Classification of Mixtures of Spatial Point Processes via Partial Bayes Factors

Daniel C. I. WALSH and Adrian E. RAFTERY

Motivated by the problem of minefield detection, we investigate the problem of classifying mixtures of spatial point processes. In particular we are interested in testing the hypothesis that a given dataset was generated by a Poisson process versus a mixture of a Poisson process and a hard-core Strauss process. We propose testing this hypothesis by comparing the evidence for each model by using *partial Bayes factors*. We use the term partial Bayes factor to describe a Bayes factor, a ratio of integrated likelihoods, based on only part of the available information, namely that information contained in a small number of functionals of the data. We applied our method to both real and simulated data, and considering the difficulty of classifying these point patterns by eye, our approach overall produced good results.

Key Words: Minefield detection; Poisson process; Strauss process.

1. INTRODUCTION

We investigate the problem of comparing competing models for spatial point process data. In particular we are interested in testing the hypothesis that the data were generated by a Poisson process (i.e., complete spatial randomness) versus a mixture of a Poisson process and an inhibited process. The motivation behind this methodology is the problem of minefield detection.

The U.S. Marine Corps developed the Coastal Battlefield Reconnaissance and Analysis (COBRA) program to detect minefields in coastal areas before troop deployment. The program consists of an unmanned aerial vehicle imaging a possible minefield. This image can then be processed into a set of object locations. Each object is either a mine or can be considered clutter or noise. Various problems can be addressed with these processed images. Muise and Smith (1992), Dasgupta and Raftery (1998), and Byers and Raftery

Daniel C. I. Walsh is Lecturer, Institute of Information and Mathematical Sciences, Massey University, Albany Campus, Private Bag 102-904, North Shore Mail Centre, Auckland, New Zealand (E-mail: d.c.walsh@massey.ac.nz). Adrian E. Raftery is Professor of Statistics and Sociology, University of Washington, Seattle, WA 98195-4320 (E-mail: raftery@stat.washington.edu; Web: www.stat.washington/raftery).

Journal of Computational and Graphical Statistics, Volume 14, Number 1, Pages 139–154
DOI: 10.1198/106186005X27149

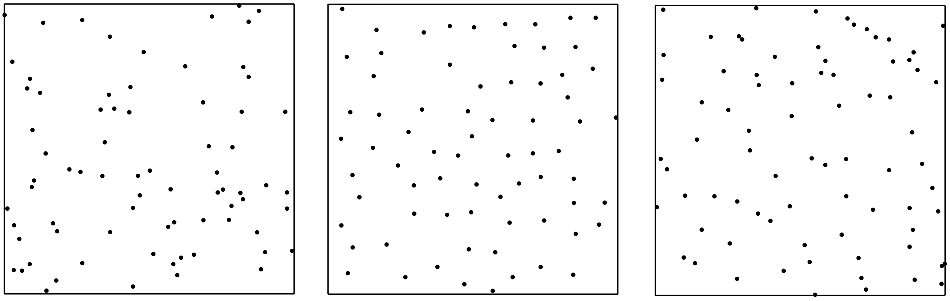


Figure 1. Examples of Simulated Spatial Point Patterns from Different Models.

(1998) considered the problem of localizing the mined region within the image of a large area; Cressie and Lawson (1998), Lake, Sadler, and Casey (1997), and Walsh and Raftery (2002) considered identifying each object as a mine or non-mine. In this article, we assume that, if present, the minefield extends throughout the study region; our goal is to classify the region as a minefield with clutter, or pure clutter.

The mines are assumed to be laid out in such a way that two mines are unlikely to be close together. A hard-core Strauss process is one way to model this constraint. The noise points are assumed to follow a Poisson process throughout the study region. The inherent difficulty of distinguishing between pure Poisson processes, and different mixtures of the Poisson and Strauss process can be seen in Figures 1(a), 1(b), and 1(c). Here, realizations of three different point processes are shown, and yet the human eye detects few visual cues to reliably differentiate them. However, statistical techniques can produce surprisingly good results.

The problem of comparing a simple model for inhibition or clustering versus complete spatial randomness was considered by Diggle (1983) and Cressie (1993). The problem of classifying mixtures of spatial point processes was tackled by Raghavan, Goel, and Ghosh (1997, 1998). Their approach was to develop a supervised pattern recognition scheme using functionals based on nearest neighbor distances, second-order statistics, and spatial tessellations. We propose comparing the evidence for each model directly by using *partial Bayes factors*. We use the term partial Bayes factor to describe a Bayes factor based on only part of the available information, namely the information contained in a small number of functionals of the data. In other words, our inferences are based on statistics that are informative and tractable, but that are not sufficient statistics. The likelihood of these statistics is estimated via simulation.

Our use of "partial Bayes factors" should not be confused with the sense in which O'Hagan (1995) and others used the phrase partial Bayes factors; they use likelihoods that include only "partial information" from the data in the sense of actually excluding a subset of data points (this subset is used in computing a prior).

In the next two sections we describe the spatial point process models we use and formulate the minefield problem as a hypothesis testing problem. We then briefly review Bayes factors and define partial Bayes factors in Section 4. In Section 5 we discuss possible summary statistics one could use. Results of applying our method to simulated and real data are presented in Section 6 and are discussed in Section 7.

2. POINT PROCESS MODELS

First we give some notation. To avoid possible ambiguities associated with the word "point," we shall refer to locations of objects as *events*, and let the word *point* refer to any location in the sample space. The sample space or study region will be denoted by $A \in \mathbb{R}^2$, $|A|$ will denote the area of this region, and A^n will denote the space of all locations of n points in A. We will consider each event to be one of two types, "noise events" and "mines." Let N be the total number of events, n_0 be the number of noise events, and m be the number of mines. Let d_{ij} be the distance between the ith and jth events, and let $d_i = \min_{j,j \neq i} d_{ij}$. Let the spatial location of the ith event be denoted by $y_i = \{y_i^1, y_i^2\}$, and the location of all events be $Y = (y_1, \ldots, y_N) \in A^N$. We shall condition on the number of events, N, and the study region, A, throughout.

2.1 Noise Process

As was mentioned in the introduction, the noise events are considered to be scattered randomly throughout A. Under the hypothesis that no minefield is present, and given that we are conditioning on the number of events in A, the distribution of Y is uniform over A^N, that is,

$$P_u(Y) = \frac{1}{|A|^N}.$$

We call this a *uniform process*, and we denote it by $Y \sim \text{Uniform}(N, A)$.

2.2 Minefield Process

The mines are assumed to be spread evenly over A. This implies that the minefield process displays *inhibition*. A simple model for an inhibited process is the *Strauss* process (Strauss 1975; Kelly and Ripley 1976). The likelihood for the Strauss process is:

$$P_s(Y \mid \theta) = C \prod_{i<j} g_\theta(d_{ij}),$$

where $g(\cdot)$ is the *interaction* function, given by:

$$g_\theta(r) = \begin{cases} \gamma, & 0 \leq r < \rho, \qquad \text{where} \qquad \gamma \in [0, 1] \\ 1, & r \geq \rho, \end{cases}$$

and where the parameters of the Strauss process are denoted by $\theta = \{\rho, \gamma\}$. The interaction function defines the relationship between pairs of events. The extent of the interactions between two events is controlled by ρ, and the nature of these interactions is determined by γ. If $\gamma = 0$, the process is known as a *hard-core* process. In this process, two events are forbidden to be within distance ρ of each other. Alternatively, if $\gamma = 1$, the process is simply a uniform process on A. Values of γ between 0 and 1 discourage but do not forbid events to be within distance ρ of each other. Note that the normalizing constant, C, of the Strauss process can be difficult to calculate, especially for processes demonstrating strong inhibition (see Diggle et al. 1994).

2.3 MIXTURE PROCESS

Consider a superposition of a Strauss process upon a uniform process. Let $Y = Y_u \cup Y_s$, where Y_u are the events generated by the uniform process and Y_s are the events generated by the Strauss process. Let $Z = \{Z_1, \ldots, Z_N\}$ be a variable indicating to which group each observation belongs, that is,

$$Z_i = \begin{cases} 0, & \text{if} \quad y_i \in Y_u \\ 1, & \text{if} \quad y_i \in Y_s, \end{cases}$$

for $i = 1, \ldots, N$. Note that $\sum_{i=1}^{N} Z_i = m$, the number of Strauss events (mines). If Z is known, then the likelihood for the mixture process can be written as

$$P_m(Y \mid Z, \theta) = P_u(Y_u \mid Z, \theta) \times P_s(Y_s \mid Z, \theta).$$

If Z is unknown (as would be the case in practice), then we must sum over all the values of Z, multiplied by their respective probabilities, that is,

$$\begin{aligned} P_m(Y \mid \theta) &= \sum_{z \in Z} P_m(Y \mid Z, \theta) \pi(Z \mid \theta) \\ &= \sum_{m=0}^{N} \sum_{z \in Z \mid \sum z = m} P_m(Y \mid Z, \theta) \pi(Z \mid m, \theta) \pi(m \mid \theta). \end{aligned}$$

Given the problem of obtaining the normalizing constant for a Strauss process, this sum is extremely difficult to compute.

3. FORMULATION OF THE MINEFIELD PROBLEM AS A HYPOTHESIS TESTING PROBLEM

We cast the minefield problem in terms of two competing hypotheses. Here we model the minefield process as a hard-core Strauss model. Thus, for a given point pattern Y, the competing hypotheses of interest are:

H_0: No minefield present, $Y \sim$ Uniform (N, A)

H_1: Minefield present, $Y = Y_u \cup Y_s$, where $Y_u \sim$ Uniform(n_0, A), $Y_s \sim$ Strauss$(m, A, \rho, \gamma = 0)$, and $m + n_0 = N$.

3.1 PRIOR SPECIFICATION

Under H_1, there are two unknown model parameters, ρ and m. In a Bayesian framework, one must specify a prior distribution $\pi(\rho, m)$ on ρ and m. The prior could be decomposed in the following ways: (1) $\pi(\rho, m) = \pi(\rho) \times \pi(m)$ (Assuming ρ and m are independent.), (2) $\pi(\rho, m) = \pi(\rho|m) \times \pi(m)$, and (3) $\pi(\rho, m) = \pi(m|\rho) \times \pi(\rho)$.

If good prior information about the number of mines and inhibition distance is available, then the independence assumption of the first prior may be reasonable. However, given that we are conditioning on the study region, A, and the total number of events, there exists a constraint on the maximum separation between two events and the total number of mines in A. Diggle (1983) noted that the maximum proportion of a finite region, A, that can be covered by nonoverlapping discs, of radius ρ, is achieved when the discs are packed in an equilateral triangular lattice. This suggests that the maximum value of ρ, given that there are m points in A, (ignoring edge effects) is:

$$\rho_{\max} = \sqrt{\frac{2}{\sqrt{3}} \frac{|A|}{m}}.$$

This bound will be useful in setting a prior for ρ conditional on m. For instance, if one has only a vague idea of the number of mines, but knows that they are closely packed together, then one could use priors of the following form:

$$\begin{aligned} \pi(m) &= \text{Discrete Uniform}(m_1, m_2) \\ \pi(\rho \mid m) &= \text{Uniform}(\alpha_1 \rho_{\max}, \alpha_2 \rho_{\max}), \text{ where } 0 \le \alpha_1 < \alpha_2 \le 1. \end{aligned}$$

This is the form of the prior distributions we used in our simulation study in Section 6.

4. PARTIAL BAYES FACTORS

In this section we briefly introduce Bayes factors and define what we mean by partial Bayes factors. Consider data Y that are assumed to have arisen under one of the two competing hypotheses, H_0 or H_1. Let θ_i be a q_i-dimensional vector of parameters associated with hypothesis H_i $(i = 1, 2)$, and let $\pi_i(\theta_i \mid H_i)$ denote its prior distribution. Let the probability density of Y given the value of θ_i, that is, the likelihood function, be denoted by $P(Y \mid \theta_i, H_i)$. The Bayes factor for H_1 against H_0 is the ratio of the posterior to the prior odds for H_1 against H_0, namely:

$$\begin{aligned} \mathrm{BF}_{10} &= \frac{P(H_1 \mid Y)}{P(H_0 \mid Y)} \Big/ \frac{P(H_1)}{P(H_0)} = \frac{P(Y \mid H_1)}{P(Y \mid H_0)} \\ &= \frac{\int P(Y \mid \theta_1, H_1) \pi_1(\theta_1 \mid H_1)\, d\theta_1}{\int P(Y \mid \theta_0, H_0) \pi_0(\theta_0 \mid H_0)\, d\theta_0}. \end{aligned}$$

Table 1. Guide for Interpreting Bayes Factors

2 $\log_e(B_{10})$	B_{10}	*Evidence for* H_1
0 to 2	1 to 3	Weak
2 to 5	3 to 12	Positive
5 to 10	12 to 50	Strong
$>$ 10	$>$ 150	Decisive

In other words, the Bayes factor is the ratio of integrated likelihoods. The Bayes factor provides evidence for one hypothesis over another. Kass and Raftery (1995) reviewed the history, development, and use of Bayes factors. A guide for interpreting Bayes factors, proposed by Kass and Raftery (based on Jeffreys 1961), is given in Table 1.

In the mixture models we consider, it is possible to simulate realizations from each model, but it is difficult to write down the likelihood explicitly because the normalizing constant and the group memberships, Z, are unknown. Instead, we use the *partial Bayes factor*, defined as the ratio of integrated likelihoods for a summary statistic, X (or a vector of several summary statistics, X), rather than for the complete data:

$$\mathrm{PBF}_{10} = \frac{P(X \mid H_1)}{P(X \mid H_0)}.$$

This can be written as

$$\mathrm{PBF}_{10} \quad = \quad \frac{\int P(x \mid \theta_1, H_1)\pi_1(\theta_1 \mid H_1)\, d\theta_1}{\int P(x \mid \theta_0, H_0)\pi_0(\theta_0 \mid H_0)\, d\theta_0} = \frac{I_1(x)}{I_0(x)}.$$

We can calculate these integrated likelihoods by quadrature methods or by Monte Carlo integration. If $\theta_i = \{\theta_i^{(1)}, \ldots, \theta_i^{(K)}\}$ is a random sample of size K from the prior under hypothesis i, and $\widehat{P}(X \mid \theta_i^{(j)}, H_i)$ is an estimate of $P(X \mid \theta_i^{(j)}, H_i)$, then the Monte Carlo estimate of I_i is

$$\widehat{I}_i(x) = \frac{1}{K} \sum_{j=1}^{K} \widehat{P}\left(x \mid \theta_i^{(j)}, H_i\right).$$

To obtain the estimated density function $\widehat{P}(X \mid \theta_i^{(j)}, H_i)$, we simulate K point patterns from H_i with parameters $\theta_i^{(j)}$, and calculate their summary statistics. Let these K summary statistics be denoted by $X_i^{(j)}$. A standard density estimation procedure, such as kernel density estimation (Silverman 1986), is then applied to X_i to obtain $\widehat{P}(X \mid \theta_i^{(j)}, H_i)$. For the value of K, in practice we found that $K = 100$ was usually sufficient although occasionally $K = 10{,}000$ was needed to estimate tail densities accurately.

Obviously the selection of X is important. We discuss choices of X in the following. Note that nowhere do we assume that X is univariate. A bivariate or higher dimensional statistic may give better discrimination between the hypotheses. However, this may lead to excessive computation as density estimation in more than one dimension can be difficult.

5. SUMMARY STATISTICS

The types of summary statistics considered by Raghavan, Goel, and Ghosh (hereafter RGG) for their supervised pattern recognition scheme fell into three main categories: nearest neighbor distances, second-order statistics, and spatial tessellations. In this article we consider the latter two categories.

5.1 SECOND-ORDER STATISTICS

The K-function (Bartlett 1964; Ripley 1976, 1977; Cressie 1993, chap. 8), has been used extensively as an exploratory tool for the analysis of point patterns, in particular their second-order statistics. For a spatial point process of intensity λ, it is defined as:

$$K(r) \quad = \quad \lambda^{-1}\, \mathrm{E}\,(\#\text{ of events within distance } r \text{ of an arbitrary event}).$$

An estimator that corrects for edge effects was given by Ripley (1976):

$$\widehat{K}(r) \quad = \quad \frac{|A|}{N^2} \sum_{i=1}^{N} \sum_{j=1, i \neq j}^{N} w_{ij} 1_{\{d_{ij} < r\}}, \qquad r > 0,$$

where w_{ij} is the proportion of the circumference of a circle centered at event i that passes through event j, that is, inside the study region A. The intensity of the spatial point process, λ, is estimated by $\widehat{\lambda} = N/|A|$.

If the underlying process over region A is uniform, then the distribution of events within a ball of radius r around a given event, assuming the ball is contained in A, is binomial with mean $N\pi r^2/|A|$. Thus, the K-function is given by $K(r) = \pi r^2$, and plotting $\sqrt{K(r)/\pi}$ versus r gives a line of unit slope through the origin. Plotting this function with $\hat{K}$ instead of K gives one a graphical means to detect deviations from complete spatial randomness. In a similar vein, RGG proposed the following two statistics based on the K-function:

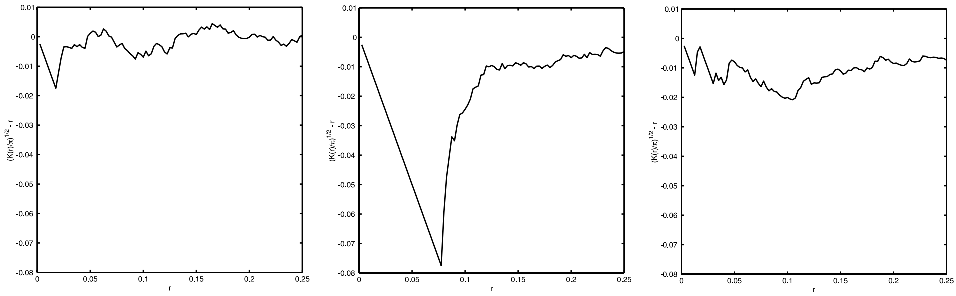
1. The difference between the area under $\sqrt{\widehat{K}(r)/\pi}$ and the 45° line over the initial part (from $\min_i(d_i)$ to $\max_i(d_i)$) of the curve: that is,

$$\int_{\min_i(d_i)}^{\max_i(d_i)} \left(\sqrt{\widehat{K}(u)/\pi} - u \right) du.$$

2. The slope of $\sqrt{\widehat{K}(r)/\pi}$ from $\min_i(d_i)$ to $\max_i(d_i)$.

We propose another statistic based on the K-function. Under strict inhibition, Isham (1984) showed that in the plane the K-function for the Strauss process with $\gamma = 0$ is approximately

$$K(r) = \begin{cases} 0, & 0 \leq r \leq \rho \\ \pi r^2 - \pi \rho^2, & r > \rho. \end{cases}$$



(a) Uniform process (b) Strauss process (c) Mixture process

Figure 2. K-function plots of each spatial point pattern of Figure 1.

For this process, clearly there is a change in the K-function at the point ρ which defines the inhibition process. Even for the K-function of a mixture process, we expect a change in the behavior of the estimated K-function, because it is a mixture of the inhibited K-function and the uniform K-function. We can estimate ρ by

$$\hat{\rho} = \arg\min_{r} \sqrt{\widehat{K(r)}/\pi} - r,$$

which we use as a summary statistic. This summary statistic has the advantage of being an estimate of an interpretable quantity, namely the inter-mine inhibition distance.

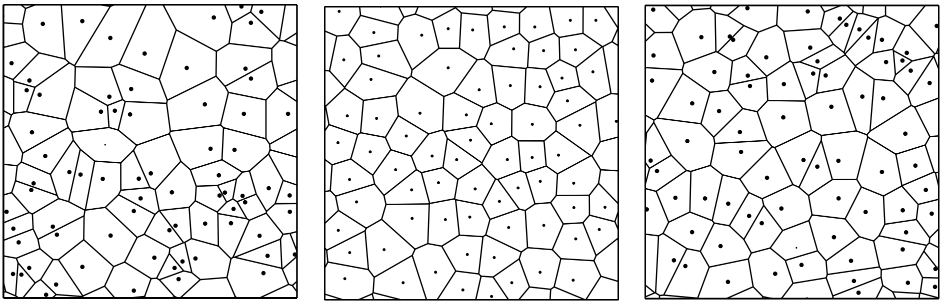
Figure 2 shows the K-functions for the spatial point patterns of Figure 1. The differences between the plots are apparent: the K-functions of the Strauss process and the mixture process both have sharp drops at $r = \rho$, while the K-function of the Poisson process is stationary with mean zero.

5.2 Spatial Tessellations

RGG also investigated using spatial tessellations (e.g., see Okabe, Boots, and Sugihara 1992) to distinguish between point process models. The simplest spatial tessellation is the Voronoï tessellation. Here every point in A is associated with the nearest event in A. This results in the study region, A, being partitioned into polygonal tiles (or Voronoï cells) (see Figure 3). RGG found the second central moment of the areas of the Voronoï cells to be a good statistic for differentiating between uniform and mixed uniform-Strauss processes. The inhibition between some events in the mixed processes results in less variable tile areas.

6. SIMULATION STUDY AND DATA ANALYSIS

We performed a simulation study to assess the performance of partial Bayes factors in the minefield problem. The simulation study is a simple 2^2 factorial design. The two factors we considered were: the number of noise events, n_0, and the amount of prior information. The parameters used in the simulation study are given in Table 2.



(a) Uniform process (b) Strauss process (a) Mixture process

Figure 3. Voronoï tessellations of each spatial point pattern of Figure 1.

The minefields simulated were either low noise ($m = 50, n_0 = 30$), or high noise ($m = 50, n_0 = 50$), and were compared with the corresponding uniform point process ($N = 80$, $N = 100$). Neither of these point patterns is easily distinguished by eye from a realization of a uniform process.

6.1 PRIORS

We decomposed the prior distribution on ρ and m in the following way.

$$\begin{aligned}\pi(\rho, m) &= \pi(\rho \mid m) \times \pi(m) \\ \pi(\rho \mid m) &= \text{Uniform}(\alpha_1 \rho_{\max}, \alpha_2 \rho_{\max}) \\ \pi(m) &= \text{Discrete Uniform}(\lfloor \beta_1 N, \beta_2 N \rfloor),\end{aligned}$$

where $\lfloor . \rfloor$ is the floor function.

We selected three sets of values for α_1, α_2, β_1, and β_2, that would correspond to "diffuse," "compact," and "tight" priors. These are shown in Table 3. We also created a prior corresponding to "perfect" prior information, that is, a prior with point mass on $m = 50, \rho = \frac{1}{2}\rho_{\max}$.

Table 2. Parameters of Simulation Study

Variable	*Value*
A	$(0,1)^2$
N	80, 100
m	0, 50
ρ	$\frac{1}{2}\rho_{\max}$

Table 3. Parameters of Prior Distributions

Prior	α_1	α_2	β_1	β_2
Diffuse	.3	.7	.10	.90
Compact	.4	.6	.30	.70
Tight	.45	.55	.40	.60

6.2 SUMMARY STATISTICS

We considered two different summary statistics on which to base the partial Bayes factors. The first summary statistic was based on the K-function, and the second (due to RGG) on the Voronoï tessellation. We shall refer to them as X_K and X_V, respectively:

$$\begin{aligned} X_K &= \arg\min_{d:\min_i(d_i)<r<\max_i(d_i)} \sqrt{\widehat{K(r)}/\pi} - r, \\ X_V &= \text{second central moment of the areas of the Voronoï cells.} \end{aligned}$$

6.3 EDGE EFFECTS

Both of the above statistics can suffer from edge effects (particularly X_V). Edge effects occur because events near the boundary have fewer neighbors than events in the central part of the study area. We accounted for edge effects by generating all point patterns on a region with a 20% border. Thus, instead of generating a point pattern with N events on the unit square, we generated $\lfloor 1.96 \times N \rfloor$ events on $(-.2, 1.2)^2$. The factor 1.96 is the ratio of the areas of the two regions.

6.4 RESULTS

We simulated 100 point patterns on the unit square (accounting for edge effects as described earlier) under each hypothesis, and for each value of n_0 (i.e., a total of 400

Table 4. Simulation Study: Percentage of Datasets Misclassified. These are broken down by prior information, noise level, summary statistics used, and data generation method (H_0 or H_1).

	Noise	*K-function*		Voronoï	
Prior	*Level*	H_1	H_0	H_1	H_0
Diffuse	Low	4	40	40	10
Compact	Low	6	32	33	13
Tight	Low	10	31	33	14
Perfect	Low	11	22	34	12
Diffuse	High	13	39	47	14
Compact	High	23	28	35	16
Tight	High	28	25	34	20
Perfect	High	28	25	33	23
Average Error Rate		15	30	36	15
Total Error Rate		22.5		25.5	

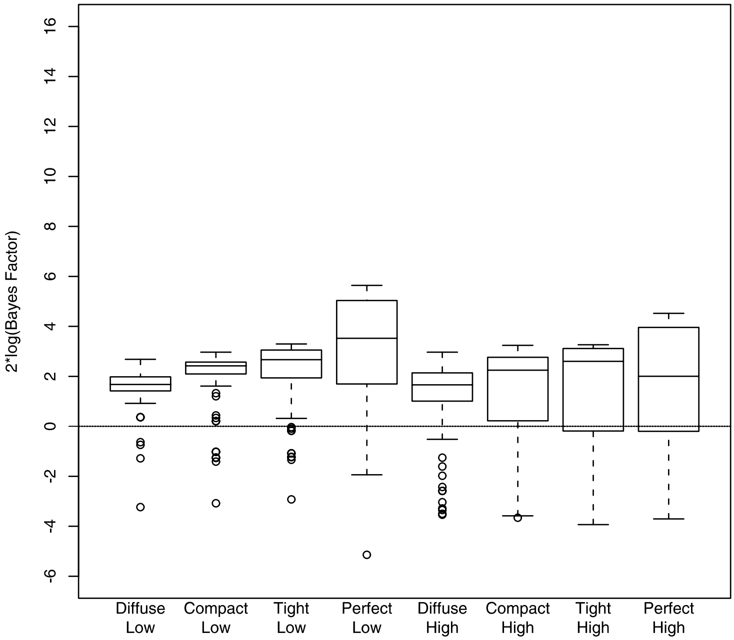


Figure 4. Partial Bayes factors for H_1, based on X_K, under H_1.

datasets). We calculated the partial Bayes factors for each dataset using both summary statistics, X_K and X_V, and the prior distributions given in Table 3. The partial Bayes factors are calculated in terms of evidence for H_1 over H_0. The integration was performed using simple numerical quadrature. The misclassification rates are given in Table 4. From this table we can see that the total misclassification rate (assuming that each hypothesis is equally likely a priori) was 22.5% for the partial Bayes factor based on the K-function, and 25.5% for the partial Bayes factor based on the Voronoï tiling. While these total error rates are similar, the partial Bayes factor based on X_K was more successful at correctly classifying minefields than noise processes. The opposite was true of the partial Bayes factor based on X_V. As one would expect, an increase in the amount of (correct) information contained in priors improves the discrimination in both cases.

Boxplots summarizing these partial Bayes factors are shown in Figures 4, 5, 6, and 7. The results under the X_K statistic, may be summarized as follows:

- Under H_1, the partial Bayes factors typically provide weak to positive evidence for the (correct) minefield hypothesis.
- Under H_0, the partial Bayes factors typically range from positive evidence for H_0 to weak evidence for H_1. The median partial Bayes factor is approximately 1.7 in favor of the correct hypothesis.
- The increase in noise had a small negative effect on the performance of the partial Bayes factors under both hypotheses.

Using the X_V statistic the results were:

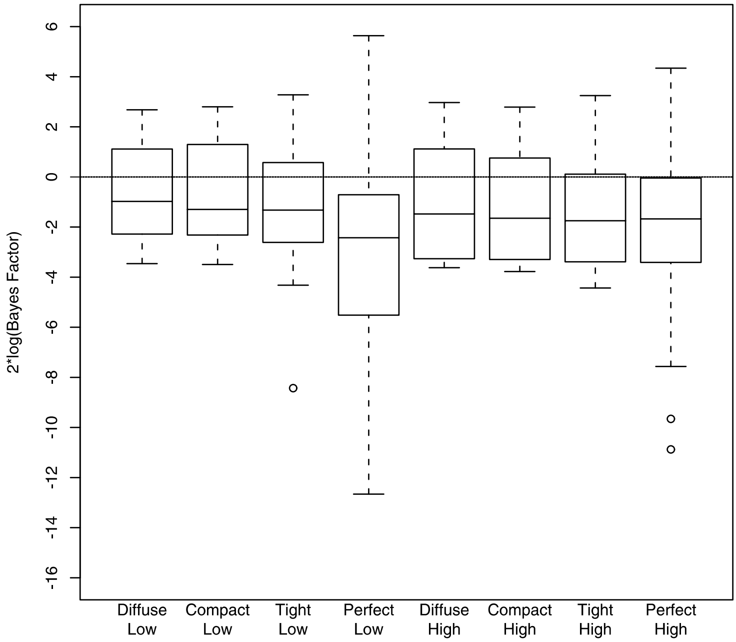


Figure 5. *Partial Bayes factors for H_1, based on X_K, under H_0.*

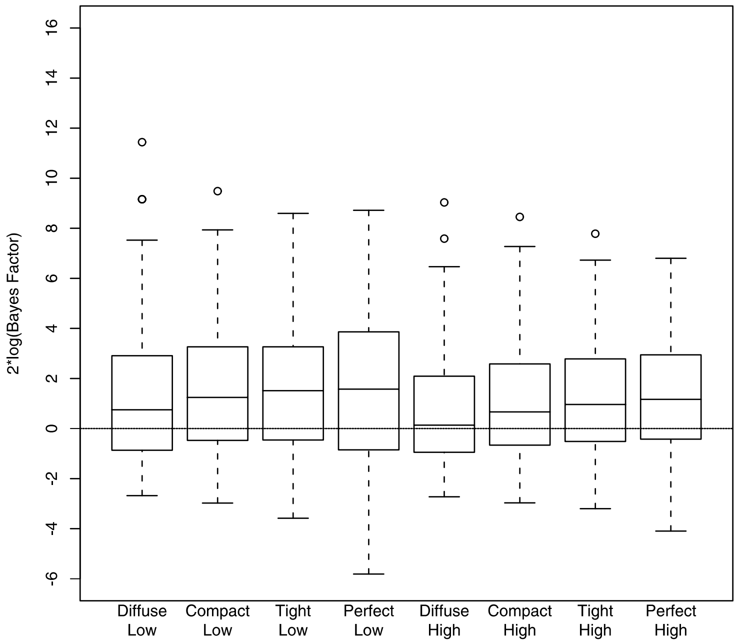


Figure 6. *Partial Bayes factors for H_1, based on X_V, under H_1.*

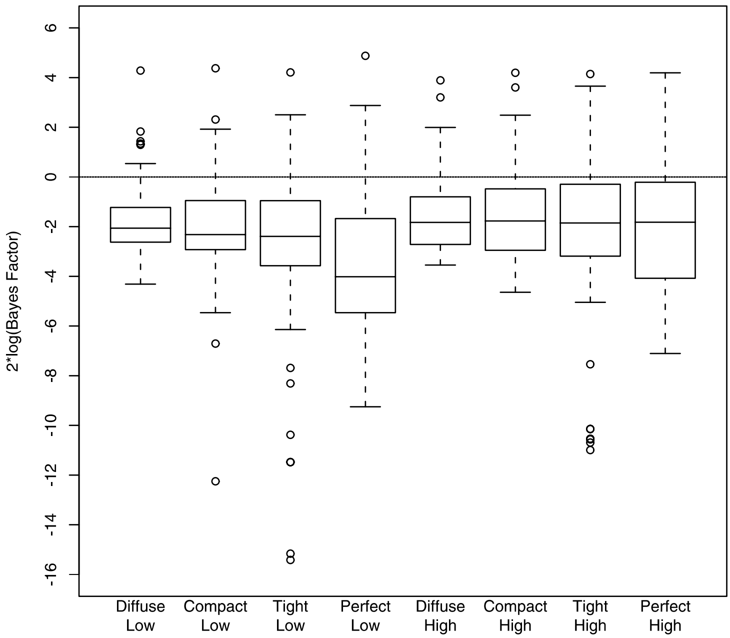


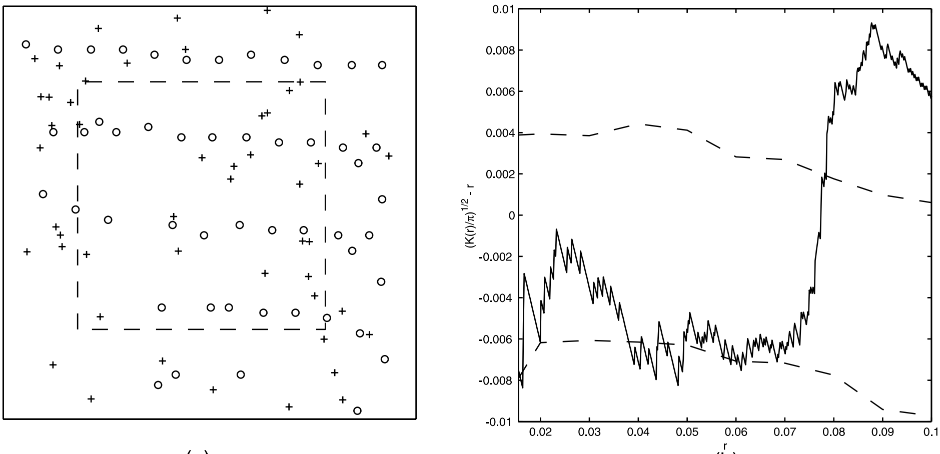
Figure 7. *Partial Bayes factors for H_1, based on X_V, under H_0.*

- Under H_1, the partial Bayes factors typically range from positive evidence for H_1 to weak evidence for H_0. The median partial Bayes factor is again approximately 1.7 in favor of the correct hypothesis.
- Under H_0, the evidence for H_0 typically ranges from weak to positive.
- There is a negligible negative effect in the performance of the partial Bayes factors due to the increase in noise.

6.5 MINEFIELD DATA

Figure 8(a) shows the locations of mines (o) and noise events (+) on a surf beach. The dataset, which we shall refer to as the Surf Zone dataset, was described and analyzed by Lake and Keenan (1995). One should note that the mines are approximately arranged evenly spaced in parallel rows. Because the rows are farther apart than consecutive mines within a row, the mines do not resemble a typical Strauss process. However since the minefield does display inhibition, the Strauss model is a useful first approximation to the minefield process.

In order to account for edge effects we analyzed the points lying in the central square region shown in Figure 8(a). This region contains 40 events, 20 mines, and 20 noise events. The inter-mine distance is approximately half of $\rho_{\max}$, so we used the same priors to analyze this dataset as were used in the simulation study. The K-function for the dataset is shown in Figure 8(b).



(a) (b)

Figure 8. Surfzone minefield dataset. (a) Surf zone dataset; o = mines, + = noise. (b) K-Function for the surf zone dataset. The dotted lines represent the 10th and 90th percent quantiles for a K-function of a uniform process.

The partial Bayes factors calculated based on X_K and X_V for each of the three priors are shown in Table 5. We can see that the partial Bayes factors based on X_K provide weak evidence for the minefield hypothesis. Because the mines are not actually laid as Strauss process this result is reasonably good.

However, the partial Bayes factors based on X_V provide weak evidence against the minefield hypothesis. It is apparent from these results that the K-function was better than the Voronoï tiling variance at capturing the regularities in this dataset. This result, along with the simulation study, show that the partial Bayes factors based on X_V are more prone to Type II error; that is, failing to detect a minefield, than the partial Bayes factor based on X_K. Since failing to detect a minefield is obviously worse than falsely detecting one, we conclude that the X_K is preferable to X_V in this situation.

7. DISCUSSION

In this article we investigated the feasibility of using partial Bayes factors to classify mixtures of spatial point processes. We limited our attention to two different summary statistics on the basis of which to calculate the partial Bayes factors. One summary statistic was based on the K-function, and the other on the Voronoï tessellation. We performed a simulation study, and found that partial Bayes factors based on both statistics provided good

Table 5. Twice the Log Partial Bayes factors for H_1, for the Minefield Dataset

Prior	*K-function*	*Voronoï*
Diffuse	1.64	−1.47
Compact	1.87	−1.26
Tight	2.02	−1.31

discrimination between the competing hypotheses we considered. The partial Bayes factor based on the K-function had lower Type II error, while the partial Bayes factor based on the Voronoï tessellation had lower Type I error. In the case of point processes arising in the minefield context, the penalty for Type II errors is high, therefore we prefer the K-function based partial Bayes factor. When we applied our method to real minefield data, we found that the statistic based on the K-function provided some evidence for the presence of a minefield.

A previous approach to this problem using a supervised pattern recognition scheme based on summary statistics of the point pattern was developed by Raghavan, Goel, and Ghosh (1997, 1998). Our approach has the advantage of providing a natural framework within which to include prior information about each competing hypothesis, which can be very useful in this kind of application when it is available.

Our approach is motivated by the problem of having a statistical model from which we can simulate data, but which has a likelihood that is difficult to evaluate. The task of parameter estimation in this setting was investigated by Diggle and Graton (1984). Their approach was to use simulated realizations from an "implicit" statistical model, and kernel estimation, to estimate the log-likelihood function and then to maximize this function via a modified simplex algorithm.

More recently, Harshman and Clark (1998) used a simulation-based maximum likelihood method for estimation of parameters in a sperm competition model. As in our method, they reduced the data to an approximately sufficient summary statistic. In this article we have limited our attention to the problem of classifying spatial point processes. However, the partial Bayes factor methodology is clearly applicable in other situations: it is potentially helpful whenever the full likelihood of the data is intractable.

ACKNOWLEDGMENTS

This research was supported by the Office of Naval Research grants N00014-96-1-0192. N00014-97-1-0736, and N00014-01-10745, and by NIH Grant 1R01CA094212-01.

[Received December 2002. Revised February 2004.]

REFERENCES

Bartlett, M. S. (1964), "The Spectral Analysis of Two-Dimensional Point Processes," *Biometrika*, 51, 299–311.

Byers, S. D., and Raftery, A. E. (1998), "Nearest Neighbor Clutter Removal for Estimating Features in Spatial Point Processes," *Journal of the American Statistical Association*, 93, 557–584.

Cressie, N., and Lawson, A. B. (1998), "Bayesian Hierarchical Analysis of Minefield Data," in *Detection and Remediation Technologies for Mines and Minelike Targets III*, Volume 3392 of *Proceedings of SPIE*, eds. A. C. Dubey, J. F. Harvey, and J. T. Broach, pp. 930–940.

Cressie, N. A. C. (1993), *Statistics for Spatial Data*, New York: Wiley.

Dasgupta, A., and Raftery, A. E. (1998), "Detecting Features in Spatial Point Processes with Clutter via Model-Based Clustering," *Journal of the American Statistical Association*, 93, 294–302.

Diggle, P. J. (1983), *Statistical Analysis of Spatial Point Patterns*, London: Academic Press.

Diggle, P. J., Fiksel, T., Grabarnik, P., Ogata, Y., Stoyan, D., and Tanemura, M. (1994), "On Parameter Estimation for Pairwise Interaction Point Processes," *International Statistical Review*, 62, 99–117.

Diggle, P. J., and Graton, R. J. (1984), "Monte Carlo Methods of Inference for Implicit Statistical Models" (with discussion), *Journal of the Royal Statistical Society*, Ser. B, 46, 193–227.

Harshman, L. G., and Clark, A. G. (1998), "Inference of Sperm Competition From Broods of Field-Caught, *drosophila*," *Evolution*, 52, 1334–1341.

Isham, V. (1984), "Multitype Markov Point Processes: Some Approximations," *Proceedings of the Royal Society of London*, Ser. A, 391, 39–53.

Jeffreys, H. (1961), *Theory of Probability* (3rd ed.), Oxford: Clarendon Press.

Kass, R. E., and Raftery, A. E. (1995), "Bayes Factors," *Journal of the American Statistical Association*, 90, 773–795.

Kelly, F. P., and Ripley, B. D. (1976), "A Note on Strauss's Model for Clustering," *Biometrika*, 63, 357–360.

Lake, D. E., and Keenan, D. M. (1995), "Identifying Minefields in Clutter via Collinearity and Regularity Detection," in *Detection and Remediation Technologies for Mines and Minelike Targets*, Volume 2496 of *Proceedings of SPIE*, eds. A. C. Dubey, I. Cindrich, J. M. Ralston, and K. Rigano, pp. 519–530.

Lake, D. E., Sadler, B., and Casey, S. (1997), "Detecting Regularity in Minefields Using Collinearity and a Modified Euclidean Algorithm," in *Detection and Remediation Technologies for Mines and Minelike Targets II*, Volume 3079 of *Proceedings of SPIE*, eds. A. C. Dubey and R. Barnard, pp. 508–518.

Muise, R. R., and Smith, C. M. (1992), "Nonparametric Minefield Detection and Localization," Technical Report CSS-TM-591-91, Naval Surface Warfare Center, Coastal Systems Station.

O'Hagan, A. (1995), "Fractional Bayes Factors for Model Comparison" (with discussion), *Journal of the Royal Statistical Society*, Ser. B, 57, 99–138.

Okabe, A., Boots, B., and Sugihara, K. (1992), *Spatial Tesselations: Concepts and Applications of Voronoï Diagrams*, Chichester: Wiley.

Raghavan, N., Goel, P., and Ghosh, S. (1997), "Classification of Mixtures of Spatial Point Processes," in *Computing Science and Statistics. Mining and Modeling Massive Data Sets in Science, Engineering, and Business with a Subtheme in Environmental Statistics. Proceedings of the 29th Symposium on the Interface*, Fairfax Station, VA: Interface Foundation of North America, pp. 112–116.

——— (1998), "Computer Experiments for the Classification of Spatial Point Processes," Technical Report 600, Department of Statistics, Ohio State University.

Ripley, B. D. (1976), "The Second-Order Analysis of Stationary Point Processes," *Journal of Applied Probability*, 13, 255–266.

——— (1977), "Modelling Spatial Patterns" (with discussion) *Journal of the Royal Statistical Society*, Ser. B, 39, 172–212.

Silverman, B. W. (1986), *Density Estimation for Statistics and Data Analysis*, London: Chapman & Hall.

Strauss, D. J. (1975), "A Model for Clustering," *Biometrika*, 62, 467–475.

Walsh, D. C. I., and Raftery, A. E. (2002), "Detecting Mines in Minefields with Linear Characteristics," *Technometrics*, 41, 34–44.