Ann Inst Stat Math (2014) 66:687–702
DOI 10.1007/s10463-013-0431-z

Identification and estimation of superposed Neyman–Scott spatial cluster processes

Ushio Tanaka · Yosihiko Ogata

Received: 15 November 2012 / Revised: 9 June 2013 / Published online: 14 February 2014

Abstract This paper proposes an estimation method for superposed spatial point patterns of Neyman–Scott cluster processes of different distance scales and cluster sizes. Unlike the ordinary single Neyman–Scott model, the superposed process of Neyman–Scott models is not identified solely by the second-order moment property of the process. To solve the identification problem, we use the nearest neighbor distance property in addition to the second-order moment property. In the present procedure, we combine an inhomogeneous Poisson likelihood based on the Palm intensity with another likelihood function based on the nearest neighbor property. The derivative of the nearest neighbor distance function is regarded as the intensity function of the rotation invariant inhomogeneous Poisson point process. The present estimation procedure is applied to two sets of ecological location data.

Keywords Contact distances · Likelihood functions · Multi-type Neyman–Scott processes · Nearest neighbor distance function · Palm intensity

U. Tanaka (✉)
Rikkyo University, 3-34-1 Nishi-Ikebukuro, Toshima-ku,
Tokyo 171-8501, Japan
e-mail: utanaka@rikkyo.ac.jp

Y. Ogata
Institute of Industry Science, University of Tokyo, 4-6-1 Komaba, Meguro-ku,
Tokyo 153-8505, Japan

Y. Ogata
The Institute of Statistical Mathematics, 10-3 Midori-cho,
Tachikawa, Tokyo 190-8562, Japan
e-mail: ogata@ism.ac.jp

1 Introduction

The Neyman–Scott process (see Neyman and Scott (1958)), originally proposed as the model of galaxy distribution, is well-known cluster point process. The model first generates unobservable parent points according to a homogeneous Poisson process. Then, each parent point generates a random number of descendants that scatter around the parent location according to a spatial density function. The parameter estimation method usually uses the least squares of the discrepancies between the values of the empirical L-function and the theoretical L-function corresponding to a parameterized Neyman–Scott process model (e.g., Diggle (1983, p. 74), Cressie (1993, p. 666) and Stoyan and Stoyan (1996)), where the L-function is the normalized square root of the K-function of Ripley (1977). As an alternative to the L-function, Stoyan and Stoyan (1996) recommend the use of the pair-correlation function $g(r)$ of the point process, essentially the derivative of the K-function of Ripley (1977), to reduce the dependencies of residuals in the sum of squares of the residuals.

For sensitive parameter estimation and model selection, it would be advantageous to obtain the maximum likelihood estimates. However, this has not been possible owing to the following difficulties: (1) the data-set does not specify what events are the parents (cluster centers), (2) the relationship between the clustered points (descendants) and the attribution of their cluster center are not specified in the given data-set, and (3) the ranges of clusters are overlapping with each other so that their ranges are not specific. Indeed, Baudin (1981) showed that the likelihood function cannot be described in an analytically closed form. Therefore, instead of the ordinary maximum likelihood estimation, Tanaka et al. (2008b) proposed a maximum likelihood procedure based on the Palm intensity function, which is proportional to the pair-correlation function between descendant points. Roughly speaking, the Palm intensity does not address the configurations of given point coordinates of data but their difference vectors.

Now, suppose that we have a few Neyman–Scott processes that are independent of one another, that is, they have different parents (cluster centers) intensities, mean cluster sizes, and location distributions of the descendants relative to their parent. In this study, we focus on estimating all parameters of each component process by observing the superposed configuration of these descendants.

However, the Palm intensity alone cannot identify such a superposed Neyman–Scott process (see Tanaka et al. (2008b)). In this study, we shall overcome this difficulty by the additional use of a likelihood function based on the nearest neighbor distance (NND) function, that is, the shortest distance from a given location to the nearest point.

As applications, we will apply the present estimation procedure to two plant location data-sets obtained from Cressie (1993) and Diggle (1983).

2 Maximum Palm-likelihood estimation

2.1 Preliminaries on clustering point process models

For the remainder of this study, we assume that the processes have the following characteristics: we consider two Neyman–Scott processes with different parameter values. We restrict our work to two-dimensional Euclidean space (see Tanaka et al.

(2008b) for a detailed account). The superposed Neyman–Scott spatial cluster process is defined to be the union of all descendant points in both processes. Furthermore, the observed window is prescribed to be a unit square with periodic boundary conditions (a torus), on which the considered processes are stationary and isotropic. Finally, we restrict ourselves to the case where the density functions are rotation invariant, two-dimensional Gaussian distributions with different scale parameters relative to each other, each of which is called a Thomas process (see Thomas (1949)).

2.2 Thomas processes and their superposition

Let the parents of the two processes be distributed according to homogeneous Poisson processes with intensity rates μ_1 and μ_2, and the numbers of descendants have a Poisson distribution with mean values ν_1 and ν_2. Then, each descendant is located close to its parent (cluster center) and is distributed independently according to density functions $q_{\sigma_1}(x, y)$ and $q_{\sigma_2}(x, y)$ with parameters σ_1 and σ_2, respectively, where (x, y) is the location relative to the corresponding parent. Here, we restrict ourselves to the case in which the density functions q_{σ_1} and q_{σ_2} are two-dimensional Gaussian distributions $N(\mathbf{0}, {\sigma_i}^2 I)$, where I is a two-dimensional identity matrix. These are called Thomas processes (see Thomas (1949)). Because the distribution is rotation invariant, the polar coordinate representation with respect to the distance r is given as

$$q_{\sigma_i}(r) = \frac{r}{{\sigma_i}^2} \exp\left(-\frac{r^2}{2{\sigma_i}^2}\right), \quad 0 \le \theta < 2\pi, \quad i = 1, 2.$$

We consider the superposed point pattern of these two Thomas processes (see Fig. 1). Here, we should note that the superposed Thomas process is different from the Neyman–Scott process with the mixture of Gaussian distributions $\alpha q_{\sigma_1}(r) + (1 - \alpha) q_{\sigma_2}(r)$, $0 < \alpha < 1$ (see Tanaka et al. (2008b)).

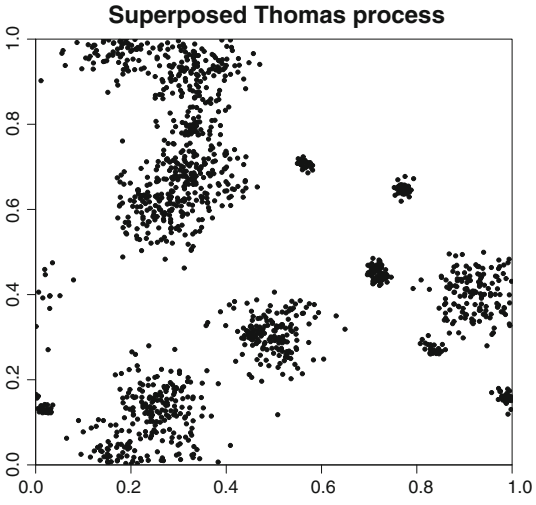

Fig. 1 A simulated realization of a superposed Thomas process with different parameter sets

2.3 Palm intensities of the cluster process models

The statistical methods of the present paper are based on the second-order properties of point processes (e.g., Daley and Vere-Jones (2003), Chapter 8). Of particular importance is the Palm intensity function $\lambda_o(\cdot)$ (see Ogata and Katsura (1991)), or the second-order intensity. The Palm intensity function can be heuristically described as follows: let $\mathbf{x}$ be any point in $\mathbb{R}^2$ at a distance r from the origin $\mathbf{o}$. Then, the occurrence rate at $\mathbf{x}$, provided that a point is at $\mathbf{o}$, is

$$\lambda_o(\mathbf{x})\mathbf{dx} = \Pr(\{N(\mathbf{dx}) = 1 | N(\{\mathbf{o}\}) = 1\}) \tag{1}$$

for an infinitesimal set $\mathbf{dx}$, where N stands for a counting measure. Given stationarity and isotropy, $\lambda_o(\mathbf{x})$ depends only on the distance r of $\mathbf{x}$ from $\mathbf{o}$, and the function is then written as $\lambda_o(r)$.

The relationships between the Palm intensity function and both the pair-correlation function $g(r)$ and the K-function are $\lambda_o(r) = \lambda g(r)$ and $\lambda_o(r) = \lambda K'(r)/(2\pi r)$, respectively, where K' is the first-order derivative of K with respect to the distance r.

For both Thomas models, from (Daley and Vere-Jones, 1988, Sect. 8.1), we know that

$$\lambda_o{}^i(r) = \mu_i \nu_i + \frac{\nu_i}{4\pi {\sigma_i}^2} \exp\left(-\frac{r^2}{4{\sigma_i}^2}\right), \quad i = 1, 2. \tag{2}$$

Figure 2 shows a simulated realization of the Palm intensity function, that is, the superposition of all difference vectors between point coordinates.

Moreover, by a simple calculation, the Palm intensity functions of superposed Thomas processes with distance density functions is obtained as

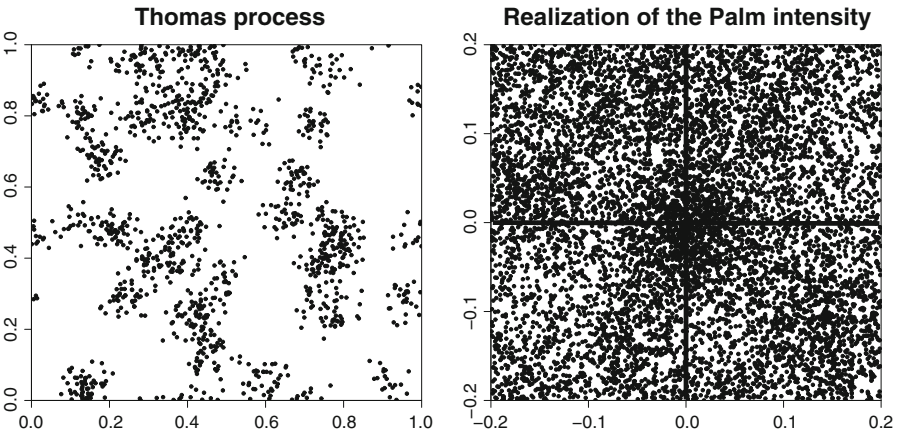

Fig. 2 A simulated point pattern of a Thomas process (*left*) and the realization of its Palm intensity (*right*). The *right panel* is obtained by superposing the point patterns, and all point coordinates are shifted so that each point is at the origin. In the case of a Neyman–Scott clustering process, the Palm intensity is most dense around the origin of the resulting point pattern

$$\lambda_o(r) = \lambda + \frac{a\nu_1}{4\pi{\sigma_1}^2}\exp\left(-\frac{r^2}{4{\sigma_1}^2}\right) + \frac{(1-a)\nu_2}{4\pi{\sigma_2}^2}\exp\left(-\frac{r^2}{4{\sigma_2}^2}\right), \tag{3}$$

where $\lambda = \mu_1\nu_1 + \mu_2\nu_2$ is the total intensity of the process and $a = \mu_1\nu_1/\lambda$ is the fraction of the first Thomas process relative to the total intensity of the superposed process.

2.4 log-Palm-likelihood

Let $\{\psi_i\}$ be each individual vector coordinate of a Neyman–Scott process on the torus $W = [0, 1]^2$. Tanaka et al. (2008b) assumed that the distribution of the differences $\Psi_{i,j} \equiv \psi_j - \psi_i$ for $i = 1, \ldots, n$, $j = 1, \ldots, n$, $i \neq j$ were well approximated by an inhomogeneous Poisson process that is rotation invariant and has intensity $N(W)\lambda_o(r)$ centered at $\mathbf{o}$, as illustrated in Fig. 2. The corresponding log-likelihood function of the point pattern, called the *Palm-likelihood function*, is then

$$\log L(\mu, \nu, \tau) = \sum_{\{i,j;\, i\neq j,\, r_{ij}<R\}} \log\{N(W)\cdot\lambda_o(r_{ij})\} - N(W)\int_0^R \lambda_o(r)\cdot 2\pi r\, dr. \tag{4}$$

Here, the sum is taken over all pairs i, j such that the distance r_{ij} between ψ_i and ψ_j is smaller than R, where R is sufficiently greater than the range of correlation of the process, i.e., the minimum value of r such that $\lambda_o(r) = \lambda$. The parameter R must be estimated before fitting the parametric model, which is not difficult if one uses a non-parametric estimate of the pair-correlation function. Assuming the distances are measured with respect to the periodic boundary condition, in the numerical likelihood calculations of the present study, the window is always transformed to the standardized rectangle $W = [0, 1] \times [0, A]$, where $A \geq 1$ and R is taken as 1/2, which means that the non-constant range of the Palm intensity function is assumed to be less than 1/4 of the rectangle's shorter side. Tanaka et al. (2008b) demonstrated the accuracy and robustness of the maximum Palm-likelihood estimates (MPLEs) through simulations of various Neyman–Scott models with general distance distributions. The computing and simulation codes are provided in Tanaka et al. (2008a).

For the present superposed Thomas model, the log-Palm-likelihood is given by

$$\begin{aligned}
&\log L(\mu_1, \nu_1, \sigma_1, \mu_2, \nu_2, \sigma_2)\\
&= \sum_{\{i,j;\, i\neq j,\, r_{ij}<R\}} \log\left\{\lambda + \frac{a\nu_1}{4\pi{\sigma_1}^2}\exp\left(-\frac{{r_{ij}}^2}{4{\sigma_1}^2}\right) + \frac{(1-a)\nu_2}{4\pi{\sigma_2}^2}\exp\left(-\frac{{r_{ij}}^2}{4{\sigma_2}^2}\right)\right\}\\
&\quad -N(W)\left[\pi\lambda R^2 + a\nu_1\left\{1-\exp\left(-\frac{R^2}{4{\sigma_1}^2}\right)\right\} + (1-a)\nu_2\left\{1-\exp\left(-\frac{R^2}{4{\sigma_2}^2}\right)\right\}\right],
\end{aligned} \tag{5}$$

where $\lambda = \mu_1\nu_1 + \mu_2\nu_2$ is the total intensity of the process and $a = \mu_1\nu_1/\lambda$ is the fraction of the first Thomas process relative to the total intensity. The MPLEs of the parameters $\mu_1, \nu_1, \sigma_1, \mu_2, \nu_2$, and σ_2 of the Thomas models are those which maximize the function given in (5).

However, non-unique MPLE solutions are anticipated because of the identification problem of the Palm intensity (3). This problem arises from the fact that the respective values λ, $a\nu_1$ and $(1-a)\nu_2$ can take the same values for different sets of μ_1, μ_2, ν_1, ν_2 and a with $a = \mu_1\nu_1/\lambda$ (see Table 1 and Table 3 for numerical examples), while the MPLEs of the scaling parameters σ_1 and σ_2 are uniquely determined. This means that the MPLE of $\hat{\mu}_1(a)$, $\hat{\nu}_1(a)$, $\hat{\mu}_2(a)$ and $\hat{\nu}_2(a)$ are uniquely determined once the ratio a is fixed.

Therefore, another criterion is needed to estimate the ratio a in order to determine each component of the Thomas processes. In the next section, we will use the NND function for this purpose.

3 Nearest neighbor distance maximum likelihood estimation

3.1 Nearest neighbor distance likelihood

Let $\mathbf{X}$ be a Neyman–Scott process, and let $\mathrm{dist}(\psi, \mathbf{X})$ denote the shortest distance from an arbitrary location ψ to the nearest point of the process $\mathbf{X}$. As such, the cumulative distribution function $F(r) = \Pr(\{\,\mathrm{dist}(\psi, \mathbf{X}) \le r\,\})$ is denoted as the spherical contact distance function or empty space function. Then, the location of a point $\psi \in \mathbf{X}$, $G(r) = \Pr(\{\,\mathrm{dist}(\psi, \mathbf{X}\backslash\psi) \le r | \{\psi\}\,\})$ is denoted as the NND function (see Baddeley et al. (2007)).

Let ${F_a}^i(r)$ and ${G_a}^i(r)$ for $i = 1, 2$ be a spherical contact distance function and an NND function, respectively, for each individual Neyman–Scott process associated with the set of MPLE parameters restricted for an arbitrary $a = \mu_1\nu_1/\lambda$, as described in the previous section. Then, following Van Lieshout and Baddeley (1996), $G_a(r)$ of the superposed process satisfies the relation

$$\begin{aligned} 1 - G_a(r) = a\left\{1 - {G_a}^1(r)\right\}\left\{1 - {F_{1-a}}^2(r)\right\} \\ + (1-a)\left\{1 - {F_a}^1(r)\right\}\left\{1 - {G_{1-a}}^2(r)\right\} \end{aligned} \tag{6}$$

for any $r \ge 0$.

Let g_a be the derivative of G_a, which plays a central role in the maximum NND-likelihood procedure, as described later.

For parametric statistical analysis, we assume that the difference coordinates of all nearest neighboring pairs are well approximated by an inhomogeneous Poisson process with the intensity function $g_a(r)$, which is centered at the origin. The approximation relies on limit theorems that demonstrate that properly scaled superposition (stacking) of nearly independent point patterns results in a Poisson process (Daley and Vere-Jones, 2008, Sect. 11.2).

The function $g_a(r)$ can be regarded as the inhomogeneous Poisson intensity function of r in the normalized distance space $[0, 1]$. Then, the log-likelihood function $\log L(a)$ is of the following form:

$$\log L(a) = \sum_{j=1}^{N(W)} \log g_a(r_j) - G_a(1/2), \quad 0 \le a \le 1, \tag{7}$$

where r_j denotes the NND (contact distance) for each individual, and the number $1/2$ is owing to the periodic boundary condition over the unit square. Here, we assume that the range of the NND is sufficiently less than 1/2 to satisfy $G_a(1/2) = N(W)$ for all a. We call the function given in (7) the log-NND-likelihood function.

In this procedure, we assume that the log-NND-likelihood function is smooth and unimodal with respect to the parameter a, at least in the neighborhood of the maximum log-NND-likelihood estimate. The case where ${F_a}^i(r)$ and ${G_a}^i(r)$ for $i = 1, 2$ are independent of a meets the required conditions. However, it is not so easy to show this accurately, especially under the restricted parameter space of the MPLE solution, as stated in the previous section. Through simulation experiments, we will see that the conditions of smoothness and local concavity hold at least in the neighborhood of the local maximum, even under such MPLE restrictions. In the following section, we describe the algorithm to attain the maximum log-NND-likelihood function under the restricted parameter space of the MPLE solution.

3.2 Calculation of the log-NND-likelihood function

For a superposed Neyman–Scott process, analytic calculation of the log-NND-likelihood function is difficult owing to its complicated explicit form of an NND function associated with the ratio a. The log-NND-likelihood function requires a complicated analytic expression of F^i and J^i for Neyman–Scott processes (see Stoyan and Stoyan (1994) and Van Lieshout and Baddeley (1996)). Thus, we are forced to numerically evaluate the log-NND-likelihood function given in (7). The proposed estimation procedure is summarized in the following steps:

1. Obtain the unique solution of the MPLEs $(\hat{\lambda}, \hat{c}_1, \hat{c}_2, \hat{\sigma}_1, \hat{\sigma}_2)$, where $c_1 = a\nu_1$ and $c_2 = (1 - a)\nu_2$, as described in Sect. 2.4.
2. Set a value for the ratio $0 \le a \le 1$, so that $\{\mu_i(a), \nu_i(a), \sigma_i\}$ for $i = 1, 2$ are uniquely determined.
3. Generate two Thomas configurations from the above parameters, and superpose them. Then, calculate the NND r_j for respective points j.
4. Repeat step 3 on the order of 100 times until the histogram of the estimate of the NND density function $\hat{g}_a(r_j)$ and cumulative function $\hat{G}_a(r_j)$ start to show a well defined and consistent shape. First, the number of NND points are calculated for the estimation of $\hat{g}_a(r_j)$ in the disjoint intervals of the distances centered at $0.05j$. Then, the points are summed up for the estimation of $\hat{G}_a(r_j)$ for the calculation of the log-NND-likelihood function.

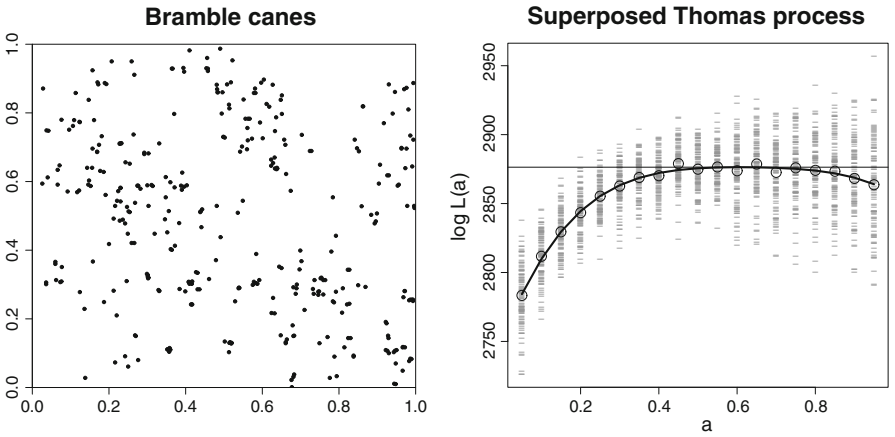

Fig. 3 The *left panel* shows locations of BC data points. In the right panel, signs "−" mark the log-NND-likelihood values of the superposed Thomas model estimated from 100 simulations at each a, and the circle marks their mean value. The *curved* and *horizontal lines* are the best-fitted polynomial function and its maximum, respectively

5. Calculate the log-NND-likelihood function given in (7) for the NND r_j for respective points of the data j.
6. Repeat steps 3–5 to determine the variability of the log-NND-likelihood function and to use these for the estimation of the log-NND-likelihood function at the value a.
7. Go to step 2 and repeat the above steps for different a-values to search for the maximum log-NND-likelihood function.

To delineate the smooth log-NND-likelihood function, $\log L(a)$, with respect to a given in (7), the least squares method is applied for the simulated samples in step 6 by fitting polynomials to the data, the optimal order of which is determined by the AIC (see Akaike (1974)).

4 Applications to ecological data sets

4.1 Case study 1: Bramble canes data

The left panel of Fig. 3 shows the locations of 359 newly emergent bramble canes (BCs), as discussed in Diggle (1983). The BC data points in the figure are scaled in the unit square while the original data were collected in a 9m × 9m square (see Hutchings (1979)).

We first apply the MPLE method to the data-set to restrict the most likely parameter subspace, which is given in Table 1. Table 1 lists the estimates of $\lambda = \mu_1\nu_1 + \mu_2\nu_2$, $a\nu_1$, $(1-a)\nu_2$, σ_1 and σ_2 for the superposed Thomas model. In particular, $\hat{\lambda} = 349.37$, which is close to the total number of data points ($N(W) = 359$). Here, Tanaka et al. (2008b) showed that regardless of the non-identifiability problem, this superposed Thomas model is better fitted by the AIC using the MPLE than the single Neyman–Scott process using the mixture of Gaussian distributions $\alpha q_{\sigma_1}(r) + (1-\alpha)q_{\sigma_2}(r)$ with any $0 < \alpha < 1$.

Table 1 MPLE of the superposed Thomas processes applied to the BC data points

Model	Superposed Thomas process				
Parameters	λ	$a\nu_1$	$(1-a)\nu_2$	σ_1	σ_2
MPLE	349.37	0.91	4.57	0.00355	0.0477

Table 2 Estimates by the maximum NND-likelihood estimate $\hat{a}$ together with the MPLE values (see Table 1) for the BC data points

Model	Superposed Thomas process						
Parameters	μ_1	μ_2	ν_1	ν_2	σ_1	σ_2	a
Estimates	137.6	12.3	1.52	11.4	0.00355	0.0477	0.60

Now, to obtain unique values of μ_1, ν_1, μ_2, and ν_2, we need to determine the a-value in $[0, 1]$. For this, we applied the maximum NND-likelihood estimation procedure described above to derive the most likely unique values of the superposed Thomas processes. Thus, we calculated the log-NND-likelihood function $\log L(a_i)$ for $a_i = 0.05i$, $i = 1, 2, \ldots, 19$, as explained in Sect. 3.

In the right panel of Fig. 3, signs "−" at each a_i mark the estimates of $\log L(a_i)$ given in (7) from 100 simulations for the superposed Thomas model with constraint of parameters as given in Table 1, and each circle indicates their mean value. Because these vary considerably, we fit a polynomial function by the least squares method to all of the simulated data $\{\log L_j(a_i);\ j = 1, \ldots, 100,\ i = 1, \ldots, 19\}$. The best-fitted order of the polynomial was determined by the AIC (see Akaike (1974)). The horizontal line indicates the maximum value of the polynomial curve, which was attained at approximately $\hat{a} = 0.60$. From this, we derived the solution of all parameters of the superposed Thomas model provided in Table 2.

To examine the reproducibility of the estimated model, we repeat the same procedures for a simulated point pattern of the superposed Thomas process with the parameters given in Table 2. Figure 4 shows one of the simulated point patterns (left panel) and the log-NND-likelihood values, as described above. The simulated point pattern looks similar to the BC data points in Fig. 3, and the maximum log-NND-likelihood function for this data is attained at $\hat{a} = 0.6$ again.

Tanaka et al. (2008b) graphically demonstrated the goodness-of-fit of the estimated model using the empirical and theoretical Palm intensity instead of the K-statistics. Here, we show density of the NND function $\hat{g}_a(r)$ to confirm how the NND-statistics of the estimated model are consistent with that of the BC data points. Figure 5 shows the estimated densities of the nearest neighbor distances $\{\hat{g}_a\}$ from simulated data using the superposed Thomas models with the MPLE estimates given in Table 1 for respective a-values, as indicated in the figure caption. Comparing with the empirical NND estimate circles from the BC data points, we see that the maximum NND estimate solid white line is fairly unbiased. From these, the model with $\hat{a} = 0.6$ seems better fitted to the data than other a-values.

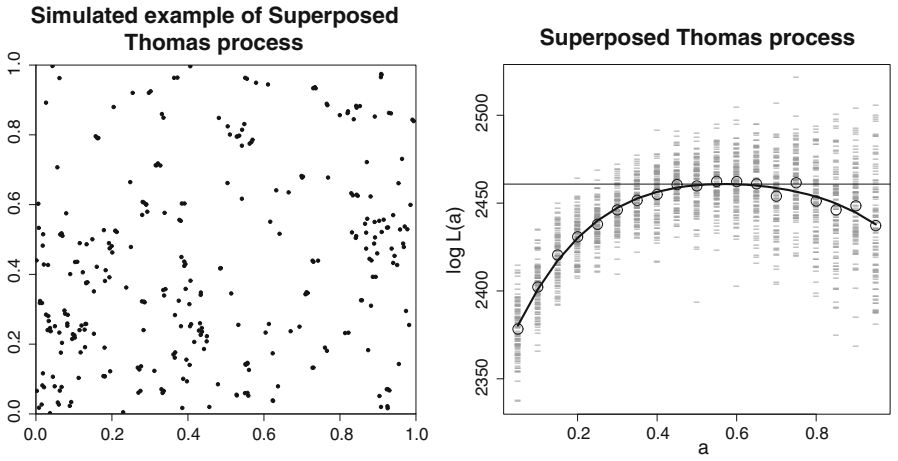

Fig. 4 The *left panel* shows one of the simulated point patterns in Table 2, and the *right panel* displays a re-estimation summary for the simulated data, where the layout is equivalent to that shown in Fig. 3

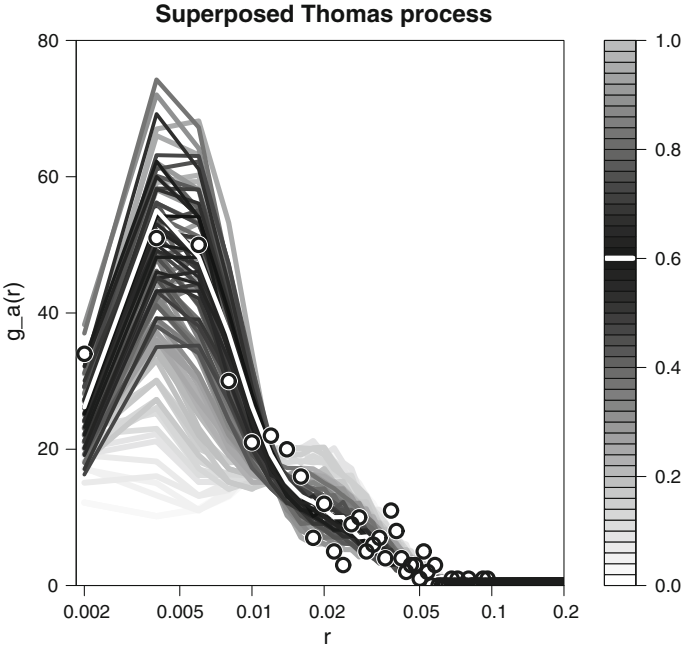

Fig. 5 Estimated densities of nearest neighbor distances $\{g_a\}$ against r in log scale from simulated data using the superposed Thomas models with the MPLE estimates given in Table 1, for respective a-values raging $a = 0.01i$ for $i = 1, \ldots, 99$. The *gray scales of the lines* correspond to the a-values in the gray scale table shown to the right of the figure. The *circles* are histograms from the BC data points, and the *white solid line* in the panel is from the maximum log-NND-likelihood estimate $\hat{a} = 0.6$ given in Table 2

4.2 Case study 2: The Longleaf Pine data

Figure 6 shows the locations of 584 Longleaf Pine Trees taken from Cressie (1993). Also pertinent to this study is Rathbun and Cressie (1994). The longleaf pine (LLP) data point coordinates were scaled to a unit square from the original scale size (200m $\times$ 200m).

As in the previous case study, we first applied the MPLE method to the data-set in order to restrict the most likely parameter subspace. Table 3 lists the MPLEs of $\lambda = \mu_1\nu_1 + \mu_2\nu_2$, $a\nu_1$, $(1-a)\nu_2$, σ_1 and σ_2 for the superposed Thomas model. Here, we should note that this superposed Thomas process with the MPLE given in Table 3 is better fitted than the Neyman–Scott process using the mixture of Gaussian distributions $\alpha q_{\sigma_1}(r) + (1-\alpha)q_{\sigma_2}(r)$ for any $0 < \alpha < 1$ in the sense of the AIC (see Tanaka et al. (2008b)).

To obtain a set of unique values of μ_1, ν_1, μ_2, and ν_2, we need to determine the a-value in [0, 1]. Thus, we calculated the log-NND-likelihood function $\log L(a_i)$ for $a_i = 0.05i$, $i = 1, 2, \ldots, 19$, as explained in Sect. 3.

In the right panel of Fig. 6, signs "$-$" at each a_i indicate the estimates of $\log L(a_i)$ given in (7) from 100 simulations for the superposed Thomas model with the constraint of parameters, as given in Table 3. Because these vary considerably, we fit a polynomial function by the least squares method to all simulated data

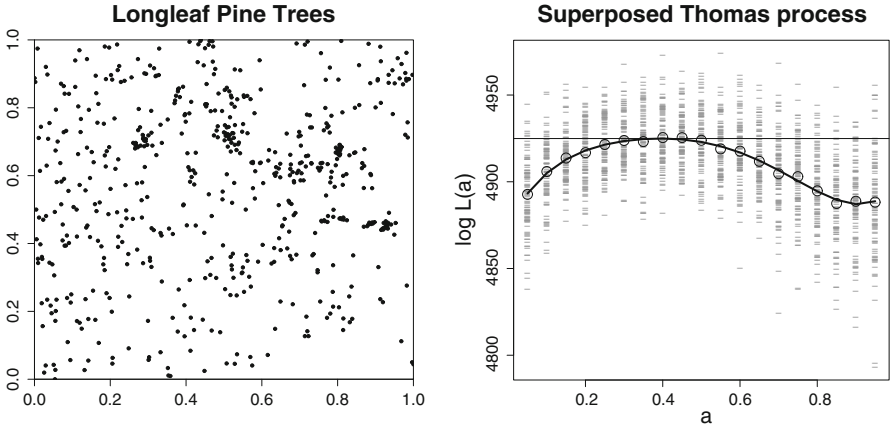

Fig. 6 The *left panel* shows locations of the LLP data points. In the *right panel*, signs "$-$" marks the log-NND-likelihood values of the superposed Thomas model estimated from 100 simulations at each a, and the circle indicates their mean value. The *curved* and *horizontal lines* mark the best-fitted polynomial function and its maximum, respectively

Table 3 MPLE of the superposed Thomas processes applied to the LLP data points

Model	Superposed Thomas process				
Parameter	λ	$a\nu_1$	$(1-a)\nu_2$	σ_1	σ_2
MPLE	562.11	2.93	24.0	0.0134	0.136

Table 4 Estimates by the maximum NND-likelihood estimate $\hat{a}$ together with the MPLE values (see Table 3) for the LLP data points.

Model	Superposed Thomas process						
Parameter	μ_1	μ_2	ν_1	ν_2	σ_1	σ_2	a
Estimates	30.9	8.45	7.28	39.9	0.0134	0.136	0.40

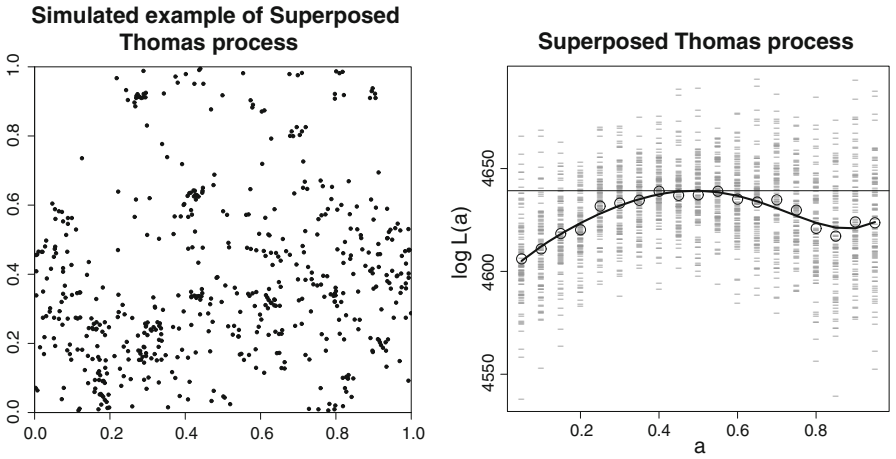

Fig. 7 The *left panel* shows a simulated point pattern using the data given in Table 4, and the *right panel* is the re-estimation for the data where the signs and curves are equivalent to that described in Fig. 6

$\{\log L_j(a_i);\ j = 1, \ldots, 100, i = 1, \ldots, 19\}$. The best-fitted order of the polynomial is determined by the AIC (see Akaike (1974)). Here, we see that the mean values of the estimates at respective a_j are on the estimated curve. The horizontal line indicates the maximum value of $\log L(a)$. From computing the corresponding value of the estimated $\log L(a)$ at each a, the maximum of the polynomial curve is attained at $\hat{a} = 0.40$. This value is used to determine the solution of all parameters of the superposed Thomas model, which are listed in Table 4.

To evaluate the reproducibility of the estimated model, we apply the same analysis procedures to a simulated point pattern of the superposed Thomas process using the parameters given in Table 4. Figure 7 shows the point pattern (*left panel*) and the re-estimated log-NND-likelihood values as described in Fig. 4. The simulated point pattern looks similar to the LLP data points given in Fig. 6.

We show the density of the NND function to confirm how the NND-statistics of the estimated model are consistent with that of the LLP data points. Figure 8 shows the estimated densities of nearest neighbor distances $\{\hat{g}_a\}$ from the simulated data using the superposed Thomas models with the MPLE estimates given in Table 3 for respective a-values, as indicated in the figure caption. Comparison with the empirical NND estimate (circles) derived from the LLP data points reveals that the maximum NND-likelihood estimate (solid white line) is fairly unbiased. From these results, the model with $\hat{a} = 0.4$ seems better fitted to the data than the other a-values.

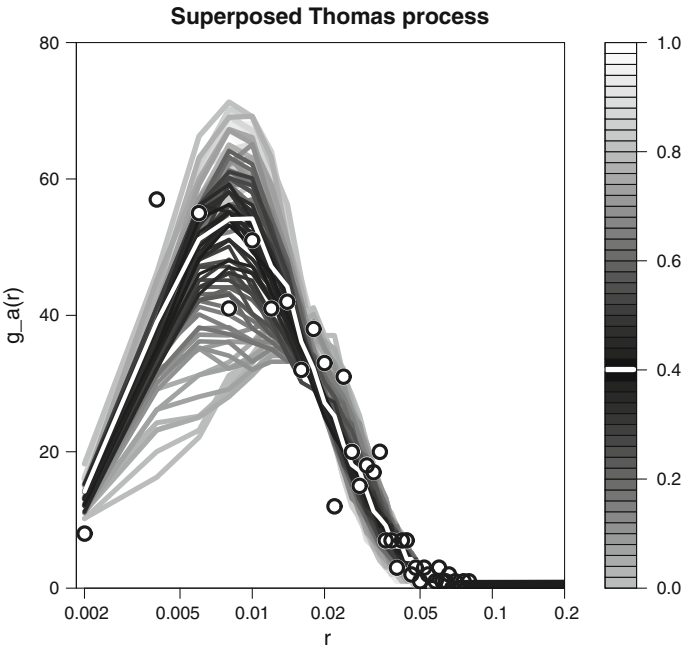

Fig. 8 The estimated density function of nearest neighbor distances of the LLP data points. The lines are same as that in Fig. 5 except that the *solid white line* represents the NND density for $\hat{a} = 0.4$

5 Concluding remarks

This paper summarizes estimation difficulties associated with superposed Neyman–Scott cluster models with different sizes and scales of clusters. Because not all parameters can be uniquely determined by the Palm intensity function method or second-order characteristics (see Tanaka et al. (2008b)), additional criteria for model fitting must be employed. Here, the likelihood function based on the NND distribution was considered and implemented.

The Palm-likelihood is not the only alternative to maximum likelihood estimation. For example, the composite likelihood method is a pseudo-likelihood approach for estimating the spatial point patterns; see Baddeley and Turner (2000), Mrkvička and Molchanov (2005), Guan (2006), Møller and Waagepetersen (2007), and Guan et al. (2011). The Palm-likelihood estimation method is closely related to the composite likelihood method suggested in Waagepetersen (2007), but the Palm-likelihood method is numerically simpler than the composite likelihood method in Guan (2006) because of a simpler form of the normalization term. Furthermore, the maximum likelihood estimation procedure was attempted through a simulation-based Bayesian method in, for example, McKeague and Loizeaux (2002), and Van Lieshout and Baddeley (2002). The Bayesian method is not feasible at present owing to its highly intensive computational burden.

Prokesova and Vedel Jensen (2013) proved the consistency and asymptotic normality of the Palm-likelihood method. Hence, the error estimates of the parameters

are obtained by the inverse Hessian matrix of the log-Palm-likelihood function. However, it is not easy for us to evaluate the efficiency relative to the maximum likelihood estimate at present. It might be possible to evaluate the efficiency numerically using a Markov chain Monte Carlo simulation in the future. The error of the maximum log-NND-likelihood estimate can be observed in Figs. 3, 4, 6, and 7, which does not show very high sensitivity owing to the low number of simulated points. By the same reason as given above, it is also difficult to obtain the efficiency of the parameter.

Once we can fix the estimated parameters by the methods in Tanaka et al. (2008b) and the proposed method, we can proceed to estimate the posterior distribution of locations of cluster centers of the different types (Ogata and Tanaka (2013)). Then, this Bayesian inference can be directly applied to the problem of probabilistic discrimination of spatial clusters.

The dimension of the space of the sample points and boundary conditions can be generalized, as discussed in Tanaka et al. (2008b).

6 Appendix: Strong consistency of the maximum NND-likelihood estimator

In this appendix, we formulate the strong consistency of the maximum NND-likelihood estimator of a given in (7). Its detailed proof is omitted here because the argument is lengthy and it is almost a literal translations of that provided by Prokesova and Vedel Jensen (2013), who showed that the maximum Palm-likelihood estimator satisfied the strong consistency under certain conditions. We recall the notion needed in the statement of the theorem on the strong consistency of the maximum NND-likelihood estimator, in which a convex averaging sequence of windows $\{W_n\}_{n\in\mathbb{N}}$ is founded and stated as follows: all the windows W_n are bounded and convex Borel sets, $W_n \subset W_{n+1}$ for all n and the inradii

$$\rho(W_n) := \sup\{\, \rho;\, B_\rho(x) \subset W_n \; \exists x \,\},$$

where $B_\rho(x)$ is a ball with its center at x and of radius $\rho > 0$, converges to ∞ as $n \to \infty$ (refer Daley and Vere-Jones (1988), Daley and Vere-Jones (2003), and Daley and Vere-Jones (2008)).

We denote by $a_0 \in (0, 1)$ a true parameter. Under the assumption that the log-NND-likelihood function of (7) is differentiable, we indicate by $U_n(a)$ the score function computed from observations in the window W_n for a given homogeneous point process X, namely

$$U_n(a) = \frac{1}{N(W_n \ominus B_R(\mathbf{o}))} \sum_{x,y\in X\cap(W_n\ominus B_R(\mathbf{o}))} \frac{d}{da}\log g_a(x-y) - \frac{N(X\cap(W_n\ominus B_R(\mathbf{o})))}{N(W_n\ominus B_R(\mathbf{o}))}\frac{d}{da}\int_0^R g_a(r)\,dr,$$

where $\ominus$ signifies Minkowski subtraction, $R > 0$ is some constant, and x and y run such that $\mathbf{dist}(x, y) \le r$ for each $r \in [0, R]$, in which **dist** is the nearest neighbor

contact distance. We denote by $\hat{a}_n$ the NND-likelihood estimator obtained from $U_n(a)$, i.e., $U_n(\hat{a}_n) = 0$. Then, we can state the desired result of the strong consistency as follows:

Theorem 1 *Let X be a stationary ergodic point process observed in a convex averaging sequence of windows $\{W_n\}_{n\in\mathbb{N}}$ of $\mathbb{R}^d$. Assume that $|\mathbb{E}_{a_0} U_n(a)| = 0$ only when $a = a_0$ and that $g_a(r)$ is of class C^1 with respect to $a \in [0, 1]$ and $r \in B_R(\mathbf{o})$, where $\mathbb{E}_{a_0}$ denotes the mean value with respect to the distribution of the point process with $a = a_0$. Then $\hat{a}_n$ is a strongly consistent estimator of a_0, i.e., $\hat{a}_n \to a_0$ Pr-a.s.*

Because we consider the case where the intensity function given in (7) to be approximated by that of an inhomogeneous Poisson process is g_a, the proof of Theorem 1 shall be parallel to that of Prokesova and Vedel Jensen (2013), who considered the case where the Palm intensity was approximated by the intensity function of an inhomogeneous Poisson process.

Acknowledgments The authors are grateful to Koichi Katsura for his generous help in programming matters and to Jancang Zhuang for his helpful advice related to the graphical software R. This study is partially supported by Grant-in-Aid 23240039 for Scientific Research (A), Ministry of Education, Science, Sports and Culture. Y.O. was supported by the Aihara Innovative Mathematical Modelling Project and the Japan Society for the Promotion of Science (JSPS) through the 'Funding Program for World-Leading Innovative R&D on Science and Technology (FIRST Program)', initiated by the Council for Science and Technology Policy (CSTP).

References

Akaike, H. (1974). A new look at the statistical model identification. IEEE Transactions on Automatic Control, AC-19, 716–723.

Baddeley, A., Turner, R. (2000). Practical maximum pseudolikelihood for spatial point patterns. *Australian & New Zealand Journal of Statistics*, *42*, 283–322.

Baddeley, A., Barany, I., Schneider, R., Weil, W. (2007). *Stochastic Geometry Lecture Notes in Mathematics 1892*. Berlin: Springer.

Baudin, M. (1981). Likelihood and nearest-neighbor distance properties of multidimensional Poisson cluster processes. *Journal of Applied Probability*, *18*, 879–888.

Cressie, N. (1993). *Statistics for Spatial Data*. New York, Chichester: John Wiley & Sons.

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

Daley, D. J., Vere-Jones, D. (2003). *An Introduction to the Theory of Point Processes, Volume I: Elementary Theory and Methods* (2nd ed.). New York: Springer.

Daley, D. J., Vere-Jones, D. (2008). *An Introduction to the Theory of Point Processes, Volume II: General Theory and Structure* (2nd ed.). New York: Springer.

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

Guan, Y. (2006). A Composite Likelihood Approach in Fitting Spatial Point Process Models. *Journal of the American Statistical Association*, *101*, 1502–15012.

Guan, Y., Jalilian, A., Waagepetersen, R. P. (2011). Optimal Estimation of the Intensity Function of a Spatial Point Process, Research Report, Centre for Stochastic Geometry and Advanced Bioimaging, No. 07, Department of Mathematical Sciences, Aarhus University, Denmark.

Hutchings, M. J. (1979). Standing crop and pattern in pure stands of Mercurialis perennis and Rubus fruticosus in mixed deciduous woodland. *Oikos*, *31*, 351–357.

McKeague, I.W., Loizeaux, M. (2002) *Perfect sampling for point process cluster modeling,* Spatial Cluster Modelling, eds. A. B. Lawson and G. T. Denison, 87–107, Florida: Chapman & Hall/CRC Press.

Møller, J., Waagepetersen, R. P. (2007). Modern statistics for spatial point processes. *Scandinavian Journal of Statistics*, *34*, 643–684.

Mrkvička, T., Molchanov, I. (2005). Optimisation of linear unbiased intensity estimators for point processes. *Annals of the Institute of Statistical Mathematics*, *57*, 71–81.

Neyman, J., Scott, E. (1958). Statistical approach to problems of cosmology. *Journal of the Royal Statistical Society*. *Series. B*, *20*, 1–43.
Ogata, Y., Katsura, K. (1991). Maximum likelihood estimates of the fractal dimension for random spatial patterns. *Biometrika*, *78*, 463–474.
Ogata Y., Tanaka U. (2013). Searching for locations of parents from superposed clusters, Research Memo., No.1175, The Institute of Statistical Mathematics, Tokyo.
Prokešová, M., Vedel Jensen, E. B. (2013). Asymptotic Palm likelihood theory for stationary point processes. *Annals of the Institute of Statistical Mathematics*, *65*, 387–412.
Rathbun, S. L., Cressie, N. (1994). A space-time survival point process for a longleaf pine forest in Southern Georgia. *Journal of the American Statistical Association*, *89*, 1164–1174.
Ripley, B. D. (1977). Modelling spatial patterns. *Journal of the Royal Statistical Society*. *Series B*, *39*, 172–212.
Stoyan, D., Stoyan, H. (1994). *Fractals, Random Shapes and Point Fields*. Chichester: John Wiley & Sons.
Stoyan, D., Stoyan, H. (1996). Estimating pair-correlation functions of planar cluster processes. *Biometrical Journal*, *38*, 259–271.
Tanaka, U., Ogata, Y., Katsura, K. (2008a). Simulation and Estimation of the Neyman–Scott Type Spatial Cluster Models. *Computer Science Monographs*, *34*, 1–44.
Tanaka, U., Ogata, Y., Stoyan, D. (2008b). Parameter estimation and model selection for Neyman–Scott point processes. *Biometrical Journal*, *50*, 43–57.
Thomas, M. (1949). A generalization of Poisson's binomial limit for use in ecology. *Biometrika*, *36*, 18–25.
Van Lieshout, M. N. M., Baddeley, A. J. (1996). A nonparametric measure of spatial interaction in point patterns. *Statistica Neerlandica*, *50*, 344–361.
Van Lieshout, M. N. M., Baddeley, A. J. (2002). *Extrapolating and interpolating spatial patterns,* Spatial Cluster Modelling, eds. A. B. Lawson and G. T. Denison, 61–86, Florida: Chapman & Hall/CRC Press.
Waagepetersen, R. (2007). An estimating function approach to inference for inhomogeneous Neyman–Scott processes. *Biometrics*, *63*, 252–258.