

Eckardt, Matthias; Comas, Carles; Mateu, Jorge

**Article — Published Version**

## Summary characteristics for multivariate function-valued spatial point process attributes

International Statistical Review

*Suggested Citation:* Eckardt, Matthias; Comas, Carles; Mateu, Jorge (2024) : Summary characteristics for multivariate function-valued spatial point process attributes, International Statistical Review, ISSN 1751-5823, Vol. 93, Iss. 1, pp. 150-178, <https://doi.org/10.1111/insr.12582>

This Version is available at:

<https://hdl.handle.net/10419/319336>

### Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

### Terms of use:

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*



<http://creativecommons.org/licenses/by/4.0/>

# Summary characteristics for multivariate function-valued spatial point process attributes

Matthias Eckardt<sup>1</sup> , Carles Comas<sup>2</sup> and Jorge Mateu<sup>3</sup>

<sup>1</sup>Department of Statistics, Humboldt-Universität zu Berlin, Spandauer Strasse 1, Berlin, Germany

<sup>2</sup>Department of Mathematics, Universitat de Lleida, Av. Alcalde Rovira Roure 191, Lleida, Spain

<sup>3</sup>Department of Mathematics, Universitat Jaume I, E-12071 Castellón, Spain

Correspondence Matthias Eckardt, Department of Statistics, Humboldt-Universität zu Berlin, Spandauer Strasse 1, Berlin, Germany. Email: [m.eckardt@hu-berlin.de](mailto:m.eckardt@hu-berlin.de)

## Summary

Prompted by modern technologies in data acquisition, the statistical analysis of spatially distributed function-valued quantities has attracted a lot of attention in recent years. In particular, combinations of functional variables and spatial point processes yield a highly challenging instance of such modern spatial data applications. Indeed, the analysis of spatial random point configurations, where the point attributes themselves are functions rather than scalar-valued quantities, is just in its infancy, and extensions to function-valued quantities still remain limited. In this view, we extend current existing first- and second-order summary characteristics for real-valued point attributes to the case where, in addition to every spatial point location, a set of distinct function-valued quantities are available. Providing a flexible treatment of more complex point process scenarios, we build a framework to consider points with multivariate function-valued marks, and develop sets of different cross-function (cross-type and also multi-function cross-type) versions of summary characteristics that allow for the analysis of highly demanding modern spatial point process scenarios. We consider estimators of the theoretical tools and analyse their behaviour through a simulation study and two real data applications.

*Key words:* cross-function mark correlation; forest monitoring data; functional-marked point processes; mark variogram; mark weighted second order summary characteristics; nearest neighbour mark indices; urban economics.

## 1 Introduction

Introducing general ideas from functional data analysis (Ferraty & Vieu, 2006; Horváth & Kokoszka, 2012; Hsing & Eubank, 2015; Ramsay & Silverman, 1997) into the field of spatial statistics, the statistical analysis of functional spatial data has attracted a lot of attention in recent years (see Delicado *et al.*, 2010; Mateu & Romano, 2017, Martínez-Hernández & Genton, 2020, for a general review). Potential applications from the literature include the analysis of regional penetration resistance profiles (Giraldo *et al.*, 2011), air pollution monitoring data (Bohorquez *et al.*, 2017) and regional gross domestic product dynamics (Pineda-Ríos *et al.*, 2019). Different from more classical spatial statistics (see, e.g. Cressie, 1993), the objects of interest in any such

data are the realised trajectories, that is, curves, of some underlying continuous mechanism which are collected over some spatial domain  $\mathcal{X} \subset \mathbb{R}^d$ , usually  $d = 2$ . As such, the functional observations themselves are assumed to be spatially dependent relative to the distance between the spatial entities, which needs to be accounted for in any statistical analysis. The observed functions at nearby stations might be more similar or dissimilar in shape depending on where the measurements are recorded. Clearly, in an urban context the amount of gases and particulates in the air depends on the location of the monitoring stations, and nearby stations are likely to have similar air pollution profiles. This spatial structure among the functions is, however, not accounted for in any non-spatial methods yielding potentially biased or misleading results. To this end, various approaches from classical spatial statistics were extended to functional outcomes yielding an ever-increasing methodological toolbox of different functional spatial data analysis techniques. Predetermined by the exact nature of  $\mathcal{X}$ , these techniques help to investigate the spatial interrelations between the individual functional objects in geostatistical, areal or point process data contexts. While a relatively large body of contributions exists on geostatistical functional data and functional areal data, the analysis of spatial random point configurations is just in its infancy. In particular, different from geostatistical functional data/functional kriging approaches (see Bohorquez *et al.*, 2022; Franco-Villoria & Ignaccolo, 2022; Nerini *et al.*, 2022, for general reviews) or functional areal data regression models (Aw & Cabral, 2020; Pineda-Ríos *et al.*, 2019; Zhang *et al.*, 2016), the point locations are treated as random and the attributes themselves are functions rather than scalar-valued quantities.

Despite the notable progress in spatial point process methodology with extensions to more challenging non-Euclidean domains for the points including the sphere, linear networks and graphs with Euclidean edges, extensions to more complicated non-scalar marks have not been covered much so far. For the scalar-valued marks setting, there already exist various mark summary characteristics and nearest-neighbour versions (Stoyan & Stoyan, 1994). Here, prominent tools for the characterisation of real-valued marks include the mark covariance (Stoyan, 1984), mark correlation (Isham, 1985; Stoyan & Stoyan, 1994), mark weighted  $K$  (Penttinen *et al.*, 1992), mark variogram (Cressie, 1993; Stoyan & Wälder, 2000; Wälder & Stoyan, 1996), and mark differentiation (Hui & Pommerening, 2014; Pommerening *et al.*, 2011) functions.

Literature currently offers methodology for (marked) spatio-temporal point processes (González *et al.*, 2016; Rathbun, 1993; Vere-Jones, 2009) where different clustered point process (González *et al.*, 2016), Gibbsian processes (Redenbach & Särkkä, 2013; Renshaw *et al.*, 2009; Renshaw & Särkkä, 2001; Särkkä & Renshaw, 2006), log-Gaussian (Serra *et al.*, 2014; Siino *et al.*, 2018) and shot-noise (Brix & Chadœuf, 2002; Møller & Díaz-Avalos, 2010) Cox model specifications are developed. We can also find corresponding second-order summary characteristics (Iftimi *et al.*, 2019; Stoyan *et al.*, 2017) used to characterise the temporal evolution for a set of (marked) point locations. However, contributions to function-valued marks remain elusive. In particular, advances to sets of distinct function-valued attributes, that is, multivariate curves, do not exist.

The mark correlation function for function-valued point attributes originates from the pioneering paper of Comas *et al.* (2008) and subsequent works by Comas *et al.* (2011); Comas *et al.* (2013). Ghorbani *et al.* (2021) were the first to provide a mathematically rigorous treatment on the subject. Instead of the set  $\{(\mathbf{x}_i, m(\mathbf{x}_i))\}_{i=1}^n$  with points  $\mathbf{x}_i \in \mathcal{X}$  and scalar-valued marks  $m(\mathbf{x}_i)$  on some suitable mark space  $\mathbb{M}$ , these authors considered the set  $\{(\mathbf{x}_i, (f(\mathbf{x}_i)(t), l(\mathbf{x}_i)))\}_{i=1}^n$  where each point  $\mathbf{x}_i$  is augmented by a function-valued quantity  $f(\mathbf{x}_i)(t) \in \mathbb{F}$  and, potentially, an additional Euclidean auxiliary mark  $l(\mathbf{x}_i)$  living on a suitable latent mark space  $\mathbb{L}$ . As such, apart from the trivial case when no auxiliary mark is available,

this formulation allows for (i) function-valued marked multivariate point patterns, where different types of points with one function-valued point attribute are observed, and (ii) function-valued marked point patterns where at each point additional real-valued information is available. While providing a flexible treatment of more complex point process scenarios which include unmarked (resp. scalar-valued marked) point processes as special case when the  $(f(\mathbf{x}_i)(t), l(\mathbf{x}_i))$  (resp.  $f(\mathbf{x}_i)(t)$ ) argument is ignored, extensions to points with multiple distinct function-valued marks have not been covered to the very best of our knowledge. Considering at least two distinct function-valued point attributes for each point location, this paper aims to fill this gap. In particular, sets of different cross-function summary characteristics for points with two distinct function-valued marks are introduced and extended to cross-type and also multi-function cross-type versions. Overall, the proposed tools will allow for the analysis of highly demanding modern spatial point process scenarios. All data and R code to reproduce the proposed auto- and cross-function mark characteristics are made publicly available in a github repository <https://github.com/carlescomas/SppFDA>.

The remainder of the paper is structured as follows. After a general introduction to spatial point processes with multivariate function-valued point attributes, Section 2 establishes different cross-function mark characteristics and potential extension to multitype point processes. In particular, extensions of classical test functions to the function-valued mark setting are discussed in Section 2.2. Estimators of the proposed mark characteristics are presented in Section 3. The proposed characteristics are evaluated through a simulation study in Section 4. Section 5 presents an application of the proposed tools to two different data sources originating from forestry and urban economic contexts. The paper concludes with a discussion in Section 6.

## 2 Spatial Point Processes With Multivariate Function-Valued Marks

To extend the theory and methodology of function-valued marked spatial point processes to multivariate function-valued point attributes, let  $\mathcal{X}$  denote a subset of  $\mathbb{R}^2$  equipped with Borel sets  $\mathcal{B}(\mathcal{X})$ , and  $d(\cdot)$  an Euclidean metric on  $\mathcal{X}$ . On  $\mathcal{X}$ , define  $\Psi_G = \{\mathbf{x}_i\}_{i=1}^n$  as ground, that is, unmarked, spatial point process with intensity measure  $\Lambda_G$ . As such,  $\Psi_G$  is well embedded into the theory of spatial point processes and a rich body of different tools can directly be applied to investigate the structural properties of the points (see Møller & Waagepetersen, 2004; Illian *et al.*, 2008; Chiu *et al.*, 2013, for general references to spatial point processes). Associated with  $\Psi_G$ , denote by  $\Psi = \{\mathbf{x}_i \mathbf{f}(\mathbf{x}_i)(t)\}_{i=1}^n$  a marked spatial point process on  $\mathcal{X} \times \mathbb{F}^p$  with locations  $\mathbf{x}_i \in \mathcal{X}$  and  $p$ -variate associated function-valued marks  $\mathbf{f}(\mathbf{x}_i)(t) = (f_1(\mathbf{x}_i)(t), \dots, f_p(\mathbf{x}_i)(t))$  on  $\mathbb{F}^p$  where each  $f_h(\mathbf{x}_i)(t): \mathcal{T} \subseteq \mathbb{R} \mapsto \mathbb{R}$ ,  $h = 1, \dots, p$  with  $\mathcal{T} = (a, b)$ ,  $-\infty \leq a \leq b \leq \infty$ . In general,  $\mathbb{F}^p$  is assumed to be a Polish, that is, complete separable metric space equipped with  $\sigma$ -algebra  $\mathcal{F}^p = \otimes_{h=1}^p \mathcal{F}_h$  (Daley & Vere-Jones, 2008). For  $\Psi$ , the expected number of points  $N_h(\cdot)$  in  $B \in \mathcal{B}(\mathcal{X})$  with function-valued attribute in  $F_h \in \mathcal{F}_h$  corresponds to the intensity measure  $\Lambda_h(B \times F_h)$  which simplifies to

$$\mathbb{E}[N_h(B \times F_h)] = \Lambda_h(B \times F_h) = \int_{B \times F_h} \lambda_G(\mathbf{x}) d\mathbf{x} P(dF_h)$$

for fixed  $F_h$  in  $\mathcal{F}_h$  with  $\lambda_G$  being the intensity function of  $\Psi_G$  and  $P(dF_h)$  a reference measure on  $(\mathbb{F}^p, \mathcal{F}^p)$ . For stationary  $\Psi$ , that is, if  $\Psi = \Psi_x$  with  $\Psi_x = \{\mathbf{x}_i + x, \mathbf{f}(\mathbf{x}_i)(t)\}_{i=1}^n$  for any translation  $x$  and fixed  $f_h(t)$ , the intensity measure  $\Lambda_h(B \times F_h)$  equals  $\lambda_h v(B)$ , with  $\lambda_h$  denoting the intensity of  $\Psi$  with respect to  $F_h$  and  $v(\cdot)$  the Lebesgue measure, that is, the volume, of its argument. Similarly,  $\Psi$  is called isotropic if  $\Psi = \tau\Psi$  with  $\tau\Psi = \{\{\tau\mathbf{x}_i, \mathbf{f}(\mathbf{x}_i)(t)\}\}_{i=1}^n$  for any rotation  $\tau$ .

To account for additional integer-valued marks, that is, when different types of points are available,  $\Psi$  can be generalised to a  $m$ -variate (i.e. multitype) spatial point process  $\mathbf{\Psi}$  with  $n = n_1 + \dots + n_m$  points and multivariate function-valued point attributes on  $\mathcal{X}^m \times \mathbb{F}^p$  with corresponding component processes  $\Psi_d$ ,  $d = 1, \dots, m$  and associated ground process  $\Psi_G$ . We note that the above point processes could also be extended by additional real-valued mark information, for example, through additional auxiliary mark terms  $l(\mathbf{x}_i) \in \mathbb{R}$ , which allows for the formulation of doubly-marked (multitype) point processes where each point is augmented by multivariate function-valued marks and one (resp.  $w$  distinct) real-valued mark living on  $\mathcal{X}^m \times \mathbb{F}^p \times \mathbb{R}$  (resp.  $\mathcal{X}^m \times \mathbb{F}^p \times \mathbb{R}^w$ ).

### 2.1 Functional Summary Characteristics for Unmarked Point Processes

Before discussing various mark characteristics, a succinct overview of the standard functional summary characteristics commonly associated with unmarked point patterns is presented. In the absence of any mark information, the spatial arrangement of points is routinely examined via the assumption of complete spatial randomness, wherein the absence of discernible structure is assessed, against clustering or regularity of the points. Using empty-space, nearest-neighbour or pairwise distances, commonly applied characteristics include the empty space function  $F(r)$ , the nearest neighbour distance distribution function  $G(r)$ , the pair correlation function  $g(r)$  and Ripley's  $K(r)$  function (Ripley, 1977). The empty space function  $F(r)$  considers the distance from a typical point to a point in the pattern, whereas the nearest neighbour distance distribution function  $G(r)$  is constructed from the distance of any neighbouring points. Under complete spatial randomness, corresponding to a homogeneous Poisson processes, both the  $F(r)$  and  $G(r)$  functions are equal to  $1 - e^{-\lambda\pi r^2}$  where  $\lambda$  is the intensity of the points. The pair correlation function is constructed from the pairwise distance between the locations and becomes equal to 1 under the complete spatial randomness assumption. The reader is referred to the complete essays of Illian *et al.* (2008) and Chiu *et al.* (2013).

### 2.2 Cross-Function Second-Order Mark Summary Characteristics and Nearest Neighbour Indices

Apart from the first-order properties, a variety of second-order mark summary characteristics and their related nearest-neighbour versions have become useful methodological tools for the analysis of classical (real-valued) marked spatial point process scenarios. They help to investigate the heterogeneity and interrelation between the observed point attributes, and decide on the independent mark hypothesis as a function of the distance between pairs of two points. To extend the methodological toolbox to the function-valued marks setting and define suitable cross-function characteristics, let  $f_h(\mathbf{x})(t)$  and  $f_l(\mathbf{x}')(t)$  denote two distinct function-valued marks for a pair of distinct point locations in  $\Psi$  with interpoint distance  $d(\mathbf{x}, \mathbf{x}') = r$ ,  $r > 0$ . Adopting the core principles from classical mark characteristics and applying a pointwise evaluation first, different cross-function mark characteristics can be defined by introducing a test function  $t_f$  (Penttinen & Stoyan, 1989), that is, a map  $t_f: \mathbb{F} \times \mathbb{F} \rightarrow \mathbb{R}^+$ , which itself takes the marks  $f_h(t)$  and  $f_l(t)$  for a pair of distinct points in  $\Psi$  as arguments. In what follows, we assume  $\Psi$  to be second-order stationary and isotropic such that the characteristics solely depend on the distance  $r$ ,  $r > 0$  and it suffices to consider the marks at the origin  $\circ$  and the distance  $\mathbf{r}$  where  $d(\circ\mathbf{r}) = r$ . Depending on the precise specification of  $t_f$ , different mark characteristics can be constructed by taking the expectation  $\mathbb{E}_{\circ, \mathbf{r}}$  of  $t_f$  under the condition that  $\Psi$  has indeed points at locations  $\circ$  and  $\mathbf{r}$ . Writing  $f_h(\circ)(t)$  and  $f_l(\mathbf{r})(t)$  to denote the  $h$ -th and  $l$ -th function-valued marks at the origin  $\circ$  and at a distance  $\|\mathbf{r}\| = r$  apart, different summary

characteristics can be obtained by generalising their classic form to the function-valued marks setting. Focusing on the most prominent test functions from the literature, we yield the following five specifications, noting that each  $t$  has as input arguments  $(f_h(\circ)(t), f_l(\mathbf{r})(t))$ :  $t_1 = 1/2(f_h(\circ)(t) - f_l(\mathbf{r})(t))^2$  and  $t_2 = (\min(f_h(\circ)(t), f_l(\mathbf{r})(t)))/(\max(f_h(\circ)(t), f_l(\mathbf{r})(t)))$ , which are based on the difference or the ratio between the pair of distinct marks, and  $t_3 = f_h(\circ)(t) \cdot f_l(\mathbf{r})(t)$ ,  $t_4 = f_h(\circ)(t)$  and  $t_5 = f_l(\mathbf{r})(t)$ , which are based on the product of the arguments (see Schlather, 2001; Illian *et al.*, 2008, for detailed discussion). All of the above test functions address only certain aspects of the mark distribution. Taking the conditional expectation of test function  $t_1$  yields the mark variogram which depicts the variation of the marks as a function of the distance  $r$ . The mark variogram could be used to decide on the average (dis)similarity for pairs of marks at distinct points. If the marks are on average similar, their variation is on average also small leading small values of the mark variogram. A similar characteristic, the mark differentiation function, is obtained by taking the conditional expectation of  $1 - t_2$ . If the marks are similar in values, the ratio in  $t_2$  will become close to one such that  $1 - t_2$  is also close to zero. Different from these two test functions,  $t_3$  and  $t_4$  or  $t_5$  can be used to compute Stoyan's mark correlation and the  $\mathbf{r}$ -mark functions. Although different from Pearson's correlation, Stoyan's mark correlation (i.e. the conditional expectation of  $t_3$ ) reflects the average pairwise association of the marks at any two distinct points. Under independence of the marks, this characteristic coincides with the squared mean mark such that normalising Stoyan's mark correlation by the squared mean mark yields a value of one. If the marks are on average larger (resp. smaller) than the squared mean mark, the estimated products will also be clearly different from the squared mean mark. Finally, the conditional expectation of either  $t_4$  or  $t_5$  relates to the conditional mean of the first or the second mark for any pair of points given that there are points at both locations, respectively. The conditional mean could be used to detect dependence between the marks and the points. We note that only under independent marks, the conditional expectation of  $t_4$  or  $t_5$  coincides with the mark mean. Obviously, the above formulations include auto-mark characteristics as special cases for  $h = l$ . We note that apart from a concurrent setting, alternative pointwise test functions may be defined for the marks  $f_h(t)$  and  $f_l(s)$  with  $s < t$ .

### 2.2.1 Cross-function variation and differentiation characteristics

As a first cross-function mark characteristic, the pointwise cross-function mark variogram  $\gamma_{hl}(r, t)$ , which helps to investigate the strength and range of the variation in the mark differences with respect to the distance  $\mathbf{r}$ , can be derived by taking the conditional expectation  $\mathbb{E}_{\circ, \mathbf{r}}$  of  $t_1$ . This pointwise characteristic can then be turned into a global cross-function mark variogram  $\gamma_{hl}(r)$ , which corresponds to the  $L_2$  metric, by integrating  $\gamma_{hl}(r, t)$  over  $\mathcal{T}$ ,

$$\gamma_{hl}(r) = \int_{\mathcal{T}} \mathbb{E}_{\circ, \mathbf{r}}[t_1(f_h(\circ)(t), f_l(\mathbf{r})(t))] dt =: \int_{\mathcal{T}} \gamma_{hl}(r, t) dt.$$

The limit of this characteristic equals the non-spatial variance, and the normalised version yields a straight line that is constantly one under the independent mark assumption. In contrast, large values of this cross-function characteristic will indicate a strong heterogeneity between the function-valued attributes at a distance  $r$ .

Different from the cross-function mark variogram, taking  $t_3$  as argument of  $\mathbb{E}_{\circ, \mathbf{r}}$  yields a pointwise cross-function version of Stoyan's mark covariance function (Stoyan, 1984),

$$\text{Cov}_{hl}^{\text{Sto}}(r, t) = \mathbb{E}_{\circ, \mathbf{r}}[t_3] - \mu_h(t) \cdot \mu_l(t)$$

where  $\mu_h(t) = \mathbb{E}[f_h(t)]$  and  $\mu_l(t) = \mathbb{E}[f_l(t)]$  are the non-spatial means of  $f_h(t)$  and  $f_l(t)$ ,

respectively. Again, a corresponding global characteristic is achieved as  $\text{Cov}_{hl}^{\text{Sto}}(r) = \int \text{Cov}_{hl}^{\text{Sto}}(r, t) dt$ . Alternatively, a cross-function mark covariance can also be obtained by rewriting Cressie's (Cressie, 1993) covariance function  $\text{Cov}_{hl}^{\text{Cre}}$  as  $\text{Cov}_{hl}^{\text{Cre}}(r, t) = \mathbb{E}_{\circ, \mathbf{r}}[t_3] - \mathbb{E}_{\circ, \mathbf{r}}[t_4]\mathbb{E}_{\circ, \mathbf{r}}[t_5]$ , with  $\text{Cov}_{hl}^{\text{Cre}}(r) = \int \text{Cov}_{hl}^{\text{Cre}}(r, t) dt$ .

Inserting the ratio of  $t_2$  instead of the difference between the paired marks into the conditional expectation  $\mathbb{E}_{\circ, \mathbf{r}}[\cdot]$  yields a pointwise cross-function version of the mark differentiation function (Hui & Pommerening, 2014; Pommerening *et al.*, 2011)  $\tau_{hl}(r, t)$  for function-valued marks, defined by  $\tau_{hl}(r, t) = 1 - \mathbb{E}_{\circ, \mathbf{r}}[t_2]$  with global characteristic  $\tau_{hl}(r) = \int \tau_{hl}(r, t) dt$ . Obviously, values of  $\tau_{hl}(r, t)$  equal or close to zero imply that the function-valued point attributes at distance  $r$  are equal or almost identical while increasing non-zero values indicate an increase in heterogeneity of the marks.

### 2.2.2 Cross-function correlation characteristics

Different from the difference and ratio based characteristics, an alternative set of cross-function mark characteristics can be defined through the product of the function-valued marks. Taking  $t_3$  as argument of  $\mathbb{E}_{\circ, \mathbf{r}}$  yields a pointwise cross-function version of the conditional mean product of marks  $c_{hl}(r, t)$  within a distance  $r$  at  $t \in \mathcal{T}$ . We note that  $c_{hl}(r, t)$  translates again into a global cross-function characteristic  $c_{hl}(r)$  by integration of the pointwise one over  $\mathcal{T}$ ,

$$c_{hl}(r) = \int_{\mathcal{T}} \mathbb{E}_{\circ, \mathbf{r}}[f_h(\circ)(t) \cdot f_l(\mathbf{r})(t)] dt =: \int_{\mathcal{T}} c_{hl}(r, t) dt. \tag{1}$$

Further, normalising  $c_{hl}(r, t)$  by  $\mu_h(t) \cdot \mu_l(t)$ , that is, the product of non-spatial means, yields a cross-function pointwise version of Stoyan's mark cross-correlation function  $\kappa_{hl}(r, t)$  (Stoyan, 1987) from which the global characteristic  $\kappa_{hl}(r)$  follows analogous to (1) by integration of  $\kappa_{hl}(r, t)$  over  $\mathcal{T}$ . We note that Stoyan's mark covariance function is indeed a linear transformation of the mark correlation function such that  $\text{Cov}_{hl}^{\text{Sto}}(r)$  and  $\kappa_{hl}(r)$  are essentially the same (Schlather, 2001).

Apart from Stoyan's mark correlation function, Isham (1985) and Beisbart and Kerscher (2000) introduced two alternative mark correlation functions that could also be extended to the function-valued mark setting. Beisbart and Kerscher (2000) proposed a simpler version of the above formulation of the mark correlation function in which the product of the mark values is replaced by the normalised sum of marks. Using this formulation allows for a straightforward extension to a pointwise version for function-valued marks  $f_h(\circ)(t)$  and  $f_l(\mathbf{r})(t)$  defined by

$$\kappa_{hl}^{\text{Bei}}(r, t) = \frac{f_h(\circ)(t) + f_l(\mathbf{r})(t)}{\mu_h(t) + \mu_l(t)}$$

with  $\kappa_{hl}^{\text{Bei}}(r) = \int \kappa_{hl}^{\text{Bei}}(r, t) dt$ . As the nominator approaches the product (resp. the sum) of means  $\mu_h(t)$  and  $\mu_l(t)$  as limits, both  $\kappa_{hl}$  and  $\kappa_{hl}^{\text{Bei}}$  are constantly equal to one for all  $t \in \mathcal{T}$  in case of independent marks. In contrast, positive or negative mark correlations could easily be identified by positive or negative deviations from one, respectively. Opposite to the above formulations, Isham (1985) introduced a different type of mark correlation function which is closely related to Pearson's correlation. Using the cross-function version of Cressie's mark covariance function, a pointwise cross-function analogue to Isham's mark correlation function can be defined as

$$\text{Cor}_{hl}^{\text{Ish}}(r, t) = \frac{\text{Cov}_{hl}^{\text{Cre}}(r, t)}{\sqrt{\text{Var}_{hh}(r, t)} \cdot \sqrt{\text{Var}_{ll}(r, t)}}$$

where  $\text{Var}_{hh}(r, t) = \mathbb{E}_{\circ, \mathbf{r}}[(f_h(\circ)(t) \cdot f_h(\mathbf{r})(t))] - \mathbb{E}_{\circ, \mathbf{r}}[f_h(\circ)(t)]\mathbb{E}_{\circ, \mathbf{r}}[f_h(\mathbf{r})(t)]$  and  $\text{Var}_{ll}(r, t)$  is defined analogously, and  $\text{Cor}_{hl}^{\text{Ish}}(r) = \int \text{Cor}_{hl}^{\text{Ish}}(r, t) dt$ .

Apart from the extended cross-function mark correlation characteristics outlined above, taking  $t_4$  or  $t_5$  as arguments of  $\mathbb{E}_{\circ, \mathbf{r}}$  leads to pointwise  $\mathbf{r}$ -mark functions  $c_{h\bullet}(r, t)$  and  $c_{\bullet l}(r, t)$ , respectively, where  $c_{h\bullet}(r, t) = c_{\bullet l}(r, t)$ . As before, both pointwise  $\mathbf{r}$ -mark functions translate into global characteristics  $c_{h\bullet}(r)$  and  $c_{\bullet l}(r)$  by integration of  $c_{h\bullet}(r, t)$  and  $c_{\bullet l}(r, t)$  over  $\mathcal{T}$ , respectively. Further, normalisation of  $c_{h\bullet}(r, t)$  and  $c_{\bullet l}(r, t)$  by  $\mu_h(t)$  and  $\mu_l(t)$  yields the pointwise  $\mathbf{r}$ -mark correlation functions  $\kappa_{h\bullet}(r, t)$  and  $\kappa_{\bullet l}(r, t)$ , respectively, where  $\kappa_{h\bullet}(r) = \int \kappa_{h\bullet}(r, t) dt$  and  $\kappa_{\bullet l}(r) = \int \kappa_{\bullet l}(r, t) dt$ .

We note that the cross-function mark correlation and  $\mathbf{r}$ -correlation functions can also be used to define a counterpart version of the  $U(r)$  function for function-valued marks, this  $U(r)$  being the mean product of marks sited at distance  $r$  apart,

$$U(r) = \int \lambda^2 g(r) \kappa_{hl}(r) da da', \tag{2}$$

where  $\lambda \equiv \lambda_G$  is the intensity of the points,  $g(r)$  the pair correlation function, and  $a$  and  $a'$  are two infinitesimal small areas containing points  $\mathbf{x}$  and  $\mathbf{x}'$  which are separated by a distance  $r$  (Capobianco & Renshaw, 1998; Renshaw, 2002). Including second-order summary characteristics for both the points and the function-valued marks, these characteristics account jointly for spatial variation of the point locations and the marks. Under the independent marks assumption,  $\kappa_{hl}(r) = 1$  whereas  $g(r) = 1$  under the complete spatial randomness hypothesis, that is, the homogeneous Poisson point process case. Alternative formulation of  $U(r)$  can be achieved by substituting  $\kappa_{hl}(r)$  by the  $\mathbf{r}$ -correlation functions  $\kappa_{h\bullet}(r)$  and  $\kappa_{\bullet l}(r)$ , the mark variogram  $\gamma_{hl}(r)$  and the mark differentiation function  $\tau_{hl}(r)$ , or alternatively by rewriting  $U(r)$  into polar coordinates allows for handling anisotropic behaviour.

### 2.2.3 Cross-function nearest-neighbour indices and $k$ -nearest neighbour characteristics

While second-order cross-characteristics provide functional summary characteristics of the pairwise interrelations between the function-valued point attributes against the distance  $r$ , nearest-neighbour indices are essentially numerical mark summary characteristics which help to quantify the local variation between the marks for a pair of nearest-neighbouring points. Similar to the previous sections, different cross-function nearest-neighbour characteristics can be constructed by taking the conditional expectation of particular test functions. In contrast to the above sections these, however, only consider the function-valued marks  $f_h(\circ)(t)$  and  $f_l(\mathbf{z}(\circ))(t)$  at the origin  $\circ$  and its nearest neighbouring point  $\mathbf{z}(\circ)$  (Stoyan & Stoyan, 1994).

Rewriting the test function  $t_3$  into a nearest-neighbour version  $t_3^{\text{nn}} = f_h(\circ)(t) \cdot f_l(\mathbf{z}(\circ))(t)$  and taking the conditional expectation  $\mathbb{E}_{\circ, \mathbf{z}(\circ)}[t_3^{\text{nn}}]$  leads to a pointwise cross-function nearest-neighbour mark product index  $c_{hl}^{\text{nn}}(t)$ . The corresponding pointwise nearest-neighbour mark product correlation index  $\kappa_{hl}^{\text{nn}}(t)$  derives directly from  $c_{hl}^{\text{nn}}(t)$  by normalising  $c_{hl}^{\text{nn}}(t)$  by the product of means  $\mu_h(t) \cdot \mu_l(t)$ . Likewise, taking the conditional expectation of  $t_4^{\text{nn}} = f_l(\mathbf{z}(\circ))(t)$  yields a pointwise nearest-neighbour mark index  $c_{\bullet l}^{\text{nn}}(t)$  which transforms into the pointwise nearest-neighbour mark correlation index by normalising  $c_{\bullet l}^{\text{nn}}(t)$  by  $\mu_l$ . Similarly, pointwise cross-function nearest neighbour mark variogram and mark differentiation indices  $\gamma_{hl}^{\text{nn}}(t)$  and

$\tau_{hl}^{nn}(t)$  can be constructed by taking the conditional expectation  $\mathbb{E}_{\circ, \mathbf{z}(\circ)}$  of the test functions  $\mathfrak{t}_1^{nn} = 1/2(f_h(\circ)(t) - f_l(\mathbf{z}(\circ))(t))^2$  and  $\mathfrak{t}_2^{nn} = \min_{nn}/\max_{nn}$  where  $\min_{nn} = \min(f_h(\circ)(t), f_l(\mathbf{z}(\circ))(t))$  and  $\max_{nn} = \max(f_h(\circ)(t), f_l(\mathbf{z}(\circ))(t))$ , respectively. We note that all pointwise indexes translate into global numerical summary characteristics by integration of the pointwise version over  $\mathcal{T}$ .

Apart from considering only the function-valued mark of the nearest neighbouring point location  $\mathbf{z}(\circ)$ , the nearest neighbour indices can also be used to compute cumulative cross-function  $k$ -nearest neighbour summary characteristics from the marks  $f_h$  and  $f_l$  at the origin  $\circ$  and its  $k$ -th nearest neighbouring point  $\mathbf{z}_v(\circ)$  with  $v = 1, \dots, k$ . Substituting  $\mathbf{z}(\circ)$  by  $\mathbf{z}_v(\circ)$ , a corresponding cumulative mark correlation index can be computed from the pointwise cross-function  $k$ -th nearest neighbouring index  $\mathcal{K}_k(t)$ ,

$$\mathcal{K}_k(t) = \frac{1}{k} \mathbb{E}_{\circ, \mathbf{z}_v} \left( \sum_{v=1}^k f_h(\circ)(t) \cdot f_l(\mathbf{z}_v(\circ))(t) \right) / \mu_h(t) \mu_l(t),$$

with  $\mathcal{K}_k = \int \mathcal{K}_k(t) dt$ . Likewise, a cumulative mark variogram index can be defined as  $\Gamma_k = \int \Gamma_k(t) dt$  with

$$\Gamma_k(t) = \frac{1}{k} \mathbb{E}_{\circ, \mathbf{z}_v} \left( \sum_{v=1}^k \frac{1}{2} (f_h(\circ)(t) - f_l(\mathbf{z}_v(\circ))(t))^2 \right).$$

In addition, a pointwise counterpart version of Hui's mark dominance index  $D_k$  (Hui *et al.*, 1998) for function-valued marks can be defined as

$$D_k(t) = \frac{1}{k} \mathbb{E}_{\circ, \mathbf{z}_v} \left( \sum_{v=1}^k 1(f_h(\circ)(t) > f_l(\mathbf{z}_v(\circ))(t)) \right)$$

which translates into a global characteristics by computing  $D_h = \int D_k(t) dt$ .

### 2.3 Cross-Function Mark-Weighted Summary Characteristics

A different useful set of cross-function cumulative summary characteristics can be defined by adjusting classical functional point process summary characteristics for the function-valued marks by introducing a test function as weight into the specific functional point process summary expression. Although the principal idea also applies for the empty space and nearest neighbour contact distribution functions and related quantities, we explicitly only cover extension of second-order summary characteristics to the function-valued mark scenario including the mark-weighted pair correlation,  $K$  and  $L$  functions.

#### 2.3.1 Mark-weighted characteristics for unitype point processes with multivariate function-valued marks

To define a suitable pair correlation function for function-valued marks  $f_h(\mathbf{x})(t)$  and  $f_l(\mathbf{x})(t)$  in  $\Psi$ , let  $\alpha_{\mathfrak{t}_f}^{(2)}(t)$  denote the pointwise cross-function second-order factorial moment measure with density  $\varrho_{\mathfrak{t}_f}^{(2)}(t)$ , that is, the pointwise cross-function second-order product density functions. For  $\mathfrak{t}_f = \mathfrak{t}_3$ ,  $\alpha_{hl}^{(2)}$  becomes

$$\alpha_{hl}^{(2)}(B_1 \times B_2)(t) = \sum_{\substack{\neq \\ (\mathbf{x}, f_h(\mathbf{x})), \\ (\mathbf{x}', f_l(\mathbf{x}')) \in \Psi}} \mathbb{E}[1_{B_1}(\mathbf{x})1_{B_2}(\mathbf{x}')f_h(\mathbf{x})(t) \cdot f_l(\mathbf{x}')(t)]$$

with density  $\varrho_{hl}^{(2)}(t)$ . Then, a pointwise mark-weighted pair correlation function  $g_{hl}(r, t)$  can be defined as  $g_{hl}(r, t) = \varrho_{hl}^{(2)}(r, t)/(\lambda^2\mu_h(t)\mu_l(t))$  and  $g_{hl}(r) = \int g_{hl}(r, t)dt$ . Specifying  $t_f = t_5$  instead, the pair correlation function equals  $g_{\bullet l}(r, t) = \varrho_{\bullet l}^{(2)}(r, t)/(\lambda^2\mu_l(t))$  where  $\varrho_{\bullet l}^{(2)}$  is the density of  $\alpha_{\bullet l}^{(2)}(t)$ .

Similarly, setting  $t_f = t_3$  a cross-function pointwise extension of the mark-weighted  $K$  function (Penttinen *et al.*, 1992) can be defined as

$$\lambda\mu_h(t)\mu_l(t)K_{hl}(r, t) = \mathbb{E}_{\circ} \left[ \sum_{(\mathbf{x}, f_l(\mathbf{x})) \in \Psi} f_h(\circ)(t) \cdot f_l(\mathbf{x})(t) 1_{b(\circ\mathbf{r})}\{\mathbf{x}\} \right],$$

where  $b(\circ\mathbf{r})$  is a ball of radius  $\mathbf{r}$  centred at the origin,  $\lambda$  is the intensity of the points and  $\mu_h(t)$  and  $\mu_l(t)$  are the non-spatial means of  $f_h(t)$  and  $f_l(t)$ , respectively. Likewise, for  $t_4$  and  $t_5$ , the mark weighted pointwise  $K$  function changes to

$$\lambda\mu_h(t)K_{h\bullet}(r, t) = \mathbb{E}_{\circ} \left[ \sum_{(\mathbf{x}, f_l(\mathbf{x})) \in \Psi} f_h(\circ)(t) 1_{b(\circ\mathbf{r})}\{\mathbf{x}\} \right]$$

and

$$\lambda\mu_l(t)K_{\bullet l}(r, t) = \mathbb{E}_{\circ} \left[ \sum_{(\mathbf{x}, f_l(\mathbf{x})) \in \Psi} f_l(\mathbf{x})(t) 1_{b(\circ\mathbf{r})}\{\mathbf{x}\} \right],$$

respectively (see Illian *et al.*, 2008). Analogous to the classical scalar-valued case, the mark-weighted  $L$  function  $L_{t_f}(r, t)$  is preferable to use in practice instead of the mark-weighted  $K_{t_f}(r, t)$  function, where  $L_{t_f}(r, t) = \sqrt{K_{t_f}(r, t)/\pi}$ . As for the pair correlation function, both  $K_{t_f}(r, t)$  and  $L_{t_f}(r, t)$  translate into global cross-function characteristics by integration of the pointwise versions over  $\mathcal{T}$ .

Finally, applying the principal idea of the above mark-weighted second-order summary characteristics to local summary characteristics, a localised version of the pointwise cross-function mark-weighted  $K$  function for the  $u$ -th point location of  $\Psi$  can be defined for  $t_f = t_3$  as

$$\lambda\mu_h(t)\mu_l(t)K_u(r, t) = \mathbb{E} \left[ \sum_{(\mathbf{x}', f_l(\mathbf{x}')) \in \Psi} f_h(\mathbf{x}_u)(t) \cdot f_l(\mathbf{x}')(t) 1_{b(\mathbf{x}_u\mathbf{r})}\{\mathbf{x}'\} \right].$$

### 2.3.2 Extensions to multitype points with multivariate function-valued marks

In what follows, let  $\Psi_i = \{(\mathbf{x}_i, \mathbf{f}(\mathbf{x}_i)(t))\}_{i=1}^{n_i}$  and  $\Psi_j = \{(\mathbf{x}_j, \mathbf{f}(\mathbf{x}_j)(t))\}_{j=1}^{n_j}$  denote two component processes of  $\Psi$  where  $i \neq j$  and  $\mathbf{f}(\mathbf{x}_i)(t)$  and  $\mathbf{f}(\mathbf{x}_j)(t)$  are  $p$ -variate function-valued marks on  $\mathbb{F}^p$ . For  $t_f = t_3$  the pointwise cross-function second-order factorial moment measure can be rewritten into a cross-function cross-type measure  $\alpha_{ij, hl}^{(2)}(t)$ ,

$$\alpha_{ij,hl}^{(2)}(B_1 \times B_2)(t) = \sum_{\substack{\neq \\ (\mathbf{x}_i, f_h(\mathbf{x}_i)) \in \Psi_i, \\ (\mathbf{x}_j, f_l(\mathbf{x}_j)) \in \Psi_j}} \mathbb{E} [1_{B_1}(\mathbf{x}_i) 1_{B_2}(\mathbf{x}_j) f_h(\mathbf{x}_i)(t) \cdot f_l(\mathbf{x}_j)(t)],$$

with density  $\varrho_{ij,hl}^{(2)}(B_1 \times B_2)(t)$ . This product density, in turn, allows to define a pointwise cross-function cross-type pair correlation function  $g_{ij,hl}(r, t) = \varrho_{ij,hl}(t) / (\lambda_i \lambda_j \mu_h(t) \mu_l(t))$  with  $g_{ij,hl}(r) = \int g_{ij,hl}(r, t) dt$ . Further, a pointwise mark-weighted cross-function cross-type  $K$  function can be defined as

$$\lambda_i \mu_h(t) \mu_l(t) K_{ij,hl}(r, t) = \mathbb{E}_{\circ, i} [t_f(f_h(\circ)(t), f_l(\mathbf{r}(t)) 1(b(\circ), r) \{ \mathbf{x}_j \})]$$

from which a pointwise cross-function dot-type version can be obtained through  $\lambda K_{i^*,hl} = \sum_j \mu_h(t) \mu_l(t) \lambda_j K_{ij,hl}(r)$ .

### 2.4 Test Functions for Points With Multiple Function-Valued Marks

Previous sections were restricted to the derivation of mark summary characteristics defined through a set of different test functions with at most two distinct function-valued point attributes. Here, we now discuss potential extensions to marked point process scenarios with  $p \geq 3$  distinct function-valued marks. For simplicity, we restrict to the trivariate case which could be extended naturally to sets of  $p$  with  $p \geq 3$  distinct function-marks.

For points with function-valued marks  $\mathbf{f}(\mathbf{x})(t) = (f_d(\mathbf{x})(t), f_h(\mathbf{x})(t), f_l(\mathbf{x})(t))$ , three different generalised test functions can be defined through unconditional and also conditional formulations. As a general first approach, function  $f_d(\mathbf{x})(t)$  could be related to the set  $\{f_h(\mathbf{x})(t), f_l(\mathbf{x})(t)\}$  by specifying  $t(f_d(\circ)(t), \{f_h f_l\}(\mathbf{r})(t))$ . Using this formulation and writing  $\mu_{hl}(\mathbf{r})(t)$  for the mean of functions  $h$  and  $l$  at  $t \in \mathcal{T}$  and distance  $\mathbf{r}$ , trivariate versions of  $t_1$  and  $t_3$  can be expressed as  $t_1(f_d(\circ)(t), \{f_h f_l\}(\mathbf{r})(t)) = 1/2(f_d(\circ)(t) - \mu_{hl}(\mathbf{r})(t))^2$  and  $t_3(f_d(\circ)(t), \{f_h f_l\}(\mathbf{r})(t)) = (f_d(\circ)(t) \cdot \mu_{hl}(\mathbf{r})(t))$ . Instead of  $\mu_{hl}$ , alternative formulations might be defined through the sum of pairwise operations, for example,

$$t_1(f_d(\circ)(t), \{f_h f_l\}(\mathbf{r})(t)) = 1/2((f_d(\circ)(t) - f_h(\mathbf{r})(t))^2 + (f_d(\circ)(t) - f_l(\mathbf{r})(t))^2).$$

Instead of only the  $h$ -th and  $l$ -th function-valued marks, a second general approach could be defined by relating the  $d$ -th function to all three functions, that is,  $t(f_d(\circ)(t), \mathbf{f}(\mathbf{r})(t))$ . Similar to the above formulation, suitable specifications might be defined through  $\mu_{dhl}(\mathbf{r})(t)$ , the mean of set  $\mathbf{f}(\mathbf{x})(t)$  at  $t \in \mathcal{T}$  and distance  $\mathbf{r}$ , or alternatively using a pairwise formulation which, in turn, would combine both auto- and cross-type terms. While both of the above versions are specified through unconditional formulations, an alternative approach might derive from the conditional function-valued marks  $f_d | f_l$  and  $f_h | f_l$  yielding  $t(f_d | f_l, |f_l(\circ)(t), f_h | f_l, |f_l(\mathbf{r})(t))$ . Both conditional marks could be derived by partialising out the effect of  $f_l(\mathbf{x})(t)$  from  $f_d(\mathbf{x})(t)$  and  $f_h(\mathbf{x})(t)$  using standard functional regression methods. Although interesting, this conditional formulation will not be pursued in this paper to make it more concise and focused.

## 3 Estimation of Cross-Function Summary Characteristics

After having discussed extensions of various mark characteristics for (multitype) spatial point processes with multiple function-valued marks, their estimation from observed spatial point patterns is presented next. As before, the empirical cross-function estimators for unitype point patterns are first described. To this end, let  $\psi$  denote a spatial point pattern of  $n$  points observed in a

bounded observation window  $W$ , where each point is augmented by a multivariate function-valued point attribute. Denote by  $\psi$  the corresponding multitype point pattern with multivariate function-valued marks and components  $\psi_1, \dots, \psi_d$ . Further, let  $\text{card}(\cdot)$  denote the cardinality, that is, the number of points, in the argument.

### 3.1 Estimation of Cross-Function Summary Characteristics

Using the results of Section 2.2 and writing  $c_t(r)$  to denote the conditional expectation for any specific test function  $t_f$ , both variation and product related cross-function mark summary characteristics can be derived through a generic function  $\kappa_t(r)$ . Again, the specific form of  $\kappa_t(r)$  itself depends on the specification of  $t_f$ . We can estimate the second-order  $t_f$ -product density  $\varrho_t^{(2)}(r)$ , and its analogue version  $\varrho^{(2)}(r)$  of the ground pattern  $\psi_G$ , by

$$\widehat{\varrho}_t^{(2)}(r) = \frac{1}{2\pi r\nu(W)} \sum_{\substack{\neq \\ (\mathbf{x}f_h), \\ (\mathbf{x}'f_l) \in \psi}} \ell(t_f(f_h(\mathbf{x})(t), f_l(\mathbf{x}')(t))) \mathfrak{K}_b(\|\mathbf{x} - \mathbf{x}'\| - r) e(\|\mathbf{x} - \mathbf{x}'\|), \tag{3}$$

and

$$\widehat{\varrho}^{(2)}(r) = \frac{1}{2\pi r\nu(W)} \sum_{\mathbf{x}, \mathbf{x}' \in \psi_G} \mathfrak{K}_b(\|\mathbf{x} - \mathbf{x}'\| - r) e(\|\mathbf{x} - \mathbf{x}'\|), \tag{4}$$

respectively, where

$$\ell(t_f(f_h(\mathbf{x}), f_l(\mathbf{x}'))) = \int_a^b t_f(f_h(\mathbf{x})(t), f_l(\mathbf{x}')(t)) dt$$

and  $\mathfrak{K}_b(\cdot)$  is a kernel function with bandwidth  $b$ ,  $\nu(\cdot)$  the area of its argument, and  $e(\cdot)$  is an edge correction factor. Then,  $\kappa_t(r)$  can be estimated by

$$\widehat{\kappa}_t(r) = \frac{\widehat{\varrho}_t^{(2)}(r)}{\widehat{\varrho}^{(2)}(r)} \widehat{c}_t, \quad \text{for } r > 0, \tag{5}$$

where  $\widehat{c}_t = \sum_{\mathbf{x}} \sum_{\mathbf{x}'} \ell(t_f(f_h(\mathbf{x})(t), f_l(\mathbf{x}')(t))) / n^2$  is an estimator of  $c_t = c_t(\infty)$ .

We note that  $\kappa_t(r)$  can alternatively also be estimated by

$$\widehat{\kappa}_t(r) = \frac{1}{2\pi r\nu(W)} \sum_{\substack{\neq \\ (\mathbf{x}f_h), \\ (\mathbf{x}'f_l) \in \psi}} \frac{\ell(t_f(f_h(\mathbf{x}), f_l(\mathbf{x}'))) \mathfrak{K}_b(\|\mathbf{x} - \mathbf{x}'\| - r) e(\|\mathbf{x} - \mathbf{x}'\|)}{\widehat{\lambda}^2 \widehat{g}(r) \widehat{c}_t} \tag{6}$$

where  $\widehat{g}(r) = \widehat{\varrho}^{(2)}(r) / \widehat{\lambda}^2$ ,  $r \geq 0$ , and  $\widehat{\lambda} = \text{card}(W) / \nu(W)$  are estimators for the pair correlation function and the intensity of the ground process, respectively. Specifying  $t_f$  by  $t_1$  and  $t_3$  in the above formulation of  $\widehat{\kappa}_t$ , the cross-function mark variogram and mark correlation function can be estimated by

$$\widehat{\gamma}_{hl}(r) = \frac{1}{2\pi r\nu(W)} \sum_{\substack{\neq \\ (\mathbf{x}f_h), \\ (\mathbf{x}'f_l) \in \psi}} \frac{\ell(t_1(f_h(\mathbf{x})(t), f_l(\mathbf{x}')(t))) \mathfrak{K}_b(\|\mathbf{x} - \mathbf{x}'\| - r) e(\|\mathbf{x} - \mathbf{x}'\|)}{\widehat{\lambda}^2 \widehat{g}(r) \widehat{c}_t}. \tag{7}$$

and

$$\widehat{\kappa}_{hl}(r) = \frac{1}{2\pi r v(W)} \sum_{\substack{\neq \\ (\mathbf{x}'f_i), \\ (\mathbf{x}'f_l) \in \psi}} \frac{\ell(\mathbf{t}_3(f_h(\mathbf{x})(t), f_l(\mathbf{x}')'(t))) \mathfrak{R}_b(\|\mathbf{x} - \mathbf{x}'\| - r) e^{(\mathbf{x}\|\mathbf{x} - \mathbf{x}'\|)}}{\widehat{\lambda}^2 \widehat{g}(r) \widehat{\mu}_h \widehat{\mu}_l}, \quad (8)$$

respectively, where  $\widehat{\mu}_l$  is the empirical functional mean of mark  $f_l$ . Similarly, estimators for  $\kappa_{h*}(r)$  (resp.  $\kappa_{*l}(r)$ ) can be obtained by setting  $\mathbf{t}_f$  to  $\mathbf{t}_4$  (resp.  $\mathbf{t}_5$ ) and substituting  $\widehat{\mu}_h$  (resp.  $\widehat{\mu}_l$ ) for  $\widehat{\mu}_h \widehat{\mu}_l$  in (8).

### 3.2 Estimation of Cross-Function Nearest-Neighbour Indices

Estimators for the cross-function nearest-neighbour indices can be derived analogous to Section 3.1 by replacing the above test functions by the nearest neighbour counterpart versions. Using the nearest neighbour test functions  $\mathbf{t}_1^{\text{nn}}$  and  $\mathbf{t}_3^{\text{nn}}$ , the nearest-neighbour mark variogram and mark product correlation index can be estimated through

$$\widehat{\gamma}_{hl}^{\text{nn}} = \frac{1}{n} \sum_{i=1}^n \ell(\mathbf{t}_1^{\text{nn}}(f_h(\mathbf{x}_i)(t), f_l(\mathbf{z}(i))(t))) / \widehat{c}_t$$

and

$$\widehat{\kappa}_{hl}^{\text{nn}} = \frac{1}{n} \sum_{i=1}^n \ell(\mathbf{t}_3^{\text{nn}}(f_h(\mathbf{x}_i)(t), f_l(\mathbf{z}(i))(t))) / \widehat{\mu}_h \widehat{\mu}_h,$$

respectively.

### 3.3 Estimation of Cross-Function Mark-Weighted Summary Characteristics

Estimators of the cross-function mark weighted  $K$  function can be obtained by normalising the function  $\widehat{k}_{hl}$ ,

$$\widehat{k}_t(r) = \sum_{\mathbf{x}, \mathbf{x}' \in W}^{\neq} \frac{\ell(\mathbf{t}_f(f_h(\mathbf{x})(t), f_l(\mathbf{x}')'(t))) 1\{\|\mathbf{x} - \mathbf{x}'\| \leq r\}}{v(W)}$$

by the empirical versions of the intensity  $\widehat{\lambda}^2$  and a suitable normalising factor  $\widehat{c}_t$  corresponding to the specific test function used with  $\widehat{c}_t = \widehat{\mu}_h \widehat{\mu}_l$  for  $\mathbf{t}_f = \mathbf{t}_3$ . The normalised estimator can then be transformed in the corresponding cross-function mark weighted  $L$  function by taking the square root of  $\widehat{K}(r)$ . Likewise, the cross-function mark weighted pair correlation function can be computed as  $\widehat{g}_t(r) = \widehat{\varrho}_t^{(2)}(r) / \widehat{\lambda}^2 \widehat{c}_t$  which becomes  $\widehat{g}_{hl}(r) = \widehat{\varrho}_{hl}^{(2)}(r) / \widehat{\lambda}^2 \widehat{\mu}_h \widehat{\mu}_l$  for choosing  $\mathbf{t}_3$  as test function.

For the multitype point pattern scenario with components  $\psi_i$  and  $\psi_j$  and function-valued marks  $f_h$  and  $f_l$ , the cross-function cross-type mark weighted  $K$  function can be estimated by dividing

$$\widehat{k}_{ij,t}(r) = \sum_{\mathbf{x}_i, \mathbf{x}_j \in W}^{\neq} \frac{\ell(\mathbf{t}_f(f_h(\mathbf{x}_i)(t), f_l(\mathbf{x}_j)(t))) 1\{\|\mathbf{x}_i - \mathbf{x}_j\| \leq r\}}{v(W)}$$

by  $\widehat{\lambda}_i \widehat{\lambda}_j \widehat{c}_t$  with  $\widehat{\lambda}_i$  denoting the intensity of the  $i$ -th component of  $\boldsymbol{\psi}$ . Again, this function translates into the corresponding estimator of the  $L$  function through the square root of  $\widehat{K}$ . Similarly to the unitype case, a cross-function cross-type mark weighted pair correlation function can be calculated by computing  $\widehat{g}_{ij,t}(r) = \widehat{Q}_{ij,t}^{(2)}(r) / \widehat{\lambda}_i \widehat{\lambda}_j \widehat{c}_t$ .

#### 4 A Simulation Study

We conducted a simulation study to investigate how our estimators of the cross-function summary characteristics behave not only under several point configurations (including random, cluster and regular structures) but, also under different mark scenarios including spatial independence and function-valued marks, and positive or negative inter-dependencies between functions of type  $h$  and type  $l$ . For the case when we have positive interaction between functions, functions of type  $l$  grow or decrease when interacting with functions of type  $h$ , and vice versa, whilst for negative inter-dependencies, functions of type  $l$  grow or decrease when interacting with functions of type  $h$ , but functions of type  $h$  decrease or grow when interacting functions of type  $l$ .

##### 4.1 Generating Point Patterns With Function-Valued Marks

To control for the effect of the inherent point configuration on the proposed estimators, we considered three distinct point process configurations including Poisson, cluster and regular point process scenarios. Each of these three cases were generated on the unit torus to avoid edge effects with an expected number of points  $n = 200$ . To obtain a clustered point process structure, we simulated a Thomas process (Thomas, 1949) with offspring dispersion parameter  $\sigma = 0.04$ , parent intensity  $\lambda_p = 40$  and  $\mu = 5$  expected offsprings per parent, yielding an average number of points of around 200 in the unit square with a moderate clustered configuration. The regular point process scenario was constructed using a Strauss process (Kelly & Ripley, 1976; Strauss, 1975) with interaction effect parameter  $q = 0.05$ , and a radius of interaction  $R_{int} = 0.025$  which ensures strong inhibition effects for short scales of interaction, with an average number of points of around 200. Note that the Thomas and the Strauss processes have been considered to model and analyse clustered and aggregated point patterns, respectively. In particular, these models have been very suitable to model forest patterns. For instance, Asefa *et al.* (2020) considered a Thomas process to analyse the spatial distribution of tropical, subtropical and subalpine forests in southwest China, while Obiang *et al.* (2010) considered the Strauss process to analyse the strength of competition between trees in central African rainforests.

To generate spatial point patterns in which each point location is augmented by a set of function-valued quantities, we consider the continuous space–time stochastic process developed by Renshaw and Särkkä (2001). In this model, marked points located on the unit torus grow and interact with each other in terms of a suitable growth–interaction scheme. We adapt this approach to avoid point mortality and point immigration. In this way, we keep the same point pattern over time and their associated growth curves. Technically, the Renshaw and Särkkä algorithm generates a spatial point pattern with function-valued marks  $h$  and  $l$  through

$$\begin{aligned}
 f_h(\mathbf{x})(t + dt) &= f_h(\mathbf{x})(t) + \beta_h f_h(\mathbf{x})(t) (1 - f_h(\mathbf{x})(t) / S_h) dt \\
 &+ \sum_{\substack{\neq \\ (\mathbf{x}, f_h(\mathbf{x})), \\ (\mathbf{x}', f_l(\mathbf{x}')) \in \boldsymbol{\psi}}} J_h(f_h(\mathbf{x})(t), f_l(\mathbf{x}')(t); \|\mathbf{x} - \mathbf{x}'\|) dt
 \end{aligned} \tag{9}$$

and

$$\begin{aligned}
 f_l(\mathbf{x})(t + dt) &= f_l(\mathbf{x})(t) + \beta_l(1 - f_l(\mathbf{x})(t)/S_l)dt \\
 &+ \sum_{\substack{\neq \\ (\mathbf{x}, f_l(\mathbf{x})), \\ (\mathbf{x}', f_l(\mathbf{x}')) \in \psi}} J_l(f_l(\mathbf{x})(t), f_l(\mathbf{x}')(t); \|\mathbf{x} - \mathbf{x}'\|)dt
 \end{aligned}
 \tag{10}$$

where  $f_h(\mathbf{x})(t)$  and  $f_l(\mathbf{x})(t)$  are two functions of point  $\mathbf{x}$  at time  $t$ ,  $\psi$  is a realisation of  $\Psi$ ,  $\beta_h$  the intrinsic rate of growth,  $S_h$  the non-spatial carrying capacity,  $\|\mathbf{x}_1 - \mathbf{x}_2\|$  the Euclidean distance between a pair of points, and  $J_h(\cdot)$  a suitable interaction function between points. Note that functions of type  $h$  and  $l$  grow in terms of the classic logistic growth and the immigration-death process, respectively. These two simple growth functions ensure that both functions remain bounded.

To generate positive correlation between functions, we use a Strauss like symmetric interaction function (Renshaw & Särkkä, 2001) adapted to the case where the interaction is between the marks  $f_h(\mathbf{x})(t)$  and  $f_l(\mathbf{x})(t)$ ,

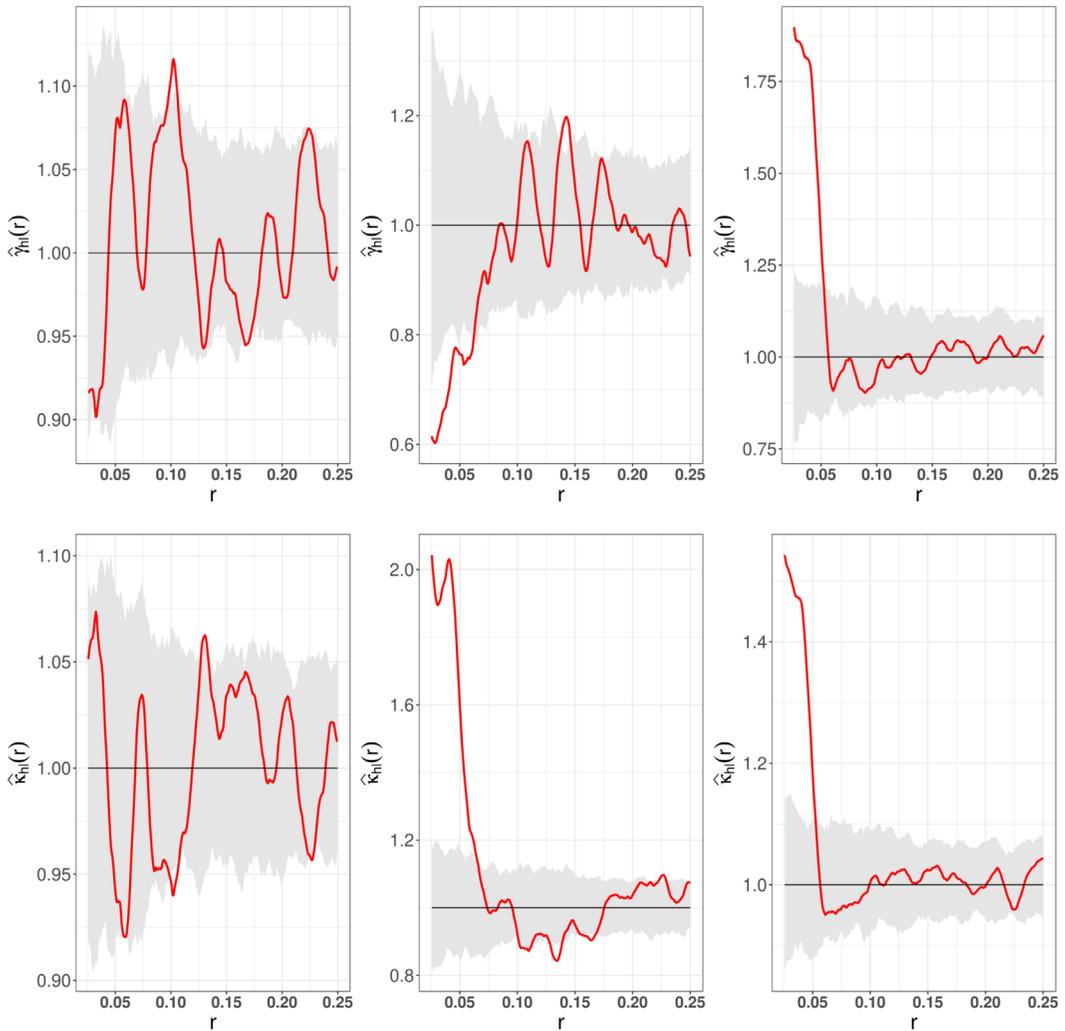
$$J(f_h(\mathbf{x})(t), f_l(\mathbf{x}')(t); \|\mathbf{x} - \mathbf{x}'\|) = \begin{cases} c & \text{if } \|\mathbf{x} - \mathbf{x}'\| < D \\ 0 & \text{otherwise,} \end{cases}
 \tag{11}$$

where  $c \in \mathbb{R}$  is a constant interaction effect. Here, points start to interact with each other with constant value  $c$  as soon as their distance is less than  $D$ . To ensure a symmetric interacting structure, we set  $J_h(\cdot) = J_l(\cdot) = J(\cdot)$ , so that smaller function values affect the growth of larger ones in the same way as larger function values affects smaller ones. To avoid interacting effects to decrease function values, we set  $c > 0$ ;  $c < 0$  implies function reduction and eventually negative function values.

Moreover, to generate negative correlation between functions, we take  $J_h(\cdot) = J(\cdot)$  and  $J_l(\cdot) = 0$ . Now functions of type  $h$  take an advantage when interacting with functions of type  $l$  (faster grow), whilst functions of type  $l$  are not affected by the interaction with functions of type  $h$ . This promotes an asymmetric function interaction, resulting in negative spatial correlation between growth functions of distinct type.

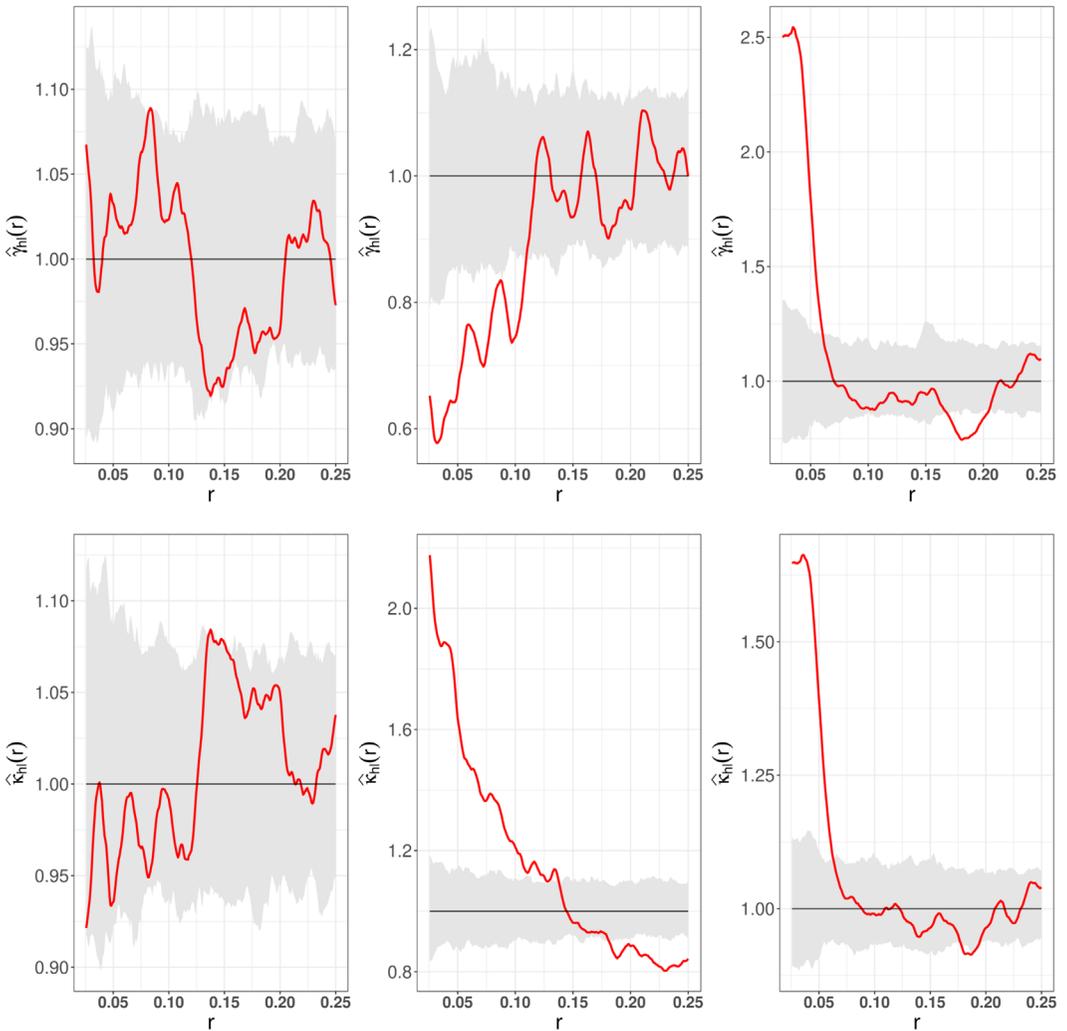
To generate spatial point patterns with function-valued marks, we consider expressions (9) and (10) with growth carrying capacity  $S_h = S_l = 5$ , intrinsic rates of growth  $\beta_h = 0.05$ ,  $\beta_l = 0.2$  and interaction distance  $D = 0.05$ . These scenario parameters are chosen as they give rise to functions that are convenient as illustrative examples. Moreover, to obtain the desired marked point patterns with spatially independent and/or positive correlation between function-valued marks, we consider the interaction mechanism (11) for  $J_h(\cdot) = J_l(\cdot) = J(\cdot)$ , with interaction parameter  $c = 0$  and  $c = 0.5$ , respectively. Whilst to generate negative spatial correlation between functions we assume the same interaction function (11), but for  $J_h(\cdot) = J(\cdot)$  and  $J_l(\cdot) = 0$  with  $c = 0.5$ .

Figure 1 shows the results for the homogeneous Poisson point process scenario with intensity  $\lambda = 200$ . The red lines are the empirical cross-function mark summary characteristics from a single simulation. The grey shading shows the fifth-largest and smallest envelope values based on 199 random simulations according to the null hypothesis of random labeling of functions over fixed point locations. Here, we consider three correlation function scenarios, namely, spatial independence between functions (left), positive (middle) and negative (right) correlation between functions. This highlights that in absence of interaction between functions, the resulting estimators of both the cross-function mark variogram (7) (top panels) and the cross-function mark correlation (8) (bottom panels) lie within the grey shading area, confirming the spatial independence between functions of type  $h$  and  $l$ . In direct contrast, when assuming spatial positive



**Figure 1.** Cross-function mark summary characteristics for a simulated homogeneous Poisson process on the unit torus with point intensity  $\lambda = 200$ . Cross-function mark variogram (top) and cross-function mark correlation (bottom) with no-interaction effects ( $J_h(\cdot) = J_l(\cdot) = J(\cdot)$ , with  $c = 0$ ) (left), positive inter-function interaction ( $J_h(\cdot) = J_l(\cdot) = J(\cdot)$ , with  $c = 0.5$ ) (central), and negative inter-function correlation ( $J_h(\cdot) = J(\cdot)$  and  $J_l(\cdot) = 0$  with  $c = 0.5$ ) (right). Empirical versions of both characteristics are highlighted in red, theoretical values in black. Grey shading shows the fifth-largest and smallest envelope values based on 199 random simulations according to the null hypothesis of random labeling of functions over fixed point locations.

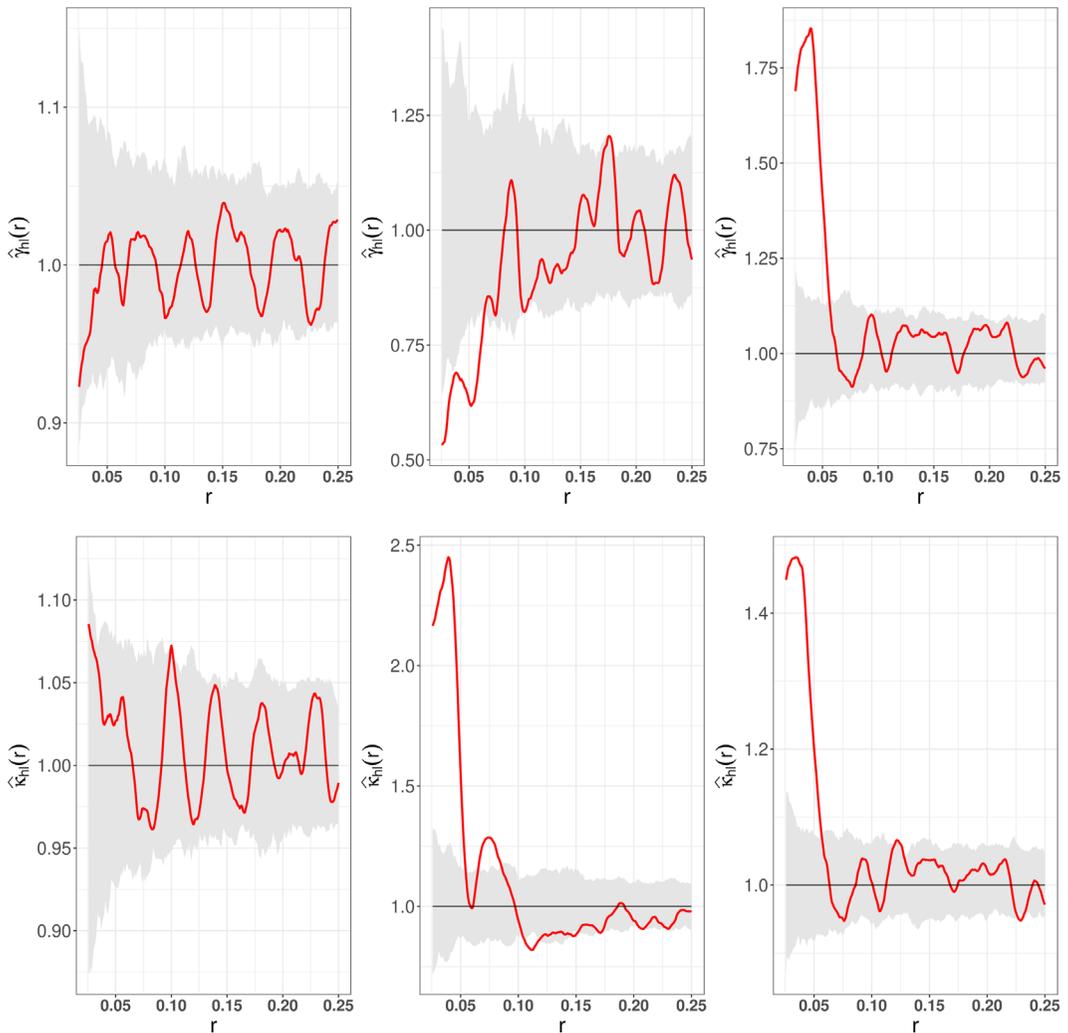
or negative interaction between curves (central and right panels, respectively), these estimators lie outside this grey shading area, confirming the presence of inter-function dependencies. In particular, under positive correlation of the function-valued marks, the empirical cross-function mark variogram lies outside this grey shading area with values smaller than the smallest envelope values, for small  $r$  values. This suggests that the positive interacting function-valued marks have less variability than under the independent mark setting. Similarly, under negative correlation between functions, estimators of both the cross-function mark variogram and the mark correlation lie outside the grey shading area with values larger than



**Figure 2.** Cross-function mark summary characteristics for a simulated Thomas process with offspring dispersion parameter  $\sigma = 0.04$ , parent intensity  $\lambda_p = 40$ , and  $\mu = 4$  expected offsprings per parent. Cross-function mark variogram (top) and cross-function mark correlation (bottom) with no-interaction effects ( $J_h(\cdot) = J_l(\cdot) = J(\cdot)$ , with  $c = 0$ ) (left), positive inter-function interaction ( $J_h(\cdot) = J_l(\cdot) = J(\cdot)$ , with  $c = 0.5$ ) (central), and negative inter-function correlation ( $J_h(\cdot) = J(\cdot)$  and  $J_l(\cdot) = 0$  with  $c = 0.5$ ) (right). Empirical versions of both characteristics are highlighted in red, theoretical values in black. Grey shading shows the fifth-largest and smallest envelope values based on 199 random simulations according to the null hypothesis of random labeling of functions over fixed point locations.

the largest envelope values, for small  $r$  values, suggesting negative interactions between functions.

Similar results can be found for the Thomas (Figure 2) and the Strauss process scenarios (Figure 3). In absence of inter-function dependencies both estimators (cross-function mark variogram and mark correlation) lie within the grey shading area, whilst for positive or negative correlation effects between functions these functions lie outside these envelopes. This confirms that the new cross-function mark summary characteristics can detect spatial dependencies between functions of distinct type independently of the spatial structure of the underlying point pattern.



**Figure 3.** Cross-function mark summary characteristics for a simulated Strauss process with interaction distance  $R_{int} = 0.05$  and interaction parameter  $q = 0.05$ . Cross-function mark variogram (top) and cross-function mark correlation (bottom) with no-interaction effects ( $J_h(\cdot) = J_l(\cdot) = J(\cdot)$ , with  $c = 0$ ) (left), positive inter-function interaction ( $J_h(\cdot) = J_l(\cdot) = J(\cdot)$ , with  $c = 0.5$ ) (central), and negative inter-function correlation ( $J_h(\cdot) = J(\cdot)$  and  $J_l(\cdot) = 0$  with  $c = 0.5$ ) (right). Empirical versions of both characteristics are highlighted in red, theoretical values in black. Grey shading shows the fifth-largest and smallest envelope values based on 199 random simulations according to the null hypothesis of random labeling of functions over fixed point locations.

For completeness, we consider a simulation study based on the simulation of 100 datasets for the above scenarios to illustrate the performance of our second-order characteristics. In particular, we simulated 100 realisations of the scenario defined above to show the percentage of times the resulting empirical function lies outside the simulated envelopes based on 199 random realisations according to the null hypothesis of random labeling of functions over fixed point locations. Table 1 shows the percentage of times the simulated empirical cross-function mark variogram and correlation lie outside the envelopes for the scenarios defined above. This highlights that under the scenarios with absence of interaction between functions, the percentage of times both empirical function lie outside the simulated envelopes is around 25% of the

Table 1. Percentage of times the simulated empirical cross-function mark variogram (Mark Var.) and correlation (Mark Cor.) lie outside the envelopes, based on the simulation of 100 datasets for the homogeneous Poisson process (Poisson) (as defined in Figure 1, the Thomas process (as defined in Figure 2) and the Strauss process (as defined in Figure 3), assuming no interaction effects, and, positive and negative inter-function interaction effects (as defined in Figure 1). A and B denote that the empirical function lies above or below the simulated envelopes, respectively.

	No interaction				Positive interaction				Negative interaction			
	Mark Var.		Mark Cor.		Mark Var.		Mark Cor.		Mark Var.		Mark Cor.	
	A	B	A	B	A	B	A	B	A	B	A	B
	Poisson	24	28	25	21	44	97	100	76	100	63	100
Thomas	28	25	30	24	49	78	100	66	100	60	100	58
Strauss	31	29	25	28	32	100	100	59	100	50	100	44

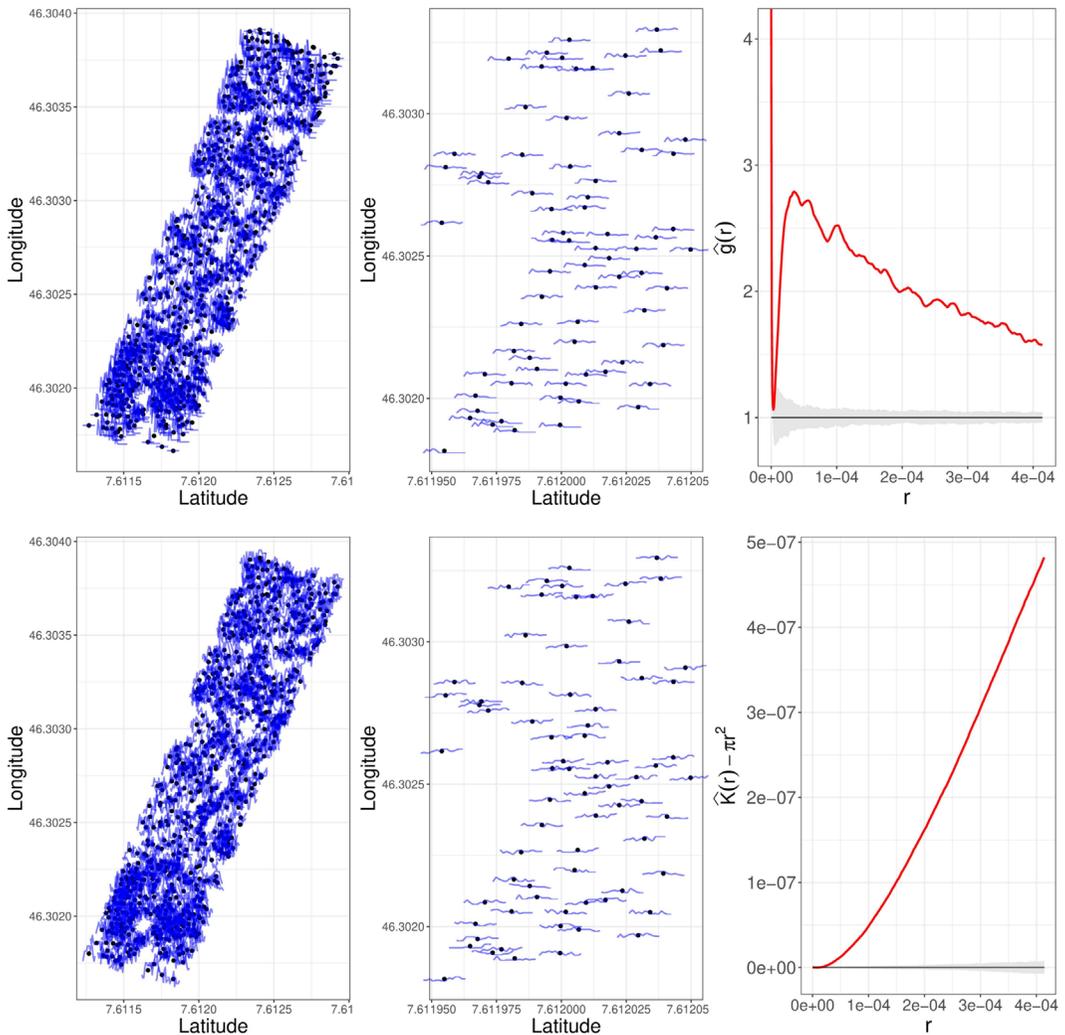
realisations for any of the initial point configurations. This suggests that the rejection of the null hypothesis is unlikely in absence of inter-function interactions. Note that these significant cases, where the empirical function lies outside the envelopes, correspond to random excursions of the empirical function outside the envelopes rather than significant trends of this function. In direct contrast, for the scenarios with positive/negative interaction effects between functions, the percentage of times both empirical functions lie outside the envelopes is always larger than 78% (assuming the empirical function lying above or below the envelopes) showing that our new summary characteristics detect the presence of these inter-function dependencies. For instance, for the scenarios with negative interaction effects, and any initial point configuration, empirical cross-function mark variogram and correlation functions lie 100% of the simulations above the envelopes, confirming the validity of our new approach to detect spatial dependencies between functions.

## 5 Applications

### 5.1 Application to Swiss Tree Data

We first consider tree measurements recorded at an annual basis over 14 years that originates from a long-term irrigation experiment located in *Pfynwald*, the central part of the *Pfyn-Finges* national park in Switzerland (Schaub *et al.*, 2016). Initiated in 2003, the experiment aimed to investigate the effect of increased water availability on the individual trees and the ecosystem in a naturally dry Scots pine (*Pinus sylvestris* L.) forest. The study region covers an area of 1.2 ha and is located in one of the driest inner-Alpine valleys of the European Alps (see Bose *et al.*, 2022, for detailed summary). The data at hand was provided as open data under an Open Database Licence and has been made available publicly at <https://opendata.swiss>. It covers the tree-specific spatial coordinates, the initial assignment into the treatment or control group and different tree characteristics for 900 trees. From this source, we initially selected the annual total crown defoliation (TCD) from the provided list of tree characteristics and also the exact point locations of the individual tree stands. The TCD parameter is a commonly used parameter in forest monitoring studies to quantify the loss of needles or leaves of a given tree relative to a local reference tree. Within the application, we considered the retrieved TCD information as function-valued tree attribute and assigned it as a mark to the tree locations in a subsequent action. Restricting the data to complete cases, we excluded any trees with incomplete or missing TCD information from the data yielding a final sample of 799 trees with annual TCD records over all 14 years. In a next step, we computed the local pairwise correlation function for all trees of the reduced sample which describes the contribution of the individual point to the empirical

pair correlation function, that is, its pair correlation function based local indicator of spatial association. The local information was then used as a second function-valued mark in our application such that each tree was marked by two distinct function-valued quantities. The resulting point pattern with both function-valued marks and classic second-order summary characteristics of the points are shown in Figure 4. While not considered here, we note that the data also allows for cross-function cross-type versions as outlined on Section 2.3.2 by taking additionally the tree-specific assignment into treatment or control group into account. Such advanced mark

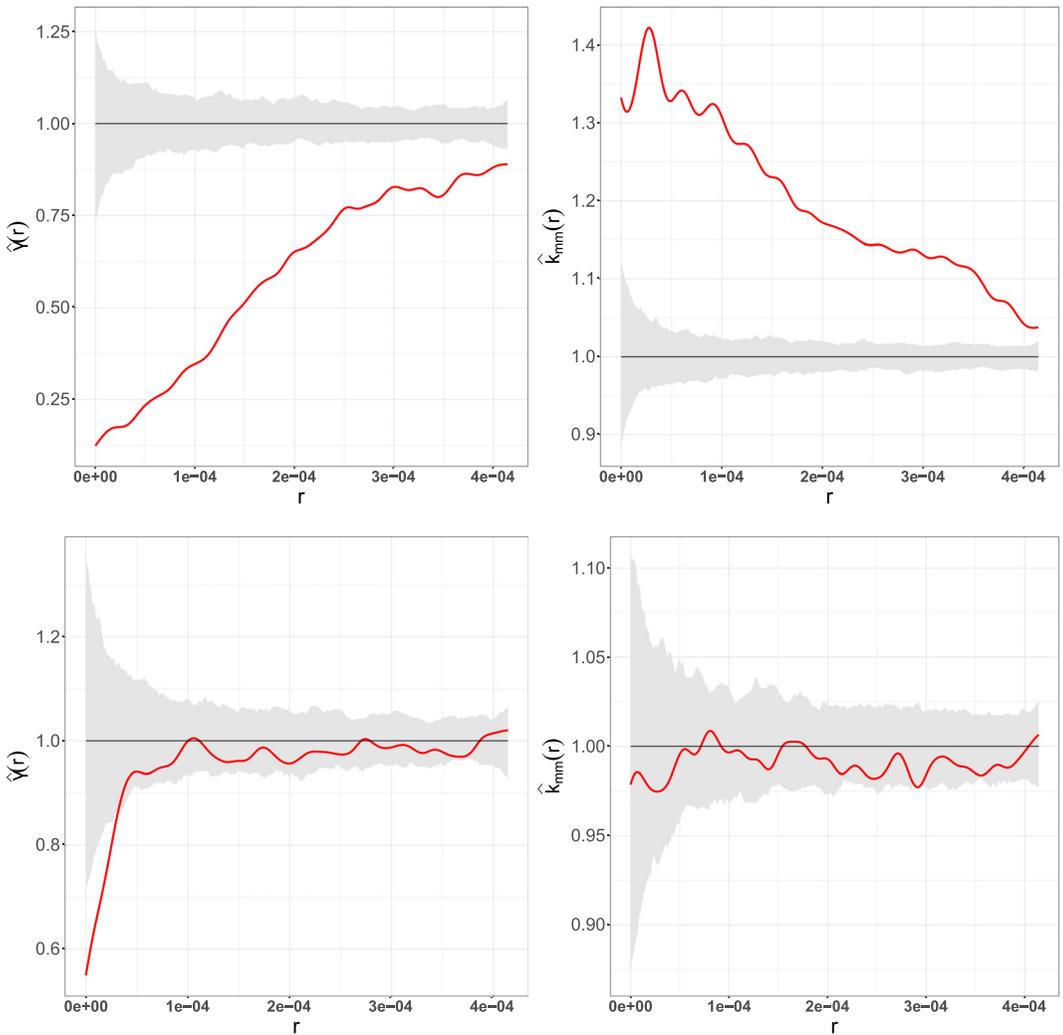


**Figure 4.** Observed function-valued marks and classic second-order point process summary characteristics of the Swiss tree patterns. Top panel: spatial distribution of Scots pines of the Pfywald data with observed total crown defoliation (left), a magnification of the study region (center), and pair correlation function and theoretical envelopes under the independent mark hypothesis (left). Bottom panel: spatial distribution of Scots pines of the Pfywald data with local pair correlation functions as function-valued marks (left), a magnification of the study region (center), and Ripley's  $K(r)$  function minus  $\pi r^2$  and theoretical envelopes (right) computed from the point locations. Empirical versions of both characteristics are highlighted in red, theoretical values in black, and blue lines are the function-valued marks. Grey shading shows the fifth-largest and smallest envelope values based on 199 random simulations according to the null hypothesis of complete spatial randomness (Poisson point randomisations).

characteristics might help to investigate the complex interplay of the TCD and local pair correlation function curves with the effect of additional water supply.

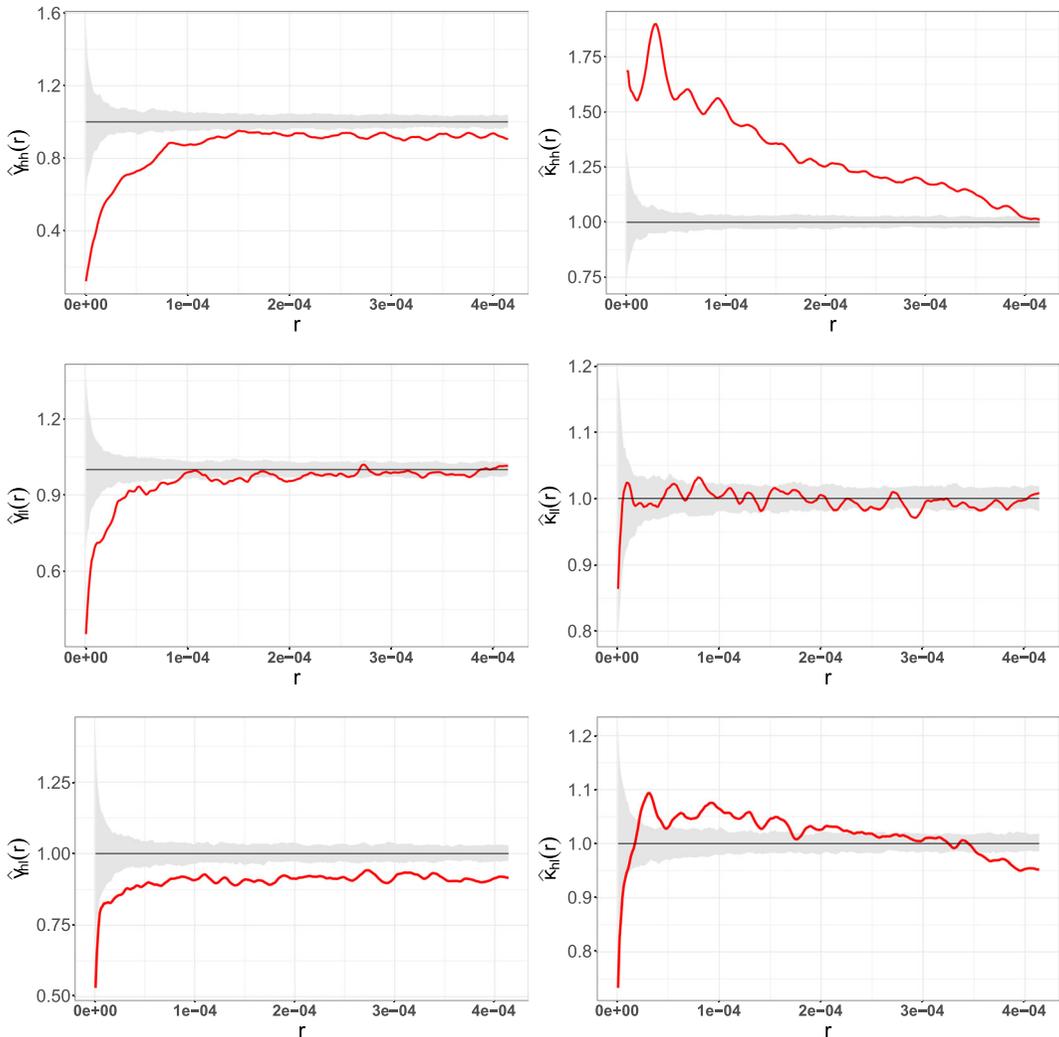
As expected by the large number of trees, the sampled point pattern reflects some clear structure and a tendency of clustering among the points. This impression is supported by the pair correlation function (top-right panel) and also Ripley's  $K$  function (bottom-right panel) which show a clear positive shift of the empirical curves from the theoretical lines under the complete spatial randomness hypothesis which indicates a clear tendency of clustering.

Next, to evaluate the findings of the proposed auto- and cross-function summary characteristics with the classic summary characteristics for scalar-valued marks commonly used at present, we transformed the function-valued marks into function-wise averages and computed the mark variogram and Stoyan's mark correlation function from the averaged quantities (see Figure 5).



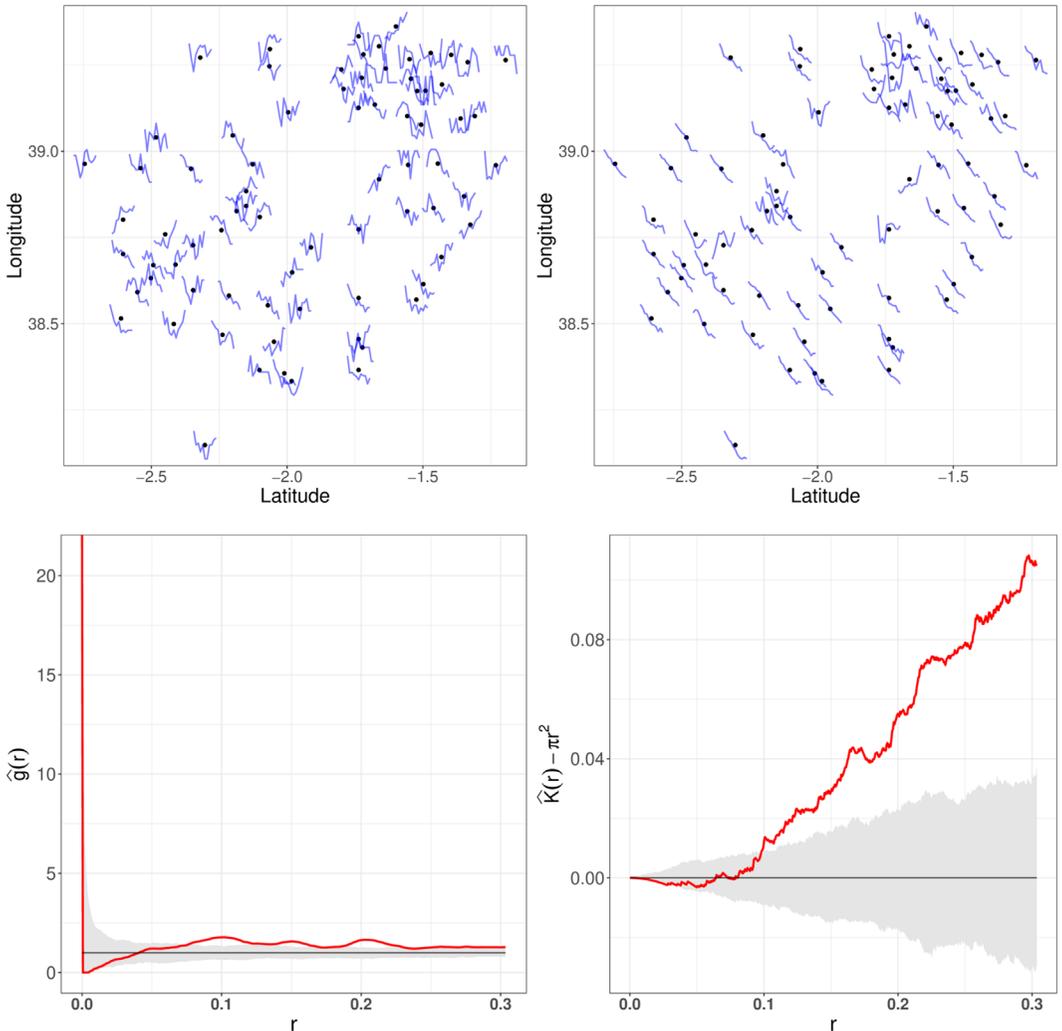
**Figure 5.** Classic mark summary characteristics for the Pfynwald tree data with averaged function-valued point attributes treated as scalar-valued marks. Mark variogram and mark correlation functions for the mean TCD (top) and mean local pair correlation function (bottom). Empirical versions of both characteristics are highlighted in red, theoretical values in black. Grey shading shows the fifth-largest and smallest envelope values based on 199 random simulations according to the null hypothesis of random labeling of marks over fixed point locations.

The empirical versions of both mark characteristics show a clear deviation from the theoretical envelopes for the average TCD (top panel). While the mark variogram (top-left panel) suggests that the mean TCD values exhibit less pairwise variation as expected under the independent mark hypothesis, we found a clear positive shift of the empirical pairwise product of TCD averages as considered by the mark correlation function (top-right panel) from the theoretical envelopes. In comparison with the TCD, both empirical mark characteristics show almost no deviations from the independent mark hypothesis in case of the averaged local pair correlation function (bottom panels). Except for only some negative shift of the mark variogram (left) at small distances, both estimated characteristics are covered by the envelopes.



**Figure 6.** Auto- and cross-function mark summary characteristics computed from the Swiss tree data. Top: auto-function mark variogram (left) and mark correlation function (right) of the total crown defoliation curves. Central panel: auto-function mark variogram (left) and mark correlation function (right) of the local pair correlation functions. Bottom: cross-function (left) and mark correlation function (right) between the total crown defoliation and the local pair correlation functions. Empirical versions of both characteristics are highlighted in red, theoretical values in black. Grey shading shows the fifth-largest and smallest envelope values based on 199 random simulations according to the null hypothesis of random labeling of functions over fixed point locations.

Different from the classic mark characteristics, all auto- and cross-function mark variograms and correlation functions of Figure 6, except the auto-function mark correlation of the local pair correlation (central-right panel), show significant results. As already indicated by the classic characteristics, the top panel corresponding to the TCD curves reflects again a negative deviation of the empirical auto-function mark variogram (top-leftpanel) contrasted with a clear positive shift of the empirical auto-function mark correlation function (top-right panel) from the theoretical lines under the independent mark hypothesis. This indicates that the observed TCD curves show less spatial variation among pairs of neighbouring points. At the same time,

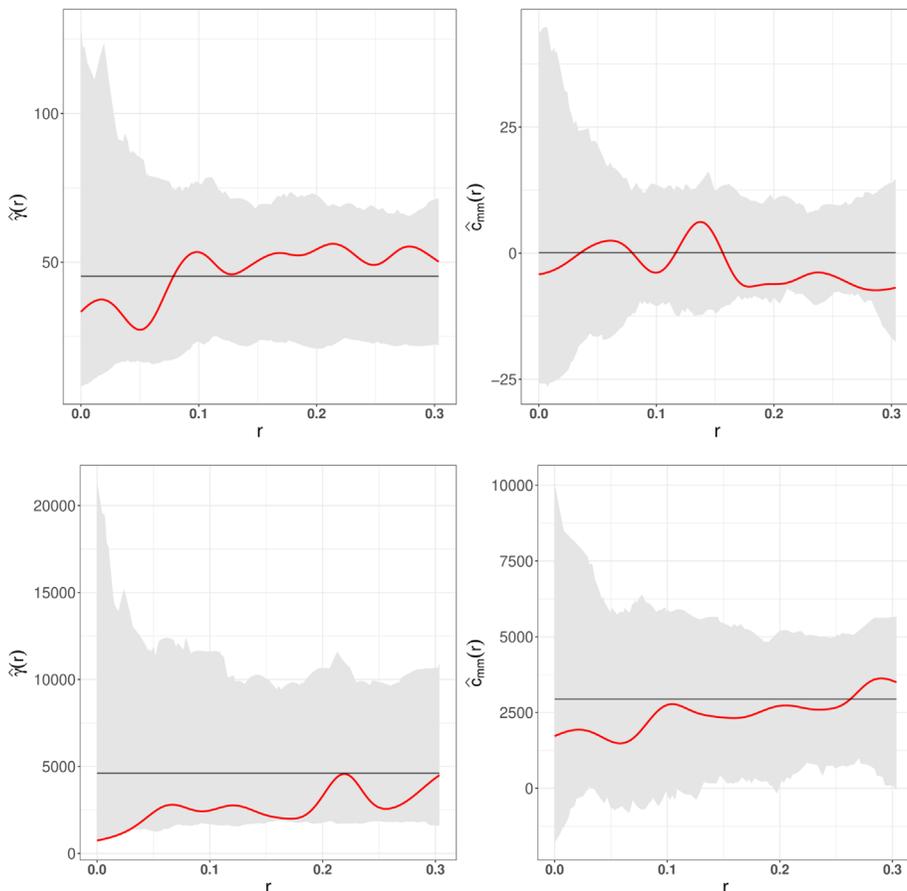


**Figure 7.** Observed function-valued marks and classic second-order point process summary characteristics computed from 84 municipalities of the province of Albacete. Top panel: spatial distribution of Spanish municipalities and the yearly differences to the reference year 2022 for the business (top-left) and population (top-right) records as function-valued marks. Bottom panel: pair correlation function and theoretical envelopes under the independent mark hypothesis (left), and Ripley's  $K(r)$  function minus  $\pi r^2$  and theoretical envelopes (right) computed from the point locations. Empirical versions of both characteristics are highlighted in red, theoretical values in black, and blue lines are the function-valued marks. Grey shading shows the fifth-largest and smallest envelope values based on 199 random simulations according to the null hypothesis of complete spatial randomness (Poisson point randomisations).

the product of the TCD curves clearly exceeds the expected case, that is, the non-spatial functional mean squared. For the central panels showing the auto-function characteristics computed from the local pair correlation functions, the auto-function mark variogram (left) again suggests smaller variation between the function-valued marks compared with the independent mark setting for some small distances. Finally, looking at the cross-function characteristics of the TCD and local pair correlation curves, both results show a clear variation from the independent mark envelopes. This would imply that the pairwise spatial variation of both functions is smaller than under the limiting case where the cross-function variogram is equal to the covariance, whereas the pairwise product of the two marks exceeds the limiting case in which the pairwise product of the two marks approaches the product of the functional means  $\mu_h$  and  $\mu_l$ .

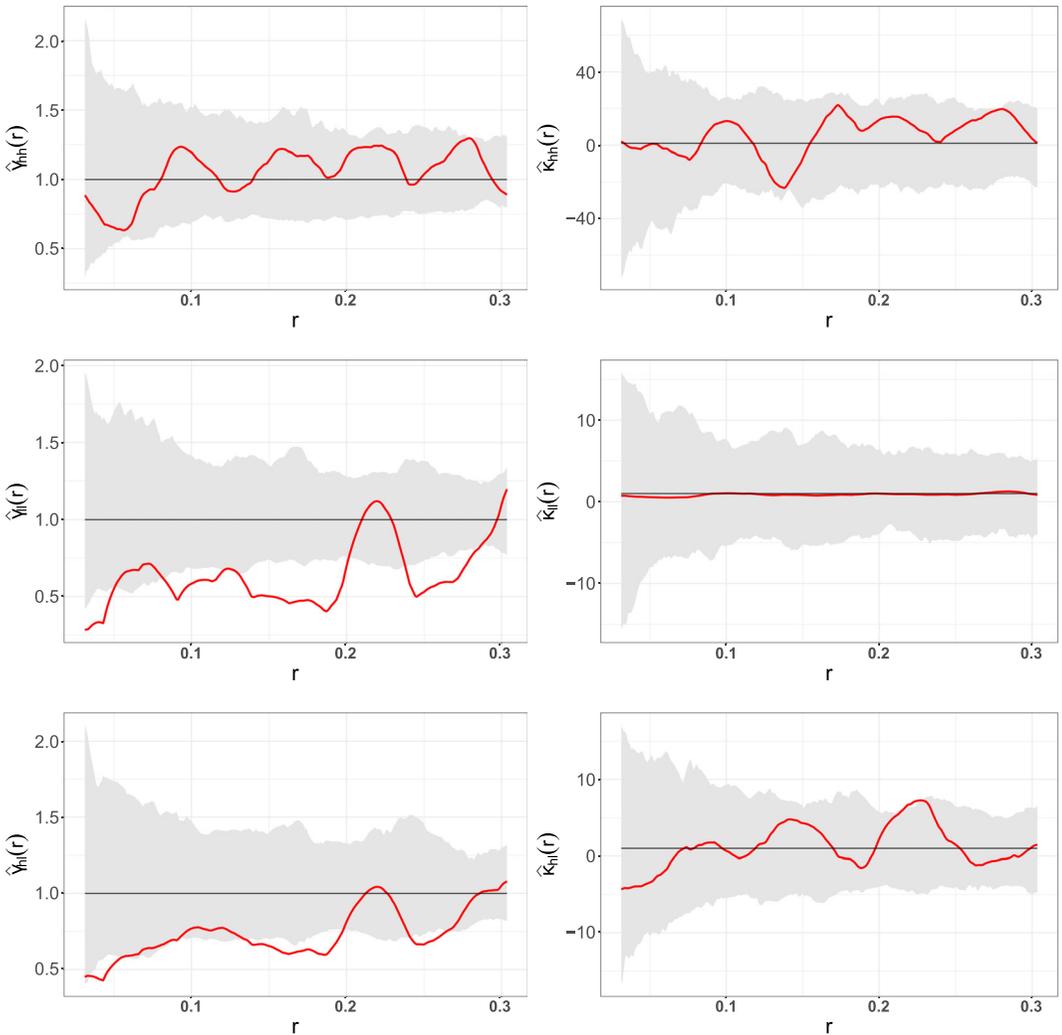
## 5.2 Application to Spanish Labour Data

As second example of a spatial point pattern with bivariate function-valued marks, we consider data on the total number of companies as of 1 January and the number of residents



**Figure 8.** Classic mark summary characteristics for the Spanish municipality data computed from the averaged business and population information with averaged function-valued point attributes treated as scalar-valued marks. Mark variogram and conditional mean product of marks for the mean business (top) and mean population function (bottom). Empirical versions of both characteristics are highlighted in red, theoretical values in black. Grey shading shows the fifth-largest and smallest envelope values based on 199 random simulations according to the null hypothesis of random labeling of marks over fixed point locations.

recorded annually at municipality level for the period from 2012 to 2022. The data originated from the official data reports released by the National Statistics Institute of Spain (INE) and was made publicly available at [www.ine.es](http://www.ine.es). The business information was derived from the official business register of INE and corresponds to the total number of local companies over different economic sectors. The local assignment of the companies to exactly one municipality was performed by INE in a pre-processing step using the registered business address information to avoid potential inconsistencies in case of regionally wide spreading business locations, for example, factories or business facilities of one company in several distinct municipalities. From the provided data, we initially selected a sample of 87 municipalities that fall into the boundaries of *Albacete*, a Spanish province on *La Mancha* (the Spanish Plateau). The area of *La Mancha* is



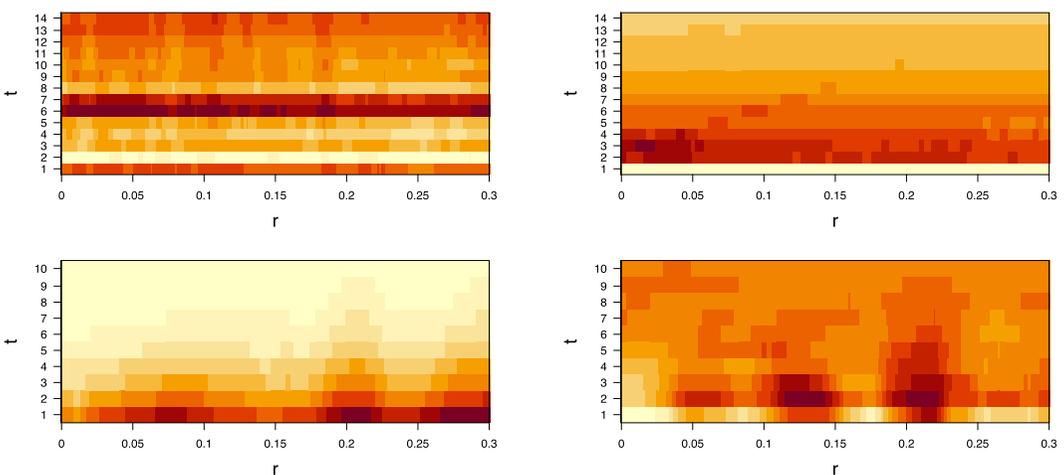
**Figure 9.** Auto- and cross-function mark summary characteristics computed from the Spanish municipality data. Top panel: auto-function mark variogram (left) and mark correlation function (right) of the business curves. Central panel: auto-function mark variogram (left) and mark correlation function (right) of the population curves. Bottom panel: cross-function (left) and mark correlation function (right) between the business and population curves. Empirical versions of both characteristics are highlighted in red, theoretical values in black. Grey shading shows the fifth-largest and smallest envelope values based on 199 random simulations according to the null hypothesis of random labeling of functions over fixed point locations.

located southeast of Madrid and is characterised by a homogeneous climate and population density, and as such, has been treated as a particular instance of a homogeneous spatial point process in the literature (see, e.g. Glass & Tobler, 1971; Ripley, 1977, 1988, Chiu *et al.*, 2013, for some applications). From the selected file we excluded the municipalities of *Masegoso*, *Montalvos* and *Villa de Ves* in a subsequent action for which no economic information was available. This yielded a final sample of 84 Spanish municipalities to which we applied the following pre-processing. In a first step, we derived the exact spatial location of the centroids for each municipality in the sample and assigned the corresponding pair of coordinates to the data. Next, we generated two function-valued attributes from the provided local business and population statistics by computing the pointwise yearly differences between the values from 2012 to 2021 and the reference records of 2022. As such, both generated marks express the annual change in the size (resp. number) of the local business sector (resp. population) with respect to 2022. All information was then transformed into a spatial point process with function-valued marks in a final step. The generated point pattern and classic second-order point process summary characteristics of the points are shown in Figure 7.

Different from the Swiss tree data example, the Spanish point pattern appears to be less dense and clustered with both constructed business (7, top-left panel) and population (7, top-right panel) showing some heterogeneity. Both the empirical pair correlation function (bottom-left panel) and Ripley's  $K$  function minus  $r\pi^2$  indicate a clear tendency to clustering for the point locations, which supports the visual impression.

As for Section 5.1, we computed the means from the business and population curves and used the scalar information as input for classic mark summary characteristics (see Figure 8). Due to the presence of negative mark values, the unnormalised version, that is, the conditional mean product of marks  $c_{mm}(r)$ , was computed instead of the mark correlation function  $k_{mm}(r)$ . For the averaged business variation (top panels) and the mean variation of the population (bottom panels) both the mark variogram and mark correlation function are completely covered by the envelopes, supporting the independent mark hypothesis.

Comparing these findings with the results of the auto- and cross-function mark characteristics depicted in Figure 9, the independence hypothesis is not supported by the auto-function mark



**Figure 10.** Cross-function mark variogram (left) and mark correlation (right) summary characteristics computed from the Swiss tree data (top) and the Spanish municipality data (bottom) adapted to show time and inter-point distances: dark and white colours correspond to large and small values of these functions, respectively.

variogram for population (central-left panel) and the cross-function mark variogram of business and population (bottom-left). Both functions would instead suggest less pairwise variation between the generated populations differences resp. business and population differences as expected under the independent mark assumption. As both mark summary characteristics for the change in size of the business sectors (top panels) are included in the envelopes, the significant results of the cross-function mark variogram seem to be driven mostly by the pairwise variation of the population curves.

For completeness, Figure 10 shows the cross-function mark summary characteristics for the two analysed datasets adapted to show time and inter-point distances. This modification of the original functions permits to observe the resulting spatial structure for a given inter-point distance and time interval of these functional objects. This figure highlights that, as expected, the resulting spatial structure depends on the time interval, and the spatial configuration of points affects the resulting interaction of these functions for each time interval. In general, resulting values of these functions are quite similar between time intervals suggesting a temporal dependency of these spatial structures. This is expected because the values that form each function are dependent.

## 6 Conclusions

This paper proposes an immense variety of different mark summary characteristics which allow to decide on potential structure of the marks within highly challenging spatial point process scenarios. Including cross-function, multi-function and corresponding mark weighted versions of well-established mark summary characteristics, the extended methods are providing a suitable statistical toolbox for the analysis of spatially aligned function-valued quantities for a plethora of potential applications. Formalised through generalisations of classical test functions to the (multivariate) function-valued mark scenario, the proposed characteristics are well embedded into the statistical literature and methodology for spatial point patterns with real-valued marks and allow for similar interpretations.

The considered estimators are natural extensions to the complex function-valued cases of those proposed in Comas *et al.* (2008, 2011, 2013) and are technically supported by the theoretical treatments in Ghorbani *et al.* (2021). We rely on this latter contribution to support the behaviour of the proposed estimators. However, there are a number of doors open in both, theoretical and inferential aspects, that starting from our developments can go further in providing tools for complex mark structures. One such example we can think of is the case when we have trajectories restricted to a network-based topology. Here more topological arguments are needed to be considered when developing further tools and their estimators. While we have discussed extensions to multivariate functions, it would be intriguing to further expand our concept by offering additional technical insights into the construction and estimation of general test functions involving more than two arguments. Specifically, comparing classical test functions with two arguments to more intricate versions, as delineated in Section 2.3, could yield valuable and thought-provoking results for advancing the development of such multi-argument test functions. In addition to the spatial domain formulation, it would be intriguing to derive appropriate frequency domain characteristics that consider the structural interdependence of both the points and the function-valued marks (see, e.g. Eckardt & Mateu, 2019a, 2019b, for frequency domain approaches to spatial point processes with scalar-valued marks). Lastly, beyond point processes in planar regions, there is a requirement for appropriate mark characteristics for point processes occurring in structured domains featuring function-valued marks.

## Acknowledgments

The authors gratefully acknowledge financial support through the German Research Association and the Spanish Ministry of Science. Matthias Eckardt was funded by the Walter Benjamin grant 467634837 from the Deutsche Forschungsgemeinschaft. Jorge Mateu was partially funded by grant PID2019-107392RB-I00, from the Spanish Ministry of Science. Carles Comas was partially funded by grant PID2020-115442RB-I00 from MCIN/AEI/10.13039/501100011033 (Spanish Ministry of Science). Open access funding enabled and organized by Projekt DEAL.

## Data Availability Statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

## References

- Asefa, M., Wen, H.-D., Brown, C., Cao, M., Xu, K. & Hu, Y.-H. (2020). Ecological drivers of tree assemblage in tropical, subtropical and subalpine forests. *J. Veg. Sci.*, **31**, 107–117. DOI [10.1111/jvs.12819](https://doi.org/10.1111/jvs.12819).
- Aw, A. & Cabral, E.N. (2020). Bayesian estimation of the functional spatial lag model. *J. Time Ser. Econom.*, **12**, 20190047.
- Beisbart, C. & Kerscher, M. (2000). Luminosity- and morphology-dependent clustering of galaxies. *Astrophys. J.*, **545**, 6–25.
- Bohorquez, M., Giraldo, R. & Mateu, J. (2017). Multivariate functional random fields: Prediction and optimal sampling. *Stoch. Environ. Res. Risk A.*, **31**, 53–70.
- Bohorquez, M., Giraldo, R. & Mateu, J. (2022). Spatial prediction and optimal sampling for multivariate functional random fields. In *Geostatistical Functional Data Analysis*, Eds. J. Mateu & R. Giraldo, chap. 13, pp. 329–349. John Wiley & Sons, Ltd.
- Bose, A.K., Rigling, A., Gessler, A., Hagedorn, F., Brunner, I., Feichtinger, L., Bigler, C., Egli, S., Etzold, S., Gossner, M.M., Guidi, C., Lévesque, M., Meusburger, K., Peter, M., Saurer, M., Scherrer, D., Schleppi, P., Schönbeck, L., Vogel, M.E., ... Schaub, M. (2022). Lessons learned from a long-term irrigation experiment in a dry Scots pine forest: Impacts on traits and functioning. *Ecol. Monogr.*, **92**, e1507.
- Brix, A. & Chadœuf, J. (2002). Spatio-temporal modelling of weeds by shot-noise G Cox processes. *Biom. J.*, **44**, 83–99.
- Capobianco, R. & Renshaw, E. (1998). The autocovariance function for marked point processes: A comparison between two different approaches. *Biom. J.*, **40**, 431–446.
- Chiu, S.N., Stoyan, D., Kendall, W.S. & Mecke, J. (2013). *Stochastic Geometry and Its Applications*, 3rd ed. Chichester: John Wiley & Sons.
- Comas, C., Delicado, P. & Mateu, J. (2008). Analysing spatial point patterns with associated functional data. In *Statistics for Spatio-Temporal Modelling, Proceedings of the 4th International Workshop on Spatio-Temporal Modelling (METMA-4)*, pp. 157–163.
- Comas, C., Delicado, P. & Mateu, J. (2011). A second order approach to analyse spatial point patterns with functional marks. *TEST*, **20**, 503–523.
- Comas, C., Mehtätalo, L. & Miina, J. (2013). Analysing space–time tree interdependencies based on individual tree growth functions. *Stoch. Environ. Res. Risk A.*, **27**, 1673–1681.
- Cressie, N. (1993). *Statistics for Spatial Data*. Wiley.
- Daley, D. & Vere-Jones, D. (2008). *An Introduction to the Theory of Point Processes: General Theory and Structure*. Berlin-Heidelberg: Springer.
- Delicado, P., Giraldo, R., Comas, C. & Mateu, J. (2010). Statistics for spatial functional data: Some recent contributions. *Environmetrics*, **21**, 224–239.
- Eckardt, M. & Mateu, J. (2019a). Analysing multivariate spatial point processes with continuous marks: A graphical modelling approach. *Int. Stat. Rev.*, **87**, 44–67.
- Eckardt, M. & Mateu, J. (2019b). Partial characteristics for marked spatial point processes. *Environmetrics*, **30**, e2565.
- Ferraty, F. & Vieu, P. (2006). *Nonparametric Functional Data Analysis: Theory and Practice*. New York: Springer.
- Franco-Villoria, M. & Ignaccolo, R. (2022). Universal, residual, and external drift functional kriging. In *Geostatistical Functional Data Analysis*, Eds. J. Mateu & R. Giraldo, chap. 3, pp. 55–72. John Wiley & Sons, Ltd.

- 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*.
- Giraldo, R., Delicado, P. & Mateu, J. (2011). Ordinary kriging for function-valued spatial data. *Environ. Ecol. Stat.*, **18**, 411–426.
- Glass, L. & Tobler, W.R. (1971). General: Uniform distribution of objects in a homogeneous field: Cities on a plain. *Nature*, **233**, 67–68.
- González, J.A., Rodríguez-Cortés, F.J., Cronie, O. & Mateu, J. (2016). Spatio-temporal point process statistics: A review. *Spat. Stat.*, **18**, 505–544.
- Horváth, L. & Kokoszka, P. (2012). *Inference for Functional Data With Applications*. New York: Springer.
- Hsing, T. & Eubank, R. (2015). *Theoretical Foundations of Functional Data Analysis, With an Introduction to Linear Operators*. Chichester, West Sussex, United Kingdom: Wiley.
- Hui, G. & Pommerening, A. (2014). Analysing tree species and size diversity patterns in multi-species uneven-aged forests of Northern China. *Forest Ecol. Manag.*, **316**, 125–138.
- Hui, G.Y., Albert, M. & Gadow, K.v. (1998). Das Umgebungsmaß als Parameter zur Nachbildung von Bestandesstrukturen. *Forstw. Cbl.*, **117**, 258–266.
- Iftimi, A., Cronie, O. & Montes, F. (2019). Second-order analysis of marked inhomogeneous spatiotemporal point processes: Applications to earthquake data. *Scand. J. Stat.*, **46**, 661–685.
- Illian, J., Penttinen, A., Stoyan, H. & Stoyan, H. (2008). *Statistical Analysis and Modelling of Spatial Point Patterns*. Chichester: John Wiley & Sons.
- Isham, V. (1985). Marked point processes and their correlations. In *Spatial Processes and Spatial Time Series Analysis*, pp. 63–75. Publications des Facultes Universitaires Saint-Louis Brussels.
- Kelly, F.P. & Ripley, B.D. (1976). A note on Strauss's model for clustering. *Biometrika*, **63**, 357–360.
- Martínez-Hernández, I. & Genton, M.G. (2020). Recent developments in complex and spatially correlated functional data. *Braz. J. Probab. Stat.*, **34**, 204–229.
- Mateu, J. & Romano, E. (2017). Advances in spatial functional statistics. *Stoch. Environ. Res. Risk. A.*, **31**, 1–6.
- Møller, J. & Díaz-Avalos, C. (2010). Structured spatio-temporal shot-noise Cox point process models, with a view to modelling forest fires. *Scand. J. Stat.*, **37**, 2–25.
- Møller, J. & Waagepetersen, R. (2004). *Statistical Inference and Simulation for Spatial Point Processes*. Boca Raton: Chapman & Hall.
- Nerini, D., Manté, C. & Monestiez, P. (2022). Extending geostatistical kriging when data are multivariate curves: Some technical considerations and operational solutions. In *Geostatistical Functional Data Analysis*, Eds. J. Mateu & R. Giraldo, chap. 4, pp. 73–103. John Wiley & Sons, Ltd.
- Obiang, N.L.E., Ngomanda, A., Mboma, R., Nzabi, T., Ngoye, A., Atsima, L., Ndjélé, L., Mate, J.-P., Lomba, C. & Picard, N. (2010). Spatial pattern of central African rainforests can be predicted from average tree size. *Oikos*, **119** (10), 1643–1653. DOI [10.1111/j.1600-0706.2010.18440.x](https://doi.org/10.1111/j.1600-0706.2010.18440.x).
- Penttinen, A. & Stoyan, D. (1989). Statistical analysis for a class of line segment processes. *Scand. J. Stat.*, **16**, 153–168.
- Penttinen, A., Stoyan, D. & Henttonen, H.M. (1992). Marked point processes in forest statistics. *Forest Sci.*, **38**, 806–824.
- Pineda-Ríos, W., Giraldo, R. & Porcu, E. (2019). Functional SAR models: With application to spatial econometrics. *Spat. Stat.*, **29**, 145–159.
- Pommerening, A., Gonçalves, A.C. & Rodríguez-Soalleiro, R. (2011). Species mingling and diameter differentiation as second-order characteristics. *Allg. Forst. Jagdztg.*, **182**, 115–129.
- Ramsay, J. & Silverman, B. (1997). *Functional Data Analysis*. Springer.
- Rathbun, S.L. (1993). Modeling marked spatio-temporal point patterns. *Bull. Int. Stat. Inst.*, **55**, 379–396.
- Redenbach, C. & Särkkä, A. (2013). Parameter estimation for growth interaction processes using spatio-temporal information. *Comput. Statist. Data Anal.*, **57**, 672–683.
- Renshaw, E. (2002). Two-dimensional spectral analysis for marked point processes. *Biom. J.*, **44**, 718–745.
- Renshaw, E., Comas, C. & Mateu, J. (2009). Analysis of forest thinning strategies through the development of space–time growth–interaction simulation models. *Stoch. Environ. Res. Risk. A.*, **23**, 275–288.
- Renshaw, E. & Särkkä, A. (2001). Gibbs point processes for studying the development of spatial-temporal stochastic processes. *Comput. Statist. Data Anal.*, **36**, 85–105.
- Ripley, B.D. (1977). Modelling spatial patterns. *J. R. Stat. Soc. Ser. B. Stat. Methodol.*, **39**, 172–212.
- Ripley, B.D. (1988). *Statistical Inference for Spatial Processes*. Cambridge University Press.
- Särkkä, A. & Renshaw, E. (2006). The analysis of marked point patterns evolving through space and time. *Comput. Statist. Data Anal.*, **51**, 1698–1718.
- Schaub, M., M. Haeni, C. Hug, A. Gessler, and A. Rigling (2016): “Tree measurements 2002–2016 from the long-term irrigation experiment Pfywald, Switzerland.”

- Schlather, M. (2001). On the second-order characteristics of marked point processes. *Bernoulli*, **7**, 99–117.
- Serra, L., Saez, M., Mateu, J., Varga, D., Juan, P., Díaz-Ávalos, C. & Rue, H. (2014). Spatio-temporal log-Gaussian Cox processes for modelling wildfire occurrence: The case of Catalonia, 1994–2008. *Environ. Ecol. Stat.*, **21**, 531–563.
- Siino, M., Adelfio, G. & Mateu, J. (2018). Joint second-order parameter estimation for spatio-temporal log-Gaussian Cox processes. *Stoch. Environ. Res. Risk. A.*, **32**, 3525–3539.
- Stoyan, D. (1984). Correlations of the marks of marked point processes - Statistical inference and simple models. *J. Inf. Process. Cybern.*, **20**, 285–294.
- Stoyan, D. (1987). Statistical analysis of spatial point processes: A soft-core model and cross-correlations of marks. *Biom. J.*, **29**, 971–980.
- Stoyan, D., Rodríguez-Cortés, F.J., Mateu, J. & Gille, W. (2017). Mark variograms for spatio-temporal point processes. *Spat. Stat.*, **20**, 125–147.
- Stoyan, D. & Stoyan, H. (1994). *Fractals, Random Shapes, and Point Fields : Methods of Geometrical Statistics*. Chichester, New York: Wiley.
- Stoyan, D. & Wälder, O. (2000). On variograms in point process statistics, II: Models for markings and ecological interpretation. *Biom. J.*, **42**, 171–187.
- Strauss, D.J. (1975). A model for clustering. *Biometrika*, **62**, 467–475.
- Thomas, M. (1949). A generalization of Poisson's binomial limit for use in ecology. *Biometrika*, **36**, 18–25.
- Vere-Jones, D. (2009). Some models and procedures for space-time point processes. *Environ. Ecol. Stat.*, **16**, 173–195.
- Wälder, O. & Stoyan, D. (1996). On variograms in point process statistics. *Biom. J.*, **38**, 895–905.
- Zhang, L., Baladandayuthapani, V., Zhu, H., Baggerly, K.A., Majewski, T., Czerniak, B.A. & Morris, J.S. (2016). Functional CAR models for large spatially correlated functional datasets. *J. Amer. Statist. Assoc.*, **111**, 772–786.

[Received June 2023; accepted May 2024]