Pair correlation functions and limiting distributions of iterated cluster point processes

Jesper Møller and Andreas D. Christoffersen
Department of Mathematical Sciences, Aalborg University

Abstract
We consider a Markov chain of point processes such that each state is a superposition of an independent cluster process with the previous state as its centre process together with some independent noise process and a thinned version of the previous state. The model extends earlier work by Felsenstein and Shimatani describing a reproducing population. We discuss when closed term expressions of the first and second order moments are available for a given state. In a special case it is known that the pair correlation function for these type of point processes converges as the Markov chain progresses, but it has not been shown whether the Markov chain has an equilibrium distribution with this, particular, pair correlation function and how it may be constructed. Assuming the same reproducing system, we construct an equilibrium distribution by a coupling argument.

Keywords: Coupling; equilibrium; independent clustering; Markov chain; pair correlation function; reproducing population; weighted determinantal and permanental point processes.

2010 Mathematics Subject Classification: Primary 60G55; 60J20
Secondary 60D05; 62M30

1 Introduction
This paper deals with a discrete time Markov chain of point processes $G_0, G_1, \ldots$ in the $d$-dimensional Euclidean space $\mathbb{R}^d$, where the chain describes a reproducing population and we refer to $G_n$ as the $n$th generation (of points). We
make the following assumptions. Any point process considered in this paper will be viewed as a random subsets of $\mathbb{R}^d$ which is almost surely locally finite, that is, the point process has almost surely a finite number of points within any bounded subset of $\mathbb{R}^d$ (for measure theoretical details, see e.g. [3] or [16]). Recall that a point process $X \subset \mathbb{R}^d$ is stationary if its distribution is invariant under translations in $\mathbb{R}^d$, and then its intensity $\rho_X \in [0, \infty]$ is the mean number of points in $X$ falling in any Borel subset of $\mathbb{R}^d$ of unit volume. Now, for generation 0, $G_0$ is stationary with intensity $\rho_{G_0} \in (0, \infty)$. Further, for generation $n = 1, 2, \ldots$, conditional on the previous generations $G_0, \ldots, G_{n-1}$, we obtain $G_n$ by four basic operations for point processes:

(a) Independent clustering: To each point $x \in G_{n-1}$ is associated a (non-centred) cluster $Y_{n,x} \subset \mathbb{R}^d$. These clusters are independent identically distributed (IID) finite point processes and they are independent of $G_0, \ldots, G_{n-1}$. The cardinality of $Y_{n,x}$ has finite mean $\beta_n$ and finite variance $\nu_n$ and is independent of the points in $Y_{n,x}$ which are IID, with each point following a probability density function (PDF) $f_n$. We refer to $x + Y_{n,x}$ (the translation of $Y_{n,x}$ by $x$) as the offspring/children process generated by the ancestor/parent $x$, and we let

$$Y_n = \bigcup_{x \in G_{n-1}} (x + Y_{n,x})$$

be the independent cluster process given by the superposition of all offspring processes generated by the points in the previous generation $G_{n-1}$.

(b) Independent thinning: For all $y \in \mathbb{R}^d$, let $B_{n,y}$ be IID Bernoulli variables which are independent of $Y_n, G_0, \ldots, G_{n-1}$, and all previously generated Bernoulli variables. Let $p_n = P(B_{n,y} = 1)$. For all $x \in G_{n-1}$, let

$$W_{n,x} = \{y \in x + Y_{n,x} : B_{n,y} = 1\}$$

be the independent $p_n$-thinned point process of $x + Y_{n,x}$, and let

$$W_n = \bigcup_{x \in G_{n-1}} W_{n,x}$$

be the independent $p_n$-thinned point process of $Y_n$. Note that with probability one, $W_n \cap G_{n-1} = \emptyset$, since by assumption on the cluster points the origin is not contained in $Y_{n,x}$. 

2
(c) Independent retention: For all $x \in \mathbb{R}^d$, let $Q_{n,x}$ be IID Bernoulli variables which are independent of $Y_n, G_0, \ldots, G_{n-1}$, and all previously generated Bernoulli variables. Let $q_n = P(Q_{n,x} = 1)$ and let

$$G_{n-1}^{\text{thin}} = \{ x \in G_{n-1} : Q_{n,x} = 1 \}$$

be the independent $q_n$-thinned point process of $G_{n-1}$.

(d) Independent noise: Let $Z_n \subset \mathbb{R}^d$ be a stationary point process with finite intensity $\rho_{Z_n}$ and which is independent of $W_n, G_0, \ldots, G_{n-1}$, and $G_{n-1}^{\text{thin}}$. Finally, let

$$G_n = W_n \cup G_{n-1}^{\text{thin}} \cup Z_n \quad (3)$$

where we interpret $Z_n$ as noise. For ease of presentation we assume with probability one that $W_n \cup G_{n-1}^{\text{thin}}$ and $Z_n$ are disjoint. Thus $W_n, G_{n-1}^{\text{thin}},$ and $Z_n$ are pairwise disjoint almost surely.

When we later interpret our results, for any point $x \in G_{n-1}^{\text{thin}}$, since $x \in G_{n-1} \cap G_n$, we consider $x$ both as its own ancestor and its own child.

Our model is an extension of the model in Shimatani’s paper [19], which in turn is an extension of Malécot’s model studied in [4] (we return to this in Section 2, item (vii) and (viii)). In particular, our extension allows us to model cluster centres exhibiting clustering or regularity, points from previous generations can be retaining, and the noise processes can also exhibit clustering or regularity (i.e., they are not assumed to be Poisson processes). For statistical applications, we have in mind that $G_n$ may be observable (at least for some values of $n \geq 1$) whilst $G_0$ and the cluster, thinning, and superpositioning procedures in item (a)–(b) and (d) are unobservable. Our model may be of relevance for applications in population genetics and community ecology (see [19] and the references therein), for analyzing tropical rain forest point pattern data with multiple scales of clustering (see [23]), and for modelling proteins with multiple noisy appearances in PhotoActivated Localization Microscopy (PALM) (see [1]). However, we leave it for other work to study the statistical applications of our model and results.

The paper is organized as follows. A discussion of the assumptions in items (a)–(d) and the related literature is given in Section 2. Section 3 focuses on the first and second order moment properties of $G_n$, that is, its intensity and pair correlation function (PCF); we extend model cases and results in Shimatani’s paper [19] and show that tractable model cases for the PCF of $G_0$ are meaningful in terms of Poisson, weighted permanental, and weighted determinantal point processes (which was not observed in [19]). Section 4 discusses limiting cases of the PCF of $G_n$ as $n \to \infty$ when we have the same reproduction system and under weaker conditions than in
In particular, when natural conditions are satisfied, we establish ergodicity of the Markov chain by using a coupling construction and by giving a constructive description of the chain’s unique invariant distribution when extending the Markov chain backwards in time. Finally, Appendix A provides background knowledge on weighted permanental and determinantal point processes, Appendix B verifies some technical details, and Appendix C specifies an algorithm for approximate simulation of the Markov chain’s invariant distribution.

2 Assumptions and related work

Items (i)–(iv) below comment on the model assumptions in items (a)–(d).

(i) The process $Y_n$ is a stationary independent cluster process [3] and we have the following special cases: If $G_{n-1}$ is a stationary Poisson process, $Y_n$ is a Neyman-Scott process [18]. If in addition $\#Y_{n,x}$ follows a Poisson distribution, then $\beta_n = \nu_n$ and $Y_n$ is a shot-noise Cox process (SNCP; [13]) driven by

$$\Lambda_n(x) = \beta_n \sum_{y \in G_{n-1}} f_n(x - y), \quad x \in \mathbb{R}^d. \quad (4)$$

This is a (modified) Thomas process [21] if $f_n$ is the density of $d$ IID zero-mean normally distributed variates with variance $\sigma_n^2$ – we denote this distribution by $N_d(\sigma_n^2)$ – and it is a Matérn cluster process [8, 9] if instead $f_n$ is a uniform density of a $d$-dimensional ball with centre at the origin. However, in many applications a Poisson centre process is not appropriate. For instance, Van Lieshout & Baddeley (2002) considered a repulsive Markov point process model for the centre process, whereby it is easier to identify the clusters than under a Poisson centre process.

(ii) When $\beta_n \leq \nu_n$, we may consider $Y_n$ as a stationary generalised shot-noise Cox process (GSNCP) as studied in [14]: In this model [1] is extended to the case where $G_{n-1}$ is a general stationary point process and $Y_n$ is a Cox process driven by

$$\Lambda_n(x) = \sum_{y \in G_{n-1}} \gamma_y k_n[(x - y)/b_y]/b_y^d, \quad x \in \mathbb{R}^d,$$

where $k_n$ is a PDF on $\mathbb{R}^d$, the $\gamma_y$ and the $b_y$ for all $y \in G_{n-1}$ are independent positive random variables which are independent of $G_{n-1}$, and the $\gamma_y$ are identically distributed with mean $\beta_n$ and variance $\nu_n$ –
\(\beta_n\) (as \(#Y_{n,x}\) has mean \(\beta_n\) and variance \(\nu_n = \text{E}\{\text{var}(#Y_{n,x} | y)\} + \text{var}(\text{E}(#Y_{n,x} | y)) = \beta_n + \text{var}(\gamma_n)\). Further, \(b_y\) has an interpretation as a random band-width and \(f_n(x) = \text{E}\{k_n(x/b_y)/b_y^d\}\). The general results for the intensity and PCF of \(G_n\) in Section 3 will be unchanged whether we consider this stationary GSNCP or the more general case in item (a).

(iii) Clearly, there is no noise (\(Z_n\) is empty with probability one) if \(\rho_{Z_n} = 0\). The case \(\rho_{Z_n} > 0\) may be relevant when not all points in a generation can be described as resulting from independent clustering and thinning as in (a)–(c). Note that in item (d) we could without loss of generality assume \(Z_1, Z_2, \ldots\) are independent. Further, we introduce the thinning of \(Y_n\) in item (b) only for modelling purposes and for comparison with \(\text{[19]}\); from a mathematical point of view this thinning could be omitted if in item (a) we replace each cluster \(Y_{n,x}\) by what happens after the independent thinning: Namely that independent thinned clusters \(Y_{n,x}^{th}\) appear so that \(#Y_{n,x}^{th}\) has mean \(\beta_n^{th} = \beta_np_n\) and variance \(\nu_n^{th} = \beta_n p_n - \beta_n p_n^2 + \nu_n p_n^2\) and is independent of the points in \(Y_{n,x}^{th}\) which are IID with PDF \(f_n\), whereby \(W_n\) and \(Y_{n,x}^{th} : = \bigcup_{x \in G_{n-1}} (x + Y_{n,x}^{th})\) are identically distributed.

(iv) Assuming for \(n = 1, 2, \ldots\) no thinning of \(Y_n\) (\(p_n = 1\)), an equivalent description of items (a) and (c)–(d) is given in terms of the Voronoi tessellation generated by \(G_{n-1}\): For \(x \in G_{n-1}\), let \(C(x | G_{n-1})\) be the Voronoi cell associated to \(x\) and consisting of all points in \(\mathbb{R}^d\) which are at least as close to \(x\) than to any other point in \(G_{n-1}\) (with respect to usual distance in \(\mathbb{R}^d\)). With probability one, since \(G_{n-1}\) is stationary and non-empty, each Voronoi cell is bounded and hence its volume is finite (see e.g. \(\text{[11, 12]}\)). Thus we can set

\[G_n = \bigcup_{x \in G_{n-1}} (x + G_{n,x}) \bigcup G_{n-1}^{\text{thin}}\]

where conditional on \(G_{n-1}\), for all \(x \in G_{n-1}\), the \(G_{n,x}\) are independent of \(G_{n-1}^{\text{thin}}\) and they are IID finite point processes with a distribution as follows: \(#G_{n,x}\) has mean \(\beta_n + |C(x | G_{n-1})|/\rho_{Z_n}\), variance \(\nu_n + |C(x | G_{n-1})|/\rho_{Z_n}\), and is independent of the points in \(G_{n,x}\), where \(|\cdot|\) denotes volume; and the points are IID, each following a mixture distribution so that with probability \(\beta_n/\{\beta_n + |C(x | G_{n-1})|/\rho_{Z_n}\}\) the PDF is \(f_n\) and else it is a uniform distribution on \(C(x | G_{n-1})\).

In items (v)–(vi) below we discuss earlier work on the model for \(G_0, G_1, \ldots\), where \(G_0\) is a stationary Poisson process, all \(G_n = Y_n\) for \(n \geq 1\) (i.e., no thin-
ning, no retention, and no noise), \( f_n = f \) and \( \beta_n = \beta \) do not depend on \( n \geq 1 \). We may refer to this as a replicated SNCP. Frequently in the literature, a so-called replicated Thomas process is considered, that is, \( f \sim N_d(\sigma^2) \).

(v) Apparently this replicated SNCP was originally studied by Malécot, see the discussion and references in Felsenstein’s paper [4] where the following three conditions are stated: “(I) individuals are distributed randomly on the line with equal expected density everywhere; (II) each individual reproduces independently, the number of offspring being drawn from a Poisson distribution with a mean of one; and (III) each offspring migrates independently, the displacements being drawn from some distribution \( m(x) \), which we will take to be a normal distribution.” (In our notation, \( d = 1 \), \( \beta = 1 \), and \( f \sim N_1(\sigma^2) \), but [4] considered also more general offspring densities \( f \) and the cases \( d = 2, 3 \).) [4] showed that “(I) is incompatible with (II)–(III)” because \( G_1, G_2, \ldots \) are not stationary Poisson processes and “a model embodying (II) and (III) will lead to the formation of larger and larger clumps of individuals separated by greater and greater distances”, and then he concluded “This model is therefore biologically irrelevant”.

(vi) Kingman in [5] considered the case where \( \beta \) is replaced by a non-negative function \( b \) which is allowed to depend on the cluster centre \( x \) and the previous generation, so a cluster with centre \( x \) is a Poisson process with intensity function \( b(x, G_{n-1})f(\cdot - x) \); e.g., as in the Voronoi case discussed in item (iv), \( b(x, G_{n-1}) \) may depend on \( G_{n-1} \) in a neighbourhood of \( x \). Then \( G_n \) is a Cox process: \( G_n \) conditional on \( G_{n-1} \) is a Poisson process with intensity function

\[
\Lambda_n(x) = \sum_{y \in G_{n-1}} b(y, G_{n-1})f(x-y), \quad x \in \mathbb{R}^d. \tag{5}
\]

In this setting [5] verified that it is impossible for \( G_n \) to be a stationary Poisson process, however, replacing \( f(x-y) \) in (5) by a more general density which may depend on \( G_{n-1} - x \), [5] noticed that it is possible for \( G_n \) to be a stationary Poisson process. A trivial example is the Voronoi case in item (iv) when \( G_n = Z_n \) for \( n \geq 1 \).

Recently, Shimantani in [19] considered first the case of items (a)–(b) and no noise, when \( d = 2 \) and there is the same reproduction system so that \( f_n = f, \beta_n = \beta > 0, \nu_n = \nu, \) and \( p_n = p \in (0, 1] \) do not depend on \( n \geq 1 \).

(vii) In particular, [19] considered the case \( f \sim N_2(\sigma^2) \) and when \( \beta p = 1 \) or equivalently when the intensities \( \rho_{G_0} = \rho_{G_1} = \ldots \) are invariant
over generations, and then he showed that as \( n \to \infty \), the PCF for \( G_n \) diverges and “all offspring will eventually have the same ancestor”. It follows from item (iii) that the model is equivalent to a replicated Neyman-Scott process; this becomes a replicated Thomas process when each cluster size is Poisson distributed, and hence the result in [19] agrees with the results in [4] and [5]. Note that [19] implicitly assumed that a cluster can have more than one point. Otherwise the PCF of \( G_n \) becomes equal to 1; we discuss this rather trivial case again in Section 3.2.2 and [1], see also Section 3 in [5].

Next, Shimantani in [19] extended the model by including noise as in item (d) and by making the following assumptions: The noise processes \( Z_n \) are stationary Poisson processes, satisfying \( 0 < \rho_{Z_1} = \rho_{Z_2} = \ldots \) and \( \rho_{G_0} = \rho_{G_1} = \ldots \), meaning that \( \beta p \leq 1 \). As there is no noise if \( \beta p = 1 \) it is also assumed that \( \beta p < 1 \).

(viii) Then [19] showed that the PCF of \( G_n \) converges uniformly as \( n \to \infty \) and he argued that this limiting case may be “biologically valid” [19, Section 2.4]. However, there are some flaws in [19] which we address:

- He did not show that there exists an underlying point process having this limiting case as its PCF, although he claimed that “this modified replicated Neyman-Scott process reaches an equilibrium state”. In Section 4 for our more general model, we prove the existence of such an underlying point process.
- When \( G_0 \) is not a stationary Poisson process but its PCF is of a particular form (which we specify later in connection to (9)), he did not argue that there exists an underlying point process and what it could be. In Section 3 we verify this existence under our more general model.

Finally, we remark on a few related cases.

(ix) Whilst we study the processes \( G_n \) for all \( n = 1, 2, \ldots \), often in the spatial point process literature the focus is on either \( G_1 \) or \( G_2 \), assuming \( p_n = 1 \) and \( \rho_{Z_n} = 0 \) for \( n = 1 \) or \( n = 1, 2 \), respectively. [23] studied this in the special case of a double Thomas cluster process \( G_2 \) when \( d = 2 \), i.e., when \( G_0 \) is a stationary Poisson process, (4) holds for both \( G_1 = Y_1 \) and \( G_2 = Y_2 \), and \( f_n \sim N_2(\sigma_n^2) \) for \( n = 1, 2 \); see also [1] for more general functions \( f_n \). Moreover, [23] extended the double Thomas process to the case where \( \rho_{Z_1} = 0 \) and \( \rho_{Z_2} > 0 \); this type of model is also considered in [1]. In any case, our general results for intensities and PCFs in Section 3 will cover all these cases.
If for each generation we assume no thinning \((p_1 = p_2 = \ldots = 1)\), no noise \((\rho Z_1 = \rho Z_2 = \ldots = 0)\), no retention \((q_1 = q_2 = \ldots = 0)\) as well as \(\beta_1 = \beta_2 = \ldots \) and \(f_1 = f_2 = \ldots\), then the superposition \(\bigcup_{n=0}^{\infty} G_n\) is known as a spatial Hawkes process, see [15] and the references therein.

3 First and second order moment properties

In this section we determine the intensity and the PCF of \(G_n\) for \(n = 1, 2, \ldots\), under more general assumptions than in Shimatani’s paper [19]. Specifically, points from one generation can be retained in the next generation, the noise is an arbitrary stationary point process (not necessarily a stationary Poisson process as in [19]), and we do not assume the same reproduction system.

3.1 Intensities

By induction \(G_n\) is seen to be stationary for \(n = 0, 1, \ldots\). Its intensity is determined in the following proposition where for notational convenience we define \(Z_0 = G_0\) so that \(\rho Z_0 = \rho G_0\).

**Proposition 3.1.** For \(n = 1, 2, \ldots\), we have that \(G_n\) is stationary with a positive and finite intensity given by

\[
\rho_{G_n} = \rho_{G_{n-1}}(\beta_n p_n + q_n) + \rho Z_n = \rho Z_n + \sum_{i=0}^{n-1} \rho Z_i \prod_{j=i+1}^{n} (\beta_j p_j + q_j). \tag{6}
\]

**Proof.** Using induction for \(n = 1, 2, \ldots\), the proposition follows immediately from items (a)–(d), where the term \(\rho Z_i \prod_{j=i+1}^{n} (\beta_j p_j + q_j)\) is the contribution to the intensity caused by the clusters with centres \(Z_i\) and after applying the two types of independent thinnings. \(\square\)

3.2 Pair correlation functions

3.2.1 Preliminaries

Recall that a stationary point process \(X \subset \mathbb{R}^d\) with intensity \(\rho_X \in (0, \infty)\) has a translation invariant PCF (pair correlation function) \((u, v) \rightarrow g_X(u - v)\) with \((u, v) \in \mathbb{R}^d \times \mathbb{R}^d\) if for any bounded Borel function \(h : \mathbb{R}^d \times \mathbb{R}^d \rightarrow [0, \infty)\) with compact support,

\[
\mathbb{E} \sum_{x_1, x_2 \in X : x_1 \neq x_2} h(x_1, x_2) = \rho_X^2 \int \int h(x_1, x_2) g_X(x_1 - x_2) \, dx_1 \, dx_2 < \infty. \tag{7}
\]
Equivalently, for any bounded and disjoint Borel sets $A, B \subset \mathbb{R}^d$, denoting $N(A)$ the cardinality of $X \cap A$, the covariance between $N(A)$ and $N(B)$ exists and is given by
\[
\text{cov}\{N(A), N(B)\} = \rho_X^2 \int_A \int_B \{g_X(x_1 - x_2) - 1\} \, dx_1 \, dx_2.
\]

Some remarks are in order. Note that $g_X$ is uniquely determined except for nullsets with respect to Lebesgue measure on $\mathbb{R}^d$, but we ignore such nullsets in the following. Thus the translation invariance of the PCF is implied by the stationarity of $X$. Our results below are presented in terms of the reduced PCF $g_X - 1$ rather than $g_X$, and $g_X = 1$ if $X$ is a Poisson process. It is convenient when $g_X$ is isotropic, meaning that there is a function $g_{X,o}$ so that for all $x \in \mathbb{R}^d$, $g_X(x) = g_{X,o}(\|x\|)$ depends only on $x$ through $\|x\|$. With a slight abuse of terminology, we also refer to $g_X$ and $g_{X,o}$ as PCFs.

For a PDF $h$ on $\mathbb{R}^d$, let $\tilde{h}(x) := h(-x)$ and let
\[
h * \tilde{h}(x_1 - x_2) = \int h(x_1 - y)h(x_2 - y) \, dy
\]
be the convolution of $h$ and $\tilde{h}$. Note that if $U$ and $V$ are IID random variables with PDF $h$, then $U - V$ has PDF $h * \tilde{h}$. In the following section we consider the case
\[
g_X - 1 = a \ h * \tilde{h}
\]
for real constants $a$, where $X$ in particular, may refer to the initial generation process, $G_0$, or the noise process, $Z_n$. This corresponds to $X$ being a Poisson process if $a = 0$, a point process with positive association between its points (attractiveness, clustering, or clumping) if $a > 0$, and a point process with negative association between its points (repulsiveness or regularity) if $a < 0$. In [19], for the initial generation process $G_0$, Shimatani briefly discussed the special case of (9) when $h \sim N_2(\tau^2/2)$ (so $h * \tilde{h} \sim N_2(\tau^2)$) whilst the noise processes are stationary Poisson processes. However, if $a \neq 0$ he did not argue if an underlying point process with PCF $g_X$ exists. Indeed, as detailed in Appendix A there exist $\alpha$-weighted determinantal point processes satisfying (9) if $\alpha = -1/a$ is a positive integer, and there exist Cox processes given by $\alpha$-weighted permanental point processes satisfying (9) if $\alpha = 1/a$ is a positive half-integer. Additionally, $h$ needs not to be Gaussian when dealing with weighted determinantal and permanental point processes; e.g. $h$ may be the density of a normal-variance mixture distribution [2]. Also generalized shot-noise Cox processes [15] have PCFs of the form (9) with $a > 0$. Moreover, (9) holds for many other cases of point process models for
If the Fourier transform $\mathcal{F}(g_X - 1)$ is well-defined and non-negative, if $h = \tilde{h}$, and if $a := \int (g_X - 1) \in (0, \infty)$, then (9) holds with

$$h = \mathcal{F}^{-1}\{\sqrt{\mathcal{F}(g_X - 1)}\}/\sqrt{a}$$

provided this inverse transform is well-defined.

The following lemma is needed in Section 3.2.3.

**Lemma 3.2.** Suppose $g_X$ is of the form (9). Then

$$\iint \{g_X(x_1 - x_2) - 1\} f(u - x_1) f(v - x_2) \, dx_1 \, dx_2 = ah \ast \tilde{h} \ast f \ast \tilde{f}(u - v)$$

for any integrable real function $f$ defined on $\mathbb{R}^d$ and for any $u, v \in \mathbb{R}^d$.

**Proof.** Follows from (8) and (9) using Fubini’s theorem and the fact that the convolution operation is commutative and associative.

### 3.2.2 First main result

This section concerns our first main result, Theorem 3.3, which is verified in Section 3.2.3. We use the following notation. Define

$$c_n = \mathbb{E}\{\#Y_{n,x}(\#Y_{n,x} - 1)\}/\beta_n^2 = (\nu_n + \beta_n^2 - \beta_n)/\beta_n^2 \quad \text{if } \beta_n > 0, \quad (10)$$

with $c_n = 0$ if $\beta_n = 0$. If $\beta_n = \nu_n > 0$, as in the case when $\#Y_{n,x}$ follows a (non-degenerated) Poisson distribution, then $c_n = 1$. The case of overdispersion (underdispersion), that is, $\nu_n > \beta_n$ ($\nu_n < \beta_n$) corresponds to $c_n > 1$ ($c_n < 1$). Denote by $\delta_0$ the Dirac delta function defined on $\mathbb{R}^d$. Recall that for any integrable real function $h$ defined on $\mathbb{R}^d$, $h * \delta_0 = \delta_0 * h = h$, and for any $a \in \mathbb{R}$, $a \delta_0 * a \delta_0 = a^2 \delta_0$, where we understand $0 \delta_0$ as 0. Finally, let $h_1 * \cdots * h_n$ where each $h_i$ is either of the form $a_i \delta_0$, with $a_i$ a real constant, or it is an integrable real function defined on $\mathbb{R}^d$.

**Theorem 3.3.** Suppose $g_{G_0}$ and $g_{G_Z}$ are of the form (9), that is, $g_{G_0} - 1 = a f_0 \ast \tilde{f}_0$ and $g_{G_Z} - 1 = b_n f_{Z_n} \ast \tilde{f}_{Z_n}$ for $n = 1, 2, \ldots$. Then, for all $u \in \mathbb{R}^d$ and $n = 1, 2, \ldots$,

$$g_{G_n}(u) - 1 = \left(\frac{\rho_{G_0}}{\rho_{G_n}}\right)^2 a f_0 \ast \tilde{f}_0 \ast \hat{\ast}_{i=1}^n \left\{ (\beta_i p_i f_i + q_i \delta_0) \ast (\beta_i p_i \tilde{f}_i + q_i \delta_0) \right\}(u)$$

(11)
\[
+ \sum_{i=1}^{n} \frac{\rho_{G_{i-1}}}{\rho_{G_n}} \left\{ c_i(b_i p_i)^2 f_i * \tilde{f}_i + \beta_i p_i q_i (f_i + \tilde{f}_i) \right\} \tag{12}
\]
\[
* \prod_{j=i+1}^{n} \left\{ (\beta_j p_j f_j + q_j \delta_0) * (\beta_j p_j \tilde{f}_j + q_j \delta_0) \right\}(u)
\]
\[
+ \sum_{i=1}^{n-1} \left( \frac{\rho_{Z_i}}{\rho_{G_n}} \right)^2 b_i f_{Z_i} * \tilde{f}_{Z_i} \tag{13}
\]
\[
* \prod_{j=i+1}^{n} \left\{ (\beta_j p_j f_j + q_j \delta_0) * (\beta_j p_j \tilde{f}_j + q_j \delta_0) \right\}(u)
\]
\[
+ \left( \frac{\rho_{Z_n}}{\rho_{G_n}} \right)^2 b_n f_{Z_n} * \tilde{f}_{Z_n}(u) \tag{14}
\]

where the \((n-i)\)-th fold convolution in (12) is interpreted as 1 if \(i = n\) and the sum in (13) is interpreted as 0 if \(n = 1\).

The terms in (11)-(14) have the following interpretation when \(u \neq 0\): The right side of (11) corresponds to pairs of \(n\)-th generation points with different 0-th generation ancestors; the \(i\)-th term in (12) corresponds to pairs of \(n\)-th generation points when they have a common \((i-1)\)-th generation ancestor initiated by \(Z_{i-1}\) if \(i > 1\) or by \(G_0\) if \(i = 1\); the \(i\)-th term in (13) corresponds to pairs of \(n\)-th generation points with different \(i\)-th generation ancestors initiated by \(Z_i\); and (14) corresponds to point pairs in \(Z_n\).

Later in Section 4.1, our main interest is in the behaviour of \(g_{G_n}\) as \(n \to \infty\) when we have the same reproduction system, but for the moment, it is worth noticing the flexibility of our model for \(G_1\) and the effect of the choice of its centre process \(G_0\): For simplicity, suppose there is no noise and no retention of points from one generation to the next (i.e., \(\rho_{Z_n} = q_n = 0\) for \(n = 1, 2, \ldots\)), and \(G_0\) is stationary and either a Poisson or an weighted determinantal or permanental point process with a Gaussian kernel. Specifically, suppose \(d = 2\), \(G_0\) has intensity \(\rho_{G_0} = 100\), and using a notation as in Appendix A, the Gaussian kernel has an auto-correlation function of the form \(R(x) = \exp(-\|x/\tau\|^2)\), where the value of \(\tau\) depends on the type of process: For the \(\alpha\)-weighted determinantal point process, we consider the most repulsive case, that is, a determinantal point process (\(\alpha = 1\)) and \(\tau = 1/\sqrt{\rho_{G_0}}\) is largest possible to ensure existence of the process [6]; for the \(\alpha\)-weighted permanental point process, \(\alpha = 1/2\) (the most attractive case when it is also a Cox process, see Appendix A) and \(\tau = 0.1\) is an arbitrary value (any positive number can be used). Note that \(R^2 = (\sqrt{\pi} \tau)^2 f_0 * \tilde{f}_0\) where \(f_0 \sim N_2(\tau^2/8)\), which by (30) and (31) mean that (9) is satisfied with \(a = 2(\sqrt{\pi} \tau)^2\) and \(a = -2(\sqrt{\pi} \tau)^2\) for the weighted permanental and determinantal point processes, respectively, and \(a = 0\) in case of the Poisson process. Moreover, let
the number of points in a cluster be Poisson distributed with mean $\beta_1 = 10$, $p_1 = 1$, and $f_1 \sim N_2(\sigma^2)$, with $\sigma = 0.01$. Then, by Theorem 3.3

$$g_{G_1}(u) - 1 = \frac{a}{\sqrt{2\pi(2\sigma^2 + \tau^2/4)}} \exp \left\{ -\frac{||u||^2}{2(2\sigma^2 + \tau^2/4)} \right\}$$

$$+ \frac{1}{\rho_{G_0} \sqrt{4\pi\sigma^2}} \exp \left( -\frac{||u||^2}{4\sigma^2} \right).$$

Figure 1 shows the isotropic PCF $g_{G_1,o}(r) = g_{G_1}(u)$ as a function of the inter-point distance $r = ||u||$ in case of each of the three models of $G_0$, where using an obvious notation, $g_{G_1,o}^{\text{det}} < g_{G_1,o}^{\text{Pois}} < g_{G_1,o}^{\text{wper}}$. Most notable is the fact that $g_{G_1,o}^{\text{det}}(r)$ exhibits repulsion at midrange distances $r$. For $g_{G_1,o}^{\text{wper}}$, we see a high degree of clustering, which is persistent for large values of $r$; this will of course be even more pronounced if we increase the value of $\tau$; whilst decreasing $\sigma$ will increase the peak at small values of $r$. Figure 2 shows simulations of $G_1$ in each of the three cases of the model of $G_0$. As expected, we clearly see a higher degree of repulsion when $G_0$ is a determinantal point process (the left most plot) and a higher degree of clustering when $G_0$ is a weighted permanental point process (the right most plot). In particular, the clusters are more distinguishable when $G_0$ is a determinantal point process, and this will be even more pronounced if decreasing $\sigma$ because the spread of clusters then decrease. When $G_0$ is a weighted permanental point process, the clusters overlap more.

Figure 1: The PCFs of $G_1$ when $G_0$ is a determinantal, Poisson, or weighted permanental point process (dashed, solid, and dotted respectively), with parameters and Gaussian offspring PDF as specified in the text. The solid horizontal line is the PCF for a Poisson process.
Figure 2: Simulations of $G_1$ restricted to a unit square when $G_0$ is a determinantal (left panel), Poisson (middle panel), or weighted permanental (right panel) point process, see Figure 1 and the text.

3.2.3 Proof of Theorem 3.3

Shimatani in [19] verified Theorem 3.3 in the special case where $q_1 = q_2 = \cdots = 0$, $b_1 = b_2 = \cdots = 0$ (as is the case if $Z_1, Z_2, \ldots$ are stationary Poisson processes), and $c_1 = c_2 = \cdots > 0$, in which case the terms in (13)–(14) are zero. If $c_1 = c_2 = \cdots = 0$, then (12) is zero and by (10), with probability one, $\# Y_{n,x}$ is in the set $\{0, 1\}$ for all $x \in G_{n-1}$ and $n = 1, 2, \ldots$. Consequently, the proof of Theorem 3.3 is trivial if $c_1 = c_2 = \cdots = 0$ and both $G_0$ and $Z_1, Z_2, \ldots$ are stationary Poisson processes, because then $a = 0$, $b_1 = b_2 = \cdots = 0$, $G_1, G_2, \ldots$ are stationary Poisson processes, and the class of stationary Poisson processes is closed under IID random shifts of the points, thinning, and superposition. The general proof of Theorem 3.3 follows by induction from the following lemma together with Lemma 3.2.

**Lemma 3.4.** If $\rho G_{n-1} > 0$, $\rho G_n > 0$, and $g_{G_{n-1}}$ and $g_{Z_n}$ exist, then $g_{G_n}$ exists and is given by

$$g_{G_n}(u - v) - 1 = \left(\frac{\rho G_{n-1} \beta B_n}{\rho G_n}\right)^2 \left[\int \int \left\{ g_{G_{n-1}}(x_1 - x_2) - 1 \right\} f_n(u - x_1) f_n(v - x_2) \, dx_1 \, dx_2 + \frac{c_n}{\rho G_{n-1}} f_n * \tilde{f}_n(u - v) \right]$$

$$+ \left(\frac{\rho G_{n-1} q_n}{\rho G_n}\right)^2 \left\{ g_{G_{n-1}}(u - v) - 1 \right\} + \left(\frac{\rho Z_n}{\rho G_n}\right)^2 \left\{ g_{Z_n}(u - v) - 1 \right\}$$

$$+ \frac{\rho^2_{G_{n-1}} \rho G_n \rho G_n}{\rho^2 G_n} \int \left\{ g_{G_{n-1}}(v - x) - 1 \right\} \left\{ f_n(u - x) + \tilde{f}_n(u - x) \right\} \, dx$$
for any \( u, v \in \mathbb{R}^d \).

**Proof.** Note that \( Y_n \) is stationary with intensity

\[
\rho_{Y_n} = \rho_{G_{n-1}} \beta_n. \tag{15}
\]

It follows straightforwardly from (1), (7), and Fubini’s theorem that its PCF is given by

\[
\rho_{Y_n}^2 g_{Y_n}(u - v) = \rho_{G_{n-1}}^2 \beta_n^2 \int \int g_{G_{n-1}}(x_1 - x_2) f_n(u - x_1) f_n(v - x_2) \, dx_1 \, dx_2 
+ \rho_{G_{n-1}} c_n \beta_n^2 f_n * \tilde{f}_n(u - v) \tag{16}
\]

for any \( u, v \in \mathbb{R}^d \), where the two terms on the right hand side correspond to pairs of points from \( Y_n \) belonging to different clusters and the same cluster, respectively. Hence by (2) and (15), \( W_n \) is stationary with intensity

\[
\rho_{W_n} = p_n \rho_{Y_n} = \rho_{G_{n-1}} \beta_n p_n \tag{17}
\]

and PCF

\[
g_{W_n}(u - v) = g_{Y_n}(u - v) 
= \int \int g_{G_{n-1}}(x_1 - x_2) f_n(u - x_1) f_n(v - x_2) \, dx_1 \, dx_2 
+ \frac{c_n}{\rho_{G_{n-1}}} f_n * \tilde{f}_n(u - v) \tag{18}
\]

where the first identify follows from the fact that PCFs are invariant under independent thinning, and where (16) is used to obtain the second identity. Also, for disjoint Borel sets \( A_1, A_2 \subseteq \mathbb{R}^d \), it follows from items (a)–(c) that

\[
E\left\{ \#(W_n \cap A_1) \#(G_{n-1}^{\text{thin}} \cap A_2) \right\} = \rho_{W_n} \rho_{G_{n-1}^{\text{thin}}} \int_{A_1} \int_{A_2} \left\{ \frac{1}{\rho_{G_{n-1}}} f_n(x_1 - x_2) + g_{G_{n-1}^{\text{thin}}}(x_1 - x_2) \right\} \, dx_1 \, dx_2.
\]

Furthermore, by (3), (7), and Fubini’s theorem it is readily seen that \( G_n \) has PCF given by

\[
\rho_{G_n}^2 g_{G_n}(x) = \rho_{G_{n-1}^{\text{thin}}}^2 g_{G_{n-1}^{\text{thin}}}(x) + \rho_{G_{n-1}^{\text{thin}}}^2 g_{G_{n-1}^{\text{thin}}}(x) + \rho_{G_{n-1}^{\text{thin}}}^2 g_{G_{n-1}^{\text{thin}}}(x)
+ 2 \rho_{G_{n-1}^{\text{thin}}} \rho_{Z_n}^2 + 2 \rho_{Z_n} \rho_{G_{n-1}^{\text{thin}}}
+ 2 \rho_{W_n} \rho_{G_{n-1}^{\text{thin}}} \left\{ g_{G_{n-1}^{\text{thin}}} * f_n(x) + \frac{1}{\rho_{G_{n-1}^{\text{thin}}}} f_n(x) \right\}
\]
where the terms on the right hand side correspond to pairs of points from $W_n, Z_n, G_{n-1}^{\text{thin}}, W_n$ and $Z_n, Z_n$ and $G_{n-1}^{\text{thin}},$ and $W_n$ and $G_{n-1}^{\text{thin}},$ respectively, where the latter three cases can be ordered in two ways. Combining all this with the first identity in (6) and (17), we easily obtain

$$g_{G_n}(u - v) - 1 = \left( \frac{\rho_{G_{n-1}} \beta_n p_n}{\rho_{G_n}} \right)^2 \{ g_{W_n}(u - v) - 1 \} + \left( \frac{\rho_{Z_n}}{\rho_{G_n}} \right)^2 \{ g_{Z_n}(u - v) - 1 \}$$

$$+ \rho_{W_n} \rho_{G_{n-1}} \left[ \int (g_{G_{n-1}}(v - x) - 1) \left\{ f_n(u - x) + \tilde{f}_n(u - x) \right\} \mathrm{d}x \right.$$  

$$+ \frac{1}{\rho_{G_{n-1}}} \left\{ f_n(u - v) + \tilde{f}_n(u - v) \right\} \right].$$

This combined with (18) imply the result in Lemma 3.4.

3.2.4 Extension

More generally than in Section 3.2.2 we may consider the case where the PCF of the initial generation $G_0$ and the noise $Z_n$ are affine expressions:

$$g_{G_0} - 1 = a_0 + a_1 f_{0,1} * \tilde{f}_{0,1} + \cdots + a_k f_{0,k} * \tilde{f}_{0,k}$$

and

$$g_{Z_n} - 1 = b_{n,0} + b_{n,1} f_{Z,n,1} * \tilde{f}_{Z,n,1} + \cdots + b_{n,l} f_{Z,n,l} * \tilde{f}_{Z,n,l}, \quad n = 1, 2, \ldots, (20)$$

for real constants $a_0, \ldots, a_k, b_{n,1}, \ldots, b_{n,l}$ and PDFs $f_{0,1}, \ldots, f_{0,k}, f_{Z,n,1}, \ldots, f_{Z,n,l}$. For instance, the superposition of $k$ independent Poisson, weighted permanental, or weigthed determinantal point processes has a PCF of the form (19) or (20). Also Theorem 3.3 provides examples of PCFs of the form (19) or (20). Assuming (19) and (20), Theorem 3.3 is immediately generalised by replacing $a f_0 * f_0$ in (11) by (19), $b_n f_{Z,n} * f_{Z,n}$ in (14) by (20), and similarly for $b_n f_{Z,i} * \tilde{f}_{Z,i}$ in (13).

4 Same reproduction system

Throughout this section we assume the same reproduction system over generations, that is, in items (a)–(d), $\beta_n = \beta, \nu_n = \nu, f_n = f, p_n = p, q_n = q$ do not depend on $n, Z_1, Z_2, \ldots$ are IID stationary point processes, so $\rho_{Z_n} = \rho_Z$
for \( n = 1, 2, \ldots \), and \( \rho_{G_0} = \rho_{G_1} = \cdots = \rho_G > 0 \). Note that the noise process \( Z_n \) and the initial generation process \( G_0 \) need not be Poisson processes and the offspring densities need not be Gaussian as in Shimatani’s paper [19]. By (6), we have either

\[
\beta p + q = 1 \quad \text{and} \quad \rho_Z = 0,
\]

or

\[
\beta p + q < 1 \quad \text{and} \quad \rho_Z > 0.
\]

In case of (22),

\[
\rho_G = \frac{\rho_Z}{(1 - \beta p - q)}.
\]

### 4.1 Limiting pair correlation function

Under the assumptions above and in Theorem 3.3, setting \( 0^0 = 1 \), the PCF simplifies after a straightforward calculation to

\[
g_{G_n}(u) - 1 = a f_0 * \tilde{f}_0 * \sum_{k_1=0}^{n} \sum_{k_2=0}^{n} \binom{n}{k_1} \binom{n}{k_2} q^{2n-k_1-k_2}(\beta p)^{k_1+k_2} f^{*k_1} * \tilde{f}^{*k_2}(u) + \frac{\rho Z}{\rho G} \cdot \sum_{i=0}^{n-1} \sum_{k_1=0}^{i} \sum_{k_2=0}^{i} \binom{i}{k_1} \binom{i}{k_2} q^{2i-k_1-k_2}(\beta p)^{k_1+k_2} f^{*k_1} * \tilde{f}^{*k_2}(u),
\]

for \( n = 1, 2, \ldots \), where

\[
c = (\nu + \beta^2 - \beta)/\beta^2 \quad \text{if} \quad \beta > 0, \quad c = 0 \quad \text{if} \quad \beta = 0,
\]

\( f^{*n} \) is the \( n \)-th convolution power of \( f \) if \( n > 0 \), and \( f^{*0} * \tilde{f}^{*0} = \delta_0 \). For instance, consider the case \( f \sim N_d(\sigma^2) \) and \( f_Z \sim N_d(\kappa^2) \), and suppose \( d \geq 3 \) in case of (21). Then the binomial formula combined with either (21) or (22), imply that the first double sum in (24) tends to 0 as \( n \to \infty \), and hence

\[
g_G(u) - 1 := \lim_{n \to \infty} g_{G_n}(u) - 1
\]

\[
= \frac{c}{\rho G} \sum_{i=0}^{\infty} \sum_{k_1=0}^{i} \sum_{k_2=0}^{i} \binom{i}{k_1} \binom{i}{k_2} \frac{q^{2i-k_1-k_2}(\beta p)^{2i-k_1-k_2}}{\{2\pi(2+k_1+k_2)\sigma^2\}^{d/2}} \cdot \exp \left\{ -\frac{\|u\|^2}{2(2+k_1+k_2)\sigma^2} \right\}
\]

16
is finite. Shimatani in [19] only showed that this is finite under the assumption \( d = 2, \ b = q = 0, \) and \( c > 0. \) Then Shimatani noticed that \( \beta p = 1 \) and \( \rho Z = 0 \) (which is (21) for \( q = 0 \)) imply divergence of \( g_{G_n} \) as \( n \to \infty, \) whilst \( \beta p < 1 \) and \( \rho Z > 0 \) (which is (22) for \( q = 0 \)) imply convergence. Further, in the case of convergence and when \( \beta p \approx 1, \) he discussed an approximation of \( g_{g}(u) \) that depends on whether \( \|u\| \) is close to 0 or not.

In general (i.e., without making the assumption of normal distributions and so on), if we assume \( g_{G_n} - 1 \) has a finite limit and (22) is satisfied, then

\[
g_{G}(u) - 1 = \left\{ \frac{c(\beta p)^2 f * \tilde{f} + \beta pq(f + \tilde{f})}{\rho G} + \left( \frac{\rho Z}{\rho G} \right)^2 b f_{Z} * \tilde{f}_{Z} \right\} + \sum_{i=0}^{\infty} \sum_{k_1=0}^{i} \sum_{k_2=0}^{i} \left( \frac{i}{k_1} \right) \left( \frac{i}{k_2} \right) q^{2i-k_1-k_2} \frac{(\beta p)_{k_1+k_2}}{2 \pi ((k_1 + k_2)\sigma^2 + 2\kappa^2)} d^{i/2}
\]

which does not depend on \( a \) or \( f_0. \) Here, as \( \beta p + q \uparrow 1, \) \( \rho Z/\rho G \) goes to 0, meaning that the less noise we consider, the less it matters which type of PCF for the noise process \( Z_n \) we choose. On the other hand, as \( \beta p + q \downarrow 0, \) \( g_{G} - 1 \) tends to \( b f_{Z} * \tilde{f}_{Z}, \) which simply is the PCF of \( Z. \)

Considering the situation at the end of Section 3.2.2 assume that \( d = 2, \ q = 0, \ f \sim N_d(\sigma^2), \) and \( g_{Z_n} - 1 = b f_{Z} * \tilde{f}_{Z} \) (corresponding to (9)) with \( f_{Z} \sim N_d(\kappa^2/8) \) and \( b = 0, \ b = -(\sqrt{\pi}\kappa)^2, \) and \( b = 2(\sqrt{\pi}\kappa)^2 \) for the Poisson, determinantal, and weighted permanental point process, respectively. Then \( g_{G}(u) \) is given by (25), where \( d = 2 \) and \( \kappa^2 \) is replaced by \( \kappa^2/8. \) Also assume that \( p = 1, \ \sigma = 0.1, \ \rho_G = 100, \) and the number of points in a cluster is Poisson distributed (implying \( c = 1 \)) with mean \( \beta = 0.8, \) so \( \rho Z = 20. \) Finally, assume \( \kappa = 0.1 \) in case of weighted permanent noise and \( \kappa = 1/\sqrt{\rho Z \pi} \) in case of determinantal noise (the most repulsive Gaussian determinantal point process). Shimatani in [19] discussed the case where \( \beta p = 0.99 \) – a plot (omitted here) shows that the limiting PCFs corresponding to the three

\[
+ \frac{2}{\rho G} \sum_{i=0}^{\infty} \sum_{k_1=0}^{i} \sum_{k_2=0}^{i} \left( \frac{i}{k_1} \right) \left( \frac{i}{k_2} \right) q^{2i-k_1-k_2+1} \frac{(\beta p)_{1-k_1-k_2}}{2 \pi ((1 + k_1 + k_2)\sigma^2)} d^{i/2}
\]

\[
+ \left( \frac{\rho Z}{\rho G} \right)^2 \sum_{i=0}^{\infty} \sum_{k_1=0}^{i} \sum_{k_2=0}^{i} \left( \frac{i}{k_1} \right) \left( \frac{i}{k_2} \right) q^{2i-k_1-k_2} \frac{(\beta p)_{k_1+k_2}}{2 \pi ((k_1 + k_2)\sigma^2 + 2\kappa^2)} d^{i/2}
\]
models of noise processes are then effectively equal. By lowering $\beta p$, the reproduction system is diminished, and hence depending on the model type, a higher degree of regularity or clustering is obtained. This will also increase the rate of convergence because the number of generations initialized by a single point will be fewer. Note that in Figure 3 the convergence is already rapid as $g_{G_8}$ and $g_{G_{16}}$ are practically indistinguishable. Figure 3 further shows that it is only for small or moderate inter-point distances that the three limiting PCFs clearly differ.

Figure 3: The reduced PCFs $g_{G_n} - 1$ when the noise processes are either determinantal, Poisson or weighted permanental point processes (left to right), with parameters and Gaussian offspring PDF as specified in the text. The solid horizontal line is the PCF for a Poisson process.

4.2 Second main result

Although Shimatani in [19] showed convergence of $g_{G_n}$ in the special case considered above, he did not clarify whether the Markov chain $G_0, G_1, \ldots$ converges in distribution to a limit so that this limiting distribution (also called the equilibrium, invariant, or stationary distribution) has a PCF given by (26). In order to show that $G_0, G_1, \ldots$ is indeed converging to a limiting distribution under more general conditions, and to specify what this is, we construct in accordance with items (a)–(d) a Markov chain $\ldots, G_{s-1}^{st}, G_0^{st}, G_1^{st}, \ldots$ with times given by all integers $n$ and so that this chain is time-stationary (its distribution is invariant under discrete time shifts), as follows. First, we generate noise processes as in item (d): Let $\ldots, Z_{-1}, Z_0, Z_1, \ldots$ be independent stationary Poisson processes on $\mathbb{R}^d$ with intensity $\rho_Z$. Second, for any integer $n$ and point $x \in Z_n$, we consider the family of all generations initiated by the ancestor $x$, that is, the family

$$F_{n,x} = \bigcup_{m=1}^{\infty} W_{n,x}^{(m)}$$
where

$$W_{n,x}^{(1)} = \begin{cases} W_{n,x} \cup \{ x \} & \text{if } Q_{1,x} = 1, \\ W_{n,x} & \text{if } Q_{1,x} = 0, \end{cases}$$

is defined by the reproduction mechanism of independent clustering, independent thinning, and independent retention given in items (a)–(c) (with $\beta_n = \beta$ and $\nu_n = \nu$), $W_{n,x}^{(2)}$ is the offspring and retained points generated by the points in $W_{n,x}^{(1)}$ (using the same reproduction mechanism as before), and so on. In other words, $W_{n,x}^{(m)}$ is the set of $(m + n)$-th generation points with common $n$-th generation ancestor $x \in Z_n$. Moreover, we assume that conditional on $\ldots, Z_{-1}, Z_0, Z_1, \ldots$, the families $F_{n,x}$ for all integers $n$ and $x \in Z_n$ are independent (and hence IID). Finally, for all integers $n$, we let

$$G_{n,x}^{st} = W_{n,x}^{st} \cup Z_n \quad \text{with } W_{n,x}^{st} = \bigcup_{m=1}^{\infty} \bigcup_{x \in Z_{n-m}} W_{n-m,x}^{(m)}.$$  

(27)

It will be evident from Theorem 4.2 below that any $G_{n,x}^{st}$ has intensity $\rho_G$ given by (23) and PCF $g_G$ given by (26) (provided $g_G(u - v)$ is a locally integrable function of $(u, v) \in \mathbb{R}^d \times \mathbb{R}^d$); a formal proof is given in Appendix B. The proof of Theorem 4.2 is based on a coupling construction between $G_1^{st}, G_2^{st}, \ldots$ and $G_1^{st}, G_2^{st}, \ldots$ together with the following result.

**Lemma 4.1.** Suppose $\beta_n = \beta$, $\nu_n = \nu$, $f_n = f$, $p_n = p$, $q_n = q$, and $\rho_{Z_n} = \rho_Z$ do not depend on $n \geq 1$, where $\beta p + q < 1$ and $\rho_Z > 0$. Let $K \subset \mathbb{R}^d$ be a compact set and let

$$T_{0,K}^{st} = \sup \{ m \in \{1, 2, \ldots \} : W_{0,x}^{(m)} \cap K \neq \emptyset \text{ for some } x \in G_{0,x}^{st} \}$$

be the last time a point in $K$ is a member of a family initiated by some point in the $0$-th generation $G_{0,x}^{st}$. Then

$$E(T_{0,K}^{st}) \leq |K| \rho_G \frac{\beta p + q}{1 - \beta p - q}$$

is finite, and so $T_{0,K}^{st} < \infty$ almost surely.

**Proof.** Let $K \subset \mathbb{R}^d$ be compact and define

$$N = \sum_{x \in G_{0,x}^{st}} \#(F_{0,x} \cap K).$$

By the law of total expectation, conditioning on $G_{0,x}^{st}$ and using Campbell’s theorem, we obtain

$$E(N) = \rho_G \int \sum_{m=1}^{\infty} \int_K (\beta p f + q \delta_0)^m (y - x) \, dy \, dx.$$
Theorem 4.2. \( n > T \) and converges in distribution to random time is a coupling between for any compact set \( n \)

Consequently, \( P(T_{0,K}^{st} \leq N) = 1 \), whereby the lemma follows. \( \square \)

We are now ready to state our second main result.

**Theorem 4.2.** Suppose \( \ldots , Z_{-1}, Z_0, Z_1 \ldots \) are IID stationary point processes and \( \beta_n = \beta, \nu_n = \nu, f_n = f, p_n = p, q_n = q, \) and \( \rho_{Z_n} = \rho_Z \) do not depend on \( n \geq 1 \), where \( \beta p + q < 1 \) and \( \rho_Z > 0 \). Then \( \ldots , G_{n-1}^{st}, G_0^{st}, G_1^{st} \ldots \) is a time-stationary Markov chain constructed in accordance to items (a)–(d). Let \( \Pi \) be the distribution of any \( G_n^{st} \) and let \( \mathcal{N} \) be the space of all locally finite subsets of \( \mathbb{R}^d \). Then there exists a (measurable) subset \( \Omega \subseteq \mathcal{N} \) so that \( \Pi(\Omega) = 1 \) and for any compact set \( K \subset \mathbb{R}^d \) and all \( \omega \in \Omega \), conditional on \( G_0 = \omega \), there is a coupling between \( G_1, G_2, \ldots \) and \( G_{n-1}^{st}, G_0^{st}, G_1^{st} \ldots \), and there exists a random time \( T_K(\omega) \in \{0, 1, \ldots \} \) so that \( G_n \cap K = G_n^{st} \cap K \) for all integers \( n > T_K(\omega) \). In particular, for any \( \omega \in \Omega \) and conditional on \( G_0 = \omega \), \( G_n \) converges in distribution to \( \Pi \) as \( n \to \infty \), and so \( \Pi \) is the unique invariant distribution of the chain \( G_0, G_1, \ldots \).

**Proof.** Obviously, \( \ldots , G_{n-1}^{st}, G_0^{st}, G_1^{st} \ldots \) is a time-stationary Markov chain constructed in accordance to items (a)–(d). To verify the remaining part of the theorem, we may assume that \( G_0 \) and \( G_0^{st} \) are independent. Then, conditional on \( G_0 \), we have a coupling between \( G_1, G_2, \ldots \) and \( G_{n-1}^{st}, G_0^{st}, G_1^{st} \ldots \), because \( G_1^{st}, G_2^{st} \ldots \) and \( G_1, G_2, \ldots \) are generated by the same noise processes \( Z_1, Z_2, \ldots \), the same offspring processes \( Y_{n,x} \) for all times \( n = 1, 2, \ldots \) and all ancestors \( x \in G_{n-1} \cap G_{n-1}^{st} \), the same Bernoulli variables \( B_{n,y} \) for all times \( n = 1, 2, \ldots \) and all offspring \( y \in Y_{n,x} \) with ancestor \( x \in G_{n-1} \cap G_{n-1}^{st} \), and the same Bernoulli variables \( Q_{n,x} \) for all times \( n = 1, 2, \ldots \) and all retained points \( x \in G_{n-1} \cap G_{n-1}^{st} \). Let \( K \subset \mathbb{R}^d \) be compact. In accordance with (28), for \( \omega \in \mathcal{N} \), let

\[
T_K(\omega) = \sup \{ m \in \{1, 2, \ldots \} : W_{0,x}^{(m)} \cap K \neq \emptyset \text{ for some } x \in \omega \}
\]

be the last time a point in \( K \) is a member of a family initiated by some point in \( \omega \), and let \( \Omega = \{ \omega \in \mathcal{N} : T_K(\omega) < \infty \} \). By Lemma 4.1 and the coupling
construction, \( \Pi(\Omega) = 1 \) and \( G_n \cap K = G_{st}^n \cap K \) whenever \( n > T_K(\omega) \), so for any \( \omega \in \Omega \),

\[
\lim_{n \to \infty} P(G_n \cap K = \emptyset | G_0 = \omega) = \lim_{n \to \infty} P(G_{st}^n \cap K = \emptyset, n > T_K(\omega))
\]

because \( G_0 \) is independent of \( (G_{st}^0, T_K(\omega)) \). Since the sequence of events \( \{ \omega : 1 > T_K(\omega) \} \subseteq \{ \omega : 2 > T_K(\omega) \} \subseteq \ldots \) increases to \( \Omega \), we obtain

\[
\lim_{n \to \infty} P(G_n \cap K = \emptyset | G_0 = \omega) = \lim_{n \to \infty} P(G_{st}^n \cap K = \emptyset) = P(G_{st}^0 \cap K = \emptyset).
\]

Thus, recalling that the distribution of a random closed set \( X \subseteq \mathbb{R}^d \) (including a locally finite point process) is uniquely characterized by the void probabilities \( P(X \cap K = \emptyset) \) for all compact sets \( K \subset \mathbb{R}^d \), we have verified that conditional on \( G_0 = \omega \), the chain \( G_1, G_2 \ldots \) converges in distribution towards \( \Pi \). In turn, this implies uniqueness of the invariant distribution \( \Pi \).

In Theorem 4.2, under mild conditions, we can take \( \Omega = \mathbb{N} \). For instance, this is easily seen to be the case if there exists \( \epsilon > 0 \) so that \( f(x) > 0 \) whenever \( \|x\| \leq \epsilon \). In the special case \( c = 0 \), \( \Pi \) is just a stationary Poisson process, and so \( \Omega = \mathbb{N} \). Moreover, the integral

\[
\gamma := \int \left( g_G - 1 \right)
\]

is a rough measure of the amount of positive/negative association between the points in \( G_{st}^n \). Note that comparing \( \gamma \) with the corresponding measure for another stationary point process makes only sense if the processes have equal intensities, see [6]. Under the assumptions in both Theorem 3.3 and 4.2 by (26),

\[
\gamma = \frac{c(\beta p)^2 + 2\beta pq}{\rho_G \{1-(\beta p + q)^2\}} + \frac{b\rho_Z^2}{\rho_G \{1-(\beta p + q)^2\}}
\]

\[
= \frac{1}{1 + \beta p + q} \left\{ \frac{c(\beta p)^2 + 2\beta pq}{\rho_Z} + b(1 - \beta p - q) \right\}
\]

which does not depend on \( f \) or \( f_Z \). Furthermore, \( \gamma \) may take any positive value and some negative values depending on how we choose the values of the parameters. This means we may have an equilibrium distribution exhibiting any degree of clustering or some degree of regularity. In fact, \( \gamma \) can only be negative when \( b \) is negative, e.g. when \( Z_n \) is a determinantal point process. In this case \( b \) has a lower bound, \( b_{\min} \), that ensures the existence of the determinantal point process [6] and consequently, \( \gamma \geq b_{\min} \). The case \( \gamma = b_{\min} \).
happens exactly when \( \beta p + q = 0 \) (i.e., when offspring are never produced or no points are retained after the thinning procedures in items (b) and (c)) and thus \( G_n = Z_n \) is a determinantal point process.

For approximate simulation of \( G_0^{st} \) under each of the three models of the noise processes, we use the algorithm described in Appendix C. Simulation was initially done with parameters and set-up corresponding to that of Figure 3. However, the resulting point patterns were not distinguishable from a stationary Poisson process when comparing empirical estimates of the PCF, \( L \)-function, or \( J \)-function of the simulations to 95% global rank envelopes under each model (for definition of \( L \)- and \( J \)-functions, see e.g. [16], and for the envelopes, see [17]). Therefore, in order to better distinguish the three models, we consider two cases as follows.

**Case 1:**
This case is based on minimizing \( \gamma \) under determinantal noise and on maximizing \( \gamma \) under weighted permanental noise. Let \( d = 2, f \sim N_d(\sigma^2) \), with \( \sigma = 0.1, f_Z \sim N_d(\kappa^2/8), \rho_G = 100, p = 1, \beta = 0.3, q = 0 \), and consequently \( \rho_Z = 70 \).

- In case of determinantal noise: Let \( \kappa = 1/\sqrt{\rho_Z\pi} \) (the most repulsive Gaussian determinantal point process) and the number of points in a cluster be Bernoulli distributed with parameter \( \beta \), implying \( c = 0 \) (each point has at most one offspring). Then \( \gamma \approx -5.38 \times 10^{-3} \).
- In case of Poisson noise: Let the number of points in a cluster be Poisson distributed with intensity \( \beta \), implying \( c = 1 \). Then \( \gamma \approx 9.89 \times 10^{-4} \).
- In case of weighted permanental noise: Let \( \kappa = 1 \) and the number of points in a cluster be negative binomially distributed with probability of success equal to 0.12 and dispersion parameter equal to 0.11, implying \( c = 10 \). Then \( \gamma \approx 3.39 \).

**Case 2:**
This case is such that the clusters are more separated. Let \( d = 2, f \sim N_d(\sigma^2) \), with \( \sigma = 0.01, f_Z \sim N_d(\kappa^2/8), \rho_G = 100, p = 1, \beta = 0.95, q = 0 \), and consequently \( \rho_Z = 5 \). Also, let the number of points in a cluster be negative binomially distributed with probability of success equal to 0.208 and dispersion parameter equal to 0.25, implying \( c = 5 \).

- In case of determinantal noise: Let \( \kappa = 1/\sqrt{\rho_Z\pi} \). Then \( \gamma \approx 0.463 \).
- In case of Poisson noise: \( \gamma \approx 0.463 \).
In case of weighted permanental noise: Let $\kappa = 1$. Then $\gamma \approx 0.624$.

Figure 4 shows simulations of $G^s_0$ under each of the three models of the noise processes (left to right) in Case 1 and 2 (top and bottom). Based on these simulations, Figure 5 shows empirical estimates of functional summary statistics based on the simulated point patterns from Figure 4 along with 95% global rank envelopes based on 2499 simulations (as recommended in [17]) of a stationary Poisson process with the same intensity as used in Figure 4. The first simulated point pattern of Case 1 looks slightly less clustered than the second, whilst the last looks more clustered. This is in accordance with the values of $\gamma$ and the corresponding functional summary statistics in Figure 5. Additionally, Figure 5 reveals that the case of Poisson noise is not distinguishable from the stationary Poisson process, while the case of weighted permanental noise is more clustered. The case of determinantal noise is not distinguishable from the stationary Poisson process by the PCF or $L$-function, but is shown to be more regular by the $J$-function. In Case 2, the clusters of the point pattern simulated under determinantal noise looks more separated than the clusters of the point pattern simulated under Poisson noise. The clusters of the point pattern simulated under weighted permanental noise are clustered to such a degree that it gives the illusion of few highly separated clusters. All three models of Case 2 are as expected significantly different from the stationary Poisson process.

A Weighted determinantal and permanental point processes

When defining stationary weighted determinantal/permanental point processes, the main ingredients are a symmetric function $C : \mathbb{R}^d \to \mathbb{R}$ and a real number $\alpha$. Before giving the definitions of these point processes we recall the following.

For a real $n \times n$ matrix $A$ with $(i, j)$-th entry $a_{i,j}$, the $\alpha$-weighted permanent of $A$ is defined by

$$\text{per}_\alpha(A) = \sum_\sigma \alpha^{#\sigma} a_{1,\sigma_1} \cdots a_{n,\sigma_n}$$

where $\sigma$ denotes a permutation of $\{1, \ldots, n\}$ and $#\sigma$ is the number of its cycles. This is the usual permanent of $A$ if $\alpha = 1$. Moreover, the $\alpha$-weighted determinant of $A$ is given by

$$\det_\alpha(A) = \text{per}_{-\alpha}(-A).$$
This is the usual determinant of \( A \) if \( \alpha = -1 \). Often we just write \( \text{per}_\alpha(A) \) for \( \text{per}_\alpha(A) \), and \( \text{det}_\alpha(A) \) for \( \text{det}_\alpha(A) \).

For any \( G_1, \ldots, G_n \in \mathbb{R}^d \), the \( n \times n \) matrix with \((i, j)\)-th entry \( C(G_i - G_j) \) is denoted by \( [C](G_1, \ldots, G_n) \). Thus

\[
\text{per}_\alpha[C](G_1, \ldots, G_n) = \sum_\sigma \alpha^\# \sigma C(G_1 - G_{\sigma_1}) \cdots C(G_n - G_{\sigma_n}).
\]

Note that the weighted permanent/determinant can be negative if the mapping \( \mathbb{R}^d \times \mathbb{R}^d \ni (u, v) \rightarrow C(u - v) \) is not positive semi-definite. When this mapping is positive semi-definite, \( C \) is an auto-covariance function, with corresponding auto-correlation function \( R(x) = C(x)/C(0) \) provided \( C(0) > 0 \).

A locally finite point process \( X \subset \mathbb{R}^d \) has \( n \)-th order joint intensity \( \rho_X^{(n)} \) for \( n = 1, 2, \ldots \) if for any bounded and pairwise disjoint Borel sets \( A_1, \ldots, A_n \subset \mathbb{R}^d \),

\[
\mathbb{E}[N(A_1) \cdots N(A_n)] = \int_{A_1} \cdots \int_{A_n} \rho_X^{(n)}(x_1, \ldots, x_n) \, dx_1 \cdots \, dx_n < \infty.
\]
Figure 5: Empirical PCFs, \( L \)-functions, and \( J \)-functions (left to right) based on the simulations of \( G_{\text{st}}^0 \) from Figure 4 when the noise processes are either determinantal (dashed), Poisson (solid), or weighted permanental (dotted). The rows corresponds to Case 1 and 2, respectively. The grey regions are 95\% global rank envelopes based on 2499 simulations of a stationary Poisson process with the same intensity as \( G_{\text{st}}^0 \).

Note that \( \rho^{(n)}_X \) is unique except for a Lebesgue nullset in \( \mathbb{R}^{dn} \) (we ignore nullsets in the following). Thus, if \( X \) is stationary, \( \rho^{(1)}_X \) is constant and agrees with the intensity \( \rho_X \), and \( \rho_X > 0 \) implies that \( g_X(u - v) = \rho^{(2)}_X(u,v)/\rho^2_X \) is the PCF.

If for all \( n = 1, 2, \ldots \), the \( n \)-th order joint intensity exists and is given by

\[
\rho^{(n)}_X(G_1, \ldots, G_n) = \text{per}_\alpha[C](G_1, \ldots, G_n)
\]

we say that \( X \) is a stationary \( \alpha \)-weighted permanental point process with kernel \( C \) and write \( X \sim \text{PPP}_\alpha(C) \). Conditions are needed to ensure the existence of \( \text{PPP}_\alpha(C) \), see \cite{20} and \cite{10}. To exclude the trivial case where \( X \) is empty we assume \( \alpha C(0) > 0 \). Note that \( C \) must be an auto-covariance function and \( \alpha > 0 \) because \( \rho_X = \alpha C(0) \) and

\[
g_X(x) - 1 = R(x)^2/\alpha. \tag{30}
\]
This reflects that the process exhibits a positive association between its points. In fact, if $C$ is an auto-covariance function and $k = 2\alpha$ is a positive integer, then $X \sim \text{PPP}_\alpha(C)$ exists and it is a Cox process: Conditional on IID zero-mean stationary Gaussian processes $\Phi_1, \ldots, \Phi_k$ on $\mathbb{R}^d$ with auto-covariance function $C/2$, we can let $X$ be a Poisson process with intensity function $\Lambda(x) = \Phi_1(x)^2 + \cdots + \Phi_k(x)^2$, $x \in \mathbb{R}^d$. In particular, if $\alpha = 1$, then $X$ is the boson process introduced by [7].

If for all $n = 1, 2, \ldots$, the $n$-th order joint intensity exists and is given by

$$
\rho_X^{(n)}(G_1, \ldots, G_n) = \det_{\alpha}[C](G_1, \ldots, G_n)
$$

we say that $X$ is a stationary $\alpha$-weighted determinantal point process with kernel $C$ and write $X \sim \text{DPP}_\alpha(C)$. To exclude the trivial case where $X$ is empty we assume $\alpha C(0) > 0$. Again $C$ needs to be an auto-covariance function and $\alpha > 0$ because $\rho_X = \alpha C(0)$ and

$$
g_X(x) - 1 = -R(x)^2/\alpha.
$$

(31)

If $\alpha = 1$, then $X$ is the fermion process introduced by [7] (it is usually called the determinantal point process). We have the following existence result: If $C$ is continuous and square integrable, existence of $X \sim \text{DPP}_\alpha(C)$ is equivalent to that the Fourier transform of $C$ is bounded by 0 and 1 [6]. When $\alpha$ is a positive integer, $X \sim \text{DPP}_\alpha(C)$ can be identified with the superposition $G_1 \cup \cdots \cup G_\alpha$ of independent processes $G_i \sim \text{DPP}_\alpha(C/\alpha)$, $i = 1, \ldots, \alpha$. In general, the process is not well-defined if $0 < \alpha < 1$, cf. [10].

B The intensity and PCF of the invariant distribution

Let the situation be as in Theorem 4.2. Below we verify (23) and (26) holds for $G^*_n$ provided $g_G(u - v)$ is a locally integrable function of $(u, v) \in \mathbb{R}^d \times \mathbb{R}^d$.

Note that the $G^*_n$ are identically distributed and $G^*_0 = W^*_0 \cup Z_0$ where $W^*_0 = \bigcup_{m=1}^{\infty} \bigcup_{x \in Z_{-m}} W^*_{m,x}$, cf. (27). Hence, for Borel sets $A \subseteq \mathbb{R}^d$ with $|A| < \infty$, using similar arguments as in the derivation of (29), we obtain

$$
E\{\#(W^*_0 \cap A)\} = |A|\rho_Z \frac{\beta p + q}{1 - \beta p - q},
$$

(32)

so $W^*_0$ has intensity

$$
\rho_W = \rho_Z \frac{\beta p + q}{1 - \beta p - q}
$$

(33)
whereby it follows that \( G_{0}^{st} \) has intensity \( \rho_{G} \) as given by \( (23) \).

Let \( A_{1}, A_{2} \subseteq \mathbb{R}^{d} \) be disjoint Borel sets with \( |A_{i}| < \infty \), \( i = 1, 2 \). Using similar arguments as in the derivation of \( (29) \) (or \( (32) \)) and exploiting the fact that \( Z_{0}, Z_{-1}, \ldots \) are IID point processes with a PCF of the form \( g_{Z} = 1 + bf_{z} * \tilde{f}_{Z} \) as well as the independence between \( Z_{0} \) and \( W_{0}^{st} \), we obtain

\[
E\{ #(G_{0}^{st} \cap A_{1}) #(G_{0}^{st} \cap A_{2}) \} = \rho_{2}^{2} |A_{1}| |A_{2}| + \rho_{Z}^{2} \int_{A_{1}} \int_{A_{2}} bf_{z} * \tilde{f}_{Z}(x_{1} - x_{2}) \, dx_{1} \, dx_{2} + 2 \rho_{Z} \rho_{W} |A_{1}| |A_{2}|
\]

\[
+ \sum_{m_{1}=1}^{\infty} \sum_{m_{2}=1: m_{1} \neq m_{2}}^{\infty} \rho_{Z}^{2} (\beta p + q)^{m_{1}+m_{2}} |A_{1}| |A_{2}|
\]

\[
+ \sum_{m=1}^{\infty} \rho_{Z}^{2} (\beta p + q)^{2m} |A_{1}| |A_{2}|
\]

\[
+ \sum_{m=1}^{\infty} \rho_{Z}^{2} \int_{A_{1}} \int_{A_{2}} bf_{z} * \tilde{f}_{Z} \times \sum_{k_{1}=0}^{m} \sum_{k_{2}=0}^{m} \binom{m}{k_{1}} \binom{m}{k_{2}} q^{2m-k_{1}-k_{2}} (\beta p)^{k_{1}+k_{2}} f^{*k_{1}} * \tilde{f}^{*k_{2}}(y_{1} - y_{2}) \, dy_{1} \, dy_{2}
\]

\[
+ \sum_{m=1}^{\infty} E\left\{ \sum_{x \in Z_{-m}} #(W_{0}^{(m)} \cap A_{1}) #(W_{0}^{(m)} \cap A_{2}) \right\}.
\]

Here,

- the first two terms of \( (34) \) corresponds to pairs of points from \( Z_{0} \) with one point falling in \( A_{1} \) and the other in \( A_{2} \);
- the third term corresponds to pairs of points either from \( Z_{0} \cap A_{1} \) and \( W_{0}^{st} \cap A_{2} \) or from \( Z_{0} \cap A_{2} \) and \( W_{0}^{st} \cap A_{1} \);
- the term in \( (35) \) corresponds to pairs of points, with one point falling in \( A_{1} \) and the other in \( A_{2} \) of two families initiated by ancestors from different generations;
- the two terms in \( (36) \) corresponds to pairs of points, with one point falling in \( A_{1} \) and the other in \( A_{2} \) from two different families initiated by ancestors from the same generation;
- the term in \( (37) \) corresponds to pairs of points from the same family, falling in \( A_{1} \) and \( A_{2} \), respectively.
Using (23) and (33), we observe that (34)–(36) simplify to

\[ \rho^2 G |A_1||A_2| + \sum_{m=0}^{\infty} \rho Z \int A_1 \int A_2 b f_Z * \tilde{f}_Z \]

\[ \times \sum_{k_1=0}^{m} \sum_{k_2=0}^{m} (m \choose k_1) (m \choose k_2) q^{2m-k_1-k_2} (\beta p)^{k_1+k_2} f^{*k_1} \tilde{f}^{*k_2} (y_1 - y_2) \, dy_1 \, dy_2 \]

and the term in (37) is equal to

\[ \rho Z \sum_{m=1}^{\infty} \int \int \int \int_{A_1} \int_{A_2} ((\beta p f + q \delta_0)^{*i}(y - x) \]

\[ \cdot \left[ c(\beta p)^2 f(\tilde{y}_1 - y)f(\tilde{y}_2 - y) + \beta pq \{f(\tilde{y}_1 - y)\delta_0(\tilde{y}_2 - y) + \delta_0(\tilde{y}_1 - y)f(\tilde{y}_2 - y)\} \right] \]

\[ \cdot (\beta p f + q \delta_0)^{*m-1-i}(y_1 - \hat{y}_1) \]

\[ \cdot (\beta p f + q \delta_0)^{*m-1-i}(y_2 - \hat{y}_2) \, dy_1 \, dy_2 \, d\hat{y}_1 \, d\hat{y}_2 \, dy \, dx. \]

(38)

In (39), \( y \) corresponds to an \( i \)-th generation point in the family initiated by \( x \in Z - m \), \( c(\beta p)^2 + 2\beta pq \) is the expected number of pairs of points \( \tilde{y}_1 \) and \( \tilde{y}_2 \) which are children of \( y \), and \( y_1 \) and \( y_2 \) are the \( (m - 1 - i) \)-th generation offspring of \( \tilde{y}_1 \) and \( \tilde{y}_2 \), respectively. Using Fubini’s theorem together with (23), after straightforward calculations, (39) reduces to

\[ \rho G \int_{A_1} \int_{A_2} \sum_{i=0}^{\infty} \left\{ c(\beta p)^2 f * \tilde{f} + \beta pq (f + \tilde{f}) \right\} \]

\[ \times \sum_{k_1=0}^{i} \sum_{k_2=0}^{i} (i \choose k_1) (i \choose k_2) q^{2i-k_1-k_2} (\beta p)^{k_1+k_2} f^{*k_1} \tilde{f}^{*k_2} (y_1 - y_2) \, dy_1 \, dy_2 \]

Combining this result with (38) we finally see that \( G^t \) has PCF \( g_G \) as given by (26).

C Simulating the limiting process

This appendix presents an approximate simulation procedure for simulating a special case of \( G^t \) on a bounded region \( R \subset \mathbb{R}^d \). It is available in \( \mathbb{R} \) through the package icpp, which can be obtained at [https://github.com/adchSTATS/icpp](https://github.com/adchSTATS/icpp). The implementation utilizes existing functions from the packages spatstat and RandomFields to simulate the noise process.
For simplicity and specificity we make the following assumptions. Let the situation be as in Theorem 4.2, but with \( q = 0 \) and let \( f \sim N_d(\sigma^2) \) with \( \sigma > 0 \). Also, without loss of generality, assume no thinning (i.e., \( p = 1 \)). Let \( R_{\bowtie r} = \{ \xi \in \mathbb{R}^d : b(\xi, r) \cap R \neq \emptyset \} \) where \( b(\xi, r) \) is a closed ball with centre \( \xi \) and radius \( r \geq 0 \). Denote \( n \) the number of iterations in our approximate simulation algorithm, that is, \( -n \) is the starting time when ignoring what happens previously. Note that \( \sqrt{n}\sigma \) is the standard deviation of the \( n \)th convolution power of \( f \). To account for edge effects, let \( r = 4\sqrt{n}\sigma \) where 4 is an arbitrary non-negative value ensuring that a point of \( G_{\text{st}} - n \) \( \bowtie R \) would generate a \( n \)th generation offspring in \( R \) with very low probability, at most \( 1/15787 \). In the approximate simulation procedure, we ignore those points of \( G_{\text{st}} \cap R \) which are generated by an \( i \)th generation ancestor \( x \) when \( i < -n \) or both \( -n \leq i < 0 \) and \( x \notin R_{\bowtie 4\sqrt{-i}\sigma} \). This is our algorithm in pseudocode where “parallel-for” means a parallel for loop:

```plaintext
parallel-for \( i = -n \) to 0 do
    simulate \( Z_i' := Z_i \cap R_{\bowtie 4\sqrt{-i}\sigma} \)
end parallel-for
set \( O := Z_{-n} \)
if \( n \neq 0 \) then
    for \( i = -(n - 1) \) to 0 do
        parallel-for \( x \in O \) do
            simulate the 1st generation offspring process, \( O_x \), with parent \( x \)
        end parallel-for
        set \( O := Z_i' \cup (\bigcup_{x \in O} O_x \cap R_{\bowtie 4\sqrt{-i}\sigma}) \)
    end for
end if
return \( O \)
```

Note that \( \rho Z \sum_{i=0}^n (\beta p)^i \) is the intensity of the stationary point process obtained by ignoring those points of \( G_{\text{st}}^0 \) which are generated by an \( i \)th generation ancestor with \( i < -n \). We base the choice of \( n \) on this fact by considering a precision parameter \( \varepsilon > 0 \) and letting

\[
n = \sup \left\{ m \in \{1, 2, \ldots\} : \left\| \rho Z \sum_{i=0}^m (\beta p)^i - \rho G \right\| \leq \varepsilon \right\}.
\]

To exemplify, let \( \rho G = 100 \) and \( \beta p = 0.8 \) implying that \( \rho Z = 20 \), and let \( \varepsilon = 2.22 \times 10^{-16} \), then \( n = 159 \). If instead \( \beta p = 0.99 \), then \( n = 3609 \).
Acknowledgements

Supported by The Danish Council for Independent Research | Natural Sciences, grant DFF – 7014-00074 “Statistics for point processes in space and beyond”, and by the “Centre for Stochastic Geometry and Advanced Bioimaging”, funded by grant 8721 from the Villum Foundation. We thank Ina Trolle Andersen, Yongtao Guan, Ute Hahn, Henrike Häbel, Eva B. Vedel Jensen, and Morten Nielsen for helpful comments.

References

[1] I. T. Andersen, U. Hahn, E. A. Christensen, L. N. Nejsum, and E. B. V. Jensen. Double Cox cluster processes - with applications to photoactivated localization microscopy. Spatial Statistics, 2018. To appear.

[2] O. E. Barndorff-Nielsen, J. Kent, and M. Sørensen. Normal variance-mean mixtures and z-distributions. International Statistical Review, 50:145–159, 1982.

[3] D. J. Daley and D. Vere-Jones. An Introduction to the Theory of Point Processes. Volume I: Elementary Theory and Methods. Springer-Verlag, New York, second edition, 2003.

[4] J. Felsenstein. A pain in the torus: Some difficulties with models of isolation by distance. The American Naturalist, 109:359–368, 1975.

[5] J. F. C. Kingman. Remarks on the spatial distribution of a reproducing population. Journal of Applied Probability, 14:577–583, 1977.

[6] F. Lavancier, J. Møller, and E. Rubak. Determinantal point process models and statistical inference. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 77:853–877, 2015.

[7] O. Macchi. The coincidence approach to stochastic point processes. Advances in Applied Probability, 7:83–122, 1975.

[8] B. Matérn. Spatial Variation, Meddelanden från Statens Skogsforskningsinstitut, 49 (5). 1960.

[9] B. Matérn. Spatial Variation. Lecture Notes in Statistics 36, Springer-Verlag, Berlin, 1986.

[10] P. McCullagh and J. Møller. The permanental process. Advances in Applied Probability, 38:873–888, 2006.
[11] J. Møller. Random tessellations in $\mathbb{R}^d$. Advances in Applied Probability, 21:37–73, 1989.

[12] J. Møller. Lectures on Random Voronoi Tessellations. Lecture Notes in Statistics 87. Springer-Verlag, New York, 1994.

[13] J. Møller. Shot noise Cox processes. Advances in Applied Probability, 35:614–640, 2003.

[14] J. Møller and G. L. Torrisi. Generalised shot noise Cox processes. Advances in Applied Probability, 37:48–74, 2005.

[15] J. Møller and G. L. Torrisi. The pair correlation function of spatial hawkes processes. Statistics and Probability Letters, 77:995–1003, 2007.

[16] J. Møller and R. Waagepetersen. Statistical Inference and Simulation for Spatial Point Processes. Chapman & Hall/CRC, Boca Raton, Florida, 2004.

[17] M. Myllymäki, T. Mrkvička, P. Grabarnik, H. Seijo, and U. Hahn. Global envelope tests for spatial processes. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 79:381 – 404, 2016.

[18] J. Neyman and E. L. Scott. Statistical approach to problems of cosmology. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 20:1–43, 1958.

[19] I. K. Shimatani. Spatially explicit neutral models for population genetics and community ecology: Extensions of the Neyman-Scott clustering process. Theoretical Population Biology, 77:32–41, 2010.

[20] T. Shirai and Y. Takahashi. Random point fields associated with certain Fredholm determinants I: fermion, Poisson and boson point processes. Journal of Functional Analysis, 205:414–463, 2003.

[21] M. Thomas. A generalization of Poisson’s binomial limit for use in ecology. Biometrika, 36:18–25, 1949.

[22] M. N. M. van Lieshout and A. J. Baddeley. Extrapolating and interpolating spatial patterns. In A. B. Lawson and D. Denison, editors, Spatial Cluster Modelling, pages 61–86. Chapman & Hall/CRC, Boca Raton, Florida, 2002.
[23] T. Wiegand, S. Gunatileke, N. Gunatileke, and T. Okuda. Analyzing the spatial structure of a Sri Lankan tree species with multiple scales of clustering. *Ecology*, 88:3088–3102, 2007.