Application of Hybrid Multivariate Functional Principal Component Analysis for the Analysis of Multivariate Spatial Point Process Summary Characteristics

Master's thesis

for acquiring the degree of
Master of Science (M.Sc.)

in Statistics

at the School of Business and Economics
of Humboldt-Universität zu Berlin

submitted by

Bianca Neubert

Student no. 604466

Examiners: Prof. Dr. Sonja Greven and Dr. Matthias Eckardt

Berlin, 23 December 2021

Abstract

Due to an increasing amount of multivariate spatial point process data with both a large number of points and component processes, there is a great demand for appropriate methods to deal with them. Most existing approaches concentrate on the case of two or three component processes, but here is a shortage of methods analyzing a large number of component processes simultaneously. Thus, the aim of this thesis is to structure multivariate spatial point processes based on the spatial behavior of component processes. The proposed approach extends the idea of using principal component analysis methods on either numerical or functional summary characteristics to the combination of both. To include both types, hybrid multivariate functional principal component analysis is introduced for the simultaneous analysis of multivariate functional and vector data. Estimation methods are discussed. The proposed approach is evaluated in a simulation study and subsequently applied to the Duke forest data set to gain further understanding of the spatial behavior of tree species.

Contents

Index of Notations

List of Abbreviations

PC, PC1	Principal component, first principal component
PCA	Principal component analysis
FPCA	Functional principal component analysis
MFPCA	Multivariate functional principal component analysis
CSR	Complete spatial randomness
HFV-PCA	Principal component analysis of hybrid functional and vector data

List of Formulae and Symbols

Spaces and σ-algebras

$\mathbb{A}$, $\mathcal{A}$	Space of locally finite simple sequences, σ-algebra of locally finite point figuration sets
$\mathbb{N}$, $\mathbb{N}_0$	Natural numbers and natural numbers including zero
$\mathbb{R}^d$, $\mathcal{B}(\mathbb{R}^d)$	d-dimensional Euclidean space and corresponding Borel-σ-algebra
$\mathbb{X}$, $\mathcal{B}(\mathbb{X})$	Point space and corresponding Borel-σ-algebra
$(\Omega, \mathcal{F}, \mathbb{P})$	Probability space
$\mathcal{B}(\mathbb{X}) \otimes \mathcal{A}$	Product Borel-σ-algebra
$\mathbb{M}$, $\mathcal{B}(\mathbb{M})$	Mark space and corresponding Borel-σ-algebra
$\mathbb{S}$	Hilbert space
$\mathcal{T}$	Cartesian product of compact domains $\mathcal{T}_i \subset \mathbb{R}^{d_i}, i = 1, \ldots, q$
$L^2(\mathcal{T}, \mathcal{B}(\mathcal{T}), \nu)$	Space of L^2-integrable functions on domain $\mathcal{T}$ with respect to its Borel-σ-algebra and the corresponding Lebesgue measure
$\mathbb{F}$	Multivariate function space, $\mathbb{F} = L^2(\mathcal{T}_1) \times \ldots L^2(\mathcal{T}_q)$
$\mathbb{H}$	Space of hybrid data
$S([0, p])$	Space of step functions on $[0, p]$

Operators

$\lvert B \rvert$	Cardinality of a set B
$\times$	Cartesian product
$\otimes$	Tensor product on the corresponding space
$(\cdot)^{\mathrm{T}}$	Transpose of a vector or a matrix
$\approx$	Approximately equal to
$:=$	Defined as
$\sum^{\neq}$	Sum over tuples with distinct entries
$\lVert \cdot \rVert_{\mathbb{R}^d}$	Euclidean norm on $\mathbb{R}^d$
$\langle \cdot, \cdot \rangle_{\mathbb{R}^d}$	Euclidean scalar product on $\mathbb{R}^d$
$\lVert \cdot \rVert_{\mathbb{S}}$	Norm on Hilbert space $\mathbb{S}$
$\langle \cdot, \cdot \rangle_{\mathbb{S}}$	Scalar product on Hilbert space $\mathbb{S}$
$\lVert \cdot \rVert_{\mathbb{F}}$	L^2-norm on function space $\mathbb{F}$
$\langle \cdot, \cdot \rangle_{\mathbb{F}}$	L^2- scalar product on function space $\mathbb{F}$
$\lVert \cdot \rVert_{L^2(\mathcal{T}_r)}$	Norm on $L^2(\mathcal{T}_r)$
$\langle \cdot, \cdot \rangle_{L^2(\mathcal{T}_r)}$	Scalar product on $L^2(\mathcal{T}_r)$

$\|\cdot\|_{\mathbb{H}}$	Norm on hybrid space $\mathbb{H}$
$\langle\cdot,\cdot\rangle_{\mathbb{H}}$	Scalar product on hybrid space $\mathbb{H}$

Functions and Objects

$\mathbb{E}$	Expectation function
$\mathbb{E}_0^!$	Expectation with respect to the Palm distribution
$\mathbb{V}\mathrm{ar}$	Variance of a random variable
$\mathbb{C}\mathrm{ov}$	Covariance of two random variables
δ_{ij}	Kronecker delta
N	Spatial point process
$\mathbf{N}$	Multivariate spatial point process, $\mathbf{N} = (N_1, \ldots, N_n)$
φ, $\boldsymbol{\varphi}$, φ_i	Point patterns, that is, locally finite and simple sequence, element of $\mathbb{A}$, realizations of $N, \mathbf{N}, N_i$
$\mathbb{P}_N$	Distribution of a point process
$N(B)$	Number of points of a point process N in Borel set B
N_x	Point process translated by $x \in \mathbb{X}$
$\mathbf{r}N$	Point process rotated by $\mathbf{r}$
x, x_i	Elements of the point space $\mathbb{X}$
ω	Element of the sample space Ω
A	Configuration set, element of $\mathcal{A}$
B	Borel-set of the Euclidean space, element of $\mathcal{B}(\mathbb{R}^d)$
B_x	Borel-set translated by $x \in \mathbb{X}$
A_x	Configuration set consisting of the sequences translated by $x \in \mathbb{X}$
W	Observation window
$\hat{\lambda}$, $\widehat{\mathrm{D}}$, $\widehat{\mathrm{K}}$	Estimator of the corresponding functions
$b(x, r)$	Sphere around x with radius r
ν, ν_d	Lebesgue measure on $\mathcal{T}$ and $\mathbb{R}^d$
Λ, λ	Intensity measure and intensity function
$\mu^{(k)}$	k-th moment measure
$\alpha^{(k)}$, $\varrho^{(k)}$	k-th factorial measure and its density with respect to the Lebesgue measure
$\mathscr{C}, \mathscr{C}^!$	Campbell measure and reduced Campbell measure
$\mathcal{P}_0$, P_0, $\mathrm{P}_0^!$	Radon-Nikodym derivative of the Campbell measure, Palm distribution and reduced Palm distribution
$\mathbb{1}_A(x)$	Indicator function, equal to one if $x \in A$ and else zero
D	Nearest-neighbor distance distribution function
K, K′	Ripley's K-function and its first derivative
L	Besag's L-function
CEI	Clark-Evans index
$\overline{R}$	Mean directional index
SK	Ratio of mean and median of the nearest-neighbor distribution
g	Pair correlation function
$\hat{g}(r)$	Estimated pair correlation function at value r
$r^{(1)}$	Minimal r value for which $\hat{g}(r) = 1$
$n_i(r)$	Number of points within distance r from x_i of some point pattern
$\bar{n}(r)$	Average number of points within distance r of a point pattern
$\mathscr{K}$	Covariance operator

$\mathcal{K}$	Covariance function/kernel
μ	Mean function of a process
b_d	Factor for the volume of the d-dimensional sphere
V	Random vector on $\mathbb{R}^p$, $V = (V^{(1)}, \ldots, V^{(p)})^{\mathrm{T}}$
v, v_i	Element of $\mathbb{R}^p$, i-th realization of V, i.e. $v_i = (v_i^{(1)}, \ldots, v_i^{(p)})^{\mathrm{T}}$
C_V	Covariance matrix of random vector V
w, w_j	PC vectors for standard PCA
β_j	Eigenvalues
J, M, M_r, L	Truncation lags for numerical, (multivariate) functional and hybrid PCA methods
$V^{[J]}$	Truncated vector
TVE	Total variance explained
ε	Variance explained by the first J PCs
γ, γ_j	Vector PC scores
$\hat{\mu}_V$, $\widehat{C}_V$	Sample mean and covariance matrix for vector V
$\hat{w}_j$, $\hat{\beta}_j$, $\hat{\gamma}_{ij}$	Estimated vector PCs, eigenvalues and PC scores for the j-th PC and observation i
S, μ_S	Random element in Hilbert space $\mathbb{S}$ and its mean
s_j	Eigenobjects of a random element S of Hilbert space $\mathbb{S}$
F	(Multivariate) functional object, $F = (F^{(1)}, \ldots, F^{(q)})^{\mathrm{T}}$
t	Elements of the domain space, $t = (t_1, \ldots, t_q)^{\mathrm{T}} \in \mathcal{T}$
f, f_i	Elements of the function space, realizations of random elements in $\mathbb{F}$, $f_i = (f_i^{(1)}, \ldots, f_i^{(q)})^{\mathrm{T}} \in \mathbb{F}$
e, e_j	Eigenfunctions of random elements of $\mathbb{F}$
Z, Z_j, Z_{ij}	Functional PC scores
$F^{[M]}$	Truncated Karhunen-Loève decomposition of random F
$\hat{\mu}$, $\widehat{\mathcal{K}}$, $\hat{e}_j$, $\hat{Z}_j$	Estimators of the mean function, the covariance kernel, the functional PC and PC score, respectively
H, μ_H	Random object in $\mathbb{H}$ and its mean, $H = (F, V)^{\mathrm{T}}$
$H^{[L]}$	Truncated Karhunen-Loève expansion of random hybrid object H
$\widetilde{H}$	Hybrid object consisting of truncated functional and truncates vector PC decomposition, $\widetilde{H} = (F^{[M]}, V^{[J]})^{\mathrm{T}}$
h, h_i	Realizations of H, $h_i = (f_i, v_i)^{\mathrm{T}} = (f_i^{(1)}, \ldots, f_i^{(p)}, v^{(1)}, \ldots, v^{(p)})^{\mathrm{T}}$
ξ, ξ_j	Hybrid eigenobjects, hybrid PCs
θ_j, ψ_j	Functional and vector part of a hybrid eigenobject
ρ_j, ρ_{ij}	Hybrid PC score
κ	Weighting constant for the hybrid scalar product
d	Dimension of the point space
d_i	Dimension of domain $\mathcal{T}_i$
p	Dimension of vectors
q	Dimension of multivariate functional objects
n	Number of component processes
$\phi_i^{(r)}$	i-th basis function for component r of a multivariate function
a_i, a	Concatenated vector of coefficients from pre-smoothing of the functional data and the vector data
C_A	Covariance matrix of the vector a resulting from pre-smoothing

$\mathcal{K}_F$, $\mathcal{K}_{FV}$, $\mathcal{K}_V$	Covariance kernels corresponding to the combination of different data modalities
$\widetilde{\mathscr{K}}$	Covariance operator of $\widetilde{H}$
$\widetilde{\mathcal{K}}_F$, $\widetilde{\mathcal{K}}_{FV}$, $\widetilde{\mathcal{K}}_V$	Covariance kernels corresponding to the combination of different data modalities for the truncated PC decompositions
$\widetilde{H}^{[L]}$	Truncation of the hybrid object based on the truncated functional and vector PC decompositions
$\tilde{\beta}_l$, $\tilde{\xi}_l$, $\tilde{\rho}_l$	Eigenvalues, eigenobjects and scores corresponding to $\widetilde{\mathscr{K}}$, $\tilde{\xi}_l = (\tilde{\psi}_l, \tilde{\theta}_l)^{\mathrm{T}}$
C_F, C_V	Covariance matrices of the functional and vector score vectors (separate data modalities)
C_{FV}, C_{VF}	Covariance matrices of vector and functional score vectors (mixed data modalities)
C, $\widehat{C}$	Matrix containing the covariance matrices of score vectors as blocks and corresponding sample covariance matrix
$b_l = (c_l^{\mathrm{T}}, d_l^{\mathrm{T}})^{\mathrm{T}}$	Eigenvectors of matrix score vector covariance matrix C, c_l for the functional part and d_l for the vector part
$\hat{b}_l = (\hat{c}_l^{\mathrm{T}}, \hat{d}_l^{\mathrm{T}})^{\mathrm{T}}$	Corresponding estimators
$\hat{\chi}_i$	Concatenated vector of estimated vector and functional PC scores
$\hat{\psi}_l$	Estimated values of the functional part of the l-th hybrid PC
$\hat{\theta}_l$	Estimated values of the numerical part of the l-th hybrid PC
$\hat{\rho}_{il}$	Estimated l-th hybrid PC score for observation i
f^V, f^v, f_i^v	Step function corresponding to V and to realizations v, v_i
$\mathcal{K}_{f^V}$	Covariance kernel of f^V

List of Figures

List of Tables

1 Introduction

Studying random sets of points in some space, called spatial point processes, has become increasingly important. Spatial point process data sets become more and more available, containing on the one hand data for a large number of processes and on the other hand a lot of points for each process. In turn, the general aim is to reduce the dimension of the data to be able to recognize patterns between the processes, for example by grouping them with respect to their spatial behavior.

Existing methods mostly concentrate on the pairwise comparison of point processes, which gets tedious very fast when the amount of observed patterns is high. To analyze the different processes simultaneously, there are existing approaches based on the multivariate or functional analysis of spatial point process characteristics. Using numerical or functional characteristics separately, these approaches use only parts of the information given in a sample. Consequently, the motivation of this thesis is to consider a joint analysis of functional and numerical spatial point process summary characteristics through multivariate functional principal component analysis (MFPCA) extended to include numerical features as well.

In applications of point process statistics the simultaneous analysis of multiple point processes of a similar nature or origin gains more and more in importance. Often the patterns are observed on the same or congruent windows. This motivates the consideration of multivariate point processes. For example, in the field of ecosystem diversity and functioning it is of interest to study patterns from ecological communities to promote and sustain biodiversity (Illian et al. (2006)). Samples contain the locations of individuals from different plant species in the same observation area. More specifically, the data set considered in this thesis contains the locations of different tree species in the Duke forest in Durham, Orange Alamance counties in North Carolina (USA).

As another example, in material sciences one might deal with a number of point patterns derived from different, potentially unknown, materials. Samples contain some spatial information of objects that are of different materials but of the same form. That is, the observed windows are congruent (Illian et al. (2008)). Further examples include the analysis of galaxies in the universe or occurrences of diseases in certain regions.

In most applications the number of processes as well as the number of points in each observed point pattern may be extremely large. Understanding these high dimensional objects is difficult and, hence, grouping the processes by their spatial behavior simplifies the analysis. That is, one may relate the grouping structure to characteristics that form the point pattern (Illian et al. (2008)).

For instance, grouping species by their spatial behavior potentially reveals which species benefit from one another and which distribution of species increases biodiversity. Analogously, grouping the patterns of different material samples by their spatial behavior may reveal that the patterns originate from similar material structure groups.

The status quo for analyzing multiple point patterns is to consider bivariate versions of summary characteristics such as the cross-K-function of Ripley's K. In this way, only two patterns are compared and analyzed at once. When the multivariate point process has more than one subprocess, one needs to analyze the components pairwise. This can get quite cumbersome when there are many subprocesses.

As an alternative, Illian et al. (2008) proposes two approaches for grouping point patterns by their spatial behavior and, consequently, analyzing many patterns simultaneously. The first is based on numerical summary characteristics. The idea is to use classical multivariate analysis considering the patterns as objects and the summary characteristics as variables. The second approach is based

on functional summary characteristics. Here, functional principal component analysis is applied to one function for each pattern. Patterns are then grouped with respect to the first few principal components. Naturally, therefrom arises the question if one could combine both methods.

Happ and Greven (2018) introduce multivariate functional principal component analysis for data observed on different domains. The concept can be further extended to data combining functional data and a vector using the scalar product proposed by Ramsay and Silverman (1997). This kind of data is also called hybrid data.

The aim of this master's thesis is to present the concept of hybrid functional principal component analysis and to combine the two approaches by Illian et al. (2008) for the analysis of multivariate point patterns.

The thesis is structured as follows. In Chapter 2, the theory of multivariate point processes is presented. Focus lies on the different summary characteristics, numerical and functional. Based on this, the approach proposed by Illian et al. (2008) can be formulated more precisely.

In Chapter 3, we derive the concept of hybrid functional principal component analysis. For this, we review the concept of standard principal component analysis for vector data. Generalizing the concept for Hilbert spaces, principal component analysis for (multivariate) functional data and, finally, hybrid data can be derived. Furthermore, a detailed presentation of three estimation schemes for hybrid multivariate principal component analysis is given.

This theoretical basis enables to extend the approach of Illian et al. (2008). The extension is evaluated in a simulation study in Chapter 4.

In Chapter 5, we apply the derived method to forestry data collected in a part of the Duke forest. That is, hybrid multivariate functional principal component analysis is carried out on numerical and functional summary characteristics of the different tree species in this forest. In this way, tree species can be grouped by their spatial behavior and the analysis yields insights into the mechanisms of tree allocation in the forest.

The thesis is concluded with a discussion and an outlook on further extensions in Chapter 6.

2 General Theory of Point Processes

To provide context for the methodology, we introduce in this chapter point processes in general and the extension to multivariate point processes and review important results.

Heuristically speaking, point processes model the random allocation of points in some space or area. There is extensive literature on point processes, both application and theory oriented including Chiu et al. (2013), Daley and Vere-Jones (1998), Møller and Waagepetersen (2004), Stoyan and Stoyan (1994) and Illian et al. (2008). In this chapter notations and concepts are compiled citations from these references unless otherwise stated.

2.1 Basic Properties of Point Processes

Before defining a point process, the point space of interest is shortly discussed. One can define point processes on general spaces, e.g. in Daley and Vere-Jones (1998) they are introduced on complete separable metric spaces. For clarity, we restrict the definition of point processes on subspaces $\mathbb{X}$ of the d-dimensional Euclidean space $\mathbb{R}^d$, for some natural number $d \in \mathbb{N}$. Most applications consider the two or three dimensional Euclidean space. For $d = 2$ this means considering points on a plane. Denote by $\mathcal{B}(\mathbb{R}^d)$ and $\mathcal{B}(\mathbb{X})$ the corresponding Borel-σ-algebras.

There are two ways to define point processes. Firstly, a point process N can be defined as a random counting measure on $\mathcal{B}(\mathbb{X})$. That is, a realization of N is a function taking sets $B \in \mathcal{B}(\mathbb{X})$ as arguments and returning a number in the natural numbers including zero $\mathbb{N}_0 := \mathbb{N} \cup \{0\}$. This number is interpreted as the number of points the realization of N has in set B. See for example Daley and Vere-Jones (1998) for an introduction to this approach.

The second approach defines point processes as a random set of points. For a set B let $|B|$ denote its cardinality. More precisely, we define point processes in this thesis adopting the definition of Chiu et al. (2013, p. 108-109).

A *point process* N on $\mathbb{X} \subset \mathbb{R}^d$ is a random variable taking values in a measurable space $(\mathbb{A}, \mathcal{A})$. Here, let $\mathbb{A}$ be the family of all sequences $\varphi = \{x_i\}_{i \in \mathbb{N}} \subset \mathbb{X}$ satisfying the two regularity conditions that

(i) the sequence φ is locally finite, that is, each bounded set $B \in \mathcal{B}(\mathbb{X})$ must contain only a finite number of points in φ, $|B \cap \varphi| < \infty$,

(ii) the sequence is simple, i.e. for two different elements $x_i, x_j \in \varphi$ it holds $x_i \neq x_j$ if $i \neq j$.

$\mathcal{A}$ denotes the σ-algebra defined as the smallest σ-algebra on $\mathbb{A}$ such that the mappings $\varphi \mapsto \varphi(B) := |B \cap \varphi|$ are measurable. The elements of $\mathcal{A}$ are called *locally finite point configuration sets*, in short, *configuration sets*.

Consequently, a point process is defined as a measurable mapping from some probability space into the above specified space of sequences $N \colon (\Omega, \mathcal{F}, \mathbb{P}) \longrightarrow (\mathbb{A}, \mathcal{A})$, that is, N is a random element with realization $\varphi \in \mathbb{A}$, which is also called *point pattern*. An element x_i of the sequence φ is a *point* in the point pattern. Elements of the point space $\mathbb{X}$ are often referred to as *locations*.

Sometimes the simplicity assumption (ii) is not included in the definition but defined as a separate property of point processes. In most applications it is an appropriate assumption that

there cannot occur a case of two points at exactly the same location. Therefore, we include it in the definition here.

Analogously to the distribution of random variables the distribution of a point process N is defined as the measure $\mathbb{P}_N$ on $(\mathbb{A}, \mathcal{A})$ given by the following relation

$$\begin{aligned}\mathbb{P}_N(A) &= \mathbb{P}(N \in A) \\ &= \mathbb{P}(\{\omega \in \Omega \colon N(\omega) \in A\}), \quad A \in \mathcal{A}.\end{aligned}$$

That is, the measure assigns each configuration set $A \in \mathcal{A}$ the probability that the point process N is a sequence $\varphi = \{x_i\}_{i \in \mathbb{N}}$ in this set. Note that $N(\omega)$ denotes the evaluation of the mapping N from the probability space to $(\mathbb{A}, \mathcal{A})$. Therefore, $N(\omega)$ represents one realization of the point process. In contrast to that, when writing $N(B)$, one refers to the random variable $N(B) \colon (\Omega, \mathcal{F}, \mathbb{P}) \longrightarrow \mathbb{N}_0, \omega \mapsto N(\omega)(B) := |B \cap N(\omega)|$. In other words, it gives the random number of points of the point process located in set B.

The remaining part of this chapter deals with finding alternative ways of describing the distribution of a point process.

An important point configuration set for given $B \in \mathcal{B}(\mathbb{X})$ is the set $\{\varphi \colon \varphi(B) = 0\}$, i.e. the set of sequences that have no point in B. The corresponding values of the distribution of N are called *void probabilities*, that is,

$$\begin{aligned}\mathbb{P}_N(\{\varphi \colon \varphi(B) = 0\}) &= \mathbb{P}(N \in \{\varphi \colon \varphi(B) = 0\}) \\ &= \mathbb{P}(N(B) = 0).\end{aligned}$$

One can show that the void probabilities for all compact sets $B \in \mathcal{B}(\mathbb{X})$ uniquely determine the distribution of a point process; for a proof see, e.g. Møller and Waagepetersen (2004), Theorem B.1. This illustrates the importance of the random variables $N(B)$ for a compact Borel set B. In particular, it is of interest to consider their mean and higher moments, which is the topic of the next paragraphs.

Considering the mean of $N(B)$ as a function in B results in a measure on $\mathcal{B}(\mathbb{X})$ which yields a characteristic of point processes analogous to the mean of random variables. That is, for $B \in \mathcal{B}(\mathbb{X})$ the measure Λ is defined as

$$\Lambda(B) \coloneqq \mathbb{E}[N(B)]$$

and gives the mean number of points in a given set. Throughout this thesis let $\mathbb{E}$ denote the expectation with respect to corresponding distributions. The measure Λ is called the *intensity measure* of N. If the intensity measure is absolutely continuous with respect to the Lebesgue-measure ν_d on $\mathbb{R}^d$, the density function or, in other words, the Radon-Nikodym derivative λ is called the *intensity function*. This can be written as

$$\Lambda(B) = \int_B \lambda(x) \, d\nu_d(x).$$

Analogously, higher-order moments of the random variables $N(B)$ generate the so called moment measure of a point process N. For more details on the derivation we refer to Daley and Vere-Jones (2008). Following Chiu et al. (2013) the *k-th moment measure* $\mu^{(k)}$ can be defined as the measure

taking k elements of $\mathcal{B}(\mathbb{X})$ as arguments and satisfying that

$$\int_{\mathbb{R}^{kd}} f(x_1,\dots,x_k)d\mu^{(k)}(x_1,\dots,x_k) = \mathbb{E}\left(\sum_{x_1,\dots,x_k\in N} f(x_1,\dots,x_k)\right),$$

for any non-negative measurable function f on $\mathbb{R}^{kd}$. This yields for sets $B_1,\dots,B_k \in \mathcal{B}(\mathbb{X})$ that

$$\mu^{(k)}(B_1\times\ldots\times B_k) = \mathbb{E}[N(B_1)\cdot\ldots\cdot N(B_k)].$$

Therefore, choosing $B_1 = \ldots = B_k = B$, for random variable $N(B)$ the k-th moment is given by $\mathbb{E}[N(B)^k] = \mu^{(k)}(B^k)$. For $k = 1$ it follows that the moment measure is equivalent to the intensity measure, since $\mu^{(1)}(B) = \mathbb{E}[N(B)] = \Lambda(B)$. Also, the variance of $N(B)$ can be written with moment measures. More precisely, it holds that

$$\mathbb{V}\mathrm{ar}(N(B)) = \mu^{(2)}(B\times B) - \Lambda(B)^2.$$

Additionally, for the analysis of point processes, the *factorial moment measure* $\alpha^{(k)}$ is of interest. It is similarly defined as the measure satisfying

$$\int_{\mathbb{R}^{kd}} f(x_1,\dots,x_k)d\alpha^{(k)}(x_1,\dots,x_k) = \mathbb{E}\left(\sum_{x_1,\dots,x_k\in N}^{\neq} f(x_1,\dots,x_k)\right),$$

again for any non-negative measurable functions f. Here, the unequal sign $\neq$ above the summation sign indicates that the sum is taken over all distinct k-tuples in the point process N. This is exactly the difference between the moment measure and factorial moment measure. This also means that if the sets $B_1,\dots,B_k$ are disjoint, i.e. all points are distinct, the two measures are equal, i.e.

$$\mu^{(k)}(B_1\times\ldots\times B_k) = \alpha^{(k)}(B_1\times\ldots\times B_k).$$

Consequently, for $k = 1$ the measures coincide. For $k = 2$, the relationship can be written down explicitly. More precisely, it holds that for two sets $B_1, B_2 \in \mathcal{B}(\mathbb{X})$

$$\mu^{(2)}(B_1\times B_2) = \Lambda(B_1\cap B_2) + \alpha^{(2)}(B_1\times B_2).$$

Roughly speaking, the measures of the points that are in both Borel-sets are accounted for by adding their intensity measure.

If the measure $\alpha^{(k)}$ is absolutely continuous with respect to the Lebesgue-measure on $\mathbb{R}^{kd}$, the resulting density $\varrho^{(k)}$ of the factorial moment measure is called *k-th order product density.*

In practice, random objects can only be analyzed based on realizations. Thus, it makes sense to shortly discuss how realizations of point processes are structured and which additional assumptions might be appropriate.

Often, it is only possible to have one observation of a point process, that is one point pattern. For example, if the point process of interest is the process leading to the allocation of trees of a certain forest at a specific point of time, obviously, it will not be possible to observe other realizations of this process. In this example the sample is truly unique due to the nature of the phenomenon, but there are also examples where collecting more than one sample is too complicated.

Also, the area for which the process can be observed is often restricted due to practical reasons.

For instance, the whole forest might be too large to measure the location of each tree so that only a small part of it is observed. This region is called *observation window* and denoted by $W \subset \mathbb{X}$. Since many applications are in the two dimensional space and it is easier to visualize, the word *area* is often used in the following to describe a subset of the point space $\mathbb{X}$.

Though one can only observe a finite set of points in a restricted area, in some applications it makes sense to assume that the observed point pattern is part of a much larger pattern. In this larger pattern, the points are then assumed to be distributed according to the same law as in the observation window in which one also assumes some kind of homogeneity of the distribution.

Considering the forest example again, then, on the one hand the growth conditions for the trees might to some extent be the same for all trees within the observation window. On the other hand the observation window is somewhat arbitrary and has no biological meaning. This means, that trees behave the same whether they are close to the edge of the window or closer to the center. Also, a reasonable assumption can be that the trees interact in the same way everywhere and, thus, one could extend the observation window in all directions without observing a different behavior of the points. As these assumptions are very strong, they should be discussed before every spatial point process analysis.

To formalize the idea of homogeneity, we consider the definition of a stationary and isotropic point process. A point process N is call *stationary* if its distribution is translation invariant. Namely, for the point process $N = \{x_i\}_{i=1}^{\infty}$ it holds that for any $x \in \mathbb{X}$ the translated point process $N_x := \{x_i + x\}_{i=1}^{\infty}$ has the same distribution as N itself. So, for any $A \in \mathcal{A}$ and any $x \in \mathbb{X}$ it holds that

$$\mathbb{P}(N \in A) = \mathbb{P}(N_x \in A).$$

Analogously, a point process is called *isotropic* if it is invariant under any rotation, i.e. for any rotation $\mathbf{r} \in \mathbb{R}^{d\times d}$ the rotated process $\mathbf{r}N = \{\mathbf{r}x_i\}_{i=1}^{\infty}$ has the same distribution as N. That is, for any $A \in \mathcal{A}$ and any rotation $\mathbf{r} \in \mathbb{R}^{d\times d}$ it holds that

$$\mathbb{P}(N \in A) = \mathbb{P}(\mathbf{r}N \in A).$$

Since it often simplifies the analysis of point processes and it is a sensible assumption in many applications, such as the forest example, we assume from now on that the considered point processes are isotropic and stationary. However, the method derived for analyzing multivariate point processes in later chapters can be extended to non-stationary and non-isotropic point processes.

As an example of how the assumption of stationarity simplifies the analysis, we consider again the intensity measure Λ of a stationary point process N. Since the translation-invariance of N transfers to the intensity measure, the intensity measure is proportional to the Lebesgue measure ν_d (see e.g. Chiu et al. (2013), Chapter 4.1.5); this means that

$$\Lambda(B) = \lambda\nu_d(B).$$

Consequently, the intensity function λ is constant over the whole point space. The amount of points in an area depends on the size of that area. In other words, the density of points is homogeneous for the point process N.

Another important assumption for the analysis of point processes is *ergodicity*, which is needed for asymptotic results. It guarantees that, if one considers spatial averages over observation windows converging to the point space $\mathbb{X}$, the limits exist and are deterministic. This justifies basing the analysis of spatial point processes on a sample consisting only of one pattern if the observation

window is appropriately large. For a more detailed explanation of the concept and why it is needed here see e.g. Chiu et al. (2013, Sec. 4.1.6).

An important but complex concept in spatial point process theory is the Palm distribution of a point process. Intuitively, it describes the distribution of the point process close to a typical point in the process. Fundamental numerical and functional summary characteristics introduced later in this chapter, such as the mean nearest-neighbor distance or Ripley's K, build on this concept. There are different approaches to introduce the Palm distribution of a point process. Here, we present the approach via the Radon-Nikodym derivative of the Campbell measure, following Chiu et al. (2013) and Daley and Vere-Jones (1998). For this, first define the Campbell measure of a point process. The *Campbell measure* $\mathscr{C}$ is defined for $B \in \mathcal{B}(\mathbb{X})$ and $A \in \mathcal{A}$ as

$$\mathscr{C}(B \times A) \coloneqq \mathbb{E}[N(B)\mathbb{1}_A(N)] = \int_{\mathbb{A}} \varphi(B)\mathbb{1}_A(\varphi) d\mathbb{P}_N(\varphi),$$

where $\mathbb{1}_A$ denotes the indicator function on a set A. Note again that $\mathbb{P}_N$ is the distribution of a point process N. In other words, the Campbell measure gives the expected number of points in a set B for a point process N given that this process is a sequence of the configuration set A. Using basic measure theoretic methods, the Campbell measure can be uniquely extended to a σ-finite measure on the product Borel-σ-algebra $\mathcal{B}(\mathbb{X}) \otimes \mathcal{A}$; that is, the σ-algebra generated by sets $B \times A \subset \mathbb{X} \times \mathbb{A}$ with $B \in \mathcal{B}(\mathbb{X}), A \in \mathcal{A}$. Alternatively, one can define the Campbell measure as the measure on $\mathcal{B}(\mathbb{X}) \otimes \mathcal{A}$ satisfying

$$\int_{\mathbb{A}} \sum_{x \in \varphi} f(x, \varphi) d\mathbb{P}_N(\varphi) = \int_{\mathbb{A}} f(x, \varphi) d\mathscr{C}(x, \varphi)$$

for every non-negative measurable function f on $\mathbb{X} \times \mathbb{A}$. This way, there is no need for an extension and the first equation follows by setting $f(x, \varphi) = \mathbb{1}_B(x)\mathbb{1}_A(\varphi)$, since it holds that $\sum_{x \in \varphi} \mathbb{1}_B(x) = \varphi(B)$. Also, the *reduced Campbell measure* $\mathscr{C}^!$ can be defined analogously as the measure satisfying

$$\int_{\mathbb{A}} \sum_{x \in \varphi} f(x, \varphi \setminus \{x\}) d\mathbb{P}_N(\varphi) = \int_{\mathbb{A}} f(x, \varphi) d\mathscr{C}^!(x, \varphi).$$

Recall that a point pattern $\varphi \in \mathbb{A}$ is a set of points and therefore $\varphi \setminus \{x\}$ is the set excluding the point $x \in \varphi$.

There is a close connection between the Campbell measure and the intensity measure. Whenever the intensity measure exists, it holds that

$$\mathscr{C}(B \times \mathbb{A}) = \mathbb{E}[N(B)] = \Lambda(B).$$

This relation also yields an approach to introduce the Palm distribution. If the intensity measure exists and is σ-finite, then for each fixed $A \in \mathcal{A}$ the measure $\mathscr{C}(\cdot \times A)$ is absolutely continuous with respect to the intensity measure. Consequently, the Radon-Nikodym derivative exists and defines a $\mathcal{B}(\mathbb{X})$-measurable function $\mathcal{P}_0(\cdot, A)$ satisfying for each $B \in \mathcal{B}(\mathbb{X})$

$$\mathscr{C}(B \times A) = \int_B \mathcal{P}_0(x, A) d\Lambda(x).$$

Since $\mathcal{P}_0$ is the Radon-Nikodym derivative it is defined uniquely up to Λ-null sets. The derivatives

can be chosen such that for fixed $x \in \mathbb{X}$ it holds that $\mathcal{P}_0(x, \cdot)$ is a probability measure on the σ-algebra of configuration sets $\mathcal{A}$ and each such measure is called *local Palm distribution* (cf. Daley and Vere-Jones (1998, Ch. 12.1)). In the case of stationary point processes the Palm distribution simplifies considerably. Under stationarity the point processes N and the translated point process N_x have the same distribution for any $x \in \mathbb{X}$ and the intensity measure has constant density λ with respect to the Lebesgue measure ν_d. Consequently, for any $B \in \mathcal{B}(\mathbb{X}), A \in \mathcal{A}, x \in \mathbb{X}$ it holds that

$$\lambda \int_B \mathcal{P}_0(z, A) d\nu_d(z) = \lambda \int_B \mathcal{P}_0(x + z, A_x) d\nu_d(z).$$

Here, A_x and B_x denote, analogously to N_x, the sets where each sequence and each point is translated by x, respectively. From this follows that $\mathcal{P}_0(z, A) = \mathcal{P}_0(x + z, A_x)$ for any $x, z \in \mathbb{X}$, $A \in \mathcal{A}$. Setting $z = 0$ justifies the definition of the Palm distribution P_0 of a point process N as the measure satisfying for all $x \in \mathbb{X}$

$$\mathrm{P}_0(A) \coloneqq \mathcal{P}_0(x, A_x), \quad A \in \mathcal{A}.$$

Since the measure does not depend on x as long as the process is translated accordingly, one can simply choose the measure with $x = 0$, which is why one uses the index zero. Alternatively, in some literature the index o is used indicating the origin of the point space. In the case of vector spaces such as $\mathbb{R}^d$ the origin often is the zero. This also motivates the usage of the phrase distribution of a typical point in this context. Defining the Palm distribution as the density of the Campbell measure also yields the interpretation as a distribution of the point process close to a typical point in the process.

Analogously, the reduced Palm distribution $\mathrm{P}_0^!$ is defined using the reduced Campbell measure $\mathscr{C}^!$ instead of the Campbell measure. Since in the integral equation that defines the reduced Campbell measure the point of consideration is excluded, the reduced Palm distribution can be interpreted as the distribution close to a typical point in a point process, excluding the point.

In this chapter, we discussed a set of different means of describing the distribution of a point process. Based on these concepts, the following chapter examines how to describe certain aspects of the distribution of a point process.

2.2 Summary Characteristics

The distribution of a point process N is rather complex and not intuitive to understand or visualize. Summary characteristics are a means to facilitate the analysis of point processes describing certain aspects of the distribution. In this chapter, we introduce a set of numerical and functional summary characteristics together with corresponding estimation methods.

In general, the standard approach in analyzing point processes is to compare properties of a point pattern to corresponding properties of a Poisson process. In a Poisson process, all points are independent from each other, which is why Poisson processes are used to describe so called *complete spatial randomness* (CSR). Refer to Illian et al. (2008, Ch. 2) for a more precise and detailed introduction to the Poisson process in the case of stationarity.

In particular, it is explored if in the given pattern points are allocated completely randomly or if there is a tendency towards either clustering or regularity. Most summary characteristics have a simple form for the Poisson process due to the independence between points.

A simple summary characteristic already discussed in the previous chapter is the intensity function λ. Recall that throughout this thesis stationarity is assumed. Consequently, the intensity λ is a constant. It contains important information about the point process and is an essential summary characteristic when analyzing a stationary point process. The interpretation of the intensity as the point density, i.e. the number of points per area, yields an obvious estimator. Let W be the observation window and φ a realization of N, then an estimator $\hat{\lambda}$ of the intensity λ is given by

$$\hat{\lambda} \coloneqq \frac{\varphi(W)}{\nu_d(W)},$$

that is, the number of points in the observation window divided through the area of the window. One can easily see that this estimator is unbiased. It is furthermore consistent in the sense that, if the window size increases towards the point space $\mathbb{X}$ as a convex averaging sequence, then the estimator converges almost surely to the true value (Chiu et al. (2013, Sec. 4.7.3)). The intensity can be seen as the scale of the process and point patterns are usually compared to Poisson processes with its estimated intensity.

Since the intensity is a scalar, it is also called a *numerical summary characteristic*. For stationary point processes, one often also considers *functional summary characteristics* that describe the behavior of a process in dependence of the distance between points. As a first example, consider the *nearest-neighbor distance distribution function* D for a point process N. It is defined as

$$\mathrm{D}(r) \coloneqq \mathrm{P}_0^!(N(b(0,r)) > 0), \quad r \geq 0.$$

Here, $b(0,r)$ denotes the ball around 0 with radius r. Note that the reduced Palm distribution is used in the definition. Recall that it gives the distribution of the point process close to a typical point in the process excluding the point itself. Consequently, the function D gives the probability that for a typical point the distance to its nearest neighbor is less than r, so D can be interpreted as the distribution function from the typical point to its nearest neighbor. Therefore, it is also considered an interpoint-distance based characteristic.

A very intuitive estimator for $\mathrm{D}(r)$ given a realization of the point process is to count for how many points there is at least one other point within distance r or closer and to divide through the number of points in the observation window. Define the function d as the function giving for each point the distance to its nearest neighbor. Then, given realization $\varphi = \{x_i\}_{i \in \mathbb{N}}$ of the point process on observation window W, an estimator is given by

$$\widehat{\mathrm{D}}(r) \coloneqq \frac{\sum_{i \in \mathbb{N}} \mathbb{1}_{[0,r]}(d(x_i))}{\varphi(W)}.$$

One problem of this estimator is that for points near the border of the observation window the nearest neighbor might not be observed. In other words, no point within distance r is observed, even though its actual nearest neighbor is closer than r, but happens to lie outside the observation window. There are multiple methods to deal with this problem called *border corrections*. Since the estimation of point process characteristics can become quite complicated due to the need for methods such as border corrections, the reader is referred to Illian et al. (2008) or Chiu et al. (2013) for a detailed coverage of this topic. In the following, the most straightforward estimator for each summary characteristic is stated. They are often not very appropriate, but they give a starting point for the development of more sensible estimators and problems such as border corrections are neglected. However, in the simulation study and application chapters of this thesis, the border corrected versions of the estimators are used unless otherwise stated. Further theoretical discussions

of the estimators are omitted since it is out of scope of this thesis.

Often, functional summary characteristics can be used as a basis for numerical summary characteristics; for example, evaluating the function at a certain value r. In the case of the nearest-neighbor distance distribution, a meaningful numerical summary characteristic is the mean m_{D} with respect to this distribution. It is referred to as *mean nearest-neighbor distance*. It may be preferred to describe the distribution independent from its scale. If a characteristic does not depend on the number of points in the pattern, it is called *scale-invariant*. The Clark-Evans-index CEI can be used as a scale-invariant version of the mean nearest-neighbor distance. It is defined as the mean nearest-neighbor distance of the process divided through the mean nearest-neighbor distance for a Poisson process of the same intensity λ. For the case $d = 2$, the latter is given by $(2\sqrt{\lambda})^{-1}$ which leads to

$$\mathrm{CEI} = 2\sqrt{\lambda} m_{\mathrm{D}}.$$

Alternatively, a scale-invariant measure SK is defined as the ratio of the mean nearest-neighbor distance to the median nearest-neighbor distance. Analogously to m_{D}, the *median nearest-neighbor distance* $\widetilde{m}_{\mathrm{D}}$ is defined as the median of the nearest-neighbor distance distribution. Then, SK is given by

$$\mathrm{SK} = \frac{m_{\mathrm{D}}}{\widetilde{m}_{\mathrm{D}}}.$$

The measure SK focuses on the shape of the nearest-neighbor distance distribution and its degree of skewness. A value close to one indicates regularity and that the distribution is symmetric, while a larger value indicates that the mean is larger than the median and hence the distribution has a positive skew. In Illian et al. (2008), this measure is denoted by μ.

A different approach for summary characteristics is to use directions between points instead of distances. An example for an index based on directions is the *mean directional index* $\overline{R}$ introduced in Corral-Rivas et al. (2006). For a point x_i in the given point pattern $\varphi = \{x_i\}_{i=1}^n$, define $e_{i1}, \ldots, e_{ik}$ as the vectors of norm one that are directed from point x_i to the $k \in \mathbb{N}$ nearest neighbors of x_i in the pattern. Then, the value R_i is defined as

$$R_i = \|e_{i1} + \ldots + e_{ik}\|_{\mathbb{R}^d}. \tag{1}$$

That is, it gives the length of the sum of these vectors. The mean directional index is then defined as the average of the values R_i over all points, $\overline{R} = \frac{1}{n}\sum_{i=1}^n R_i$. Applications have shown that four neighbors is an appropriate choice ($k = 4$), see Corral-Rivas et al. (2006) for references. Therefore, in the following, we refer to the mean directional index as the index with four neighbors. As shown in the simulation study in Corral-Rivas et al. (2006), this index can be used to classify a point pattern as clustered, regular or random. In the extreme case of a regular grid, it holds that $\overline{R} = 0$ since opposing vectors cancel each other out. For the Poisson process it holds that $\overline{R} = 1.799$. For clustered patterns, the mean directional index has a tendency to take values larger than 1.799.

The functional summary characteristic D is considered short-sighted (Illian et al. (2008, Ch. 4.2.6)) as it only considers the nearest neighbor and any behavior of the point process over larger distances is not taken into account. Therefore, other functional summary characteristics are introduced, sometimes referred to as *second-order summary characteristics*. Some authors refer to all functional summary characteristics as second-order summary characteristics and not only the

following functions.

Ripley's K-function measures the average number of points around a typical point in a point process within some distance r. Since it describes the distribution of a typical point, the reduced Palm distribution is used to define the K-function. Let $\mathbb{E}_0^!$ denote the expectation with respect to the reduced Palm distribution P_0 of the point process N. Then, Ripley's K-function for this process is defined as

$$\mathrm{K}(r) \coloneqq \frac{1}{\lambda}\mathbb{E}_0^![N(b(0,r))], \quad r \geq 0.$$

Note that the expectation is also divided through the intensity λ of the process. This is done to separate out the global point density and it further makes the K-function of different point processes easier to compare.

As discussed above, the K-function of a Poisson process serves as a starting point for the analysis. Recall that a Poisson process can be used to decide if a pattern has a tendency towards complete spatial randomness, clustering or regularity. The K-function of a Poisson process has the following form

$$\mathrm{K}(r) = b_d r^d,$$

where b_d denotes the factor corresponding to dimension d for the volume of a unit sphere. In the case of $d = 2$ it holds that $b_d = \pi$. The equality holds since for homogeneous Poisson processes the Palm distribution coincides with that of the point process obtained by adding 0 to the point process; see Illian et al. (2008, Ch. 2.5.2) for a detailed explanation.

The K-function of a point process is proportional to r^d. In the case of clustering, for a small r one expects more points close to a typical point, resulting in the K-function having higher values than the Poisson K-function. Conversely, in the case of regularity, for small distances there are few points resulting in the K-function having lower values. A straightforward estimate of the expectation term of the K-function is to count for each point in a pattern the amount of neighbors within distance r and take the average. That is, given a point pattern $\{x_i\}_{i=1}^n$, let $n_i(r) = N(b(x_i, r) \setminus \{x_i\})$ denote the number of points around each point x_i in the point pattern within distance r. Then,

$$\bar{n}(r) \coloneqq \frac{1}{n}\sum_{i=1} n_i(r)$$

gives the average of points within distance r in the point pattern and therefore yields an estimator of the reduced Palm expectation term in the K-function. For the intensity one could then insert the estimator mentioned above which yields an estimator for the K-function, i.e. $\widehat{\mathrm{K}}(r) = \bar{n}(r)/\hat{\lambda}$.

To simplify the analysis, one often considers a linearization of Ripley's K-function. This function is called *Besag's* L*-function* and for the d-dimensional case it is defined as

$$\mathrm{L}(r) \coloneqq \sqrt[d]{\frac{\mathrm{K}(r)}{b_d}}, \quad r \geq 0.$$

The L-function is then proportional to r instead of r^d. In the case of $d = 2$ this means one analyzes lines instead of parabolic curves. In the Poisson case, it holds that $\mathrm{L}(r) = r$. According to Illian et al. (2008), experience shows that with increasing r the fluctuations of Ripley's K-functions increase. The linearization to the L-function stabilizes fluctuations and can make them independent of r. Estimations of the L-function can be based on estimation methods for the K-function.

A third functional summary characteristic containing the same information as the K- and the L-function is the pair correlation function. It can be seen as the density of the K-function relative to the density of the K-function of the Poisson process with respect to Lebesgue measure, respectively. Assuming that the K-function has a first derivative K' with respect to r, the *pair correlation function* g can be defined as

$$g(r) \coloneqq \frac{\mathrm{K}'(r)}{db_d r^{d-1}}, \quad r \geq 0.$$

It is given by the derivative of the K-function divided through the derivative of the K-function of the Poisson process. Obviously, it follows that the pair correlation function of the Poisson process is constantly one. In the case of $d = 2$, the pair correlation function can equivalently be defined via the product density $\varrho^{(2)}$; see Illian et al. (2008) for reference. Recall that the function $\varrho^{(2)}$ is defined as the density of the second moment measure $\alpha^{(2)}$ with respect to the Lebesgue measure. Heuristically, for points $x, y \in \mathbb{X}$, the term $\varrho^{(2)}(x, y)dxdy$ gives the probability that there is a point both in the infinitesimal small ball of volume dx around x and simultaneously a point in the ball of volume dy around y. In the case of stationarity, the density $\varrho^{(2)}$ is only dependent on the distance $d(x, y) = r$ between points and can therefore be seen as a function in $r \geq 0$. Then, one can show that for the pair correlation function it holds that

$$g(r) = \frac{\varrho^{(2)}(r)}{\lambda^2}, \quad r \geq 0.$$

Thus, the pair correlation function can be seen as a version of the density $\varrho^{(2)}$ scaled by the intensity λ for each argument. For clustered processes we observe a pair correlation function larger than one for small r and for regular processes less than one because of properties of corresponding K-functions. For the estimation of the pair correlation function, methods analogously to the K-function can be used.

Summarizing, detailed analysis of single point processes, also considered as univariate point processes, is possible based on numerical and functional summary characteristics. In the next chapter, we consider the extension of spatial point processes to the multivariate case.

2.3 Multivariate Point Processes

In applications multiple different point processes on the same area are often observed, for example, different types of trees in a forest. More precisely, the object of interest is then a n-dimensional *multivariate spatial point process* which can be seen as a tuple $\mathbf{N} = (N_1, \ldots, N_n)$ of point processes $N_1, \ldots, N_n$ for $n \in \mathbb{N}$ on the same point space $\mathbb{X}$ and the same probability space. Measurability constraints need to be adapted accordingly. Equivalently, multivariate point processes can be defined as a specific kind of marked point processes.

A marked point process is a point process with additional information attached to each point. In general, this information can be any element of a complete separable metric space $\mathbb{M}$. The information attached is called *mark* and $\mathbb{M}$ is called *mark space*. Denote by $\mathcal{B}(\mathbb{M})$ the corresponding Borel-σ-algebra. Most important examples of mark spaces are subsets of the real and natural numbers.

In the case of $\mathbb{M} \subset \mathbb{R}$, the marks are called *quantitative* marks. An example would be considering the trees in a forest together with their diameter at breast height. The locations of the trees form

the point pattern and the breast height the corresponding quantitative marks.

In the case of $\mathbb{M} \subset \mathbb{N}$, that is, being either discrete or categorical, the marks are called *qualitative* marks. In the forest example, the mark could indicate the type of tree from a finite set of tree species.

Following the notation of the definition of spatial point processes in Chapter 2.1 where a point process was defined as a random sequence $\varphi = \{x_i\}_{i \in \mathbb{N}}$ satisfying certain properties, a *marked point process* M is a corresponding random sequence of the form

$$\{(x_i, m_i)\}_{i \in \mathbb{N}} \subset \mathbb{X} \times \mathbb{M}.$$

More precisely, a marked point process can be seen as a point process on the product space $\mathbb{X} \times \mathbb{M}$, where the marginal point processes also need to be point processes themselves.

Consequently, a n-dimensional multivariate point process $\mathbf{N}$ can be equivalently defined as a qualitative marked point processes with mark set $\mathbb{M} = \{1, \ldots, n\}$. Marginal point processes are immediately well defined due to the finiteness of $\mathbb{M}$ (c.f. Daley and Vere-Jones (1998)) and each marginal point process corresponds to one element of the multivariate point process. Considering X as the set of all points in the multivariate point process combined, each component can be written as

$$N_l = \{(x_i, m_i) \colon x_i \in X, m_i = l\}_{i \in \mathbb{N}}, \quad l \in \mathbb{M}.$$

Therefore, the example of different tree species can both be seen as a qualitative marked point process with each species being one mark, and as a multivariate point process $\mathbf{N} = (N_1, \ldots, N_n)$ with each of the l different species corresponding to one element N_l of the process. The processes N_l are referred to as *subprocess* or *component processes*.

The aim of this thesis is to analyze multivariate point processes without further marks which is discussed in the next chapter.

2.4 Analyzing Multivariate Point Processes

There are several approaches for analyzing multivariate point processes. However, the literature so far is mostly limited to the bivariate case. For example, Diggle and Milne (1983) discuss different models for bivariate spatial point patterns. Alternatively, the above definition of multivariate point processes as qualitatively marked point processes suggests that one could use methods derived for marked point fields, such as second-order characteristics (e.g. see Stoyan and Stoyan (1994, Sec. 14.6)). Other works consider summary characteristics for multivariate spatial point processes. These can mostly be divided into two types of characteristics, *dot-type* or *cross-type* characteristics. For cross-type characteristics, summary characteristics for univariate spatial point processes such as the nearest-neighbor distance distribution function D can be defined based on distances between points of two different component processes. Dot-type characteristics on the other hand describe the relationship of points in one component process to the points of all other component processes.

All above approaches only make it possible to compare two component processes at a time. Analyzing multivariate spatial point processes can then become tedious very fast when the amount of subprocesses n becomes large. If one can only compare two point processes simultaneously, one would need to do $n(n-1)/2$ comparisons which quickly becomes computational burdensome. For example, comparing and analyzing the first 8 tree species in the Duke forest data set based on the cross pair correlation function 28 plots need to be compared (see Appendix, Figure 26) which

makes it difficult to detect groups of similar spatial behavior.

The analysis of a large number of point processes is restricted to the work of Eckardt and Mateu (2019) and an approach originally presented in Illian (2006). The latter was further developed in a series of contributions including Illian et al. (2006) and Illian et al. (2008). When referring to this approach in the following, a contribution is cited that contains the aspect discussed in the corresponding paragraph and when referring to the idea in general, Illian (2006) is cited.

The approach by Eckardt and Mateu (2019) is based on graphical modeling. Subpatterns are identified as nodes of a graph and connected if similar, where similarity is based on spherical densities. In this thesis we propose an extension for the latter approach; therefore, it is presented in the following in more detail.

The approach proposed in Illian (2006) is based on using summary characteristics. For each subprocess consider certain summary characteristics and perform data analysis on the resulting data set. Two approaches are proposed, one based on numerical and one based on functional summary characteristics.

Given a realization $\boldsymbol{\varphi} = (\varphi_1, \dots, \varphi_n)$ of a multivariate spatial point process $\mathbf{N} = (N_1, \dots, N_n)$, the idea of the first approach is to estimate a fixed set of $p \in \mathbb{N}$ appropriate numerical summary characteristics $v^{(1)}, \dots, v^{(p)}$ for each subpattern $\varphi_i, i \in \{1, \dots, n\}$. We view the subpatterns as observation units and the summary characteristics as variables. This yields vectors $v_i = (v_i^{(1)}, \dots, v_i^{(p)})^{\mathrm{T}} \in \mathbb{R}^p$ for each pattern $i \in \{1, \dots, n\}$. By v^{T} denote the transpose of a vector $v \in \mathbb{R}^p$.

Then, the data set consisting of the vectors v_i can be analyzed using standard multivariate data analysis methods such as principal component analysis and cluster analysis. See Chapter 3.1 for details on standard principal component analysis.

For example, if the first two principal components explain most of the variation in the numerical summary characteristics, the two dimensional score vectors can be used as a low dimensional representation for each point pattern. Based on the scores, using for example Ward's algorithm, the patterns can then also be grouped according to their spatial behavior. For an introduction to cluster analysis and a definition of Ward's algorithm, we refer to Everitt et al. (2001). In Illian et al. (2008), there is a detailed application and discussion of this method for grouping multivariate spatial point patterns on a data set from community ecology.

Alternatively, Illian (2006) proposes to estimate one functional summary characteristic for each pattern, such as the pair correlation function g, and then applies functional data analysis on the functional data $g_1, \dots, g_n$, again considering each pattern as one observation unit. For an introduction to functional data analysis, we refer to Ramsay and Silverman (1997).

More precisely, the idea is to perform functional principal component analysis and to use the first few scores as a low dimensional representation for the spatial behavior of each pattern, for example for grouping the point patterns, e.g. using Ward's algorithm. See Chapter 3.3 for details of functional principal component analysis. In Illian et al. (2006), there is a detailed simulation study illustrating the good performance of this method on several different simulated data sets.

Note that summary characteristics are analyzed describing the spatial behavior of the point processes. Thus, the spatial component of the data is automatically controlled for and there is no need to explicitly model the spatial structure any further.

However, each numerical and functional summary characteristic contains information on certain aspects of the distribution of a point process, both are needed to get an in depth understanding of the spatial behavior of a point process. Therefore, a straightforward extension to both approaches

proposed by Illian (2006) is to combine functional and numerical summary characteristics to form a hybrid data object and then perform hybrid multivariate functional principal component analysis (MFPCA) on the data. Finally, use the resulting first few hybrid scores as a representation to group the patterns based on their spatial behavior, measured with the aspects of the point process distribution contained in both numerical and functional summary characteristics.

Before this method can be studied in a simulation study in Chapter 4 and applied to a data set in Chapter 5, an important question needs to be answered: How can one perform hybrid MFPCA? This problem is topic of Chapter 3, where we discuss hybrid data and the derivation of hybrid MFPCA in general, not only in the context of analyzing multivariate spatial point processes.

3 Hybrid Multivariate Functional Principal Component Analysis

In this chapter, the theory of hybrid data is introduced. Hybrid data describes the combination of different data modalities. For the purpose of this thesis, functional and numerical data is combined. Before discussing the theory in more detail, the following gives a motivation for functional and subsequently hybrid data analysis.

The importance of functional data is illustrated by the PhD thesis by Illian (2006) as already indicated in the last chapter. The contribution of this thesis is to model and analyze multivariate point processes by applying principal component analysis (PCA) on summary characteristics and using the scores to facilitate grouping the point processes according to their spatial behavior. Specifically, the approach is based on functional summary characteristics which make it possible to compare point processes based on their behavior over the whole space at once. Therefore, Illian (2006) proposes the use of functional principal component analysis (FPCA) on a second-order functional summary characteristic estimated on a discrete grid, such as the pair correlation function.

The importance of functional data analysis exceeds the application of this thesis by far. Recent advances in data collection techniques allow for the collection of big amounts of data; especially functional data gains importance in research. But what is functional data? In general, functional data is considered to be the observation of discrete values which reflect smooth variation, e.g. over space or time.

To define functional data more precisely, there are two approaches. Firstly, functional data can be considered to be realizations of random variables that take values in a Hilbert space such as the space of L^2-integrable functions. Secondly, functional data can be viewed as sample paths of stochastic processes, which may have various properties, such as continuity or square integrability with probability one (Hsing and Eubank (2015)). For an introduction to both approaches and how they are related, see Hsing and Eubank (2015, Ch. 7).

A main challenge in modeling functional data is the inherent infinite dimensionality of the observed objects. The key technique for reducing dimensionality in functional data while preserving most variability is functional principal component analysis, analogously to principal component analysis in multivariate data analysis.

There is a large and steadily growing body of literature on FPCA. For a summary see Jang (2021).

However, analyzing multivariate point processes based on functional summary characteristics still has limitations, since they only reflect certain aspects of the complex distribution of a point process; recall Chapter 2.2 for the discussion of summary characteristics. This motivates the aim of this thesis to extend the method to a combination of multiple numerical and functional summary characteristics which form a hybrid data object.

Again, analyzing different data modalities is gaining importance in many other fields as well, e.g. in clinical data where an increasing amount of medical devices becomes capable of collecting multiple data modalities, e.g. taking images (Jang (2021)). The most common combination is that of vector and (multivariate) functional data. Therefore, principal component analysis for this kind of hybrid data is derived in this chapter.

Advantages of principal component analysis for hybrid data are, analogously to vector and functional data, that the principal components are intepretable, they have the same structure as

the data and result in a single set of scores that provide a low dimensional representation for hybrid data (Jang (2021)).

The remaining part of this chapter is structured as follows. For the derivation of hybrid MFPCA we review main aspects of standard principal component analysis (PCA) in Chapter 3.1 to gain a sense of the principal component analysis concept. Then, the concept is generalized for Hilbert spaces in Chapter 3.2 to better understand the theoretical background needed. In Chapter 3.3 MFPCA is discussed based on this theoretical background. All this together yields the theoretical foundation needed for the definition of hybrid MFPCA and appropriate estimation methods in Chapter 3.4. The following chapters summarize notation and concepts of (Ramsay and Silverman; 1997, Ch. 6, 8), Hsing and Eubank (2015, Ch. 1, 7-9), Jang (2021) and Happ and Greven (2018). Notations were adapted when necessary to be consistent within this thesis.

3.1 Standard Principal Component Analysis

In this chapter, principal component analysis is reviewed for vector data to outline the main ideas in an easily interpretable context. For this, first we consider the setting.

In standard PCA multivariate data is considered, that is, a random vector $V = (V^{(1)}, \dots, V^{(p)})^{\mathrm{T}}$ of dimension $p \in \mathbb{N}$ is analyzed. The first two moments need to exist so that the covariance matrix is well-defined (Jang (2021)). Realizations v of V are in $\mathbb{R}^p$, therefore, this is the space of interest and appropriate operators need to be defined. On $\mathbb{R}^p$ consider the standard Euclidean scalar product $\langle \cdot, \cdot \rangle_{\mathbb{R}^p}$ and the induced norm $\| \cdot \|_{\mathbb{R}^p}$. That is, for $v_1 = (v_1^{(1)}, \dots, v_1^{(p)})^{\mathrm{T}}, v_2 = (v_2^{(1)}, \dots, v_2^{(p)})^{\mathrm{T}} \in \mathbb{R}^p$ it holds that

$$\langle v_1, v_2 \rangle_{\mathbb{R}^p} = v_1^{\mathrm{T}} v_2 = \sum_{r=1}^{p} v_1^{(r)} v_2^{(r)} \qquad \text{and} \qquad \|v_1\|_{\mathbb{R}^p} = \sqrt{\langle v_1, v_1 \rangle_{\mathbb{R}^p}}.$$

Without loss of generality, assume that the vector has mean zero, i.e. $\mathbb{E}[V] = 0$, otherwise consider the centered random vector $V - \mathbb{E}[V]$. The goal of principal component analysis is to reduce dimensionality while still displaying types of variation that are very strongly represented in the data. Hence, it might be reasonable to standardize V before the analysis if the scales of the components vary considerably.

The idea of standard PCA is to search the linear combination of V that accounts for the maximum of total variation. For this, let $C_V \coloneqq \mathbb{E}[VV^{\mathrm{T}}]$ denote the covariance matrix of V. Then, search for the vector w_1 that attains the maximum

$$\max_{w \in \mathbb{R}^p, w^{\mathrm{T}} w = 1} \mathbb{V}\mathrm{ar}(w^{\mathrm{T}} V) = \max_{w \in \mathbb{R}^p, w^{\mathrm{T}} w = 1} w^{\mathrm{T}} C_V w.$$

w_1 is called first *principal component* (PC). In the j-th subsequent step, $j \leq p$, search for the linear combination of V that accounts for the remaining total variation after removing the portion explained by $w_1, \dots, w_{j-1}$. That is, search for w_j that attains

$$\max_{\substack{w \in \mathbb{R}^p, w^{\mathrm{T}} w = 1, \\ w^{\mathrm{T}} w_i = 0, i = 1, \dots, j-1}} \mathbb{V}\mathrm{ar}(w^{\mathrm{T}} V) = \max_{\substack{w \in \mathbb{R}^p, w^{\mathrm{T}} w = 1, \\ w^{\mathrm{T}} w_i = 0, i = 1, \dots, j-1}} w^{\mathrm{T}} C_V w.$$

w_j is called j-th principal component. One can show that the principal components $w_1, \dots, w_p$ are equivalent to orthonormal eigenvectors of C_V. Therefore, instead of solving the maximum problems one could perform an eigen decomposition fo C_V. An eigen decomposition of the covariance matrix

C_V exists, since it is symmetric positive semi-definite. More precisely, solve

$$C_V w_j = \beta_j w_j, \quad j = 1, \ldots, p,$$

where β_j is the j-th eigenvalue of C_V and $w_j = (w_j^{(1)}, \ldots, w_j^{(p)})^{\mathrm{T}}, j \in \{1, \ldots, p\}$, is the associated eigenvector. In addition, the eigenvectors are chosen to be orthonormal, i.e. $\langle w_j, w_i \rangle_{\mathbb{R}^p} = \delta_{ij}$, where δ_{ij} denotes the Kronecker delta. The eigenvalues are sorted in descending order such that $\beta_1 \geq ... \geq \beta_p \geq 0$. Then, the eigen decomposition of the covariance matrix is given by

$$C_V = \sum_{j=1}^{p} \beta_j w_j w_j^{\mathrm{T}}.$$

Clearly, if C_V has full rank, the eigenvectors $w_1, \ldots, w_p$ provide a complete orthonormal basis for $\mathbb{R}^p$. Therefore, the vector V can be written as a weighted sum of the eigenvectors, that is,

$$V = \sum_{j=1}^{p} \gamma_j w_j.$$

The coefficients $\gamma_j, j = 1, \ldots, p$, can be calculated in the standard way by the scalar product of V and the eigenvector, i.e.

$$\gamma_j = \langle w_j, V \rangle_{\mathbb{R}^p} = w_j^{\mathrm{T}} V.$$

These coefficients γ_j are also called *principal component scores*. The PC scores have the important properties that they have mean zero, are uncorrelated and have variances equal to the corresponding eigenvalues of the covariance matrix β_j due to the orthonormality of the eigenvectors and bilinearity properties of the scalar product. By weighting each eigenvector of random vector V, the PC scores become the coefficients in the linear combinations of the original variable that maximize the variance in the data.

Since the goal is to reduce dimensionality, especially if the dimension p of the vector V is large, only a number $J < p$ of the components is retained summarizing the covariance relationship between the variables in V. The truncated vector PC decomposition is given by

$$V \approx V^{[J]} := \sum_{j=1}^{J} \gamma_j w_j.$$

Clearly, one looses some information by using only the first J summands. The extend of loss can be measured by considering the proportion of total variance explained $\mathrm{TVE} = \mathrm{trace}(C_V) = \sum_{j=1}^{p} \beta_j$. Since it holds that $\beta_1 \geq \beta_2 \geq \ldots$, the first principal component w_1 captures most of the variance in the data; subsequent PCs capture successively smaller fractions of the TVE. Often β_j becomes small very fast.

The j-th component accounts for β_j/TVE of the total variance. Therefore, the first J components explain $\varepsilon = \sum_{j=1}^{J} \beta_j/\mathrm{TVE}$ of the total variance. This provides a mean to choose J, namely, to decide on the value J such that ε exceeds some prespecified value, e.g. 0.9.

Finally, the truncated score vector $\gamma = (\gamma_1, \ldots, \gamma_J)^{\mathrm{T}}$ is used as a lower dimensional representation of V. With this one gets a form of dimension reduction while preserving most variation in the data, which was the goal. This is especially practical if the amount of variables p is large. Note that PCA is not unique, e.g. it is always possible to change signs of all the values in any vector w_j.

In practice, one has a sample of $n \in \mathbb{N}$ independent and identically distributed observations of V denoted by $v_1, \ldots, v_n$ with $v_i = (v_i^{(1)}, \ldots, v_i^{(p)})^{\mathrm{T}}, i \in \{1, \ldots, n\}$. Eigenvalues and eigenvectors can be estimated based on the eigen decomposition of the sample covariance matrix

$$\widehat{C}_V = \frac{1}{n-1} \sum_{i=1}^{n} (v_i - \hat{\mu}_V)(v_i - \hat{\mu}_V)^{\mathrm{T}}, \quad \text{with } \hat{\mu}_V = \frac{1}{n} \sum_{i=1}^{n} v_i$$

yielding estimates for the eigenvectors $\hat{w}_j$ and eigenvalues $\hat{\beta}_j$. The first J estimated eigenvectors can be interpreted as the directions indicating the most variation in the data. The j-th PC score of the i-th subject can be estimated by $\hat{\gamma}_{ij} = (v_i - \hat{\mu}_V)^{\mathrm{T}} \hat{w}_j$, $i \in \{1, \ldots, n\}$ and $j \in \{1, \ldots, p\}$. Consequently, for each observation one obtains a score vector $\hat{\gamma}_i = (\hat{\gamma}_{i1}, \ldots, \hat{\gamma}_{ip})^{\mathrm{T}}$ and uses the first J elements as a lower dimensional representation of each observation.

3.2 Principal Component Analysis on Hilbert Spaces

More generally, PCA can be defined on Hilbert spaces. It is then is possible to consider functional PCA as the case when the observed object is a random element of a Hilbert space. In the case of this thesis, the Hilbert space of interest for modeling functional data is the space of square integrable functions on some domain $\mathcal{T}$ with respect to its Borel-σ-algebra and the Lebesgue measure $L^2(\mathcal{T}, \mathcal{B}(\mathcal{T}), \nu)$. In addition, if one defines an appropriate Hilbert space for hybrid data, hybrid MFPCA can be derived. This chapter closely follows Hsing and Eubank (2015, Ch. 7-9). All definitions and theorems cited in this chapter are taken from this work.

First, we generalize the concept of random variables to random elements of a Hilbert space. Let $\mathbb{S}$ be a Hilbert space and $S \in \mathbb{S}$ be a random element of this space with $\mathbb{E}\|S\|_{\mathbb{S}}^2 < \infty$. Here, $\|\cdot\|_{\mathbb{S}}$ denotes the norm on the Hilbert space $\mathbb{S}$. The *covariance operator* $\mathscr{K}$ is given by

$$\mathscr{K} = \mathbb{E}[(S - \mu_S) \otimes (S - \mu_S)] \coloneqq \int_{\Omega} (S - \mu_S) \otimes (S - \mu_S) d\mathbb{P},$$

where μ_S is the mean of S, $(\Omega, \mathcal{A}, \mathbb{P})$ the probability space on which S is defined and $\otimes$ denotes the corresponding tensor product. Note that, in contrast to the last chapter, it is not assumed that the random element of the Hilbert space has mean zero. Instead, the mean is subtracted in the definition of the covariance kernel. The integral is the so called *Bochner integral* which is a generalization of the Lebesgue integral for Banach spaces and, therefore, is also defined for Hilbert spaces. Refer to Hsing and Eubank (2015, Ch. 2.6) for a derivation of this integral. The covariance operator maps an object of the Hilbert space to another object of this Hilbert space, that is, $\mathscr{K} : \mathbb{S} \longrightarrow \mathbb{S}, s \mapsto \mathbb{E}[(S - \mu_S) \otimes (S - \mu_S)](s) = \mathbb{E}[\langle (S - \mu_S), s \rangle_{\mathbb{S}} (S - \mu_S)]$. Accordingly, the Hilbert-Schmidt Theorem gives that the covariance operator $\mathscr{K}$ admits the eigen decomposition

$$\mathscr{K} = \sum_{j=1}^{\infty} \beta_j (s_j \otimes s_j).$$

The coefficients $\{\beta_j\}_{j \in \mathbb{N}}$ are the non-negative eigenvalues. Either the set is finite or the sequence tends towards zero. The set of eigenobjects $\{s_j\}_{j \in \mathbb{N}} \subset \mathbb{S}$ forms an orthonormal basis of the image of the covariance operator. They are the Hilbert space principal components. Then, the *Karhunen-*

Loève decomposition of a random element $S \in \mathbb{S}$ is given with probability one as

$$S = \mu_S + \sum_{j=1}^{\infty} \langle S, s_j \rangle_{\mathbb{S}} s_j,$$

where $\langle \cdot, \cdot \rangle_{\mathbb{S}}$ denotes the scalar product of $\mathbb{S}$ and $\langle S, s_j \rangle_{\mathbb{S}}$ are uncorrelated random variables with mean zero and variances β_j; we call them principal component scores. We term the process of evaluating the above equations principal component analysis. For dimension reduction we consider the truncated decomposition (e.g. only the first $L \in \mathbb{N}$ terms).

Consider the case $\mathbb{S} = \mathbb{R}^p$. Since the tensor product in the Euclidean space is equal to the outer product, the covariance operator becomes $\mathbb{E}[(S - \mu_S)(S - \mu_S)^{\mathrm{T}}]$ which is the definition of the covariance matrix. Then, the eigen decomposition of the covariance operator is the eigen decomposition of the covariance matrix. Consequently, standard PCA coincides with PCA on Hilbert spaces. Since the dimension of $\mathbb{R}^p$ is finite, the sums in the decomposition are finite as well.

In the next chapter, we consider the case of functional data.

3.3 (Multivariate) Functional Principal Component Analysis

Now, we extend the concept of principal component analysis for (multivariate) functional data. First, we introduce the setting and important notations.

The object of interest is $F = (F^{(1)}, \ldots, F^{(q)})^{\mathrm{T}}$, a stochastic process of dimension $q \in \mathbb{N}$ on compact sets with finite Lebesgue measure $\mathcal{T}_1, \ldots, \mathcal{T}_q$, and $\mathcal{T}_r \subset \mathbb{R}^{d_r}$, $d_r \in \mathbb{N}$, $r = 1, \ldots, q$. Assume that realizations of each process $F^{(r)}$ are in $L^2(\mathcal{T}_r)$, $r = 1, \ldots, q$, that is, if the $F^{(r)}$ are defined on the probability space $(\Omega, \mathbb{F}, \mathbb{P})$, then it holds that $F^{(r)}(\omega) : \mathcal{T}_r \longrightarrow \mathbb{R}$ with

$$F^{(r)}(\omega) \in L^2(\mathcal{T}_r), \text{ i.e. } \quad \int_{\mathcal{T}_r} F^{(r)}(\omega)(t_r)^2 d\nu_{d_r}(t_r) < \infty, \qquad \forall \omega \in \Omega,\ r = 1, \ldots, q.$$

In the following, we use the notations $F(t) \coloneqq (F^{(1)}(t_1), \ldots, F^{(q)}(t_q))^{\mathrm{T}} \in \mathbb{R}^q$, $t \coloneqq (t_1, \ldots, t_q)^{\mathrm{T}}$ element of $\mathcal{T} \coloneqq \mathcal{T}_1 \times \ldots \times \mathcal{T}_q$. Denote the functional space by $\mathbb{F} \coloneqq L^2(\mathcal{T}_1) \times \ldots \times L^2(\mathcal{T}_q)$, the Cartesian product of individual L^2-spaces. For $\mathbb{F}$ consider the L^2-scalar product and -norm, i.e. for $f_1 = (f_1^{(1)}, \ldots, f_1^{(q)})^{\mathrm{T}}, f_2 = (f_2^{(1)}, \ldots, f_2^{(q)})^{\mathrm{T}} \in \mathbb{F}$, we have

$$\begin{aligned} \langle f_1, f_2 \rangle_{\mathbb{F}} &= \sum_{r=1}^{q} \langle f_1^{(r)}, f_2^{(r)} \rangle_{L^2(\mathcal{T}_r)} = \sum_{r=1}^{q} \int_{\mathcal{T}_r} f_1^{(r)}(t_r) f_2^{(r)}(t_r) d\nu_{d_r}(t_r) \\ \|f_1\|_{\mathbb{F}} &= \sqrt{\langle f_1, f_1 \rangle_{\mathbb{F}}}. \end{aligned} \tag{2}$$

The random multivariate functional object F taking values in $\mathbb{F}$ is infinite dimensional, therefore, the aim for the analysis is to effectively reduce dimension of F by projecting it onto a space spanned by finitely many functional principal components, as is the typical goal in principal component analysis. Here, the first functional principal component (PC) should again identify the strongest and most important mode of variation in F while having norm one. Later PCs should define the most important mode of variation in the function subject to each mode being orthogonal to all previously defined modes.

To specify the space of interest more precisely, for $\mathbb{F}$ consider the Borel-σ-algebra and the Lebesgue measure $(\mathbb{F}, \mathcal{B}(\mathcal{T}), \nu)$. Together with the scalar product defined in Equation 2 this defines a Hilbert space as a direct sum of Hilbert spaces. Refer to Happ and Greven (2018), Hsing and

Eubank (2015) for L^2 being a Hilbert space.

In functional PCA one considers a stochastic process $F = \{F(t) : t \in \mathcal{T}\}$ which is also an element of a Hilbert space, that is $\mathbb{S} = \mathbb{F}$. Then, define the *covariance kernel* $\mathcal{K}$ and the mean function μ as

$$\mathcal{K}(s,t) \coloneqq \mathbb{C}\mathrm{ov}(F(s), F(t)) \quad \text{and } \mu(t) \coloneqq \mathbb{E}[F(t)], \qquad t, s \in \mathcal{T},$$

if the expectations exist. Stochastic processes with existing covariance kernel and mean function are also called *second-order processes* in Hsing and Eubank (2015). From now on assume that considered processes are of second-order.

Since we consider a Hilbert space, results of Chapter 3.2 hold. Specifically, the Hilbert-Schmidt-Theorem gives that an eigen decomposition of the covariance operator exists. For the case of $\mathbb{S} = \mathbb{F}$ it is equivalent to know the covariance operator and to know the covariance kernel. More precisely, it holds that the covariance operator coincides with the operator $\mathscr{K}$ defined for $f \in \mathbb{F}$, $t \in \mathcal{T}$ by

$$(\mathscr{K} f)(t) = \int_{\mathcal{T}} \mathcal{K}(t,s) f(s) d\nu(s). \tag{3}$$

Define a *mean-square continuous* spatial point process X as a process with the following property

$$\lim_{n\to\infty} \mathbb{E}[X(t_n) - X(t)]^2 = 0$$

for any $t \in \mathcal{T}$ and any sequence $\{t_n\} \subset \mathcal{T}$ converging to t. Mercer's theorem tells us that if F is mean-square continuous, an eigen sequence $\{(\beta_j, e_j)\}_{j=1}^{\infty}$ for the integral operator $\mathscr{K}$ also yields an eigen decomposition for $\mathcal{K}$, that is,

$$\mathcal{K}(s,t) = \sum_{j=1}^{\infty} \beta_j e_j(s) e_j(t)$$

and that the sequence converges absolutely and uniformly in $s, t \in \mathcal{T}$. Analogously to PCA in Hilbert spaces, the Karhunen-Loève decomposition follows for F, i.e.

$$F(t) = \mu(t) + \sum_{j=1}^{\infty} Z_j e_j(t), \quad \text{and } Z_j = \langle F - \mu, e_j \rangle_{\mathbb{F}} = \int_{\mathcal{T}} (F(t) - \mu(t)) e_j(t) d\nu(t).$$

Consequently, functional PCA is a way to perform inference on the properties of the covariance kernel. For the dimension reduction consider the truncated Karhunen-Loève decomposition

$$F(t) \approx F^{[M]}(t) = \mu(t) + \sum_{j=1}^{M} Z_j e_j(t), \quad t \in \mathcal{T}$$

$$F^{[M](r)}(t_r) = \mu^{(r)}(t_r) + \sum_{j=1}^{M} Z_j e_j^{(r)}(t_r), \quad t_r \in \mathcal{T}_r,\ r = 1, \ldots, q.$$

$M \in \mathbb{N}$ is chosen to explain most variability in F. Here, the same truncation parameter is chosen for every element $F^{(r)}$ of F. One could also consider different truncation lags M_r for every element, $r = 1, \ldots, q$.

Z_j are the functional PC scores. Again, they are uncorrelated random variables with mean zero and variance β_j. Finally, the M-dimensional vector $Z = (Z_1, \ldots, Z_M)^{\mathrm{T}}$ represents the functional

data in the most effective way in that corresponding functional PCs explain more variation than M elements of any other basis. It is important to note that the eigenfunctions are again only uniquely defined up to the sign.

One main difference between vector and functional PCA is the maximal number of principal components, that is, $\min(p, n-1)$ and ∞, respectively.

Now, the theoretical basis for multivariate functional PCA is established. It remains the question of how to estimate the decomposition. There are different estimation strategies depending on properties of the data, see Jang (2021) for a summary of FPCA estimation strategies. Here, consider the case of univariate functional data ($q = 1$) observed on a dense grid. Assume to observe $n \in \mathbb{N}$ independent copies of F denoted by $F_1, \ldots, F_n$ and corresponding realizations by $f_1, \ldots, f_n$. We consider the case of univariate functional data ($q = 1$) observed on a dense grid following Jang (2021). The mean and covariance kernel can be consistently estimated for $s, t \in \mathcal{T}$ via

$$\hat{\mu}(t) = \frac{1}{n} \sum_{i=1}^{n} f_i(t)$$
$$\widehat{\mathcal{K}}(s,t) = \frac{1}{n-1} \sum_{i=1}^{n} (f_i(s) - \hat{\mu}(s))(f_i(t) - \hat{\mu}(t)).$$

Approximating integrals numerically by a quadrature rule, eigenvalues and -functions can be estimated solving the eigen equation

$$\int_{\mathcal{T}} \widehat{\mathcal{K}}(s,t)\hat{e}_j d\nu(t) = \hat{\beta}_j \hat{e}_j(s).$$

By numerical integration estimation of the functional PC scores can be performed via

$$\widehat{Z}_{ij} = \int_{\mathcal{T}} (f_i(t) - \hat{\mu}(t))\hat{e}_j(t) d\nu(t).$$

For estimating multivariate functional data the method proposed in Happ and Greven (2018) is used. Here, FPCA components are estimated separately for each multivariate component and are combined in a way that produces consistent estimates for MFPCA components. This method is the cornerstone of the approach to hybrid MFPCA proposed by Jang (2021), which is explained in more detail in Chapter 3.4.2. Chapter 3.4 discusses the definition of hybrid data and the extension of the PCA concept to this type of data.

3.4 Hybrid Multivariate Functional Principal Component Analysis

For the derivation of hybrid PCA, one needs to establish a theoretical basis for modeling and analyzing functional and vector data simultaneously. The approach is to introduce a Hilbert space of hybrid data and define an appropriate scalar product. Then, one can define hybrid MFPCA.

The most straightforward idea is to use FPCA or MFPCA and standard PCA separately. Possible problems of this approach are multicollinearity due to possibly correlated scores and information loss since the maximal joint variance is not considered (Jang (2021)). Three alternative approaches are presented in the following chapters.

First, Ramsay and Silverman (1997) propose a method using pre-smoothing and standard PCA on vector data combining coefficients for basis functions and vector data (Chapter 3.4.1). Second, Jang (2021) proposes a method extending the approach by Happ and Greven (2018) performing separately PCA and MFPCA and then combing the results taking the correlation into account

(Chapter 3.4.2). Last, we present a method based on rewriting the vector data as step functions and then using the MFPCA method from Happ and Greven (2018) in Chapter 3.4.3.

First of all, hybrid data needs to be defined more precisely. Define the space of hybrid data $\mathbb{H} := \mathbb{F} \times \mathbb{R}^p$. Therefore, hybrid data is defined as random objects taking values both in a (multivariate) functional space and $\mathbb{R}^p$. In other words, objects in $\mathbb{H}$ have elements that are scalar as well as function valued. The goal is to analyze a random object $H = (F, V)^{\mathrm{T}} \in \mathbb{H}$ given independent $n \in \mathbb{N}$ realizations, denoted by lower case, i.e.

$$h_i = (f_i, v_i)^{\mathrm{T}} := (f_i^{(1)}, \ldots, f_i^{(q)}, v_i^{(1)}, \ldots, v_i^{(p)})^{\mathrm{T}} \in \mathbb{H},$$

for $i \in \{1, \ldots, n\}$. Recall that $f_i = (f_i^{(1)}, \ldots, f_i^{(q)})^{\mathrm{T}} \in \mathbb{F}$ and $v_i = (v_i^{(1)}, \ldots, v_i^{(p)})^{\mathrm{T}} \in \mathbb{R}^p$. For the extension of PCA for hybrid objects H, first consider the theoretical foundation. For using the results for Hilbert spaces (see page 19), an appropriate Hilbert space needs to be defined, that is, an appropriate scalar product for $\mathbb{H}$.

Consider the scalar product proposed by Ramsay and Silverman (1997): For $h_1 = (f_1, v_1)^{\mathrm{T}}$, $h_2 = (f_2, v_2)^{\mathrm{T}} \in \mathbb{H}$ define

$$\langle h_1, h_2 \rangle_{\mathbb{H}} := \langle v_1, v_2 \rangle_{\mathbb{R}^p} + \langle f_1, f_2 \rangle_{\mathbb{F}}, \tag{4}$$

where $\langle \cdot, \cdot \rangle_{\mathbb{R}^p}$ denotes the Euclidean scalar product and $\langle \cdot, \cdot \rangle_{\mathbb{F}}$ the scalar product defined for multivariate functional data. This scalar product induces the norm $\|h\|_{\mathbb{H}}^2 = \langle h, h \rangle_{\mathbb{H}}$ for $h \in \mathbb{H}$.

Since $\mathbb{H}$ is the direct sum of two separable Hilbert spaces, it is again a separable Hilbert space which contains the given Hilbert spaces as mutually orthogonal subspaces. Now, one can use the results for general Hilbert spaces: Given that for $H \in \mathbb{H}$ we have that $\mathbb{E}\|H\|_{\mathbb{H}} < \infty$, we define the mean function and covariance operator $\mathscr{K}$ as in Chapter 3.2 and have that the latter admits an eigen decomposition, i.e.

$$\mathscr{K} = \sum_{j=1}^{\infty} \beta_j (\xi_j \otimes \xi_j),$$

where $\{\beta_j\}_{j=1}^{\infty} \subset \mathbb{R}^p$ is the set of non-negative and non-decreasing eigenvalues and $\{\xi_j\}_{j=1}^{\infty} \subset \mathbb{H}$ the set of eigenobjects forming a complete orthonormal basis. This means that the eigenobjects satisfy $\langle \xi_j, \xi_l \rangle_{\mathbb{H}} = \delta_{jl}$. The word eigenobject is used here to refer to the elements that have the same purpose as eigenvectors in PCA and eigenfunctions in FPCA. Since they have the same structure as the data, the eigenobjects are elements of the Hilbert space and, therefore, consist of a vector and a multivariate functional part. Consequently, a typical principal component object consists of a pair $(e, w)^{\mathrm{T}}$ where $e \in \mathbb{F}$ is a multivariate function and $w \in \mathbb{R}^p$ a vector.

Analogously to previous PCA methods, this leads to the hybrid PC decomposition with probability one

$$H = \mu_H + \sum_{j=1}^{\infty} \rho_j \xi_j,$$

where μ_H denotes the expectation of H and the hybrid PC scores ρ_j are uncorrelated with zero mean and variances β_j. Analogously to the other PCA concepts, the first eigenobject explains the most important mode of variation and the subsequent eigenobjects explain the most important mode of the remaining variation.

The PC score of a particular observation $(f_i, v_i)^{\mathrm{T}}$ for principal component $(e, w)^{\mathrm{T}}$ is given by

$$\rho_i = \langle v_i, w \rangle_{\mathbb{R}^p} + \langle f_i, e \rangle_{L^2(\mathcal{T})} = v_i^{\mathrm{T}} w + \sum_{r=1}^{q} \int_{\mathcal{T}_r} f_i^{(r)}(t_r) e^{(r)}(t_r) d\nu(t_r)$$

and a typical observation from the distribution of the data could be modeled as

$$(f_i, v_i)^{\mathrm{T}} = \mu_H + \sum_{j=1}^{\infty} \rho_{ij} (e_j, w_j)^{\mathrm{T}},$$

where $(e_j, w_j)^{\mathrm{T}}$ is the j-th hybrid principal component. To find the leading principal component, find $\xi = (e, w)^{\mathrm{T}} \in \mathbb{H}$ maximizing the sample variance of the $\langle \xi, h_i \rangle_{\mathbb{H}}$ subject to $\|\xi\|_{\mathbb{H}} = 1$. Subsequent principal components should maximize the sample variance subject to the additional condition of orthogonality (defined by the hybrid scalar product from Equation 4) to the principal components already found. Principal components found this way yield principal component scores that are uncorrelated, just as for conventional multivariate PCA.

One problem of this method is that units of functional and vector data are not comparable and it might not be appropriate to weight them equally. One approach to solve this problem is to redefine the scalar product from Equation 4 by including a weighting factor, that is for some constant κ define

$$\langle h_1, h_2 \rangle_{\mathbb{H}} = \langle v_1, v_2 \rangle_{\mathbb{R}^p} + \kappa^2 \langle f_1, f_2 \rangle_{L^2(\mathcal{T})}.$$

Ramsay and Silverman (1997) and Jang (2021) give approaches for choosing κ and discuss those choices. In general, a sensible choice depends on the application. For the sake of clarity we choose $\kappa = 1$ in the following.

As for functional data, the difficulty now lies in estimating the theoretically derived eigen decomposition. In the following, three different approaches are described in more detail.

3.4.1 Approach using Pre-smoothing

The idea is to estimate the functional data by pre-smoothing. The approach is proposed in Ramsay and Silverman (1997) and is given for the univariate case for the functional part of the hybrid data, i.e. $q = 1$. There is a straightforward extension to the multivariate case, which is presented in the following.

For $r = 1, \ldots, q$ suppose that $\{\phi_1^{(r)}, \ldots, \phi_{M_r}^{(r)}\}$, $M_r \in \mathbb{N}$, is a set of functions that approximates the univariate functional part $f^{(r)}$ of a hybrid data object well. For any $h = (v, f)^{\mathrm{T}} \in \mathbb{H}$ and $M \coloneqq (M_1 + \ldots + M_q)$ define the vector $u = (u^{(1)}, \ldots, u^{(M)})^{\mathrm{T}} \in \mathbb{R}^M$ to be the combined vector of coefficients of $f^{(r)}$ relative to the functions $\phi_m^{(r)}$ for every $m = 1, \ldots, M_r$ and $r = 1, \ldots, q$. That is, it holds that

$$f^{(r)}(t_r) \approx \sum_{m=1}^{M_r} u^{(m + \sum_{i=1}^{r-1} M_i)} \phi_m^{(r)}(t_r), \quad t_r \in \mathcal{T}_r.$$

Consider the vector $a = (u, v)^{\mathrm{T}} \in \mathbb{R}^{M+q}$. Suppose the $\phi_m^{(r)}$ are orthonormal for each $r \in \{1, \ldots, q\}$, e.g. Fourier functions, then the inner product of any two hybrid elements $h_1, h_2 \in \mathbb{H}$ is approxi-

mately equal to the Euclidean scalar product of corresponding vectors a_1 and a_2, i.e.

$$\begin{aligned}\langle h_1, h_2\rangle_{\mathbb{H}} &\approx \sum_{r=1}^{q} \left\langle \sum_{m=1}^{M_r} u_1^{(m+\sum_{i=1}^{r-1} M_i)} \phi_m^{(r)}, \sum_{m=1}^{M_r} u_2^{(m+\sum_{i=1}^{r-1} M_i)} \phi_m^{(r)} \right\rangle_{L^2(\mathcal{T}_r)} + \langle v_1, v_2\rangle_{\mathbb{R}^p} \\ &= \langle a_1, a_2\rangle_{\mathbb{R}^{M+p}}.\end{aligned}$$

This yields a representation of the hybrid data as an $(M+q)$-dimensional multivariate object. Given an independent and identically distributed sample $h_1, \ldots, h_n$, $n \in \mathbb{N}$, of a random hybrid object H, for each component r of the multivariate functional part a set of functions $\{\phi_m^{(r)}\}_{m=1}^{M_r}$ can be chosen. However, we decide on one set of functions for all observations together. Then, the sample can be represented by vectors $a_1, \ldots, a_n \in \mathbb{R}^{M+q}$. Hence, the standard multivariate vector PCA method can be used. That is, find the eigenvalues and eigenvectors of the matrix $C_A = \frac{1}{n}\sum_{i=1}^{n} a_i a_i^{\mathrm{T}}$.

If w is any resulting eigenvector, the corresponding first M elements are the Fourier coefficients of the functional part resulting in eigenfunctions and the remaining elements are the vector part of the hybrid eigenobjects. Smoothing can be incorporated by adapting algorithms for functional smoothed principal components. The key idea is to define the roughness of hybrid data as the roughness of the functional part. For more details see Ramsay and Silverman (1997).

This method has two main drawbacks. First, difficulties arise when the method is applied to sparse data since basis coefficients cannot be estimated reliably. Second, the method is sensitive to the choice of the orthogonal basis system (Jang (2021)).

3.4.2 Principal Component Analysis of Hybrid Functional and Vector Data

The following approach provides a solution to the problems of the previous approach. The method is presented in Jang (2021) and called PCA of hybrid functional and vector data, in short HFV-PCA. It extends the MFPCA method of Happ and Greven (2018) to hybrid data.

First, the structure of hybrid data is discussed in more detail to be able to then derive the explicit forms of the covariance operator and the Karhunen-Loève expansion in the following. As discussed at the beginning of Chapter 3.4, hybrid data can be seen as random elements in the Hilbert space $\mathbb{H}$. To simplify estimation and asymptotic analysis, functional data can also be considered as sample paths of stochastic processes. For hybrid data this results in considering H as a multivariate stochastic process with respect to a multi-dimensional argument. Treat $\{H(t) : t \in \mathcal{T}\}$ as a $(q+p)$-dimensional stochastic process, that is, for $t = (t_1, \ldots, t_q)^{\mathrm{T}} \in \mathcal{T}$

$$H(t) = (F(t), V)^{\mathrm{T}} = (F^{(1)}(t_1), \ldots, F^{(q)}(t_q), V^{(1)}, \ldots, V^{(p)})^{\mathrm{T}}.$$

Only the first q elements, the functional part, depend on the argument t. This is different to how stochastic processes are usually defined. Going through the resulting changes in classic stochastic process analysis gives a better understanding of the method proposed in this chapter. For example, define the mean function as

$$\mu(t) \coloneqq \mathbb{E}[H(t)] = (\mathbb{E}[F(t)], \mathbb{E}[V])^{\mathrm{T}}.$$

Without loss of generality, assume μ to be constantly zero. Then, both the functional and the vector part are centered in advance. For defining the covariance operator, first, define the following three covariance kernels for the three different possible combinations of data modalities. For $s_u \in \mathcal{T}_u$,

$t_r \in \mathcal{T}_r$ with $u, r \in \{1, \ldots, q\}$ and $w, z \in \{1, \ldots, p\}$ define

$$\begin{aligned}
\mathcal{K}_F^{(ur)}(s_u, t_r) &:= \mathbb{C}\text{ov}(F^{(u)}(s_u), F^{(r)}(t_r)) \\
\mathcal{K}_{FV}^{(r)}(t_r, w) &:= \mathbb{C}\text{ov}(F^{(r)}(t_r), V^{(w)}) \\
\mathcal{K}_V(z, w) &:= \mathbb{C}\text{ov}(V^{(z)}, V^{(w)}).
\end{aligned}$$

Note that the kernels take values in different spaces corresponding to the combination of data modalities for which they are defined. Next, derive the explicit form of the covariance operator $\mathscr{K} : \mathbb{H} \longrightarrow \mathbb{H}$. As operator on the hybrid Hilbert space $\mathbb{H}$, the covariance operator has the following elements

$$(\mathscr{K}h)(t) = ((\mathscr{K}h)^{(1)}(t_1), \ldots, (\mathscr{K}h)^{(q)}(t_q), (\mathscr{K}h)_1, \ldots, (\mathscr{K}h)_p)^{\mathrm{T}} \in \mathbb{R}^{q+p}.$$

Derive functional and vector parts of the covariance kernel separately. Recall that in the functional case, the covariance operator of some function is equal to the integral of the covariance kernel multiplied with this function (Equation 3). This can be transferred to hybrid data, resulting in integrals over covariance kernels including functional parts and sums in the case of purely vector parts. More precisely, let $r \in \{1, \ldots, q\}$ and $w \in \{1, \ldots, p\}$, and $h = (f^{(1)}, \ldots, f^{(q)}, v^{(1)}, \ldots, v^{(p)})^{\mathrm{T}} \in \mathbb{H}$. Then write the functional element r and vector element w of the covariance operator as

$$\begin{aligned}
(\mathscr{K}h)^{(r)}(t_r) &= \sum_{u=1}^{q} \int_{\mathcal{T}_u} \mathcal{K}_F^{(ur)}(s_u, t_r) f^{(u)}(s_u) ds_u + \sum_{z=1}^{p} \mathcal{K}_{FV}^{(r)}(t_r, z) v^{(z)} \\
(\mathscr{K}h)_w &= \sum_{u=1}^{q} \int_{\mathcal{T}_u} \mathcal{K}_{FV}^{(u)}(s_u, w) f^{(u)}(s_u) ds_u + \sum_{z=1}^{p} \mathcal{K}_V(z, w) v^{(z)}.
\end{aligned}$$

With this specified, boundedness and continuity conditions can be derived so that $\mathscr{K}$ is a positive self-adjoint compact operator. Summarizing the conditions, every covariance kernel needs to be uniformly bounded in squared norm by a constant in every argument and the operators $\mathcal{K}_F$ and $\mathcal{K}_{FV}$ need to be uniformly continuous in the functional arguments.

Then, the Hilbert-Schmidt theorem can be applied and gives an eigen decomposition of the covariance operator analogously to previous PCA methods. That is, a complete orthonormal system of eigenobjects $\{\xi_j\}_{j=1}^{\infty}$ of $\mathscr{K}$ exists, each ξ_j consisting of a functional part ψ_j and a vector part θ_j, such that $(\mathscr{K}\xi_j)(t) = \beta_j \xi_j(t)$, where $\{\beta_j\}_{j=1}^{\infty}$ is the corresponding sequence of eigenvalues that converges to zero as j goes to infinity and $\beta_1 \geq \beta_2 \geq \ldots \geq 0$. This yields the eigen decomposition of the covariance operator, i.e. for $h \in \mathbb{H}$,

$$(\mathscr{K}h)(t) = \sum_{j=1}^{\infty} \beta_j (\xi_j \otimes \xi_j) h(t).$$

Subsequently, in Jang (2021) an elementwise Hilbert-Schmidt theorem is derived for future applications; refer to the paper for details.

In addition, an adaption of Mercer's theorem and the Karhunen-Loève theorem is proposed. Mercer's theorem gives the resulting decomposition of the covariance kernel. In the case of hybrid data, this means a decomposition of the covariance kernel for the functional part $\mathcal{K}_F$ and the vector

part $\mathcal{K}_V$. More precisely, for $r = 1, \ldots, q$ and $s_r, t_r \in \mathcal{T}_r$ it holds that

$$\mathcal{K}_F^{(rr)}(s_r, t_r) = \sum_{j=1}^{\infty} \beta_j \psi_j^{(r)}(s_r) \psi_j^{(r)}(t_r),$$

where the convergence is absolute and uniform. This means that the functional covariance kernel can be decomposed with the same set of eigenvalues and eigenfunctions being equal to the functional part of the set of eigenobjects. Analogously, the decomposition of the vector covariance kernel consists of the same eigenvalues and the vector part of the eigenobjects. That is, for $r = 1, \ldots, p$ we have that

$$\mathcal{K}_V(r, r) = \sum_{j=1}^{\infty} \beta_j \theta_{jr}^2,$$

where the convergence is absolute. The Karhunen-Loève theorem then gives the decomposition of the random object. To be precise, the hybrid random variable H can be written as

$$H(t) = \sum_{j=1}^{\infty} \rho_j \xi_j(t), \qquad t \in \mathcal{T},$$

where $\rho_j = \langle H, \xi_j \rangle_{\mathbb{H}}$ denote the hybrid scores which have mean zero, are uncorrelated and have variances that are equal to corresponding eigenvalues, i.e. $\mathbb{E}[\rho_j] = 0$ and $\mathbb{E}[\rho_j \rho_l] = \beta_j \delta_{jl}$. Moreover, we have

$$\lim_{L \to \infty} \sup_{t \in \mathcal{T}} \mathbb{E}\left[\left\| H(t) - \sum_{j=1}^{L} \rho_j \xi_j(t) \right\|^2 \right] = 0.$$

I.e., the truncated hybrid decomposition converges to the random object as the truncation lag L converges to infinity in L^2-norm.

The Karhunen-Loève expansion and hybrid PC decomposition have essentially the same form and interpretation as before. An advantage of the Karhunen-Loève expansion is that it is constructed in the more familiar space of mean-square continuous stochastic processes. This makes it easier to formulate necessary conditions on covariance kernels. In addition, the relationship between the hybrid decomposition and the standard functional and vector PC decomposition can be given which facilitates the development of estimation methods.

In practice, as before, the focus is on deriving the truncated PC decomposition for $L \in \mathbb{N}$, that is,

$$H^{[L]}(t) \coloneqq \sum_{j=1}^{L} \rho_j \xi_j(t), \qquad t \in \mathcal{T}.$$

The first L hybrid scores and hybrid PCs are optimal L-dimensional approximations of H in the sense that it minimizes the norm in $\mathbb{H}$ of the difference to a representation with any complete orthogonal system in $\mathbb{H}$.

Next, one approximates the functional and vector parts of a hybrid object $H = (F, V)^{\mathrm{T}}$ with

truncated functional and vector PC decompositions, respectively, that is,

$$\widetilde{H} \coloneqq (F^{[M]}, V^{[J]})^{\mathrm{T}} = \left(\sum_{j=1}^{M} Z_j e_j, \sum_{j=1}^{J} \gamma_j w_j \right)^{\mathrm{T}}. \tag{5}$$

Let $\widetilde{\mathscr{K}} \coloneqq \mathbb{E}[\widetilde{H} \otimes \widetilde{H}]$ denote the covariance operator for the approximated hybrid object. Analogously to the covariance operator of H, for $h = (f, v)^{\mathrm{T}} \in \mathbb{H}$ each element of $\widetilde{\mathscr{K}} h$ is given by

$$\begin{aligned}
(\widetilde{\mathscr{K}} h)^{(r)}(t_r) &= \sum_{u=1}^{q} \int_{\mathcal{T}_u} \widetilde{\mathcal{K}}_F^{(ur)}(s_u, t_r) f^{(u)}(s_u) ds_u + \sum_{z=1}^{p} \widetilde{\mathcal{K}}_{FV}^{(r)}(t_r, z) v^{(z)}, \\
(\widetilde{\mathscr{K}} h)_w &= \sum_{u=1}^{q} \int_{\mathcal{T}_u} \widetilde{\mathcal{K}}_{FV}^{(u)}(s_u, w) f^{(u)}(s_u) ds_u + \sum_{z=1}^{p} \widetilde{\mathcal{K}}_V(z, w) v^{(z)},
\end{aligned}$$

where $\widetilde{\mathcal{K}}_F, \widetilde{\mathcal{K}}_{FV}, \widetilde{\mathcal{K}}_V$ denote the covariance kernels to the corresponding truncated objects. Let $\tilde{\beta}_j, \tilde{\xi}_j$, $j \in \mathbb{N}$ denote the eigenvalues and eigenobjects of $\widetilde{\mathscr{K}}$. The eigen analysis yields the approximate truncated hybrid PC decomposition

$$\widetilde{H}^{[L]}(t) = \sum_{j=1}^{L} \tilde{\rho}_j \tilde{\xi}_j(t), \quad t \in \mathcal{T},$$

where $\tilde{\xi}_j$ and $\tilde{\rho}_j = \langle \tilde{H}, \tilde{\xi}_j \rangle_{\mathbb{H}}$ represent the approximated hybrid principal components and scores, respectively. Again scores are uncorrelated, mean zero and have variance $\tilde{\beta}_j$.

This leads to the analytical relationship between the decomposition of the vector part and the functional part of the hybrid data described in the following.

First, perform a standard PCA on the PC scores obtained. Suppose that truncated vector and functional PC decompositions are given as in Equation 5. Compile covariances within and between functional and vector PC scores in the following four matrices

$$\begin{aligned}
C_F &\coloneqq \{\mathbb{C}\mathrm{ov}(Z_h, Z_l)\}_{h=1,\dots,M}^{l=1,\dots,M} \in \mathbb{R}^{M \times M} \\
C_V &\coloneqq \{\mathbb{C}\mathrm{ov}(\gamma_h, \gamma_l)\}_{h=1,\dots,J}^{l=1,\dots,J} \in \mathbb{R}^{J \times J} \\
C_{FV} &\coloneqq \{\mathbb{C}\mathrm{ov}(Z_h, \gamma_l)\}_{h=1,\dots,M}^{l=1,\dots,J} \in \mathbb{R}^{M \times J} \\
C_{VF} &\coloneqq C_{FV}^{\mathrm{T}} \in \mathbb{R}^{J \times M}.
\end{aligned}$$

The $(M + J) \times (M + J)$-dimensional block matrix containing these matrices is again symmetric and positive semi-definite, so we can define

$$C \coloneqq \begin{bmatrix} C_F & C_{FV} \\ C_{VF} & C_V \end{bmatrix}. \tag{6}$$

In the next step, we perform eigen analysis of this matrix. Then, for $l = 1, \dots, L$ with $L \leq M + J$ the l-th eigenvalue of C is equal to the l-th eigenvalue of $\widetilde{\mathscr{K}}$ denoted by $\tilde{\beta}_l$. Denote the l-th $(M+J)$-dimensional eigenvector of C as $b_l = (c_l^{\mathrm{T}}, d_l^{\mathrm{T}})^{\mathrm{T}}$, where $c_l = (c_{l1}, \dots, c_{lM})^{\mathrm{T}}$ denotes the first M elements of b_l and $d_l = (d_{l1}, \dots, d_{lJ})^{\mathrm{T}}$ the last J elements of b_l. With the eigenvectors of C it is possible to write down the analytical relationship. Namely, for the l-th approximate hybrid PC

$\tilde{\xi}_l = (\tilde{\psi}_l, \tilde{\theta}_l)$ the functional parts $\tilde{\psi}$ and vector parts $\tilde{\theta}$ are given by

$$\tilde{\psi}_l(t) = \sum_{j=1}^{M} c_{lj} e_j(t) \quad \text{and} \quad \tilde{\theta}_l = \sum_{j=1}^{J} d_{lj} w_j,$$

respectively. I.e., for $\tilde{\psi}_l$, the functional part of an approximate hybrid PC consists of the functional PCs obtained by a separate MFPCA on the functional part where the correlation to the vector PCs is incorporated using the results from a standard PCA on the concatenated score vectors. The approximate hybrid PC score is given by

$$\tilde{\rho}_m = \sum_{h=1}^{L} Z_h c_{mh} + \sum_{j=1}^{J} \gamma_j d_{mj}.$$

In Jang (2021), it is proven that the approximate HFV-PCA components converge to the true components for each $l \in \mathbb{N}$ as $M \to \infty$ and $J \to p$. That is, as more functional and vector PCs are used to approximate H with $\tilde{H}$ the approximations converge to the true components. In other words, the analytical relationship holds in an asymptotic sense.

This analytical relationship can be used for deriving an estimation method. Consider a random sample $\{(f_1, v_1)^{\mathrm{T}}, \ldots, (f_n, v_n)^{\mathrm{T}}\}$ drawn independently from the same distribution $(F, V)^{\mathrm{T}}$. The proposed estimation method is the following.

1. Perform FPCA/MFPCA on the observed functional data $f_i, i = 1, \ldots, n$, to estimate functional PCs and PC scores, where the truncation lag M is chosen data-adaptively, e.g. by percentage of variance explained. Chose an appropriate method of existing MFPCA/FPCA methods depending on the data.

2. Perform classical PCA on the observed vector data $v_i, i = 1, \ldots, n$ to estimate vector PCs and PC scores, where the truncation lag J is chosen data-adaptively.

3. For each $i \in \{1, \ldots, n\}$ create a $(L+J)$-dimensional vector $\hat{\chi}_i$ that concatenates the functional and vector PC scores and compute the sample covariance matrix of the $\hat{\chi}_i$, that is,

$$\widehat{C} = \frac{1}{n-1} \sum_{i=1}^{n} \hat{\chi}_i \hat{\chi}_i^{\mathrm{T}},$$

   which estimates the matrix C given in Equation 6.

4. Perform an eigen analysis of $\widehat{C}$ to obtain eigenvalues $\hat{\beta}_l$ and orthonormal eigenvectors given by $\hat{b}_l = (\hat{c}_l^{\mathrm{T}}, \hat{d}_l^{\mathrm{T}})^{\mathrm{T}}$, $l = 1, \ldots, L$, $L \leq M + J$, where L is chosen data-adaptively and $\hat{c}_l$ and $\hat{d}_l$ denote the first M and last J elements of $\hat{b}_l$, respectively. Estimate the functional and vector parts of the l-th hybrid PC $\xi_l = (\psi_l, \theta_l)$ and the l-th hybrid PC score of the i-th observation ρ_{il} by

$$\hat{\psi}_l(t) = \sum_{h=1}^{M} \hat{c}_{lh} \hat{e}_h(t), \quad \hat{\theta}_l = \sum_{j=1}^{J} \hat{d}_{lj} \hat{w}_j \quad \text{and} \quad \hat{\rho}_{il} = \sum_{h=1}^{M} \widehat{Z}_{ih} \hat{c}_{lh} + \sum_{j=1}^{J} \hat{\gamma}_{ij} \hat{d}_{lj}, \tag{7}$$

   where $t \in \mathcal{T}$.

The choice of L, J and M is crucial and the general recommendation in Jang (2021) is to choose the lag so that 0.99 of the variance is explained.

For asymptotic results, the two common sources of errors are considered separately. On the one hand, there is the approximation error. This error accrues based on the fact that the truncated decomposition is estimated and is sometimes referred to as deterministic error. On the other hand, there is the estimation error which is the usual sampling error, also referred to as stochastic error. Due to the triangle inequality, one can look at both errors separately for asymptotic results. For example, for the eigenvalues β_j of the covariance operator, the eigenvalues $\tilde{\beta}_j$ of the truncated covariance vector and the estimated eigenvalues $\hat{\beta}_j$ it holds that

$$|\beta_j - \hat{\beta}_j| \leq |\beta_j - \tilde{\beta}_j| + |\tilde{\beta}_j - \hat{\beta}_j|.$$

The left hand side gives the total estimation error, that is the difference between the actual eigenvalue and its estimation. The first summand on the right side gives the approximation error and the second summand the estimation error. So the total estimation error can be upper bounded by the sum of the approximation and the estimation error.

To derive convergence rates for the estimation error, some assumptions are needed, e.g. to guarantee that eigenvalues have multiplicity one. For more details, refer to Jang (2021).

In this setting, the usual trade-off between approximation error and estimation error arises: For a fixed number of observations, increasing M reduces the approximation error but increases the estimation error. J does not affect the estimation error in an asymptotic sense since it is bounded by the dimension of the vector part p. Jang (2021) gives assumptions for consistency results as well as a simulation study demonstrating its improved performance compared to the pre-smoothing approach discussed in Chapter 3.4.1 in case of wrongly specified basis functions or sparse data.

3.4.3 Approach using Step Functions

As a version of the HFV-PCA approach proposed by Jang (2021), the idea of the approach proposed in this thesis is to write the vector data as step functions to get a purely functional multivariate object. This would enable us to use existing and already implemented MFPCA methods, in particular the MFPCA methods proposed in Happ and Greven (2018) and the functions in the R package `MFPCA`. In the following, we refer to this approach as step function approach.

Essentially, the idea of the step function approach is to transform the vector data into an object of just one data modality and then apply the corresponding PCA concept. In contrast to the pre-smoothing approach of Chapter 3.4.1, where the functional part of the hybrid data is transformed into a vector and then standard PCA is performed, in this approach the vector part is transformed into a function.

This way, instead of reducing the dimension of the higher dimensional part of the hybrid data, namely the functional part before performing PCA, the lower dimensional part, namely, the vector part, is embedded into the functional space. If one wishes to be able to transform the corresponding functions back to vectors, the transformation function needs to be bijective. Clearly, this means that the image of the transformation function has the same dimension as the domain, that is, the vector space which is finite dimensional. Maybe the most straightforward idea to represent vector data as functions would be to represent each entry of the vector as a constant function and, therefore, get a multivariate functional object representing the vector data consisting of constant functions. For a better comparison to previous methods consider a representation of the vector part in one function. The combination of functions that are constant on some interval yields a step function, which is the choice of representation.

To formalize this idea, consider the space of step functions of the following form

$$f(t) = \sum_{i=1}^{p} v^{(i)} \mathbb{1}_{(i-1,i]}(t), \quad t \in (0,p]$$

and $f(0) = v^{(1)}$ for some $v^{(1)}, \ldots, v^{(p)} \in \mathbb{R}$. Denote this space by $S([0,p])$. Obviously, it holds that $S([0,p]) \subset L^2([0,p])$, and we can use the scalar product on $L^2([0,p])$ for functions in this space. Let $f_1, f_2 \in S([0,p])$ be two functions with corresponding coefficients $v_1^{(i)}$ and $v_2^{(i)}$. Define $v_1 \coloneqq (v_1^{(1)}, \ldots, v_1^{(p)})^{\mathrm{T}}, v_2 \coloneqq (v_2^{(1)}, \ldots, v_2^{(p)})^{\mathrm{T}} \in \mathbb{R}^p$. Then, we have that the corresponding scalar products of f_1, f_2 and v_1, v_2 are equal. For more details see the appendix, Proof 1. Especially, it holds that induced distances coincide and completeness of $S([0,p])$ follows by completeness of $\mathbb{R}^p$. Consequently, $S([0,p])$ is a Hilbert space.

Now, for a random vector $V = (V^{(1)}, \ldots, V^{(p)})^{\mathrm{T}} \in \mathbb{R}^p$ define the corresponding random step function in $S([0,p])$ by

$$f^V(t) = \sum_{i=1}^{p} V^{(i)} \mathbb{1}_{(i-1,i]}(t), \quad t \in (0,p]$$

and $f^V(0) = V^{(1)}$. The covariance matrix of V is denoted by $C_V = \mathbb{E}[(V - \mathbb{E}[V])(V - \mathbb{E}[V])^{\mathrm{T}}]$ and the covariance operator of f^V is denoted by $\mathcal{K}_{f^V}(s,t) = \mathbb{C}\mathrm{ov}(f^V(s), f^V(t))$ for $s,t \in [0,p]$. With these notations the following assertion holds for the relationship between the standard PCA method for C_V and functional PCA for f^V:

$\{\beta_j, w_j\}_{j=1}^p$ are orthonormal eigenvalues and eigenvectors of C_V if and only if $\{\beta_j, e_j\}_{j=1}^p$ are orthonormal eigenvalues and eigenfunctions of $\mathcal{K}_{f^V}$, where

$$e_j(t) = \sum_{i=1}^{p} w_j^{(i)} \mathbb{1}_{(i-1,i]}(t), \quad t \in (0,p]\,.$$

For a proof of this assertion see Appendix, Proof 2. This means that functional PCA of the step functions in $S([0,p])$ is equivalent to standard PCA of the corresponding vector. The PC scores are also equal because of the relationship between the scalar products, that is,

$$\langle e_j, f^V \rangle_{L^2([0,p])} = \langle w_j, V \rangle_{\mathbb{R}^p}.$$

Finally, this yields the step function approach proposed in this thesis. Go through the following steps for hybrid data $(f_i^{(1)}, \ldots, f_i^{(q)}, v_i)^{\mathrm{T}}, i = 1, \ldots, n$.

1. Transform vector data $v_1, \ldots, v_n$ into step functions $f_1^v, \ldots, f_n^v$.
2. Apply MFPCA on multivariate functional objects $(f_1, f_1^v)^{\mathrm{T}}, \ldots, (f_n, f_n^v)^{\mathrm{T}}$.
3. Transform the components of the multivariate functional PCs corresponding to f^v back to the corresponding vectors to obtain hybrid PCs.

A possible problem of this method is that standard FPCA implementations assume that the functions are smooth, which is not the case for step functions at the interval boundaries. In addition, FPCA is computationally more complex than standard PCA which is unnecessary for the p-dimensional vector data.

Comparing the HVF-PCA method and the step function approach, they mostly differ in their motivation. HFV-PCA focuses on the hybrid aspect of the data and emphasizes its multi-modality. Here, it is more straightforward to extend the method including other data modalities even when these data modalities possibly cannot be written as a functional object. For this, it is necessary that one can define and estimate PCA on this type of data. The step function approach emphasizes the similarity of functional and vector data and gives a relationship between these data modalities instead of a relationship between the different PCA decompositions.

Otherwise, the estimation with the step function approach yields the same results as the HFV-PCA method. That is, for a sample $(f_i^{(1)}, \ldots, f_i^{(q)}, f_i^v)^{\mathrm{T}}$ of $(F, V)^{\mathrm{T}}$, $i = 1, \ldots, n$, and the corresponding sample $(f_i^{(1)}, \ldots, f_i^{(q)}, v_i)^{\mathrm{T}}$ the estimation results are the same. Note that the eigenfunctions and vectors are only unique up to the sign. Therefore, the equality of the results only holds when one adapts the signs accordingly.

Consequently, due to the lower computational cost the approach in Jang (2021) is to be preferred in practice. It is used for the estimation of hybrid MFPCA in the simulation study in Chapter 4 as well as in the application in Chapter 5.

4 Simulation Study

In this chapter the approach for analyzing multivariate spatial point processes using both functional and numerical summary characteristics and applying hybrid MFPCA is evaluated. Given a multivariate point pattern $\boldsymbol{\varphi} = (\varphi_1, \ldots, \varphi_n)$ consisting of $n \in \mathbb{N}$ component patterns, the approach consists of the following steps.

1. Calculate $q \in \mathbb{N}$ functional and $p \in \mathbb{N}$ numerical summary characteristics for each of the n different component point patterns. Consider each point pattern as an observation and the summary characteristics as variables. Get a resulting $(n \times p)$-matrix and for each component a q-dimensional functional object.

2. Perform hybrid MFPCA on this data to obtain hybrid MFPCA scores for each pattern.

3. Group the patterns using the first few scores obtained, e.g. with Ward's algorithm.

Before evaluating this method we need to determine which summary characteristics should be used. The following discussion is built on the results from the studies in Illian et al. (2006) and Illian et al. (2008). For the definition and discussion of the summary characteristics we refer to Chapter 2.2. For practicality, we hereafter assume a two dimensional point space, that, is $\mathbb{X} \subset \mathbb{R}^2$.

In Illian et al. (2008), the following numerical summary characteristics are used to analyze a natural plant community in the heath lands of Western Australia. In the initial analysis the number of points in each pattern was included. It turns out that this does not yield informative and clear results since this number often varies considerably between patterns. In addition, in Illian et al. (2008) it is argued that in the context of indices, scale invariant results are to prefer. The reason is that the analysis is focused on analyzing different spatial behavior of point patterns, i.e. if the patterns are clustered, regular or random, independent of the number of points. Consequently, summary characteristics that are scale-invariant are preferred and the total number of points as well as the intensity are not chosen for the analysis.

As a scale-invariant measure of aggregation the Clark-Evans index CEI is chosen. The characteristic SK measures a different aspect of the aggregation since it gives the skewness of the nearest-neighbor distribution and is therefore included as second scale-invariant numerical summary characteristic. The mean directional index is included as index based on directions instead of distances between points.

To include information based on second-order summary characteristics, numerical values derived from the pair correlation function are also included in the analysis. More specifically, $\hat{g}(2)$ and the minimal r with $\hat{g}(r) = 1$, denoted by $r^{(1)}$, are chosen. Here, the choice of exact values depends on the application and size of the observation window. Note that for the CEI a lower value indicates clustering whereas for all other characteristics a higher value generally indicates clustering. In the following, these summary characteristics are chosen for the approaches containing numerical summary characteristics.

Concerning functional summary characteristics, Illian et al. (2006) provides an analysis of which characteristic should be used for the analysis of multivariate point processes. In various settings, the pair correlation function yields better results than the L-function. This could be explained due to the fact that the pair correlation function is not of the cumulative nature as the L-function. This means that spatial behavior for higher r values is better represented. However, the focus of the analysis including functional summary characteristics is restricted to those two functions. It

is not discussed how functions such as the nearest-neighbor distance function perform. The choice of characteristics is not straightforward. A thorough evaluation of different methods in a variety of settings is necessary. However, this is out of scope of this thesis. Since the type of application in Illian et al. (2008) is similar to the data discussed in the application of this thesis, the pair correlation function is used in the following as functional summary characteristic. One could include more functions in the analysis but for better understanding only one functional part is included in this thesis ($q = 1$).

For the evaluation of the different approaches we consider the following two settings. To begin with, in Chapter 4.1 we compare the effect on the analysis using either one functional or multiple numerical summary characteristics or both on a set of clearly clustered, regular and random point patterns similarly to the feasibility study in Illian et al. (2006). In Chapter 4.2 the effect of dependence between patterns on the hybrid approach is analyzed.

There is a large number of point process models which can be used to simulate point patterns that have certain properties. A discussion is out of scope of this thesis, we refer to one of the references for general point processes such as Illian et al. (2008) or Baddeley et al. (2016) for details.

4.1 Clustered, Regular and Random Patterns

First, we consider an example to evaluate the feasibility of the approaches. This chapter is structured as follows. We generate point patterns which can be categorized in clustered, regular and random point patterns. Then, we perform PCA on numerical summary characteristics to be able to group the patters based on the first few PC scores. This is followed by performing FPCA on one functional summary characteristic. Lastly, both types of summary characteristics are analyzed simultaneously by performing hybrid MFPCA. To compare the approaches, we evaluate the misclassification percentage.

We start with the generation of the component patterns. Consider the unit square as the observation window, that is set $W = [0, 1] \times [0, 1]$. For representing complete spatial randomness, 20 independent point patterns are generated from a Poisson process with intensity 100 using the R function `rpoispp` from the package `spatstat` (Baddeley et al. (2016)). R functions referred to throughout the rest of this thesis are from this package unless stated otherwise.

For regular point patterns, the R function `rMaternII` is used to generate 20 independent patterns of a Matérn model II inhibition process. It is one of two point process models introduced in Matérn (1960) and Matérn (1986). We refer to these publications for more details on the theory. For patterns of this kind, first, proposal points are generated from a Poisson point process. For two points that are closer than some inhibition distance, one of the points is deleted. To decide which of the two points should be deleted, all points are marked with a number uniformly distributed in $[0, 1]$. Then, all points which lie closer than the inhibition distance to a point in the pattern with a lower mark value are deleted. Here, the intensity of the Poisson process of proposal points is chosen as 200 and the inhibition distance as 0.05, giving the degree of regularity.

A Thomas cluster process is chosen to represent aggregation. For this process, parent points are generated from a Poisson process with some intensity that needs to be specified. Then, for each parent point, a random cluster of offspring points is generated where the number of points in a cluster is Poisson distributed and the positions are isotropic Gaussian displacements from the cluster parent location. Then, the pattern consisting of the offsprings is a Thomas cluster point pattern.

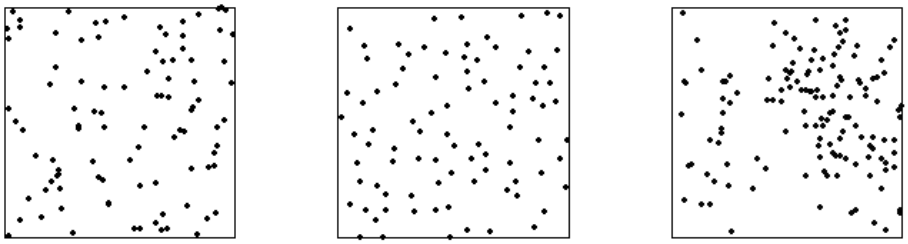

Figure 1: One pattern of each group of patterns generated in this case setting, a Poisson process (left plot), a Matérn model II inhibition process (middle plot) and a Thomas cluster process (right plot).

For this example, 20 patterns are drawn using the corresponding R function `rThomas`. As intensity for the Poisson process of cluster centers the value 10 is chosen, for the random displacement 0.1 and 10 as the mean number of points per cluster. We refer to Thomas (1949) for more details.

Choosing the parameters in this way yields point patterns containing about 100 points per pattern and results in patterns that can be classified into regular, clustered and random. All 60 patterns together can then be considered as a multivariate point pattern. In Figure 1 for each type of spatial behavior one generated point pattern is depicted as a representative. The left plot is a point pattern representing a random pattern. There are points which are very close to each other and points with some distance to the nearest neighbor. In general, the points are spread over the whole observation window with about the same number of points in each part of the window.

In contrast, in the pattern depicted in the middle the distance between points appears to be approximately the same over the observation window and the points are homogeneously spread. This pattern is an example of a regular pattern.

In the right plot, a point pattern generated from the above specified cluster process is depicted. Points are generally closer to their nearest neighbor. In addition, there are some regions in the observation window without points, for example the lower middle part.

In the next step, we calculate for each point pattern the numerical summary characteristics discussed in the beginning of this chapter. That is, determine the Clark-Evans index CEI, the mean directional index $\overline{R}$, the index SK, $\hat{g}(0.02)$ and $r^{(1)}$.

For the CEI the function `clarkevans` with the cumulative distribution function method is used. This means that the estimation is based on an estimation of the nearest neighbor distance distribution function. The mean directional index is calculated without border correction via the Equation 1 in Chapter 2.2. This means, the estimator is biased. The average value of the estimator for patterns generated from a Poisson process is close to 1.96 instead of the theoretical value of 1.799. However, for clustered patterns the estimator still gives larger values and lower values for more regular patterns.

For the estimator $\hat{g}$ of the pair correlation function the isotropic border corrections of the function `pcf` is used as in Illian et al. (2006). This means, that for points in the pattern close to the border a weight is introduced based on the proportion of the ball around the point that is inside the observation window. The value $r = 0.02$ for the characteristic $\hat{g}(0.02)$ was chosen since for lower r values the pair correlation function is either extremely high or zero for many patterns. For the variable $r^{(1)}$ the values of the estimated pair correlation function are rounded to two decimal places. Since the functions are estimated on a grid, the function values may otherwise cross the value one without attaining it. Especially for clustered patterns it may occur that the

Point pattern	CEI	$\overline{R}$	SK	$\hat{g}(0.02)$	$r^{(1)}$
Random	1.01	1.96	1.10	1.07	0.03
Regular	1.26	1.75	1.03	0.00	0.06
Clustered	0.88	1.99	1.10	1.40	0.25

Table 1: Numerical summary characteristics for the three patterns depicted in Figure 1, that is, generated from a Poisson process (Random), a Matérn model II inhibition process (Regular) and a Thomas cluster process (Clustered).

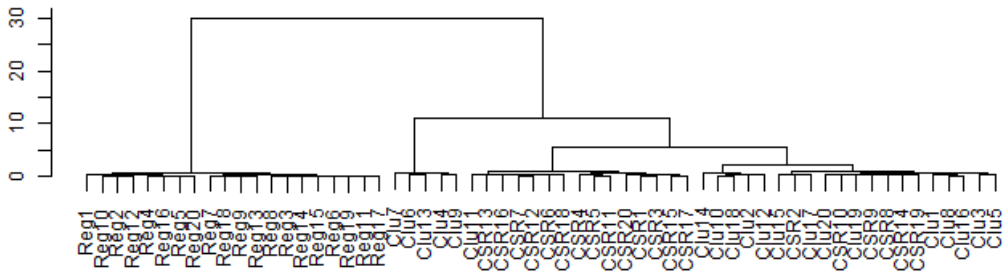

Figure 2: Dendrogram resulting from applying Ward's algorithm on the numerical summary characteristics of the patterns directly. The prefix Reg in the nodes indicates a regular pattern, Clu a clustered and CSR a random pattern.

pair correlation function does not attain value one at all. In these cases, $r^{(1)}$ is set to the maximal r value for which the pair correlation function is estimated; in this example it is a value of 0.25.

Table 1 contains the corresponding numerical summary characteristics for the patterns depicted in Figure 1. The theoretical value of the Clark-Evans index (CEI) for a Poisson process is equal to one. A value less than one indicates clustering and a value larger than one regularity. This matches with the estimated values for the patterns, i.e a CEI of 1.01 for the pattern that is generated from a Poisson process (Random), a CEI of 1.26 for the pattern generated from a Matérn model II inhibition process (Regular) and 0.88 for the pattern generated from a Thomas cluster process (Clustered).

For the mean directional index $\overline{R}$ both the random and the clustered pattern have values of approximately 1.96; though the value for the clustered pattern is slightly higher. For the regular pattern $\overline{R}$ has a lower value.

The SK value is close to one for all patterns indicating some kind of regularity in contrast to irregularity. For the regular pattern the value is the closest to one.

The values for $\hat{g}(0.02)$ and $r^{(1)}$ reflect the theoretical properties of the processes since the pair correlation function is constantly one for the Poisson process, less than one in case of regularity and greater than one for clustered processes. This fits with the observation that $r^{(1)}$ is the smallest for the random pattern. In contrast, for the clustered pattern a value of 0.25 is estimated. Recall that this was the maximal r and indicates that the corresponding estimated pair correlation function does not attain the value one. In Figure 5 the pair correlation functions for the three example patterns are plotted for reference.

The idea proposed by Illian et al. (2008) is to perform multivariate data analysis on the data

Summary Characteristic	1PC	2PC	3PC
CEI	0.49	-0.30	0.17
$\overline{R}$	-0.46	0.42	-0.28
SK	-0.42	-0.24	0.83
$\hat{g}(0.02)$	-0.50	0.08	0.02
$r^{(1)}$	-0.35	-0.82	-0.45

Table 2: Loadings of the numerical summary characteristics on the first three principal components.

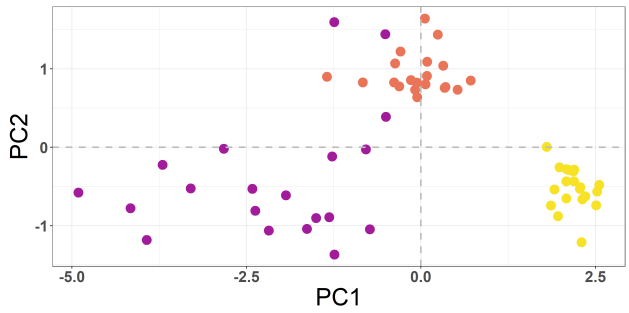

Figure 3: First two PC scores for all patterns resulting from standard PCA on the numerical summary characteristics. The colors indicate the type of the pattern; regular patterns in yellow, clustered in violet and random in red. The gray lines indicate the x- and y-axis.

obtained by calculating the numerical summary characteristic for each pattern. Since the goal is to group patterns based on their spatial behavior, one could apply cluster analysis approaches directly on the numerical summary characteristics. Applying e.g Ward's algorithm yields the dendrogram pictured in Figure 2. Clearly, the algorithm is able to group the regular patterns. They appear to be distinguishable in this setting from clustered or random patterns. If the patterns are grouped into three groups based on this dendrogram, one group consists of all regular patterns, indicated by the prefix Reg. The second group is quite small and only consists of clustered patterns. However, the third group is large, containing both clustered and random patterns.

To be able to use the information in the numerical summary characteristics more effectively, an alternative idea would be to search for the main directions of variation in the data. For this the classical approach is to perform standard PCA. Consequently, in the next step we perform PCA on the values for the CEI, $\overline{R}$, SK, $\hat{g}(0.02)$ and $r^{(1)}$ for each of the generated point patterns. Whenever standard PCA on vectors is performed in this thesis, the correlation matrix is used instead of the covariance matrix due to possibly different scales in the variables.

The first three principal components explain about 97% of the variation of the data and are considered in more detail. The loadings are given in Table 2. The first principal component consists of about the same amount of each variable. The CEI has a different direction than the other four variables. This is consistent with the results in Illian et al. (2008), where all variables besides the CEI are negatively associated with the first PC, but no variable dominates it. This makes sense since the CEI behaves contrary to all other variables for clustered or regular patterns. No variable dominates the first PC. This can be interpreted as being the result of informative summary characteristics that all explain variation between the spatial behavior of the patterns.

A large negative first PC score can be interpreted as a tendency towards clustering, since e.g. the CEI is less than the mean value over all patterns, which indicates clustering. Being above average for the other four variables also indicates clustering. It turns out that all clustered patterns have a negative first score. For random patterns the first scores are close to zero and for all regular

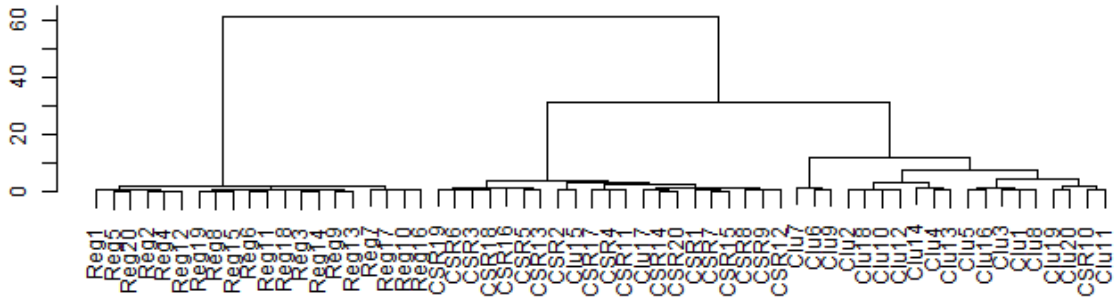

Figure 4: Dendrogram illustrating the clustering of the point patterns based on the first three PC scores from performing PCA on numerical summary characteristics.

patterns positive first score values are observed. The first two PC scores are plotted against each other in Figure 3.

The second PC is most strongly correlated with $r^{(1)}$. A large negative score for the second PC can be interpreted as a pattern for which the value one is observed later than the average estimated pair correlation function. This indicates a higher degree of clustering for larger r values. For $\hat{g}(0.02)$ the loading is rather low; $\hat{g}(0.02)$ was the variable most strongly associated with the first PC. The third PC is dominated by SK which measures the irregularity of a pattern and describes a different aspect of the distribution of a point pattern than the other four variables.

Applying Ward's algorithm to the first three scores for each point pattern yields the dendrogram pictured in Figure 4. Since Ward's algorithm is based on the Euclidean distance between points, clustering is based on the results one could already see in the score plot. The groupings are very close to true separation into clustered, regular and random patterns. Not only are the regular patterns clearly distinguished from the other patterns, but also the clustered and the random patterns are mostly divided in two different clusters. Only two clustered patterns (Clu15 and Clu17) are put in the group of random patterns and one random pattern is grouped within the clustered patterns (CSR10). Looking at the plot of the scores in Figure 3, these two patterns correspond to the two violet points on the left above the cloud of red points. Consequently, applying PCA first on the patterns yielded an improvement in the grouping of the patterns in contrast to clustering directly based on the numerical summary characteristics. It also gives a better understanding of how the spatial behavior of the patterns differs. Specifically, it gives a first impression on the distribution of the generated multivariate spatial point pattern.

In the next step, consider the approach using only one functional summary characteristic. Use the pair correlation function as discussed above. In Figure 5 the pair correlation for the three patterns is plotted for r values below 0.15 and pair correlation function values below 5. The yellow line gives the function for the regular pattern. As expected for low r values the pair correlation function is zero. Only right before the value 0.05 it takes positive values and after 0.05 stays around 0.05. This fits the expectations since the pattern was generated with inhibition distance 0.05. The points of the Matérn II inhibition process are generated from thinning a Poisson pattern. Thus, for larger r values the pair correlation function moves around value one. The estimated pair correlation function for the example random pattern (green line) moves around value one as expected, starting

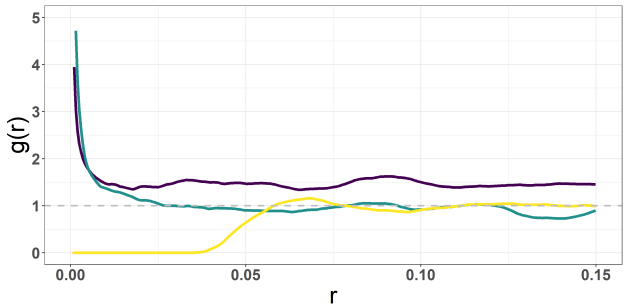

Figure 5: The pair correlation function for the three patterns in Figure 1. The yellow line corresponds to the regular pattern, the dark blue to the clustered pattern and the lighter blue line to the random pattern. The dotted gray line indicates the value one.

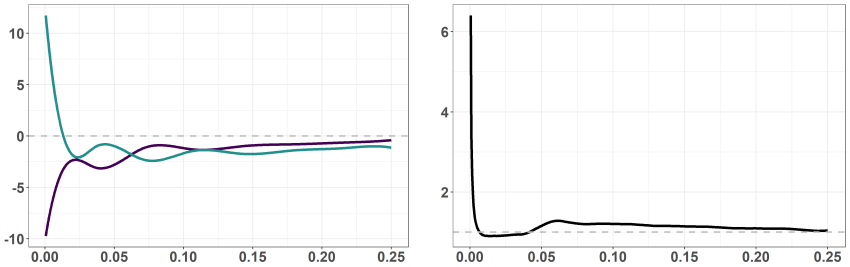

Figure 6: Left plot: First two functional principal components. The darker blue line indicates the first principal component of FPCA on the pair correlation functions of the patterns, the lighter blue line the second functional PC, and the gray dotted line indicates the zero. Right plot: Estimated mean function over all patterns (black line), the gray line indicating the value one.

from around $r = 0.025$. The theoretical pair correlation function is constantly one from $r = 0$. However, since the pair correlation function is estimated from one pattern, deviations may appear, which explains the peak for low values. For the clustered pattern (dark violet line) the estimated pair correlation function is larger than one for all r values but especially for very small values the function takes very high values.

Applying FPCA on the pair correlation functions yields the following results. When estimating functional principal components in this thesis the function `MFPCA` of the R package `MFPCA` (Happ-Kurz (2020)) is used. As recommended in Illian et al. (2006) B-splines are used to smooth the pair correlation functions. Here, the first four functional PCs are estimated and the first two components account for more than 90% of the variance explained. Therefore, only the first two PCs are considered for the analysis of the point patterns. In the left panel of Figure 6 the two functions are depicted. The lighter blue line gives the first functional PC. High positive score values of this component mean that the observation has a higher pair correlation for r values in the interval from zero to around 0.25 than average, but for later r values the pair correlation function has lower values than the mean function of the pair correlation functions. In contrast, high positive scores on the second PC indicate a lower than average pair correlation function over the whole domain; but the difference is more extreme for lower r values. In the right panel the estimated mean function over all patterns is depicted. For very small r clustered behavior seems to dominate since the mean function has high values. In the interval from around 0.01 to 0.05 the regular patterns dominate since the mean function is less than one. Then again, the mean function rises above value one and slowly decreases to one.

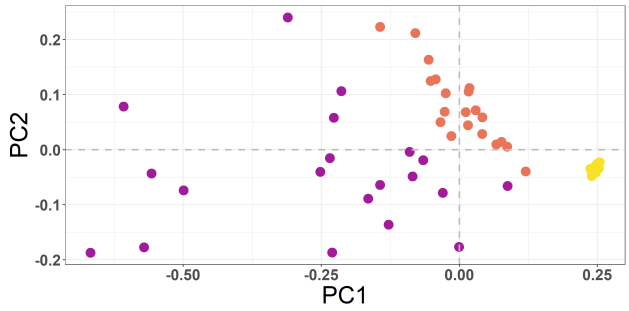

Figure 7: First two functional PC scores for each pattern resulting from a FPCA of the estimated pair correlation functions. The violet points are the score values from clustered patterns, the yellow points from the regular patterns and the red points from the random patterns. The dotted gray lines indicate the coordinate axes

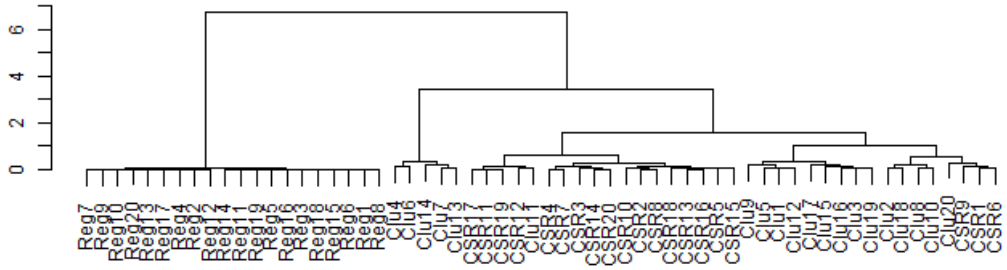

Figure 8: Dendrogram based on Ward's algorithm on the first two functional PC scores of the FPCA on the pair correlation function.

Consider Figure 7 for the analysis of the functional PC scores. All regular patterns (yellow points) have very similar first two scores, where the first score has a positive sign and the second score a negative sign. Looking at the PC functions this means that for low r values the patterns have pair correlation functions below the average. For higher r values the PC functions have opposing influences. The absolute value of the score value for the first PC is higher for all regular patterns and the sign is positive. This means that the patterns have lower function values than the mean function. Looking at the mean function which is depicted in the right panel of Figure 6, this is consistent with the observation that the pair correlation function is approximately one for regular patterns, which is lower than the mean function for most higher values.

The clustered and random patterns (CSR) have scores that are more spread over the space. The random patterns mostly have a positive score for the second principal component, but the sign of the first principal component is about equally often positive as it is negative. For these patterns, the higher the score for the second PC the lower is the score for the first PC, and vice versa. For the clustered patterns both score values are mostly negative though there is great variation in the score vectors compared to the other two point pattern types.

The three pattern types are clearly separable based on their first two scores. However, the variability in the scores differs greatly between the patterns and the scores for the clustered patterns make it difficult for clustering algorithms to detect the right groups. Using Ward's algorithm again on these scores yields the dendrogram in Figure 8. As expected, the regular patterns are grouped in one cluster. Five clustered patterns have comparably extreme scores in Figure 7 in the lower

	PC1	PC2	PC3
F1	0.70	0.11	-0.00
F2	0.11	-0.70	-0.00
V1	0.70	0.02	0.15
V2	0.08	-0.60	-0.52
V3	0.08	0.37	-0.84

Table 3: First three principal components of the PCA on the concatenated score vectors. F1 and F2 are the first two scores from the FPCA on the pair correlation function; V1, V2 and V3 are the first three scores for the PCA on the numerical summary characteristics.

left corner. When grouping the patterns in three groups, these five patterns build one group. This yields a third group that consists of clustered and regular patterns. However, looking at this group in more detail one can see that in this group the clustered and the regular patterns are separated with exception of three random patterns (CSR1, CSR6 and CSR9) that are among the clustered patterns and the one clustered pattern (Clu11) that is among the random patterns. Consequently, when dividing the patterns into four groups, the different types of patterns are mostly in different groups and only two score values were considered. Nevertheless, the approach using the numerical summary characteristics which also included information on the pair correlation function was able to find the right grouping within three groups. For comparison, the amount of misclassified patterns is calculated when dividing the patterns into three groups (see Table 5).

Finally, we investigate the approach using both functional and numerical summary characteristics and performing hybrid MFPCA on the data with the approach proposed by Jang (2021). See Chapter 3.4.2 for details. For this, the numerical and (one) functional summary characteristics are used as above, that is, we consider for each pattern the numerical summary characteristics CEI, $\overline{R}$, SK, $\hat{g}(0.02)$, $r^{(1)}$ as well as the estimated pair correlation function $\hat{g}$. Obviously, the information of the last two numerical summary characteristics is also included in the pair correlation function. However, they are kept in the analysis as numerical summary characteristics for better comparison to the above results.

For hybrid MFPCA the first step is to perform standard PCA and FPCA on each component, respectively. This was already done in the previous approaches and the results are discussed in detail. The truncation lags for both principal component analyses are chosen as above. For the numerical part the first three scores are considered, that is, $J = 3$, and for the functional part the first two scores ($M = 2$). The next step is to concatenate the score vectors to one vector which then consists of five entries for each pattern. On the score vectors, PCA is performed again.

The first three PCs explain 90% of the variation of the score vectors and are given in Table 3. The variability in the first PC scores for the numerical (V1) and the functional part (F1) is the highest and they dominate the first PC of the score vector data to an equal amount. The second scores of both the functional (F2) and the vector part (V2) dominate the second PC; the third PC of the score data is dominated by the third PC of the numerical part (V3). This is not surprising since the first principal component explains most variability of the data. With the PCs from the score vectors hybrid PCs and PC scores can be estimated; refer to Chapter 3.4.2, Equation 7 for the corresponding formulas. Each hybrid PC in this case consists of a vector of dimension five and one function, corresponding to the five summary characteristics and the one functional summary characteristic, since principal components have the same structure as the data.

The numerical part of the first hybrid PC is given in Table 4. Since it is dominated by the first

Summary characteristic	PC1	PC2	PC3
CEI	0.33	0.26	0.09
$\overline{R}$	-0.31	-0.37	-0.06
SK	-0.25	0.44	-0.64
$\hat{g}(0.02)$	-0.34	-0.05	-0.14
$r^{(1)}$	-0.34	0.32	0.75

Table 4: Numerical part of the first three hybrid PCs.

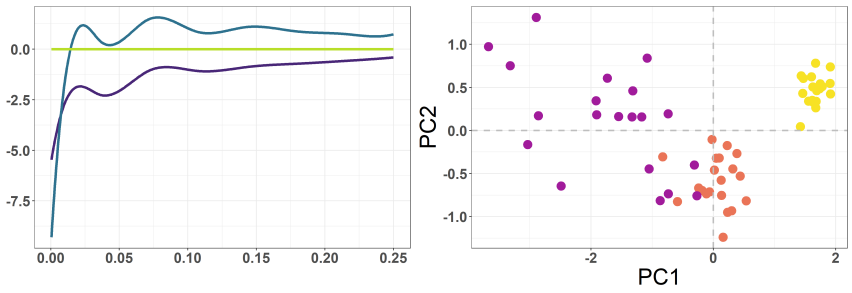

Figure 9: Left panel: Estimated functional part of the first three hybrid PCs; first PC in dark blue, second PC in lighter blue and third PC in green. Right panel: Estimated hybrid scores for every pattern, colors indicating the type of pattern: regular (yellow), random (red) or clustered (violet). The dotted gray lines indicate the coordinate axes.

PC from the PCA on numerical summary characteristics (Table 1), it consists about equally of all numerical summary characteristics and the CEI is associated with the first hybrid PC contrary to all other variables. Analogously, the numerical parts of the following hybrid PCs are similarly structured as the corresponding PCs of the standard PCA on the numerical summary characteristics. Recall that each PC vector is only unique up to the sign. For example, in the numerical part of the second hybrid PC the signs are all reversed in comparison to the second numerical PC.

The functional parts of the first two hybrid PCs are depicted in the left panel of Figure 9. They consist mostly of the first two functional PCs of the FPCA on the pair correlation functions; besides that the signs of the second PC are swapped as already indicated in the numerical part of the hybrid PCs. The darkest line gives the functional part of the first hybrid PC, the lighter blue line for the second hybrid PC and the green line for the third hybrid PC. Since only two functional PCs were included ($M = 2$), it is not surprising that the third functional part is constantly zero.

For interpreting the hybrid PCs the vector and functional part of the PCs need to be interpreted jointly. For example, if a pattern has a high positive first hybrid PC score, this means that the pattern has a behavior described by an higher than average CEI, a lower than average $\overline{R}$, SK, $\hat{g}(0.02)$ and $r^{(1)}$ and a pair correlation function that is lower than the mean function of pair correlation functions in the sample but with less difference to the mean function as the distance between points increases. All these properties are typical for regular patterns. In the right panel of Figure 9 the first two hybrid scores are given for each pattern; the different colors indicating again from which type of point process the patterns were generated. Here, one can see that the observed first PC scores fit to the interpretation since all regular patterns have a positive first hybrid PC score, all clustered patterns have a negative first PC score and within the random patterns both signs appear.

Since hybrid scores are given by the weighted scores of the numerical and the functional PC scores, the plot also looks like a combination of the two previous score plots. For the regular patterns the scores are clustered close to $(1.7, 0.5)^{\mathrm{T}}$. The hybrid scores for the random patterns are not as clustered but all lie around $(0, -0.5)^{\mathrm{T}}$. For the hybrid scores of clustered patterns it holds

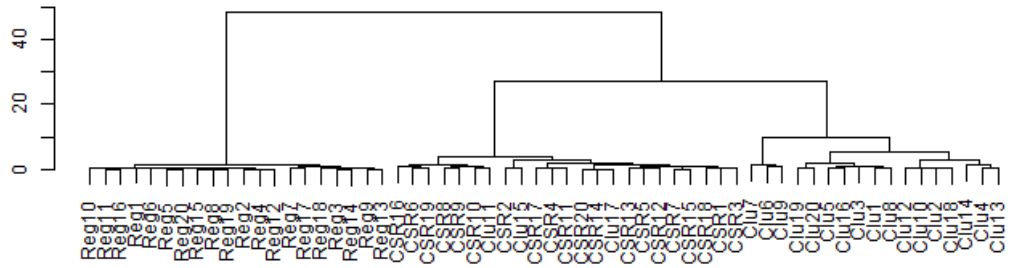

Figure 10: Dendrogram based on Ward's algorithm on the first two hybrid PC scores of the hybrid MFPCA on the pair correlation function and the numerical summary characteristics.

Approach	Percentage of misclassified patterns
Numerical	0.05
Clustering	0.25
Functional	0.25
Hybrid	0.05

Table 5: Percentage of misclassification for the approaches based on numerical summary characteristics (Clustering), performing PCA on the numerical summary characteristics (Numerical), preforming FPCA on the pair correlation function as functional summary characteristics (Functional) and the combination of both (Hybrid).

again that the scores are spread the most over space. Properties of both the numerical and the functional scores are visible; this indicates that the hybrid scores are not dominated by one data modality.

The dendrogram based on Ward's algorithm on the first three hybrid scores in Figure 10 confirms this observation. All the regular patterns and the random patterns are grouped together each in a cluster. As was already indicated in the score plot, there are three clustered patterns that are grouped with the random patterns. All three of those patterns were grouped with random patterns in either of the first two approaches, that is, based on either only numerical or functional summary characteristics. This might indicate a tendency towards random behavior in these patterns, which can be the case since the patterns are single realizations of a process. It also shows that mistakes made by either the functional or the numerical approach may be passed on to the hybrid approach since it is a weighted result of the single approaches. However, for twelve patterns the hybrid approach classifies patterns correctly, though only in one of the functional and the numerical approaches these patterns are correctly grouped. This indicates that the hybrid approach seems to combine the information from both approaches in a way that is desired rather than accumulating mistakes.

For measuring the ability of grouping the patterns and comparing the different approaches, one can consider the percentage of misclassified patterns. For this, in each case the patterns are grouped into three clusters based on the given dendrogram. Then, each group is assigned to the type that dominated the group. Finally, consider the proportion of patterns that are in the group corresponding to the process from which they were generated. The results are summarized in Table 5.

Summarizing, all three approaches based on the principal component analysis of summary characteristics perform quite well in detecting the spatial behavior of point patterns in the case of

patterns generated from processes that are random, clustered or regular. However, in practice the point process underlying the data might be very complex and not specifically assigned to one type of point process. The principal component analysis of summary characteristics, both numerical and functional, yields an approach of describing this distribution in more detail. The extension to the hybrid approach has the advantage of being able to consider both kinds of summary characteristics simultaneously to find out which spatial behavior explains most of the variation in the data.

One important point in principal component analysis that should be discussed is scaling. First, the scales within one data modality could differ. For vector data the standard approach is to normalize the vectors and perform eigen analysis of the correlation matrix instead of the covariance matrix. This was done in the analysis above. Since only a one-dimensional functional object was considered there is no need to correct for variations in the scales of different functions.

Second, it should be noted that the scalar product for hybrid data was chosen arbitrarily. That is, it is not straightforward that the scalar product of the vector part should have the same weight as the scalar product of the function part. For example, if they have completely different scales, one part might dominate the scalar product and the successive analysis. To deal with this problem there are multiple ways for defining an appropriate weighting factor. To incorporate a weighted scalar product in the estimation procedure one can transform the data in the first step accordingly. More precisely, for weight κ^2, consider the objects $(\kappa f_i, v_i)^{\mathrm{T}}$ instead of $(f_i, v_i)^{\mathrm{T}}$ for $i \in \{1, \ldots, n\}$. Jang (2021) recommends to use a data-driven approach analogously to standardization in PCA and utilize

$$\kappa^2 = \frac{\sum_{i=1}^{n} \|v_i - \hat{\mu}_v\|_{\mathbb{R}^p}^2}{\sum_{i=1}^{n} \|f_i - \hat{\mu}_f\|_{\mathbb{F}}^2}, \tag{8}$$

where $\hat{\mu}_f$ and $\hat{\mu}_v$ denote the estimated mean function and the estimated mean vector of the data.

However, in the simulated data set of this chapter the numerical and functional characteristics have similar scales, and using the weight defined in Equation 8 does not change the results considerably. The discussion of other weighting factors is neglected in this thesis. In the following analyses, the scalar product with $\kappa = 1$ is used unless otherwise stated.

In the next chapter we consider how dependence between component patterns effects the analysis of multivariate spatial point patterns.

4.2 Dependence between Point Patterns

An important assumption for principal component analysis is that the observations are independent from each other. However, in many applications independence between the component patterns of a multivariate point pattern is not necessarily given. Especially when all points are observed in the same observation window the occurrence of a point at a certain location may influence the occurrence of a point of a different pattern in the area close to this location. For example, considering how different tree species are distributed in a forest, the different patterns, i.e. tree species, may not be independent from each other. Trees have influence on the growing conditions around them and other tree species might profit or benefit from them, e.g. due to a change in soil conditions. Therefore, the probability of observing a tree from a different pattern may depend on the occurrence of the tree. When the distributions of point processes are dependent, summary characteristics most likely are as well. Consequently, the independence assumption for performing PCA on the summary characteristics is violated.

There are many models for multivariate point processes that describe some kind of dependency between component processes. See Baddeley et al. (2016, Ch. 14) for an introduction to multivari-

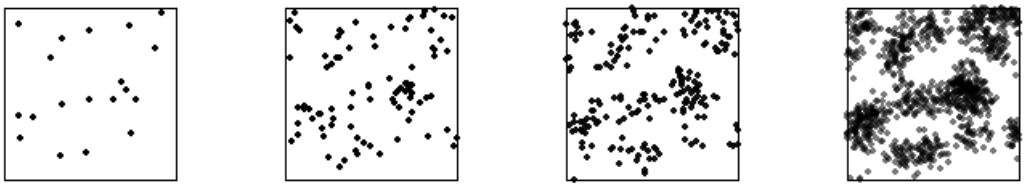

Figure 11: Thomas cluster patterns with average number of 5 (second from the left plot), 10 (second from the right) and 50 points (right plot) per cluster and corresponding parent pattern (left plot).

ate point process models. Here, we focus on one example of attraction between points of different component processes.

Again, choose the unit square as observation window, $W = [0,1] \times [0,1]$. To get dependent patterns, generate two Thomas cluster processes with the same parent pattern. The parent pattern is generated from a Poisson process with intensity 20. For this parent pattern two Thomas cluster processes of offsprings are generated with the average number of points per cluster given by `mu1` for the first pattern X and `mu2` for the second pattern Y and the same deviation parameter set to 0.05. For the generation of patterns from the Thomas cluster process the parent pattern is generated on an extended window. Denote by Z the parent pattern restricted on the observation window W. In Figure 11 three patterns are given generated from Thomas cluster processes with average number of points set to 5, 10 and 50 and the corresponding restricted parent pattern. If the average number of points in a cluster is very low, the patterns resemble the parent pattern in their distribution. The higher the average number of points per cluster, the more clustered is the pattern.

Then, all three patterns are pairwise dependent but the underlying point processes have different distributions. For instance, whenever there is a point of the parent process Z, the probability of having a cluster in X and Y close to this location is high. However, the parent pattern is drawn from a Poisson process and therefore follows complete spatial randomness. In contrast, the other two processes are both clustered but to a different degree. Generating 20 such triplets of point patterns, one gets a multivariate point pattern consisting of patterns with three different kinds of spatial behavior while each pattern has two component patterns it is dependent from. Performing the multivariate spatial point process analysis approach based on both functional and numerical summary characteristics, one would hope that the patterns with the same spatial behavior are still grouped together. To see how the dependence between patterns affects the analysis this was done for the parameters `mu1` and `mu2` in $\{5, 10, 20, 50\}$ with the same summary characteristics as in Chapter 4.1. For two cases we discuss the results in more detail.

Dendrograms are given resulting from performing Ward's algorithm on the first four hybrid scores for the cases `mu1` $= 5$ and `mu2` $= 50$ (Figure 12) and `mu1` $= 10$ and `mu2` $= 50$ (Figure 13). The number behind each letter in the name of the patterns denotes the number of the triplet. For example Z1, X1 and Y1 are dependent. Indeed, the patterns are grouped by their spatial behavior and not the dependence on other patterns. In the first case, the patterns denoted by X are generated with average number of points per cluster set to 5. This means that the pattern is clustered but is still quite similarly distributed as a Poisson process (Z). The patterns denoted by Y on the other hand have an average number of 50 points per pattern making them very clustered. In the corresponding dendrogram (Figure 12), the Y patterns are all in one group. There is a large group of all Z patterns, but in this group there are also 7 patterns that belong to the X patterns. This reflects the fact that the distribution underlying the X patterns is similar to that of the Z patterns. Conversely, in the second case both average numbers per cluster are high, though

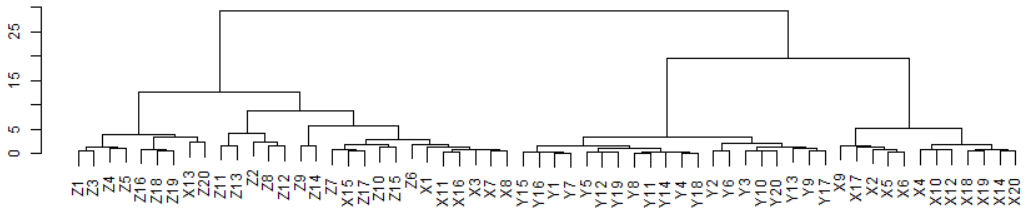

Figure 12: Dendrogram for clustering 60 patterns consisting of triplets of dependent patterns based on Ward's algorithm on the first four hybrid PC scores ($\texttt{mu1} = 5, \texttt{mu2} = 50$).

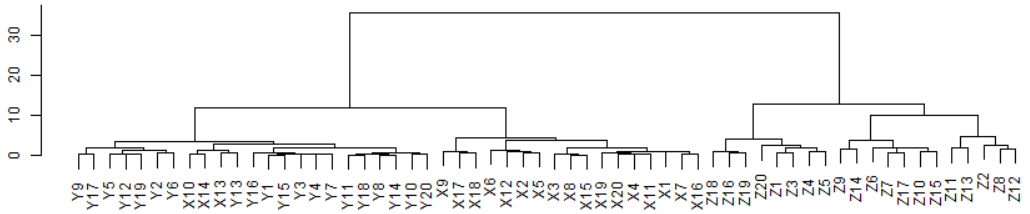

Figure 13: Dendrogram for clustering 60 patterns consisting of triplets of dependent patterns based on Ward's algorithm on the first four hybrid PC scores ($\texttt{mu1} = 10, \texttt{mu2} = 50$).

different. That is, in this case the distributions of the Y and X patterns are more similar than to the Z distribution. In the corresponding dendrogram (Figure 13) one can see that the algorithm is able to group the patterns accordingly. Only three X patterns are grouped with the Y patterns.

Generally, it was observed that the smaller the parameters `mu1` and `mu2`, the less the algorithm was able to distinguish between the three patterns, and the closer the parameters were together, the more the algorithm was not able to distinguish between them. Also, when the parameters were both very high, the differences between Z patterns based on the clustering results became higher than the differences between the X and Y patterns. Then, grouping the patterns in three groups results in two groups of Z patterns and one group of X and Y patterns. However, in further subgroups the spatial behavior can still be detected.

The analysis was also done with the weighted scalar product as given in Equation 8, but it did not change the results considerably. In addition, generating the patterns according to the same distributions but independent from each other the resulting groups are different, but not consistently better. Summarizing, the dependence between patterns does not seem to affect the results of the approach grouping the components of multivariate point patterns based on scores of a hybrid MFPCA on numerical and functional summary characteristics.

All in all, the discussed approach is a reasonable method for analyzing multivariate spatial point patterns and finding groups of patterns with similar spatial behavior. In the next chapter this approach is applied to the Duke forest data set to get a detailed analysis of the spatial distribution of the tree species in this forest.

5 Application to the Duke Forest Data

In this chapter we apply the method to analyze multivariate spatial point processes proposed in this thesis to the Duke forest data set.

The Duke forest lies in Durham, Orange Alamance counties in North Carolina (United States). It was established in 1931 for the purpose of sylvicultural management and demonstration. Now, it is managed by the School of Forestry and Environmental Studies at Duke University; among other reasons for educational, research, wildlife and water quality purposes (Palmer (1990)). It covers an area of about 7000 acres of forested land and open fields (Eckardt and Mateu (2019)). The given data set contains information on the location of 14,992 trees in an area of approximately 200 by 200 meters and a precision grid of 10 by 10 cm of the Blackwood Division, which is a part of the Duke forest. The locations are given in the Universal Transverse Mercator (UTM) projection system. That is, the x-coordinate gives the distance to the central meridian of zone 17S and the y-coordinate gives the distance to the equator, both measured in meters. Therefore, distances between locations are given approximately in meters by the Euclidean distance of the corresponding coordinates and the point space is a subset of $\mathbb{R}^2$. In addition to the location, the tree species and the diameter at breast height measured in 2014 are documented in the data set for each tree. 38 different botanic types of trees are documented with their common English name which will also be used in the following for simplicity instead of the scientific name. The aim of the analysis is to understand more about the occurrence of tree species in the Duke forest.

It has been analyzed by several authors, among them Palmer (1990), Shirota and Gelfand (2016) and Eckardt and Mateu (2019). The contribution of Palmer (1990) is a vouchered checklist of the vascular plants of the Duke forest. Mostly, this data set was spatially analyzed based on a subset of at most three botanic tree species. For example, Shirota and Gelfand (2016) analyze the three species Red Maple, Carolina Buckthorn and Sweetgum applying a log Gaussian Cox process for capturing spatial dependence of the point patterns. Eckardt and Mateu (2019) mark an exception, analyzing the tree species and applying their approach to the analysis of quantitatively marked multivariate spatial point processes based on graphical modeling also considering the diameter at breast height. There, also the approach proposed by Illian et al. (2008) using numerical summary characteristics is applied using summary characteristics for qualitatively marked point processes. The analysis of qualitatively marked point processes is the focus of this publication.

In this thesis, we omitted the consideration of possible qualitative marks of multivariate spatial point processes to focus on the simultaneous consideration of functional and numerical summary characteristics which is also the focus of this chapter. Before applying the proposed extension of the approach by Illian (2006) to analyze multivariate spatial point processes based on hybrid multivariate functional principal component analysis of numerical and functional summary characteristics, we consider basic spatial analysis methods for a more detailed data description, following the chapters on basics and data exploration in Baddeley et al. (2016).

5.1 Evaluation of Assumptions

First of all, it should be mentioned why it makes sense to treat the Duke forest data set as generated from a multivariate point process considering each tree species as a component of the multivariate object. The two-dimensional vector containing the x- and the y-coordinate of a tree can be seen as a random point and the set of all trees as an unordered finite subset of points in $\mathbb{R}^2$. The underlying process of interest is the process that determines the occurrence of trees in the Duke forest and the data set can be seen as one realization of this process. That is, the forest with the given trees

form one point pattern consisting of 38 subpatterns. In addition, the physical size of objects can be neglected at the scale of interests. Recall that for the analysis of spatial point patterns there are a few assumptions typically made. One of them is the assumption of simplicity; see Chapter 2.1 for an introduction to the assumptions discussed in this Chapter. To see if this assumption is actually satisfied, check the data for duplicated points. It turns out that the data contains 2,298 entries with exactly the same coordinates, that is a proportion of about 0.15 of the total amount of trees. Obviously, in this application simplicity is a reasonable assumption since it is unlikely that two trees fit in an 10 by 10 cm area. Reasons for the duplicates could be that due to human error or technical problems

- the exact coordinate of a tree was not documented correctly or the grid is not fine enough and the duplicates consist of trees very close to each other, or
- trees were falsely included more than once in the data set.

81 of the trees with exactly the same coordinates have different documented tree species. Though this suggests that there might be an error, for this analysis we assume that for these trees the coordinates were not documented correctly and leave these points in the data set. Summary characteristics are calculated for each component pattern separately, and these trees do not lead to duplicates in component processes. In addition, it is only a neglectable proportion of the data set and it is not possible to know which tree should be excluded.

There are 1,709 trees in the data set that have the same location, the same species but different diameter recorded. The differing diameters at breast height suggest that these entries also consist of trees that are very close to each other and refer to different trees. A further indication for this is that the tree species which has the highest proportion of duplicates is Possumhaw which needs less than one meter to grow while tree species that need a lot of space such as Willow oak have mostly no duplicates. Consequently, these trees are also included in further analysis. To ensure that there are no duplicates, in each component pattern for the estimation of the summary characteristics, in this case independent random displacement is applied to the coordinate of each duplicate entry using the function `rjitter`.

However, there are still 508 cases where at least two entries are completely identical, that is, the coordinate, the tree species and the diameter at breast height agree. For these entries assume that they were more than once falsely included in the data set and exclude the duplicates from further analysis.

A detailed simulation study in Illian et al. (2008) revealed that the method based on functional summary characteristics is stable as long as the random noise added to the location of a point is not very strong, that is, more than 20% of the window size and the probability of misspecification of the species is below 0.24. Both are reasonable assumptions in this setting.

With these adaptions of the data set, the simplicity assumption is satisfied for each component process.

Consider now the multivariate point pattern in more detail. Figure 14 depicts the locations of the trees in the data set. Each point in the plot corresponds to the UTM-coordinates of one tree and the color indicates the tree species. Due to the large number of trees as well as species it is difficult to make out any pattern visually.

A second very important assumption for spatial point processes is stationarity. This assumption is necessary for the definition of most of the summary characteristics introduced in Chapter 2, since it extremely facilitates the analysis of point processes. In theory, it makes sense to assume that the arrangement of trees does not depend on the location when ignoring other covariates that might

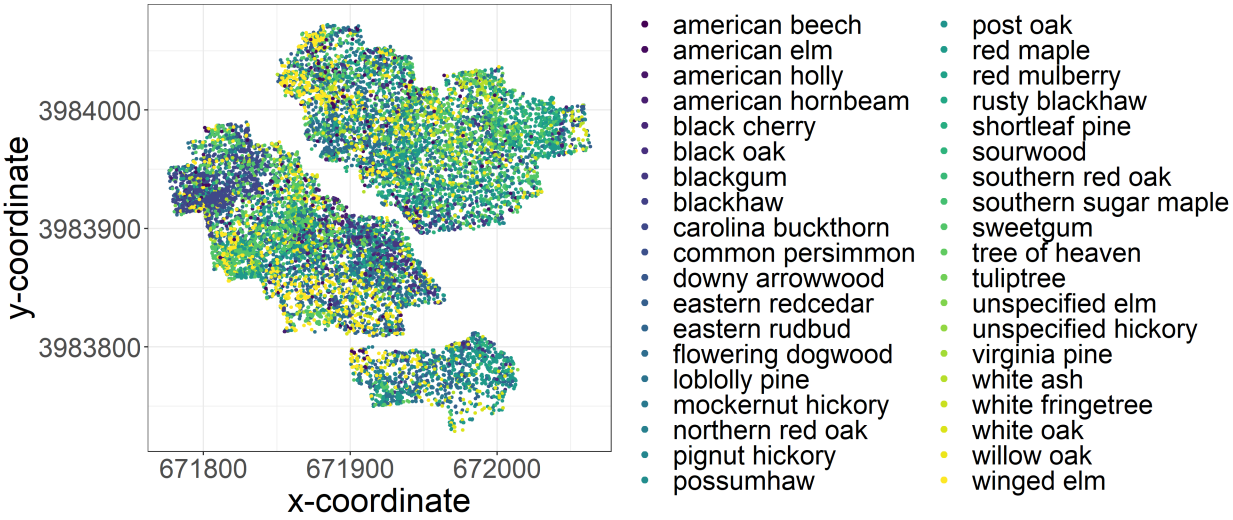

Figure 14: Duke forest data set; each point represents the UTM-coordinate of one tree. The color of a point indicates its tree species which are listed at the right.

influence the growth conditions such as soil characteristics. However, since the tree locations are only collected in a small part of the forest, it may be that for some species the observation window is too small to detect stationarity; for example, if a tree species is clustered with cluster centers having a large distance so that only one cluster falls inside the window. For a plot of the trees in the observation window for all tree species, we refer to Figure 25 in the appendix. On the other hand, it may also be that the pattern is non-stationary. For simplicity, assume that all the component processes are stationary. The summary characteristics for stationary point processes can be estimated and analyzed also for non-stationary patterns.

Ergodicity needs to be assumed so that from the results based on this data one can conclude on the distribution of the whole process. Based on one point pattern one cannot check for this assumption. But it is a reasonable assumption if the behavior of the trees is not exceptional in this specific part of the forest. Consequently, assume that the underlying process is ergodic. For the PCA on numerical or functional summary characteristics the point patterns need to be also independent. This is an assumption that is very likely violated for this data set since most trees species, especially the ones where the trees are large, have an influence on the surroundings. It may well be that certain tree species improve growing conditions for other species, so their spatial behavior has some dependence. Or conversely, tree species may not grow together within a certain distance, which again implies dependence in their spatial behavior. The simulation study in Chapter 4.2 suggests that patterns will still be grouped according to their spatial behavior.

When analyzing point patterns, the observation window W needs to be discussed since it is part of many analysis methods, for example when considering empty spaces in the pattern or points per area. In other words, it is also important to know if there are actually no points in an area or if this area is outside the observation window so there could be points which are not observed. The true observation window is often specified in the sampling process. For this thesis, further information on the sampling of this data set is not at hand, consequently the observation window needs to be estimated from the data; see Figure 15 for the bounding box and the convex hull of the Duke forest data set.

One approach to estimate W is to take the largest rectangle including all points with sides parallel to the x- and y-axis, respectively. The resulting observation window is called *bounding box* and is often taken as a default if the observation window is not specified. However, for this data set, looking at Figure 14, this is not an appropriate choice since the trees were obviously not

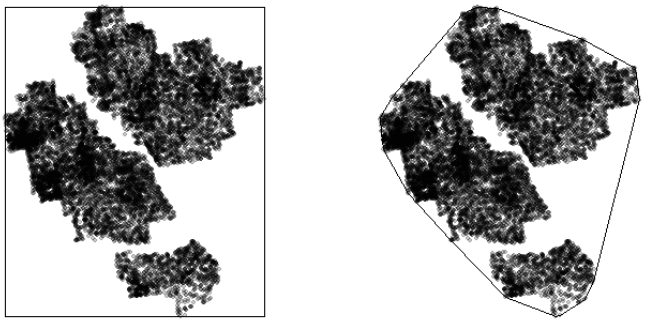

Figure 15: Observation windows for the Duke forest data.The black line indicates the observation window and each black point corresponds to one tree in the data set. Left plot: Bounding box. Right plot: Convex hull.

recorded in a rectangle. In the `spatstat` package there is also the option to use a point-and-click interface to set the observation window by hand for example on pictures available of the observation area. Looking at the coordinates given in this data set on a map, using OpenStreetMaps revealed that there are no obvious boundaries of this part of the forest and that the forest continues in all directions. However, there are two large streets cutting through that area which explains the clear separation of areas with recorded trees. An alternative approach is to consider the convex hull. Obviously, the points are located in a polygon, which is not convex. Therefore, this approach is not ideal but it is considerably better than the bounding box and the click-and-point approach. We use in the following the convex hull calculated with the `ripras` function which is recommended by Baddeley et al. (2016).

Having discussed the necessary assumptions, the proposed method is now applied to the Duke forest data set. For this, first numerical summary characteristics are estimated and standard PCA is performed.

5.2 Approach using Numerical Summary Characteristics

In Illian et al. (2006) when analyzing plant communities all species with an abundance of above 20 were included in the analysis. Though the settings differ, this number is taken as reference. All species in the Duke forest data set have more than 20 occurrences; Willow Oak and Unspecified Hickory, having the smallest number of trees with 22 trees in the data set. Consequently, for each species in the Duke forest data set we consider summary characteristics.

For each tree species the Clark-Evans index CEI, the mean directional index $\overline{R}$, the measure SK, the pair correlation function at approximately $r = 6.87$ and $r^{(1)}$ are estimated (see Appendix, Table 8). These are the numerical summary characteristics used in Illian et al. (2008) and in Chapter 4.1; except that the r value for $\hat{g}(r)$ was adapted to this setting.

Considering the CEI, all values are below one. This indicates that there is a tendency for all patterns towards clustering. However, the values vary between 0.16 (Downy Arrowwood) and 0.93 (Unspecified Elm). Therefore, the patterns appear to be clustered to very different degrees. The mean CEI lies at around 0.48 with a standard deviation of 0.19. This means that the average pattern has a mean nearest-neighbor distance that is half of the mean nearest-neighbor distance of

Summary Characteristics	PC1	PC2	PC3	PC4
CEI	0.57	-0.33	0.06	0.11
$\overline{R}$	-0.48	-0.35	0.18	-0.72
SK	-0.47	-0.22	-0.73	0.36
$\hat{g}(6.87)$	-0.45	0.41	0.53	0.41
$r^{(1)}$	0.14	0.74	-0.39	-0.42

Table 6: First four principal components of the standard PCA on the numerical summary characteristics CEI, $\overline{R}$, SK, $\hat{g}(6.87)$ and $r^{(1)}$ of each tree species in the Duke forest data set.

a comparable Poisson process.

The mean directional index $\overline{R}$ is for all patterns beside one (Loblolly Pine) above a value of 2, and all values are above the mean of 1.96 for random patterns. This indicates irregularity since the nearest neighbors of a typical point seem on average to have similar directions in contrast to being evenly distributed around the point. The latter would lead to a mean directional index of zero. The mean of $\overline{R}$ for all patterns is 2.32 and the standard deviation is approximately 0.17.

The measure SK is above one for all tree species except Tree of Heaven. For this species it is only slightly below one (0.99). Recall that SK is the ratio of the mean nearest-neighbor distance and its median and therefore indicates the skewness of the nearest-neighbor distance distribution. This means that there is a tendency towards a positive skew. Some species have relatively high SK values of above five (American Hornbeam, Blackhaw, Downy Arrowwood, Possumhaw), indicating that for these tree species there are some outliers that are comparatively far away from the other trees of that species. These are all species with relatively few trees in the observation window so that few outliers have a large impact on the mean nearest-neighbor distance. The mean SK over all patterns lies at 2.42.

For the two variables based on the pair correlation for all patterns the values indicate clustering; since $\hat{g}(6.87)$ is above one and $r^{(1)}$ is quite high. The estimated pair correlation function for a distance between trees of 6.87 meters is on average over all patterns 7.77. However, there are again some outliers. There are three species with values above 20 indicating a high degree of clustering for these r values, namely Carolina Buckthorn, Southern Sugar Maple and Virginia Pine. Interestingly, tree species with the most extreme values differ for the first four numerical summary characteristics. This suggests that the four numerical summary characteristics all measure different clustering behavior, and there are species in this data set that have different spatial behavior though all patterns have a tendency towards clustering.

For 22 out of the 38 tree species it holds that $r^{(1)} = 71.75$ which is the highest r value for which the pair correlation function is estimated in this setting. Hence, for most of the patterns the pair correlation function does not attain a value of one. The mean value for $r^{(1)}$ is equal to 64.42. This could indicate that this variable is not very informative in this setting of only clustered point patterns.

Considering the number of points in each pattern, there is variation over the patterns. There are four tree species with over 1000 trees (Carolina Buckthorn, Flowering Dogwood, Red Maple and Sweetgum) and 20 tree species with less than 100 trees. This is also reflected in the great difference between the mean number of points (376.68) and the median number of points (79.50). As argued in Chapter 4 and in Illian et al. (2008), the number of trees and the intensity are not considered in the further analysis since the scale-invariant spatial behavior of the patterns is of interest.

Performing standard PCA on the summary characteristics discussed, the first four principal

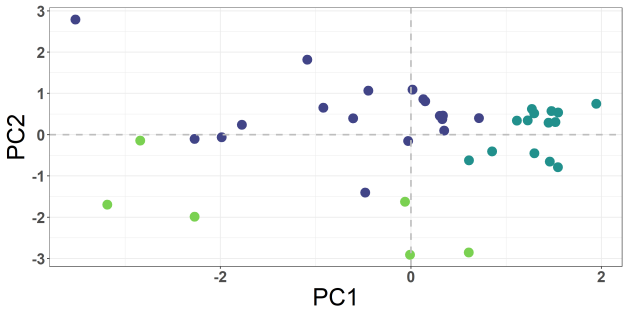

Figure 16: First two scores of the standard PCA on the numerical summary characteristics on the tree species in the Duke forest data set. The colors indicate the three groups that result from clustering based on Ward's algorithm on the first four PC scores and the gray lines give the coordinate axes.

components explain 94% of the variation in the data and are given in Table 6. Similar to the results in Chapter 4.1, the CEI is associated in the opposite direction as $\overline{R}$, SK and $\hat{g}(6.87)$ with the first PC, and they all have about the same loadings. The first PC is slightly dominated by the CEI but could be interpreted as generally measuring the clustering, regularity or randomness of the patterns. In contrast to the simulation study $r^{(1)}$ has a low positive loading. This could be explained by the above observation that for this data set the variable is not as informative. However, the second PC is dominated by $r^{(1)}$ indicating that it still explains an important part of the variation in the data. The third PC is dominated by SK and the fourth PC by $\overline{R}$. That is, the second, third and fourth PC could be interpreted as indicating a certain spatial behavior based on the variable that is dominant.

In Figure 16 for each tree species the first two PC scores are depicted. It is difficult to make out any clear clusters. Clustering the tree species into three groups based on Ward's algorithm on the first four PC scores yields groups indicated by the colors in Figure 16.

Consider the group with the darker green color in more detail. It consists of 14 patterns. Most patterns in that group have a positive first PC score close to one and a second PC score of absolute value less than one. For these patterns the CEI is higher than on average which means that the mean distance between points is higher. A positive first PC score indicates that the value for the three variables $\overline{R}$, SK and $\hat{g}(6.87)$ is lower than the average. This indicates that within all the patterns these patterns are less clustered.

The group indicated by the lighter green color consists of patterns either with a strongly negative first or second PC score. Since the second PC score is strongly associated with $r^{(1)}$, this indicates that these patterns have lower than average values for this variable. Indeed, all five tree species are in this group for which it holds $r^{(1)} < 50$ indicating that these patterns have a tendency towards random behavior or even regularity at a certain distance. However, as indicated in the score plot the spatial behavior is again split in this group. Three of the patterns (Blackhaw, Downy Arrowwood and Possumhaw) have also very high values for SK and $\hat{g}(6.87)$, a slightly higher $\overline{R}$ value and a low CEI. This can be interpreted as a high degree of clustering at short distances but some regular behavior at higher distances. The other tree patterns (Southern Red Oak, Unspecified Hickory, Willow Oak) are the three outliers in the variable SK with very high values.

The third group that is colored in blue in Figure 16 contains trees that are not as strongly categorized as the groups before since the scores vary within this group.

The results of the clustering based on Ward's algorithm can also be depicted in the corresponding dendrogram (Figure 17).

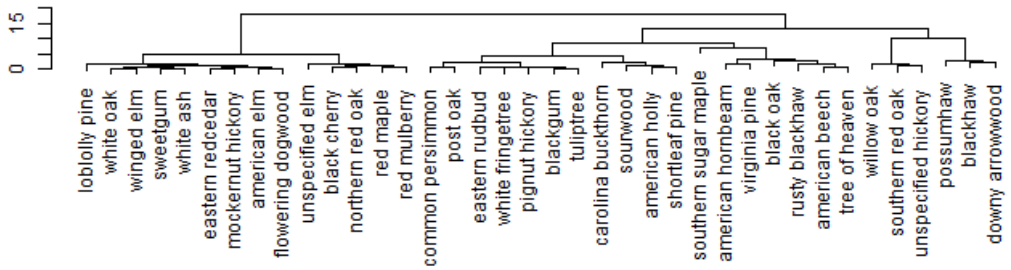

Figure 17: Dendrogram based on Ward's algorithm on the first four scores of the PCA on numerical summary characteristics for the tree species in the Duke forest data set.

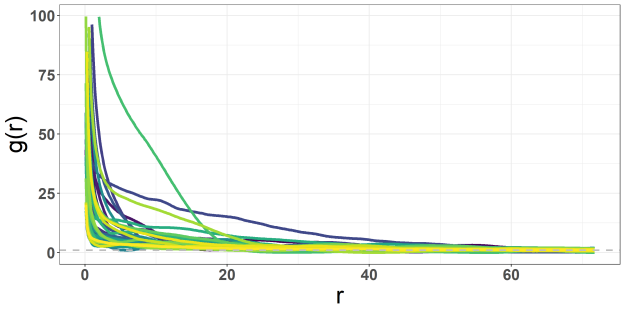

Figure 18: Estimated pair correlation function for each tree species in the Duke forest data set given by the colored lines. The gray dotted line indicates the one.

In this figure the difference discussed regarding the light green group is illustrated. The group consists of the six tree species in the nodes at the right of the dendrogram which can be split into two clearly different sub clusters. It also shows that the large blue group which is the middle cluster also can be split into two sub clusters which have different spatial behavior. Also, Southern Sugar Maple seems to be a general outlier.

5.3 Approach using one Functional Summary Characteristic

Having discussed the numerical summary characteristics, now we turn to the functional summary characteristics. More specifically, we consider the pair correlation for each tree species which is depicted in Figure 18. The plot also confirms the results from the numerical summary characteristics that all patterns have a tendency towards clustering since the pair correlation functions have mostly values above one but the degree of clustering differs greatly between the tree species. In this plot, function values above 100 are cut to be able to better see how the functions behave when they are slowly decreasing towards one. It becomes apparent that especially for distances between zero and 20 the pair correlation functions differ greatly. There are three lines that stand out due to their high values.

In the next step FPCA is performed on the pair correlation functions estimating the first four functional principal components. Here, all values besides for $r = 0$ are included where the values of all pair correlation functions are estimated as infinite. B-splines were used for smoothing as proposed in Illian et al. (2006). The first functional PC then describes more than 90% of the explained variation, and the first two PCs explain even around 99%. Figure 19 depicts the first two estimated functional principal components in the left panel and the mean function in the right panel.

The mean function reflects the behavior already described. The first functional PC (darker blue line) indicates a curve that has higher values than the mean function for r values below 20. Then

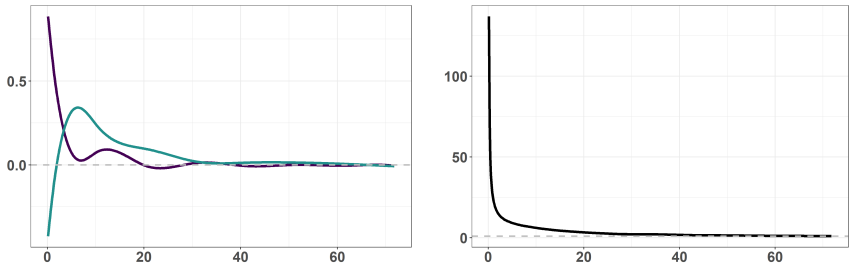

Figure 19: Left plot: First two functional PCs that result from performing FPCA on the pair correlation functions for each tree species. The gray dotted line indicates the zero, the dark blue line the first PC and the lighter blue line the second PC. Right plot: Estimated mean function. The dotted gray line indicates the value one.

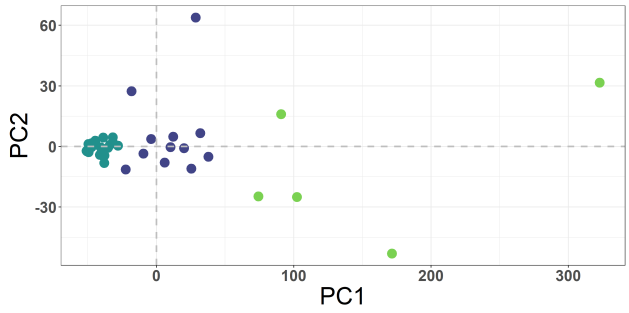

Figure 20: First two functional PC scores. The colors indicate groups that result from applying Ward's algorithm on the first two PC scores for each pattern in the Duke forest data set and the dotted gray lines indicate the coordinate axes.

the sign of the function changes but the function moves around one from then on. A very high first functional PC score indicates a higher than average degree of clustering for very low r values.

The second functional PC has negative values for very small r values, then changes the sign and increases until an r value around six. Then it slowly decreases until reaching zero around $r = 30$. A high positive PC score on the second PC indicates less than average clustering for very small r but higher than average clustering for values for middle distance points. This indicates a generally flatter pair correlation function.

There does not seem to be much variation in the pair correlation functions for r values above 30 which could already be seen in the plot of the pair correlation functions.

Figure 20 gives the first two functional PC scores for each tree species in the Duke forest data set. The colors indicate the grouping of the patterns in three clusters based on Ward's algorithm. The first PC explains most of the variation in the data and, therefore, it is not surprising that the grouping of the patterns mostly is explained by differences in the first PC score. There is a group of patterns (dark green) with a negative first PC score and a second PC score close to zero. This means that these patterns are less clustered for low distances between trees than the average. There are again five patterns with very high first PC score values (light green color). Three of those are the scores of the tree species Possumhaw, Downy Arrowwood and Blackhaw which already stood out in the numerical summary characteristics analysis due to their high values for $\hat{g}(6.87)$ and SK. The pattern having the extreme high first PC score of above 300 belongs to the tree species Southern Sugar Maple which also stood out in the previous analysis and corresponds to the lighter green line in the plot of the pair correlation functions in Figure 18 that stands out on the interval from one to 20 with very high values.

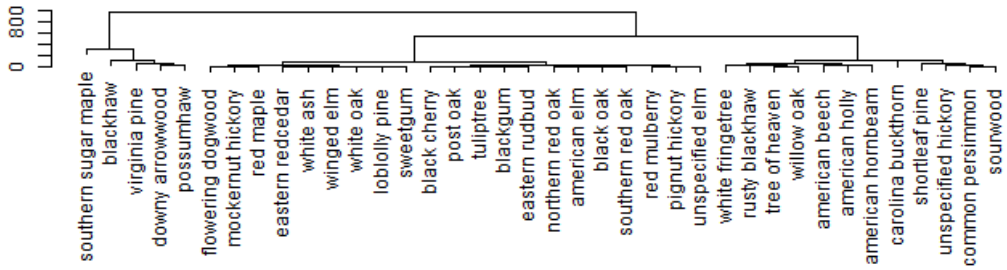

Figure 21: Dendrogram based on Ward's algorithm on the first four scores of FPCA on the pair correlation function for the tree species in the Duke forest data set.

Summary Characteristics	PC1	PC2	PC3	PC4
CEI	0.37	-0.17	0.28	0.16
$\overline{R}$	-0.16	0.29	0.03	0.41
SK	0.08	0.57	-0.18	-0.30
$\hat{g}(6.87)$	-0.52	-0.14	-0.17	0.19
$r^{(1)}$	-0.07	-0.18	-0.30	-0.75

Table 7: Numerical part of the first four hybrid PCs for the Duke forest data set.

In Figure 21 the corresponding dendrogram is given. All tree species that were categorized as less clustered in the numerical summary characteristics analysis are now again in the group that is considered containing less than average clustered patterns. In the dendrogram this corresponds to the group in the middle starting on the left with Flowering Dogwood.

5.4 Hybrid Approach for Summary Characteristics

Having considered the numerical and functional summary characteristics separately, we apply hybrid MFPCA in the next step to analyze them simultaneously. First, the numerical and functional PC scores are concatenated for each pattern to one score vector and on these vectors standard PCA is performed. The resulting first four PCs explain about 96% of the variation in the score vectors. Consequently, the first four hybrid scores are calculated based on the functional and vector PCs and weights given by the score vector PCs. The vector parts of the first four hybrid PCs are given in Table 7 and the functional parts in Figure 22.

The numerical part of the first hybrid PC is dominated by the variable $\hat{g}(6.87)$ and the CEI. They are oppositely associated with the first hybrid PC, and a high positive first hybrid PC score indicates a less clustered point pattern. The numerical part of the second hybrid PC is dominated by the measure SK, therefore, a high second hybrid score indicates a higher than average degree of irregularity.

Since only two functional PCs were included, the third and fourth functional part of the hybrid PCs are constantly zero, indicated by the green line in Figure 22. The functional part of the first hybrid PCs is given by the darker blue line. A high positive first hybrid score then indicates a pair correlation function that has lower values than the mean pair correlation function for low r values. In other words, the corresponding point pattern shows less clustering for short distances between trees. In contrast, interpreting the functional part of the second hybrid PC indicated by the lighter blue line, a high positive second PC indicates a higher degree of clustering for very low r values

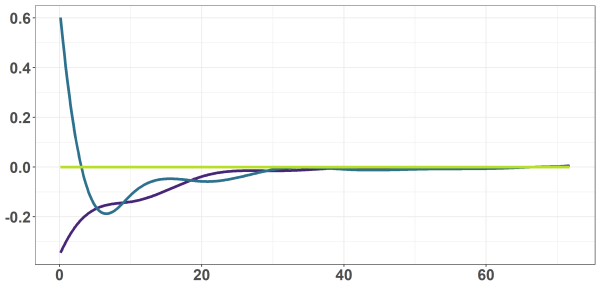

Figure 22: Functional part of the first four hybrid PCs for the Duke forest data set. The dark blue line corresponds to the first PC, the lighter blue line to the second PC, and the third and fourth PC are constantly zero indicated by the green line which overlaps with the third line.

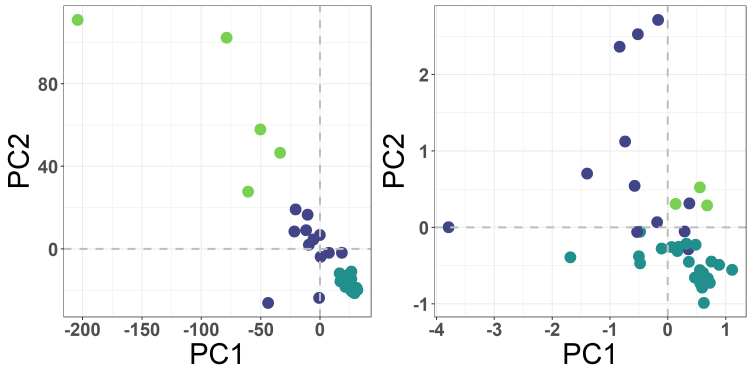

Figure 23: Left panel: First two hybrid PC scores for the Duke forest data set, colors indicating groups resulting from Ward's algorithm. Right panel: First two hybrid PC scores, using a weighted scalar product. The gray dotted lines indicate the coordinate axes in both panels.

but for r values larger than three until 30, less than average clustering would be expected.

The hybrid scores for each pattern are given in Figure 23 in the left panel. Again, colors indicate three groups resulting from clustering based on Ward's algorithm. Similar to the results from Chapter 5.3 analyzing the pair correlation function, the grouping mostly happens based on the first score. This is due to the extremely large values in the functional part of the first hybrid PC scores. Some patterns have mostly points very close to each other which leads to an extremely high pair correlation function for very small r and corresponding functional PC scores. Consequently, the first functional PC scores vary between -50.88 and 322.70 with an estimated standard deviation of 73.72. In contrast, the numerical PC scores have an estimated standard deviation of 1.45 and attain values between -3.52 and 1.94. This means the hybrid scores which are weighted sums of the numerical and functional PC scores are dominated by the functional PC scores. It is no surprise that grouping the tree species based on the first four hybrid scores yields exactly the same groups as in the case when grouping based on the functional PC scores.

However, the idea of hybrid MFPCA is to simultaneously analyze the numerical and functional summary characteristics. It should not be that one part dominates the analysis due to a different scale.

Therefore, an appropriate weighted scalar product for the hybrid data objects should be used. The first two hybrid scores are depicted in the right plot of Figure 23 using the weight defined in Equation 8 for scaling. The values are not as extreme as in the case without scaling.

Grouping the tree species based on these scores yields groups indicated again by three different colors in the plot. Points of different colors overlap indicating that later hybrid PCs also play a

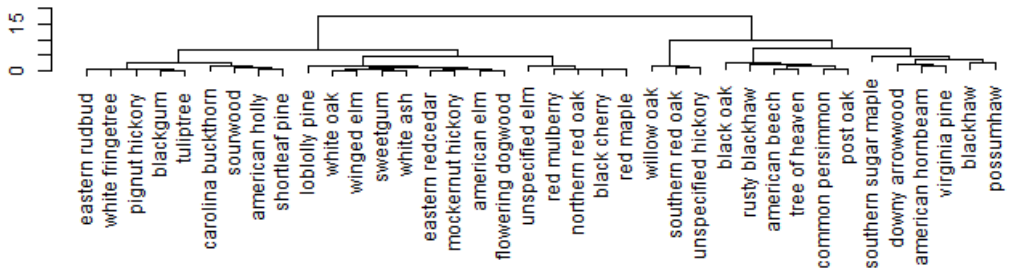

Figure 24: Clustering of the tree species in the Duke forest data set based on Ward's algorithm on the first four hybrid scores with normalized functional and numerical scores.

role in the grouping. The group indicated by the light green color is composed of the tree species Southern Red Oak, Unspecified Hickory and Willow Oak, which stood out in the analysis based on numerical summary characteristics. Thus, it is not surprising, that in the simultaneous analysis they mark an outlier group as well.

Considering the corresponding dendrogram in Figure 24, the groupings appear to contain aspects of both previous analyses. Looking at the three groups at the right hand side of the dendrogram, the most left is the group of outliers from the numerical summary characteristics analysis mentioned already. The group consisting of six tree species at the right consists of the five tree species that had extreme values in the functional summary characteristic analysis and the tree species American Hornbeam which is similar to Southern Sugar Maple and Virginia Pine when considering numerical summary characteristics. The group between has four tree species (Black Oak, Rusty Blackhaw, American Beech, Tree of Heaven) that were in the numerical summary characteristics analysis grouped with Southern Sugar Maple and Virginia Pine. Consequently, the grouping based on the hybrid data combines information of both analyses. The tree species with corresponding extreme behavior in one variable, functional or numerical, are again grouped together. This is clear since the hybrid scores are the sums of weighted numerical and functional scores, so extreme values in either type of score also yield extreme values in the hybrid score. This is an advantage of the hybrid approach. Patterns that have a certain particular spatial behavior that can either be measured with a numerical or functional summary characteristic can be detected.

However, one should pay attention to estimation errors which could be a reason for extreme values in summary characteristics, for example the tree species Possumhaw has nearly 50% duplicates in the original data set. Consequently, the way duplicates were handled could have a serious impact on the results. It is necessary to look into the patterns with extreme values of summary characteristics in more detail. For simplicity, a detailed analysis of each pattern is omitted here.

Additionally, one could consider alternative weights for the scalar product. It might be appropriate to weight the different summary characteristics differently, based on the research question. For a more specific choice of summary characteristics and their weight as well as a detailed analysis of the found groups and how they can be interpreted, profound knowledge about botanic properties of the tree species is necessary.

6 Discussion and Outlook

This thesis discussed the analysis of multivariate spatial point processes. For the simultaneous analysis of a large number of component processes there are only two approaches in existing literature proposed by Eckardt and Mateu (2019) and Illian (2006). We further investigated and executed the approach established in Illian (2006) that applies principal component analysis on summary characteristics of the component processes.

To this end, general spatial point process theory was reviewed in Chapter 2 and important summary characteristics, both functional and numerical, were discussed. In the approach established in Illian (2006) numerical and functional summary characteristics were only analyzed separately. To extend the approach to analyzing both data modalities simultaneously, principal component analysis for hybrid data was discussed in Chapter 3. First, the general idea of principal component analysis was presented by reviewing standard PCA on vector data. Then, the generalization on Hilbert spaces was discussed in order to get the theoretical requirements for principal component analysis on general spaces. This yielded the theoretical basis for (multivariate) functional PCA and finally hybrid MFPCA.

For the estimation of hybrid MFPCA, three approaches were discussed. The first approach is based on pre-smoothing. The second and third approach extend the estimation method for MFPCA proposed by Happ and Greven (2018). The first approach depends on the choice of basis functions and performs not very well on sparse data. The second of those approaches is introduced in Jang (2021) and extends the approach to include both vector and functional data. The third approach builds on the idea of transforming vector data into step functions and then applies MFPCA on the purely functional object. However, the last approach has the disadvantage of being computationally more expensive. The approach proposed by Jang (2021) has the advantages of being fast, performing well on sparse data and not being susceptible to bad performance in case of misspecification of the basis.

After the theoretical derivation of the approach using hybrid MFPCA on numerical and functional summary characteristics of the components of multivariate spatial point processes, the approach was evaluated in a small simulation study.

The simulation study was kept simple for a better understanding of the presented methods. That is, in the first part it was investigated if the approach is able to distinguish the three standard types of spatial behavior, CSR, clustering, and regularity. In the second part, we considered dependence between patterns. In both cases the proposed extension was able to group patterns according to their spatial behavior.

In Chapter 4.1 three simple cases of spatial point process models were studied. There is a large amount of other models which describe different aspects of spatial behavior that could also be looked at. In addition, the effect of the degree of difference between patterns could be investigated in more detail. In this thesis the hybrid scalar product with weight $\kappa = 1$ and the weight given in Equation 8 were considered. Other weights could be discussed as well. Depending on the application, it might also be sensible to give more weight to different numerical or functional summary characteristics. Considering the choice of summary characteristics in more detail opens up a large field of questions. In this thesis the choice of summary characteristics was restricted to the choices made in Illian et al. (2006) and Illian et al. (2008) due to the similarity in the applications and lack of knowledge of biological processes. Specifically only one functional summary characteristic was included though the approach allows for the inclusion of more than one. There is a large amount of

other numerical and functional summary characteristics that could be useful to describe different spatial behavior of component patterns and therefore could further be explored in this regard. This might strongly depend on the application at hand.

Finally, the approach, when applied to the Duke forest data set, was be able to group tree species based on their spatial behavior. All tree species have a tendency towards clustering. However, the approach enabled to specify the difference in the spatial behaviors in more detail. The application also showed that a certain spatial behavior detected in one summary characteristic can also be observed in the hybrid approach. Consequently, the hybrid approach can be seen as a summary of the spatial behavior described in all summary characteristics together.

Note that the proposed approach can be easily extended to a large variety of other settings. Since the only premise is to have a set of sensible summary characteristics, this approach can be extended to any setting in which such characteristics exist. For example, if the stationarity assumption is not satisfied, one could include summary characteristics that are derived for non-stationary point processes. These include several inhomogeneous functional summary characteristics such as the inhomogeneous K-function implemented in the `spatstat` package as function `Kinhom`. For reasons of clarity, the case of non-stationarity was not discussed in this thesis. However, stationarity is a strong assumption and, for example, in the Duke forest data set there are tree species that very well might not be stationarily distributed.

A second example for an extension of the approach would be to consider (quantitatively) marked multivariate spatial point processes and use summary characteristics such as the mean mark and the interquartile-range of the mark as numerical summary characteristics as well as the mark correlation function and the mark variogram as functional summary characteristics. For numerical summary characteristics applied to the Duke data set this was already done in Eckardt and Mateu (2019). It was also pointed out that using agglomerative cluster methods such as Ward's algorithm can become computationally expensive for large data sets.

Summarizing, the approach proposed in this thesis yields a sensible method to describe and analyze the spatial behavior of multivariate spatial point patterns when the amount of component patterns is large. In addition, it facilitates the consideration of different types of summary characteristics simultaneously and gives a detailed description of the distribution of the multivariate spatial point pattern. The main properties of the given distribution can be given via the first few principal components describing the most variability in the data.

References

Baddeley, A., Rubak, E. and Turner, R. (2016). *Spatial Point Patterns - Methodology and Applications with R*, Interdisciplinary Statistics Series, Chapman and Hall/CRC, Boca Raton.

Chiu, S. N., Stoyan, D., Kendall, W. and Mecke, J. (2013). *Stochastic geometry and its applications*, Wiley series in probability and statistics, 3rd ed. edn, Wiley, Chichester.

Corral-Rivas, J. J., Pommerening, A., von Gadow, K. and Stoyan, D. (2006). An analysis of two directional indices for characterizing the spatial distribution of forest trees, *Models of tree growth and spatial structure for multi-species, uneven-aged forests in Durango (Mexico)*, Cuvillier Verlag, pp. 106–121.

Daley, D. J. and Vere-Jones, D. (1998). *An Introduction to the Theory of Point Processes*, Springer Series in Statistics, New York, NY.

Daley, D. and Vere-Jones, D. (2008). *An Introduction to the Theory of Point Processes Volume II: General Theory and Structure*, Probability and Its Applications, 2nd ed. 2008. edn, New York, NY.

Diggle, P. J. and Milne, R. K. (1983). Bivariate cox processes: Some models for bivariate spatial point patterns, *Journal of the Royal Statistical Society. Series B (Methodological)* **45**(1): 11–21.

Eckardt, M. and Mateu, J. (2019). Analysing multivariate spatial point processes with continuous marks: A graphical modelling approach, *International Statistical Review* **87**(1): 44–67.

Everitt, B. S., Landau, S. and Leese, M. (2001). *Cluster analysis*, 4. ed. edn, London.

Happ, C. and Greven, S. (2018). Multivariate functional principal component analysis for data observed on different (dimensional) domains, *Journal of the American Statistical Association* **113**(522): 649–659.

Happ-Kurz, C. (2020). *MFPCA: Multivariate Functional Principal Component Analysis for Data Observed on Different Dimensional Domains*. R package version 1.3-6.

Hsing, T. and Eubank, R. (2015). *Theoretical foundations of functional data analysis, with an introduction to linear operators*, Chichester, West Sussex, United Kingdom.

Illian, J. (2006). *Spatial point process modelling of a biodiverse plant community*, PhD thesis, Abertay University.

Illian, J., Benson, E., Crawford, J. and Staines, H. (2006). Principal component analysis for spatial point processes — assessing the appropriateness of the approach in an ecological context, *Case Studies in Spatial Point Process Modeling*, Lecture Notes in Statistics, Springer New York, New York, NY, pp. 135–150.

Illian, J., Penttinen, A., Stoyan, H. and Stoyan, D. (2008). *Statistical analysis and modelling of spatial point patterns*, Statistics in practice, John Wiley, Chichester, England ; Hoboken, NJ.

Jang, J. H. (2021). Principal component analysis of hybrid functional and vector data, *Statistics in Medicine* **22**(1).

Matérn, B. (1960). Spatial variation : Stochastic models and their application to some problems in forest surveys and other sampling investigations, *Meddelanden från Statens Skoforskingsinstitut* **49**(5).

Matérn, B. (1986). *Spatial Variation*, Lecture Notes in Statistics, 36, second edition. edn, New York, NY.

Møller, J. and Waagepetersen, R. P. (2004). *Statistical inference and simulation for spatial point processes*, Monographs on statistics and applied probability BV002494005 100, Boca Raton ; London ; New York ; Washington D.C.

Palmer, M. W. (1990). Vascular flora of the duke forest, north carolina, *Castanea* **55**(4): 229–244.

Ramsay, J. and Silverman, B. W. (1997). *Functional Data Analysis*, Springer Series in Statistics, New York, NY.

Shirota, S. and Gelfand, A. E. (2016). Inference for log gaussian cox processes using an approximate marginal posterior, *arXiv: Computation* .

Stoyan, D. and Stoyan, H. (1994). *Fractals, Random Shapes and Point Fields*, Wiley, Chichester.

Thomas, M. (1949). A generalization of poisson's binomial limit for use in ecology, *Biometrika* **36**(1-2): 18–25.

Appendix

Proofs

Proof 1 It holds that

$$\begin{aligned}
\langle f_1, f_2\rangle_{S([0,p])} &= \langle f_1, f_2\rangle_{L^2([0,p])} \\
&= \int_{[0,p]} f_1(t) f_2(t) d\nu(t) \\
&= \int_{[0,p]} \left(\sum_{i=1}^{p} v_1^{(i)} \mathbb{1}_{(i-1,i]}(t)\right) \left(\sum_{i=1}^{p} v_2^{(i)} \mathbb{1}_{(i-1,i]}(t)\right) d\nu(t) \\
&= \int_{[0,p]} \left(\sum_{i=1}^{p} v_1^{(i)} v_2^{(i)} \mathbb{1}_{(i-1,i]}(t)\right) d\nu(t) \\
&= \sum_{i=1}^{p} v_1^{(i)} v_2^{(i)} \int_{(i-1,i]} 1 d\nu(t) \\
&= \sum_{i=1}^{p} v_1^{(i)} v_2^{(i)} \\
&= \langle v_1, v_2\rangle_{\mathbb{R}^p},
\end{aligned}$$

where definitions, the linearity of the integral and properties of the indicator function are used.

Proof 2 To see that the assertion holds, first, note that $S[0,p]$ is a space of dimension p, so there are p eigenfunctions. Also, $\{w_j\}$ are orthonormal if and only if the $\{e_j\}$ are since scalar products are equal. For $s,t \in [0,p]$ there are $r,z \in \{1,\ldots,p\}$ such that $s \in (r-1,r], t \in (z-1,z]$ and we have that

$$\begin{aligned}
\mathcal{K}_{f^V}(s,t) &= \mathbb{C}\mathrm{ov}(f^V(s), f^V(t)) \\
&= \mathbb{C}\mathrm{ov}(V^{(r)}, V^{(z)}) \\
&= (C_V)_{rz}.
\end{aligned}$$

Let $\{\beta_j, w_j\}_{j=1}^p$ be the eigenvalues and orthonormal eigenvectors of C_V. It follows that

$$\begin{aligned}
C_V = \sum_{j=1}^{p} \beta_j w_j w_j^{\mathrm{T}} &\Leftrightarrow (C_V)_{rz} = \sum_{j=1}^{p} \beta_j w_j^r w_j^z \\
&\Leftrightarrow \mathcal{K}_{f^V}(s,t) = \sum_{j=1}^{p} \beta_j w_j^r w_j^z \\
&\Leftrightarrow \mathcal{K}_{f^V}(s,t) = \sum_{j=1}^{p} \beta_j \left(\sum_{i=1}^{p} w_j^{(i)} \mathbb{1}_{(i-1,i]}(s)\right) \left(\sum_{i=1}^{p} w_j^{(i)} \mathbb{1}_{(i-1,i]}(t)\right) \\
&\Leftrightarrow \mathcal{K}_{f^V}(s,t) = \sum_{j=1}^{p} \beta_j e_j(s) e_j(t).
\end{aligned}$$

That is, $\{\beta_j, e_j\}_{j=1}^p$ are orthonormal eigenvalues and eigenfunctions of $\mathcal{K}_{f^V}$.

Figures and Tables

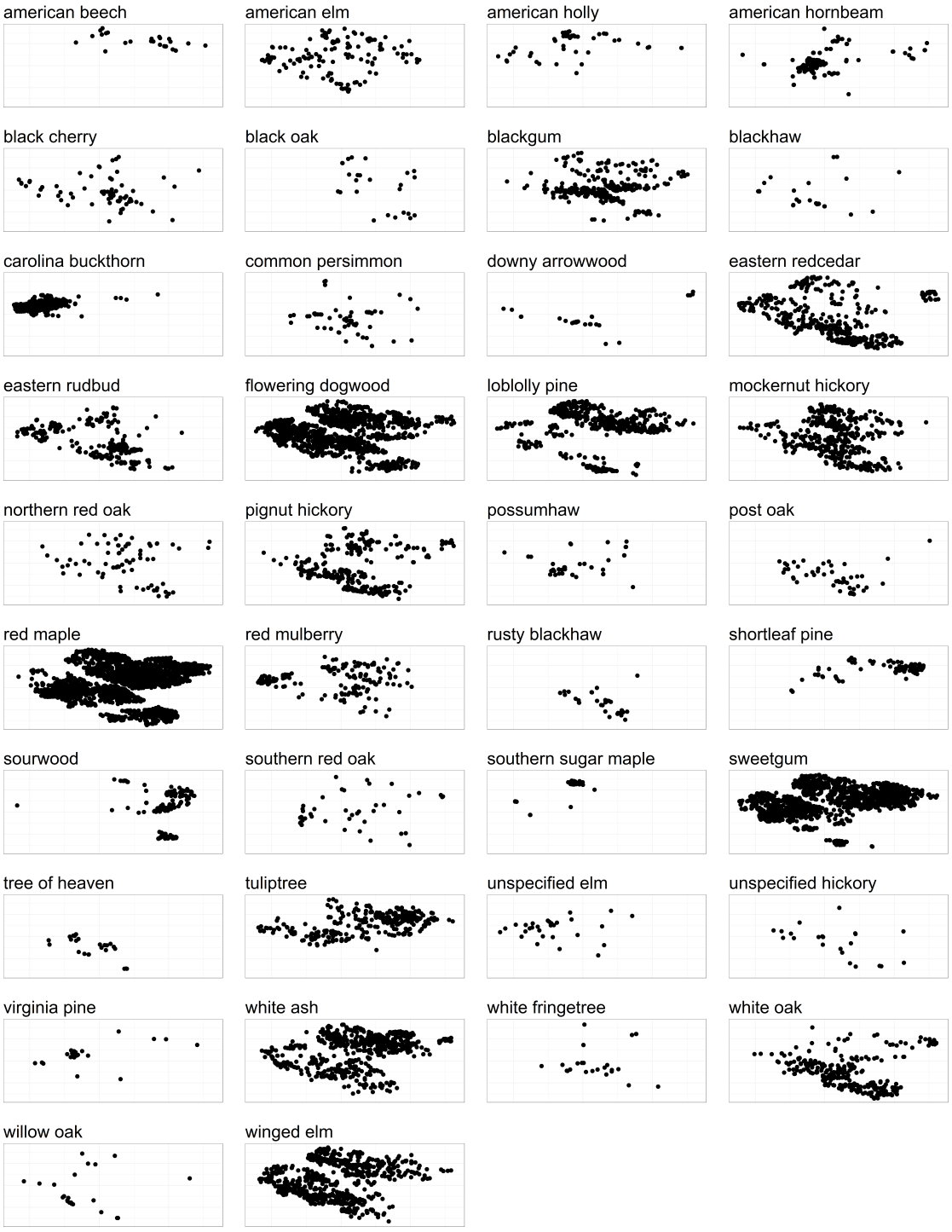

Figure 25: Coordinates of all tree species in the Duke forest data set.

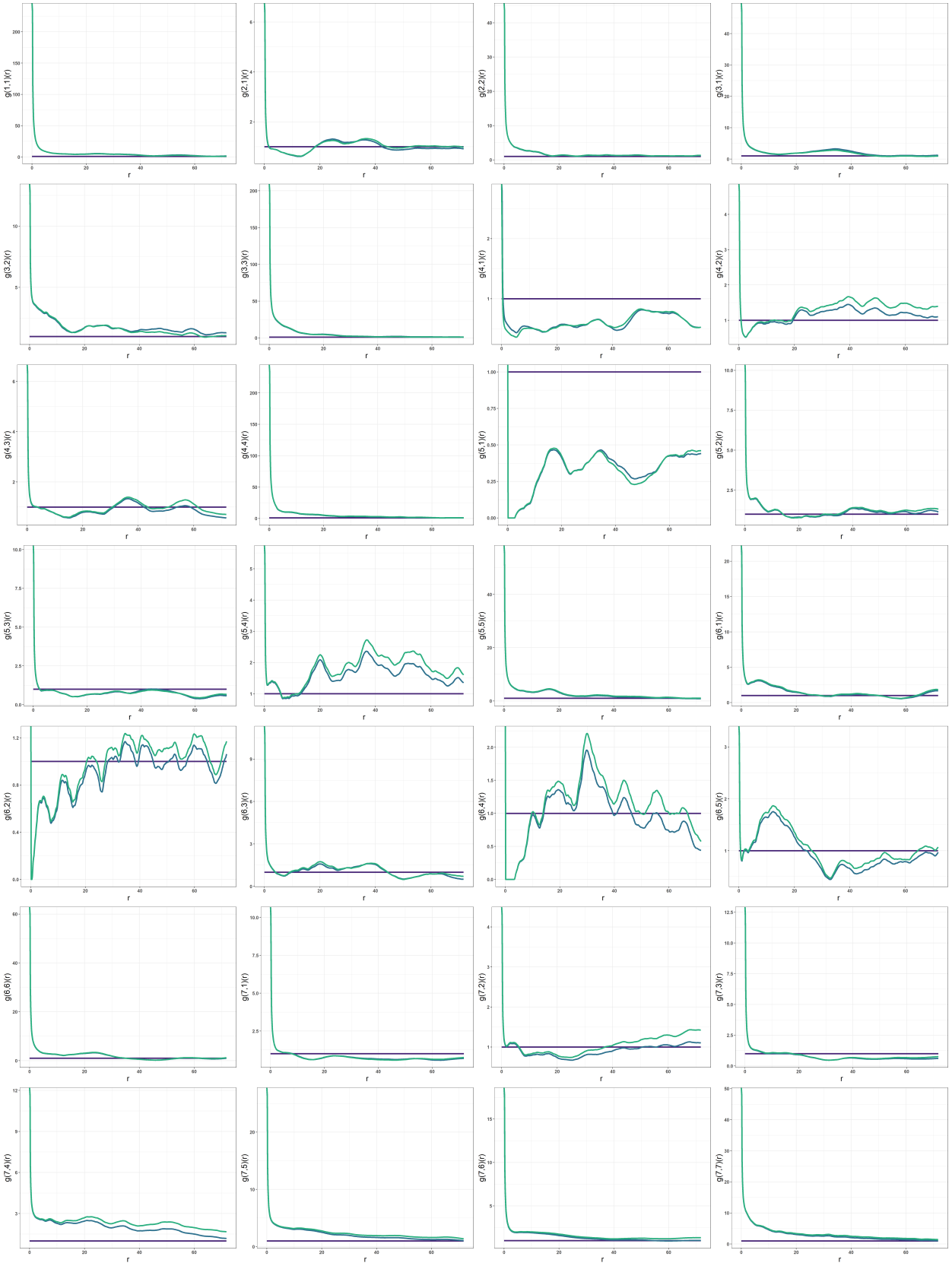

Figure 26: Cross pair correlation functions for the first eight tree species in the Duke forest data set (American Beech to Blackhaw).

Tree species	CEI	$\overline{R}$	SK	$\hat{g}(6.87)$	$r^{(1)}$	n
American Beech	0.33	2.48	1.72	6.50	71.75	30
American Elm	0.59	2.22	1.35	2.85	71.75	139
American Holly	0.42	2.20	2.08	13.65	71.75	54
American Hornbeam	0.16	2.48	5.05	9.94	61.66	191
Black Cherry	0.75	2.24	1.69	3.74	63.76	70
Black Oak	0.51	2.66	1.00	2.68	54.94	25
Blackgum	0.50	2.32	1.91	5.73	71.75	382
Blackhaw	0.49	2.44	12.04	9.00	48.07	32
Carolina Buckthorn	0.19	2.16	1.69	24.89	67.13	1085
Common Persimmon	0.58	2.53	1.89	6.34	71.75	59
Downy Arrowwood	0.16	2.64	8.02	10.23	43.72	33
Eastern Redcedar	0.66	2.17	1.41	2.73	71.47	472
Eastern Rudbud	0.41	2.24	1.73	5.24	68.25	294
Flowering Dogwood	0.60	2.20	1.34	2.34	71.75	1610
Loblolly Pine	0.65	1.98	1.09	3.61	71.75	468
Mockernut Hickory	0.66	2.17	1.15	2.27	71.75	457
Northern Red Oak	0.73	2.19	1.09	1.21	57.18	77
Pignut Hickory	0.52	2.27	1.30	3.73	71.75	313
Possumhaw	0.19	2.47	11.86	7.24	71.75	56
Post Oak	0.54	2.42	1.27	4.32	71.75	48
Red Maple	0.54	2.18	1.70	2.27	58.02	3394
Red Mulberry	0.56	2.26	1.33	5.10	54.94	180
Rusty Blackhaw	0.21	2.70	2.27	9.12	71.75	36
Shortleaf Pine	0.38	2.16	1.30	12.25	65.45	82
Sourwood	0.24	2.24	2.35	8.69	71.75	153
Southern Red Oak	0.72	2.34	1.37	3.31	27.05	44
Southern Sugar Maple	0.17	2.43	3.22	54.50	71.75	50
Sweetgum	0.56	2.11	1.45	3.25	71.75	1999
Tree of Heaven	0.27	2.51	0.99	10.21	71.75	24
Tuliptree	0.46	2.31	1.55	5.21	71.75	409
Unspecified Elm	0.93	2.44	1.41	3.48	71.75	29
Unspecified Hickory	0.64	2.50	1.30	2.82	29.01	22
Virginia Pine	0.24	2.57	3.06	21.37	55.22	25
White Ash	0.58	2.14	1.34	3.07	71.75	762
White Fringetree	0.37	2.24	1.74	6.45	71.75	27
White Oak	0.63	2.10	1.22	3.47	71.75	276
Willow Oak	0.63	2.40	2.32	9.18	43.44	22
Winged Elm	0.61	2.09	1.37	3.24	71.75	885

Table 8: Numerical summary characteristics CEI, $\overline{R}$, SK, $\hat{g}(6.87)$, $r^{(1)}$ and the number of points n in the pattern for each species in the Duke forest data set.

Declaration of Academic Honesty

I, Bianca Neubert, hereby declare that I have not previously submitted the present work for other examinations. I wrote this work independently. All sources, including sources from the Internet, that I have reproduced in either an unaltered or modified form (particularly sources for texts, graphs, tables and images), have been acknowledged by me as such. I understand that violations of these principles will result in proceedings regarding deception or attempted deception.

Berlin, December 23, 2021

Bianca Neubert