A Comparison of Multivariate Log Gaussian Cox Process and Saturated Pairwise Interaction Gibbs Point Process

Chathuri Samarasekara[1], Yan Wang[1], and Ian Flint[2]

[1]RMIT University
[2]The University of Melbourne School of Agriculture Food and Ecosystem Sciences

September 13, 2024

Abstract

The study of the spatial point patterns in ecology, such as the records of the observed locations of trees, shrubs, nests, burrows, or documented animal presence, relies on multivariate point process models. This study aims to compare the efficacy and applicability of two prominent multivariate point process models, the multivariate log Gaussian Cox process (MLGCP) and the Saturated Pairwise Interaction Gibbs Point Process model (SPIGPP) , highlighting their respective strengths and weaknesses in various scenarios. Using synthetic and real datasets, we assessed both models based on their predictive accuracy of the empirical K function (can we say this?). Our analysis revealed that both MLGCP and SPIGPP effectively identify and capture mild to moderate attractions and regulations. MLGCP struggles to capture repulsive associations as they intensify. In contrast, SPIGPP can well estimates both the direction and magnitude of interactions even when the model is miss-specified. Both models present unique advantages: MLGCP is particularly effective when there is a need to account for complex, unobserved heterogeneities that vary across space, while SPIGPP is suitable when interactions between points are the primary focus. The choice between these models should be guided by the specific needs of the research question and data characteristics.

A Comparison of Multivariate Log Gaussian Cox Process and Saturated Pairwise Interaction Gibbs Point Process

Chathuri L. Samarasekara[1,*], Ian Flint[2], and Yan Wang[1]

[1]*School of Science, RMIT University, Melbourne, VIC, Australia*

[2]*School of Agriculture, Food and Ecosystem Sciences, the University of Melbourne, Parkville, VIC, Australia*

Abstract

The study of the spatial point patterns in ecology, such as the records of the observed locations of trees, shrubs, nests, burrows, or documented animal presence, relies on multivariate point process models. This study aims to compare the efficacy and applicability of two prominent multivariate point process models, the multivariate log Gaussian Cox process (MLGCP) and the Saturated Pairwise Interaction Gibbs Point Process model (SPIGPP) , highlighting their respective strengths and weaknesses in various scenarios. Using synthetic and real datasets, we assessed both models based on their predictive accuracy of the empirical K function (can we say this?). Our analysis revealed that both MLGCP and SPIGPP effectively identify and capture mild to moderate attractions and regulations. MLGCP struggles to capture repulsive associations as they intensify. In contrast, SPIGPP can well estimates both the direction and magnitude of interactions even when the model is miss-specified. Both models present

*Corresponding author. Email: chathuri.l.samarasekara@gmail.com

unique advantages: MLGCP is particularly effective when there is a need to account for complex, unobserved heterogeneities that vary across space, while SPIGPP is suitable when interactions between points are the primary focus. The choice between these models should be guided by the specific needs of the research question and data characteristics.

Keywords log-Gaussian Cox process, saturated pairwise interaction Gibbs point process, semi-parametric, pair correlation function, point process, multivariate

1 Introduction

Spatial point patterns in ecology record are a common object of study. Point Process Models (PPMs) offer a theoretical foundation for the understanding and analysis ofthe spatial arrangement of trees or animals. PPMs also play a crucial role in understanding species distributions across continuous space. The majority of multivariate spatial point process applications in ecology so far have predominantly taken descriptive approaches, relying on cross summary statistics such as cross K, cross pair correlation, or cross J functions (Baddeley et al., 2014; Cronie and van Lieshout, 2016; Møller and Waagepetersen, 2003) if consistent estimates of the intensity functions are available. Parametric estimation of cross associations is also possible. Jalilian et al. (2015); Waagepetersen et al. (2016) and Choiruddin et al. (2020) used parametric models of intensity and pair correlation functions, while Rajala et al. (2018) specified a full multivariate Markov point process model.

To address this limitation, two primary multivariate point process models have emerged, the Multivariate Log Gaussian Cox Process (Waagepetersen et al., 2016) and the Saturated Pairwise Interaction Gibbs Point Process (Flint et al., 2022; Rajala et al., 2018). In a recent development, Hessellund et al. (2022a) replaced the parametric model in Waagepetersen et al. (2016) with a semi-parametric model from Hessellund et al. (2022b), deriving a second-order conditional composite likelihood function for Multivariate Log Gaussian Cox Process (MLGCP). Hessellund et al. (2022a) combines semi-parametric composite likelihood with a Lasso penalization. A similar technique was applied by Choiruddin et al. (2020) to explore least squares estimation for a MLGCP, where a full parametric model determined the multivariate intensity function.

Cox processes struggle to model negative interactions and interactions of varying scales

(Waagepetersen et al., 2016). In contrast, the saturated models, which are a type of Gibbs processes, address these limitations by introducing a saturation parameter that allow them to model either attraction or repulsion (C.J.Geyer, 1999). Rajala et al. (2018) extended this process to the setting so as to study a larger species subset from the Barro Colorado Island dataset.

However, Rajala et al. (2018) models interactions as being driven by step-function potentials. To overcome this, Flint et al. (2022) introduced the Saturated Pairwise Interaction Gibbs Point Process (SPIGPP) model, building upon Rajala et al. (2018). This model introduces a unified framework to model multi-species marked point patterns, by allowing for a range of potential shapes, enabling ecologically grounded potential functions that account for individual characteristics such as size or diameter.

While these models have seen widespread use, there has been a notable absence of direct comparative studies between the two types of point process models of MLGCP and SPIGPP. This may be due to their different theoretical foundations, which make direct comparisons challenging. Our research addresses this gap by developing statistical measures that facilitate the systematic evaluation of these two distinct types of point process models. Through our comprehensive simulation study and the examination of real data examples, we not only highlight the advantages and disadvantages of both models, but also provide novel insights into where they excel and their limitations. This comparative analysis is essential for advancing our understanding of multi-type point pattern modelling in ecology, offering clear context-dependent guidance on selecting and comparing these models.

The paper is organized as follows: Section 2 includes an overview of multivariate log Gaussian Cox processes and saturated pairwise Gibbs processes and the detailed protocol for comparison of fitted models. Then in Sections 3 and 4 we applied the methodologies to the simulation studies and case analyses. Section 5 includes a detailed discussion of the results obtained from both the simulation study and the case study. Finally, Section 6 concludes with some closing remarks.

2 Materials and Methodology

In this section, we provide a brief overview of the MLGCP and SPIGPP models.

2.1 Multivariate Log Gaussian Cox Process

This section describes the theoretical underpinnings of the MLGCP as introduced by Hessellund et al. (2022a), which builds upon the groundwork laid by Waagepetersen et al. (2016). Choiruddin et al. (2020) and Jalilian et al. (2020) have additionally contributed to its expansion.

Following the definition outlined in Waagepetersen et al. (2016), we denote by $X = (X_1, \dots, X_p)$, a multivariate spatial point process, where X_i is a spatial point process on $\mathbb{R}^d$ (in ecology we will be using $d = 2$) representing points of type $i = 1, \dots, p$. The point pattern X_i for $i = 1, 2, \dots, p$ is modelled as a Cox process with random intensity function;

$$\Lambda_i(\mathbf{u}) = \rho_0(\mathbf{u}) \exp(\gamma_i^T \mathbf{z}(\mathbf{u})) \exp\Big(\mu_i + \sum_{k=1}^{q} \alpha_{ik} \mathbf{Y}_k(\mathbf{u}) + \sigma_i \mathbf{U}_i(\mathbf{u})\Big). \tag{1}$$

A Cox process simply is a Poisson point process in which the intensity is random because of the Gaussian filed introduced. Note that we will define and interpret the various terms in the following paragraphs.

In the approach outlined in Hessellund et al. (2022a), a semi-parametric model is employed. The background intensity function ρ_0, aims to capture intricate variations in intensity functions common to all point processes $X_1, \dots, X_p$. The intensity of X_i is determined by a regression parameter vector γ_i alongside a vector of spatial covariates denoted as $\mathbf{z}(\mathbf{u})$ at location u.

The formulation involves independent zero-mean unit-variance Gaussian random fields $\mathbf{Y}_k$ and $\mathbf{U}_i$ with $\mu_i = -\sum_{k=2}^{q} \frac{\alpha_{ik}^2}{2} - \frac{\sigma_i^2}{2}$. Y_k acts as a latent factor influencing all point types, potentially creating correlations among different types due to their simultaneous dependence on $\mathbf{Y}_k$. Conversely, each $\mathbf{U}_i$ is a type-specific factor exclusively impacting the ith point type, modeling clustering within X_i. The parameter q, representing the number of latent common fields, governs the model's complexity.

When $\mathbf{Y}_k$ is observed (i.e., non-random), constraints such as $\alpha_{pl} = 0$ or $\sum_{i=1}^{p} \alpha_{il} = 0$, $l = 1, \ldots, q$ are necessary for identifiability. With unobserved $\mathbf{Y}_k$ and less information, a sum-to-zero constraint, $\sum_{i=1}^{p} \alpha_{il} = 0, i = 1, \ldots, q$, ensures symmetrical treatment across all X_i. The cross pair correlation function (pcf) of X_i and X_j are given by (Hessellund et al., 2022a);

$$g_{ij}(\mathbf{r};\theta) = \exp\left[\sum_{k=1}^{q} \alpha_{ik}\alpha_{jk} \exp\left(\frac{-r}{\xi_k}\right) + 1[i=j]\sigma_i^2 \exp\left(\frac{-r}{\psi_i}\right)\right], \tag{2}$$

where θ is the concatenation of $\alpha_{.k} = (\alpha_{1k}, \ldots, \alpha_{pk})^T$ $(k = 1, \ldots, q)$, $\xi = (\xi_1, \ldots, \xi_q)^T$, $\sigma^2 = (\sigma_1^2, \ldots, \sigma_p^2)^T$ and $\psi = (\psi_1, \ldots, \psi_p)^T$. If $\sum_{k=1}^{q} \alpha_{ik}\alpha_{jk} \exp\left(\frac{-r}{\xi_k}\right)$ is positive (negative), it indicates positive (negative) spatial correlation between points from X_i and X_j at distance r. The parameters ξ_k and ψ_i are the exponential correlation scale parameters of $\mathbf{Y}_k$ and $\mathbf{U}_i$, respectively.

In Hessellund et al. (2022a), β_i, the coefficients of the covariates are estimated first using the first order conditional likelihood as used in Hessellund et al. (2022b). Then, estimating θ is done by maximizing the second-order conditional composite likelihood function in equation (7) in Hessellund et al. (2022a). The cross Pair Correlation Functions (PCFs) in equation 2 and the second-order conditional composite likelihood function (equation

(7) in Hessellund et al. (2022a)) remain invariant to specific transformations, as noted by Hessellund et al. (2022a). The lack of identifiability isn't a significant concern, given the focus on the correlation structure rather than individual α_{ij}'s. Further optimization details can be found in Sections 3.1 and 3.2 of Hessellund et al. (2022a).

2.2 Saturated Pairwise Interaction Gibbs Point Process

This section recall the definition of the Saturated Pairwise Interaction Gibbs Point Process (SPIGPP) as introduced in Flint et al. (2022). The model is specified by its density

$$j(X) = C \exp \Bigg[\sum_{(x,i,m)\in X} (\beta_{0,i} + \sum \beta_{i,k} Z_k(x)) \\ + \sum_{i=1}^{p} \sum_{z=(x_1,i_1,m_1)\in X} \alpha_{p_{i_1,i_2}} u(z,(X \setminus \{z\})_{i_2}) + \sum_{i=1}^{p} \sum_{z=(x_1,i_1,m_1)\in X} \gamma_{i_1,i_2} v(z,(X \setminus \{z\})_{i_2}) \Bigg]. \tag{3}$$

In the equation above, X is a spatial pattern and $C > 0$ is a normalization constant and the other parameters are interpreted as (Flint et al., 2022):

(a) An intercept vector $(\beta_{1,0}, \beta_{2,0}, ..., \beta_{p,0})^T \in \mathbb{R}^p$, representing the log-intensities of distinct species in the absence of interactions.

(b) Environmental covariates $Z_1, \ldots, Z_K$, assumed to have bounded values.

(c) For $1 \leq i \leq p$ and $1 \leq k \leq K$, coefficients $\beta_{i,k}$ indicating the response of species i to environmental covariate k.

(d) A function $u(z,(X \setminus \{z\})_{i_2})$ modeling short-range interactions between species i_2 in X and an individual $z = (x, i_1, m)$ of species i_1 with mark m at location x.

(e) A function $v(z,(X \setminus \{z\})_{i_2})$representing medium-range interactions between species i^2 in X and an individual z as in (d).

(f) Coefficients $\alpha_{p_{i_1,i_2}}$ for $1 \leq i_1, i_2 \leq p$, denoting the magnitude of short-range interactions between species i_1 and i_2. Positive values signify attraction, while negative values denote repulsion. The assumption of symmetry holds ($\alpha_{p_{i_1,i_2}} = \alpha_{p_{i_2,i_1}}$).

(g) Symmetric coefficients γ_{i_1,i_2} for $1 \leq i_1, i_2 \leq p$, representing the magnitude of medium-range interactions between each pair of species i_1 and i_2. Similar to (f), the sign of γ_{i_1,i_2} indicates attraction or repulsion.

The Papangelou conditional intensity π is directly derived from equation 3 using the formula: $\pi((x, i, m), X) := j(X \cup (x, i, m))/j(X)$ for $(x, i, m) \in X$. Furthermore, the definitions of short, medium and long range interactions distances can be found equations $2 - 5$ of Flint et al. (2022).

2.3 Protocol/Algorithm for Comparison of fitted PPMs

The primary objective of this study is to compare the performance of different point process models. However, due to the different nature of MLGCP and SPIGPP models, direct comparison is not feasible. To allow for their comparison, we propose a step-by-step procedure.

As discussed earlier, the pair correlation function of a MLGCP (equation 2) has a closed form whereas in SPIGPP there exists only a series of expansion which is difficult to compute in practice. Therefore, a comparison of the two methods using the theoretical pair correlations functions is not feasible. However, estimates of summary statistics of both SPIGPP and MLGCP can be computed through using Monte-Carlo (MC) simulations. In the following we focus on the K function which can be estimated more reliably by this MC procedure than alternatives.

Therefore, we propose simulating N samples from the fitted model, and subsequently computing MC estimates of the K function. This process can be easily implemented with the 'spatstat' R package. Using this method, the K functions will be comparable across models, regardless of the model used. As a further step, we compute the mean Integrated Squared Errors (MISE) (Hessellund et al., 2022a) aggregated over all cross-type K functions, that is;

$$MISE_{between}(\hat{\theta}) = \sum_{i<j} E\left[\int_{0.01}^{0.1} (K_{ij}(r;\widehat{\theta_{ij}}) - K_{ij}(r;\theta_{ij}))^2 dr\right]. \tag{4}$$

Where for any pair of types i and j, the multitype K-function $K_{ij}(r,)$, also called the bivariate or cross- type K-function (Baddeley et al., 2016). We also extend this definition to $MISE_{within}$ and $MISE_{total}$, which are similar to $MISE_{between}$ but with summation over $i = j$ or $i \leq j$. It is important to note that this proposed method is applicable to any summary statistic, including cross pair correlation functions, cross J functions, cross L functions, cross F functions, as well as cross K functions. As mentioned previously, we have used cross-type K functions due to their stable nature, which facilitates clearer interpretation.

3 Simulation Study

In this section, we describe the framework of our simulation study. In the first subsection, we discuss the thorough analysis of the SPIGPP (Flint et al., 2022) and MLGCP models (Hessellund et al., 2022a) under various scenarios with two species.

In the Appendix A, we expand on the simulation study introduced in Waagepetersen et al. (2016) and revisited in Hessellund et al. (2022a). We have also assessed the SPIGPP model fit performance when data are simulated from MLGCP, extending beyond the bi-variate case using this simulation study given in the Appdenix B.

To carry out this investigation, we used R (version 4.3.1) statistical software, and the packages ‘Multilogreg’, ‘randomField’, ‘spatstat’, ‘ppjsdm’, and ‘ggplot2’.

3.1 Comparative Simulation Study: Assessing MLGCP and SPIGPP Models Under Various Scenarios

In this section, we describe the comprehensive simulation study, utilising both MLGCP and SPIGPP models with two species. The main objective here is to discern the strengths and weaknesses of each model across various scenarios. The simulation study is organized into two parts: 1) MLGCP Scenarios and 2) SPIGPP Scenarios. Subsequent discussions address each part separately, providing a detailed exploration of the performance of each model under diverse conditions.

3.1.1 MLGCP Scenarios

In each part of the simulation, we explored the association between two different species in various ways, focusing on both within and between species associations. When generating MLGCP scenarios, our emphasis was on understanding the underlying model behaviour. Given that MLGCP cannot model repulsion within a species, we design four distinct scenarios in this section, including mild to strong attractions between and within species as well as mild to strong repulsion between species. The scenarios were defined as follows:

1. MLGCP Scenario 1 - Mild-moderate attraction between and within species (mild “+” b/w species)

2. MLGCP Scenario 2 - Strong attraction between and within species (strong “+” b/w species)

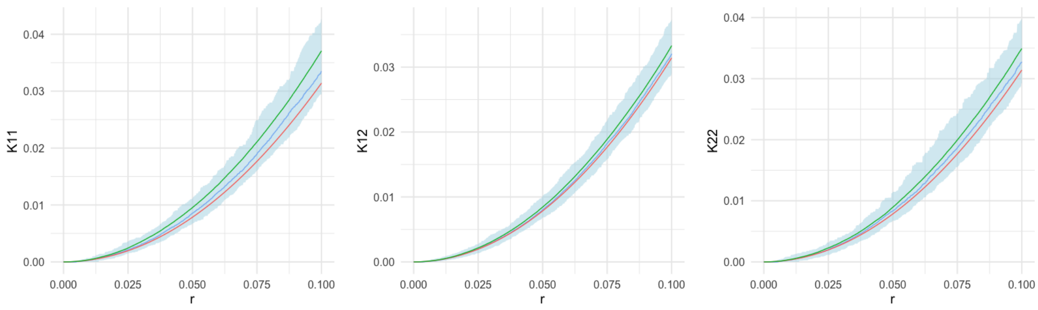
3. MLGCP Scenario 3 - mild-moderate repulsion between and mild to moderate attractions within species (mild "-" b & mild "+" w)

4. MLGCP Scenario 4 - Strong repulsion between and strong attractions within species (strong "-" b & mild "+" w)

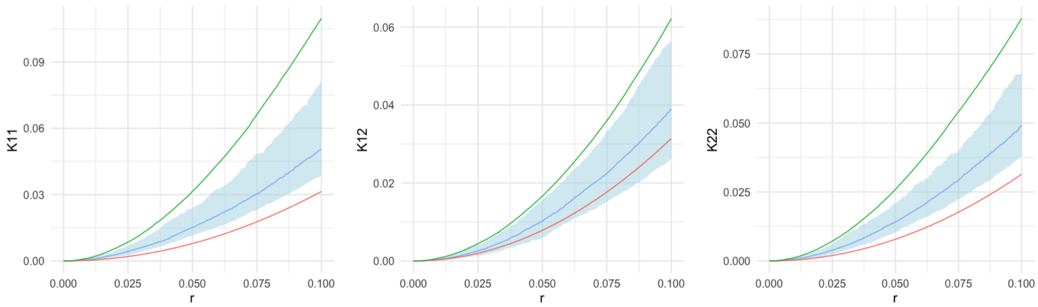
As the initial step of our analysis, we simulated 100 MLGCP processes following the principles outlined in Hessellund et al. (2022a) using the parameters specified in Table A.1 in Appendix A. We then fitted these MLGCP scenarios using SPIGPP to evaluate the fit of SPIGPP when used for mis-specified models (A detailed description of the simulation and fitting procedure can be found in Appendix A). To assess the model fit, we compared the empirical K functions with the fitted K functions, along with their respective confidence bands. Retrieving the model parameters for the MLGCP models was not emphasized, given the identifiability issues discussed in Hessellund et al. (2022a); Jalilian et al. (2020); Choiruddin et al. (2020). Therefore, our primary focus was on the K functions when evaluating the model performance.

In the Figure 1, we compare the fitted and empirical K functions against the baseline K function (given in red), representing the value of K for a homogeneous Poisson point process, defined as $K(r) = \pi \cdot r^2$. If the empirical K function deviates above (below) from this baseline K function, it indicates attraction (repulsion) within/between species.

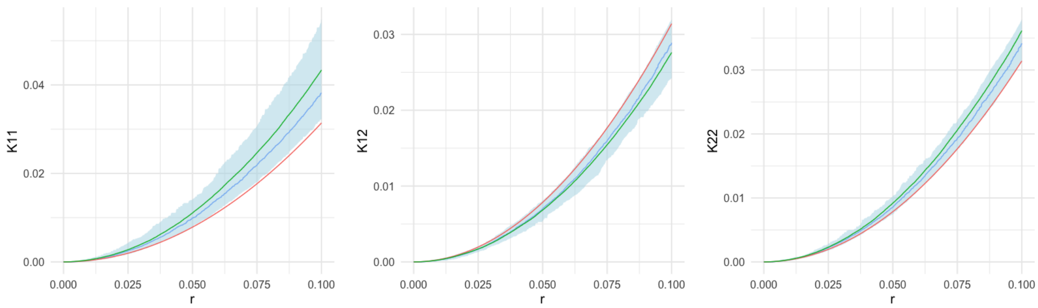
The K functions from scenario 1, featuring mild-moderate attractions between and within species, are depicted on the top row of Figure 1. The fitted SPIGPP model performs admirably in this scenario, with the fitted K functions (blue) closely aligning with all empirical MLGCP K functions (green) and falling well within the estimated confidence bands. We expect differences in the curve shapes of MLGCP and SPIGPP K functions, as they originate from two distinct underlying processes and are not anticipated to overlap.



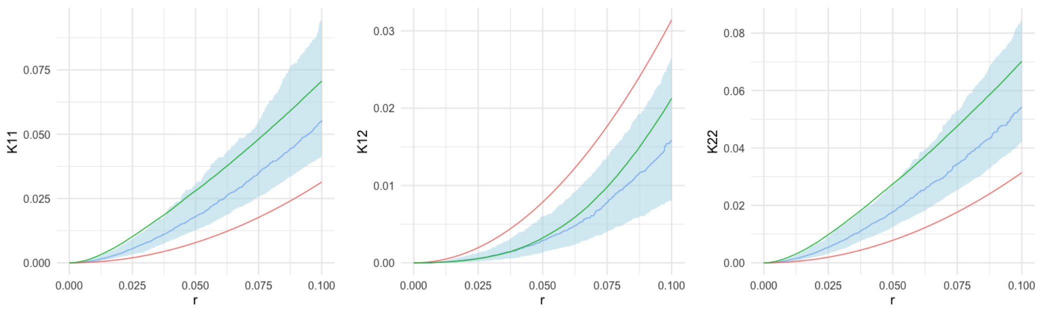
(a) Scenario 1 - mild “+” b/w



(b) Scenario 2 - strong “+” b/w



(c) Scenario 3 - mild “-” b & mild “+” w



(d) Scenario 4 - strong “-” b & mild “+” w

Figure 1: Comparison of K functions across simulated MLGCP scenarios using SPIGPP models. The red line represents the baseline K function (πr^2), while the blue and green lines represent the estimated SPIGPP K function and the empirical K function derived from the simulated MLGCP data, respectively. Each row in the figure corresponds to a distinct scenario, labelled from 1 to 4, showcasing variations across different simulation setups.

The two bottom rows in Figure 1 depict MLGCP scenarios 3 and 4 respectively, involving between-species repulsion and within-species attractions ranging from mild-moderate to strong. SPIGPP adeptly captures the mild-moderate repulsion between species (middle graph in third row from top of Figure 1) as well as the moderate attraction within the species (left and right graphs in third row from top). In MLGCP scenarios 4 (bottom row of Figure 1), characterized by strong between-species repulsive associations and strong within-species attractions, the SPIGPP model effectively captures the between-species repulsive associations. It appropriately captures the strong within-species attractions at longer distances, although at shorter distances the SPIGPP fit slightly falls outside the confidence bounds.

However, MLGCP scenarios 2 (given in the second row from top of Figure 1), characterized by strong attractions within and between species, present a different challenge. The point patterns exhibit notable instability, with a fluctuating number of points for each species during simulation from MLGCP under this scenario. In a lot of cases, SPIGPP underestimated the fitted α_p. Even so, a SPIGPP with mild to large α_p values (i.e., interaction coefficients) is difficult to simulate from. Indeed, the Metropolis-Hastings algorithms in this case regularly fails to converge, with one species dying out and never reappearing. Filtering out some of the samples was thus required. To address this issue, we generated 150 samples of MLGCP processes under these scenarios and removed 50 troublesome samples to obtain a final set of 100 samples. These refined samples were then used to fit SPIGPP models. However, even with this pre-processing, the number of points for each species still varied significantly within the 100 samples, making inference challenging for the SPIGPP.

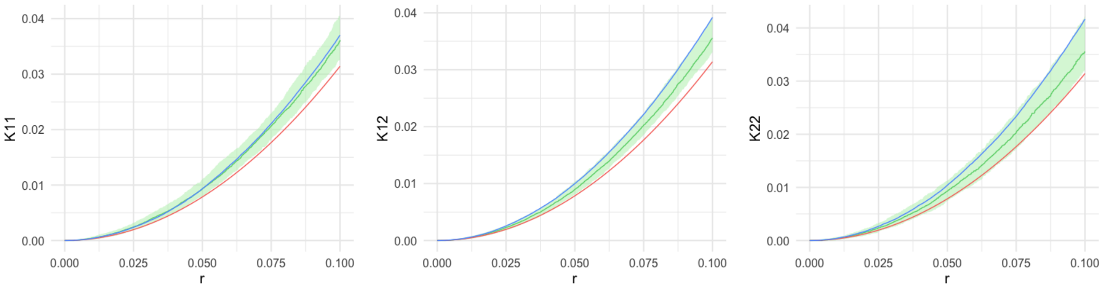
3.1.2 SPIGPP Scenarios

Here, we once again consider two different species, focusing on both within and between species associations. Since SPIGPP models can handle repulsion within species, we create five distinct scenarios in this section, covering mild to strong attractions and repulsive associations within and between species. The scenarios are defined as follows:

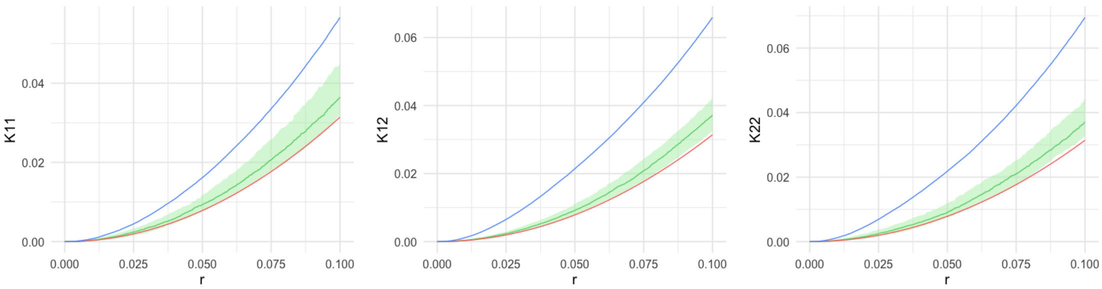
1. SPIGPP Scenario 1 - mild-moderate attraction between and within species (mild "+" b/w)
2. SPIGPP Scenario 2 - strong attraction between and within species (strong "+" b/w)
3. SPIGPP Scenario 3 - mild repulsion between species and mild-moderate attractions within species (mild "-" b & mild "+" w)
4. SPIGPP Scenario 4 - mild-moderate repulsion between species and mild-moderate attractions within species $(2, 2)$ and mild repulsion within species $(1, 1)$ (mild "-" w/b & mild "+" w)
5. SPIGPP Scenario 5 - Strong repulsion between and strong attractions within species (strong "-" b & strong "+" w)

A detailed description of the simulation and fitting procedure is given in Appendix A. Consistent with the approach outlined in the previous section, we assessed the model performance of MLGCP fit in mis-specified scenarios by comparing the empirical and fitted K functions along with the respective confidence bands.

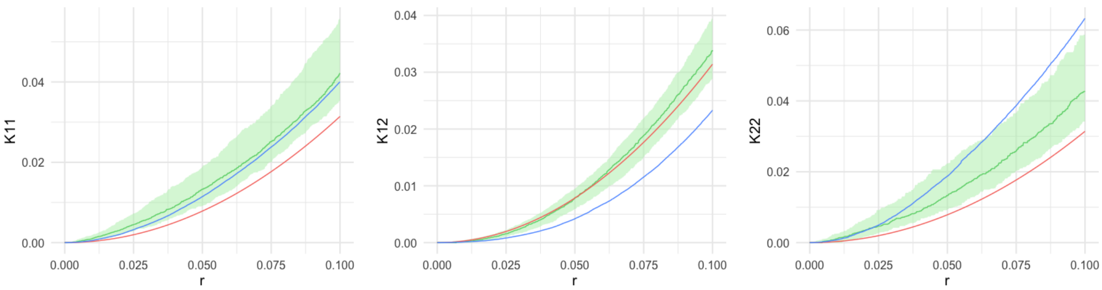
The top row in Figure 2 shows the comparison of K functions for the SPIGPP scenario 1, where there was mild-moderate attractions within and between the two species. The K functions on the top row of Figure 2 show that the MLGCP model captured the mild



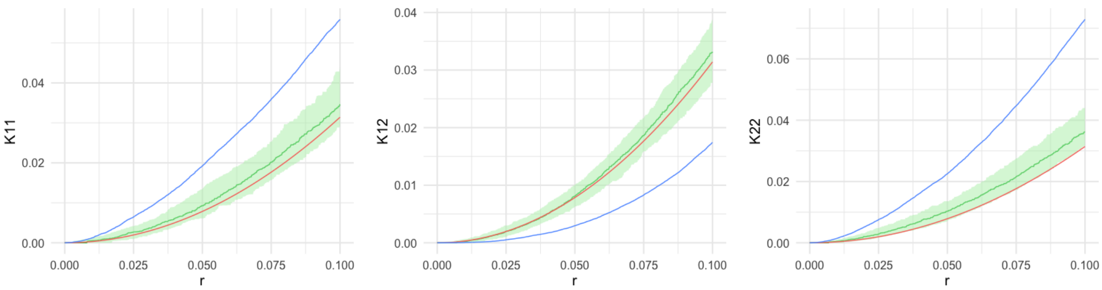
(a) Scenario 1 - mild “+” b/w



(b) Scenario 2 - strong “+” b/w

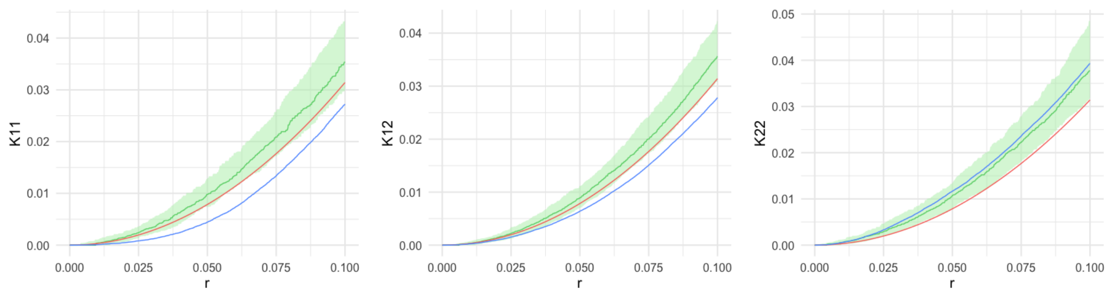


(c) Scenario 3 - mild “-” b & mild “+” w



(d) Scenario 5 - strong “-” b & strong “+” w

Figure 2: Comparison of K functions across simulated MLGCP scenarios using SPIGPP models. The red line represents the baseline K function (πr^2), while the blue and green lines represent the estimated SPIGPP K function and the empirical K function derived from the the simulated MLGCP data, respectively. Each row in the figure corresponds to a distinct scenario, labelled from 1 to 3 and 5, showcasing variations across different simulation setups.



(a) Scenario 4 - mild “-” w/b & mild “+” w

Figure 3: Comparison of K functions across simulated MLGCP scenarios using SPIGPP models. The red line represents the baseline K function (πr^2), while the blue and green lines represent the estimated SPIGPP K function and the empirical K function derived from the the simulated MLGCP data, respectively for scenario 4 where there is mild repulsion within species and mild attraction and repulsion within species.

to moderate within-species attractions in the SPIGPP scenario well (the left and right graphs display the empirical K function in blue within the estimated confidence bands). However, the top middle plot, representing the between-species interaction, showed the empirical K function at the upper bound of the confidence band, indicating that the fit was not very accurate.

It was observed that the stronger the attractions generated by SPIGPP, the more challenging it became for the MLGCP fit to achieve the required magnitude of attraction both within and between, even though it captured the presence of an attraction in the scenario (second row of Figure 2). This was similar to what we observed in the previous section with MLGCP scenarios.

In the third row in Figure 2, we observed mild to moderate repulsion between species and moderate attractions within each species (SPIGPP scenario 3). The MLGCP fit performed well when the attraction was mild, as seen in the left graph in third row plots of Figure 2, and it also accurately estimated the attraction within the second species (right graph in third row) at short distances. While it identified the repulsion between species at shorter distances, it was challenging for the MLGCP fit to accurately estimate the magnitude of the moderate repulsion.

Similarly, in scenarios with strong repulsion between species and strong attractions within each species (bottom row of Figure 2), such as Scenario 5, the MLGCP fit struggled to identify the repulsion. It also found it challenging to accurately model the magnitude of the attractions as well as the repulsive associations in this scenario.

In Figure 3, we observed the K functions generated for SPIGPP scenario 4, which featured a moderate attraction within species 2, mild repulsion within species 1, and strong repulsion between species $(1, 2)$ (represented by the blue solid line). The right plot in Figure 3 indicates a good fit for species 2, as the blue and green solid lines closely align and within the confidence bands. However, this accuracy was not observed in the other two K functions (left and middle plots in Figure 3), where the repulsion between the two species and within species 1 are inaccurately modeled as attractions by MLGCP model. While it was expected that the MLGCP fit may struggle to capture within-species repulsion, it should theoretically identify between-species repulsion, which was not the case in this scenario.

4 Case Study

In this section, we revisit the South Carolina Savannah river site study conducted in Flint et al. (2022). Studying the spatial patterns of plants is of significant interest to ecologists as it provides a better understanding of the community structure.

Seven different plots of South Carolina Savannah river site were originally created by Bill Good (Good and Whipple, 1982) and several analyses have been conducted thereafter

(Good and Whipple, 1982; Jones et al., 1994; Dixon, 2002; Flint et al., 2022). In this study, we study one of the plots from the original experiment (Figure 4). The dataset can be obtained using the R language (R Core Team, 2019) as `ecespa::swamp` from the `ecespa` package available on CRAN.

The dataset, as shown in Table C.1 in Appendix C, contains four species of trees and another (OT) group of eight additional tree species with their arrangement shown in the Figure 4. There are no known environmental covariates related to this dataset, however the (unmeasured) water level is thought to be important for the spatial distribution. Therefore, we have introduced an artificial horizontal covariate that is proportional to water level for this analysis (Flint et al., 2022).

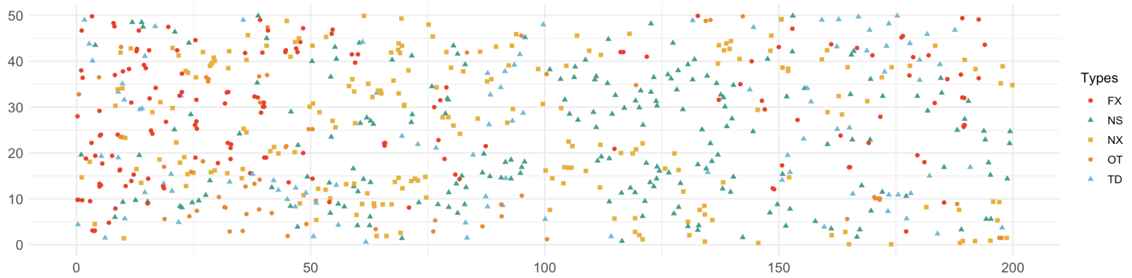
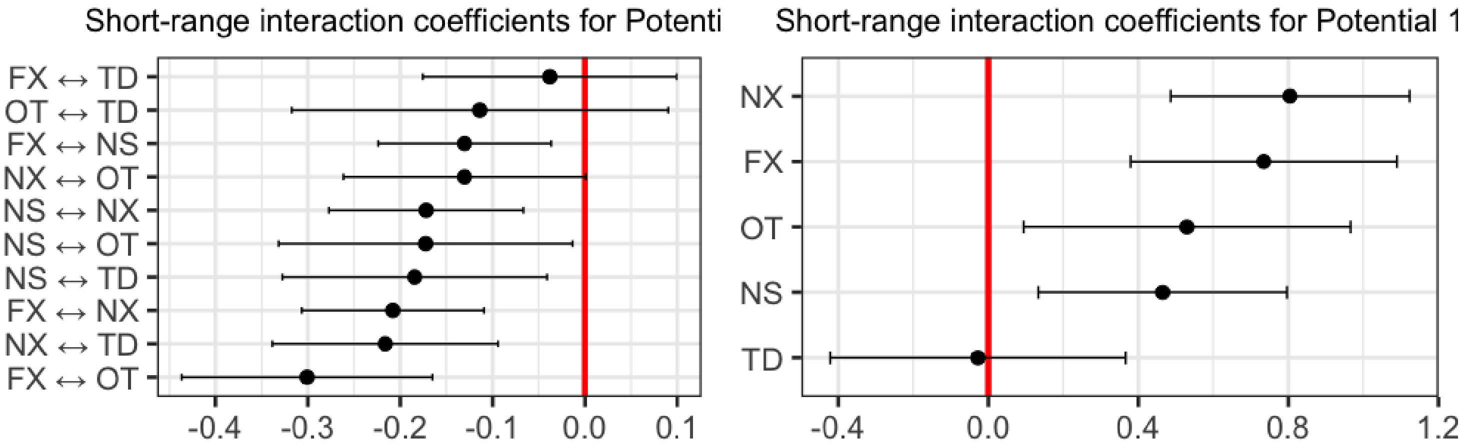


Figure 4: Trees in the Savannah River South Carolina, USA

We also utilise the K functions approximations computed through standard cross K functions methods provided in the ‘spatstat’ R package (Baddeley et al., 2016), where all the effects of covariates and the intensity function are included. This approach enables us to compare the performance of MLGCP and SPIGPP fits using these functions as explained in Section 2.3. The fitting procedure used in the analysis is explained in detail in Appendix C.

The parameters ϕ and σ govern the volatility of the Gaussian random fields in the MLGCP (Table C.3 in Appendix C). The estimates of ϕ_i for tree species are small, with

Carolina Ash having the smallest value and Bald Cypress the largest. The estimates for σ_i are generally small to large depending on the tree species. For example, there is important clustering within Carolina Ash and the other tree category, while the clustering within Swamp Tupelo is the smallest. All other tree species exhibit moderate clustering.



(a) Between tree species

(b) Within tree species

Figure 5: Estimated Short range interaction coefficients for the Tree types of SPIGPP fitted model.

The coefficients and their significance for estimated short-range interactions in the SPIGPP are presented in Figure 5. Notably, most of the coefficients of the short-range interactions (α_p) are found to be statistically significant at 0.05 level of significance. Interaction coefficients for within species are given in the right hand side while the left side shows the between species interaction coefficients. Within species interactions of Bald Cypress, between species interactions of Bald Cypress and Other tree species, Water Tupelo and other species as well as Carolina Ash and Bald Cypress, are the interaction coefficients that were not found to be statistically significant.

The within-species short-range interaction coefficients other than Bald Cypress are all positive, and larger than that of between interaction coefficients, while all between tree species interactions are negative. The smallest repulsion (negative) is between in Carolina Ash and Other tree species (-0.301) and largest between Carolina Ash and Bald Cypress (-0.038). This suggests that similar species of trees tend to occur together more frequently than different species of trees occurring together. Similar findings were reported in the analysis by Flint et al. (2022).

The response to the background intensity estimated from the data is statistically significant for almost all (except for Water Tupelo) of the tree species. It is always positive and this is expected since it captures the general area where trees occur.

	FX	NS	NX	OT	TD
Intercept	-4.60	-3.88	-4.93	-5.64	-5.42
Water level	$-0.88***$	-0.22	$-0.43**$	$-0.88***$	$-0.57**$
Background Intensity	$0.20*$	0.04	$0.23***$	$0.44***$	$0.52***$

Table 1: Significance of covariates in SPIGPP Model

Log-Papangelou conditional intensities (Baddeley et al., 2016; Daley and Vere-Jones, 2003) of a given species in the SPIGPP model, conditional on all other species, are given in Figure C.2 in Appendix C.

The fitted model has effectively captured the spatial inhomogeneity, with its conditional intensity appropriately delineating the area into regions of high and low tree density. The clustering within the points as given in the conditional predictions show similar results as given by $\widehat{\sigma_i}$ of the MLGCP model. The rather large corresponding AUC values for these species [Carolina Ash (0.703), Swamp Tupelo (0.605), Water Tupelo (0.609), Other tree species (0.728) and Bald Cypress (0.679)] corroborate this result.

Figures 6 - 7 display the respective K functions for the fitted models: MLGCP (green) and SPIGPP (blue). We computed the envelopes of the K-function based on simulations from the fitted models of MLGGP and SPIGPP, and they are given in light green and light blue respectively . Additionally, the empirical (purple) and base K (red) functions are shown for comparison.

In Figure 6, we display all the within associations of the five tree species, which show attractions (positive associations). For Carolina Ash (top left) and Other tree species (bottom left) both SPIGPP and MLGCP fit the data well at shorter distances ($<$ 4m). However, at longer distances (4m - 12m), SPIGPP continues to capture species interactions effectively, while MLGCP fails to do so. For Bald Cypress (bottom right), the SPIGPP model gives a better fit compared to the MLGCP model. For Bald Cypress (bottom right), the empirical (purple) K function is zero up until 2m, as trees closer than 2m to each other had been cut down by people at the time of measurement. Unfortunately, none of the models have been able to accurately capture this change in the K functions. However, the SPIGPP is able to well capture the intra-species interaction beyond distance of 2m. For Swamp Tupelo (top middle), the MLGCP model shows a slightly better fit. Both MLGCP and SPIGPP models perform exceptionally well at modeling Water Tupelo (top right).

In Figures 7, the between species associations are presented. Here, we observe that the SPIGPP model provides a better fit than the MLGCP model for most of the between tree associations shown in Figure 7. Most of the repulsive associations/negative associations (top middle, top right graphs, third row graphs, second row middle and right graphs) are either estimated as attractions/positive associations by the MLGCP model or are not accurately identified, defaulting to the baseline K function, while SPIGPP accurately models them. For the top left graph of association between Carolina Ash and Swamp Tupelo for

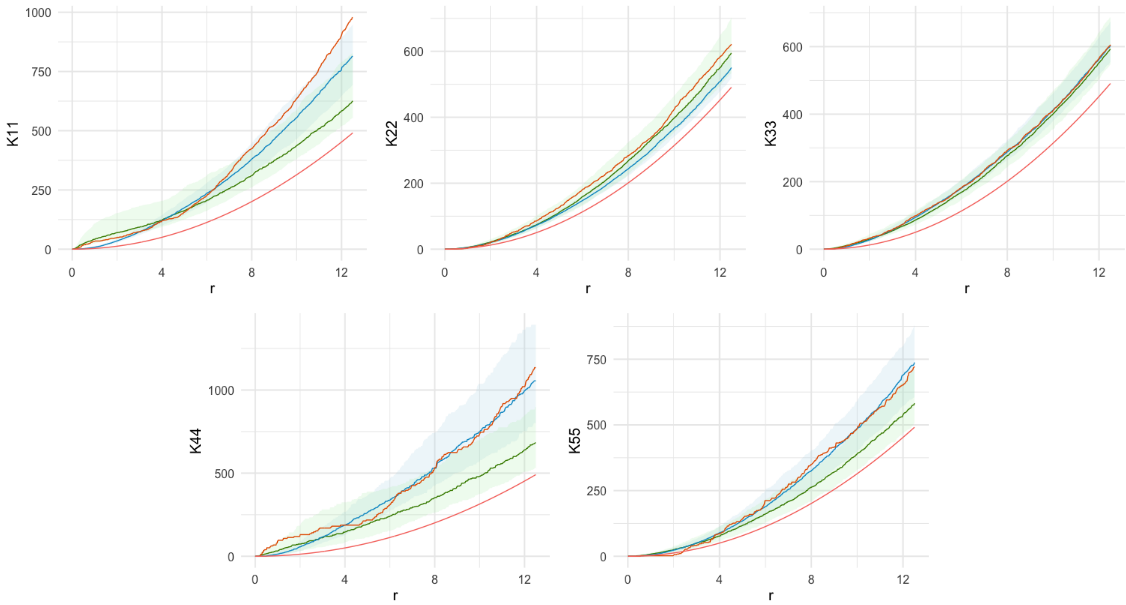


Figure 6: Comparison of fitted estimated K functions of the models using SPIGPP (blue) and MLGCP (green) for the Savannah river study. The empirical K function is given in orange while the red solid line indicates the baseline K function of πr^2. K11 represents the estimated K function of FX and similarly, K22, K33, K44 and K55 represent the estimated K functions of NS, NX, OT and TD respectively.

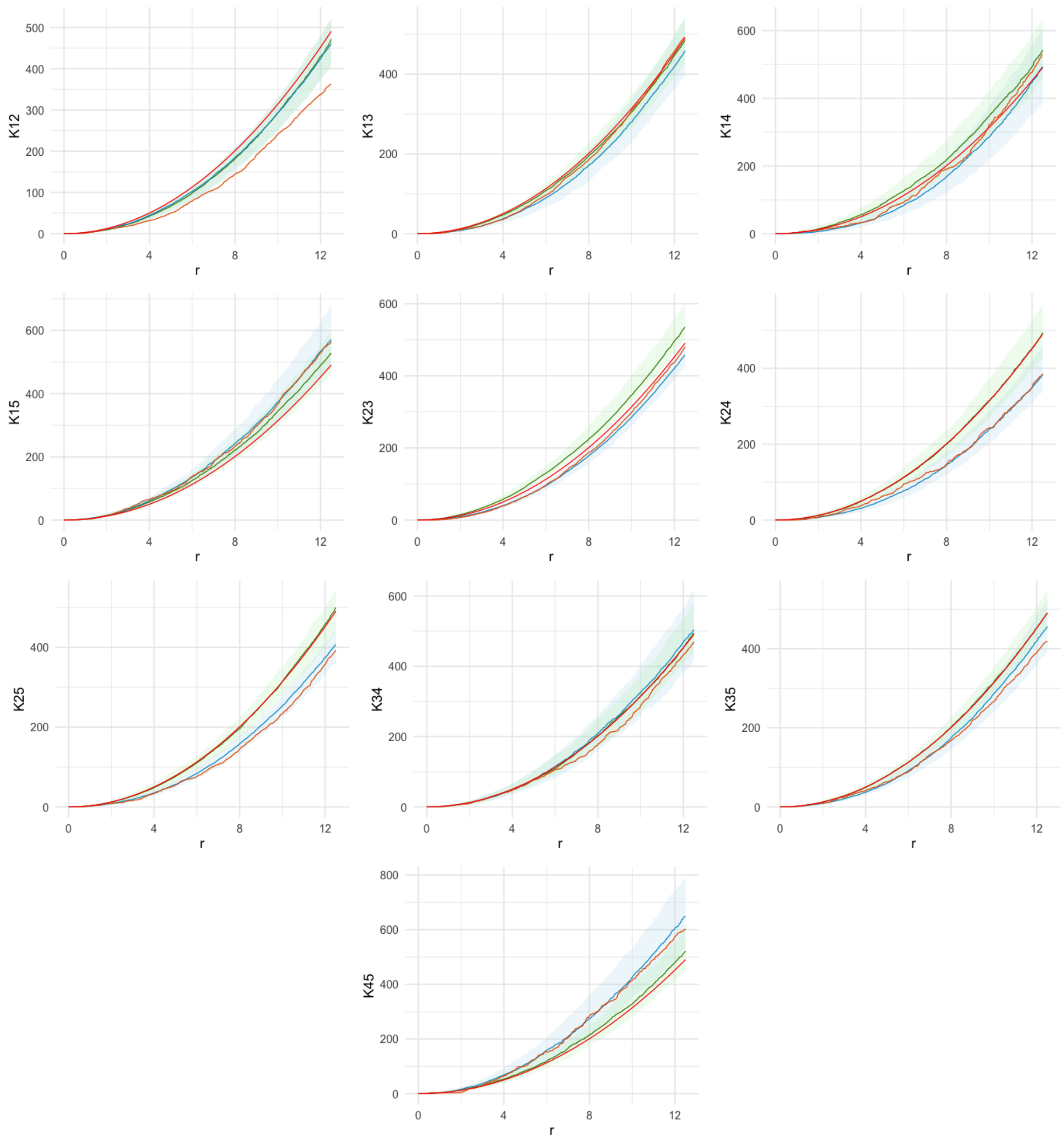


Figure 7: Comparison of fitted estimated K functions of the models using SPIGPP (blue) and MLGCP (green) for the Savannah River study. The empirical K function is given in orange while the red solid line indicates the baseline K function of πr^2. K12, K13, K14, K15, K23, K24, K25, K34 and K45 represent the estimated K functions between (FX,NS), (FX,NX), (FX,OT), (FX,TD),(NS,NX), (NS,OT), (NS,TD), (NX,OT), (NX,TD) and (OT,TD) respectively.

which both MLGCP and SPIGPP fails to capture the maginitude of the repulsion accurately. For the positive associations between Carolina Ash and Bald Cypress (second row left) and Other tree species and Bald Cypress (bottom), the MLGCP model does identify the attraction accurately but fails to estimate the magnitude effectively while SPIGPP accurately estimates the associations.

As shown in Table 2, the MISEs for SPIGPP are much smaller for both within and between species interactions. SPIGPP performs much better at modeling both between and within tree species associations. While the MLGCP models do a fair job of modeling within species associations compared to the baseline, they are not as effective as SPIGPP. As a summary, our findings indicate that the SPIGPP offers a superior fit for the K function compared to the MLGCP model in this case study. Specifically, the Gibbs process more accurately captures the spatial interactions and dependencies present in the data, leading to more reliable and interpret-able results. This improved fit is evident across various distances, highlighting the robustness of the Gibbs process in modeling spatial point patterns.

	SPIPP	MLGCP	Base
$MISE_{total}$	715.76	3861.69	13012.67
$MISE_{within}$	1448.38	8841.47	43036.64
$MISE_{between}$	316.15	1145.44	1003.09

Table 2: MISE of fitted SPIPP and MLGCP models for the Savannah Trees

5 Discussion

In this paper, we specifically focus on the Log Gaussian Cox process proposed by Waagepetersen et al. (2016) and Hessellund et al. (2022a) and the saturated pairwise interaction Gibbs Point process model introduced by Flint et al. (2022). This study is the first comparison of these two models through extensive simulation studies and an illustrative case study, aiming to identify the conditions under which the models excel or fall short.

Based on our simulation study outlined in Sections 3.1, we observe that MLGCP models perform well in scenarios involving mild attractions between and within species. Additionally, MLGCP models maintain a good fit for scenarios with moderate and strong attractions between and within species as well. MLGCP correctly detect positive associations even though they sometimes fail to precisely model the magnitude of attractions. These models excel in cases of mild to moderate repulsive associations between species coupled with mild to moderate attractions within species. However, their performance diminishes in capturing true repulsion when confronted with strong to extremely strong repulsion between species, accompanied by strong attractions within species. Furthermore, MLGCP models can not identify within-species repulsion, as the model is inherently not designed for this aspect.

In contrast, SPIGPP models perform well in scenarios with mild to moderate attractions and/or repulsion between species, along with mild to moderate attractions within species. They particularly excel in modeling repulsive associations between species, spanning from mild to extremely strong. Challenges arise for SPIGPP models when confronted with strong to extremely strong attractions within and between species. Notably, the strong or extremely strong attractions between and/or within species generated from the MLGCPS show considerable fluctuations in the number of points for each species across different realizations. The SPIGPP, however, is designed to model a roughly constant number of points between samples, making it challenging to handle such situations (Baddeley et al., 2016). Thus, SPIGPP models have difficulty in accurately fitting processes with strong and extremely strong attractions between and within species. In spite of this, the SPIGPP models are able to consistently identify the direction of attractions and/or repulsion accurately. A summary of these findings is provided in Table 3, which evaluates the situations in which each model (SPIGPP and/or MLGCP) should be used, considering inter- and intra-species interactions (within and between) and the ground truth.

Scenarios	Fit with MLGCP		Fit with SPIGPP	
	within	between	within	between
mild "+" b/w	good	good	good	good
strong "+" b/w	poor	poor	poor	poor
mild "-" b/w & mild "+" w	good (attraction)	poor	good	good
	-	-	good (repulsion)	-
mild "-" b & mild "+" w	good	good	good	good
strong "-" b & strong"+" w	poor	poor	good	good

Table 3: Summary of comparative simulation study.

Furthermore, based on our investigation into the five-variate LGCP simulation (in Appendix B), we observe that the SPIGPP model accurately identifies attractions and repulsive associations when there are no transitions from attraction to repulsion or vice versa within a single species. However, when there are fluctuations with distance between attractions and/or repulsive associations, the SPIGPP model effectively captures the interaction in short ranges but struggles to accurately represent the transitions in the interaction, while MLGCP tends to capture the attractions at the longer distances. This limitation in SPIGPP may arise from the disparities in the underlying Cox and Gibbs processes between MLGCP and SPIGPP models. We may be able to get a better fit by using `medium_range` and/or `long_range` in SPIGPP.

In our examination of real data, we observe that while the MLGCP models yield adequate results for the within-species associations, they are unable to accurately model the between-species associations. In contrast, the SPIGPP models perform admirably in fitting the data, as evidenced by the low MISE values as well as the estimated conditional predictions shown in Figure C.2 in Appendix C.

When deciding on the use of SPIGPP and MLGCP for fitting data, we can take the following into consideration.

- Gibbs model is suitable when interactions between points are the primary focus. If the intensity of points varies significantly over space and this variation is crucial to your analysis, MLGCPs provide a natural framework for incorporating complex unobserved heterogeneities.
- Gibbs processes often offer more direct interpretability regarding interaction terms. In contrast, MLGCPs, while more flexible and capable of capturing more complex patterns, can sometimes offer less direct interpretability due to the latent Gaussian field.
- Both MLGCP and SPIGPP effectively identify and capture mild to moderate attractions and repulsive associations. MLGCP struggles to capture repulsive associations as they intensify. In contrast, SPIGPP can well estimates both the direction and magnitude of interactions generated by MLGCP. A limitation of SPIGPP, however, is its difficulty in modeling fluctuating interactions that transition between attrac-

tions and repulsive associations. (This may be addressed by fitting more advanced SPIGPP models.)

- SPIGPP is highly effective in handling many species and points, accommodating approximately hundreds of species and up to $\sim 100,000$ points. Such a scale is challenging for MLGCP models, particularly when species has complex correlation structure involving within-species repulsion and attractions/repulsive associations at various distances.

- It is also worthwhile to remember that Rajala et al. (2018) says "For longer spatial scales the log-Gaussian Cox process is a well-suited modelling framework, but it is not a good framework for studying small-scale interactions. Instead we shall use the multivariate Gibbs point process model to discover small scale point-to-point interactions..."

- Ultimately, the choice between models depends on the setting of the scenario. For instance, if there is an expectation of a missing unmeasured covariate distributed as an approximate Gaussian field, MLGCP models are more reliable for inferring missing covariates and explaining clustering as a result of the covariate. In contrast, if interactions between points are not important, SPIGPP would be the preferable option.

6 Conclusions

This paper demonstrates that both MLGCP and SPIGPP excel within their own distinct contexts, despite their unique underlying character. The performance of each model is comparable when dealing with mild to moderate attractions/repulsive associations, as both are proficient in identifying and appropriately capturing these patterns. Notably, SPIGPP models are better at identifying and modeling repulsive associations compared to MLGCP models, while MLGCP models excel at capturing strong attractions. SPIGPP models consistently identify the direction of the interaction type accurately, even when faced with challenges in modeling their magnitude appropriately. A limitation of MLGCP models is their inability to identify repulsive associations as they intensify, often modeling them as attractions.

Acknowledgements

This work was supported by Australian Research Council Grant No. DP190100613.

Conflicts of Interest

There are no conflicted of interest from the authors.

Authors' Contributions

Chathuri L. Samarasekara : Conceptualization (supporting); data curation (lead); formal analysis (lead); methodology (lead); resources (supporting); visualization (lead); writing original draft (lead); writing, review and editing (equal). **Ian Flint** : Conceptualization (lead); data curation (supporting); resources (lead); methodology (lead); supervision (lead); visualization (supporting); writing, review and editing (equal). **Yan Wang** : Conceptualization (lead); data curation (supporting), supervision (lead); methodology (lead); resources (lead); writing, review and editing (equal); funding acquisition (lead).

Data Availability

R scripts used to generate simulated data will be shared at Github, should the manuscript be accepted.

ORCID

Chathuri L. Samarasekara: https://orcid.org/0000-0001-7414-497X

Ian Flint: https://orcid.org/0000-0002-8721-5340

Yan Wang: https://orcid.org/0000-0003-1635-5554

References

Baddeley, A., Jammalamadaka, A., and Nair, G. (2014). Multitype point process analysis of spines on the dendrite network of a neuron. *Journal of the Royal Statistical Society: Series C*, 63:673–694.

Baddeley, A., Rubak, E., and Turner, R. (2016). *Spatial Point Patterns – Methodology and Applications with R*, volume 101(4). CRC Press, 3rd edition.

Choiruddin, A., Cuevas-Pacheco, F., Coeurjolly, J.-F., and Waagepetersen, R. (2020). Regularized estimation for highly multivariate log gaussian cox processes. *Statistics and Computing*, 71(5):1721–1752.

C.J.Geyer (1999). *Likelihood inference for spatial point processes: Likelihood and computation. In Stochastic Geometry: Likelihood and Computation (eds. W. Kendall, O. Barndroff-Nielsen and M. N. van Lieshout).* London: Chapman and Hall/CRC.

Cronie, O. and van Lieshout, M. (2016). Summary statistics for inhomogeneous marked point processes. *Annals of the Institute of Statistical Mathematics*, 68:905–928.

Cronie, O. and van Lieshout, M. (2018). A non-model-based approach to bandwidth selection for kernel estimators of spatial intensity functions. *Biometrika*, 105:455–462.

Daley, D. and Vere-Jones, D. (2003). *An introduction to the theory of point processes, vol. 1.* New York: Probability and its Applications. Springer-Verlag.

Dixon, P. (2002). Nearest-neighbor contingency table analysis of spatial segregation for several species. *Ecoscience*, 9:142–151.

Flint, I., Golding, N., Vesk, P., Wang, Y., and Xia, A. (2022). The saturated pairwise interaction gibbs point process as a joint species distribution model. *Journal of the Royal Statistical Society, series C: Applied Statistics*, 71(5):1721–1752.

Good, B. and Whipple, S. (1982). Tree spatial patterns: South carolina bottomland and swamp forests. *Bulletin of the Torrey Botanical Club*, 109:529–536.

Hessellund, K. B., Xu, G., Guan, Y., and Waagepetersen, R. (2022a). Second-order semi-parametric inference for multivariate log gaussian cox processes. *Journal of the Royal Statistical Society Series C: Applied Statistics*, 71(1):244–268.

Hessellund, K. B., Xu, G., Guan, Y., and Waagepetersen, R. (2022b). Semiparametric multinomial logistic regression for multivariate point pattern data. *Journal of the American Statistical Association*, 117(539):1500–1515.

Jalilian, A., Guan, Y., Mateu, J., and Waagepetersen, R. (2015). Multivariate product-shot-noise cox point process models. *Biometrics*, 71(4):1022–1033.

Jalilian, A., Safari, A., and Sohrabi, H. (2020). Modeling spatial patterns and species associations in a hyrcanian forest using a multivariate log-gaussian cox process. *Journal of Statistical Modelling: Theory and Applications*, (1),(2):59–76.

Jones, R. H., Sharitz, R. R., James, S. M., and Dixon, P. (1994). Tree population dynamics in seven south carolina mixed-species forests. *Bulletin of the Torrey Botanical Club*, 121:360–368.

Møller, J. and Waagepetersen, R. P. (2003). *Statistical Inference and Simulation for Spatial Point Processes*. Chapman and Hall.

Rajala, T., Murrell, D., and Olhede, S. (2018). Detective multivariate interactions in spatial point patterns with gibbs models and variable selection. *Journal of the Royal Statistical Society Series C, Royal Statistical Society*, 67(5):1237–1273.

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

Appendix A : Comparative Simulation Study: Assessing MLGCP and SPIGPP models under various scenarios

In this section, we will delve into the simulation and fitting procedures of the comprehensive two-species simulation study discussed in Section 3.1. We will follow the same format as in Section 3.1 and explain the fitting procedure in two parts: 1) MLGCP Scenarios and 2) SPIGPP Scenarios.

textcolorred(can you describe all the simulation studies using past tense?)

MLGCP Scenarios

All processes under MLGCP Scenarios were confined within a unit square window, featuring a single covariate $Z(.)$ and a background intensity $\rho_0(.) = 150\exp(0.5V(.) - \frac{(0.5)^2}{2})$. Here, the covariate, Z and the background intensity V, were zero-mean unit variance Gaussian random fields with exponential and Gaussian correlation functions with following parameter choices. For all scenarios, $Corr(Z(u), Z(v)) = \exp(\frac{-||u-v||}{0.5})$ and $Corr(V(u), V(v))$ was set to $\exp(-\big(\frac{||u-v||}{0.8}\big)^2)$.

When fitting SPIGPP models to these MLGCP scenarios, we used two different initial distances for `short\_range` distances of 0.5 and 0.8 with `exponential` models. The exponential model here means that the interaction potential is given by $\varphi(x) = \exp(\frac{-\log(2)x}{R})$ where R was the aforementioned `short_range` distance. Moreover, the initial values `min\_dummy = 4000`, `dummy\_factor = 5` and `dummy\_distribution = 'stratified'` were used in all the SPIGPP fits for the scenarios. This meant that the dummy points were distributed as a stratified point process where each species with n number of points had `max(5*n,4000)` dummy points. The `fitting\_package` used was 'glmnet' with a saturation parameter of 4. In the SPIGPP fitting procedure, both the simulated covari-

	α	σ	ϕ	ξ
MLGCP Scenario 1	$\begin{bmatrix} -0.1 & 0.5 \\ 0.1 & -0.5 \end{bmatrix}$	$\begin{bmatrix} 0.5 \\ 0.2 \end{bmatrix}$	$\begin{bmatrix} 0.5 \\ 0.5 \end{bmatrix}$	0.5
MLGCP Scenario 2	$\begin{bmatrix} -0.6 & 0.1 \\ 0.6 & -0.1 \end{bmatrix}$	$\begin{bmatrix} 0.8 \\ 0.6 \end{bmatrix}$	$\begin{bmatrix} 0.5 \\ 0.1 \end{bmatrix}$	$\begin{bmatrix} 0.5 \\ 0.1 \end{bmatrix}$
MLGCP Scenario 3	$\begin{bmatrix} 0.1 & -0.9 \\ -0.1 & 0.9 \end{bmatrix}$	$\begin{bmatrix} 0.5 \\ 0.02 \end{bmatrix}$	$\begin{bmatrix} 0.5 \\ 0.08 \end{bmatrix}$	0.01
MLGCP Scenario 4	$\begin{bmatrix} -1.5 & -0.4 \\ -1.5 & 0.4 \end{bmatrix}$	$\begin{bmatrix} 0.5 \\ 0.02 \end{bmatrix}$	$\begin{bmatrix} 0.5 \\ 0.8 \end{bmatrix}$	0.05

Table A.1: Initial parameter choices of scenarios when simulating with MLGCP

ate and the background intensity were regarded as covariates. The background intensity served as a covariate due to the absence of analogous settings in SPIGPP compared to MLGCP. Incorporating the background intensity allowed for the comprehensive utilisation of available data without any loss of information during the model fitting process. Other initial parameter choices for α, σ, ϕ and ξ are listed in Table A.1.

SPIGPP Scenarios

Similar to the MLGCP scenarios, the SPIGPP processes are generated within a unit window, utilising a shared normalised covariate for both species. In SPIGPP models, the parameters β_0 and α_p jointly control the number of points generated. The expected number of points for each species in all scenarios is set to 100. We employ 10^5 steps in the Metropolis-Hastings algorithm for all scenarios with a saturation parameter of 2. Table A.2 list the initial parameter choices for the simulated SPIGPP scenarios.

	β_0	β	model	short_range	α_p
Scenario 1	$(4.8, 4.5)$	$(1.5, 2)$	exponential	$\begin{bmatrix} 0.05 & 0.05 \\ 0.05 & 0.05 \end{bmatrix}$	$\begin{bmatrix} 0.02 & 0.05 \\ 0.05 & 0.04 \end{bmatrix}$
Scenario 2	$(3.2, 2.2)$	$(1.5, 2)$	square_exponential	$\begin{bmatrix} 0.05 & 0.05 \\ 0.05 & 0.05 \end{bmatrix}$	$\begin{bmatrix} 0.4 & 0.6 \\ 0.6 & 0.9 \end{bmatrix}$
Scenario 3	$(4.8, 4.5)$	$(1.5, 2)$	exponential	$\begin{bmatrix} 0.05 & 0.05 \\ 0.05 & 0.05 \end{bmatrix}$	$\begin{bmatrix} 0.2 & -0.5 \\ -0.5 & 0.4 \end{bmatrix}$
Scenario 4	$(5, 4.1)$	$(-1, 1)$	square_bump	$\begin{bmatrix} 0.05 & 0.05 \\ 0.05 & 0.05 \end{bmatrix}$	$\begin{bmatrix} -0.4 & -0.1 \\ -0.1 & 0.3 \end{bmatrix}$
Scenario 5	$(3.6, 3.5)$	$(1.5, 2)$	exponential	$\begin{bmatrix} 0.05 & 0.05 \\ 0.05 & 0.05 \end{bmatrix}$	$\begin{bmatrix} 0.9 & -0.5 \\ -0.5 & 0.9 \end{bmatrix}$

Table A.2: Initial parameter choices of scenarios when simulating with SPIGPP

When fitting the MLGCP models to the SPIGPP scenarios outlined in Table A.2, we employ the initial values specified in Table A.3 without any regularization ($\lambda = 0$). We also estimate the background intensity (ρ_0) from data using the approach stated in Hessellund et al. (2022a). Then, we move on to compare the fitted models as discussed in section 3.1.

Scenarios	ξ	σ	ϕ	latent
Scenarios 1/2	$(0.05, 0.01)$	$(1, 0.01)$	$(0.05, 0.01)$	1
Scenarios 2	$(0.03, 0.01)$	$(0.8, 0.1)$	$(0.03, 0.01)$	1
Scenario 5	$(0.03, 0.01)$	$(0.8, 0.1)$	$(0.03, 0.01)$	2

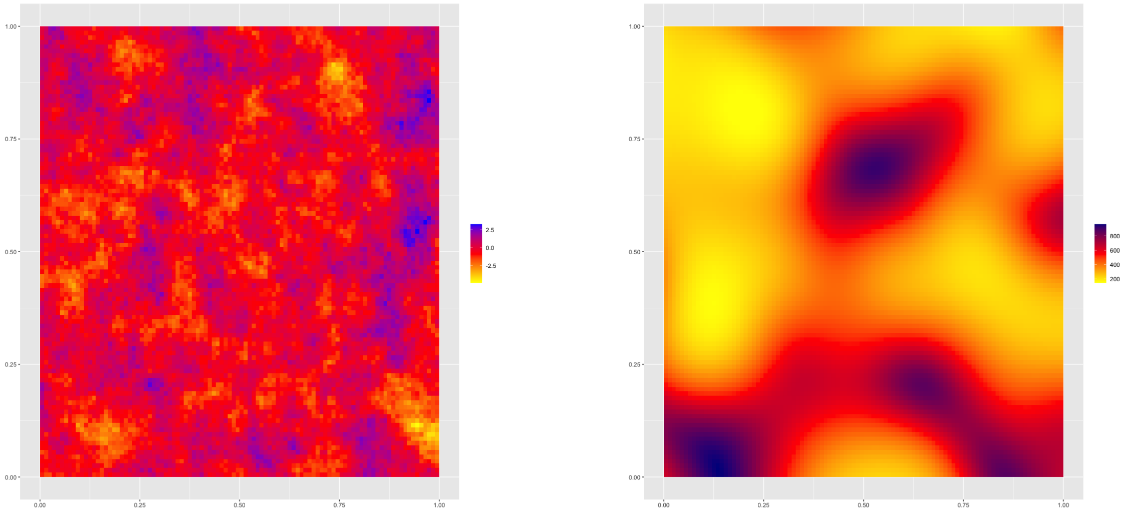
Table A.3: Initial values chosen to fit the generated SPIGPP processes with MLGCP models

Appendix B : Evaluating SPIGPP model performance on Five-variate LGCP

This segment of our simulation follows the study conducted in Waagepetersen et al. (2016), later revisited by Choiruddin et al. (2020); Jalilian et al. (2020), and extensively explored in Hessellund et al. (2022a). Our objective in this phase is to utilise these simulations to gain a comprehensive understanding of the joint performance of SPIGPP models when used with mis-specified models. Additionally, we opt to compare the performance of the SPIGPP fit with the second-order conditional composite likelihood, as outlined in Hessellund et al. (2022a).

We generated a five-variant point process, denoted as $X = (X_1, X_2, X_3, X_4, X_5)^T$, over the spatial domain $W = [0,1]^2$. This simulation is based on two distinct settings, where X is modeled as a multivariate LGCP. We also generate a single covariate $Z(.)$ and a background intensity $\rho_0(.) = 400 \exp(0.5V(.) - \frac{0.5^2}{2})$, where Z and V represent zero-mean unit-variance Gaussian random fields with exponential and Gaussian correlation functions, respectively, as employed in Hessellund et al. (2022a). The realizations of Z and ρ_0 are illustrated in Figure B.1, and these realizations remain constant throughout the entire simulation study.

Table B.1 provides the values used for the intensity function regression parameters γ, along with the standard deviation σ and correlation scale parameters ϕ for the type-specific latent fields Z and V. We set $q = 2$, and $\xi_1 = 0.02$ and $\xi_2 = 0.03$, with $\alpha = \begin{bmatrix} 0.5 & 0.5 & -1 & 0 & 0 \\ -1 & 0 & 0 & 0.5 & 0.5 \end{bmatrix}^T$. In this case, a positive spatial dependence exists between X_1 and X_2, and between X_4 and X_5, while negative spatial dependence is observed between X_3 and (X_1, X_2) and between X_1 and (X_4, X_5).



(a) simulated covariate

(b) simulated ρ_0

Figure B.1: Realizations of covariate and ρ_0 for the simulated five-variant MLGCP

X	γ_0	γ_1	σ	ϕ
X_1	0.1	-0.1	0.71	0.02
X_2	0.2	-0.2	0.71	0.02
X_3	0.3	0	0.71	0.03
X_4	0.4	0.1	0.71	0.03
X_5	0.5	0.2	0.71	0.04

Table B.1: Simulation settings for X in each setup $q = 0, 2$ (excluding α and ξ)

Subsequently, we applied SPIGPP models to the simulated samples from MLGCP, utilising a `Rectangle\_window(c(0,1),c(0,1))` with initial values for the model set to `dummy\_factor = 5`, `min\_dummy = 1000`, `dummy\_distribution = ‘‘stratified"`, and `saturation = 5`. We also use a `fitting\_package` of `‘‘glmnet"` and two potentials for model and `short\_range` interaction distances as given in Table B.2. As detailed previously, our focus lies on using the K functions to assess the model performance. Additionally, we evaluate the performance of each selected model using the mean integrated squared error (MISE) computed based on the K functions.

	model	*short_range*
Potential 1	“exponential”	`matrix(0.05, nrow=5,ncol=5)`
Potential 2	“exponential”	`matrix(0.2, nrow=5,ncol=5)`

Table B.2: Initial Potentials chosen for SPIGPP model parameters of 5 species MLGCP simulation

The empirical and fitted K functions from MLGCP and SPIGPP models for the within-species associations with respective confidence bands are presented in Figure B.2. Out of the five species, the K function of the MLGCP fit (green), and the SPIGPP fit (blue) closely follows that of the empirical K function in purple in species 1, 2 and 5. Both MLGCP and SPIGPP fits show similar deviations from the empirical K functions for species 3 and 4, which is interesting as it is expected for the MLGCP to perform better since the scenario is simulated from MLGCP.

Similar observations apply to the between-species attractions depicted by the cross K functions in Figure B.4. Empirically, these inter-species attractions are relatively smaller compared to the intra-species attractions discussed earlier. The blue solid lines in the figure illustrate that the SPIGPP model adequately captures most of the between-species attractions, denoted as $(1,2),(2,5),(3,4),(3,5),(4,5)$, despite being a mis-specified model while MLGCP fits them better. Notably, between-species attraction $(2,4)$ stand out as slightly larger than the others and SPIGPP fails to capture the magnitude of the attraction accurately, similarly for MLGCP fitting algorithm proposed by (Hessellund et al., 2022b).

The interactions between species $(1,3)$, $(1,4)$, $(1,5)$, and $(2,3)$ exhibit empirical K functions showcasing repulsion initially, transitioning into attraction around $r = 0.1$. In Figure B.3, we demonstrate the close alignment of the SPIGPP model with these empirical dynamics. The blue solid line representing the SPIGPP fit closely overlays the empirical (purple) K functions during the repulsion phase at the beginning as can be seen in Figure B.3, while MLGCP only captures (over-estimates) the attraction between them and is unable to identify the repulsive associations at the beginning.

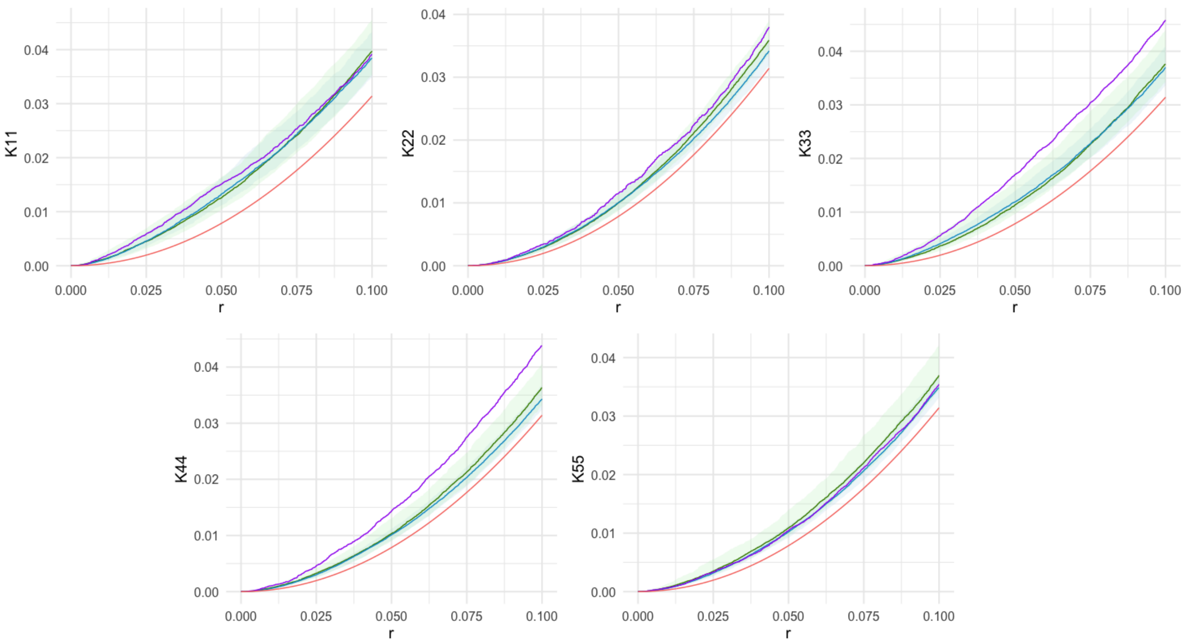


Figure B.2: Comparison of within K functions across simulated five species simulation study. The red line represents the baseline K function (πr^2), while the blue and green lines represent the estimated SPIGPP and MLGCP K functions respectively. The empirical K function derived from the the simulated MLGCP data, is given in purple.

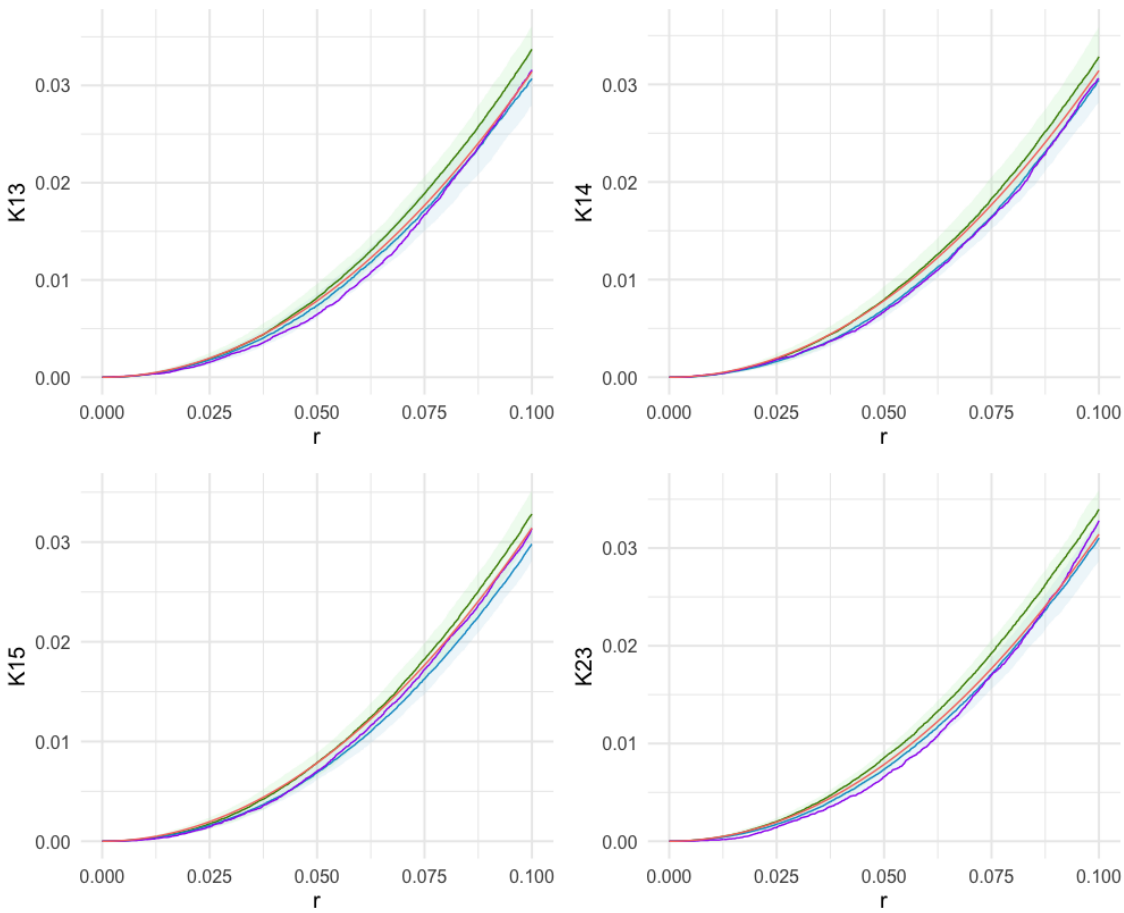


Figure B.3: Comparison of between K functions across simulated five species simulation study. The red line represents the baseline K function (πr^2), while the blue and green lines represent the estimated SPIGPP and MLGCP K functions respectively. The empirical K function derived from the the simulated MLGCP data, is given in purple.

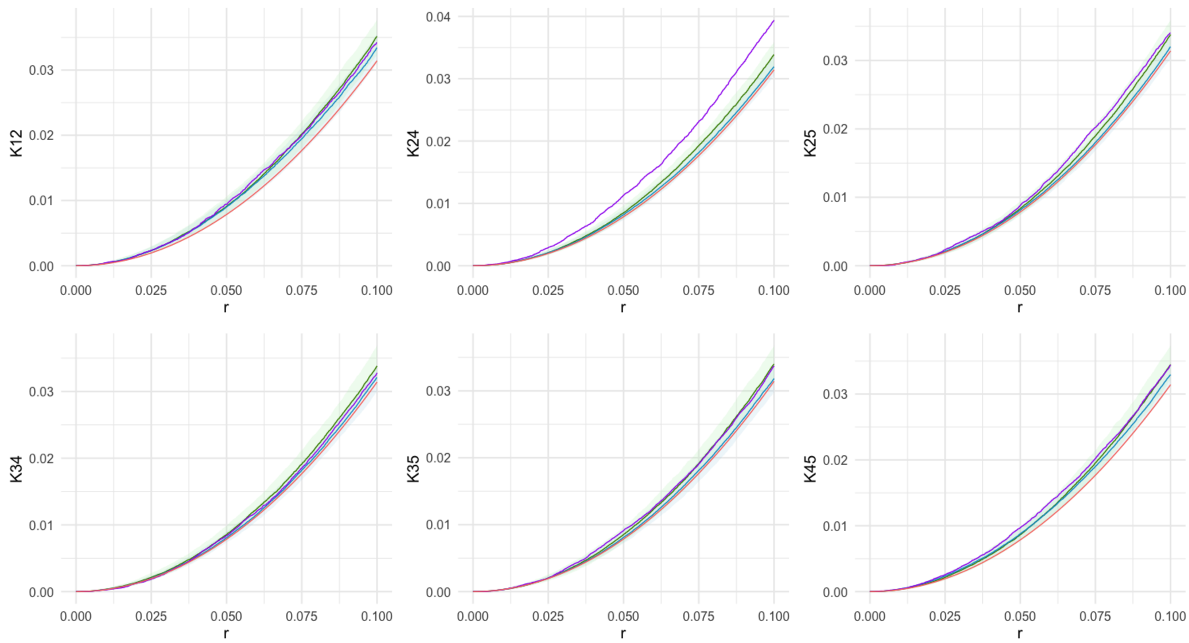


Figure B.4: Comparison of between K functions across simulated five species simulation study. The red line represents the baseline K function (πr^2), while the blue and green lines represent the estimated SPIGPP and MLGCP K functions respectively. The empirical K function derived from the the simulated MLGCP data, is given in purple.

Appendix C: Case Study - Fitting Procedure

In this section, we delve into the details of the fitting procedure used in the case study discussed in Section 3.2.

As displayed in section 3.2, and Table C.1, we can see that the dataset contains four types of trees and another (OT) for a group of eight additional tree species.

Trees	Number
FX - Carolina ash (Fraxinus caroliniana)	156
NS - Swamp tupelo (Nyssa sylvatica)	205
NX - Water tupelo (Nyssa aquatica)	215
OT - stems of 8 additional species	60
TD - Bald cypress (Taxodium distichum)	98

Table C.1: Trees in a plot in the Savannah River South Carolina, USA.

In the MLGCP fitting process, we set $q = 2$ and a regularization parameter $\lambda = 2.5$. For the second-order composite likelihood, a distance parameter of $R = 200$ meters was used with the covariate for water level. Both models assume a rectangle window of $(0, 200) \times (0, 50)$ around the area where the points were distributed.

When fitting SPIGPP models, the covariate for water level and the estimated background intensity ρ_0 was used as covariates to ensure a fair comparison with the MLGCP models fitted. We also use the parameter choices provided in Table C.2.

To estimate ρ_0, we employ the semi-parametric kernel estimator outlined in Section 5 of the supplementary documents in Hessellund et al. (2022a). This involves sub-setting the dataset for each tree type and fitting regression models, incorporating an intercept and the covariate (water level), utilising the function `ppm` in `spatstat` package in R statis-

	model
`short_range`	`matrix(5, 5, 5)`
`model`	`square_exponential`
`dummy_factor`	1
`min_dummy`	5000
`dummy_distribution`	`stratified`
`fitting_package`	`glm`
`saturation`	2

Table C.2: SPIGPP models' initial value choices

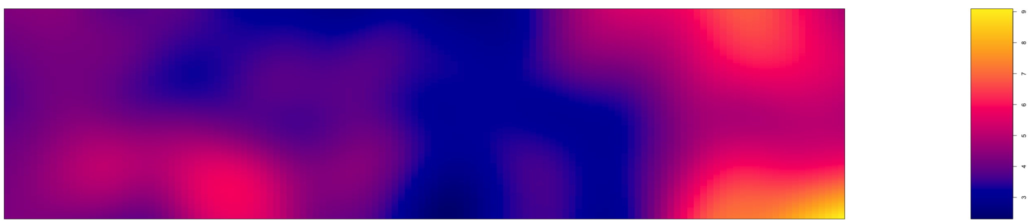


Figure C.1: Estimated background intensity (ρ_0)

tical software. Then the intensity at each tree location is predicted using the `intensity` function. Following this, the density for each tree is computed, incorporating the intensity as weights in the `density` function. In this step, we select the bandwidth based on the criterion inspired by Cronie and van Lieshout (2018), implemented through the `bw.CvL`. Finally, the density across all tree types are averaged to obtain the estimated background intensity, ρ_0 and is shown in Figure C.1.

The estimates derived from the MLGCP model using $(q, \lambda) = (2, 2.5)$ are summarized in Table C.3. The correlation scale parameter estimates for the common latent fields, denoted as ξ, are reported as $(1.44, 21.05)$. Lasso regularization has driven the estimates of the Y_1 latent field, $\widehat{\alpha_{\cdot 1}}$, to 0, similar to the results derived in Hessellund et al. (2022a) while the latent field Y_2 exhibits fluctuations in $\widehat{\alpha_{\cdot 2}}$ from moderate to large. Swamp Tupelo and Water Tupelo respond negatively to Y_2, and they are negatively correlated with Carolina Ash, Bald Cypress and Other tree species.

Tree type	$\widehat{\alpha_{\cdot 2}}$	$\widehat{\sigma}$	$\hat{\phi}$
Carolina Ash	0.565	2.236	0.668
Swamp Tupelo	−0.645	0.660	3.891
Water Tupelo	−0.356	1.401	2.295
Other	0.230	2.0327	1.999
Bald Cypress	0.205	1.091	5.659

Table C.3: MLGCP Parameter estimates for each Tree type for $(q, \lambda) = (2, 2.5)$

The below, α_p matrix supports the result given in Figure 5 in section 3.2. In the estimated α_p matrix, we observed repulsive associations between all species. However, when computing the K functions, we also find a few attractions. These attractions occur between the tree species Carolina Ash and Bald Cypress, and between Other species and Bald Cypress. These K functions show attractions, due to the fact that they were not significant at the 0.05 level of significance when estimating the short_range α_p matrix.

$$\alpha_p = \begin{array}{c} \\ FX \\ NS \\ NX \\ OT \\ TD \end{array} \begin{array}{c} \begin{array}{ccccc} FX & NS & NX & OT & TD \end{array} \\ \begin{pmatrix} 0.735 & -0.130 & -0.208 & -0.301 & -0.0380 \\ -0.130 & 0.465 & -0.172 & -0.172 & -0.184 \\ -0.208 & -0.172 & 0.805 & -0.130 & -0.216 \\ -0.301 & -0.172 & -0.130 & 0.529 & -0.114 \\ -0.038 & -0.184 & -0.216 & -0.114 & -0.027 \end{pmatrix} \end{array}$$

Figure C.2 displays the log-papangelou conditional intensities (Flint et al., 2022) of the tree species in the fitted SPIGPP model (Which is explained in section 3.2).

For the comparison of the two methods, we compute the K functions. To derive these K functions, we simulated 100 fitted MLGCPs using the estimated parameters such as α, σ, ξ, ϕ, estimated βs, and $log(\widehat{\rho_0})$. These simulated fitted MLGCP samples are then used to compute the MLGCP K functions. Similarly, utilising the fitted parameters from each model, we simulated 100 fitted SPIGPP samples to generate the fitted approximate K functions for the SPIGPP models. We then compared the K functions of the fitted models (MLGCP and SPIGPP) with the empirical K function computed from the data.

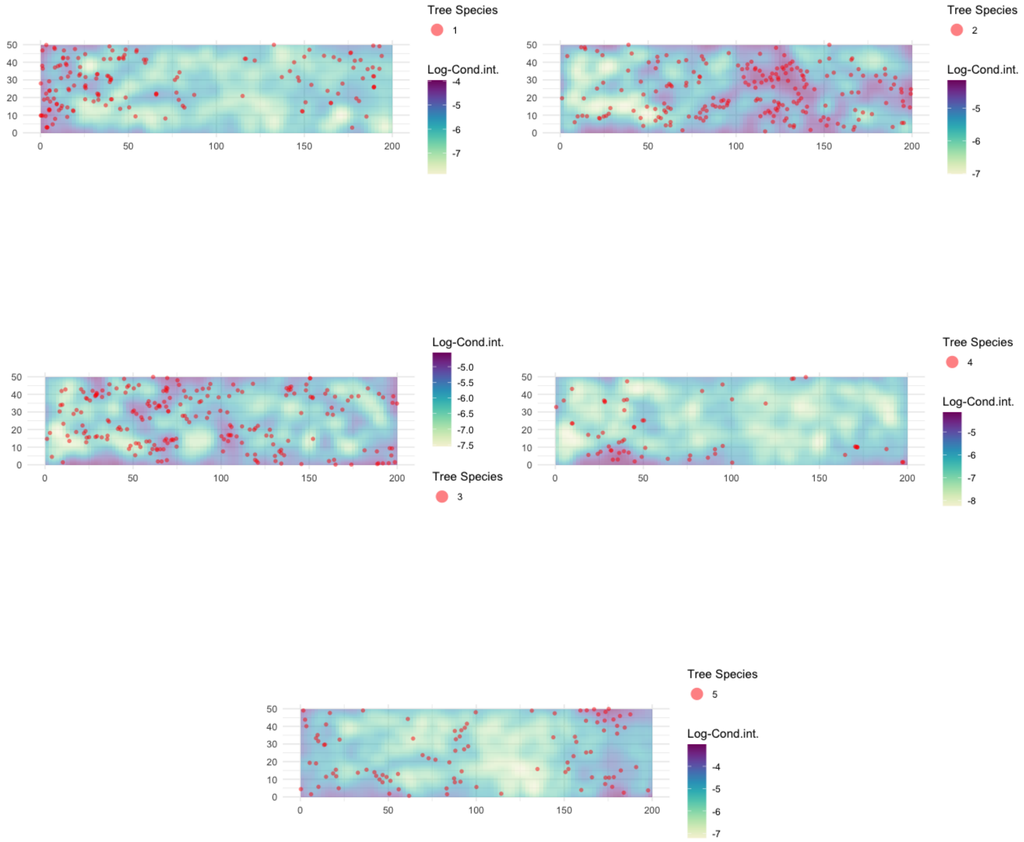


Figure C.2: SPIGPP fitted model - Conditional predictions