tilde{\mathcal{X}}\}$ is strictly positive (Van Lieshout 1995, theorem 29). For Cox cluster processes, the formulae (4.7)–(4.10) can be greatly simplified. Since Z_ξ has density

$$g(\mathbf{z}|\xi) = \exp\left[\int_{\tilde{\mathcal{X}}} (1 - h(t + \xi|\xi))\, dt\right] \prod_{z\in\mathbf{z}} h(z + \xi|\xi)$$

with respect to a unit rate Poisson process on $\tilde{\mathcal{X}}$, (4.9) reduces to

$$\begin{aligned} p_{U|X}(\mathbf{u} \mid \mathbf{x}) &= \exp\left[\int_{\mathcal{X}\oplus\tilde{\mathcal{X}}} (1 - \lambda(t \mid \mathbf{x}))\, dt\right] \sum_\varphi \prod_{i=0}^n \prod_{t\in\mathbf{u}_{\varphi^{-1}(\{i\})}} h(t|x_i) \\ &= \exp\left[\int_{\mathcal{X}\oplus\tilde{\mathcal{X}}} (1 - \lambda(t \mid \mathbf{x}))\, dt\right] \prod_{j=1}^m \lambda(u_j \mid \mathbf{x}) \end{aligned} \tag{4.11}$$

coding $h(\cdot|x_0) \equiv \epsilon$ for the dummy parent x_0. Thus, (4.11) is in accordance with the fact that the independent superposition of Poisson processes is again a Poisson process, here with intensity $\lambda(\cdot \mid \mathbf{x})$ (cf. (4.4) and the discussion in Section 4.3.2). As for the offspring labelling, (4.10) for a Cox cluster process equals

$$P(\mathbf{z}_0, \ldots, \mathbf{z}_n \mid x_0, \ldots, x_n, \mathbf{u}) = \frac{\prod_{i=0}^{n} \prod_{t \in \mathbf{z}_i} h(t + x_i | x_i)}{\prod_{j=1}^{m} \lambda(u_j \mid \mathbf{x})} \tag{4.12}$$

whenever $\cup_i (x_i + \mathbf{z}_i) = \mathbf{u}$, see van Lieshout and Baddeley (1995) or corollary 30 in van Lieshout (1995). In terms of the label allocation function $\varphi : \{1, \ldots, m\} \mapsto \{0, 1, \ldots, n\}$ allocating each daughter point to its parent, equation (4.12) implies that the daughters are ascribed to a cluster centre x_I *independently* of one another, with probabilities

$$P(\varphi(j) = I) = \frac{h(u_j | x_I)}{\lambda(u_j \mid \mathbf{x})}.$$

The analogue of this result for finite mixtures with m and n fixed was called the *Random Imputation Principle* by Deibolt and Robert (1994). It was taken as an assumption by Binder (1978) (see page 32). Note the statement holds only for Cox cluster processes, i.e. when the clusters are Poisson.

4.4 Bayesian Cluster Analysis

From Section 4.2.2, recall that the prime object of spatial cluster analysis is to evaluate conditional expectations of quantities such as the number of clusters and the mean number of points per cluster based on the posterior distribution (4.7) of the complete data W given $\mathbf{y}$. In the previous section, we derived the densities associated with cluster formation. In Section 4.4.1 below, we discuss the prior, and investigate properties of the posterior distribution in Section 4.4.2. Then we turn to the problems of generating realisations of (4.7) by Markov chain Monte Carlo methods, and of estimating the model parameters (Sections 4.4.3–4.4.4). In Section 4.4.5, we propose an adaptive coupling from the past algorithm that yields exact samples from (4.7). Throughout, the redwood data set (Strauss 1975, Ripley 1977) is used as an illustration.

4.4.1 Markov Point Processes

In this section we focus on the prior term $p_X(\mathbf{x})$ in (4.7), which we shall assume to be the density of a Markov point process (Ruelle 1969, Preston 1976, Ripley 1977, Ripley 1988, Baddeley and Møller 1989, van Lieshout 2000). Following is a brief summary of the facts we need. Let X be a point process on a compact subset $\mathcal{X} \subseteq \mathbb{R}^d$ of positive volume, whose distribution

is absolutely continuous with respect to a unit rate Poisson process on $\mathcal{X}$, say with density $p_X(\cdot)$. Then X is *Markov* at range R in the sense of Ripley (1977) if the ratio

$$\lambda_X(\xi; \mathbf{x}) = \frac{p_X(\mathbf{x} \cup \{\xi\})}{p_X(\mathbf{x})} \tag{4.13}$$

is well-defined for all $\xi \in \mathcal{X}$ (i.e. $p_X(\mathbf{x} \cup \{\xi\}) = 0$ implies $p_X(\mathbf{x}) = 0$; in this case we will also say that $p_X(\cdot)$ is *hereditary*) and depends only on those $x_i \in \mathbf{x}$ for which $d(x_i, \xi) \leq R$. More generally, the fixed range dependence may be replaced by an arbitrary symmetric neighbourhood relation $\sim$ (so that (4.13) depends on $x_i \sim \xi$ only). Even more general Markov point processes are considered by Baddeley and Møller (1989), and the Markovianity of spatial cluster processes is studied in Baddeley et al. (1996). A (Markov) point process defined by its density with respect to a unit rate Poisson process is said to be *locally stable* if its conditional intensity (4.13) is well-defined and uniformly bounded in both its arguments. To model patterns in which the points tend to avoid coming too close together, it is convenient to consider *pairwise-interaction* processes with densities of the form

$$p_X(\mathbf{x}) = \alpha \prod_{x \in \mathbf{x}} \beta(x) \prod_{x \sim x' \in \mathbf{x}} \gamma(x, x') \tag{4.14}$$

where $\beta : \mathcal{X} \to [0, \infty)$ (the 'intrinsic activity') and $\gamma : \mathcal{X} \times \mathcal{X} \to [0, \infty)$ (the 'pairwise interaction') are measurable functions, γ is symmetric, and $\alpha > 0$ is the normalising constant. This model is well-defined (i.e. the density is integrable) at least whenever $\beta(\cdot)$ is uniformly bounded and $\gamma(\cdot, \cdot) \leq 1$. A standard example of (4.14) is the *Strauss process* (Strauss 1975) with $\beta(\cdot) \equiv \beta > 0$ and

$$\gamma(x, x') = \begin{cases} \gamma & \text{if } d(x, x') \leq R \\ 1 & \text{otherwise} \end{cases} \tag{4.15}$$

where $0 \leq \gamma \leq 1$, which has density

$$p_X(\mathbf{x}) = \alpha \beta^{n(\mathbf{x})} \gamma^{s(\mathbf{x})}$$

where $n(\mathbf{x})$ is the number of points in $\mathbf{x}$ and $s(\mathbf{x})$ is the number of pairs x, x' with $d(x, x') \leq R$. The model favours realisations $\mathbf{x}$ that tend to have more points at distances larger than R than under the Poisson model, that is there is *repulsion* between the points. The special case $\gamma = 0$ in which no R-close point pairs are permitted is known as the *hard core process*; $\gamma = 1$ corresponds to a Poisson process with intensity β. More formally, a point process density $p_X(\cdot)$ is called *anti-monotone* (or repulsive) if

$$\lambda_X(\xi; \mathbf{x}') \leq \lambda_X(\xi; \mathbf{x})$$

for all ξ whenever $\mathbf{x} \subseteq \mathbf{x}'$ and *monotone* (or attractive) if its conditional intensity satisfies

$$\lambda_X(\xi; \mathbf{x}') \geq \lambda_X(\xi; \mathbf{x}).$$

The reader may verify that the Strauss process is repulsive for all $\gamma \leq 1$.

4.4.2 Sampling Bias for Independent Cluster Processes

Note that the restriction Y of an independent cluster process U to some compact observation window A is itself an independent cluster process. Indeed,

$$Y = U \cap A = \bigcup_{x \in X} (x + Z_x) \cap A = \bigcup_{x \in X} (x + (Z_x \cap (A - x))).$$

The distribution $Q_{\xi,A}$ of the grain $Z_\xi \cap (A - \xi)$ associated with ξ in the A-clipped process has density

$$g_A(\mathbf{z}|\xi) = \sum_{k=0}^{\infty} \frac{1}{k!} \int \cdots \int_{(\tilde{\mathcal{X}} \setminus (A-\xi))^k} g(\mathbf{z} \cup \{ v_1, \ldots, v_k \} | \xi) \, dv_1 \ldots dv_k \quad (4.16)$$

with respect to a unit rate Poisson process on $\tilde{\mathcal{X}}$. It follows that the posterior distribution of X given Y is analogous to (4.8), except for the fact that $g_A(\cdot|\cdot)$ features instead of $g(\cdot|\cdot)$. As before, a ghost parent is added to account for scatter noise. For Cox cluster processes, (4.16) simplifies to

$$g_A(\mathbf{z}|\xi) = \exp\left[\int_{\tilde{X}} (1 - h(t + \xi|\xi) \, \mathbf{1}\{t \in A - \xi\}) \, dt \right] \prod_{z \in \mathbf{z}} h(z + \xi|\xi)$$

for $\mathbf{z} \subseteq A - \xi$, the density of a Poisson point process with intensity function $h(\cdot + \xi|\xi) \, \mathbf{1}\{\cdot \in A - \xi\}$. Hence, conditionally on $X = \mathbf{x} = \{x_1, \ldots, x_n\}$, Y is an inhomogeneous Poisson process on A with intensity function

$$\lambda(a \mid \mathbf{x}) = \epsilon + \sum_{i=1}^{n} h(a|x_i), \qquad a \in A,$$

where $\epsilon > 0$ is the background clutter term (cf. Section 4.3.2). As for the prior, one could simply assume the parents to be distributed as a Poisson point process, but it seems more natural to incorporate repulsion at short range to avoid 'over fitting' in the sense of many close parents. Thus, one might take as prior for example a hard core process (cf. Section 4.4.1) with density

$$p_X(\mathbf{x}) = \begin{cases} \alpha \beta^{n(\mathbf{x})} & \text{if } d(x_i, x_j) > R \text{ for all pairs} \\ 0 & \text{otherwise} \end{cases} \quad (4.17)$$

with respect to a unit rate Poisson process on $\mathcal{X}$. Upon observing $Y = \mathbf{y} = \{y_1, \ldots, y_m\}$, the analogue of (4.8) for the A-clipped process is

$$p_{X|Y}(\mathbf{x} \mid \mathbf{y}) = c(\mathbf{y}) \, p_X(\mathbf{x}) \, \exp\left[\int_A (1 - \lambda(a \mid \mathbf{x})) \, da \right] \prod_{j=1}^{m} \lambda(y_j \mid \mathbf{x}) \quad (4.18)$$

which has posterior conditional intensity

$$\lambda_{X|Y}(\xi; \mathbf{x} \mid \mathbf{y}) = \lambda_X(\xi; \mathbf{x}) \exp\left[-\int_A h(a|\xi)\, da\right] \prod_{j=1}^{m} \left[1 + \frac{h(y_j|\xi)}{\lambda(y_j \mid \mathbf{x})}\right]. \tag{4.19}$$

If the prior density $p_X(\cdot)$ is that of a repulsive Markov point process, the posterior distribution specified by (4.18) is hereditary and repulsive too. The posterior range of interaction depends on the support of the family $h(\cdot|\xi)$, $\xi \in \mathcal{X}$, of intensity functions. If furthermore the prior density $p_X(\cdot)$ is locally stable with bound λ for its conditional intensity and $h(\cdot|\cdot)$ is uniformly bounded in both its arguments by H, then $\lambda_{X|Y}(\xi; \mathbf{x} \mid \mathbf{y}) \leq \lambda(1 + H/\epsilon)^m$ implying local stability of (4.18).

4.4.3 Spatial Birth-and-Death Processes

In this section, we address the problem of sampling from the posterior distribution of the complete data W given partial observations $Y = U \cap A = \{y_1, \ldots, y_m\}$ of a Cox cluster process U within some compact observation window A. Note that since the offspring allocation labels are discrete and distributed according to (4.12), and by the Poisson assumption any daughters in A^c are conditionally independent of those in A, the problem reduces to sampling from the conditional distribution (4.18) of X given Y. Since direct sampling does not seem feasible, we apply Markov chain Monte Carlo techniques. Perhaps the oldest such technique is based on *spatial birth-and-death processes* (Preston 1977), continuous time Markov processes whose transitions are either births or deaths. The traditional choice (Ripley 1977, Baddeley and Møller 1989, Møller 1989) is to take a birth rate proportional to the posterior conditional intensity and a constant death rate. Under mild non-explosion conditions, this procedure converges to the target distribution and hence yields approximate samples if run for long enough (Preston 1977, Møller 1989). A disadvantage is that the total birth rate is difficult to compute, and the product over data points in (4.19) may be very large. For these reasons, we prefer to work with the alternative birth rate

$$b(\mathbf{x}, \xi) = \lambda_X(\xi; \mathbf{x}) \left[1 + \sum_{j=1}^{m} \frac{h(y_j|\xi)}{\epsilon}\right] \tag{4.20}$$

which is less peaked than the posterior conditional intensity, while retaining the desirable property of placing most new-born points in the vicinity of points of $\mathbf{y}$. In order to satisfy the detailed balance equations

$$p_{X|Y}(\mathbf{x} \mid \mathbf{y})\, b(\mathbf{x}, \xi) = p_{X|Y}(\mathbf{x} \cup \{\xi\} \mid \mathbf{y})\, d(\mathbf{x} \cup \{\xi\}, \xi),$$

the death rate for deleting ξ from configuration $\mathbf{x} \cup \{\xi\}$ is

$$d(\mathbf{x} \cup \{\xi\}, \xi) = \frac{\exp\left[\int_A h(a|\xi)\, da\right]}{\prod_{j=1}^{m}\left[1 + \frac{h(y_j|\xi)}{\lambda(y_j|\mathbf{x})}\right]} \left[1 + \sum_{j=1}^{m} \frac{h(y_j \mid \xi)}{\epsilon}\right]. \tag{4.21}$$

Note that for any locally stable prior distribution for which $\lambda_X(\xi; \mathbf{x}) \leq \lambda$ uniformly in $\mathbf{x}$ and ξ, and any $h(\cdot|\cdot)$ that is uniformly bounded in both its arguments by H, the total birth rate

$$B(\mathbf{x}) = \int_{\mathcal{X}} b(\mathbf{x}, \xi)\, d\xi \leq \lambda \left[|\mathcal{X}| + \frac{1}{\epsilon} \sum_{j=1}^{m} \int_{\mathcal{X}} h(y_j|\xi)\, d\xi\right] := B$$

is bounded from above by a constant $B \leq \lambda|\mathcal{X}|(1 + mH/\epsilon)$ that is easy to evaluate for typical choices of $h(\cdot|\cdot)$ such as (4.5) or (4.6). The total death rate from parent configuration $\mathbf{x}$ satisfies

$$D(\mathbf{x}) = \sum_i d(\mathbf{x}, x_i) \geq n(\mathbf{x})(1 + H/\epsilon)^{-m}.$$

Hence, by the Preston theorem (Preston 1977) (see e.g. Baddeley and Møller 1989, Møller 1989), there exists a unique spatial birth-and-death process with transition rates given by (4.20) and (4.21). It has unique equilibrium distribution $p_{X|Y}(\cdot \mid \mathbf{y})$, to which it converges in distribution from any initial state. From an algorithmic point of view, if the current state is $\mathbf{x}$, after an exponentially distributed sojourn time of rate $B + D(\mathbf{x})$, with probability $D(\mathbf{x})/(B + D(\mathbf{x}))$ a point of $\mathbf{x}$ is deleted according to the distribution $d(\mathbf{x}, x_i)/D(\mathbf{x})$; a birth is proposed with the complementary probability $B/(B + D(\mathbf{x}))$ by sampling a candidate ξ from the mixture density $\frac{\lambda}{B}\left[1 + \sum_{j=1}^{m} \frac{h(y_j|\xi)}{\epsilon}\right]$, which is then accepted with probability $\lambda_X(\xi; \mathbf{x})/\lambda$.

Example: Redwood Seedlings

Figure 4.1 shows the locations of 62 redwood seedlings in a square of side approximately 23 m. The data were extracted by (Ripley 1977) from a larger data set in Strauss (1975). The K-function (Ripley 1979, Ripley 1981) for these data is given in Ripley (1977) and suggests aggregation. As noted by Strauss this is caused by the presence of stumps known to exist in the plot, but whose position has not been recorded. Previous analyses of this data set include that of Strauss (1975), who fitted a model later criticised by Kelly and Ripley (1976). Ripley (1977) concluded we should reject the Poisson hypothesis and remarked that there appears to be both clustering and inhibition between clusters. Diggle (1983) fitted a Poisson cluster process of Thomas type and reported least squares estimates $(25.6, 0.042)$ for the parent intensity and the standard deviation of daughter–parent distances. A goodness of fit test showed adequate fit, but, from a biological

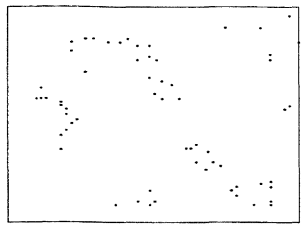

Figure 4.1 *Positions of 62 redwood seedlings in a unit square*

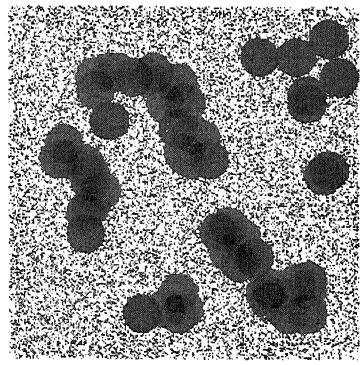

Figure 4.2 *Empirical log posterior parent intensity surface based on a Cox–Matérn cluster process with* $R_h = 0.061$ *and on average* 2.14 *points per cluster, noise intensity* $\epsilon = 10.0$ *and a hard core prior with* $R = 0.03$ *and* $\beta = 1.0$ *by spatial birth-and-death over* 2.0×10^4 *time units. Black corresponds to high values, white to small ones*

point of view, a mean number of 26 stumps seems implausible. In Diggle (1978), a Poisson cluster process of Matérn type was fitted with similar results (radius 0.061 and 29 clusters). None of the above have looked at cluster centre location. This was first studied in Baddeley and van Lieshout (1993) and by Lawson (1993a) who fitted a Poisson–Thomas cluster process and reported 16 parents. An approach based on variational methods can be found in Van Lieshout et al. (2001). In earlier work (Baddeley and van Lieshout 1993, van Lieshout 1995, van Lieshout and Baddeley 1995), we analysed the redwood data using a modified Thomas displacement function (4.6) and a Strauss prior (4.15) with interaction distance 0.084 (Diggle 1983) and $\log\beta = \log\gamma = -10$. Simulation was based on a constant death rate spatial birth-and-death process. Initialising with parameter values $\mu = 7$, $\sigma = 0.042$ and an empty list of cluster centres, we ran the birth-and-death process for 2 time units and found maximum likelihood estimates $\mu = 6.5$ and $\sigma = 0.05$. Here, we use the spatial birth-and-death process with rates (4.20)–(4.21) to sample from the posterior distribution of cluster centres for a Cox model with a Matérn style intensity function given by (4.5) with $R_h = 0.061$ and $\mu = 2.14/(\pi R_h^2)$ as in Diggle (1978), orphan intensity $\epsilon = 10.0$, and a hard core prior with $R = 0.03$ and $\beta = 1.0$. The posterior intensity surface of parents in $\mathcal{X} = [-R_h, 1.0 + R_h]^2$ over 2.0×10^4 time units after a burn-in period of 200.0 units with empty initial state is

plotted in Figure 4.2; for the posterior histogram of the number of cluster see Figure 4.3 (*right*). To indicate the effect of the choice of parameter the posterior histogram for $\beta = 0.052$ and an average cluster size of 4.3 shown in Figure 4.3 (*left*). It can be seen that the latter choice shifts th posterior histogram towards fewer cluster centres.

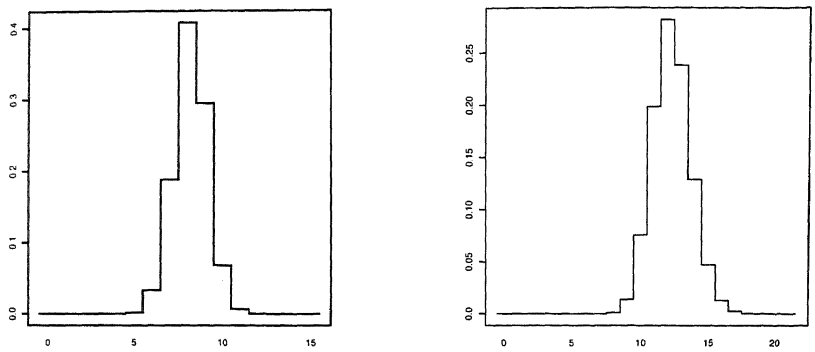

Figure 4.3 *(Left) Posterior histogram for the number of parents given the data of Figure 4.1 based on a Cox-Matérn cluster process with* $R_h = 0.061$ *and on average* 4.3 *points per cluster, noise intensity* $\epsilon = 10.0$ *and a hard core prior with* $R = 0.03$ *and* $\beta = \exp(4.3 + 2.0\log(\epsilon/(\epsilon + \mu))$ *by spatial birth-and-death over* 2.0×10^5 *time units and (right) for on average* 2.14 *points and* $\beta = 1.0$ *over* 2.0×10^4 *time units as in Figure 4.2.*

4.4.4 Parameter Estimation

In general, the independent cluster model $g(\cdot|\cdot)$ will contain parameters θ that must be estimated. For the Cox cluster model, the parameters are the clutter intensity ϵ as well as parameters of the displacement function $h(\cdot|\cdot)$ specifying the shape, the spread and the number of daughters in each cluster. Moreover, the prior model $p_X(\cdot)$ also contains parameters, but since these are merely used as regularisation to avoid over fitting, we will treat these as fixed. We shall use the Monte Carlo maximum likelihood method for missing data models (Gelfand and Carlin 1993, Geyer 1999). In the context of detecting the centres in an independent cluster process, the observed data consists of a point pattern $\mathbf{y}$, the combined offspring in the window A. The missing data are both the parents and their associated

grains. In terms of the cluster formation density derived in Section 4.3.3, the log likelihood ratio with respect to a fixed reference parameter θ_0 can be written as

$$l(\theta) = \log E_{\theta_0}\left[\prod_{i=0}^{n(X)} \frac{g_\theta(Z_i|X_i)}{g_{\theta_0}(Z_i|X_i)} \middle| Y = \mathbf{y}\right] \tag{4.22}$$

by importance sampling theory. The Monte Carlo analogue $l_k(\theta)$ of (4.22) is obtained by replacing the expectation by the average in a sample $W_1, \ldots, W_k$ from the complete model under the conditional distribution with parameter θ_0. Differentiating with respect to θ, the parameter of interest, we obtain

$$\nabla l_k(\theta) = \frac{1}{k}\sum_{j=1}^{k} w_{j,\theta_0,\theta} \frac{\nabla\left[\prod_{(X_i,Z_i)\in W_j} g_\theta(Z_i|X_i)\right]}{\prod_{(X_i,Z_i)\in W_j} g_\theta(Z_i|X_i)} \tag{4.23}$$

where

$$w_{j,\theta_0,\theta} = \left(\prod_{(X_i,Z_i)\in W_j} \frac{g_\theta(Z_i|X_i)}{g_{\theta_0}(Z_i|X_i)}\right) \Big/ \left(\frac{1}{k}\sum_{j=1}^{k} \prod_{(X_i,Z_i)\in W_j} \frac{g_\theta(Z_i|X_i)}{g_{\theta_0}(Z_i|X_i)}\right)$$

are the importance weights. The well-known EM algorithm (Dempster et al. 1977b) is an iterative procedure based on (4.23) that consists of two steps: the E-step computes the conditional log likelihood given the data and current estimates of the parameters, the M-step maximises the result with respect to the parameter. Thus, the importance weights reduce to 1, but resampling is needed at each step. For a critical evaluation of these and other parameter estimation methods, the reader is referred to Geyer (1999); see also Diggle et al. (1994), Geyer and Møller (1994) and Huang and Ogata (2001). For Cox cluster processes, (4.22) simplifies to

$$l(\theta) = \log E_{\theta_0}\left[e^{\sum_{i=0}^{n(X)}(H_{\theta_0}(X_i)-H_\theta(X_i))} \prod_{i=0}^{n(X)} \prod_{t\in X_i+Z_i} \frac{h_\theta(t|X_i)}{h_{\theta_0}(t|X_i)} \middle| Y = \mathbf{y}\right]$$

hence the Monte Carlo score vector (4.23) is

$$\nabla l_k(\theta) = \frac{1}{k}\sum_{j=1}^{k}\left\{w_{j,\theta_0,\theta} \sum_{(X_i,Z_i)\in W_j} [-\nabla H_\theta(X_i) + \sum_{t\in X_i+Z_i} \nabla \log h_\theta(t|X_i)]\right\}$$

where as before $\{W_j\}$ is a sample of size k from the conditional distribution of the complete data given $\mathbf{y}$ under the reference parameter θ_0, and $w_{j,\theta_0,\theta}$ are the importance weights.

Example: Cox–Matérn Cluster Process

Consider the Cox–Matérn cluster process on $\mathbb{R}^2$ with offspring governed by (4.5) and independent Poisson background clutter. Treating the range R_h as fixed, the parameter vector is $\theta = (\epsilon, \mu)$. The grain is a finite point process on $\tilde{\mathcal{X}} = B(0, R_h)$, and $H(\xi) = \mu\pi R_h^2$ for each genuine parent $\xi \in \mathcal{X}$. For the dummy parent, $H(x_0) = \epsilon|\mathcal{X} \oplus B(0, R_h)|$. If $\mathcal{X}$ is a convex set, the Steiner formula may be used to find an explicit expression of this area, see Section 4.2.3. By differentiation with respect to the parameter vector, it follows that the components of $\nabla l_k(\theta)$ are the weighted averages of $-|\mathcal{X} \oplus B(0, R_h)| + n(Z_0^j)/\epsilon$ and $-n(W_j)\,\pi\, R_h^2 + \sum_{i=1}^{n(W_j)} n(Z_i^j)/\mu$ where $n(W_j)$ denotes the number of genuine parents in W_j, and Z_0^j its orphan cluster. The EM-updates are easily derived:

$$\begin{aligned} \epsilon^{(n+1)} &= \frac{E_{\theta^{(n)}}\left[n(Z_0) \mid Y = \mathbf{y}\right]}{|\mathcal{X} \oplus B(0, R_h)|}; \\ \mu^{(n+1)} &= \frac{E_{\theta^{(n)}}\left[\sum_{i=1}^{n(X)} n(Z_i) \mid Y = \mathbf{y}\right]}{E_{\theta^{(n)}}\left[\pi R_h^2 n(X) \mid Y = \mathbf{y}\right]}. \end{aligned}$$

For the example on redwood seedlings (Section 4.4.3) with a unit rate hard core prior at range 0.03 and reference parameter vector $(10.0, 183.06)$ as in Figure 4.2, the Monte Carlo log likelihood ratio for $\epsilon \in (5, 30)$ and $\mu\pi R_h^2 \in (1.14, 7.14)$ is given in Figure 4.4; the solution of the Monte Carlo score equations is $(\hat{\epsilon}_{100}, \hat{\mu}_{100}) = (19.65, 354.15)$. For comparison, the Monte Carlo EM-updates would be $\epsilon = 15.12$ and $\mu = 311.61$ corresponding to 3.64 daughters on average in a cluster.

4.4.5 Adaptive Coupling from the Past

Remarkably, the spatial birth–and–death approach described in Section 4.4.3 can be adapted to yield an *exact* sample from the desired posterior distribution using coupling from the past (Propp and Wilson 1996, Kendall and Møller 2000). Such algorithms are particularly efficient when there is some monotonicity in the state space, and the sampler respects this order. In the context of this chapter, the prior distribution of X is a repulsive Markov point process. Whether the same is true for the posterior distribution depends on the grain distributions Q_ξ. However, for Cox cluster processes, we showed in Section 4.4.2 that the posterior distribution is repulsive and hereditary too. Moreover, (4.20)–(4.21) reverse the inclusion ordering in the sense that if $\mathbf{x} \subseteq \mathbf{x}'$ then $b(\mathbf{x}, \xi) \geq b(\mathbf{x}', \xi)$ for all $\xi \in \mathcal{X}$, while $d(\mathbf{x}, x_i) \leq d(\mathbf{x}', x_i)$ for $x_i \in \mathbf{x}$. Our proof can be found in Section 4.9.3 of van Lieshout (2000). If the displacement functions $h(\cdot|\cdot)$ are uniformly bounded by H, the posterior inherits local stability from the prior. Such properties are particularly pleasing for Bayesian analysis, as they imply

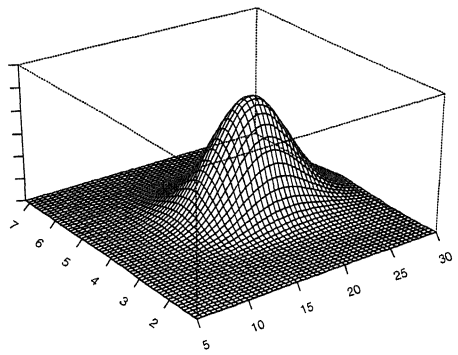

Figure 4.4 *Monte Carlo log likelihood ratio surface as a function of the noise intensity* $\epsilon \in (5, 30)$ *and the mean cluster size* $\mu\pi R_h^2 \in (1.14, 7.14)$ *for the redwood seedlings data (Figure 4.1) based on a Cox–Matérn cluster process with* $R_h = 0.061$ *and reference parameter values such that the average number of points per cluster is* 2.14, *the noise intensity* $\epsilon = 10.0$. *We used a hard core prior with* $R = 0.03$ *and* $\beta = 1.0$. *One hundred realisations were subsampled from a run of a spatial birth-and-death process over* 2.0×10^4 *time units after burn-in.*

that the choice of prior is not crucial in these respects. Hence the coupling from the past algorithm of Kendall and Møller (2000) for locally stable point processes in principle applies. Those authors presented their method for the constant death rate dynamics, with a dominating process that is Poisson with an upper bound to the conditional intensity of the distribution to be sampled as intensity. In our context, such a method would be impractical, as in most cases the upper bound will be orders of magnitude too large. For this reason, we present an adaptive coupling from the past algorithm based on (4.20)–(4.21). Suppose a spatial birth-and-death process with transition rates $b(\cdot, \cdot)$ and $d(\cdot, \cdot)$ is available to sample from the posterior density of cluster centres $p_{X|Y}(\cdot \mid \mathbf{y})$, and we have upper and

lower bounds

$$b(\mathbf{x},\xi) \leq \bar{b}(\xi) \tag{4.24}$$

$$d(\mathbf{x}\cup\{\xi\},\xi) \geq \underline{d}(\mathbf{x}\cup\{\xi\},\xi) \tag{4.25}$$

holding for all configurations $\mathbf{x}$ and all $\xi \in \mathcal{X}$. Suppose furthermore unique probability density $\pi(\cdot)$ solving the detailed balance equations

$$\pi(\mathbf{x})\,\bar{b}(\xi) = \pi(\mathbf{x}\cup\{\xi\})\,\underline{d}(\mathbf{x}\cup\{\xi\},\xi)$$

exists. For the classical constant death rate process, $\underline{d}(\mathbf{x}\cup\{\xi\},\xi) \equiv 1$, $\bar{b}(\xi$ is an upper bound to the posterior conditional intensity at ξ that doe not depend on the configuration to which ξ is added, and $\pi(\cdot)$ defines a inhomogeneous Poisson process with intensity function $\bar{b}(\cdot)$. The generic adaptive choice in our context is

$$\bar{b}(\xi) = \lambda \exp\left[-\int_A h(a|\xi)da\right] \prod_{j=1}^{m}\left[1+\frac{h(y_j|\xi)}{\epsilon}\right] \leq \lambda \prod_{j=1}^{m}\left[1+\frac{h(y_j|\xi)}{\epsilon}\right]$$

where λ is the prior local stability bound. If a uniform bound is required (Chapter 5), the right hand side above may be replaced by

$$\lambda \sup_{\xi\in\mathcal{X}} \prod_{j=1}^{m}\left[1+\frac{h(y_j|\xi)}{\epsilon}\right].$$

Similarly, for the transition rates given by (4.20)–(4.21), generic bounds are

$$\bar{b}(\xi) = \lambda\left[1+\sum_{j=1}^{m}\frac{h(y_j|\xi)}{\epsilon}\right]$$

and

$$\underline{d}(\mathbf{x}\cup\{\xi\},\xi) = \exp\left[\int_A h(a|\xi)\,da\right]\left[1+\sum_{j=1}^{m}\frac{h(y_j|\xi)}{\epsilon}\right] \Big/ \prod_{j=1}^{m}\left[1+\frac{h(y_j|\xi)}{\epsilon}\right].$$

The corresponding equilibrium distribution is that of a Poisson process with intensity function $\lambda e^{-\int_A h(a|\xi)\,da}\prod_{j=1}^{m}\left[1+\frac{h(y_j|\xi)}{\epsilon}\right]$. However, one may often do better by exploiting specific model characteristics, as we shall illustrate in Section 4.4.6 below. If we couple the spatial birth-and-death process defined by $\bar{b}(\cdot)$ and $\underline{d}(\cdot,\cdot)$ to processes defined by $b(\cdot,\cdot)$ and $d(\cdot,\cdot)$ as in Kendall and Møller (2000), we obtain the following algorithm.

Algorithm 1 Let $V_{t,\xi}$, $t \leq 0$, $\xi \in \mathcal{X}$, be a family of independent, uniformly distributed random variables on $(0,1)$. Initialise $T = 1$, and let $D(0)$ be a sample from $\pi(\cdot)$. Repeat

- extend $D(\cdot)$ backwards until time $-T$ by means of a spatial birth-and-death process with birth rate $\bar{b}(\cdot)$ and death rate $\underline{d}(\cdot,\cdot)$;

- generate a lower process $L_{-T}(\cdot)$ and an upper process $U_{-T}(\cdot)$ on $[-T, 0]$ as follows:
  - initialise $L_{-T}(-T) = \emptyset$, $U_{-T}(-T) = D(-T)$;
  - to each forward transition time $t \in (-T, 0]$ of $D(\cdot)$ correspond updates of the upper and lower processes;
  - in case of a death (i.e. a backwards birth), say $D(t) = D(t-) \setminus \{d\}$ where $D(t-)$ denotes the state just prior to time t, the point d is deleted from $L_{-T}(t-)$ and $U_{-T}(t-)$ as well;
  - in case of a birth, say $D(t) = D(t-) \cup \{\xi\}$, the point ξ is added to $U_{-T}(t-)$ only if

$$V_{t,\xi} \leq \max \left\{ \frac{b(\mathbf{x}, \xi)\, \underline{d}(\mathbf{x} \cup \{\xi\}, \xi)}{\bar{b}(\xi)\, d(\mathbf{x} \cup \{\xi\}, \xi)} : L_{-T}(t-) \subseteq \mathbf{x} \subseteq U_{-T}(t-) \right\}.$$

    similarly, ξ is added to $L_{-T}(t-)$ only if $V_{t,\xi}$ does not exceed the above expression with a minimum instead of a maximum;

- if $U_{-T}(0) = L_{-T}(0)$, return the common value $U_{-T}(0)$; otherwise set $T := 2T$;

until the upper and lower processes have coalesced.

The next theorem gives conditions for algorithm 1 to output unbiased samples from the posterior distribution of cluster centres.

Theorem 1 Let $p_X(\cdot)$ be an anti-monotone, locally stable Markov point process density with respect to a unit rate Poisson process on a compact set $\mathcal{X} \subseteq \mathbb{R}^d$, and $h(\cdot|\cdot)$ a uniformly bounded displacement function of a Cox cluster process U observed in a bounded window A. Suppose the birth rates $b(\cdot, \cdot)$ and death rates $d(\cdot, \cdot)$ define a unique spatial birth-and-death process converging in distribution to the posterior density of cluster centres $p_{X|Y}(\cdot \mid \mathbf{y})$, and there exist upper and lower bounds (4.24)–(4.25) also defining a unique spatial birth-and-death process that converges in distribution to a probability density $\pi(\cdot)$ for which $\pi(\emptyset) > 0$ and detailed balance between births and deaths holds. Then the coupling from the past algorithm 1 almost surely terminates and outputs an unbiased sample from $p_{X|Y}(\cdot \mid \mathbf{y})$.

The proof is an adaptation to the inhomogeneous case of the proof in Kendall and Møller (2000).

Proof. First, note that by assumption the dominating process $D(\cdot)$ is in equilibrium, its distribution being defined by $\pi(\cdot)$. Clearly, for all $T > 0$,

$$\emptyset = L_{-T}(-T) \subseteq U_{-T}(-T) = D(-T)$$

and by construction the updates respect the inclusion order. Hence $L_{-T}(t) \subseteq U_{-T}(t)$ for all $t \in [-T, 0]$. Moreover, the processes funnel, i.e.

$$L_{-T}(t) \subseteq L_{-S}(t) \subseteq U_{-S}(t) \subseteq U_{-T}(t) \tag{4.26}$$

whenever $-S \leq -T \leq t \leq 0$. The first inclusion can be verified by noting that $L_{-T}(-T) = \emptyset \subseteq L_{-S}(-T)$ and recalling that the transitions respect the inclusion order. Since $U_{-T}(-T) = D(-T) \supseteq U_{-S}(-T)$, the last inclusion in (4.26) follows by the same argument. If $L_{-T}(t_0) = U_{-T}(t_0)$ for some $t_0 \in [-T, 0]$, as the processes are coupled, $L_{-T}(t) = U_{-T}(t)$ for all $t \in [t_0, 0]$. Next, set $X_{-T}(-T) = \emptyset$ and define a process $X_{-T}(\cdot)$ on $[-T, 0]$ in analogy to the upper and lower processes, except that if $X_{-T}(t-) = \mathbf{x}$ the birth at time t of a point ξ is accepted if $V_{t,\xi} \leq \frac{b(\mathbf{x},\xi)\, \underline{d}(\mathbf{x}\cup\{\xi\},\xi)}{b(\xi)\, d(\mathbf{x}\cup\{\xi\},\xi)}$. In other words, $X_{-T}(\cdot)$ exhibits the dynamics of a spatial birth-and-death process with birth rate $\tilde{b}(\mathbf{x}, \xi) = b(\mathbf{x}, \xi)\frac{\underline{d}(\mathbf{x}\cup\{\xi\},\xi)}{d(\mathbf{x}\cup\{\xi\},\xi)}$ and death rate $\tilde{d}(\mathbf{x} \cup \{\xi\}, \xi) = \underline{d}(\mathbf{x} \cup \{\xi\}, \xi)$. Thus, its detailed balance equations coincide with those for $b(\cdot, \cdot)$ and $d(\cdot, \cdot)$. Furthermore, $\tilde{b}(\cdot, \cdot) \leq b(\cdot, \cdot)$, hence explosion is prevented so that the process converges in distribution to its equilibrium distribution defined by $p_{X|Y}(\cdot \mid \mathbf{y})$. The inclusion properties derived above imply $L_{-T}(0) \subseteq X_{-T}(0) \subseteq U_{-T}(0)$, so that – provided the sampler terminates almost surely – with probability 1 the limit $\lim_{T\to\infty} X_{-T}(0)$ is well-defined. Since $D(\cdot)$ is in equilibrium, $X_{-T}(0)$ has the same distribution as if the X-process were run forward from time 0 (coupled to the dominating process as before) over a time period of length T, the limit distribution of which is $p_{X|Y}(\cdot \mid \mathbf{y})$. We conclude that the algorithm outputs an unbiased sample from the posterior distribution of parents. It remains to show that coalescence occurs almost surely. Recall that by assumption $\pi(\emptyset) > 0$. Set, for $n \in \mathbb{N}_0$, $E_n = \mathbf{1}\{D(-n) \neq \emptyset\}$. Now $(E_n)_n$ is an irreducible aperiodic Markov chain on $\{0, 1\}$ for which the equilibrium probability $\pi(\emptyset)$ of state 0 is strictly positive. Hence state 0 will be reached with probability 1, which implies that the dominating process $D(t)_{t\leq 0}$ will almost surely be empty for some t. But then (4.26) and the coupling imply that the algorithm terminates almost surely, and the proof is complete. □

4.4.6 Example: Cox-Matérn Cluster Process

To describe a tailor-made coupling from the past algorithm, consider a Cox cluster process with intensity function given by (4.4)–(4.5) and prior density (4.17). For this model, the birth rate (4.20) satisfies

$$b(\mathbf{x}, \xi) \leq \beta \left[1 + \sum_{j=1}^{m} \frac{h(y_j|\xi)}{\epsilon} \right] = \bar{b}(\xi). \tag{4.27}$$

In order to derive a lower bound for the death rate (4.21), note that

$$1 + \frac{h(y_j|\xi)}{\lambda(y_j \mid \mathbf{x})} \leq 1 + \frac{\mu}{\epsilon + \mu} \leq 2$$

if $y_j \in B(\xi, R_h) \cap U_{\mathbf{x}}$, where $U_{\mathbf{x}} = \cup_{x_i \in \mathbf{x}} B(x_i, R_h)$ denotes the union of balls centred at the points of $\mathbf{x}$. It follows that the death rate $d(\mathbf{x} \cup \{\xi\}, \xi)$ is bounded below by

$$\underline{d}(\mathbf{x} \cup \{\xi\}, \xi) = d(\mathbf{x} \cup \{\xi\}, \xi) \prod_{j: y_j \in B(\xi, R_h) \cap U_{\mathbf{x}}} \left(\frac{1}{2} + \frac{\mu}{2\,\lambda(y_j \mid \mathbf{x})} \right). \quad (4.28)$$

By the Preston theorem, the transition rates $\bar{b}(\xi)$ and $\underline{d}(\mathbf{x} \cup \{\xi\}, \xi)$ define

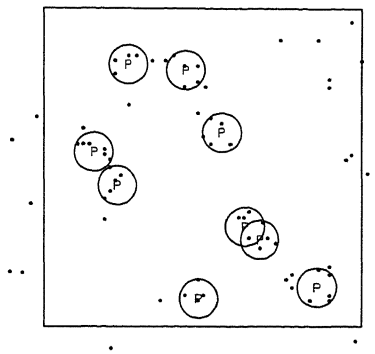

Figure 4.5 *Realisation of extrapolated redwood seedling pattern on* $\mathcal{X} \oplus B(0, R_h)$ *from observations in unit square (box) and interpolated parent pattern ('P') based on a Cox–Matérn cluster process with* $R_h = 0.061$ *and* $(\hat{\epsilon}_{100}, \hat{\mu}_{100})$ *obtained by coupling from the past. We used a hard core prior with* $R = 0.03$ *and* $\beta = 1.0$.

a unique spatial birth-and-death process, whose limit distribution is given by

$$\pi(\mathbf{x}) \propto \gamma^{-n(\mathbf{y} \cap U_{\mathbf{x}})} \prod_{i=1}^{n(\mathbf{x})} \beta(x_i), \quad (4.29)$$

a generalised area-interaction process (Widom and Rowlinson 1970, Baddeley and van Lieshout 1995, Kendall 1998, Häggström et al. 1999) with intensity function

$$\beta(\xi) = \beta \exp\left[- \int_A h(a|\xi)\, da \right] 2^{n(\mathbf{y} \cap B(\xi, R_h))}$$

and interaction parameter $\gamma = (\frac{1}{2} + \frac{\mu}{2\epsilon})^{-1}$. Regarding the implementation of algorithm 1, note that

$$\frac{b(\mathbf{x},\xi)\,\underline{d}(\mathbf{x}\cup\{\xi\},\xi)}{\overline{b}(\xi)\,d(\mathbf{x}\cup\{\xi\},\xi)} = \mathbf{1}\{d(\xi,\mathbf{x}) > R\} \prod_{j:y_j\in B(\xi,R_h)\cap U_{\mathbf{x}}} \left(\frac{1}{2} + \frac{\mu}{2\lambda(y_j \mid \mathbf{x})}\right)$$

is decreasing in $\mathbf{x}$, so the sampler is anti-monotone, and the births in the upper and lower processes may be implemented by simply considering the current state of the other process at each transition; see Kendall (1998). We applied the above algorithm to the redwood seedlings data of Figure 4.1 for the Matérn parameter vector (ϵ, μ) equal to its Monte Carlo maximum likelihood estimate (cf. Section 4.4.4) and a hard core prior with $\beta = 1.0$ and $R = 0.03$ as before. A typical realisation from the posterior distribution of parents can be seen in Figure 4.5 as well as an extrapolation of the redwood pattern to the set $\mathcal{X} \oplus B(0, R_h)$.

4.5 Summary and Conclusion

We discussed issues arising when a spatial pattern is observed within some bounded region of space, and one wishes to predict the process outside of this region (extrapolation) as well as to perform inference on features of the pattern that cannot be observed (interpolation). We focused on spatial cluster analysis. Here the interpolation arises from the fact that the centres of clustering are not observed. We took a Bayesian approach with a repulsive Markov prior, derived the posterior distribution of the complete data, i.e. cluster centres with associated offspring marks, and proposed an adaptive coupling from the past algorithm to sample from this posterior. The approach was illustrated by means of the redwood data set (Ripley 1977).

Acknowledgements

We gratefully acknowledge the expert programming support of Adri Steenbeek, very helpful comments from Yih Chong Chin, Nick Fisher, Richard Gill, Ilya Molchanov and Elke Thönnes, and the influence of previous unpublished collaboration with Andrew Lawson and Henry Y.W. Cheng. The research for this chapter was carried out under CWI project PNA4.3 'Stochastic Geometry', ARC large grant A69941083 'Extrapolating and Interpolating Spatial Patterns', and NWO grant CIMS 613.003.100.