Journal of Statistical Modelling: Theory and Applications
Vol. 1, No. 2, 2020, pp. 59-76
Yazd University Press 2020

JSMTA

Research Paper

Modeling spatial patterns and species associations in a Hyrcanian forest using a multivariate log-Gaussian Cox process

Abdollah Jalilian*[1], Amir Safari[2], Hormoz Sohrabi[3]
[1]Department of Statistics, Razi University, Kermanshah, Iran
[2]Natural resources and watershed management general office of Kermanshah, Iran
[3]Department of forestry, Tarbiat Modares University, Tehran, Iran

Received: October 26, 2019/ Revised: December 18, 2019 / Accepted: February 11, 2020

Abstract: This paper aims to conduct a model-based analysis of the spatial patterns of three tree species in a Hyrcanian forest and investigate their associations. There are many known and unknown mechanisms that influence the spatial forest structure and species associations. These complex and mainly unobservable mechanisms can be modeled by hidden Gaussian random fields and log-Gaussian Cox process models are appropriate for linking them to the spatial patterns of tree species. We consider a multivariate log-Gaussian Cox process model that can take into account the overall mixed effects of all influential factors on spatial distributions of species and quantify species associations in terms of some parameters. This construction provides a suitable framework for modeling and analyzing spatial patterns of several species. We also discuss modeling tree diameters, parameter estimation and goodness of fit methods and apply them to the data. Results from fitting the model to the data show that there is a significant negative association between two light-demanding species. Finally, a Gamma intensity-dependent model is considered to model spatial correlation in tree diameters of one of the species.

Keywords: Cross-pair correlation function; Hidden Gaussian random field; Intensity-dependent marking model; Mark variogram.
Mathematics Subject Classification (2010): 62H11, 62M30, 60G55, 60G15.

1 Introduction

Statistical analysis of spatial point patterns of trees can extract information about forest dynamics and mechanisms (Stoyan and Penttinen, 2000; Comas and Mateu, 2007).

*Corresponding author: jalilian@razi.ac.ir

A model-based approach in statistical analysis of spatial patterns of tree locations tries to construct and develop parametric spatial point process models for the observed data and explain the underlying process that leads the spatial forest structure (see e.g. Waagepetersen, 2007; Picard et al., 2009; Grabarnik and Särkkä, 2009; Illian et al., 2013). This approach reveals identifiable random and non-random spatial structures in the data and describes the behaviour of the whole system in terms of a set of parameters and functional relationships (Law et al., 2009; Illian et al., 2009; Jalilian et al., 2012). In fact, a suitable multivariate point process model can provide a mathematical explanation for the complex unobservable underlying mechanisms that create the spatial forest structure and cause species associations (Comas and Mateu, 2007; Picard et al., 2009; Grabarnik and Särkkä, 2009). In addition, such a model can be used to assess relevant biological theories and hypotheses (Wiegand et al., 2007; Illian and Burslem, 2007). For these reasons, spatial point process models recently gained much attention in forestry and plant ecology.

Environmental variables and direct plant-to-plant interactions are known to be the most influential factors in distribution of trees in a given forest stand and hence they must be included in any model for the spatial forest structure (Wiegand et al., 2007; Illian and Burslem, 2007; Law et al., 2009). However, modeling the impact of environment on distribution of trees and measuring interactions between trees of the same (intra-specific interactions) and different (inter-specific interactions) species are not straightforward. Soil and topographical conditions, dispersal mechanisms and limitations, infectious diseases, competition or facilitation among species and other known and unknown factors can change the environment and direct interactions among trees in many complex ways and gradually form the spatial forest structure and species associations (Luo et al., 2012).

Nevertheless, spatial Cox processes provide a framework for modeling the overall effect of a complex set of factors in terms of a hidden random field (Møller and Waagepetersen, 2007). Among Cox process models, the log-Gaussian Cox processes, introduced by Møller et al. (1998), are mathematically flexible models that can be used for modeling spatial patterns of single or several tree species. A log-Gaussian Cox process is a special case of latent Gaussian models and the integrated nested Laplace approximation (INLA) method (Rue et al., 2009) can be used to fit log-Gaussian Cox processes with complex dependence structures at small and larger spatial scales (Illian et al., 2012, 2013). Although flexible and computationally efficient, the log-Gaussian Cox processes considered in Illian et al. (2012) and Illian et al. (2013) do not have pair correlation functions with closed analytical forms. Instead, as suggested by Møller et al. (1998) and Waagepetersen et al. (2016), we consider parametric multivariate log-Gaussian Cox models that are constructed by linear combinations of independent hidden Gaussian random fields. This structure provides models with tractable auto- and cross-pair correlation functions and allows for parametric explanation of associations within and among species using the correlation functions of the hidden Gaussian random fields (Møller et al., 1998; Brix and Møller, 2001).

In the present work, we use such a multivariate log-Gaussian Cox process to simultaneously model spatial patterns of three tree species in a forest park in the north of Iran. Our main objective is to assess and find significant associations among the species. We also investigate spatial correlations between diameters of trees of each species and

propose models for tree diameters. The remainder of the paper is organized as follows: Sections 2 and 3 introduce the study area, available data and modelling approach. Section 4 gives some background on log-Gaussian Cox processes and introduces the considered model. Modelling the diameter of trees is discussed in Section 5. The results from fitting the model to the data are presented in Section 6. The paper concludes in Section 7 with some discussions on limitations and advantages of the fitted model.

2 Data

The broad-leaved deciduous Hyrcanian (Caspian) forests cover an area of about 1.9 million hectares between the northern slopes of the Alborz Mountains and the southern shores of the Caspian Sea in North of Iran (Ramezani et al., 2013). These mixed forests are the most important natural forests in Iran and different aspects of them have been studied in the past two decades, see e.g. Sheykholislami et al. (2009); Esmailzadeh et al. (2011); Akhavan et al. (2012); Salehi et al. (2013); Ramezani et al. (2013).

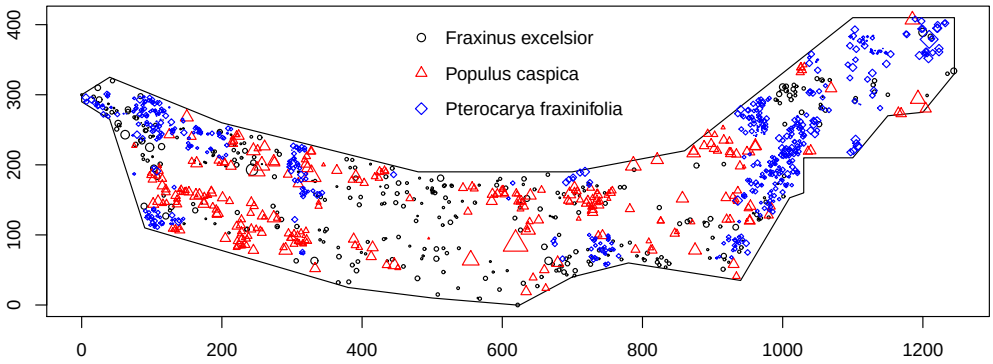

Figure 1: Spatial point pattern of locations of *Fraxinus excelsior*, *Populus caspica* and *Pterocarya fraxinifolia* trees in a polygonal plot in the Nour forest park, Mazandaran, Iran. The sizes of representing points are proportional to the diameter at breast height (DBH) of trees.

The study area is located in the South-East of the Nour forest park, forest reserve of Populus caspica, Mazandaran, Iran (36°34′N, 51°50′E). It is situated in the central Hyrcanian forests where the elevation varies between 20 m and 43 m from the sea level and the slope gradient is 3-5%, which suggests a rather topographically homogeneous landscape. The area has a humid subtropical climate. The annual mean precipitation and temperature at the nearest meteorological station (Noshahr, Iran) are 1310 mm and 16.1°C, respectively.

In order to investigate spatial distributions and coexistence of species, a 18-hectare forest plot (see 1) inside the study area was established and surveyed in 2011. The irregular polygonal plot was chosen due to lack of human disturbance and ease of operation. In the survey, Cartesian (rectangular) coordinates for all trees were determined using distance and azimuth from a starting point. In addition to spatial location, diameter at breast height (DBH) and species of each tree were recorded. The study is

focused on three species: *Pterocarya fraxinifolia* (Lam.) Spach. (Caucasian wingnut), an Arcto-Tertiary relict species (Ramezani et al., 2013), *Populus caspica* Born, an endemic and endangered Popolar species (Salehi et al., 2013), and *Fraxinus excelsior* L. (common ash), a major charismatic species in Asia and Europe (Pautasso et al., 2013). 1 shows locations of trees of these species in the plot and 1 summaries their observed characteristics. For these species, we focus on analysis and modeling species associations and distribution and spatial correlation of tree diameters.

Table 1: The number of observed trees, N_i, estimated intensity (tree per ha), $\widehat{\rho}_i$, and average, $\overline{m}_i$, variance, $\widehat{\sigma}^2_{m_i}$, and maximum, $\max_j m_{ij}$, of DBH (cm) and mean basal area (m^2) for each species.

Species	N_i	$\widehat{\rho}_i$	$\overline{m}_i$	$\widehat{\sigma}^2_{m_i}$	$\max_j m_{ij}$	basal area
Fraxinus excelsior	368	20.4	38.55	298.8	136	7.8×10^{-3}
Populus Caspica	286	15.8	72.95	694.5	220	26.2×10^{-3}
Pterocarya Fraxinifolia	613	34.1	34.16	269.2	153	6.3×10^{-3}

3 Preliminary analysis

The first step in analyzing spatial patterns of several species in a forest stand is to use empirical summary statistics, such as intensities, nearest neighbor functions, Ripley's K-function and the pair correlation function, and essentially describe features of the observed data (see e.g. Hou et al., 2004; Hao et al., 2007; Martínez et al., 2010). Let $W \subset \mathbb{R}^2$ denote the polygonal study plot (observation window) in 1 and $\boldsymbol{u}_{ij} = (u_{ij1}, u_{ij2})$ and $m(\boldsymbol{u}_{ij})$ be, respectively, the location and DBH of the j-th tree, $j = 1, \ldots, N_i$, from the i-th species, $i = 1, 2, 3$. For $i = 1, 2, 3$, we assume that X_i is the underlying spatial point process that governs the spatial distribution of tree locations $\boldsymbol{u}_{ij}$, $j = 1, \ldots, N_i$, over W, see Comas and Mateu (2007) and Law et al. (2009). Thus X_i is the stochastic mechanism that has generated the observed spatial point pattern $\boldsymbol{x}_i = \{\boldsymbol{u}_{i1}, \ldots, \boldsymbol{u}_{iN_i}\}$ of the i-th species. Since the study area is topographically homogeneous and no data on soil conditions of the study area were collected, we let X_i to be a stationary point process with constant intensity $\rho_i > 0$ which controls the abundance of X_i. Estimates of ρ_i's are given in 1.

The overall pairwise intra- and inter-specific associations can be quantified and analyzed using second-order point process characteristics such as Ripley's K-function and the pair correlation function (Stoyan and Penttinen, 2000; Comas and Mateu, 2007). Due to the cumulative nature of K-function, the pair correlation function is preferable (Wiegand and Moloney, 2004; Illian et al., 2008). The auto- (or univariate) pair correlation function $g_{ii}(r)$ of the i-th species is a suitable tool in assessing and modeling pairwise correlations between points of X_i and the cross-pair correlation function (also known as bivariate or partial pair correlation function) $g_{ij}(r)$ of the i-th and j-th species is useful in assessing pairwise correlations between points of X_i and X_j (Comas and Mateu, 2007; Illian et al., 2008; Law et al., 2009).

Plots of the empirical (nonparametric) estimates (see e.g. Møller and Waagepetersen, 2004, Sections 4.3.5 and 4.4.3) of auto- and cross-pair correlation functions for *Fraxinus excelsior*, *Populus Caspica* and *Pterocarya Fraxinifolia* species are shown in 2. The

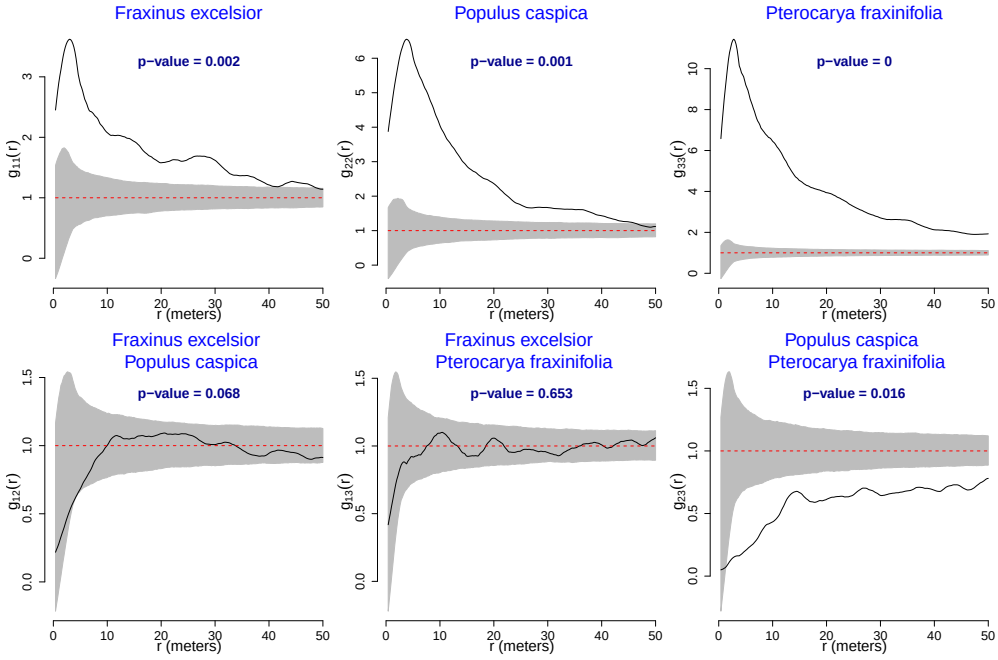

Figure 2: Empirical estimates of auto- (upper panels) and cross- (lower panels) pair correlation functions along with 95% simulation envelopes (grey areas) under the CSR model and p-values of corresponding deviation tests.

plots also show 95% pointwise simulation envelopes obtained under the null model of complete spatial randomness (CSR). The simulation envelopes are useful to investigate departures of $\widehat{g}_{ij}(r)$ values from the reference value $g_{ij}(r) \equiv 1$ (Wiegand and Moloney, 2004). The agreement between the empirical $\widehat{g}_{ij}(r)$ and their corresponding theoretical values under the null model of CSR, $g_{ij}(r) \equiv 1$, is also checked using deviation tests with studentized scaling integral deviation measure (see Myllymäki et al., 2015)

$$\int_0^{50} \left[\frac{\widehat{g}_{ij}(r) - 1}{\sqrt{\mathbb{V}\mathrm{ar}\widehat{g}_{ij}(r)}} \right]^2 \mathrm{d}r,$$

where $\mathbb{V}\mathrm{ar}\widehat{g}_{ij}(r)$ is the variance of $\widehat{g}_{ij}(r)$ under the null model of CSR and is estimated using 5000 simulations. The p-values of the deviation tests are presented at the top of the plots in 2.

The p-values and plots of the auto-pair correlation functions show that $\widehat{g}_{ii}(r)$, $i = 1, 2, 3$, is significantly larger than one. This, in turn, indicates that there are positive intra-specific correlations, or aggregation, within trees of each species, although the range and strength of clustering is different for each species. This fact can also be seen from clustered patterns of the three species in 1 and it is a common pattern for species in naturally regenerated forests (see Hao et al., 2007; Zhang et al., 2010; Lee et al., 2012; Luo et al., 2012). The cross-pair correlation function for species *Populus*

Caspica and *Pterocarya Fraxinifolia* is constantly below the lower envelope which means that $\widehat{g}_{23}(r)$ is significantly (with p-value=0.001) less than one, and hence there are negative inter-specific association, or competition, between *Populus Caspica* trees and *Pterocarya Fraxinifolia* trees when they are less than 50 meters apart. The cross-pair correlation functions for *Fraxinus excelsior* and *Populus Caspica* and *Fraxinus excelsior* and *Pterocarya Fraxinifolia* lie inside the envelopes which suggests no significant (with p-values > 0.05) inter-specific correlations.

4 Modelling tree locations

The null model of CSR assumes that there are no intra- and inter-specific associations (see Illian et al., 2008, Section 7.4.1). This model implies that X_1, X_2 and X_3 are independent stationary Poisson processes which is clearly not a suitable model for the clustered data in 1. Instead of Poisson processes, Cox processes can be employed to model clustering behaviour in spatial patterns of species (Møller and Waagepetersen, 2004, 2007). Thus we assume that the spatial distribution of trees of each species follows a Cox process. Because of possible inter-specific associations, these Cox processes are not necessarily independent. A Cox process model assumes that for each species there is an unobservable random field (surface) that creates the spatial pattern of the species and it is the variations and associations of these random fields that generate the spatial forest structure and control species coexistence mechanisms. The overall mixed impact of ecological processes such as reproduction, growth, mortality, interactions, dispersal, resource use, gap formation and understory development (see Wiegand et al., 2007; Zhang et al., 2010) and other influential factors can be represented by the hidden random fields of the Cox processes.

4.1 Cox model

If X_i, $i = 1, 2, 3$, be a Cox process then there is a hidden (unobservable) random field $\Lambda_i(\boldsymbol{u})$ on W and X_i is a Poisson process with the random intensity function $\Lambda_i(\boldsymbol{u})$ (Møller and Waagepetersen, 2004, 2007). The random field $\Lambda_i(\boldsymbol{u})$ represents the overall mixed impact of all influential factors in distribution of X_i, including unobserved environmental variabilities, seed dispersal and intra- and inter-specific associations. 3 illustrates the effect of $\Lambda_i(\boldsymbol{u})$ on points of X_i.

For a Cox process X_i with stationary and isotropic $\Lambda_i(\boldsymbol{u})$, the intensity is given by the mean of $\Lambda_i(\boldsymbol{u})$; i.e. $\rho_i = \mathbb{E}\Lambda_i(\boldsymbol{u})$, and

$$g_{ij}(r) = 1 - \frac{\mathbb{C}\mathrm{ov}\big[\Lambda_i(\boldsymbol{u}), \Lambda_j(\boldsymbol{v})\big]}{\rho_i \rho_j}, \quad i, j = 1, 2, 3, \tag{1}$$

where $r = \|\boldsymbol{u} - \boldsymbol{v}\|$ is the distance between $\boldsymbol{u}, \boldsymbol{v} \in W$ and $\mathbb{C}\mathrm{ov}\big[\Lambda_i(\boldsymbol{u}), \Lambda_j(\boldsymbol{v})\big]$ is the covariance of $\Lambda_i(\boldsymbol{u})$ and $\Lambda_j(\boldsymbol{v})$. This means that the dependence structure between and within random fields $\Lambda_1(\boldsymbol{u})$, $\Lambda_2(\boldsymbol{u})$ and $\Lambda_3(\boldsymbol{u})$ determines intra- and inter-specific associations and hence define the structure of the auto- and cross-pair correlation functions. Therefore, to explain species associations, it suffices to determine and model the spatial variations and associations of the hidden random fields $\Lambda_1(\boldsymbol{u}), \Lambda_2(\boldsymbol{u})$ and $\Lambda_3(\boldsymbol{u})$.

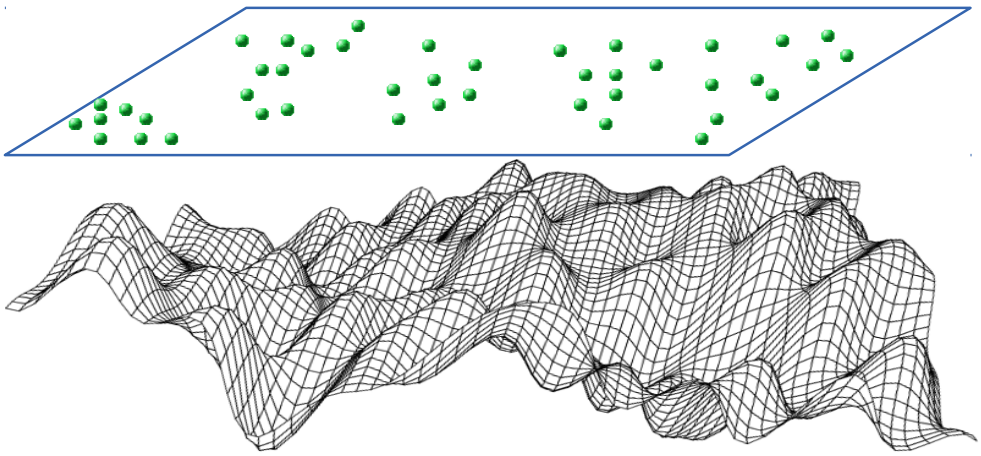

Figure 3: The hidden random field (surface) $\Lambda_i(\boldsymbol{u})$ controls the abundance and distribution of points of the Cox process X_i.

4.2 Log-Gaussian Cox model

If $(\log\Lambda_1(\boldsymbol{u}), \log\Lambda_2(\boldsymbol{u}), \log\Lambda_3(\boldsymbol{u}))$ is a multivariate Gaussian random field, then X_1, X_2 and X_3 are called log-Gaussian Cox processes (Møller et al., 1998; Brix and Møller, 2001). To obtain a valid multivariate log-Gaussian Cox process model for $\boldsymbol{X} = (X_1, X_2, X_3)$, an approach is to use a log-Gaussian Cox process for the spatial pattern of all tree locations, regardless of species, as a plant establishment process and then, given the spatial pattern, fit a multinomial regression model to the species with the spatial coordinates as covariates (Illian et al., 2013). The INLA method (Rue et al., 2009) can be used to fit such models to the data using the R package R-INLA, see Illian et al. (2012). However, this approach does not provide explicit expressions for the auto- and cross-pair correlation functions of species which are convenient tools in forestry and ecological studies.

As Møller et al. (1998) and Waagepetersen et al. (2016) suggest, another approach is to let $\log\Lambda_i(\boldsymbol{u})$ be a linear combination of independent Gaussian random fields. Let $Z_1(\boldsymbol{u})$, $Z_2(\boldsymbol{u})$ and $Z_3(\boldsymbol{u})$ be independent zero-mean Gaussian random fields with covariance functions

$$R_i(r;\omega_i) = \mathbb{C}\mathrm{ov}(Z(\boldsymbol{u}), Z(\boldsymbol{v})) = \exp\left(-\frac{\|\boldsymbol{u}-\boldsymbol{v}\|}{\omega_i}\right) \quad , \quad i = 1,2,3,$$

where $r = \|\boldsymbol{u}-\boldsymbol{v}\|$ is the distance between $\boldsymbol{u}, \boldsymbol{v} \in W$. This implies that $Z_i(\boldsymbol{u})$ is a stationary and isotropic Gaussian process with unit variance and exponential correlation function $R_i(r;\omega_i)$ with the scale parameter $\omega_i > 0$. We assume that X_1, X_2 and X_3 are Cox processes driven by random fields

$$\Lambda_i(\boldsymbol{u}) = \exp\left(\beta_i + \sum_{k=1}^{3}\alpha_{ik}Z_k(\boldsymbol{u})\right), \quad i = 1,2,3,$$

where

$$\boldsymbol{\alpha} = \begin{bmatrix} \alpha_{11} & \alpha_{12} & \alpha_{13} \\ \alpha_{21} & \alpha_{22} & \alpha_{23} \\ \alpha_{31} & \alpha_{22} & \alpha_{33} \end{bmatrix}$$

is the matrix of coefficients. The random fields $Z_1(\boldsymbol{u})$, $Z_2(\boldsymbol{u})$ and $Z_3(\boldsymbol{u})$ can be thought of as three independent (orthogonal) unknown linear combinations of all sources of variations in spatial distributions of species, which summarize all influential factors at different spatial scales; i.e. ω_1, ω_2 and ω_3. The coefficient α_{ik} represents the contribution of the Gaussian random field $Z_k(\boldsymbol{u})$ in the spatial distribution of trees of the i-th species. 4 illustrates how under this model the multivariate point process $\boldsymbol{X} = (X_1, X_2, X_3)$ is constructed by the hidden Gaussian random fields $Z_1(\boldsymbol{u})$, $Z_2(\boldsymbol{u})$ and $Z_3(\boldsymbol{u})$ and the coefficient matrix $\boldsymbol{\alpha}$. Here $\beta_i = \log \rho_i - \sum_{k=1}^{3} \alpha_{ik}^2$ is a constant term related to the intensity of X_i, ρ_i.

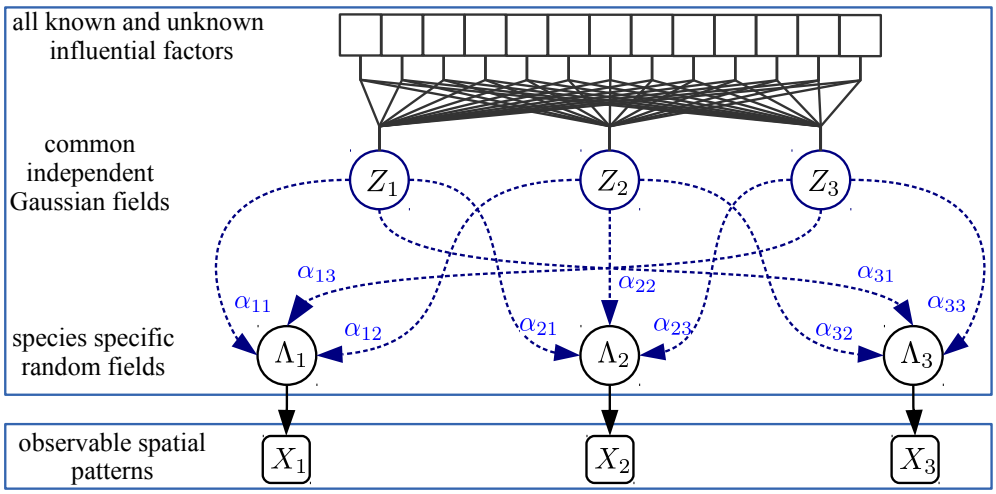

Figure 4: A schematic diagram of relations among the observable point processes X_1, X_2 and X_3 and the hidden Gaussian random fields $Z_1(\boldsymbol{u})$, $Z_2(\boldsymbol{u})$ and $Z_3(\boldsymbol{u})$.

According to this model, X_1, X_2 and X_3 are log-Gaussian Cox processes with (Møller et al., 1998)

$$g_{ij}(r; \boldsymbol{\alpha}, \boldsymbol{\omega}) = \exp\left(\sum_{k=1}^{3} \alpha_{ik}\alpha_{jk} R_k(r; \omega_k)\right), \tag{2}$$

where $\boldsymbol{\omega} = (\omega_1, \omega_2, \omega_3)$ is the vector of scale parameters. Particularly, the auto-pair correlation functions

$$g_{ii}(r; \boldsymbol{\alpha}, \boldsymbol{\omega}) = \exp\left(\sum_{k=1}^{3} \alpha_{ik}^2 R_k(r; \omega_k)\right) > 1, \tag{3}$$

for all $r \geq 0$, which means intra-specific correlations are always positive and hence X_i, $i = 1, 2, 3$, produces clustered point patterns. For $i \neq j$, the cross-pair correlation function $g_{ij}(r; \boldsymbol{\alpha}, \boldsymbol{\omega}) < 1$ if $\sum_{k=1}^{3} \alpha_{ik}\alpha_{jk} < 0$ and $g_{ij}(r; \boldsymbol{\alpha}, \boldsymbol{\omega}) > 1$ if $\sum_{k=1}^{3} \alpha_{ik}\alpha_{jk} > 0$.

Thus, the sign and magnitude of $\sum_{k=1}^{3}\alpha_{ik}\alpha_{jk}$, which is the ij entry of the matrix $\boldsymbol{\Sigma} = \boldsymbol{\alpha}\boldsymbol{\alpha}^{\mathsf{T}}$, determines the type and strength of associations between the i-th and j-th species. Furthermore, the parameters ω_i along with α_{ik} and α_{jk}, $k = 1, 2, 3$, control the correlation range between species i and j. Therefore, the above considered log-Gaussian Cox model formulates the type, strength and range of intra- and inter-specific associations in terms of the parameters $\boldsymbol{\alpha}$ and $\boldsymbol{\omega}$. The model imposes positive intra-specific associations, or clustering, for each species but allows positive, negative or even neutral inter-specific associations between pairs of species. This provides a flexibility in modeling the possible associations among species and makes the model a very suitable model for the data.

In general, the number of hidden Gaussian random fields $Z_i(\boldsymbol{u})$'s is unknown and it may not be equal to the number of species. In fact, when the number of species is large, considering smaller number of hidden Gaussian random fields leads to a more parsimonious model (Waagepetersen et al., 2016). As discussed in Waagepetersen et al. (2016), for large number of species, the number of hidden Gaussian random fields can be estimated using, for example, a cross-validation criterion. However, since there are only three species in the considered forest plot, here we let the number of hidden Gaussian random fields to be equal to the number of species.

5 Modelling tree DBHs

Besides tree locations, we need to analyze tree diameters. 5 shows the DBH histograms of species in 5 cm classes. The histograms suggest that the DBH distributions of species are not symmetric (normal or bell shape) but skewed to the right. We use the conditional Gamma intensity-dependent marking suggested by Myllymäki and Penttinen (2009) to model the DBH of the species. Given the spatial tree locations $\boldsymbol{x}_i = \{\boldsymbol{u}_{i1}, \ldots, \boldsymbol{u}_{iN_i}\}$ and random intensity function $\Lambda_i(\boldsymbol{u})$, the DBHs $m_i(\boldsymbol{u}_{i1}), \ldots, m_i(\boldsymbol{u}_{iN_i})$ are assumed to be independent and each $m_i(\boldsymbol{u}_{ij})$ is distributed according to a Gamma distribution with shape parameter $\nu_i > 0$ and scale parameter $1/\left[\tau_i + \eta_i\Lambda_i(\boldsymbol{u}_{ij})\right] > 0$. Thus given $\Lambda_i(\boldsymbol{u}_{i1}), \ldots, \Lambda_i(\boldsymbol{u}_{iN_i})$, the model for $m_i(\boldsymbol{u}_{i1}), \ldots, m_i(\boldsymbol{u}_{iN_i})$ can be fitted as a generalized linear model (GLM) with Gamma distributed errors and inverse link function

$$\frac{1}{\mathbb{E}[m_i(\boldsymbol{u}_{ij})|\Lambda_i(\boldsymbol{u}_{ij}), X_i = \boldsymbol{x}_i]} = \frac{\tau_i}{\nu_i} + \frac{\eta_i}{\nu_i}\Lambda_i(\boldsymbol{u}_{ij}). \tag{4}$$

If $\eta_i = 0$, then $\Lambda_i(\boldsymbol{u}_{ij})$ does not affect $m_i(\boldsymbol{u}_{ij})$ and hence conditional on $\boldsymbol{x}_i$, $m_i(\boldsymbol{u}_{i1}), \ldots, m_i(\boldsymbol{u}_{iN_i})$ are independent and identically distributed Gamma random variables with shape parameter ν_i and scale parameter $1/\tau_i$. But $\eta_i > 0$ implies that $m_i(\boldsymbol{u}_{ij})$ is reciprocally related to the random intensity $\Lambda_i(\boldsymbol{u}_{ij})$ and hence $m_i(\boldsymbol{u}_{i1}), \ldots, m_i(\boldsymbol{u}_{iN_i})$ are, unconditionally, spatially correlated due to interactions with the conspecific neighboring trees through the random field $\Lambda_i(\boldsymbol{u})$.

The mark variogram is a useful tool in studying spatial correlation between the sizes of trees at close proximity (Stoyan and Penttinen, 2000; Pommerening and Särkkä, 2013). When $m_i(\boldsymbol{u}_{i1}), \ldots, m_i(\boldsymbol{u}_{iN_i})$ are uncorrelated ($\eta_i = 0$), then $\gamma_{m_i}(r) \equiv \sigma^2_{m_i}$ where $\sigma^2_{m_i} = \nu_i/\tau_i^2$ is the variance of $m_i(\boldsymbol{u}_{i1}), \ldots, m_i(\boldsymbol{u}_{iN_i})$ (see Illian et al., 2008, Section 5.3.3). In order to check the spatial correlation of $m_i(\boldsymbol{u}_{i1}), \ldots, m_i(\boldsymbol{u}_{iN_i})$, the empirical mark variograms of the three species and their corresponding 95% pointwise sim-

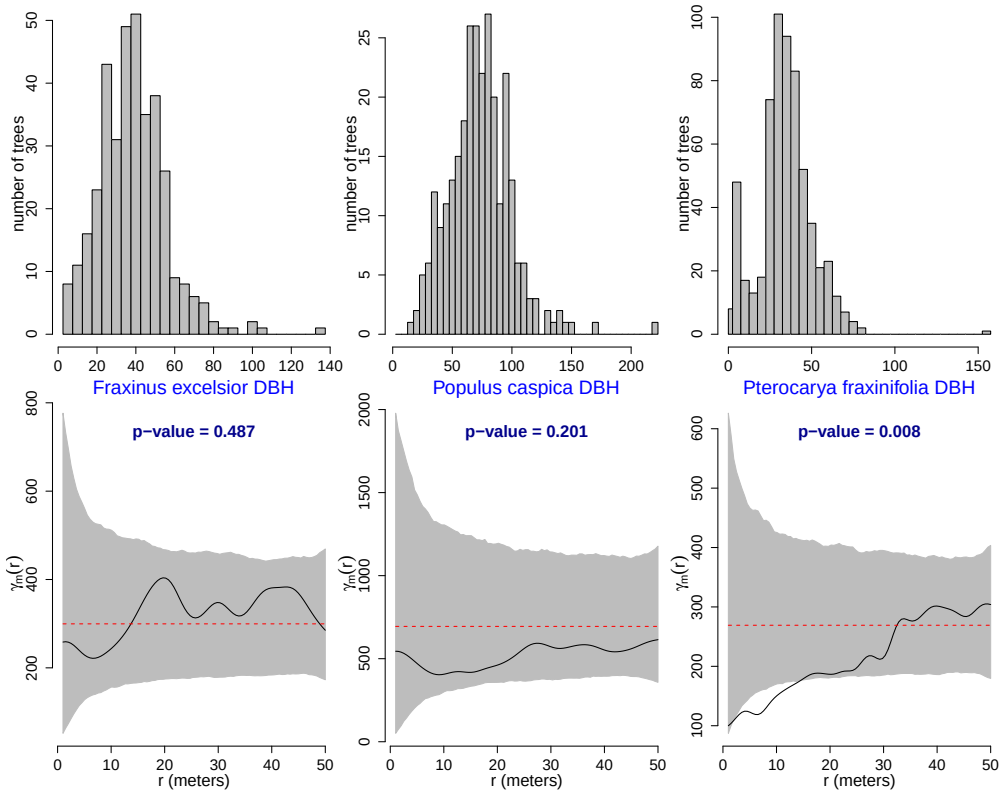

Figure 5: Histograms (upper panels) and mark variograms (lower panels), their 95% pointwise simulation envelopes (grey area) and p-values of corresponding deviation tests for DBH of each species.

ulation envelops under the assumption of uncorrelated DBHs are shown in lower panels of 5. Also, the agreement between the empirical variograms $\widehat{\gamma}_{m_i}(r)$ and their corresponding theoretical values under the uncorrelated DBH assumption, $\gamma_{m_i}(r) = \sigma^2_{m_i}$, are checked using deviation tests with studentized scaling integral deviation measure (see Myllymäki et al., 2015)

$$\int_0^{50} \left[\frac{\widehat{\gamma}_{m_i}(r) - \mathbb{E}\widehat{\gamma}_{m_i}(r)}{\sqrt{\mathbb{V}\mathrm{ar}\widehat{\gamma}_{m_i}(r)}} \right]^2 \mathrm{d}r,$$

where $\mathbb{E}\widehat{\gamma}_{m_i}(r)$ and $\mathbb{V}\mathrm{ar}\widehat{\gamma}_{m_i}(r)$ are the mean and variance of $\widehat{\gamma}_{m_i}(r)$ under the assumption of uncorrelated DBHs and are estimated using 5000 simulations. The p-values of the deviation tests are presented in 5. The plots and p-values of the deviation tests indicate that the DBH of *Fraxinus excelsior* and *Populus Caspica* trees can be considered spatially uncorrelated ($\eta_1 = \eta_2 = 0$) because their empirical mark variograms lie inside the envelopes. The mark variogram of *Pterocarya Fraxinifolia* is below the lower envelope for $0 < r < 25$ and hence the DBH of trees have significant (with p-value=0.008) positive spatial correlation (see Pommerening and Särkkä, 2013). This

means that the DBH of *Pterocarya Fraxinifolia* trees depend on the DBH of their conspecific neighboring trees ($\eta_3 > 0$).

6 Results

6.1 Estimates of model parameters

The model parameters $\boldsymbol{\alpha}$ and $\boldsymbol{\omega}$ can be estimated by minimizing the contrast function (see Møller and Waagepetersen, 2004, Section 10.1)

$$M(\boldsymbol{\alpha}, \boldsymbol{\omega}) = \int_1^{50} \left(\sum_{i=1}^{3} \sum_{j=1}^{3} \left\{ \left[\widehat{g}_{ij}(r)\right]^{1/4} - \left[g_{ij}(r; \boldsymbol{\alpha}, \boldsymbol{\omega})\right]^{1/4} \right\}^2 \right) \mathrm{d}r$$

with respect to $\boldsymbol{\alpha}$ and $\boldsymbol{\omega}$. Here $\widehat{g}_{ij}(r)$ is the empirical estimates of the pair correlation functions and $g_{ij}(r; \boldsymbol{\alpha}, \boldsymbol{\omega})$ is the theoretical pair correlation function of the model, given in (2). Also, the maximum pairwise dependence range between trees of the same and different species is assumed to be 50 m. In addition, since the empirical estimates $\widehat{g}_{ij}(r)$ are not reliable at r values near zero (Illian et al., 2008, pp. 234-235), we only considered $1 < r < 50$ in the integral of the contrast function. Using this minimum contrast method, the estimated parameters are

$$\widehat{\boldsymbol{\alpha}} = \begin{bmatrix} 1.13\ (.24) & -0.29\ (.51) & 0.09\ (.42) \\ 0.23\ (.62) & 1.52\ (.32) & -0.22\ (.40) \\ -0.30\ (.56) & -0.77\ (.42) & 1.41\ (.31) \end{bmatrix}$$

and

$$\widehat{\boldsymbol{\omega}} = \big(21.25\ (8.98), 16.72\ (7.43), 35.44\ (15.74)\big),$$

where the small numbers in the parenthesis are parametric bootstrap (also known as Monte Carlo) standard errors of the estimates which are obtained by 500 simulations from the fitted log-Gaussian Cox model.

Given $\widehat{\boldsymbol{\alpha}}$, estimate of the association parameter matrix

$$\widehat{\boldsymbol{\Sigma}} = \widehat{\boldsymbol{\alpha}}\widehat{\boldsymbol{\alpha}}^{\mathsf{T}} = \begin{bmatrix} 1.38\ (.38) & -0.21\ (.51) & 0.01\ (.38) \\ -0.21\ (.51) & 2.42\ (.55) & -1.55\ (.48) \\ 0.01\ (.38) & -1.55\ (.48) & 2.67\ (.52) \end{bmatrix}$$

is obtained. According to $\widehat{\sigma}_{ij}$ values, there is a considerable negative association ($\widehat{\sigma}_{23} = -1.55$) between *Populus Caspica* and *Pterocarya Fraxinifolia* species but the association between *Fraxinus excelsior* and *Pterocarya Fraxinifolia* species ($\widehat{\sigma}_{13} = -0.21$) and specially *Fraxinus excelsior* and *Populus Caspica* ($\widehat{\sigma}_{12} = 0.01$) species are small. To explicitly check the hypothesis $\mathrm{H}_0 : \sigma_{ij} = 0$, $i \neq j$, we used the parametric bootstrap with 500 simulations from the fitted model and obtained the 95% bootstrap percentile confidence intervals for the association parameters σ_{ij}. The confidence intervals are reported in 2. The association between *Fraxinus excelsior* and *Populus Caspica* and *Fraxinus excelsior* and *Pterocarya Fraxinifolia* are not significant at the 5% level because their corresponding confidence intervals contain zero. However, the confidence interval of σ_{23} does not contain zero which implies that the association between *Populus Caspica* and *Pterocarya Fraxinifolia* species is significantly negative. Therefore, as

Table 2: The 95% bootstrap percentile confidence intervals for the association parameters σ_{ij} obtained using 500 simulations from the fitted log-Gaussian Cox model.

species	parameter	95% CI
Fraxinus excelsior & *Populus Caspica*	σ_{12}	(-1.54, 0.50)
Fraxinus excelsior & *Pterocarya Fraxinifolia*	σ_{13}	(-0.75, 0.66)
Populus Caspica & *Pterocarya Fraxinifolia*	σ_{23}	(-2.81, -0.98)

Table 3: The maximum-likelihood estimates of Gamma model parameters for DBH of *Fraxinus excelsior*, *Populus Caspica* and *Pterocarya Fraxinifolia*. The standard errors are given in the parenthesise.

species	parameters
Fraxinus excelsior	$\widehat{\nu}_1 = 4.38$ (0.31), $\widehat{\tau}_1 = 0.11$ (0.009)
Populus Caspica	$\widehat{\nu}_2 = 7.38$ (0.60), $\widehat{\tau}_2 = 0.10$ (0.009)
Pterocarya Fraxinifolia	$\widehat{\nu}_3 = 7.63$ (0.42), $\widehat{\tau}_3 = 0.0194$ (0.0008), $\widehat{\eta}_3 = 0.393$ (0.029)

anticipated from plots of the empirical pair correlation functions in 2, the associations between *Fraxinus excelsior* and *Populus Caspica* and *Fraxinus excelsior* and *Pterocarya Fraxinifolia* species are not strong enough to be distinguished from the neutral case but there is a distinguishable negative association between *Populus Caspica* and *Pterocarya Fraxinifolia* species. Both *Populus Caspica* and *Pterocarya Fraxinifolia* are hygrophyte species and hence they are expected to co-occur in moist sites. However, the two species are also both light-demanding species and their competition for light is a source of the negative associations.

The maximum-likelihood estimates of parameters of the Gamma models for DBH distributions of species *Fraxinus excelsior* and *Populus Caspica* are reported in 3. In order to fit the GLM model (4) to the DBHs of *Pterocarya Fraxinifolia* trees, the latent random intensity function $\Lambda_3(\boldsymbol{u}_{3j})$ needs to be predicted by $\mathbb{E}[\Lambda_3(\boldsymbol{u}_{3j})|X_3 = \boldsymbol{x}_3]$ which can not be expressed in a closed analytical form (Møller and Waagepetersen, 2004). Thus, the observation window W is discretized by a 512×256 grid and a MCMC algorithm, described in Taylor et al. (2013), with 100000 iterations and 10000 burn-in is used to approximate $\mathbb{E}[\Lambda_3(\boldsymbol{u})|X_3 = \boldsymbol{x}_3]$ over W (see middle panel of 7). Then the GLM model (4) is used to obtain maximum-likelihood estimates of parameters ν_3, τ_3 and η_3 (see 3).

6.2 Model checking

To assess the goodness of fit of the fitted log-Gaussian Cox model, we compare empirical and model based estimates of the auto- and cross-pair correlation functions. Plots of these functions, their 95% simulation envelopes under the fitted model and p-values of the corresponding deviation tests are shown in 6. The plots and p-values show that there is no significant disagreement between empirical auto- and cross-pair correlation functions and their corresponding theoretical values from the fitted model. Thus the fitted log-Gaussian Cox model can adequately explain the intra- and inter-specific associations in the observed point patterns $\boldsymbol{x}_1$, $\boldsymbol{x}_2$ and $\boldsymbol{x}_3$ and it is likely that the fitted model generates point patterns with spatial structures similar to the observed point pattens in Figure 1.

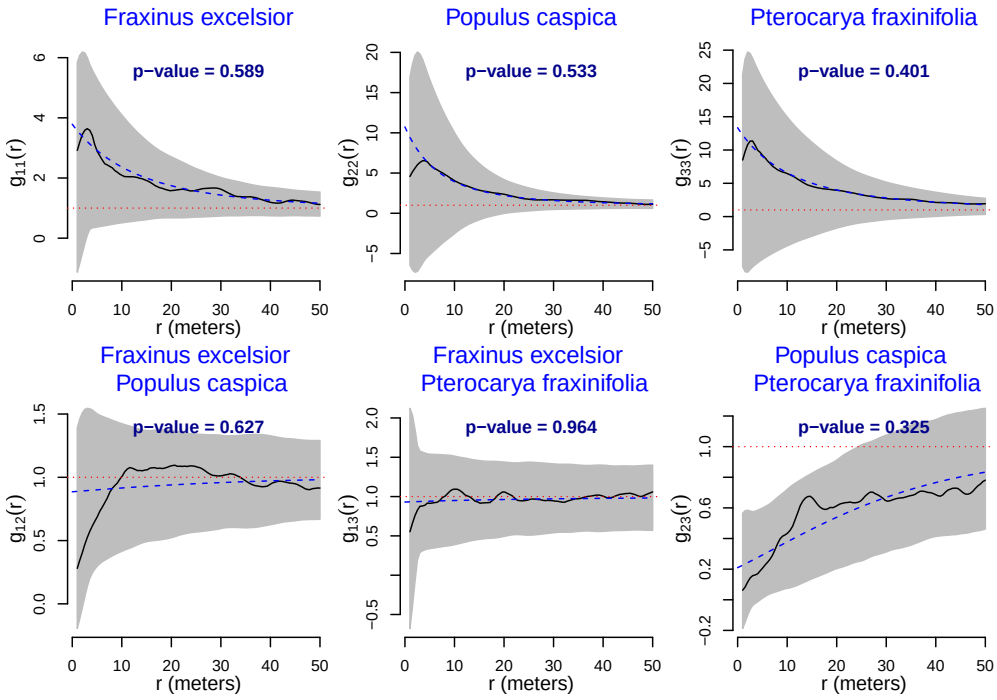

Figure 6: Estimated auto- and cross-pair correlation functions along with 95% Monte Carlo envelopes (grey area) under the fitted log-Gaussian Cox model and p-values of corresponding deviation tests.

To check the DBH models, we use the Q-Q plot of the fitted Gamma models to the DBH distributions of *Fraxinus excelsior* and *Populus Caspica* species. These plots are shown in upper panels of 7. Except for several outliers, the plotted quantile points lie approximately on a straight line which means that it is plausible that the DBH data follow the fitted Gamma distributions. The scatter plot of the DBH of *Pterocarya Fraxinifolia* trees versus the predicted random intensity function $\mathbb{E}[\Lambda_3(\boldsymbol{u}_{3j})|X_3 = \boldsymbol{x}_3]$ and the fitted Gamma GLM model with inverse link function (4) are shown in the left lower panel of 7. The scatter plot shows a negative nonlinear relation between the DBH and predicted random intensity of *Pterocarya Fraxinifolia* and the model seems to reasonably capture this relationship. The right lower panel of 7 shows the empirical variogram and the mean and 95% confidence intervals of variograms of 5000 simulations from the fitted GLM model. The p-value of the corresponding deviation test is also reported at the top of the plot. The empirical variogram lies between the 95% simulation envelopes meaning that the model can be said to adequately describe the spatial correlation in the DBH of *Pterocarya Fraxinifolia* trees.

All calculations were carried out with the R language (R Core Team, 2013) using the spatial statistics package spatstat (Baddeley and Turner, 2005) and the package lgcp (Taylor et al., 2013).

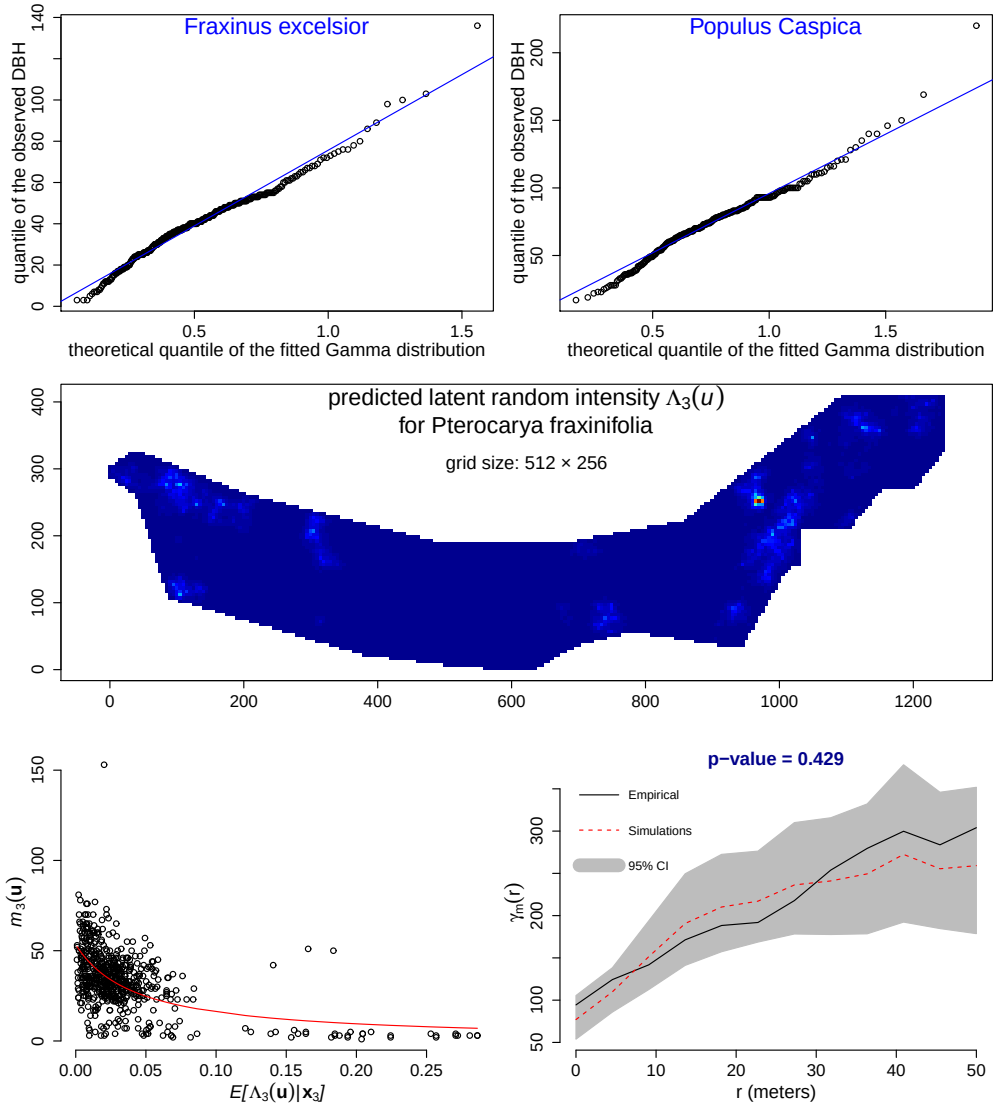

Figure 7: Q-Q plots of the fitted Gamma distributions to the DBH of *Fraxinus excelsior* (left upper panel) and *Populus Caspica* (right upper panel), predicted random intensity function for *Pterocarya Fraxinifolia* (middle panel) on a grid over W, DBH of *Pterocarya Fraxinifolia* versus its predicted random intensity function at tree locations and the fitted GLM model (left lower panel) and empirical, mean and 95% simulation envelopes under the fitted Gamma intensity-dependent model to the DBH of *Pterocarya Fraxinifolia* species (right lower panel).

7 Conclusion

In this article, we studied the spatial distributions and associations of species *Fraxinus excelsior*, *Populus caspica* and *Petrocarya fraxinifolia* in a Hyrcanian forest in North

of Iran. *Fraxinus excelsior* is a flexible species that can be found over a range of growing conditions. It often occurs in mixed forests but rarely become a dominant species. The seeds of *Fraxinus excelsior* are lightweight and can be carried by the wind. However, they only grow in areas with appropriate conditions and create clustered patterns (Pautasso et al., 2013). *Pterocarya fraxinifolia* is a fast growing species and prefers flat ground and deep moist soils (Sheykholislami et al., 2009). Regeneration of *Populus caspica* and *Petrocarya fraxinifolia* are mostly by root propagation and seed dispersal and thus their seedlings establish near the parent trees, which cause clustered patterns.

We used a multivariate log-Gaussian Cox process model with tractable and flexible parametric auto- and cross-pair correlation functions which explain intra- and inter-specific correlations. The associations among species were quantified in terms of some parameters. Using this model, we were able to detect a significant negative associations, or competition, between *Populus caspica* and *Petrocarya fraxinifolia* species. The negative association was expected because the two species compete for the light.

After modeling tree locations, we used a mark variogram to analyze spatial correlation in diameters of trees of each species. The results showed that DBH of *Petrocarya fraxinifolia* trees are influenced by their conspecific neighboring trees but the DBH of trees of other species can be considered independent and identically distributed. We used a conditional Gamma intensity-dependent marking model for the DBH of *Petrocarya fraxinifolia* trees and fitted Gamma distributions to DBH of *Fraxinus excelsior* and *Populus caspica* trees. All fitted models were checked for goodness of fit.

Through this analysis we showed that the considered multivariate log-Gaussian Cox process model, suggested by Møller et al. (1998), can be a flexible and powerful tool in modeling spatial point patterns of several species. We discussed how the model can take into account the overall mixed effect of all known and unknown influential factors in spatial forest structure and species associations using hidden Gaussian random fields. The multivariate log-Gaussian model can be extended and modified to cover spatial patterns of several species with different features. For example, the model can be extended to the inhomogeneous case if the environment is inhomogeneous and spatial covariates on soil and topographical conditions are available. In the inhomogeneous case, the intensity function of each species can be a parametric function of the covariates. Also, for data with more complex spatial dependence structures the exponential correlation functions of the hidden Gaussian random fields of the model can be replaced with other correlation functions; for instance the general class of Matérn correlation functions.

References

Akhavan, R., Sagheb-Talebi, K., Zenner, E. and Safavimanesh, F. (2012). Spatial patterns in different forest development stages of an intact old-growth Oriental beech forest in the Caspian region of Iran. *European Journal of Forest Research*, **131**(5):1355–1366.

Baddeley, A. and Turner, R. (2005). spatstat: An r package for analyzing spatial point patterns. *Journal of Statistical Software*, **12**(6):1–42.

Brix, A. and Møller, J. (2001). Space-time multi type log Gaussian Cox processes with a view to modelling weeds. *Scandinavian Journal of Statistics*, **28**(3):471–488.

Comas, C. and Mateu, J. (2007). Modelling forest dynamics: a perspective from point process methods. *Biometrical journal*, **49**(2):176–196.

Esmailzadeh, O., Hosseini, S. M., Tabari, M., Baskin, C. C. and Asadi, H. (2011). Persistent soil seed banks and floristic diversity in Fagus orientalis forest communities in the Hyrcanian vegetation region of Iran. *Flora - Morphology, Distribution, Functional Ecology of Plants*, **206**(4):365 – 372.

Grabarnik, P. and Särkkä, A. (2009). Modelling the spatial structure of forest stands by multivariate point processes with hierarchical interactions. *Ecological Modelling*, **220**(9):1232–1240.

Hao, Z., Zhang, J., Song, B., Ye, J. and Li, B. (2007). Vertical structure and spatial associations of dominant tree species in an old-growth temperate forest. *Forest Ecology and Management*, **252**(1):1–11.

Hou, J., Mi, X., Liu, C. and Ma, K. (2004). Spatial patterns and associations in a quercus-betula forest in northern china. *Journal of Vegetation Science*, **15**(3):407–414.

Illian, J. and Burslem, D. (2007). Contributions of spatial point process modelling to biodiversity theory. *Journal de la société française de statistique*, **148**:9–29.

Illian, J., Penttinen, A., Stoyan, H. and Stoyan, D. (2008). *Statistical Analysis and Modelling of Spatial Point Patterns*. Wiley, London.

Illian, J. B., Martino, S., Sørbye, S. H., Gallego-Fernández, J. B., Zunzunegui, M., Esquivias, M. P. and Travis, J. M. (2013). Fitting complex ecological point process models with integrated nested laplace approximation. *Methods in Ecology and Evolution*.

Illian, J. B., Møller, J. and Waagepetersen, R. P. (2009). Hierarchical spatial point process analysis for a plant community with high biodiversity. *Environmental and Ecological Statistics*, **16**(3):389–405.

Illian, J. B., Sørbye, S. H. and Rue, H. (2012). A toolbox for fitting complex spatial point process models using integrated nested laplace approximation (inla). *The Annals of Applied Statistics*, **6**(4):1499–1530.

Jalilian, A., Guan, Y. and Waagepetersen, R. (2012). Decomposition of variance for spatial Cox processes. *Scandinavian Journal of Statistics*, **40**(1):119–137.

Law, R., Illian, J., Burslem, D. F., Gratzer, G., Gunatilleke, C. and Gunatilleke, I. (2009). Ecological information from spatial patterns of plants: insights from point process theory. *Journal of Ecology*, **97**(4):616–628.

Lee, K., Kim, S., Shin, Y. and Choung, Y. (2012). Spatial pattern and association of tree species in a mixed Abies holophylla-broadleaved deciduous forest in Odaesan National Park. *Journal of Plant Biology*, **55**(3):242–250.

Luo, Z. R., Ming, J. Y., Chen, D. L., Wu, Y. G. and Ding, B. Y. (2012). Spatial associations of tree species in a subtropical evergreen broad-leaved forest. *Journal of Plant Ecology*, **5**(3):346–355.

Martínez, I., Wiegand, T., González-Taboada, F. and Obeso, J. R. (2010). Spatial associations among tree species in a temperate forest community in north-western spain. *Forest Ecology and Management*, **260**(4):456–465.

Møller, J., Syversveen, A. R. and Waagepetersen, R. P. (1998). Log Gaussian Cox processes. *Scandinavian Journal of Statistics*, **25**:451–482.

Møller, J. and Waagepetersen, R. P. (2004). *Statistical inference and simulation for spatial point processes*. Chapman and Hall/CRC, Boca Raton.

Møller, J. and Waagepetersen, R. P. (2007). Modern statistics for spatial point processes*. *Scandinavian Journal of Statistics*, **34**(4):643–684.

Myllymäki, M., Grabarnik, P., Seijo, H. and Stoyan, D. (2015). Deviation test construction and power comparison for marked spatial point patterns. *Spatial Statistics*, **11**:19–34.

Myllymäki, M. and Penttinen, A. (2009). Conditionally heteroscedastic intensity-dependent marking of log gaussian cox processes. *Statistica Neerlandica*, **63**(4):450–473.

Pautasso, M., Aas, G., Queloz, V. and Holdenrieder, O. (2013). European ash (Fraxinus excelsior) dieback – a conservation biology challenge. *Biological Conservation*, **158**(0):37 – 49.

Picard, N., BAR-HEN, A., Mortier, F. and Chadœuf, J. (2009). The multi-scale marked area-interaction point process: A model for the spatial pattern of trees. *Scandinavian Journal of Statistics*, **36**(1):23–41.

Pommerening, A. and Särkkä, A. (2013). What mark variograms tell about spatial plant interactions. *Ecological Modelling*, **251**:64–72.

R Core Team (2013). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Ramezani, E., Mohadjer, M. R. M., Knapp, H.-D., Theuerkauf, M., Manthey, M. and Joosten, H. (2013). Pollen–vegetation relationships in the central Caspian (Hyrcanian) forests of northern Iran. *Review of Palaeobotany and Palynology*, **189**(0):38 – 49.

Rue, H., Martino, S. and Chopin, N. (2009). Approximate bayesian inference for latent gaussian models by using integrated nested laplace approximations. *Journal of the royal statistical society: Series b (statistical methodology)*, **71**(2):319–392.

Salehi, A., Ghorbanzadeh, N., Kahneh, E., et al. (2013). Earthworm biomass and abundance, soil chemical and physical properties under different poplar plantations in the north of Iran. *Journal of Forest Science*, **59**(6):223–229.

Sheykholislami, A., Ahmadi, T., et al. (2009). The study of Caucasian walnut (Pterocarya frexinifolia (Lam.) Spach.) in forests of Mashelak (Noshahr, Iran). *Botany Research Journal*, **2**(2):28–33.

Stoyan, D. and Penttinen, A. (2000). Recent applications of point process methods in forestry statistics. *Statistical Science*, **15**(1):61–78.

Taylor, B. M., Davies, T. M., Rowlingson, B. S. and Diggle, P. J. (2013). lgcp: An R package for inference with spatial and spatio-temporal log-Gaussian Cox processes. *Journal of Statistical Software*, **52**(4):1–40.

Waagepetersen, R., Guan, Y., Jalilian, A. and Mateu, J. (2016). Analysis of multivariate spatial point patterns by using multivariate log-gaussian cox processes. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, **65**(1):77–96.

Waagepetersen, R. P. (2007). An estimating function approach to inference for inhomogeneous neyman–scott processes. *Biometrics*, **63**(1):252–258.

Wiegand, T., Gunatilleke, S. and Gunatilleke, N. (2007). Species associations in a heterogeneous Sri Lankan dipterocarp forest. *The American Naturalist*, **170**(4):E77–E95.

Wiegand, T. and Moloney, K. A. (2004). Rings, circles, and null-models for point pattern analysis in ecology. *Oikos*, **104**(2):209–229.

Zhang, J., Song, B., Li, B.-H., Ye, J., Wang, X.-G. and Hao, Z.-Q. (2010). Spatial patterns and associations of six congeneric species in an old-growth temprate forest. *Acta Oecologica*, **36**(1):29–38.