Biometrika (2026), **113**, 2, asag021 https://doi.org/10.1093/biomet/asag021
Advance Access publication 28 March 2026

Palm distributions of superposed point processes for statistical inference

BY M. BERAHA, F. CAMERLENGHI

Department of Economics, Management and Statistics, University of Milano-Bicocca, Via Bicocca degli Arcimboldi 8, 20126 Milan, Italy
mario.beraha@unimib.it federico.camerlenghi@unimib.it

AND L. GHILOTTI

Department of Statistical Science, Duke University, 218 Old Chemistry, Durham, North Carolina 27708, U.S.A.
lorenzo.ghilotti@duke.edu

SUMMARY

Palm distributions play a central role in the study of point processes and their associated summary statistics. In this work, we characterize the Palm distributions of the superposition of independent point processes, establishing a simple mixture representation depending on the point processes' Palm distributions and moment measures. We explore two statistical applications enabled by our main result. First, we consider minimum contrast estimation for corrupted point processes. Second, we investigate the class of shot noise Cox processes and derive explicit expressions for their higher-order Palm distributions. In the finite case, we further obtain a tractable expression for the Janossy density, which plays the role of a likelihood function and thus can be used for new likelihood-based inference strategies. Extensions to the superposition of multiple point processes and to higher-order Palm distributions are also presented.

Some key words: Matérn cluster process; Minimum contrast estimation; Noisy observation; Shot noise Cox process; Summary statistic; Superposition of point processes.

1. INTRODUCTION

Real-world point patterns often combine several structured components, some regular or inhomogeneous and others explicitly clustered, and may also include random noise. Semiconductor wafer defect maps (Borgoni et al., 2021), disease-case locations in epidemiology (Meyer et al., 2017), cellular network base-station layouts (Choi et al., 2017), mixed-age tree stands in ecology (Ngo Bieng et al., 2011) and earthquake aftershock sequences (Ogata, 1998) can all be regarded as superpositions of two or more independent point processes. While the superposition operation is trivial at the model level, from an inferential standpoint, superposed point processes are notoriously awkward to deal with. Standard tools such as minimum contrast estimation rely on closed-form expressions for second-order summaries, such as the J-function, Ripley's K-function and Besag's L-function. However, those summaries remain unknown for a generic superposition (Møller & Waagepetersen, 2003; Diggle, 2013). Hence, practitioners need to rely on complex algorithms often developed on a case-by-case basis (e.g., Tanaka & Ogata, 2014; Xu et al., 2018).

In this work, we first focus on the Palm distributions of the superposition of two independent point processes. Palm distributions (Baccelli et al., 2020) are key mathematical objects in the study of point processes, describing the conditional behaviour of a process given the location of one or more of its points or *atoms*. We establish a mixture representation for the Palm distributions of the superposed process, in which the mixture components are sums of the two original processes and their Palm versions, weighted by their respective mean measures. We also show that our analysis easily extends to the superposition of more than two independent processes.

We demonstrate the practical usefulness of our result for statistical inference in two settings. First, we consider fitting a point process contaminated by random background noise via minimum contrast estimation. Indeed, through the Palm distributions, it is straightforward to obtain functional summary statistics, enabling robust and fast inference via minimum contrast. This method is particularly relevant in applied contexts such as the analysis of spatial defect structures in semiconductor manufacturing (Borgoni et al., 2021).

Second, we investigate structural and distributional properties of the shot noise Cox process (SNCP; Møller, 2003), a prominent class of cluster processes. Cluster processes are routinely employed in several applied areas, including astronomy, materials science and plant ecology; see, for example, Illian et al. (2008) and Møller & Waagepetersen (2003). Despite their flexibility and wide applicability, several distributional aspects of SNCP models remain only partially understood. As an application of our main result, we derive the higher-order Palm distributions of the SNCP, which were previously unavailable in the literature. In the case of finite SNCPs, we further obtain an explicit expression for the Janossy density. Since the Janossy density plays the role of a likelihood function for finite point processes, this result paves the way for new likelihood-based inference strategies in the context of SNCP models. We also discuss the use of our results in other statistical contexts, both frequentist and Bayesian.

2. Superposition of point processes

2.1. *Background and notation for point processes*

Let $\mathbb{X}$ be a Polish space equipped with corresponding Borel σ-algebra $\mathscr{X}$. A point process Φ on $\mathbb{X}$ can be represented as $\Phi = \sum_{j\geqslant 1} \delta_{X_j}$, where $(X_j)_{j\geqslant 1}$ is a sequence of random variables (atoms) taking values in $\mathbb{X}$ and δ_x denotes the Dirac delta mass at x. The probability distribution of Φ is denoted by P_Φ. The number of atoms in Φ could be either finite or infinite. We will follow the approach of Baccelli et al. (2020), where Φ is regarded as a random counting measure; see the Supplementary Material for further mathematical details, including the σ-algebra on the space of counting measures.

Let $M_\Phi(B) = E\{\Phi(B)\}$, for $B \in \mathscr{X}$, be the mean measure of Φ, and define the kth factorial moment measure $M_\Phi^{(k)}$ to be the mean measure of the kth factorial power of Φ, i.e., of the point process $\Phi^{(k)}$ defined as

$$\Phi^{(k)} := \sum_{(j_1,\ldots,j_k)}^{\neq} \delta_{(X_{j_1},\ldots,X_{j_k})},$$

where the symbol $\neq$ means that the sum is taken over all pairwise distinct indexes.

To introduce the notion of *Palm distribution*, let us define the Campbell measure of Φ, $\mathscr{C}_\Phi(B \times L) := E\{\Phi(B)\mathbb{1}(\Phi \in L)\}$, where $B \in \mathscr{X}$ and L is an element of the appropriate σ-algebra on the space of random counting measures (see the Supplementary Material). Under the assumption that M_Φ is σ-finite, it can be shown that $\mathscr{C}_\Phi$ admits the representation

$$\mathscr{C}_\Phi(B \times L) = \int_B \mathsf{P}_\Phi^x(L) M_\Phi(\mathrm{d}x),$$

where $\{\mathsf{P}_\Phi^x\}_{x\in\mathbb{X}}$ is the almost surely unique disintegration probability kernel of $\mathscr{C}_\Phi$ with respect to M_Φ and is referred to as the Palm kernel of Φ. For fixed $x \in \mathbb{X}$, P_Φ^x is a probability distribution over the space of counting measures, termed the Palm distribution of Φ at x, and thus it can be identified with the law of a point process $\Phi_x \sim \mathsf{P}_\Phi^x$, which is consequently called a *Palm version* of Φ at x. By Proposition 3.1.12 in Baccelli et al. (2020), for M_Φ-almost all $x \in \mathbb{X}$, the point process Φ_x contains the atom x with probability 1. This justifies the interpretation of the Palm distribution of Φ at x as the law of Φ conditional on Φ having an atom at x. In addition, the point process $\Phi_x^! := \Phi_x - \delta_x$ is well-defined, and $\Phi_x^!$ is called a *reduced Palm version* of Φ at x. Finally, it is possible to extend the definition of Palm distributions to multiple conditioning points $\underline{x} = (x_1, \ldots, x_k) \in \mathbb{X}^k$. In this case, the Palm distribution of Φ at $\underline{x}$ is interpreted as the probability distribution of Φ conditional on Φ having k atoms at locations $x_1, \ldots, x_k$. See the Supplementary Material.

2.2. *Palm distributions of the superposition of independent processes*

From now on, consider independent simple point processes Φ_i $(i = 1, \ldots, m)$. Their superposition Φ is defined as $\Phi := \sum_{i=1}^m \Phi_i$, i.e., the union of all the point patterns. The following theorem is the main theoretical result of this work and characterizes the Palm distributions of the superposition of two independent point processes.

THEOREM 1. *Let Φ_1 and Φ_2 be two independent point processes on $\mathbb{X}$. Then for any $x \in \mathbb{X}$, the Palm version $(\Phi_1 + \Phi_2)_x$ can be expressed as the following mixture:*

$$(\Phi_1 + \Phi_2)_x \overset{\mathrm{d}}{=} \begin{cases} \Phi_{1x} + \Phi_2 \textit{ with probability } \dfrac{\mathrm{dM}_{\Phi_1}}{\mathrm{dM}_\Phi}(\mathrm{x}), \\ \Phi_1 + \Phi_{2x} \textit{ with probability } \dfrac{\mathrm{dM}_{\Phi_2}}{\mathrm{dM}_\Phi}(\mathrm{x}). \end{cases}$$

The proof is given in the Appendix. Theorem 1 admits an intuitive interpretation of the Palm distributions of the superposed process. Conditioning the superposed process $\Phi = \Phi_1 + \Phi_2$ on having a point at x, its distribution depends on whether x originates from Φ_1 or from Φ_2, thus reflecting the respective contributions of each source. The mixture structure in the Palm version $(\Phi_1 + \Phi_2)_x$ explicitly accounts for the two mutually exclusive scenarios. The mixing probabilities correspond precisely to the probabilities that point x originates from one or the other process. Furthermore, exploiting the relation $\Phi_{ix} \overset{\mathrm{d}}{=} \Phi_{ix}^! + \delta_x$ for M_{Φ_i}-almost all x and $i = 1, 2$, we can replace all the Palm versions appearing in the statement of Theorem 1 with their reduced counterparts.

To generalize Theorem 1 to m processes and k points, we introduce latent allocation variables $T_1, \ldots, T_k \in \{1, \ldots, m\}$, with the interpretation that $T_j = i$ if and only if x_j was generated from Φ_i.

THEOREM 2. *Let $\Phi = \sum_{i=1}^m \Phi_i$ be the superposition of the independent simple point processes $\Phi_1, \ldots, \Phi_m$, and let $\underline{x} = (x_1, \ldots, x_k) \in \mathbb{X}^k$ with pairwise distinct coordinates. Assume that the kth factorial moment measure $M_\Phi^{(k)}$ is σ-finite and that $M_{\Phi_i}^{(r)}$ admits a density $\rho_{\Phi_i}^{(r)}$ with respect to $\mu^{\otimes r}$, the r-fold product of a reference measure μ, for all $i = 1, \ldots, m$ and $r \geqslant 1$. For $\underline{T} := (T_1, \ldots, T_k)$, define $\underline{x}_i = (x_j : T_j = i)$ and denote by n_i the cardinality of $\underline{x}_i$. Then for $M_\Phi^{(k)}$-almost all $\underline{x}$,*

$$\Phi_{\underline{x}}^! \mid (\underline{T} = \underline{t}) \overset{\mathrm{d}}{=} \sum_{i=1}^m (\Phi_i)_{\underline{x}_i}^!, \quad \mathrm{pr}(\underline{T} = \underline{t}) \propto \prod_{i=1}^m \rho_{\Phi_i}^{(n_i)}(\underline{x}_i),$$

where, conditionally on $\underline{T} = \underline{t}$, the $(\Phi_i)_{\underline{x}_i}^!$ are mutually independent and, when $n_i = 0$, $(\Phi_i)_{\underline{x}_i}^! = \Phi_i$ and $\rho_{\Phi_i}^{(n_i)} \equiv 1$.

3. Inference for corrupted processes via minimum contrast

3.1. *Summary statistics for superposed point processes*

An application of Theorem 1 yields closed-form and interpretable expressions for commonly used summary statistics. Here we concentrate on two summary statistics for stationary point processes; several additional examples are provided in the Supplementary Material. Throughout this section, we assume that $\Phi = \Phi_1 + \Phi_2$, where the Φ_j are stationary with intensity ρ_j, and we let $\rho = \rho_1 + \rho_2$.

Perhaps the most commonly used summary statistic is Ripley's K-function (Ripley, 1976). Thanks to Theorem 1, for a superposed point process Φ, this takes the following form for $r \geqslant 0$:

$$K_\Phi(r) := \frac{1}{\rho} E[\Phi^!_o\{B(o,r)\}] = \frac{1}{\rho}\left\{ K_{\Phi_1}(r)\frac{\rho_1^2}{\rho} + K_{\Phi_2}(r)\frac{\rho_2^2}{\rho} + 2|B(o,r)|\,\frac{\rho_1\rho_2}{\rho}\right\}, \tag{1}$$

where o denotes a generic point of $\mathbb{X}$ and $|B(o,r)|$ is the volume of the ball $B(o,r)$ with radius r centred at o, given that $B(o,r) \subset \mathbb{X}$. Thanks to stationarity, K_Φ is invariant with respect to the choice of the point o, which is called the *typical point* of Φ; the quantity $\rho K_\Phi(r)$ represents the expected number of points that are r-close to a generic point o, given that Φ has an atom at o. As pointed out by a reviewer, expression (1) can also be obtained by ad hoc computations, without applying Theorem 1. A second example of a summary statistic is the reduced Palm distribution generating function A, defined by Chiu (2008) and with expression

$$A_\Phi(s,r) := E\left[s^{\Phi^!_o\{B(o,r)\}}\right] = \frac{\rho_1}{\rho} A_{\Phi_1}(s,r) E\left[s^{\Phi_2\{B(o,r)\}}\right] + \frac{\rho_2}{\rho} A_{\Phi_2}(s,r) E\left[s^{\Phi_1\{B(o,r)\}}\right]$$

by virtue of Theorem 1. In contrast to the K-function, the A-function captures higher-order characteristics of the point process and, as shown in Chiu (2008), is particularly suited to cluster point processes with regularity. Finally, in the Supplementary Material we provide an explicit expression for the nearest-neighbour distance distribution function in the stationary case, and we also exploit Theorem 1 to evaluate summary statistics for nonstationary point processes.

3.2. *Fitting a corrupted Matérn cluster process via minimum contrast estimation*

Minimum contrast estimation (MCE) methods (Møller & Waagepetersen, 2003) constitute a class of techniques for fitting parametric point process models to observed point patterns. In a nutshell, given a functional summary statistic, such as $K_\Phi(r)$, and its nonparametric estimator $\hat{K}(r)$, the parameters of Φ are chosen to minimize a suitable distance between the theoretical functional summary statistic and its realized estimator, integrated over a specified interval for r. Typically, the distance is taken to be an L_q distance between functions; see Diggle (2013) for further details. Similarly, one can use other summary statistics, such as the A-function instead of the more common Ripley's K-function. Here, we describe an application of MCE in which Φ_1 is a Matérn cluster process, corrupted by a background noise Φ_2. An additional example is presented in the Supplementary Material, where we fit a corrupted repulsive point process.

For the sake of illustration, let $\mathbb{X} = \mathbb{R}^2$. Define the kernel function $\kappa(\xi; c) = \mathbb{1}_{B(c,R)}(\xi)/(\pi R^2)$, where $R > 0$. A Matérn cluster process Φ_1 is a Cox process defined in a hierarchical fashion: $\Phi_1 \mid \Phi_p$ is a Poisson process with intensity measure $\int_{\mathbb{R}^2} \mu\kappa(\xi; c)\Phi_p(\mathrm{d}c)\,\mathrm{d}\xi$, and $\Phi_p \sim \mathrm{PP}(\lambda\,\mathrm{d}c)$, where $\mathrm{PP}(\lambda\,\mathrm{d}c)$ denotes a Poisson process with intensity λ on $\mathbb{X}$, and $\lambda, \mu > 0$. Equivalently, Φ_1 is a cluster process: the parent points are generated by the latent parent process Φ_p, while each cluster of Φ_1 consists of points uniformly distributed in $B(c_j, R)$, where $\Phi_p = \sum_{j\geqslant 1}\delta_{c_j}$.

Assume that we observe a realization of the Matérn cluster process Φ_1 corrupted by a background noise independent of Φ_1. We further suppose that the corrupting noise Φ_2 is a homogeneous Poisson point process with intensity ρ_2. We propose to fit the superposition $\Phi_1 + \Phi_2$ to the observed point pattern via MCE. We compare three different approaches: two based on the A- and K-functions of the superposed process, and one where we ignore the superposition and fit MCE based on the

Table 1. *Median and interquartile range (in parentheses) of the estimates for the superposed Matérn–Poisson model over* 200 *independent replicated datasets*

Window	Method	ρ_1	μ	R	ρ_2
	A-MCE (correct)	53.7 (24.5)	5.42 (2.05)	0.05 (0.01)	19.3 (30.8)
$[0,1]^2$	K-MCE (correct)	36.1 (16.4)	7.09 (3.65)	0.05 (0.02)	57.5 (16.8)
	A-MCE (misspecified)	67.5 (29.9)	4.60 (1.37)	0.05 (0.01)	—
	A-MCE (correct)	50.1 (4.9)	5.03 (0.39)	0.05 (0.00)	20.2 (6.7)
$[0,5]^2$	K-MCE (correct)	42.1 (6.3)	5.54 (0.70)	0.05 (0.00)	36.8 (4.9)
	A-MCE (misspecified)	64.2 (3.7)	4.28 (0.24)	0.05 (0.00)	—

A-function of Φ_1 (this is equivalent to setting $\rho_2 \equiv 0$). See the Supplementary Material for the explicit expressions of A and K for our Matérn–Poisson superposition. We generate data from a Matérn point process with parameters $\rho_1 = 50$, $\mu = 5$ and $R = 0.05$ and corrupt it with a Poisson process with $\rho_2 = 20$. We consider two observation windows, $[0,1]^2$ and $[0,5]^2$.

The results reported in Table 1 highlight clear differences between the three estimators. The A-MCE under the correct superposed model is well centred around the true parameter values, and its dispersion decreases significantly as the observation window grows, as expected from increasing information. In contrast, the A-MCE that ignores the Poisson background exhibits substantial bias in estimating ρ_1, while still delivering accurate estimates of μ and R. Finally, the K-MCE tends to overestimate ρ_2 and hence underestimate ρ_1, underscoring that capturing features beyond second-order characteristics can be crucial for reliable inference in cluster processes, especially under superposition.

4. Shot noise Cox processes: Palm distributions and Janossy densities

SNCPs constitute a class of general models for clustered point patterns. To define an SNCP, let $\nu(\mathrm{d}\theta\,\mathrm{d}\gamma)$ be a locally finite diffuse intensity measure on $\mathbb{X} \times \mathbb{R}_+$. Let $\{\kappa(\cdot\,;\theta)\}_{\theta\in\mathbb{X}}$ be a family of parametric probability density functions on $\mathbb{X}$, called the kernel of the SNCP, where the parameter space is $\mathbb{X}$ itself. Assuming that $\int_{\mathbb{X}\times\mathbb{R}_+} \gamma\kappa(x;\theta)\nu(\mathrm{d}\theta\,\mathrm{d}\gamma) < \infty$ for any $x \in \mathbb{X}$, we have that Φ is a SNCP directed by ν with kernel κ if

$$\Phi|\Lambda \sim \mathrm{PP}\left\{\int_{\mathbb{X}\times\mathbb{R}_+} \gamma\kappa(x;\theta)\Lambda(\mathrm{d}\theta\,\mathrm{d}\gamma)\mathrm{d}x\right\}, \quad \Lambda \sim \mathrm{PP}(\nu). \tag{2}$$

We write $\Phi \sim \mathrm{SNCP}(\kappa,\nu)$. An SNCP is a *cluster process*: the colouring theorem of Poisson processes entails that Φ equals in distribution the sum $\sum_{i\geqslant 1}\Phi_i$, where $\Phi_i|\Lambda \overset{\text{ind}}{\sim} \mathrm{PP}\{\gamma_i\kappa(x;\theta_i)\,\mathrm{d}x\}$ and $\Lambda = \sum_{i\geqslant 1}\delta_{(\theta_i,\gamma_i)}$. Then for each point X_j of Φ, it is possible to introduce a latent cluster allocation variable T_j such that $\Phi_i(\{X_j\}) = 1$ if and only if $T_j = i$ and $\Phi_i(\{X_j\}) = 0$ otherwise, i.e., $T_j = i$ if X_j arose from Φ_i. The number of distinct values across $\underline{T} := (T_j)_{j\geqslant 1}$, denoted by $|\underline{T}|$, represents the number of clusters among the points. Wang et al. (2023) exploited this construction to draw a connection with Bayesian mixtures of finite mixtures (Lijoi et al., 2008; Miller & Harrison, 2018) in the case where Λ is a finite Poisson process and the γ_i are independent and gamma-distributed.

The Palm distributions of an SNCP at a single point x were obtained by Møller (2003), but higher-order Palm and reduced Palm distributions for SNCPs have not appeared in the literature. We fill this gap with the next theorem, which follows from a recursive application of Theorem 1. Define $\eta(x_1,\ldots,x_p) = \int_{\mathbb{X}\times\mathbb{R}_+}\gamma^p\prod_{j=1}^p\kappa(x_j;\theta)\nu(\mathrm{d}\theta\,\mathrm{d}\gamma)$ and assume that $\eta(x_1,\ldots,x_p) < \infty$ for any $x_1,\ldots,x_p$. The latter assumption imposes constraints on the choices of κ and ν.

Theorem 3. *Let* $\Phi \sim \mathrm{SNCP}(\kappa, \nu)$ *and* $\underline{x} = (x_1, \ldots, x_k) \in \mathbb{X}^k$. *Then the reduced Palm version of* Φ *at* $\underline{x}$ *admits the representation*

$$\Phi^!_{\underline{x}} | \underline{T} \overset{\mathrm{d}}{=} \Phi + \sum_{h=1}^{|\underline{T}|} \Phi_{\zeta_{\underline{x}_h}}, \quad \mathrm{pr}(\underline{T} = \underline{t}) \propto \prod_{h=1}^{|\underline{t}|} \eta(\underline{x}_h),$$

where $\underline{T} := (T_1, \ldots, T_k)$ *are latent indicators describing a partition of* $\underline{x}$ *into* $|\underline{T}|$ *clusters and*

$$\Phi_{\zeta_{\underline{x}_h}} | \zeta_{\underline{x}_h} = (\theta_{\underline{x}_h}, \gamma_{\underline{x}_h}) \sim \mathrm{PP}\{\gamma_{\underline{x}_h} \kappa(x; \theta_{\underline{x}_h})\, \mathrm{d}x\},$$
$$\zeta_{\underline{x}_h} \sim f_{\underline{x}_h}(\mathrm{d}\theta\, \mathrm{d}\gamma) \propto \gamma^{n_h} \prod_{x_j \in \underline{x}_h} \kappa(x_j; \theta) \nu(\mathrm{d}\theta\, \mathrm{d}\gamma),$$

with $\underline{x}_h = (x_j : T_j = h)$ *and* n_h *denoting the cardinality of* $\underline{x}_h$. *Finally, the processes* Φ *and* $\Phi_{\zeta_{\underline{x}_h}}$ $(h = 1, \ldots, |\underline{T}|)$ *are mutually independent conditional on* $\underline{T}$.

See the Supplementary Material for the precise definition of the space where $\underline{T}$ takes its values.

Theorem 3 plays a key role in deriving the likelihood of an SNCP, thereby paving the way for new strategies in maximum likelihood estimation. Assume that Φ has almost surely a finite number of points. Following Daley & Vere-Jones (2003), we define the likelihood of Φ to be its Janossy density viewed as a function of the parameters of Φ. Briefly, we recall that for a regular finite point process Φ with associated Janossy density $j_k : \mathbb{X}^k \to \mathbb{R}_+$, $j_k(x_1, \ldots, x_k)\, \mathrm{d}x_1 \ldots \mathrm{d}x_k$ represents the probability that Φ consists of exactly k points located in infinitesimal neighbourhoods of x_j $(j = 1, \ldots, k)$. Under suitable regularity conditions, the family of Janossy densities $j_k(\cdot)$ $(k \geqslant 0)$ characterizes its probability distribution (Daley & Vere-Jones, 2003, Proposition 5.3.II). The next theorem, derived from Theorem 3, gives an explicit expression for the Janossy densities when Φ is a general finite SNCP. We remark that a sufficient condition for the finiteness of Φ is $\int_{\mathbb{X} \times \mathbb{R}_+} \gamma \nu(\mathrm{d}\theta\, \mathrm{d}\gamma) < \infty$. For simplicity, we present here only the case where $\nu(\mathrm{d}\theta\, \mathrm{d}\gamma) = \rho(\mathrm{d}\gamma) G_0(\mathrm{d}\theta)$; see the Supplementary Material for the general statement and the proof.

Theorem 4. *Let* $\Phi \sim \mathrm{SNCP}(\kappa, \nu)$ *such that* $\Phi(\mathbb{X}) < \infty$ *almost surely and* $\nu(\mathrm{d}\theta\, \mathrm{d}\gamma) = \rho(\mathrm{d}\gamma) G_0(\mathrm{d}\theta)$. *Then*

$$j_k(x_1, \ldots, x_k) = k!\, \mathrm{pr}(\Phi(\mathbb{X}) = k)\, E\left\{ \prod_{h=1}^{|\underline{T}|} \int_{\mathbb{X}} \prod_{j : T_j = h} \kappa(x_j; \theta) G_0(\mathrm{d}\theta) \right\}, \tag{3}$$

where the expectation is taken with respect to the indicators $\underline{T} = (T_1, \ldots, T_k)$ *with distribution*

$$\mathrm{pr}(\underline{T} = \underline{t}) \propto \prod_{h=1}^{|\underline{t}|} \int_{\mathbb{R}_+} \exp(-\gamma) \gamma^{n_h} \rho(\mathrm{d}\gamma). \tag{4}$$

The expression of the Janossy density in Theorem 4 naturally enables maximum likelihood estimation for $\Phi \sim \mathrm{SNCP}(\kappa, \nu)$, obtained by selecting the parameters that maximize j_k in (3) and (4). In particular, the structure revealed by Theorem 4 suggests the potential use of expectation-maximization algorithms for parameter estimation, thereby providing a viable alternative to existing inferential methods (Møller & Waagepetersen, 2003). This line of research is currently under active investigation and will be the subject of future studies.

5. Discussion

The implications of Theorem 1 and its generalizations extend well beyond the two concrete applications presented in this work. As discussed in § 3.1, our main result enables the evaluation of summary statistics for both stationary and nonstationary point processes, which is useful for both exploratory analysis and model checking via, say, envelope tests (Chiu et al., 2013).

We also envision applications in the field of Bayesian nonparametrics, where superpositions of point processes are used to define prior distributions when data are divided into groups (Griffin et al., 2013; Lijoi et al., 2014); Theorem 1 can be used for posterior analysis and numerical computations in such models. Finally, Theorem 3 plays a key role in deriving posterior representations in settings where the SNCP serves as a Bayesian nonparametric prior. Relevant examples include Bayesian mixture modelling (Beraha et al., 2025a) and extended feature allocation models, which are extensions of Bayesian clustering models that allow individuals to share multiple features (Beraha et al., 2025b).

Acknowledgement

The authors are grateful to two referees for their insightful comments and suggestions, which have helped to substantially improve the paper. F.C. gratefully acknowledges the support of MUR - Prin 2022 (grant no. 2022CLTYP4), funded by the European Union – Next Generation EU. L.G. gratefully acknowledges support from the U.S. Office of Naval Research (N00014-24-1-2626-P00002).

Supplementary material

The Supplementary Material contains further details on point processes and proofs of the theorems.

Appendix

Proof of Theorem 1

Let $\Phi = \Phi_1 + \Phi_2$, with Φ_1 independent of Φ_2. By Baccelli et al., (2020, Proposition 3.2.1), the Palm distribution of a point process Φ is uniquely characterized by the following relation: for all measurable $f, g : \mathbb{X} \to \mathbb{R}_+$ such that g is M_Φ-integrable,

$$\frac{\partial}{\partial t}\mathcal{L}_\Phi(f + tg)\Big|_{t=0} = -E\left[\Phi(g)\exp\{-\Phi(f)\}\right] = -\int_{\mathbb{X}} g(x)\mathcal{L}_{\Phi_x}(f)M_\Phi(\mathrm{d}x), \tag{A1}$$

where $\mathcal{L}_\Phi(f)$ denotes the Laplace functional of Φ evaluated at f, i.e., $\mathcal{L}_\Phi(f) = E\big[\exp\{-\int_{\mathbb{X}} f(x)\Phi(\mathrm{d}x)\}\big]$, and $\Phi(f) := \int_{\mathbb{X}} f(x)\Phi(\mathrm{d}x)$. Therefore, we have

$$\begin{aligned} E\left[\Phi(g)\exp\{-\Phi(f)\}\right] &= E\left[\int_{\mathbb{X}} g(x)\exp\left\{-\int_{\mathbb{X}} f(y)(\Phi_1+\Phi_2)(\mathrm{d}y)\right\}\Phi_1(\mathrm{d}x)\right] \\ &\quad + E\left[\int_{\mathbb{X}} g(x)\exp\left\{-\int_{\mathbb{X}} f(y)(\Phi_1+\Phi_2)(\mathrm{d}y)\right\}\Phi_2(\mathrm{d}x)\right] \\ &= E\left[\Phi_1(g)\exp\{-\Phi_1(f)\}\right]E[\exp\{-\Phi_2(f)\}] \\ &\quad + E\left[\Phi_2(g)\exp\{-\Phi_2(f)\}\right]E[\exp\{-\Phi_1(f)\}], \end{aligned}$$

where the last equality follows from the independence of Φ_1 and Φ_2. Then, applying Proposition 3.2.1 from Baccelli et al. (2020) again, we obtain

$$E[\Phi(g)\exp\{-\Phi(f)\}] = \left\{\int_{\mathbb{X}} g(x)\mathcal{L}_{\Phi_{1x}}(f)M_{\Phi_1}(\mathrm{d}x)\right\}\mathcal{L}_{\Phi_2}(f) + \left\{\int_{\mathbb{X}} g(x)\mathcal{L}_{\Phi_{2x}}(f)M_{\Phi_2}(\mathrm{d}x)\right\}\mathcal{L}_{\Phi_1}(f).$$

Since $M_\Phi(\mathrm{d}x) = M_{\Phi_1}(\mathrm{d}x) + M_{\Phi_2}(\mathrm{d}x)$ and M_{Φ_i} is absolutely contintuous with respect to M_Φ, we write

$$E[\Phi(g)\exp\{-\Phi(f)\}] = \int_{\mathbb{X}} g(x) \left\{ \mathcal{L}_{\Phi_2}(f)\mathcal{L}_{\Phi_{1x}}(f)\frac{\mathrm{d}M_{\Phi_1}}{\mathrm{d}M_\Phi}(x) + \mathcal{L}_{\Phi_1}(f)\mathcal{L}_{\Phi_{2x}}(f)\frac{\mathrm{d}M_{\Phi_2}}{\mathrm{d}M_\Phi}(x) \right\} M_\Phi(\mathrm{d}x).$$

Therefore, from (A1), it follows that

$$\mathcal{L}_{\Phi_x}(f) = \mathcal{L}_{\Phi_2}(f)\mathcal{L}_{\Phi_{1x}}(f)\frac{\mathrm{d}M_{\Phi_1}}{\mathrm{d}M_\Phi}(x) + \mathcal{L}_{\Phi_1}(f)\mathcal{L}_{\Phi_{2x}}(f)\frac{\mathrm{d}M_{\Phi_2}}{\mathrm{d}M_\Phi}(x),$$

which is an alternative expression of the statement of the theorem.

REFERENCES

BACCELLI, F., BŁASZCZYSZYN, B. & KARRAY, M. (2020). Random measures, point processes, and stochastic geometry. HAL preprint, available at https://hal.inria.fr/hal-02460214/.

BERAHA, M., ARGIENTO, R., CAMERLENGHI, F. & GUGLIELMI, A. (2025a). Bayesian mixture models with repulsive and attractive atoms. *J. R. Statist. Soc. B* **87**, 1481–507.

BERAHA, M., CAMERLENGHI, F. & GHILOTTI, L. (2025b). Bayesian calculus and predictive characterizations of extended feature allocation models. *arXiv:* 2502.10257.

BORGONI, R., GALIMBERTI, C. & ZAPPA, D. (2021). Identification of spatial defects in semiconductor manufacturing. *Appl. Stoch. Mod. Bus. Indust.* **37**, 878–93.

CHIU, S. N. (2008). Reduced Palm distribution generating function for a spatial point process. *J. Statist. Comp. Simul.* **78**, 1241–50.

CHIU, S. N., STOYAN, D., KENDALL, W. S. & MECKE, J. (2013). *Stochastic Geometry and Its Applications*. Wiley Series in Probability and Statistics. Chichester: John Wiley & Sons, 3rd ed.

CHOI, C.-S., WOO, J. O. & ANDREWS, J. G. (2017). An analytical framework for modeling a spatially repulsive cellular network. *arXiv:* 1701.02261v3.

DALEY, D. J. & VERE-JONES, D. (2003). *An Introduction to the Theory of Point Processes: Volume I: Elementary Theory and Methods*. New York: Springer, 2nd ed.

DIGGLE, P. J. (2013). *Statistical Analysis of Spatial and Spatio-Temporal Point Patterns*. Boca Raton, Florida: CRC Press.

GRIFFIN, J. E., KOLOSSIATIS, M. & STEEL, M. F. (2013). Comparing distributions by using dependent normalized random-measure mixtures. *J. R. Statist. Soc. B* **75**, 499–529.

ILLIAN, J., PENTTINEN, A., STOYAN, H. & STOYAN, D. (2008). *Statistical Analysis and Modelling of Spatial Point Patterns*. Statistics in Practice. Chichester: John Wiley & Sons.

LIJOI, A., NIPOTI, B. & PRÜNSTER, I. (2014). Bayesian inference with dependent normalized completely random measures. *Bernoulli* **20**, 1260–91.

LIJOI, A., PRÜNSTER, I. & WALKER, S. G. (2008). Bayesian nonparametric estimators derived from conditional Gibbs structures. *Ann. Appl. Prob.* **18**, 1519–47.

MEYER, S., HELD, L. & HÖHLE, M. (2017). Spatio-temporal analysis of epidemic phenomena using the `R` package `surveillance`. *J. Statist. Software* **77**, 1–55.

MILLER, J. W. & HARRISON, M. T. (2018). Mixture models with a prior on the number of components. *J. Am. Statist. Assoc.* **113**, 340–56.

MØLLER, J. (2003). Shot noise Cox processes. *Adv. Appl. Prob.* **35**, 614–40.

MØLLER, J. & WAAGEPETERSEN, R. P. (2003). *Statistical Inference and Simulation for Spatial Point Processes*. Boca Raton, Florida: CRC Press.

NGO BIENG, M. A., GINISTY, C. & GOREAUD, F. (2011). Point process models for mixed sessile forest stands. *Ann. Forest Sci.* **68**, 267–74.

OGATA, Y. (1998). Space–time point-process models for earthquake occurrences. *Ann. Inst. Statist. Math.* **50**, 379–402.

RIPLEY, B. D. (1976). The second-order analysis of stationary point processes. *J. Appl. Prob.* **13**, 255–66.

TANAKA, U. & OGATA, Y. (2014). Identification and estimation of superposed Neyman-Scott spatial cluster processes. *Ann. Inst. Statist. Math.* **66**, 687–702.

WANG, Y., DEGLERIS, A., WILLIAMS, A. & LINDERMAN, S. W. (2023). Spatiotemporal clustering with Neyman-Scott processes via connections to Bayesian nonparametric mixture models. *J. Am. Statist. Assoc.* **119**, 2382–95.

XU, H., LUO, D., CHEN, X. & CARIN, L. (2018). Benefits from superposed Hawkes processes. In *Proc. 21st Int. Conf. Artificial Intelligence and Statistics, vol. 84 of Proc. Mach. Learn. Res.* MLResearch Press, pp. 623–31.

[*Received on* 28 *August* 2025. *Accepted on* 16 *March* 2026]