Modelling Spatial Patterns

By B. D. RIPLEY

Imperial College, London

[Read before the ROYAL STATISTICAL SOCIETY at a meeting organized by the RESEARCH SECTION on Wednesday, March 16th, 1977, Professor S. D. SILVEY in the Chair]

SUMMARY

Spatial point processes may be analysed at two levels. Quadrat and distance methods were designed for the sampling of a population in the field. In this paper we consider those situations in which a map of a spatial pattern has been produced at some cost and we wish to extract the maximum possible information. We review the stochastic models which have been proposed for spatial point patterns and discuss methods by which the fit of such a model can be tested. Certain models are shown to be the equilibrium distributions of spatial–temporal stochastic processes. The theory is illustrated by several case studies.

Keywords: SPATIAL PATTERN; POINT PROCESS; HARD-CORE MODEL; MARKOV POINT PROCESS; MATHEMATICAL MORPHOLOGY; CELL SHAPES

1. INTRODUCTION

PATTERNS of "small" objects in two or three dimensions or on the surface of the terrestrial or celestial spheres are commonplace; some examples are towns in a region, trees in a forest and galaxies in space. We will see that other spatial patterns such as a sheet of biological cells can be reduced to a pattern of points.

Such spatial patterns are often tested for "randomness", i.e. the null hypothesis of an underlying Poisson process is tested. We can divide the methods used into four classes:

(i) those based on quadrat counts;
(ii) "distance" or nearest-neighbour methods;
(iii) second-order methods (including spectra);
(iv) the "test-set" approach.

Methods of type (i) and (ii) seem to be the most appropriate in preliminary fieldwork. Here we suppose that a complete map of the spatial pattern is available. Our aim will be to find useful non-Poisson models and to establish methods of types (iii) and (iv) to test the fit of any proposed model. The main technique will be the estimation of the first two moment measures of the supposed underlying point process.

One of the drawbacks of existing methods is their inability to test for interactions at different scales *simultaneously* in formal significance tests (*cf.* Mead, 1974). Even in the analysis of contiguous quadrat counts the choice of simultaneous scales is restricted (Hill, 1973). We attempt to overcome this problem by validating tests with artificial data.

Usually spatial patterns will have evolved over a period of time; we would like our models to reflect this temporal aspect. In Section 4 we show that some of the models reviewed in Section 3 can arise as the equilibrium distributions of spatial–temporal stochastic processes. We use this fact in Section 5 to design procedures for the Monte Carlo simulation of some of our models.

It is impossible to write down the complete likelihood function for most of our models, so the usual methods for the estimation of unknown parameters cannot be used. This and other difficulties are discussed in several case studies.

2. Spatial Point Processes

Our models will represent the "small" objects by points or by spheres of fixed radius. (The latter may be specified by their centres; this is a special case of the basic transformation of stochastic geometry (Kendall, 1974a).) We assume that for a realization of the spatial pattern we are able to observe all the objects within a known set E which we will call the *sampling window*. (Alternatively E might be natural and unknown; it is then a separate interesting pattern recognition problem to estimate E from the data (*cf.* Grenander, 1973, 1977).) In practice it may be difficult or impossible to collect the data in this form; for instance, in a mature forest it may be much easier to measure nearest-neighbour distances.

Occasionally E may be a natural region but usually the pattern extends beyond E to an unknown extent. We assume E to be a homogeneous region (as far as causal factors are concerned) and that there is no preferred origin or direction. This assumption forces us to postulate that the pattern extends throughout the space (two- or three-dimensional space or a sphere, denoted by $\mathscr{X}$).

Formally a *model* is the distribution of a simple second-order point process strictly stationary under rigid motions (in the terminology of Ripley, 1976a). An informal description follows. Each realization is a countable set of points from $\mathscr{X}$ (a finite set if $\mathscr{X}$ is sphere). The assumption of simplicity excludes the possibility that two points might coincide in space. The important random variables are the $Z(A)$ counting the number of points in each Borel subset A of $\mathscr{X}$. If A is bounded then $Z(A)$ must be finite; we are assuming that $E(Z(A)^2)$ is also finite. The statistical properties are invariant under both translations and rotations.

The simplest parameter of a model is the *intensity* λ, the expected number of points per unit area; this summarizes the first moments of the family $\{Z(A)\}$. The second moment structure may be reduced to a function K defined on $(0, \infty)$ (or $(0, \pi]$ for a sphere). The following intuitive definitions of this function are explained in the Appendix.

(a) $\lambda^2 K(t)$ is the expected number of ordered pairs of distinct points less than distance t apart with the first point in a given set of unit area.

(b) $\lambda K(t)$ is the expected number of further points within t of an arbitrary point of the process.

(c) Under additional assumptions $g(t) = (\lambda^2 dK/dt)/c(t)$ is a joint density for the occurrence of two points distance t apart (where $c(t) = 2\pi t$, $4\pi t^2$ or $2\pi \sin t$ for R^2, R^3 or the unit sphere respectively). Thus g is a *product density*. We can express the second moments of $\{Z(A)\}$ by

$$E\{Z(A)\,Z(B)\} = \lambda \nu(A \cap B) + \lambda^2 \int_{(0,\infty)} \nu_t(A \times B)\, dK(t)$$

where ν is Lebesgue measure on $\mathscr{X}$ and

$$\nu_t(A \times B) = \int \sigma_t[\{y - x\colon x \in A, y \in B, d(x,y) = t\}]\, d\nu(x).$$

Here $d(x, y)$ is the distance between x and y (the great-circle distance for a sphere) and σ_t is the uniform probability on the surface of the sphere of radius t centred at the origin. (The existence of this decomposition and the uniqueness of K are established in Ripley, 1976a.)

We avoid technical difficulties by considering the distribution function K of a measure rather than its density; we shall see later that it is easier to use the cumulative function in statistical analyses (as we might have expected from real random variables). The choice of the power of λ in the definition of K is explained below; notice that K has the dimensions of an area or volume.

We may also reduce the higher moments of a model; we shall discuss third moments in an example.

We may specify a model by giving $P\{Z(A_1) = 0, \ldots, Z(A_r) = 0\}$ for all finite collections $\{A_1, \ldots, A_r\}$ of spheres of the form $b(x, t) = \{y\colon d(x, y) < t\}$ for any $x \in \mathscr{X}$ and $t > 0$ (Ripley,

1976b). The "first-order" part of this description would be to consider $p(t) = P\{Z(b(x, t)\} > 0)$ which is independent of the choice of x by stationarity. This provides another function on $(0, \infty)$ summarizing part of the model.

In Section 6 we will construct estimators of both K and p. In the next three sections we study classes of models.

3. MODELS

The basic model is the Poisson process defined (within our framework) by either or both of the following properties: $Z(A)$ has a Poisson distribution for all bounded Borel A and $\{Z(A_i)\}$ are independent whenever $\{A_i\}$ is a disjoint class of such sets. This model is specified by its intensity λ; $K(t) = \pi t^2$ irrespective of λ (hence the scaling of K). Models which give "clustered" patterns will have K greater than this function whereas "self-inhibiting" models will have a smaller K.

We will describe our models on R^2; they may be defined on R^3 or a sphere with trivial modifications.

3.1. *Cluster Processes*

One of the earliest and most intensively studied classes of models is the class of cluster processes (see Neyman and Scott, 1972, for a historical survey.) Recently cluster processes have played an important role in the theory of infinitely divisible point processes (cf. Matthes, 1972). We have a parent process (usually Poisson) on a group G, either R^2 or the group of rigid motions, and a daughter process ϕ. We sample the parent process; with each point we associate an independent copy of ϕ translated by the chosen transformation. The cluster process is the superposition of the daughter processes. For a Poisson parent process the use of the whole group of rigid motions amounts to giving the daughter process an independent uniformly distributed rotation. For conditions for such a model to be well defined see Goldman (1967); Fisher (1972) gives an elementary survey.

The most important subclass consists of Neyman–Scott processes. Here the daughter process ϕ is a random number N of independent identically distributed points and the parent process is Poisson of intensity α.

If the density of the distribution of the distance between two arbitrary points of the daughter process exists and is denoted by f, we have:

$$\lambda = \alpha E(N),$$

$$g(t) = \lambda^2 + \alpha E\{N(N-1)\} f(t)/c(t).$$

The first term in the last expression comes from the consideration of two points from different clusters, the second from two points within a cluster.

3.2. *Hard-core Models*

Various models have been proposed for processes of the centres of non-overlapping circles or spheres of radius R. Matérn (1960) defined two such models. The first is obtained by sampling a Poisson process of intensity α and deleting any point which is within $2R$ of any other whether or not this has already been deleted. For this process

$$\lambda = \alpha \exp(-4\pi\alpha R^2)$$

and $g(t)$ exists and is given by

$$g(t) = \begin{cases} 0, & t \leqslant 2R, \\ \alpha^2 \exp\{-\alpha U(t)\}, & t > 2R, \end{cases}$$

where

$$U(t) = \text{meas}\{b(0, 2R) \cup b((t, 0), 2R)\},$$

$$b(x, r) = \{y\colon d(x, y) < r\}.$$

This model has a maximum intensity (as α is varied) of $(4\pi eR^2)^{-1}$ and so cannot model densely packed cores. This is 10 per cent of the theoretical bound $(\sqrt{12}R^2)^{-1}$ which is attained by a triangular lattice packing (see Rogers, 1964).

Matérn's second process was one of the first marked point processes; the points of a Poisson process of intensity α are independently marked with a uniformly distributed birth time on $(0, 1)$, so that we have a Poisson process on $R^2 \times (0, 1)$. A point is retained if no point within $2R$ has an earlier birth time. We can define two variants of this process by considering points which have already been deleted or not. Matérn took the former view; although this is less natural it does enable us to find

$$\lambda = (1 - e^{-\alpha c})/c, \quad c = U(0) = 4\pi R^2,$$

$$g(t) = \begin{cases} 0, & t \leqslant 2R, \\ \dfrac{2U(t)(1 - e^{-\alpha c}) - 2c(1 - e^{-\alpha(t)})}{cU(t)(U(t) - c)}, & t > 2R. \end{cases}$$

These formulae are summarized in Paloheimo (1971). This second model may be appropriate in botanical problems but not in competitive zoological situations such as nesting colonies of seabirds. We shall meet this dichotomy again.

These models may also be defined on a bounded set. If we wish to simulate their restrictions to a bounded set A we may do so by simulating the Poisson process on $\{y\colon \exists x \in A, d(x, y) \leqslant 2R\}$; notice that this is impossible for the second variant of the second model.

If we do consider a point process on a bounded subset of the plance we will have to impose boundary conditions. The natural choice is to assume that there are no points outside this set but this will be appropriate only if we know the true extent of the process to be modelled. In general, a more satisfactory solution is to take a sufficiently large square B and impose periodic boundary conditions. Formally we regard B as a torus. If we envisage the plane packed by a lattice of translations of B we will obtain a process which is strictly stationary under the group of translations parallel to the sides of B (if the process on the torus is strictly stationary, as all our examples are); however, this process is never stationary under rotations. We will only consider this process restricted to a subset of B; it is then plausible that this directionality (which is the price we have to pay on a bounded set) will be negligible. It may be possible to obtain a limiting distribution for the restriction of the process to a bounded set A as the square B expands (*cf.* Ruelle, 1970). We ignore this directionality even though it strictly invalidates our reduction of the second moments.

Our third hard-core model may only be defined on a bounded set A. We sample a Poisson process on A and retain only those *samples* which contain no pair of points less than $2R$ apart.

Our three hard-core models differ even on spaces such as a torus or sphere where edge-effects are absent. This can be grasped most easily by considering their analogues on the discrete space of twelve sites arranged on a clock-face. The equivalent of hard-cores is to forbid the occupation of adjacent sites. It is possible to calculate the probabilities of various triples under all these processes (by enumerating all outcomes). Rejecting those samples which contain forbidden combinations generates all triples equiprobably. The analogues of both Matérn's models give different probabilities to $\{2, 4, 6\}$ and $\{2, 6, 10\}$. Of course Matérn's two models differ; their second moment properties expressed by g are different.

Bartlett (1974) considered an extension of Matérn's second model with variable core size. Suppose each point is marked independently with a birth time and a core size R chosen from

a prescribed distribution. Various definitions are possible; perhaps the most plausible is that a point is deleted if there is another with an earlier birth time (deleted or not) such that the discs of marked size centred on the points overlap. Bartlett ignored deleted points, deleting a point if an earlier point fell within the disc of marked size centred on the point.

The SSI process of Diggle, *et al.* (1976) is the variant of Matérn's second model conditioned on the total number of points.

3.3. *Fixed-range Interactions*

Throughout the rest of this section we will define point processes on a bounded set A by the Radon–Nikodym derivatives f of their distributions with respect to that of a Poisson process. Thus f is a function from the class of finite subsets of A to $(0, \infty)$. For each realization x we can think of $f(x)$ as a measure of how much more frequently the finite set of points x occurs in this process than in the Poisson process, in the same way that the density of a continuously distributed random variable heuristically measures the frequency of occurrence of a particular value. Almost always f will be defined up to a constant factor which cannot be given explicitly. We demand that f be *hereditary*, that is if $f(x) > 0$ then $f(y) > 0$ for all $y \subset x$.

Pairs of points whose distance apart are less than $r = 2R$ are called *neighbours*. In a hard-core model such pairs are excluded. Consider a Borel set B and its complement $A \backslash B$. Let

$$B^* = \{\xi; \xi \notin B, \exists \eta \in B \text{ with } d(\xi, \eta) < r\}.$$

If for every B the distribution of the process restricted to B conditional on the realization in $A \backslash B$ is the same as that conditional on that part of the realization in B^* the process is said to be *Markov* of range r. (In statistical mechanics such a process is said to have finite range (Preston, 1976b)). This concept extends that of a Markov random field on a finite lattice or graph (e.g. Besag, 1974; Preston, 1974). We have the following extension of the Markov field theorem:

$$f(x) = \prod_{y \subset x} \phi(y),$$

where $\phi(y) = 1$ unless every pair of points in y are neighbours (Ripley and Kelly, 1977, cf. Hammersley, 1975). Notice that the requirement that f be hereditary is a natural relaxation of what Hammersley (1972) called the "positivity condition".

Suppose the conditional density of a point ξ given the realization in $A \backslash \{\xi\}$ depends only on the *number* of neighbours of ξ. Then

$$f(x) = a b^{\#(x)} c^{s(x)}$$

where $\#(x)$ is the number of points in x and $s(x)$ is the number of pairs of neighbours. Here b may be any positive number (which may be suppressed by increasing the intensity of the underlying Poisson process by a factor of b) but $0 \leqslant c \leqslant 1$ unless the total number of points is limited. The case $c = 0$ is our third hard-core model. For a fixed sample size and $c > 0$ this formula was given by Strauss (1975); the quoted result is taken from Kelly and Ripley (1976).

3.4. *Pairwise Interactions*

In the last class of models the interaction between pairs of points depends only on whether or not they are neighbours. In this subsection we let this interaction depend on the distance between the pair. We define

$$f(x) = a b^{\#(x)} \prod_{\substack{\xi, \eta \in x \\ \xi \neq \eta}} h\{d(\xi, \eta)\}$$

for constants a and b and an interaction function $h : (0, \infty) \to [0, \infty)$. If we consider realizations with two points we see that $h\{d(\,,\,)\}$ is a joint density for the pair of points. The product

form allows us to combine all these pairwise interactions; it was suggested by the formulae of statistical mechanics (e.g. Ruelle, 1969) where $-\log h(s)$ is proportional to the interaction energy between two particles distance s apart. We have to check whether for a given b and h this function is the Radon–Nikodym derivative of a probability for some choice of a. In general this need not be so. We shall use these models only for bounded functions h which vanish on $(0, r)$. Then $f(x) = 0$ if $\#(x)$ exceeds the finite number N of balls of radius $(r/2)$ which may be packed into A, and $f(x) \leqslant ab^N (\sup h)^{N^2/2}$, so f is integrable.

Note that the final models of the last subsection are obtained by setting

$$h(d) = \begin{cases} c, & d < r, \\ 1, & d \geqslant r. \end{cases}$$

3.5. *Mixtures of Models*

Classes of models have been proposed which are mixtures of simpler models, the most important being mixed Poisson and doubly stochastic Poisson (or Cox) processes. Such models cannot be fitted from a single sample unless parametric assumptions are made about the random mean measure; see Grandell (1972).

4. EVOLVING POINT PROCESSES

In this section we show how point processes on a bounded set A may arise as the equilibrium distributions of spatial–temporal stochastic processes. These are of two types; we may think of them as modelling botanical and zoological phenomena.

4.1. *Spatial Birth-and-death Processes*

This class of Markov jump processes was introduced by Preston (1976a). Suppose we are interested in a point process whose distribution is specified by a Radon–Nikodym derivative f with respect to that of the Poisson process of unit intensity on a bounded set A. The state space of our stochastic process will be $\mathscr{X} = \{x : f(x) > 0\}$, i.e. the class of finite sets which have positive "density". Thus viewed at any fixed time the stochastic process is a point process. The process evolves in jumps, a finite number of which occur in each bounded time interval. These jumps are of two types, a "birth" in which a single point is added, or a "death" in which one of the points is deleted. The probability that a point within F, a measurable subset of A, is added to the finite set x between times t and $t+s$ is $s\int_F b(x, \xi)\, d\nu(\xi) + o(s)$, and the probability that the point ξ is deleted from $x \cup \xi$ in this time interval is $D(x, \xi)s + o(s)$. The functions b and D characterize a spatial birth-and death process; of course we have to impose conditions on b and D to ensure that the process exists, in particular that we do not have an infinite sequence of jumps in a finite time. Sufficient conditions and a formal construction are given by Preston.

We would like the point process which results from viewing our stochastic process at each fixed time to be that specified by f, i.e. we would like this to be the equilibrium distribution. Suppose the following condition holds:

$$b(x, \xi) f(x) = D(x, \xi) f(x \cup \xi) > 0 \quad \text{if } x \cup \xi \in \mathscr{X}. \tag{1}$$

This is a "detailed balance" condition which ensures that births from x to $x \cup \xi$ match deaths from $x \cup \xi$ to x. The stochastic process is then time-reversible and the point process specified by f is its unique equilibrium distribution. (This may be checked directly. The uniqueness is not given by Preston but follows from the indecomposability of $\mathscr{X}$ ensured by (1) and the assumption that f is positive on the hereditary set $\mathscr{X}$).

Thus we can consider our point process evolving in time by the addition and deletion of points which remain fixed throughout their lifetimes. The ratio

$$b(x, \xi)/D(x, \xi) = f(x \cup \xi)/f(x)$$

of the birth rate to the death rate is a measure of the "viability" of a point at ξ given the rest of the points x. For a Markov point process this ratio is $\prod_{y \subset x \phi}(y \cup \xi)$ which depends only on the interaction of ξ with the set of its neighbours in x. For a pairwise interaction process we have $\prod_{\eta \in x} h\{d(\eta, \xi)\}$ which is a measure of the "crowding near ξ".

We have not yet shown that (1) imposes sufficient conditions on b and D for the existence of our stochastic process; indeed it does not. We can, however, always take $b(x, \xi) = 1$ for $x \cup \xi \in \mathscr{X}, 0$ otherwise. Then proposition 5.1 of Preston (1976a) shows that the process *is* well defined. Preston also considers conditions under which $P\{X(t) \in F\}$ converges to the probability given to the measurable subset F of $\mathscr{X}$ by the equilibrium point process. If this point process has a bounded number of points, for instance any hard-core model, then the process converges in this sense to its unique equilibrium distribution given by f.

For the Strauss process, Kelly and Ripley (1976) considered the spatial birth-and-death processes given by $b(x, \xi) = 1$, $D(x, \xi) = \exp\{-t(x, \xi)\}$ and $b(x, \xi) = \exp\{t(\xi, x)\}$, $D(x, \xi) = 1$, where $t(\xi, x)$ denotes the number of neighbours of ξ in x. Both processes are well defined and converge to the Strauss process.

4.2. *Diffusions*

For pairwise interaction processes conditioned on the total number of points N we can find a model with continuously moving particles. The N particles perform (mathematical) diffusions with constant infinitesimal variance but with drift the vector sum of $\nabla_j \log h(d(x_j, x_i))$ over all the other points. We have to impose boundary conditions on A; these will be either reflecting or periodic. This process is defined at least if h is positive and infinitely differentiable, when with periodic boundaries there is a unique equilibrium density f which is the solution of

$$\nabla_i \log f = \sum_{j \neq i} \nabla_j \log h\{d(x_j, x_i)\}, \quad i = 1, \ldots, N$$

(J. T. Kent, personal communication); this is satisfied by our product formula for f. This process converges weakly to the pairwise interaction process conditioned on a total of N points. These assertions are very plausible for reflecting boundaries and more general interaction functions h.

We can combine this process with a spatial birth-and-death process to obtain a process with equilibrium distribution the complete pairwise interaction process. However, when using this type of model we will usually be interested in time-scales short compared with the lifetimes of the points.

5. SIMULATION

Our methods of testing the goodness of the fit of a model will depend on our ability to simulate a proposed model.

It is clear how one would simulate a Poisson point process on a bounded set A; one chooses the number of points from a Poisson distribution with mean $\lambda\nu(A)$ and distributes these uniformly and independently. To obtain a uniformly distributed point on A we can enclose A in a rectangle and generate a uniformly distributed point on this rectangle by choosing its co-ordinates uniformly. If this point lies within A we accept it; if it does not we repeat the procedure until we generate a point within A.

Those point processes defined by a derivative f with respect to a Poisson process may in principle be simulated by rejection sampling if f is bounded (which might be achieved by a suitable choice of the Poisson intensity α or by restricting attention to samples of a given size). In this procedure f is calculated for a sample of the Poisson process and compared with a uniformly distributed random variable X on $(0, \sup f)$; the sample is accepted if X is smaller otherwise the process is repeated. If f has a non-continuous distribution we must refine this procedure. The only case we shall meet is that in which f takes the value 0 with positive probability; of course we reject all such samples. An example is our third hard-core

model; in this case the probability of an acceptable sample is about $\exp(-\alpha^2\pi r^2)$, about 10^{-24} for the example of Section 7. (The number of neighbour pairs per unit area has mean about $\alpha^2\pi r^2$ and is approximately Poisson, cf. Brillinger, 1976). In this case the rejection method is impracticable; this appears to be a common conclusion.

Our simulations were obtained from the spatial birth-and-death processes of Section 4. We could choose a stochastic spatial–temporal process with the required equilibrium distribution and simulate this over a long time interval to obtain a sample of a point process with distribution "close" to that required. We will, however, be interested in samples containing a fixed number of points N. *Either* we choose $D(x, \xi) = 1$ and

$$b(x, \xi) = f(x \cup \xi)/f(x) \quad \text{if } \#(x) = N-1,$$

both 0 otherwise, *or* if f is positive we may choose $b(x, \xi) = 1$ and $D(x, \xi) = f(x)/f(x \cup \xi)$ if $\#(x) = N$, 0 otherwise. Suppose we observe these processes only when jumps occur. We obtain a discrete-time Markov process in which births and deaths alternate. In the continuous-time processes all states with N points are left at unit rate, so an equilibrium distribution for the discrete-time process has a density proportional to f over states with N points (with respect to Lebesgue measure on $(R^2)^N$). It is easily checked that these Markov chains are indecomposable and so converge to a unique equilibrium distribution (e.g. Breiman, 1968, chapter 7).

Although it was convenient to use spatial birth-and-death processes to show that these Markov chains converge to the required point process distribution we can describe them independently. The first is defined by selecting one of the points at random and deleting it, then adding a point with a density proportional to $p(\xi) = f(x \cup \xi)/f(x)$. This random point might be generated by rejection sampling. This process (every second step of the Markov chain) converges to the point process defined by f conditional on N points. The second process is defined similarly; we add a uniformly distributed point then delete one point chosen at random according to a distribution with $P(\xi$ is deleted from $x)$ proportional to $f(x \backslash \xi)/f(x)$. Because this process does not involve rejection sampling it might seem preferable to use it; empirically it tends to add a point and delete it immediately, so converging to equilibrium very slowly. In our examples the first process was used.

To use any of these stochastic processes we have to know that they are in equilibrium. We can generate a starting pattern of points from a Poisson process or from a lattice of points if as in a hard-core model only certain finite sets of points are allowed. We can use the methods of the next section to decide whether the stochastic process has for practical purposes reached equilibrium. We can then sample the process at sufficiently long intervals; for our examples this was every $4N$ steps.

6. FITTING MODELS

In the introduction we mentioned several methods which have been used to test the fit of a Poisson model. In this section we consider how we might test the fit of a quite general model. Statistical inference from just one sample will be possible because our assumption of stationarity provides in-built replication.

Nearest-neighbour methods can essentially only test for the presence of pattern on a single scale (in the ecologists' terminology). The use of second or third nearest neighbours or Besag and Gleaves' (1974) T-square methods can extend the range of scales slightly, but not for highly clustered data (see Section 8). It seems to be virtually impossible to obtain anaytical information on the distribution of distance method test statistics under most non-Poisson models. (For exceptions see Besag and Gleaves, 1974; Diggle, 1975; Diggle *et al.*, 1976; Holgate, 1965a, b, c.) Bartlett's (1974, 1976) example of herring gull nest data is a case in which the scale of distance methods is appropriate, but ecologists are often interested in the interaction of *clusters* of objects (cf. Greig–Smith's contribution to Bartlett, 1971). Thus distance methods did not seem appropriate for the intensive analysis of a map.

Finding the distribution of the numbers of points in each of a set of quadrats is straightforward; it amounts to partially specifying the model. However, even a test of the Poisson null hypothesis for a collection of scales has been described only recently (Mead, 1974). The use of counts on a grid of contiguous quadrats disregards the fine structure of the pattern. A different type of model and analysis is appropriate to such a lattice process (Besag, 1974; Bartlett, 1976).

Bartlett (1964, 1976) considered the two-dimensional spectral analysis of a spatial point pattern. Spectral analysis is closely tied to the use of the space itself as the group governing stationarity and so is most appropriate when we assume stationarity under translations only (cf. Ripley, 1976a). The spectral estimate must be smoothed; the choice of the degree of smoothing seems necessarily to introduce an arbitrary factor into the analysis.

The two methods proposed here are based on estimators of the two one-dimensional summaries of a model described in Section 2.

An unbiased estimator of $\lambda^2 K(t)$ was described in Ripley (1976a). We give each ordered pair (x, y) of objects a weight $k(x, y)$ which is the reciprocal of the proportion of the perimeter of the circle centred on x and passing through y which is within the sampling window. Thus $k(x, y)$ is inversely proportional to the chance that a point the distance of y from x had of being recorded. Let

$$\lambda^2 \hat{K}(t) = \sum k(x, y)/\nu(E),$$

the sum being over ordered pairs of distinct data points less than t apart. (Remember $\nu(E)$ is the area of the sampling window.) This is an unbiased estimator of $\lambda^2 K(t)$ for

$$t \leqslant t_0 = \inf\{s\colon \exists x \in E, \partial b(x, s) \cap E = \emptyset\}.$$

Thus if the intensity λ is *known*, $\hat{K}(t)$ is an unbiased estimator of $K(t)$ for small enough t. If λ is unknown we estimate it by the obvious quantity $\hat{\lambda} = Z(E)/\nu(E)$ and define

$$\hat{K}(t) = \sum k(x, y)\, \nu(E)\, N^{-2}, \quad \text{if } Z(E) = N.$$

In practice the possible bias of $\hat{K}$ seems to be small.

For a Poisson process we can calculate this bias exactly. Suppose we have a sample of N points. If Y is uniformly distributed on E then

$$E\{1_{(0<d(x,Y)<t)}\, k(x, Y)\} = \int_0^t \left\{ \int_{E \cap \partial b(x,s)} k(x, y)\, dm \right\} ds/\nu(E)$$

$$= \pi [\![\min [t, \inf\{s : \partial b(x, s) \cap E = \phi\}]^2/\nu(E)]\!],$$

where m is Lebesgue measure on $\partial b(x, s)$, the circumference of $b(x, s)$. Thus for $t \leqslant t_0$ the conditional expectation of each term in the sum is $\pi t^2/N^2$, so

$$E\{\hat{K}(t) \,|\, Z(E) = N\} = \pi t^2 (N-1)/N.$$

This correction factor is possibly related to the use of $\{N/\nu(E)\}^2$ as an estimate of λ^2 whereas $N(N-1)\,\nu(E)^{-2}$ is unbiased.

The advantage of an unbiased estimator of K is that one can develop an intuition about the process from K without taking into account the shape of the sampling window. Bartlett (1964) suggested the use of the distribution of the interpoint distances uncorrected for edge effects. Usually we cannot even calculate this distribution or the mean of its empirical version.

The *test-set* method is based on the function

$$\hat{p}(t) = \nu[\{x\colon (x + tB) \cap (\text{data}) \neq \phi\}]/\nu(E)$$

for a *test set* or *structuring element*, B, usually a circle, hexagon or square centred at the origin. For a process that is stationary under translations, in particular for a model, this is a nearly

unbiased estimator of the probability that tB is not empty, the bias being introduced by edge-effects. Because $P(Z(B)>0)$ is not additive in B it is not possible to introduce an edge-correction analogous to $\hat{K}$. We can obtain an unbiased estimator by considering only points within

$$E_{tB} = \{x: x+tB \subset E\}.$$

Then

$$E\{\hat{p}(t)\} = E(\nu[\{x: x \in E_{tB}, (x+tB) \cap (\text{data}) \neq \phi\}])/\nu(E_{tB})$$
$$= \int P\{Z(x+tB)>0\}\, d\nu/\nu(E_{tB}) = P\{Z(tB)>0\}$$

by stationarity. Thus if we take B to be the unit circle we obtain an unbiased estimator of $p(t)$ (defined in Section 2).

The computation of $\hat{p}$ is best done by an image analyser such as the *texture analyser* developed at Fontainebleau by Serra (Klein and Serra, 1972; Serra, 1972; Watson, 1973). These machines compute the proportion of those x's on a fine grid for which $(x+tB)$ contains a datum point. This method can also be applied to general (non-point) spatial patterns; some of the relevant theory is given by Matheron (1967, 1975).

Both $\hat{K}$ and $\hat{p}$ provide useful summaries of the data even when no model is envisaged. Very little is known about the sampling fluctuations of either statistic even for a Poisson model. Some asymptotic results for $\hat{K}$ have been given by Silverman (1976).

We are forced to use simulation to explore the sampling fluctuations, and we will use these simulations to generate the critical regions of our "goodness of fit" tests. Suppose we plot $\hat{K}$ for the data, K for the model (if known) and $\hat{K}$ for m simulations of this model. Sometimes it will be obvious that K for the data is radically different from the other plots and no formal test is then needed. If we had a particular interpoint distance in mind *before* observing the data we could compare the ordinates of $\hat{K}$ at this distance for the data and the simulations. On our null hypothesis that the data were generated by the model we have $(m+1)$ independent samples of a random variable, so the probabilities that the data gave the upper or lower extreme value are each $1/(m+1)$ (*cf.* Hope, 1968, and Barnard's contribution to the discussion of Bartlett, 1963).

It is more reasonable to assume that we are interested in a certain range of interpoint distances. We can define the acceptance region of a test by requiring $\hat{K}$ for the data to be within the envelope of $\hat{K}$ for the simulations throughout this range. The size of this test must exceed $2/(m+1)$; it may be estimated by the proportion of a further batch of simulations which are rejected. We may consider one-sided tests with acceptance region the area above the lower envelope or below the upper envelope. These will be appropriate if we have alternatives which are "self-inhibiting" or "clustered" respectively. In general our two-sided tests are against all other models.

This discussion ignores the fact that we will have to fit some parameters of a family of models. Even in the Poisson case we have to choose the intensity. We side-step the problem of unknown intensity by conditioning on the total number of data points and generating simulations with the same number of points as the data. For Poisson models we can justify this conditioning by an appeal to a conditionality principle. In general all we can say is that conditional inference seems reasonable and that our revised procedures yield valid conditional tests.

There may be other parameters which we will have to fit from the data in a formal or informal way; of course this will affect our significance levels. As a general principle we should only fit families of models with few parameters compared to the number of data points, so the effect should be small. A counsel of perfection is to divide the data into two parts, using one to establish a suitable model which is then tested against the other.

We can use $\hat{p}$ in exactly the same way. With either statistic we are essentially using the combination of a large number of tests at different ordinates which are highly inter-dependent. The suggested procedure is not ideal but needs little human judgement.

Both K and p are essentially cumulative functions; the reader may be wondering why we have chosen to estimate these rather than their densities. One reason is that we would have to smooth these estimates which it seems impossible to do in an objective way. A further reason is that the author's limited experiments suggest that tests based on such smoothed estimates are considerably less powerful than their cumulative analogues. This point is discussed further in the examples.

7. Cell Shapes and Compartments

In this section we model a patch of biological cells in order to provide a solution to the following problem posed by Dr Francis Crick; the biological background is given by Crick and Lawrence (1975). At an early stage in the development of an adult insect there is a roughly circular patch of about 20 two-dimensional contiguous cells. These divide simultaneously, then the patch of daughter cells is split into two roughly equal compartments by an approximately straight line following the cell boundaries. The problem is to find the proportion of sister cell pairs which are in different compartments.

We approach this problem by first analysing the patch of cells at a single point in time. A sample patch of cells (with somewhat idealized boundaries) is shown in Fig. 1. Inspection

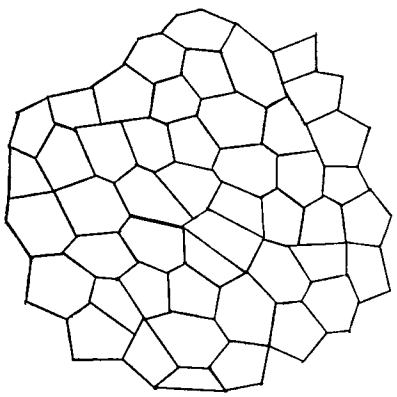
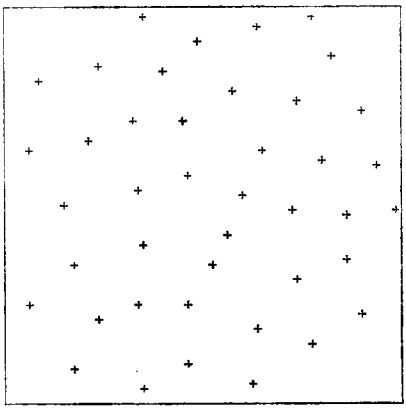
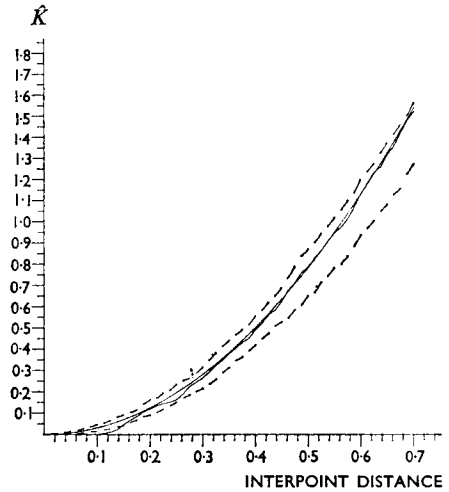


FIG. 1. A typical patch of cells.

of this figure shows that it is reasonable to approximate the cells by the Dirichlet cells of their centres of gravity, except at the edges of the patch where the boundary is undefined. (The Dirichlet cell or Voronoi polygon of a cell centre is the set of points nearer to this centre than to any other.)

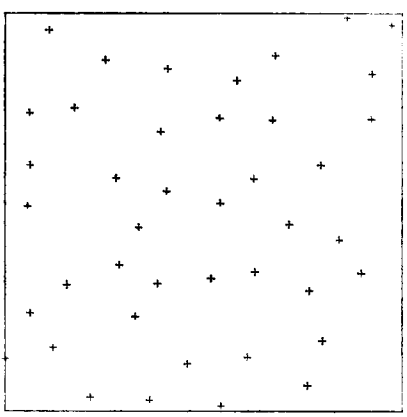
To circumvent problems introduced by biological edge-effects I took a centrally located sample square E containing the 42 cell centres shown in Fig. 2. The units were chosen to make E the unit square. The initial analysis is based on this sample. Fig. 3 shows $\hat{K}$ for the data and K for the Poisson process. These are clearly different up to 0·1 but similar throughout the range (0·2, 0·7). The dashed lines are the envelope of $\hat{K}$ for 99 simulations of the Poisson process conditioned to produce 42 points. The area above the lower envelope (for $0 < t < 0{\cdot}7$) provides the acceptance region of a significance test of size about 7 per cent (estimated by further simulation); the Poisson null hypothesis is rejected; the part of the band in the range (0, 0·3) is shown enlarged on Fig. 5; the form of $\hat{K}$ for the data shows that there are no pairs of cell centres closer than 0·1 apart, which suggests we should fit a hard-core model with inhibition distance $r = 0{\cdot}1$.

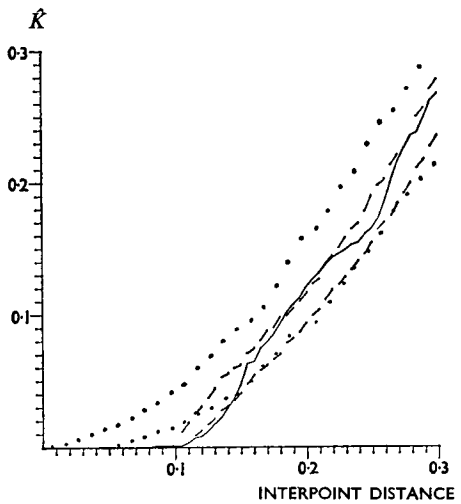




Left. FIG. 2. The sample square containing 42 points derived from the patch of biological cells.

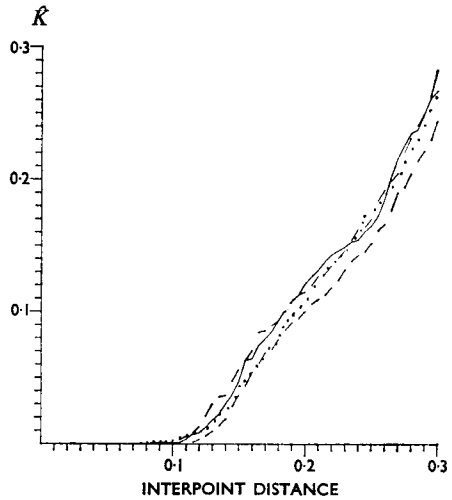
Right. FIG. 3. $\hat{K}$ plotted for the cell data and the envelope of the plots of $\hat{K}$ for 99 simulations of the Poisson process conditioned to produce 42 points. The smooth curve is the theoretical value of K for the Poisson process.

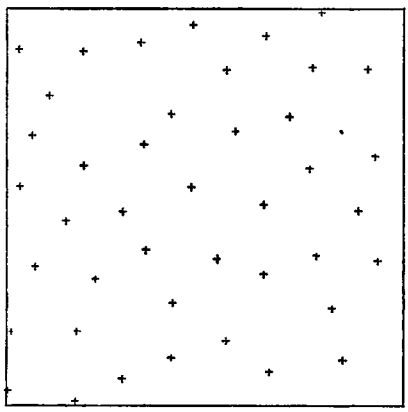




Left. FIG. 4. A typical simulation of the third hard-core model with periodic boundary conditions, containing 42 points.

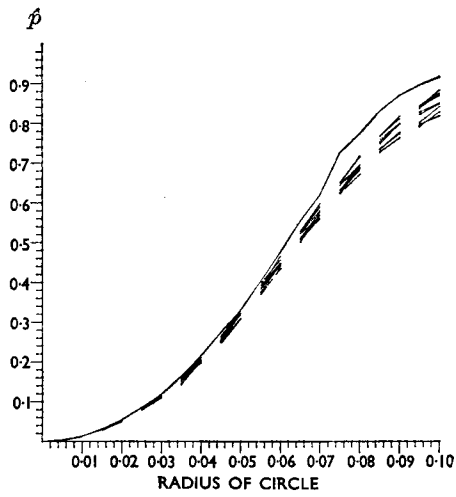
Right. FIG. 5. $\hat{K}$ plotted for the cell data and the envelopes of the plots of 49 simulations of the hard-core model (the dashed curves) and of 99 simulations of the Poisson process (the dotted curves); all samples contained 42 points.

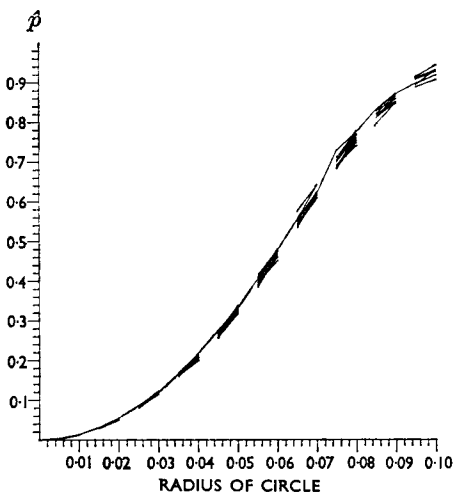




Left. FIG. 6. $\hat{K}$ plotted for the cel data and the envelope of 20 simulations of the pairwise interaction process with periodic boundary conditions. The dotted curve is $\hat{K}$ for another sample of 89 cell centres.

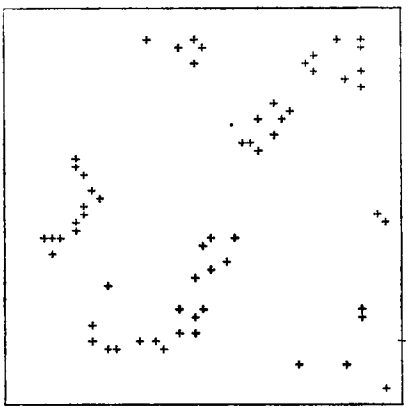
Right. FIG. 7. A typical simulation of the pairwise interaction process with repulsion in the range 0·10–0·14.

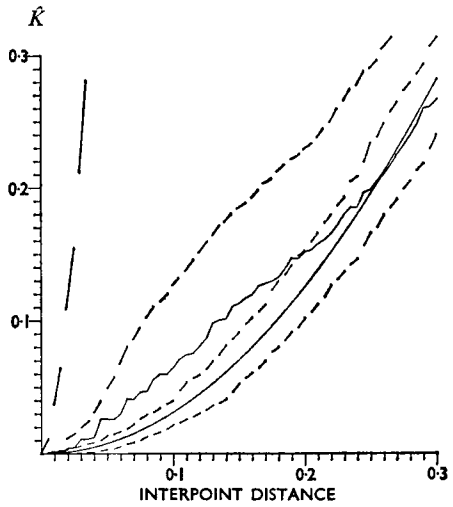




Left. FIG. 8. Plots of $\hat{p}$ for the cell data and 10 simulations (the dashed lines) of the hard-core model.

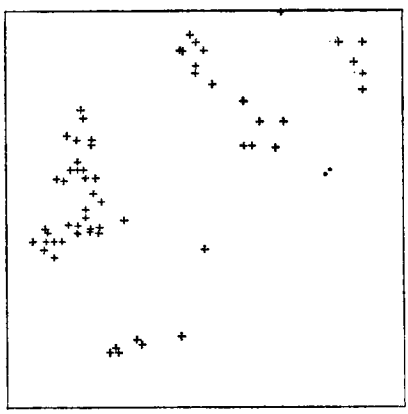
Right. FIG. 9. Plots of $\hat{p}$ for the cell data and 10 simulations (the dashed lines) of the pairwise interaction model.

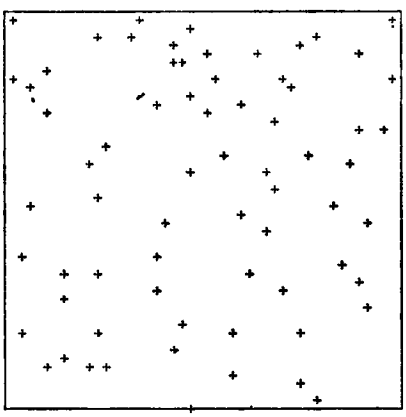




Left. FIG. 10. The positions of 62 Redwood seedlings.

Right. FIG. 11. The solid curves are the plots of $\hat{K}$ for the Redwood data and of K for the Poisson process (the parabolic curve). The lower and upper pairs of dashed curves are the envelopes of the plots of $\hat{K}$ for 99 simulations of the Poisson process and 20 samples of Strauss' model.





Left. FIG. 12. A typical simulation of Strauss' model, containing 62 points.

Right. FIG. 13. The centres of 69 towns in a 40-mile square in Spain.

For this value of r, Matérn's first process has a maximum intensity of about 11·7; his second model is biologically inappropriate. We fit the third hard-core model with periodic boundary conditions on E. A typical simulation of this process with 42 points is shown in Fig. 4, and Fig. 5 shows $\hat{K}$ for the data compared with 49 simulations of this process. It is clear that the formal significance test rejects the hard-core model (at an estimated 20 per cent), on the basis of Fig. 5 one might be prepared to accept it. However, the simulations appear less regular than the data (compare Figs 2 and 4), having "holes" which would give rise to some unacceptably large Dirichlet cells.

Thus we will continue to search for a satisfactory model; we try a model with more general pairwise interactions. Unfortunately there seems to be no general way to estimate the interaction function h from the data. The curves of Fig. 5 show that the data have fewer pairs of points with distances in the range (0·1, 0·14) than the simulations of the hard-core model. This suggested we should introduce a repulsive force in this range; it is convenient to make this force constant and take $h(s) = \exp\{k(s-0{\cdot}14)\}$ in this range, $h(s) = 0$ for $s \leqslant 0{\cdot}1$, and $h(s) = 1$ for $s \geqslant 0{\cdot}14$. There is a point process defined by h for each positive value of k. The "constant" in the product formula for f is an unknown function of k, hence so is the likelihood function for k and no formal statistical method may be used to estimate k. The value $k = 80$ was chosen to obtain a good fit of $\hat{K}$ for the simulations to $\hat{K}$ for the data. Fig. 6 shows the plots of $\hat{K}$ for the data and the envelope of 20 simulations. Fig. 7 shows a typical simulation; notice that the "holes" have disappeared. This model corresponds to cells with a hard core and soft edge.

It might be objected that it is unreasonable to fit such a complex model with so many parameters to just one small sample. This objection is countered by the plot of $\hat{K}$ on Fig. 6 for another sample of a patch of cells containing 89 centres. The agreement between the two samples and the simulations of our model is sufficient for us to accept this as our model.

The existence of "holes" shows up well in the analysis based on $\hat{p}$. Figs 8 and 9 show $\hat{p}$ for the cell data and ten simulations (the dashed lines) of the hard core and final models respectively. The test set B here was a circle. The superior fit of the pairwise interaction model is clearly shown. The small number of simulations is due to the expense of this procedure in computer time; these figures took about 5 minutes compared with 30 seconds for Fig. 5.

7.1. *The Next Generation*

The exact nature of the transition of a patch of cells from one generation to the next is unknown. We seek a plausible model which will preserve the statistical properties of the pattern of the cell centres. We assume that each cell splits into two daughter cells and then we rescale all distances by $\sqrt{2}$ to allow for the expansion of the patch following the doubling of the number of cells.

In the first phase we replace each cell centre by two centres at the ends of a line segment centred on the parent cell centre. This segment has a uniformly distributed orientation independent for each parent, and a length equal to half the average cell diameter. This models the division of a cell along a "diameter". This set of $2N$ cell centres is then allowed to move under forces which are constant direct repulsions between each pair of neighbouring centres (less than 0·1 apart) until there are no neighbours.

The second phase is an approximation to the diffusion discussed in Section 4. We replace the diffusion by a random walk performed by the cells which are assumed to have hard cores which behave as perfectly elastic spheres should they collide. The motion proceeds in discrete steps. First each centre is moved a distance d_1 away from any other centre which was between 0·10 and 0·14 distant, and then a distance d_2 in a random uniformly distributed direction independently for each centre. If this perturbation causes the hard cores to collide then they bounce apart elastically. (This process is closely related to the Manx motion of Kendall, 1974b).

The essential parameter of this second phase is the relationship between d_1 and d_2 which must be connected with the parameter k of our model. Consider just two points distance d apart away from the boundary of E. Let $D = 2(d_1 + d_2)$; we will consider the distribution of d in the range $(0{\cdot}10 + D, 0{\cdot}14 - D)$. Let $p(d)$ denote the equilibrium density for d which must satisfy

$$p(d) = \int p(d - 2d_1 - 2d_2 \cos\theta)\, d\theta/2\pi;$$

this is satisfied by $h(d) = c \exp\{k(d - 0{\cdot}14)\}$ if

$$\exp(2kd_1) = I_0(2kd_2) \sim 1 + (kd_2)^2,$$

where I_0 is the modified Bessel function of the first kind of order zero. Since the equilibrium distribution conditional on two points must be proportional to h for our model, we assume the approximate relation $d_2^2 = 2d_1/k$. This process was simulated with $d_2 = 1/200$, $d_1 = 1/1{,}000$; the resulting pattern of cell centres after 50 steps produced functions $\hat{K}$ which accorded well with those for the original data and our model. This process is an indecomposable Markov chain which thus converges to its equilibrium distribution (in the sense described for spatial birth-and-death processes).

7.2. *Compartments*

This modelling enables us to mark on a simulated patch of cells which pairs of cells are sisters. There seems to be no way in which this could have been done directly on Fig. 1. We divide the patch of daughter cells by a random straight line passing near the centre of the patch; cells are allocated to a compartment according to the side of this line on which their centres fall. Suppose we have a roughly circular patch of $2N$ daughter cells. The length of the dividing line between compartments will be proportional to $N^{\frac{1}{2}}$, which suggests that the average proportion of sister cell pairs which are in different compartments will be $cN^{-\frac{1}{2}}$ where c is approximately constant. From simulations on $N = 40$, c was estimated as about 0·4.

Another model for the formation of compartments which yielded the estimates quoted by Crick and Lawrence is discussed in Ripley (1976c).

8. OTHER EXAMPLES

8.1. *A Redwood Plot*

The data for this example were taken from Strauss (1975). The 62 points are shown in Fig. 10; they were taken from the clustered group of seedlings of his Fig. 1. This part of the plot was not square so for computational convenience I took a square mainly within the region of supposed clustering.

The plot of $\hat{K}$ compared with the envelope of 99 simulations of the Poisson process producing 62 points is shown in Fig. 11. Again we choose our units to make E the unit square. Strauss suggested that the plants might interact up to $r = 0{\cdot}08$ (about 6 feet), so the scale of the plot was chosen to cover (0, 0·3). The ordinates at 0·08 provide an exact 1 per cent one-sided significance test which clearly rejects the Poisson model. The whole upper envelope gives a test of size about 3 per cent.

Strauss fitted the density $f(x) = ac^{s(x)}$ by an approximation method and estimated $\log c$ as 1·03. Samples from this density were simulated by the method described in Section 5; a typical sample is shown in Fig. 12. It is evident that this model does not fit the data, as the plots of $\hat{K}$ on Fig. 11 also show. Simulations of this model for other values of c and r showed that it cannot be fitted to the data. It seems to produce tightly bound clusters rather than the loose clusters in the data.

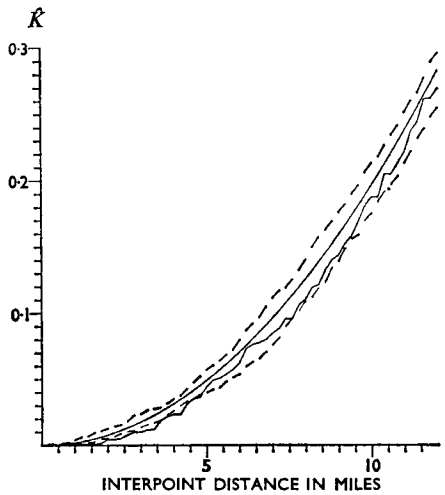
Notice that the plot of $\hat{K}$ for the data crosses K for the Poisson process at about 0·25. In fact $\hat{K}$ and K are very close throughout (0·25, 0·7). For a cluster process the function $K(t) - \pi t^2$

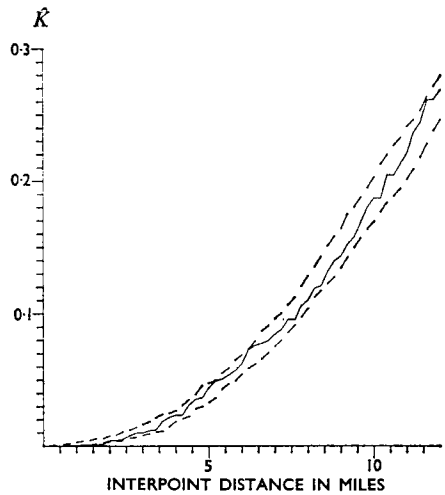
is an increasing function if the parent process is Poisson (because $\text{cov}\{Z(E), Z(F)\} \geq 0$ for any E and F). Thus no such cluster process can fit the data. It appears that there is both clustering and inhibition between clusters. In addition, $\hat{K}$ for the data is zero up to a distance corresponding to about 1 foot, suggesting some local inhibition (which is certainly not statistically significant).

I have been unable to find any model which fits these data. The number of *clusters* involved is too small to merit detailed analysis.

8.2. *Central Places*

Geographers are interested in the distribution of towns in homogeneous regions; Cliff and Ord (1975) give a survey and further references. The Poisson random pattern is usually contrasted with some form of uniform spacing. Fig. 13 shows the centres of 69 towns in a geographically homogeneous square on the Spanish plateau. Fig. 14 shows the plots of $\hat{K}$





Left. FIG. 14. $\hat{K}$ for the Spanish towns compared with K for the Poisson process and the envelope of $\hat{K}$ for 99 simulations of that process.

Right. FIG. 15. $\hat{K}$ for the Spanish towns and the envelope of 20 simulations of Strauss' model with 69 points, $c = 0{\cdot}5$ and an inhibition distance of 3·5 miles.

for the data and 99 simulations of the Poisson process. Here we are interested in a one-sided test so we take the acceptance region to be the set above the lower envelope; this has an estimated size of 3 per cent. There is a marginal deviation from the Poisson null hypothesis at an interpoint distance of about 3·5 miles.

Glass and Tobler plotted a histogram estimate of $g(r)/2\pi r$ with a bin size of 1 mile and an edge correction similar to ours. They failed to consider the sampling fluctuations of their estimate but claimed to reject the Poisson hypothesis and to fit a hard-core model with $r = 2R$ equal to 3·46 miles. This is absurd since Fig. 14 shows that there are pairs of towns closer than this distance (as does their Fig. 2!). Simulations showed that the unsmoothed histogram is too rough to be useful and that the non-Poisson behaviour is swamped by the sampling fluctuations (the details are given in Ripley, 1976c).

The shape of $\hat{K}$ for the data suggested that we should fit an interaction model with interaction distance $r = 3{\cdot}5$ miles. Strauss' model was fitted by comparing $\hat{K}$ for the data with simulations. The plots for $c = 0{\cdot}5$ are shown in Fig. 15; this model fits well, the choice of c being not at all critical.

The analysis of the towns in Winsconsin used by Brush (1953) gave very similar results with an interaction range of about 6 miles.

8.3. *Galaxies*

The observed positions of galaxies in the celestial sphere and their estimated positions in space form vast data-sets of which parts may be regarded as homogeneous. Parts of various galaxy catalogues have been analysed by Peebles and his colleagues. They estimate the two-point correlation function w which is interpreted in the Appendix as $(g/\lambda^2 - 1)$. A power law $w(r) = cr^{-\alpha}$ was fitted to a histogram estimate of w (Peebles, 1974). There appears to be no general way to decide when a postulated λ and K summarize the first and second moments of a model. We note that for a Neyman–Scott cluster model $c(r)\,w(r)$ is a power law proportional to the density of the distribution between two cluster points, which cannot be of power law form (this is for galaxies in space). Thus no Neyman–Scott process corresponds to this choice of K.

Peebles and Groth (1975) considered three-point correlation functions for a point process on a sphere. This is a symmetric function z of three arguments such that

$$E\{Z(A)Z(B)Z(C)\}$$

$$= \lambda^3 \int_{A\times B\times C} \{1 + w(\theta_1 - \theta_2) + w(\theta_2 - \theta_3) + w(\theta_3 - \theta_1) + z(\theta_1 - \theta_2, \theta_2 - \theta_3, \theta_3 - \theta_1)\}\, d\theta_1\, d\theta_2\, d\theta_3$$

for all disjoint sets A, B and C. The function

$$\lambda^3\{1 + w(\theta_1) + w(\theta_2) + w(\theta_3) + z(\theta_1, \theta_2, \theta_3)\}$$

is a density of the reduced third moment measure (cf. Krickeberg, 1973; Ripley, 1976a). This is a function of three arguments and is difficult to visualize. Nevertheless Peebles and Groth fitted a function $z(\theta_1, \theta_2, \theta_3)$ proportional to

$$w(\theta_1)\,w(\theta_2) + w(\theta_2)\,w(\theta_3) + w(\theta_3)\,w(\theta_1).$$

It is difficult to see what this would mean in terms of a possible model; the Gauss–Poisson processes (Newman, 1970; Milne and Westcott, 1972) have $z = 0$ but a similar form for the analogous four-point correlation function.

9. CONCLUSIONS

Our examples have indicated the usefulness of the statistics K and p in testing the fit of models. Analytical information on even their asymptotic distributions would be enlightening and might suggest more economical ways of generating significance tests. It might also be possible to obtain some results on the power of these tests; however, it would seem to be optimistic to hope that this might be possible for realistic alternatives. There is a shortage of suitable models, although the interaction models described here have broadened the field. These models would seem to be suitable for biological competition situations such as the nesting territories of birds (cf. the maps in Tubbs, 1974; and in Brown, 1976); indeed in this case the pattern is thought to result from pairwise competition.

The test set method of mathematical morphology has never before to my knowledge been applied to point processes. The theoretical unification between point processes and more general random sets expounded in Ripley (1976b) will I hope encourage applications; it would also be interesting to analyse patterns such as Fig. 1 directly as a random set.

ACKNOWLEDGEMENTS

I am indebted to all who have commented on this work, especially Julian Besag, Pierre Delfiner, Peter Diggle, Frank Kelly, John Kent, Bernard Silverman and Simon White. Francis Crick kindly provided me with the cell data; David Kendall introduced me to this problem and he and the referees made many useful comments on an earlier draft.

The computations were carried out by Fortran programs on the IBM 370/165 of the University of Cambridge Computing Service. Most of this work was carried out during the tenure of a Science Research Council studentship at the Statistical Laboratory, Cambridge.

APPENDIX

The interpretations of the reduced second moment measure $\lambda^2 K$ given in the text were obtained independently by Krickeberg (1976; lectures of 1974) and the author. We define the mth moment measure μ^m of a point process Z by

$$\mu^m\left(\prod_1^m A_i\right) = E\left\{\prod_1^m Z(A_i)\right\},$$

where A_i are bounded Borel sets. Let $P(\cdot \,\|\, x)$ be a Palm probability (Papangelou, 1974); this is a probability for each x such that

$$\int_E P(A \,\|\, x)\, d\mu^1(x)$$

exists and equals $\int_A Z(E)\, dP(z)$ for any Borel set E and measurable event A. We interpret $P(\cdot \,\|\, x)$ as the probability "given a point at x". It follows immediately that

$$\mu^m\left(\prod_1^m A_i\right) = \int_{A_1} \mu^{m-1}\left(\prod_2^m A_i \| x\right) d\mu^1(x).$$

Now suppose our point process is a *model*; we choose the unique invariant Palm probability considered by Papangelou. Then we have

$$\begin{aligned} K(t) &= \lambda^{-2}\mu^2[\{(x,y): x \in I, 0 < d(x,y) < t\}] \\ &= \lambda^{-1}\int_I \mu^1(b(x,t) \,\|\, x)\, dv(x) = \lambda^{-1}\mu^1(b(0,t) \,\|\, 0), \end{aligned}$$

where I is the unit cube. This is our interpretation (b).

Let K_0 be the distribution function of the measure representing $\nu \times \nu$. Suppose $\mu^2 \ll \nu \times \nu$ on $E_2 = \{(x,y): x \neq y\}$. (We know from Krickeberg (1973) that μ^2 is singular on the diagonal.) Then K is absolutely continuous with respect to K_0, and if $g = \lambda^2 dK/dK_0$ then $f(x,y) = g\{d(x,y)\}$ is a version of $d\mu^2/d(\nu \times \nu)$ on E_2. The factor $c(t) = dK_0/dt$, so

$$g(t) = (\lambda^2 dK/dt)/c(t).$$

Suppose A and B are cubes decreasing to the distinct points x and y respectively. Then

$$\mu^2(A \times B)/\nu(A)\,\nu(B) \to f(x,y), \quad \text{a.e.}$$

Thus if $d(x,y) = t$ then

$$g(t) = \lim_{A \downarrow x\, B \downarrow y} \frac{E\{Z(A)\,Z(B)\}}{\nu(A)\,\nu(B)}$$

If we knew $E\{Z(A)\,Z(B)\} \sim P\{Z(A) > 0, Z(B) > 0\}$ we could regard $g(t)$ as the joint probability density for the occurrence of a pair of points distance t apart. Whereas we always have $E\{Z(A)\} \sim P\{Z(A) > 0\}$ (e.g. Leadbetter, 1972) this formula does *not* follow from our assumptions; a counter-example can be constructed by taking a Poisson cluster process on the group of rigid motions and a daughter process defined by the random set

$$(\{0, 1 + n^2 X: n = 1, 2, \ldots\} \cap [0, 2]) \times \{0\}^{d-1},$$

where X is a uniformly distributed random variable on $(0, 1)$. This fact is often overlooked.

Under these assumptions we have

$$P\{Z(A) > 0, Z(B) > 0\} \sim \lambda^2\{1 + w(t)\}\,\nu(A)\,\nu(B)$$

for "small" sets A and B about t apart, where $w = (g/\lambda^2 - 1)$ is Peebles' "two-point correlation function". Similar assumptions must be made to interpret the three-point analogue.

REFERENCES

BARTLETT, M. S. (1963). The spectral analysis of point processes (with Discussion). *J. R. Statist. Soc.* B, **25**, 264–296.

—— (1964). The spectral analysis of two-dimensional point processes. *Biometrika*, **51**, 299–311.

—— (1971). Two-dimensional nearest-neighbour systems and their ecological application. In *Statistical Ecology*, Vol. I (G. P. Patil, ed.), pp. 179–194. Pennsylvania University Press.

—— (1974). The statistical analysis of spatial pattern. *Advances Appl. Prob.*, **6**, 336–358.

—— (1976). *The Statistical Analysis of Spatial Pattern.* London: Chapman and Hall.

BESAG, J. (1974). Spatial interaction and the statistical analysis of lattice systems (with Discussion). *J. R. Statist. Soc.* B, **36**, 192–236.

BESAG, J. and GLEAVES, J. T. (1974). On the detection of spatial pattern in plant communities. *Bull. Inst. Int. Statist.*, **45**, 153–158.

BREIMAN, L. (1968). *Probability.* Reading, Mass.: Addison-Wesley.

BRILLINGER, D. R. (1976). Estimation of the second-order intensities of a bivariate stationary point process. *J. R. Statist. Soc.* B, **38**, 60–66.

BROWN, L. (1976). *British Birds of Prey.* London: Collins.

BRUSH, J. E. (1953). The hierarchy of central places in Southwestern Wisconsin. *Geogr. Rev.*, **43**, 380–402.

CLIFF, A. D. and ORD, J. K. (1975). Model building and the analysis of spatial pattern in human geography (with Discussion). *J. R. Statist. Soc.* B, **37**, 297–348.

CRICK, F. H. C. and LAWRENCE, P. A. (1975). Compartments and polyclones in insect development. *Science*, **189** (4200), 340–347.

DIGGLE, P. J. (1975). Robust estimation using distance methods. *Biometrika*, **62**, 39–48.

DIGGLE, P. J., BESAG, J. E. and GLEAVES, J. T. (1976). Statistical analysis of spatial patterns by means of distance methods. *Biometrics*, **32**, 659–667.

FISHER, L. (1972). A survey of the mathematical theory of multidimensional point processes. In *Stochastic Point Processes* (P. A. W. Lewis, ed.), pp. 468–513. New York: Wiley.

GLASS, L. and TOBLER, W. R. (1971). Uniform distribution of objects in a homogeneous field: cities on a plain. *Nature*, **233** (5314), 67–68.

GOLDMAN, J. R. (1971). Infinitely divisible point processes in R^n. *J. Math. Anal. Appl.*, **17**, 133–146.

GRANDELL, J. (1972). Statistical inference for doubly stochastic Poisson processes. In *Stochastic Point Processes* (P. A. W. Lewis, ed.), pp. 90–121. New York: Wiley.

GRENANDER, U. (1973). Statistical geometry; a tool for pattern analysis. *Bull. Amer. Math. Soc.*, **79**, 829–856.

—— (1977). *Pattern Analysis; Lectures on Pattern Theory*, Vol. 2. Berlin: Springer-Verlag.

HAMMERSLEY, J. M. (1972). Stochastic models for the distribution of particles in space. *Suppl. Adv. Appl. Prob.*, 47–68.

—— (1975). Rumination on infinite Markov systems. In *Perspectives in Probability and Statistics* (J. Gani, ed.), pp. 195–200. New York: Academic Press.

HILL, M. O. (1973). The intensity of spatial pattern in plant communities. *J. Ecol.*, **61**, 225–236.

HOLGATE, P. (1965a). Some new tests of randomness. *J. Ecol.*, **53**, 261–266.

—— (1965b). Tests of randomness based on distance methods. *Biometrika*, **52**, 345–353.

—— (1965c). The distance from a random point to the nearest point of a closely packed lattice. *Biometrika*, **52**, 261–263.

HOPE, A. C. A. (1968). A simplified Monte Carlo significance testing procedure. *J. R. Statist. Soc.* B, **30**, 582–598.

KELLY, F. P. and RIPLEY, B. D. (1976). A note on Strauss' model for clustering. *Biometrika*, **63**, 357–360.

KENDALL, D. G. (1974a). An introduction to stochastic geometry. In *Stochastic Geometry* (E. F. Harding and D. G. Kendall, eds.), pp. 1–9. London: Wiley.

—— (1974b). Pole-seeking Brownian motion and bird navigation (with Discussion). *J. R. Statist. Soc.* B, **36**, 365–402.

KLEIN, J. C. and SERRA, J. (1972). The texture analyser. *J. Micr.*, **95**, 349–356.

KRICKEBERG, K. (1973). Moments of point processes. In *Stochastic Geometry* (E. F. Harding and D. G. Kendall, eds.), pp. 89–113. London: Wiley.

—— (1976). The statistics of random processes. Lecture notes. Hanoi: Societé Mathématique du Vietnam. (In Vietnamese.)

LEADBETTER, M. R. (1972). On the basic results of point process theory. *Proc. 6th Berkeley Symp. Math. Statist. & Prob.*, Vol. III, pp. 449–462.

MATÉRN, B. (1960). Spatial variation. *Meddelanden från Statens Skogsforskningsinstitut*, **49**, 5.

MATHERON, G. (1967). *Éléments pour une Théorie des Milieux Poreux.* Paris: Masson.

MATHERON, G. (1975). *Random Sets and Integral Geometry.* New York: Wiley.

MATTHES, K. (1972). Infinitely divisible point processes. In *Stochastic Point Processes* (P. A. W. Lewis, ed.), pp. 384–404. New York: Wiley.

MEAD, R. (1974). A test for spatial pattern at several scales using data from a grid of contiguous quadrats. *Biometrics*, **30**, 295–307.
MILNE, R. K. and WESTCOTT, M. (1972). Further results on Gauss–Poisson processes. *Advances Appl. Prob.*, **4**, 151–176.
NEWMAN, D. S. (1970). A new family of point processes which are characterized by their second-moment properties. *J. Appl. Prob.*, **7**, 338–358.
NEYMAN, J. and SCOTT, E. L. (1972). Processes of clustering and applications. In *Stochastic Point Processes* (P. A. W. Lewis, ed.), pp. 646–681. New York: Wiley.
PALOHEIMO, J. E. (1971). On a theory of search. *Biometrika*, **58**, 61–75.
PAPANGELOU, F. (1974). On the Palm probabilities of processes of points and processes of lines. In *Stochastic Geometry* (E. F. Harding and D. G. Kendall, eds.), pp. 114–147. London: Wiley.
PEEBLES, P. J. E. (1974). The nature of the distribution of galaxies. *Astron. & Astrophys.*, **32**, 197–202.
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. *Astrophys. J.*, **196**, 1–11.
PRESTON, C. J. (1974). *Gibbs States on Countable Sets*. London: Cambridge University Press.
—— (1976a). Spatial birth-and-death processes. *Bull. Inst. Int. Statist.*, **46** (in press).
—— (1976b). *Random Fields*. Lecture Notes in Math. **534**. Berlin: Springer-Verlag.
RIPLEY, B. D. (1976a). The second-order analysis of stationary point processes. *J. Appl. Prob.*, **13**, 255–266.
—— (1976b). Locally finite random sets: foundations for point process theory. *Ann. Prob.*, **4**, 983–994.
—— (1976c). *Stochastic Geometry and the Analysis of Spatial Pattern*. Ph.D. Thesis, Cambridge University.
RIPLEY, B. D. and KELLY, F. P. (1977). Markov point processes. *J. London Math. Soc.* (in press).
ROGERS, C. A. (1964). *Packing and Covering*. London: Cambridge University Press.
RUELLE, D. (1969). *Statistical Mechanics*. New York: Benjamin.
—— (1970). Superstable interaction in classical statistical mechanics. *Commun. Math. Phys.*, **18**, 127–159.
SERRA, J. (1972). Stereology and structuring elements. *J. Micr.*, **95**, 93–103.
SILVERMAN, B. W. (1976). Limit theorems for dissociated random variables. *Advances. Appl. Prob.*, **8**, 806–819.
STRAUSS, D. J. (1975). A model for clustering. *Biometrika*, **62**, 467–475.
TUBBS, C. R. (1974). *The Buzzard*. Newton Abbot, Devon: David and Charles.
WATSON, G. (1973). Mathematical morphology. Tech. Report no. 21, Series 2, Dept. of Statistics, Princeton University.

DISCUSSION ON DR RIPLEY'S PAPER

Professor PHILIP HOLGATE (Birkbeck College, University of London): In his recent work and in the paper that he has presented this afternoon Dr Ripley has combined three of the main interests motivating this Society. He has contributed to the analytic probability theory of random point patterns. He has developed and studied models reflecting interesting physical and biological phenomena and he has considered the statistical problem of analysing data arising from them.

Fifteen years ago it was a common if not quite justified complaint among quantitative plant ecologists that statisticians ignored their problems. They were left alone to deal with a range of special techniques including nearest neighbour methods and the analysis of variance of blocks of continguous quadrats that did not form part of everyday statistical currency and were consequently not supported by any accumulated theoretical knowledge and practical experience. As well as presenting his own substantial contribution to the subject, Dr Ripley's paper reminds us of the radical change that has taken place since then. An ecological botanist who is prepared to consider the pattern of his plants as a realization of a point process will have no difficulty in finding papers discussing the problem although he may have difficulty in understanding them. It is important that a genuine effort is made to convey findings such as those Dr Ripley has presented this afternoon to the practical scientists involved.

One problem raised by the paper is the adequacy of second-order analysis of point patterns. Such methods belong properly to the domain of processes that are linear and normal, which point patterns are not, as well as stationary. In another paper Dr Ripley (1966a) refers to psychological evidence that the human eye can only see the second-order aspects of a pattern, but that seems to me to be an added incentive to try to analyse the finer aspects by mathematical statistics.

Methods of doing this are indicated in the work cited in section 8.3. The term $z(\theta_1, \theta_2, \theta_3)$ is related to that part of the probability of finding points at angles θ_1, θ_2, θ_3 that cannot be explained by the overall density or the pairwise interaction between infinitesimal areas at those angles, and should