Space-time multi type log Gaussian Cox processes with a view to modeling weed data

Anders Brix*
Royal Veterinary and Agricultural University

Jesper Møller†
MaPhySto

Summary. Inhomogeneous log Gaussian Cox process models for multi type point patterns are studied and extended to spatial-temporal models called log Gaussian Cox birth-death processes. This is used in a detailed statistical analysis of a particular agricultural experiment concerning the development of weeds on an organic barley field.

Keywords: Inhomogeneous point patterns; Log Gaussian Cox birth-death process; Metropolis adjusted Langevin algorithm; Minimum contrast estimation; Pair correlation function; Prediction; Third order summary statistic; Time series of bivariate point patterns.

1 Introduction

So far most applications of spatial point processes have concerned homogeneous and relatively small point patterns observed at a fixed time, see, for example, Diggle (1983) and Stoyan *et al.* (1995). The purpose of this paper is to show how log Gaussian Cox processes as introduced in Møller *et al.* (1998) can be extended to the inhomogeneous case and used in the statistical analysis of a time series of multi type point patterns. For simplicity we consider only the case where $X(t) = (X_1(t), X_2(t))$ is a bivariate point process and $t \geq 0$ denotes time; however, everything is easily extended to cases with three or more types of points. The discussion is related to a particular agricultural experiment concerning the development of weeds on an organic barley field, see Section

*Address for correspondence: Department of Mathematics and Physics, Thorvaldsensvej 40, DK-1871 Frederiksberg C, Denmark
E-mail brix@dina.kvl.dk

†MaPhySto – Centre for Mathematical Physics and Stochastics, funded by a grant from The Danish National Research Foundation.

2. We consider the two most common species, which are observed at eight different dates in a design consisting of 45 frames, and demonstrate that a certain log Gaussian Cox birth process model for $X(t)$, $t \geq 0$, fits the data very well (no plants are "dying" during the experiment). Briefly, under that log Gaussian Cox birth process model, the marginal distribution of $X(t)$ is a bivariate log Gaussian Cox process defined as follows. For $i = 1, 2$, $s \in \mathbf{R}^2$, $t \geq 0$, consider a random "intensity surface" given by $s \mapsto \exp(Y_i^{(t)}(s))$ with $Y_i^{(t)}(s) = G(s) + G_i(s) + m_i^{(t)}(s)$, where G, G_1, G_2 are independent Gaussian processes with mean 0 and $m_i^{(t)}$ is a mean function so that $m_i^{(t_1)} \leq m_i^{(t_2)}$ whenever $0 \leq t_1 < t_2$. Here G can be thought of as a common environmental factor, while G_1 and G_2 may reflect the influence of the seed bank for the two species. Then, conditional on $(Y_1^{(t)}, Y_2^{(t)})$, $X_1(t)$ and $X_2(t)$ are independent Poisson point processes with intensity functions $\exp(Y_1^{(t)})$ and $\exp(Y_2^{(t)})$, respectively. Moreover, for $0 \leq t_1 < t_2$, we can obtain $X(t_1)$ from $X(t_2)$ by independent thinning of the points in $X(t_2)$, where the thinning probability for a point located at $s \in \mathrm{R}^2$ is given by $p_{i,t_1,t_2}(s) = 1 - m_i^{(t_1)}(s)/m_i^{(t_2)}(s)$. Actually, we show that we can model $m_i^{(t)}$ so that $p_{i,t_1,t_2}(s)$ does not depend on s.

Useful properties of bivariate log Gaussian Cox processes are considered in Section 3.1. Bivariate log Gaussian Cox birth processes as briefly introduced above are discussed further in Section 3.2. Section 4 is concerned with the statistical analysis of the dataset for the development of weeds: We investigate to what extent there is non-stationarity and anisotropy in the observed point patterns, propose parametric forms for the mean and covariance functions of the underlying Gaussian processes, estimate the unknown parameters, check the estimated bivariate log Gaussian Cox birth process model, and discuss how the unobserved intensity surfaces can be predicted. Finally, in Section 5, we report briefly on an alternative model for the data and discuss how our model easily can be extended to include death of plants using a so-called log Gaussian Cox birth-death process model.

2 Description of data

The data considered in this paper originates from an experiment made on a Danish organic barley field in May 1996, where $30\,cm \times 20\,cm$ metal frames were placed in a design consisting of 9 plots each containing 5 frames as shown in Figure 1. The size of the experimental area is $750\,cm \times 500\,cm$, where the longest side agrees with the ploughing direction. The distance between two neighboring plots is $150\,cm$ in the ploughing direction and $100\,cm$ perpendicular to the ploughing direction. Two neighboring frames in the same plot are $30\,cm$ apart in the ploughing direction and $20\,cm$ apart perpendicular to the ploughing direction.

Photos of every metal frame were taken at eight different dates (7/5, 9/5, 13/5, 16/5, 20/5, 23/5, 27/5, and 30/5 1994), every plant (weed and barley) was identified to the most specific taxonomical level based on Haas and Laursen (1994), and the approximate

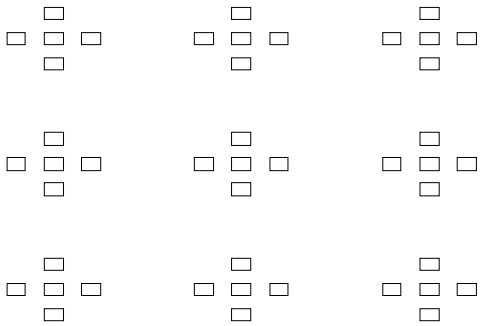

Figure 1: Sampling design for the mapped point pattern data set.

Species	7/5	9/5	13/5	16/5	20/5	23/5	27/5	30/5
Veronica spp.	16	58	174	268	321	339	378	406
Trifolium spp.	105	234	503	717	819	860	949	976

Table 1: Counts of Trifolium spp. and Veronica spp. at the eight dates.

position of the plant was noted. The photos were developed to Kodak* photo CD and these digitized images were used to mark the exact position of each plant by help of the notes made in the field. Due to the roughness of the field it was not possible to place the camera at exactly the same position nor in exactly the same height at each photo. The positions of weed plants used in this paper is thus the rectified coordinates for the weeds, where the rectification is based on translation, rotation, scaling and skewness estimated from the corners of the metal frames, see the documentation of the ArcInfo* function `transform` (Anonymous (1993)) for more details.

In the sequel we restrict ourselves to considering two species, Trifolium spp. (clover) and Veronica spp. (speedwell), which are the two most frequent species in the survey. The bivariate point patterns $(x_1(t), x_2(t))$ observed at the eight dates $t_1 < \ldots < t_8$ form an increasing sequence so that $x_i(t_{j-1}) \subset x_i(t_j),\ i = 1, 2,\ j = 2, \ldots, 8$. Table 1 shows the increasing number of plants for the two species at the different dates. Barley was sown 26/4 1996, and due to ploughing of the field immediately before sowing, no plants were present in the field at this date. The bivariate point pattern at date 30/5 for each of the nine plots is shown in Figure 2. In order to get an idea of the temporal development, Figure 3 shows the bivariate point pattern in the central plot for the eight dates.

3 Statistical model

In this section we study bivariate log Gaussian Cox processes and relate these to the data in Section 2. In Section 3.1 we consider the fixed time case and extend some results in

*Kodak and ArcInfo are registered trademarks.

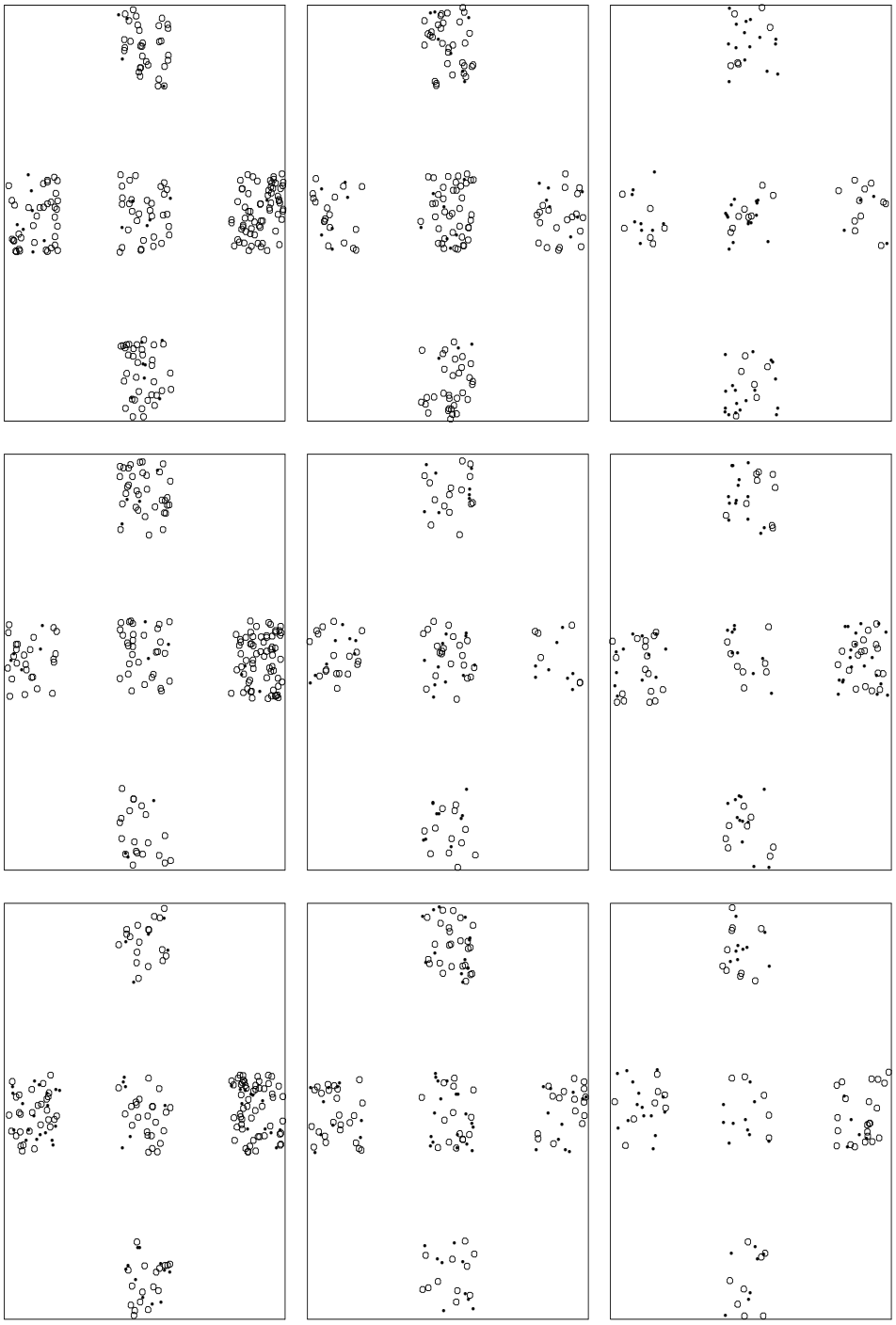

Figure 2: The mapped bivariate point pattern at date 30/5 within each plot. Small circles correspond to locations of Trifolium spp. and dots to Veronica spp. The lower right plot corresponds to the lower left plot in Figure 1 (the distances between plots are reduced in order to clarify the display of the weed positions).

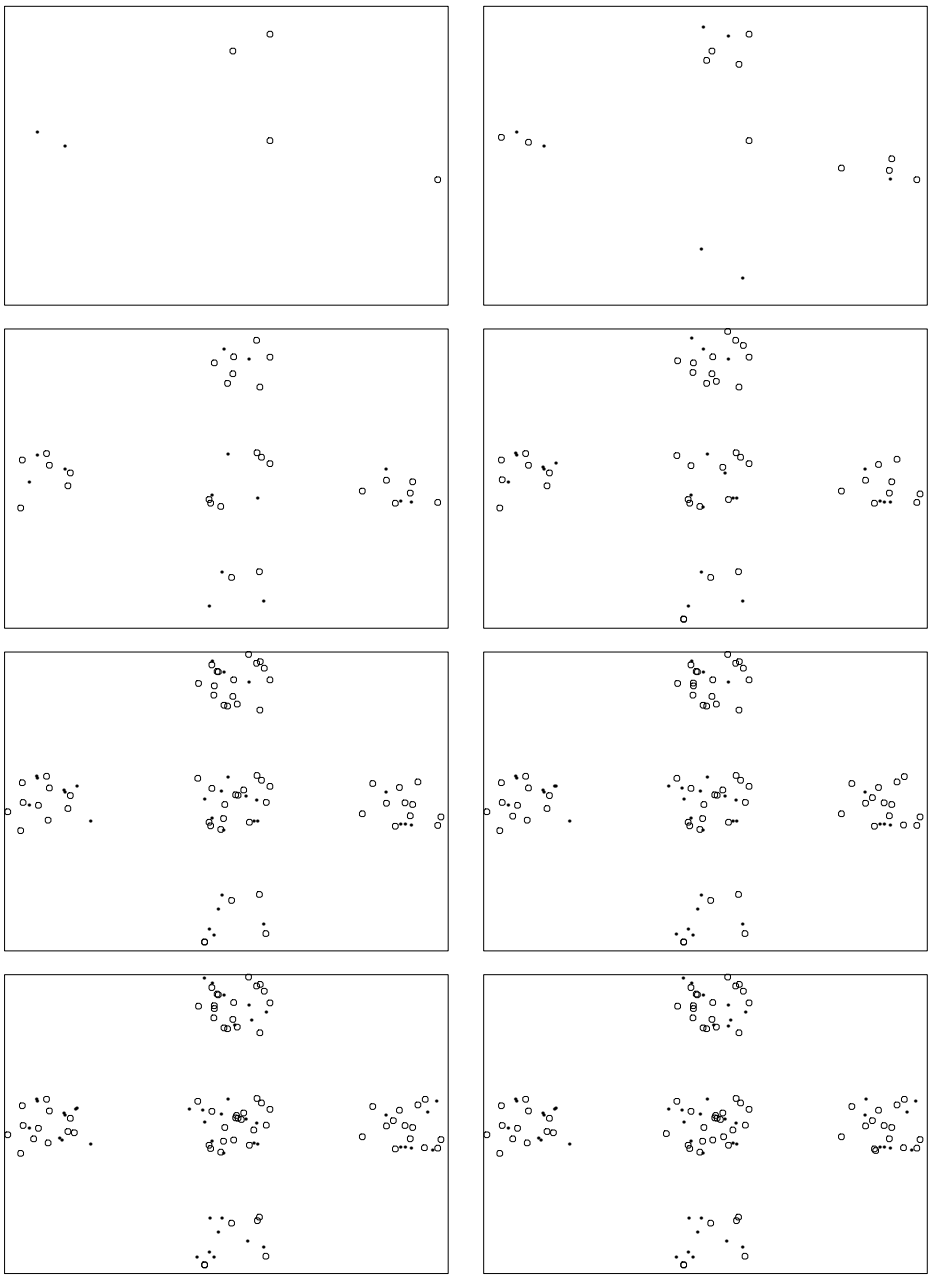

Figure 3: The bivariate point pattern in the central plot for the different dates. Small circles correspond to locations of Trifolium spp. and dots to Veronica spp.

Møller *et al.* (1998) to the inhomogeneous case in space, while extensions to the temporal domain are considered in Section 3.2.

3.1 Inhomogeneous bivariate log Gaussian Cox processes

As explained in more detail below we consider bivariate log Gaussian Cox processes where the intensity is allowed to be inhomogeneous while the pair correlation functions are assumed to be stationary and isotropic.

Recall first the definition of a planar bivariate Cox process $X = (X_1, X_2)$ directed by a random intensity process $\Lambda = (\Lambda_1, \Lambda_2)$ which is locally integrable so that $\nu_i(B) := \int_B \Lambda_i(s)\,\mathrm{d}s < \infty$ for all bounded Borel sets $B \subset \mathbf{R}^2$ and $i = 1, 2$. Conditional on Λ, X_1 and X_2 are independent Poisson processes with intensity functions Λ_1 and Λ_2, respectively. If $\Lambda_i(s) = \exp(Y_i(s))$, $i = 1, 2$, where $Y = (Y_1, Y_2)$ is a two-dimensional real-valued Gaussian process defined on $\mathbf{R}^2$, then X is called a *bivariate log Gaussian Cox process*. As discussed in Møller *et al.* (1998) it is necessary to impose weak conditions on the mean functions $m_i(s) = EY_i(s)$ and covariance functions $c_{ii}(s_1, s_2) = Cov(Y_i(s_1), Y_i(s_2))$, $i = 1, 2$, in order to ensure that each Λ_i is integrable and that the distribution of ν_i is uniquely determined by the distribution of Y_i.

Henceforth, we impose the natural condition that the mean functions m_1 and m_2 are continuous. Further, we assume that the covariance functions are stationary and isotropic so that $c_{ii}(s_1, s_2) = \sigma_i^2 r_{ii}(||s_1 - s_2||)$, where σ_i^2 is the variance of Y_i and r_{ii} its correlation function; the cross covariance functions $c_{ij}(a) = Cov(Y_i(s_1), Y_j(s_2))$, $i \neq j$, are also assumed to depend only on the distance $a = ||s_1 - s_2||$ whereby $c_{12} = c_{21}$. Then, if there exist some positive numbers K, ϵ, δ so that $1 - r_{ii}(a) < K(-\log a)^{-(1+\epsilon)}$ whenever $a < \delta$, we can assume that Y has continuous sample paths whereby Λ_i is integrable and ν_i is uniquely determined. Moreover, the covariance function matrix $c(a) = \{c_{ij}(a)\}_{i,j=1,2}$ has to be positive semi-definite. If the c_{ij} are absolutely integrable, then c is positive semi-definite if and only if

$$C_{ii}(a) \geq 0, \quad C_{12}(a)^2 \leq C_{11}(a)C_{22}(a), \quad i = 1, 2, \ a \geq 0, \tag{1}$$

where

$$C_{ij}(a) = \frac{1}{2\pi} \int_0^\infty J_0(ab) c_{ij}(b) b \,\mathrm{d}b$$

is the Hankel transform of c_{ij} and J_0 is the Bessel function of first kind and order zero. In the particular case where

$$Y_i = \sum_{j=1}^{k} \alpha_{ij} Z_j + m_i \tag{2}$$

with parameters $\alpha_{ij} \in \mathbf{R}$ and the Z_j being mutually independent Gaussian processes with mean 0, variance 1, and stationary isotropic correlation functions r_j, we have immediately

that c is well-defined and specified by

$$c_{ii} = \sum_{j=1}^{k} \alpha_{ij}^2 r_j\,, \quad c_{12} = \sum_{j=1}^{k} \alpha_{1j}\alpha_{2j} r_j\,. \tag{3}$$

There is a simple one-to-one relationship between the mean and covariance functions of the Gaussian field Y and the first and second order characteristics of the bivariate point process X. Let $\rho_i(s) = E\Lambda_i(s)$ be the intensity function of X_i, $i = 1, 2$, and define the pair correlation functions

$$g_{ij}(a) = E[\Lambda_i(s_1)\Lambda_j(s_2)]/(\rho_i(s_1)\rho_j(s_2)), \quad i, j \in \{1, 2\},\ a = ||s_1 - s_2||. \tag{4}$$

Roughly speaking, $g_{ij}(a)\rho_i(s_1)\rho_j(s_2)\mathrm{d}s_1\mathrm{d}s_2$ is the probability that X has a point of type i and another of type j at the infinitesimally small regions of volumes $\mathrm{d}s_1$ and $\mathrm{d}s_2$, respectively. We have that

$$\rho_i(s) = \exp(m_i(s) + \sigma_i^2/2), \quad g_{ij}(a) = \exp(c_{ij}(a)). \tag{5}$$

This means that the distribution of (X, Y) is specified by the functions ρ_i, g_{ij}, $1 \le i \le j \le 2$.

For the estimation of g_{ij} we need a definition of the second reduced moment function K_{ij}. Suppose we have observed a bivariate point pattern (x_1, x_2) of X restricted to some bounded window W with area $|W| > 0$. In the inhomogeneous case we follow Baddeley *et al.* (1998) and define, for $a \ge 0$,

$$K_{ij}(a) = \frac{1}{|W|}\mathrm{E} \sum_{\xi \in X_i \cap W, \eta \in X_j, \xi \neq \eta} \frac{\mathbf{1}[||\xi - \eta|| \le a]}{\rho_i(\xi)\rho_j(\eta)} = 2\pi \int_0^a r g_{ij}(r)\,\mathrm{d}r\,, \tag{6}$$

where $\mathbf{1}[\cdot]$ denotes the indicator function. Note that this is in agreement with the definition in the stationary case where the ρ_i are constant and $\rho_i^2 K_{ij}(a)$ is interpreted as the mean number of points of type j within distance a of a typical point of type i (see, e.g., Ripley (1977) and Diggle (1983)). Further, $K_{12} = K_{21}$. We use (6) for estimating g_{ij} in Section 4.2.

Furthermore, higher order product densities for X can be expressed in terms of ρ_i, g_{ij}, $1 \le i \le j \le 2$, see Theorem 1 in Møller *et al.* (1998). In particular, the third order product density $\rho_i^{(3)}(\xi, \eta, \kappa) = E[\Lambda_i(\xi)\Lambda_i(\eta)\Lambda_i(\kappa)]$ is given by

$$\rho_i^{(3)}(\xi, \eta, \kappa) = \rho_i(\xi)\rho_i(\eta)\rho_i(\kappa) g_{ii}(||\xi - \eta||) g_{ii}(||\xi - \kappa||) g_{ii}(||\eta - \kappa||).$$

Hence, the third order characteristic $z_i(a)$ defined by the expression

$$E \sum_{\substack{\xi,\eta,\kappa \in X_i\,:\,\xi \in W,\\ \xi \neq \eta, \xi \neq \kappa, \eta \neq \kappa,\\ 0 < ||\xi - \eta|| \le a, 0 < ||\xi - \kappa|| \le a}} 1/[\pi^2 a^4 |W| \rho_i(\xi)\rho_i(\eta)\rho_i(\kappa) g_{ii}(||\xi - \eta||) g_{ii}(||\xi - \kappa||) g_{ii}(||\eta - \kappa||)] \tag{7}$$

is easily seen to be equal to 1 for all distances $a > 0$ and $i = 1, 2$, when X_i is a log Gaussian Cox process. An unbiased estimator of $z_i(a)$ given by (7) is easily obtained by modifying Theorem 2 in Møller *et al.* (1998) as follows. Let $\psi(\xi, \eta, \kappa)$ denote the angle (anti clockwise) between $\eta - \xi$ and $\kappa - \xi$. For $\xi \in W,\ a > 0,\ b > 0,\ 0 \le \psi < 2\pi$, set $2\pi/\omega_{\xi,a,b,\psi}$ equal to the length of

$$U_{\xi,a,b,\psi} = \{\varphi \in [0, 2\pi) \quad | \quad \xi + a(\cos\varphi, \sin\varphi) \in W, \\ \xi + b(\cos(\varphi + \psi), \sin(\varphi + \psi)) \in W\},$$

taking $\omega_{\xi,a,b,\psi} = \infty$ if the length of $U_{\xi,a,b,\psi}$ is 0. Then

$$\sum_{\substack{\xi \in X_i \cap W, \\ \{\eta,\kappa\} \subseteq X_i \cap W: \\ \xi \neq \eta, \xi \neq \kappa, \eta \neq \kappa, \\ ||\xi - \eta|| \le a, ||\xi - \kappa|| \le a}} \frac{2\omega_{\xi,||\xi-\eta||,||\xi-\kappa||,\psi(\xi,\eta,\kappa)}}{\pi^2 a^4 |W| \rho_i(\xi)\rho_i(\eta)\rho_i(\kappa) g_{ii}(||\xi - \eta||) g_{ii}(||\xi - \kappa||) g_{ii}(||\eta - \kappa||)} \tag{8}$$

is an unbiased estimator of $z_i(a)$ whenever $a < a_i^*$, where

$$a_i^* = \inf\{a > 0 | \int_{\xi \in W} \int_{a \in (0,t]} \int_{b \in (0,t]} \int_{\psi \in [0,2\pi)} \mathbf{1}[\omega_{\xi,a,b,\psi} = \infty] \, \mathrm{d}\psi \, \mathrm{d}a \, \mathrm{d}b \, \mathrm{d}\xi > 0\}.$$

This is used in Section 4.4 for model checking.

Having estimated the mean and covariance functions of Y using the methods to be discussed in Sections 4.1–4.2, it is possible to predict Y on any rectangular region containing a sub-window T of W. This is further considered in Section 4.5.

3.2 Spatio-temporal bivariate log Gaussian Cox processes

We now return to the data in Section 2 and specify the general space-time model for the increasing sequence of bivariate point patterns $(x_1(t), x_2(t))$ observed at the eight different times $t_1 < \ldots < t_8$. Letting W denote the union of the 45 frames in Figure 1, we consider the data as a realization of a time-continuous bivariate point process $X(t) = (X_1(t), X_2(t))$ defined on $\mathbf{R}^2$ so that the realizations $X_i(t) \cap W = x_i(t),\ i = 1, 2$, are observed at times $t = t_1, \ldots, t_8$.

It seems natural to model $\{X(t) : t \ge 0\}$ as a spatial birth process. Specifically, conditional on a bivariate Gaussian process $Y = \{Y(s) : s \in \mathbf{R}^2\}$ as considered in Section 3.1, we let $\{X_i(t) : t \ge 0\},\ i = 1, 2$, be independent Poisson birth processes: conditional on Y, X_i is a Poisson process on $\mathbf{R}^2 \times [0, \infty)$ with intensity measure $\nu_i \times \gamma_i$, where ν_i is defined in the beginning of section 3.1, and γ_i is a measure defined on the Borel sets of $[0, \infty)$ so that $0 < \gamma_i([0, t]) < \infty$ for all $t > 0$. We call $\{X(t) : t \ge 0\}$ a *bivariate log Gaussian Cox birth process*.

If $t = 0$ corresponds to the date 26/4 1996, where no plants were present, then $X(0) = (\emptyset, \emptyset)$, where $\emptyset$ denotes the empty point configuration. For any fixed $t > 0$ we have that conditional on Y, $X_i(t)$ is a Poisson process with intensity measure $\nu_{i,t}(B) =$

$\nu_i(B)\gamma_i([0,t])$, and so $X(t)$ is a bivariate log Gaussian Cox process driven by the log Gaussian process $\{(\exp(Y_1(s))\gamma_1([0,t]), \exp(Y_2(s))\gamma_2([0,t])) : s \in \mathbf{R}^2\}$. Hence, by (4) and (5), the intensity functions at time t becomes

$$\rho_i^{(t)}(s) = \gamma_i([0,t]) \exp(m_i(s) + \sigma_i^2/2), \quad i = 1, 2,$$

while the pair correlation functions of $X(t)$ do not depend on the time t. In Section 4 we utilize these properties together with the fact that $X(t_8), \ldots, X(t_1)$ is a Markov chain where $X(t_{j-1})$ is obtained from $X(t_j)$ by independent thinning, i.e. when the points in $X_i(t_j)$ are independently deleted with probability

$$p_{ij} = \gamma_i((t_{j-1}, t_j])/\gamma_i([0, t_j]). \tag{9}$$

We consider mainly the following model for Y. Let Z, U_1, U_2 be three independent univariate Gaussian processes with mean 0, variance 1, and correlation functions r, r_1, r_2, respectively. Then

$$Y_i = \omega Z + \sigma_i U_i + m_i, \quad i = 1, 2, \tag{10}$$

where $\omega > 0$ and $\sigma_i > 0$ are parameters and m_i is the mean function of Y_i. This model has the advantage of being well defined for any choice of correlation functions r, r_1, r_2. Furthermore, it is easily interpretable as the common component Z of Y_1 and Y_2 may reflect unobserved environmental factors such as soil nutrients or soil humidity, while U_i could reflect the bank of seeds of species i. Combining (3) and (10) we have that

$$c_{12} = \omega^2 r, \quad c_{ii} = \omega^2 r + \sigma_i^2 r_i, \quad i = 1, 2. \tag{11}$$

In Section 5 we comment briefly on an alternative model for Y which is neither of the type (10) nor (2).

4 Statistical analysis

Throughout this section we let the situation be as in Section 3.2, i.e. X denotes a bivariate log Gaussian Cox birth process with Y specified by (10) and the γ_i being arbitrary locally finite measures with $\gamma_i([0,t]) > 0$ for $t > 0$. The statistical analysis in Sections 4.1–4.2 is mainly based on the data $x(t_8)$, while the entire dataset is considered in Sections 4.3–4.4. In Section 4.1 we examine inhomogeneity and isotropy properties, and we estimate the intensity functions. In Section 4.2 we consider non-parametric estimation of the covariance functions, propose a parametric model for the correlation functions in (11), and estimate the unknown parameters by a minimum contrast method. The thinning probabilities p_{ij} in (9) are estimated in Section 4.3 by non-parametric methods. Validation of the estimated parametric model from Section 4.2 is considered in Section 4.4. Finally, in Section 4.5 we discuss prediction of the unobserved Gaussian process.

4.1 Analysis of non-stationarity and anisotropy

In many applications of point process models, stationarity and isotropy is often assumed if the observation window is too small to asses whether apparent heterogeneity is due to anisotropy or non-stationarity or if it is caused by random variation. However, if a parametric form is assumed, even a small observation window may be used to determine if anisotropy or non-stationarity is present. Then the problem is to choose a parametric form which is reasonable in relation to the problem considered. We will here analyze some possible parametric forms of non-stationarity and anisotropy which, for our data, can be explained by ploughing and sowing in the same direction year after year.

Ploughing furrows may cause non-stationarity in form of periodicity of the intensity perpendicular to the ploughing direction due to the relief of the field. This can be checked by projecting the intensity onto the axis perpendicular to the ploughing direction. For neither of the two species such a periodic non-stationarity could be observed. However, the projected intensity for Trifolium spp. seemed to be increasing, while it appeared to be constant for Veronica spp. For Trifolium spp., Figure 4 shows the linear regression of the logaritmized intensity of points in each frame against the position of the frame in the direction perpendicular to the ploughing direction at time t_8. Logarithms have been taken in order to stabilize the variances. The straight line in Figure 4 is the estimated regression line, and a t-test for a horizontal line turned out to be highly significant ($p < 10^{-8}$).

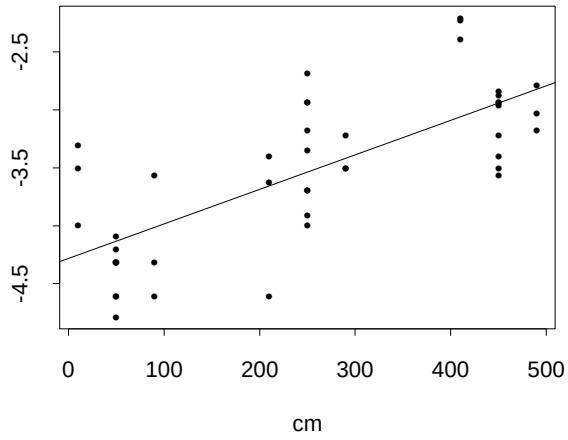

Figure 4: Trifolium spp. at the latest date; linear regression of the logaritmized intensity of points in each frame against the position of the frame in the direction perpendicular to the ploughing direction.

We will thus assume that the positions of Veronica spp. constitutes a stationary point pattern, whereas the intensity for Trifolium spp. can be parameterized as

$$\rho_2^{(t_8)}(s) = \exp(a + bs_2), \quad s = (s_1, s_2) \in \mathbf{R}^2. \tag{12}$$

Estimation of a and b is done by ordinary linear regression of the logaritmized intensities, giving $\hat{\rho}_2^{(t_8)}(s) = \exp(\hat{a} + \hat{b}s_2)$ with $\hat{a} = -4.28$ and $\hat{b} = 0.0030$. The intensity for Veronica

spp. is estimated by the average of plants per $1\,cm^2$ giving $\hat{\rho}_1^{(t_8)} = 0.015$. At the other 7 dates (7/5–27/5) we have checked non-stationarity by the same methods and found the same pattern; Veronica spp. seems to form a stationary point process, while Trifolium spp. appears to have a log-linear trend perpendicular to the ploughing direction.

Ploughing and sowing may also cause anisotropy which causes larger correlations in the ploughing direction. Possible parametric forms are provided by covariance functions with elliptic contour curves and main axis in the ploughing direction. This type of anisotropy is called geometric anisotropy (see e.g. Cressie (1987)). It can be investigated by an orientation analysis using directional K-functions defined as follows (see also Stoyan and Stoyan (1995)). For $t = t_8$ (or any other fixed time), $-\pi/2 \leq \phi < \pi/2$, $\phi < \psi \leq \phi + \pi$, and $a \geq 0$, define

$$K_{i,\phi,\psi}(a) = E \sum_{\substack{\xi \in X_i(t_8) \cap W, \\ \eta \in X_i(t_8) \cap S_{\xi,a,\phi,\psi}: \\ \xi \neq \eta}} \mathbf{1}[\|\xi - \eta\| \leq a]/(|W|\rho_i(\xi)\rho_i(\eta)), \quad i = 1, 2, \tag{13}$$

where $S_{\xi,a,\phi,\psi}$ is the 'sector' given by the union of the two segments of a disc with center ξ and radius a when the angles of the first segment is between ϕ, ψ and the second is between $\pi + \phi, \pi + \psi$; here we let the angle 0 correspond to the ploughing direction. Under the assumption of isotropy, $K_{i,\phi,\psi}$ depends only on $\psi - \phi$. We have considered two sectors with $(\phi, \psi) = (\pi/4, 3\pi/4)$ and $(\phi, \psi) = (-\pi/4, \pi/4)$, respectively. Note that $K_{ii} = K_{i,\pi/4,3\pi/4} + K_{i,-\pi/4,\pi/4}$, so setting $M_{i1} = K_{i,\pi/4,3\pi/4}/K_{ii}$ and $M_{i2} = K_{i,-\pi/4,\pi/4}/K_{ii}$, isotropy implies that $M_{ij} = 0.5$, $i, j = 1, 2$.

We estimated $K_{i,\phi,\psi}$ by

$$\hat{K}_{i,\phi,\psi}(a) = \frac{1}{|W|} \sum_{\substack{\xi \in x_i(t_8), \\ \eta \in x_i(t_8) \cap S_{\xi,a,\phi,\psi}: \\ \xi \neq \eta}} \mathbf{1}[\|\xi - \eta\| \leq a] \frac{w_{\xi,\eta}}{\hat{\rho}_i(\xi)\hat{\rho}_i(\eta)}, \quad i = 1, 2, \tag{14}$$

where $w_{\xi,\eta}$ is Ripley's (1977) edge correction factor given by the proportion of the circumference of the circle with center ξ, radius $\|\xi - \eta\|$, and lying within W. Figure 5 shows the estimates of M_{i1} and M_{i2}. The envelopes in Figure 5 (and Figures 8–11 to be considered later on) are constructed as in Diggle (1983); as we are using 19 simulations under the estimated parametric model in Section 4.2, the envelopes are pointwise approximate 90%-confidence intervals. The plot for Veronica spp. does not give any reason to believe that geometric anisotropy is present, while the plot for Trifolium spp. indicates that some anisotropy may be present. However, since it seems difficult to model this possible departure from isotropy, we will continue the analysis assuming isotropy for both species.

4.2 Estimation of the covariance functions

Under the log Gaussian Cox birth process model, the pair correlation functions (and thereby also the covariance functions) are identical at all times. Therefore we start by

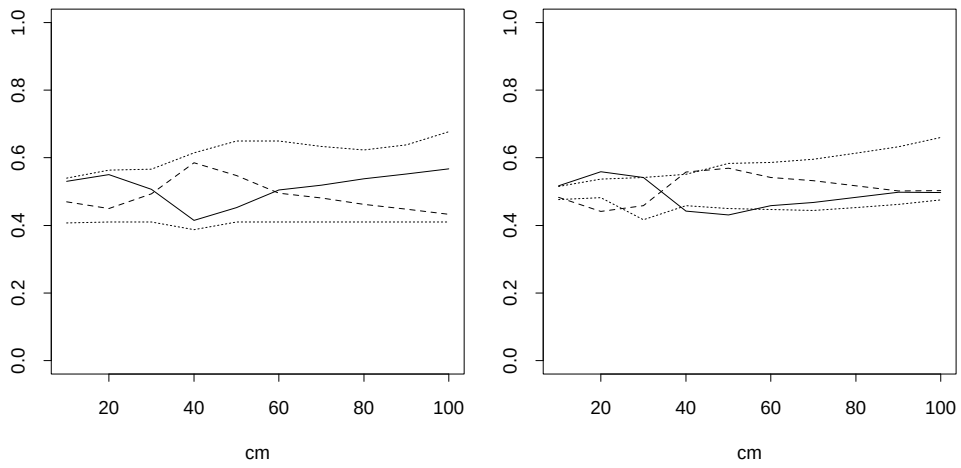

Figure 5: Estimated M_{ij}-functions in the ploughing direction (solid line) and perpendicular to it (dashed line) using a sector width of $\psi - \phi = \pi/2$ for Veronica spp. (left) and Trifolium spp. (right) at the latest date. Envelopes (dotted lines) are based on 19 simulations using the estimated parametric model in Section 4.2.

analyzing the latest bivariate point pattern $x(t_8)$; the other patterns are analyzed in Sections 4.3–4.4.

Inspired by the definition (6) of the K_{ij}-functions for an inhomogeneous bivariate point process, we estimate the pair correlation functions by

$$\tilde{g}_{ij}(a) = \frac{1}{|W|} \sum_{\substack{\xi \in x_i(t_8), \eta \in x_i(t_8): \\ \xi \neq \eta}} \frac{k(\|\xi - \eta\| - a)}{2\pi a} \frac{w_{\xi,\eta}}{\hat{\rho}_i^{(t_8)}(\xi)\hat{\rho}_j^{(t_8)}(\eta)}, \quad i,j \in \{1,2\}, \tag{15}$$

where $w_{\xi,\eta}$ is the edge correction from (14) and k is the Epanečnikov kernel with bandwidth $0.15/(\tilde{\rho}_i\tilde{\rho}_j)^{1/4}$ for all $i,j \in \{1,2\}$, where $\tilde{\rho}_i$ is the average of points in $x_i(t_8)$ per $1\,cm^2$ (see Stoyan and Stoyan (1994)). Since (15) is not symmetric in i and j, the cross pair correlation function is finally estimated by $\bar{g}_{12} = \tilde{\rho}_2/(\tilde{\rho}_1+\tilde{\rho}_2)\tilde{g}_{12} + \tilde{\rho}_1/(\tilde{\rho}_1+\tilde{\rho}_2)\tilde{g}_{21}$. Hence by (5), non-parametric estimators $\bar{c}_{11} = \log \tilde{g}_{11}$, $\bar{c}_{12} = \log \bar{g}_{12}$, $\bar{c}_{22} = \log \tilde{g}_{22}$ of the covariance functions are obtained.

Figure 6 shows the non-parametric estimates of the covariance functions at time t_8. Though the estimates for Trifolium spp. have been corrected for the log-linear trend found in Section 4.1, we observed that they did not differ much from the uncorrected estimates obtained by assuming stationarity, since the estimates of the pair correlation functions based on each plot were very alike at distances less than $20\,cm$ (as considered in Figure 6). For distances about $15\,cm$ and more, a simulation study showed that the estimators $\bar{c}_{ij}$ were heavily biased downwards (see also Figure 6).

The non-parametric estimates of covariance functions in Figure 6 suggest to use exponential covariance function for Z, U_1, U_2 given by $c_{12}(a) = \omega^2 \exp(-a/\beta)$ and

$c_{ii}(a) - c_{12}(a) = \sigma_i^2 \exp(-a/\beta_i)$, $i = 1, 2$, cf. (11). Here $\omega^2, \sigma_i^2, \beta, \beta_i > 0$ are parameters estimated by minimizing integrals on the form

$$\int_{\epsilon_{ij}}^{a_0} (\bar{c}_{ij}(a)^\alpha - c_{ij}(a)^\alpha)^2 \, \mathrm{d}a, \tag{16}$$

where $0 < \epsilon_{ij} < a_0$ and $\alpha > 0$ are user specified parameters. More precisely, we propose first to obtain estimates $\hat{\omega}^2$ and $\hat{\beta}$ from (16) when $i = 1, j = 2$, and then for each $i = 1, 2$, estimate σ_i^2 and β_i by replacing $c_{ii}(a)$ in (16) by $\hat{\omega}^2 \exp(-a/\hat{\beta}) + \sigma_i^2 \exp(-a/\beta_i)$. For this minimum contrast method we found it appropriate to use $a_0 = 10\,cm$, $\alpha = 0.5$, and $\epsilon_{ij} = \min\{\|\xi - \eta\| : \xi \in x_i(t_8), \eta \in x_j(t_8), \xi \neq \eta\}$; similar values were used in Møller *et al.* (1998). .

The minimum contrast method gave $\hat{\beta} = 44.11$, $\hat{\omega}^2 = 0.22$, $\hat{\beta}_1 = 1.24$, $\hat{\sigma}_1^2 = 0.5$, $\hat{\beta}_2 = 23.22$, $\hat{\sigma}_2^2 = 0.26$; inserting these values in the parametric expressions for c_{11}, c_{12}, c_{22}, we obtained the estimated covariance functions $\hat{c}_{11}$, $\hat{c}_{12}$, $\hat{c}_{22}$ for the fitted parametric model, see Figure 6. Finally, the mean functions at time t_8 can be estimated by $\hat{m}_i^{(t_8)}(s) = \log \hat{\rho}_i^{(t_8)}(s) - (\hat{\omega}^2 + \hat{\sigma}_i^2)/2$.

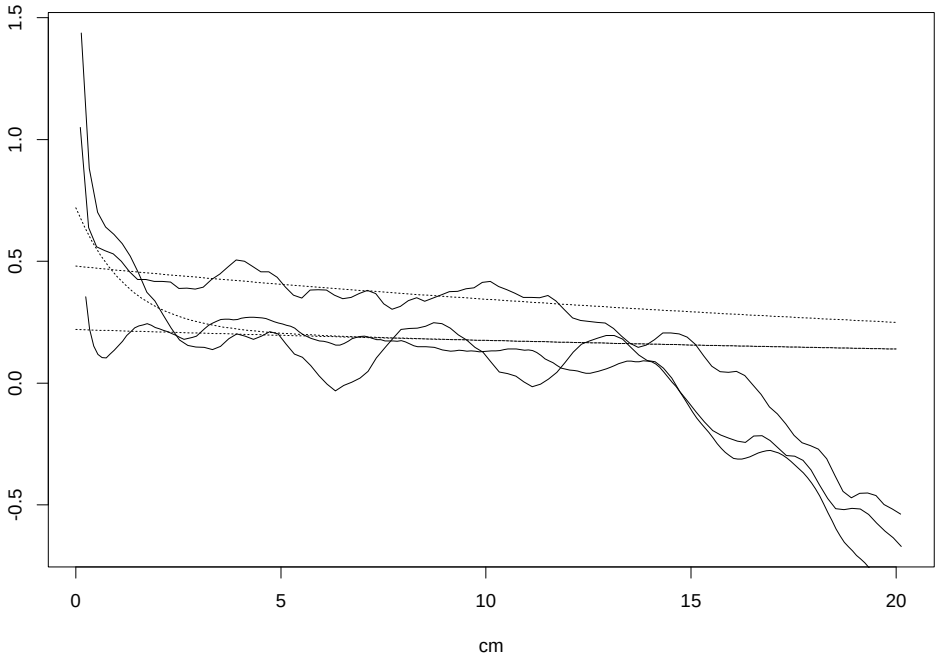

Figure 6: Non-parametric estimates $\bar{c}_{11}$, $\bar{c}_{12}$, $\bar{c}_{22}$ (solid lines) and parametric estimates $\hat{c}_{11}$, $\hat{c}_{12}$, $\hat{c}_{22}$ (dotted lines) for the covariance functions at the latest date. From above (at $0\,cm$) $\bar{c}_{11}$, $\bar{c}_{22}$, $\hat{c}_{11}$, $\hat{c}_{22}$, $\bar{c}_{12}$, $\hat{c}_{12}$.

4.3 Estimation of thinning probabilities

We do not make any parametric assumptions about the birth intensity measures γ_i introduced in Section 3.2. Instead we estimate the thinning probabilities in (9) by

$$\hat{p}_{ij} = [n_i(t_j) - n_i(t_{j-1})]/n_i(t_j), \quad i = 1, 2, \; j = 1, ..., 8,$$

where $n_i(t_j)$ denotes the number of points in $x_i(t_j)$ (we set $t_0 = 0$, so $n_i(0) = 0$). Note that $\hat{p}_{ij}$ is the ratio of the maximum likelihood estimates of $\theta_{ij1} = \nu_i(W)\gamma_i((t_{j-1}, t_j])$ and $\theta_{ij2} = \nu_i(W)\gamma_i([0, t_j])$; actually Y is an ancillary statistic for $\theta = (\theta_{ijk})$. The estimated thinning probabilities are shown in Figure 7.

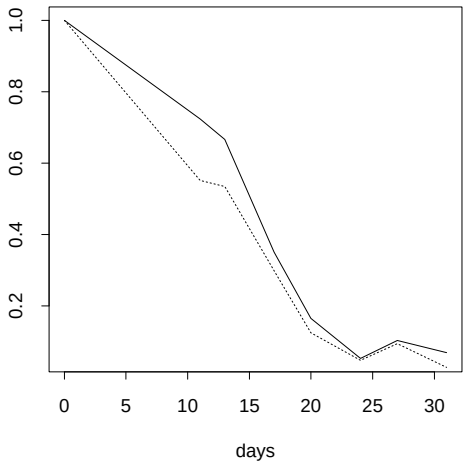

Figure 7: Estimated thinning probabilities for Veronica spp. (solid) and Trifolium spp. (dotted) at times $t_0 = 0$, $t_1 = 12$, $\ldots$, $t_7 = 32$ corresponding to the dates 26/4, 7/5, ..., 27/5 1994.

The estimated thinning probabilities for Veronica spp. and Trifolium spp. are very similar. This may indicate that the birth intensities mainly are controlled by external factors such as light intensity, temperature, and soil humidity, which are common to the two species.

4.4 Model validation

Several summary statistics are available for model checking. These are often defined so that their expectations have a clear interpretation in the homogeneous case (Veronica spp.), but plots of the summary statistics together with simulated envelopes can of course also be used in the inhomogeneous case (Trifolium spp.). For the latest date and for each species $i = 1, 2$, we consider below the empty space function F_i (see Diggle (1983)) and the third order characteristic z_i defined by (7). Moreover, in order to check the modeling of the time development, we compare the covariance functions at the eight dates.

We used the Kaplan-Meier estimator introduced in Baddeley and Gill (1997) for non-parametric estimation of F_i — the common estimator for F_i based on minus sampling (see e.g. Stoyan *et al.* (1995)) is not appropriate due to the relatively low number of points in each frame (see Figure 2). Figure 8 shows the estimated empty space functions for the two species. The plot for Veronica spp. shows a fine agreement between data and the estimated parametric model from Section 4.2, while the estimated F_2 (Trifolium spp.) is very close to the upper envelope after about $5\,cm$.

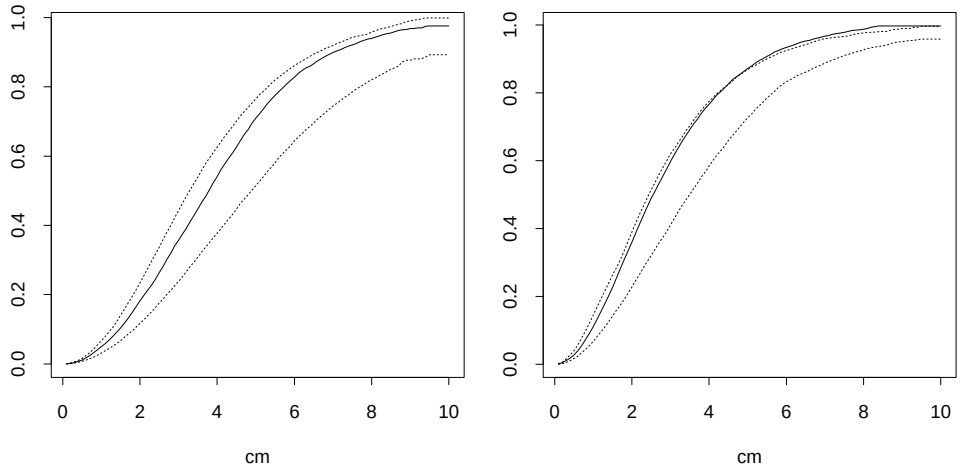

Figure 8: Non-parametric estimates of empty space functions for Veronica spp. (left) and Trifolium spp. (right) with envelopes from 19 simulations under the estimated parametric model at the latest date.

When simulating the log Gaussian Cox processes used to calculate the envelopes in Figure 8, we discovered that it was important to simulate on the entire $750 \times 500\,cm^2$ area rather than simulating nine independent Gaussian processes corresponding to the nine plots. A simulation study revealed that if nine independent Gaussian processes are simulated, then the covariance functions will be heavily under estimated, even if the correlation function decays relatively fast to zero (for example being 0.005 at $100\,cm$, which correspond to the minimum distance between plots).

Using (8) with ρ_i replaced by $\hat{\rho}_i^{(t_8)}$ from Section 4.1 and g_{ij} estimated by $\hat{g}_{ij} = \exp(\hat{c}_{ij})$ as obtained under the parametric model from Section 4.2, we got an estimate for the third order statistic z_i as shown in Figure 9. These plots give no reason to doubt the estimated parametric model at the latest date for both species. The nearly constant and relatively wide envelopes for Trifolium spp. is caused by the much larger value of $\hat{\beta}_2$ compared to $\hat{\beta}_1$, as β_2 increases, c_{22} becomes more heavytailed, whereby extra variation is introduced into X_2.

We now turn to the model validation for the time development. Recall that the pair correlation functions are unchanged over time under the bivariate log Gaussian Cox birth process model. Figures 10 and 11 show the estimated log pair correlation functions at the

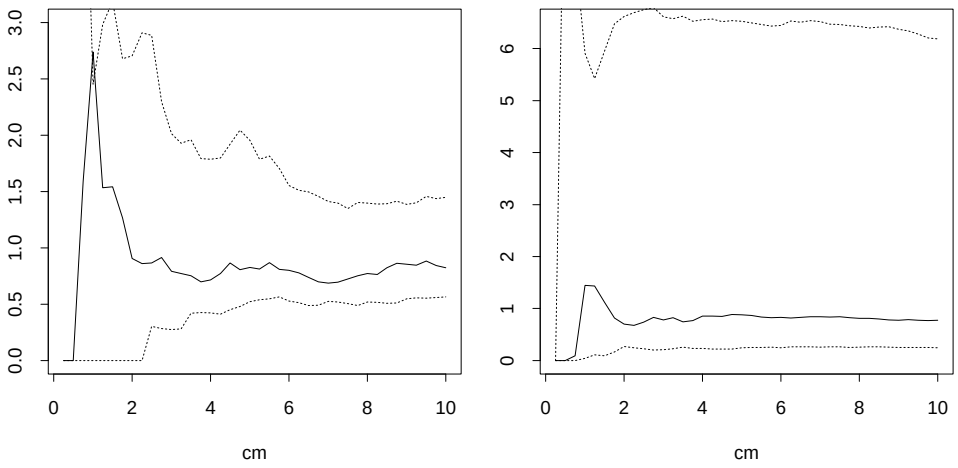

Figure 9: The third order statistic z_i for Veronica spp. (left) and Trifolium spp. (right) with envelopes from 19 simulations under the estimated parametric model at the latest date.

six latest times $t_3, ..., t_8$. The envelopes are based on 19 simulations of $X_i(t_8)$ under the estimated parametric model. If instead simulations corresponding to the lower intensities at the dates 13/5–27/5 had been used, wider envelopes would be expected. Figures 10 and 11 rise no doubt about the estimated bivariate log Gaussian Cox birth process model. The $\bar{c}_{ij}$ in Figures 10–11 are heavily biased downwards for distances about 15 cm and more, cf. Section 4.2, and the envelopes in Figure 11 are wider than in Figure 10 since $\hat{\beta}_2 >> \hat{\beta}_1$.

4.5 Prediction

At each of the eight dates $t \in \{t_1, \ldots, t_8\}$, it is possible to predict the underlying Gaussian process $Y^{(t)}(s) = (Y_1^{(t)}(s), Y_2^{(t)}(s))$ by simulating from the conditional distribution of $Y^{(t)}$ given $(X_1(t) \cap T, X_2(t) \cap T) = (x_1(t) \cap T, x_2(t) \cap T)$, where $Y_i^{(t)}(s) = (Y_i(s) + \log \gamma_i([0,t])$, $i = 1, 2$, and $T \subseteq W$ is a sub-window. For example, suppose we want to predict $Y^{(t)}$ on a rectangular region S. We may then extend the Metropolis adjusted Langevin algorithm (MALA) studied in Møller *et al.* (1998) from the univariate case to the bivariate case considered in the present paper (see also Roberts and Tweedie (1997)); the choice of $T \subseteq W$ may depend both on W, S, and the performance of the MALA as exemplified below. However, we found it much easier and faster to use the MALA for the univariate case directly by considering for each $i = 1, 2$, the Gaussian process $Y_i^{(t)}$ conditional on the data $X_i(t) \cap T = x_i(t) \cap T$ only. Then we have of course ignored the possible dependence between $Y_1^{(t)}$ and $Y_2^{(t)}$; indeed we have in any case ignored the information provided by the data concerning the other species of weeds as well.

We illustrate the method by letting T be the central plot and the plot right below it, and predict $Y_1^{(t_8)}$ on a 300 $cm \times 300\,cm$ square S containing T as indicated in Figure 12 (upper plot). The only reason for not considering e.g. $T = W$ and S a 500 $cm \times 750\,cm$ area

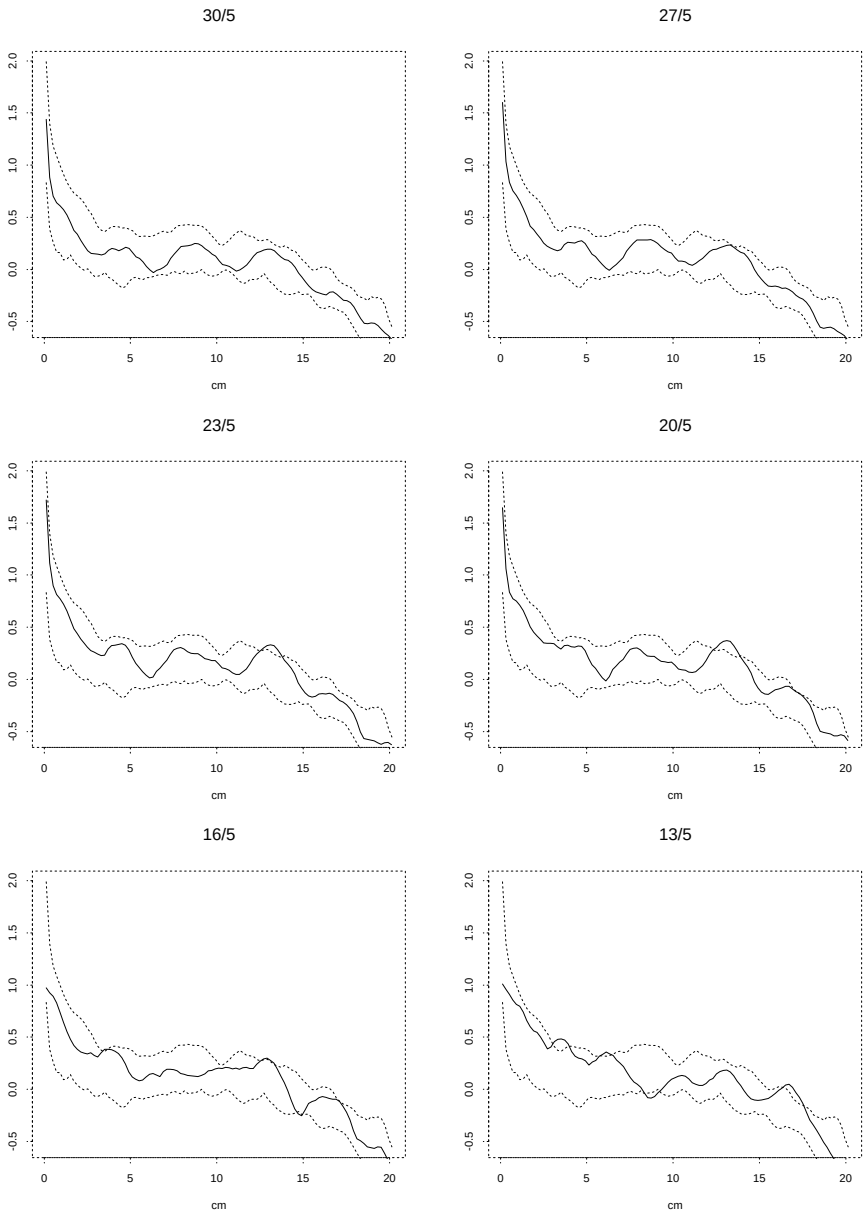

Figure 10: Non-parametric estimates of the log pair correlation functions for Veronica spp. (solid lines) at the six latest dates with envelopes based on 19 simulations under the estimated parametric model (dotted lines).

containing W is the amount of computer power required. The variance and correlation function of $Y_1^{(t_8)}$ are specified by the estimated parametric model from Section 4.2, and using the variance estimate together with the estimate $\hat{\rho}_1^{(t_8)}$ (see Section 4.1) we obtained from (5) an estimate $\hat{m}_1^{(t_8)} = -4.56$ of the mean. The Markov chain generated by the MALA appeared to reach equilibrium after 1000 iterations. For safety we used a burn-in of 10000 iterations before collecting a sample of size 40000. Thereby we obtained the Monte

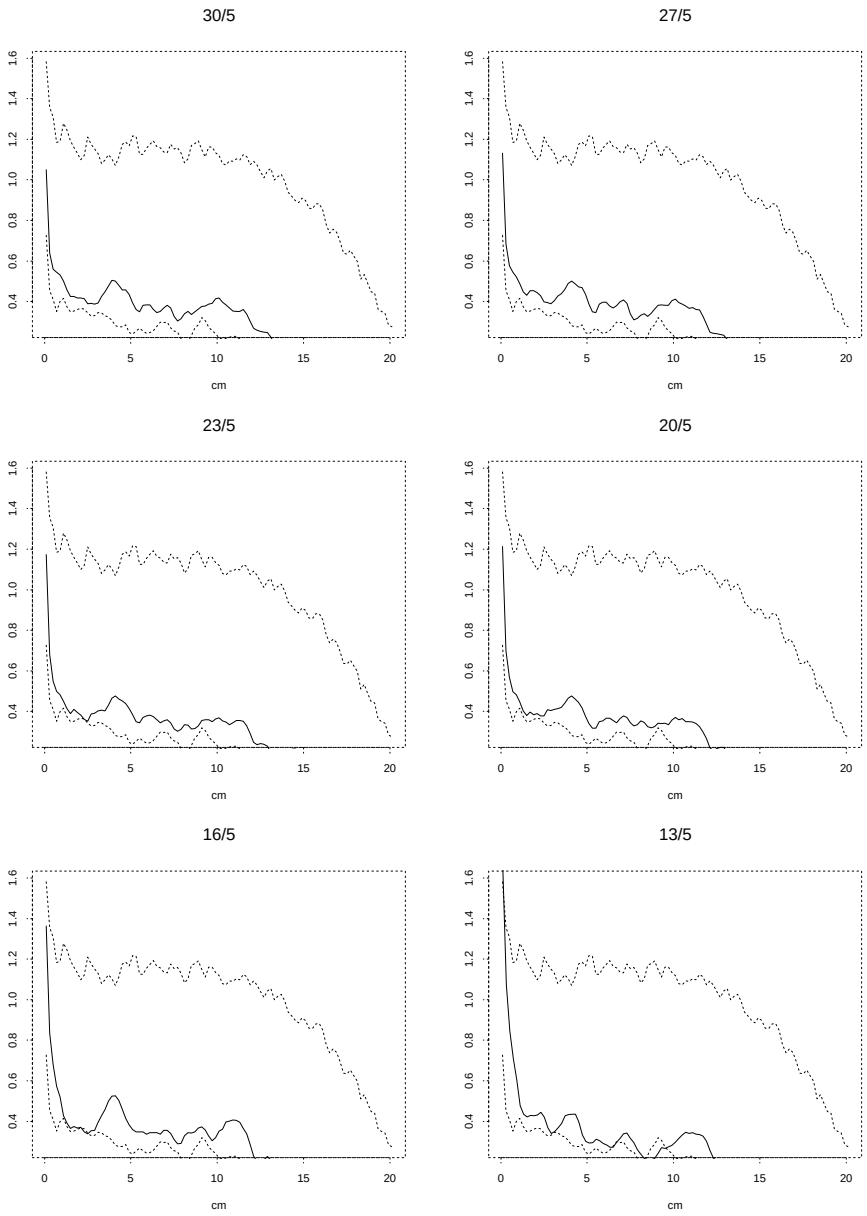

Figure 11: Non-parametric estimates of the log pair correlation functions for Trifolium spp. (solid lines) at the six latest dates with envelopes based on 19 simulations under the estimated parametric model (dotted lines).

Carlo estimate $\tilde{Y}_1^{(t_8)}(s)$ of $\mathrm{E}(Y_1^{(t_8)}(s)|x_1(t_8) \cap T)$, $s \in S$, as shown in Figure 12. Note how the high predicted intensity in the lower part of Figure 12 is in fine correspondence with the fact that the lower plot has the highest occurrence of Veronica spp. among all plots. Furthermore, in contrast to Figure 12 in Møller *et al.* (1998), it is not possible to recognize the original point pattern represented by high peaks of the predicted intensity. This is because our estimated covariance functions have much heavier tails than the ones

considered in Møller *et al.* (1998).

Incidentally, we mention that it is also possible to predict $X(t_8)$ on $S \setminus T$ from the conditional distribution of $X(t_8)$ given an estimate of the intensity surface on $S \setminus T$, for example as provided by $\exp(\tilde{Y}^{(t_8)})$ (with obvious notation).

5 Concluding remarks

We have also tried to fit another model which is not of the form (2), but with c_{11} a stable covariance function and c_{12} and c_{22} being exponential covariance functions. First, we estimated the parameters of the mean and covariance functions in a similar way as in Sections 4.1–4.2. Secondly, we used (1) to check and see that under the estimated model we have a well-defined covariance structure. Then we proceeded as in Sections 4.3– 4.4 and concluded that the present model fits the data just as well as the model (10) analyzed in Section 4 as regards the empty space functions and the third order characteristics. However, we prefer the model (10) because of its easy interpretation as discussed at the end of Section 3.2, and since the temporal development seemed better described by this model when considering the covariance functions at the different dates (as in Figures 10–11).

Finally, we notice that the log Gaussian Cox birth processes considered so far can be extended in many ways. For example, death of plants can be incorporated in a straight forward manner to obtain what we naturally would call a *bivariate log Gaussian Cox birth-and-death process*. The construction will be so that the pair correlation functions do not depend on the time, whereby the statistical analysis will be very similar to what we have discussed in Section 4. Another possibility is to incorporate covariate information in the mean functions and the birth intensity measures γ_i. Such covariates could e.g. be light intensity or temperature which is known to trigger germination or patches of humid soil which can effect the mean functions. If covariate information is used only for modeling the mean functions, the statistical analysis will again follow the same lines as in Section 4.

Acknowledgments

This research was supported by the European Union's network "Statistical and Computational Methods for the Analysis of Spatial Data. ERB-FMRX-CT96-0096", by The Danish Informatics Network in the Agricultural Sciences, and by MaPhySto – Centre for Mathematical Physics and Stochastics, funded by a grant from The Danish National Research Foundation. The paper is a part of the first authors Ph.D. thesis. We thank Mats Rudemo, Anne Randi Syversveen and Rasmus Plenge Waagepetersen for their helpful comments.

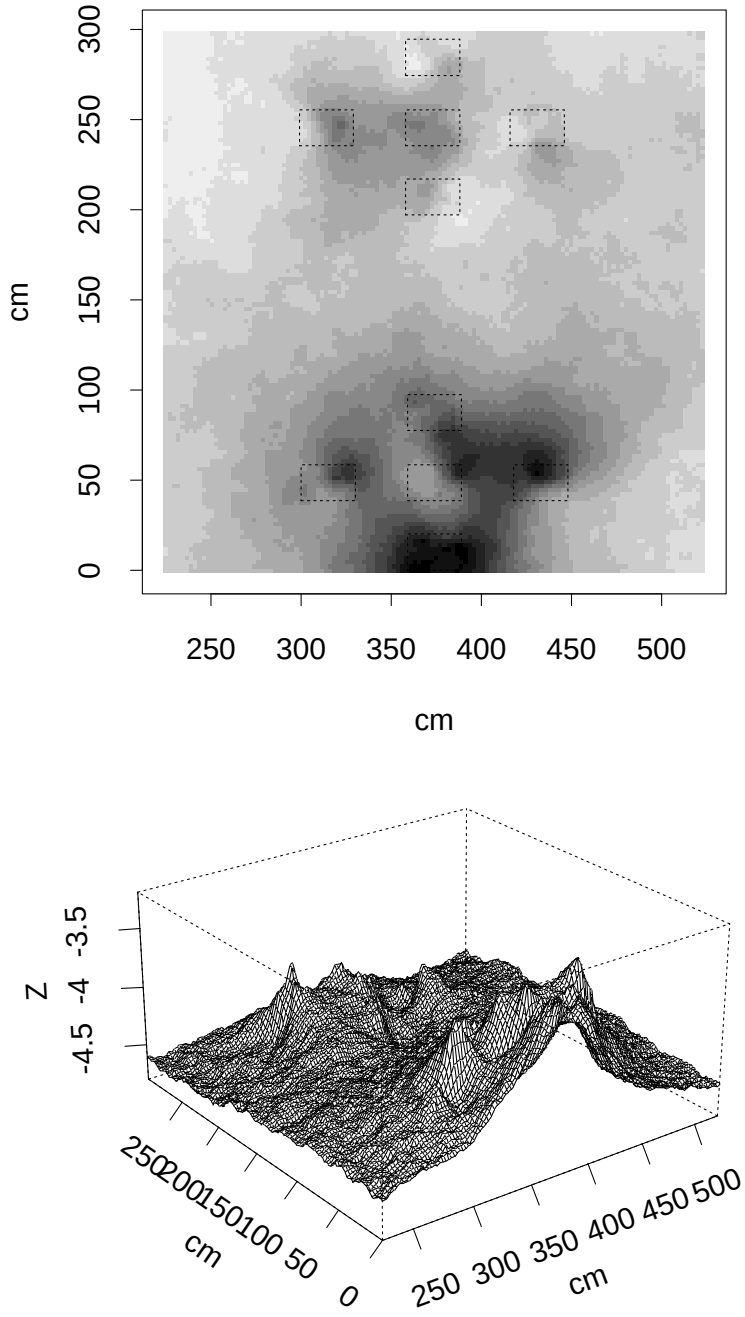

Figure 12: Prediction of $Y_1^{(t_8)}$ on the $300\,cm \times 300\,cm$ square S centered around the central plot and the plot right below it (as indicated by the dashed lines in the upper plot).

References

Anonymous (1993). *Understanding GIS: The ARC/INFO method.* Environmental Systems Research Institute.

Baddeley, A. and Gill, R.D. (1997) Kaplan-Meier estimators of distance distributions for spatial point processes. *The Annals of Statistics*, **25**, 263–292.

Baddeley, A., Møller, J. and Waagepetersen, R.P. (1998) Non and semi-parametric estimation of local interaction in inhomogeneous point patterns. (In preparation).

Cressie, N.A.C. (1991) *Statistics for spatial data.* New York: Wiley & Sons.

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

Haas, H. and Laursen, F. (1994) *Ukrudtskimplanter.* Copenhagen: Jordbrugsforlaget ApS.

Møller, J., Syversveen, A.R. and Waagepetersen, R.P. (1998) Log Gaussian Cox processes. *Scandinavian Journal of Statistics*, **25.** (To appear).

Ripley, B.D. (1977) Modelling spatial point patterns (with discussion). *J.R. Statist. Soc. B*, **41**, 368–374.

Roberts, G.O. and Tweedie, R.L. (1996) Exponential convergence of Langevin distributions and their discrete approximations. *Bernoulli*, **2**, 341–363.

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

Stoyan, D. and Stoyan, H. (1994) *Fractals, random shapes and point fields.* Chichester: Wiley & Sons.