Ann. Inst. Statist. Math.
Vol. 41, No. 3, 583-600 (1989)

LIKELIHOOD ESTIMATION OF SOFT-CORE INTERACTION POTENTIALS FOR GIBBSIAN POINT PATTERNS

YOSIHIKO OGATA AND MASAHARU TANEMURA

The Institute of Statistical Mathematics, 4-6-7 Minami-Azabu, Minato-ku, Tokyo 106, Japan

(Received December 4, 1987; revised September 1, 1988)

Abstract. The likelihood method is developed for the analysis of so-called regular point patterns. Approximating the normalizing factor of Gibbs canonical distribution, we simultaneously estimate two parameters, one for the scale and the other which measures the softness (or hardness), of repulsive interactions between points. The approximations are useful up to a considerably high density. Some real data are analyzed to illustrate the utility of the parameters for characterizing the regular point pattern.

Key words and phrases: Gibbs canonical distribution, Hard-Core model, Monte Carlo method, ψ-value, reduced density, Soft-Core model, spline surface fitting.

1. Introduction

Suppose we are given a mapped spatial pattern of points in a finite planar region. The points may be individuals of animals or plants, and the region may be their habitat. When a mass of territorial animals are settled in a certain habitat, they usually show a pattern where individuals are rather equally spaced from one another. Such a pattern of points is called "regular". It is easily surmised that this regular pattern is due to territorial aggression. It will be interesting to know how strong and hard this aggression, or repulsive interaction, is in this species of animals.

We suppose that these patterns are realized after some adjustment of positions. Thus, these may be in a sort of equilibrium system under a certain mechanism. The simplest examples of self-organized mechanisms include mutual interaction, which is described by a function of distances between points or individuals. It is customary to assume that an equilibrium system is statistically characterized by a Gibbs distribution of a total potential energy. Then, what sort of interaction potential is working

among the points? And how can we estimate such an interaction from the mapped data of points? If the likelihood function of the parameters which characterize the potential can be written by the use of the Gibbs distribution, then maximum likelihood methods may well provide a sensible estimation of parameters and sensitive testing of the models. However, this has so far hardly been implemented, because of the very high multiplicity of the integral for the normalizing factor of the distribution.

Ogata and Tanemura (1984) discussed a class of repulsive potential functions including so-called Soft-Core potentials, and provided an approximation method of the log normalizing factor of the Gibbs distribution by the use of Monte Carlo experiments. The principal aim of the present paper is to extend the method of approximation for a two-parameter family of potential models representing the softness of repulsion, including the Hard-Core potential, and also to show the utility of such parameters and related important measures which characterize the regular point patterns.

2. Likelihood of equilibrium point patterns

Suppose that we are given a mapped spatial pattern of N points in a finite planar region V. Let the coordinates of the points of a pattern in equilibrium be $X = \{x_i = (x_i, y_i) \in V, i = 1, 2, \ldots, N\}$. The simplest physically justified supposition consists of reducing the total potential energy to a sum of pairwise interaction potential energy. Let us represent such a mutual interaction between two points i and j by a potential energy function $\Phi(r_{ij})$ of the distance $r_{ij} = |x_i - x_j|$. We suppose that the points in the region are realized after adjusting themselves to make the total potential energy

$$U_N(X) = \sum_{i<j}^{N} \Phi(r_{ij}) \tag{2.1}$$

being stationary. Thus the points X can be regarded as being distributed according to the Gibbs canonical distribution

$$f(X) = \exp\{-U_N(X)\}/Z(\Phi; N, V)\,, \tag{2.2}$$

where the *normalizing factor* is given by the $2N$-fold integral

$$Z(\Phi; N, V) = \int_{V^N} \exp\{-U_N(X)\}dx_1 \cdots dx_N\,. \tag{2.3}$$

We have been interested in estimating the form of $\Phi(r)$ from a single instantaneous spatial realization of points X in equilibrium (Ogata and

Tanemura (1981*a*, 1984, 1985 and 1986)). Thus, consider a family of parametrized pairwise potential functions $\{\Phi_\theta(r); \theta \in \Theta\}$ defined in some parameter space Θ. If the data X is substituted in (2.2), then the likelihood of the potential $\Phi_\theta(r)$ is obtained as a function of θ. The difficulty, however, is to obtain the explicit form of the normalizing factor (2.3) as a function of θ, except for the case of the *Poisson model*. By the Poisson model we mean that the N points are uniformly and independently distributed in the region V. In this case there is no interaction, i.e., $\Phi_\theta(r) \equiv 0$, and we have $|V|^N$ for (2.3). Considering the Poisson case as the standard, we hereafter use the log likelihood (ratio) function

$$L(\Phi_\theta; X) = -\sum_{i<j}^{N} \Phi_\theta(r_{ij}) - \log \bar{Z}(\Phi_\theta; N, V) , \tag{2.4}$$

where $\bar{Z} = Z/|V|^N$.

We may consider another type of Gibbs distribution which assumes the number of points in a region to be a random variable, called the grand canonical distribution in statistical mechanics, which is related to the so-called full likelihood. On the other hand, the above likelihood may be strictly called the conditional likelihood on the fixed number N of points in a particular region. Nevertheless, it is known that these do not make much practical difference for a not small mean number of points in a region.

3. Monte Carlo method for simulating the Gibbsian patterns

To give some feeling of the relation between a pairwise potential and its equilibrium point patterns, let us review a simulation method which uses a particular type of random walk known as a Markov chain. The simulation was originally devised by Metropolis *et al.* (1953) and developed by Wood (1968) and others for the study of atomic systems.

Consider a set of particles, interacting according to a certain potential function, on a square V with a periodic boundary (i.e., V is identical to a torus). The most commonly used algorithm is described in the following manner: Assume that at time t, the state of the N particle system is $X(t) = \{(x_n(t), y_n(t)) \in V; n = 1,\ldots, N\}$. A trial state $X'(t) = \{(x_n'(t), y_n'(t))\}$ is then chosen in such a way that the coordinates $(x_r'(t), y_r'(t))$ of a randomly chosen particle r lie in some square with vertices at the points $(x_r(t) \pm \delta, y_r(t) \pm \delta)$ while all other $N - 1$ particles have the same position as in state $X(t)$, where $\delta > 0$ is a parameter to be discussed below. The corresponding total potential energy $U_N(X'(t))$ in (2.1) is then calculated and compared with $U_N(X(t))$ as follows.

1. If $U_N(X'(t)) \leq U_N(X(t))$, then without further ado the next state $X(t + 1)$ of the realization is taken as the trial state $X'(t)$.

2. If $U_N(X'(t)) > U_N(X(t))$, then we obtain a uniform random number

ξ and (a) if $\xi \leq \exp\{U_N(X(t)) - U_N(X'(t))\}$, state $X(t+1)$ is taken to be the trial state $X'(t)$; (b) otherwise, state $X(t+1)$ is taken to be the previous state $X(t)$.

It should be noticed that the normalizing factor Z in (2.3) has not been used in the simulation. In essence, the Monte Carlo procedure here is nothing but selecting the transition probabilities

$$g(X, Y)dY = \text{Prob}\{X(t+1) \in dY \mid X(t) = X\}, \tag{3.1}$$

which satisfy $\int f(X)g(X, Y)dX = f(Y)$ for all the state Y of the N-particle system in V^N, and further satisfy the condition that the n-step transition probability density $g^{(n)}(X, Y)$ converges to the given equilibrium probability $f(Y)$ in (2.2).

The parameter δ, the maximum single step displacement allowed in passing from one state to the next, ought in principle to be adjusted for an optimum rate of convergence of the Markov chain. Wood (1968), according to his experiments, suggests that a reasonable choice for the adjusting parameter δ has been found to be a value leading to rejection of the trial configuration on about half of the time-steps. This is a trade-off, especially in the case of high density, between effective transition of the state and avoiding unnecessary repetition of the same state. In connection to this issue, an alternative simulation method provided by Ripley (1979) on the mathematical basis of the spatio-temporal birth and death process should be modified in choosing the trial points in order to be useful for high density cases.

In addition to the selection of δ, in order to attain the equilibrium state within fewer time-steps in Monte Carlo simulation, the initial configuration should be suitably chosen. We devised a method of generating points sequentially using the rejection method, for the construction of such an initial state (Ogata and Tanemura (1981*b*)).

4. Potential function with scaling property and the normalizing factor

We here review the basic idea which is the clue to the use of the Monte Carlo method for the estimation of the normalizing constant (see Ogata and Tanemura (1981*c*, 1982 and 1984)). Consider the case where the pairwise potential function is characterized by a scale parameter σ such that

$$\Phi_\sigma(r) = \Phi_1(r/\sigma). \tag{4.1}$$

The standardized function Φ_1 exhibits the shape of the potential function and may be characterized by some other parameters. Because of the scaling

property of Φ_σ in (2.4), we can only consider the situation in the standard scale; introducing the parameter $\tau = (N/V)\sigma^2$, called the *reduced density*, we have the following log likelihood ratio

$$\log L(\tau; X) = -\sum_{i<j}^{N} \Phi_1(r_{ij}^*/\sqrt{\tau}) - \log \bar{Z}(\tau, \Phi_1; N) , \tag{4.2}$$

where the distances $r_{ij}^* = r_{ij}/\sqrt{V/N}$. Under very broad regularity conditions we can expect $E[\partial L(\tau; X)/\partial\tau] = 0$, which implies the following equality

$$-\frac{1}{2N} E\left[\sum_{i<j}^{N} (r_{ij}^*/\sqrt{\tau})\Phi_1'(r_{ij}^*/\sqrt{\tau})\right] = -\frac{\tau}{N} \cdot \frac{\partial}{\partial\tau} \log \bar{Z}(\tau, \Phi_1; N) , \tag{4.3}$$

where $\Phi_1'(\cdot)$ indicates the derivative function of $\Phi(\cdot)$. Now both sides of this equality are known to be distinctive definitions of $P|V|/N - 1$ in kinematics (the left-hand side) and in statistical mechanics (the right-hand side), where P is the *pressure*. This is a function of the reduced density and we here define $\psi(\tau) = P|V|/N - 1$, which measures a degree of deviation from the ideal gas (Poisson patterns). A great advantage to using this function for the approximation of the normalizing factor is that $\psi(\tau)$ is almost constant with respect to the number of points N if N is not very small. Indeed, once the function $\psi(\tau)$ is known for any τ, we get

$$\frac{1}{N} \log \bar{Z}(\tau, \Phi_1; N) = -\int_0^\tau \frac{\psi(t)}{t}\, dt , \tag{4.4}$$

from the right-hand side of (4.2). Note here that any information of $\log \bar{Z}$, such as the first or second derivatives of $\log \bar{Z}$ with respect to τ or any other parameter if included in Φ_1, is completely included in the ψ-function. Therefore, we concentrate our attention on getting the explicit form of the function $\psi(\tau)$ for each standardized potential model Φ_1.

The same idea for the derivation of the equation (4.3) was used and extended by Penttinen (1984) to estimate the Hessian matrix (the second derivative of the log likelihood), as well as the efficient score, for the use of a Newton-Raphson type algorithm in solving the likelihood equation. The equation for one parameter is solved for the example of his estimation procedure, but the Newton-Raphson equation with some random effects seems to have practical numerical difficulty, especially in determining the convergence criterion. Our procedure, as is seen later, eventually avoided this kind of difficulty by the use of a certain smoothing technique. Furthermore, it should be stressed that the value of the maximum log likelihood is available in our procedure for the model comparison, or goodness-of-fit of models.

5. ψ-values through computer experiments

According to the other definition of ψ-function given in the left-hand side of (4.3), we consider a consistent and unbiased estimator of $\psi(\tau)$; replacing the Gibbs ensemble average $E\{\cdot\}$ by the time average

$$\hat{\psi}(\tau) = -\frac{1}{2N} \cdot \frac{1}{M} \sum_{t=1}^{M} \sum_{i<j}^{N} \frac{r_{ij}^*(t)}{\sqrt{\tau}} \cdot \Phi_1'\left(\frac{r_{ij}^*(t)}{\sqrt{\tau}} \right). \tag{5.1}$$

Here $r_{ij}^*(t) = |x_i^*(t) - x_j^*(t)|$ and the patterns $[X^*(t) = \{x_i^*(t);\ i = 1,\ldots, N\}$;

Table 1. Results of the Monte Carlo experiments for Soft-Core potentials $\Phi(r) = (\sigma/r)^n$ for patterns of 500 points in 2-dimensional space: sample mean (5.3) of $\psi(\tau_i)$ for each τ_i and standard deviation of the corresponding time series. M in (5.1) and (5.3) is obtained by multiplying M_0 by 500.

Soft-Core: $n = 4$

τ	$\hat{\psi}$	s.d.	M_0
0.05	0.14732	0.03433	1900
0.10	0.31426	0.04022	1900
0.15	0.50447	0.04460	1900
0.20	0.71244	0.03848	1900
0.25	0.94399	0.03510	1900
0.30	1.20016	0.03980	1900
0.35	1.47182	0.06585	1900
0.40	1.77189	0.08588	1900
0.45	2.10233	0.10426	1900
0.50	2.45329	0.12014	1900
0.55	2.82461	0.13728	1900
0.60	3.22863	0.13189	1900
0.65	3.66457	0.15078	1900
0.70	4.11131	0.16969	1900
0.75	4.59497	0.18301	1900

Soft-Core: $n = 6$

τ	$\hat{\psi}$	s.d.	M_0
0.05	0.11160	0.03465	1900
0.10	0.24241	0.04257	1900
0.15	0.39641	0.05092	1900
0.20	0.56114	0.05251	1900
0.25	0.76378	0.05344	1900
0.30	0.97851	0.05486	1900
0.35	1.23053	0.06097	1900
0.40	1.51912	0.08610	1900
0.45	1.84457	0.10523	1900
0.50	2.21532	0.12409	1900
0.55	2.62742	0.14441	1900
0.60	3.08346	0.14469	1900
0.65	3.60767	0.16457	1900
0.70	4.18188	0.18587	1900
0.75	4.80052	0.20387	1900

Soft-Core: $n = 8$

τ	$\hat{\psi}$	s.d.	M_0
0.05	0.10050	0.09194	1900
0.10	0.22150	0.10179	1900
0.15	0.35722	0.05870	1900
0.20	0.51460	0.06040	1900
0.25	0.69386	0.06151	1900
0.30	0.90108	0.06862	1900
0.35	1.15119	0.07002	1900
0.40	1.43336	0.09208	1900
0.45	1.75532	0.11273	1900
0.50	2.13349	0.13080	1900
0.55	2.58963	0.15406	1900
0.60	3.08778	0.15785	1900
0.65	3.66157	0.17943	1900
0.70	4.35107	0.20440	1900
0.75	5.15353	0.21721	1900

Soft-Core: $n = 12$

τ	$\hat{\psi}$	s.d.	M_0
0.05	0.09449	0.04169	3800
0.10	0.20593	0.23038	1900
0.15	0.32557	0.06657	1900
0.20	0.47426	0.07625	1900
0.25	0.63616	0.07732	1900
0.30	0.83994	0.08743	1900
0.35	1.06117	0.09485	1900
0.40	1.34459	0.11061	1900
0.45	1.65606	0.12608	1900
0.50	2.04039	0.14937	1900
0.55	2.51227	0.17327	1900
0.60	3.04830	0.18545	1900
0.65	3.70757	0.21162	1900
0.70	4.53279	0.23985	1900
0.75	5.50700	0.24745	1900

Table 1. (continued).

Soft-Core: $n = 16$				Soft-Core: $n = 24$			
τ	$\hat{\psi}$	s.d.	M_0	τ	$\hat{\psi}$	s.d.	M_0
0.05	0.09007	0.04527	3800	0.05	0.08655	0.06609	4900
0.10	0.19153	0.05939	1900	0.10	0.18794	0.08311	4900
0.15	0.30946	0.07519	3500	0.15	0.30043	0.10050	4900
0.20	0.45513	0.09179	3500	0.20	0.43486	0.11779	4900
0.25	0.61515	0.10789	3500	0.25	0.58728	0.13439	4900
0.30	0.80768	0.12359	3500	0.30	0.77549	0.15518	4900
0.35	1.02964	0.14158	3500	0.35	0.98483	0.17704	4900
0.40	1.29252	0.16182	3500	0.40	1.23539	0.19313	4900
0.45	1.60127	0.18019	3500	0.45	1.54246	0.21890	4900
0.50	1.96820	0.19766	3500	0.50	1.92613	0.23700	4900
0.55	2.44003	0.22684	3500	0.55	2.36671	0.26682	4900
0.60	3.00689	0.22334	3500	0.60	2.91175	0.27558	4900
0.65	3.69256	0.22841	3500	0.65	3.61088	0.25331	4900
0.70	4.54177	0.23143	3500	0.70	4.49158	0.27324	4900
0.75	5.59701	0.24767	3500	0.75	5.58680	0.31645	4900

$t = 1,\ldots, M]$, are obtained by the Monte Carlo simulation procedure (see Section 2) which generates the Gibbs canonical ensemble characterized by $\Phi_1(r)$. Thus, for some reduced densities $\tau_1, \tau_2,\ldots, \tau_p$, we get the averages $\hat{\psi}(\tau_i)$ and the variances $\hat{s}_i^2, i = 1,\ldots, p$, of the time series.

For example, consider the so-called Soft-Core models (see Fig. 1)

$$\Phi_\sigma(r) = (\sigma/r)^n, \quad n > 2\,, \tag{5.2}$$

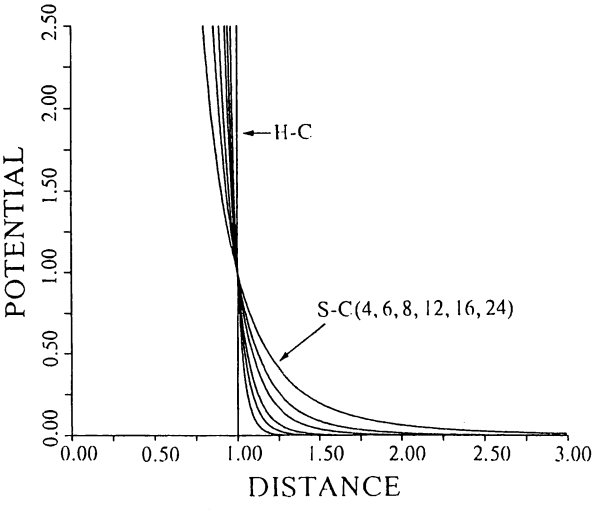


Fig. 1. Curves of potential models for $\sigma = 1$. The model corresponding to each curve is nearby indicated where H-C and S-C(n) correspond to Hard-Core and Soft-Core with index n, respectively.

where the restriction of n in (5.2) is to ensure a stable potential in the sense of Ruelle (see Gates and Westcott (1986), for example). Then (5.1) is written as

$$\hat{\psi}(\tau) = \frac{n}{2N} \cdot \frac{1}{M} \sum_{l=1}^{M} \sum_{i<j}^{N} \left(\frac{\sqrt{\tau}}{r_{ij}^{*}(t)} \right)^{n} . \tag{5.3}$$

After some computer simulations with $M \simeq 10^6$ for the number of points $N = 500$ in the rescaled square region $V_1 = V/(|V|/N)$ with a periodic boundary (identical to the torus), we obtained the sample means $\hat{\psi}(\tau_k)$, on $\tau_k = 0.05k$, $k = 1, 2, \ldots, 15$ for $n = 4, 6, 8, 12, 16$ and 24 together with the standard deviations $\hat{s}_{k,n}$ of the time series. These, together with the step size $M_0 = M/N$ used to take the average in (5.3), are listed in Table 1, which we use in the next section. This table itself is useful in physics for the comparison and further accumulation of such simulation experiments; see Swol *et al.* (1980) for the $n = 12$ case of the Soft-Core in 2-dimensional space and Hoover *et al.* (1971) for $n = 4, 6, 9, 12$, and ∞ of the Soft-Core in 3-dimensional space.

It is crucial to choose a well-tested random number generation scheme for such large-scale Monte Carlo experiments, in order to avoid the biases of the present statistics. Throughout this work, we used a physical random number generator (200Kbytes per second) installed in the Institute of Statistical Mathematics.

6. Approximation of the log likelihood

Reparametrize the set of Soft-Core potential model (5.2) in the form

$$\Phi(r; \sigma, \alpha) = (\sigma/r)^{2/\alpha}; \quad 0 \le \sigma < \infty, \quad 0 \le \alpha < 1 . \tag{6.1}$$

Note that $\alpha = 0$ corresponds to the Hard-Core potential such that $\Phi(r; \sigma, 0) = \infty$ for $r \le \sigma$ and $= 0$ for $r > \sigma$, and that $\sigma = 0$ corresponds to the Poisson model. Our goal in this section is to give an approximated function for the logarithm of the normalizing factor in (4.2) by using the relation (4.4) and Table 1.

In addition to the ψ-value estimates obtained in Table 1, we used the ψ-values for the Hard-Core model by the Padé approximant (Ree and Hoover (1967)),

$$\psi(\tau) = \frac{1.570796\tau - 0.498612\tau^2 + 0.021662\tau^3}{1 - 1.545797\tau + 0.599371\tau^2} . \tag{6.2}$$

This approximant is known to be useful in the range up to about $\tau = 0.88$

(Wood (1968)). Standard errors, for the present ψ-values, are assumed to be similar to those for the Soft-Core $n = 24$ for each $\tau_k = 0.05k$, $k = 1, 2, \ldots, 15$.

Before performing the surface fitting for ψ-values on (α, τ) for carrying out the integral in (4.4), we should note that the data set of the form $\{\psi_n(\tau_k)/\tau_k\}$ is convenient, rather than $\{\psi_n(\tau_k)\}$ themselves. Thus the estimate and standard errors obtained in Table 1 are recalculated for the form $\psi(\tau)/\tau$. Note that we can further add the values of $\psi_n(\tau)/\tau$ for $\tau = 0$, because $\lim_{\tau \to 0+} \psi_n(\tau)/\tau$ provides the second order virial coefficient of each model and is equal to $(\pi/2)\Gamma(1 - 2/n)$, where $\Gamma(\cdot)$ is the gamma function (Ogata and Tanemura (1981*a*, 1984)). Thus we use the value $(\pi/2)\Gamma(1 - 2/n)$; $n = 4, 6, 8, 12, 16, 24$ and ∞ (Hard-Core) with similar standard error to $\psi_n(0.05)/0.05$ for each n.

Unlike the curve fitting implemented in Ogata and Tanemura (1981*b*, 1982 and 1984), low dimension polynomials are not very good for surface fitting, since these cause bias at the sample points. On the other hand, high order polynomials which needs lots of parameters usually provide unnecessarily rapid fluctuations of the surfaces. Therefore, some restriction for the smoothness of the surface and then trade-off with the goodness-of-fit to the data are required (see Good and Gaskins (1971), for example). The situation is the same for the bicubic spline functions which we here prefer to deal with.

The 2-dimensional cubic spline function is defined as follows. Consider that the rectangular area $A = [x_0, x_M] \times [y_0, y_N]$ for the domain of definition of the function, and the sequences of points $x_0 < x_1 < \cdots < x_{M-1} < x_M$ and $y_0 < y_1 < \cdots < y_{N-1} < y_N$ are all equally spaced. Both of the segments $[x_0, x_M]$ and $[y_0, y_N]$ are extended to $[x_{-3}, x_{M+3}]$ and $[y_{-3}, y_{N+3}]$ where $\{x_m; m = -3, -2, \ldots, M+3\}$, and $\{y_n; n = -3, -2, \ldots, N+3\}$ are again equally spaced in the distance of $d_x = (x_M - x_0)/M$ and $d_y = (y_N - y_0)/N$, respectively. Consider the cubic B-spline basis $\{B_i(r); i = 1, 2, 3, 4\}$ on $[0, 1]$ such that

$$
\begin{aligned}
B_1(r) &= r^3/6 , \\
B_2(r) &= (-3r^3 + 3r^2 + 3r + 1)/6 , \\
B_3(r) &= (3r^3 - 6r^2 + 4)/6 , \\
B_4(r) &= (-r^3 + 3r^2 - 3r + 1)/6 .
\end{aligned} \tag{6.3}
$$

Thus for (x, y) in a subdivided rectangle $[x_m, x_{m+1}] \times [y_n, y_{n+1}]$, the spline function is given by

$$
h(x, y \mid C) = \sum_{i=0}^{3} \sum_{j=0}^{3} c_{m+i, n+j} B_{4-i}(r_x) B_{4-j}(r_y) , \tag{6.4}
$$

where $C = \{c_{ij}\}$ are coefficients, $r_x = (x - x_m)/d_x$ and $r_y = (y - y_n)/d_y$. Thus we consider a linear regression

$$\psi(a_i, \tau_j)/\tau_j = h(\alpha_i, \tau_j | C) + \varepsilon_{ij} , \tag{6.5}$$

where $\varepsilon_{ij} \sim N(0, \sigma^2 s_{ij}^2)$ is assumed and σ^2 is a parameter to be minimized.

Since quite a few parameters C are required to represent the 2-dimensional spline function, the maximum likelihood estimate usually produces a rapidly fluctuating surface. Thus we had to resolve two conflicting aims in surface estimation, which are to produce a good fit to the data but to avoid too much rapid local variation. A measure of the rapid local variation of a surface can be given by roughness penalties, such as the integrated squared first or second derivatives,

$$\Phi_1(h) = \int_A \left\{ \left(\frac{\partial h}{\partial x} \right)^2 + \left(\frac{\partial h}{\partial y} \right)^2 \right\} dxdy \tag{6.6}$$

and

$$\Phi_2(h) = \int_A \left\{ \left(\frac{\partial^2 h}{\partial x^2} \right)^2 + 2 \left(\frac{\partial^2 h}{\partial x \partial y} \right)^2 + \left(\frac{\partial^2 h}{\partial y^2} \right)^2 \right\} dxdy . \tag{6.7}$$

Thus for the suitable weights w_1 and w_2, the estimates of the parameters $C = (c_{ij})$ are obtained so as to maximize the penalized log likelihood

$$\log L(C) - [w_1 \Phi_1(h) + w_2 \Phi_2(h)] , \tag{6.8}$$

where

$$\log L(C) = -(IJ/2) \log \sigma^2 + \sum_{i,j} \{\psi(\alpha_i, \tau_j)/\tau_j - h(\alpha_i, \tau_j | C)\}^2 / (2\sigma^2 s_{ij}^2) , \tag{6.9}$$

for total number $IJ = 7 \times 16$ of data set (α_i, τ_j). Here we have adopted the case where $M = 3$ and $N = 4$; that is, $x_m = m/3$, $(m = -3, -2, \ldots, 6)$, and $y_n = n/4$, $(n = -3, -2, \ldots, 7)$, from which the number of parameters C is 42.

The crucial point here is the determination of weights w_1 and w_2 for penalties in (6.8). To obtain the optimal weights w_1 and w_2, we are led to a Bayesian interpretation. That is to say, the weighted sum of penalties in (6.8) is considered to be proportionate to the logarithm of the prior distribution $\pi(C | w_1, w_2)$ characterized by the hyperparameters w_1 and w_2. Since the penalties Φ_1 and Φ_2 are quadratic with respect to the parameters $C = \{c_{ij}\} = (c_k)_{k=1,2,\ldots,IJ}$, the prior π is a multivariate Gaussian distribution. To avoid the difficulty in the case where the prior is improper (this actually

takes place, since the covariance matrix Σ such that $C\Sigma C^t/2 = w_1\Phi_1(h) + w_2\Phi_2(h)$ is degenerated and $r = \text{rank}(\Sigma) = IJ - 1$ in the present case), the parameter vector $C = (c_k)_{k=1,2,\ldots,IJ}$ is divided into $(\mathbf{c}^r, c_{IJ})$ so that $\pi(\mathbf{c}^r, c_{IJ}|w_1, w_2)$ is proper with respect to $\mathbf{c}^r$, where c_{IJ} is the last (scalar) component of C and $\mathbf{c}^r$ the rest. Then we consider the marginal,

$$\mathbf{L}(w_1, w_2, c_{IJ}) = \int L(C)\pi(\mathbf{c}^r, c_{IJ}|w_1, w_2)\mathbf{dc}^r , \tag{6.10}$$

of the posterior to obtain w_1, w_2 and c_{IJ} which maximize $\mathbf{L}$. This is called the method of *type II maximum likelihood* due to Good (1965). Akaike (1980) justifies and develops this method based on the *entropy maximization principle* and defined

$$\text{ABIC} = (-2)\max \log \mathbf{L}(w_1, w_2, c_{IJ}) + 2 \times (\text{number of parameters}) , \tag{6.11}$$

in relation to the Akaike Information Criterion (AIC), here the number of parameters for the last term in the present case is 3 which stands for w_1, w_2 and c_{IJ}.

By the similar calculation method described in Ogata and Katsura (1988), we get $w_1 = 4.42 \times 10^{-6}$ and $w_2 = 5.53 \times 10^{-3}$. Under these values, the estimated coefficients $\hat{C} = (\hat{c}_{ij})$ in (6.4) are obtained and listed in Table 2. Also the graphs of $h(\alpha, \tau|\hat{C}) = \hat{\psi}(\alpha, \tau)/\tau$ are given in Fig. 2 together with the plots of the experimental results $\psi(\alpha_i, \tau_j)/\tau_j$ obtained above.

Table 2. Coefficient matrix $C = \{c_{ij}\}$ of the 2-dimensional spline function in (6.4).

j \ i	1	2	3	4	5	6
1	− 6.2628	3.1550	0.41372	3.4324	8.5895	11.245
2	3.3230	0.93849	2.1801	3.0206	9.5866	10.733
3	1.0112	2.2701	2.6552	4.7195	9.1765	13.105
4	2.7570	3.1222	4.4722	5.1175	9.8713	10.652
5	− 3.2697	8.1699	5.7940	6.2979	8.7423	11.957
6	− 7.8934	21.519	6.1032	7.8280	8.8784	12.984
7	69.293	4.8399	19.569	4.1605	13.848	1.1280

After all we have obtained the log likelihood function for the Soft-Core models

$$\log L(\alpha, \sigma; X) = -\sum_{i<j}^{N}\left(\frac{\sigma}{r_{ij}}\right)^{2/\alpha} + N\int_0^{\tau} h(\alpha, t|\hat{C})\,dt , \tag{6.12}$$

where $\tau = N\sigma^2/V$.

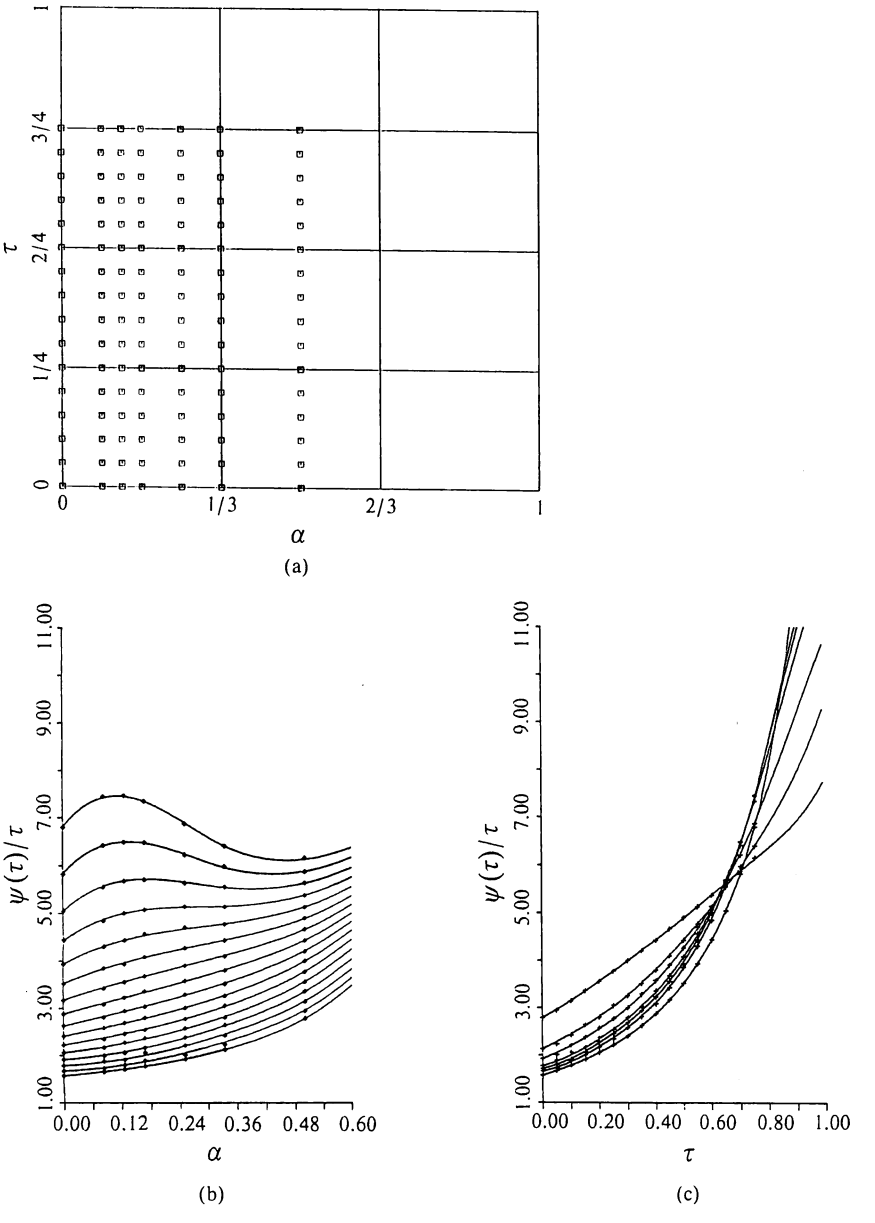


Fig. 2. Fitted $h(\alpha, \tau)$ function to $\psi(\alpha, \tau)/\tau$ in the integrand in (4.4) for reparametrized Soft-Core potential (6.1). (a) Sample points (square dots in the figure) for ψ-values in the parameter space $[0, 1) \times [0, 1)$ of (α, τ) which is divided into 3×4 identical rectangles for the definition of spline function (6.4). (b) Profiles of the spline surface along sample points for fixed τ's (solid lines) and related experimental values (sign $\diamond$). (c) Profiles of the spline surface along sample points for fixed α's (solid lines) and related experimental values (sign $+$).

7. Simulation study and analysis of observed data

In order to check the usefulness of the approximated log likelihood, we fitted this to some simulated patterns for each (α_i, τ_j). We examined whether the likelihood ratio statistics with respect to the true values distributed according to chi-square distribution with d.f. = 2. Note here for the Hard-Core case ($\alpha = 0$) that the chi-square also has 2 d.f. in spite of the single parameter τ (Ogata and Tanemura (1984), p. 507). We also examined whether the maximum likelihood estimate $(\hat{\alpha}, \hat{\tau})$ is distributed according to the normal distribution $N(0, J^{-1})$, where $J = (J_{ij})$, $i, j = 1, 2$, is the negative of the Hessian matrix which is computed instead of the Fisher information matrix.

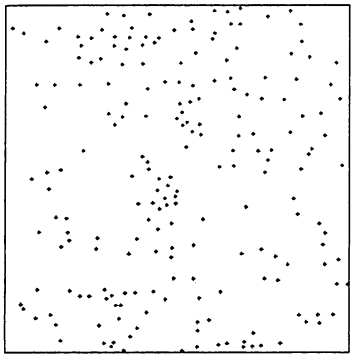
After all it is confirmed that the maximum likelihood method by the approximation works very well for $\tau = 0.05k$, $(k = 1, 2, \ldots, 15)$ and $\alpha = 2/n$, $(n = 6, 8, 12, 16, 24, \infty)$, but does not work very well for $n = 4$. This might be due to the lack of data set beyond $\alpha = 1/2$. Eventually, we can say that the approximated log likelihood is useful for the area $0 \leq \alpha < 1/2$ and $0 \leq \tau < 0.8$ (see Fig. 2).

For the analysis of observed data, we consider 6 point patterns in Fig. 3 (Ogata and Tanemura (1981*c*, 1982)). Here, for the convenience of our analysis, we calculated the total potential with the assumption of the periodic boundary. The pattern of Fig. 3(a) is a map of natural stands of seedlings and saplings of the Japanese black pine *Pinus Thunbergii* (Numata (1964)). The number of pines is $N = 204$ and the area observed is $A = 10\ \text{m} \times 10\ \text{m}$. We have used this data in some other methods (Ogata and Tanemura (1981*a*, 1984, 1985 and 1986)).

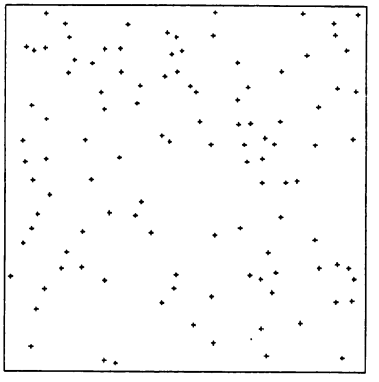
A nesting pattern of Gray Gulls, *Larus modestus*, in a 100 meter square area (Howell *et al.* (1974)) is also analyzed in Ogata and Tanemura (1981*a*) under the assumption of low density gas, and self-inhibiting interaction was indicated. It may nevertheless be interesting to investigate the same set of data by the present approximated likelihood which should be feasible in the wider range of density. The distribution of nests ($N = 110$) is shown in Fig. 3(b).

Figure 3(c) exhibits geographical data of Iowa county seats investigated by Dacey (1972). This shows a considerably regular pattern, because the county seats seem to have been placed with some planning. Dacey demonstrated his imperfect central place theory from the data.

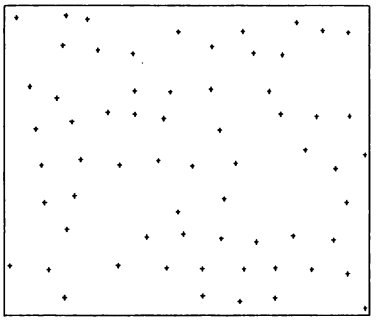
Next, a map of small steel balls ($N = 271$) in an area $28.1\ \text{mm} \times 27.9$ mm in Fig. 3(d) is investigated. The map was obtained in the following manner. We first put a number of steel balls of diameter 0.5 mm in a transparent plastic box and shook it violently by hand. Then each ball became charged with electricity and repelled the others. Therefore, the set of steel balls showed a stable and regular pattern as is shown in Fig. 3(d). The interaction between steel balls is not as simple as is expected by the



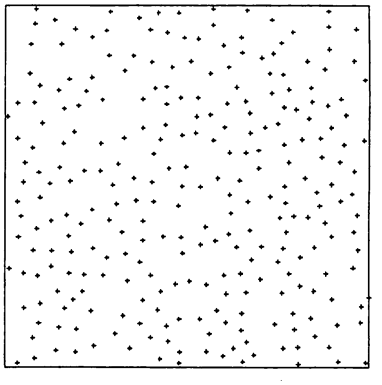
(a) Black pines: $N = 204$



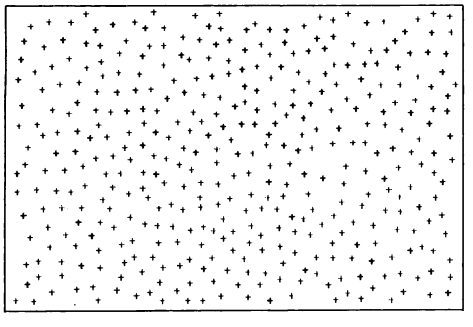
(b) Gray gull



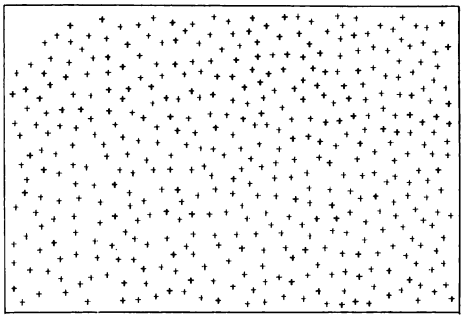
(c) Iowa county seats



(d) Steel balls



(e) P6T13



(f) M6T10

Fig. 3. Examples of point patterns. (a) The map of the seedlings and saplings of Japanese black pine *Pinus Thunbergii*: $N = 204$ in a $V = 10$ m $\times$ 10 m area (Numata (1964)). (b) The map of the nests of Gray Gull *Larus modestus*: $N = 110$ in a $V = 100$ m $\times$ 100 m area (Howell *et al.* (1974)). (c) The map of the county seats of the state of Iowa, U.S.A., $N = 65$ in a $V = 240.0$ miles $\times$ 205.714 miles area (Dacey (1972)). (d) The map of the charged steel balls: $N = 271$ in a $V = 2.81$ cm $\times$ 2.79 cm area. (e) and (f) Plots of the spatial patterns of blue cones of macaque retina obtained by Shapiro *et al.* (1985): (e) 398 cones identified P6T13 and (f) 427 cones identified M6T10, respectively, in the same volume $V = 125$ mm $\times$ 85 mm in the scale of the photomicrographs.

Coulombic law, because the wall of the plastic box itself also generates static electricity. Therefore, it will be useful to estimate a repulsive interaction potential by our procedure.

Finally, we fit the spatial patterns of blue cones in a macaque retina which are discussed in Shapiro *et al.* (1985), who recommended an elastic-ball model rather than a hard ball model. We have read the locations of the data from the photomicrographs in the paper, and Figs. 3(e) and 3(f) illustrate the identified patterns P6T13 and M6T10, respectively.

Table 3 gives the summarized results of the Soft-Core model fitting. By assessing the Hessian matrix of the log likelihood, we can also obtain the estimates of respective standard errors of the maximum likelihood estimates α and σ. It is noted that, in spite of the very different sizes of the areas and corresponding estimated parameter σ, the parameters τ, α and ψ are independent of the size of the areas and therefore compatible with each other; that is, α (or n) measures the *softness* or *hardness* of the potential, $\tau = N\sigma^2/V$ measures the *crampedness* of a pattern, and $\psi = \psi(\alpha, \tau)$ measures the *strength of repulsive force*, or *degree of the deviation from Poisson pattern.* It seems that the impression of the regularity of the patterns are consistently measured by these parameters. It is also interesting that the two patterns P6T13 and M6T10 have similar α and τ values. This is consistent with the results in second-order distance properties (L-function) and angular structure (A-function) obtained in Shapiro *et al.* (1985).

Table 3. Model fitting to the various point patterns: values in the parentheses are standard errors.

Data	σ	τ	α	n	ψ	log L
Pines 10 × 10	0.14 meters	.04	0	∞	.07	13.2
Gulls 100 × 100	2.26 (0.21) meters	.06	.29 (.22)	6.9	.12	9.0
Iowa 240 × 205.7	16.8 (1.0) miles	.37	.32 (.09)	6.2	1.31	37.5
Balls 28.1 × 27.9	1.10 (0.03) millimeters	.42	.35 (.04)	5.8	1.64	178.3
P6T13 1250 × 850	38.8 (1.0) millimeters	.56	.41 (.03)	4.9	2.81	338.7
M6T10 1250 × 850	37.3 (0.8) millimeters	.56	.36 (.03)	5.6	2.73	385.6

8. Discussions

It should be stressed that throughout the present paper we have used a region with a periodic boundary (identical to a torus) in both the simulation of point patterns and estimation of the potential function. That is to say, the distance r_{ij} between points (x_i, y_i) and (x_j, y_j) on $V = [0, T_x] \times [0, T_y]$ is given by

$$r_{ij}^2 = \min \{|x_i - x_j|, T_x - |x_i - x_j|\}^2 + \min \{|y_i - y_j|, T_y - |y_i - y_j|\}^2 .$$

This periodic boundary condition approximately realizes a state of an infinite particle system and enables the stable convergence of ψ-values or pressure at any reduced densities τ. Now it can occur in analyzing the data that a pair of points i and j crossing the predetermined periodic boundary has a very short distance r_{ij} which can affect the estimation of potential for the given equilibrium point configuration. This can be avoided by moving the boundary slightly to change the region V. Also, a suitable choice of a rectangular region from an arbitrarily shaped region of the original point pattern is sometimes useful for carrying out the present estimation procedure with a periodic boundary.

Thus a problem arises as to how to adjust the suitable boundary of a rectangular region. Let us specifically formulate the problem and suggest a solution to it in the following way. Let $V = [0, T_x] \times [0, T_y]$ be a tentative region including points $X = \{(x_i, y_i), i = 1, 2, \ldots, N\}$, and consider a perturbation of the region $V_{\xi,\eta} = [0, T_x + \xi] \times [0, T_y + \eta]$. Then a distance r_{ij} on the torus becomes $r_{ij}(\xi, \eta)$ whose square is

$$\begin{aligned} r_{ij}(\xi, \eta)^2 = {} & \min \{|x_i - x_j|, T_x + \xi - |x_i - x_j|\}^2 \\ & + \min \{|y_i - y_j|, T_y + \eta - |y_i - y_j|\}^2 , \end{aligned}$$

where choices of the minimum above depend on the configuration of the pair of points. Then the log likelihood is defined, using the formula (6.12), by

$$\log L(\alpha, \sigma, \xi, \eta; X) = -\sum_{i<j}^{N} \left(\frac{\sigma}{r_{ij}(\xi, \eta)} \right)^{2/\alpha} + N \int_0^{\tau(\xi,\eta)} h(\alpha, t \mid \hat{C}) dt ,$$

where $\tau(\xi, \eta) = N\sigma^2 / V_{\xi,\eta}$ and $h(\alpha, t \mid \hat{C})$ is the spline function in (6.4) with coefficient matrix $\hat{C}$ given in Table 2. Thus the suitable estimates of the α, σ, ξ and η are obtained simultaneously by maximizing the above likelihood function.

Let us just see the performance of the above mentioned method. We have a simulated pattern of $N = 500$ points in a $V = 1000.0 \times 1000.0$ region

by the Soft-Core potential with $\alpha = 0.25$ and $\tau = 0.20$ so that $\sigma = \sqrt{\tau V/N} = 20$. The maximum likelihood estimates are $\hat{\alpha} = 0.191$, $\hat{\sigma} = 20.0$, $\hat{\xi} = -0.147$ and $\hat{\eta} = 0.198$. The maximum log likelihood is 161.46 (AIC = -314.9). On the other hand, by the restriction of $\xi = 0$ and $\eta = 0$, the maximum log likelihood is 160.66 (AIC = -317.3) with $\hat{\alpha} = 0.185$ and $\hat{\sigma} = 19.9$. This means that the original boundary for the periodicity is appropriate for the equilibrium pattern of the considered potential function. This result is sensible because the data is simulated in area V.

Finally, there may be a question as to why we seek approximations which are globally applicable, instead of either locally fitting quadratic surfaces near the maximum likelihood estimates to solve Newton-Raphson's equations such as in Penttinen's method (1984) or using the pseudo-likelihood method (Besag *et al.* (1982)). The reason is simply to get value of the log likelihood itself for comparing the goodness-of-fit with alternative models other than Soft-Core models; see related papers of the present authors.

Acknowledgements

The authors would like to express their gratitude to their colleague Koichi Katsura for his generous help in preparing Fig. 2. They also would like to thank the referee for comments which were useful for the revision of the present paper.

REFERENCES

Akaike, H. (1980). Likelihood and Bayes procedure, *Bayesian Statistics*, (eds. J. M. Bernardo *et al.*), 141–166, University Press, Valencia, Spain.

Besag, J., Milne, R. and Zachary, S. (1982). Point process limits of lattice processes, *J. Appl. Probab.*, **19**, 210–216.

Dacey, M. F. (1972). Regularity in spatial distributions: a stochastic model of the imperfect central place plane, *Statistical Ecology*, (eds. G. Patil, E. C. Pielou and W. E. Waters), Vol. 1, 287–309.

Gates, D. J. and Westcott, M. (1986). Clustering estimates for spatial point distributions with unstable potentials, *Ann. Inst. Statist. Math.*, **38**, 123–135.

Good, I. J. (1965). *The Estimation of Probabilities*, M.I.T. Press, Cambridge, Massachusetts.

Good, I. J. and Gaskins, R. A. (1971). Nonparametric roughness penalties for probability densities, *Biometrika*, **58**, 255–277.

Hoover, W. G., Gray, S. G. and Johnson, K. W. (1971). Thermodynamic properties of the fluid and solid phases for inverse power potentials, *J. Chem. Phys.*, **55**, 1128–1136.

Howell, T. R., Araya, B. and Millie, W. R. (1974). Breeding biology of the Gray Gull, *Larus modestus*, *Univ. Calif. Publ. Zool.*, **104**, 1–57.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H. and Teller, E. (1953). Equation of state calculations by fast computing machines, *J. Chem. Phys.*, **21**, 1087–1092.

Numata, M. (1964). Forest vegetation, particularly pine stems in the vicinity of Choshi—flora and vegetation at Choshi, Chiba Prefecture, IV, *Bull. Choshi Marine Laboratory*, No. 6, 27–37, Chiba University (in Japanese).

Ogata, Y. and Katsura, K. (1988). Likelihood analysis of spatial inhomogeneity for marked point patterns, *Ann. Inst. Statist. Math.*, **40**, 29–39.

Ogata, Y. and Tanemura, M. (1981*a*). Estimation of interaction potentials of spatial point patterns through the maximum likelihood procedure, *Ann. Inst. Statist. Math.*, **33**, 315–338.

Ogata, Y. and Tanemura, M. (1981*b*). A simple simulation method for quasi-equilibrium point patterns, Research Memorandum No. 210, The Institute of Statistical Mathematics, Tokyo, Japan.

Ogata, Y. and Tanemura, M. (1981*c*). Approximation of likelihood function in estimating the interaction potentials from spatial point patterns, Research Memorandum No. 216, The Institute of Statistical Mathematics, Tokyo, Japan.

Ogata, Y. and Tanemura, M. (1982). Likelihood analysis of spatial point patterns, Research Memorandum No. 241, The Institute of Statistical Mathematics, Tokyo, Japan.

Ogata, Y. and Tanemura, M. (1984). Likelihood analysis of spatial point patterns, *J. Roy. Statist. Soc. Ser. B*, **46**, 496–518.

Ogata, Y. and Tanemura, M. (1985). Estimation of interaction potentials of marked spatial point patterns through the maximum likelihood method, *Biometrics*, **41**, 421–433.

Ogata, Y. and Tanemura, M. (1986). Likelihood estimation of interaction potentials and external fields of inhomogeneous spatial point patterns, *Pacific Statistical Congress*, (eds. I.S. Francis, B. F. J. Manly and F. C. Lam), 150–154, North-Holland, Amsterdam.

Penttinen, A. (1984). *Modelling Interactions in Spatial Point Patterns: Parameter Estimation by the Maximum Likelihood Method*, Jyväskylä Studies in Computer Science, Economics and Statistics, **7**, University of Jyväskylä, Jyväskylä, Finland.

Ree, F. H. and Hoover, W. G. (1967). Seventh virial coefficients for hard spheres and hard disks, *J. Chem. Phys.*, **46**, 4181–4197.

Ripley, B. D. (1979). Simulating spatial patterns: dependent samples from a multivariate density, *Appl. Statist.*, **28**, 109–112.

Shapiro, M. B., Schein, S. J. and Monasterio, F. M. (1985). Regularity and structure of the spatial pattern of blue cones of macaque retina, *J. Amer. Statist. Assoc.*, **80**, 803–814.

Swol, F. van, Woodcock, L. V. and Cape, J. N. (1980). Melting in two dimensions: determination of phase transition boundaries, *J. Chem. Phys.*, **73**, 913–922.

Wood, W. W. (1968). Monte Carlo studies of simple liquid models, *Physics of Simple Liquids*, (eds. H. N. V. Temperley, J. S. Rowlinson and G. S. Rushbrooke), 115–230, North-Holland, Amsterdam.