

On Spatial Point Processes With Composition-Valued Marks

Matthias Eckardt¹ , Mari Myllymäki² and Sonja Greven¹

¹Humboldt-Universität zu Berlin, Berlin, Germany

²Natural Resources Institute Finland (Luke), Helsinki, Finland

Corresponding to: Matthias Eckardt, Humboldt-Universität zu Berlin, Berlin, Germany.

Email: m.eckardt@hu-berlin.de

Summary

Methods for marked spatial point processes with scalar marks have seen extensive development in recent years. While the impressive progress in data collection and storage capacities has yielded an immense increase in spatial point process data with highly challenging non-scalar marks, methods for their analysis are not equally well developed. In particular, there are no methods for composition-valued marks, that is, vector-valued marks with a sum-to-constant constrain (typically 1 or 100). Prompted by the need for a suitable methodological framework, we extend existing methods to spatial point processes with composition-valued marks and adapt common mark characteristics to this context. The proposed methods are applied to analyse spatial correlations in data on tree crown-to-base and business sector compositions.

Key words: business sector composition; compositional data analysis; crown-to-base ratios; mark correlation function; mark variogram; marked spatial point processes.

1 Introduction

In recent years, there has been substantial interest in the analysis of marked spatial point processes. This type of data involves random locations of n events with marks that provide additional specific details about each event. Although spatial point process theory is mature, and various summaries for marked points have been proposed (see Illian *et al.*, 2008; Chiu *et al.*, 2013), there is still a gap in addressing non-scalar marks. Particularly, spatial point processes with compositional marks, where marks are made up of D non-negative components that together sum to a constant, have not yet been addressed. Examples include pest infestation ratios in wood samples, local proportions of business sectors, and fractions of tree biomass. Here, marks present relative information with interdependent D components; an increase in one by necessity results in a decrease in the others, due to the fixed sum. This constrained nature demands analysis techniques beyond traditional Euclidean methods. An analysis of individual components with existing methods disregards these constraints, risking biased or spurious results (Chayes, 1960). Thus, a methodological framework for analysing such composition-valued marks remains crucial in modern applications. This paper seeks to address this gap by introducing a new class of spatial point processes with composition-valued marks. By applying compositional data analysis principles (Aitchison, 1986) to marked spatial point processes, this work goes beyond existing methods, establishing novel summary characteristics

for complex mark scenarios. To our knowledge, composition-valued marked point processes are examined here for the first time.

Various summary characteristics are standard tools to characterise both the properties of points and marks (Illian *et al.*, 2008; Chiu *et al.*, 2013). Of all established methods, functional summary characteristics, in which the characteristic is defined as a function of the distance between points r , play a particularly important role. They are used to investigate the marked point pattern, to determine a suitable model and to evaluate the goodness-of-fit of such models (e.g. Illian *et al.*, 2008; Myllymäki *et al.*, 2017).

Despite the growing availability of highly demanding spatial marked point process scenarios, the literature focused, so far, almost exclusively on the analysis of scalar-valued attributes (see Baddeley, 2010, for a general treatment). For integer-valued marks, where the points are assigned to exactly one out of $k \geq 2$ distinct types, classic nearest-neighbour and pairwise distance based summary characteristics were extended to so-called *cross*- and *dot*-type versions (Lotwick & Silverman, 1982; Harkness & Isham, 1983; Van Lieshout & Baddeley, 1999). For real-valued marks, the aim has mainly been to quantify the average association/variation among the marks. This includes mark-to-mark (Stoyan, 1984b; 1987; Stoyan & Wälder, 2000; Schlather, 2001), point-to-mark (Schlather *et al.*, 2004; Guan, 2006; Guan & Afshartous, 2007; Ho & Stoyan, 2008), and mark-weighted summary characteristics (Penttinen *et al.*, 1992; van Lieshout, 2006) as well as frequency domain approaches (Eckardt & Mateu, 2019b). Stoyan (1987) considered the analysis of two distinct real-valued marks, and (Penttinen *et al.*, 1992), (Wiegand & Moloney, 2013) and (Eckardt & Mateu, 2019a; 2019b) discussed methods for mixtures of integer- and real-valued marks. Further, function-valued marks (Comas *et al.*, 2008; Ghorbani *et al.*, 2021) and generalisations to multivariate function-valued marks (Eckardt *et al.*, 2025) have been investigated (see Eckardt & Moradi, 2024, for a general review). However, there are no methods available for the joint analysis of (constrained) vector-valued marks.

This paper aims to characterise the spatial association for pairs of points with D -part composition-valued marks. Instead of investigating the component parts separately, the underlying idea is to treat the observed marks as part of a total, and to transfer the main principles of compositional data analysis to marked spatial point processes. Composition-valued geostatistical and areal data has already received much attention, including various structural analysis and kriging approaches (Pawlowsky-Glahn & Ricardo, 2004; Tolosana Delgado, 2006) and areal regression models (Leininger *et al.*, 2013). However, no extensions to marked point processes with composition-valued marks exist so far. As our main contribution, we introduce a general framework of composition-valued marks and define appropriate mark characteristics for them.

The remainder of the paper is structured as follows. Section 2 summarises the present methodological toolbox for real-valued marks we build on. Section 3 introduces composition-valued marks, suitable mark transformations and first-order tools (Sections 3.1–3.3), different mark summary characteristics (Sections 3.4–3.5), extensions to composition-valued marks with total information (Section 3.6) and introduces estimation for the new methods (Section 3.7). Section 4 considers testing of the basic random labelling hypothesis for composition-valued marks. Applications of the proposed methods to a Finnish tree pattern and Spanish business sector data are provided in Section 5. The paper closes with a conclusion in Section 6.

2 Summary Characteristics for Real-Valued Point Attributes

As compositions can be seen as a constrained vector of real-valued marks, the following discussion mostly restricts to methods for the real-valued marks scenario, which are to be extended for composition-valued marks in Section 3.

2.1 Preliminaries

Let $X = \{x_i, m(x_i)\}_{i=1, \dots, n}$ denote a marked spatial point process on $\mathbb{R}^2 \times \mathbb{M}$ with points x_i in $\mathbb{R}^2$ and associated marks $m(x_i)$ living in a Polish space $\mathbb{M}$ equipped with a σ -algebra $\mathcal{M}$ and an appropriate reference measure ϖ . The mark distribution will be denoted by M . The Borel σ -algebra of $\mathbb{R}^2$ is denoted by $\mathcal{B}$. The observed point pattern and the related unmarked point process will be denoted by $\mathbf{x}$ and $\check{X}$, respectively. In what follows, X is assumed to be simple, where simplicity means that multiple coincident points do not occur. For X , the expected number $N(\cdot)$ of points in $B \in \mathcal{B}$ with marks in $L \in \mathcal{M}$ is $\Lambda(B \times L) = \mathbb{E}(N(B \times L))$.

Further, X is said to be stationary if $\{x_i, m(x_i)\} \stackrel{d}{=} \{x_i + s, m(x_i)\}$ for all $s \in \mathbb{R}^2$ and isotropic if $\{x_i, m(x_i)\} \stackrel{d}{=} \{\mathcal{R}x_i, m(x_i)\}$ for any rotation $\mathcal{R} \in \mathbb{R}^2$, $\stackrel{d}{=}$ denoting equality in distribution. For stationary X , $\Lambda(B \times L)$ simplifies to $\Lambda(B \times L) = \lambda v(B)M(L)$, where λ is the intensity of $\check{X}$, and $v(\cdot)$ is the Lebesgue measure. If $\mathbb{M} = \mathbb{R}$, M is completely determined by the mark distribution function $F_M(m) = M((-\infty, m])$ for $-\infty \leq m \leq \infty$, where

$$F_M(m) = \int_{-\infty}^m f_M(\tilde{m}) d\tilde{m},$$

with $f_M(\cdot)$ denoting the mark density function if it exists. Here, $f_M(L) := \int_L f_M(l) dl$ can be interpreted as the probability that at an arbitrarily chosen point the mark is in L . Useful quantities of the mark distribution are the mean mark μ_M and the mark variance σ_M^2 . Finally, the second-order factorial moment measure of $\check{X}$ plays an important role in the second-order summary characteristics. It is defined as

$$\alpha^{(2)}(B_1 \times B_2) = \int_{B_1} \int_{B_2} \varrho^{(2)}(x_1, x_2) dx_1 dx_2 \quad (1)$$

where $\varrho^{(2)}$ is the second-order product density. Heuristically, for any $x_1, x_2 \in \mathbb{R}^2$, $\varrho^{(2)}(x_1, x_2) dx_1 dx_2$ can be interpreted as the probability of observing exactly one point in each of the infinitesimal areas dx_1 and dx_2 .

2.2 Functional Summary Characteristics for Real-Valued Marks

Within the last decades various mark characteristics were introduced. These characteristics either describe the average pairwise variation or association of the marks at two point locations as a function of the interpoint distance r . Thus, these characteristics are useful tools to detect and quantify spatial dependencies in the marks as well as dependencies of the marks on the points. Prominent cases include Stoyan's mark correlation and covariance functions (Stoyan, 1984a), the mark variogram (Cressie, 1993), Isham's mark correlation function (Isham, 1985), and Schlather's (Schlather *et al.*, 2004) and Shimatani's (Shimatani, 2002) I functions. For example, the mark variogram, a widely used instance of mark summary characteristics, investigates the squared difference of marks of two points at a distance r apart from each other, and relates it to the same quantity under independent marks. It has small values if the two marks are similar and large values if they differ strongly. While the name and character of the mark variogram is similar to the geostatistical variogram, their definitions differ. Namely, mathematically, all the mark characteristics are defined exclusively for stationary point processes and are conditional quantities (in a Palm sense, see Chiu *et al.*, 2013), that is, conditional on that there are indeed points at location $\circ$ and $\mathbf{r}$ in $\mathbb{R}^2$. For example, the conditional expectation of the mark (say tree height) at the first point, another characteristic, may be smaller than under independence if there is a second point at distance r (say another tree at small distance r) due to competition. They are commonly constructed by taking the conditional expectation $\mathbb{E}_{\circ, r}$ of a so-called *test function*

$\mathcal{T}_f: \mathbb{M} \times \mathbb{M} \rightarrow \mathbb{R}^+$, which takes the marks $m(\circ)$ and $m(\mathbf{r})$ at the origin $\circ$ and any alternative points at distance $\|\mathbf{r}\| = r > 0$ from $\circ$ as its arguments (Penttinen & Stoyan, 1989). Without imposing any invariance assumptions, the conditional expectation is formally defined with respect to the joint distribution of the marks $m(x_1)$ and $m(x_2)$ for any two points $x_1, x_2 \in \mathbb{R}^2$, that is, the so-called two-point mark distribution $M_{x_1, x_2}(dm(x_1)dm(x_2))$. For motion-invariant X , M_{x_1, x_2} depends on the points only through the Euclidean distance $\|x_1 - x_2\| = r$ and can be written as M_r . For $L_1, L_2 \in \mathcal{M}$, $M_r(L_1 \times L_2)$ corresponds to the probability of having $m(\circ) \in L_1$ and $m(\mathbf{r}) \in L_2$ under the condition that there are indeed points at $\circ$ and $\mathbf{r}$ at a distance r from each other in $\check{X}$. Note that $M_r(L_1 \times L_2) = \varrho^{(2)}(r, L_1, L_2)/\varrho^{(2)}(r)$ simplifies to $(\varrho^{(2)}(r)M(L_1)M(L_2))/\varrho^{(2)}(r) = M(L_1)M(L_2)$ under independent marks for all L_1 and L_2 in $\mathcal{M}$, where $\varrho^{(2)}(r, L_1, L_2)$ is the second-order product density of

$$\alpha_m^{(2)}(B_1 \times L_1 \times B_2 \times L_2) = \int_{B_1 \times B_2} M_r(L_1 \times L_2) \alpha^{(2)}\{d(x_1, x_2)\}$$

and $\varrho^{(2)}(r)$ is the second-order product density as in (1) for x_1, x_2 at distance r (Penttinen & Stoyan, 1989). Less formally, $\alpha_m^{(2)}$ measures how many point pairs are expected to occur, on average, with one point in area B_1 and mark in L_1 and another in area B_2 with mark in L_2 . Similarly, $\varrho^{(2)}(r)$ tells us how likely it is to find a pair of marked points x_1 and x_2 of distance r , simultaneously. Using the above notation, functional mark characteristics are constructed as follows. Denoting by $\nabla_{\mathcal{T}_f}(r) = \mathbb{E}_{\circ, r}[\mathcal{T}_f(m(\circ), m(\mathbf{r}))]$ and writing $\nabla_{\mathcal{T}_f} = \nabla_{\mathcal{T}_f}(\infty)$ for the expected value of the chosen test function $\mathcal{T}_f$ at very large distances, that is, when the marks are expected to be independent, the specific mark characteristic itself is determined by the specific choice of the test function $\mathcal{T}_f$ (see Illian *et al.*, 2008). Formally, $\nabla_{\mathcal{T}_f}(r)$ can be expressed as the ratio of two product density functions,

$$\nabla_{\mathcal{T}_f}(r) = \varrho_{\mathcal{T}_f}^{(2)}(r)\varrho^{(2)}(r), \quad (2)$$

that is, the densities of the $\mathcal{T}_f$ -factorial moment measure $\alpha_{\mathcal{T}_f}^{(2)}(B_1 \times B_2)$,

$$\alpha_{\mathcal{T}_f}^{(2)}(B_1 \times B_2) = \mathbb{E} \left[\sum_{\substack{(x_1, m(x_1)), \\ (x_2, m(x_2)) \in X}}^{\neq} \mathcal{T}_f(m(x_1), m(x_2)) 1_{B_1}(x_1) \sum_{B_2}^{\neq}(x_2) \right], \quad (3)$$

and of the factorial moment measures $\alpha^{(2)}(B_1 \times B_2)$ of (1), respectively, where $B_1, B_2 \in \mathcal{B}$, 1 is an indicator function and $\sum^{\neq}$ denotes the sum over distinct pairs of points. When $r \rightarrow \infty$, the marks are assumed to be independent, $\nabla_{\mathcal{T}_f}(r)$ becomes independent of the points and simplifies to

$$\nabla_{\mathcal{T}_f} = \int_{\mathbb{M}} \int_{\mathbb{M}} \mathcal{T}_f(m_1, m_2) F_M(dm_1) F_M(dm_2). \quad (4)$$

An overview of the most prominent specifications for $\mathcal{T}_f$ is presented in Table 1.

Evaluating the test functions of Table 1 immediately yields different functional mark summary characteristics, which we briefly discuss next. With the exception of $\mathcal{T}_5$ and $\mathcal{T}_6$, the normalising factor $\nabla_{\mathcal{T}_f}$ (3rd column) corresponds to the expected value of the test function if $r \rightarrow \infty$ and follows directly from the solution of (4) with $\mathcal{T}_f(m_1, m_2)$ replaced by the specific test function as given in the 2nd column of Table 1 (see Illian *et al.*, 2008). For $\mathcal{T}_5$ and $\mathcal{T}_6$, however,

Table 1. Overview of prominent test function specifications with μ_M and σ_M^2 denoting the unconditional mark mean and mark variance, $\mu_M(\mathbf{r})$ the mean for the second point at location $\mathbf{r}$ under the condition that there are points at $\circ$ and $\mathbf{r}$.

Name for $\mathcal{T}_f$	Test function $\mathcal{T}_f$	Normalising factor $\nabla_{\mathcal{T}_f}$	Notation for $\nabla_{\mathcal{T}_f}(r)$	Notation for $\kappa_{\mathcal{T}_f}(r)$
$\mathcal{T}_1$	$m(\circ)m(\mathbf{r})$	μ_M^2	$\tau_{mm}(r)$	$\kappa_{mm}(r)$
$\mathcal{T}_2$	$m(\circ)$	μ_M	$\tau_{m\bullet}(r)$	$\kappa_{m\bullet}(r)$
$\mathcal{T}_3$	$m(\mathbf{r})$	μ_M	$\tau_{\bullet m}(r)$	$\kappa_{\bullet m}(r)$
$\mathcal{T}_4$	$0.5(m(\circ) - m(\mathbf{r}))^2$	σ_M^2	$\gamma_{mm}(r)$	$\gamma_{mm}^n(r)$
$\mathcal{T}_5$	$(m(\circ) - \mu_M)(m(\mathbf{r}) - \mu_M)$	σ_M^2	$I_{mm}^{\text{Shi}}(r)$	$I_{mm}^{\text{Shi}}(r)$
$\mathcal{T}_6$	$(m(\circ) - \mu_M(\mathbf{r}))(m(\mathbf{r}) - \mu_M(\mathbf{r}))$	σ_M^2	$I_{mm}^{\text{Sch}}(r)$	$I_{mm}^{\text{Sch}}(r)$

$\nabla_{\mathcal{T}_f}$ is set to σ_M^2 in close analogy to Moran's I (Moran, 1950). Instead of $\nabla_{\mathcal{T}_f}(r)$, it is sometimes preferable to compute the $\mathcal{T}_f$ -correlation function (Penttinen & Stoyan, 1989),

$$\kappa_{\mathcal{T}_f}(r) = \frac{\nabla_{\mathcal{T}_f}(r)}{\nabla_{\mathcal{T}_f}} = \frac{\varrho_{\mathcal{T}_f}^{(2)}(r)}{\varrho^{(2)}(r)} \bigg/ \nabla_{\mathcal{T}_f}, \quad (5)$$

which normalises the conditional expectation of the test function $\nabla_{\mathcal{T}_f}(r)$ by its expectation $\nabla_{\mathcal{T}_f}$ for $r \rightarrow \infty$ (or by σ_M^2 for $\mathcal{T}_5$ and $\mathcal{T}_6$) such that $\kappa_{\mathcal{T}_f}(r) = 1$ for all r under mark independence. For completeness, Table 1 covers both the unnormalised ($\nabla_{\mathcal{T}_f}(r)$) and related normalised ($\kappa_{\mathcal{T}_f}(r)$) expressions.

A classic summary is the conditional mean product of marks of two points being a distance r apart (cf. test function $\mathcal{T}_1$ of Table 1), $\tau_{mm}(r)$, and its scaled version $\kappa_{mm}(r)$ often termed the mark correlation function (Stoyan & Stoyan, 1994). If large (resp. small) marks systematically co-occur at interpoint distance r , their pairwise products will also be large (resp. small) and deviate from the independent mark assumption, that is, the mark mean squared. Thus, $\kappa_{mm}(r)$ provides a useful tool to decide if points that are close together in space tend to have marks that are smaller (or larger) than the mean mark (Myllymäki *et al.*, 2015), and how that relationship changes with distance. Both $\mathbf{r}$ -mark functions $\tau_{m\bullet}$ and $\tau_{\bullet m}$ and the related $\mathbf{r}$ -mark correlation functions $\kappa_{m\bullet}$ and $\kappa_{\bullet m}$ can be interpreted as the conditional expectation of the mark of either the first or second point for any pair of points with distance r . This expectation often deviates from the mark mean μ_M when the marks are dependent on the existence of other points (Schlather *et al.*, 2004; Myllymäki *et al.*, 2015), for example, a smaller average tree height conditional on the existence of another tree at small distance. The mark variogram $\gamma_{mm}(r)$ is a measure of the average dispersion (Cressie, 1993). It helps to detect situations where the marks of points close together tend to be more similar (or different) than expected under mark independence. If nearby points tend to have similar marks, the mark variogram has small values at small distances r . On the other hand, if the mark variogram decreases with distance, the marks of nearby points are less similar than the marks of farther apart points. Shimantani's (Shimantani, 2002) and Schlather's (Schlather, 2001) I -functions I_{mm}^{Shi} and I_{mm}^{Sch} and their normalised versions I_{mm}^{Shi} and I_{mm}^{Sch} can be seen as adaptations of Moran's I (Moran, 1950) to spatial point processes. They are helpful for identifying spatial autocorrelation among the marks. Although these test functions are similar in spirit, Schlather applies a centering not by the unconditional mean but by the conditional mean $\mathbb{E}_{\circ, r}[m(\mathbf{r})] = \mu_M(\mathbf{r})$ of the second point.

3 Composition-Valued Marked Point Processes

3.1 Composition-Valued Marks

To extend spatial point processes to composition-valued marks, let $\{(x_i, \mathbf{c}(x_i))\}_{i=1}^n$ denote a set of n points $x_i \in \mathbb{R}^2$ with associated marks $\mathbf{c}(x_i) = (c_1(x_i), \dots, c_D(x_i))^T$ living in a D -part simplex $\mathbb{S}^D \subset \mathbb{R}^D$, where for some $\mathcal{W} \in \mathbb{R}$

$$\mathbb{S}^D = \left\{ \mathbf{c} = (c_1, c_2, \dots, c_D)^T \in \mathbb{R}^D \mid c_j \geq 0, j = 1, 2, \dots, D; \sum_{j=1}^D c_j = \mathcal{W} \right\}.$$

Common choices of $\mathcal{W}$ include $\mathcal{W} = 1$ and $\mathcal{W} = 100$ (per cent) depending on the marks at hand. Thus, the mark at each location is a composition of D non-negative parts summing to a constant and we call any such mark composition-valued. That is, composition-valued marked spatial point processes are a particular type of spatial compositional data (Aitchison, 1986), where each mark $\mathbf{c}(x_i)$ quantitatively describes the relative importance of the individual parts with respect to a given total. We note that composition-valued marks could have two potential origins: they might (i) directly arise from the data collection or (ii) be constructed from non-simple spatial point process scenarios with integer-valued marks, by summarising the absolute numbers $\tilde{c}_j$, $j = 1, \dots, D$, at a point location a-posteriori into relative numbers for distinct categories. Generally, any such absolute information can always be transformed into a composition-valued mark by applying the closure operation $\mathbf{c} = \text{cls}(\tilde{\mathbf{c}}) = \left(\tilde{c}_1 / \sum_{j=1}^D \tilde{c}_j, \dots, \tilde{c}_D / \sum_{j=1}^D \tilde{c}_j \right)^T$. Consequently, both the absolute and the relative, composition-valued marks can be analysed depending on the question of interest, providing two different views of the marks as discussed below in Section 3.6.

The space $\mathbb{S}^D$ can be equipped with a finite $(D - 1)$ dimensional Euclidean vector space structure, that is, the Aitchison geometry, with the perturbation $\mathbf{c} \oplus \mathbf{c}' = \text{cls}(c_1 c'_1, c_2 c'_2, \dots, c_D c'_D)$, a commutative group operation on the simplex with neutral element $\mathbf{n} = \text{cls}(1, 1, \dots, 1)$ and inverse operation $\mathbf{c} \ominus \mathbf{c}' = \mathbf{c} \oplus ((-1) \odot \mathbf{c}')$, and the powering operation $\xi \odot \mathbf{c} = \text{cls}(c_1^\xi, c_2^\xi, \dots, c_D^\xi)$, where $\mathbf{c}, \mathbf{c}' \in \mathbb{S}^D$ and $\xi \in \mathbb{R}$ (Aitchison, 2001). Additionally, the inner product is defined as

$$\langle \mathbf{c}, \mathbf{c}' \rangle_A = \frac{1}{2D} \sum_{l=1}^D \sum_{j=1}^D \log \left(\frac{c_l}{c_j} \right) \log \left(\frac{c'_l}{c'_j} \right), \quad (6)$$

yielding the norm $\|\mathbf{c}\|_A = \sqrt{\langle \mathbf{c}, \mathbf{c} \rangle_A}$ and the associated distance $d_A(\mathbf{c}, \mathbf{c}') = \|\mathbf{c} \ominus \mathbf{c}'\|_A$. Note that the composition-valued marks can be transformed to real-valued coordinates through a map function $\psi: \mathbb{S}^D \rightarrow \mathbb{R}^{\tilde{D}}$, $\mathbf{c} \mapsto \psi(\mathbf{c})$ where $\tilde{D}$ is determined by the particular choice of ψ . For particular useful choices, an isometric isomorphism exists between $\mathbb{S}^D$ and $\mathbb{R}^{\tilde{D}}$ (Billheimer *et al.*, 2001; Pawlowsky-Glahn & Egozcue, 2001), which allows expressing the composition-valued marks in real coordinates such that their relations are correspondent to those in the Aitchison geometry. After transformation, statistical analysis methods can be performed in $\mathbb{R}^{\tilde{D}}$ (Mateu-Figueras *et al.*, 2011).

3.2 Transformations

Writing $\psi_j(\mathbf{c})$ for the j -th element of $\psi(\mathbf{c}) = (\psi_1(\mathbf{c}), \dots, \psi_{\tilde{D}}(\mathbf{c}))^\top$, different coordinate representations can be derived (Pawlowsky-Glahn & Buccianti, 2011). Apart from the *log-ratio* (lr) transformation early specifications of ψ include the *additive log-ratio* (alr) transformation (Aitchison & Shen, 1980) where $\psi_j(\mathbf{c}) = \log(c_j/c_D)$, $j = 1, \dots, \tilde{D} = D - 1$ that is, the log-ratios relative to the D -th component, and the *centered log-ratio* (clr) (Aitchison, 1983) transformation $\psi_j(\mathbf{c}) = \log(c_j/g(\mathbf{c}))$, $j = 1, \dots, \tilde{D} = D$, that is, the log-ratios relative to the geometric mean $g(\mathbf{c}) = \left(\prod_{j=1}^D c_j\right)^{1/D}$. While the alr transformation yields an isomorphic but non-isometric relation between the two spaces, that is, it does not preserve distances, the clr transformation results in an isometric isomorphism by mapping the marks from the simplex to a hyperplane $\mathbb{H} \subset \mathbb{R}^D$ that is orthogonal to the vector of ones. However, the clr imposes a sum-to-zero constraint on the transformed marks, which can make analysis more difficult, for example, due to degenerated distributions and singular covariance matrices.

We observe that neither the alr nor the clr transformation can be directly linked to an orthonormal coordinate system on the simplex. However, an orthonormal basis $(\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_{D-1})$ on the simplex $\mathbb{S}^D$ with respect to the inner product can be derived using the Gram-Schmidt procedure giving

$$\mathbf{c} = \bigoplus_{j=1}^{D-1} \langle \mathbf{c}, \mathbf{e}_j \rangle_A \odot \mathbf{e}_j.$$

This coordinate representation corresponds to the *isometric log-ratio* (ilr) transformation, a class of orthonormal coordinate representations, defined by

$$\text{ilr}(\mathbf{c}) = (\langle \mathbf{c}, \mathbf{e}_1 \rangle_A, \langle \mathbf{c}, \mathbf{e}_2 \rangle_A, \dots, \langle \mathbf{c}, \mathbf{e}_{D-1} \rangle_A)$$

and establishes an isometric isomorphism through the map between $\mathbb{S}^D$ and $\mathbb{R}^{D-1}$. Further, the ilr transformation is related to the clr and log transformations through $\text{ilr}(\mathbf{c}) = \text{clr}(\mathbf{c})\mathbf{H}_D^\top = \log(\mathbf{c})\mathbf{H}_D^\top$ where $\mathbf{H}_D$ denotes a $((D-1) \times D)$ -dimensional Helmert matrix with rows $\mathbf{h}_j = \text{clr}(\mathbf{e}_j)$, $j = 1, \dots, D-1$ satisfying $\mathbf{H}_D\mathbf{H}_D^\top = \mathbf{I}_{D-1}$ and $\mathbf{H}_D^\top \mathbf{H}_D = \mathbf{G}_D$ and $\mathbf{G}_D$ is the D -dimensional centering matrix $\mathbf{G}_D = \mathbf{I}_D - D^{-1}\mathbf{1}_D\mathbf{1}_D^\top$, $\mathbf{I}_D$ is the identity matrix of dimension $(D \times D)$, and $\mathbf{1}_D$ a $(D \times 1)$ vector of ones. Due to the isometric isomorphism established between the Aitchison and the Euclidean geometry by the clr and the ilr transformations, the Aitchison inner product, distances and metrics coincide with their Euclidean counterparts on the transformed quantities, such that for any $\mathbf{c}, \mathbf{c}' \in \mathbb{S}^D$

$$d_A(\mathbf{c}, \mathbf{c}') = d_E(\text{clr}(\mathbf{c}), \text{clr}(\mathbf{c}')) = d_E(\text{ilr}(\mathbf{c}), \text{ilr}(\mathbf{c}')) \quad (7)$$

with d_E the Euclidean distance. The ilr transformation is equivalent to the logit-function used in logistic regression if $D = 2$. When $D > 2$, infinitely many orthonormal basis systems exist and the concrete choice has a crucial impact on the interpretation of the projected data. Particular choices of an orthonormal basis are coordinate representations using (i) *balances* (Egozcue & Pawlowsky-Glahn, 2005) in which each *balancing element* can be interpreted as the normalised log-ratio of the geometric means (centers) of two groups, and (ii) *pivot coordinates* (Fišerová & Hron, 2011; Hron *et al.*, 2017). Balances originate from a *sequential binary partition* method, which involves dividing the composition into two parts. The j -th ilr coefficient using pivot coordinates can be expressed as

$$\text{ilr}_j(\mathbf{c}) = \sqrt{\left(\frac{D-j}{D-j+1}\right)} \log \left\{ \frac{c_j}{\sqrt[D-j]{\prod_{k=j+1}^D c_k}} \right\}, j = 1, \dots, D-1$$

(Fišerová & Hron, 2011). The initial ilr coefficient is akin to the first clr coefficient, scaled by $\sqrt{D/(D-1)}$, and is easily interpreted as the log-ratio of the respective component to the geometric mean. In contrast, interpreting subsequent coefficients is more complex. To address this, the literature proposes generalised pivot coordinates using permuted compositions and symmetric pivot coordinates (see, e.g. Kynčlová *et al.*, 2017; Hron *et al.*, 2021).

While the above transformations allow for a representation of the composition-valued marks in coordinates in $\mathbb{R}$, the underlying log-operations are undefined for zero values. In what follows, we assume that the proportions for all D -parts are non-zero. However, as zeros might be present in some marked point process scenarios when potentially not all components are observed at each location, we also provide a treatment of different transformations in the presence of structural zeros, see Section S1.

3.3 First-Order Tools for Composition-Valued Marks

We first review first-order mark characteristics for the composition-valued marks. General characteristics for a strictly-positive sample $\mathbf{c}_1, \dots, \mathbf{c}_n$ of compositions commonly applied in the literature include the geometric center, that is, the closed geometric mean,

$$\text{cen}(\mathbf{c}) = \frac{1}{n} \odot \bigoplus_{j=1}^n \mathbf{c}_j = \text{clr}^{-1} \left(\frac{1}{n} \sum_{j=1}^n \text{clr}(\mathbf{c}_j) \right), \quad (8)$$

where $\text{clr}^{-1}(\tilde{\mathbf{c}}) = \text{cls}(\exp(\tilde{\mathbf{c}}))$. It serves as the mean composition-valued mark. Further, the variation of the composition-valued marks can be described through the variation matrix $\mathbf{T}$ with elements $t_{jl} = \mathbb{V} \text{ar}[\log(c_j/c_l)]$, $j, l = 1, \dots, D$, and the total, that is, metric, variance

$$m\mathbb{V} \text{ar}[\mathbf{c}] = \frac{1}{2D} \sum_{j,l=1}^D t_{jl} = \frac{1}{n-1} \sum_{j=1}^n d_A^2(\mathbf{c}_j, \text{cen}(\mathbf{c})),$$

serving as a global measure of dispersion (Aitchison, 1986; Pawlowsky-Glahn & Egozcue, 2001). Instead of $\mathbf{T}$, the normalised variation matrix $\bar{\mathbf{T}} = 0.5 \cdot \mathbf{T}$ can be used.

3.4 Componentwise Summary Characteristics for Composition-Valued Marks

We now define novel componentwise characteristics for composition-valued marks analogously to the mark characteristics of Section 2.2. These are useful to look at the spatial behaviour of all components individually, while we propose characteristics for the whole composition in the next subsection. We here explicitly focus on the case where each point is augmented by exactly one composition while multivariate extensions are outlined in Section S4. Recall that $\mathbf{c}(\circ)$ and $\mathbf{c}(\mathbf{r})$ denote the composition-valued marks for a pair of points at locations $\circ$ and $\mathbf{r}$ at distance $\|\mathbf{r}\| = r$, and $\psi(\mathbf{c})$ is the transformed composition-valued mark with j th element or component $\psi_j(\mathbf{c})$, where $\psi: \mathbb{S}^D \mapsto \mathbb{R}^{\tilde{D}}$. We then define

Table 2. Specifications for the componentwise test function $\mathcal{T}_f^{\psi,jl}(\psi_j(\mathbf{c}(\mathbf{o})), \psi_l(\mathbf{c}(\mathbf{r})))$ with μ_j^ψ denoting the unconditional mark mean and $\mu_j^\psi(\mathbf{r})$ the conditional mark mean of the j -th element of the ψ -transformed composition-valued marks, where $\zeta_{jl}^\psi = 0.5[\sigma_{jj}^\psi + \sigma_{ll}^\psi + (\mu_j^\psi - \mu_l^\psi)^2]$ with σ_{jl}^ψ denoting the covariance of the j -th and the l -th element of the ψ -transformed composition-valued mark.

Name for $\mathcal{T}_f^{\psi,jl}$	Test function $\mathcal{T}_f^{\psi,jl}$	Normalising factor $\nabla_{\mathcal{T}_f}^{\psi,jl}$	Notation for $\nabla_{\mathcal{T}_f}^{\psi,jl}(r)$	Notation for $\kappa_{\mathcal{T}_f}^{\psi,jl}(r)$
$\mathcal{T}_1^{\psi,jl}$	$\psi_j(\mathbf{c}(\mathbf{o}))\psi_l(\mathbf{c}(\mathbf{r}))$	$\mu_j^\psi \mu_l^\psi$	$\tau_{jl}^\psi(r)$	$\kappa_{jl}^\psi(r)$
$\mathcal{T}_2^{\psi,jl}$	$\psi_j(\mathbf{c}(\mathbf{o}))$	μ_j^ψ	$\tau_j^\psi(r)$	$\kappa_j^\psi(r)$
$\mathcal{T}_3^{\psi,jl}$	$\psi_l(\mathbf{c}(\mathbf{r}))$	μ_l^ψ	$\tau_l^\psi(r)$	$\kappa_l^\psi(r)$
$\mathcal{T}_4^{\psi,jl}$	$0.5(\psi_j(\mathbf{c}(\mathbf{o})) - \psi_l(\mathbf{c}(\mathbf{r})))^2$	ζ_{jl}^ψ	$\gamma_{jl}^\psi(r)$	$\gamma_{jl}^{\psi,n}(r)$
$\mathcal{T}_5^{\psi,jl}$	$(\psi_j(\mathbf{c}(\mathbf{o})) - \mu_j^\psi)(\psi_l(\mathbf{c}(\mathbf{r})) - \mu_l^\psi)$	σ_{jl}^ψ	$i_{jl}^{\psi, \text{Shi}}(r)$	$I_{jl}^{\psi, \text{Shi}}(r)$
$\mathcal{T}_6^{\psi,jl}$	$(\psi_j(\mathbf{c}(\mathbf{o})) - \mu_j^\psi(\mathbf{r}))(\psi_l(\mathbf{c}(\mathbf{r})) - \mu_l^\psi(\mathbf{r}))$	σ_{jl}^ψ	$i_{jl}^{\psi, \text{Sch}}(r)$	$I_{jl}^{\psi, \text{Sch}}(r)$

$$\nabla_{\mathcal{T}_f}^{\psi,jl}(r) = \mathbb{E}_{\mathbf{o},r} \left[\mathcal{T}_f^{\psi,jl}(\psi_j(\mathbf{c}(\mathbf{o})), \psi_l(\mathbf{c}(\mathbf{r}))) \right],$$

where $\mathcal{T}_f^{\psi,jl}$ denotes a test function specific to the transformed composition-valued marks. The test functions of Table 1 can be employed as $\mathcal{T}_f^{\psi,jl}$; for clarity, they are re-expressed for the transformed marks $\psi(\mathbf{c})$ in Table 2. Further, we let $\nabla_{\mathcal{T}_f}^{\psi,jl}$ stand for the limiting case of $\nabla_{\mathcal{T}_f}^{\psi,jl}(r)$ when $r \rightarrow \infty$, that is,

$$\nabla_{\mathcal{T}_f}^{\psi,jl} = \int_{\mathbb{R}^D} \int_{\mathbb{R}^D} \mathcal{T}_f^{\psi,jl}(\psi_j(\mathbf{c}_1), \psi_l(\mathbf{c}_2)) \varpi(d\psi_j(\mathbf{c}_1)) \varpi(d\psi_l(\mathbf{c}_2)), \quad (9)$$

and define the $\mathcal{T}_f^{\psi,jl}$ -correlation function $\kappa_{\mathcal{T}_f}^{\psi,jl}(r)$ as

$$\kappa_{\mathcal{T}_f}^{\psi,jl}(r) = \nabla_{\mathcal{T}_f}^{\psi,jl}(r) \nabla_{\mathcal{T}_f}^{\psi,jl}, \quad (10)$$

where (9) is again replaced by the corresponding variance in the case of $\mathcal{T}_5^{\psi,jl}$ and $\mathcal{T}_6^{\psi,jl}$. The choice of the test function $\mathcal{T}_f^{\psi,jl}$ determines the focus of the analysis, as the different characteristics highlight different properties of marks (see Section 2.2).

Although in the following we restrict our discussion to only some characteristics, the same principle applies to all test functions in Table 2. Consider as an example test function $\mathcal{T}_1^{\psi,jl}$. It yields a componentwise conditional mean product of marks $\tau_{jl}^\psi(r)$,

$$\tau_{jl}^\psi(r) = \mathbb{E}_{\mathbf{o},r} \left[\psi_j(\mathbf{c}(\mathbf{o}))\psi_l(\mathbf{c}(\mathbf{r})) \right], \quad (11)$$

which describes the average product of two components of the ψ -transformed marks for any pair of points as a function of the distance r .

Restricting to pairs of the j -th element of $\psi(\mathbf{c})$ at two locations, $j = l$, and specifying ψ in (11) as log-ratio between parts j_1 and j_2 , to directly focus on their relative contribution to the total, yields the mean pairwise product of mark log-ratios $\tau_{jj}^{\text{lr}}(r)$,

$$\tau_{jj}^{\text{lr}}(r) = \mathbb{E}_{\mathbf{o},r} \left[\log \left(\frac{c_{j_1}(\mathbf{o})}{c_{j_2}(\mathbf{o})} \right) \log \left(\frac{c_{j_1}(\mathbf{r})}{c_{j_2}(\mathbf{r})} \right) \right] \quad (12)$$

for j indexing the $\tilde{D} = D^2$ ordered pairs $(j_1, j_2) = (1, 1), \dots, (1, D), \dots, (D, D)$. This characteristic can highlight if the log-ratios at nearby points are correlated and their products thus tend to be larger (smaller) than expected under independence, $\nabla_{\mathcal{F}_1}^{\text{lr}, j}$. It is easy to see that τ_{jj}^{lr} equals the squared mean $\left(\mu_j^{\text{lr}}\right)^2$ of the log-ratios of the j_1 -th and j_2 -th parts under independent marks.

Similarly, the componentwise log-ratio $\mathbf{r}$ -mark functions τ_j^{lr} and τ_j^{lr} result from taking the conditional expectation of either $\mathcal{F}_2^{\text{lr}, j}$ or $\mathcal{F}_3^{\text{lr}, j}$, which both coincide with the mean of the log-ratio of the j -th part, μ_j^{lr} , under independent marks. Reflecting the mean behaviour of the transformed mark at either the first or second point, deviations of the empirical curves from μ_j^{lr} at some distances indicate the presence of spatial structure in the mark parts, that is, changes in the average mark ratios if a second point is present at distance r . As such, both quantities could be helpful tools to identify potential interrelations of the points and marks.

Replacing $\mathcal{F}_1^{\psi, jl}$ by $\mathcal{F}_4^{\psi, jl}$ in (11) yields a generic componentwise mark variogram

$$\gamma_{jl}^{\psi}(r) = \mathbb{E}_{\circ, r} \left[0.5(\psi_j(\mathbf{c}(\circ)) - \psi_l(\mathbf{c}(\mathbf{r})))^2 \right]. \quad (13)$$

The precise form again depends on the specific choice of ψ . Using a transformation into log-ratios and $j = l$ leads to a componentwise log-ratio mark variogram γ_{jj}^{lr} for composition-valued marks defined by

$$\gamma_{jj}^{\text{lr}}(r) = \mathbb{E}_{\circ, r} \left[\frac{1}{2} \left(\log \left(\frac{c_{j_1}(\circ)}{c_{j_2}(\circ)} \right) - \log \left(\frac{c_{j_1}(\mathbf{r})}{c_{j_2}(\mathbf{r})} \right) \right)^2 \right] \quad (14)$$

for j indexing $(j_1, j_2) = (1, 1), \dots, (1, D), \dots, (D, D)$. Similar to the classic mark variogram, $\gamma_{jj}^{\text{lr}}(r)$ measures the average pairwise variation of the j -th log-ratio of $\psi(\mathbf{c})$ at two distinct points at distance r and tends to the variance σ_{jj}^{lr} of the log-ratios of the j_1 -st and j_2 -nd parts for $r \rightarrow \infty$. The mark variogram can be used to investigate the heterogeneity among the marks, that is, if the proportions of the specific parts for any pair of points are on average more similar in value for small distances. We note that for each r all of the above log-ratio characteristics could be stored in local $(D \times D)$ matrices with $D(D - 1)$ log-ratios of different parts in the off-diagonal entries and zeros for the log-ratios of all parts with themselves on its diagonal. In particular, collecting all $\gamma_{jj}^{\text{lr}}(r)$ into a local mark variogram matrix $\mathbf{I}^{\text{lr}}(r)$ is similar in spirit to a local variation matrix $\mathbf{T}(r)$ which captures the spatial dispersion of the composition-valued mark at the distance r .

While the above log-ratio characteristics are most useful for autocovariances and autocorrelations, the clr and ilr transformations allow for both auto- and cross-characteristic formulations. For $\psi = \text{clr}$, evaluation of the conditional expectation of $\mathcal{F}_1^{\psi, jl}$ and $\mathcal{F}_4^{\psi, jl}$ yields the conditional mean product of clr marks, $\tau_{jl}^{\text{clr}}(r)$, and the clr mark variogram, γ_{jl}^{clr} , defined by

$$\tau_{jl}^{\text{clr}}(r) = \mathbb{E}_{\circ, r} \left[\log \left(\frac{c_j(\circ)}{g(\mathbf{c})(\circ)} \right) \cdot \log \left(\frac{c_l(\mathbf{r})}{g(\mathbf{c})(\mathbf{r})} \right) \right] \quad (15)$$

and

$$\gamma_{jl}^{\text{clr}}(r) = \mathbb{E}_{\circ, r} \left[\frac{1}{2} \left(\log \left(\frac{c_j(\circ)}{g(\mathbf{c})(\circ)} \right) - \log \left(\frac{c_l(\mathbf{r})}{g(\mathbf{c})(\mathbf{r})} \right) \right)^2 \right], \quad (16)$$

respectively. Due to the construction of (15) and (16), both functions describe the average spatial association/variation of the j -th and l -th parts relative to the geometric mean, including both auto- (for $j = l$) and cross-relations (for $j \neq l$). In particular, cross-relations might be useful to analyse the association between different parts, for example, the proportions of two distinct parasite species for neighbouring trees. Both quantities allow for similar interpretations as their log-ratio counterparts and could provide useful information on the distributional behaviour of the clr transformed parts. Inserting the corresponding terms into (9) immediately implies that $\tau_{jl}^{\text{clr}}(r)$ converges to the product of means $\mu_j^{\text{clr}} \cdot \mu_l^{\text{clr}}$ of the clr transformed parts j, l for $r \rightarrow \infty$ as shown in Section S2. Likewise, for $\gamma_{jl}^{\text{clr}}(r)$, inserting $\mathcal{T}_4^{\text{clr}, jl}$ into (9), $\nabla_{\mathcal{T}_4}^{\text{clr}, jl}$ becomes ζ_{jl}^{clr} .

The clr characteristics for all D parts can efficiently be stored in local $(D \times D)$ matrices including the local clr mark variogram matrix $\mathbf{I}^{\text{clr}}(r) = [\gamma_{jl}^{\text{clr}}(r)]_{j, l=1, \dots, D}$. Similarly, using ilr coordinates yields

$$\tau_{jl}^{\text{ilr}}(r) = \mathbb{E}_{\circ, r} [\text{ilr}_j(\mathbf{c}(\circ)) \text{ilr}_l(\mathbf{c}(\mathbf{r}))] \quad (17)$$

and

$$\gamma_{jl}^{\text{ilr}}(r) = \mathbb{E}_{\circ, r} \left[\frac{1}{2} (\text{ilr}_j(\mathbf{c}(\circ)) - \text{ilr}_l(\mathbf{c}(\mathbf{r})))^2 \right], \quad (18)$$

where ilr_j and ilr_l are the j -th and l -th ilr coordinates of the composition-valued marks.

3.5 Compositional Summary Characteristics for Composition-Valued Marks

Instead of describing the spatial behaviour of one component of ψ -transformed compositions using a componentwise test function specification, we now discuss an extension which allows us to evaluate the spatial properties of whole D -part compositions. This allows us to assess the variation and correlation of the complete composition of nearby trees or business sectors. Such compositional summary characteristics, which summarise the variation and interrelation between the D -part compositions for a pair of points as a function of the interpoint distance r , can be derived using central concepts from the Aitchison geometry. In particular, defining $\nabla_{\mathcal{T}_f}^{\mathbb{S}}(r)$ to denote the conditional mean of the test function $\mathcal{T}_f^{\mathbb{S}}$ of a D -part composition-valued mark $\mathbf{c}$ at locations $\circ$ and $\mathbf{r}$ where $\|\circ - \mathbf{r}\| = r$, and $\kappa_{\mathcal{T}_f}^{\mathbb{S}}(r)$ as

$$\kappa_{\mathcal{T}_f}^{\mathbb{S}}(r) = \frac{\nabla_{\mathcal{T}_f}^{\mathbb{S}}(r)}{\nabla_{\mathcal{T}_f}^{\mathbb{S}}}, \quad (19)$$

where $\nabla_{\mathcal{T}_f}^{\mathbb{S}}$ extends (9) to

$$\nabla_{\mathcal{T}_f}^{\mathbb{S}} = \int_{\mathbb{S}^D} \int_{\mathbb{S}^D} \mathcal{T}_f^{\mathbb{S}}(\mathbf{c}(\circ), \mathbf{c}(\mathbf{r})) \varpi(d\mathbf{c}(\circ)) \varpi(d\mathbf{c}(\mathbf{r})) \quad (20)$$

and denotes the conditional expectation of the test function for $r \rightarrow \infty$, allows us to define different compositional mark summary characteristics. An overview of different test functions $\mathcal{T}_f^{\mathbb{S}}$ and the corresponding characteristics is provided in Table 3 where here only those test functions are considered, which allow a direct relation to the corresponding componentwise test functions of Table 2. We note that instead of applying (9) to compute $\nabla_{\mathcal{T}_f}^{\mathbb{S}}$ for the compositional versions of Schlather's and Shimantani's I functions, $\nabla_{\mathcal{T}_f}^{\mathbb{S}}$ is set to $\sigma_c^2 = \omega \sum_{j=1}^{\bar{D}} \zeta_{jj}^{\psi}$ where $\omega = 1/2D$ if a

Table 3. Compositional test function specifications with $\mu_{\mathbf{c}}^2 = \omega \sum_{j=1}^{\tilde{D}} \mu_j^{\psi} \cdot \mu_j^{\psi}$ and $\sigma_{\mathbf{c}}^2 = \omega \sum_{j=1}^{\tilde{D}} \zeta_{jj}^{\psi}$ with $\omega = 1/2D$ if a transformation into logratios is applied and $\omega = 1$ under ilr and clr transformations, $\text{cen}(\mathbf{c})$ is the center of (8), and $\text{cen}(\mathbf{c})(\mathbf{r})$ is the conditional center computed over all pairs of points at a distance $\|\mathbf{r}\| = r$.

Name for $\mathcal{T}_f^{\mathbb{S}}$	Test function $\mathcal{T}_f^{\mathbb{S}}$	Normalising factor $\nabla_{\mathcal{T}_f}^{\mathbb{S}}$	Notation for $\nabla_{\mathcal{T}_f}^{\mathbb{S}}(r)$	Notation for $\kappa_{\mathcal{T}_f}^{\mathbb{S}}(r)$
$\mathcal{T}_1^{\mathbb{S}}$	$\langle \mathbf{c}(\circ), \mathbf{c}(\mathbf{r}) \rangle_A$	$\mu_{\mathbf{c}}^2$	$\tau_{\text{cc}}(r)$	$\kappa_{\text{cc}}(r)$
$\mathcal{T}_4^{\mathbb{S}}$	$0.5 \ \mathbf{c}(\circ) \ominus \mathbf{c}(\mathbf{r})\ _A^2$	$\sigma_{\mathbf{c}}^2$	$\gamma_{\text{cc}}(r)$	$\gamma_{\text{cc}}^n(r)$
$\mathcal{T}_5^{\mathbb{S}}$	$\langle \mathbf{c}(\circ) \ominus \text{cen}(\mathbf{c}), \mathbf{c}(\mathbf{r}) \ominus \text{cen}(\mathbf{c}) \rangle_A$	$\sigma_{\mathbf{c}}^2$	$i_{\text{cc}}^{\text{Shi}}(r)$	$I_{\text{cc}}^{\text{Shi}}(r)$
$\mathcal{T}_6^{\mathbb{S}}$	$\langle \mathbf{c}(\circ) \ominus \text{cen}(\mathbf{c})(\mathbf{r}), \mathbf{c}(\mathbf{r}) \ominus \text{cen}(\mathbf{c})(\mathbf{r}) \rangle_A$	$\sigma_{\mathbf{c}}^2$	$i_{\text{cc}}^{\text{Sch}}(r)$	$I_{\text{cc}}^{\text{Sch}}(r)$

transformation into logratios is applied and $\omega = 1$ under ilr and clr transformations to allow for a close analogy to Moran's I . Like the componentwise mark characteristics, all of these test functions can be used to highlight particular aspects of the distributional properties. However, providing different insights into the underlying structure of the composition-valued marks, the componentwise and the compositional mark characteristics both provide useful tools to investigate the individual contribution of the components to the results and to assess the overall variation and correlation.

Noting the isometry of the Aitchison norm and distance to their Euclidean counterpart versions of the clr or ilr transformed compositions, the test functions of Table 2 and Table 3 are related as follows. The compositional conditional expectation of the inner product of marks $\tau_{\text{cc}}(r)$ can be constructed by using $\mathcal{T}_1^{\mathbb{S}}$, which is related to the componentwise test function $\mathcal{T}_1^{\psi, \text{ilr}}$ through

$$\begin{aligned}
 \mathcal{T}_1^{\mathbb{S}} &= \langle \mathbf{c}(\circ), \mathbf{c}(\mathbf{r}) \rangle_A = \langle \text{clr}(\mathbf{c}(\circ)), \text{clr}(\mathbf{c}(\mathbf{r})) \rangle_E \\
 &= \sum_{j=1}^D \text{clr}_j(\mathbf{c}(\circ)) \text{clr}_j(\mathbf{c}(\mathbf{r})) = \sum_{j=1}^D \mathcal{T}_1^{\text{clr}, jj} \\
 &= \sum_{j=1}^{D-1} \text{ilr}_j(\mathbf{c}(\circ)) \text{ilr}_j(\mathbf{c}(\mathbf{r})) = \sum_{j=1}^{D-1} \mathcal{T}_1^{\text{ilr}, jj} \\
 &\stackrel{(6)}{=} \frac{1}{2D} \sum_{j_1} \sum_{j_2} \log \left(\frac{c_{j_1}(\circ)}{c_{j_2}(\circ)} \right) \log \left(\frac{c_{j_1}(\mathbf{r})}{c_{j_2}(\mathbf{r})} \right) = \frac{1}{2D} \sum_{j=1}^{D^2} \mathcal{T}_1^{\text{lr}, jj}
 \end{aligned} \tag{21}$$

with $\langle \cdot, \cdot \rangle_E$ denoting the Euclidean inner product. Thus

$$\tau_{\text{cc}}(r) = \sum_{j=1}^D \tau_{jj}^{\text{clr}}(r) = \sum_{j=1}^{D-1} \tau_{jj}^{\text{ilr}}(r) = \frac{1}{2D} \sum_{j=1}^{D^2} \tau_{jj}^{\text{lr}}(r), \tag{22}$$

which allows us to evaluate how individual components contribute to the overall mark characteristic. The corresponding compositional mark correlation function κ_{cc} is obtained by normalising τ_{cc} by $\mu_{\mathbf{c}}^2 = \omega \sum_{j=1}^{\tilde{D}} \mu_j^{\psi} \cdot \mu_j^{\psi}$, the limiting case which follows by substituting $\mathcal{T}_1^{\mathbb{S}}$ for $\mathcal{T}_f^{\mathbb{S}}$ in (20) with $\omega = 1/2D$ under the lr-transformation and $\omega = 1$ for ilr or clr transformations (see Section S2). Hence, for independent composition-valued marks, κ_{cc} becomes constant one.

Selection of $\mathcal{T}_4^{\mathbb{S}}$ results in a compositional mark variogram, $\gamma_{\text{cc}}(r)$,

$$\gamma_{\mathbf{c}\mathbf{c}}(r) = \mathbb{E}_{\circ, r} \left[\frac{1}{2} \|\mathbf{c}(\circ) \ominus \mathbf{c}(\mathbf{r})\|_A^2 \right] = \mathbb{E}_{\circ, r} \left[\frac{1}{2} d_A(\mathbf{c}(\circ), \mathbf{c}(\mathbf{r}))^2 \right]. \quad (23)$$

By the decomposition of the Aitchison squared distance into the sum of the squared distances between the clr or ilr transformed marks as stated in (7), the compositional mark variogram can be decomposed into the sum of the componentwise mark variogram terms. In particular, we have that

$$\gamma_{\mathbf{c}\mathbf{c}}(r) = \sum_{j=1}^D \gamma_{jj}^{\text{clr}}(r) = \sum_{j=1}^{D-1} \gamma_{jj}^{\text{ilr}}(r).$$

This functional quantity describes the average dispersion between the D -part composition-valued marks for a focal and a second point at distance r . If composition-valued marks are correlated for small distances, the compositional mark variogram $\gamma_{\mathbf{c}\mathbf{c}}(r)$ will show a corresponding decrease for small spatial distances. The decomposition allows us to evaluate which mark components contribute how to the overall compositional mark variogram $\gamma_{\mathbf{c}\mathbf{c}}(r)$. Again using (20), the mark variogram under independent marks is equal to $\sigma_{\mathbf{c}}^{\mathbb{S}} = \omega \sum_{j=1}^D \zeta_{jj}^{\psi}$ (see Section S2).

Test functions $\mathcal{T}_5^{\mathbb{S}}$ and $\mathcal{T}_6^{\mathbb{S}}$ yield adaptations of Schlather's and Shimatani's I functions, respectively, which can be useful tools to investigate the spatial autocorrelation of the composition-valued marks. Writing $\bar{\mathbf{c}} = \mathbf{c} - \text{cen}(\mathbf{c})$ to denote the centered composition, $i_{\mathbf{c}\mathbf{c}}^{\text{Shi}}$ can be decomposed into the weighted sum over the componentwise Shimantani's functions $i_{\mathbf{c}\mathbf{c}}^{\psi, jj}$ analogously to (22), using the equivalences of the Aitchinson inner product as in (21). The unnormalised $i_{\mathbf{c}\mathbf{c}}^{\text{Sch}}$ function of Schlather can be obtained from its componentwise versions in a similar manner by applying a centering of the composition by the conditional center $\text{cen}(\mathbf{c})(\mathbf{r})$. The decomposition of $i_{\mathbf{c}\mathbf{c}}^{\text{Sch}}(r)$ into individual contributions is also analogous to that for $\gamma_{\mathbf{c}\mathbf{c}}(r)$.

3.6 Extensions to Composition-Valued Marks With Total Information

Finally, extensions of the proposed mark characteristics to combinations of relative, that is, composition-valued, and absolute point-specific information are outlined. Adapting the results of Pawlowsky-Glahn *et al.* (2015) to the present context, consider $\tilde{\mathbf{c}} \in \mathbb{R}^D$ such that $\mathbf{c} = \text{cls}(\tilde{\mathbf{c}}) \in \mathbb{S}^D$ and denote by $y = \sum_{j=1}^D \tilde{c}_j$ the total. We then call $\boldsymbol{\eta} = (y, \mathbf{c})$ a mixed mark with real-valued and composition-valued components y and $\mathbf{c}$ living on $\mathbb{T} = \mathbb{R}_+ \times \mathbb{S}^D$, $\mathbb{R}_+ = [0, \infty)$. Focusing on the process $\{x_i, \boldsymbol{\eta}(x_i)\}$ instead of $\{x_i, \mathbf{c}(x_i)\}$, all the above mark characteristics can be extended to mixed marks on $\mathbb{T}$. To this end, consider first the scalars $y, y' \in \mathbb{R}_+$ and let $\oplus_+$ and $\odot_+$ denote the plus-perturbation and plus-powering operations, respectively, where $y \oplus_+ y' = yy'$ and $\zeta \odot_+ y = y^{\zeta}$ for $\zeta \in \mathbb{R}$. Further, denote by $\langle y, y' \rangle_+ = \langle \log(y), \log(y') \rangle_E$ the plus-inner product and by $d_+(y, y') = \text{abs}(\log(y) - \log(y'))$ the plus-distance on $\mathbb{R}_+$. The above results can then be used to establish a vector space structure on $\mathbb{T}$ with $\mathbb{T}$ -perturbation $\boldsymbol{\eta} \oplus_{\mathbb{T}} \boldsymbol{\eta}' = (y \oplus_+ y', \mathbf{c} \oplus \mathbf{c}')$ and $\mathbb{T}$ -powering operations $\zeta \odot_{\mathbb{T}} \boldsymbol{\eta}' = (y^{\zeta}, \zeta \odot \mathbf{c})$ where $\boldsymbol{\eta}, \boldsymbol{\eta}' \in \mathbb{T}$ and $\zeta \in \mathbb{R}$ as before. Both variation and correlation related mark characteristics from Section 3.5 can be redefined through the $\mathbb{T}$ -inner product and $\mathbb{T}$ -squared distance $\langle \boldsymbol{\eta}, \boldsymbol{\eta}' \rangle_{\mathbb{T}} = \langle \mathbf{c}, \mathbf{c}' \rangle_A + \beta \langle y, y' \rangle_+$ and $d_{\mathbb{T}}(\boldsymbol{\eta}, \boldsymbol{\eta}') = d_A(\mathbf{c}, \mathbf{c}') + \beta (\text{abs}(\log(y) - \log(y')))$, respectively, where β is a weight which can be chosen as one or as the ratio of the variances of the composition $\mathbf{c}$

and of the total y (Happ & Greven, 2018). Using the above extensions and reformulating the compositional mark variogram for mixed composition- and real-valued marks, we obtain

$$\gamma_{\mathbf{cc}, y}(r) = \mathbb{E}_{\mathbf{o}, r} \left[\frac{1}{2} d_A(\mathbf{c}(\mathbf{o}), \mathbf{c}(\mathbf{r}))^2 + \beta \left(\frac{1}{2} d_+(y(\mathbf{o}), y(\mathbf{r})) \right)^2 \right] \quad (24)$$

The other summary characteristics of Table 3 can be extended analogously.

3.7 Estimation

Recalling the representation of $\nabla_{\mathcal{F}_f}^{\psi, jl}(r)$ as the ratio of the two second-order product density functions $\varrho_{\mathcal{F}_f}^{\psi, jl(2)}(r)$ and $\varrho^{(2)}(r)$, $\nabla_{\mathcal{F}_f}^{\psi, jl}(r)$ can be estimated in close analogy to classic spatial point processes by

$$\hat{\nabla}_{\mathcal{F}_f}^{\psi, jl}(r) = \hat{\varrho}_{\mathcal{F}_f}^{\psi, jl, (2)}(r) \hat{\varrho}_{\mathcal{F}_f}^{(2)}(r), \quad (25)$$

where

$$\hat{\varrho}_{\mathcal{F}_f}^{\psi, jl, (2)}(r) = \frac{1}{2\pi r v(W)} \sum_{x_1, x_2 \in W}^{\neq} \mathcal{F}_f^{\psi, jl}(\psi_j(\mathbf{c}(x_1)), \psi_l(\mathbf{c}(x_2))) \mathfrak{K}_b(\|x_1 - x_2\| - r) \quad (26)$$

and

$$\hat{\varrho}^{(2)}(r) = \frac{1}{2\pi r v(W)} \sum_{x_1, x_2 \in W}^{\neq} \mathfrak{K}_b(\|x_1 - x_2\| - r). \quad (27)$$

Here, $\mathfrak{K}_b$ denotes a kernel function of bandwidth b and $v(W)$ the area of the observation window W . We note that as (26) and (27) are estimated using the same estimation principle, an edge correction factor can be ignored in both expressions (Illian *et al.*, 2008). Similarly, an estimator of $\kappa_{\mathcal{F}_f}^{\psi, jl}$ in (10) can be obtained from normalising $\hat{\nabla}_{\mathcal{F}_f}^{\psi, jl}(r)$ by $\hat{\nabla}_{\mathcal{F}_f}^{\psi, jl}$,

$$\hat{\kappa}_{\mathcal{F}_f}^{\psi, jl}(r) = \hat{\nabla}_{\mathcal{F}_f}^{\psi, jl}(r) \hat{\nabla}_{\mathcal{F}_f}^{\psi, jl}, \quad (28)$$

where $\hat{\nabla}_{\mathcal{F}_f}^{\psi, jl}$ in (28) can be estimated analogously to the scalar case from the transformed marks of the n points $x_1, \dots, x_n$ by

$$\hat{\nabla}_{\mathcal{F}_f}^{\psi, jl} = \frac{1}{n^2} \sum_{i=1}^n \sum_{h=1}^n \mathcal{F}_f^{\psi, jl}(\psi_j(\mathbf{c}(x_i)), \psi_l(\mathbf{c}(x_h))).$$

For example, specifying ψ_j as the j -th component of the clr transformation of $\mathbf{c}$, that is, $\log(c_j(\mathbf{o})/g(\mathbf{c}))$, an estimator of the clr mark variogram γ_{jl}^{clr} of (16) can be obtained from the ratio of second-order density functions $\hat{\varrho}_{\mathcal{F}_4}^{\text{clr}, jl, (2)}(r) / \hat{\varrho}_{\mathcal{F}_4}^{(2)}(r)$,

$$\hat{\gamma}_{jl}^{\text{clr}} = \frac{\sum_{x_1, x_2 \in W}^{\neq} 0.5 \left(\log \left(\frac{c_j}{g(\mathbf{c})} \right) (x_1) - \log \left(\frac{c_l}{g(\mathbf{c})} \right) (x_2) \right)^2 \mathfrak{R}_b(\|x_1 - x_2\| - r)}{\sum_{x_1, x_2 \in W}^{\neq} \mathfrak{R}_b(\|x_1 - x_2\| - r)},$$

as constant terms cancel.

Proposition 1. Equations 26 and 27 provide unbiased estimators and (25) thus yields a ratio-unbiased estimator as $b \rightarrow 0$.

To proof that the above proposition holds, consider the unmarked case first. Applying the Campbell theorem (Chiu *et al.*, 2013) we have $\mathbb{E}[\hat{\varrho}^{(2)}(r)] = \int \mathfrak{R}_b(s) \varrho^{(2)}(r + bs) ds$. Noting that $\mathbb{E}[\hat{\varrho}^{(2)}(r)] \rightarrow \varrho^{(2)}(r)$ as $b \rightarrow 0$ it follows that (27) is an unbiased estimator for $b \rightarrow 0$ of the second-order product density function. Analogously, $\varrho_{\mathcal{F}_f}^{\psi, \hat{j}l, (2)}$ of (26) can be shown to be an unbiased estimator of $\varrho_{\mathcal{F}_f}^{\psi, j\hat{l}, (2)}$ for $b \rightarrow 0$ by applying the Campbell theorem to the marked case (Daley & Vere-Jones, 2003) such that $\mathbb{E}[\varrho_{\mathcal{F}_f}^{\psi, \hat{j}l, (2)}(r)] \rightarrow \varrho_{\mathcal{F}_f}^{\psi, j\hat{l}, (2)}$. As both (26) and (27) yield unbiased estimators, (25) yields a ratio-unbiased estimator for $b \rightarrow 0$.

4 Test of Random Labelling Hypothesis for Composition-Valued Marked Point Processes

To test for deviations from the null hypothesis of random labels, that is, marks that are *i.i.d.* and thus independent of each other and the points, we adopt global envelope tests. These are non-parametric tests based on s simulations of the test statistic under the null model, originally introduced by Myllymäki *et al.* (2017) to solve multiple testing problems in spatial statistics. In the case of the random labelling hypothesis, the simulations can be obtained simply by permuting the marks of the points (e.g. Myllymäki *et al.*, 2015). In the case of composition-valued marks, we take the same approach, that is, permute the composition-valued marks.

Thus, first, the marks are permuted s times. The next step then is to compute the test statistic from the observed marked point pattern $\{(x_i, \mathbf{c}(x_i))\}_{i=1}^n$ and the s simulated patterns with permuted composition-valued marks. Let $\vartheta_1(r)$ stand for the empirical functional test statistic computed from the observed pattern and let $\vartheta_2(r), \dots, \vartheta_{s+1}(r)$ be the s functional test statistics computed from the s simulated patterns. Then a Monte Carlo test is done based on $\vartheta_1(r), \dots, \vartheta_{s+1}(r)$. If the functional test statistics can be ordered from the least extreme to the most extreme, a Monte Carlo p -value can be computed for the test in a similar manner as in the classical Monte Carlo test (Barnard, 1963). Here, to order the statistics, we use the extreme rank length (ERL) measure (Myllymäki *et al.*, 2017; Mrkvička *et al.*, 2020) as a particular instance of rank-based measures which also allow for the graphical interpretation of the test in terms of a global envelope. Please refer to the publications cited above and Myllymäki & Mrkvička (2023, Appendix A) for the definition of ERL and a discussion of alternative rank measures.

More precisely, let E_i , $i = 1, \dots, s + 1$ denote the measure associated with the i -th functional test statistic. Further, $<$ represents an ordering for E_i such that $E_i < E_j$ whenever ϑ_i is more extreme than ϑ_j with respect to the measure E . The critical value $E_{(\alpha)}$ under a given significance level α can then be found as the largest E_i which satisfies $\sum_{i=1}^{s+1} 1(E_i < E_{(\alpha)}) \leq \alpha(s + 1)$. Given

the set $I_{(\alpha)}$ of test statistics ϑ_i that are less than or as extreme as $E_{(\alpha)}$ as measured by their associated E_i , the $100(1 - \alpha)\%$ global envelope band is defined by the two functions $\vartheta_{(\alpha)}^l(r) = \min_{i \in I_{(\alpha)}} \vartheta_i(r)$ and $\vartheta_{(\alpha)}^u(r) = \max_{i \in I_{(\alpha)}} \vartheta_i(r)$. If $\vartheta_1(r)$ goes outside of $(\vartheta_{(\alpha)}^l(r), \vartheta_{(\alpha)}^u(r))$ for any of its argument values r , there is evidence to reject the null hypothesis at the given significance level α . The values of r at which $\vartheta_1(r)$ leaves the envelope show the reasons for the rejection of the test. There is a one-to-one correspondence between the graphical interpretation by the global envelope and the Monte Carlo p -value of the test, given by

$$p = \frac{1}{s+1} \left\{ 1 + \sum_{i=2}^{s+1} \mathbf{1}(E_i < E_1) \right\}. \quad (29)$$

That is, assuming that there are no pointwise ties in $\vartheta_i(r)$, $i = 1, \dots, s+1$, with probability 1, the empirical statistic $\vartheta_1(r)$ leaves the envelope if and only if $p \leq \alpha$, and $\vartheta_{(\alpha)}^l(r) \leq \vartheta_1(r) \leq \vartheta_{(\alpha)}^u(r)$ if $p > \alpha$ (Mrkvička *et al.*, 2022, Theorem 1). The size of the test based on the p -value (29), and thus the global envelope, is α when the test statistics $\vartheta_i(r)$ can be strictly ordered and $\alpha(s+1)$ is an integer (Myllymäki *et al.*, 2017, Lemma 1).

In practice, the functional test statistics are estimators of functional summary characteristics computed on a chosen finite but dense set of argument values r . If the random labelling hypothesis concerns all D parts of the composition-valued marks, the functional test statistics could be specified by $\vartheta(r) = \kappa_{\mathcal{F}_f}^{\hat{S}}(r)$ for any of the functional test statistics in Table 3. Such a test can point out distances r which are responsible for the potential rejection of the test, but does not focus on which components in the composition-valued marks show the strongest dependence.

Alternatively, the functional test statistic can be constructed from the componentwise mark summary characteristics using the combining procedure of Myllymäki & Mrkvička (2023, Appendix B). For example, for the componentwise mark variograms, that is, $\gamma_{jj}^w(r)$, $j = 1, \dots, \tilde{D}$, and a set of d r -values $r_1, \dots, r_D$, the test vector is constructed from the estimated mark variograms $\hat{\gamma}_{jj}^w(r)$, $j = 1, \dots, \tilde{D}$:

$$\boldsymbol{\vartheta} = ((\gamma_{11}^{\hat{w}}(r_1), \dots, \gamma_{11}^{\hat{w}}(r_d)), \dots, (\gamma_{\tilde{D}\tilde{D}}^{\hat{w}}(r_1), \dots, \gamma_{\tilde{D}\tilde{D}}^{\hat{w}}(r_d))). \quad (30)$$

This test summarises the information from all the components of the composition-valued marks using the chosen componentwise characteristics. This test holds the global significance level for the complete test vector (30) and it can point out both the components j and distances r which are responsible for the potential rejection of the test.

5 Applications

We use our new mark characteristics to analyse two marked point patterns from forestry and urban economics. The forestry data is an example of a spatial point pattern with 2-part composition-valued marks and additional absolute information (totals), while the urban economics data includes compositional marks with 4 parts. A comparison of the compositional analyses to separate analyses of the absolute raw marks is provided in the Supplement.

5.1 Application to Tree Data With Crown-to-Base Proportion

We analysed a pattern of 349 trees located in a plot of size 40 m $\times$ 40 m in Vesijako, southern Finland (see Figure 1). The data originate from a forest development study of managed, uneven-aged Norway spruce forests conducted under the ERIKA research project at the Natural

Table 4. Summary statistics for the relative height from the ground to the first living branch of the tree (h_b/h), the relative height of the crown (h_{cr}/h), and the total height (h) in meters for the Finnish tree data.

Mark	Min	1st quantile	Mean	Median	3rd quantile	Max
h_b/h	0.11	0.30	0.39	0.40	0.48	0.81
h_{cr}/h	0.19	0.52	0.61	0.60	0.70	0.89
h	1.36	3.19	6.30	8.10	12.50	26.2

Resources Institute Finland (Luke) (Eerikäinen *et al.*, 2007; Eerikäinen *et al.*, 2014; Saksa & Valkonen, 2011). The tree locations and associated tree characteristics, including the total height of the tree, h , and the height from the ground to the first living branch (of the crown), h_b , were recorded for all trees with $h \geq 1.3$ m. The height of the crown was obtained by $h_{cr} = h - h_b$. In what follows, we call h_b 'base' for short. Instead of the absolute values of h_{cr} and h_b (which clearly depend on the individual age of the trees), we considered the relative base height h_b/h (with geometric mean 0.37), relative crown height h_{cr}/h (with geometric mean 0.58) and the total height h . A summary of these three marks (two relational, one absolute) is provided in Table 4.

The total tree heights varied up to 26.2 meters with an average crown proportion of 61%. The spatial distribution of the marks and the corresponding ilr coordinates are depicted in Figure 1. Crown proportions tended to be rather large, with only some small values. Tree heights varied more. The ilr coordinates corresponding to the log-ratio transformation of the crown-to-base compositions support the above impressions: Most of the ilr coordinates are positive (indicated in red), which occurs when crown proportions are larger than base proportions.

Initially, we considered only the crown-to-base ratios through their ilr coordinates and conducted a separate analysis where the total height was examined at its original scale. Next, to extend the composition-valued mark by the absolute height information, we additionally included the log transformation of the total heights in a vector-valued mark in our computations according to Section 3.6. For both steps of our analysis and each mark characteristic, we computed the 95% global envelope under the random labelling hypothesis (see Section 4 for details), based on 3000 permutations. While the interplay of tree crowns and total heights, and competition between neighbouring trees have been of interest in different studies (Hui *et al.*, 2018; Pitkänen *et al.*, 2022), we are not aware of studies on the interdependencies of crown-to-base proportions and, additionally, the total heights over space. In particular, different from the existing approaches, the proposed rescaling of the absolute information into relative proportions allows for the characterisation of the structural properties of the marks without being affected by any heterogeneity in the size or types of the trees under study.

The top row of Figure 2 shows the compositional mark variogram γ_{cc} (left), the conditional mean product of marks τ_{cc} (central) and Shimantani's ι_{cc} (right) together with 95% global envelopes. We note that for $D = 2$ as in this application, all three compositional characteristics coincide with their componentwise analogues using the ilr transformation. While the first two empirical characteristics for the crown-to-base composition leave the global envelope for some distances r , the third one is completely inside the global envelope. The mark variogram suggests that the average dispersion of the crown-to-base log-ratios is smaller than expected under the random labelling hypothesis for any pair of points at interpoint distances of about $r = 6$ m. This would correspond to above random similarity in the crown-to-base log-ratios for neighbouring trees at intermediate distances.

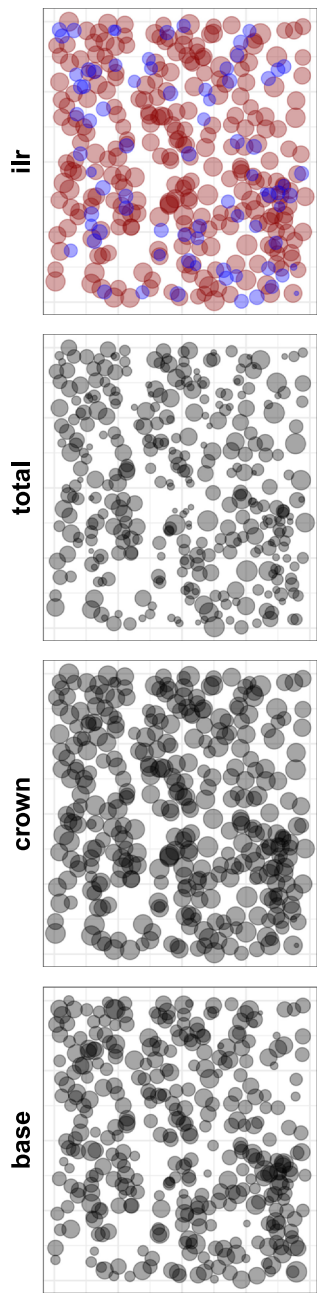


FIGURE 1. Spatial distribution of the original marks and *ilr*-transformed composition-valued marks for the Finnish forest stand data. Form left to right with the diameter of the discs proportional to the mark values and with negative values shown in blue and positive values highlighted in red: Base proportions, crown proportions, absolute height in metres per tree, and *ilr*-transformed crown-to-base proportions.

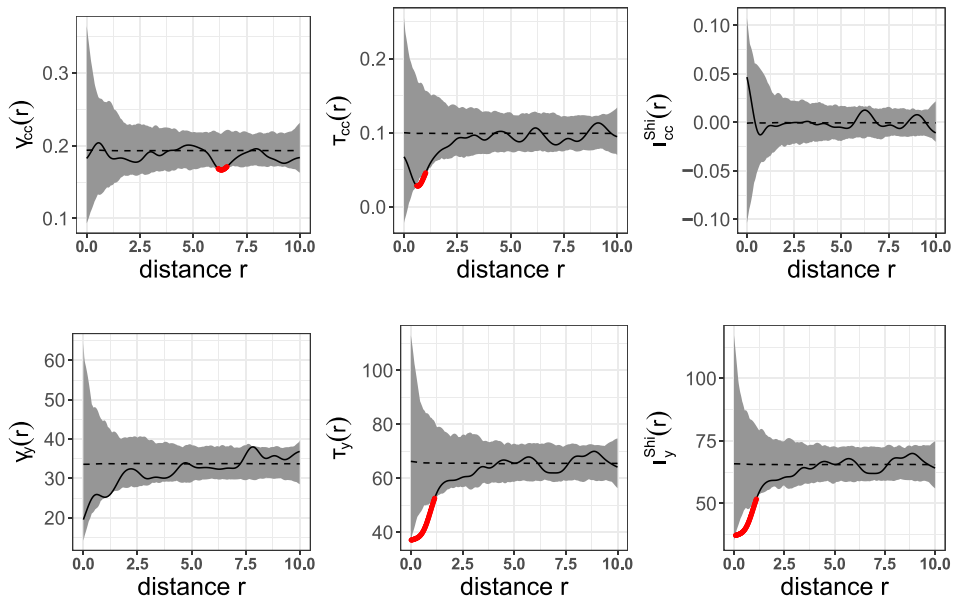


FIGURE 2. Selected compositional mark summary characteristics and 95% global envelopes (shaded) based on 3000 permutations of marks computed from the ilr transformed crown-to-base proportions (top row) and the total height (bottom row): mark variogram γ_{cc} (top left), the conditional mean product of marks τ_{cc} (top central), Shimantani's I_{cc}^{Shi} (top right), mark variogram γ_y (bottom left), the conditional mean scalar product of marks τ_y (bottom central) and Shimantani's I_y^{Shi} (bottom right). The dashed line corresponds to the mean function under the random labelling hypothesis and the solid curve to the test function on the observed data. Distances are given in meters.

The results for τ_{cc} suggest that the product of the crown-to-base log-ratios of two points at distance $r \approx 1$ m apart from each other tend to be smaller than expected under the random labelling hypothesis. Recall that small values of $\tau_{cc}(r)$ occur for distance r if the transformed tree compositions for any two points at a distance r are more different, for example, trees with large crown proportions (i.e. positive coordinates) are surrounded by trees with small crown proportions (i.e. negative coordinates). This finding might be explained by crown competition and growth restrictions due to space limitations for closely neighbouring trees. Reinspecting the ilr scores of Figure 1, positive values (red, i.e. large crown to small base ratio) co-occur closely to negative ilr coordinates (blue, i.e. small crown to large base ratio) which could explain the results.

Comparing the findings with the results for the total information depicted in the bottom row of Figure 2, both the conditional mean product of marks τ_y and Shimantani's I_y^{Shi} show clear negative deviations from the global envelopes for distances $r < 1.25$ m. These findings indicate that on average large trees are surrounded by smaller trees at shorter distances and vice versa implying negative autocorrelation. On the other hand, the mark variogram γ_y is completely within the 95% global envelope, even though it is rather close to the lower boundary for small r . Thus, no significant dependence was detected using the squared difference of the tree heights as the test function.

Additionally, specifying the weight β introduced in Section 3.6 as the ratio of the variances for the ilr-transformed composition and the log-transformed totals, we computed all three mark characteristics following the concepts outlined in Section 3.6. Accounting for the total information in the analysis of the crown-to-base composition, all three characteristics are completely

covered within the global envelopes (see Figure S1). This result means that when taking both mark components $\mathbf{c}$ and $\log(y)$ jointly into account, the marks did not show any significant spatial dependence or autocorrelation.

5.2 Application to Spanish Municipalities Data With Local Business Sector Compositions

For a second analysis, we evaluated data sourced from the National Statistics Institute of Spain (INE), focusing on the segmentation of the local economy into four sectors at the municipal level originating from the Spanish central business register. The analysis covers Spanish municipalities with populations over 1000, detailing the total number of economic entities, such as companies, situated within a municipality. To ensure consistency in data allocation and mitigate issues arising from businesses with extensive locations, INE matched each company to a single municipality during pre-processing, based on its headquarters' address. Hereafter, based on their primary economic activity as of January 1, 2022, the corporations were assigned to the categories (a) *industry* (encompassing extractive and manufacturing sectors, utilities, sanitation, waste management, and remediation), (b) *construction*, (c) *commerce* (covering both wholesale and retail trade, motor vehicle and motorcycle repairs, transportation, logistics, and hospitality), and (d) *services* (including communication, financial, insurance services, administrative, educational, healthcare, arts, recreation, and entertainment).

By focusing on the Spanish regions of *Albacete*, *Cuenca*, *Ciudad Real*, and *Toledo*, we gathered data from 66 municipalities with complete information on the business sector decomposition. By focusing on the Spanish regions of *Albacete*, *Cuenca*, *Ciudad Real*, and *Toledo*, we gathered data from 66 municipalities with complete information on the business sector decomposition. These provinces, located southeast of Madrid, form part of *La Mancha*, a plateau known for its uniform climate and population density. For these reasons, municipality locations in La Mancha are recognised as an example of a homogeneous spatial point process in the literature (see, e.g. Glass & Tobler, 1971; Ripley, 1977; Chiu *et al.*, 2013). The point pattern

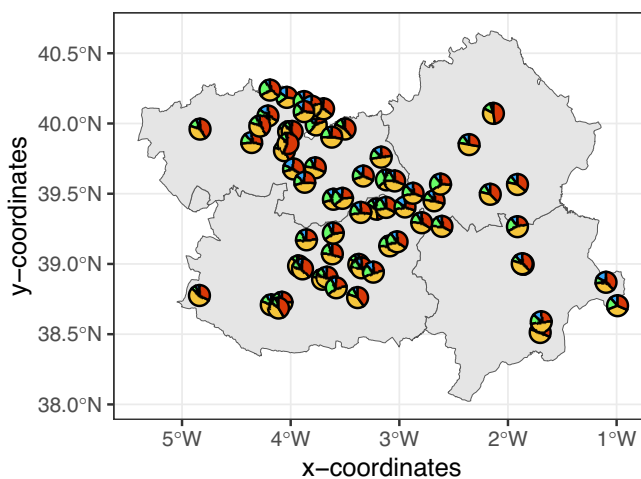


FIGURE 3. Distribution of four-part business sector composition on the Spanish Plateau for municipalities with at least 1000 inhabitants. Pie charts show the local decomposition of the economy into the business sector proportions of industry (blue), construction (green), commerce (orange) and service (red).

studied here is shown in Figure 3, with the four-part composition-valued marks depicted as pie charts.

The observed configuration of the marked points reflects a clear tendency of clustering for the point locations in combination with some variation over the individual pie charts, which indicate a clear predominance of the *commerce* and *service* sectors within the four-part compositions. While strong heterogeneity of the business sector decomposition appears at larger inter-point distances, the composition seems to become more homogeneous for closely neighbouring points.

This observed variation of the marks is also supported by the numerical summary statistics of the business sector proportions reported in Table 5, which again reflect a clear predominance of the commerce and service sectors contrasted with only smaller proportions of the industry and construction sectors. The geometric means of the four parts highlight clear differences between the sectors *industry* (0.09), *construction* (0.16), *commerce* (0.43) and *services* (0.32).

We computed the same three compositional mark summary characteristics with global envelopes as before to investigate the spatial joint variation, association and autocorrelation of the complete four-part composition (see Figure 4), using simple Euclidean distances on Longitude/Latitude due to the negligibility of the earth's curvature at this small spatial scale. While global envelopes for the mark variogram γ_{cc} (left) and Shimantani's i_{cc}^{Shi} (right) suggest deviations from the independent mark assumption for some distances r , the conditional mean of the inner product of marks τ_{cc} (central) is completely within the global envelope.

The mark variogram γ_{cc} indicates that the average dispersion between the transformed four-part compositions of any pairs of points is greater than expected under the independent mark assumption at distances of around 0.3 units. For distances $r \leq 0.2$ units of Longitude/Latitude, the mark variogram is smaller than expected under the random labelling

Table 5. Summary statistics for the closed 4-part business sector composition computed from the Spanish business sector data for 66 municipalities with at least 1000 inhabitants on the Spanish Plateau.

Sector (in %)	Min	1st quantile	Mean	Median	3rd quantile	Max
Industry	2.81	7.35	8.83	9.56	11.72	19.78
Construction	6.69	11.96	15.61	15.90	18.49	29.30
Commerce	31.15	38.28	41.73	42.02	44.96	56.07
Service	18.18	27.11	31.47	32.52	37.08	56.72

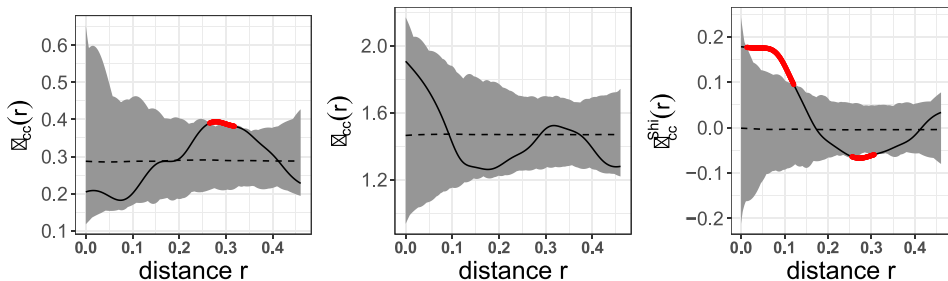


FIGURE 4. Compositional mark summary characteristics computed from the four-part economic sector compositions: Mark variogram γ_{cc} (left), the conditional mean of the product of marks τ_{cc} (central) and Shimantani's i_{cc}^{Shi} (right). The grey areas are 95% global envelopes constructed from 3000 simulations under the random labelling hypothesis. The dashed line corresponds to the mean function under the random labelling hypothesis and the solid curve to the empirical test function on the observed data.

hypothesis, although the empirical function stays within the envelope. This suggests similarity of the marks for pairs of nearby municipalities (although not significant), with increasing variability as the distances between point locations become larger, at least until about 0.3 units, where there is the most data. This might be explained by a strong variation in and clustering of the contribution of sectors such as tourism or banking to the local economy which, in turn, would affect the relative size of the service and commerce sectors. For Shimantani's γ_{cc}^{Shi} the findings suggest positive conditional spatial autocorrelation of the compositions at any nearby points with $r < 0.15$ units, and negative autocorrelation for $r \approx 0.3$ units, consistent with the results of the variogram.

To investigate the effect of each of the four parts on the above results, we additionally computed the componentwise clr mark variograms γ_{jj}^{clr} and componentwise Shimantani's $\gamma_{jj}^{clr, Shi}$ functions (see Figures S2 and S3). Of all four γ_{jj}^{clr} functions, only the mark variogram of clr(services) highlights deviations from the independent mark setting. This would suggest that the service sector proportions are more heterogeneous at larger distances, which is consistent with the visual impression from Figure 3. By contrast, empirical $\gamma_{jj}^{clr, Shi}$ functions show deviations from the random labelling hypothesis for all components except the construction sector (see Figure S3). While we found positive autocorrelations for clr(commerce) and clr(services) at distances $r \approx 0.1$ units, both clr(industry) and clr(services) are below the 95% envelopes at distances $r \approx 0.3$ units corresponding to negative autocorrelation. This again would relate to similarity among the business sectors for closely neighbouring municipalities and an increasing dissimilarity with increasing distances, but with some differences between sectors.

6 Conclusion

This paper presents the new class of composition-valued marked spatial point processes by integrating methodological principles for compositional data and spatial point processes. We propose various (functional) mark summary characteristics to assess mark independence and examine pairwise dependencies for this novel marked spatial point process. The framework is formalised using extended test functions that adapt established interpretations to this context. By converting composition-valued marks to the Euclidean space, the proposed methodologies leverage existing techniques for real-valued marks and utilise current computational implementations. Allowing for the characterisation of the spatial variation and association between both the complete composition-valued marks as well as their distinct compositional parts, the proposed extensions are helpful tools to highlight different aspects of the mark pattern. While the overall measures characterise the global interdependencies of the marks as a function of the interpoint distance, their componentwise counterparts provide useful insights into the contribution of the distinct parts to the overall results.

In addition to methods for purely composition-valued marks, we explored mixed marks combining composition-valued and absolute attributes. These techniques further enable separating vector-valued marks into their absolute and relative components, emphasising patterns in relative information while preventing results from being dominated by absolute data. For an integrated analysis of both types of information, we introduced weights within the scalar product to manage variation differences between the mark types (Happ & Greven, 2018).

We specifically addressed spatial point processes with composition-valued marks, which can be considered as a particular type of spatial point processes with so-called object-valued marks. Further extensions in this direction might also include alternative non-scalar marks such as density-valued or shape-valued marks. Note that as a by-product of our developed methods, we also showed how to handle vector-valued marks and derived both componentwise and

(full-vector) compositional summary characteristics as well as their relationship, a result which is of independent interest in its own right.

Acknowledgements

The authors gratefully acknowledge financial support through the German Research Foundation and Research Council of Finland. Matthias Eckardt and Sonja Greven were funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) - project numbers 467634837 (Walter Benjamin grant for ME) and 459422098 (SG), respectively. Mari Myllymäki was financially supported by the Research Council of Finland (Grant numbers 295100, 327211) and the European Union – NextGenerationEU in the Research Council of Finland project (Grant number 348154) under flagship ecosystem for Forest-Human-Machine Interplay – Building Resilience, Redefining Value Networks and Enabling Meaningful Experiences (UNITE) (Grant numbers 337655 and 357909). We thank Sauli Valkonen (Luke) for providing the Norway spruce data (ERIKA), and Hilkka Ollikainen and Juhani Korhonen for measuring the plots.

Open Access funding enabled and organized by Projekt DEAL.

References

- Aitchison, J. (1983). Principal component analysis of compositional data. *Biometrika*, **70**(1), 57–65.
- Aitchison, J. (1986). *The statistical analysis of compositional data*. Chapman & Hall: GBR.
- Aitchison, J. 2001. Simplicial inference. In *Algebraic methods in statistics and probability (Notre Dame, IN, 2000)*, Contemp. Math., Vol. **287**, Amer. Math. Soc.: Providence, RI, pp. 1–22.
- Aitchison, J. & Shen, S.M. (1980). Logistic-normal distributions: some properties and uses. *Biometrika*, **67**(2), 261–272.
- Baddeley, A. 2010. Handbook of spatial statistics. In *Chapman & Hall/CRC Handbooks of Modern Statistical Methods*, CRC Press, pp. 371–402.
- Barnard, G.A. (1963). Discussion on ‘the spectral analysis of point processes’ (by M. S. Bartlett). *J. R. Stat. Soc. Ser. B Stat. Method.*, **25**, 264–296.
- Billheimer, D., Guttorp, P. & Fagan, W.F. (2001). Statistical interpretation of species composition. *J. Am. Stat. Assoc.*, **96**(456), 1205–1214.
- Chayes, F. (1960). On correlation between variables of constant sum. *J. Geophys. Res.*, **65**(12), 4185–4193.
- Chiu, S.N., Stoyan, D., Kendall, W.S. & Mecke, J. (2013). *Stochastic geometry and its applications*, Third. John Wiley & Sons.
- Comas, C., Delicado, P. & Mateu, J. (2008). Analysing spatial point patterns with associated functional data. In *Proceedings of the 4th International Workshop on Spatio-Temporal Modelling (metma-4) Statistics for Spatio-Temporal Modelling Edited by 157-163*.
- Cressie, N. (1993). *Statistics for Spatial Data*. Wiley.
- Daley, D.J. & Vere-Jones, D. (2003). *An Introduction to the Theory of Point Processes. Volume I*. Springer: Berlin-Heidelberg.
- Eckardt, M., Comas, C. & Mateu, J. (2025). Summary characteristics for multivariate function-valued spatial point process attributes. *International Statistical Review*, **93**, 150–178.
- Eckardt, M. & Mateu, J. (2019a). Analysing multivariate spatial point processes with continuous marks: a graphical modelling approach. *Int. Stat. Rev.*, **87**(1), 44–67.
- Eckardt, M. & Mateu, J. (2019b). Partial characteristics for marked spatial point processes. *Environmetrics*, **30**(6), e2565.
- Eckardt, M. & Moradi, M. (2024). Marked spatial point processes: current state and extensions to point processes on linear networks. *J. Agric. Biol. Environ. Stat.*, **29**(2), 346–378.
- Eerikäinen, K., Miina, J. & Valkonen, S. (2007). Models for the regeneration establishment and the development of established seedlings in uneven-aged, Norway spruce dominated forest stands of Southern Finland. *For. Ecol. Manag.*, **242**(2), 444–461.
- Eerikäinen, K., Valkonen, S. & Saksa, T. (2014). Ingrowth, survival and height growth of small trees in uneven-aged picea abies stands in Southern Finland. *For. Ecosyst.*, **1**(1), 5.

- Egozcue, J.J. & Pawłowsky-Glahn, V. (2005). Groups of parts and their balances in compositional data analysis. *Math. Geol.*, **37**(7), 795–828.
- Fišerová, E. & Hron, K. (2011). On the interpretation of orthonormal coordinates for compositional data. *Math. Geosci.*, **43**(4), 455.
- Ghorbani, M., Cronie, O., Mateu, J. & Yu, J. (2021). Functional marked point processes: a natural structure to unify spatio-temporal frameworks and to analyse dependent functional data. *TEST*, **30**, 529–568.
- Glass, L. & Tobler, W.R. (1971). General: Uniform distribution of objects in a homogeneous field: cities on a plain. *Nature*, **233**(5314), 67–68.
- Guan, Y. (2006). Tests for independence between marks and points of a marked point process. *Biometrics*, **62**(1), 126–134.
- Guan, Y. & Afshartous, D.R. (2007). Test for independence between marks and points of marked point processes: a subsampling approach. *Environ. Ecol. Stat.*, **14**, 101–111.
- Happ, C. & Greven, S. (2018). Multivariate functional principal component analysis for data observed on different (dimensional) domains. *J. Am. Stat. Assoc.*, **113**(522), 649–659.
- Harkness, R.D. & Isham, V. (1983). A bivariate spatial point pattern of ants' nests. *J. R. Stat. Soc., C: Appl. Stat.*, **32**(3), 293–303.
- Ho, L.P. & Stoyan, D. (2008). Modelling marked point patterns by intensity-marked Cox processes. *Stat. Probab. Lett.*, **78**(10), 1194–1199.
- Hron, K., Engle, M., Filzmoser, P. & Fišerová, E. (2021). Weighted symmetric pivot coordinates for compositional data with geochemical applications. *Math. Geosci.*, **53**(4), 655–674.
- Hron, K., Filzmoser, P., de Caritat, P., Fišerová, E. & Gardlo, A. (2017). Weighted pivot coordinates for compositional data and their application to geochemical mapping. *Math. Geosci.*, **49**(6), 797–814.
- Hui, G., Wang, Y., Zhang, G., Zhao, Z., Bai, C. & Liu, W. (2018). A novel approach for assessing the neighborhood competition in two different aged forests. *For. Ecol. Manag.*, **422**, 49–58.
- Illian, J., Penttinen, A., Stoyan, H. & Stoyan, D. (2008). *Statistical analysis and modelling of spatial point patterns*. John Wiley & Sons: New York.
- Isham, V. 1985. Marked point processes and their correlations. Publications des Facultes Universitaires Saint-Louis Brussels, Spatial processes and spatial time series analysis.
- Kynčlová, P., Hron, K. & Filzmoser, P. (2017). Correlation between compositional parts based on symmetric balances. *Math. Geosci.*, **49**(6), 777–796.
- Leininger, T.J., Gelfand, A.E., Allen, J.M. & Silander, J.A. (2013). Spatial regression modeling for compositional data with many zeros. *J. Agric. Biol. Environ. Stat.*, **18**(3), 314–334.
- Lotwick, H.W. & Silverman, B.W. (1982). Methods for analysing spatial processes of several types of points. *J. R. Stat. Soc. Ser. B Stat. Method.*, **44**(3), 406–413.
- Mateu-Figueras, G., Pawłowsky-Glahn, V. & Egozcue, J.J. 2011. The principle of working on coordinates. In *Compositional data analysis*, John Wiley & Sons, Ltd, pp. 29–42.
- Moran, P.A.P. (1950). Notes on continuous stochastic phenomena. *Biometrika*, **37**(1/2), 17–23.
- Mrkvička, T., Myllymäki, M., Kuronen, M. & Narisetty, N.N. (2022). New methods for multiple testing in permutation inference for the general linear model. *Stat. Med.*, **41**(2), 276–297.
- Mrkvička, T., Myllymäki, M., Jilek, M. & Hahn, U. (2020). A one-way ANOVA test for functional data with graphical interpretation. *Kybernetika*, **56**(3), 432–458.
- Myllymäki, M., Grabarnik, P., Seijo, H. & Stoyan, D. (2015). Deviation test construction and power comparison for marked spatial point patterns. *Spat. Stat.*, **11**, 19–34.
- Myllymäki, M. & Mrkvička, T. 2023. Get: Global envelopes in R. arXiv:1911.06583 [stat.ME].
- Myllymäki, M., Mrkvička, T., Grabarnik, P., Seijo, H. & Hahn, U. (2017). Global envelope tests for spatial processes. *J. R. Stat. Soc. Ser. B Stat. Method.*, **79**(2), 381–404.
- Pawłowsky-Glahn, V. & Buccianti, A. (2011). *Compositional data analysis*. John Wiley & Sons, Ltd.
- Pawłowsky-Glahn, V. & Egozcue, J. (2001). Geometric approach to statistical analysis on the simplex. *Stoch. Environ. Res. Risk Assess.*, **15**(5), 384–398.
- Pawłowsky-Glahn, V., Egozcue, J.J. & Lovell, D. (2015). Tools for compositional data with a total. *Stat. Model.*, **15**(2), 175–190.
- Pawłowsky-Glahn, V. & Ricardo, A.O. (2004). *Geostatistical analysis of compositional data*, IAMG studies in Mathematical Geology, Vol. 7. Oxford University Press: United Kingdom.
- Penttinen, A. & Stoyan, D. (1989). Statistical analysis for a class of line segment processes. *Scand. J. Stat.*, **16**(2), 153–168.
- Penttinen, A., Stoyan, D. & Henttonen, H.M. (1992). Marked point processes in forest statistics. *For. Sci.*, **38**, 806–824.

- Pitkänen, T.P., Bianchi, S. & Kangas, A. (2022). Quantifying the effects of competition on the dimensions of Scots pine and Norway spruce crowns. *Int. J. Appl. Earth Obs. Geoinf.*, **112**, 102941.
- Ripley, B.D. (1977). Modelling spatial patterns. *Journal of the Royal Statistical Society. Series B*, **39**(2), 172–212.
- Saksa, T. & Valkonen, S. (2011). Dynamics of seedling establishment and survival in uneven-aged boreal forests. *For. Ecol. Manag.*, **261**(8), 1409–1414.
- Schlather, M. (2001). On the second-order characteristics of marked point processes. *Bernoulli*, **7**(1), 99–117.
- Schlather, M., Riberio, P. & Diggle, P. (2004). Detecting dependence between marks and locations of marked point processes. *J. R. Stat. Soc., B: Stat. Methodol.*, **66**, 79–93.
- Shimatani, K. (2002). Point processes for fine-scale spatial genetics and molecular ecology. *Biom. J.*, **44**(3), 325–352.
- Stoyan, D. (1984a). Correlations of the marks of marked point processes - statistical inference and simple models. *J. Inf. Process. Cybern.*, **20**(5/6), 285–294.
- Stoyan, D. (1984b). On correlations of marked point processes. *Math. Nachr.*, **116**(1), 197–207.
- Stoyan, D. (1987). Statistical analysis of spatial point processes: a soft-core model and cross-correlations of marks. *Biom. J.*, **29**(8), 971–980.
- Stoyan, D. & Stoyan, H. (1994). *Fractals, random shapes, and point fields: methods of geometrical statistics*. Wiley: Chichester, New York.
- Stoyan, D. & Wälder, O. (2000). On variograms in point process statistics, II: models for markings and ecological interpretation. *Biom. J.*, **42**, 171–187.
- Tolosana Delgado, R. (2006). Geostatistics for constrained variables: positive data, compositions and probabilities. Applications to environmental hazard monitoring. Ph.D. Thesis, University of Girona.
- van Lieshout, M.N.M. (2006). A j-function for marked point patterns. *Ann. Inst. Stat. Math.*, **58**(2), 235.
- Van Lieshout, M.N.M. & Baddeley, A.J. (1999). Indices of dependence between types in multivariate point patterns. *Scand. J. Stat.*, **26**(4), 511–532.
- Wiegand, T. & Moloney, K.A. (2013). *Handbook of spatial point-pattern analysis in ecology*. Chapman and Hall/CRC.

[Received May 2025; accepted November 2025]