Bayesian analysis of spatial point patterns from noisy observations

Jens Lund* Antti Penttinen† Mats Rudemo‡

October 27, 1999

Abstract

A Bayesian analysis of point patterns degraded by thinning, random displacement and superposition of 'ghost' points is suggested. Pairwise interaction Gibbs point processes are used as prior models for the unobserved true point pattern, and a Metropolis-Hastings type algorithm is constructed for simulation from the posterior distribution of the unobserved point pattern and associated information. The simulations are used for estimation of statistical summaries such as the K-function, the nearest neighbour distribution function and the empty space statistic for the underlying point pattern. Analysis of degraded point patterns is relevant in a variety of situations with indirect observations or inverse problems in e.g. high level image analysis. We illustrate the method by a forestry example where tree maps are constructed from aerial photographs and the observed tree positions are disturbed by the image analysis process.

Keywords: point process, noise, disturbance, Gibbs process, Bayesian analysis.

Running headline: Analysis of noisy point patterns

1 Introduction

Statistical analysis of point patterns is typically based on observations of positions $\{Y_j, j = 1, \ldots, n\}$ of objects observed within a sampling window. The only commonly

*The Royal Veterinary and Agricultural University, Department of Mathematics and Physics, Thorvaldsensvej 40, 1871 Frederiksberg C, Denmark, e-mail: jlund@dina.kvl.dk.

†University of Jyväskylä, Department of Statistics, P.O. Box 35 (MaD), 40351 Jyväskylä, Finland, e-mail: penttine@stat.jyu.fi.

‡Chalmers University of Technology and Gothenburg University, Department of Mathematical Statistics, 41296 Gothenburg, Sweden, e-mail: rudemo@math.chalmers.se.

applied correction is for edge effects caused by the unobserved objects which are outside the sampling window but affect the data, see e.g. Ripley (1988) and Baddeley & Gill (1997). Less attention has been paid to the possibility that the positions of objects within the sampling window may be contaminated with points not belonging to the true point pattern, that points may be lost and that the positions may be subject to displacement. These errors imply that the 'true' point process $X = \{X_i, i = 1, \ldots, m\}$ is not observed. Instead, the inference is based on $Y = \{Y_j, j = 1, \ldots, n\}$, a random transformation of X. Evidently, the number of points in X and Y may differ. Diggle (1993) deals with independent random displacement of points. In the present paper the random transformation consists of random thinning, displacement of points, and superposition with additional points.

We are for instance interested in point patterns originating from tree positions. Extracting a tree position map of a forest area is severely restricted by the need of extensive field work. In case a large data set of this type is needed, recently developed cost efficient data collection methods such as GPS (Global Positioning Systems) can be applied. The choice of the observation method can be viewed as a compromise between cost and precision.

A further step in cost efficiency is the use of indirect observation methods such as airborne photography as well as spectrometer, laser, and radar imaging where the observed process can be highly degraded. An example can be found in Dralle & Rudemo (1997) where positions of trees in a forest area are estimated from a panchromatic aerial photograph. There the tree-top positions are estimated from a grey-level image. The obtained point pattern is subject to errors caused, for instance, by the irregular forms of the tree-tops and by motion due to wind. An idea towards rigorous modelling is further developed in Larsen & Rudemo (1998) based on template modelling and in Lund & Rudemo (1998) which gives a likelihood model for tree-top positions observed from a grey-level image through templates. Empirical results on estimation of stem numbers from digitized aerial photographs for stands with regular and clustered tree patterns are given in Uuttera *et al.* (1998).

We pose two statistical problems which appear in the analysis of point patterns observed with errors and suggest a Bayesian approach to treat them.

The first problem is to make statements about the unobserved point pattern $X = \{X_i, i = 1, \ldots, m\}$ given the noise-corrupted pattern $Y = \{Y_j, j = 1, \ldots, n\}$. This part can be interpreted as a further development of Lund & Rudemo (1998) which gives a statistical model and an approximation to the likelihood for efficient calculation useful

in the analysis of training data where both X and Y are observed. However, in forest inventories typically only the Y-process is observed and the investigator is faced with an inverse problem resembling image restoration, see e.g. Besag (1986). We advocate a Bayesian approach: construct a prior model $L(X)$ for the unobserved point pattern, an error model $L(Y|X)$ for the measurement noise, and give a Markov chain Monte Carlo (MCMC) estimate for the posterior $L(X|Y) \propto L(X)L(Y|X)$.

The second problem deals with point process statistics for the unobserved point pattern X. Commonly used data summaries are the K-function of Ripley (1976), the nearest neighbour distribution G and the empty space statistic F, see e.g. Diggle (1983). The uncertainty of these estimators for an observed mapped point pattern is e.g. derived from simulation of a hypothetical model as in Ripley (1977). The observed uncertainty may be expressed in terms of pointwise confidence limits. For indirectly observed point patterns degradation adds an extra source of uncertainty. Some analytical progress in calculation of the K statistic is possible. Diggle (1993) derived the connection between the K-function of the X- and Y-processes with independent random perturbations of the X-points. This connection is intractable for G and F functions. We bypass this problem by suggesting posterior estimates for K, G, and F derived through MCMC sampling.

Bayesian analysis of an unobserved point pattern from a related noisy observation is studied in Baddeley & van Lieshout (1993, Sec. 5) too, where cluster centres are estimated from a cluster process. Our problem may be formulated in a similar way with the number of offsprings being one or zero. In addition, motivated by our forestry application we superimpose an independent point process. We focus on the statistical properties (interpretation) of the unobserved point pattern rather than reconstruction (prediction). Cressie & Lawson (1998, 1999) use cluster processes with two types of points and MCMC in their application for detecting mine fields from noisy observations. Related methods for mine field detection with model-based clustering are studied by Dasgupta & Raftery (1998).

Our methods are applied to a forestry dataset where indirect observation of the tree positions are obtained by image analysis. For this particular dataset the true positions are known too, and we compare them with the results from our Bayesian method applied to the noisy observation. We study the performance of both a Poisson prior and Gibbs priors with Strauss and logistic pair interaction functions in the estimation of statistical summaries.

Section 2 describes the model for the disturbances of the original X point process,

and Section 3 describes prior models for the point pattern X. The analysis of the posterior model is done by MCMC sampling and statistical summaries as outlined in Section 4. Section 6 contains the application in forestry and image analysis, and some final remarks are made in Section 7.

2 A model for point pattern disturbances

Let $X = \{X_i, i \in M\}$, where $M = \{1, \ldots, m\}$ is the index set of X, be a finite point process on a bounded set $A \subset \mathbb{R}^d$. We consider X as the 'true' point process and we have observed a disturbed version $Y = \{Y_j, j \in N\}$, $N = \{1, \ldots, n\}$, of X.

The conditional distribution of the observed Y-process given the X-process is denoted by $L(Y|X)$. It defines the connection between the true process and the observations allowing both missing and extra points, and disturbances of positions. Parameters in this measurement model may be estimated from training data where both X and Y have been observed. In detail, the model consists of the following three independent components:

(a) *Thinning.* The points X_i, $i \in M$, are retained with probability $p(X_i)$ and thinned with probability $1 - p(X_i)$ independently of each other.

(b) *Displacement.* An unthinned point X_i is displaced according to a probability distribution with density $k(\cdot|X_i)$, $i \in M$. The displacements are independent, and points displaced outside A are censored.

(c) *Superposition of ghost points.* Points from an independent Poisson process with intensity $\lambda(\cdot|X)$ are superimposed on the set of displaced points inside A.

We will be more specific and make some simplifying assumptions. Assume that (*i*) the thinning probability is homogeneous, $1 - p(X_i) = 1 - p$, and (*ii*) the intensity of the superimposed Poisson process is homogeneous as well, $\lambda(\cdot|X) = \lambda$. Furthermore, we assume (*iii*) that $k(\cdot|X_i)$ corresponds to a normal d-dimensional distribution $N_d(X_i + \mu, \Sigma)$. Here μ is a systematic displacement common to the X-points. In the planar case ($d = 2$) this leads, with $\mu = (\mu_1, \mu_2)$ and Σ equal to the two-by-two matrix with diagonal elements σ_1^2 and σ_2^2 and off-diagonal elements $\rho\sigma_1\sigma_2$, to the parameter vector $\theta = (p, \lambda, \mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$.

Let $|B|$ denote the number of elements in a finite set B and let $|B|_d$ denote the d-dimensional volume of B when B is a measurable subset of $\mathbb{R}^d$. For two finite sets

M_1 and N_1 with $|M_1| = |N_1|$ we let $\mathcal{P}(M_1, N_1)$ denote the set of all bijections π from M_1 to N_1. Introduce $S_{m,n} = \{(M_1, N_1, \pi) : M_1 \subseteq M = \{1, \ldots, m\},\ N_1 \subseteq N = \{1, \ldots, n\},\ |M_1| = |N_1|,\ \pi \in \mathcal{P}(M_1, N_1)\}$ for $m, n \geq 0$. In case $m = 0$ ($n = 0$) we let $M = \emptyset$ ($N = \emptyset$) in the definition. Further, define $S = \cup_{m=0}^{\infty} \cup_{n=0}^{\infty} S_{m,n}$, and provide S with the σ-algebra consisting of all subsets. We let $s = (M_1, N_1, \pi) \in S$ specify the correspondence between X- and Y-points and we call s a 'matching' with the interpretation that $(X_i, i \in M_1)$ and $(Y_j, j \in N_1)$ consist of matched points. More precisely, $Y_{\pi(i)}$ is matched to X_i for $i \in M_1$, that is, $Y_{\pi(i)}$ is obtained from X_i by displacement.

Define

$$T(X, Y, \theta, s) = L_1\, L_2\, L_3\, L_4$$

for $s = (M_1, N_1, \pi)$ where

$$\begin{aligned}
L_1 &= p^{|M_1|} \prod_{i \in M_1} k(Y_{\pi(i)} | X_i, \mu, \Sigma), \\
L_2 &= \prod_{i \in M \setminus M_1} \left(p \int_{A^c} k(y | X_i, \mu, \Sigma)\, dy + 1 - p \right), \\
L_3 &= \lambda^{|N \setminus N_1|} \exp((1 - \lambda)|A|_d), \\
L_4 &= 1\big(s \in S_{|X|,|Y|}\big)\,.
\end{aligned}$$

Here L_1 corresponds to unthinned points displaced within A, L_2 to thinned points and unthinned points displaced out of A, and L_3 to superpositioned 'ghost' points. We have used the notation A^c for the complementary set $\mathbb{R}^d \setminus A$ and $1(\cdot)$ for the indicator function. As discussed in Lund & Rudemo (1998) the approximation $\int_{A^c} k(y | X_i, \mu, \Sigma)\, dy = 0$ is reasonable in our application as the size of the displacements is very small compared to the region A. With this approximation we have that $L_2 = (1 - p)^{|M \setminus M_1|}$.

Theorem 1 in Lund & Rudemo (1998) states that the likelihood for (X, θ) based on the observation Y is

$$L(Y | X, \theta) = \sum_{s \in S} T(X, Y, \theta, s), \tag{1}$$

with the reference measure taken as the Poisson process with intensity 1 on the set A. One should note that the sum typically contains a very large number of non-zero terms but the number of leading terms may be small and thus approximations are possible. In the sequel we will consider the joint likelihood of the point processes Y and the matching s given by

$$L(Y, s | X, \theta) = T(X, Y, \theta, s), \tag{2}$$

with the reference measure taken as the product of the measure of a Poisson process on A with intensity 1 and the counting measure on S. The connection between (1) and (2) is that (1) is the marginal distribution of the point process Y in (2). The matching s is unobserved and thus (2) could be considered as a missing data model (Smith & Roberts, 1993, Sec. 6).

3 The prior distribution for X

The pattern of tree positions in a forest typically shows some kind of structure. For example, if the forest is thinned we do not expect trees to be very close to each other. In addition, planted forests typically have a row structure visible during a considerable time after plantation provided the forest is not too heavily thinned. It seems useful to include such information in the form of a prior distribution $L(X)$ for the unobserved point pattern X. This leads us to concentrate on prior distributions giving low weight to point patterns with close neighbours, that is, clustering is avoided.

We will consider Gibbs point processes on A with a pairwise interaction function H as point process priors. The density is defined by

$$f(x) = c \prod_{x_i} \alpha(x_i) \prod_{x_i, x_j, i<j} H(||x_i - x_j||) \tag{3}$$

with respect to the unit rate Poisson measure on A, where c is the normalizing constant, $\alpha(x)$ describes the spatial inhomogeneity, H is the pair interaction function, and $||x - x'||$ is the d-dimensional distance between x and x', see e.g. Stoyan *et al.* (1995). We have no a priori indication of inhomogeneity, and we take $\alpha(x_i) = \alpha$, a constant. The interpretation of α is that $\alpha|A|_d$ equals the expected number of points in A — had it been a Poisson process. We assume that $H(r) \leq 1$ and that $H(\cdot)$ is nondecreasing which leads to a repulsive local behaviour. We consider three choices of the pair interaction function: (*i*) a Poisson process with $H(r) = 1$ for all distances r, (*ii*) a Strauss process prior where

$$H(r) = \gamma^{1(0 \leq r \leq R)}$$

for fixed R with $0 \leq \gamma < 1$, where $\gamma = 0$ corresponds to a hardcore process , and (*iii*) a logistic pair interaction function

$$H(r) = \frac{1}{1 + \exp(-\beta(r - R))}.$$

The interpretation of the parameters in the logistic function is as follows: $\beta > 0$ describes the steepness of the curve; R defines the distance where $H(R) = 0.5$. Instead

of (β, R) we use the parameters $(H(0), R)$. Note that $H(r) \to 1$ as $r \to \infty$, and that $\beta \to \infty$ (i.e. $H(0) \to 0$ when $R > 0$) leads to a hardcore Strauss process in the limit.

The main disadvantage of using a non-Poisson Gibbs distribution prior for X is that there is no simple expression for the expected number of points in the prior and the normalizing constant c of the distribution is unknown. Here we suggest to fix the parameters in such a way that the expected number of points in $L(X)$ approximately corresponds to the number of trees we expect to find and that there is a reasonable repulsion between points. Suitable values for the prior parameters are found by sampling from the prior distribution by the algorithm described in Geyer & Møller (1994). The number of trees may be inferred from exploratory analyses as in Dralle & Rudemo (1996). The discussion in Section 7 has comments on alternatives to this approach and problems associated with the unknown normalizing constant.

4 Sampling from $L(X|Y,\theta)$

Maximum likelihood estimation of θ in the model (1) when both X and Y are observed was treated in Lund & Rudemo (1998). In the present paper we will focus on the conditional distribution of X given Y when the parameter θ is assumed to be known. The knowledge of θ can, for instance, be obtained through training data sets observed under similar conditions as the current data.

We suggest a simulation based Bayesian analysis with the model (2) as the measurement error model. The matching s is introduced as missing data and the simulation method moves randomly among matchings. Thus calculation of the large sum over $s \in S$ in (1) is avoided and replaced by a large number of simulations.

4.1 Outline of the Bayesian analysis

The basic idea is to get information on the unobserved quantities by looking at the conditional distribution of the unobserved X-points given the observed Y-points with a fixed model parameter θ. This type of inference is typically based on the posterior $L(X|Y,\theta)$. However, it is more attractive to simulate the joint posterior $L(X,s|Y,\theta)$ of X and s, and we thus consider s as missing data and derive the X marginal from $L(X,s|Y,\theta)$. In addition, the posterior distribution of s may give further information on the data and additional model assessment possibilities such as diagnostics of the displacement distribution. We construct a sampler of Metropolis-Hastings type

along the lines described in Geyer & Møller (1994) and Green (1995). The details are presented in Section 4.2.

Given the prior distribution $L(X)$ for X we get $L(X, s|Y, \theta) = L(X, Y, s|\theta)/L(Y|\theta)$ with $L(X, Y, s|\theta) = L(Y, s|X, \theta)L(X)$. Notice that in the sampling procedure it is not necessary to know the normalizing constant, but it suffices to know that $L(X, s|Y, \theta) \propto T(X, Y, \theta, s)L(X)$. However, see the discussion in Section 7 for some potential problems associated with unknown normalizing constants in the prior distribution.

4.2 MCMC algorithm for posterior sampling

In this section we suggest an MCMC algorithm for sampling of the posterior $L(X, s|Y, \theta)$. It is based on the Metropolis-Hastings sampler for point processes suggested by Geyer & Møller (1994), which is a special case of the variable-dimension sampler of Green (1995). A detailed description of a related reversible jump algorithm can be found in Heikkinen (1998). Our algorithm defines a Markov chain with transitions $(X_t, s_t) \to (X_{t+1}, s_{t+1})$ having the posterior $L(X, s|Y, \theta)$ as the stationary distribution. In an update of the current state (X_t, s_t) at time t, first a suggestion using a 'proposal kernel' $Q(X_t, s_t; X'_{t+1}, s'_{t+1}|Y, \theta)$ is made. The obtained proposal (X'_{t+1}, s'_{t+1}) is then accepted with probability $\alpha = \min\{1, r\}$ where $r = r_{\text{posterior}} \times r_{\text{proposal}} \times |J|$ is the Hastings-Green ratio with

$$r_{\text{posterior}} = \frac{L(X'_{t+1}, s'_{t+1}|Y, \theta)}{L(X_t, s_t|Y, \theta)},$$
$$r_{\text{proposal}} = \frac{Q(X'_{t+1}, s'_{t+1}; X_t, s_t|Y, \theta)}{Q(X_t, s_t; X'_{t+1}, s'_{t+1}|Y, \theta)},$$

and J is the Jacobian due to the change of dimension of the state space. (In our case the Jacobian is identically one because we do not make any changes to points in the configuration but just add and delete points.) We put $(X_{t+1}, s_{t+1}) = (X'_{t+1}, s'_{t+1})$ in case of acceptance and $(X_{t+1}, s_{t+1}) = (X_t, s_t)$ in case of rejection.

From a current state (X_t, s_t), where X_t has index set M_t and $s_t = (M_{1,t}, N_{1,t}, \pi_t)$, we obtain a proposal (X'_{t+1}, s'_{t+1}), where X'_{t+1} has index set M'_{t+1} and $s'_{t+1} = (M'_{1,t+1}, N'_{1,t+1}, \pi'_{t+1})$, in a two-step procedure. First we choose to add or delete a point with probabilities $b(X_t)$ and $1 - b(X_t)$, respectively. Then, we select a matched or an unmatched point with probability $h(X_t, s_t)$ and $1 - h(X_t, s_t)$, respectively, in the add case, and $1 - h(X_t, s_t)$ and $h(X_t, s_t)$, respectively, in the delete case. More precisely, we get four possible proposal types with probabilities as follows:

(*i*) *Add a matched point with probability* $b(X_t)\ h(X_t, s_t)$.

Choose an unmatched free Y-point uniformly among the $N \backslash N_{1,t}$ available points, say Y_j. Generate $\xi \sim N(Y_j - \mu, \Sigma)$, and put $X'_{t+1} = X_t \cup \{\xi\}$. Form s'_{t+1} from s_t by adding ξ and Y_j with Y_j matched to ξ. We find that

$$r_{\text{posterior}} = \frac{L(X'_{t+1}, s'_{t+1}|Y, \theta)}{L(X_t, s_t|Y, \theta)} = \frac{L(X'_{t+1})}{L(X_t)} \frac{p}{\lambda} k(Y_j|\xi, \mu, \Sigma)$$

and

$$r_{\text{proposal}} = \frac{(1 - b(X_t \cup \{\xi\}))(1 - h(X_t \cup \{\xi\}, s'_{t+1}))/|M'_{1,t+1}|}{b(X_t)h(X_t, s_t)k(Y_j|\xi, \mu, \Sigma)/|N \setminus N_{1,t}|}$$

with $|M'_{1,t+1}| = |M_{1,t}| + 1$. Note that the factors from the normal density cancel each other in r because we generated $\xi \sim N(Y_j - \mu, \Sigma)$ in the proposal, which has density $k(Y_j|\xi, \mu, \Sigma)$ as a function of ξ.

(*ii*) *Add an unmatched point with probability* $b(X_t)\ (1 - h(X_t, s_t))$.

Put $(X'_{t+1}, s'_{t+1}) = (X_t \cup \{\xi\}, s_t)$ with ξ uniformly distributed on A, and make no new connections between X- and Y-points. We find that

$$r_{\text{posterior}} = \frac{L(X'_{t+1}, s'_{t+1}|Y, \theta)}{L(X_t, s_t|Y, \theta)} = \frac{L(X'_{t+1})}{L(X_t)}(1 - p)$$

and

$$r_{\text{proposal}} = \frac{(1 - b(X_t \cup \{\xi\}))h(X_t \cup \{\xi\}, s'_{t+1})/|M'_{t+1} \setminus M'_{1,t+1}|}{b(X_t)(1 - h(X_t, s_t))/|A|_d}$$

with $|M'_{t+1} \setminus M'_{1,t+1}| = |M_t \setminus M_{1,t}| + 1$ and $s'_{t+1} = s_t$.

(*iii*) *Delete a matched point with probability* $(1 - b(X_t))\ (1 - h(X_t, s_t))$.

Choose, say X_i, uniformly among the $M_{1,t}$ paired X-points, and put $X'_{t+1} = X_t \setminus \{X_i\}$. Delete the connection between X_i and $Y_{\pi(i)}$ in s_t to obtain s'_{t+1}. Delete-moves have r-values as reciprocals of the r-values found above in add-moves. We see that

$$r_{\text{posterior}} = \frac{L(X'_{t+1}, s'_{t+1}|Y, \theta)}{L(X_t, s_t|Y, \theta)} = \frac{L(X'_{t+1})}{L(X_t)} \frac{\lambda}{p} \frac{1}{k(Y_{\pi(i)}|X_i, \mu, \Sigma)},$$

$$r_{\text{proposal}} = \frac{b(X_t \setminus \{X_i\})h(X_t \setminus \{X_i\}, s'_{t+1})k(Y_{\pi(i)}|X_i, \mu, \Sigma)/(|N \setminus N_{1,t}| + 1)}{(1 - b(X_t))(1 - h(X_t, s_t))/|M_{1,t}|}.$$

(*iv*) *Delete an unmatched point with probability* $(1 - b(X_t))\ h(X_t, s_t)$.

Choose, say X_i, uniformly among the $|M_t \setminus M_{1,t}|$ unpaired X-points, and put $(X'_{t+1}, s'_{t+1}) = (X_t \setminus \{X_i\}, s_t)$. Then

$$r_{\text{posterior}} = \frac{L(X'_{t+1}, s'_{t+1}|Y, \theta)}{L(X_t, s_t|Y, \theta)} = \frac{L(X'_{t+1})}{L(X_t)} \frac{1}{1 - p},$$

$$r_{\text{proposal}} = \frac{b(X_t \setminus \{X_i\})(1 - h(X_t \setminus \{X_i\}, s_t))/|A|_d}{(1 - b(X_t))h(X_t, s_t)/|M_t \setminus M_{1,t}|}.$$

A number of special cases are not covered by the above description. For example, if we propose to add a matched point and there are no free Y-points, or if we propose to delete a point and the current X configuration is empty, then we do nothing and put $(X_{t+1}, s_{t+1}) = (X_t, s_t)$. This is the same as modifying the proposal kernel slightly in those situations and propose a "do nothing" step instead of the impossible update. Such "do nothing" proposals are always accepted. In a similar fashion, we put $(X_{t+1}, s_{t+1}) = (X_t, s_t)$ when the proposed point ξ in (i) is outside A. This can also be thought of as the introduction of a "do nothing" step in the proposal kernel and a slight modification when the point is matched. For instance, we could say that we propose to match the point Y_j with probability $b(X_t)h(X_t, s_t)/|N \setminus N_{1,t}| \cdot \int_A k(Y_j|\xi, \mu, \Sigma)\, d\xi$ and then generate ξ as $N(Y_j - \mu, \Sigma)$ conditioned to be inside A. In any case, the above acceptance probabilities ensure that our Markov chain has the correct stationary distribution.

In practice we cannot suggest forms of $b(X_t)$ and $h(X_t, s_t)$ that generally seem to work better than $b(X_t) = h(X_t, s_t) = 1/2$, and we use this simple choice. In this case the factors corresponding to $b(X_t)$ and $h(X_t, s_t)$ in the r ratios cancel out.

We just propose the basic moves of adding or deleting points. Geyer & Møller (1994) find in the simpler situation of simulation of a Strauss process, that their sampler has the best mixing properties when they only consider additions and deletions of points. In the example in Section 6 this turns out to be sufficient in our case too. Additional move types like "displacing a point" or switches in the matching s could have been considered if mixing of the chain had turned out to be too slow. If the chain tends to get stuck in local maxima of the density a simulated tempering scheme could also be useful (Geyer & Thompson, 1995).

Convergence of the Markov chain $\{(X_t, s_t), t = 1, 2, \dots\}$ to the stationary distribution $L(X, s|Y, \theta)$ for any initial distribution can be established in a way similar to Geyer & Møller (1994). Our Markov chain is irreducible as the displacement distribution $k(\cdot|\mu, \Sigma)$ is Gaussian and we assume that both $b(X)$ and $h(X, s)$ have values in $(0, 1)$ for all (X, s). Then the empty configuration $(\emptyset, \emptyset)$ can be reached with positive probability from any configuration (X, s) in a finite number of steps. Our Markov chain has invariant distribution $L(X, s|Y, \theta)$ so Theorem 1 in Tierney (1994) says that the Markov chain is positive recurrent, and it is further possible to show that the chain is Harris recurrent. Theorem 3 in Tierney (1994) gives the convergence of sample averages of measurable integrable real-valued functions g on the state space of X,

$$\frac{1}{T}\sum_{t=1}^{T} g(X_t) \to \int g(X)L(X, s|Y, \theta)d(X, s) = \int g(X)L(X|Y, \theta)dX$$

with $L(X, s|Y, \theta)$-probability one. Furthermore the transition $(\emptyset, \emptyset) \to (\emptyset, \emptyset)$ takes place with positive probability for any number of steps, and thus our chain is aperiodic. This means that the n-step transition matrix P^n converges in total variation norm to the stationary distribution $L(X, s|Y, \theta)$. Central limit theorems for the sample averages may be obtained from geometric ergodicity of the Markov chain, cf. Geyer & Møller (1994).

5 Statistics derived from the MCMC samples

The algorithm of Section 4.2 results in a sequence $X^1, X^2, \ldots, X^T$ of point patterns with a variable number of points and where T is chosen to be large (in our application $T = 100000$). These patterns can after a burn in period be considered as a (dependent) sample of the X-marginal from the posterior $L(X, s|Y, \theta)$.

A useful summary plot is a kernel-smoothed version of the union $\cup_{j=1}^{T} X^j$. This gives information on the conditional intensity of X-points given Y and θ. A feature, to be discussed below, is that the set of matched points and the set of unmatched points call for quite different smoothing bandwidths.

We also suggest explanatory statistics and figures which give information on the posterior distribution beyond the (marginal) posterior intensity. One such figure shows the probability that a Y-point gives rise to a matched X-point as a function of the distance to the nearest neighbour Y-point and whether the Y-point is matched or not.

For our purposes it is informative to calculate point process summaries which supply information on the neighbour structure of the point positions of X. Conventional point process statistics are the K, G and F-functions defined as follows, see Ripley (1981), Diggle (1983):

(*i*) $K(r) = (1/\lambda_X)\,\mathrm{E}\,[|X \cap b(x, r) \setminus \{x\}| \;\; \| \; x]$. Here $b(x, r)$ is the ball with radius r and centered at x, λ_X is the intensity for the X-points, and the expectation is conditional given that there is an X-point at x (conditional in the Palm measure sense with the notation $\mathrm{E}[\cdot \;\|\; \cdot]$). The interpretation of K is that $\lambda_X K(r)$ is the expected number of additional points within distance r from a typical point.

(*ii*) $G(r) = 1 - \mathrm{P}(|X \cap b(x, r)| - 1 = 0 \;\; \| \; x)$ is the distribution function of the point to nearest point distance called the *nearest neighbour distribution function.*

(*iii*) $F(r) = 1 - \mathrm{P}(|X \cap b(x, r)| = 0)$ where x is an arbitrary location. This is the

distribution function of an arbitrary location to nearest point distance and is called the *empty space distribution function.*

Note that *(i)* and *(ii)* are conditional Palm type characteristics. Nonparametric edge corrected estimators exist for K, G and F, see e.g. Diggle (1983). The functions are estimated from $X^1, X^2, \ldots, X^T$ by pointwise medians of the estimator derived from each sample.

Let us from now on consider planar point processes. For a stationary Poisson process with intensity λ the F and G functions are $F(r) = G(r) = 1 - \exp(-\lambda\pi r^2)$ for $r \geq 0$, and $K(r) = \pi r^2$, $r > 0$. Hence the transformation $L(r) = \sqrt{K(r)/\pi}$ is commonly applied.

The effect of the three sources of error described in Section 2 on the K-function is known or can be derived when the point process is stationary and of second order:

(a) *Thinning.* If the transformation $X \to Y$ is independent thinning, then

$$K_Y(r) = K_X(r) \tag{4}$$

for all $r > 0$, see e.g. Theorem 2.2 in Diggle (1993).

(b) *Displacement.* Assume that Y is generated from X through independent displacements using a radially symmetric perturbation distribution $h(\cdot)$. In the computation of K_Y independent perturbations of a pair of points in X may be treated as a one-point shift with the perturbation distribution $f(x) = \int_{R^2} h(x-y)h(y)\, dy$, the convolution of two h-functions. For $||x|| = u$ we put $R(r,u) = \int_{||z||\leq r} f(z - x)\, dz$ observing that the integral depends on x only through $||x||$. Note that $R(r, ||x||)$ is the probability that the distribution $f(\cdot)$ will move a point at x to any location z with $||z|| \leq r$. Then K_Y can be obtained from K_X as

$$K_Y(r) = \pi r^2 + \frac{2\pi}{\lambda_X^2} \int_0^\infty R(r,u)\gamma_X(u)u\, du \tag{5}$$

where $\gamma_X(r)$ is the (isotropic) covariance density of X given by

$$\gamma_X(r) = \lambda_X^2 \left[\frac{1}{2\pi r} \frac{d}{dr} K_X(r) - 1 \right].$$

This connection is due to Diggle (1993).

(c) *Superposition of ghost points.* If the transformation $X \to Y$ is the result of a superposition, $Y = X \cup Z$, of X and an independent homogeneous Poisson process Z with intensity $\lambda_Z = \alpha\lambda_X$, then

$$K_Y(r) = \frac{1}{(1+\alpha)^2} \left[K_X(r) + \alpha(2+\alpha)\pi r^2 \right]. \tag{6}$$

This follows from the fact that if x is a randomly chosen point of $X \cup Z$ then $x \in X$ with probability $\lambda_X/(\lambda_X + \lambda_Y)$, and the formula

$$K_Y(r) = \frac{1}{\lambda_X + \lambda_Z}\left[\frac{\lambda_X}{\lambda_X + \lambda_Z}\lambda_X K_X(r) + \frac{\lambda_Z}{\lambda_X + \lambda_Z}\lambda_X \pi r^2 + \lambda_Z \pi r^2\right].$$

Summarizing, if K_Y is estimated from the degraded data Y, then the corresponding estimate of the K_X function for the unobserved X process can be found under any of the three transformations when the transformation parameters are known.

The equations (4), (5), and (6) can be solved recursively in the case where the three sources of errors appear simultaneously: Applying (6) we obtain the K function of the displaced points, wherefrom (5) gives the K function of X because the K function is invariant under independent thinning due to (4).

Despite the possibility of solving equations for K, our Bayesian approach seems preferable. First, using the connections above we are not able to express the sampling variation of the estimator of the K function. Second, matching between the X- and Y-points gives further information for interpretation. Finally, similar connections for the G and F functions are not known; more generally, our Bayesian method can be applied for any point process summary.

6 Example: tree-top positions from aerial photography

6.1 The observed point process obtained from template matching

The data analysed originate from a series of aerial photographs of a thinning experiment in Norway spruce (*Picea abies* (L.) Karst.), see Dralle & Rudemo (1996, 1997) for a detailed description of the experiment and image aquisition. For one of the photographs and one of the subplots in the experiment an observed point process $Y = (Y_1, \ldots, Y_n)$ with $n = 206$ points was obtained by template matching (Larsen & Rudemo, 1998). This data set together with the 'true' tree-top positions $X = (X_1, \ldots, X_m)$ with $m = 171$ was analysed in Lund & Rudemo (1998) with focus on parameter estimation in the conditional likelihood $L(Y|X, \theta)$ when both X and Y are observed. The X- and Y-points are shown in Figure 1. In the present paper we study the posterior distribution of X from an observation of Y, given the parameters θ and a prior distribution for X. We use the θ-parameters found in Lund & Rudemo (1998) for the reconstruction of X in this paper. The parameters are given in Table 1.

6.2 Prior distributions

We have used a Poisson process, a Strauss process, and a further Gibbs point process with logistic pair interaction function as prior distributions for X. The parameters, which are given in Table 2, are chosen such that the expected number of points in the prior distribution is approximately 171 — the true number of trees. We have also tried other parameter combinations with similar results, but they are not shown here.

The Poisson process can be regarded as a neutral prior with respect to the interaction between points as there is neither repulsion nor attraction between points.

6.3 Comments on simulations

The X- and Y-points in the region A are shown in Figure 1 together with the border of A. The region A is almost convex and, for computational convenience, the points from the posterior distribution are generated in the convex hull of A and subsequently points outside A are discarded in the figures.

Figure 2 shows the number of simulated X-points and the number of points that are matched for a run of length 102000. The prior used in this particular run is the logistic prior of Table 2, but the graphs are similar for all the priors. The simulations are started from the empty state (i.e. the lower left corner) and as indicated by the figure the number of simulated points stabilizes quickly. From time series graphs of the two marginal distributions and figures like Figure 3 with an individual sample it is judged to be sufficient with a burn in of 2000 steps. The two basic add and delete moves used here seem to be sufficient for a satisfactory mixing of the chain. The following 100000 simulations are used in the estimation of mean values etc. However, only every 200th simulation is used in the estimation of the L, G, and F, functions in Section 6.5.

Table 3 shows some summary statistics for simulations of the priors and posteriors. In general the acceptance probabilities of the chains seem very reasonable. As the moves just go up or down one point at a time there is a fairly strong autocorrelation in the chain. There are two approaches to remedy this. The first one is to let the proposals add, delete, and move several points in each step, which, however, might lead to lower acceptance rates. Instead we prefer to subsample the original Markov chain at evenly spaced points. From Table 3 we see that the number of X-points sampled at every 200th simulation are fairly weakly correlated. It is worth noticing that it was difficult to fine-tune the parameters of the prior to obtain a mean number of points of 171 as

we aimed at. In the Poisson prior it is possible to adjust the parameter such that the theoretical mean number of points is exactly 171 points, but still the sample mean is a bit higher. A study of plots of the number of X-points versus simulation number indicates that the reason for the fine-tuning difficulties are some weak but long term correlation components.

6.4 Realizations from the posterior distribution, intensity surfaces

Figure 3 shows one realization from the posterior X distribution. The simulated points are marked with crosses, and the matched ones have a line from the center of the cross to the Y-point it is matched to. In most cases the corresponding Y-point is so close to the simulated point, that it is impossible to see this line. In general most of the simulated points are matched.

Estimated intensity surfaces are shown in Figures 4 and 5. Figure 4 shows the estimated intensity surface for all simulated points, while Figure 5 shows the estimated intensity surface for the free points alone. Both figures are based on 100000 simulations and are smoothed by isotropic Gaussian kernels. One problem is that the intensity surface for the matched points which are heavily clustered around the Y-points call for a small smoothing bandwidth, while the free points which are distributed over larger areas (and which are fewer in number) call for a large bandwidth. In Figure 4 which is based on both matched and free points we have thus smoothed the two types of points separately with variances 6 and 100 (unit pixel squares) in the isotropic kernels. The intensity surfaces are then added after the smoothing. Thus we use here the simulated matchings $s^1, \ldots, s^{100000}$ as well as the simulated point patterns. Note that the maximum of the intensity estimate in Figure 4 is more than 200 times larger than the maximum of the intensity estimate in Figure 5.

By comparison with Figure 1 we note the main feature of the posterior intensity in Figure 4: it is concentrated around the observed Y-points with pronounced local maxima. A closer look shows that these maxima attain lower levels at Y-points with close neighbours; look in particular at the area in the lower part of Figure 4 around $(450, 150)$, where there is a cluster of Y-points. In contrast, Figure 5 shows that the posterior intensity for the free X-points has elevated levels in areas away from the Y-points.

Figure 6 shows the probability that a Y-point is matched given the distance to the nearest Y-point. The probability of being matched increases with the distance. Furthermore, if the nearest neighbour is matched (not matched), then the probability of

being matched is lower (higher) than the overall probability of being matched. The difference is greater for small distances than for large distances. For large distances it seems that the "advantage" of having a free nearest neighbour is greater than the "disadvantage" of having an occupied neighbour.

6.5 L, G, and F, functions

Figures 7, 8, and 9 show the L, G, and F functions estimated on the basis of the true X-points, the Y-points, and posterior estimates. The posterior estimates are pointwise medians of the corresponding functions from the simulated point patterns. On these graphs that contain several simulations, a particular simulation is identified by the name of the prior distribution with an indication of whether it is a posterior or prior distribution and the parameters in the prior distribution.

The L functions in Figure 7 show that the estimate based on Y is clearly biased towards a Poisson process compared to the curve based on the true X-points. The estimate based on the Poisson process prior is (as expected) even more biased towards the Poisson process. The estimate from the Strauss prior is in reasonably good agreement with the truth, but the cut point in the Strauss process is unfortunately very evident in the posterior distribution. The estimate based on the prior distribution with a logistic pair interaction function is smooth. It is also clearly better than the raw estimate based on Y.

The G function in Figure 8 based on Y is again clearly biased compared to the function estimated on the basis of X. In X there is a hard core distance between points at approximately 10 pixels. The estimate based on a Poisson prior is similar to the raw estimate based on the Y-points, although more smooth. Again, the Strauss based estimate has the cut point, and the logistic based posterior is smooth, and both are definitely better than the raw estimate.

For the F functions in Figure 9 the curves are more similar than for the L and G functions. The bias in the Y estimate is not large, and in fact the Poisson prior gives the largest departure from the true curve. Note that the cutpoint in the Strauss process is not evident in this graph.

Figure 10 shows the 95% credibility regions for the L-function from the MCMC runs. Note that these bands only express variability from the posterior distribution and not any uncertainty in the estimation of the 'true' curve due to finite window. The uncertainty of the median estimate is of course a lot smaller than the shown credibility

bands.

In Figure 11 we compare for a logistic prior the posterior L-curve with the prior L-curve and the L-curves for the Y- and X-points. We note that the posterior curve is closer to the true curve of the X-points compared both to the prior curve and to the observed curve.

7 Discussion

7.1 Comments on the suggested approach and possible modifications

The Strauss process is useful as a prior distribution due to its simplicity. Even though the cut point chosen in the prior distribution also turns up in the posterior distribution, the estimates from posterior distributions are clearly better than the raw estimates. However, in general we recommend the logistic prior because of the smoothness of the posterior estimates. It would be useful to have a larger class of prior distribution than those considered here, as one might then tailor the prior distribution more accurately to prior knowledge. The drawback of using a larger class of prior distributions is that is becomes even more difficult to decide on the parameters.

As mentioned in Section 3 the main disadvantage of using a non-Poisson Gibbs distribution prior for X is that there is no simple expression for the expected number of points in the prior. We fixed the parameters in the prior by a simulation study of the prior. Another suggestion would be to factorize the prior as $L(X) = L(X| \, |X| = n)L(n)$. However, this leads to difficulties in the Metropolis-Hastings updates described in Section 4.2. Due to the intractable normalizing constant Z_n of $L(X| \, |X| = n)$ we cannot calculate the fraction Z_n/Z_{n+1} needed in the acceptance probability for each proposal of an addition or deletion of a point. Baddeley & van Lieshout (1993) did as we do here — fix the parameters in the prior distribution.

If a fully Bayesian approach should be used then a hyperprior for the parameters θ_{prior} in the prior distribution should be introduced. There are two problems with this approach. First, it is difficult to specify meaningful hyperpriors as there is no simple connection between the parameters in the prior and basic summaries as the mean number of points in the prior. Second, the fraction between two unknown normalizing constants c of the density (3) should be computable in each update of the θ_{prior} parameter. Cressie & Lawson (1999) used a Poisson approximation (Ripley, 1988) for the normalizing constant in a Strauss process prior to circumvent the second problem.

The assumption of a constant θ_{prior} in the prior distribution $L(X \mid \theta_{\text{prior}})$ could be partially relaxed using the following procedure: First, for a fixed value $\theta^{\star}_{\text{prior}}$, a realization X_0 from $L(X|Y, \theta^{\star}_{\text{prior}})$ is simulated. Then θ_{prior} is estimated from X_0. This new θ_{prior}, or a θ_{prior} obtained after a few iterations of the procedure, is used in the final estimation of X. This is now an empirical Bayes type procedure used in image analysis in connection with Markov field priors, see e.g. Besag (1986). The advantage of this extension is to adapt the choice of θ_{prior} to the data when prior information on X is lacking.

We have assumed that the model parameter vector $\theta = (p, \lambda, \mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$ is known, but in principle it is reasonably straightforward to include uncertainty of θ in the calculations. A natural way to do it would be to put a prior distribution on θ as well as on X. Then we would sample from the posterior distribution of (X, s, θ) given Y. This sampler could contain a Gibbs like step alternating between sampling from (X, s) given (Y, θ) and sampling from θ given (X, Y, s). The first sampler needed is the same as described in Section 4.2, and the second one is reasonable standard.

In general the posterior distribution $L(X|Y, \theta)$ is not stationary, which somewhat complicates the interpretation of our estimates of the L, G, and F functions which are defined for stationary point patterns. However, point patterns generated from $L(X|Y, \theta)$ resemble the realization from the underlying true stationary distribution of X that generated our observation Y. Thus the functions estimated from these generated point patterns are useful as estimates of the functions for X, and this is what we are aiming at.

For forestry applications a particularly attractive characterization of a tree position point pattern is the corresponding Voronoi tesselation, which for each tree would indicate the "space" available. Compare Figures 5 and 6 in Berndt & Stoyan (1997) which give tesselations for manually and automatically determined point fields in an application from metallurgy.

7.2 Conclusions

We have described a method to analyse degraded point patterns. The method is a sample based Bayesian approach, and we are able to compute posterior versions of several important point process statistics. By this method we can make general posterior statements which are focused on the particular problems of interest whereas the existing analytical methods (Diggle, 1993) are limited to statements about the K function only. This work can also be considered as a theoretical and practical step in

the development of a high precision airborne forest inventory system, simultaneously reducing expensive field work.

Acknowledgements

Part of this work was done while Jens Lund visited Antti Penttinen at University of Jyväskylä. This visit was supported by a grant from NorFA. The Academy of Finland has financed a part of this study (ASED project 37206 and Senior Researchers's grant 385). The research was also supported by Dina, Danish Informatics Network in the Agricultural Sciences and the Gothenburg Stochastic Centre. We are grateful to Helle Sørensen for her comments on the manuscript.

References

Baddeley, A. & Gill, R. D. (1997). Kaplan-Meier estimators of distance distributions for spatial point processes. *Ann. Statist.* **25**, 263–292.

Baddeley, A. J. & van Lieshout, M. N. M. (1993). Stochastic geometry models in high-level vision. *Advances in Applied Statistics: Statistics and Images* **1**, 231–255. Supplement to Journal of Applied Statistics vol. 20 Nos. 5/6 1993, *Statistics and images*, eds. K. V. Mardia and G. K. Kanji.

Berndt, S. & Stoyan, D. (1997). Automatic determination of dendritic arm spacing in directionally solidified matters. *Z. Metallkd.* **88**, 758–763.

Besag, J. (1986). On the analysis of dirty pictures. *J. Roy. Statist. Soc. Ser. B* **48**, 259–302. With discussion.

Cressie, N. & Lawson, A. B. (1998). Bayesian hierarchical analysis of minefield data. In *Detection and Remediation Technologies for Mines and Minelike Targets III* (eds A.C. Dubey, J.F. Harvey & J.T. Broach) Proc. SPIE Vol. 3392, 930–940 .

Cressie, N. & Lawson, A. B. (1999). Hierarchical probability models and Bayesian analysis of mine locations. Preprint. Department of Statistics. Ohio State University and University of Aberdeen.

Dasgupta, A. & Raftery, A. E. (1998). Detecting features in spatial point processes with clutter via model-based clustering. *J. Amer. Statist. Assoc.* **93**, 294–302.

Diggle, P. J. (1983). *Statistical analysis of spatial point patterns.* Academic Press, London.

Diggle, P. J. (1993). Point process modelling in environmental epidemiology. In *Statistics for the Environment* (eds V. Barnett & F. Turkman) 89–110. Wiley, Chichester .

Dralle, K. & Rudemo, M. (1996). Stem number estimation by kernel smoothing of aerial photos. *Canad. J. Forest Res.* **26**, 1228–1236.

Dralle, K. & Rudemo, M. (1997). Automatic estimation of individual tree positions from aerial photos. *Canad. J. Forest Res.* **27**, 1728–1736.

Geyer, C. J. & Møller, J. (1994). Simulation procedures and likelihood inference for spatial point processes. *Scand. J. of Statist.* **21**, 359–373.

Geyer, C. J. & Thompson, E. A. (1995). Annealing Markov chain Monte Carlo with applications to ancestral inference. *J. Amer. Statist. Assoc.* **90**, 909–920.

Green, P. J. (1995). Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika* **82**, 711–732.

Heikkinen, J. (1998). Curve and surface estimation using dynamic step functions. In *Practical Nonparametric and Semiparametric Bayesian Statistics* (eds D. Dey, P. Müller & D. Sinha) 255–272. Springer-Verlag, New York .

Larsen, M. & Rudemo, M. (1998). Optimizing templates for finding trees in aerial photographs. *Pattern Recognition Letters* **19**, 1153–1162.

Lund, J. & Rudemo, M. (1998). Models for point processes observed with noise. Report. Department of Mathematics and Physics. The Royal Veterinary and Agricultural University. Submitted.

Ripley, B. D. (1976). The second order analysis of stationary point processes. *J. Appl. Probab.* **13**, 255–266.

Ripley, B. D. (1977). Modelling spatial patterns. *J. Roy. Statist. Soc. Ser. B* **39**, 172–212. With discussion.

Ripley, B. D. (1981). *Spatial statistics.* Wiley Series in Probability and Mathematical Statistics. Wiley, New York.

Ripley, B. D. (1988). *Statistical inference for spatial processes.* Cambridge Univ. Press, Cambridge.

Smith, A. F. M. & Roberts, G. O. (1993). Bayesian computation via the Gibbs sampler and related Markov chain Monte Carlo methods. *J. Roy. Statist. Soc. Ser. B* **55**, 3–23.

Stoyan, D., Kendall, W. S. & Mecke, J. (1995). *Stochastic geometry and its applications.* Second edn. Wiley, Chichester.

Tierney, L. (1994). Markov chains for exploring posterior distributions. *Ann. Statist.* **22**, 1701–1762. With discussion.

Uuttera, J., Haara, A., Tokola, T. & Maltamo, M. (1998). Determination of the spatial distribution of trees from digital aerial photographs. *Forest Ecology and Management* **110**, 275–282.

Tables and figures

θ	p	λ	μ_1	μ_2	σ_1^2	σ_2^2	$\sigma_1\sigma_2\rho$
Value	0.941	0.000275	-0.342	0.0815	1.047	2.028	-0.0489

Table 1: Parameter vector $\theta = (p, \lambda, \mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$ used in the analysis.

Distribution	Parameters
Poisson	$\alpha\|A\|_d = 171$
Strauss	$(\alpha\|A\|_d, \gamma, R) = (455, 0.14, 17)$
Logistic	$(\alpha\|A\|_d, H(0), R) = (530, 0.05, 15)$

Table 2: Parameters used in the prior distributions.

Run	Mean $\|X\|$	Mean $\|M_1\|$	Acceptance probability	Autocorrelation of $\|X\|$, 200 steps apart
Poisson, prior	174.1		0.97	0.58
Poisson, posterior	171.6	161.3	0.90	0.18
Strauss, prior	173.3		0.50	0.60
Strauss, posterior	176.2	167.0	0.51	0.26
Logistic, prior	176.7		0.67	0.47
Logistic, posterior	178.1	169.0	0.69	0.14

Table 3: Summaries of simulations from both the prior and posterior distributions. After a burn in of 2000 steps, the posterior chain was run for 100000 steps, and the prior chain for 50000 steps. "Mean $|M_1|$" is the mean number of matched X-points and does not make sense for prior simulations.

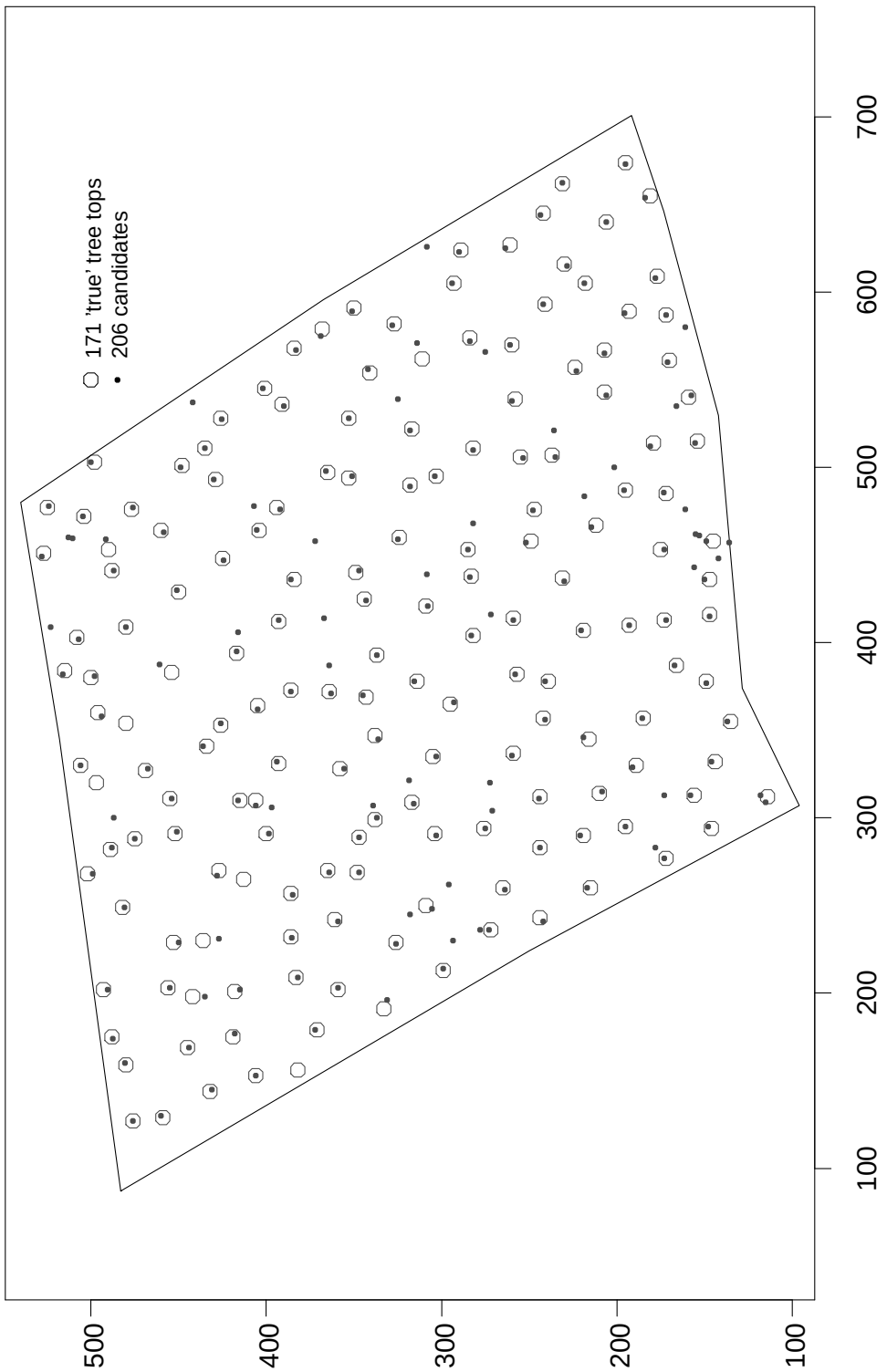

Figure 1: 171 X-points (centres of circles) corresponding to 'true' tree-tops and 206 Y-points (dots) corresponding to template matching. The area of the delineated subplot is 4454 m^2, and the unit of the axes is linear pixel size, 0.15 m.

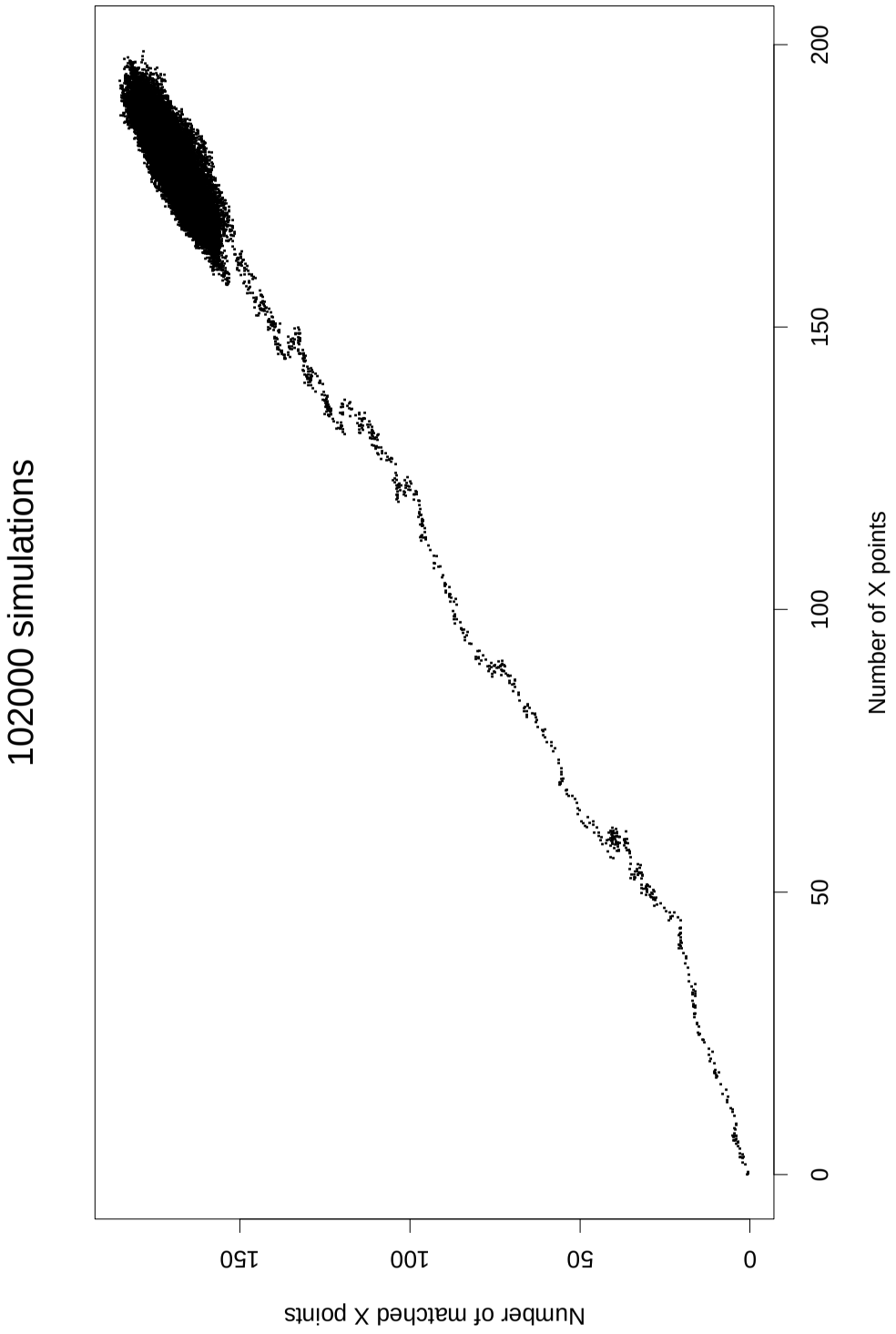

Figure 2: The number of matched X-points versus the number of simulated X-points for the posterior with the logistic prior. The simulations are started from the empty state, and converge reasonably fast towards a stationary distribution. The points in the figure are jittered by independent uniform $(0, 1)$ variables in the horisontal and vertical directions to facilitate interpretation of the figure.

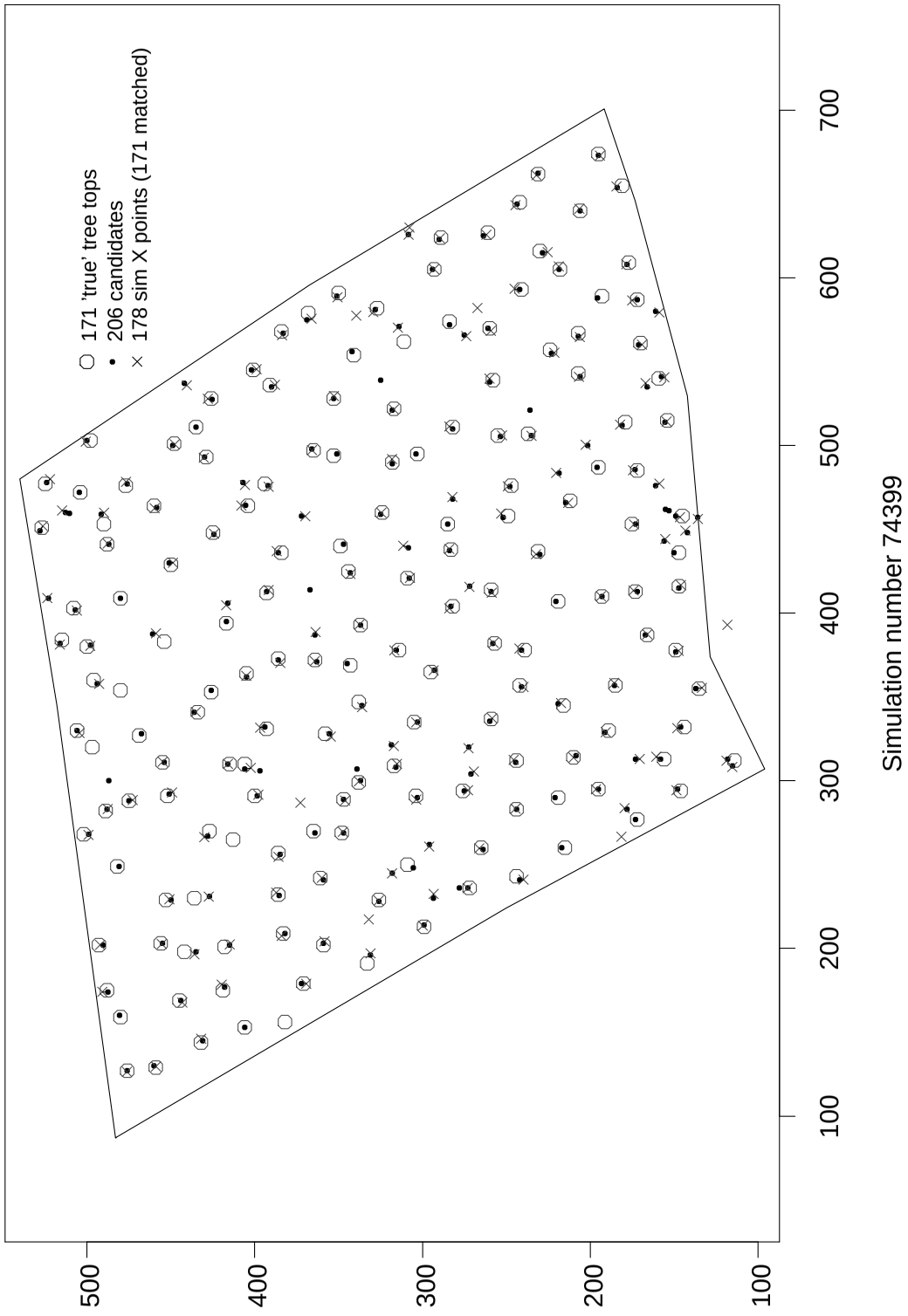

Figure 3: A sample realization from the posterior distribution when the prior with the logistic pair interaction function is used. The crosses $\times$ are the simulated X-points. The matched X-points are connected to the corresponding Y-point by a line.

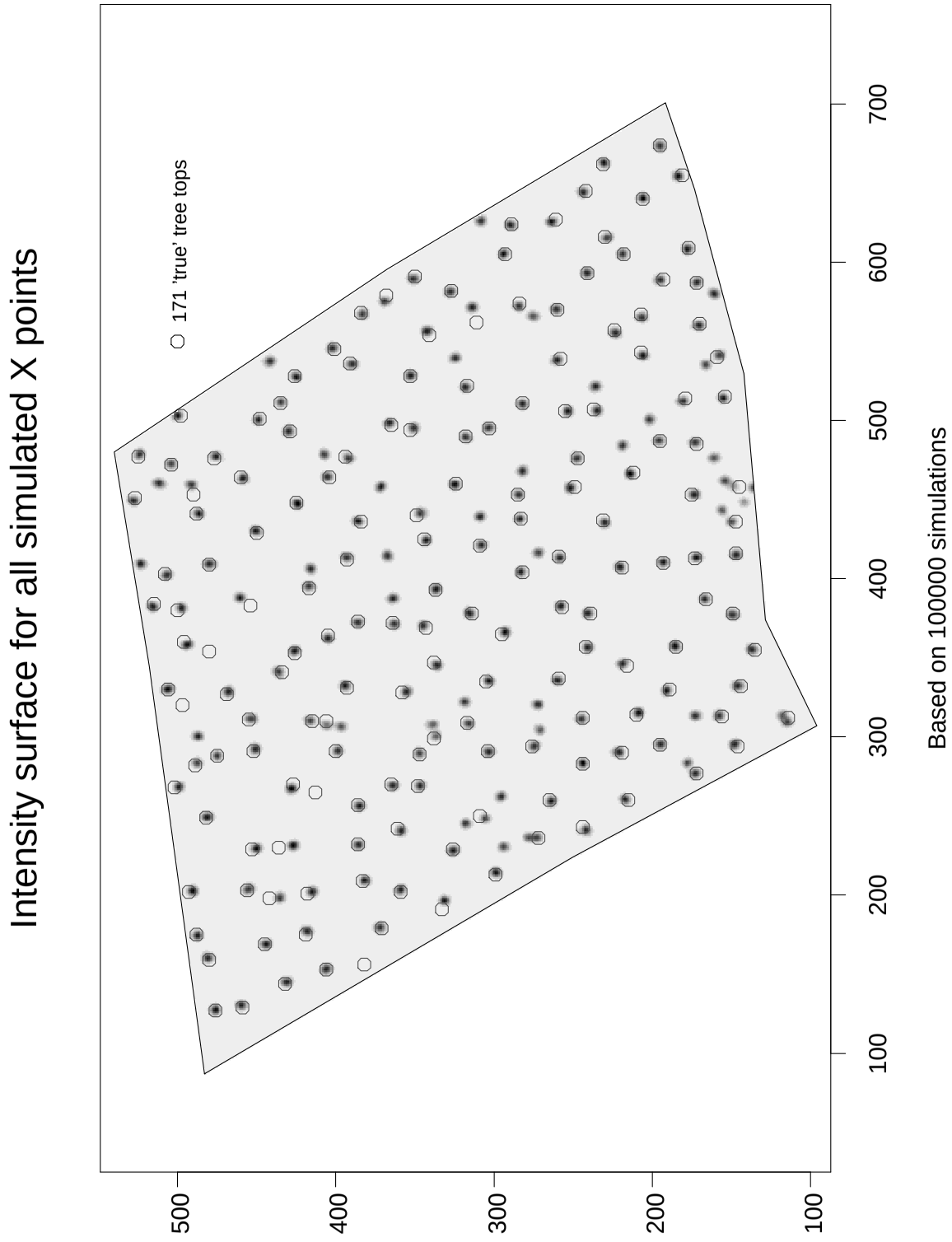

Figure 4: Estimated posterior intensity surface (black corresponds to high intensity) for all X-points together with the 'true' X-points (centres of circles). The prior with logistic pair interaction function is used. The maximum of the intensity estimate in this figure is 0.0351.

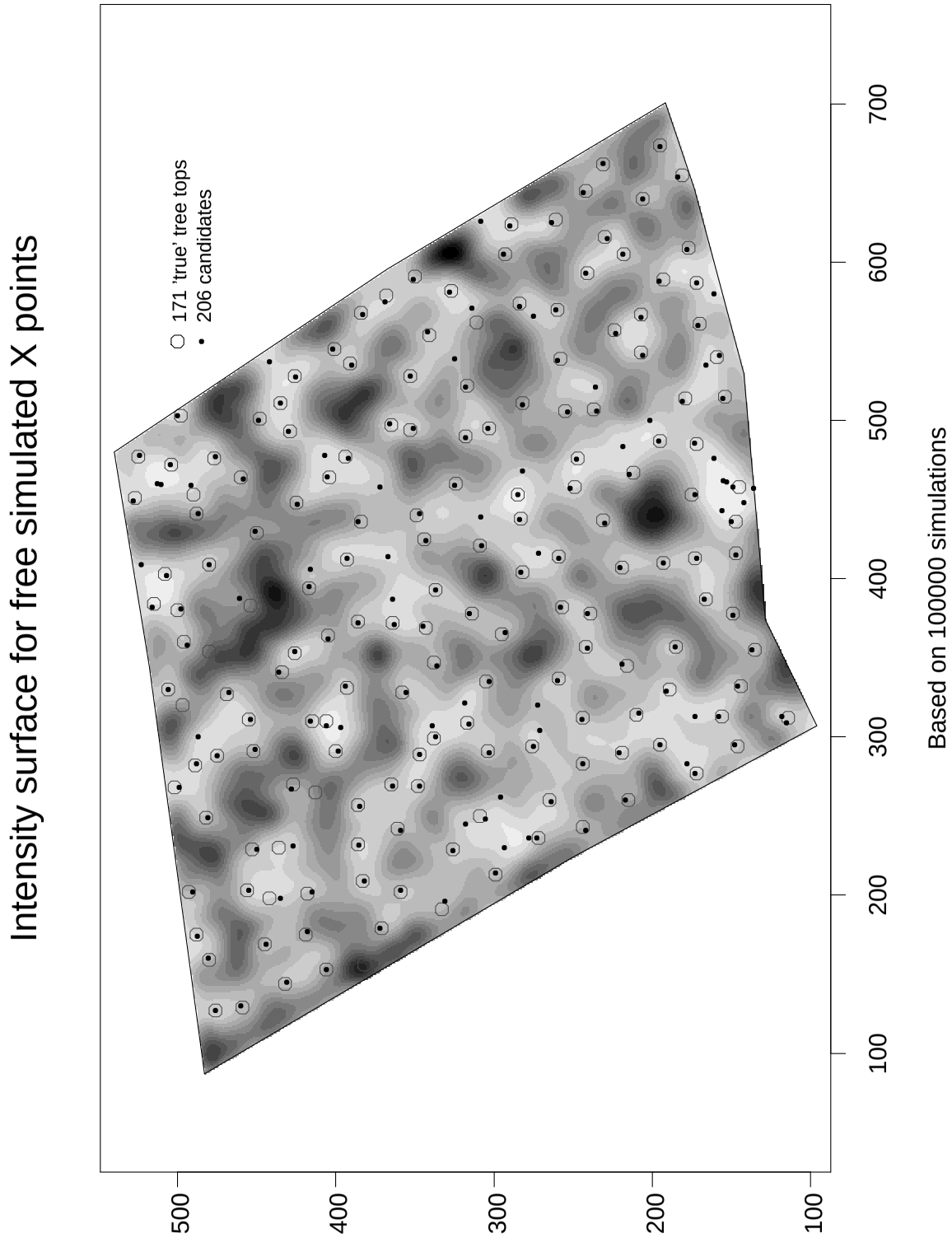

Figure 5: Estimated posterior intensity surface (black corresponds to high intensity) for free X-points together with the 'true' X-points (centres of circles) and the observed Y-points (dots). This figure is based on the same data as Figure 4. The maximum of the intensity estimate in this figure is 0.000154.

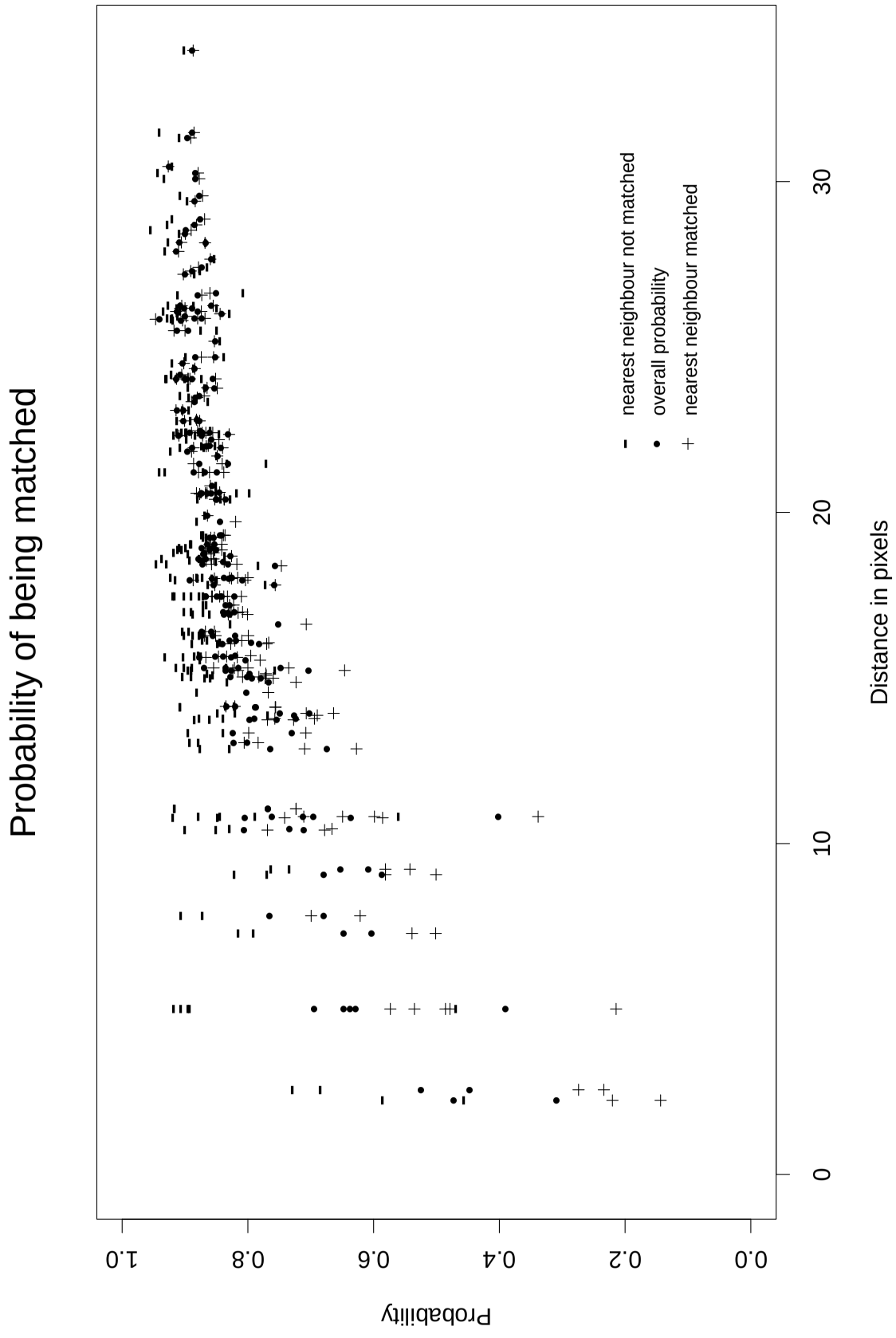

Figure 6: The estimated probability that one Y-point gives rise to a matched X-point as a function of the distance to the nearest neighbour Y-point and whether it is matched or not. The prior with logistic pair interaction function is used.

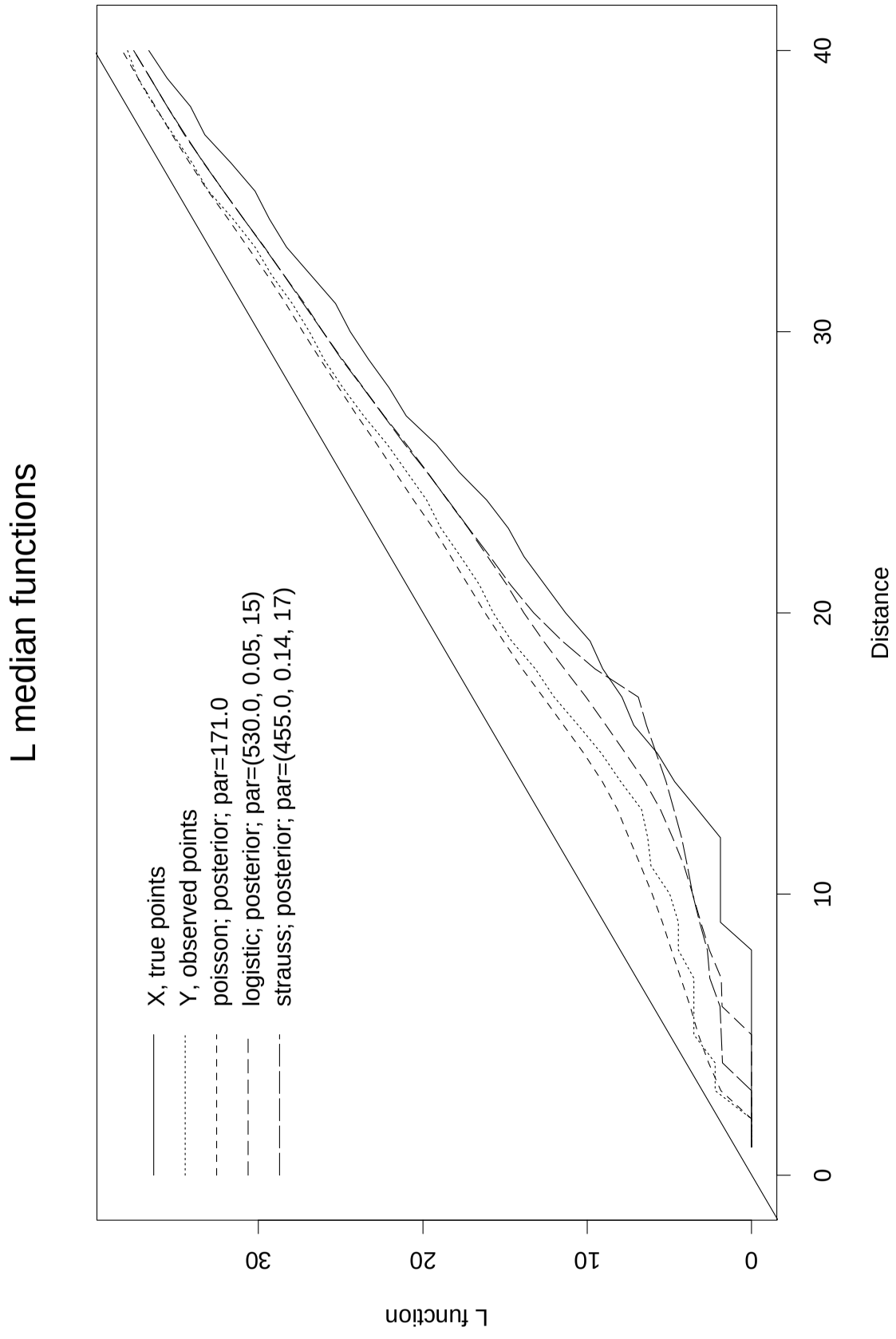

Figure 7: Posterior L function estimates. The straight unbroken line is the estimate for a Poisson process. The other curves are for the true data X, the observed data Y, and for the posterior distributions with a Poisson, Strauss, and Logistic prior, respectively.

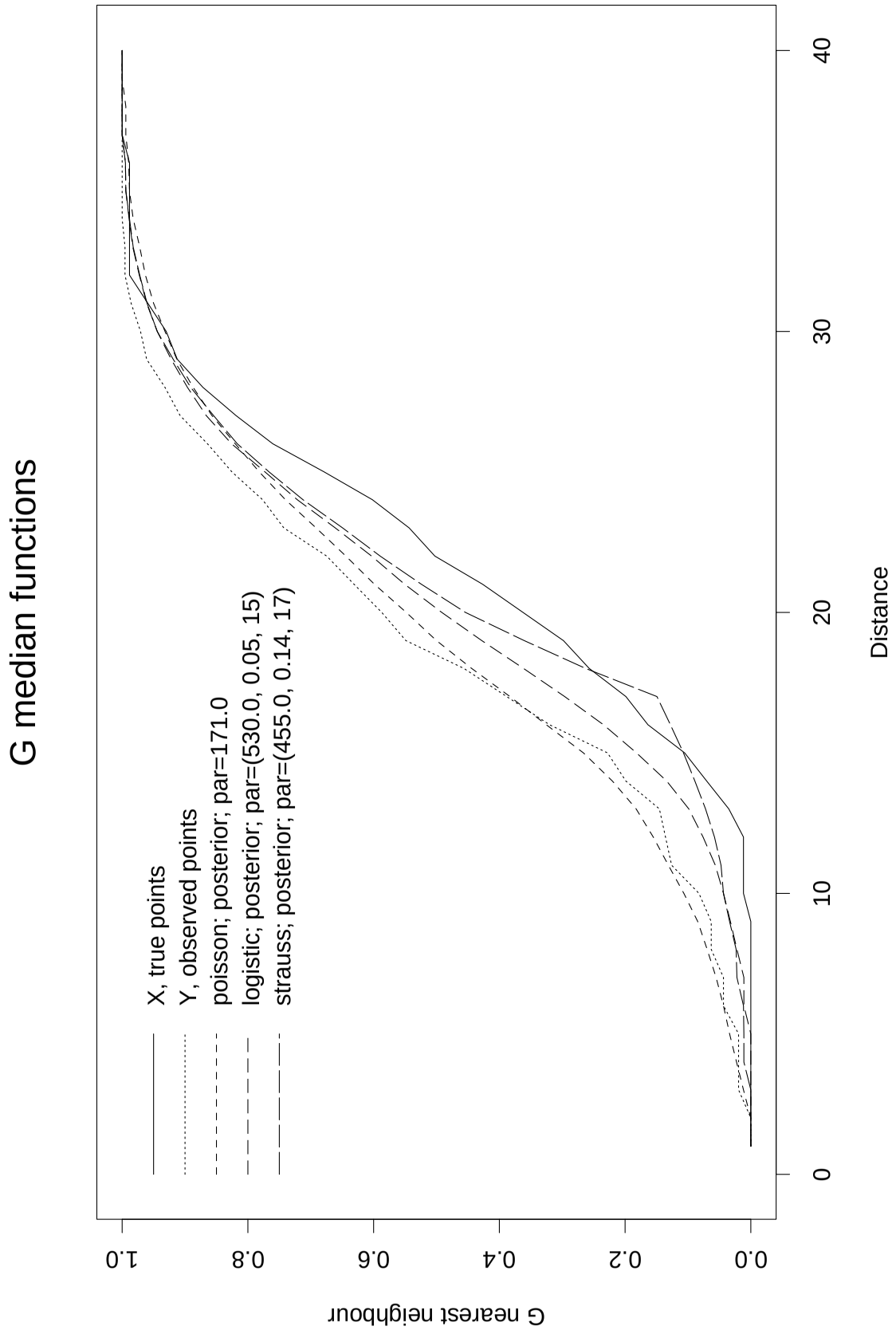

Figure 8: Posterior G function estimates.

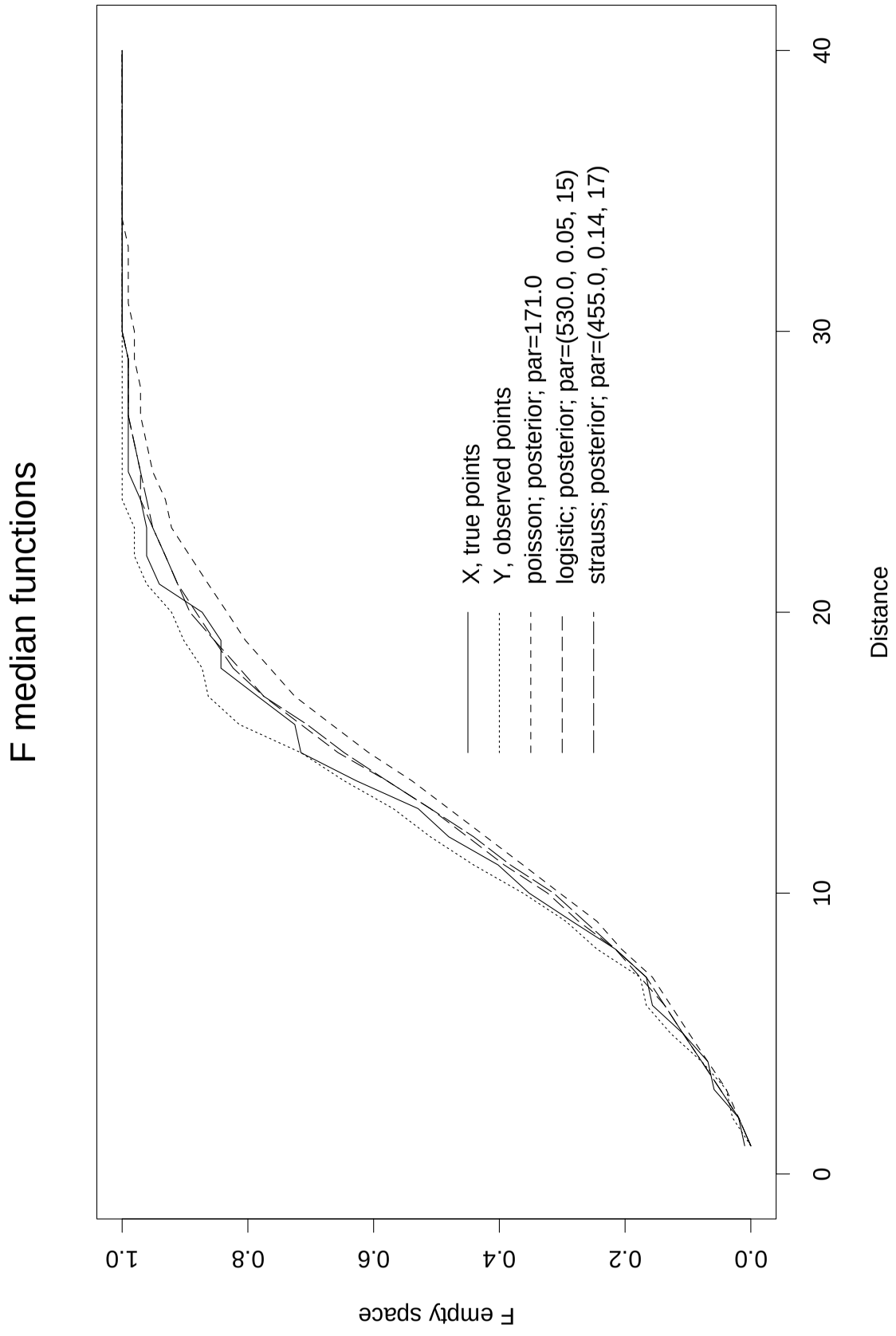

Figure 9: Posterior F function estimates.

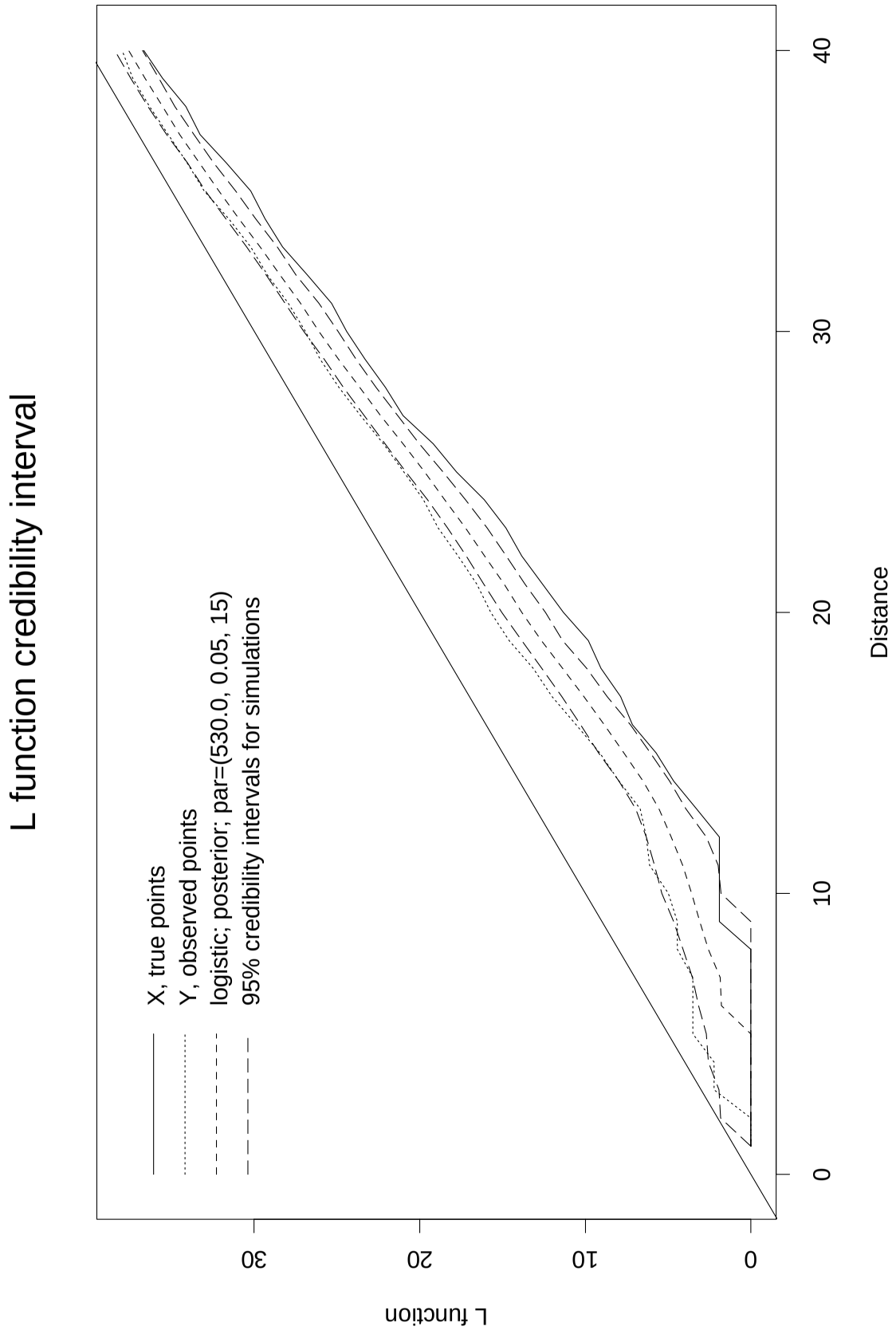

Figure 10: Posterior L function estimate with 95% credibility intervals.

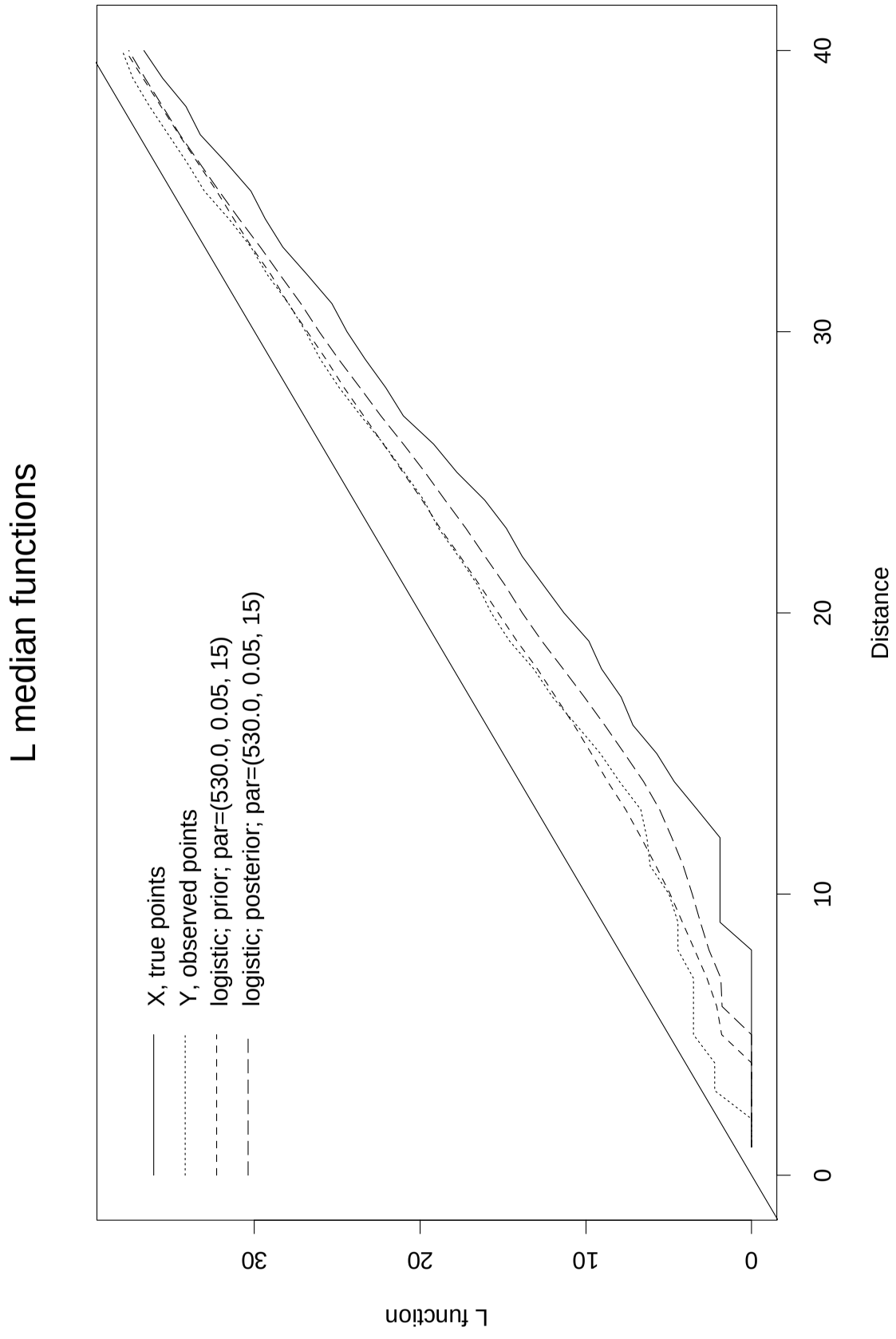

Figure 11: Comparison of L-functions for the observed points Y, the true X-points, the logistic prior and the corresponding posterior estimate.