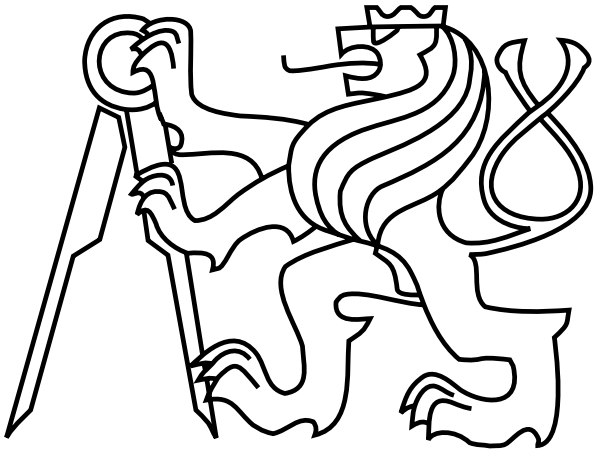
Czech Technical University in Prague
Faculty of Nuclear Sciences and Physical Engineering

Department of Physics



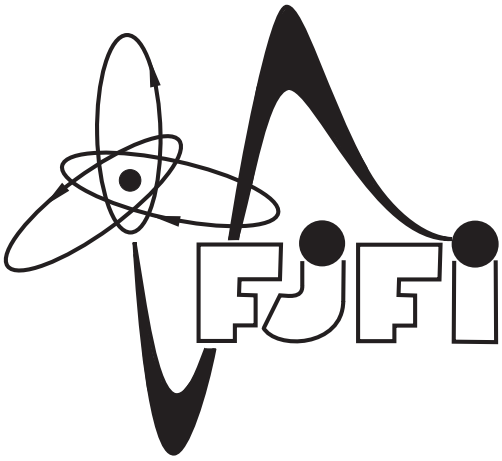
Built-up Structure Analysis Using Stochastic Geometry

DOCTORAL THESIS

November 24, 2014 Ing. Daniel Vašata

Czech Technical University in Prague
Faculty of Nuclear Sciences and Physical Engineering

Department of Physics



Built-up Structure Analysis Using Stochastic Geometry

by

Ing. Daniel Vašata

Supervisor: Prof. RNDr. Petr Šeba, DrSc.

Dissertation submitted to
the Faculty of Nuclear Sciences and Physical Engineering of
Czech Technical University in Prague
in partial fulfillment of the requirements
for the degree of

Doctor of Philosophy

in the branch of study Mathematical Physics
of the study program Mathematical Engineering

Prague, September 2014

This thesis is the result of my own work, except where explicit reference is made to the work of others and has not been submitted for another qualification to this or any other university.

Daniel Vašata

Acknowledgement

I would like to thank my supervisor Prof. Petr Šeba for an inspiring motivation, discussions and a patience throughout the whole time of my Ph.D. studies. Besides him, I also wish to thank Prof. Pavel Exner for his encouragement. Next, I would like to acknowledge the support from the grant No. 202/08/H072 of the Czech Science Foundation and grant No. LC06002 of the Czech Ministry of Education, Youth and Sport.

Last but not least, I would like to thank my wife, Hana and the rest of my family, who supported and encouraged me during my long studies.

Contents

List of Symbols

This index contains only the notation used throughout the thesis. Symbols with localized usage are omitted, as are standard symbols such as e and π.

Special sets

Symbol	Description	Page
$\mathbb{N}$	Natural numbers, $\mathbb{N} = \{1, 2, 3, \dots\}$.	
$\mathbb{R}$	Real numbers.	
(a, b)	Open interval in $\mathbb{R}$ from a to b.	
$[a, b]$	Closed interval in $\mathbb{R}$ from a to b.	
A°	Topological interior of a set A.	
$\bar{A}$	Topological closure of a set A.	
∂A	Topological boundary of a set A.	
$\mathbb{R}^d$	d-dimensional Euclidean space	5
B^d	Closed unit ball in $\mathbb{R}^d$.	5
U^d	Open unit ball in $\mathbb{R}^d$.	5
$B_r(\boldsymbol{x})$	Closed ball in $\mathbb{R}^d$ with radius r centred at $\boldsymbol{x}$.	5
$U_r(\boldsymbol{x})$	Open ball in $\mathbb{R}^d$ with radius r centred at $\boldsymbol{x}$.	5
$(\boldsymbol{a}, \boldsymbol{b}]$	Interval in $\mathbb{R}^d$.	6
C_0	Half-open unit cube in $\mathbb{R}^d$.	9
C_ε	The ε neighbourhood of a set C.	51, 188
$\operatorname{supp} \mu$	Support of a Borel measure μ.	19
$E^k_{\neq}$	Complete non-diagonal part of E^k	21
$\mathcal{B}(E)$	Borel σ-algebra of a topological space E.	6, 178
$C_b(S)$	Bounded, continuous real functions on a metric space S.	179

Measures

Symbol	Description	Page
$\mathbb{P}$	A probability measure.	6
$\mathbb{E}$	Expectation with respect to given probability measure.	6
ν_d	d-dimensional Lebesgue measure.	6
$\mathbb{P}_X$	Distribution of a random closed set X.	6
T_X	Capacity functional of a random closed set X.	7
$\mathbb{P}_\xi$	Distribution of a random measure ξ.	19
ν_X	Volume measure of a random closed set X in $\mathbb{R}^d$.	27
Γ	Bartlett spectrum.	13, 28
$\mathcal{H}^\alpha(A)$	α-dimensional Hausdorff measure.	187
$\mathcal{M}^\alpha(A)$	α-dimensional Minkowski content.	188
$\mu_i \xrightarrow{\text{w}} \mu$	Weak limit of finite measures.	179

Characteristics of random closed sets and random measures

Symbol	Description	Page
$m(\boldsymbol{x})$	Volume fraction of a random closed set.	10
p	Volume fraction of a stationary random closed set.	10
$C(\boldsymbol{x},\boldsymbol{y}), C(\boldsymbol{r})$	Covariance.	10
$\mathrm{cov}(\boldsymbol{x},\boldsymbol{y}), \mathrm{cov}(\boldsymbol{r})$	Covariance function.	10
$\kappa(\boldsymbol{x},\boldsymbol{y}), \kappa(\boldsymbol{r})$	Correlation function.	11
Λ	Intensity measure of a random measure.	21
$\Lambda^{(k)}$	kth order moment measure.	21
$\breve{\Lambda}^{(k)}$	Reduced kth order moment measure.	23
$\breve{\Lambda}^{[k]}$	Reduced kth order factorial moment measure.	23
$C^{(2)}$	Second order covariance measure.	22
$\breve{C}^{(2)}$	Reduced second order covariance measure.	23
$\breve{C}^{[2]}$	Reduced second order factorial covariance measure.	23

Statistics

Symbol	Description	Page
$\hat{p}_v$	Empirical volume fraction.	71
$\hat{p}_p$	Discrete empirical volume fraction.	79
$\hat{m}_h$	Kernel estimator of the volume fraction.	82
$\hat{m}_{A;h}$	Approximative volume fraction kernel estimator.	133
$\hat{m}_{p;h}$	Discrete volume fraction kernel estimator.	142
$\hat{C}_p(\boldsymbol{r})$	Discrete empirical covariance.	79
$\hat{\kappa}_p(\boldsymbol{r})$	Discrete empirical correlation function.	79
$\hat{\kappa}_p^{\bullet}(\boldsymbol{r})$	Discrete intrinsically balanced empirical correlation function.	80
$\hat{\kappa}_p^{AI}(r)$	Adapted isotropic empirical correlation function.	81
$\hat{\kappa}_p^{\bullet AI}(r)$	Adapted isotropic int. balanced covariance function estimator.	111
$\hat{\kappa}_{p;h}(\boldsymbol{r})$	Discrete correlation function estimator.	90
$\hat{\kappa}_{p;h}^{AI}(\boldsymbol{r})$	Adapted isotropic correlation function estimator.	134
$\overline{\kappa}_W(\boldsymbol{r})$	Spatially averaged correlation function.	146
$\overline{\kappa}_W^I(\boldsymbol{r})$	Isotropised, spatially averaged correlation function.	147
$\mathrm{MISE}(\hat{m}_h)$	Mean integrated square error of $\hat{m}_h$.	83
$\mathrm{MISE}_{h_0}(\hat{m}_h)$	Approximate mean integrated square error of $\hat{m}_h$ based on h_0.	136
$\overline{\hat{f}}$	Sample mean of an estimator $\hat{f}$.	91
$s_N(\hat{f})$	Sample standard deviation of an estimator $\hat{f}$.	91
$d_N(\hat{f}; f)$	Deviation of an estimator $\hat{f}$ from the true value f.	91
$I_{0.9}(\hat{f})$	Two sided 90% confidence interval of $\hat{f}$.	91
$\hat{\mathrm{var}}_v^M\big(\nu_X(B_r)\big)$	Discrete estimator of the variance in balls.	119

Miscellaneous

Symbol	Description	Page
$\boldsymbol{x} \leq \boldsymbol{y}$	Inequality in $\mathbb{R}^d$.	5
$\gamma_B(\boldsymbol{r})$	Set covariance of B.	25
$\overline{\gamma}_B(\boldsymbol{r})$	Isotropised set covariance of B.	26, 77
$\dim_H A$	Hausdorff dimension of a set A.	188
c_α	Generalized volume of an α dimensional unit ball.	188
$\mathcal{O}$	Landau big O notation.	
$\sim$	Landau asymptotic equality.	44

Chapter 1

Introduction

Cities certainly represent one of the most significant fingerprints of human activity on the Earth. The creation and development of their structure is influenced by cultural, sociological, economic, political, and other conditions. There were several attempts in the past to understand the structure of cities and their neighbourhood. Despite the apparent complexity, some simple universal properties and rules were found. The classic example is the rank size distribution of cities which was according to [1] firstly mentioned by Auerbach in 1913 and later discussed by Zipf [2]. They claimed that if the cities are ranked by the number of inhabitants, then the rank-size distribution follows a power law with the exponent close to -1 (see also [1, 3]).

In the last 20 years much attention was drawn by the spatial analysis of urban structures. A comprehensive review about both static and dynamic properties is given by Batty in [4] and an overview of structural properties is given by Schweitzer in [1].

The probably most visible morphological property of urban structure is the clustering of the built-up area and free space that appears over various scales. As many researchers agree, the prominent role in describing those properties is played by the fractal analysis.

1.1 Fractal properties of the built-up structure

The connection between fractal properties and urban systems goes back to 1980's (see [5, 6, 7]). Systematic development of qualitative characterization of fractal properties is given in more recent publications, see e.g. [8, 9, 10, 11, 1, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21].

The authors usually argue that fractals appearing in urban analysis are self-similar, repeating their structure over different scales, i.e. from parts of a city through whole cities and metropolitan areas to wide urban areas. An urban structure is usually analysed using the binary lattice representation given by black-and-white maps of a certain resolution with the best values around 16 m^2 per one pixel in recent publications, [20]. Black pixels stand for occupied, built-up sites and white pixels for empty, non building sites. The difference between various types of buildings is not taken into account. Fractal properties of such maps are characterized by fractal dimensions.

There are generally three different fractal dimensions and four methods how to estimate them used in connection with the built-up structure. In the following parsimonious inventory they are called exactly as in referred publications, which is consistent with the terminology used in the overview [22]. Those are the box-counting dimension estimated by a grid method [8, 10, 11, 15] and by a dilation method [10, 12], the correlation dimension [9, 10, 1, 12, 14, 17, 20], and the radial dimension [8, 9, 10, 1, 15, 18, 21]. Numerical values of fractal dimensions obtained using those methods are always between 1 and 2. The most common values are between 1.5 and 1.9. Moreover, it is sometimes noted that the correlation dimension is more reliable than other methods, [12, 20].

Radial analysis provides specific information about a spatial organisation around a specific point, usually the centre of a city. The number $N_r(R)$ of black pixels which lies within a circle of radius R centred at that point is counted. The fractal law then takes the following form

$$N_r(R) = CR^{D_r}$$

for some range of R, where D_r is called the radial (mass) dimension. Its estimation is performed using linear regression on the logarithm of the previous relation. It is often mentioned by some authors [9, 10, 1, 18] that varying R, there are regions with different slopes corresponding to different fractal dimensions. In particular there appears to be a region, around the city core, where the radial dimension $D_r = 2$ (see [1, 23]).

It should be noted that the above terminology used in the urban context often differs from the terminology used in mathematical publications like [24, 25]. Also the power law relations that lead to fractal dimensions are not analysed in the limit as the length factor goes to 0 but only within a certain range of values.

1.2 Physical properties and existing models

From the physical point of view the built-up structure shows significant analogies to critical systems in statistical physics, that is of systems that are at critical points of second order phase transitions. It is often mentioned, e.g. in [9, 1, 10, 26, 18, 16], that the built-up structure in a city is self-organized since it results from a large number of decisions on various scales (individuals, private sector groups, local authorities). The self-organization then leads to typical features of critical systems (see for instance [27, 28]) like cluster formation, aggregation, or percolation phenomena.

Observed fractal properties lead many authors to simulate the growth of urban clusters by means of physical growth models like the diffusion-limited aggregation and dielectric breakdown model, [29, 30, 31, 8]. The structure generated by those models follows the fractal properties. However, they can only produce connected clusters, which is not sufficient for the full description of the observed properties.

To overcome this problem, several methods were developed. In [1, 9] Schweitzer and Steinbing introduced the kinetic model of Brownian agents. They use three types of agents representing the already constructed built-up area, the free space, and the demand for built-up growth. The existing built-up area creates a spatio-temporal attractive field that follows the reaction-diffusion dynamics. The agents representing the demand diffuse from the centres, meet there the free space, and eventually transform into a constructed built-up area. This reaction-diffusion approach can reflect the attraction of the existing built-up area, the movement of growth zones into outer regions, and thus the depletion of free space in the centres.

Another model is a correlated percolation model introduced by Makse et al. in [32, 33]. They model urban population occupancy using percolation in the presence of the exponential radial decrease of the population density $\rho(r)$ from the city centre and under long-range power law decaying correlations between points.

Besides physical models, White and Engelen in [34] developed a model of the urban land-use dynamics based on the cellular automata (CA) technique, which was later studied extensively and generalized (for a references see e.g. [35, 36]). In this approach the land use in each cell is qualitatively represented (unlike the binary representation in previous models) in terms of the land use type. The cell state is updated each time step with respect to transition rules depending on the states of neighbouring cells. The neighbourhood is typically small, but regardless of that fact, after sufficient iteration time the interaction is able to propagate over long distances and therefore the typical properties of urban systems emerge.

Instead of analysing the spatial distribution of one particular land type it is also possible to focus on the morphology of the division of the whole land into land parcels as was done in [37, 38]. One usually observes the power law distribution of the parcel sizes in the urban

areas. In rural areas the power law holds for the distribution of the overall land area owned by individuals.

1.3 Outline of the thesis

Our aim is to analyse the built-up structure in cities using the framework of stochastic geometry and especially of random closed sets. Such an approach is suitable since stochastic geometry is well known for its ability to characterise randomly generated patterns and it is natural to understand the built-up structure of a given city as randomly generated.

Throughout the analysis the built-up structure is represented by the collection of individual buildings that are taken as two dimensional subsets of the plane. Hence, unlike in the previous studies, there is no rough coarse graining that yields grid maps of certain resolution. The accuracy of our data is sufficient enough when compared to individual building sizes and thus we may analyse the structure also on small scales.

The thesis is organized as follows: After this introduction Chapter 2 is devoted to the theoretical background of stochastic geometry. The random closed sets are recalled together with their basic characteristics and several important models. Then random measures are introduced together with special cases given by point and particle processes. The notion of the Bartlett spectrum is defined and the long-range dependence of random measures and random closed sets is described.

Next, Chapter 3 provides a detailed construction of the special centre function that can be used for some particle processes. It is given by the centroid and its two generalizations to certain sets of zero Lebesgue measure.

Chapter 4 deals with statistical issues concerning the estimation of basic characteristics of random closed sets. We focus especially on second order properties of the stationary case and on the volume fraction estimation for the non-stationary case. The numerical simulations are performed to analyse the properties of several proposed estimators. For the convenience of the reader selected mathematical issues important for deriving results in Chapters 2 and 3 are reviewed in Appendix A, thus making our exposition self-contained.

Chapter 5 is devoted to the analysis of the built-up structure. We focus especially on the second order characteristics of both stationary and non-stationary approach. The long-range dependence is here clearly observed and discussed. We also study the radial density dependence and distribution of building sizes.

Finally, we recapitulate major findings of the thesis and discuss a possible future research.

Chapter 2

Preliminaries from stochastic geometry

In this chapter the basic mathematical notation and background from stochastic geometry is introduced. In particular, we focus on random closed sets, random measures, and their special cases like random point processes and particle processes. Known facts are presented with references and only new assertions are proved. Although the built-up area is studied as a subset of $\mathbb{R}^2$, the theory in the following is introduced for general $\mathbb{R}^d$. Moreover, even more general space is used when necessary.

In the first section, some basic notation is summarized. Then in Section 2.2 random closed sets are introduced and theirs important properties and characteristics are defined. As a first important model of random closed sets we introduce the level excursion set in Section 2.4. Then the generalization given by random measures is reviewed. We especially focus on the precise definition of the covariance measure and corresponding Bartlett spectrum. In Sections 2.6 and 2.7 the marked point process and the particle process are presented and their relationship is discussed. Finally, Section 2.8 deals with long-range dependence of random measures and random closed sets. The connection to the Bartlett spectrum is shown and a particularly important sub-case of isotropic-long range dependence is introduced.

2.1 Basic notation

Let $\mathbb{R}^d$ denote the d-dimensional Euclidean space of real numbers. The notation $\boldsymbol{x} \cdot \boldsymbol{y}$ is used for the standard Euclidean scalar product defined by

$$\boldsymbol{x} \cdot \boldsymbol{y} = \sum_{i=1}^{d} x_i y_i,$$

for each $\boldsymbol{x} = (x_1, \ldots, x_d)$ and $\boldsymbol{y} = (y_1, \ldots, y_d)$ in $\mathbb{R}^d$. Instead of $\boldsymbol{x} \cdot \boldsymbol{x}$ we write $\boldsymbol{x}^2$. The Euclidean norm $\|\boldsymbol{x}\|$ is as usual given by

$$\|\boldsymbol{x}\| = \sqrt{\boldsymbol{x} \cdot \boldsymbol{x}}$$

for each $\boldsymbol{x} \in \mathbb{R}^d$. The distance between $\boldsymbol{x}$ and $\boldsymbol{y}$ is $d(\boldsymbol{x}, \boldsymbol{y}) = \|\boldsymbol{x} - \boldsymbol{y}\|$. The distance from $\boldsymbol{x}$ to a set $A \subset R^d$ is defined to be $d(\boldsymbol{x}, A) = \inf\{d(\boldsymbol{x}, \boldsymbol{y}) | \boldsymbol{y} \in A\}$. The d-dimensional open and closed unit ball centred at the origin $\mathbf{0} = (0, \ldots, 0)$ are given by

$$U^d = \{\boldsymbol{x} \in \mathbb{R}^d \mid \|\boldsymbol{x}\| < 1\} \quad \text{and} \quad B^d = \{\boldsymbol{x} \in \mathbb{R}^d \mid \|\boldsymbol{x}\| \leq 1\},$$

respectively. The d-dimensional open and closed ball of radius r centred at $\boldsymbol{x}$ are

$$U_r(\boldsymbol{x}) = \{\boldsymbol{y} \in \mathbb{R}^d \mid \|\boldsymbol{x} - \boldsymbol{y}\| < r\} \quad \text{and} \quad B_r(\boldsymbol{x}) = \{\boldsymbol{y} \in \mathbb{R}^d \mid \|\boldsymbol{x} - \boldsymbol{y}\| \leq r\},$$

respectively. By $<, \leq, >, \geq$ on $\mathbb{R}^d$ we mean the simultaneous inequalities in all coordinates, i.e. $\boldsymbol{x} < \boldsymbol{y}$ if and only if $x_i < y_i$ for all $i = 1, \ldots, d$. For $\boldsymbol{a}, \boldsymbol{b} \in \mathbb{R}^d$ we write $(\boldsymbol{a}, \boldsymbol{b}) =$

$(a_1, b_1) \times \ldots \times (a_d, b_d) \subset \mathbb{R}^d$ and analogously for all other types of intervals in $\mathbb{R}^d$. In particular we write $(-\infty, \boldsymbol{a}] = \{\boldsymbol{x} \in \mathbb{R}^d | \boldsymbol{x} \leq \boldsymbol{a}\}$.

The **Minkowski addition** $A + B$ and the **Minkowski reflection** $\check{A}$ of $A, B \subset \mathbb{R}^d$ are defined to be

$$A + B = \{\boldsymbol{x} + \boldsymbol{y} | \boldsymbol{x} \in A, \boldsymbol{y} \in B\} \quad \text{and} \quad \check{A} = \{-\boldsymbol{x} | \boldsymbol{x} \in A\},$$

respectively.

The family of all Borel sets of $\mathbb{R}^d$ is denoted by $\mathcal{B}(\mathbb{R}^d)$. By ν_d (or ν) we denote the d-dimensional Lebesgue measure on the measurable space $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$. Note, that up to a normalizing constant, ν_d is the only translation-invariant measure on $\mathcal{B}(\mathbb{R}^d)$. In integrals the notation $\nu_d(\mathrm{d}\boldsymbol{x}) = \mathrm{d}\boldsymbol{x}$ is used. If f is a non-negative measurable function then by $f\nu_d$ we mean a measure defined by $(f\nu_d)(B) = \int_B f(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}$ for all $B \in \mathcal{B}(\mathbb{R}^d)$. Clearly, $f\nu_d$ is absolutely continuous with respect to ν_d, and f is the corresponding density. When we work with a probability space $(\Omega, \mathcal{A}, \mathbb{P})$, the expectation of a random variable with respect to the probability measure $\mathbb{P}$ is denoted by $\mathbb{E}$.

2.2 Random closed sets

For further connection to particle processes it is useful to define random closed sets not only in $\mathbb{R}^d$ but in more general space. In the following, the standard construction of random closed sets is briefly presented. For a fuller treatment we refer the reader to [39, 40, 41, 42].

Let E be a locally compact, Hausdorff topological space with a countable base. Let $\mathcal{G}, \mathcal{F}$, and $\mathcal{C}$ denote the class of all open, closed, and compact subsets of E, respectively. For any $A \subset E$ we define

$$\mathcal{F}_A = \{F \in \mathcal{F} | F \cap A \neq \emptyset\} \quad \text{and} \quad \mathcal{F}^A = \{F \in \mathcal{F} | F \cap A = \emptyset\}.$$

In Section A.4 is shown that the collection of sets of the form

$$\mathcal{F}^C \cap \mathcal{F}_{G_1} \cap \ldots \cap \mathcal{F}_{G_n},$$

where $C \in \mathcal{C}$ is compact and $G_1, \ldots, G_n \in \mathcal{G}$ are open, is a base of a topology on $\mathcal{F}$ called the Fell topology which is locally compact, Hausdorff separable, and with a countable base. Let us denote by $\mathcal{B}(\mathcal{F})$ the Borel σ-algebra generated by the Fell topology on $\mathcal{F}$.

Definition 2.2.1. Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space and $(\mathcal{F}, \mathcal{B}(\mathcal{F}))$ be a measurable space introduced above. The map

$$X : \Omega \to \mathcal{F}$$

is called the **random closed set** if X is $\mathcal{A} - \mathcal{B}(\mathcal{F})$ measurable, i.e.

$$X^{-1}(F) = \{\omega \in \Omega | X(\omega) \in F\} \in \mathcal{A}$$

for each $F \in \mathcal{B}(\mathcal{F})$.

The definition means that probabilities can be assigned to the statement that X hits the countable collection of open sets and miss any of another countable collection of compact sets. The measurability property of X implies the probability measure $\mathbb{P}_X$ on $(\mathcal{F}, \mathcal{B}(\mathcal{F}))$ according to $\mathbb{P}_X(F) \equiv \mathbb{P}(X \in F) = \mathbb{P}(X^{-1}(F))$ for all $F \in \mathcal{B}(\mathcal{F})$ which is called the **distribution** of X. Two random closed set X and X' are **stochastically equivalent**, $X \overset{\mathcal{D}}{\sim} X'$, if they have the same distribution.

2.2.1 Capacity functional and Choquet's theorem

The distribution of a random closed set can be effectively characterised by the restriction of the probability to particular subclass $\mathcal{F}_C$ of $\mathcal{B}(\mathcal{F})$. The result is the **capacity functional** $T_X : \mathcal{C} \to [0,1]$ of a random closed set X defined by

$$T_X(C) = \mathbb{P}_X(\mathcal{F}_C) = \mathbb{P}(X \cap C \neq \emptyset)$$

for all $C \in \mathcal{C}$.

The capacity functional returns the probability of hitting a given compact set. It can be regarded as the analogon of a distribution function of a random variable in $\mathbb{R}$. In particular, it determines the distribution of X uniquely. This result is known as Choquet's theorem.

Theorem 2.2.1 (Choquet's theorem)**.** *A functional $T : \mathcal{C} \to [0,1]$ such that $T(\emptyset) = 0$ is the capacity functional of a necessarily unique random closed set if and only if T satisfies the following conditions:*

- *$T(C_n) \searrow T(C)$ as $C_n \searrow C$ in $\mathcal{C}$*
- *$\Delta_{C_n} \cdots \Delta_{C_1} T(C) \leq 0$ for $n \geq 1$ and $C, C_1, \ldots, C_n \in \mathcal{C}$, where the successive differences are for $n = 1$ given by*

  $$\Delta_{C_1} T(C) = T(C) - T(C \cup C_1)$$

  and for $n \geq 2$ by

  $$\Delta_{C_n} \cdots \Delta_{C_1} T(C) = \Delta_{C_{n-1}} \cdots \Delta_{C_1} T(C) - \Delta_{C_{n-1}} \cdots \Delta_{C_1} T(C \cup C_n).$$

Proof. Theorems 2.1.2 and 2.2.1 in [40] or Section 1.3 in [42]. □

2.2.2 Properties of random closed sets

The joint distribution of random closed sets $X_1, \ldots, X_k$ defined on the same probability space $(\Omega, \mathcal{A}, \mathbb{P})$ is the probability measure

$$\mathbb{P}_{X_1,\ldots,X_k}(F_1, \ldots, F_k) = \mathbb{P}(X_1 \in F_1, \ldots, X_k \in F_k)$$

for all $F_1, \ldots, F_k \in \mathcal{B}(\mathcal{F})$.

Definition 2.2.2. Random closed sets $X_1, \ldots, X_k$ are **independent** if for arbitrary $F_1, \ldots, F_k \in \mathcal{B}(\mathcal{F})$,

$$\mathbb{P}_{X_1,\ldots,X_k}(F_1, \ldots, F_k) = \mathbb{P}(X_1 \in F_1) \cdots \mathbb{P}(X_k \in F_k).$$

As follows from Choquet's theorem 2.2.1, the independence condition can be equivalently written as

$$T_{X_1,\ldots,X_k}(C_1, \ldots, C_k) = T_{X_1}(C_1) \cdots T_{X_k}(C_k)$$

for all compact $C_1, \ldots, C_k$.

Let G be a topological group that acts measurably on the space E. Then G acts in a natural way on $\mathcal{F}$ by letting

$$S_g F = \{gx | x \in F\} \quad \text{for} \quad F \in \mathcal{F}. \tag{2.1}$$

From Lemma A.4.1 follows that the mapping $F \mapsto S_g F$ from $(\mathcal{F}, \mathcal{B}(\mathcal{F}))$ into itself is continuous and consequently measurable. Hence, if X is a random closed set and $g \in G$, then $S_g X$ is a random closed set.

In the following let E be the d-dimensional Euclidean space $\mathbb{R}^d$. Important cases arise if G is the group of all translations or all rotations of $\mathbb{R}^d$. Note that in the case of the group of translations, if $\boldsymbol{y} \in \mathbb{R}^d$ then $\boldsymbol{y}(\boldsymbol{x}) = \boldsymbol{x} + \boldsymbol{y}$ and so $S_{\boldsymbol{y}} F = F + \boldsymbol{y}$ for all $F \in \mathcal{F}$. Since $S_{\boldsymbol{y}} X$ is a random closed set we obtain $\mathbb{P}_{X+\boldsymbol{y}}(G) = \mathbb{P}_X(G - \boldsymbol{y})$ for all $G \in \mathcal{B}(\mathcal{F})$.

Definition 2.2.3. A random closed set X is called **stationary** if $X \overset{\mathcal{D}}{\sim} (X+\boldsymbol{y})$ for all $\boldsymbol{y} \in \mathbb{R}^d$. It is called **isotropic** if $X \overset{\mathcal{D}}{\sim} \theta X$ for all rotations $\theta \in SO_d$.

The probabilities $\mathbb{P}_X$ and $\mathbb{P}_{X+\boldsymbol{y}}$ induced by a stationary random closed set X satisfy

$$\mathbb{P}_X(F) = \mathbb{P}_{X+\boldsymbol{y}}(F)$$

for every $F \in \mathcal{B}(\mathcal{F})$ and $\boldsymbol{y} \in \mathbb{R}^d$. This particularly gives

$$\begin{aligned}\mathbb{P}(\boldsymbol{y} \in X) = \mathbb{P}(X \cap \{\boldsymbol{y}\} \neq \emptyset) = \mathbb{P}_X(\mathcal{F}_{\{\boldsymbol{y}\}}) = \mathbb{P}_{X+\boldsymbol{y}}(\mathcal{F}_{\{\boldsymbol{y}\}}) = \\ = \mathbb{P}(X+\boldsymbol{y} \cap \{\boldsymbol{y}\} \neq \emptyset) = \mathbb{P}(\{\boldsymbol{0}\} \cap X \neq \emptyset) = \mathbb{P}(\boldsymbol{0} \in X). \quad (2.2)\end{aligned}$$

The stationarity is a property that implies strong consequences on the structure of random closed set as will be also seen later.

Both stationarity and isotropy of a random closed set X can be equivalently formulated using the capacity functional T_X.

Theorem 2.2.2. *The random closed set X is stationary if and only if its capacity functional T_X is translation invariant, i.e.*

$$T_X(C) = T_X(S_{\boldsymbol{y}}C) = T_X(C+\boldsymbol{y})$$

for every $C \in \mathcal{C}$ and $\boldsymbol{y} \in \mathbb{R}^d$, and it is isotropic if and only if T_X is rotation invariant, i.e.

$$T_X(C) = T_X(S_\theta C) = T_X(\theta C)$$

for every $C \in \mathcal{C}$ and $\theta \in SO_d$.

Proof. Theorem 2.4.5 in [40]. □

Corollary 2.2.1. *Let X be a stationary random closed set in $\mathbb{R}^d$ and $\boldsymbol{r} \in \mathbb{R}^d$. Then $X \cap (X-\boldsymbol{r})$ is a stationary random closed set.*

Proof. The mapping $\omega \mapsto X(\omega) \cap (X(\omega)-\boldsymbol{r})$ from $(\Omega, \mathcal{A})$ to $(\mathcal{F}, \mathcal{B}(\mathcal{F}))$ is measurable as follows from Lemma A.4.1 and hence $X \cap (X-\boldsymbol{r})$ is a random closed set. Stationarity follows from

$$\begin{aligned}T_{X\cap(X-\boldsymbol{r})}(C) &= \mathbb{P}\left(X \cap (X-\boldsymbol{r}) \cap C \neq \emptyset\right) = \mathbb{P}\left(X \cap C \neq \emptyset, (X-\boldsymbol{r}) \cap C \neq \emptyset\right) \\ &= \mathbb{P}\left(X \cap C \neq \emptyset, X \cap C + \boldsymbol{r} \neq \emptyset\right) = \mathbb{P}\left(X \cap C \cap (C+\boldsymbol{r}) \neq \emptyset\right) = T_X\big(C \cap (C+\boldsymbol{r})\big),\end{aligned}$$

which holds for every $C \in \mathcal{C}$ and from the obvious property $S_{\boldsymbol{y}}C \cap (S_{\boldsymbol{y}}C + \boldsymbol{r}) = S_{\boldsymbol{y}}\big(C \cap (C+\boldsymbol{r})\big)$. □

2.2.3 Ergodicity

The important property of random closed sets is the ergodicity, which allows one to express statistical expectations by limits of arithmetic or spatial averages. For every Borel A we define the **inner radius** of A by

$$r(A) = \sup\{r \geq 0 | B_r(\boldsymbol{x}) \subset A \text{ for some } \boldsymbol{x} \in A\}. \quad (2.3)$$

A stationary random closed set X is said to be **metrically transitive** if the condition

$$\mathbb{P}_X(F \setminus S_{\boldsymbol{y}}F \cup S_{\boldsymbol{y}}F \setminus F) = 0 \quad \text{for all} \quad \boldsymbol{y} \in \mathbb{R}^d \quad (2.4)$$

on $F \in \mathcal{B}(\mathcal{F})$ implies that $\mathbb{P}_X(F) = 0$ or $\mathbb{P}_X(F) = 1$.

Definition 2.2.4. A stationary random closed set X is

(a) **mixing**, if for all $F, G \in \mathcal{B}(\mathcal{F})$

$$\mathbb{P}_X(S_{\boldsymbol{x}}F \cap G) \to \mathbb{P}_X(F)\,\mathbb{P}_X(G) \quad \text{as} \quad \|\boldsymbol{x}\| \to \infty,$$

(b) **weakly mixing**, if for all $F, G \in \mathcal{B}(\mathcal{F})$

$$\frac{1}{\nu_d(K_n)} \int\limits_{K_n} \left| \mathbb{P}_X(S_{\boldsymbol{x}}F \cap G) - \mathbb{P}_X(F)\,\mathbb{P}_X(G) \right| \mathrm{d}\boldsymbol{x} \to 0 \quad \text{as} \quad n \to \infty,$$

(c) **ergodic**, if for all $F, G \in \mathcal{B}(\mathcal{F})$

$$\frac{1}{\nu_d(K_n)} \int\limits_{K_n} \mathbb{P}_X(S_{\boldsymbol{x}}F \cap G)\,\mathrm{d}\boldsymbol{x} \to \mathbb{P}_X(F)\,\mathbb{P}_X(G) \quad \text{as} \quad n \to \infty,$$

for any **convex averaging sequence** $\{K_n, n \in \mathbb{N}\}$, i.e. a sequence of non-empty convex compact sets, non-decreasing in the sense of inclusion, $K_n \subset K_{n+1}$ for $n \in \mathbb{N}$, such that $r(K_n) \to \infty$ as $n \to \infty$.

It is easy to see that mixing implies weak mixing and this further implies ergodicity. It can also be shown that metrical transitivity is equivalent to ergodicity, [43, Proposition 12.3.III]. All properties from previous definition can be equivalently formulated using the capacity functional, for details see [44].

To present the results for ergodic random closed sets, some notation must be introduced to describe the kind of spatial averaging that will be used. Let C_0 denote the half-open cube,

$$C_0 = \left[-\frac{1}{2}, \frac{1}{2}\right) \times \cdots \times \left[-\frac{1}{2}, \frac{1}{2}\right) \subset \mathbb{R}^d, \tag{2.5}$$

and let $\mathcal{B}_0(\mathbb{R}^d)$ denote the family of all bounded Borel sets of $\mathbb{R}^d$.

Theorem 2.2.3. *Let X be a stationary ergodic random closed set and the mapping $h : \mathcal{B}_0(\mathbb{R}^d) \to \mathbb{R}$ be a measurable, translation invariant, additive set-function defined on bounded Borel sets. Furthemore, if there is a non-negative random variable ξ of finite mean such that $|h(X \cap K)| < \xi$ almost surely for all non-empty, convex $K \subset C_0$ then*

$$\lim_{n\to\infty} \frac{h(X \cap K_n)}{\nu_d(K_n)} = \mathbb{E}\, h(X \cap C_0) \qquad \mathbb{P}\text{-almost surely}$$

for each convex averaging sequence $\{K_n, n \in \mathbb{N}\}$.

Proof. See [45] or Theorem 12.2.IV in [43]. □

The theorem states that the spatial average $h(X \cap K)/\nu_d(K)$ converges to the expectation $\mathbb{E}\, h(X \cap K)$ under conditions of stationarity and ergodicity. The most useful version is obtained when one uses the d-dimensional Lebesgue measure as a mapping h.

Corollary 2.2.2. *Let X be a stationary ergodic random closed set. Then*

$$\lim_{n\to+\infty} \frac{\nu_d(X \cap K_n)}{\nu_d(K_n)} = \mathbb{E}\, \nu_d(X \cap C_0) \qquad \mathbb{P}\text{-almost surely.}$$

2.3 Characteristics of random closed sets

2.3.1 Volume fraction

A random closed set X in $\mathbb{R}^d$ is given by its indicator function $\mathbb{1}_X(\boldsymbol{x})$. The measurability of $\mathbb{1}_X(\boldsymbol{x})$ follows from Theorem A.4.5. Therefore we can take its mean

$$m(\boldsymbol{x}) = \mathbb{E}\,\mathbb{1}_X(\boldsymbol{x}) = \mathbb{P}(\boldsymbol{x} \in X) \tag{2.6}$$

which is called the **volume fraction**. As follows from (2.2) the volume fraction of a stationary random closed set X is constant,

$$m(\boldsymbol{x}) = m(\boldsymbol{0}) = p.$$

By the Fubini-Tonelli theorem

$$p = \frac{\int_B p\,\mathrm{d}\boldsymbol{x}}{\nu_d(B)} = \frac{\int_B \mathbb{E}\,\mathbb{1}_X(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}}{\nu_d(B)} = \frac{\mathbb{E}\int_B \mathbb{1}_X(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}}{\nu_d(B)} = \frac{\mathbb{E}\,\nu_d(X\cap B)}{\nu_d(B)} \tag{2.7}$$

for every Borel set $B \subset \mathbb{R}^d$ with $0 < \nu_d(B) < +\infty$.

2.3.2 Second order characteristics

Given a random closed set X the **covariance function** cov of X is defined by

$$\mathrm{cov}(\boldsymbol{x},\boldsymbol{y}) = \mathbb{E}(\mathbb{1}_X(\boldsymbol{x}) - m(\boldsymbol{x}))(\mathbb{1}_X(\boldsymbol{y}) - m(\boldsymbol{y})) \qquad \text{for all} \boldsymbol{x},\boldsymbol{y} \in \mathbb{R}^d,$$

where $m(\boldsymbol{x})$ is the volume fraction of X. Obviously

$$\mathrm{cov}(\boldsymbol{x},\boldsymbol{y}) = \mathbb{E}\,\mathbb{1}_X(\boldsymbol{x})\mathbb{1}_X(\boldsymbol{y}) - m(\boldsymbol{x})m(\boldsymbol{y})$$

and because $\mathbb{E}\,\mathbb{1}_X(\boldsymbol{x})\mathbb{1}_X(\boldsymbol{y}) = \mathbb{P}(\boldsymbol{x} \in X, \boldsymbol{y} \in X)$ we obtain

$$\mathrm{cov}(\boldsymbol{x},\boldsymbol{y}) = \mathbb{P}(\boldsymbol{x} \in X, \boldsymbol{y} \in X) - m(\boldsymbol{x})m(\boldsymbol{y}).$$

The first term is called the **covariance** of X and is usually denoted by $C(\boldsymbol{x},\boldsymbol{y})$. Thus

$$C(\boldsymbol{x},\boldsymbol{y}) = \mathrm{cov}(\boldsymbol{x},\boldsymbol{y}) + m(\boldsymbol{x})m(\boldsymbol{y}).$$

Obviously $C(\boldsymbol{x},\boldsymbol{y}) = C(\boldsymbol{y},\boldsymbol{x})$. If X is stationary then

$$\begin{aligned} C(\boldsymbol{x},\boldsymbol{y}) &= \mathbb{P}(\boldsymbol{x} \in X, \boldsymbol{y} \in X) = \mathbb{P}(\boldsymbol{x} \in X + \boldsymbol{y}, \boldsymbol{y} \in X + \boldsymbol{y}) \\ &= \mathbb{P}(\boldsymbol{x} - \boldsymbol{y} \in X, \boldsymbol{0} \in X) = C(\boldsymbol{x} - \boldsymbol{y}, \boldsymbol{0}) \end{aligned}$$

and the same relation holds for the covariance function $\mathrm{cov}(\boldsymbol{x},\boldsymbol{y})$. Both the covariance and covariance function thus depend only on one argument

$$C(\boldsymbol{x},\boldsymbol{y}) = C(\boldsymbol{x} - \boldsymbol{y}, \boldsymbol{0}) = C(\boldsymbol{r}),$$

where $\boldsymbol{r} = \boldsymbol{x} - \boldsymbol{y}$. Moreover, $C(\boldsymbol{r}) = C(-\boldsymbol{r})$ for all $\boldsymbol{r} \in \mathbb{R}^d$. If X is isotropic, it can be easily shown, then the covariance (and covariance function) depends only on the distance $r = \|\boldsymbol{r}\|$.

For stationary RACS we have

$$C(\boldsymbol{r}) = \mathbb{P}(\boldsymbol{r} \in X, \boldsymbol{0} \in X) = \mathbb{P}\left(\boldsymbol{0} \in (X - \boldsymbol{r}), \boldsymbol{0} \in X\right) = \mathbb{P}\left(\boldsymbol{0} \in X \cap (X - \boldsymbol{r})\right).$$

The covariance at $\boldsymbol{r}$ is thus the volume fraction of $X \cap (X - \boldsymbol{r})$ that is from Corollary 2.2.1 also stationary RACS. A relation similar to (2.7) can now be easily derived:

$$\begin{aligned} C(\boldsymbol{r}) =& \frac{\int_B \mathbb{P}\left(\boldsymbol{0} \in X \cap (X - \boldsymbol{r})\right)\mathrm{d}\boldsymbol{x}}{\nu_d(B)} = \frac{\int_B \mathbb{P}\left(\boldsymbol{x} \in X \cap (X - \boldsymbol{r})\right)\mathrm{d}\boldsymbol{x}}{\nu_d(B)} \\ =& \frac{\int_B \mathbb{E}\,\mathbb{1}_{X\cap(X-\boldsymbol{r})}(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}}{\nu_d(B)} = \frac{\mathbb{E}\,\nu_d\big(X \cap (X - \boldsymbol{r}) \cap B\big)}{\nu_d(B)} \end{aligned} \tag{2.8}$$

for every Borel $B \subset \mathbb{R}^d$.

Let assume that the volume fraction is positive. Then the covariance function $\mathrm{cov}(\boldsymbol{x}, \boldsymbol{y})$ can be normalized by variances of the process at $\boldsymbol{x}$ and $\boldsymbol{y}$. The result is called the **correlation function** and is given by

$$\kappa(\boldsymbol{x}, \boldsymbol{y}) = \frac{\mathrm{cov}(\boldsymbol{x}, \boldsymbol{y})}{\sqrt{\mathrm{cov}(\boldsymbol{x}, \boldsymbol{x})}\sqrt{\mathrm{cov}(\boldsymbol{y}, \boldsymbol{y})}}.$$

Since

$$\mathrm{cov}(\boldsymbol{x}, \boldsymbol{x}) = \mathbb{E}(\mathbb{1}_X(\boldsymbol{x}) - m(\boldsymbol{x}))^2 = \mathbb{E}\,\mathbb{1}_X(\boldsymbol{x}) - m(\boldsymbol{x})^2 = m(\boldsymbol{x}) - m(\boldsymbol{x})^2,$$

the relation between the covariance and the correlation function is

$$\kappa(\boldsymbol{x}, \boldsymbol{y}) = \frac{C(\boldsymbol{x}, \boldsymbol{y}) - m(\boldsymbol{x})m(\boldsymbol{y})}{\sqrt{m(\boldsymbol{x}) - m(\boldsymbol{x})^2}\sqrt{m(\boldsymbol{y}) - m(\boldsymbol{y})^2}}.$$

Similarly to the correlation function of two random variables, $\kappa(\boldsymbol{x}, \boldsymbol{y}) \in [-1, 1]$ for all $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^d$. The covariance may be expressed using the correlation function and the volume fraction as

$$C(\boldsymbol{x}, \boldsymbol{y}) = \kappa(\boldsymbol{x}, \boldsymbol{y})\sqrt{m(\boldsymbol{x}) - m(\boldsymbol{x})^2}\sqrt{m(\boldsymbol{y}) - m(\boldsymbol{y})^2} + m(\boldsymbol{x})m(\boldsymbol{y}) \tag{2.9}$$

for all $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^d$.

Sometimes, analogously to point processes, the **pair correlation function** of X is defined to be

$$g(\boldsymbol{x}, \boldsymbol{y}) = \frac{C(\boldsymbol{x}, \boldsymbol{y})}{m(\boldsymbol{x})m(\boldsymbol{y})}$$

for all $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^d$.

In many cases the covariance has nice continuity properties. First we show the following simple implication.

Proposition 2.3.1. *If C is continuous at $\mathbf{0}$ then $C(\boldsymbol{r})$ is uniformly continuous everywhere.*

Proof. Let $\boldsymbol{r} \in \mathbb{R}^d$ be fixed. Further let $f = \mathbb{1}_X(\boldsymbol{r}) - \mathbb{1}_X(\boldsymbol{r} - \boldsymbol{y})$ and $g = \mathbb{1}_X(-\boldsymbol{y})$ for $\boldsymbol{y} \in \mathbb{R}^d$. Using the Schwartz inequality (e.g. [46, (5.36)]), $\mathbb{E}(|fg|)^2 \le \mathbb{E}\,f^2 \cdot \mathbb{E}\,g^2$, we obtain

$$\big(C(\boldsymbol{r}, -\boldsymbol{y}) - C(\boldsymbol{r} - \boldsymbol{y}, -\boldsymbol{y})\big)^2 \le \big(C(\boldsymbol{r}, \boldsymbol{r}) + C(\boldsymbol{r} - \boldsymbol{y}, \boldsymbol{r} - \boldsymbol{y}) - 2C(\boldsymbol{r}, \boldsymbol{r} - \boldsymbol{y})\big)C(-\boldsymbol{y}, -\boldsymbol{y}).$$

The stationarity yields

$$\big(C(\boldsymbol{r} + \boldsymbol{y}) - C(\boldsymbol{r})\big)^2 \le 2\big(C(\mathbf{0}) - C(\boldsymbol{y})\big)C(\mathbf{0})$$

and the statement follows by letting $\boldsymbol{y} \to \mathbf{0}$. □

Many important continuity properties holds for the so called P-continuous random closed sets, see [39] for further details.

Definition 2.3.1. A random closed set X is **P-continuous at a point $\boldsymbol{x} \in \mathbb{R}^d$** if

$$\lim_{\boldsymbol{y} \to \boldsymbol{x}} \mathbb{P}(\boldsymbol{y} \in X, \boldsymbol{x} \notin X) = \lim_{\boldsymbol{y} \to \boldsymbol{x}} \mathbb{P}(\boldsymbol{x} \in X, \boldsymbol{y} \notin X) = 0.$$

We say that X is **P-continuous** if it is P-continuous at every $\boldsymbol{x} \in \mathbb{R}^d$.

Almost all models of random closed sets are P-continuous and in particular, stationary random closed sets are P-continuous.

Lemma 2.3.1. *Every stationary random closed set X is P-continuous.*

Proof. Let $\boldsymbol{y}-\boldsymbol{x}=\boldsymbol{z}$. From stationarity it follows that

$$\mathbb{P}(\boldsymbol{x}+\boldsymbol{z}\in X,\boldsymbol{x}\notin X)=\mathbb{P}(\boldsymbol{z}\in X,\mathbf{0}\notin X)\leq\mathbb{P}(\|\boldsymbol{z}\|\,B^d\cap X\neq\emptyset,\mathbf{0}\notin X),$$
$$\mathbb{P}(\boldsymbol{x}\in X,\boldsymbol{x}+\boldsymbol{z}\notin X)=\mathbb{P}(-\boldsymbol{z}\in X,\mathbf{0}\notin X)\leq\mathbb{P}(\|\boldsymbol{z}\|\,B^d\cap X\neq\emptyset,\mathbf{0}\notin X),$$

where $\|\boldsymbol{z}\|\,B^d$ is a closed ball of radius $\|\boldsymbol{z}\|$ centred at $\mathbf{0}$. Since X is closed, we have $\{\|\boldsymbol{z}\|\,B^d\cap X\neq\emptyset,\mathbf{0}\notin X\}\downarrow\emptyset$ as $\|\boldsymbol{z}\|\downarrow 0$. The assertion follows from the continuity of the probability measure (Theorem A.1.1 (c)). □

Proposition 2.3.2. *The covariance of a P-continuous RACS X is continuous in each argument, uniformly with respect to the other argument, and is also continuous in $\mathbb{R}^d\times\mathbb{R}^d$. Moreover, if X is stationary, then it is uniformly continuous.*

Proof. We follow the idea of the proof given in [47] for stationary RACS. For any two events A and B one can easily show the inequality

$$|\mathbb{P}(A)-\mathbb{P}(B)|\leq\max\{\mathbb{P}(A\setminus B),\mathbb{P}(B\setminus A)\}.$$

Let now $A=\{\boldsymbol{x}\in X\}\cap\{\boldsymbol{y}\in X\}$ and $B=\{\boldsymbol{x}\in X\}\cap\{\boldsymbol{y}+\boldsymbol{z}\in X\}$. In that case

$$A\setminus B=\{\boldsymbol{x}\in X,\boldsymbol{y}\in X,\boldsymbol{y}+\boldsymbol{z}\notin X\}\subset\{\boldsymbol{y}\in X,\boldsymbol{y}+\boldsymbol{z}\notin X\},$$
$$B\setminus A=\{\boldsymbol{x}\in X,\boldsymbol{y}+\boldsymbol{z}\in X,\boldsymbol{y}\notin X\}\subset\{\boldsymbol{y}+\boldsymbol{z}\in X,\boldsymbol{y}\notin X\}.$$

Hence, for the covariance we obtain

$$\begin{aligned}|C(\boldsymbol{x},\boldsymbol{y})-C(\boldsymbol{x},\boldsymbol{y}+\boldsymbol{z})|&=|\mathbb{P}(A)-\mathbb{P}(B)|\\&\leq\max\{\mathbb{P}(\boldsymbol{y}\in X,\boldsymbol{y}+\boldsymbol{z}\notin X),\mathbb{P}(\boldsymbol{y}+\boldsymbol{z}\in X,\boldsymbol{y}\notin X)\}.\end{aligned}$$

From the P-continuity of X follows that $|C(\boldsymbol{x},\boldsymbol{y})-C(\boldsymbol{x},\boldsymbol{y}+\boldsymbol{z})|\to 0$ as $\|\boldsymbol{z}\|\to 0$ independently of $\boldsymbol{x}$. The continuity in the first argument of $C(\boldsymbol{x},\boldsymbol{y})$ can be proved in the same way and is again uniform in the second argument. The continuity of $C(\boldsymbol{x},\boldsymbol{y})$ in $\mathbb{R}^d\times\mathbb{R}^d$ follows from

$$|C(\boldsymbol{x},\boldsymbol{y})-C(\boldsymbol{x}+\boldsymbol{u},\boldsymbol{y}+\boldsymbol{v})|\leq|C(\boldsymbol{x},\boldsymbol{y})-C(\boldsymbol{x}+\boldsymbol{u},\boldsymbol{y})|+|C(\boldsymbol{x}+\boldsymbol{u},\boldsymbol{y})-C(\boldsymbol{x}+\boldsymbol{u},\boldsymbol{y}+\boldsymbol{v})|\,.$$

In the case of a stationary RACS X we get that $C(\boldsymbol{r})=C(\boldsymbol{r},\mathbf{0})$ is continuous in $\boldsymbol{r}$ and in particular at $\boldsymbol{r}=\mathbf{0}$. The uniform continuity then follows from Proposition 2.3.1. □

The previous assertion does not hold generally. If we take a constant RACS X in $\mathbb{R}$ given as the closed interval $X=[0,1]$, it is obvious that the covariance $C(0,x)$ has a discontinuity at point $x=1$.

Proposition 2.3.3. *The volume fraction $m(\boldsymbol{x})$ of a P-continuous RACS X is continuous.*

Proof. For any two events A and B it holds

$$\mathbb{P}(B)=\mathbb{P}(A)+\mathbb{P}(A^c\cap B)-\mathbb{P}(A\cap B^c).$$

Let now tek $A=\{\boldsymbol{x}\in X\}$ and $B=\{\boldsymbol{y}\in X\}$. It follows that

$$\mathbb{P}(\boldsymbol{y}\in X)=\mathbb{P}(\boldsymbol{x}\in X)+\mathbb{P}(\boldsymbol{y}\in X,\boldsymbol{x}\notin X)-\mathbb{P}(\boldsymbol{x}\in X,\boldsymbol{y}\notin X).$$

Letting $\boldsymbol{y}\to\boldsymbol{x}$ on both sides and using the definition of P-continuity completes the proof. □

The trivial consequence of last two propositions is the continuity of the covariance function and also the correlation function, if exists.

Corollary 2.3.1. *Let X be a P-continuous RACS. Then the covariance function* cov *of X is continuous. If moreover $m(\boldsymbol{x})>0$ for all $\boldsymbol{x}\in\mathbb{R}^d$ then the correlation function κ is continuous.*

The last important property of the covariance function is its positive semi-definiteness.

Definition 2.3.2. We say that a real valued function f defined on $\mathbb{R}^d$ is **positive semi-definite** if

$$\sum_i \sum_j \lambda_i \overline{\lambda_j} f(\boldsymbol{x}_i - \boldsymbol{x}_j) \geq 0$$

for all $n \in \mathbb{N}$, $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n \in \mathbb{R}^d$ and $\lambda_1, \ldots, \lambda_n \in \mathbb{C}$.

It is a simple consequence of the definition that $f(\boldsymbol{0}) \geq 0$, $f(\boldsymbol{x}) = f(-\boldsymbol{x})$ and $|f(\boldsymbol{x})| \leq f(\boldsymbol{0})$ for all $\boldsymbol{x} \in \mathbb{R}^d$. It is also clear that convex combinations and scalar multiplications of positive semi-definite functions produce again positive semi-definite functions.

The positive semi-definiteness is usually formulated for stationary random closed sets where the covariance function depends only on the difference $\boldsymbol{x} - \boldsymbol{y}$ (see e.g. [48, 49]). In the following we prove the more general version of this property that holds also for non-stationary random closed sets.

Theorem 2.3.1. *Let X be a random closed set with correlation function $\kappa(\boldsymbol{x}, \boldsymbol{y})$. Then for every $n \in \mathbb{N}$, $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n \in \mathbb{R}^d$ and $\lambda_1, \ldots, \lambda_n \in \mathbb{C}$*

$$\sum_i \sum_j \lambda_i \overline{\lambda_j} \kappa(\boldsymbol{x}_i, \boldsymbol{x}_j) \geq 0.$$

Proof. It is easy to see the following chain of equalities.

$$\begin{aligned}
\sum_{i=1}^n \sum_{j=1}^n \lambda_i \overline{\lambda_j} \kappa(\boldsymbol{x}_i, \boldsymbol{x}_j) &= \sum_{i=1}^n \sum_{j=1}^n \lambda_i \overline{\lambda_j} \frac{\mathbb{E}\, \mathbb{1}_X(\boldsymbol{x}_i)\mathbb{1}_X(\boldsymbol{x}_j) - m(\boldsymbol{x}_i)m(\boldsymbol{x}_j)}{\sqrt{m(\boldsymbol{x}_i) - m(\boldsymbol{x}_i)^2}\sqrt{m(\boldsymbol{x}_j) - m(\boldsymbol{x}_j)^2}} \\
&= \sum_{i=1}^n \sum_{j=1}^n \lambda_i \overline{\lambda_j} \frac{\mathbb{E}\left(\mathbb{1}_X(\boldsymbol{x}_i)\mathbb{1}_X(\boldsymbol{x}_j) - \mathbb{1}_X(\boldsymbol{x}_i)m(\boldsymbol{x}_j) - \mathbb{1}_X(\boldsymbol{x}_j)m(\boldsymbol{x}_i) + m(\boldsymbol{x}_i)m(\boldsymbol{x}_j)\right)}{\sqrt{m(\boldsymbol{x}_i) - m(\boldsymbol{x}_i)^2}\sqrt{m(\boldsymbol{x}_j) - m(\boldsymbol{x}_j)^2}} \\
&= \mathbb{E} \sum_{i=1}^n \sum_{j=1}^n \lambda_i \overline{\lambda_j} \frac{\mathbb{1}_X(\boldsymbol{x}_i)\mathbb{1}_X(\boldsymbol{x}_j) - \mathbb{1}_X(\boldsymbol{x}_i)m(\boldsymbol{x}_j) - \mathbb{1}_X(\boldsymbol{x}_j)m(\boldsymbol{x}_i) + m(\boldsymbol{x}_i)m(\boldsymbol{x}_j)}{\sqrt{m(\boldsymbol{x}_i) - m(\boldsymbol{x}_i)^2}\sqrt{m(\boldsymbol{x}_j) - m(\boldsymbol{x}_j)^2}} \\
&= \mathbb{E} \left| \sum_{i=1}^n \lambda_i \frac{\mathbb{1}_X(\boldsymbol{x}_i) - m(\boldsymbol{x}_i)}{\sqrt{m(\boldsymbol{x}_i) - m(\boldsymbol{x}_i)^2}} \right|^2 \geq 0.
\end{aligned}$$

□

Analogous argumentation holds for the covariance function $\mathrm{cov}(\boldsymbol{x}, \boldsymbol{y})$ and the covariance $C(\boldsymbol{x}, \boldsymbol{y})$.

Corollary 2.3.2. *If the correlation function $\kappa(\boldsymbol{x}, \boldsymbol{y})$ (or the covariance function, or the covariance) of a random closed set X depends only on $\boldsymbol{x} - \boldsymbol{y}$, then it is positive semi-definite, i.e. for all $n \in \mathbb{N}$, $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n \in \mathbb{R}^d$ and $\lambda_1, \ldots, \lambda_n \in \mathbb{C}$*

$$\sum_i \sum_j \lambda_i \overline{\lambda_j} \kappa(\boldsymbol{x}_i - \boldsymbol{x}_j) \geq 0.$$

2.3.3 The Bartlett spectrum

Since the covariance function of a stationary random closed set X is positive semi-definite then by Bochner's theorem A.2.2, there exists a finite Borel measure Γ on $\mathbb{R}^d$ such that

$$\mathrm{cov}(\boldsymbol{x}) = (2\pi)^{-d/2} \int_{\mathbb{R}^d} e^{i\boldsymbol{x}\cdot\boldsymbol{\xi}}\, \Gamma(\mathrm{d}\boldsymbol{\xi}).$$

The measure Γ is called the **Bartlett spectrum** of X.

Recall that the Fourier transform $\mathscr{F}h$ and the inverse Fourier transform $\mathscr{F}^{-1}h$ of a function $h \in L^1(\mathbb{R}^d, \nu_d)$ are given by

$$(\mathscr{F}h)(\boldsymbol{\xi}) = (2\pi)^{-d/2} \int_{\mathbb{R}^d} e^{-i\boldsymbol{\xi}\cdot\boldsymbol{x}} h(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} \quad \text{and} \quad (\mathscr{F}^{-1}h)(\boldsymbol{x}) = (2\pi)^{-d/2} \int_{\mathbb{R}^d} e^{i\boldsymbol{\xi}\cdot\boldsymbol{x}} h(\boldsymbol{\xi})\,\mathrm{d}\boldsymbol{\xi},$$

respectively. If the covariance function cov of X is absolutely integrable, i.e. $\mathrm{cov} \in L^1(\mathbb{R}^d, \nu_d)$, then from the inversion formula A.2.1 follows that the Bartlett spectrum of X is absolutely continuous with respect to ν_d, and a density f_Γ is given by

$$f_\Gamma(\boldsymbol{\xi}) = (2\pi)^{-d/2} \int_{\mathbb{R}^d} e^{-i\boldsymbol{\xi}\cdot\boldsymbol{x}} \mathrm{cov}(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}.$$

More details about the Bartlett spectrum of random closed sets can be found in [48, 50]. The extension to random measures is presented in Subsection 2.5.5.

2.4 Random fields and Excursion sets

In the following we introduce the important class of models of random closed sets. We start with a brief review of the theory of random fields. For a more detailed treatment we refer the reader to [51].

Definition 2.4.1. A **random field** Z in $\mathbb{R}^d$ is a collection of real random variables $\{Z(\boldsymbol{x}) | \boldsymbol{x} \in \mathbb{R}^d\}$ defined on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$.

The distribution of random field is usually described by its finite-dimensional distributions defined for every $k \in \mathbb{N}$ and every $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k \in \mathbb{R}^d$ by $\mu_{\boldsymbol{x}_1,\ldots,\boldsymbol{x}_k}(B) = \mathbb{P}\left((Z(\boldsymbol{x}_1), \ldots, Z(\boldsymbol{x}_k)) \in B\right)$ for all Borel $B \in \mathbb{R}^k$. The Kolmogorov's existence theorem (see [46, Theorem 36.2]) implies that for a family of probability measures $\{\mu_{\boldsymbol{x}_1,\ldots,\boldsymbol{x}_k} | k \in \mathbb{N},\ \boldsymbol{x}_1, \ldots, \boldsymbol{x}_k \in \mathbb{R}^d\}$ there exists a random field with finite-dimensional distributions given by those measures if and only if they satisfy the additional properties of symmetry,

$$\mu_{\boldsymbol{x}_1,\ldots,\boldsymbol{x}_k}(A_1 \times \ldots \times A_k) = \mu_{\boldsymbol{x}_{\pi(1)},\ldots,\boldsymbol{x}_{\pi(k)}}(A_{\pi(1)} \times \ldots \times A_{\pi(k)})$$

for every $k \in \mathbb{N}$, $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k \in \mathbb{R}$, Borel sets $A_1, \ldots, A_k \subset \mathbb{R}$, and every permutation π, and consistency,

$$\mu_{\boldsymbol{x}_1,\ldots,\boldsymbol{x}_k}(A_1 \times \ldots \times A_k) = \mu_{\boldsymbol{x}_1,\ldots,\boldsymbol{x}_{k+1}}(A_1 \times \ldots \times A_k \times \mathbb{R})$$

for every $k \in \mathbb{N}$, $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k \in \mathbb{R}$, and all Borel sets $A_1, \ldots, A_k \subset \mathbb{R}$.

Definition 2.4.2. A **Gaussian random field** Z in $\mathbb{R}^d$ is a random field with finite-dimensional distributions that are all multivariate normal.

This means that the finite-dimensional distributions are determined by the probability density functions

$$f_{\boldsymbol{x}_1,\ldots,\boldsymbol{x}_k}(z_1, \ldots, z_k) = \frac{1}{(2\pi)^{k/2}\sqrt{\det \boldsymbol{V}}} e^{-\frac{1}{2}(\boldsymbol{z}-\boldsymbol{\mu})\boldsymbol{V}^{-1}(\boldsymbol{z}-\boldsymbol{\mu})^T},$$

where $\boldsymbol{z} = (z_1, \ldots, z_k) \in \mathbb{R}^k$, $\boldsymbol{\mu} = \left(\mathbb{E}\, Z(\boldsymbol{x}_1), \ldots, \mathbb{E}\, Z(\boldsymbol{x}_1)\right)$ and V is a positive-semi definite matrix with elements $(\boldsymbol{V})_{ij} = \mathrm{cov}\left(Z(\boldsymbol{x}_i), Z(\boldsymbol{x}_j)\right)$ for all $i, j = 1, \ldots, k$.

For a random field Z we define the **mean function** by $\mu_Z(\boldsymbol{x}) = \mathbb{E}\, Z(\boldsymbol{x})$, $\boldsymbol{x} \in \mathbb{R}^d$ and the **covariance function** by $\mathrm{cov}_Z(\boldsymbol{x}, \boldsymbol{y}) = \mathrm{cov}\left(Z(\boldsymbol{x}), Z(\boldsymbol{y})\right)$, $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^d$. Clearly $\mathrm{cov}_Z(\boldsymbol{x}, \boldsymbol{y}) =$

$\mathrm{cov}_Z(\boldsymbol{y}, \boldsymbol{x})$. Moreover, it can be easily shown that cov_Z satisfies the condition similar to positive-semi definiteness:

$$\sum_i \sum_j \lambda_i \overline{\lambda_j}\, \mathrm{cov}_Z(\boldsymbol{x}_i, \boldsymbol{x}_j) \geq 0 \tag{2.10}$$

for every $n \in \mathbb{N}$, $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n \in \mathbb{R}^d$ and $\lambda_1, \ldots, \lambda_n \in \mathbb{C}$. A Gaussian random field Z is therefore completely determined by its mean function and covariance function. Moreover, for any function $\mu : \mathbb{R}^d \to \mathbb{R}$ and $\mathrm{cov} : \mathbb{R}^d \times \mathbb{R}^d$ satisfying $\mathrm{cov}(\boldsymbol{x}, \boldsymbol{y}) = \mathrm{cov}(\boldsymbol{y}, \boldsymbol{x})$ and (2.10) it is possible to construct a (Gaussian) random field having μ and cov as its mean and covariance functions, respectively.

Let G be a topological group acting measurably on $\mathbb{R}^d$. Then we define the action of G on a random field Z by

$$S_g Z(\boldsymbol{x}) = Z(g^{-1}\boldsymbol{x}) \quad \text{for all} \quad \boldsymbol{x} \in \mathbb{R}^d, g \in G.$$

For the finite-dimensional distributions $\mu_{S_g Z;\boldsymbol{x}_1,\ldots,\boldsymbol{x}_k}$ of $S_g Z$ it follows

$$\mu_{S_g Z;\boldsymbol{x}_1,\ldots,\boldsymbol{x}_k} = \mu_{Z;g^{-1}\boldsymbol{x}_1,\ldots,g^{-1}\boldsymbol{x}_k},$$

where $\mu_{Z;\boldsymbol{x}_1,\ldots,\boldsymbol{x}_k}$ of $S_g Z$ are finite dimensional distributions of Z. For the group of translations of $\mathbb{R}^d$ we use the notation $S_{\boldsymbol{y}} Z = Z + \boldsymbol{y}$ for all $\boldsymbol{y} \in \mathbb{R}^d$.

Definition 2.4.3. A random field Z in $\mathbb{R}^d$ is called **stationary** if $Z \overset{\mathcal{D}}{\sim} Z + \boldsymbol{x}$ for all $\boldsymbol{x} \in \mathbb{R}^d$. It is called **isotropic** if $Z \overset{\mathcal{D}}{\sim} \theta Z$ for all rotations $\theta \in SO_d$.

For a stationary random field Z the mean function is constant, $\mu(\boldsymbol{x}) = \mu$ for all $\boldsymbol{x} \in \mathbb{R}^d$, and the covariance function $\mathrm{cov}_Z(\boldsymbol{x}, \boldsymbol{y})$ depends only on $\boldsymbol{x} - \boldsymbol{y}$, $\mathrm{cov}_Z(\boldsymbol{x}, \boldsymbol{y}) = \mathrm{cov}_Z(\boldsymbol{x} - \boldsymbol{y})$ for all $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^d$. The property (2.10) now corresponds to the standard positive semi-definiteness from Definition 2.3.2. It easily follows that a Gaussian random field Z is stationary if and only if it has constant mean function and its covariance function $\mathrm{cov}_Z(\boldsymbol{x}, \boldsymbol{y})$ depends only on $\boldsymbol{x} - \boldsymbol{y}$. Similarly, for a stationary isotropic random field Z the covariance function $\mathrm{cov}_Z(\boldsymbol{x}, \boldsymbol{y})$ depends only on $\|\boldsymbol{x} - \boldsymbol{y}\|$, $\mathrm{cov}_Z(\boldsymbol{x}, \boldsymbol{y}) = \mathrm{cov}_Z(\|\boldsymbol{x} - \boldsymbol{y}\|)$ for all $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^d$. A stationary Gaussian random field Z is isotropic if and only if its covariance function $\mathrm{cov}_Z(\boldsymbol{x}, \boldsymbol{y})$ depends only on $\|\boldsymbol{x} - \boldsymbol{y}\|$.

Now we introduce the connection to random closed set. For a random field Z, $u \in \mathbb{R}^d$ and Borel B let us define

$$A_u(Z, B)(\omega) = \{\boldsymbol{x} \in B | Z(\boldsymbol{x}, \omega) \geq u\} \quad \text{for all} \quad \omega \in \Omega. \tag{2.11}$$

Our aim is to take $A_u(Z, \mathbb{R}^d)$ as a random closed set. This is, however, a problem in general since $A_u(Z, \mathbb{R}^d)$ may not be closed and measurable in sense of Definition 2.2.1. For measurability the concept of separability introduced by Doob in [52] have to be used.

Definition 2.4.4. A real random field Z in $\mathbb{R}^d$ on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$ is **separable**, if there exists a countable set $D \subset \mathbb{R}^d$ dense in $\mathbb{R}^d$ such that for any closed interval (finite or infinite) $I \subset \mathbb{R}$ and any open set $G \subset \mathbb{R}^d$,

$$\{\omega | Z(\boldsymbol{x}, \omega) \in I, \boldsymbol{x} \in G\} = \{\omega | Z(\boldsymbol{x}, \omega) \in I, \boldsymbol{x} \in G \cap D\}$$

holds. The set D is called a **separating set** for Z.

For a separable random field Z with separating set D the following two relations hold, [53, Lemma 2.2]:

$$\inf_{\boldsymbol{y} \in G \cap D} Z(\boldsymbol{y}) = \inf_{\boldsymbol{x} \in G} Z(\boldsymbol{x}) \quad \text{and} \quad \sup_{\boldsymbol{y} \in G \cap D} Z(\boldsymbol{y}) = \sup_{\boldsymbol{x} \in G} Z(\boldsymbol{x}),$$

where $G \subset \mathbb{R}^d$ is an arbitrary open set. Furthermore, it can be shown (see [53, Theorem 2.6]) that to every random field there is an equivalent separable random field. Thus in the following we always assume that random fields are separable.

Let us now deal with the continuity. Clearly, for a random field with almost surely continuous sample functions and a closed set F is the set $A_u(Z,F)$ almost surely closed. Therefore random fields with almost surely continuous sample functions are of primary importance. The following assertion gives a sufficient condition for a Gaussian random field to have this property.

Theorem 2.4.1. *Let Z be a zero-mean Gaussian random field with a continuous covariance function and such that for some $0 < K < \infty$, some $\varepsilon > 0$, and every compact $C \subset \mathbb{R}^d$,*

$$\mathbb{E}(Z(\boldsymbol{x}) - Z(\boldsymbol{y}))^2 \leq \frac{K}{|\log \|\boldsymbol{x}-\boldsymbol{y}\||^{1+\varepsilon}} \quad \text{for all} \quad \boldsymbol{x},\boldsymbol{y} \in C.$$

Then Z has almost surely continuous sample functions.

Proof. Follows from Theorem 3.4.1 in [51]. □

For a stationary random field the expectation on the left side of the condition can be written as $2\,\mathrm{cov}_Z(\mathbf{0}) - 2\,\mathrm{cov}_Z(\boldsymbol{x}-\boldsymbol{y})$.

Corollary 2.4.1. *Let Z be a stationary zero-mean Gaussian random field with a continuous covariance function and such that for some $0 < K < \infty$ and some $\varepsilon > 0$,*

$$\mathrm{cov}_Z(\mathbf{0}) - \mathrm{cov}_Z(\boldsymbol{r}) \leq \frac{K}{|\log \|\boldsymbol{r}\||^{1+\varepsilon}}$$

for all $\|\boldsymbol{r}\| < 1$. Then Z has almost surely continuous sample functions.

A similar condition can be formulated for non-Gaussian random fields as was done in [54]. Now we may proof the main theorem of this part.

Theorem 2.4.2. *Let $A_u(Z,\mathbb{R}^d)$ be defined by (2.11) for a random closed set Z with almost surely continuous sample functions. Then $X_u(Z)$ given by $A_u(Z,\mathbb{R}^d)$ whenever $A_u(Z,\mathbb{R}^d)$ is a closed set and by $\emptyset$ otherwise is a random closed set. If Z is stationary (isotropic) then $X_u(Z)$ is stationary (isotropic).*

Proof. Let $u \in \mathbb{R}$ be fixed. By construction, $X_u(Z)$ is always closed. Now we show the measurability. Without loss of generality we may assume that the set $A = \{\omega | X_u(Z)(\omega) \neq A_u(Z,\mathbb{R}^d)(\omega)\} \subset \{\omega | X_u(Z)(\omega) = \emptyset\}$ is measurable and $\mathbb{P}(A) = 0$. Let $C \in \mathcal{C}'$. Then $\{X_u(Z) \in \mathcal{F}^C\} = A \cup \{\omega | Z(\boldsymbol{x},\omega) < u, \boldsymbol{x} \in C\}$. It is enough to prove the measurability of $\{Z(\boldsymbol{x},\omega) < u, \boldsymbol{x} \in C\} \cap A^c$. Since C is compact and Z is continuous on A^c this is equivalent to $\{\sup_{\boldsymbol{x}\in C} Z(\boldsymbol{x},\omega) < u\} \cap A^c$. Let G_n be a decreasing sequence of open sets such that $C = \bigcap_{n=1}^{\infty} G_n$ given e.g. by $G_n = C + 1/nU^d$. From the separability of Z follows that for each n and for all $\omega \in B$, $\sup_{\boldsymbol{x}\in G_n} Z(\boldsymbol{x},\omega) = \sup_{\boldsymbol{x}\in G_n \cap D} Z(\boldsymbol{x},\omega)$, where D is a separating set for Z. Since D is countable it follows that $\sup_{\boldsymbol{x}\in G_n} Z(\boldsymbol{x},\omega)$ as a mapping from Ω to $\mathbb{R}$ is measurable. Hence the set $\{\omega | \sup_{\boldsymbol{x}\in G_n} Z(\boldsymbol{x},\omega) < u\}$ is measurable for each n. Since $\{\omega | \sup_{\boldsymbol{x}\in C} Z(\boldsymbol{x},\omega) < u\} = \bigcup_{n=1}^{\infty} \{\omega | \sup_{\boldsymbol{x}\in G_n} Z(\boldsymbol{x},\omega) < u\}$, it is also measurable and thus also the set $\{Z(\boldsymbol{x},\omega) < u, \boldsymbol{x} \in C\} \cap A^c$ is measurable. Taking all together, we have shown that $\{X_u(Z) \in \mathcal{F}^C\}$ is measurable. The measurability of $X_u(Z)$ now follows from Lemma A.4.3. The stationarity and isotropy follow from obvious relations $X_u(Z) + \boldsymbol{x} = X_u(Z + \boldsymbol{x})$, $\boldsymbol{x} \in \mathbb{R}^d$ and $\theta X_u(Z) = X_u(\theta Z)$, $\theta \in SO_d$, respectively. □

Definition 2.4.5. Let Z be a random closed set with almost surely continuous sample functions. The random closed set $X_u(Z)$ is called the **u-level excursion set** of a random field Z.

Finally we present some basic characteristics of u-level excursion set $X_u(Z)$ of a Gaussian random field determined by a mean function μ_Z and a covariance function $\mathrm{cov}_Z(\boldsymbol{x},\boldsymbol{y})$. For the volume fraction $m(\boldsymbol{x})$ of $X_u(Z)$ we obtain

$$\begin{aligned} m(\boldsymbol{x}) = \mathbb{P}(Z(\boldsymbol{x}) \geq u) =& \mathbb{P}\left(\frac{Z(\boldsymbol{x}) - \mu_Z(\boldsymbol{x})}{\sqrt{\mathrm{cov}_Z(\boldsymbol{x},\boldsymbol{x})}} \geq \frac{u - \mu_Z(\boldsymbol{x})}{\sqrt{\mathrm{cov}_Z(\boldsymbol{x},\boldsymbol{x})}}\right) \\ =& 1 - \Phi\left(\frac{u - \mu_Z(\boldsymbol{x})}{\sqrt{\mathrm{cov}_Z(\boldsymbol{x},\boldsymbol{x})}}\right), \end{aligned} \tag{2.12}$$

where Φ is a distribution function of a standard normal random variable. In order to obtain the covariance function we first show an important relation that holds for a bivariate normal distribution. If

$$\phi(x,y;\rho) = \frac{1}{2\pi\sqrt{1-\rho^2}} e^{-\frac{x^2+y^2-2\rho xy}{2(1-\rho^2)}}$$

is the joint probability density function of two standard normal variables with mutual covariance ρ, then it can be easily checked that

$$\frac{\partial\phi}{\partial\rho} = \frac{\partial^2\phi}{\partial x \partial y}.$$

Similar relation holds for a general multivariate normal distribution, see [55, p. 26]. Integrating the relation with respect to x, y and ρ one obtains

$$\int_a^\infty \int_b^\infty \phi(x,y;\rho)\,\mathrm{d}x\mathrm{d}y = (1-\Phi(a))(1-\Phi(b)) + \int_0^\rho \phi(a,b;z)\,\mathrm{d}z.$$

The covariance of $X_u(Z)$ can be written as

$$\begin{aligned} C(\boldsymbol{x},\boldsymbol{y}) &= \mathbb{P}(Z(\boldsymbol{x}) \geq u, Z(\boldsymbol{y}) \geq u) \\ &= \frac{1}{2\pi\sqrt{\det \boldsymbol{V}}} \int_u^\infty \int_u^\infty e^{-\frac{1}{2}(s-\mu_Z(\boldsymbol{x}),t-\mu_Z(\boldsymbol{y}))\boldsymbol{V}^{-1}(s-\mu_Z(\boldsymbol{x}),t-\mu_Z(\boldsymbol{y}))^T}\,\mathrm{d}s\mathrm{d}t, \end{aligned}$$

where

$$\boldsymbol{V} = \begin{pmatrix} \mathrm{cov}_Z(\boldsymbol{x},\boldsymbol{x}) & \mathrm{cov}_Z(\boldsymbol{x},\boldsymbol{y}) \\ \mathrm{cov}_Z(\boldsymbol{x},\boldsymbol{y}) & \mathrm{cov}_Z(\boldsymbol{y},\boldsymbol{y}) \end{pmatrix}.$$

The inverse of V is

$$\boldsymbol{V}^{-1} = \frac{1}{\det \boldsymbol{V}} \begin{pmatrix} \mathrm{cov}_Z(\boldsymbol{y},\boldsymbol{y}) & -\mathrm{cov}_Z(\boldsymbol{x},\boldsymbol{y}) \\ -\mathrm{cov}_Z(\boldsymbol{x},\boldsymbol{y}) & \mathrm{cov}_Z(\boldsymbol{x},\boldsymbol{x}) \end{pmatrix}$$

and $\det \boldsymbol{V} = \mathrm{cov}_Z(\boldsymbol{x},\boldsymbol{x})\,\mathrm{cov}_Z(\boldsymbol{y},\boldsymbol{y}) - \mathrm{cov}_Z^2(\boldsymbol{x},\boldsymbol{y}) = \mathrm{cov}_Z(\boldsymbol{x},\boldsymbol{x})\,\mathrm{cov}_Z(\boldsymbol{y},\boldsymbol{y})\big(1-\kappa_Z^2(\boldsymbol{x},\boldsymbol{y})\big)$, where

$$\kappa_Z(\boldsymbol{x},\boldsymbol{y}) = \frac{\mathrm{cov}_Z(\boldsymbol{x},\boldsymbol{y})}{\sqrt{\mathrm{cov}_Z(\boldsymbol{x},\boldsymbol{x})}\sqrt{\mathrm{cov}_Z(\boldsymbol{y},\boldsymbol{y})}}$$

is the **correlation function** of Z. Taking all together and after the substitution $x = (s -$

$\mu_Z(\boldsymbol{x}))/\sqrt{\mathrm{cov}_Z(\boldsymbol{x},\boldsymbol{x})}$, $y = (s-\mu_Z(\boldsymbol{y}))/\sqrt{\mathrm{cov}_Z(\boldsymbol{y},\boldsymbol{y})}$ we obtain

$$\begin{aligned} C(\boldsymbol{x},\boldsymbol{y}) &= \frac{1}{2\pi\sqrt{(1-\kappa_Z^2(\boldsymbol{x},\boldsymbol{y}))}} \int\limits_{\frac{u-\mu_Z(\boldsymbol{x})}{\sqrt{\mathrm{cov}_Z(\boldsymbol{x},\boldsymbol{x})}}}^{\infty} \int\limits_{\frac{u-\mu_Z(\boldsymbol{y})}{\sqrt{\mathrm{cov}_Z(\boldsymbol{y},\boldsymbol{y})}}}^{\infty} e^{-\frac{x^2+y^2-2xy\kappa_Z(\boldsymbol{x},\boldsymbol{y})}{2(1-\kappa_Z^2(\boldsymbol{x},\boldsymbol{y}))}} \,\mathrm{d}x\mathrm{d}y \\ &= \int\limits_{\tilde{x}}^{\infty}\int\limits_{\tilde{y}}^{\infty} \phi(x,y;\kappa_Z(\boldsymbol{x},\boldsymbol{y}))\,\mathrm{d}x\mathrm{d}y = (1-\Phi(\tilde{x}))(1-\Phi(\tilde{y})) + \int\limits_0^{\kappa_Z(\boldsymbol{x},\boldsymbol{y})} \phi(\tilde{x},\tilde{y};z)\,\mathrm{d}z \\ &= m(\boldsymbol{x})m(\boldsymbol{y}) + \int\limits_0^{\kappa_Z(\boldsymbol{x},\boldsymbol{y})} \phi(\tilde{x},\tilde{y};z)\,\mathrm{d}z, \end{aligned}$$

where $\tilde{x} = \frac{u-\mu_Z(\boldsymbol{x})}{\sqrt{\mathrm{cov}_Z(\boldsymbol{x},\boldsymbol{x})}}$ and $\tilde{y} = \frac{u-\mu_Z(\boldsymbol{y})}{\sqrt{\mathrm{cov}_Z(\boldsymbol{y},\boldsymbol{y})}}$. The covariance function is thus

$$\mathrm{cov}(\boldsymbol{x},\boldsymbol{y}) = \int\limits_0^{\kappa_Z(\boldsymbol{x},\boldsymbol{y})} \phi(\tilde{x},\tilde{y};z)\,\mathrm{d}z. \tag{2.13}$$

If Z is a stationary Gaussian random field determined by $\mu_Z = 0$ and $\mathrm{cov}_Z(\boldsymbol{r})$ with $\mathrm{cov}_Z(\boldsymbol{0}) = 1$ we particularly obtain

$$p = 1-\Phi(u) \tag{2.14}$$

and

$$\mathrm{cov}(\boldsymbol{r}) = \frac{1}{2\pi}\int\limits_0^{\mathrm{cov}_Z(\boldsymbol{r})} \frac{e^{-\frac{u^2}{1+z}}}{\sqrt{1-z^2}}\,\mathrm{d}z.$$

If moreover $u = 0$, then

$$\mathrm{cov}(\boldsymbol{r}) = \frac{1}{2\pi}\arcsin\big(\mathrm{cov}_Z(\boldsymbol{r})\big). \tag{2.15}$$

Thus for the correlation function κ of $X_0(Z)$ holds

$$\kappa(\boldsymbol{r}) = \frac{2}{\pi}\arcsin\big(\mathrm{cov}_Z(\boldsymbol{r})\big). \tag{2.16}$$

Finally, it should be noted that those formulas are already known and in the stationary case also usually mentioned (see e.g. [41]). However, hard to find their precise derivation, which is the reason why it was shown here.

2.5 Random measures and Point processes

We introduce random measures and point processes in the similar fashion as was done in [40] and in [43]. Let E be locally compact, Hausdorff topological space with a countable base and let $\mathcal{B}(E)$ be its Borel σ-algebra. By M we denote the set of all locally finite Borel measures on E. Let $\mathcal{M}$ be the smallest σ-algebra for which all mappings $\varphi \mapsto \varphi(A), A \in \mathcal{B}(E)$ are measurable.

A **counting measure** on E is a measure $\eta \in \mathsf{M}$ with $\eta(A) \in \mathbb{N}_0 \cup \{\infty\}$ for all $A \in \mathcal{B}(E)$. Let N be the set of all counting measures on E and let $\mathcal{N}$ be the trace σ-algebra of $\mathcal{M}$ on N.

Lemma 2.5.1. *The set* N *is a measurable subset of* M*, i.e.* $\mathsf{N} \in \mathcal{M}$.

Proof. Lemma 3.1.2 in [40]. □

The last important lemma before the definition of a random measure and a point process shows that every counting measure can be measurably represented as a sum of Dirac measures. The Dirac measure δ_x at $x \in E$ is for every $A \in \mathcal{B}(E)$ defined by

$$\delta_x(A) = \begin{cases} 1, & \text{if } x \in A, \\ 0, & \text{if } x \notin A. \end{cases}$$

Lemma 2.5.2. *There exist measurable mappings $\zeta_i : \mathsf{N} \to E$ such that*

$$\eta = \sum_{i=1}^{\eta(E)} \delta_{\zeta_i(\eta)}$$

for $\eta \in \mathsf{N}$.

Proof. Lemma 3.1.3 in [40]. □

The counting measure η is **simple** if $\eta(\{x\}) \leq 1$ for all $x \in E$. Let N_s be the set of all simple counting measures on E and let $\mathcal{N}_s$ be the trace σ-algebra of $\mathcal{N}$ on N_s. From the previous lemma follows that for $i, j \in \mathbb{N}$, $\zeta_i \cap \zeta_j$ is measurable, and since E is a Hausdorff topological space, that the set $\{\eta \in \mathsf{N} | \zeta_i(\eta) \neq \zeta_j(\eta)\}$ is measurable in N. Now, because η is simple if and only if $\zeta_i(\eta) \neq \zeta_j(\eta)$, for all pairs $i \neq j$, we see, that N_s can be written as a countable intersection of measurable subsets of N. Hence N_s is a measurable subset of $\mathcal{N}$.

Definition 2.5.1. Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space and $(\mathsf{M}, \mathcal{M})$ be a measurable space of locally finite measures. A map

$$\xi : \Omega \to M$$

is called a **random measure** if ξ is $\mathcal{A} - \mathcal{M}$ measurable. The image measure $\mathbb{P}_\xi$ on $(\mathsf{M}, \mathcal{M})$ given by $\mathbb{P}_\xi(A) = \mathbb{P}(\xi \in A) = \mathbb{P}(\xi^{-1}(A))$ for all $A \in \mathcal{M}$ is called the **distribution** of ξ. Two random measures ξ and ξ' are **stochastically equivalent**, $\xi \overset{\mathcal{D}}{\sim} \xi'$, if they have the same distribution.

A **point process** is a random measure N which is almost surely concentrated on N. If it is concentrated on N_s, we call it a **simple** point process.

Given a random measure ξ we associate a particular realization with every sample point $\omega \in \Omega$, which is a locally finite measure on E. We denote such realization by $\xi(\omega, \cdot)$ or just $\xi(\cdot)$ and sometimes just ξ. The same notation is used for point processes and simple point processes. From the construction of $\mathcal{M}, \mathcal{N}$, and $\mathcal{N}_s$ it follows that $\xi(\omega, A)$ as a function of ω is a random variable for every Borel A. The following lemma states that this condition is also sufficient.

Lemma 2.5.3. *The mapping $\xi : (\Omega, \mathcal{A}, \mathbb{P}) \to (\mathsf{M}, \mathcal{M})$ is a random measure if and only if $\{\omega \in \Omega | \xi(\omega, G) \leq r\}$ is measurable for all open, relatively compact G and all $r \geq 0$.*

The mapping $N : (\Omega, \mathcal{A}, \mathbb{P}) \to (\mathsf{N}, \mathcal{N})$ is a point process if and only if $\{\omega \in \Omega | \xi(\omega, G) = k\}$ is measurable for all open, relatively compact G and all $k \in \mathbb{N}_0$.

Proof. Lemma 3.1.5 in [40] or Proposition 9.1.III in [43]. □

For a measure $\varphi \in \mathsf{M}$, the **support** of φ, $\operatorname{supp} \varphi$, is the smallest closed set F such that $\varphi(E \setminus F) = 0$. This leads to the following connection between random measures and random closed sets.

Proposition 2.5.1. *The mapping $\varphi \to \operatorname{supp} \varphi$ from M to $\mathcal{F}$ is measurable. Moreover for every random measure ξ, $\operatorname{supp} \xi$ is a random closed set whose distribution is uniquely determined by the distribution of ξ.*

Proof. Proposition 8.16 in [42] or Lemma 3.1.4 in [40]. □

Random measures can be constructed from random sets. One particular way is given by the following proposition.

Proposition 2.5.2. *Let X be a random closed set in E and let μ be a locally finite measure on E. Then $\mu_X : \Omega \to \mathsf{M}$ defined by $\mu_X(\omega, B) = \mu(X(\omega) \cap B)$ for each $B \in \mathcal{B}(E)$ is a random measure.*

Proof. Let us take an open, relatively compact $G \subset E$ and $r \geq 0$. The restriction $\mu|_G$ of μ on G is a finite measure. From Theorem A.4.4 follows that it is upper-semi continuous on $\mathcal{F}$ and therefore measurable. The composition $\mu|_G \circ X : \Omega \to \mathbb{R}$ given by $(\mu|_G \circ X)(\omega) = \mu(X(\omega) \cap B)$ is also measurable. Hence the set $\{\omega \in \Omega | \mu_X(\omega, B) \leq r\} = (\mu|_G \circ X)^{-1}([0, r])$ is measurable, which together with Lemma 2.5.3 completes the proof. □

In the case of simple point processes the correspondence is much stronger.

Proposition 2.5.3. *N is a simple point process if and only if its support* $\operatorname{supp} N$ *is a locally finite random closed set.*

Proof. Lemma 3.1.4 in [40]. □

Finally we present the result similar to the Choquet theorem 2.2.1 for random closed sets.

Theorem 2.5.1. *The mapping $N : (\Omega, \mathcal{A}, \mathbb{P}) \to (\mathsf{N}_s, \mathcal{N}_s)$ is a simple point process if and only if $\{\omega \in \Omega | N(\omega, C) = 0\}$ is measurable for all $C \in \mathcal{C}$.*

Let N, N' be simple point processes in E. If $\mathbb{P}(N(C) = 0) = \mathbb{P}(N'(C) = 0)$ for all $C \in \mathcal{C}$, then $N \overset{\mathcal{D}}{\sim} N'$.

Proof. Theorem 3.1.1 in [40]. □

Let G be a topological group that acts measurably on the space E. Then G acts in a canonical way on M by letting

$$S_g \xi(B) = \xi(g^{-1} B) \quad \text{for} \quad B \in \mathcal{B}(E), g \in G. \tag{2.17}$$

It can be shown (see Section 3.1 in [40]) that the mapping $\xi \mapsto S_g \xi$ on $(\mathsf{M}, \mathcal{M})$ is measurable. Therefore if ξ is a random measure and $g \in G$, then $S_g \xi$ is a random measure. Important examples are when $E = \mathbb{R}^d$ or $E = \mathcal{F}'(\mathbb{R}^d)$. For the group G of rigid motions the acting of G on E is continuous in both cases. In the case of the group of translations, if $\boldsymbol{y} \in \mathbb{R}^d$ then $S_{\boldsymbol{y}} \xi(B) = \xi(B - \boldsymbol{y})$ for all $B \in \mathcal{B}(\mathbb{R}^d)$ and we use the notation $S_{\boldsymbol{y}} \xi \equiv \xi + \boldsymbol{y}$.

Note that the action S_g is for random closed sets and random measures defined consistently in the sense that

$$\operatorname{supp} S_g \xi = S_g \operatorname{supp} \xi. \tag{2.18}$$

Moreover, if μ is a locally finite measure on E invariant with respect to S_g and X is a random closed set, then

$$S_g \mu_X = \mu_{S_g X}, \tag{2.19}$$

where μ_X is defined as in Proposition 2.5.2.

Definition 2.5.2. The random measure (point process) ξ on $E = \mathbb{R}^d$ or $E = \mathcal{F}'(\mathbb{R}^d)$ is **stationary** if $\xi \overset{\mathcal{D}}{\sim} \xi + \boldsymbol{x}$ for all $\boldsymbol{x} \in \mathbb{R}^d$. It is **isotropic** if $\xi \overset{\mathcal{D}}{\sim} \theta \xi$ for all rotations $\theta \in SO_d$.

2.5.1 Moment measures

Here we focus on the properties of random measures (point processes) that are analogous to moments of random variables.

Definition 2.5.3. Let ξ be a random measure on E and let Λ be a measure on E defined by

$$\Lambda(B) = \mathbb{E}\,\xi(B)$$

for every Borel B in E. If Λ is locally finite, then we call it the **intensity measure (first moment measure)**.

It follows easily that Λ is a Borel measure: since $\Lambda(\emptyset) = 0$, it is finitely additive and the σ-additivity follows from Lebesgue's monotone convergence theorem. Generally the expectation $\mathbb{E}\,\xi(B)$ may be infinite even for bounded Borel sets, but we take into account only random measures with locally finite Λ.

By standard arguments of a measure theory it is easy to proof the following result.

Theorem 2.5.2 (Campbell)**.** *Let ξ be a random measure on E with intensity measure Λ, and let $f : E \to \mathbb{R}$ be a non-negative measurable function. Then $\int f\,\mathrm{d}\xi$ is measurable and*

$$\mathbb{E}\int_E f\,\mathrm{d}\xi = \int_E f\,\mathrm{d}\Lambda.$$

Proof. Theorem 3.1.2 in [40]. □

Similarly to the intensity measure we can construct high-order measures. Let $E^k \equiv E \times \cdots \times E$ be equipped with the usual product Borel σ-algebra. For a random measure ξ we use the notation $\xi^{(k)}$ for a Borel measure on E^k given almost surely by

$$\xi^{(k)}(B_1 \times \cdots \times B_k) = \xi(B_1)\cdots\xi(B_k)$$

for all $B_1, \ldots, B_k \in \mathcal{B}(E)$.

Definition 2.5.4. The **kth order moment measure** $\Lambda^{(k)}$ of ξ is the intensity measure of $\xi^{(k)}$, whenever this intensity measure exists.

Therefore if the kth order moment measure exists, we have

$$\Lambda^{(k)}(B_1 \times \cdots \times B_k) = \mathbb{E}\,\xi(B_1)\cdots\xi(B_k)$$

for all $B_1, \ldots, B_k \in \mathcal{B}(E)$, and $\Lambda^{(k)}$ is locally finite. A special role is played by the subset $E^k_{\neq}$ of E^k obtained by removing all sub-diagonals of E^k,

$$E^k_{\neq} = \{(x_1, \ldots, x_k) \in E^k | x_i \neq x_j \text{ for all } i \neq j\}.$$

Definition 2.5.5. Let ξ be a random measure with kth order moment measure $\Lambda^{(k)}$. The restriction $\Lambda^{[k]} = \Lambda^{(k)}\,|_{E^k_{\neq}}$ of $\Lambda^{(k)}$ on $E^k_{\neq}$ is called the **kth order factorial moment measure** of ξ.

Thus the kth order factorial moment measure is a Borel measure given by

$$\Lambda^{[k]}(B_1 \times \cdots \times B_k) = \mathbb{E}\,\xi^{(k)}\big(B_1 \times \cdots \times B_k \cap E^k_{\neq}\big)$$

for all $B_1, \ldots, B_k \in \mathcal{B}(E)$. Moment measures and factorial moment measures again satisfy the Campbell theorem.

Theorem 2.5.3. *Let ξ be a random measure on E with kth order moment measure $\Lambda^{(k)}$, and let $f : E^k \to \mathbb{R}$ be a non-negative measurable function. Then $\int_{E^k} f \,\mathrm{d}\xi^{(k)}$ and $\int_{E^k} f \,\mathrm{d}\xi^{(k)} \,|_{E^k_{\neq}}$ are measurable, and*

$$\mathbb{E}\int_{E^k} f \,\mathrm{d}\xi^{(k)} = \int_{E^k} f \,\mathrm{d}\Lambda^{(k)} \quad \text{and} \quad \mathbb{E}\int_{E^k} f \,\mathrm{d}\xi^{(k)} \,|_{E^k_{\neq}} = \int_{E^k} f \,\mathrm{d}\Lambda^{[k]}.$$

Finally, the **second order covariance measure** $C^{(2)}$ is for every bounded $A, B \in \mathcal{B}(E)$ given by

$$C^{(2)}(A \times B) = \Lambda^{(2)}(A \times B) - \Lambda(A)\Lambda(B) = \operatorname{cov}\big(\xi(A), \xi(B)\big),$$

whenever both $\Lambda^{(2)}$ and Λ are locally finite. The second order factorial covariance measure $C^{[2]}$ is defined analogously. It should be noted that both $C^{(2)}$ and $C^{[2]}$ are signed measures in the sense of Section A.2 rather than ordinary Borel measures. Moreover, both can be regarded as second order moment measures of the mean-corrected random signed measure

$$\tilde{\xi} = \xi - \Lambda,$$

again in the functional meaning.

In the following, we consider the case $E = \mathbb{R}^d$. If the intensity measure Λ of a random measure ξ is absolutely continuous with respect to the d-dimensional Lebesgue measure ν_d, we call its density the **intensity** of ξ and denote it by λ. Thus

$$\Lambda(B) = \int_B \lambda \,\mathrm{d}\nu_d = \int_B \lambda(\boldsymbol{x}) \,\mathrm{d}\boldsymbol{x}$$

for every $B \in \mathcal{B}(\mathbb{R}^d)$. If ξ is a stationary random measure on $\mathbb{R}^d$, then Λ is clearly invariant under translations. Moreover, since it is well known that the only translation invariant locally finite measure on $\mathbb{R}^d$ is, up to a constant factor, the Lebesgue measure ν_d, we claim that

$$\Lambda = \lambda\nu_d$$

for some non-negative constant λ, which is then the intensity of ξ.

Similarly to the intensity measure one can study densities of high-order moment measures and of high-order factorial moment measures. Let ξ be a random measure with kth order intensity measure $\Lambda^{(k)}$ that is absolutely continuous with respect to the Lebesgue measure ν_{dk} on $\mathbb{R}^{dk}$. The density $\rho^{(k)}$ of $\Lambda^{(k)}$ is called the **kth order product density**. Analogously we define the **kth order factorial product density** $\rho^{[k]}$. Thus

$$\Lambda^{(k)}(B_1 \times \cdots \times B_k) = \int_{B_1} \cdots \int_{B_k} \rho^{(k)}(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k) \,\mathrm{d}\boldsymbol{x}_1 \ldots \,\mathrm{d}\boldsymbol{x}_k,$$

for all $B_1, \ldots, B_k \in \mathcal{B}(\mathbb{R}^d)$, and equivalently for $\Lambda^{[k]}$ and $\rho^{[k]}$.

Now we turn the attention to reduced moment measures that appear in high order ergodic theorems.

Definition 2.5.6. A random measure ξ on $\mathbb{R}^d$ is **kth order stationary**, if its kth moment measure exists, and for each $j = 1, \ldots, k$, bounded Borel sets $B_1, \ldots, B_j$, and $\boldsymbol{x} \in \mathbb{R}^d$,

$$\Lambda^{(j)}\big((B_1 + \boldsymbol{x}) \times \cdots \times (B_j + \boldsymbol{x})\big) = \Lambda^{(j)}(B_1 \times \cdots \times B_j).$$

It is easy to see that if ξ is stationary random measure for which the kth order moment measure exists, it is kth order stationary. The converse is not true in general (see Section 12.6 in [43] for a deeper discussion).

Proposition 2.5.4. *Let ξ be a second order stationary random measure. Then there exists $\lambda_D \geq 0$ such that*

$$\Lambda^{(2)}(B_1 \times B_2) = \lambda_D \nu_d(B_1 \cap B_2) + \Lambda^{[2]}(B_1 \times B_2)$$

for all $B_1, B_2 \in \mathcal{B}(\mathbb{R}^d)$.

Proof. Since $(\mathbb{R}^d)^2_{\neq} = \{(\boldsymbol{x}, \boldsymbol{y}) \in (\mathbb{R}^d)^2 | \boldsymbol{x} \neq \boldsymbol{y}\}$ we have $D = \{(\boldsymbol{x}, \boldsymbol{x}) \in (\mathbb{R}^d)^2 | \boldsymbol{x} \in \mathbb{R}^d\} = (\mathbb{R}^d)^2 \setminus (\mathbb{R}^d)^2_{\neq}$. Thus $\Lambda^{(2)} - \Lambda^{[2]} = \Lambda^{(2)} \mid_D$ and

$$\Lambda^{(2)} \mid_D (B_1 \times B_2) = \Lambda^{(2)}((B_1 \times B_2) \cap D) = \Lambda^{(2)}\big(\{(\boldsymbol{x}, \boldsymbol{x}) | \boldsymbol{x} \in B_1 \cap B_2\}\big),$$

for all Borel sets $B_1, B_2 \in \mathbb{R}^d$. Let us define new Borel measure μ on $\mathbb{R}^d$ by setting $\mu(B) = \Lambda^{(2)}\big(\{(\boldsymbol{x}, \boldsymbol{x}) | \boldsymbol{x} \in B\}\big)$ for all Borel B. Clearly, μ is locally finite. From second order stationarity of ξ follows that $\mu(B + \boldsymbol{x}) = \mu(B)$ for all Borel B and all $\boldsymbol{x} \in \mathbb{R}^d$. Hence μ is a multiple of the Lebesgue measure, $\mu = \lambda_D \nu_d$, and the proof is complete. □

Now we proceed to reduced measures.

Proposition 2.5.5. *For kth order (second order) stationary random measures on $\mathbb{R}^d$, there exist reduced measures $\breve{\Lambda}^{(k)}, \breve{\Lambda}^{[k]}$ and $\breve{C}^{(2)}, \breve{C}^{[2]}$ related to the corresponding kth order measures $\Lambda^{(k)}, \Lambda^{[k]}$ and second order measures $C^{(2)}, C^{[2]}$, respectively, through equations valid for any bounded Borel functions of bounded support on $(\mathbb{R}^d)^k$, of the type*

$$\int_{(\mathbb{R}^d)^k} f(\boldsymbol{x}_1, \dots, \boldsymbol{x}_k) \Lambda^{(k)}(\mathrm{d}\boldsymbol{x}_1 \times \cdots \times \mathrm{d}\boldsymbol{x}_k)$$

$$= \int_{\mathbb{R}^d} \mathrm{d}\boldsymbol{x} \int_{(\mathbb{R}^d)^{k-1}} f(\boldsymbol{x}, \boldsymbol{x} + \boldsymbol{y}_1, \dots, \boldsymbol{x} + \boldsymbol{y}_{k-1}) \breve{\Lambda}^{(k)}(\mathrm{d}\boldsymbol{y}_1 \times \cdots \times \mathrm{d}\boldsymbol{y}_{k-1}).$$

Proof. Proposition 12.6.III in [43]. □

We call $\breve{\Lambda}^{(k)}, \breve{\Lambda}^{[k]}, \breve{C}^{(2)}$, and $\breve{C}^{[2]}$ the **reduced kth order measure**, the **reduced kth order factorial measure**, the **reduced second order covariance measure**, and the **reduced second order factorial covariance measure**, respectively. Note that for $k = 1$ the reduced measure and the reduced factorial measure coincide and are equal to the constant intensity λ. The measures $\breve{C}^{(2)}$ and $\breve{C}^{[2]}$ are signed measures on $\mathbb{R}^d$ in the sense of Section A.2. The density of $\breve{C}^{[2]}$, if it is absolutely continuous with respect to ν_d, is called the **covariance function** and denoted by cov.

Reduced moment measures have some important properties that are, for the particular choice of $\breve{\Lambda}^{(k)}$, summarized in the following proposition, which holds for all other reduced moment measures introduced above.

Proposition 2.5.6. *Let $\breve{\Lambda}^{(k)}$ be the kth order reduced moment measure of a kth order stationary random measure ξ on $\mathbb{R}^d$.*

*(a) $\breve{\Lambda}^{(k)}$ is a **symmetric measure** on $(R^{d(k-1)}$, i.e. $\breve{\Lambda}^{(k)}(A) = \breve{\Lambda}^{(k)}(-A)$ for all Borel sets $A \in \mathbb{R}^{d(k-1)}$.*

(b) $\breve{\Lambda}^{(k)}$ is invariant under the shift reflection transformation mapping $(\boldsymbol{u}_1, \boldsymbol{u}_2, \dots, \boldsymbol{u}_{k-1}) \mapsto (-\boldsymbol{u}_1, \boldsymbol{u}_2 - \boldsymbol{u}_1, \dots, \boldsymbol{u}_{k-1} - \boldsymbol{u}_1)$.

(c) When $\Lambda^{(k)}$ is absolutely continuous with respect to the dk-dimensional Lebesgue measure ν_{dk}, with density $\rho^{(k)}$, then $\breve{\Lambda}^{(k)}$ is absolutely continuous with respect to $\nu_{d(k-1)}$, and its density $\breve{\rho}^{(k)}$ is related to $\rho^{(k)}$ by

$$\rho^{(k)}(\boldsymbol{x}_1, \dots, \boldsymbol{x}_k) = \breve{\rho}^{(k)}(\boldsymbol{x}_2 - \boldsymbol{x}_1, \dots, \boldsymbol{x}_k - \boldsymbol{x}_1).$$

(d) If $k = 2$, then $\breve{\Lambda}^{(2)}$ is positive semi-definite according to Definition A.2.1, i.e. for every continuous function f of compact support on $\mathbb{R}^d$,

$$\int_{\mathbb{R}^d} (f * f^*)(\boldsymbol{x})\ \breve{\Lambda}^{(2)}(\mathrm{d}\boldsymbol{x}) \geq 0,$$

where $f * f^*$ is defined by

$$(f * f^*)(\boldsymbol{x}) = \int_{\mathbb{R}^d} f(\boldsymbol{y})\overline{f(\boldsymbol{y}-\boldsymbol{x})}\,\mathrm{d}\boldsymbol{y}.$$

*(e) If $k = 2$, then $\breve{\Lambda}^{(2)}$ is **translation bounded**, i.e. for every bounded Borel set A in $\mathbb{R}^d$, there exists a finite constant K_A such that*

$$\breve{\Lambda}^{(k)}(\boldsymbol{x} + A) \leq K_A \quad \text{for all } \boldsymbol{x} \in \mathbb{R}^d.$$

Proof. Proposition 12.6.IV in [43] and Proposition 8.1.II in [56]. □

The condition of positive semi-definiteness can be equivalently formulated using bounded functions of bounded support as is shown in Proposition A.2.3.

Corollary 2.5.1. *The reduced second order (factorial) covariance measure $\breve{C}^{(2)}$ ($\breve{C}^{[2]}$) of a second order stationary random measure ξ on $\mathbb{R}^d$ is symmetric, positive semi-definite and translation bounded signed measure.*

Proof. Corollary 8.1.III in [56]. □

2.5.2 Characteristics of point processes

Throughout this subsection $E = \mathbb{R}^d$. For the intensity measure of a point process N we have

$$\Lambda(B) = \mathbb{E}\sum_{\boldsymbol{x}\in N} \mathbb{1}_B(\boldsymbol{x})$$

for every Borel B. If Λ is absolutely continuous with respect to ν_d, then

$$\Lambda(B) = \mathbb{E}\sum_{\boldsymbol{x}\in N} \mathbb{1}_B(\boldsymbol{x}) = \int_B \lambda(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}. \tag{2.20}$$

For stationary N the intensity is constant, $\lambda(\boldsymbol{x}) = \lambda$.

Let N be the point process with second order moment measure $\Lambda^{(2)}$. The relation between $\Lambda^{(2)}$ and $\Lambda^{[2]}$ can be easily derived. For $A, B \in \mathcal{B}(E)$ we obtain

$$\begin{aligned}
\Lambda^{(2)}(A \times B) &= \mathbb{E}\sum_{\boldsymbol{x}_1,\boldsymbol{x}_2\in N} \mathbb{1}_A(\boldsymbol{x}_1)\mathbb{1}_B(\boldsymbol{x}_2) \\
&= \mathbb{E}\sum_{\boldsymbol{x}_1,\boldsymbol{x}_2\in N}^{\neq} \mathbb{1}_A(\boldsymbol{x}_1)\mathbb{1}_B(\boldsymbol{x}_2) + \mathbb{E}\sum_{\boldsymbol{x}\in N} \mathbb{1}_A(\boldsymbol{x})\mathbb{1}_B(\boldsymbol{x}) \\
&= \Lambda^{[2]}(A \times B) + \Lambda(A \cap B).
\end{aligned}$$

Note that this relation holds for arbitrary E and not just $\mathbb{R}^d$. It is clear that $\Lambda^{(2)}(A\times B)$ cannot be absolutely continuous with respect to ν_{2d} whenever is Λ, and hence N, non-trivial. On the

other side, for the factorial moment measure $\Lambda^{[2]}$ the absolute continuity is often assumed. For such a point process we can write Theorem 2.5.3 in the form:

$$\mathbb{E}\left(\sum_{\boldsymbol{x},\boldsymbol{y}\in N}^{\neq} f(\boldsymbol{x},\boldsymbol{y})\right) = \int_{\mathbb{R}^d\times\mathbb{R}^d} f(\boldsymbol{x},\boldsymbol{y})\rho^{[2]}(\boldsymbol{x},\boldsymbol{y})\,\mathrm{d}\boldsymbol{x}\,\mathrm{d}\boldsymbol{y}$$

for any non-negative, measurable function on $\mathbb{R}^d \times \mathbb{R}^d$.

The second order factorial product density $\rho^{[2]}$ normalized by the intensities $\lambda(\boldsymbol{x})$ and $\lambda(\boldsymbol{y})$,

$$g(\boldsymbol{x},\boldsymbol{y}) = \frac{\rho^{[2]}(\boldsymbol{x},\boldsymbol{y})}{\lambda(\boldsymbol{x})\lambda(\boldsymbol{y})} \quad \text{for all } \boldsymbol{x},\boldsymbol{y}\in\mathbb{R}^d,$$

is called the **pair correlation function**. If N is second order stationary then from Proposition 2.5.6 (c) follows

$$g(\boldsymbol{x},\boldsymbol{y}) = \frac{\rho^{[2]}(\boldsymbol{x},\boldsymbol{y})}{\lambda^2} = \frac{\breve{\rho}^{[2]}(\boldsymbol{x}-\boldsymbol{y})}{\lambda^2} = g(\boldsymbol{r}),$$

where $\boldsymbol{r} = \boldsymbol{x}-\boldsymbol{y}$, and the pair correlation function is thus dependent only on the difference of $\boldsymbol{x}$ and $\boldsymbol{y}$. Moreover, if N is isotropic it can be easily shown that $g(\boldsymbol{r}) = g(r)$, where $r = \|\boldsymbol{r}\|$.

Important integral characteristic of a second order stationary point process N is the reduced second order factorial moment measure $\breve{\Lambda}^{[2]}$ that is applied on the ball $B_r(\boldsymbol{0})$ and normalized by λ^2. The result is called **Ripley's K-function** and denoted by $K(r)$. Thus we have

$$K(r) = \frac{\breve{\Lambda}^{[2]}(B_r(\boldsymbol{0}))}{\lambda^2} \quad \text{for all } r\geq 0.$$

For a second order stationary point process that is also isotropic one can, using spherical coordinates, directly derive the connection between the K-function and the pair correlation function that is given by

$$g(r) = \frac{1}{dc_d r^{d-1}}\frac{\mathrm{d}K(r)}{\mathrm{d}r} \quad \text{for all } r\geq 0,$$

where c_d is the d-dimensional volume of the unit sphere in $\mathbb{R}^d$.

Finally, we show the useful connection between the variance of a number of points in a Borel set, the pair correlation function, and the intensity. Let N be a second order stationary point process with intensity λ, pair correlation function g, and let $B\subset\mathbb{R}^d$ be a Borel set. From previous considerations and Propositions 2.5.5 and 2.5.6 follows

$$\begin{aligned}
\operatorname{var}(N(B)) &= \Lambda^{(2)}(B,B) - \Lambda(B)^2 = \Lambda^{[2]}(B,B) + \Lambda(B) - \Lambda(B)^2 \\
&= \int_{\mathbb{R}^d\times\mathbb{R}^d} \mathbb{1}_B(\boldsymbol{x})\mathbb{1}_B(\boldsymbol{y})\rho^{[2]}(\boldsymbol{x},\boldsymbol{y})\,\mathrm{d}\boldsymbol{x}\,\mathrm{d}\boldsymbol{y} + \lambda\nu_d(B) - \lambda^2\nu_d^2(B) \\
&= \int_{\mathbb{R}^d\times\mathbb{R}^d} \mathbb{1}_B(\boldsymbol{x})\mathbb{1}_B(\boldsymbol{x}+\boldsymbol{y})\breve{\rho}^{[2]}(\boldsymbol{y})\,\mathrm{d}\boldsymbol{x}\,\mathrm{d}\boldsymbol{y} + \lambda\nu_d(B) - \lambda^2\nu_d^2(B) \\
&= \lambda^2\int_{\mathbb{R}^d} \gamma_B(\boldsymbol{y})g(\boldsymbol{y})\,\mathrm{d}\boldsymbol{y} + \lambda\nu_d(B) - \lambda^2\nu_d^2(B), \qquad (2.21)
\end{aligned}$$

where $\gamma_B(\boldsymbol{y})$ is the **set covariance** of B defined by

$$\gamma_B(\boldsymbol{r}) = \nu_d\big(B\cap(B-\boldsymbol{r})\big) \qquad (2.22)$$

for all $\boldsymbol{r} \in \mathbb{R}^d$. If N is isotropic, we may use the spherical coordinates and further obtain

$$\begin{aligned}\operatorname{var}(N(B)) &= \lambda^2 \int_{\mathbb{R}^d} \gamma_B(\boldsymbol{y}) g(\boldsymbol{y}) \,\mathrm{d}\boldsymbol{y} + \lambda\nu_d(B) - \lambda^2\nu_d^2(B) \\ &= \lambda^2 \int_0^\infty \int_{S^{d-1}} \gamma_B(r\boldsymbol{u}) g(r) r^{d-1} \,\sigma_{d-1}(\mathrm{d}\boldsymbol{u})\,\mathrm{d}r + \lambda\nu_d(B) - \lambda^2\nu_d^2(B) \\ &= \lambda^2 dc_d \int_0^\infty \overline{\gamma}_B(r) g(r) r^{d-1}\,\mathrm{d}r + \lambda\nu_d(B) - \lambda^2\nu_d^2(B),\end{aligned}$$

where S^{d-1} is the unit sphere in $\mathbb{R}^d$, σ_{d-1} is the usual spherical measure (non-normalized and thus equal to the $d-1$ dimensional Hausdorff measure) on S^{d-1}, dc_d is the surface area of S^{d-1}, $dc_d = \sigma_{d-1}(S^{d-1})$, and $\overline{\gamma}_B(r)$ is the isotropised set covariance given by

$$\overline{\gamma}_B(r) = \frac{1}{dc_d} \int_{S^{d-1}} \gamma_B(r\boldsymbol{u}) \,\sigma_{d-1}(\mathrm{d}\boldsymbol{u})$$

for all $r \geq 0$.

2.5.3 The Poisson point process

In this part we introduce the most important point process model. We follow the definition in [40] Section 3.2. Let E be a locally compact, Hausdorff topological space with a countable base.

Definition 2.5.7. Let N be a simple point process in E with intensity measure Λ. We say that N is a **Poisson point process** in E if the two following conditions are satisfied:

(a) $\mathbb{P}\left(N(A) = k\right) = e^{-\Lambda(A)} \frac{\Lambda^k(A)}{k!}$ for all $k \in \mathbb{N}_0$ and all bounded Borel $A \subset E$,

(b) $N(A_1), \ldots, N(A_n)$ are independent random variables for every $n \in \mathbb{N}$ and all pairwise disjoint Borel sets $A_1, \ldots, A_n$ in E.

It can be shown that the intensity measure of the Poisson point process has no atoms (see e.g. [40, Lemma 3.2.1]). The following theorem shows the wideness of the class of Poisson point processes.

Theorem 2.5.4. *Let Λ be a locally finite measure on E without atoms. Then there exists a Poisson process in E with intensity measure Λ and it is (up to equivalence) uniquely determined.*

Proof. Theorem 3.2.1 in [40]. □

There is an important corollary of the previous theorems for stationary Poisson point processes in $\mathbb{R}^d$.

Corollary 2.5.2. *Let $\lambda \geq 0$. Then there is (up to equivalence) precisely one stationary Poisson process N in $\mathbb{R}^d$ with intensity λ. Moreover, N is also isotropic.*

Proof. Corollary 3.2.1 in [40]. □

The following theorem states the most important properties of the Poisson point process.

Theorem 2.5.5. *Let N be a Poisson point process in E with intensity measure Λ.*

(a) Let $A_1, A_2, \ldots$ be pairwise disjoint Borel sets in E. Then the point processes given by the restrictions $N\,|_{A_1}, N\,|_{A_2}, \ldots$ are independent.

(b) Let A be a Borel set in E with $0 < \Lambda(A) < \infty$ and let $k \in \mathbb{N}$. Then under the condition $N(A) = k$ the points $\zeta_1, \ldots, \zeta_k$ of $N\,|_A$ are independent, identically distributed random points in E with distribution

$$\mathbb{P}_{\zeta_i} = \frac{\Lambda\,|_A}{\Lambda(A)}, \quad i = 1, \ldots, k.$$

Proof. Theorem 3.2.2 in [40]. □

The last statement shows that the factorial moment measures of a Poisson point process are the powers of its intensity measure.

Proposition 2.5.7. *For a Poisson point process N in E with intensity measure Λ and for $m \in \mathbb{N}$, the mth factorial moment measure $\Lambda^{(m)}$ of N satisfies*

$$\Lambda^{[m]} = \Lambda^m.$$

Proof. Corollary 3.2.4 in [40]. □

2.5.4 Relation to random closed sets

We again assume $E = \mathbb{R}^d$ throughout this subsection. In proposition 2.5.2 we have seen how to construct a random measure from a random closed set. The canonical way is to use the d-dimensional Lebesgue measure ν_d.

Definition 2.5.8. Let X be a random closed set in $\mathbb{R}^d$. The random measure ν_X defined by $\nu_X(B) = \nu_d(X \cap B)$ for each $B \in \mathcal{B}(\mathbb{R}^d)$ is called the **volume measure** of X.

By Fubini-Tonelli's theorem, it follows that for every $B \in \mathcal{B}(\mathbb{R}^d)$

$$\Lambda(B) = \mathbb{E}\,\nu_X(B) = \mathbb{E}\,\nu(X \cap B) = \int_B \mathbb{E}\,\mathbb{1}_X(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} = \int_B m(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x},$$

where $m(\boldsymbol{x})$ is the volume fraction of X. Therefore Λ is locally finite and hence it is the intensity measure with intensity $\lambda(\boldsymbol{x})$ that is ν_d-almost surely equal to the volume fraction $m(\boldsymbol{x})$ of X. For a stationary RACS X, both λ and m are constant and therefore $\lambda = m = p$.

By the same arguments as for the intensity measure we obtain that the kth order moment measure $\Lambda^{(k)}$ exists and is absolutely continuous with respect to the Lebesgue measure ν_{dk} on $(\mathbb{R}^d)^k$. Thus the stationary RACS X is also kth order stationary for every $k \geq 1$. The corresponding kth order product density $\rho^{(k)}$ is ν_{dk}-almost surely given by

$$\rho^{(k)}(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k) = \mathbb{E}\,\mathbb{1}_X(\boldsymbol{x}_1) \cdot \ldots \cdot \mathbb{1}_X(\boldsymbol{x}_k)$$

for all $\boldsymbol{x}_1, \dots, \boldsymbol{x}_k \in \mathbb{R}^d$. Since obviously $\nu_{dk}\big((\mathbb{R}^d)^k \setminus (\mathbb{R}^d)^k_{\neq}\big) = 0$, the definition of the kth order factorial moment measure yields

$$\begin{aligned}
\Lambda^{[k]}(B_1 \times \cdots \times B_k) &= \mathbb{E} \int_{(B_1 \times \ldots \times B_k) \cap (\mathbb{R}^d)^k_{\neq}} \mathbb{1}_X(\boldsymbol{x}_1) \dots \mathbb{1}_X(\boldsymbol{x}_k) \, \mathrm{d}\boldsymbol{x}_1 \dots \, \mathrm{d}\boldsymbol{x}_k \\
&= \mathbb{E} \int_{B_1 \times \ldots \times B_k} \mathbb{1}_X(\boldsymbol{x}_1) \dots \mathbb{1}_X(\boldsymbol{x}_k) \, \mathrm{d}\boldsymbol{x}_1 \dots \, \mathrm{d}\boldsymbol{x}_k \\
&= \int_{B_1 \times \ldots \times B_k} \rho^{(k)}(\boldsymbol{x}_1, \dots, \boldsymbol{x}_k) \, \mathrm{d}\boldsymbol{x}_1 \dots \, \mathrm{d}\boldsymbol{x}_k \\
&= \int_{(B_1 \times \ldots \times B_k) \cap (\mathbb{R}^d)^k_{\neq}} \rho^{(k)}(\boldsymbol{x}_1, \dots, \boldsymbol{x}_k) \, \mathrm{d}\boldsymbol{x}_1 \dots \, \mathrm{d}\boldsymbol{x}_k \\
&= \Lambda^{(k)}(B_1 \times \cdots \times B_k)
\end{aligned}$$

for all $B_1, \dots, B_k \in \mathcal{B}(\mathbb{R}^d)$. Hence

$$\Lambda^{(k)} = \Lambda^{[k]} \quad \text{and} \quad \rho^{(k)} = \rho^{[k]}.$$

For the particular value $k = 2$ we have

$$\rho^{(2)}(\boldsymbol{x}, \boldsymbol{y}) = \rho^{[2]}(\boldsymbol{x}, \boldsymbol{y}) = \mathbb{E}\, \mathbb{1}_X(\boldsymbol{x}) \mathbb{1}_X(\boldsymbol{y}) = C(\boldsymbol{x}, \boldsymbol{y})$$

for all $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^d$, where $C(\boldsymbol{x}, \boldsymbol{y})$ is the covariance of X. The second order covariance measure $C^{(2)}$ equals the second order factorial covariance measure $C^{[2]}$ and both are also absolutely continuous with density given by the covariance function $\operatorname{cov}(\boldsymbol{x}, \boldsymbol{y})$ of X. Hence λ_D in Proposition 2.5.4 equals 0.

2.5.5 The Bartlett spectrum

The Bartlett spectrum of a stationary random closed set was defined in Subsection 2.3.3. Here we extend the notion to stationary random measures. The construction is adapted from [56]. For the majority of statements we provide full direct proofs.

Let ξ be a second order stationary random measure (point process) and $\breve{C}^{[2]}$ the reduced second order factorial covariance measure of ξ. The measure $\breve{C}^{[2]}$ is understood as a signed measure on $\mathbb{R}^d$ in the sense of Section A.2. From Proposition 2.5.5 defining $\breve{C}^{[2]}$ follows that

$$\breve{C}^{[2]} = \breve{\Lambda}^{[2]} - \lambda^2 \nu_d. \tag{2.23}$$

Thus $\breve{C}^{[2]}$ is a difference of two positive measures. Corollary 2.5.1 implies that $\breve{C}^{[2]}$ is positive semi-definite. Thus from Theorem A.2.3 there exists a uniquely given Fourier transform of $\breve{C}^{[2]}$, here denoted by Γ, such that

$$\int_{\mathbb{R}^d} (f * f^*)(\boldsymbol{x}) \; \breve{C}^{[2]}(\mathrm{d}\boldsymbol{x}) = (2\pi)^{d/2} \int_{\mathbb{R}^d} |\check{f}(\boldsymbol{\omega})|^2 \; \Gamma(\mathrm{d}\boldsymbol{\omega})$$

for all $f \in C_c(\mathbb{R}^d)$, where f^* is the involution defined by (A.6) and $\check{f}(\boldsymbol{\omega}) = \hat{f}(-\boldsymbol{\omega})$ is given by (A.8).

Definition 2.5.9. The **Bartlett spectrum** of a second order stationary random measure ξ on $\mathbb{R}^d$ is the Fourier transform Γ of the reduced second order factorial covariance measure $\breve{C}^{[2]}$ of ξ.

Note that the Fourier transform of a positive semi-definite measure is a positive measure and hence a locally finite Borel measure in the usual sense. Due to specific properties of $\breve{C}^{[2]}$, the Bartlett spectrum has several specific properties as well.

Proposition 2.5.8. *Let Γ be the Bartlett spectrum of a second order stationary random measure ξ on $\mathbb{R}^d$. Then*

(a) Γ is symmetric, positive, and translation-bounded (see Proposition 2.5.6),

(b) for all bounded measurable functions f of bounded support,

$$\int_{\mathbb{R}^d} (f * f^*)(\boldsymbol{x})\ \breve{C}^{[2]}(\mathrm{d}\boldsymbol{x}) = (2\pi)^{d/2} \int_{\mathbb{R}^d} |\check{f}(\boldsymbol{\omega})|^2\ \Gamma(\mathrm{d}\boldsymbol{\omega}),$$

(c) for all $\boldsymbol{\omega} \in \mathbb{R}^d$,

$$\Gamma\big(\{\boldsymbol{\omega}\}\big) = \lim_{a\to\infty} \frac{(2\pi)^{d/2}}{(2a)^d} \int_{[-a,a]^d} e^{-i\boldsymbol{\omega}\cdot\boldsymbol{x}}\ \breve{C}^{[2]}(\mathrm{d}\boldsymbol{x}),$$

(d) if $\zeta_f = \int_{\mathbb{R}^d} f(\boldsymbol{x})\ \xi(\mathrm{d}\boldsymbol{x})$ for a bounded measurable real function f of bounded support,

$$\operatorname{var}\zeta_f = \lambda_D (f * f^*)(\boldsymbol{0}) + (2\pi)^{d/2} \int_{\mathbb{R}^d} |\check{f}(\boldsymbol{\omega})|^2\ \Gamma(\mathrm{d}\boldsymbol{\omega}) \geq 0,$$

where λ_D is the constant from Proposition 2.5.4,

(e) it holds

$$\Gamma\big(\{\boldsymbol{0}\}\big) = \lim_{a\to\infty} \frac{(2\pi)^{d/2} \operatorname{var}\xi\big([-a,a]^d\big)}{(2a)^{2d}}.$$

Proof. (a) Let $f \in C_c(\mathbb{R}^d)$. Defining $g(\boldsymbol{x}) = f(-\boldsymbol{x})$ for every $\boldsymbol{x}$ it is easy to show $\check{g}(\boldsymbol{\omega}) = \check{f}(-\boldsymbol{\omega})$ and $g * g^*(\boldsymbol{x}) = f * f^*(-\boldsymbol{x})$. From the symmetry of $\breve{C}^{[2]}$ (Corollary 2.5.1) follows

$$\int_{\mathbb{R}^d} (g * g^*)(\boldsymbol{x})\ \breve{C}^{[2]}(\mathrm{d}\boldsymbol{x}) = \int_{\mathbb{R}^d} (f * f^*)(\boldsymbol{x})\ \breve{C}^{[2]}(\mathrm{d}\boldsymbol{x}).$$

Since $g \in C_c(\mathbb{R}^d)$ then Theorem A.2.3 implies

$$\int_{\mathbb{R}^d} |\check{f}(\boldsymbol{\omega})|^2\ \Gamma(\mathrm{d}\boldsymbol{\omega}) = \int_{\mathbb{R}^d} |\check{f}(-\boldsymbol{\omega})|^2\ \Gamma(\mathrm{d}\boldsymbol{\omega}).$$

Since f was arbitrary, the uniqueness of Γ implies the symmetry of Γ. Positivity follows directly from Theorem A.2.3 and translation-boundedness is Proposition A.2.7.

(b) Proposition A.2.5.

(c) Let $\boldsymbol{\omega} \in \mathbb{R}^d$, $a > 0$ and $f_{a;\boldsymbol{\omega}}(\boldsymbol{x}) = (\pi/2)^{d/2} a^{-d} \mathbb{1}_{[-a,a]^d}(\boldsymbol{x}) e^{-i\boldsymbol{\omega}\cdot\boldsymbol{x}}$ for all $\boldsymbol{x} \in \mathbb{R}^d$, where $\mathbb{1}_{[-a,a]^d}$ is the indicator of the d-dimensional cube $[-a,a]^d \equiv [-a,a] \times \ldots \times [-a,a]$. The inverse Fourier transform reads

$$\check{f}_{a;\boldsymbol{\omega}}(\boldsymbol{\xi}) = \frac{1}{(2a)^d} \int_{[-a,a]^d} e^{i\boldsymbol{\xi}\cdot\boldsymbol{x}} e^{-i\boldsymbol{\omega}\cdot\boldsymbol{x}}\,\mathrm{d}\boldsymbol{x} = \prod_{j=1}^{d} \frac{\sin\big(a(\xi_j - \omega_j)\big)}{a(\xi_j - \omega_j)}.$$

From Proposition A.2.8 follows $\check{f}_{a;\boldsymbol{\omega}} \in L^1(\mathbb{R}^d, \Gamma)$ and

$$\frac{(2\pi)^{d/2}}{(2a)^d} \int_{[-a,a]^d} e^{-i\boldsymbol{\omega}\cdot\boldsymbol{x}}\, \breve{C}^{[2]}(\mathrm{d}\boldsymbol{x}) = \int_{\mathbb{R}^d} \check{f}_{a;\boldsymbol{\omega}}(\boldsymbol{\xi})\, \Gamma(\mathrm{d}\boldsymbol{\xi})$$

for every $a > 0$. Let

$$g = \check{f}_{1;\boldsymbol{\omega}} + \sum_{j=1}^{d} \big(\check{f}_{1;\boldsymbol{\omega}-\boldsymbol{s}_j} + \check{f}_{1;\boldsymbol{\omega}+\boldsymbol{s}_j}\big),$$

where $\boldsymbol{s}_j = (s_{j;1}, \dots, s_{j,d}) \in \mathbb{R}^d$ are such that $s_{j;\ell} = \pi/2\delta_{i,j}$ (zeros except $\pi/2$ at the jth position). Since $\check{f}_{a;\boldsymbol{\omega}} \to \mathbb{1}_{\{\boldsymbol{\omega}\}}$ point-wise as $a \to \infty$, it can be checked that $|f_{a;\boldsymbol{\omega}}| \le g$ for all $a > 2$ and $g \in L^1(\mathbb{R}^d, \hat{\mu})$, the Lebesgue dominated convergence theorem yields $\int_{\mathbb{R}^d} \check{f}_{a;\boldsymbol{\omega}}(\boldsymbol{\xi})\, \Gamma(\mathrm{d}\boldsymbol{\xi}) \to \Gamma\big(\{\boldsymbol{\omega}\}\big)$ as $a \to \infty$.

(d) Since ξ is stationary, $\mathbb{E}\,\zeta_f = \lambda \int_{\mathbb{R}_d} f(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}$, where λ is the intensity of ξ. From the symmetry of $\breve{C}^{(2)}$ (Corollary 2.5.1) follows

$$\begin{aligned}\operatorname{var}\zeta_f &= \mathbb{E}\,\zeta_f^2 - (\mathbb{E}\,\zeta_f)^2 \\ &= \iint_{\mathbb{R}^d\times\mathbb{R}^d} f(\boldsymbol{x})f(\boldsymbol{y})\; \Lambda^{(2)}(\mathrm{d}\boldsymbol{x}\times\mathrm{d}\boldsymbol{y}) - \lambda^2 \iint_{(\mathbb{R}^d)^2} f(\boldsymbol{x})f(\boldsymbol{y})\,\mathrm{d}\boldsymbol{x}\mathrm{d}\boldsymbol{y}.\end{aligned}$$

From Proposition 2.5.4 follows

$$\iint_{\mathbb{R}^d\times\mathbb{R}^d} f(\boldsymbol{x})f(\boldsymbol{y})\; \Lambda^{(2)}(\mathrm{d}\boldsymbol{x}\times\mathrm{d}\boldsymbol{y}) = \lambda_D \int_{\mathbb{R}^d} f^2(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} + \iint_{\mathbb{R}^d\times\mathbb{R}^d} f(\boldsymbol{x})f(\boldsymbol{y})\; \Lambda^{[2]}(\mathrm{d}\boldsymbol{x}\times\mathrm{d}\boldsymbol{y}).$$

Hence

$$\begin{aligned}\operatorname{var}\zeta_f =& \lambda_D \int_{\mathbb{R}^d} f^2(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} + \iint_{\mathbb{R}^d\times\mathbb{R}^d} f(\boldsymbol{x})f(\boldsymbol{y})\; \Lambda^{[2]}(\mathrm{d}\boldsymbol{x}\times\mathrm{d}\boldsymbol{y}) - \lambda^2 \iint_{(\mathbb{R}^d)^2} f(\boldsymbol{x})f(\boldsymbol{y})\,\mathrm{d}\boldsymbol{x}\mathrm{d}\boldsymbol{y} \\ =& \lambda_D \int_{\mathbb{R}^d} f^2(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} + \iint_{\mathbb{R}^d\times\mathbb{R}^d} f(\boldsymbol{x})f(\boldsymbol{y})\; C^{[2]}(\mathrm{d}\boldsymbol{x}\times\mathrm{d}\boldsymbol{y}) \\ =& \lambda_D \int_{\mathbb{R}^d} f^2(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} + \iint_{\mathbb{R}^d\times\mathbb{R}^d} f(\boldsymbol{y})f(\boldsymbol{y}-\boldsymbol{x})\,\mathrm{d}\boldsymbol{y}\; \breve{C}^{[2]}(\mathrm{d}\boldsymbol{x}) \\ =& \lambda_D (f * f^*)(\boldsymbol{0}) + \int_{\mathbb{R}^d} (f * f^*)(\boldsymbol{x})\; \breve{C}^{[2]}(\mathrm{d}\boldsymbol{x})\end{aligned}$$

and we can use (b). The non-negativity is obvious.

(e) Let $f_a = \mathbb{1}_{[-a,a]^d}$ for each $a > 0$. Then $\xi\big([-a,a]^d\big) = \int_{\mathbb{R}^d} f_a(\boldsymbol{x})\; \xi(\mathrm{d}\boldsymbol{x})$ and by (d),

$$\frac{(2\pi)^{d/2} \operatorname{var}\xi\big([-a,a]^d\big)}{(2a)^{2d}} = \frac{\pi^{d/2}\lambda_D}{2^{d/2}a^d} + \int_{\mathbb{R}^d} \frac{(2\pi)^d |\check{f}_a(\boldsymbol{\omega})|^2}{(2a)^{2d}}\; \Gamma(\mathrm{d}\boldsymbol{\omega}),$$

since $(f_a * f_a^*)(\boldsymbol{0}) = (2a)^d$. For the inverse Fourier transform $\check{f}_a(\boldsymbol{\omega})$ we have

$$\check{f}_a(\boldsymbol{\omega}) = (2\pi)^{-d/2} \int_{[-a,a]^d} e^{i\boldsymbol{\omega}\cdot\boldsymbol{x}}\,\mathrm{d}\boldsymbol{x} = (2\pi)^{-d/2} 2^d \prod_{j=1}^{d} \frac{\sin(a\omega_j)}{\omega_j},$$

where $\boldsymbol{\omega} = (\omega_1, \dots, \omega_d) \in \mathbb{R}^d$. The integrand is thus

$$\frac{(2\pi)^d |\check{f}_a(\boldsymbol{\omega})|^2}{(2a)^{2d}} = \prod_{j=1}^{d} \frac{\sin^2(a\omega_j)}{(a\omega_j)^2}.$$

Clearly $(2\pi)^d(2a)^{-2d}|\check{f}_a(\boldsymbol{\omega})|^2 \to \mathbb{1}_{\{\mathbf{0}\}}(\boldsymbol{\omega})$ point-wise as $a \to \infty$. Let $\vartheta(x) = 1$ for $x \in [-1, 1]$ and $\vartheta(x) = x^{-2}$ for $x \in (-\infty, -1) \cup (1, \infty)$. Obviously $g(\boldsymbol{\omega}) = \prod_{j=1}^{d} \vartheta(\omega_j) \geq (2\pi)^d(2a)^{-2d}|\check{f}_a(\boldsymbol{\omega})|^2$ for all $a > 1$ and all $\boldsymbol{\omega} \in \mathbb{R}^d$. Furthermore, from the translation boundedness of Γ and the convergence of series $\sum_{k=1}^{\infty} k^{-2}$ follows that $g \in L^1(\mathbb{R}^d, \Gamma)$. Finally, the Lebesgue dominated convergence theorem yields

$$\int_{\mathbb{R}^d} \frac{(2\pi)^d |\check{f}_a(\boldsymbol{\omega})|^2}{(2a)^{2d}} \, \Gamma(\mathrm{d}\boldsymbol{\omega}) \to \Gamma\big(\{\mathbf{0}\}\big) \quad \text{as} \quad a \to \infty,$$

which completes the proof. □

From the decomposition (2.23), the uniqueness of the Fourier transform stated in Theorem A.2.3, and Proposition A.2.6 follows that

$$\Gamma = \Gamma_2 - \lambda^2 \delta_{\mathbf{0}},$$

where Γ_2 is the Fourier transform of $\check{\Lambda}^{[2]}$. Thus the Bartlett spectrum may have a pole at the origin $\mathbf{0}$.

From Proposition A.2.4 follows that if the reduced factorial covariance measure $\check{C}^{[2]}$ is bounded, then the Bartlett spectrum is absolutely continuous with respect to ν_d, and its density f_Γ is given by

$$f_\Gamma(\boldsymbol{\xi}) = (2\pi)^{-d/2} \int_{\mathbb{R}^d} e^{-i\boldsymbol{\xi}\cdot\boldsymbol{x}} \, \check{C}^{[2]}(\mathrm{d}\boldsymbol{x}) \geq 0. \tag{2.24}$$

If $\check{C}^{[2]}$ is absolutely continuous with respect to ν_d, with continuous density $\mathrm{cov}(\boldsymbol{x})$, then cov is positive semi-definite function and by Bochner's theorem A.2.2 the Bartlett spectrum Γ is bounded and

$$\mathrm{cov}(\boldsymbol{x}) = (2\pi)^{-d/2} \int_{\mathbb{R}^d} e^{i\boldsymbol{\xi}\cdot\boldsymbol{x}} \, \Gamma(\mathrm{d}\boldsymbol{\xi}). \tag{2.25}$$

If moreover $\check{C}^{[2]}$ is bounded, that is $\mathrm{cov} \in L^1(\mathbb{R}^d, \nu_d)$, then

$$f_\Gamma(\boldsymbol{\xi}) = (2\pi)^{-d/2} \int_{\mathbb{R}^d} e^{-i\boldsymbol{\xi}\cdot\boldsymbol{x}} \, \mathrm{cov}(\boldsymbol{x}) \, \mathrm{d}\boldsymbol{x}. \tag{2.26}$$

From (2.25) further follows that $f_\Gamma \in L^1(\mathbb{R}^d, \nu_d)$, since $\Gamma(\mathbb{R}^d) = \mathrm{cov}(\mathbf{0})$, and the inversion formula (Theorem A.2.1) yields

$$\mathrm{cov}(\boldsymbol{x}) = (2\pi)^{-d/2} \int_{\mathbb{R}^d} e^{i\boldsymbol{\xi}\cdot\boldsymbol{x}} f_\Gamma(\boldsymbol{\xi}) \, \mathrm{d}\boldsymbol{\xi}.$$

2.5.6 Ergodic theorems

We have already discussed ergodic theorems for random closed sets in Subsection 2.2.3. Now we introduce a more general approach for random measures that particularly covers ergodic theorems for point processes and also for random closed sets. We restrict ourselves to $E = \mathbb{R}^d$. More general treatment can be found in [43] which is our main reference here.

Let ξ be a stationary random measure on $\mathbb{R}^d$ with distribution $\mathbb{P}_\xi$ and $S_{\boldsymbol{y}}$ be the action of the group of translations of $\mathbb{R}^d$ given by (2.17), i.e.

$$S_{\boldsymbol{x}}\xi(B) = \xi(B - \boldsymbol{x})$$

for all $\boldsymbol{x} \in \mathbb{R}^d$ and $B \in \mathcal{B}(\mathbb{R}^d)$. We denote by $\mathcal{I}$ the σ-algebra of events in $\mathcal{M}$, invariant under all translations, that is, of those sets $M \in \mathcal{M}$ for which

$$\mathbb{P}_\xi\left(M \setminus S_{\boldsymbol{x}}M \cup S_{\boldsymbol{x}}M \setminus M\right) = 0 \quad \text{for all } \boldsymbol{x} \in \mathbb{R}^d.$$

If $\mathcal{I}$ is trivial, that is, $\mathbb{P}_\xi(I) = 0$ or 1 for every $I \in \mathcal{I}$, we say that ξ is **metrically transitive**.

A **convex averaging sequence** $\{K_n, n \in \mathbb{N}\}$ is again a sequence of non-empty convex compact sets, non-decreasing in the sense of inclusion, $K_n \subset K_{n+1}$ for $n \in \mathbb{N}$, such that $r(K_n) \to \infty$ as $n \to \infty$, where $r(K_n)$ is given by (2.3).

Definition 2.5.10. A stationary random measure (point process) ξ on $\mathbb{R}^d$ is

(a) **mixing**, if for all $V, W \in \mathcal{M}$ ($V, W \in \mathcal{N}$ respectively)

$$\mathbb{P}_\xi(S_{\boldsymbol{x}}V \cap W) \to \mathbb{P}_\xi(V)\,\mathbb{P}_\xi(W) \quad \text{as} \quad \|\boldsymbol{x}\| \to \infty,$$

(b) **weakly mixing**, if for all such V, W,

$$\frac{1}{\nu_d(K_n)} \int_{K_n} \left|\, \mathbb{P}_\xi(S_{\boldsymbol{x}}V \cap W) - \mathbb{P}_\xi(V)\,\mathbb{P}_\xi(W)\right| \mathrm{d}\boldsymbol{x} \to 0 \quad \text{as} \quad n \to \infty,$$

(c) **ergodic**, if for all such V, W,

$$\frac{1}{\nu_d(K_n)} \int_{K_n} \mathbb{P}_\xi(S_{\boldsymbol{x}}V \cap W)\,\mathrm{d}\boldsymbol{x} \to \mathbb{P}_\xi(V)\,\mathbb{P}_\xi(W) \quad \text{as} \quad n \to \infty,$$

for any **convex averaging sequence** $\{K_n, n \in \mathbb{N}\}$.

It is easy to see that mixing implies weak mixing and this further implies ergodicity. It can also be shown that metrical transitivity is equivalent to ergodicity, [43, Proposition 12.3.III]. A direct consequence of Definition 2.5.8 of the volume measure ν_X of a random closed set X is that there is an equivalence in the ergodicity, weak mixing, and mixing, respectively, for X and for ν_X.

Proposition 2.5.9. *The stationary Poisson point process is mixing.*

Proof. Corollary 12.3.VII in [43]. □

The first ergodic theorem that we mention is based on classical ergodic theorems that were given in [57].

Theorem 2.5.6. *Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space, $\{S_{\boldsymbol{x}} | \boldsymbol{x} \in \mathbb{R}^d\}$ a group of measure preserving transformations acting measurably on $(\Omega, \mathcal{A}, \mathbb{P})$, $\{K_n, n \in \mathbb{N}\}$ a convex averaging sequence in $\mathbb{R}^d$, and $\mathcal{I}$ the σ-algebra of events in $\mathcal{A}$ that are invariant under the transformations $\{S_{\boldsymbol{x}}\}$. Let further f be a random variable on $(\Omega, \mathcal{A}, \mathbb{P})$.*

(a) If $\mathbb{E}\,|f| < \infty$ then

$$\frac{1}{\nu_d(K_n)} \int_{K_N} f(S_{\boldsymbol{x}}\omega)\,\mathrm{d}\boldsymbol{x} \to \mathbb{E}(f|\mathcal{I}) \quad \textit{as} \quad n \to \infty$$

almost surely.

(b) If $f \in L^p(\mathbb{P})$ and $p \geq 1$ then

$$\mathbb{E}\left|\frac{1}{\nu_d(K_n)}\int\limits_{K_N} f(S_{\boldsymbol{x}}\omega)\,\mathrm{d}\boldsymbol{x} - \mathbb{E}(f|\mathcal{I})\right|^p \to 0 \quad as \quad n\to\infty.$$

Proof. Theorem 6.1 and Theorem 6.2 in [57]. See also Proposition 12.2.II in [43]. □

For the definition and properties of the conditional expectation $\mathbb{E}(f|\mathcal{I})$ we refer the reader to Section 34 in [46]. The only thing we need is that if $\mathcal{I}$ is trivial, then $\mathbb{E}(f|\mathcal{I}) = \mathbb{E}(f)$ almost surely. The application of the previous theorem to random measures yields the following proposition.

Proposition 2.5.10. *Let ξ be an ergodic random measure (point process), f a measurable function on $(\mathsf{M},\mathcal{M})$ (on $(\mathsf{N},\mathcal{N})$ respectively) and $\{K_n\}$ a convex averaging sequence in $\mathbb{R}^d$. Then*

$$\frac{1}{\nu_d(K_n)}\int\limits_{K_n} f(S_{\boldsymbol{x}}\xi)\,\mathrm{d}\boldsymbol{x} \to \mathbb{E}\,f(\xi) \quad as \quad n\to\infty$$

almost surely if $\mathbb{E}\,|f(\xi)| < \infty$ and in L^p mean if $f \in L^p(\mathsf{M},\mathbb{P}_\xi)$ for $p \geq 1$.

Proof. Let us identify $(\Omega,\mathcal{A},\mathbb{P})$ with $(\mathsf{M},\mathcal{M},\mathbb{P}_\xi)$ or with $(\mathsf{M},\mathcal{M},\mathbb{P}_\xi)$ in the case of point process. The transformations $\{S_{\boldsymbol{x}}\}$ act measurably on M and since ξ is stationary, they are measure preserving. The assertion follows from the previous theorem and the fact, that $\mathcal{I}$ is trivial for ergodic ξ. □

The following important ergodic theorems deal with moment measures. Recall that $C_{\mathbf{0}}$ is the half-open unit cube given by (2.5).

Theorem 2.5.7. *Let ξ be an ergodic random measure on $\mathbb{R}^d$ and $\{K_n\}$ a convex averaging sequence in $\mathbb{R}^d$. Then*

$$\frac{\xi(K_n)}{\nu_d(K_n)} \to \mathbb{E}\,\xi(C_{\mathbf{0}}) \quad as \quad n\to\infty$$

a.s., in L^1 mean, and in a mean square if ξ is second order stationary.

Proof. Theorem 12.2.IV in [43]. □

Theorem 2.5.8. *Let ξ be a kth order stationary ergodic random measure on $\mathbb{R}^d$ with reduced kth order moment measure $\breve{\Lambda}^{(k)}$, and $B_1,\dots,B_{k-1}$ be bounded Borel sets in $\mathbb{R}^d$. Then for any convex averaging sequence $\{K_n\}$ in $\mathbb{R}^d$, as $n\to\infty$,*

$$\frac{1}{\nu_d(K_n)}\int\limits_{K_n} \xi(\boldsymbol{x}+B_1)\dots\xi(\boldsymbol{x}+B_{k-1})\xi(\,\mathrm{d}\boldsymbol{x}) \to \breve{\Lambda}^{(k)}(B_1\times\dots\times B_{k-1}) \quad a.s..$$

Proof. Theorem 12.6.VI in [43]. □

The previous theorem holds also for factorial moment measures.

Corollary 2.5.3. *Let ξ be an ergodic second order stationary random measure on $\mathbb{R}^d$. Then for any bounded Borel set B and any convex averaging sequence $\{K_n\}$, as $n\to\infty$,*

$$\frac{\breve{C}^{(2)}(K_n)}{\nu_d(K_n)} \to 0,$$

$$\frac{1}{\nu_d(K_n)}\int\limits_{K_n} \tilde{\xi}(\boldsymbol{x}+B)\tilde{\xi}(\,\mathrm{d}\boldsymbol{x}) \to \breve{C}^{(2)}(B) \qquad a.s.,$$

where $\tilde{\xi}$ is a mean-corrected random signed measure given by $\tilde{\xi} = \xi - \Lambda$.

Proof. Corollary 8.1.III in [56]. □

2.6 Marked point processes

Important extension of point processes is given by marked point processes. They are characterised by the presence of an additional information at every point of the underlying point process. In the following we present only the basic definition and refer the reader to [40, 56] for more details.

By M we denote a locally compact, Hausdorff topological space with a countable base and we assume that $\mathbb{R}^d \times M$ is equipped with the product topology.

Definition 2.6.1. A **marked point process** in $\mathbb{R}^d$ with mark space M is a simple point process Y in $\mathbb{R}^d \times M$ with intensity measure Θ satisfying

$$\Theta(C \times M) < \infty \quad \text{for all} \quad C \in \mathcal{C}.$$

If Y is a marked point process in $\mathbb{R}^d$, its image under the projection $(\boldsymbol{y}, m) \mapsto \boldsymbol{y}$ is an ordinary point process Y^0 in $\mathbb{R}^d$ that is called the **ground process** of Y.

In the following we will always assume that the ground process of a marked point process is simple, which may not be true in general.

On $\mathbb{R}^d \times M$ we define the action of the group of translations of $\mathbb{R}^d$ by setting $S_{\boldsymbol{y}}(\boldsymbol{x}, m) = (S_{\boldsymbol{y}}\boldsymbol{x}, m) = (\boldsymbol{x} + \boldsymbol{y}, m)$ for all $\boldsymbol{y} \in \mathbb{R}^d$ and all $(\boldsymbol{x}, m) \in \mathbb{R}^d \times M$. Thus $S_{\boldsymbol{y}}$ is continuous on $\mathbb{R}^d \times M$ and hence measurable. For a marked point process Y we write $S_{\boldsymbol{y}}Y \equiv Y + \boldsymbol{y}$, where $S_{\boldsymbol{y}}Y$ is defined by (2.17). Clearly, $S_{\boldsymbol{y}}Y^0 = Y^0 + \boldsymbol{y}$.

A marked point process Y is stationary if $Y \overset{\mathcal{D}}{\sim} Y + \boldsymbol{x}$ for all $\boldsymbol{x} \in \mathbb{R}^d$. It is clear that the ground process Y^0 of a stationary marked point process Y is stationary.

Theorem 2.6.1. *If Y is a stationary marked point process in $\mathbb{R}^d$ with mark space M and intensity measure $\Theta \neq 0$, then*

$$\Theta = \lambda \nu_d \otimes \mathbb{Q}$$

with $0 < \lambda < \infty$, which is the intensity of the ground process Y^0, and a (uniquely determined) probability measure $\mathbb{Q}$ on M.

Proof. Theorem 3.5.1 in [40]. □

Important subclass of marked point processes is given by independently marked point processes. If Y is a marked point process then Lemma 2.5.2 yields

$$Y = \sum_{i=1}^{Y(\mathbb{R}^d \times M)} \delta_{(\boldsymbol{y}_i, m_i)}, \tag{2.27}$$

where $\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots$ are random variables in $\mathbb{R}^d$ and $m_1, m_2, \ldots$ are random variables in M.

Definition 2.6.2. The marked point process Y is**independently marked** if the marks $m_1, m_2, \ldots$ are independently, identically distributed, and are independent of locations $\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots$, where $\{(m_i, \boldsymbol{y}_i)\}$ corresponds to Y according to the previous representation. The distribution $\mathbb{Q}$ of m_i is called the **mark distribution** of Y.

See [56, Definition 6.4.III] for a more general definition of independent marking that includes also the dependence of the mark distribution on the location.

Theorem 2.6.2. *Let Y be an independently marked point process in $\mathbb{R}^d$ with intensity measure Θ and mark distribution $\mathbb{Q}$. Then*

$$\Theta = \Lambda \otimes \mathbb{Q},$$

where Λ is the intensity measure of the ground process Y^0.

Proof. Theorem 3.5.6 in [40]. □

The last statement deals with the case where the ground process is a Poisson point process.

Theorem 2.6.3. *Let Y be an independently marked point process in $\mathbb{R}^d$ such that its ground process Y^0 is a Poisson point process. Then Y is a Poisson point process.*

Proof. Theorem 3.5.7 in [40]. □

2.7 Particle processes

In this section we introduce practically important class of models of stochastic geometry in $\mathbb{R}^d$. We follow Chapter 4 in [40]. As a consequence of Theorem A.4.2, the system $\mathcal{F}' = \mathcal{F}'(\mathbb{R}^d)$ of non-empty closed subsets of $\mathbb{R}^d$ equipped with the Fell topology is a locally compact, Hausdorff topological space with a countable base. Hence we may consider a point processes on $\mathcal{F}'$. The subset $\mathcal{C}'$ of $\mathcal{F}'$ given by non-empty compact sets is by Corollary A.5.2 a Borel subset of $\mathcal{F}'$. We always assume that $\mathcal{C}'$ is equipped with topology of the Hausdorff metric introduced in Appendix A.5 and with the Borel σ-algebra generated by this topology. By Corollary A.5.2, every Borel set in $\mathcal{C}'$ is a Borel set in $\mathcal{F}'$.

Definition 2.7.1. Let $\mathcal{W}$ be a Borel subset of $\mathcal{C}'$. A simple point process Y in $\mathcal{F}'$ concentrated on $\mathcal{W}$, i.e. $\mathbb{P}\left(Y(\mathcal{F}' \setminus \mathcal{W}) > 0\right) = 0$, for which the intensity measure exists is called a **particle process in** $\mathcal{W}$. If $\mathcal{W} = \mathcal{C}'$ we call Y a **particle process** in $\mathbb{R}^d$.

If Θ is the intensity measure of a particle process Y in $\mathcal{W}$, it follows $\Theta(\mathcal{F}' \setminus \mathcal{W}) = 0$. By Definition 2.5.3 the intensity measure Θ exists if it is locally finite. It can be shown (see [40, Lemma 2.3.1]) that this is equivalent to

$$\Theta(\mathcal{F}_C) < \infty \quad \text{for all} \quad C \in \mathcal{C}'. \tag{2.28}$$

The action of the group of translations of $\mathbb{R}^d$ on $\mathsf{N}_s(\mathcal{F}')$ is induced in a usual canonical way (2.17) from the action (2.1) on $\mathcal{F}'$. Hence

$$S_{\boldsymbol{y}} Y(F) = Y(F - \boldsymbol{y}) \quad \text{for all} \quad \boldsymbol{y} \in \mathbb{R}^d,\ F \in \mathcal{B}(\mathcal{F}),$$

where $F - \boldsymbol{y} = \{G - \boldsymbol{y} | G \in F\}$.

From now on let assume that $\mathcal{W}$ is invariant under translations, i.e. $C + \boldsymbol{x} \in \mathcal{W}$ for all $C \in \mathcal{W}$ and all $\boldsymbol{x} \in \mathbb{R}^d$. For a stationary particle process Y in $\mathcal{W}$ it is possible to find a useful decomposition of the intensity measure Λ. To do this we need a Borel measurable mapping $\boldsymbol{z} : \mathcal{W} \to \mathbb{R}^d$ called a **centre function on** $\mathcal{W}$ that satisfies $\boldsymbol{z}(C + \boldsymbol{x}) = \boldsymbol{z}(C) + \boldsymbol{x}$ for every $C \in \mathcal{W}$ and every $\boldsymbol{x} \in \mathbb{R}^d$. Possible choices on $\mathcal{C}'$ are the centre of the smallest ball containing C (circumcentre of C), or the Steiner point of the convex hull of C, see [40]. In Chapter 3 we develop a new centre function motivated by the centroid (centre of mass) on certain Borel subsets of $\mathcal{C}'$.

For a centre function $\boldsymbol{z}$ on $\mathcal{W}$ we define

$$\mathcal{W}_0 = \{C \in \mathcal{C}' | \boldsymbol{z}(C) = \boldsymbol{0}\} \tag{2.29}$$

and call it the **grain space**. The measurability of $\boldsymbol{z}$ yields that $\mathcal{W}_0$ is a Borel set in $\mathcal{C}'$ and hence in $\mathcal{F}'$. Further we define the mapping $\varphi : \mathbb{R}^d \times \mathcal{W}_0 \to \mathcal{W}$ by

$$\varphi(\boldsymbol{x}, C) = C + \boldsymbol{x}, \tag{2.30}$$

which is clearly a bijection.

Theorem 2.7.1. *Let Y be a stationary particle process in $\mathcal{W}$ with intensity measure $\Theta \neq 0$. Then there exist a number $\lambda \in (0, +\infty)$ and a probability measure $\mathbb{Q}$ on $\mathcal{W}_0$ such that*

$$\Theta = \lambda\varphi(\nu_d \otimes \mathbb{Q}).$$

The number λ and the measure $\mathbb{Q}$ are uniquely determined.

Proof. Analogous to the proof of Theorem 4.1.1 in [40]. □

The number λ is called the **intensity** and the measure $\mathbb{Q}$ the **grain distribution** of the stationary particle process Y in $\mathcal{W}$. A random closed set concentrated on $\mathcal{W}_0$ with distribution $\mathbb{Q}$ is called the **typical grain** of Y. If Y is isotropic, $\mathcal{W}$ is invariant under rotations, and $\boldsymbol{z}$ is compatible with rotations, then $\mathbb{Q}$ is rotation invariant. The compatibility of $\boldsymbol{z}$ with rotations means that $\boldsymbol{z}(\theta C) = \theta\boldsymbol{z}(C)$ for all $C \in \mathcal{W}$ and all rotations $\theta \in SO_d$. The condition (2.28) gives a restriction on the grain distribution of a stationary particle process.

Theorem 2.7.2. *The probability measure $\mathbb{Q}$ on $\mathcal{W}_0$ is the grain distribution of some stationary particle process if an only if*

$$\int_{\mathcal{W}_0} \nu_d(C + rB^d)\,\mathbb{Q}(\mathrm{d}C) < \infty \quad \text{for some} \quad r > 0.$$

If $\mathbb{Q}$ satisfies the previous relation and if $\lambda > 0$ is given, then there exists (up to equivalence) precisely one Poisson particle process Y in $\mathcal{W}$ with intensity λ and grain distribution $\mathbb{Q}$. Moreover if $\mathcal{W}$ is invariant under rotations and $\boldsymbol{z}$ is compatible with rotations, then Y is isotropic if and only if $\mathbb{Q}$ is rotation invariant.

Proof. Theorem 4.1.2 in [40]. □

If Y is a particle process in $\mathcal{W}$ and $\boldsymbol{z}$ is a centre function on $\mathcal{W}$, then

$$Y^0 = \sum_{C \in Y} \delta_{\boldsymbol{z}(C)} \tag{2.31}$$

is a random counting measure on $\mathbb{R}^d$. In general Y^0 is not a point process in $\mathbb{R}^d$ since it may not be locally finite. In the stationary case the situation is simple and we may state the following connection to marked point processes.

Theorem 2.7.3. *Let Y be a stationary particle process in $\mathcal{W}$, and let $\boldsymbol{z}$ be a centre function on $\mathcal{W}$. Then Y^0 is a stationary point process in $\mathbb{R}^d$, and*

$$\tilde{Y} = \sum_{C \in Y} \delta_{\left(\boldsymbol{z}(C), C - \boldsymbol{z}(C)\right)} \tag{2.32}$$

is a stationary marked point process with mark space $\mathcal{W}_0$. The intensities of Y, Y^0, and $\tilde{Y}$ are the same. The mark distribution of $\tilde{Y}$ is the grain distribution $\mathbb{Q}$ of Y.

Proof. Theorem 4.2.1 in [40]. □

For Poisson particle processes we have the following assertion.

Theorem 2.7.4. *Let Y be a stationary Poisson particle process in $\mathcal{W}$, and let $\boldsymbol{z}$ be a centre function on $\mathcal{W}$. Then Y^0 is a stationary Poisson process and $\tilde{Y}$ is an independently marked stationary Poisson process with mark space $\mathcal{W}_0$.*

Proof. Theorems 4.1.4 and 4.2.2 in [40]. □

Let us now focus on the non-stationary case. It can be shown (see [40, Section 11.1]) that the locally finite intensity measure Θ of a particle process Y in $\mathcal{W}$ with centre function $\boldsymbol{z}$ has a decomposition

$$\Theta(A) = \int_{\mathcal{W}_0} \int_{\mathbb{R}^d} \mathbb{1}_A(C + \boldsymbol{x}) \rho(C, \mathrm{d}\boldsymbol{x}) \, \mathbb{Q}(\mathrm{d}C) \tag{2.33}$$

for all $A \in \mathcal{B}(\mathcal{C}')$, where $\mathbb{Q}$ is a probability measure on $\mathcal{W}_0$ and ρ is a function that is Borel measurable in the first variable and a locally finite measure in the second variable. Following [40], it is sometimes possible to assume that $\rho(C, \cdot)$ is for each $C \in \mathcal{W}_0$ absolutely continuous with respect to ν_d, with density $\eta(C, \boldsymbol{x})$. Then we obtain

$$\Theta(A) = \int_{\mathcal{W}_0} \int_{\mathbb{R}^d} \mathbb{1}_A(C + \boldsymbol{x}) \eta(C, \boldsymbol{x}) \, \mathrm{d}\boldsymbol{x} \, \mathbb{Q}(\mathrm{d}C) \tag{2.34}$$

for all $A \in \mathcal{B}(\mathcal{C}')$. The local finiteness of Θ implies that η is locally integrable on $\mathcal{W}_0 \times \mathbb{R}^d$ with respect to the product measure $\mathbb{Q} \otimes \nu_d$. The following proposition deals with the general case when ρ is a locally finite measure independent of the first argument.

Proposition 2.7.1. *If in the decomposition (2.33) of the intensity measure Θ of a particle process Y holds $\rho(C, \cdot) = \rho(\cdot)$ for all $C \in \mathcal{W}_0$, then*

$$\Theta = \varphi(\Lambda \otimes \mathbb{Q}),$$

where $\Lambda = \rho$ is a locally finite measure on $\mathbb{R}^d$ and $\mathbb{Q}$ is a probability measure on $\mathcal{W}_0$. The measures Λ and $\mathbb{Q}$ are uniquely determined.

Moreover Y^0 given by (2.31) is a point process in $\mathbb{R}^d$ with intensity η, and $\tilde{Y}$ given by (2.32) is a marked point process in $\mathbb{R}^d$ with mark space $\mathcal{W}_0$, ground process Y^0 and intensity measure $\tilde{\Theta}$ given by

$$\tilde{\Theta} = \Lambda \otimes \mathbb{Q}\,.$$

Proof. Let φ be defined by (2.30) and $\varphi^{-1}(\Theta)$ be the image measure of Θ on $\mathbb{R}^d \times \mathcal{W}_0$. For every Borel $A \in \mathbb{R}^d$ and every Borel $B \in \mathcal{W}_0$ we have

$$\begin{aligned}\varphi^{-1}(\Theta)(A \times B) = \Theta(\varphi(A \times B)) &= \int_{\mathcal{W}_0} \int_{\mathbb{R}^d} \mathbb{1}_{\varphi(A \times B)}(C + \boldsymbol{x}) \rho(\mathrm{d}\boldsymbol{x}) \, \mathbb{Q}(\mathrm{d}C) \\ &= \int_{\mathcal{W}_0} \int_{\mathbb{R}^d} \mathbb{1}_A(\boldsymbol{x}) \mathbb{1}_B(C) \rho(\mathrm{d}\boldsymbol{x}) \, \mathbb{Q}(\mathrm{d}C) = \int_A \rho(\mathrm{d}\boldsymbol{x}) \cdot \int_B \mathbb{Q}(\mathrm{d}C) = (\Lambda \otimes \mathbb{Q})(A \times B).\end{aligned}$$

The uniqueness is trivial. For the proof of the second part we show that $\tilde{Y}$ is a marked point process with desired properties. The result for Y^0 then follows. We proceed similarly to the proof of Theorem 4.2.1 in [40]. First let us show the measurability of $\tilde{Y}$. By Lemma 2.5.3 it is enough to show that $\{\tilde{Y}(A) = k\}$ is measurable for all Borel $A \in \mathbb{R}^d \times \mathcal{W}_0$ and all $k \in \mathbb{N}_0$. From the measurability of the centre function $\boldsymbol{z}$ follows the measurability of the mapping $\vartheta : \mathcal{W} \to \mathbb{R}^d \times \mathcal{W}_0$ defined by $\vartheta(C) = \big(\boldsymbol{z}(C), C - \boldsymbol{z}(C)\big)$ for all $C \in \mathcal{W}$. Since, $\{\tilde{Y}(A) = k\} = \{Y(\vartheta^{-1}(A)) = k\}$ the measurability of $\tilde{Y}$ follows. Let A be a Borel set in $\mathbb{R}^d$ and B be a Borel set in $\mathcal{W}_0$. Since ϑ is the inverse of previously defined φ we obtain

$$\tilde{\Theta}(A \times B) = \mathbb{E}\,\tilde{Y}(A \times B) = \mathbb{E}\,Y(\vartheta^{-1}(A \times B)) = \Theta(\varphi(A \times B)) = (\Lambda \otimes \mathbb{Q})(A \times B).$$

Hence $\tilde{\Theta} = \Lambda \otimes \mathbb{Q}$. From local finiteness of Λ and finiteness of $\mathbb{Q}$ follows that $\mathbb{E}\,\tilde{Y}(C \times \mathcal{W}_0) < \infty$ for every compact $C \in R^d$ and hence $\tilde{Y}$ is a.s. locally finite and therefore, it is a marked point process in $\mathbb{R}^d$. □

Under the assumptions of the previous theorem we analogously to the stationary case call Λ the **intensity measure** and $\mathbb{Q}$ the **grain distribution** of Y. A random closed set concentrated on $\mathcal{W}_0$ with distribution $\mathbb{Q}$ is called the **typical grain** of Y. If ρ is absolutely continuous with respect to ν_d, with density η, then clearly Θ allows the decomposition (2.34) with $\eta(C, \boldsymbol{x}) = \eta(\boldsymbol{x})$ for all $C \in \mathcal{W}_0$ and thus $\Lambda = \eta\nu_d$. In such a case we call η the **intensity** of Y.

Now let us consider the opposite situation with given marked point process $\tilde{Y}$ in $\mathbb{R}^d$ with mark space $\mathcal{W}_0$ being an arbitrary Borel subset of $\mathcal{C}'$.

Definition 2.7.2. A marked point process $\tilde{Y}$ in $\mathbb{R}^d$ with mark space $\mathcal{W}_0 \in \mathcal{B}(\mathcal{C}')$ is called a **germ-grain process** whenever a measure Y on $\mathcal{F}'$ defined by

$$Y = \sum_{(\boldsymbol{x},C)\in\tilde{Y}} \delta_{C+\boldsymbol{x}}$$

has a locally finite intensity measure. In the positive case Y defines a particle process called the **particle process generated by** $\tilde{Y}$.

If $\tilde{\Theta}$ is the intensity measure of $\tilde{Y}$ and Θ the intensity measure of Y, then clearly

$$\Theta = \varphi(\tilde{\Theta}).$$

If a germ-grain process $\tilde{Y}$ is stationary then Y is also stationary and by Theorems 2.6.1 and 2.7.3 the mark distribution $\mathbb{Q}$ is the same as the grain distribution of Y.

If $\tilde{Y}$ is an independently marked point process in $\mathbb{R}^d$ with mark space $\mathcal{W}_0$ and intensity measure $\tilde{\Theta}$, then Theorem 2.6.2 yields a decomposition

$$\tilde{\Theta} = \Lambda \otimes \mathbb{Q}$$

and thus

$$\Theta = \varphi(\Lambda \otimes \mathbb{Q}). \tag{2.35}$$

In that case we call $\tilde{Y}$ the **independent germ-grain process**. It is obvious that the conditions of Proposition 2.7.1 are satisfied for a particle process Y generated by $\tilde{Y}$. As a consequence we may use the same terminology for $\tilde{Y}$ and Y. Thus in particular, a random closed set concentrated on $\mathcal{W}_0$ with distribution $\mathbb{Q}$ is called the **typical grain** of $\tilde{Y}$.

2.7.1 Germ-grain models, Boolean models

First we mention the union set of a point process. For a point process Y in $\mathcal{F}'$ we set

$$X_Y = \bigcup_{F\in\operatorname{supp} Y(\omega)} F \quad \text{for all} \quad \omega \in \Omega.$$

Theorem 2.7.5. *The union set X_Y of a point process Y in $\mathcal{F}'$ is a random closed set. If Y is stationary (isotropic), then X_Y is stationary (isotropic).*

Proof. Theorem 3.6.2 in [40]. □

Now let us construct the union set of a germ-grain process. Let $\tilde{Y}$ be a germ-grain process and Y the particle process generated by $\tilde{Y}$. We define the union set of $\tilde{Y}$ by

$$X_{\tilde{Y}} = \bigcup_{(\boldsymbol{x},C)\in\tilde{Y}} C + \boldsymbol{x},$$

i.e. as the union set of the particle process Y. From the previous theorem follows that $X_{\tilde{Y}}$ is a random closed set.

Definition 2.7.3. A random closed set $X_{\tilde{Y}}$ of an independent germ-grain process $\tilde{Y}$ with mark space $\mathcal{W}_0 \in \mathcal{B}(\mathcal{C}')$ is called a **germ-grain model in** $\mathcal{W}_0$. A **Boolean model** is a germ-grain model for which the ground process of a corresponding germ-grain process is a Poisson process.

A Boolean model $X_{\tilde{Y}}$ is determined by the intensity measure Λ of the ground Poisson process $\tilde{Y}^0$ and by the distribution $\mathbb{Q}$ of the typical grain.

Theorem 2.7.6. *The stationary Boolean models are precisely the union sets of stationary Poisson particle processes.*

Proof. Theorem 4.3.2 in [40]. □

For a later statistical issues the next assertion is important.

Theorem 2.7.7. *A stationary Boolean model is mixing.*

Proof. Theorem 9.3.5 in [40]. □

Finally we present some relations for the most important characteristics of Boolean models. We start with a capacity functional.

Proposition 2.7.2. *Let X be a Boolean model and Θ the intensity measure of the corresponding particle process. Then*

$$T_X(C) = 1 - e^{-\Theta(\mathcal{F}_C)}$$

for each $C \in \mathcal{C}$.

Proof. Theorem 3.6.3 in [40]. □

Now we may easily derive relations for the volume fraction and the covariance. Let X be a Boolean model in $\mathcal{W}_0$ with intensity measure Θ, typical grain X_0, grain distribution $\mathbb{Q}$ and ground Poisson process X^0. Let Λ be the intensity measure of X^0. The decomposition (2.35) yields

$$\begin{aligned}
\Theta(\mathcal{F}_C) &= \int \mathbb{1}_{\mathcal{F}_C}(D + \boldsymbol{x})\ \Lambda(\mathrm{d}\boldsymbol{x})\, \mathbb{Q}(\mathrm{d}D) = \int \mathbb{1}(D + \boldsymbol{x} \cap C \neq \emptyset)\ \Lambda(\mathrm{d}\boldsymbol{x})\, \mathbb{Q}(\mathrm{d}D) \\
&= \int \mathbb{1}(\boldsymbol{x} \in C + \check{D})\ \Lambda(\mathrm{d}\boldsymbol{x})\, \mathbb{Q}(\mathrm{d}D) = \int \Lambda(C + \check{D})\ \ \mathbb{Q}(\mathrm{d}D) \\
&= \mathbb{E}\,\Lambda(C + \check{X}_0),
\end{aligned}$$

where $C + \check{D} = \{\boldsymbol{y} - \boldsymbol{z} | \boldsymbol{y} \in C, \boldsymbol{z} \in D\}$ is the combination of Minkowski addition and reflection. Hence

$$T_X(C) = 1 - e^{-\mathbb{E}\,\Lambda(C + \check{X}_0)}. \tag{2.36}$$

Taking $C = \{\boldsymbol{x}\}$ leads to

$$m(\boldsymbol{x}) = T_X\big(\{\boldsymbol{x}\}\big) = 1 - e^{-\mathbb{E}\,\Lambda(\check{X}_0 + \boldsymbol{x})}.$$

If X is stationary, then

$$p = T_X\big(\{\boldsymbol{0}\}\big) = 1 - e^{-\mathbb{E}\,\Lambda(\check{X}_0)} = 1 - e^{-\lambda\,\mathbb{E}\,\nu_d(\check{X}_0)}. \tag{2.37}$$

For the covariance we take $C = \{\boldsymbol{x}, \boldsymbol{y}\}$ which yields

$$\begin{aligned}
C(\boldsymbol{x}, \boldsymbol{y}) &= \mathbb{P}(\boldsymbol{x} \in X, \boldsymbol{y} \in X) = \mathbb{P}(\boldsymbol{x} \in X) + \mathbb{P}(\boldsymbol{y} \in X) - \mathbb{P}(\boldsymbol{x} \in X \cup \boldsymbol{y} \in X) \\
&= m(\boldsymbol{x}) + m(\boldsymbol{y}) - T_X\big(\{\boldsymbol{x}, \boldsymbol{y}\}\big) = m(\boldsymbol{x}) + m(\boldsymbol{y}) - 1 + e^{-\mathbb{E}\,\Lambda(\check{X}_0 + \boldsymbol{x} \cup \check{X}_0 + \boldsymbol{y})} \\
&= m(\boldsymbol{x}) + m(\boldsymbol{y}) - 1 + e^{-\mathbb{E}\,\Lambda(\check{X}_0 + \boldsymbol{x})} e^{-\mathbb{E}\,\Lambda(\check{X}_0 + \boldsymbol{y})} e^{\mathbb{E}\,\Lambda(\check{X}_0 + \boldsymbol{x} \cap \check{X}_0 + \boldsymbol{y})} \\
&= m(\boldsymbol{x}) + m(\boldsymbol{y}) - 1 + \big(1 - m(\boldsymbol{x})\big)\big(1 - m(\boldsymbol{y})\big) e^{\mathbb{E}\,\Lambda(\check{X}_0 + \boldsymbol{x} \cap \check{X}_0 + \boldsymbol{y})}.
\end{aligned}$$

The covariance function is

$$\operatorname{cov}(\boldsymbol{x},\boldsymbol{y}) = C(\boldsymbol{x},\boldsymbol{y}) - m(\boldsymbol{x})m(\boldsymbol{y}) = \big(1-m(\boldsymbol{x})\big)\big(1-m(\boldsymbol{y})\big)\big(e^{\mathbb{E}\,\Lambda(\check{X}_0+\boldsymbol{x}\cap\check{X}_0+\boldsymbol{y})}-1\big)$$

and the correlation function

$$\kappa(\boldsymbol{x},\boldsymbol{y}) = \sqrt{\frac{\big(1-m(\boldsymbol{x})\big)\big(1-m(\boldsymbol{y})\big)}{m(\boldsymbol{x})m(\boldsymbol{y})}}\big(e^{\mathbb{E}\,\Lambda(\check{X}_0+\boldsymbol{x}\cap\check{X}_0+\boldsymbol{y})}-1\big).$$

In the stationary case we obtain

$$\begin{aligned} C(\boldsymbol{r}) &= 2p-1+(1-p)^2 e^{\lambda\,\mathbb{E}\,\gamma_{X_0}(\boldsymbol{r})},\\ \operatorname{cov}(\boldsymbol{r}) &= (1-p)^2(e^{\lambda\,\mathbb{E}\,\gamma_{X_0}(\boldsymbol{r})}-1),\\ \kappa(\boldsymbol{r}) &= \frac{1-p}{p}(e^{\lambda\,\mathbb{E}\,\gamma_{X_0}(\boldsymbol{r})}-1), \end{aligned} \tag{2.38}$$

where $\gamma_{X_0}(\boldsymbol{r}) = \nu_d\big(X_0\cap(X_0-\boldsymbol{r})\big)$ is the set covariance of X_0.

2.8 Long-range dependence

The long-range dependence for one dimensional discrete stochastic processes is a well developed theory, see [58] and references therein. Here we present a natural extension to random measures in $\mathbb{R}^d$ that was indicated in [43]. The concept of long-range dependence is introduced through a slow decay of correlations.

Definition 2.8.1. A second order stationary random measure ξ is **long-range dependent** if

$$\limsup_{a\to\infty}\frac{\operatorname{var}\xi\big([-a,a]^d\big)}{(2a)^d} = \infty,$$

where $[-a,a]^d \equiv [-a,a]\times\ldots\times[-a,a]$ is the d-dimensional closed interval in $\mathbb{R}^d$. A random closed set X is **long-range dependent** if its volume measure ν_X is long-range dependent.

The long-range dependence is connected to the behaviour of a reduced second order factorial covariance measure. By Proposition 2.5.8 (b) and (d) with $f = \mathbb{1}_{[-a,a]^d}$,

$$\operatorname{var}\xi\big([-a,a]^d\big) = \lambda_D\big(\mathbb{1}_{[-a,a]^d} * \mathbb{1}^*_{[-a,a]^d}\big)(\boldsymbol{0}) + \int_{\mathbb{R}^d}\big(\mathbb{1}_{[-a,a]^d} * \mathbb{1}^*_{[-a,a]^d}\big)(\boldsymbol{x})\;\breve{C}^{[2]}(\mathrm{d}\boldsymbol{x}),$$

where λ_D is the constant from Proposition 2.5.4. Since

$$\big(\mathbb{1}_{[-a,a]^d} * \mathbb{1}^*_{[-a,a]^d}\big)(\boldsymbol{x}) = \prod_{i=1}^d \max\{2a-|x_i|,0\}$$

for all $\boldsymbol{x} = (x_1,\ldots,x_d)\in\mathbb{R}^d$, we obtain

$$\frac{\operatorname{var}\xi\big([-a,a]^d\big)}{(2a)^d} = \lambda_D + \int_{\mathbb{R}^d}\frac{\prod_{i=1}^d\max\{2a-|x_i|,0\}}{(2a)^d}\;\breve{C}^{[2]}(\mathrm{d}\boldsymbol{x}). \tag{2.39}$$

The reduced second order factorial covariance measure $\breve{C}^{[2]}$ of second order stationary random measure ξ is a signed measure in sense of Appendix A.2. Thus in general we cannot treat it as a usual Borel measure and try to take $\breve{C}^{[2]}(\mathbb{R}^d)$, which may be undefined. The (Jordan-Hahn) decomposition may, however, be helpful. Note that lower and upper variations of a signed measure are ordinary Borel σ-finite measures.

Proposition 2.8.1. *Let ξ be a second order stationary random measure and $C^{[2]}$ its reduced second order factorial covariance measure. Let further $C^{[2]+}$ and $C^{[2]-}$ be the upper and lower variations, respectively, of $C^{[2]}$. If $C^{[2]-}$ is finite, then ξ is long-range dependent if and only if $C^{[2]+}$ is finite.*

Proof. Let us denote the integrand in equation (2.39) by $f_a(\boldsymbol{x})$. Clearly it converges pointwise monotonously to 1 for all $\boldsymbol{x} \in \mathbb{R}^d$ as $a \to \infty$. From the dominated convergence theorem follows

$$\int_{\mathbb{R}^d} f_a(\boldsymbol{x})\, \check{C}^{[2]+}(\mathrm{d}\boldsymbol{x}) \to C^{[2]+}(\mathbb{R}^d) \quad \text{and} \quad \int_{\mathbb{R}^d} f_a(\boldsymbol{x})\, \check{C}^{[2]-}(\mathrm{d}\boldsymbol{x}) \to C^{[2]-}(\mathbb{R}^d) < \infty$$

as $a \to \infty$. Hence

$$\lim_{a\to\infty} \frac{\operatorname{var} \xi\big([-a,a]^d\big)}{(2a)^d} = \lambda_D + C^{[2]+}(\mathbb{R}^d) + C^{[2]-}(\mathbb{R}^d)$$

and the statement follows. □

There is a strong connection of the long-range dependence to the Bartlett spectrum Γ of ξ. First we mention a simple consequence of the previous proposition.

Corollary 2.8.1. *If $\check{C}^{[2]}$ is bounded, then ξ is not long-range dependent and*

$$\frac{\operatorname{var} \xi\big([-a,a]^d\big)}{(2a)^d} \to \lambda_D + (2\pi)^{d/2} f_\Gamma(\mathbf{0}) \quad \textit{as} \quad a \to \infty,$$

where f_Γ is the density of the Bartlett spectrum Γ of ξ with respect to ν_d, and

$$f_\Gamma(\mathbf{0}) = (2\pi)^{-d/2} \check{C}^{[2]}\big(\mathbb{R}^d\big).$$

Proof. By the proof of previous proposition,

$$\frac{\operatorname{var} \xi\big([-a,a]^d\big)}{(2a)^d} \to \lambda_D + C^{[2]}(\mathbb{R}^d) < \infty \quad \text{as} \quad a \to \infty.$$

The boundedness of $\check{C}^{[2]}$, by Proposition A.2.4, implies the absolute continuity of Γ with respect to the Lebesgue measure ν_d and validity of (2.24) which completes the proof. □

The existence of atom of Γ at the origin implies the long-range dependence.

Proposition 2.8.2. *If $\Gamma\big(\{\mathbf{0}\}\big) > 0$, then ξ is long-range dependent.*

Proof. Let $f_a = \mathbb{1}_{[-a,a]^d}$ for each $a > 0$. Then $\xi\big([-a,a]^d\big) = \int_{\mathbb{R}^d} f_a(\boldsymbol{x})\, \xi(\mathrm{d}\boldsymbol{x})$ and by Proposition 2.5.8 (d),

$$\frac{\operatorname{var} \xi\big([-a,a]^d\big)}{(2a)^d} \geq (2\pi)^{d/2} \int_{\mathbb{R}^d} \frac{|\check{f}_a(\boldsymbol{\omega})|^2}{(2a)^d}\, \Gamma(\mathrm{d}\boldsymbol{\omega}).$$

The integrand is (see the proof of Proposition 2.5.8 (e))

$$\frac{|\check{f}_a(\boldsymbol{\omega})|^2}{(2a)^d} = \prod_{j=1}^d \frac{\sin^2(a\omega_j)}{\pi a \omega_j^2}.$$

Since Γ is a positive measure, we obtain

$$\int_{\mathbb{R}^d} \frac{|\check{f}_a(\boldsymbol{\omega})|^2}{(2a)^d}\, \Gamma(\mathrm{d}\boldsymbol{\omega}) = \int_{\mathbb{R}^d\setminus\{\mathbf{0}\}} \frac{|\check{f}_a(\boldsymbol{\omega})|^2}{(2a)^d}\, \Gamma(\mathrm{d}\boldsymbol{\omega}) + \frac{a^d}{\pi^d}\Gamma\big(\{\mathbf{0}\}\big) \geq \frac{a^d}{\pi^d}\Gamma\big(\{\mathbf{0}\}\big).$$

Taking the limit $a \to \infty$ yields the result. □

The case of a positive atom at the origin cannot happen for ergodic random measures as stated by the following theorem.

Proposition 2.8.3. *Let ξ be an ergodic second order stationary random measure. Then $\Gamma(\{\mathbf{0}\}) = 0$.*

Proof. Proposition 2.5.8 (c) yields

$$\Gamma(\{\mathbf{0}\}) = (2\pi)^{d/2} \lim_{a\to\infty} \frac{\breve{C}^{[2]}([-a,a]^d)}{\nu_d([-a,a]^d)}.$$

If ξ is ergodic then from Corollary 2.5.3

$$\lim_{a\to\infty} \frac{\breve{C}^{(2)}([-a,a]^d)}{\nu_d([-a,a]^d)} = 0.$$

From Proposition 2.5.4 (see also the proof of Proposition 2.5.8 (d)) follows $\breve{C}^{(2)} = \lambda_D \delta_{\mathbf{0}} + \breve{C}^{[2]}$, and hence $\breve{C}^{[2]}([-a,a]^d) \le \breve{C}^{(2)}([-a,a]^d)$. Therefore $\Gamma(\{\mathbf{0}\}) = 0$. □

Proposition 2.8.4. *Let Γ be the Bartlett spectrum of a second order stationary random measure ξ such that Γ is absolutely continuous with respect to ν_d, with spectral density f_Γ.*

(a) If f_Γ is ν_d a.e. bounded in some neighbourhood of $\mathbf{0}$, then ξ is not long-range dependent.

(b) If for every $K > 0$ there exists a neighbourhood U_K of $\mathbf{0}$ such that $f_\Gamma(\boldsymbol{x}) \ge K$ for ν_d almost all $\boldsymbol{x} \in U_K$, then ξ is long-range dependent.

Proof. Again, as in the proof of Proposition 2.5.8 (e), consider $f_a = \mathbb{1}_{[-a,a]^d}$ for each $a > 0$. Hence

$$\frac{|\check{f}_a(\boldsymbol{\omega})|^2}{(2a)^d} = \prod_{j=1}^{d} \frac{\sin^2(a\omega_j)}{\pi a \omega_j^2}.$$

With the notation $\phi_a = (2a)^{-d}|\check{f}_a|^2$, it is easy to check that $\int_{R^d} \phi_a(\boldsymbol{\omega})\,\mathrm{d}\boldsymbol{\omega} = 1$ for every $a > 0$. Moreover, $\phi_a(\boldsymbol{\omega}) = a^d \phi_1(a\boldsymbol{\omega})$ for all $\boldsymbol{\omega} \in \mathbb{R}^d$ and all $a > 0$. This means that $\{\phi_{1/a} | a > 0\}$ is an approximation of the identity (see the definition before Theorem A.2.4). Therefore $\lim_{a\to\infty}(\phi_a * g)(\mathbf{0}) = g(\mathbf{0})$ for every bounded uniformly continuous g by Theorem A.2.4. Let $\{a_n\}$ be an arbitrary sequence such that $a_n > 0$ for all n and $a_n \to \infty$ as $n \to \infty$. Since

$$\int_{\mathbb{R}^d} g(\boldsymbol{\omega})\phi_{a_n}(\boldsymbol{\omega})\,\mathrm{d}\boldsymbol{\omega} \to g(\mathbf{0}) = \int_{\mathbb{R}^d} g(\boldsymbol{\omega})\delta_{\mathbf{0}}(\mathrm{d}\boldsymbol{\omega})$$

for all bounded uniformly continuous g, Theorem A.1.4 yields $\phi_{a_n}\nu_d \xrightarrow{\mathrm{w}} \delta_{\mathbf{0}}$. Let U be a neighbourhood of $\mathbf{0}$. The complement U^c is a set with $\delta_{\mathbf{0}}(U^c) = 0$ and $\delta_{\mathbf{0}}(\partial U^c) = 0$, hence from Theorem A.1.4 (e), $\lim_{n\to\infty}(\phi_{a_n}\nu_d)(U^c) = \lim_{n\to\infty}\int_{U^c}\phi_{a_n}(\boldsymbol{\omega})\,\mathrm{d}\boldsymbol{\omega} = 0$. From this follows $\lim_{n\to\infty}\int_{U^c}\phi_{a_n}(\boldsymbol{x})f_\Gamma(\boldsymbol{\omega})\,\mathrm{d}\boldsymbol{\omega} = 0$ and hence there is $n_0 > 0$ such that $\int_{U^c}\phi_{a_n}(\boldsymbol{x})f_\Gamma(\boldsymbol{\omega})\,\mathrm{d}\boldsymbol{\omega} < 1$ for all $n > n_0$.

For the proof of (a) we take U and K such that $f_\Gamma(\boldsymbol{x}) \le K$ for ν_d almost all $\boldsymbol{x} \in U$. From previous considerations, there is n_0 such that for all $n > n_0$ we have

$$\begin{aligned}\int_{\mathbb{R}^d} \frac{|\check{f}_{a_n}(\boldsymbol{\omega})|^2}{(2a_n)^d}\,\Gamma(\mathrm{d}\boldsymbol{\omega}) &= \int_{\mathbb{R}^d} \phi_{a_n}(\boldsymbol{\omega}) f_\Gamma(\boldsymbol{\omega})\,\mathrm{d}\boldsymbol{\omega}\\ &= \int_U \phi_{a_n}(\boldsymbol{x}) f_\Gamma(\boldsymbol{\omega})\,\mathrm{d}\boldsymbol{\omega} + \int_{U^c} \phi_{a_n}(\boldsymbol{x}) f_\Gamma(\boldsymbol{\omega})\,\mathrm{d}\boldsymbol{\omega}\\ &< K \int_U \phi_{a_n}(\boldsymbol{x})\,\mathrm{d}\boldsymbol{\omega} + 1\\ &< K + 1.\end{aligned}$$

Now we use Proposition 2.5.8 (d). Since $(f_a * f_a^*)(\mathbf{0}) = (2a)^d$ for all $n > n_0$,

$$\frac{\operatorname{var}\xi\big([-a_n, a_n]^d\big)}{2^d a_n^d} < \lambda_D + (2\pi)^{d/2}(K+1).$$

Since K is independent on the choice of $\{a_n\}$ the statement follows.

For the proof of (b) let take arbitrary $K > 0$ and corresponding U_K such that $f_\Gamma(\boldsymbol{x}) \ge K$ almost everywhere with respect to ν_d. Again by Theorem A.1.4 (e), $\lim_{n\to\infty}(\phi_{a_n}\nu_d)(U_K) = \lim_{n\to\infty}\int_{U_K}\phi_{a_n}(\boldsymbol{\omega})\,\mathrm{d}\boldsymbol{\omega} = 1$. From this follows $\liminf_{n\to\infty}\int_{U_K}\phi_{a_n}(\boldsymbol{x}) f_\Gamma(\boldsymbol{\omega})\,\mathrm{d}\boldsymbol{\omega} > K$ and hence there is $m_0 > 0$ such that $\int_{U^c}\phi_{a_n}(\boldsymbol{x}) f_\Gamma(\boldsymbol{\omega})\,\mathrm{d}\boldsymbol{\omega} > K/2$ for all $n > m_0$. Therefore for all $n > \max\{n_0, m_0\}$,

$$\begin{aligned}\int_{\mathbb{R}^d} \frac{|\check{f}_{a_n}(\boldsymbol{\omega})|^2}{(2a_n)^d}\,\Gamma(\mathrm{d}\boldsymbol{\omega}) &= \int_U \phi_{a_n}(\boldsymbol{x}) f_\Gamma(\boldsymbol{\omega})\,\mathrm{d}\boldsymbol{\omega} + \int_{U^c} \phi_{a_n}(\boldsymbol{x}) f_\Gamma(\boldsymbol{\omega})\,\mathrm{d}\boldsymbol{\omega}\\ &\ge \int_U \phi_{a_n}(\boldsymbol{x}) f_\Gamma(\boldsymbol{\omega})\,\mathrm{d}\boldsymbol{\omega} > \frac{K}{2}.\end{aligned}$$

Finally, for all sufficiently large n,

$$\frac{\operatorname{var}\xi\big([-a_n, a_n]^d\big)}{(2a_n)^d} > \lambda_D + (2\pi)^{d/2}\frac{K}{2},$$

which completes the proof. □

In the remainder of this section we assume $\check{C}^{[2]}$ of ξ to be absolutely continuous with respect to ν_d, with density $\operatorname{cov}(\boldsymbol{x})$ called the **covariance function**, and the Bartlett spectrum Γ of ξ to be absolutely continuous with **spectral density** f_Γ.

In most practical applications the long-range dependence is formulated using the power law decay of the covariance, or equivalently the power law dependence of the spectral density in the neighbourhood of $\mathbf{0}$. We present here the multidimensional extension motivated by [59] (see also [58, 60]).

Definition 2.8.2. Let ξ be a second order stationary random measure with covariance function cov. We say that ξ exhibits **isotropic long-range dependence** if

$$\operatorname{cov}(\boldsymbol{x}) \sim \|\boldsymbol{x}\|^{-\alpha}\,\ell(\|\boldsymbol{x}\|)\,g\left(\frac{\boldsymbol{x}}{\|\boldsymbol{x}\|}\right) \qquad (\|\boldsymbol{x}\| \to \infty),$$

where $\alpha \in (0, d)$, ℓ is slowly varying at infinity, and g is a positive function on the unit sphere in $\mathbb{R}^d$. A random closed set X exhibits **isotropic long-range dependence** if its volume measure ν_X exhibits isotropic long-range dependence.

We use $\sim$ for the asymptotic equality defined by: $f(\boldsymbol{x}) \sim g(\boldsymbol{x})$ ($\|\boldsymbol{x}\| \to \infty$) if and only $\lim_{\|\boldsymbol{x}\|\to\infty} f(\boldsymbol{x})/g(\boldsymbol{x}) = 1$. The **slowly varying function** $f : (0,\infty) \to \mathbb{R}$ is a positive measurable function satisfying

$$\frac{f(ut)}{f(t)} \to 1 \quad \text{as} \quad t \to \infty$$

for every $u > 0$. If the limit equals u^α we say that f is **regularly varying of degree** α. The slowly varying function ℓ in the previous definition can be actually considered continuous, see [61, Proposition 1.3.4] or [62, p. 56] for details, and we may write

$$\operatorname{cov}(\boldsymbol{x}) = h(\boldsymbol{x})\,\|\boldsymbol{x}\|^{-\alpha}\,\ell\big(\,\|\boldsymbol{x}\|\,\big)\,g\left(\frac{\boldsymbol{x}}{\|\boldsymbol{x}\|}\right) \tag{2.40}$$

for all $\boldsymbol{x} \in \mathbb{R}^d \setminus \{\boldsymbol{0}\}$, where h is bounded and $h(\boldsymbol{x}) \to 1$ as $\|x\| \to \infty$.

Proposition 2.8.5. *Let ℓ be a slowly varying function and $\beta > 0$. Then*

$$x^\beta \ell(x) \to \infty \qquad \text{and} \qquad x^{-\beta}\ell(x) \to 0$$

as $x \to \infty$.

Proof. Proposition 1.3.6 in [61]. □

The next result concerns the asymptotic behaviour of the special integrals of regularly varying functions.

Theorem 2.8.1 (Karamata's theorem)**.** *Let f be a regularly varying of degree γ, and locally bounded in $[x_0,\infty)$ for some x_0. Then*

(a) for any $\beta \ge -(\gamma+1)$,

$$x^{\beta+1} f(x) \Big/ \int_{x_0}^{x} t^\beta f(t)\,\mathrm{d}t \to \gamma+\beta+1 \qquad (x\to\infty),$$

(b) for any $\beta < -(\gamma+1)$,

$$x^{\beta+1} f(x) \Big/ \int_{x}^{\infty} t^\beta f(t)\,\mathrm{d}t \to \gamma+\beta+1 \qquad (x\to\infty).$$

Conversely, if f is positive, locally integrable in $[x_0,\infty)$ for some x_0 and for some β satisfies (a) with sharp inequality, or (b), then it is regularly varying of degree γ.

Proof. Theorems 1.5.11 and 1.6.1 in [61]. □

Let us now discuss the range $(0,d)$ of the coefficient α in Definition 2.8.2. Assume that the covariance function cov of ξ is given by (2.40). The case $\alpha < 0$ cannot occur, since otherwise Proposition 2.8.5 implies that the reduced second order factorial covariance measure $\breve{C}^{[2]}$ cannot be translation bounded as stated by Corollary 2.5.1.

In order to analyse the case $\alpha > 0$, let us take r_0 such that $|h(\boldsymbol{x})| < 2$ for all $\boldsymbol{x}$ with $\|\boldsymbol{x}\| > r_0$. Then for $a > r_0$,

$$\left|\breve{C}^{[2]}\right|\left(aB^d\right) = \int_{aB^d} |\operatorname{cov}(\boldsymbol{x})|\,\mathrm{d}\boldsymbol{x} = \left|\breve{C}^{[2]}\right|\left(r_0B^d\right) + \int_{aB^d\setminus r_0B^d} |\operatorname{cov}(\boldsymbol{x})|\,\mathrm{d}\boldsymbol{x}.$$

Using spherical coordinates $\boldsymbol{x} = r\boldsymbol{u}$, $r \in (0,\infty), \boldsymbol{u} \in S^{d-1}$,

$$\begin{aligned}\left|\breve{C}^{[2]}\right|\left(aB^d\right) &= \left|\breve{C}^{[2]}\right|\left(r_0B^d\right) + \int_{r_0}^{a}\int_{S^{d-1}} |h(r\boldsymbol{u})|\, r^{d-1-\alpha}\ell(r)\,|g(\boldsymbol{u})|\;\sigma_{d-1}(\mathrm{d}\boldsymbol{u})\,\mathrm{d}r \\ &\le \left|\breve{C}^{[2]}\right|\left(r_0B^d\right) + 2\tilde{G}\int_{r_0}^{a} r^{d-1-\alpha}\ell(r)\,\mathrm{d}r,\end{aligned}$$

where σ_{d-1} is the spherical measure (the $d-1$ dimensional Hausdorff measure) on the unit sphere S^{d-1} in $\mathbb{R}^d$ and $G = \int_{S^{d-1}} g(\boldsymbol{u})\sigma_{d-1}(\mathrm{d}\boldsymbol{u}) > 0$. By Proposition 2.5.8 (c),

$$\begin{aligned}\Gamma\big(\{\mathbf{0}\}\big) &\le (2\pi)^{d/2}\limsup_{a\to\infty}\frac{\left|\breve{C}^{[2]}\right|\left([-a,a]^d\right)}{(2a)^d} \le (2\pi)^{d/2}\limsup_{a\to\infty}\frac{\left|\breve{C}^{[2]}\right|\left(a\sqrt{d}B^d\right)}{(2a)^d} \\ &\le (2\pi)^{d/2}\limsup_{a\to\infty}\frac{\breve{C}^{[2]}\left(r_0B^d\right)}{(2a)^d} + (2\pi)^{d/2}\limsup_{a\to\infty}\frac{2\tilde{G}}{(2a)^d}\int_{r_0}^{a\sqrt{d}} r^{d-1-\alpha}\ell(r)\,\mathrm{d}r.\end{aligned}$$

Now Theorem 2.8.1 (a) (for $d=1$ only when $\alpha<0$) yields

$$\frac{G}{2(2a)^d}\int_{r_0}^{a} r^{d-1-\alpha}\ell(r)\,\mathrm{d}r \sim \frac{Ga^{-\alpha}\ell(a)}{2^{d+1}(d-\alpha)} \qquad (a\to\infty),$$

and by Proposition 2.8.5, $a^{-\alpha}\ell(a)\to 0$ as $a\to\infty$. Thus we finally obtain $\Gamma\big(\{\mathbf{0}\}\big)=0$ which means the compatibility with possible ergodicity of ξ according to Proposition 2.8.3.

In the case $\alpha > d$ it follows from Proposition 2.8.5 with $\beta = (\alpha-d)/2$ that

$$\int_{r_0}^{\infty} r^{d-1-\alpha}\ell(r)\,\mathrm{d}r \le C\int_{r_0}^{\infty} r^{d-1-\alpha+\beta}\,\mathrm{d}r < \infty,$$

where C is the upper bound of $r^{-\beta}\ell(r)$ on $[r_0,\infty)$. Hence $\breve{C}^{[2]}$ is bounded and by Corrolary 2.8.1 it is not long-range dependent. The case $\alpha=0$ may lead to various situations, since the limit of $\ell(a)$ may be arbitrary from 0 to ∞, or it may even not exist. Only when ξ is ergodic then $\ell(a)$ must tend to 0 as $a\to\infty$ as follows from Proposition 2.8.3. The case $\alpha = d$ may again lead to various results depending on ℓ. Note, however, that all of them imply $\Gamma\big(\{\mathbf{0}\}\big)=0$ and are thus compatible with possible ergodicity of ξ. See [61, §1.5 and 1.6] for further discussion on integrals of this type.

The next result represents a very important relation between the behaviour of the covariance function at large distances and the behaviour of the spectral density in the neighbourhood of the origin.

Theorem 2.8.2. *A second order stationary random measure ξ with Bartlett spectrum Γ exhibits isotropic long-range dependence if and only if for the corresponding spectral density f_Γ holds*

$$f_\Gamma(\boldsymbol{\omega}) \sim \|\boldsymbol{\omega}\|^{\alpha-d}\,\ell\big(\|\boldsymbol{\omega}\|^{-1}\big)g\left(\frac{\boldsymbol{\omega}}{\|\boldsymbol{\omega}\|}\right) \qquad (\|\boldsymbol{\omega}\|\to 0_+),$$

where $\alpha\in(0,d)$, ℓ is slowly varying at infinity and g is positive continuous function on the unit sphere in $\mathbb{R}^d$.

Proof. The proof is based on the asymptotic behaviour of the Fourier transform of quasi-asymptotically homogeneous tempered distributions on the Schwartz space $\mathscr{S}(\mathbb{R}^d)$ of rapidly

decreasing smooth functions. First note that every translation bounded signed measure μ can be uniquely identified as a tempered distribution $\mu \in \mathscr{S}'(\mathbb{R}^d)$, since $f \in L^1(\mathbb{R}^d, \mu)$. This can be shown using the criterion derived in [63, Theorem 2.1], claiming that $\mu \in S'(\mathbb{R}^d)$ if and only if there exists real β such that $(1 + \|x\|^2)^{\beta/2} \in L^1(\mathbb{R}^d, \mu)$. However, this condition is satisfied for all translation bounded measures with $\beta = -2d$ as can be checked based on convergence of the series $\sum_{k=1}^{\infty} k^{-2}$. The result now follows from Theorem 3.1 in [64] together with Lemma 2.3 in [64] or Theorem 1 in [62, §3.3]. Since $\alpha \in (0, d)$, the asymptotic result is locally integrable. Finally to obtain the proper functional form of f_Γ (or cov in the opposite implication) one uses Theorem 1.4.1 (iii) in [61]. The non-negativity of g follows from the positivity of Γ. □

Note that the proof can be also made without the use of distributions. This is the way how it is usually done in $\mathbb{R}^1$. It is based on Theorem 4.1.5 in [61]. However, in $\mathbb{R}^d$ there appears the Hankel type transform of $t^{d/2-\alpha-1}$ that does not converge for $2\alpha < d - 1$. The general theorems that deal with such kind of dual relations are called Abellian and Tauberian theorems.

Corollary 2.8.2. *If ξ is isotropic long-range dependent then it is long-range dependent according to Definition 2.8.1.*

Proof. The assertion follows from Proposition 2.8.4 and the fact that $x^\beta \ell(x) \to \infty$ as $x \to \infty$ for $\beta > 0$ by Proposition 2.8.5. □

Finally, we prove the following proposition that will be important for statistical issues connected to estimation. Recall that the set covariance $\gamma_B(\boldsymbol{x})$ of a Borel set B is defined by $\gamma_B(\boldsymbol{x}) = \nu_d(B \cap (B + \boldsymbol{x}))$.

Proposition 2.8.6. *Let ξ be an isotropically long-range dependent random measure with covariance function of the form (2.40) and W be a compact set of positive d-dimensional Lebesgue measure. Then*

$$\operatorname{var} \xi\big(aW\big) \sim \lambda_D \nu_d(W) a^d + a^{2d-\alpha}\ F_{\alpha;W} \ell(a) \qquad (a \to \infty),$$

where

$$F_{\alpha;W} = \int_0^\infty \int_{S^{d-1}} \gamma_W(t\boldsymbol{u}) t^{d-1-\alpha} g(\boldsymbol{u})\ \sigma_{d-1}(\mathrm{d}\boldsymbol{u})\, \mathrm{d}t.$$

Proof. For the proof we start with Proposition 2.5.8 (d) and (b) which yields

$$\operatorname{var} \xi(aW) = \lambda_D (\mathbb{1}_{aW} * \mathbb{1}^*_{aW})(\mathbf{0}) + \int_{\mathbb{R}^d} (\mathbb{1}_{aW} * \mathbb{1}^*_{aW})(\boldsymbol{x})\ \breve{C}^{[2]}(\mathrm{d}\boldsymbol{x}).$$

Clearly

$$(\mathbb{1}_{aW} * \mathbb{1}^*_{aW})(\boldsymbol{x}) = \int_{\mathbb{R}^d} \mathbb{1}_{aW}(\boldsymbol{y}) \mathbb{1}_{aW}(\boldsymbol{y} - \boldsymbol{x})\, \mathrm{d}\boldsymbol{y} = \nu_d\big(aW \cap (aW + \boldsymbol{x})\big) = \gamma_{aW}(\boldsymbol{x}).$$

It is easy to see that $\gamma_{aW}(a\boldsymbol{x}) = a^d \gamma_W(\boldsymbol{x})$. Therefore

$$\operatorname{var} \xi(aW) = \lambda_D \nu_d(W) a^d + a^d \int_{\mathbb{R}^d} \gamma_W\left(\frac{\boldsymbol{x}}{a}\right)\ \breve{C}^{[2]}(\mathrm{d}\boldsymbol{x}).$$

Imposing (2.40) yields

$$a^d \int_{\mathbb{R}^d} \gamma_W\left(\frac{\boldsymbol{x}}{a}\right)\ \breve{C}^{[2]}(\mathrm{d}\boldsymbol{x}) = a^d \int_{\mathbb{R}^d} \gamma_{B^d}\left(\frac{\boldsymbol{x}}{a}\right) h(\boldsymbol{x})\, \|x\|^{-\alpha}\, \ell(\|\boldsymbol{x}\|) g\left(\frac{\boldsymbol{x}}{\|\boldsymbol{x}\|}\right)\, \mathrm{d}\boldsymbol{x}.$$

Using spherical coordinates $\boldsymbol{x} = r\boldsymbol{u}$, $r \in (0, \infty), \boldsymbol{u} \in S^{d-1}$ we obtain

$$a^d \int_{\mathbb{R}^d} \gamma_W\left(\frac{\boldsymbol{x}}{a}\right) \breve{C}^{[2]}(\mathrm{d}\boldsymbol{x}) = a^d \int_0^\infty \int_{S^{d-1}} \gamma_W\left(\frac{r\boldsymbol{u}}{a}\right) h(r\boldsymbol{u}) r^{d-1-\alpha} \ell(r) g(\boldsymbol{u})\ \sigma_{d-1}(\mathrm{d}\boldsymbol{u})\,\mathrm{d}r.$$

Using the substitution $r = at$ leads to

$$\begin{aligned} a^d \int_{\mathbb{R}^d} \gamma_W\left(\frac{\boldsymbol{x}}{a}\right) \breve{C}^{[2]}(\mathrm{d}\boldsymbol{x}) &= a^{2d-\alpha} \int_0^\infty \int_{S^{d-1}} \gamma_W(t\boldsymbol{u}) t^{d-1-\alpha} h(at\boldsymbol{u}) \ell(at) g(\boldsymbol{u})\ \sigma_{d-1}(\mathrm{d}\boldsymbol{u})\,\mathrm{d}t \\ &= a^{2d-\alpha}\ell(a) \int_0^\infty \int_{S^{d-1}} \gamma_W(t\boldsymbol{u}) t^{d-1-\alpha} h(at\boldsymbol{u}) \frac{\ell(at)}{\ell(a)}\ \sigma_{d-1}(\mathrm{d}\boldsymbol{u})\,\mathrm{d}t. \end{aligned}$$

Letting $a \to \infty$ yields $h(at\boldsymbol{u}) \to 1$ for all $\boldsymbol{u} \in S^{d-1}$ and all $t > 0$. It also yields $\frac{\ell(at)}{\ell(a)} \to 1$ for each $t > 0$ since ℓ is slowly varying function. The continuity of ℓ implies that $\frac{\ell(at)}{\ell(a)}$ is bounded for all $t \in [0, r_W]$ and all $a > 0$, where $r_W < \infty$ is determined by the condition $\gamma_W(r\boldsymbol{u}) = 0$ for all $r > r_W$ and all $\boldsymbol{u} \in S^{d-1}$. Such r_W always exists because W is compact. Since $d - 1 - \alpha > -1$ the rest of the integrand is integrable on $[0, r_W] \times S^{d-1}$ and thus by the Lebesgue dominated convergence theorem

$$\begin{aligned} &\int_0^\infty \int_{S^{d-1}} \gamma_W(t\boldsymbol{u}) t^{d-1-\alpha} h(at\boldsymbol{u}) \frac{\ell(at)}{\ell(a)}\ \sigma_{d-1}(\mathrm{d}\boldsymbol{u})\,\mathrm{d}t \\ &\qquad\qquad \to \int_0^\infty \int_{S^{d-1}} \gamma_W(t\boldsymbol{u}) t^{d-1-\alpha} g(\boldsymbol{u})\ \sigma_{d-1}(\mathrm{d}\boldsymbol{u})\,\mathrm{d}t \end{aligned}$$

as $a \to \infty$. This completes the proof. □

The previous Corollary is also implied by this proposition.

Chapter 3

Generalized centroid

Let $\mathcal{W}$ be a Borel subset of the system $\mathcal{C}'$ of non-empty compact subsets of $\mathbb{R}^d$. In Section 2.7 we saw that for a stationary particle processes in $\mathcal{W}$ it is possible to find a useful decomposition of the intensity measure Λ. To do this we need a Borel measurable mapping $\boldsymbol{z} : \mathcal{W} \to \mathbb{R}^d$ that is translation invariant, i.e. it satisfies $\boldsymbol{z}(C + \boldsymbol{x}) = \boldsymbol{z}(C) + \boldsymbol{x}$ for every $\boldsymbol{x} \in \mathbb{R}^d$ and every C in $\mathcal{W}$. For clear interpretation it is preferable to choose the mapping $\boldsymbol{z}$ so that $\boldsymbol{z}(C)$ is located at some geometrically reasonable centre of a set C. Usual choices (see [40]) are the centre of the smallest ball containing C or the Steiner point of the convex hull of C.

In this chapter we construct a different centre function that is physically more plausible and is based on the centroid of a set. We start by introducing the classical definition and discussing its properties and measurability. Then we propose a possible generalization based on the limiting procedure. The existence issues are discussed and the measurability is proven for a certain subclasses of sets. Finally we show the second possible generalization based on the Hausdorff measure and discuss their relation.

Throughout this chapter $\mathcal{C}'$ is assumed to be equipped with the Hausdorff topology defined in Section A.5. The Borel σ-algebra on $\mathcal{C}'$ is denoted by $\mathcal{B}(\mathcal{C}')$.

3.1 Classical definition

First we start with the classical definition. A **centroid** of a non-empty compact set $C \in \mathcal{C}'$ with positive d-dimensional Lebesgue measure, $\nu_d(C) > 0$, is given by

$$\boldsymbol{z}(C) = \frac{1}{\nu_d(C)} \int_C \boldsymbol{x} \, \mathrm{d}\boldsymbol{x} = \frac{1}{\nu_d(C)} \left(\int_C x_1 \, \mathrm{d}\boldsymbol{x}, \int_C x_2 \, \mathrm{d}\boldsymbol{x}, \ldots, \int_C x_d \, \mathrm{d}\boldsymbol{x} \right), \tag{3.1}$$

where $\boldsymbol{x} = (x_1, x_2, \ldots, x_d) \in \mathbb{R}^d$. The point $\boldsymbol{z}(C)$ is also called the centre of mass since it corresponds to the centre of mass of an object geometrically corresponding to set C with homogeneously distributed density, which is used in physics.

In order to be able to construct particle process with particles concentrated on sets with positive Lebesgue measure it must be shown that collection of such sets forms a Borel set in $\mathcal{C}'$.

Theorem 3.1.1. *The collection of all compact sets with positive Lebesgue measure is a Borel set in $\mathcal{C}'$.*

Proof. The Lebesgue measure ν_d is σ-finite measure and therefore upper semi-continuous as a consequence of Theorem A.5.4. According to the remark after Lemma A.4.3 it is also measurable. Hence $\nu_d^{-1}((0, \infty))$ is a Borel set. □

Theorem 3.1.2. *The centroid $\boldsymbol{z}$ is a Borel measurable mapping on $\mathcal{C}'$ with positive Lebesgue measure.*

Proof. From Theorem A.5.4 we know that every σ-finite measure is upper semi-continuous. Let us define two σ-finite measures μ_+ and μ_- on $\mathbb{R}^d$ by

$$\mu_+(A) = \int_A x_1 \mathbb{1}_{\{x_1 \geq 0\}}(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} \quad \text{and} \quad \mu_-(A) = -\int_A x_1 \mathbb{1}_{\{x_1 < 0\}}(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}.$$

Formula (3.1) for $z_1(C)$ can be now rewritten as

$$z_1(C) = \frac{\mu_+(C) - \mu_-(C)}{\nu_d(C)}.$$

Since all three elements are upper semi-continuous and therefore measurable, z_1 is also measurable. The same argumentation is valid for all other components of $\boldsymbol{z}(C)$. □

It is worth noting that the centroid is generally not continuous. To see this let take a sequence $\{C_i\}$ of finite sets converging to the unit ball B^d. A sequence $C_i \cup (B^d + \boldsymbol{x}_0)$, where $\boldsymbol{x}_0$ is an arbitrary point with $\|\boldsymbol{x}_0\| > 2$, then converges to a union $B^d \cup (B^d + \boldsymbol{x}_0)$, because union operation is continuous in $\mathcal{C}'$ according to Theorem A.5.5. Since C_i are finite sets with zero Lebesgue measure we get

$$\lim_{i \to +\infty} \boldsymbol{z}\big(C_i \cup (B^d + \boldsymbol{x}_0)\big) = \lim_{i \to +\infty} \boldsymbol{x}_0 = \boldsymbol{x}_0 \neq \boldsymbol{z}\big(B^d \cup (B^d + \boldsymbol{x}_0)\big) = \frac{\boldsymbol{x}_0}{2}.$$

The centroid is therefore not continuous and, since $\boldsymbol{x}_0$ was chosen arbitrarily, even not upper or lower semi-continuous in any component.

The continuity can be proven when restricted on a family $\mathcal{K}'$ of non-empty compact convex sets. From Theorem A.5.10 it is known that the denominator of (3.1) is continuous. In the following lemma we show that the numerator is also continuous.

Lemma 3.1.1. *The mapping $K \mapsto \int_K \boldsymbol{x}\,\mathrm{d}\boldsymbol{x}$ from $\mathcal{K}'$ to $\mathbb{R}^d$ is continuous on $\mathcal{K}'$.*

Proof. For an arbitrary compact, convex set K a measure μ_K on $\mathbb{R}^d$ can be constructed by taking

$$\mu_K(A) = \nu_d(A \cap K) \quad \text{for all Borel } A \subset \mathbb{R}^d.$$

The measure μ_K is finite since $\mu_K(\mathbb{R}^d) = \nu_d(K) < +\infty$.

Now let $K_i \to K$ be a convergent sequence in $\mathcal{K}'$ with respect to the Hausdorff metric. From the continuity of the Lebesgue measure (Theorem A.5.10) follows

$$\lim_i \mu_{K_i}(\mathbb{R}^d) = \lim_i \nu_d(K_i) \to \nu_d(K) = \mu_K(\mathbb{R}^d).$$

Since convergence in K' implies convergence in $\mathcal{F}$ (see Theorem A.5.3), by Theorem A.4.5 and Proposition A.4.1 we obtain

$$\limsup_i \mathbb{1}_{K_i}(\boldsymbol{x}) \leq \mathbb{1}_K(\boldsymbol{x}).$$

This for arbitrary closed set $F \in \mathcal{F}$ yields

$$\mu_K(F) = \nu_d(K \cap F) = \int_F \mathbb{1}_K(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} \geq \int_F \limsup_i \mathbb{1}_{K_i}(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} \geq$$

$$\geq \limsup_i \int_F \mathbb{1}_{K_i}(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} = \limsup_i \nu_d(K_i \cap F) = \limsup_i \mu_{K_i}(F)$$

by Fatou's lemma. The convergence of K_i is bounded (see Theorem A.5.3), so there exists a set $L \in \mathcal{C}'$, such that $\mathbb{1}_{K_i} \leq \mathbb{1}_L$, $\mathbb{1}_K \leq \mathbb{1}_L$, and $\int_F \mathbb{1}_L \,\mathrm{d}x < +\infty$. Hence Fatou's lemma can be applied.

For the sequence μ_{K_i} we proved all properties of Theorem A.1.4 (c) and thus $\mu_{K_i} \xrightarrow{\mathrm{w}} \mu_K$. Now let f be a bounded, continuous function from Theorem A.1.5 such that $f(\boldsymbol{x}) = 1$ for $\boldsymbol{x} \in L$, where $K, K_i \subset L$. A function g given by $g(\boldsymbol{x}) = x_1 f(\boldsymbol{x})$ is also bounded and continuous. The weak continuity implies

$$\int_{K_i} x_1 \,\mathrm{d}\boldsymbol{x} = \int_{\mathbb{R}^d} \mathbb{1}_{K_i}(\boldsymbol{x}) g(\boldsymbol{x}) \,\mathrm{d}\boldsymbol{x} = \int_{\mathbb{R}^d} g(\boldsymbol{x}) \mu_{K_i}(\mathrm{d}\boldsymbol{x}) \to \int_{\mathbb{R}^d} g(\boldsymbol{x}) \mu_K(\mathrm{d}\boldsymbol{x}) = \int_K x_1 \,\mathrm{d}\boldsymbol{x}.$$

The same holds for other components of $\boldsymbol{x}$ which completes the proof. □

Note that we have actually proven the continuity of the mapping $K \mapsto \int_K f(\boldsymbol{x}) \,\mathrm{d}\boldsymbol{x}$ for every continuous real function f bounded on compact sets. Since the Lebesgue measure ν_d is from Theorem A.5.10, or from previous remark, also continuous we obtain the following consequence.

Corollary 3.1.1. *The centroid $\boldsymbol{z}$ given by (3.1) is continuous on the system of non-empty convex compact sets with positive Lebesgue measure.*

The important property of the centroid is compatibility with euclidean transformations (i.e. translations, rotations, and reflections in $\mathbb{R}^d$).

Proposition 3.1.1. *The centroid satisfies $\boldsymbol{z}(eC) = e\boldsymbol{z}(C)$ for all $C \in \mathcal{C}', \nu_d(C) > 0$ and every Euclidean transformation e.*

Proof. An Euclidean transformation $e : \mathbb{R}^d \to \mathbb{R}^d$ can be written as

$$e\boldsymbol{x} = L\boldsymbol{x} + \boldsymbol{a},$$

where L is an orthogonal linear transformation ($L \in O_d$) and $a \in \mathbb{R}^d$. The Lebesgue measure ν_d is invariant under Euclidean transformations giving $\nu_d(eC) = \nu_d(C)$. Substituting $e\boldsymbol{y} = \boldsymbol{x}$ into the integral $\int_{eC} \boldsymbol{x} \,\mathrm{d}\boldsymbol{x}$ yields

$$\int_{eC} \boldsymbol{x} \,\mathrm{d}\boldsymbol{x} = \int_C e\boldsymbol{y} \,|\det L| \,\mathrm{d}\boldsymbol{y} = L \int_C \boldsymbol{y} \,\mathrm{d}\boldsymbol{y} + \boldsymbol{a}\nu_d(C).$$

By dividing the previous equation with $\nu_d(C)$ we conclude that

$$\boldsymbol{z}(eC) = L\frac{\int_C \boldsymbol{x} \,\mathrm{d}\boldsymbol{x}}{\nu_d(C)} + \boldsymbol{a} = e\boldsymbol{z}(C).$$ □

3.2 General definition

The usual definition of $\boldsymbol{z}$ can be applied only for sets with positive Lebesgue measure. In this section we will restrict our attention to sets with zero Lebesgue measure. Let C be non-empty closed set with $\nu_d(C) = 0$. One possible approach to overcome the problem of division by zero in (3.1) is to take an ε-neighbourhood C_ε of C defined by

$$C_\varepsilon = C + \varepsilon U^d = \{\boldsymbol{x} \mid d(\boldsymbol{x}, C) < \varepsilon\},$$

and calculate its centroid. Then one can try to take a limit of $\boldsymbol{z}(C_\varepsilon)$ as $\varepsilon \to 0+$ and if it exits then it is reasonable to call the result a centroid of C. This motivates us to the following definition.

Definition 3.2.1. Let $C \in \mathcal{C}'$ be a non-empty compact set in $\mathbb{R}^d$ such that a limit

$$\lim_{\varepsilon \to 0_+} \frac{1}{\nu_d(C_\varepsilon)} \int_{C_\varepsilon} \boldsymbol{x} \, \mathrm{d}\boldsymbol{x} \tag{3.2}$$

exists. We denote the limit by $\boldsymbol{z}_M(C)$ and call it the **generalized (Minkowski) centroid** of C.

The first natural question is whether this definition coincide with the usual centroid for sets with positive Lebesgue measure.

Proposition 3.2.1. *Let $C \in \mathcal{C}'$ be an arbitrary non-empty compact set with positive Lebesgue measure, $\nu_d(C) > 0$. Then the limit (3.2) exists and is equal to the classical definition of $\boldsymbol{z}(C)$ given by (3.1).*

Proof. Let $\{\varepsilon_n\}_{n\in\mathbb{N}}$ be a sequence such that $\varepsilon_n \to 0$ and $\varepsilon_n > 0$ for every $n \in \mathbb{N}$. Since $C \subset C_{\varepsilon_n}$ we have $\mathbb{1}_{C_{\varepsilon_n}}(\boldsymbol{x}) = 1 = \mathbb{1}_C(\boldsymbol{x})$ for all $\boldsymbol{x} \in C$. Let $\boldsymbol{x} \notin C$. Since C is compact there have to be only a finite number of i such that $\boldsymbol{x} \in C_{\varepsilon_i}$. Otherwise one can find a sequence of points in C converging to $\boldsymbol{x}$. Therefore for all $\boldsymbol{x} \notin C$ we have $\mathbb{1}_{C_{\varepsilon_n}}(\boldsymbol{x}) \to 0 = \mathbb{1}_C(\boldsymbol{x})$.

The Lebesgue dominated convergence theorem implies

$$\lim_{n\to+\infty} \int_{\mathbb{R}^d} x_1 \mathbb{1}_{C_{\varepsilon_n}}(\boldsymbol{x}) \, \mathrm{d}\boldsymbol{x} = \int_{\mathbb{R}^d} \lim_{n\to+\infty} x_1 \mathbb{1}_{C_{\varepsilon_n}}(\boldsymbol{x}) \, \mathrm{d}\boldsymbol{x} = \int_{\mathbb{R}^d} x_1 \mathbb{1}_C(\boldsymbol{x}) \, \mathrm{d}\boldsymbol{x}.$$

The same proof works for other components of $\boldsymbol{x} = (x_1, \ldots, x_d)$. By the same arguments

$$\lim_{n\to+\infty} \nu_d(C_{\varepsilon_n}) = \nu_d(C).$$

This limit also follows from the continuity of ν_d from above (Theorem A.1.1 (b)) since we can without loss of generality assume that $\{\varepsilon_n\}$ is decreasing. Thus limits of both numerator and denominator parts of (3.2) exist. Since $\nu_d(C) > 0$, the interchange of the limit and division leads to

$$\lim_{\varepsilon \to 0_+} \frac{1}{\nu_d(C_\varepsilon)} \int_{C_\varepsilon} \boldsymbol{x} \, \mathrm{d}\boldsymbol{x} = \frac{\lim_{\varepsilon\to 0_+} \int_{C_\varepsilon} \boldsymbol{x} \, \mathrm{d}\boldsymbol{x}}{\lim_{\varepsilon\to 0_+} \nu_d(C_\varepsilon)} = \frac{\int_C \boldsymbol{x} \, \mathrm{d}\boldsymbol{x}}{\nu_d(C)} = \boldsymbol{z}(C). \qquad \square$$

The other class of sets for which the centroid exists is given by centrally symmetric compact sets.

Proposition 3.2.2. *Let C be a non-empty compact set in $\mathbb{R}^d$ and $\boldsymbol{a}$ a point in $\mathbb{R}^d$ such that C possesses point symmetry with respect to $\boldsymbol{a}$, i.e. $C = r_{\boldsymbol{a}}(C)$ where $r_{\boldsymbol{a}}(\boldsymbol{x}) = 2\boldsymbol{a} - \boldsymbol{x}$. Then the generalized centroid $\boldsymbol{z}_M(C)$ of C exists and $\boldsymbol{z}_M(C) = \boldsymbol{a}$.*

Proof. Let take arbitrary $\varepsilon > 0$ and observe that C_ε is also invariant with respect to $r_{\boldsymbol{a}}$, which follows from

$$\|r_{\boldsymbol{a}}(\boldsymbol{x}) - r_{\boldsymbol{a}}(\boldsymbol{y})\| = \|\boldsymbol{y} - \boldsymbol{x}\|$$

and the fact, that $\boldsymbol{x} \in C$ implies $r_{\boldsymbol{a}}(\boldsymbol{x}) \in C$. Next, note that $r_{\boldsymbol{a}}$ is a special example of an Euclidean transformation. Moreover it has a unique fixed point $\boldsymbol{a}$. From Proposition 3.1.1 we obtain

$$\boldsymbol{z}(C_\varepsilon) = \boldsymbol{z}\big(r_{\boldsymbol{a}}(C_\varepsilon)\big) = r_{\boldsymbol{a}}\big(\boldsymbol{z}(C_\varepsilon)\big).$$

Therefore $\boldsymbol{z}(C_\varepsilon)$ is a fixed point of $r_{\boldsymbol{a}}$ and hence $\boldsymbol{z}(C_\varepsilon) = \boldsymbol{a}$ is independent of ε. The proof is completed by letting $\varepsilon \to 0_+$. $\square$

Now we are able to prove the compatibility with euclidean transformations.

Theorem 3.2.1. *Let C be a non-empty compact set for which the generalized centroid exists and e be an Euclidean transformation. Then the generalized centroid of eC exists and $\boldsymbol{z}_M(eC) = e\boldsymbol{z}_M(C)$.*

Proof. An Euclidean transformation $e : \mathbb{R}^d \to \mathbb{R}^d$ can be written as

$$e\boldsymbol{x} = L\boldsymbol{x} + \boldsymbol{a},$$

where L is an orthogonal linear transformation ($L \in O_d$) and $\boldsymbol{a} \in \mathbb{R}^d$. From the straightforward equality

$$\|e\boldsymbol{x} - e\boldsymbol{y}\| = \|L(\boldsymbol{x} - \boldsymbol{y})\| = \|\boldsymbol{x} - \boldsymbol{y}\|$$

we may conclude that $(eC)_\varepsilon = e(C_\varepsilon)$. By continuity of e and Proposition 3.1.1,

$$\lim_{\varepsilon\to 0_+} \boldsymbol{z}\big((eC)_\varepsilon\big) = \lim_{\varepsilon\to 0_+} \boldsymbol{z}\big(e(C_\varepsilon)\big) = \lim_{\varepsilon\to 0_+} e\boldsymbol{z}(C_\varepsilon) = e \lim_{\varepsilon\to 0_+} \boldsymbol{z}(C_\varepsilon) = e\boldsymbol{z}_M(C). \qquad \square$$

From the previous propositions we know that the generalized centroid exists for compact sets with positive Lebesgue measure, for point symmetric sets and the existence is invariant under Euclidean transformations. The question is whether it exists for every compact set. The following proposition shows that the answer is negative.

Proposition 3.2.3. *There are compact sets in $\mathbb{R}^d$ for which the limit (3.2) does not exist.*

Proof. We give the proof only for the case $d = 2$; the other dimensions can be treated in an analogous fashion. Throughout the proof we use notation $\boldsymbol{x} = (x, y)$. We will construct the sequence of parallel lines of length L perpendicular to the x-axis and touching it with one end at x-coordinates $\{x_i\}_{i\in\mathbb{N}_0}$.

Let $\{a_k\}_{k\in\mathbb{N}_0}$ be a sequence of integers with $a_0 = 1$ and $\{\delta_k\}_{k\in\mathbb{N}_0}$ be a strictly decreasing sequence of positive numbers such that

$$\sum_{k=0}^{+\infty} a_k\delta_k < +\infty.$$

The points x_i depend on sequences $\{a_k\}$ and $\{\delta_k\}$ so that lines form blocks of a_k lines with gaps δ_k between neighbours and between the first line of the block and the previous block. The first position is $x_0 = \delta_0$. The i-th position is therefore given by $x_i = x_{i-1} + \delta_k$, where k is such that $\sum_{l=0}^{k-1} a_l \le i < \sum_{l=0}^{k} a_l$. From the properties of sequences $\{a_k\}$ and $\{\delta_k\}$ follows

$$\lim_{i\to+\infty} x_i = \sum_{k=0}^{+\infty} a_k\delta_k < +\infty.$$

The set C given by the union of those lines and a line at $\sum_{k=0}^{+\infty} a_k\delta_k$ is obviously compact and can be schematically depicted as in the figure 3.1.

Since the sequence $\{\delta_k\}$ is strictly decreasing one can for every $\varepsilon > 0$ find $n(\varepsilon)$ such that

$$\delta_{n(\varepsilon)-1} \ge 2\varepsilon \quad \text{and} \quad \delta_{n(\varepsilon)} < 2\varepsilon.$$

Therefore the set C_ε can be divided into two parts. The first part consists of isolated objects given by first $\sum_{k=0}^{n(\varepsilon)-1} a_k$ lines covered by ε balls. The area of every such object is $2\varepsilon L + \pi\varepsilon^2$ and the x-th component of the centre of mass is the corresponding x_i. The second part is the union of the remaining lines covered by ε balls forming one connected (rectangle like) object. Its x-size is $\sum_{k=n(\varepsilon)}^{+\infty} a_k\delta_k - \delta_{n(\varepsilon)} + 2\varepsilon$. Its area is approximately

$$\left(\sum_{k=n(\varepsilon)}^{+\infty} a_k\delta_k - \delta_{n(\varepsilon)}\right)(L + 2\varepsilon) + (2\varepsilon L + \pi\varepsilon^2) + \mathcal{O}(\varepsilon) \qquad (\varepsilon \to 0_+)$$

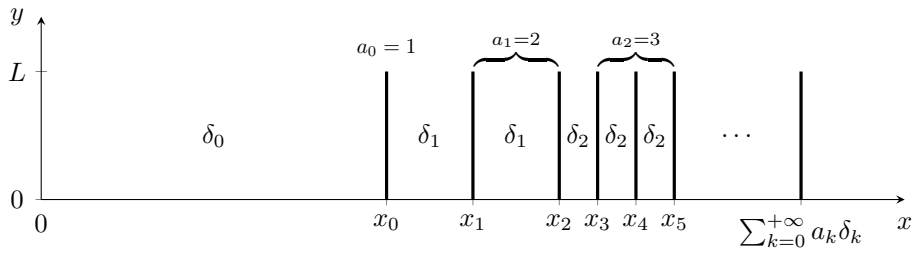


Figure 3.1: Schematic picture of the set C from the proof of Proposition 3.2.3.

and x-th component of its centroid is $\frac{1}{2}\big(\sum_{k=0}^{n(\varepsilon)-1} a_k\delta_k + \delta_{n(\varepsilon)} + \sum_{k=0}^{+\infty} a_k\delta_k\big)$.

The compact set C for which we are going to show the non-existence of the limit (3.2) is the union of two parts. The first (positive) part C^+ is formed by above described union of lines for sequences

$$a_k = 2^k \quad \text{and} \quad \delta_k = \frac{1}{4^k}.$$

The second (negative) part C^- is the analogous object mirrored to the negative x-axis ($\tilde{x}_i = -x_i$) with the choice of sequences

$$\tilde{a}_k = 2^k \quad \text{and} \quad \tilde{\delta}_k = \frac{1}{2}\cdot\frac{1}{4^k}.$$

Now let us take the sequence

$$\varepsilon_n = \frac{1}{2}\cdot\frac{1}{4^n} \quad \text{for all } n \in \mathbb{N}.$$

Such choice leads in the positive part C^+ of C to $n(\varepsilon_n) = n+1$ and in the negative part C^- of C to $\tilde{n}(\varepsilon_n) = n$. This gives

$$\int\limits_{C^+_{\varepsilon_n}} x\,\mathrm{d}\boldsymbol{x} = \frac{6L}{2^n} + \mathcal{O}\left(\frac{n}{4^n}\right), \quad \nu_2(C^+_{\varepsilon_n}) = \frac{3L}{2^n} + \mathcal{O}\left(\frac{1}{4^n}\right)$$

as $n \to \infty$ for the positive part and

$$\int\limits_{C^-_{\varepsilon_n}} x\,\mathrm{d}\boldsymbol{x} = -\frac{2L}{2^n} + \mathcal{O}\left(\frac{n}{4^n}\right), \quad \nu_2(C^-_{\varepsilon_n}) = \frac{2L}{2^n} + \mathcal{O}\left(\frac{1}{4^n}\right)$$

for the negative part. Taking all together we get

$$z_1(C_{\varepsilon_n}) = \frac{\frac{6L}{2^n} - \frac{2L}{2^n} + \mathcal{O}\left(\frac{n}{4^n}\right)}{\frac{3L}{2^n} + \frac{2L}{2^n} + \mathcal{O}\left(\frac{1}{4^n}\right)} \xrightarrow{n\to+\infty} \frac{4}{5}$$

for the x-th component.

The second sequence we take is

$$\tilde{\varepsilon}_n = \frac{1}{4^n} \quad \text{for all } n \in \mathbb{N}.$$

This in the positive part of C leads to $n(\tilde{\varepsilon}_n) = n$ and in the negative part of C to $\tilde{n}(\tilde{\varepsilon}_n) = n$. We have

$$\int\limits_{C^+_{\tilde{\varepsilon}_n}} x\,\mathrm{d}\boldsymbol{x} = \frac{8L}{2^n} + \mathcal{O}\left(\frac{n}{4^n}\right), \quad \nu_2(C^+_{\tilde{\varepsilon}_n}) = \frac{4L}{2^n} + \mathcal{O}\left(\frac{1}{4^n}\right)$$

as $n \to \infty$ for the positive part and

$$\int\limits_{C^-_{\tilde{\varepsilon}_n}} x \,\mathrm{d}\vec{x} = -\frac{3L}{2^n} + \mathcal{O}\left(\frac{n}{4^n}\right), \quad \nu_2(C^-_{\tilde{\varepsilon}_n}) = \frac{3L}{2^n} + \mathcal{O}\left(\frac{1}{4^n}\right)$$

for the negative part. Thus for the first component

$$z_1(C_{\tilde{\varepsilon}_n}) = \frac{\frac{8L}{2^n} - \frac{3L}{2^n} + \mathcal{O}\left(\frac{n}{4^n}\right)}{\frac{4L}{2^n} + \frac{3L}{2^n} + \mathcal{O}\left(\frac{1}{4^n}\right)} \xrightarrow{n\to+\infty} \frac{5}{7}.$$

Since $z_1(C_{\varepsilon_n})$ and $z_1(C_{\tilde{\varepsilon}_n})$ lead to different limits, we have constructed the compact set such that the limit (3.2) does not exist and the proof is complete. □

The previous proof can be easily modified to show that for $d > 1$ the generalized centroid may not exist even for connected compact sets.

The last assertion shows that the generalized centroid is measurable. It is easy to see that if $\mathcal{D}$ is some Borel subset of $\mathcal{C}'$ than the trace σ-algebra $\mathcal{B}(\mathcal{D})$ induced by the Borel σ-algebra $\mathcal{B}(\mathcal{C}')$ on $\mathcal{C}'$ is the same as the Borel σ-algebra generated by trace topology on $\mathcal{D}$ induced by the topology on $(\mathcal{C}', \delta)$. For more details see Section A.5.

Theorem 3.2.2. *Let $\mathcal{D}$ be a Borel subset of $\mathcal{C}'$ such that for each $C \in \mathcal{D}$ the generalized centroid of C exists. Then the generalized centroid as a mapping $\boldsymbol{z}_M : \mathcal{D} \to \mathbb{R}^d$ is Borel measurable with respect to $\mathcal{B}(\mathcal{D})$.*

Proof. First note that the generalized centroid can be equivalently defined using closed ε-neighbourhood $\overline{C_\varepsilon} = C + \varepsilon B^d$,

$$\boldsymbol{z}_M(C) = \lim_{\varepsilon \to 0_+} \frac{1}{\nu_d\big(\overline{C_\varepsilon}\big)} \int\limits_{\overline{C_\varepsilon}} \boldsymbol{x} \,\mathrm{d}\boldsymbol{x}.$$

If $\varepsilon > 0$ is fixed then from Theorem A.5.5 (b) follows that $\overline{C_\varepsilon}$ as a function from $\mathcal{D}$ to $\mathcal{C}'$ is continuous and hence Borel measurable. Moreover, from Theorem 3.1.2 we have that $\boldsymbol{z}(\cdot)$ is Borel measurable on $\mathcal{C}'$ with positive Lebesgue measure. Therefore $\boldsymbol{z}(C_\varepsilon) = \boldsymbol{z}(\overline{C_\varepsilon})$ is also Borel measurable. Since $\boldsymbol{z}_M(C)$ is a limit of $\boldsymbol{z}(\overline{C_\varepsilon})$ as $\varepsilon \to 0_+$ the measurability follows. □

3.3 Sufficient condition of existence

In this section we formulate a sufficient condition of existence of the generalized centroid. The formulation falls naturally into two parts. In the first part we observe that the limit defining generalized centroid is the special case of a general formula for weak convergence of probability measures. The second part is devoted to prove the existence of weak limit when a set is Minkowski measurable.

First we start with the weak limit. It is useful to rewrite the integral term of (3.2) as

$$\frac{1}{\nu_d(C_\varepsilon)} \int\limits_{C_\varepsilon} \boldsymbol{x} \,\mathrm{d}\boldsymbol{x} = \int\limits_{\mathbb{R}^d} \boldsymbol{x} \frac{\mathbb{1}_{C_\varepsilon}(\boldsymbol{x})}{\nu_d(C_\varepsilon)} \,\mathrm{d}\boldsymbol{x}.$$

This may be understood as an integral from $\boldsymbol{x}$ with respect to a probability measure $\mathbb{P}_{C,\varepsilon}$ defined by

$$\mathbb{P}_{C,\varepsilon}(A) = \frac{\nu_d\big(A \cap C_\varepsilon\big)}{\nu_d(C_\varepsilon)} \quad \text{for all Borel } A. \tag{3.3}$$

Clearly $\mathbb{P}_{C,\varepsilon}$ is absolutely continuous with respect to ν_d, and its density is given by $\frac{\mathbb{1}_{C_\varepsilon}(\boldsymbol{x})}{\nu_d(C_\varepsilon)}$. If we are able to show the existence of the weak limit $\mathbb{P}_{C,\varepsilon} \xrightarrow{\mathrm{w}} \mathbb{P}_C$ to some Borel probability measure $\mathbb{P}_C$, then the existence of the centroid follows. We formally state the following lemma.

Lemma 3.3.1. *Let C be a non-empty compact subset of $\mathbb{R}^d$. Let further $\mathbb{P}_{C,\varepsilon}$ be probability measures given for each $\varepsilon > 0$ by (3.3) and $\mathbb{P}_C$ be a probability measure such that for every sequence $\{\varepsilon_n\}$ of positive numbers with $\varepsilon_n \to 0_+$, a sequence $\{\mathbb{P}_{C,\varepsilon_n}\}$ converges weakly to $\mathbb{P}_C$, $\mathbb{P}_{C,\varepsilon_n} \xrightarrow{\mathrm{w}} \mathbb{P}_C$ as $n \to \infty$. Then the generalized centroid of C exists and is given by*

$$\boldsymbol{z}_M(C) = \int_{\mathbb{R}^d} \boldsymbol{x}\, \mathbb{P}_C(\mathrm{d}\boldsymbol{x}).$$

Proof. If we denote $\delta = \sup\{\varepsilon_n | n = 1, \dots\}$ then for a closure $\overline{C_\delta}$ holds $C_{\varepsilon_n} \subset \overline{C_\delta}$ for all n. From the construction (3.3) follows that $\mathbb{P}_{C,\varepsilon_n}(\overline{C_\delta}) = 1$. Since $\overline{C_\delta}$ is compact, Theorem A.1.4 (b) yields

$$\limsup_{n\to\infty} \mathbb{P}_{C,\varepsilon_n}\left(\overline{C_\delta}\right) = 1 \leq \mathbb{P}_C\left(\overline{C_\delta}\right) \leq 1.$$

Hence $\mathbb{P}_C(\overline{C_\delta}) = 1$ and $\mathbb{P}_C((\overline{C_\delta})^c) = 0$. Now we take bounded, continuous function f from Theorem A.1.5 such that $f(\boldsymbol{x}) = 1$ for all $\boldsymbol{x} \in \overline{C_\delta}$. The function g given by $g(\boldsymbol{x}) = x_1 f(\boldsymbol{x})$ is therefore also bounded and continuous. From weak continuity follows

$$\lim_{n\to\infty} \int_{\mathbb{R}^d} x_1\, \mathbb{P}_{C,\varepsilon_n}(\mathrm{d}\boldsymbol{x}) = \lim_{n\to\infty} \int_{\mathbb{R}^d} g(\boldsymbol{x})\, \mathbb{P}_{C,\varepsilon_n}(\mathrm{d}\boldsymbol{x}) = \int_{\mathbb{R}^d} g(\boldsymbol{x})\, \mathbb{P}_C(\mathrm{d}\boldsymbol{x}) = \int_{\mathbb{R}^d} x_1\, \mathbb{P}_C(\mathrm{d}\boldsymbol{x}).$$

The same holds for other components of $\boldsymbol{x}$. Since the sequence $\{\varepsilon_n\}$ was arbitrary, the proof is complete. □

Now we are ready to show the connection between the Minkowski measurability and the existence of weak limit. Let us recapitulate that the α-dimensional Minkowski content of a set $C \subset \mathbb{R}^d$ is defined to be

$$\mathcal{M}^\alpha(C) = \limsup_{\varepsilon\to 0_+} \frac{\nu_d(C_\varepsilon)}{c_{d-\alpha}\varepsilon^{d-\alpha}}$$

whenever the limit exists. The set A is called Minkowski measurable if $0 < \mathcal{M}^\alpha(A) < \infty$ for some α which is then called the Minkowski dimension of A. For further definitions and details see Section A.3. Our aim is to prove the following theorem that ensures the existence of weak limit for certain Minkowski measurable sets.

Theorem 3.3.1. *Let C be a non-empty, compact, Minkowski measurable set with Minkowski dimension α and let $\mathcal{M}^\alpha\left((-\infty, \boldsymbol{x}] \cap C\right)$ exist for all $\boldsymbol{x} \in \mathbb{R}^d$. Then there is a probability measure $\mathbb{P}_C$ such that $\mathbb{P}_{C,\varepsilon_n} \xrightarrow{\mathrm{w}} \mathbb{P}_C$ as $n \to \infty$ for every sequence $\{\varepsilon_n\}$ with $\varepsilon \to 0_+$ and $\mathbb{P}_{C,\varepsilon_n}$ given by (3.3).*

We begin by proving certain properties of Minkowski content.

Lemma 3.3.2. *Let C be a non-empty set such that $\mathcal{M}^\alpha\left((-\infty, \boldsymbol{x}] \cap C\right)$ exist for all $\boldsymbol{x} \in \mathbb{R}^d$ and some α. Then*

(a) $\mathcal{M}^\alpha\left((-\infty, \boldsymbol{x}] \cap C\right)$ is non-decreasing in each coordinate of $\boldsymbol{x}$,

(b) $0 \leq \mathcal{M}^\alpha\left((-\infty, \boldsymbol{x}] \cap C\right) \leq \mathcal{M}^\alpha\left(C\right)$ for all $\boldsymbol{x}$,

(c) for each d-dimensional rectangle $(\boldsymbol{a}, \boldsymbol{b}]$ with $\boldsymbol{a} = (a_1, \dots, a_d)$ and $\boldsymbol{b} = (b_1, \dots, b_d)$ holds

$$\sum_{(\theta_1,\dots,\theta_d)\in\{0,1\}^d} (-1)^{\sum \theta_i}\, \mathcal{M}^\alpha\left(\left(-\infty, (a_1 + \theta_1(b_1 - a_1), \dots, a_d + \theta_d(b_d - a_d))\right] \cap C\right) \geq 0.$$

(d) $\mathcal{M}^\alpha\left((-\infty, \boldsymbol{x}] \cap C\right) \to 0$ as any one coordinate of $\boldsymbol{x}$ goes to $-\infty$, and $\mathcal{M}^\alpha\left((-\infty, \boldsymbol{x}] \cap C\right) \to \mathcal{M}^\alpha\left(C\right)$ as all coordinates of $\boldsymbol{x}$ go to $+\infty$.

Proof. (a) Let $\boldsymbol{x} \le \boldsymbol{y}$. Clearly $(-\infty, \boldsymbol{x}] \subset (-\infty, \boldsymbol{y}]$ and therefore $((-\infty, \boldsymbol{x}] \cap C)_\varepsilon \subset ((-\infty, \boldsymbol{y}] \cap C)_\varepsilon$ for every $\varepsilon > 0$. By monotonicity of ν_d,

$$\frac{\nu_d\big(((-\infty, \boldsymbol{x}] \cap C)_\varepsilon\big)}{c_{d-\alpha}\varepsilon^{d-\alpha}} \le \frac{\nu_d\big(((-\infty, \boldsymbol{y}] \cap C)_\varepsilon\big)}{c_{d-\alpha}\varepsilon^{d-\alpha}}.$$

Taking the limit $\varepsilon \to 0_+$ gives

$$\mathcal{M}^\alpha\left((-\infty, \boldsymbol{x}] \cap C\right) \le \mathcal{M}^\alpha\left((-\infty, \boldsymbol{y}] \cap C\right).$$

(b) Let $\varepsilon > 0$. Obviously $\emptyset \subset ((-\infty, \boldsymbol{x}] \cap C)_\varepsilon \subset C_\varepsilon$. By monotonicity of ν_d,

$$0 \le \frac{\nu_d\big(((-\infty, \boldsymbol{x}] \cap C)_\varepsilon\big)}{c_{d-\alpha}\varepsilon^{d-\alpha}} \le \frac{\nu_d(C_\varepsilon)}{c_{d-\alpha}\varepsilon^{d-\alpha}}.$$

Letting $\varepsilon \to 0_+$ yields the desired inequality.

(c) Let $\boldsymbol{a} < \boldsymbol{b}$ and $\varepsilon > 0$. Since ν_d is a measure, it is easy to check that

$$\sum_{(\theta_1,\dots,\theta_d)\in\{0,1\}^d} (-1)^{\sum\theta_i} \nu_d\Big(\Big(\big(-\infty, (a_1+\theta_1(b_1-a_1), \dots, a_d+\theta_d(b_d-a_d))\big] \cap C\Big)_\varepsilon\Big) = \nu_d(A) \ge 0,$$

where

$$A = \Big((-\infty, \boldsymbol{b}] \cap C\Big)_\varepsilon \setminus \bigcup_{\substack{(\theta_1,\dots,\theta_d)\\ \sum\theta_i=1}} \Big(\big(-\infty, \big(a_1+\theta_1(b_1-a_1), \dots, a_d+\theta_d(b_d-a_d)\big)\big] \cap C\Big)_\varepsilon.$$

The result again follows by letting $\varepsilon \to 0_+$.

(d) Since C is a compact set, it is bounded. If any of coordinates of $\boldsymbol{x}$ is smaller than minimum from all coordinates of all points in C we obtain $(-\infty, \boldsymbol{x}] \cap C = \emptyset$. On the other hand, if all coordinates of $\boldsymbol{x}$ are larger than maximum from all coordinates of all points in C then $(-\infty, \boldsymbol{x}] \cap C = C$. This completes the proof. □

For being able to prove the existence of the weak limit we prepare two more assertions from calculus.

Lemma 3.3.3. *Let f be a real function on $\mathbb{R}^d$ that is bounded and non-decreasing in each coordinate. Then for every sequence $\{\boldsymbol{x}_n\}$ with $\boldsymbol{x}_n > \boldsymbol{a}$ for all n and $\boldsymbol{x}_n \to \boldsymbol{a}$ as $n \to \infty$ the limit*

$$\lim_{n\to\infty} f(\boldsymbol{x}_n)$$

exists and its value is independent of the sequence $\{\boldsymbol{x}_n\}$.

Proof. Suppose the assertion is false. In that case there are two sequences $\{\boldsymbol{x}_n\}$ and $\{\boldsymbol{y}_n\}$ such that $\boldsymbol{x}_n, \boldsymbol{y}_n > \boldsymbol{a}$ for all n, $\boldsymbol{x}_n, \boldsymbol{y}_n \to \boldsymbol{a}$ and $f(\boldsymbol{x}_n) \to b$, $f(\boldsymbol{y}_n) \to c$ with $b \ne c$. Without loss of generality we may assume $c < b$. Take $\varepsilon > 0$ such that $c + \varepsilon < b - \varepsilon$. Then there is n_0 such that $f(\boldsymbol{y}_{n_0}) < c + \varepsilon$ and for all $n \ge n_0$, $f(\boldsymbol{x}_n) > b - \varepsilon$. Since $\boldsymbol{a} < \boldsymbol{y}_{n_0}$ and $\boldsymbol{x}_n \to \boldsymbol{a}$, we may find n such that $\boldsymbol{a} < \boldsymbol{x}_n \le \boldsymbol{y}_{n_0}$. But from the monotonicity follows $b - \varepsilon < f(\boldsymbol{x}_n) \le f(\boldsymbol{y}_{n_0}) < c + \varepsilon$, which is a contradiction. □

Lemma 3.3.4. *Let f be a real function bounded and non-decreasing in each coordinate. Let g be defined by*

$$g(\boldsymbol{x}) = \lim_{\delta\to 0+} f(\boldsymbol{x} + \delta).$$

Then g is also bounded, non-decreasing in each coordinate, and continuous from above. Moreover, $f(\boldsymbol{x}) = g(\boldsymbol{x})$ at every continuity point $\boldsymbol{x}$ of f.

Proof. If f is continuous at $\boldsymbol{x}$ than the above limit exists and equals $f(\boldsymbol{x})$. In the following we without loss of generality assume that $0 \leq f \leq 1$ and hence $0 \leq g \leq 1$. Let us prove the monotonicity. Since $\boldsymbol{x}+\delta \leq \boldsymbol{y}+\delta$ we have $f(\boldsymbol{x}+\delta) \leq f(\boldsymbol{y}+\delta)$ and for every sequence $\{\delta_n\}$, $\delta_n \to 0+$ we get $g(\boldsymbol{x}) = \lim f(\boldsymbol{x}+\delta_n) \leq \lim f(\boldsymbol{y}+\delta_n) = g(\boldsymbol{y})$. Now we prove the continuity form above. From monotonicity of $f(\boldsymbol{x}+\delta)$ in δ follows that there is maximally countable number of values of δ such that f is discontinuous at $\boldsymbol{x}+\delta$. Therefore there is always a sequence $\{\delta_n\}$, $\delta_n > 0$ such that $f(\boldsymbol{x}+\delta_n) = g(\boldsymbol{x}+\delta_n)$ for all n and therefore $g(\boldsymbol{x}) = \lim_{n\to\infty} g(\boldsymbol{x}+\delta_n)$. From the previous Lemma 3.3.3 follows that $g(\boldsymbol{x}) = \lim_{\boldsymbol{y}_n \to \boldsymbol{x}+} g(\boldsymbol{y}_n)$ for every sequence $\{\boldsymbol{y}_n\}$ such that $\boldsymbol{y}_n > \boldsymbol{x}$ for all n. Now let $\{\boldsymbol{y}_n\}$ be a sequence such that $\boldsymbol{y}_n \geq \boldsymbol{x}$ for all n. We construct a new sequence $\{\tilde{\boldsymbol{y}}_n\}$ by setting $\tilde{\boldsymbol{y}}_n = \boldsymbol{y}_n + 1/n$ for all n. From the construction follows $\tilde{\boldsymbol{y}}_n > \boldsymbol{y}_n \geq \boldsymbol{x}$ and $\tilde{\boldsymbol{y}}_n \to \boldsymbol{x}$. Since g is non-increasing it leads to $g(\tilde{\boldsymbol{y}}_n) \geq g(\boldsymbol{y}_n) \geq g(\boldsymbol{x})$. Finally, because $g(\tilde{\boldsymbol{y}}_n) \to g(\boldsymbol{x})$ we conclude that $g(\boldsymbol{y}_n) \to g(\boldsymbol{x})$. □

Now we have all necessary assertions to prove the main result of this section.

Proof of Theorem 3.3.1. The basic idea of the proof is to construct the probability measure $\mathbb{P}_C$ such that $\mathbb{P}_{C,\varepsilon} \xrightarrow{\text{w}} \mathbb{P}_C$ as $\varepsilon \to 0_+$. First let us construct the distribution functions $F_{C,\varepsilon}$ corresponding to probability measures $\mathbb{P}_{C,\varepsilon}$ by

$$F_{C,\varepsilon}(\boldsymbol{x}) = \mathbb{P}_{C,\varepsilon}\left((-\infty,\boldsymbol{x}]\right) = \frac{\nu_d\big((-\infty,\boldsymbol{x}]\cap C_\varepsilon\big)}{\nu_d(C_\varepsilon)}.$$

The upper and lower set bounds for the intersection in the numerator are

$$\big((-\infty,\boldsymbol{x}-\varepsilon]\cap C\big)_\varepsilon \subset \big((-\infty,\boldsymbol{x}]\cap C_\varepsilon\big) \subset \big((-\infty,\boldsymbol{x}+\varepsilon]\cap C\big)_\varepsilon,$$

where $\boldsymbol{x}\pm\varepsilon = (x_1\pm\varepsilon,\ldots,x_d\pm\varepsilon) \in \mathbb{R}^d$. Let take a fixed, arbitrarily small $\delta > 0$ and assume $\epsilon < \delta$. We can enlarge the previous bounds by

$$\big((-\infty,\boldsymbol{x}-\delta]\cap C\big)_\varepsilon \subset \big((-\infty,\boldsymbol{x}]\cap C_\varepsilon\big) \subset \big((-\infty,\boldsymbol{x}+\delta]\cap C\big)_\varepsilon.$$

For the distribution functions $F_{C,\varepsilon}$ we therefore have

$$\frac{\nu_d\big(((-\infty,\boldsymbol{x}-\delta]\cap C)_\varepsilon\big)}{\nu_d(C_\varepsilon)} \leq F_{C,\varepsilon}(\boldsymbol{x}) \leq \frac{\nu_d\big(((-\infty,\boldsymbol{x}+\delta]\cap C)_\varepsilon\big)}{\nu_d(C_\varepsilon)}.$$

From the assumptions we know that

$$\mathcal{M}^\alpha\big((-\infty,\boldsymbol{y}]\cap C\big) = \lim_{\varepsilon\to 0_+} \frac{\nu_d\big(((-\infty,\boldsymbol{y}]\cap C)_\varepsilon\big)}{c_{d-\alpha}\varepsilon^{d-\alpha}} < +\infty,$$
$$0 < \mathcal{M}^\alpha(C) = \lim_{\varepsilon\to 0_+} \frac{\nu_d(C_\varepsilon)}{c_{d-\alpha}\varepsilon^{d-\alpha}} < +\infty,$$

for all $\boldsymbol{y} \in \mathbb{R}^d$. Thus we may rewrite the inequality for $F_{C,\varepsilon}$ as

$$\frac{\nu_d\big(((-\infty,\boldsymbol{x}-\delta]\cap C)_\varepsilon\big)}{c_{d-\alpha}\varepsilon^{d-\alpha}}\frac{c_{d-\alpha}\varepsilon^{d-\alpha}}{\nu_d(C_\varepsilon)} \leq F_{C,\varepsilon}(\boldsymbol{x}) \leq \frac{\nu_d\big(((-\infty,\boldsymbol{x}+\delta]\cap C)_\varepsilon\big)}{c_{d-\alpha}\varepsilon^{d-\alpha}}\frac{c_{d-\alpha}\varepsilon^{d-\alpha}}{\nu_d(C_\varepsilon)}.$$

Taking the lim inf and lim sup we obtain

$$\frac{\mathcal{M}^\alpha\big((-\infty,\boldsymbol{x}-\delta]\cap C\big)}{\mathcal{M}^\alpha(C)} \leq \liminf_{\varepsilon\to 0_+} F_{C,\varepsilon}(\boldsymbol{x}) \leq \limsup_{\varepsilon\to 0_+} F_{C,\varepsilon}(\boldsymbol{x}) \leq \frac{\mathcal{M}^\alpha\big((-\infty,\boldsymbol{x}+\delta]\cap C\big)}{\mathcal{M}^\alpha(C)}$$

Now we can take the limit $\delta \to 0_+$. At every continuity point $\boldsymbol{x}$ of $\mathcal{M}^\alpha\big((-\infty,\boldsymbol{x}]\cap C\big)$ we obtain

$$\lim_{\varepsilon\to 0_+} F_{C,\varepsilon}(\boldsymbol{x}) = \frac{\mathcal{M}^\alpha\big((-\infty,\boldsymbol{x}]\cap C\big)}{\mathcal{M}^\alpha(C)}.$$

If we define $F_C(\boldsymbol{x})$ by

$$F_C(\boldsymbol{x}) = \lim_{\delta \to 0_+} \frac{\mathcal{M}^\alpha\big((-\infty, \boldsymbol{x}+\delta] \cap C\big)}{\mathcal{M}^\alpha(C)}$$

we have

$$\lim_{\varepsilon \to 0_+} F_{C,\varepsilon}(\boldsymbol{x}) = F_C(\boldsymbol{x}) \quad \text{at all continuity points of } F_C(\boldsymbol{x})$$

since continuity points of $F_C(\boldsymbol{x})$ are the same as continuity points of $\mathcal{M}^\alpha\big((-\infty, \boldsymbol{x}] \cap C\big)$. According to Lemma 3.3.2 (a), (b) and Lemma 3.3.4, the function $F_C(\boldsymbol{x})$ is non-decreasing and continuous from above. From Lemma 3.3.2 (d) follows that $F_C(\boldsymbol{x}) \to 0$ as any one coordinate of $\boldsymbol{x}$ goes to $-\infty$, and $F_C(\boldsymbol{x}) \to 1$ as all coordinates of $\boldsymbol{x}$ go to ∞. Finally, the inequality in Lemma 3.3.2 (c) holds also for $(\boldsymbol{a}+\delta, \boldsymbol{b}+\delta]$ and taking the limit $\delta \to 0$ it consequently remains valid for $F_C(\boldsymbol{x})$. All conditions of Lemma A.1.1 are therefore satisfied. Hence $F_C(\boldsymbol{x})$ is a distribution function and there is a probability measure $\mathbb{P}_C$ such that $F_C(\boldsymbol{x})$ is its distribution function. Taking a sequence $\{\varepsilon_n\}$ such that $\varepsilon_n \to 0$ we conclude that $\mathbb{P}_{C,\varepsilon_n} \xrightarrow{\text{w}} \mathbb{P}_C$ as a consequence of Theorem A.1.6. □

Thus taking together Theorem 3.3.1 and Lemma 3.3.1 we have proven the following statement.

Corollary 3.3.1. *Let C be a non-empty, compact, Minkowski measurable set with Minkowski dimension α such that $\mathcal{M}^\alpha\left((-\infty, \boldsymbol{x}] \cap C\right)$ exists for all $\boldsymbol{x} \in \mathbb{R}^d$. Then the generalized centroid $\boldsymbol{z}(C)$ of C exists.*

3.4 Rectifiable sets and submanifolds

In the previous section we derived a sufficient condition for the existence of generalized centroid formulated with the help of Minkowski content. An important subclass of Minkowski measurable sets is given by certain rectifiable sets defined in A.3.2. First we recapitulate some known results about the connection of Hausdorff measure and Minkowski content.

Unless otherwise stated m is assumed to be a positive integer throughout this section. The following theorem shows that for m-rectifiable sets the m-dimensional Minkowski content exists and is equal to the m-dimensional Hausdorff measure.

Theorem 3.4.1. *If F is a closed m-rectifiable subset of $\mathbb{R}^n$, then $\mathcal{M}^m(F) = \mathcal{H}^m(F)$.*

Proof. Theorem 3.2.39 in [65]. □

Since every m-rectifiable set is a subset of an image of some bounded subset of $\mathbb{R}^m$ under Lipschitz function and $\nu_m = \mathcal{H}^m$ as a result of Theorem A.3.1, the following corollary follows from Theorem A.3.2.

Corollary 3.4.1. *Under the assumptions of previous theorem $\mathcal{M}^m(F) < \infty$.*

It is easy to prove that in such case the generalized centroid exists.

Corollary 3.4.2. *If C is a compact m-rectifiable subset of $\mathbb{R}^n$ such that $0 < \mathcal{H}^m(C)$, then the generalized centroid of C exists and is given by*

$$\boldsymbol{z}_M(C) = \frac{1}{\mathcal{H}^m(C)} \int_C \boldsymbol{x}\, \mathcal{H}^m(\mathrm{d}\boldsymbol{x}).$$

Proof. The previous theorem and its corollary yield $0 < \mathcal{M}^m(C) = \mathcal{H}^m(C) < \infty$. Moreover, $\mathcal{M}^m(D) = \mathcal{H}^m(D) < \infty$ for any compact subset D of C and particularly for $D = (-\infty, \boldsymbol{x}] \cap C$, where $\boldsymbol{x} \in \mathbb{R}^d$ is arbitrary. Thus $\mathcal{M}^m\left((-\infty, \boldsymbol{x}] \cap C\right) = \mathcal{H}^m\left((-\infty, \boldsymbol{x}] \cap C\right) \leq \mathcal{H}^m(C) < \infty$. Hence from Theorem 3.3.1 and Lemma 3.3.1 the generalized centroid of C exists. Moreover, because $\mathcal{H}^m$ taken on Borel sets is a measure, it follows that $\mathcal{M}^m\left((-\infty, \boldsymbol{x}] \cap C\right) = \mathcal{H}^m\left((-\infty, \boldsymbol{x}] \cap C\right)$ is continuous from above. We denote by $\mathbb{P}(C)$ the probability measure such that $\mathbb{P}_{C,\varepsilon} \xrightarrow{\text{w}} \mathbb{P}(C)$ and by F_C its distribution function. From the proof of Theorem 3.3.1 we have

$$F_C(\boldsymbol{x}) = \lim_{\delta \to 0_+} \frac{\mathcal{M}^m\left((-\infty, \boldsymbol{x} + \delta] \cap C\right)}{\mathcal{M}^m(C)} = \frac{\mathcal{H}^m\left((-\infty, \boldsymbol{x}] \cap C\right)}{\mathcal{H}^m(C)} \quad \text{for all } \boldsymbol{x} \in \mathbb{R}^d.$$

The probability measure $\mathbb{P}_C$ is therefore given by

$$\mathbb{P}_C(A) = \frac{\mathcal{H}^m(A \cap C)}{\mathcal{H}^m(C)} \quad \text{for all Borel } A.$$

This using Lemma 3.3.1 proves the second part of the assertion. □

It is easy to see that every set form the collection $\mathcal{K}'$ of non-empty compact convex sets is either an isolated point for which the generalized centroid clearly exists or an m-rectifiable set for some positive integer m. We may therefore formulate the existence particularly for sets in $\mathcal{K}'$.

Corollary 3.4.3. *The generalized centroid exists for all non-empty, compact, convex sets.*

From Theorem 3.2.2 we know that $\boldsymbol{z}_M$ is measurable on $\mathcal{K}'$. The natural question is whether $\boldsymbol{z}_M$ is continuous on $\mathcal{K}'$. The answer is negative. To see this let take isosceles triangles with equal height and length ε of the third side. For $\varepsilon > 0$ the centroid is located in the $2/3$ of the height. For $\varepsilon = 0$ the triangle collapses into a line having the centroid in the middle which proves the discontinuity.

The common extension of m-rectifiable sets is given by $(\mathcal{H}^m, m)$-rectifiable sets (see Section A.3 for a definition). The previous result, however, does not hold for $(\mathcal{H}^m, m)$-rectifiable sets. It cannot be directly extended to countable union of m-rectifiable sets even if its Hausdorff measure is finite. The counterexample of a $(\mathcal{H}^2, 2)$-rectifiable set with infinite Minkowski content is discussed in [65, 3.2.40]. However, it is still on open question if the generalized centroid of such set exists, since the conditions in Theorem 3.3.1 are only sufficient.

Let $U_r(\boldsymbol{x})$ resp. $B_r(\boldsymbol{x})$ be the open resp. closed ball in $\mathbb{R}^d$ with centre $\boldsymbol{x}$ and radius r. One of the most general results recently available is the following.

Theorem 3.4.2. *Let C be a countably $(\mathcal{H}^m, m)$-rectifiable compact set and assume that η is a Radon measure in $\mathbb{R}^d$ absolutely continuous with respect to $\mathcal{H}^m$ such that for some $\gamma > 0$, for all $\boldsymbol{x} \in C$ and for all $r, 0 < r < 1$,*

$$\eta\big(U_r(\boldsymbol{x})\big) \geq \gamma r^m.$$

Then $\mathcal{M}^m(F) = \mathcal{H}^m(F)$.

Proof. Theorem 2.104 in [66]. □

The generalized centroid therefore exists for sets satisfying conditions of the previous theorem. In the sequel we show that the condition is satisfied for compact subsets of C^1 submanifolds. First note that for $m \geq 1$ there is an equivalent formulation of Definition A.3.3 of m-dimensional C^1 submanifold of $\mathbb{R}^d$.

Proposition 3.4.1. *Let $m \geq 1$ be an integer. Then M is an m-dimensional C^1 submanifold of $\mathbb{R}^d$ if and only if for each $\boldsymbol{x} \in M$ there exists a neighbourhood T of $\boldsymbol{x}$ in $\mathbb{R}^d$, a convex open subset V of $\mathbb{R}^m$, and C^1 maps $\phi : T \to V, \psi : V \to T$ such that*

$$M \cap T = \psi(V), \quad \phi \circ \psi = \mathrm{id}_V.$$

Proof. 3.1.19 in [65]. □

Now we prepare an important result that will be used later.

Lemma 3.4.1. *Let m be a positive integer and C be a compact subset of m-dimensional C^1 submanifold M of $\mathbb{R}^d$. Then there are constants $r, K, L > 0$ such that for each $\boldsymbol{x} \in C$ there exist a neighbourhood T of $\boldsymbol{x}$ in $\mathbb{R}^d$, a convex open subset V of $\mathbb{R}^m$, and maps $\phi : T \to V, \psi : V \to T$ having the following properties:*

(a) ϕ, ψ are C^1,

(b) $M \cap T = \psi(V), \quad \phi \circ \psi = \mathrm{id}_V$,

(c) T contains a closed ball with centre $\boldsymbol{x}$ and radius r, $B_r(\boldsymbol{x}) \subset T$,

(d) ϕ restricted to $B_r(\boldsymbol{x})$ is Lipschitzian with Lipschitz constant not greater than K,

(e) ψ restricted to $\phi\big(B_r(\boldsymbol{x})\big)$ is Lipschitzian with Lipschitz constant not greater than L.

Proof. Let $\{T_{\boldsymbol{x}} | \boldsymbol{x} \in M\}$ be a collection of all neighbourhoods of all points in C with the properties from Proposition 3.4.1. Since every $T_{\boldsymbol{x}}$ is an open set, it contains with $\boldsymbol{x}$ a closed ball with radius $2r_{\boldsymbol{x}}$ for some $r_{\boldsymbol{x}}$, $B_{2r_{\boldsymbol{x}}}(\boldsymbol{x}) \subset T_{\boldsymbol{x}}$. The collection $\{U_{r_{\boldsymbol{x}}}(\boldsymbol{x}) | \boldsymbol{x} \in C\}$ is a covering of C. By compactness of C there is a finite subcovering $\{U_{r_i}(\boldsymbol{x}_i) | i = 1, \ldots, n\}$ where $r_i = r_{\boldsymbol{x}_i}$. For $i = 1, \ldots, n$ we denote $T_i = T_{\boldsymbol{x}_i}$. It follows that $\{T_i | i = 1, \ldots, n\}$ is also a finite covering of C. Let $r = \min_{1 \leq i \leq n} r_i$. This gives $B_{2r}(\boldsymbol{x}_i) \subset B_{2r_i}(\boldsymbol{x}_i) \subset T_i$ for all i. For each i we denote by V_i and ϕ_i, ψ_i the open convex set and maps corresponding to T_j and $\boldsymbol{x}_j$ by Proposition 3.4.1.

Now consider arbitrary $\boldsymbol{y} \in C$. It is possible to choose $j, 1 \leq j \leq n$, such that $\boldsymbol{y} \in U_{r_j}(\boldsymbol{x}_j) \subset B_{2r_j(\boldsymbol{x}_j)} \subset T_j$. Hence T_j contains a whole closed ball with centre $\boldsymbol{y}$ and radius r, i.e. $B_r(\boldsymbol{y}) \subset B_{2r_j}(\boldsymbol{x}_j) \subset T_j$. If we take $T = T_j, V = V_j, \phi = \phi_j, \psi = \psi_j$ the proof of parts (a), (b), (c) is complete.

To prove (d) and (e) let take arbitrary $0 \leq i \leq n$. Since the norm of the derivative $D\phi_j$ is a continuous map on a compact set $B_{2r_j}(\boldsymbol{x}_j)$ it is bounded. Hence ϕ_j restricted to $B_{2r_j}(\boldsymbol{x}_j)$ is Lipschitzian with Lipschitz constant denoted by K_j. The set $\phi_j\big(B_{2r_j}(\boldsymbol{x}_j)\big)$ is compact and consequently ψ_j restricted to $\phi_j\big(B_{2r_j}(\boldsymbol{x}_j)\big)$ is analogously Lipschitzian with Lipschitz constant denoted by L_i. If we set $K = \min_{1 \leq i \leq n} K_i$ and $L = \max_{1 \leq i \leq n} L_i$, the assertions (d) and (e) clearly follows since for every $\boldsymbol{y}$ and respective j is ϕ_j Lipschitzian also on $B_r(\boldsymbol{y}) \subset B_{2r_j}(\boldsymbol{x}_j)$ with Lipschitz constant smaller or equal than L and the same for ψ_j. □

In the following the area formula for Lipschitzian (C^1) maps is needed.

Theorem 3.4.3. *Suppose $f : \mathbb{R}^m \to \mathbb{R}^d$ is Lipschitzian with $m \leq d$. Then if A is a Borel set in $\mathbb{R}^m$, then*

$$\int_A \boldsymbol{J}_m f(\boldsymbol{x}) \nu_m(\mathrm{d}\boldsymbol{x}) = \int_{\mathbb{R}^d} \mathcal{H}^0(A \cap f^{-1}(\boldsymbol{y}))\, \mathcal{H}^m(\mathrm{d}\boldsymbol{y}).$$

Proof. Theorem 3.2.3 in [65] or Theorem 5.1.1 in [67]. The version for C^1 map is Theorem 5.1.9 in [67]. □

Note that $\mathcal{H}^0\left(A \cap f^{-1}(\boldsymbol{y})\right)$ returns the number of points of the set $A \cap f^{-1}(\boldsymbol{y})$. The $\boldsymbol{J}_m f(\boldsymbol{x})$ is the m-dimensional Jacobian defined in the usual way for instance in [65, 3.2.1].

Theorem 3.4.4. *Let C be a compact subset of an m-dimensional C^1 submanifold M of $\mathbb{R}^d$ such that $\mathcal{H}^m(M) < \infty$. Then there is a number $\gamma > 0$ and a Radon measure η in $\mathbb{R}^d$ absolutely continuous with respect to $\mathcal{H}^m$ such that for all $\boldsymbol{x} \in C$ and for all $0 < \rho < 1$*

$$\eta\big(U_\rho(\boldsymbol{x})\big) \geq \gamma\rho^m.$$

Proof. If $m = 0$ then M has a finite number of points and the theorem clearly holds for $\eta = \mathcal{H}^0 \lfloor_M$, where $\mathcal{H}^0 \lfloor_M$ is the restriction of a 0-dimensional Hausdorff measure $\mathcal{H}^0$ to M. Let $m > 0$ and $\boldsymbol{x}$ be an arbitrary point in C. From Lemma 3.4.1 and Proposition 3.4.1 there exist $r > 0$, constants $K, L > 0$, an open neighbourhood $T \subset \mathbb{R}^d$ such that $B_r(\boldsymbol{x}) \subset T$, a convex open subset V of $\mathbb{R}^m$, and C^1 maps $\phi : T \to V, \psi : V \to T$ such that

$$M \cap T = \psi(V), \quad \phi \circ \psi = \mathrm{id}_V.$$

The set $\phi\big(B_r(\boldsymbol{x})\big)$ is a compact subset of V. From Lemma 3.4.1 (e) follows that ψ restricted to $\phi\big(B_r(\boldsymbol{x})\big)$ is Lipschitzian with Lipschitz constant not greater than L. Let define $f : \mathbb{R}^m \to \mathbb{R}^d$ as a Lipschitzian extension (due to Kirszbraun's theorem e.g. 2.10.43 in [65]) of $\psi|_{\phi\left(B_r(\boldsymbol{x})\right)}$ such that the Lipschitz constant of f is the same as the Lispchitz constant of $\psi|_{\phi\left(B_r(\boldsymbol{x})\right)}$ and thus not greater than L.

Let $0 < s < r$. Therefore $U_s(\boldsymbol{x}) \subset U_r(\boldsymbol{x})$ and $\phi\big(B_s(\boldsymbol{x})\big) \subset \phi\big(B_r(\boldsymbol{x})\big)$. We may apply the area formula from Theorem 3.4.3 for $A = \phi\big(M \cap U_s(\boldsymbol{x})\big)$ and f. Since f on A is one-to-one then $A \cap f^{-1}(\boldsymbol{y})$ is exactly one point in A if $\boldsymbol{y} \in M \cap U_s(\boldsymbol{x})$ and because it is Lipschitzian then $A \cap f^{-1}(\boldsymbol{y})$ is an empty set if $\boldsymbol{y} \notin M \cap U_s(\boldsymbol{x})$. Therefore we have

$$\mathcal{H}^m\left(M \cap U_s(\boldsymbol{x})\right) = \int_{\mathbb{R}^d} \mathcal{H}^0\left(\phi\big(M \cap U_s(\boldsymbol{x})\big) \cap f^{-1}(\boldsymbol{y})\right) \mathcal{H}^m(\mathrm{d}\boldsymbol{y}) = \int_{\phi(M \cap U_s(\boldsymbol{x}))} \boldsymbol{J}_m f(\boldsymbol{z}) \nu_m(\mathrm{d}\boldsymbol{z}).$$

Since f is Lipschitz with Lipschitz constant not greater then L we have

$$\|f(\phi(\boldsymbol{x})) - f(\boldsymbol{y})\| \leq L \|\phi(\boldsymbol{x}) - \boldsymbol{y}\|, \quad \text{for all } \boldsymbol{y} \in \mathbb{R}^m.$$

This gives that if $\boldsymbol{y} \in U_{s/L}(\phi(\boldsymbol{x}))$ then $f(\boldsymbol{y}) \in U_s(\boldsymbol{x})$ and so $\boldsymbol{y} \in \phi(U_s(\boldsymbol{x})) \subset \phi\big(B_r(\boldsymbol{x})\big) \subset V$. Hence $f(\boldsymbol{y}) = \psi(\boldsymbol{y}) \in M$, which follows from the fact that $\psi(V) = M \cap T$, and finally $\boldsymbol{y} \in \phi(M \cap U_s(\boldsymbol{x}))$ because $\boldsymbol{y} = \phi(\psi(\boldsymbol{y}))$. Therefore $U_{s/L}(\phi(\boldsymbol{x})) \subset \phi(M \cap U_s(\boldsymbol{x}))$ and

$$\int_{\phi(M \cap U_s(\boldsymbol{x}))} \boldsymbol{J}_m f(\boldsymbol{z}) \nu_m(\mathrm{d}\boldsymbol{z}) \geq \int_{U_{s/L}(\phi(\boldsymbol{x}))} \boldsymbol{J}_m f(\boldsymbol{z}) \nu_m(\mathrm{d}\boldsymbol{z}).$$

The map ϕ is on $\phi(M \cap U_s(\boldsymbol{x})) \subset \phi\big(B_r(\boldsymbol{x})\big)$ Lipschitzian with Lipschitz constant smaller or equal than K. Thus

$$\|\phi(f(\boldsymbol{z})) - \phi(f(\boldsymbol{y}))\| \leq K \|f(\boldsymbol{z}) - f(\boldsymbol{y})\|, \quad \text{for all } \boldsymbol{z}, \boldsymbol{y} \in \phi(M \cap U_s(\boldsymbol{x})).$$

Since for all $\boldsymbol{y}$ in V we have $\phi(\psi(\boldsymbol{y})) = \boldsymbol{y}$, it holds particularly for all $\boldsymbol{y} \in \phi(M \cap U_s(\boldsymbol{x})) \subset V$ where $\psi(\boldsymbol{y}) = f(\boldsymbol{y})$. Therefore

$$\|f(\boldsymbol{z}) - f(\boldsymbol{y})\| \geq \frac{1}{K} \|\boldsymbol{z} - \boldsymbol{y}\|, \quad \text{for all } \boldsymbol{z}, \boldsymbol{y} \in \phi(M \cap U_s(\boldsymbol{x})).$$

From this follows that $\boldsymbol{J}_m f(\boldsymbol{x}) \geq 1/K^m$ for all $\boldsymbol{x} \in U_{s/L}(\phi(\boldsymbol{x}))$ and we obtain

$$\int_{U_{s/L}(\phi(\boldsymbol{x}))} \boldsymbol{J}_m f(\boldsymbol{z}) \nu_m(\mathrm{d}\boldsymbol{z}) \geq \frac{c_m}{L^m K^m} s^m.$$

Finally let us take $\gamma = \frac{c_m r^m}{L^m K^m}$ and $\eta = \mathcal{H}^m|_C$. Then η is clearly a Radon measure absolutely continuous with respect to $\mathcal{H}^m$ and

$$\eta\big(U_\rho(\boldsymbol{x})\big) = \mathcal{H}^m(M \cap U_\rho(\boldsymbol{x})) \geq \gamma\rho^m$$

for all $\boldsymbol{x} \in C$ and $\rho \in (0,1)$. This completes the proof. □

Corollary 3.4.4. *Let m be a non-negative integer and C be a compact subset of an m-dimensional C^1 submanifold M of $\mathbb{R}^d$ such that $0 < \mathcal{H}^m(C) < \infty$. Then the generalized centroid of C exists and is given by*

$$\boldsymbol{z}_M(C) = \frac{1}{\mathcal{H}^m(C)} \int_C \boldsymbol{x}\, \mathcal{H}^m(\mathrm{d}\boldsymbol{x}).$$

Proof. First assume $m > 0$. Then from Theorem A.3.3 follows that C is $(\mathcal{H}^m, m)$-rectifiable. This together with the previous theorem gives that C satisfies all conditions of Theorem 3.4.2. Therefore $0 < \mathcal{M}^m(C) = \mathcal{H}^m(C) < \infty$. The measure η corresponding to C from the previous theorem can be used for all subsets of C and particularly for $(-\infty, \boldsymbol{x}] \cap C$ for all $\boldsymbol{x} \in \mathbb{R}^d$. Theorem 3.4.2 yields $\mathcal{M}^m\left((-\infty, \boldsymbol{x}] \cap C\right) = \mathcal{H}^m\left((-\infty, \boldsymbol{x}] \cap C\right) \leq \mathcal{H}^m(C) < \infty$. In the case $m = 0$ we know from discussion after Definition A.3.3 that C consists only of finite number of isolated points. Thus clearly $\mathcal{M}^0\left((-\infty, \boldsymbol{x}] \cap C\right) = \mathcal{H}^0\left((-\infty, \boldsymbol{x}] \cap C\right) \leq \mathcal{H}^0(C) < \infty$ for all $\boldsymbol{x} \in \mathbb{R}^d$. Hence in both cases we may use Theorem 3.3.1 and Lemma 3.3.1 yielding that the generalized centroid of C exists. The proof of the second part is the same as proof of Corollary 3.4.2. □

It is worth noting that Theorem 3.4.4 and the previous corollary, formulated for compact subsets of C^1 submanifolds, cover many important sets that are not submanifolds as a whole. In particular it is valid for some submanifolds with boundary. For example the closed unit ball B^d is not a d-dimensional C^1 submanifold since for points on the boundary of B^d one cannot satisfy the Definition A.3.3. However, it is a closed subset of an open ball with radius larger than 1, $B^d \subset \delta U^d$ with $\delta > 1$, which is a d-dimensional C^1 submanifold. Hence the generalized centroid of B^d exists which also follows from Proposition 3.2.2.

The C^1 differentiability can be actually weaken to Lipschitz continuity. If we define m-dimensional Lipschitzian submanifold analogously to Definition A.3.3 assuming that ϕ and its inverse are Lipschitzian instead of C^1 then we can prove analogous proposition as Proposition 3.4.1 and as a consequence also the previous theorem and its corollary will be still valid.

Corollary 3.4.5. *Let C be a compact subset of an m-dimensional Lipschitzian submanifold N of $\mathbb{R}^d$ such that $0 < \mathcal{H}^m(C) < \infty$. Then the generalized centroid of C exists and is given by*

$$\boldsymbol{z}_M(C) = \frac{1}{\mathcal{H}^m(C)} \int_C \boldsymbol{x}\, \mathcal{H}^m(\,\mathrm{d}\boldsymbol{x}).$$

Finally we may easily show that the generalized centroid exists for finite unions of previously taken types of sets.

Theorem 3.4.5. *Let $k \in \mathbb{N}$, $C_1, \ldots, C_k$ be compact sets in $\mathbb{R}^d$ and $0 \leq \alpha_1, \ldots, \alpha_k \leq d$ be such that for each $i = 1, \ldots, k$ we have $0 < \mathcal{M}^{\alpha_i}(C_i) = \mathcal{H}^{\alpha_i}(C_i) < \infty$ and $\mathcal{M}^{\alpha_i}(D) = \mathcal{H}^{\alpha_i}(D)$ holds for every compact $D \subset C_i$. Then the generalized centroid of $C = \bigcup_{i=1}^k C_i$ exists and is given by*

$$\boldsymbol{z}_M(C) = \frac{1}{\mathcal{H}^\beta(C)} \int_C \boldsymbol{x}\, \mathcal{H}^\beta(\,\mathrm{d}\boldsymbol{x}), \tag{3.4}$$

where $\beta = \max\{\alpha_1, \ldots, \alpha_k\} = \dim_H C$.

Proof. It is enough to prove that $0 < \mathcal{M}^\beta(C) = \mathcal{H}^\beta(C) < \infty$ and $\mathcal{M}^\beta\left((-\infty, \boldsymbol{x}] \cap C\right) = \mathcal{H}^\beta\left((-\infty, \boldsymbol{x}] \cap C\right)$ for all $\boldsymbol{x} \in \mathbb{R}^d$. Then from Theorem 3.3.1 and Lemma 3.3.1 the generalized centroid of C exists. The second part follows in the same way as in the proof of Corollary 3.4.2.

From the assumption follows that each C_i is an α_i-set and in particular that $\dim_H C_i = \alpha_i$. By the countable stability (see e.g. [25, Section 2.2]) of the Hausdorff dimension we get $\dim_H C = \max\{\alpha_1, \ldots, \alpha_k\}$. Thus if we set $\beta = \dim_H C$ the relation $0 < \mathcal{H}^\beta(C) < \infty$ clearly follows since $\mathcal{H}^\beta(A) = 0$ whenever $\beta > \dim_H A$ and there exists i with $\beta = \alpha_i$ yielding together $0 < \mathcal{H}^\beta(C_i) \leq \mathcal{H}^\beta(C) \leq \mathcal{H}^\beta(C_1) + \ldots + \mathcal{H}^\beta(C_k) < \infty$. Next we prove $\mathcal{M}^\beta\left((-\infty, \boldsymbol{x}] \cap C\right) = \mathcal{H}^\beta\left((-\infty, \boldsymbol{x}] \cap C\right)$ form which, by taking $\boldsymbol{x}$ sufficiently large, follows $\mathcal{M}^\beta(C) = \mathcal{H}^\beta(C)$. We need to show

$$\lim_{\varepsilon \to 0_+} \frac{\nu_d\big(((-\infty, \boldsymbol{x}] \cap C)_\varepsilon\big)}{c_{d-\beta}\varepsilon^{d-\beta}} = \mathcal{H}^\beta\left((-\infty, \boldsymbol{x}] \cap C\right).$$

Let us denote $D_i = (-\infty, \boldsymbol{x}] \cap C_i \subset C_i$ and $D_{i;\varepsilon} = (D_i)_\varepsilon$ for all $i = 1, \ldots, k$. By the well known inclusion-exclusion principle

$$\nu_d\big(((-\infty, \boldsymbol{x}] \cap C)_\varepsilon\big) = \nu_d\left(\bigcup_{i=1}^k D_{i;\varepsilon}\right) = \sum_{i=1}^k (-1)^{i+1} \sum_{1 \leq m_1 < \ldots < m_i \leq k} \nu_d(D_{m_1;\varepsilon} \cap \ldots \cap D_{m_i;\varepsilon}).$$

Since $D_{m_1;\varepsilon} \cap \ldots \cap D_{m_i;\varepsilon} = (D_{m_1} \cap \ldots \cap D_{m_i})_\varepsilon$ and $D_{m_i} \subset C_{m_i}$ we obtain

$$\lim_{\varepsilon \to 0_+} \frac{\nu_d\big((D_{m_1} \cap \ldots \cap D_{m_i})_\varepsilon\big)}{c_{d-\beta}\varepsilon^{d-\beta}} = \mathcal{H}^\beta\left(D_{m_1} \cap \ldots \cap D_{m_i}\right) < \infty$$

for every $i = 1, \ldots, k$ and $1 \leq m_1 < \ldots < m_i \leq k$. Using the inclusion-exclusion principle backwards for $\mathcal{H}^\beta$ yields the result. □

Corollary 3.4.6. *Let k be a positive integer and compact sets $C_1, \ldots, C_k \subset \mathbb{R}^d$ be of the following type: for each $i = 1, \ldots, k$ there is an integer $0 \leq m_i \leq d$ such that C_i is an m_i-set and subset of either m_i-dimensional C^1 submanifold, m_i-dimensional Lipschitzian submanifold or m_i-rectifiable set. Then the generalized centroid of the union $C = \bigcup_{i=1}^k C_i$ exists and is given by*

$$\boldsymbol{z}_M(C) = \frac{1}{\mathcal{H}^n(C)} \int_C \boldsymbol{x}\, \mathcal{H}^n(\,\mathrm{d}\boldsymbol{x}),$$

where $n = \max\{m_1, \ldots, m_k\} = \dim_H C$.

Proof. Clearly follows from the previous theorem, proofs of Corollaries 3.4.2, 3.4.4, and Corollary 3.4.5. □

Corollary 3.4.7. *The generalized centroid exists for any set from the convex ring $\mathcal{R}'$, whose elements are the finite unions of non-empty compact convex sets.*

Since by Theorem A.5.11 is $\mathcal{R}'$ a Borel subset of $\mathcal{C}'$ we can, as a consequence of Theorem 3.2.2, use $\boldsymbol{z}_M$ as a centre function for particle processes in $\mathcal{R}'$ introduced in Section 2.7.

3.5 Generalized centroid based on the Hausdorff measure

Apart from Definition 3.2.1 there is another possibility how to extend the classical definition (3.1) of the centroid. It is based on the expression (3.4). Recall that an α-set is defined (see Appendix A.3) as a set $A \subset \mathbb{R}^d$ such that $\dim_H(A) = \alpha$ and $0 < \mathcal{H}^\alpha(A) < \infty$.

Definition 3.5.1. Let C be a non-empty compact α-set for some $0 \leq \alpha \leq d$. Then we define

$$\boldsymbol{z}_H(C) = \frac{1}{\mathcal{H}^\alpha(C)} \int_C \boldsymbol{x}\, \mathcal{H}^\alpha(\mathrm{d}\boldsymbol{x}) \tag{3.5}$$

and call it the **generalized Hausdorff centroid** of C.

First we show the measurability of the Hausdorff centroid on certain Borel measurable subsets of $\mathcal{C}'$. Let $\tau_\delta(\mathcal{C}')$ denote the topology induced by the Hausdorff metric δ on $\mathcal{C}'$.

Theorem 3.5.1. *Let $\alpha \in [0, d]$ be fixed. Then the class of non-empty compact α-sets $\mathcal{C}'_{\mathcal{H}^\alpha}$ is a Borel subset of $\mathcal{C}'$, and the Hausdorf centroid $\boldsymbol{z}_H$ is measurable on $\mathcal{C}'_{\mathcal{H}^\alpha}$ with respect to the Borel σ-algebra of trace topology induced by the topology $\tau_\delta(\mathcal{C}')$.*

Proof. Clearly $\mathcal{C}'_{\mathcal{H}^\alpha} = (\mathcal{H}^\alpha)^{-1}\big((0,\infty)\big)$. From Theorem A.5.7 follows that $\mathcal{C}'_{\mathcal{H}^\alpha}$ is a Borel set and that the denominator of (3.5) is measurable on $\mathcal{C}'$ and hence on $\mathcal{C}'_{\mathcal{H}^\alpha}$ since the trace σ-algebra $\mathcal{B}_\delta(\mathcal{C}') \cap \mathcal{C}'_{\mathcal{H}^\alpha}$ is the Borel σ-algebra $\mathcal{B}_\delta(\mathcal{C}'_{\mathcal{H}^\alpha})$ of the trace topology. Let us define two measures μ_+ and μ_- on $\mathbb{R}^d$ by

$$\mu_+(A) = \int_A x_1 \mathbb{1}_{\{x_1 \geq 0\}}(\boldsymbol{x})\, \mathcal{H}^\alpha(\mathrm{d}\boldsymbol{x}) \quad \text{and} \quad \mu_-(A) = -\int_A x_1 \mathbb{1}_{\{x_1 < 0\}}(\boldsymbol{x})\, \mathcal{H}^\alpha(\mathrm{d}\boldsymbol{x}).$$

Proposition A.5.1 yields that μ_+ and μ_- are measurable on $\mathcal{C}'_{\mathcal{H}^\alpha}$. The equation (3.5) for the first component of $\boldsymbol{z}_H(C)$ can be now rewritten as

$$z_{H;1}(C) = \frac{\mu_+(C) - \mu_-(C)}{\mathcal{H}^\alpha(C)}.$$

Since all three elements are measurable on $\mathcal{C}'_{\mathcal{H}^\alpha}$, $z_{H;1}$ is also measurable. The same holds for all other components of $\boldsymbol{z}_H$. □

As a consequence of the previous theorem, the generalized Hausdorff centroid can be used as a centre function for particle processes in $\mathcal{C}'_{\mathcal{H}^\alpha}$ (see Section 2.7 for details).

Important models in random geometry are given by rectifiable sets (see Definition A.3.2). The following statement shows the measurability of $\boldsymbol{z}_H$ on particular subclasses of such sets.

Theorem 3.5.2. *Let $0 < m \leq d$ be an integer and $\mathcal{X}_{(m)}$ be a class of $(\mathcal{H}^m, m)$-rectifiable non-empty compact m-sets in $\mathbb{R}^d$. Then $\mathcal{X}_{(m)}$ is a Borel subset of $\mathcal{C}'$ and $\boldsymbol{z}_H$ is measurable on $\mathcal{X}_{(m)}$.*

Proof. By Theorem A.4.7 the collection $\mathcal{X}_m$ of $(\mathcal{H}^m, m)$-rectifiable closed sets in $\mathbb{R}^d$ is a Borel set in $\mathcal{F}$. Since $\mathcal{X}_{(m)} = \mathcal{X}_m \cap \mathcal{C}'$ then by Corollary A.5.2 $\mathcal{X}_{(m)}$ is a Borel set in $\mathcal{C}'$. Let $\mathcal{X}_{(m)}$ be equipped with the Borel σ-algebra $\mathcal{B}_\delta(\mathcal{X}_{(m)})$ of trace topology induced by topology of the Hausdorff metric on $\mathcal{C}'$. Since $\mathcal{X}_{(m)} \subset \mathcal{C}'_{\mathcal{H}^m}$ and the trace σ-algebra $\mathcal{B}_\delta(\mathcal{C}'_{\mathcal{H}^\alpha}) \cap \mathcal{X}_{(m)}$ equals $\mathcal{B}_\delta(\mathcal{X}_{(m)})$, it follows from the previous theorem that $\boldsymbol{z}_H$ is measurable on $\mathcal{X}_{(m)}$. □

Corollary 3.5.1. *Let $\mathcal{X} = \cup_{m=1}^d \mathcal{X}_{(m)}$. Then $\mathcal{X}$ is a Borel subset of $\mathcal{C}'$ and $\boldsymbol{z}_H$ is measurable on $\mathcal{X}$.*

Proof. The Borel property of $\mathcal{X}$ is obvious. Let $\mathcal{X}$ be equipped with the Borel σ-algebra $\mathcal{B}_\delta(\mathcal{X})$ of trace topology induced by topology of the Hausdorff metric on $\mathcal{C}'$ which is equal to the trace σ-algebra $\mathcal{B}_\delta(\mathcal{C}') \cap \mathcal{X}$. For the measurability of $\boldsymbol{z}_H$ note that $\mathcal{X}_{(m)}$ are disjoint, i.e. $\mathcal{X}_{(i)} \cap \mathcal{X}_{(j)} = \emptyset$ for $i \neq j$, as follows from properties of the Hausdorff measure. Let $\boldsymbol{z}_H \mid_{\mathcal{X}}$ denote the restriction of $\boldsymbol{z}_H$ on $\mathcal{X}$. Hence for every open $U \subset \mathbb{R}$ is $\big(\boldsymbol{z}_H \mid_{\mathcal{X}}\big)^{-1}(U) = \bigcup_{m=1}^d \big(\boldsymbol{z}_H \mid_{\mathcal{X}_{(m)}}\big)^{-1}(U) = \bigcup_{m=1}^d C_m$, where $C_m = \big(\boldsymbol{z}_H \mid_{\mathcal{X}_{(m)}}\big)^{-1}(U) \subset \mathcal{X}_{(m)}$ is for each $m = 1, \ldots, d$ a Borel sets in $\mathcal{X}_{(m)}$ by the previous theorem. Since each Borel set in $\mathcal{X}_{(m)}$ is also a Borel set in $\mathcal{X}$, the proof is complete. □

Again the generalized Hausdorff centroid can be used as a centre function for particle processes in $\mathcal{X}$. Now let focus on the relation between the generalized Hausdorff centroid $\boldsymbol{z}_H$ and the generalized Minkowski centroid $\boldsymbol{z}_M$ discussed in previous sections.

Proposition 3.5.1. *The generalized centroids $\boldsymbol{z}_H$ and $\boldsymbol{z}_M$ coincide on compact sets with positive Lebesgue measure and on sets for which $\boldsymbol{z}_M$ exists from Corollary 3.4.6.*

Proof. The first part is a consequence of Proposition 3.2.1 and Theorem A.3.1. The second part is obvious. □

The natural question is which one of the two proposed generalizations of the centroid is more general. Clearly, the generalized Minkowski centroid may exist for sets where the generalized Hausdorff centroid is not defined. As an example consider $C = \{-n^{-1}\}_{n\in\mathbb{N}} \cup \{0\} \cup \{n^{-1}\}_{n\in\mathbb{N}} \subset \mathbb{R}$. Here the Minkowski centroid exists and $z_M(C) = 0$ by Proposition 3.2.2 since C is symmetric around 0. However, clearly $\dim_H C = 0$ and $\mathcal{H}^0(C) = \infty$ and hence C is not a 0-set (and also not an α-set for any other α). Therefore the Hausdorff centroid of C is not defined. On the other side there is an open question whether the generalized Minkowski centroid exists for each compact α-set. Clearly one may think about sets that are α-sets but are not Minkowski measurable. For those sets we cannot apply our weak limit mechanism from Theorem 3.3.1 to prove the existence of the limit. However, the limit (3.2) may still exist.

Chapter 4

Statistics of random structures

In this chapter the statistical issues of various characteristics from stochastic geometry are discussed. The outline is as follows: In the first section we present some standard estimators for point processes. In Section 4.2 we introduce several first and second order estimators for stationary random closed sets and discuss their basic properties. The use of the fast Fourier transform for the estimation is introduced. Section 4.3 is devoted to non-stationary random closed sets. Here the kernel estimator of the volume fraction is introduced. Then the problem of the second order estimation is also discussed. Finally, in Section 4.4 we present some numerical analysis of the performance of correlation function estimators in the stationary case.

4.1 Point processes

4.1.1 Stationary point processes

Here we recapitulate some well known facts from the estimation theory of point processes. The general references are [41, 68]. Let N be a stationary point process in $\mathbb{R}^d$ with intensity λ that is observed in a subset W of $\mathbb{R}^d$ called the observation window. Here $\nu_d(W) > 0$ is always assumed. The straightforward estimator of the intensity is the **empirical intensity** defined by

$$\hat{\lambda} = \frac{N(W)}{\nu_d(W)}.$$

It is unbiased as follows from (2.20).

Besides unbiasedness the important characteristic of an estimator is its consistency. We use the following definition. Let $\{W_n, n \in \mathbb{N}\}$ be a convex averaging sequence of observation windows defined in Subsection 2.5.6 and let $\hat{\lambda}_n$ be the empirical intensity corresponding to W_n for every n. We say that $\hat{\lambda}$ is **consistent** if $\hat{\lambda}_n \to p$ in probability as $n \to \infty$, i.e.

$$\mathbb{P}(|\hat{\lambda}_n - p| \geq \varepsilon) \to 0 \quad \text{as} \quad n \to \infty$$

for all $\varepsilon > 0$. Since the almost sure convergence implies convergence in probability, we obtain that if N is ergodic then from Theorem 2.5.7 follows that $\hat{\lambda}$ is a consistent estimator of λ. For the definition of different modes of convergence in probability spaces and their relations see for instance [69].

Using (2.21), the variance of $\hat{\lambda}$ can be expressed as

$$\operatorname{var}(\hat{\lambda}) = \frac{\lambda^2}{\nu_d^2(W)} \int_{\mathbb{R}^d} \gamma_W(\boldsymbol{y}) g(\boldsymbol{y}) \, \mathrm{d}\boldsymbol{y} + \frac{\lambda^2}{\nu_d^2(W)} - \lambda^2,$$

where g is the pair correlation function of N and γ_W is the set covariance of W.

The estimation of second order characteristics is strongly influenced by the boundedness of the observation window W. Since second order characteristics need joint information from at least two points of the process one must be careful in the neighbourhood of the edge of W where such a joint information may be inaccessible. Therefore it is convenient to use edge corrections (see [68] for various methods).

Probably the best results are given by Horwitz-Thompson type estimators [70, 71] with suitable weights based on W. The basic unbiased estimator of the reduced second order factorial moment measure $\breve{\Lambda}^{[2]}$ is given by

$$\widehat{\breve{\Lambda}^{[2]}_{st}}(B) = \sum_{\boldsymbol{x},\boldsymbol{y}\in N\cap W}^{\neq} \frac{\mathbb{1}_B(\boldsymbol{y}-\boldsymbol{x})}{\nu_d\big((W-\boldsymbol{x})\cap(W-\boldsymbol{y})\big)}$$

for all Borel B such that $\nu_d\big(W\cap(W-\boldsymbol{y}+\boldsymbol{x})\big)>0$ for all $\boldsymbol{z}=\boldsymbol{y}-\boldsymbol{x}\in B$. The unbiasedness can be shown easily from Campbell's theorem 2.5.3 and Proposition 2.5.5. Moreover, if N is ergodic then one can use Theorem 2.5.8 for factorial moment measures to prove consistency.

Since $\breve{\Lambda}^{[2]}(B_r)$ is $\lambda^2K(r)$, where K is the K-function, we immediately obtain the estimator of $\lambda^2K(r)$. In the isotropic case one can use Ripley's estimator

$$\widehat{\lambda^2K}_R(r) = \sum_{\boldsymbol{x},\boldsymbol{y}\in N\cap W}^{\neq} \frac{\mathbb{1}_{[0,r]}(\|\boldsymbol{x}-\boldsymbol{y}\|)\; w(\boldsymbol{x},\boldsymbol{y})}{\nu_d(W^{(\|\boldsymbol{x}-\boldsymbol{y}\|)})} \quad \text{for} \quad 0\le r<r^*,$$

where $r^*=\sup\big\{r\ge 0|\nu_d\big(W^{(r)}\big)>0\big\}$ and $W^{(r)}=\{\boldsymbol{x}\in W|\partial B_r(\boldsymbol{x})\cap W\neq 0\}$ is the set of all points $\boldsymbol{x}$ of the window W such that a circular arc of radius r centred at $\boldsymbol{x}$ is not completely outside W. Furthermore,

$$w(\boldsymbol{x},\boldsymbol{y}) = \frac{2\pi}{\alpha_{\boldsymbol{x},\boldsymbol{y}}},$$

where $\alpha_{\boldsymbol{x},\boldsymbol{y}}$ is the sum of all those central angles that belong to arcs of the circle centred at $\boldsymbol{x}$, of radius $\|\boldsymbol{x}-\boldsymbol{y}\|$ and lying in W. See [41, Chapter 4], [68, Section 4.3], [72, Chapter 15], [73] and references therein for further discussion and other possible estimators. If one wants an estimator of $K(r)$ instead of $\lambda^2K(r)$ it is strongly recommended ([68, 73]) to use an adapted estimator of λ^2, which depends on r, instead of just $(\hat{\lambda})^2$.

From an interpretation perspective either the reduced second order factorial product density $\breve{\rho}^{[2]}$ or the pair correlation function g could be regarded as a more fundamental quantities. However, the estimation is more complicated since the situation is analogous to the problem of the non-parametric density estimation in classical statistics. A standard approach is to apply kernel estimators.

In the reminder of this subsection we assume N to be isotropic. One proposed estimator for $\breve{\rho}^{[2]}$ corresponding to $\widehat{\breve{\Lambda}^{[2]}_{st}}$ is

$$\hat{\rho}_{st}(r) = \frac{1}{dc_d r^{d-1}} \sum_{\boldsymbol{x},\boldsymbol{y}\in N\cap W}^{\neq} \frac{K_h(r-\|\boldsymbol{y}-\boldsymbol{x}\|)}{\nu_d\big((W-\boldsymbol{x})\cap(W-\boldsymbol{y})\big)},$$

where dc_d is the surface area of the unit sphere in $\mathbb{R}^d$ and $K_h:\mathbb{R}\to\mathbb{R}$ is a kernel function defined by $K_h(x)=h^{-1}K(h^{-1}x)$ for some probability density function K called the **kernel** and $h>0$ called the **bandwidth**. Typically the Epanechnikov or box kernel are used. The estimator given by the previous formula has a property similar to unbiasedness:

$$\mathbb{E}\,\hat{\rho}(r) = \int_{\mathbb{R}} k_h(s)\rho(r+hs)\,\mathrm{d}s.$$

As $h\to 0_+$ one obtains $\breve{\rho}^{[2]}(r)$ if the product density is continuous at r. The crucial role is played by the choice of bandwidth. We refer the reader to [68] for a detailed discussion of this important topic.

The analogue to Ripley's estimator $\widehat{\lambda^2 K}_R(r)$ is

$$\hat{\rho}_R(r) = \frac{1}{dc_d r^{d-1}} \sum_{\boldsymbol{x},\boldsymbol{y}\in N\cap W}^{\neq} \frac{K_h(r - \|\boldsymbol{x}-\boldsymbol{y}\|)\; w(\boldsymbol{x},\boldsymbol{y})}{\nu_d(W^{(\|\boldsymbol{x}-\boldsymbol{y}\|)})} \quad \text{for} \quad 0 \le r < r^*.$$

It is convenient that the estimator has a simpler form given by Ohser's estimator ([68, 73])

$$\hat{\rho}_O(r) = \frac{1}{dc_d r^{d-1}\overline{\gamma}_W(r)} \sum_{\boldsymbol{x},\boldsymbol{y}\in N\cap W}^{\neq} K_h(r - \|\boldsymbol{x}-\boldsymbol{y}\|)$$

for all r such that $\overline{\gamma}_W(r) > 0$, where $\overline{\gamma}_W(r)$ is the isotropised set covariance function of W defined by (2.5.2).

If one wants to estimate g by division by the squared intensity λ^2 it is again important to use some adapted estimator of λ^2. For a fuller treatment of this estimation problem we refer the reader to [68, 72, 73].

4.1.2 Non-stationary point processes

Here we consider the case when the stationarity is broken even for the first moment. Thus in the following it is assumed that the intensity λ of a point process N in $\mathbb{R}^d$ exists but it is not constant. Let us again assume that N is observed in an observation window W with $\nu_d(W) > 0$.

One of the most natural ways for estimating the intensity $\lambda(\boldsymbol{x})$ is to use a non-parametric kernel estimate. The use of this method in spatial statistics of point processes goes back to 1985 when P. Diggle in [74] used kernel smoothing method to estimate the intensity of a Cox spatial point process. The basic idea is similar to kernel smoothing (see e.g. [75, 76, 77]) used in the non-parametric estimation of a probability density function in classical statistics.

The usual edge corrected kernel estimator of the intensity, [78, 79], is

$$\hat{\lambda}_h(\boldsymbol{x}) = \sum_{\boldsymbol{y}\in N\cap W} \frac{K_h(\boldsymbol{x}-\boldsymbol{y})}{\int_W K_h(\boldsymbol{u}-\boldsymbol{y})\,\mathrm{d}\boldsymbol{u}},$$

where $K_h : \mathbb{R}^d \to \mathbb{R}$ is the kernel function defined by $K_h(\boldsymbol{x}) = h^{-d}K(h^{-1}\boldsymbol{x})$ for some radially symmetric multivariate probability density function K (the kernel) and $h > 0$ (the bandwidth). The bandwidth h of the kernel involves a trade-off between bias and variance, whereas, in the common use scenarios, the choice of kernel function K is of secondary importance.

The estimator $\hat{\lambda}_h(\boldsymbol{x})$ is not unbiased since from Campbell's theorem 2.5.2 follows

$$\mathbb{E}\,\hat{\lambda}_h(\boldsymbol{x}) = \int_W \frac{K_h(\boldsymbol{x}-\boldsymbol{y})}{\int_W K_h(\boldsymbol{u}-\boldsymbol{y})\,\mathrm{d}\boldsymbol{u}}\lambda(\boldsymbol{y})\,\mathrm{d}\boldsymbol{y}.$$

If the intensity is continuous at $\boldsymbol{x}$ and $\boldsymbol{x} \in W^\circ$ then $\mathbb{E}\,\hat{\lambda}_h(\boldsymbol{x}) \to \lambda(\boldsymbol{x})$ as $h \to 0_+$.

The bandwidth is usually chosen such that it minimizes the mean square error (MSE) of $\hat{\lambda}_h(\boldsymbol{x})$ with respect to the true intensity $\lambda(\boldsymbol{x})$, which is defined by

$$\mathrm{MSE}\big(\hat{\lambda}_h(\boldsymbol{x})\big) = \mathbb{E}\,\big(\hat{\lambda}_h(\boldsymbol{x}) - \lambda(\boldsymbol{x})\big)^2.$$

For particular point models the MSE can be explicitly evaluated as a function of second order statistics. In the case of Cox process it was done by Diggle [74]. Generally the MSE depends on $\boldsymbol{x}$ and cannot be minimized simultaneously for all $\boldsymbol{x} \in W$. In that case one minimizes the mean integrated square error (MISE) defined by

$$\mathrm{MISE}\big(\hat{\lambda}_h\big) = \int_W \mathrm{MSE}\big(\hat{\lambda}_h(\boldsymbol{x})\big)\,\mathrm{d}\boldsymbol{x}.$$

Note that the problematic part of the minimization is that MISE (and also MSE) includes second order characteristics like the second order factorial product density $\rho^{[2]}$. Thus in practice the minimization without additional assumptions is quite problematic.

Sometimes, computably less demanding version (the Nadaraya-Watson estimator, [80]) $\hat{\lambda}_{NW;h}$ of $\hat{\lambda}_h$ is used,

$$\hat{\lambda}_{NW;h}(\boldsymbol{x}) = \frac{\sum_{\boldsymbol{y}\in N\cap W} K_h(\boldsymbol{x}-\boldsymbol{y})}{\int_W K_h(\boldsymbol{x}-\boldsymbol{u})\,\mathrm{d}\boldsymbol{u}}. \tag{4.1}$$

Again, it is generally not unbiased and $\mathbb{E}\,\hat{\lambda}_{NW;h}(\boldsymbol{x}) \to \lambda(\boldsymbol{x})$ as $h \to 0_+$ if λ is continuous at $\boldsymbol{x} \in W^\circ$.

For second order characteristics the situation is even more complicated since without stationarity the usually analysed characteristics like the reduced second order factorial moment measure, its density and pair correlation function generally do not exist. Here we briefly follow According to [68] it is known that, except some local methods for short range interaction processes, there are generally two ways how to deal with this problem. The first is the intensity reweighting and the second the local rescaling.

In the intensity reweighting approach one assumes that N has an everywhere positive intensity $\lambda(\boldsymbol{x})$ and constructs a new random measure ξ_N by

$$\xi_N = \sum_{\boldsymbol{y}\in N} \frac{1}{\lambda(\boldsymbol{y})}\delta_{\boldsymbol{y}},$$

where $\delta_{\boldsymbol{y}}$ is the Dirac measure defined by A.1. If ξ_N is second order stationary then N is said to be **second order intensity reweighted stationary**. In that case one analyses second order characteristics of ξ_N and estimates them by usual estimators of stationary point processes from the previous section. In particular the definition and stationarity of ξ_N implies that ξ_N has intensity equal to 1. Hence the pair correlation function of ξ_N equals the pair correlation function of N:

$$g_{\xi_N}(\boldsymbol{x},\boldsymbol{y}) = \rho^{[2]}_{\xi_N}(\boldsymbol{x},\boldsymbol{y}) = \frac{\rho^{[2]}_N(\boldsymbol{x},\boldsymbol{y})}{\lambda(\boldsymbol{x})\lambda(\boldsymbol{y})} = g_N(\boldsymbol{x},\boldsymbol{y}),$$

which consequently depends only on $\boldsymbol{r} = \boldsymbol{x}-\boldsymbol{y}$ or just on $r = \|\boldsymbol{x}-\boldsymbol{y}\|$ when ξ_N is isotropic. Intensity reweighting is suitable for inhomogeneous Poisson processes and in general for processes obtained by independent thinning from a stationary point process. On the other hand it is unsuitable for processes with varying range of correlation. For further details see [78, 79].

The other possible method is the local rescaling. The general idea is to locally rescale the metric on $\mathbb{R}^d$ in the way that the intensity with respect to corresponding locally rescaled volume measure is constant. One then studies all the characteristics in this modified metric. This method is especially useful for Gibbs point processes. For more details see [81].

It is worth noting that the main problem in non-stationary statistics of point processes is the possible confounding between intensity and interaction. It was shown (see [79, 82] and references therein) that sometimes it is not possible to recognize the difference of spatial inhomogeneity from clustering in a single realization of a point process. Thus it may be not possible to distinguish from one realization if the point process is stationary or not. In the theory of point processes this represents a fundamental limitation of the scope of statistical inference.

4.2 Stationary random closed sets

In this section we discuss possible methods for estimating the volume fraction and second order characteristics of a random closed set.

4.2.1 Volume fraction

First let us look more closely at the estimation of the volume fraction. Let assume that a stationary random closed set X is observed in an observation window W such that $0 < \nu_d(W) < \infty$. The most natural estimator of the volume fraction p is given by the **empirical volume fraction**

$$\hat{p}_v = \frac{\nu_d(X \cap W)}{\nu_d(W)}. \tag{4.2}$$

From (2.7) follows that $\hat{p}_v$ is unbiased, $\mathbb{E}\,\hat{p}_v = p$, for every window W. If X is ergodic random closed set then Corollary 2.2.2 implies the consistency of $\hat{p}_v$. Here we mean the consistency defined in part 4.1.1.

The variance of the estimator is

$$\begin{aligned}\operatorname{var}(\hat{p}_v) = \mathbb{E}(\hat{p}_v - p)^2 &= \frac{\mathbb{E}\left(\int_W \mathbb{1}_x(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}\right)^2 - \left(\int_W p\,\mathrm{d}\boldsymbol{x}\right)^2}{\nu_d^2(W)} \\ &= \frac{1}{\nu_d^2(W)}\,\mathbb{E}\int_W\int_W \mathbb{1}_x(\boldsymbol{x})\mathbb{1}_x(\boldsymbol{y}) - p^2\,\mathrm{d}\boldsymbol{x}\,\mathrm{d}\boldsymbol{y} \\ &= \frac{1}{\nu_d^2(W)}\int_W\int_W \operatorname{cov}(\boldsymbol{x}-\boldsymbol{y})\,\mathrm{d}\boldsymbol{x}\,\mathrm{d}\boldsymbol{y}. \end{aligned} \tag{4.3}$$

This can be rewritten using the correlation function κ of X as

$$\operatorname{var}(\hat{p}_v) = \frac{p(1-p)}{\nu_d^2(W)}\int_W\int_W \kappa(\boldsymbol{x}-\boldsymbol{y})\,\mathrm{d}\boldsymbol{x}\,\mathrm{d}\boldsymbol{y} = \frac{p(1-p)}{\nu_d^2(W)}\int_W\int_{\boldsymbol{v}-W} \kappa(\boldsymbol{u})\,\mathrm{d}\boldsymbol{u}\,\mathrm{d}\boldsymbol{v}. \tag{4.4}$$

For W large in all dimensions one obtains the following approximation

$$\operatorname{var}(\hat{p}_v) \approx p(1-p)\frac{A}{\nu_d(W)}, \tag{4.5}$$

where

$$A = \int_{\mathbb{R}^d} \kappa(\boldsymbol{r})\,\mathrm{d}\boldsymbol{r}$$

is called the **integral range** whenever the integral on the right side exists, [41, 49]. The dimension of integral range is the same as of a volume of $\mathbb{R}^d$. The integral range has a similar interpretation as the correlation time for stochastic processes (see e.g. [83]). If the size $\nu_d(W)$ of the observation window does not exceed very much the integral range, then the accuracy of the estimate is up to the multiplicative constant given by the standard deviation $\sqrt{p(1-p)}$ of the volume fraction. With increasing size of W the variance decreases as $\nu_d(W)^{-1}$. If we denote $N = \nu_d(W)/A$ and $\sigma^2 = p(1-p)$ then

$$\operatorname{var}(\hat{p}_v) \approx \frac{\sigma^2}{N},$$

which is the usual formula for the variance of the sample mean constructed from N i.i.d. random variables with common variance σ^2. This means that N gives the effective number of domains of size A that are analogous to independent observations.

Clearly the approximation (4.5) is inapplicable when $A = 0$, $A = \infty$, or A is not defined. In the first case it follows from (4.4) that the variance decreases more rapidly than $\nu_d(W)^{-1}$ as $\nu_d(W) \to \infty$ and hence the acceptable window sizes are generally smaller than for cases when $A > 0$. On the other hand, if $A = \infty$ then by Proposition 2.8.1 X is long-range dependent and formula (4.4) implies that $\operatorname{var}(\hat{p}_v)$ falls off more slowly than $\nu_d(W)^{-1}$ with increasing size

of W. If X is isotropically long-range dependent according to Definition 2.8.2, then we may use Proposition 2.8.6 for its volume measure ν_X. Let assume that we observe X in a family of convex compact windows $\{aW|a>0\}$. Then

$$\operatorname{var}\hat{p}_v = \frac{\operatorname{var}\nu_X(aW)}{\nu_d^2(aW)} \sim a^{-\alpha}\frac{F_{\alpha;W}\ell(a)}{\nu_d(W)} \qquad (a\to\infty)$$

since $\lambda_D = 0$ for random closed sets as was discussed in Subsection 2.5.4. Thus for large window W we have the asymptotic relation

$$\operatorname{var}\hat{p}_v \approx \nu_d(W)^{-\frac{\alpha}{d}}K_{\alpha,W}, \tag{4.6}$$

where $K_{\alpha,W}$ depends only on the shape of W and not on its scale. If A is not defined, it may still be the case of the long-range dependence but more detailed analysis is needed. For further discussion of the long-range dependence see Section 2.8.

We have seen that ergodicity was a sufficient condition for p_v to be consistent. Now we present a necessary condition for the mean square consistency, formulated using the correlation function, a result known as Slutsky's ergodic theorem [83, 49]. Let $\{W_n, n\in\mathbb{N}\}$ be a convex averaging sequence of observation windows from Definition 2.2.4 and let $\hat{p}_{v;n}$ be the empirical volume fraction corresponding to W_n for every n. We say that p_v is **mean square consistent** if for every convex averaging sequence $\{W_n\}$,

$$\mathbb{E}(\hat{p}_{v;n}-p)^2 = \operatorname{var}(\hat{p}_{v;n}) \to 0 \quad \text{as} \quad n\to\infty.$$

Classical consistency defined above is a consequence of mean square consistency ([69]).

Theorem 4.2.1 (Slutsky's ergodic theorem)**.** *Let X be a stationary random closed set with covariance function* cov*. Then the empirical volume fraction p_v is mean square consistent if and only if*

$$\frac{1}{\nu_d(W_n)}\int_{W_n}\operatorname{cov}(\boldsymbol{r})\,\mathrm{d}\boldsymbol{r}\to 0 \quad \textit{as} \quad n\to\infty \tag{4.7}$$

for every convex averaging sequence $\{W_n, n\in\mathbb{N}\}$.

Proof. The proof is similar to the one dimensional case that can be found as note 3.5 in [84]. First let assume that p_v is mean square consistent. For the expression on the left side of (4.7) we have

$$\begin{aligned}\frac{1}{\nu_d(W_n)}\int_{W_n}\operatorname{cov}(\boldsymbol{r})\,\mathrm{d}\boldsymbol{r} &= \frac{1}{\nu_d(W_n)}\,\mathbb{E}\int_{W_n}\mathbb{1}_X(\boldsymbol{r})\mathbb{1}_X(\boldsymbol{0})-p^2\,\mathrm{d}\boldsymbol{r}\\ &= \mathbb{E}\,p_{v;n}\mathbb{1}_X(\boldsymbol{0})-p^2 = \mathbb{E}(p_{v;n}-p)(\mathbb{1}_X(\boldsymbol{0})-p).\end{aligned}$$

Thus it is a scalar product of $p_{v;n}-p$ and $\mathbb{1}_X(\boldsymbol{0})-p$ in a Hilbert space $L^2(\mathcal{F},\mathbb{P}_X)$. The Cauchy-Schwartz inequality yields

$$\left|\frac{1}{\nu_d(W_n)}\int_{W_n}\operatorname{cov}(\boldsymbol{r})\,\mathrm{d}\boldsymbol{r}\right| \le \sqrt{\operatorname{cov}(\boldsymbol{0})}\sqrt{\operatorname{var}(\hat{p}_{v;n})}$$

and hence the condition (4.7) holds.

For the opposite implication let $K_n = W_n - W_n = \{\boldsymbol{x}-\boldsymbol{y}|\boldsymbol{x},\boldsymbol{y}\in W_n\}$ for every n. It is easy to see that K_n is again a convex averaging sequence. Moreover for every $n\in\mathbb{N}$ its volume satisfies $\nu_d(K_n)\le 3^d\nu_d(W_n)$ and $\boldsymbol{x}-W_n\subset K_n$ for each $\boldsymbol{x}\in W_n$. Now (4.7) yields that for $\varepsilon>0$ there exists $n_0=n_0(\varepsilon)$ such that

$$\int_{K_n}\operatorname{cov}(\boldsymbol{r})\,\mathrm{d}\boldsymbol{r}\le\varepsilon\nu_d(K_n) \quad \text{for } n>n_0.$$

On the other hand, since $|\mathrm{cov}(\boldsymbol{r})| \le \mathrm{cov}(\boldsymbol{0})$ which follows from the positive semi-definiteness,

$$\int_B \mathrm{cov}(\boldsymbol{r})\,\mathrm{d}\boldsymbol{r} \le \mathrm{cov}(\boldsymbol{0})\nu_d(B) \quad \text{for any Borel set } B.$$

Therefore if $n > n_0$,

$$\begin{aligned}
\mathrm{var}(\hat{p}_v) &= \frac{1}{\nu_d^2(W_n)} \int_{W_n} \int_{\boldsymbol{x}-W_n} \mathrm{cov}(\boldsymbol{r})\,\mathrm{d}\boldsymbol{r}\,\mathrm{d}\boldsymbol{x} \\
&= \frac{1}{\nu_d^2(W_n)} \left(\int_{W_{n_0}} \int_{\boldsymbol{x}-W_n} \mathrm{cov}(\boldsymbol{r})\,\mathrm{d}\boldsymbol{r}\,\mathrm{d}\boldsymbol{x} + \int_{W_n \setminus W_{n_0}} \int_{\boldsymbol{x}-W_n} \mathrm{cov}(\boldsymbol{r})\,\mathrm{d}\boldsymbol{r}\,\mathrm{d}\boldsymbol{x} \right) \\
&\le \frac{1}{\nu_d^2(W_n)} \left(\int_{W_{n_0}} \mathrm{cov}(\boldsymbol{0})\nu_d(\boldsymbol{x}+W_n)\,\mathrm{d}\boldsymbol{x} + \int_{W_n \setminus W_{n_0}} \int_{K_n} \mathrm{cov}(\boldsymbol{r})\,\mathrm{d}\boldsymbol{r}\,\mathrm{d}\boldsymbol{x} \right) \\
&= \frac{1}{\nu_d^2(W_n)} \left(\mathrm{cov}(\boldsymbol{0})\nu_d(W_n)\nu_d(W_{n_0}) + \int_{W_n \setminus W_{n_0}} \varepsilon\nu_d(K_n)\,\mathrm{d}\boldsymbol{x} \right) \\
&= \frac{\mathrm{cov}(\boldsymbol{0})\nu_d(W_{n_0})}{\nu_d(W_n)} + \varepsilon \frac{\nu_d(K_n)\big(\nu_d(W_n) - \nu_d(W_{n_0})\big)}{\nu_d^2(W_n)} \\
&\le \frac{\mathrm{cov}(\boldsymbol{0})\nu_d(W_{n_0})}{\nu_d(W_n)} + \varepsilon 3^d.
\end{aligned}$$

As a consequence $\mathrm{var}(\hat{p}_v)$ can be made arbitrary small by choosing n to be sufficiently large. Hence $\mathrm{var}(\hat{p}_v) \to 0$ as $n \to \infty$. □

It is easy to see that every ergodic random closed set X satisfies (4.7). This is an immediate result of Corollary 2.5.3 for the volume measure ν_X of X.

Another possible method of estimating the volume fraction p of a stationary random closed set X arises when X is observed at a specific number of points in some observation window W. For $\boldsymbol{x}_1, \dots, \boldsymbol{x}_n \in W$ we define $\hat{p}_p$ to be

$$\hat{p}_p = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_X(\boldsymbol{x}_i). \tag{4.8}$$

The estimator is clearly unbiased. For its variance we have

$$\begin{aligned}
\mathrm{var}(\hat{p}_p) = \mathbb{E}(\hat{p}_p - p)^2 &= \frac{1}{n^2}\,\mathbb{E}\left(\sum_{i=1}^{n} \mathbb{1}_X(\boldsymbol{x}_i) - np \right)^2 \\
&= \frac{1}{n^2}\,\mathbb{E} \sum_{i,j=1}^{n} \left(\mathbb{1}_X(\boldsymbol{x}_i)\mathbb{1}_X(\boldsymbol{x}_j) - p^2 \right) \\
&= \frac{2}{n^2} \sum_{i<j}^{n} \mathrm{cov}(\boldsymbol{x}_i - \boldsymbol{x}_j) + \frac{p(1-p)}{n}.
\end{aligned}$$

The exact behaviour of the first term depends on the sampling scheme, that is on the particular choice of positions $\boldsymbol{x}_1, \dots, \boldsymbol{x}_n$ in the window W. Usual possibilities are random sampling, where the points are chosen randomly with uniform distribution in W, and square grid sampling, where the points are given by the intersection of a regular point lattice in $\mathbb{R}^d$ with W. The

later situation often arise in image analysis, where the grid points are centres of pixels, see [50] for more details. For a fuller treatment of different sampling schemes and their influences to the difference $\hat{p}_p - \hat{p}_v$ we refer the reader to [49, §2.8.3], [85, §5.6.1], [86, Chapter 3] and references therein. In most situations $\mathrm{var}(\hat{p}_v) \leq \mathrm{var}(\hat{p}_p)$ and sampling on a square grid is usually the best sampling scheme, i.e. it produces the smallest variance of $\hat{p}_p$. However, it should be noted that there are situations when the estimator $\hat{p}_p$ has smaller variance than $\hat{p}_v$, see [87]. Some discussion about the connection of $\hat{p}_p$ to ergodicity can be found in [50].

4.2.2 Second order characteristics

In the following part we discuss the estimation procedure for most important second order characteristics of a stationary random closed set X that is again observed in a window W. Most estimation procedures are based on relation (2.8) for the covariance.

If the stationary random closed set X is observed in some window W, we cannot directly estimate $C(\boldsymbol{r})$ as the volume fraction of $X \cap (X - \boldsymbol{r})$ since the information from outside of W is needed. This problem can be overcome by the so called minus sampling, where the set $B = W \cap W - \boldsymbol{r}$ instead of W is used, because inside B we know values of X and also of $X - \boldsymbol{r}$. The unbiased estimator of the covariance is therefore given by the so called **empirical covariance**

$$\hat{C}_v(\boldsymbol{r}) = \frac{\nu_d\big(X \cap (X - \boldsymbol{r}) \cap W \cap (W - \boldsymbol{r})\big)}{\nu_d\big(W \cap (W - \boldsymbol{r})\big)} \tag{4.9}$$

for all $\boldsymbol{r} \in \mathbb{R}^d$ such that $\nu_d\big(W \cap (W - \boldsymbol{r})\big) > 0$. From the translation invariance of the Lebesgue measure follows $\hat{C}_v(\boldsymbol{r}) = \hat{C}_v(-\boldsymbol{r})$ for all admissible $\boldsymbol{r}$. For ergodic random closed sets the empirical covariance is consistent and also mean square consistent. Let $\{W_n, n \in \mathbb{N}\}$ be a convex averaging sequence of observation windows and let $\hat{C}_{v;n}(\boldsymbol{r})$ be the empirical covariance corresponding to W_n for every n.

Proposition 4.2.1. *Let X be a stationary ergodic random closed set. Then for every $\boldsymbol{r} \in \mathbb{R}^d$ and every convex averaging sequence $\{W_n, n \in \mathbb{N}\}$*

$$\hat{C}_{v;n}(\boldsymbol{r}) \to C(\boldsymbol{r}) \quad \textit{as} \quad n \to \infty$$

almost surely and in mean square.

Proof. Let ν_X be the volume measure of X and $\boldsymbol{r} \in \mathbb{R}^d$. We define $f_{\boldsymbol{r}}(\varphi) = \mathbb{1}_{\mathrm{supp}\,\varphi}(\boldsymbol{0})\mathbb{1}_{\mathrm{supp}\,\varphi}(\boldsymbol{r})$ for all $\varphi \in \mathsf{M}$. From Proposition 2.5.1 and Theorem A.4.5 follows that f is measurable. Let further $S_{\boldsymbol{x}}$ denote the action of a group of translations in $\mathbb{R}^d$ corresponding to $\boldsymbol{x} \in \mathbb{R}^d$. According to relations (2.19) and (2.18) we have

$$f_{\boldsymbol{r}}(S_{\boldsymbol{x}}\nu_X) = f_{\boldsymbol{r}}(\mu_{S_{\boldsymbol{x}}X}) = f_{\boldsymbol{r}}(\mu_{X+\boldsymbol{x}}) = \mathbb{1}_{X+\boldsymbol{x}}(\boldsymbol{0})\mathbb{1}_{X+\boldsymbol{x}}(\boldsymbol{r}) = \mathbb{1}_X(-\boldsymbol{x})\mathbb{1}_{X-\boldsymbol{r}}(-\boldsymbol{x})$$

for every $\boldsymbol{x} \in \mathbb{R}^d$. Now, Proposition 2.5.10 for the convex averaging sequence $\{-W_n \cap (-W_n +$

$\boldsymbol{r}), n \in \mathbb{N}\}$ yields

$$\begin{aligned}\hat{C}_{v;n}(\boldsymbol{r}) &= \frac{\nu_d\big(X \cap (X-\boldsymbol{r}) \cap W_n \cap (W_n - \boldsymbol{r})\big)}{\nu_d\big(W_n \cap (W_n - \boldsymbol{r})\big)} \\ &= \frac{1}{\nu_d\big(W_n \cap (W_n - \boldsymbol{r})\big)} \int\limits_{W_n \cap (W_n - \boldsymbol{r})} \mathbb{1}_X(\boldsymbol{x}) \mathbb{1}_{X-\boldsymbol{r}}(\boldsymbol{x})\, \mathrm{d}\boldsymbol{x} \\ &= \frac{1}{\nu_d\big(W_n \cap (W_n - \boldsymbol{r})\big)} \int\limits_{-W_n \cap (-W_n + \boldsymbol{r})} \mathbb{1}_X(-\boldsymbol{x}) \mathbb{1}_{X-\boldsymbol{r}}(-\boldsymbol{x})\, \mathrm{d}\boldsymbol{x} \\ &= \frac{1}{\nu_d\big(-W_n \cap (-W_n + \boldsymbol{r})\big)} \int\limits_{-W_n \cap (-W_n + \boldsymbol{r})} f_{\boldsymbol{r}}(S_{\boldsymbol{x}}\omega)\, \mathrm{d}\boldsymbol{x} \\ &\to \mathbb{E}(f_{\boldsymbol{r}}(\nu_X)) = C(\boldsymbol{r}) \quad \text{as} \quad n \to \infty\end{aligned}$$

almost surely and in the square mean since $|f| \leq 1$. □

Analogously to the volume fraction, the variance of $\hat{C}_{v;n}(\boldsymbol{r})$ is given by

$$\begin{aligned}\operatorname{var}\big(\hat{C}_v(\boldsymbol{r})\big) &= \mathbb{E}\big(\hat{C}_v(\boldsymbol{r}) - C(\boldsymbol{r})\big)^2 \\ &= \frac{1}{\nu_d^2\big(W \cap (W - \boldsymbol{r})\big)} \int\limits_{W \cap (W-\boldsymbol{r})} \int\limits_{W \cap (W-\boldsymbol{r})} \operatorname{cov}_{X \cap (X-\boldsymbol{r})}(\boldsymbol{x} - \boldsymbol{y})\, \mathrm{d}\boldsymbol{x}\, \mathrm{d}\boldsymbol{y},\end{aligned}$$

where $\operatorname{cov}_{X\cap(X-\boldsymbol{r})}(\boldsymbol{x}-\boldsymbol{y})$ is the covariance function of the random closed set $X \cap (X - \boldsymbol{r})$ that is also stationary as follows from Corollary 2.2.1. Again, Slutsky's theorem for mean square consistency of $\hat{C}_v(\boldsymbol{r})$ can be formulated.

Theorem 4.2.2. *Let X be a stationary random closed set and $\boldsymbol{r} \in \mathbb{R}^d$. Then the empirical covariance $\hat{C}_v(\boldsymbol{r})$ is mean square consistent if and only if*

$$\frac{1}{\nu_d(W_n)} \int\limits_{W_n} \operatorname{cov}_{X\cap(X-\boldsymbol{r})}(\boldsymbol{u})\, \mathrm{d}\boldsymbol{u} \to 0 \quad \textit{as} \quad n \to \infty$$

for every convex averaging sequence $\{W_n, n \in \mathbb{N}\}$.

Proof. Analogously as in Theorem 4.2.1. □

The condition is, as a result of the previous proposition, satisfied for every ergodic random closed set. An important modification of the empirical covariance is given by

$$\hat{C}_v^*(\boldsymbol{r}) = \frac{\nu_d\big(X \cap (X-\boldsymbol{r}) \cap W \cap (W - \boldsymbol{r})\big)}{\nu_d(W)}$$

for all $\boldsymbol{r} \in \mathbb{R}^d$. Since

$$\hat{C}_v^*(\boldsymbol{r}) = \frac{\nu_d(W)}{\nu_d\big(W \cap (W - \boldsymbol{r})\big)} \hat{C}_v(\boldsymbol{r}),$$

it is asymptotically unbiased as $\nu_d(W) \to \infty$. Moreover it is also consistent whenever $\hat{C}_v(\boldsymbol{r})$ is consistent. This estimator is usually preferred in the field of time-series analysis [83, 88]. The reason is in the fact that $\hat{C}_v^*(\boldsymbol{r})$ has despite its biasedness many favourable properties. First, note that the estimator $\hat{C}_{v;n}(\boldsymbol{r})$ does not always produce positive definite function of $\boldsymbol{r}$ which the theoretical covariance always satisfy as a consequence of Corollary 2.3.2. This can be seen on example of a set $X = [0,1] \cup [2,3] \subset \mathbb{R}$ observed in window $W = [0,3]$. In that case

$$\hat{C}_v(0) = \frac{2}{3} \quad \text{and} \quad \hat{C}_v(2) = \frac{1}{1} = 1.$$

Hence $\hat{C}_v(2) > \hat{C}_v(0)$ which is impossible for any positive semi-definite function. On the other hand we can easily prove that $\hat{C}_v^*(\boldsymbol{r})$ is always positive semi-definite.

Proposition 4.2.2. *The estimate $\hat{C}_v^*(\boldsymbol{r})$ is a positive semi-definite function of $\boldsymbol{r}$.*

Proof. For all $\boldsymbol{r}_1, \boldsymbol{r}_2 \in \mathbb{R}^d$

$$\begin{aligned}\hat{C}_v^*(\boldsymbol{r}_1 - \boldsymbol{r}_2) &= \frac{1}{\nu_d(W)} \int_{\mathbb{R}^d} \mathbb{1}_W(\boldsymbol{x}) \mathbb{1}_X(\boldsymbol{x}) \mathbb{1}_{W-\boldsymbol{r}_1+\boldsymbol{r}_2}(\boldsymbol{x}) \mathbb{1}_{X-\boldsymbol{r}_1+\boldsymbol{r}_2}(\boldsymbol{x}) \,\mathrm{d}\boldsymbol{x} \\ &= \frac{1}{\nu_d(W)} \int_{\mathbb{R}^d} \mathbb{1}_W(\boldsymbol{x}) \mathbb{1}_X(\boldsymbol{x}) \mathbb{1}_W(\boldsymbol{x}+\boldsymbol{r}_1-\boldsymbol{r}_2) \mathbb{1}_X(\boldsymbol{x}+\boldsymbol{r}_1-\boldsymbol{r}_2) \,\mathrm{d}\boldsymbol{x} \\ &= \frac{1}{\nu_d(W)} \int_{\mathbb{R}^d} \mathbb{1}_W(\boldsymbol{x}-\boldsymbol{r}_1) \mathbb{1}_X(\boldsymbol{x}-\boldsymbol{r}_1) \mathbb{1}_W(\boldsymbol{x}-\boldsymbol{r}_2) \mathbb{1}_X(\boldsymbol{x}-\boldsymbol{r}_2) \,\mathrm{d}\boldsymbol{x}.\end{aligned}$$

Hence, for all $n \in \mathbb{N}$, $\boldsymbol{r}_1, \ldots, \boldsymbol{r}_n \in \mathbb{R}^d$ and $\lambda_1, \ldots, \lambda_n \in \mathbb{C}$

$$\sum_i \sum_j \lambda_i \overline{\lambda_j} \hat{C}_v^*(\boldsymbol{r}_i - \boldsymbol{r}_j) = \frac{1}{\nu_d(W)} \int_{\mathbb{R}^d} \left(\sum_i \mathbb{1}_W(\boldsymbol{x}-\boldsymbol{r}_i) \mathbb{1}_X(\boldsymbol{x}-\boldsymbol{r}_i) \right)^2 \mathrm{d}\boldsymbol{x} \geq 0. \qquad \square$$

Another reason why the biased version is so popular is that under ergodicity condition one can prove, see [89], that $\hat{C}_v^*(\boldsymbol{r})$ converges uniformly to $C(\boldsymbol{r})$ with probability one, that is

$$\sup_{\boldsymbol{r} \in \mathbb{R}^d} \left| \hat{C}_{v;n}^*(\boldsymbol{r}) - C(\boldsymbol{r}) \right| \to 0 \quad \text{as} \quad n \to \infty$$

almost surely for every convex averaging sequence $\{W_n\}$. Moreover, it is not possible to prove analogous formula for $\hat{C}_v$. For further discussion of the properties of $\hat{C}_v^*$ and arguments for its superiority over $\hat{C}_v$ in one dimensional case and short-range dependence see [83, Section 3.17] and [88, Chapter 6].

The main problem of estimator $\hat{C}_v^*(\boldsymbol{r})$ is the bias that can produce significantly large relative error if $\boldsymbol{r}$ is not small with respect to diameter of W. This is especially a problem for long-range processes.

Now let focus on the estimation procedure of the covariance function cov and the correlation function κ. The natural estimator $\hat{\mathrm{cov}}_v(\boldsymbol{r})$ of the covariance function of stationary random closed set X is given by

$$\hat{\mathrm{cov}}_v(\boldsymbol{r}) = \hat{C}_v(\boldsymbol{r}) - (\hat{p}_v)^2,$$

whenever $\nu_d\big(W \cap (W - \boldsymbol{r})\big) > 0$, where $\hat{C}_v(\boldsymbol{r})$ is the empirical covariance and $\hat{p}_v$ is the empirical volume fraction defined by (4.2). This estimator is generally not unbiased. This follows from the fact that $\hat{C}_v(\boldsymbol{r})$ is unbiased whereas $(\hat{p}_v)^2$ is not. Since $\hat{p}_v$ is unbiased the bias of $(\hat{p}_v)^2$ can be expressed as

$$\mathbb{E}(\hat{p}_v)^2 - p^2 = \mathbb{E}\left(\hat{p}_v - p\right)^2 = \mathrm{var}(\hat{p}_v)$$

and is therefore always non-negative. For the empirical covariance function it leads to

$$\mathbb{E}\,\hat{\mathrm{cov}}_v(\boldsymbol{r}) - \mathrm{cov}(\boldsymbol{r}) = -\mathrm{var}(\hat{p}_v) \leq 0.$$

The bias is therefore independent of $\boldsymbol{r}$. The large relative errors are produced when the value of the covariance function is small which is usually the case of large $\boldsymbol{r}$. From the discussion of the consistency of $\hat{p}_v$ in the previous section follows that $\hat{\mathrm{cov}}_v(\boldsymbol{r})$ is asymptotically (with increasing W) unbiased for ergodic random closed sets and, moreover, for all random closed sets satisfying Slutsky's condition (4.7).

Several modifications that lead to estimators of better performance were proposed by Picka in [90]. Here we present only the so called **intrinsically balanced** estimator $\hat{\mathrm{cov}}_v^{\bullet}(\boldsymbol{r})$ of the covariance function defined by

$$\hat{\mathrm{cov}}_v^{\bullet}(\boldsymbol{r}) = \hat{C}_v(\boldsymbol{r}) - \hat{p}_v[\check{\mathbf{0}}, \boldsymbol{r}]\hat{p}_v[\mathbf{0}, \check{\boldsymbol{r}}], \tag{4.10}$$

where

$$\hat{p}_v[\check{\mathbf{0}}, \boldsymbol{r}] = \frac{\nu_d\big(X \cap W \cap (W - \boldsymbol{r})\big)}{\nu_d\big(W \cap (W - \boldsymbol{r})\big)} \quad \text{and} \quad \hat{p}_v[\mathbf{0}, \check{\boldsymbol{r}}] = \frac{\nu_d\big((X - \boldsymbol{r}) \cap W \cap (W - \boldsymbol{r})\big)}{\nu_d\big(W \cap (W - \boldsymbol{r})\big)}$$

for all $\boldsymbol{r} \in \mathbb{R}^d$ such that $\nu_d\big(W \cap (W - \boldsymbol{r})\big) > 0$. The general idea of this modification is to use the information from different subsets of $X \cap W$ in the same way in both parts $\hat{C}_v(\boldsymbol{r})$ and $(\hat{p}_v)^2$ of $\hat{\mathrm{cov}}_v(\boldsymbol{r})$. The observation window W can be divided into 3 disjoint sets:

$$\begin{aligned} W_0 &= W \cap (W - \boldsymbol{r})^c \cap (W + \boldsymbol{r})^c, \\ W_1 &= \big(W \cap (W - \boldsymbol{r}) \cap (W + \boldsymbol{r})^c\big) \cup \big(W \cap (W - \boldsymbol{r})^c \cap (W + \boldsymbol{r})\big), \\ W_2 &= W \cap (W - \boldsymbol{r}) \cap (W + \boldsymbol{r}). \end{aligned}$$

The information from X observed in W is used by the estimator $\hat{C}_v(\boldsymbol{r})$ in the following way. In W_0 no points of X are used, in W_1 every point of X is used exactly once, and in W_2 exactly twice. This is clearly the same way how the estimator $\hat{p}_v[\check{\mathbf{0}}, \boldsymbol{r}]\hat{p}_v[\mathbf{0}, \check{\boldsymbol{r}}]$ uses the information from X in W. Since clearly both $\hat{p}_v[\check{\mathbf{0}}, \boldsymbol{r}], \hat{p}_v[\mathbf{0}, \check{\boldsymbol{r}}]$ are unbiased, the bias of $\hat{\mathrm{cov}}_v^{\bullet}(\boldsymbol{r})$ is equal to

$$\mathbb{E}\,\hat{\mathrm{cov}}_v^{\bullet}(\boldsymbol{r}) - \mathrm{cov}(\boldsymbol{r}) = -\,\mathbb{E}\,\hat{p}_v[\check{\mathbf{0}}, \boldsymbol{r}]\hat{p}_v[\mathbf{0}, \check{\boldsymbol{r}}] + p^2 = -\,\mathrm{cov}\left(\hat{p}_v[\check{\mathbf{0}}, \boldsymbol{r}], \hat{p}_v[\mathbf{0}, \check{\boldsymbol{r}}]\right).$$

If the random closed set is assumed to be isotropic, the best performance is usually obtained by estimators modified according to this property. The isotropic modification of the intrinsically balanced covariance function estimator is for all $r > 0$ given by

$$\hat{\mathrm{cov}}_v^{I\bullet}(r) = \hat{C}_v^I(r) - \big(\hat{p}_v^I(r)\big)^2, \tag{4.11}$$

with

$$\begin{aligned} \hat{C}_v^I(r) &= \frac{\overline{\gamma}_{X\cap W}(r)}{\overline{\gamma}_W(r)}, \\ \hat{p}_v^I(r) &= \frac{\int_{S^{d-1}} \nu_d(X \cap W \cap (W - r\boldsymbol{u}))\ \sigma_{d-1}(\mathrm{d}\boldsymbol{u})}{dc_d \overline{\gamma}_W(r)}, \end{aligned} \tag{4.12}$$

where S^{d-1} is the unit sphere in $\mathbb{R}^d$, σ_{d-1} is the usual non-normalized spherical measure, which equals to $d-1$ dimensional Hausdorff measure, on S^{d-1}, dc_d is the surface area of S^{d-1}, $dc_d = \sigma_{d-1}(S^{d-1})$, and $\overline{\gamma}_B(r)$ is the isotropised set covariance defined by

$$\overline{\gamma}_B(r) = \frac{1}{dc_d} \int\limits_{S^{d-1}} \nu_d\big(B \cap (B - r\boldsymbol{u})\big)\ \sigma_{d-1}(\mathrm{d}\boldsymbol{u})$$

for all $r \geq 0$. Note that $\hat{C}_v^I(r)$ gives the isotropic modification of the empirical covariance $\hat{C}_v(\boldsymbol{r})$ and can be used to estimate the covariance $C(\boldsymbol{r})$ under the isotropy assumption of X. For other possible improvements of second order estimators and their performance we refer the reader to [90] .

The straightforward estimator $\hat{\kappa}_v(\boldsymbol{r})$ of the correlation function $\kappa(\boldsymbol{r})$ is

$$\hat{\kappa}_v(\boldsymbol{r}) = \frac{\hat{\mathrm{cov}}_v(\boldsymbol{r})}{\hat{p}_v(1 - \hat{p}_v)}, \tag{4.13}$$

whenever $0 < \hat{p}_v < 1$ and $\nu_d\big(W \cap (W - \boldsymbol{r})\big) > 0$. This estimator again suffers form the bias of $(\hat{p}_v)^2$. However, both numerator and denominator have the same bias and one can hope that they partly cancel out by division. In this case still an improvement can be made by the use of adapted estimator. Based on the previous intrinsically balanced modification we propose the following estimator of the correlation function given by

$$\hat{\kappa}_v^{\bullet}(\boldsymbol{r}) = \frac{\hat{\mathrm{cov}}_v^{\bullet}(\boldsymbol{r})}{\sqrt{\hat{p}_v[\check{\mathbf{0}}, \boldsymbol{r}]\big(1 - \hat{p}_v[\check{\mathbf{0}}, \boldsymbol{r}]\big)}\sqrt{\hat{p}_v[\mathbf{0}, \check{\boldsymbol{r}}]\big(1 - \hat{p}_v[\mathbf{0}, \check{\boldsymbol{r}}]\big)}} \tag{4.14}$$

for all $\boldsymbol{r} \in \mathbb{R}^d$ such that $\nu_d\big(W \cap (W - \boldsymbol{r})\big) > 0$ and $0 < \hat{p}_v[\check{\mathbf{0}}, \boldsymbol{r}], \hat{p}_v[\mathbf{0}, \check{\boldsymbol{r}}] < 1$. As will be seen in Section 4.4 the performance of this estimator is better than of $\hat{\kappa}_v$, especially for large $\boldsymbol{r}$.

Besides the correlation function, the pair correlation function is usually ([91]) estimated and analysed. However, use it since for random closed sets, in contrary to point processes, it cannot be consistently extended to non-stationary case as discussed in the next section.

4.2.3 Fourier transform based estimation

In practice one often has a digitalized version of part of X observed in a window W. This means that one knows the values of $\mathbb{1}_X$ at a grid of points given by the regular point lattice $L^d = a\mathbb{Z}^d + \boldsymbol{c}, a > 0, \boldsymbol{c} \in \mathbb{R}^d$ restricted to W, where $\mathbb{Z}$ denotes the set of integers, a is called the lattice distance, and $\boldsymbol{c}$ is the lattice shift. The effective way to obtain estimates of the covariance in that case is to use the Fourier transform. This method was introduced in [48] as a part of spectral theory for random closed sets.

To see how the Fourier transform appears in the estimation of the covariance let us focus on the empirical covariance $\hat{C}_v(\boldsymbol{r})$. The numerator of (4.9) can be rewritten as

$$\begin{aligned}
\nu_d\big(X \cap (X - \boldsymbol{r}) \cap W \cap (W - \boldsymbol{r})\big) &= \int_{\mathbb{R}^d} \mathbb{1}_W(\boldsymbol{x})\mathbb{1}_X(\boldsymbol{x})\mathbb{1}_{W-\boldsymbol{r}}(\boldsymbol{x})\mathbb{1}_{X-\boldsymbol{r}}(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} \\
&= \int_{\mathbb{R}^d} \mathbb{1}_W(\boldsymbol{x})\mathbb{1}_X(\boldsymbol{x})\mathbb{1}_W(\boldsymbol{x}+\boldsymbol{r})\mathbb{1}_X(\boldsymbol{x}+\boldsymbol{r})\,\mathrm{d}\boldsymbol{x} \\
&= \int_{\mathbb{R}^d} g(\boldsymbol{x})g(\boldsymbol{x}+\boldsymbol{r})\,\mathrm{d}\boldsymbol{x},
\end{aligned}$$

where $g(\boldsymbol{x}) = \mathbb{1}_W(\boldsymbol{x})\mathbb{1}_X(\boldsymbol{x})$. If we set $f(\boldsymbol{r}) = \nu_d\big(X \cap (X - \boldsymbol{r}) \cap W \cap (W - \boldsymbol{r})\big)$ for every $\boldsymbol{r} \in \mathbb{R}^d$, it is clear that both f and g are bounded and with compact support. Thus $f, g \in L^1(\mathbb{R}^d)$ and the Fourier transform $\mathscr{F}f$ of f reads

$$\begin{aligned}
(\mathscr{F}f)(\boldsymbol{\xi}) &= (2\pi)^{-d/2} \int g(\boldsymbol{x})g(\boldsymbol{x}+\boldsymbol{r})e^{-i\boldsymbol{r}\boldsymbol{\xi}}\,\mathrm{d}\boldsymbol{r}\mathrm{d}\boldsymbol{x} \\
&= (2\pi)^{-d/2} \int g(\boldsymbol{x})e^{i\boldsymbol{x}\boldsymbol{\xi}}g(\boldsymbol{v})e^{-i\boldsymbol{v}\boldsymbol{\xi}}\,\mathrm{d}\boldsymbol{v}\mathrm{d}\boldsymbol{x} \\
&= (2\pi)^{d/2}\overline{(\mathscr{F}g)}(\boldsymbol{\xi})(\mathscr{F}g)(\boldsymbol{\xi}) = (2\pi)^{d/2}\left|\mathscr{F}g\right|^2(\boldsymbol{\xi}).
\end{aligned}$$

Since g is also in $L^2(\mathbb{R}^d)$ it follows that $\mathscr{F}g \in L^2(\mathbb{R}^d)$. The Plancherel theorem (e.g. [92, Theorem IX.6]) implies $\|\mathscr{F}g\|_2 = \|g\|_2$. However, since

$$\|\mathscr{F}g\|_2 = \int_{\mathbb{R}^d} \left|\mathscr{F}g\right|^2(\boldsymbol{\xi})\,\mathrm{d}\boldsymbol{\xi},$$

we have $\left|\mathscr{F}g\right|^2 \in L^1(\mathbb{R}^d)$. Thus the inverse Fourier transform $\mathscr{F}^{-1}$ can be applied yielding

$$f(\boldsymbol{x}) = (2\pi)^{d/2}\big(\mathscr{F}^{-1}\left|\mathscr{F}g\right|^2\big)(\boldsymbol{x}).$$

The empirical covariance $\hat{C}_v(\boldsymbol{r})$ finally reads

$$\hat{C}_v(\boldsymbol{r}) = \frac{(2\pi)^{d/2}\big(\mathscr{F}^{-1}\,|\mathscr{F}g|^2\big)(\boldsymbol{r})}{\gamma_W(\boldsymbol{r})}, \tag{4.15}$$

where $\gamma_W(\boldsymbol{r})$ is the set covariance of W defined by (2.22). The same argumentation as before works for γ_W. Thus it can again be obtained using the Fourier transform as $\gamma_W = (2\pi)^{d/2}\mathscr{F}^{-1}\,|\mathscr{F}\mathbb{1}_W|^2$. The previous formula for $\hat{C}_v(\boldsymbol{r})$ may be used for effective computation of the covariance with the help of the discrete fast Fourier transform (FFT).

Now we formalize the digitalization process and the corresponding discrete estimation procedure. Let us assume that a window W is the closed d-dimensional interval with each dimension larger then some $a > 0$, that is $W = [\boldsymbol{a}, \boldsymbol{b}] \equiv [a_1, b_1] \times \cdots \times [a_d, b_d]$, where $\boldsymbol{a} + a < \boldsymbol{b}$ element-wise. The indicator $\mathbb{1}_X$ of X taken at the points of lattice $L^d = a\mathbb{Z}^d + \boldsymbol{c}$ intersected by W define a d-dimensional array $\boldsymbol{M} \in \{0,1\}^{n_1,\ldots,n_d}$ of zeros and ones with $n_j \in \mathbb{N}, (b_j - a_j)/a - 1 \le n_j \le (b_j - a_j)/a$ for all $j = 1, \ldots, d$. The lattice shift $\boldsymbol{c}$ can be, without loss of generality, chosen such that the elements of $\boldsymbol{M}$ are given by

$$(\boldsymbol{M})_{i_1,\ldots,i_d} = \mathbb{1}_X\big(a(i_1, \ldots, i_d) + \boldsymbol{c}\big)$$

for all $i_j = 1, \ldots, n_j$ and all $j = 1, \ldots, d$. The total number n of grid points equals the total number of elements of $\boldsymbol{M}$ which is given by $n = \Pi_{j=1}^d n_j$.

Let $\boldsymbol{I}$ denote the matrix of the same size as $\boldsymbol{M}$ with all elements equal to one. The discrete version of the estimator $\hat{C}_v(\boldsymbol{r})$ is given by

$$\hat{C}_p(a\boldsymbol{k}) = \frac{\sum_{\boldsymbol{i}\in A(\boldsymbol{k})}(\boldsymbol{M})_{\boldsymbol{i}}(\boldsymbol{M})_{\boldsymbol{i}+\boldsymbol{k}}}{\sum_{\boldsymbol{i}\in A(\boldsymbol{k})}(\boldsymbol{I})_{\boldsymbol{i}}(\boldsymbol{I})_{\boldsymbol{i}+\boldsymbol{k}}}, \tag{4.16}$$

where $\boldsymbol{k} = (k_1, \ldots, k_d), \boldsymbol{i} = (i_1, \ldots, i_d)$, and the index set

$$A(\boldsymbol{k}) = \{(i_1, \ldots, i_d) \in \mathbb{Z}^d | 1 \le i_j \le n_j, 1 \le i_j + k_j \le n_j \text{ for all } j = 1, \ldots, d\}$$

for all $\boldsymbol{k}$ such that the denominator is positive. Note that those $\boldsymbol{k}$ form a set $\{-n_1+1, \ldots, n_1-1\} \times \ldots \times \{-n_d+1, \ldots, n_d-1\}$. It is clear that $\hat{C}_p(a\boldsymbol{k})$ is unbiased and that it has also the desired property $\hat{C}_p(a\boldsymbol{k}) = \hat{C}_p(-a\boldsymbol{k})$ following from the simple observation $A(-\boldsymbol{k}) = A(\boldsymbol{k}) + \boldsymbol{k}$. Since $\boldsymbol{I}$ is a constant matrix, the value of the denominator can be calculated explicitly as $\Pi_1^d\big(n_j - |k_j|\big)$.

If the bounded window W is not a closed interval, then we can take a closed d-dimensional interval D such that $W \subset D$ and repeat the procedure for D. The matrix $\boldsymbol{M}$ now corresponds to the indicator of $X \cap W$ and $\boldsymbol{I}$ to the indicator of W. Formula (4.16) remains valid but now the value of the denominator cannot be in general determined explicitly and may also attain zero values. Hence there may be values of $\boldsymbol{k}$ in $\{-n_1+1, \ldots, n_1-1\} \times \ldots \times \{-n_d+1, \ldots, n_d-1\}$ for which the estimator $\hat{C}_p(a\boldsymbol{k})$ is not defined.

Similarly to $\hat{C}_p$ we may construct discrete versions of $\hat{\mathrm{cov}}_v(\boldsymbol{r})$ and of $\hat{\kappa}_v(\boldsymbol{r})$. They are given by

$$\hat{\mathrm{cov}}_p(a\boldsymbol{k}) = \hat{C}_p(a\boldsymbol{k}) - \hat{p}_p^2$$

and

$$\hat{\kappa}_p(a\boldsymbol{k}) = \frac{\hat{\mathrm{cov}}_p(a\boldsymbol{k})}{\hat{p}_p(1-\hat{p}_p)}, \tag{4.17}$$

where $\hat{p}_p$ is the discrete empirical volume fraction given according to (4.8) by

$$\hat{p}_p = \frac{\sum_{\boldsymbol{i}=(1,1)}^{\boldsymbol{n}}(\boldsymbol{M})_{\boldsymbol{i}}}{\sum_{\boldsymbol{i}=(1,1)}^{\boldsymbol{n}}(\boldsymbol{I})_{\boldsymbol{i}}} = \frac{1}{n_1 \cdot n_2}\sum_{\boldsymbol{i}=(1,1)}^{\boldsymbol{n}}(\boldsymbol{M})_{\boldsymbol{i}}. \tag{4.18}$$

To establish a version of (4.15) using discrete Fourier transform one must take care of the periodicity assumption in the discrete Fourier transform that causes an overlapping effect (edge-effect), see [93, Section 6.4]. This effect can be eliminated by expanding the matrices $\boldsymbol{M}$ and $\boldsymbol{I}$ to the window $2W$ (resp. $2D$):

$$(\tilde{\boldsymbol{M}})_{\boldsymbol{i}} = \begin{cases} (\boldsymbol{M})_{\boldsymbol{i}}, & \text{if } 1 \le \boldsymbol{i} \le \boldsymbol{n}, \\ 0, & \text{otherwise,} \end{cases} \quad \text{and} \quad (\tilde{\boldsymbol{I}})_{\boldsymbol{i}} = \begin{cases} (\boldsymbol{I})_{\boldsymbol{i}}, & \text{if } 1 \le \boldsymbol{i} \le \boldsymbol{n}, \\ 0, & \text{otherwise,} \end{cases}$$

i.e. the original matrices are padded with zeros. This increases the number of sample points to $2^d n$. The discrete version of (4.15) now reads

$$\hat{C}_p(a\boldsymbol{k}) = \frac{\mathrm{iDFT}\big(|\mathrm{DFT}(\tilde{\boldsymbol{M}})|^{\star 2}\big)(a\boldsymbol{k})}{\mathrm{iDFT}\big(|\mathrm{DFT}(\tilde{\boldsymbol{I}})|^{\star 2}\big)(a\boldsymbol{k})}, \tag{4.19}$$

where $^{\star 2}$ means the element-wise square of a matrix. The discrete Fourier transform (DFT) and its inverse (iDFT) are given by

$$\begin{aligned} (\mathrm{DFT}(\boldsymbol{M}))_{\boldsymbol{k}} &= \sum_{\boldsymbol{j} \in A(\boldsymbol{0})} e^{-2\pi i(\boldsymbol{k}-1)\cdot\frac{(\boldsymbol{j}-1)}{\boldsymbol{n}}} \boldsymbol{M}_{\boldsymbol{j}} \quad \text{for all} \quad \boldsymbol{k} \in A(\boldsymbol{0}), \\ (\mathrm{iDFT}(\boldsymbol{N}))_{\boldsymbol{j}} &= \frac{1}{\Pi_{l=1}^d n_l} \sum_{\boldsymbol{k} \in A(\boldsymbol{0})} e^{2\pi i(\boldsymbol{j}-1)\cdot\frac{(\boldsymbol{k}-1)}{\boldsymbol{n}}} \boldsymbol{N}_{\boldsymbol{k}} \quad \text{for all} \quad \boldsymbol{j} \in A(\boldsymbol{0}), \end{aligned}$$

respectively, where $\cdot$ is the standard scalar product on $\mathbb{R}^d$ and $(\boldsymbol{j}-1)/\boldsymbol{n}$ denotes the element-wise division, $(\boldsymbol{j}-1)/\boldsymbol{n} = ((j_1-1)/n_1, \dots, (j_d-1)/n_d)$.

If $\boldsymbol{M}$ is a matrix with n elements, the covariance can be computed by the use of the fast Fourier transform with a complexity in $\mathcal{O}(n \log n)$. This is a considerable gain compared to the usual estimation of the covariance with a complexity $\mathcal{O}(n^2)$.

The similar procedure can be used for all second order estimators from the previous subsection. Hence the discrete versions $\hat{\mathrm{cov}}_p^{\bullet}(\boldsymbol{r})$ and $\hat{\kappa}_p^{\bullet}(\boldsymbol{r})$ of intrinsically balanced estimators $\hat{\mathrm{cov}}_v^{\bullet}(\boldsymbol{r})$ and $\hat{\kappa}_v^{\bullet}(\boldsymbol{r})$, defined by (4.10) and (4.14), are given by

$$\hat{\kappa}_p^{\bullet}(a\boldsymbol{k}) = \hat{C}_p(a\boldsymbol{k}) - \hat{p}_p[\check{\boldsymbol{0}}, a\boldsymbol{k}]\hat{p}_p[\boldsymbol{0}, a\check{\boldsymbol{k}}]$$

and

$$\hat{\kappa}_p^{\bullet}(a\boldsymbol{k}) = \frac{\hat{C}_p(a\boldsymbol{k}) - \hat{p}_p[\check{\boldsymbol{0}}, a\boldsymbol{k}]\hat{p}_p[\boldsymbol{0}, a\check{\boldsymbol{k}}]}{\sqrt{\hat{p}_p[\check{\boldsymbol{0}}, a\boldsymbol{k}]\big(1 - \hat{p}_p[\check{\boldsymbol{0}}, a\boldsymbol{k}]\big)}\sqrt{\hat{p}_p[\boldsymbol{0}, a\check{\boldsymbol{k}}]\big(1 - \hat{p}_p[\boldsymbol{0}, a\check{\boldsymbol{k}}]\big)}}, \tag{4.20}$$

respectively, where

$$\hat{p}_p[\check{\boldsymbol{0}}, a\boldsymbol{k}] = \frac{\sum_{\boldsymbol{i} \in A(\boldsymbol{k})} (\boldsymbol{M})_{\boldsymbol{i}} (\boldsymbol{I})_{\boldsymbol{i}+\boldsymbol{k}}}{\sum_{\boldsymbol{i} \in A(\boldsymbol{k})} (\boldsymbol{I})_{\boldsymbol{i}} (\boldsymbol{I})_{\boldsymbol{i}+\boldsymbol{k}}}, \qquad \hat{p}_p[\boldsymbol{0}, a\check{\boldsymbol{k}}] = \frac{\sum_{\boldsymbol{i} \in A(\boldsymbol{k})} (\boldsymbol{I})_{\boldsymbol{i}} (\boldsymbol{M})_{\boldsymbol{i}+\boldsymbol{k}}}{\sum_{\boldsymbol{i} \in A(\boldsymbol{k})} (\boldsymbol{I})_{\boldsymbol{i}} (\boldsymbol{I})_{\boldsymbol{i}+\boldsymbol{k}}} \tag{4.21}$$

are discrete versions of $\hat{p}_v[\check{\boldsymbol{0}}, \boldsymbol{r}], \hat{p}_v[\boldsymbol{0}, \check{\boldsymbol{r}}]$. We may again use the Fourier transform to calculate $\hat{p}_p[\check{\boldsymbol{0}}, a\boldsymbol{k}]$ and $\hat{p}_p[\boldsymbol{0}, a\check{\boldsymbol{k}}]$. By the same argumentation as for the numerator of $\hat{C}_v(\boldsymbol{r})$ one obtains

$$\begin{aligned} \nu_d\big(X \cap W \cap (W - \boldsymbol{r})\big) &= (2\pi)^{d/2}\big(\mathscr{F}^{-1}(\overline{\mathscr{F} g}\, \mathscr{F} \mathbb{1}_W)\big)(\boldsymbol{r}), \\ \nu_d\big((X - \boldsymbol{r}) \cap W \cap (W - \boldsymbol{r})\big) &= (2\pi)^{d/2}\big(\mathscr{F}^{-1}(\overline{\mathscr{F} \mathbb{1}_W}\, \mathscr{F} g)\big)(\boldsymbol{r}), \end{aligned}$$

where again $g(\boldsymbol{x}) = \mathbb{1}_W(\boldsymbol{x})\mathbb{1}_X(\boldsymbol{x})$. Hence

$$\begin{aligned} \hat{p}_v[\check{\boldsymbol{0}}, \boldsymbol{r}] &= \frac{\big(\mathscr{F}^{-1}(\overline{\mathscr{F} g}\, \mathscr{F} \mathbb{1}_W)\big)(\boldsymbol{r})}{\mathscr{F}^{-1}\,|\mathscr{F} \mathbb{1}_W|^2}, \\ \hat{p}_v[\boldsymbol{0}, \check{\boldsymbol{r}}] &= \frac{\big(\mathscr{F}^{-1}(\overline{\mathscr{F} \mathbb{1}_W}\, \mathscr{F} g)\big)(\boldsymbol{r})}{\mathscr{F}^{-1}\,|\mathscr{F} \mathbb{1}_W|^2}, \end{aligned}$$

and in the discrete version:

$$\hat{p}_p[\check{\mathbf{0}}, a\boldsymbol{k}] = \frac{\mathrm{iDFT}\big(\overline{\mathrm{DFT}(\tilde{\boldsymbol{M}})} \star \mathrm{DFT}(\tilde{\boldsymbol{I}})\big)(a\boldsymbol{k})}{\mathrm{iDFT}\big(|\mathrm{DFT}(\tilde{\boldsymbol{I}})|^{\star 2}\big)(a\boldsymbol{k})}, \tag{4.22}$$

$$\hat{p}_p[\mathbf{0}, a\check{\boldsymbol{k}}] = \frac{\mathrm{iDFT}\big(\overline{\mathrm{DFT}(\tilde{\boldsymbol{I}})} \star \mathrm{DFT}(\tilde{\boldsymbol{M}})\big)(a\boldsymbol{k})}{\mathrm{iDFT}\big(|\mathrm{DFT}(\tilde{\boldsymbol{I}})|^{\star 2}\big)(a\boldsymbol{k})}, \tag{4.23}$$

where $\tilde{\boldsymbol{M}}, \tilde{\boldsymbol{I}}$ have the same meaning as before and $\star$ denotes element-wise multiplication of matrices. More detailed discussion about the use of Fourier transform in estimation of characteristics of random closed sets can be found in [50].

If the random closed set X is assumed to be isotropic, we would like to obtain estimates of the covariance, covariance function, or correlation function that depends only on the scalar parameter r. In the digitalized version of $X \cap W$ we cannot directly use the isotropic modifications as in (4.12) since we cannot calculate the isotropised set covariance. The way to overcome this problem is to use the kernel estimates analogously as for second order point processes in part 4.1.1. We demonstrate this method on the example of the covariance $C(r)$. Let assume that we have calculated the discrete version $\hat{C}_p(a\boldsymbol{k})$ of its estimator at all admissible points $a\boldsymbol{k}$. The isotropic modification is then

$$\hat{C}_p^I(r) = \frac{\sum_{\boldsymbol{k}} K_h\big(\,\|a\boldsymbol{k}\| - r\big)\hat{C}_p(a\boldsymbol{k})}{\sum_{\boldsymbol{k}} K_h\big(\,\|a\boldsymbol{k}\| - r\big)},$$

where the sum is taken over all admissible $\boldsymbol{k}$ and $K_h : \mathbb{R} \to \mathbb{R}$ is the kernel function defined by $K_h(x) = h^{-1}K(h^{-1}x)$ for some probability density function K called the **kernel** and $h > 0$ called the **bandwidth**. $\hat{C}_p^I(r)$ is defined for all r such that the denominator is positive. It is not unbiased but has the following property

$$\mathbb{E}\,\hat{C}_p^I(r) = \frac{\sum_{\boldsymbol{k}} K_h\big(\,\|a\boldsymbol{k}\| - r\big)C(a\,\|\boldsymbol{k}\|)}{\sum_{\boldsymbol{k}} K_h\big(\,\|a\boldsymbol{k}\| - r\big)}.$$

The estimator therefore depends on the choice of kernel and the bandwidth. The choice of kernel is usually of second importance since the smoothness of the estimate is mainly determined by the value of the bandwidth. The larger the bandwidth the smoother the estimator as a function of r. However, large value of bandwidth can produce larger systematic error as can be seen from the last relation.

In practical applications one can often assume that the covariance is decreasing and more flat with increasing r. In that case it is useful to use the so called adapted version of the estimator (see e.g. [76, Section 5.3])

$$\hat{C}_p^{AI}(r) = \frac{\sum_{\boldsymbol{k}} K_{h(r)}\big(\,\|a\boldsymbol{k}\| - r\big)\hat{C}_p(a\boldsymbol{k})}{\sum_{\boldsymbol{k}} K_{h(r)}\big(\,\|a\boldsymbol{k}\| - r\big)},$$

with bandwidth being a function of r chosen to properly increase with increasing r. Thus for large r, where the statistics is poor due to small number of observations included in the estimation, there is a large smoothing effect that brings acceptable systematic error, because the true covariance is flat enough. On the other hand, for small r, where the statistics is sufficiently good, the smoothing effect and the corresponding systematic error are small. In practical applications we often use the adapted isotropic covariance function estimator $\hat{\kappa}_p^{AI}(r)$ defined analogously by

$$\hat{\kappa}_p^{AI}(r) = \frac{\sum_{\boldsymbol{k}} K_{h(r)}\big(\,\|a\boldsymbol{k}\| - r\big)\hat{\kappa}_p(a\boldsymbol{k})}{\sum_{\boldsymbol{k}} K_{h(r)}\big(\,\|a\boldsymbol{k}\| - r\big)}. \tag{4.24}$$

4.3 Non-stationary random closed sets

The statistics of non-stationary random closed sets based on one sample is much more complicated compared to the stationary case. It is because one cannot use ergodicity to prove the consistency of estimators. Actually, even the unbiasedness is usually not obtained.

In this section we introduce a non-parametric kernel method for the estimation of the volume fraction and finally discuss the estimation of second order characteristics.

4.3.1 Volume fraction

Let assume that X is a random closed set observed in W with volume fraction $m(\boldsymbol{x})$. For W we further assume that $0 < \nu_d(W)$ and $\nu_d\big(W \cap B_r(\boldsymbol{x})\big) > 0$ for every $\boldsymbol{x} \in W$ and every $r > 0$, where $B_r(\boldsymbol{x})$ is the closed ball of radius r centred at $\boldsymbol{x}$. Thus every neighbourhood of a point in W has a positive volume inside W.

Without any parametric knowledge about the form of m, the most natural way to estimate m is to use a kernel based estimator. Similarly to (4.1) we take (see [94])

$$\hat{m}_h(\boldsymbol{x}) = \frac{1}{e_h(\boldsymbol{x})} \int_W \mathbb{1}_X(\boldsymbol{u}) K_h(\boldsymbol{x} - \boldsymbol{u}) \,\mathrm{d}\boldsymbol{u} \tag{4.25}$$

for all $\boldsymbol{x} \in W$, where $h > 0$ is called the **bandwidth**, $K_h : \mathbb{R}^d \to \mathbb{R}$ is the kernel function defined by $K_h(\boldsymbol{x}) = h^{-d} K(h^{-1}\boldsymbol{x})$ for some radially symmetric multivariate probability density function K called the **kernel** and $e_h(\boldsymbol{x})$ is the edge correction factor given for all $\boldsymbol{x} \in W$ by

$$e_h(\boldsymbol{x}) = \int_W K_h(\boldsymbol{x} - \boldsymbol{u}) \,\mathrm{d}\boldsymbol{u}.$$

The estimator $\hat{m}_h$ is not unbiased, except the case of constant m, since for its expectation holds

$$\mathbb{E}\,\hat{m}_h(\boldsymbol{x}) = \frac{1}{e_h(\boldsymbol{x})} \int_W m(\boldsymbol{u}) K_h(\boldsymbol{x} - \boldsymbol{u}) \,\mathrm{d}\boldsymbol{u}. \tag{4.26}$$

If X is stationary then $m(\boldsymbol{x}) = m$ is constant and $\mathbb{E}\,\hat{m}_h(\boldsymbol{x}) = m$ for all $h > 0$ and all admissible $\boldsymbol{x}$.

Proposition 4.3.1. *Let X be a random closed set with volume fraction $m(\boldsymbol{x})$ continuous at $\boldsymbol{x} \in W^\circ$. Then $\mathbb{E}\,\hat{m}_h(\boldsymbol{x}) \to m(\boldsymbol{x})$ as $h \to 0_+$.*

Proof. Since $\boldsymbol{x} \in W^\circ$ then $\boldsymbol{x}$ is in W together with its neighbourhood. Let take r such that $B_r(\boldsymbol{x}) \subset W$. Then by the integrability and symmetry, $K(\boldsymbol{x}) = K(-\boldsymbol{x})$ for all $\boldsymbol{x}$, of K one can choose h_0 such that for all $h < h_0$,

$$e_h(\boldsymbol{x}) \geq \int_{B_r(\boldsymbol{x})} K_h(\boldsymbol{x} - \boldsymbol{u}) \,\mathrm{d}\boldsymbol{u} = \int_{B_{r/h}(\boldsymbol{0})} K(\boldsymbol{y}) \,\mathrm{d}\boldsymbol{y} > \frac{1}{2}.$$

From the continuity of m at $\boldsymbol{x}$, given $\varepsilon > 0$ one can choose $\delta < r$ such that for all $\boldsymbol{u}$ with $\|\boldsymbol{u} - \boldsymbol{x}\| \leq \delta$,

$$|m(\boldsymbol{u}) - m(\boldsymbol{x})| < \frac{\varepsilon}{4}.$$

Moreover, one can choose $h_1 < h_0$ such that for all $h < h_1$,

$$\int_{\{\boldsymbol{u} \,|\, \|\boldsymbol{x} - \boldsymbol{u}\| > \delta\}} K_h(\boldsymbol{x} - \boldsymbol{u}) \,\mathrm{d}\boldsymbol{u} < \frac{\varepsilon}{8}.$$

Thus

$$\int_W K_h(\boldsymbol{x}-\boldsymbol{u})\,|m(\boldsymbol{u})-m(\boldsymbol{x})|\,\mathrm{d}\boldsymbol{u} \le 2\int_{W\cap\{\boldsymbol{u}\,|\,\|\boldsymbol{x}-\boldsymbol{u}\|>\delta\}} K_h(\boldsymbol{x}-\boldsymbol{u})\,\mathrm{d}\boldsymbol{u} + \int_{W\cap\{\boldsymbol{u}\,|\,\|\boldsymbol{x}-\boldsymbol{u}\|\le\delta\}} K_h(\boldsymbol{x}-\boldsymbol{u})\,|m(\boldsymbol{u})-m(\boldsymbol{x})|\,\mathrm{d}\boldsymbol{u} < \frac{\varepsilon}{4}+\frac{\varepsilon}{4}.$$

Finally

$$\frac{1}{e_h(\boldsymbol{x})}\left|\int_W K_h(\boldsymbol{x}-\boldsymbol{u})m(\boldsymbol{u})\,\mathrm{d}\boldsymbol{u} - e_h(\boldsymbol{x})m(\boldsymbol{x})\right| \le 2\int_W K_h(\boldsymbol{x}-\boldsymbol{u})\,|m(\boldsymbol{u})-m(\boldsymbol{x})|\,\mathrm{d}\boldsymbol{u} < \varepsilon,$$

which completes the proof. □

By Proposition 2.3.3 we obtain the following corollary.

Corollary 4.3.1. *Let X be a P-continuous random closed set. Then $\mathbb{E}\,\hat{m}_h(\boldsymbol{x}) \to m(\boldsymbol{x})$ as $h \to 0_+$ for all $\boldsymbol{x} \in W^\circ$.*

The error of the estimation at point $\boldsymbol{x} \in W$ is usually measured by a mean square error (MSE) given by

$$\mathrm{MSE}\big(\hat{m}_h(\boldsymbol{x})\big) = \mathbb{E}\,\big(\hat{m}_h(\boldsymbol{x}) - m(\boldsymbol{x})\big)^2.$$

The cumulative value over W is given by the mean integrated square error (MISE),

$$\mathrm{MISE}\big(\hat{m}_h\big) = \mathbb{E}\int_W \big(\hat{m}_h(\boldsymbol{x}) - m(\boldsymbol{x})\big)^2\,\mathrm{d}\boldsymbol{x}.$$

It is convenient to split $\mathrm{MISE}\big(\hat{m}_h\big)$ into two parts corresponding to the bias and variance of $\hat{m}_h$:

$$\mathrm{MISE}\big(\hat{m}_h\big) = \int_W \big(\,\mathbb{E}\,\hat{m}_h(\boldsymbol{x}) - m(\boldsymbol{x})\big)^2\,\mathrm{d}\boldsymbol{x} + \int_W \operatorname{var}\hat{m}_h(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}. \tag{4.27}$$

By the previous proposition, one may assume that the first part corresponding to the bias is small for small h. Let us observe the second part. We have

$$\begin{aligned}\operatorname{var}\hat{m}_h(\boldsymbol{x}) &= \frac{1}{e_h^2(\boldsymbol{x})}\,\mathbb{E}\left(\int_W \big(\mathbb{1}_X(\boldsymbol{u}) - m(\boldsymbol{u})\big)K_h(\boldsymbol{x}-\boldsymbol{u})\,\mathrm{d}\boldsymbol{u}\right)^2 \\ &= \frac{1}{e_h^2(\boldsymbol{x})}\int_W\int_W \operatorname{cov}(\boldsymbol{u},\boldsymbol{v})K_h(\boldsymbol{x}-\boldsymbol{u})K_h(\boldsymbol{x}-\boldsymbol{v})\,\mathrm{d}\boldsymbol{u}\mathrm{d}\boldsymbol{v}. \end{aligned}\tag{4.28}$$

Proposition 4.3.2. *Let X be a random closed set with $\operatorname{cov}(\boldsymbol{x},\boldsymbol{y})$ continuous in $\mathbb{R}^d\times\mathbb{R}^d$. Then $\operatorname{var}\hat{m}_h(\boldsymbol{x}) \to \operatorname{cov}(\boldsymbol{x},\boldsymbol{x})$ as $h \to 0_+$ for all $\boldsymbol{x} \in W^\circ$.*

Proof. From Theorem 2.3.1 follows $2\,|\operatorname{cov}(\boldsymbol{x},\boldsymbol{y})| \le \operatorname{cov}(\boldsymbol{x},\boldsymbol{x}) + \operatorname{cov}(\boldsymbol{y},\boldsymbol{y})$ for all $\boldsymbol{x},\boldsymbol{y} \in \mathbb{R}^d$ yielding $|\operatorname{cov}(\boldsymbol{x},\boldsymbol{y})| \le 1/4$ for all $\boldsymbol{x},\boldsymbol{y} \in \mathbb{R}^d$. Now we continue in the similar way as in the proof of Proposition 4.3.1. One can take r such that $B_r(\boldsymbol{x}) \equiv \{\boldsymbol{u}\,|\,\|\boldsymbol{u}-\boldsymbol{x}\| \le r\} \subset W$ and h_0 such that for all $h < h_0$,

$$e_h(\boldsymbol{x}) \ge \int_{B_r(\boldsymbol{x})} K_h(\boldsymbol{x}-\boldsymbol{u})\,\mathrm{d}\boldsymbol{u} = \int_{B_{r/h}(\boldsymbol{0})} K(\boldsymbol{y})\,\mathrm{d}\boldsymbol{y} > \frac{1}{2}.$$

From the continuity of cov at $(\boldsymbol{x},\boldsymbol{x})$, given $\varepsilon > 0$ one can choose $\delta < r$ such that for all $\boldsymbol{u},\boldsymbol{v}$ with $\|\boldsymbol{u}-\boldsymbol{x}\| \le \delta$ and $\|\boldsymbol{v}-\boldsymbol{x}\| \le \delta$,

$$|\mathrm{cov}(\boldsymbol{u},\boldsymbol{v}) - \mathrm{cov}(\boldsymbol{x},\boldsymbol{x})| < \frac{\varepsilon}{4}.$$

Moreover one can choose $h_1 < h_0$ such that for all $h < h_1$,

$$\int\limits_{\{\boldsymbol{u}\,|\,\|\boldsymbol{x}-\boldsymbol{u}\|>\delta\}} K_h(\boldsymbol{x}-\boldsymbol{u})\,\mathrm{d}\boldsymbol{u} < \frac{\varepsilon}{4}.$$

Thus

$$\begin{aligned}
&\int\limits_W \int\limits_W |\mathrm{cov}(\boldsymbol{u},\boldsymbol{v}) - \mathrm{cov}(\boldsymbol{x},\boldsymbol{x})|\, K_h(\boldsymbol{x}-\boldsymbol{u})K_h(\boldsymbol{x}-\boldsymbol{v})\,\mathrm{d}\boldsymbol{u}\mathrm{d}\boldsymbol{v} \\
&\qquad = \int\limits_{B_\delta(\boldsymbol{x})} \int\limits_{B_\delta(\boldsymbol{x})} |\mathrm{cov}(\boldsymbol{u},\boldsymbol{v}) - \mathrm{cov}(\boldsymbol{x},\boldsymbol{x})|\, K_h(\boldsymbol{x}-\boldsymbol{u})K_h(\boldsymbol{x}-\boldsymbol{v})\,\mathrm{d}\boldsymbol{u}\mathrm{d}\boldsymbol{v} \\
&\qquad + \frac{2}{4} \iint\limits_{\{(\boldsymbol{u},\boldsymbol{v})\in W\times W\,|\,\|\boldsymbol{u}-\boldsymbol{x}\|>\delta \text{ or } \|\boldsymbol{v}-\boldsymbol{x}\|>\delta\}} K_h(\boldsymbol{x}-\boldsymbol{u})K_h(\boldsymbol{x}-\boldsymbol{v})\,\mathrm{d}\boldsymbol{u}\mathrm{d}\boldsymbol{v} \\
&\qquad < \frac{\varepsilon}{4} + \int\limits_{W\cap\{\boldsymbol{u}\,|\,\|\boldsymbol{u}-\boldsymbol{x}\|>\delta\}} \int\limits_W K_h(\boldsymbol{x}-\boldsymbol{u})K_h(\boldsymbol{x}-\boldsymbol{v})\,\mathrm{d}\boldsymbol{u}\mathrm{d}\boldsymbol{v} < \frac{\varepsilon}{4} + \frac{\varepsilon}{4}.
\end{aligned}$$

Finally,

$$\begin{aligned}
|\mathrm{var}\,\hat{m}_h(\boldsymbol{x}) - \mathrm{cov}(\boldsymbol{x},\boldsymbol{x})| &\le \frac{1}{e_h^2(\boldsymbol{x})} \int\limits_W \int\limits_W |\mathrm{cov}(\boldsymbol{u},\boldsymbol{v}) - \mathrm{cov}(\boldsymbol{x},\boldsymbol{x})|\, K_h(\boldsymbol{x}-\boldsymbol{u})K_h(\boldsymbol{x}-\boldsymbol{v})\,\mathrm{d}\boldsymbol{u}\mathrm{d}\boldsymbol{v} \\
&\le 4 \int\limits_W \int\limits_W |\mathrm{cov}(\boldsymbol{u},\boldsymbol{v}) - \mathrm{cov}(\boldsymbol{x},\boldsymbol{x})|\, K_h(\boldsymbol{x}-\boldsymbol{u})K_h(\boldsymbol{x}-\boldsymbol{v})\,\mathrm{d}\boldsymbol{u}\mathrm{d}\boldsymbol{v} < \varepsilon
\end{aligned}$$

for all $h < h_1$, which completes the proof. □

Proposition 2.3.2 yields the following corollary.

Corollary 4.3.2. *Let X be a P-continuous random closed set. Then* $\mathrm{var}\,\hat{m}_h(\boldsymbol{x}) \to \mathrm{cov}(\boldsymbol{x},\boldsymbol{x})$ *as $h \to 0_+$ for all $\boldsymbol{x} \in W^\circ$.*

Let assume that $m(\boldsymbol{x})$ and $\mathrm{cov}(\boldsymbol{x},\boldsymbol{y})$ are continuous and $\nu_d(W) < \infty$. Then by Proposition 4.3.1 the first integrand in (4.27) have pointwise limit 0 almost surely in W and by Proposition 4.3.2 the second integrand of (4.27) have pointwise limit $\mathrm{cov}(\boldsymbol{x},\boldsymbol{x})$ almost surely in W. Both can be bounded by a constant and hence the Lebesgue dominated convergence theorem yields

$$\mathrm{MISE}\big(\hat{m}_h\big) \to \int\limits_W \mathrm{cov}(\boldsymbol{x},\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} \quad \text{as} \quad h \to 0_+.$$

On the other side under the assumption of continuity of K at $\mathbf{0}$ we obtain

$$\hat{m}_h(\boldsymbol{x}) \to \frac{1}{\nu_d(W)} \int\limits_W \mathbb{1}_X(\boldsymbol{u})\,\mathrm{d}\boldsymbol{u} \quad \text{as} \quad h \to \infty$$

by the Lebesgue dominated convergence theorem. The limit is the well known empirical volume fraction $\hat{p}_v$ given by (4.2) that was used in Subsection 4.2.1 as an estimator of the volume

fraction for stationary random closed sets. If we denote $\bar{m}_W = \nu_d^{-1}(W) \int_W m(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}$ then clearly $\mathbb{E}\,\hat{m}_h(\boldsymbol{x}) \to \bar{m}_W$ as $h \to \infty$. For the first part of (4.27) we have

$$\int_W \big(\mathbb{E}\,\hat{m}_h(\boldsymbol{x}) - m(\boldsymbol{x})\big)^2\,\mathrm{d}\boldsymbol{x} \to \int_W \big(m^2(\boldsymbol{x}) - \bar{m}_W^2\big)\,\mathrm{d}\boldsymbol{x} \quad \text{as} \quad h \to \infty.$$

Similarly

$$\operatorname{var} \hat{m}_h(\boldsymbol{x}) \to \frac{1}{\nu_d^2(W)} \int_W \int_W \operatorname{cov}(\boldsymbol{u}, \boldsymbol{v})\,\mathrm{d}\boldsymbol{u}\mathrm{d}\boldsymbol{v} \quad \text{as} \quad h \to \infty.$$

Finally, for the MISE holds

$$\operatorname{MISE}(\hat{m}_h) \to \int_W \big(m^2(\boldsymbol{x}) - \bar{m}_W^2\big)\,\mathrm{d}\boldsymbol{x} + \frac{1}{\nu_d(W)} \int_W \int_W \operatorname{cov}(\boldsymbol{u}, \boldsymbol{v})\,\mathrm{d}\boldsymbol{u}\mathrm{d}\boldsymbol{v} \quad \text{as} \quad h \to \infty.$$

Since from Theorem 2.3.1 follows $|\operatorname{cov}(\boldsymbol{u}, \boldsymbol{v})| \le \frac{\operatorname{cov}(\boldsymbol{u},\boldsymbol{u}) + \operatorname{cov}(\boldsymbol{v},\boldsymbol{v})}{2}$ we have

$$\frac{1}{\nu_d(W)} \int_W \int_W \operatorname{cov}(\boldsymbol{u}, \boldsymbol{v})\,\mathrm{d}\boldsymbol{u}\mathrm{d}\boldsymbol{v} \le \int_W \operatorname{cov}(\boldsymbol{x}, \boldsymbol{x})\,\mathrm{d}\boldsymbol{x}.$$

Moreover, the left side is for many covariance functions much smaller than the right side, since the value of $\operatorname{cov}(\boldsymbol{u}, \boldsymbol{v})$ is usually highly decreasing as $\|\boldsymbol{u} - \boldsymbol{v}\|$ growth. This suggests that the integrated variance of $\hat{m}_h$ given by the second part in (4.27) is large for small h, the phenomenon well known in usual kernel density estimation in statistics, see e.g. [76, 80]. The opposite behaviour holds for the integrated bias represented by the first part of (4.27). Thus it is small with small h and eventually approaching 0 as $h \to 0_+$.

The main task when using the estimator $\hat{m}_h$ is the proper choice of bandwidth h, since it crucially influences the estimation result. The trade-off between the large possible variation for small h and the large possible bias for large h is usually solved by taking a value that minimizes the integrated mean square error. Such value h_o is called the **optimal bandwidth**. That is

$$h_o = \arg\min_h \operatorname{MISE}(\hat{m}_h)$$

if the global minimum is attained for $h \in (0, \infty)$, $h_o = 0$ if $\lim_{h\to 0_+} \operatorname{MISE}(\hat{m}_h) < \operatorname{MISE}(\hat{m}_t)$ for all t, and $h_o = \infty$ if $\lim_{h\to\infty} \operatorname{MISE}(\hat{m}_h) < \operatorname{MISE}(\hat{m}_t)$ for all t.

In the following we present an explicit asymptotic calculation in a special case.

Proposition 4.3.3. *Let K be the Gaussian kernel in $\mathbb{R}^2$ given by*

$$K(\boldsymbol{x}) = \frac{1}{(2\pi)} e^{-\frac{\boldsymbol{x}^2}{2}}.$$

Further, let X be a random closed set in $\mathbb{R}^2$ satisfying the following conditions:

(a) the volume fraction is

$$m(\boldsymbol{x}) = Ba^2(2\pi)K_a(\boldsymbol{x}) = Be^{-\frac{\boldsymbol{x}^2}{2a^2}},$$

where $B > 0, a > 0$,

(b) the covariance function is

$$\operatorname{cov}(\boldsymbol{x}, \boldsymbol{y}) = e^{-\alpha\|\boldsymbol{x}-\boldsymbol{y}\|} \sqrt{m(\boldsymbol{x}) - m(\boldsymbol{x})^2} \sqrt{m(\boldsymbol{y}) - m(\boldsymbol{y})^2},$$

where $\alpha > 0$.

Then for $a \gg 1$, $a\alpha \gg 1$ and $W = \mathbb{R}^2$ the optimal bandwidth is approximately given by

$$h_o \approx \frac{a^{2/3}}{\alpha^{1/3}} \left(\frac{2-B}{2B} \right)^{1/6}.$$

Proof. The relation (4.27) can be further rewritten as

$$\begin{aligned}
\mathrm{MISE}\big(\hat{m}_h\big) = &\int_{\mathbb{R}^2} \frac{1}{e_h^2(\boldsymbol{x})} \int_{\mathbb{R}^2}\int_{\mathbb{R}^2} \mathrm{cov}(\boldsymbol{u},\boldsymbol{v}) K_h(\boldsymbol{x}-\boldsymbol{u}) K_h(\boldsymbol{x}-\boldsymbol{v})\,\mathrm{d}\boldsymbol{u}\mathrm{d}\boldsymbol{v}\mathrm{d}\boldsymbol{x} \\
&+ \int_{\mathbb{R}^2} \frac{1}{e_h^2(\boldsymbol{x})} \int_{\mathbb{R}^2}\int_{\mathbb{R}^2} m(\boldsymbol{u}) m(\boldsymbol{v}) K_h(\boldsymbol{x}-\boldsymbol{u}) K_h(\boldsymbol{x}-\boldsymbol{v})\,\mathrm{d}\boldsymbol{u}\mathrm{d}\boldsymbol{v}\mathrm{d}\boldsymbol{x} \\
&- \int_{\mathbb{R}^2} \frac{2m(\boldsymbol{x})}{e_h(\boldsymbol{x})} \int_{\mathbb{R}^2} m(\boldsymbol{u}) K_h(\boldsymbol{x}-\boldsymbol{u})\,\mathrm{d}\boldsymbol{u}\mathrm{d}\boldsymbol{x} + \int_{\mathbb{R}^2} m^2(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}.
\end{aligned}$$

Using assumptions (a), (b), and the fact that $e_h(\boldsymbol{x}) = 1$ for all $\boldsymbol{x} \in \mathbb{R}^2$ we obtain

$$\begin{aligned}
\mathrm{MISE}\big(\hat{m}_h\big) \approx &\iiint e^{-\alpha\|\boldsymbol{u}-\boldsymbol{v}\|} \sqrt{m(\boldsymbol{u}) - m(\boldsymbol{u})^2}\sqrt{m(\boldsymbol{v}) - m(\boldsymbol{v})^2} K_h(\boldsymbol{x}-\boldsymbol{u}) K_h(\boldsymbol{x}-\boldsymbol{v})\,\mathrm{d}\boldsymbol{u}\mathrm{d}\boldsymbol{v}\mathrm{d}\boldsymbol{x} \\
&+ B^2 4\pi^2 a^4 \iiint K_a(\boldsymbol{u}) K_a(\boldsymbol{v}) K_h(\boldsymbol{x}-\boldsymbol{u}) K_h(\boldsymbol{x}-\boldsymbol{v})\,\mathrm{d}\boldsymbol{u}\mathrm{d}\boldsymbol{v}\mathrm{d}\boldsymbol{x} \\
&- B^2 8\pi^2 a^4 \iint K_a(\boldsymbol{u}) K_a(\boldsymbol{x}) K_h(\boldsymbol{x}-\boldsymbol{u})\,\mathrm{d}\boldsymbol{u}\mathrm{d}\boldsymbol{x} \\
&+ B^2 4\pi^2 a^4 \int K_a^2(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}. \qquad (4.29)
\end{aligned}$$

The last three terms can be easily calculated using well known relations for Gaussians:

$$\begin{aligned}
K_\alpha(\boldsymbol{x}) K_\beta(\boldsymbol{x}) &= \frac{1}{2\pi(\alpha^2+\beta^2)} K_{\frac{\alpha\beta}{\sqrt{\alpha^2+\beta^2}}}(\boldsymbol{x}), \\
\sqrt{K_\alpha(\boldsymbol{x})} &= 2\sqrt{2\pi}\alpha K_{\sqrt{2}\alpha}(\boldsymbol{x}), \\
\int_{\mathbb{R}^2} K_\alpha(\boldsymbol{x}) K_\beta(\boldsymbol{y}-\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} &= K_{\sqrt{\alpha^2+\beta^2}}(\boldsymbol{y}).
\end{aligned}$$

The result is given by

$$\frac{B^2\pi a^4}{h^2+a^2} - \frac{B^2 4\pi a^4}{h^2+2a^2} + B^2\pi a^2.$$

In order to calculate the first term of (4.29) we need to approximate the inner part of the integral. It holds

$$\sqrt{m(\boldsymbol{u}) - m(\boldsymbol{u})^2}\sqrt{m(\boldsymbol{v}) - m(\boldsymbol{v})^2} = \sqrt{m(\boldsymbol{u})}\sqrt{m(\boldsymbol{v})}\big(1-m(\boldsymbol{u})\big)\sqrt{1 - \frac{m(\boldsymbol{v}) - m(\boldsymbol{u})}{1-m(\boldsymbol{u})}}.$$

Since a is large comparing to $1/\alpha$ the difference $m(\boldsymbol{v}) - m(\boldsymbol{u})$ in the volume fraction is much smaller then the decrease of $e^{-\alpha\|\boldsymbol{u}-\boldsymbol{v}\|}$ when $\boldsymbol{u}$ goes away from $\boldsymbol{v}$. The square root can be expanded into the Taylor series,

$$\sqrt{1 - \frac{m(\boldsymbol{v}) - m(\boldsymbol{u})}{1-m(\boldsymbol{u})}} = 1 - \frac{1}{2}\frac{m(\boldsymbol{v}) - m(\boldsymbol{u})}{1-m(\boldsymbol{u})} + \mathcal{O}\left(\left(\frac{m(\boldsymbol{v}) - m(\boldsymbol{u})}{1-m(\boldsymbol{u})} \right)^2 \right).$$

Hence, the first integral in (4.29) may be approximated by

$$\iiint K_h(\boldsymbol{x}-\boldsymbol{u}) K_h(\boldsymbol{x}-\boldsymbol{v}) e^{-\alpha\|\boldsymbol{u}-\boldsymbol{v}\|} \sqrt{m(\boldsymbol{u})}\sqrt{m(\boldsymbol{v})} \left(1 - \frac{m(\boldsymbol{u}) + m(\boldsymbol{v})}{2} \right) \mathrm{d}\boldsymbol{u}\mathrm{d}\boldsymbol{v}\mathrm{d}\boldsymbol{x}.$$

Since the argument of integral is symmetric with respect to interchange of $\boldsymbol{u}$ and $\boldsymbol{v}$, we get

$$\frac{B4\pi a^4}{h^2+2a^2}\int K_{\frac{2ha}{\sqrt{h^2+2a^2}}}(\boldsymbol{x})e^{-\alpha\|\boldsymbol{x}\|}\,\mathrm{d}\boldsymbol{x}-\frac{B^2 4\pi a^4}{3h^2+4a^2}\int K_{\frac{4ha}{\sqrt{2}\sqrt{3h^2+4a^2}}}(\boldsymbol{x})e^{-\alpha\|\boldsymbol{x}\|}\,\mathrm{d}\boldsymbol{x}.$$

The integral appearing in both parts can be calculated as

$$\int K_A(\boldsymbol{x})e^{-\alpha\|\boldsymbol{x}\|}\,\mathrm{d}\boldsymbol{x}=1-\frac{A\alpha}{\sqrt{2}}\sqrt{\pi}e^{\frac{A^2\alpha^2}{2}}\operatorname{erfc}\left(\frac{A\alpha}{\sqrt{2}}\right),$$

where erfc is the complementary error function given by

$$\operatorname{erfc}(x)=\frac{2}{\sqrt{\pi}}\int_x^\infty e^{-t^2}\,\mathrm{d}t.$$

Now we use the asymptotic expansion of the complementary error function for large x (see e.g. [95]) given by

$$\operatorname{erfc}(x)=\frac{e^{-x^2}}{x\sqrt{\pi}}\left(1-\frac{1}{2x^2}+\mathcal{O}(x^{-4})\right)\qquad(x\to\infty).$$

As will bee seen later, for the optimal h is $h\alpha\approx C(a\alpha)^{2/3}\gg 1$, so the argument in the erfc is large enough to use the asymptotic expansion. We obtain

$$\frac{B\pi a^2}{h^2\alpha^2}-\frac{B^2\pi a^2}{2h^2\alpha^2}=\frac{B(2-B)\pi a^2}{h^2 2\alpha^2}.$$

Taking all together, we finally have

$$\operatorname{MISE}\big(\hat{m}_h\big)\approx\frac{B(2-B)\pi a^2}{h^2 2\alpha^2}+\frac{B^2\pi a^4}{h^2+a^2}-\frac{B^2 4\pi a^4}{h^2+2a^2}+B^2\pi a^2.$$

After differentiating with respect to h one obtains

$$\frac{-B(2-B)\pi a^2}{h^3\alpha^2}-\frac{2hB^2\pi a^4}{(h^2+a^2)^2}+\frac{2hB^2 4\pi a^4}{(h^2+2a^2)^2}=0.$$

This leads to the polynomial of degree h^8. To find the approximative solution, it is worth noting that

$$\begin{aligned}-\frac{2hB^2\pi a^4}{(h^2+a^2)^2}+\frac{2hB^2 4\pi a^4}{(h^2+2a^2)^2}&=2B^2\pi a^4h^3 4^3\frac{3h^2+4a^2}{(4h^2+4a^2)^2(2h^2+4a^2)^2}\\&=2B^2\pi a^4h^3 4^3\frac{3h^2+4a^2}{\left((3h^2+4a^2)^2-h^4\right)^2}=\frac{2B^2\pi a^4h^3 4^3}{(3h^2+4a^2)^3\left(1-\frac{h^4}{(3h^2+4a^2)^4}\right)^2}.\end{aligned}$$

The largest value of $\frac{h^4}{(3h^2+4a^2)^4}$ is when $h=\frac{2}{\sqrt{3}}a$ which leads to

$$1\geq\left(1-\frac{h^4}{(3h^2+4a^2)^4}\right)^2\geq\left(1-\frac{1}{4^4 3^2 a^4}\right)^2\approx 1$$

as $a\gg 1$. The approximative equation is therefore given by

$$\frac{2B^2\pi a^4h^3 4^3}{(3h^2+4a^2)^3}-\frac{B(2-B)\pi a^2}{h^3\alpha^2}=0.$$

The solution is

$$h_o \approx \frac{2a}{\sqrt{4\left(\frac{2B\alpha^2a^2}{2-B}\right)^{\frac{1}{3}} - 3}}.$$

As $a\alpha \gg 1$, we finally obtain

$$h_o \approx \frac{a^{2/3}}{\alpha^{1/3}} \left(\frac{2-B}{2B}\right)^{1/6}.$$

□

Note that (b) means that the correlation function $\kappa(\boldsymbol{x}, \boldsymbol{y}) = e^{-\alpha\|\boldsymbol{x}-\boldsymbol{y}\|}$ depends only on $\|\boldsymbol{x} - \boldsymbol{y}\|$. This exponential approximation of the correlation function is suggested in [41].

4.3.2 Second order characteristics

As we have seen in the previous part, in order to obtain the optimal estimator of the volume fraction of a random closed set X one must minimize the integrated mean square error $\mathrm{MISE}(\hat{m}_h)$ with respect to the bandwidth h. This involves the knowledge of the volume fraction and the covariance. In most cases, however, those characteristics are unknown and must be estimated by some preliminary estimators. For the volume fraction one usually uses the estimator $\hat{m}_h$ for some subjectively chosen bandwidth h that is relatively large in order to smooth out second order fluctuations.

In the following we focus on the estimation of the unknown covariance function. First, let assume that the volume fraction m is known. In the most general situation one may similarly to the volume fraction use a kernel estimator of the covariance $C(\boldsymbol{x}, \boldsymbol{y})$ given by

$$\hat{C}_h(\boldsymbol{x}, \boldsymbol{y}) = \frac{1}{e_h(\boldsymbol{x})e_h(\boldsymbol{y})} \int_W \int_W \mathbb{1}_X(\boldsymbol{u})\mathbb{1}_X(\boldsymbol{v})K_h(\boldsymbol{x}-\boldsymbol{u})K_h(\boldsymbol{y}-\boldsymbol{v})\,\mathrm{d}\boldsymbol{u}\mathrm{d}\boldsymbol{v}$$

for all $\boldsymbol{x}, \boldsymbol{y} \in W$ with the same notation as in Subsection 4.3.1, and then set

$$\hat{\mathrm{cov}}_h(\boldsymbol{x}, \boldsymbol{y}) = \hat{C}_h(\boldsymbol{x}, \boldsymbol{y}) - m(\boldsymbol{x})m(\boldsymbol{y})$$

as the estimator for the covariance function. For $\hat{C}_h$ and $\hat{\mathrm{cov}}_h$ similar relations as in the previous part can be derived for the (integrated) mean square error now involving fourth moments of X. If m is unknown and has to be estimated one uses $\hat{m}_g$ instead of m in $\hat{\mathrm{cov}}_h$. However, the analytic expressions for the MISE of this estimator are quite complicated.

For the interpretation of second order characteristics of X it is preferable if they depend only on the difference $\boldsymbol{x} - \boldsymbol{y}$. For a random closed set with non-constant volume fraction, both the covariance $C(\boldsymbol{x}, \boldsymbol{y})$ and the covariance function $\mathrm{cov}(\boldsymbol{x}, \boldsymbol{y})$ cannot depend only on the difference $\boldsymbol{x} - \boldsymbol{y}$. For the covariance it is obvious since $C(\boldsymbol{x}, \boldsymbol{x}) = m(\boldsymbol{x})$ and for the covariance function since $\mathrm{cov}(\boldsymbol{x}, \boldsymbol{x}) = m(\boldsymbol{x}) - m(\boldsymbol{x})^2$. So the only characteristics that may depend on the difference $\boldsymbol{x} - \boldsymbol{y}$ is the correlation function $\kappa(\boldsymbol{x}, \boldsymbol{y})$.

In the theory of point processes the possible method how to deal with non-stationarity is to construct a new random measure that is stationary and analyse its second order characteristics, see Subsection 4.1.2. Two possible methods how to do that are intensity reweighting and local rescaling. For random closed sets, both methods are, however, unsuitable. The intensity reweighted measure is $\xi = m^{-1}\nu_X$, where ν_X is the volume measure of X. For its covariance holds

$$\mathrm{cov}_\xi(\boldsymbol{x}, \boldsymbol{y}) = \frac{\mathrm{cov}_X(\boldsymbol{x}, \boldsymbol{y})}{m(\boldsymbol{x})m(\boldsymbol{y})}.$$

In particular $\mathrm{cov}_\xi(\boldsymbol{x}, \boldsymbol{x}) = \frac{1-m(\boldsymbol{x})}{m(\boldsymbol{x})}$ meaning that the covariance of ξ cannot depend just on $\boldsymbol{x}-\boldsymbol{y}$ and hence ξ cannot be second order stationary. For the correlation function, the intensity

reweighting does not bring any benefit since

$$\kappa_\xi(\boldsymbol{x},\boldsymbol{y}) = \frac{\mathrm{cov}_\xi(\boldsymbol{x},\boldsymbol{y})}{\sqrt{\mathrm{cov}_\xi(\boldsymbol{x},\boldsymbol{x})}\sqrt{\mathrm{cov}_\xi(\boldsymbol{y},\boldsymbol{y})}} = \frac{\mathrm{cov}_X(\boldsymbol{x},\boldsymbol{y})}{m(\boldsymbol{x})m(\boldsymbol{y})}\frac{m(\boldsymbol{x})m(\boldsymbol{y})}{\sqrt{\mathrm{cov}_X(\boldsymbol{x},\boldsymbol{x})}\sqrt{\mathrm{cov}_X(\boldsymbol{y},\boldsymbol{y})}} = \kappa_X(\boldsymbol{x},\boldsymbol{y}).$$

Thus the correlation function of ξ depends only on $\boldsymbol{x}-\boldsymbol{y}$ if and only if the correlation function of X depends only on $\boldsymbol{x}-\boldsymbol{y}$. The similar problem is with local rescaling. Here the local change of the metric does not change the volume fraction. So the local rescaling cannot yield the stationary random measure.

In the following we focus on the case $\kappa(\boldsymbol{x},\boldsymbol{y}) = \kappa(\boldsymbol{x}-\boldsymbol{y})$. Since $\kappa(\boldsymbol{x},\boldsymbol{y}) = \kappa(\boldsymbol{x}+\boldsymbol{u},\boldsymbol{y}+\boldsymbol{u})$ for all $\boldsymbol{u},\boldsymbol{x},\boldsymbol{y} \in \mathbb{R}^d$ and by definition $\kappa(\boldsymbol{x},\boldsymbol{y}) = \kappa(\boldsymbol{y},\boldsymbol{x})$ for all $\boldsymbol{x},\boldsymbol{y} \in \mathbb{R}^d$, we have $\kappa(\boldsymbol{r}) = \kappa(\boldsymbol{0}+\boldsymbol{r},\boldsymbol{0}) = \kappa(\boldsymbol{0},\boldsymbol{0}+\boldsymbol{r}) = \kappa(\boldsymbol{0}-\boldsymbol{r},\boldsymbol{0}) = \kappa(-\boldsymbol{r})$ for all $\boldsymbol{r} \in \mathbb{R}^d$. Moreover,

$$\begin{aligned}\kappa(\boldsymbol{r}) &= \frac{1}{\nu_d(B)}\int_B \kappa(\boldsymbol{u}+\boldsymbol{r},\boldsymbol{u})\,\mathrm{d}\boldsymbol{u}\\ &= \mathbb{E}\,\frac{1}{\nu_d(B)}\int_B \frac{\big(\mathbb{1}_X(\boldsymbol{u}+\boldsymbol{r})-m(\boldsymbol{u}+\boldsymbol{r})\big)\big(\mathbb{1}_X(\boldsymbol{u})-m(\boldsymbol{u})\big)}{\sqrt{m(\boldsymbol{u}+\boldsymbol{r})-m^2(\boldsymbol{u}+\boldsymbol{r})}\sqrt{m(\boldsymbol{u})-m^2(\boldsymbol{u})}}\,\mathrm{d}\boldsymbol{u}.\end{aligned}$$

for all $\boldsymbol{r} \in \mathbb{R}^d$ and all Borel $B \subset \mathbb{R}^d$. Let suppose that X is observed in a window W with $\nu_d(W) > 0$ and $0 < m(\boldsymbol{u}) < 1$ for all $\boldsymbol{u} \in W$. A natural unbiased estimator of $\kappa(\boldsymbol{r})$ is thus given by

$$\hat\kappa_v(\boldsymbol{r}) = \frac{1}{\nu_d\big(W\cap(W-\boldsymbol{r})\big)}\int_{W\cap(W-\boldsymbol{r})} \frac{\big(\mathbb{1}_X(\boldsymbol{u}+\boldsymbol{r})-m(\boldsymbol{u}+\boldsymbol{r})\big)\big(\mathbb{1}_X(\boldsymbol{u})-m(\boldsymbol{u})\big)}{\sqrt{m(\boldsymbol{u}+\boldsymbol{r})-m^2(\boldsymbol{u}+\boldsymbol{r})}\sqrt{m(\boldsymbol{u})-m^2(\boldsymbol{u})}}\,\mathrm{d}\boldsymbol{u}$$

for all $\boldsymbol{r} \in \mathbb{R}^d$ such that $\nu_d\big(W\cap(W-\boldsymbol{r})\big) > 0$. If the volume fraction m is unknown we use the kernel estimator $\hat m_h$ and define

$$\hat\kappa_{v;h}(\boldsymbol{r}) = \frac{1}{\nu_d\big(W\cap(W-\boldsymbol{r})\big)}\int_{W\cap(W-\boldsymbol{r})} \frac{\big(\mathbb{1}_X(\boldsymbol{u}+\boldsymbol{r})-\hat m_h(\boldsymbol{u}+\boldsymbol{r})\big)\big(\mathbb{1}_X(\boldsymbol{u})-\hat m_h(\boldsymbol{u})\big)}{\sqrt{\hat m_h(\boldsymbol{u}+\boldsymbol{r})-\hat m_h^2(\boldsymbol{u}+\boldsymbol{r})}\sqrt{\hat m_h(\boldsymbol{u})-\hat m_h^2(\boldsymbol{u})}}\,\mathrm{d}\boldsymbol{u} \tag{4.30}$$

for all $\boldsymbol{r} \in \mathbb{R}^d$ such that $\nu_d\big(W\cap(W-\boldsymbol{r})\big) > 0$, whenever $0 < \hat m_h(\boldsymbol{u}) < 1$ for all $\boldsymbol{u} \in W$. Clearly, both estimators $\hat\kappa_v(\boldsymbol{r})$ and $\hat\kappa_{v;h}(\boldsymbol{r})$ have the property $\hat\kappa_v(-\boldsymbol{r}) = \hat\kappa_v(\boldsymbol{r})$ that is satisfied for κ.

Note that in the non-stationary case it is not reasonable to construct the adapted version of the previous estimator as was (4.14) in the stationary case. This is because the prime role here is played by edge effects. The benefit from balancing the bias will not be achieved since the reduced estimators of the volume fraction suffer more significantly from edge effects. The best way is to use $\hat m_h$ calculated in the window $W_0 \supset W$ that is sufficiently larger then W, which is used to estimate the covariance. This can reduce edge effects of $\hat m_h(\boldsymbol{u})$ for $\boldsymbol{u} \in W$.

The estimator $\hat\kappa_h(\boldsymbol{r})$ can be calculated using the Fourier transform. Similarly as in subsection 4.2.3,

$$\hat\kappa_{v;h}(\boldsymbol{r}) = \frac{(2\pi)^{d/2}\big(\mathscr{F}^{-1}\,|\mathscr{F}g_h|^2\big)(\boldsymbol{r})}{\gamma_W(\boldsymbol{r})}, \tag{4.31}$$

where

$$g_h(\boldsymbol{u}) = \mathbb{1}_W(\boldsymbol{u})\frac{\mathbb{1}_X(\boldsymbol{u})-\hat m_h(\boldsymbol{u})}{\sqrt{\hat m_h(\boldsymbol{u})-\hat m_h^2(\boldsymbol{u})}}.$$

Following the same approach as in Subsection 4.2.3 we can derive the discrete version when X is observed at a grid of points given by intersection of the lattice $L^d = a\mathbb{Z}^d + \boldsymbol{c}$, $a > 0$ with W being included in the closed d-dimensional interval D with each dimension larger then

a. The values of g_h taken at those points define a d-dimensional array $\boldsymbol{G}_h \in \mathbb{R}^{n_1,\ldots,n_d}$, where $n_j \in \mathbb{N}, (b_j - a_j)/a - 1 \leq n_j \leq (b_j - a_j)/a$. The lattice shift $\boldsymbol{c}$ can be, without loss of generality, chosen such that the elements of $\boldsymbol{G}_h$ are given by

$$(\boldsymbol{G}_h)_{i_1,\ldots,i_d} = g_h\big(a(i_1,\ldots,i_d) + \boldsymbol{c}\big)$$

for all $i_j = 1,\ldots,n_j$ and all $j = 1,\ldots,d$. The total number of elements of $\boldsymbol{G}_h$ is given by $n = \Pi_{j=1}^d n_j$. Let further $\boldsymbol{I}$ denote the matrix obtained from values of the indicator of W at grid points, i.e.

$$(\boldsymbol{I})_{i_1,\ldots,i_d} = \mathbb{1}_W\big(a(i_1,\ldots,i_d) + \boldsymbol{c}\big)$$

for all $i_j = 1,\ldots,n_j$ and all $j = 1,\ldots,d$. The discrete version of the estimator $\hat{\kappa}_{v;h}(\boldsymbol{r})$ is then given by

$$\hat{\kappa}_{p;h}(a\boldsymbol{k}) = \frac{\sum_{\boldsymbol{i}\in A(\boldsymbol{k})}(\boldsymbol{G}_h)_{\boldsymbol{i}}(\boldsymbol{G}_h)_{\boldsymbol{i}+\boldsymbol{k}}}{\sum_{\boldsymbol{i}\in A(\boldsymbol{k})}(\boldsymbol{I})_{\boldsymbol{i}}(\boldsymbol{I})_{\boldsymbol{i}+\boldsymbol{k}}}, \tag{4.32}$$

where $\boldsymbol{k} = (k_1,\ldots,k_d), \boldsymbol{i} = (i_1,\ldots,i_d)$ and the index set

$$A(\boldsymbol{k}) = \{(i_1,\ldots,i_d) \in \mathbb{Z}^d | 1 \leq i_j \leq n_j, 1 \leq i_j + k_j \leq n_j \text{ for all } j = 1,\ldots,d\},$$

for all $\boldsymbol{k}$ such that the denominator is positive, thus forming a subset of $\{-n_1+1,\ldots,n_1-1\}\times\ldots\times\{-n_d+1,\ldots,n_d-1\}$. The estimator has the desired property $\hat{\kappa}_{p;h}(a\boldsymbol{k}) = \hat{\kappa}_{p;h}(-a\boldsymbol{k})$ which follows from $A(-\boldsymbol{k}) = A(\boldsymbol{k}) + \boldsymbol{k}$. If $\boldsymbol{I}$ is a constant matrix, the value of the denominator can be calculated explicitly as $\Pi_1^d(n_j - |k_j|)$.

To establish a version of (4.31) using discrete Fourier transform one must again take care of the periodicity assumption in the discrete Fourier transform and expand the matrices $\boldsymbol{G}_h$ and $\boldsymbol{I}$ by padding with zeros:

$$(\tilde{\boldsymbol{G}}_h)_{\boldsymbol{i}} = \begin{cases}(\boldsymbol{G}_h)_{\boldsymbol{i}}, & \text{if } 1 \leq \boldsymbol{i} \leq \boldsymbol{n},\\ 0, & \text{otherwise,}\end{cases} \quad \text{and} \quad (\tilde{\boldsymbol{I}})_{\boldsymbol{i}} = \begin{cases}(\boldsymbol{I})_{\boldsymbol{i}}, & \text{if } 1 \leq \boldsymbol{i} \leq \boldsymbol{n},\\ 0, & \text{otherwise,}\end{cases}$$

This increases the number of sample points to $2^d n$. Thus similarly as in Subsection 4.2.3 we obtain

$$\hat{\kappa}_{p;h}(a\boldsymbol{k}) = \frac{\mathrm{iDFT}\big(|\mathrm{DFT}(\tilde{\boldsymbol{G}}_h)|^{\star 2}\big)(a\boldsymbol{k})}{\mathrm{iDFT}\big(|\mathrm{DFT}(\tilde{\boldsymbol{I}})|^{\star 2}\big)(a\boldsymbol{k})}, \tag{4.33}$$

where $^{\star 2}$ means the element-wise square of a matrix. If $\kappa(\boldsymbol{r})$ depends only on the norm $\|\boldsymbol{r}\|$ one may use the isotropic adapted kernel estimator $\hat{\kappa}_{p;h}^{AI}(r)$ defined by (4.24).

4.4 Simulation studies of second order estimators

In this section we use numerical simulations to study the performance of discrete estimators of the correlation function from Subsection 4.2.3. We focus only on the stationary case in $\mathbb{R}^2$. The non-stationary case is studied in Subsection 5.4.4 in connection to analysis of a built-up structure. Moreover, we especially focus on the performance of estimation for long-range dependent random closed sets introduced Section 2.8.

Let X be a random closed set in $\mathbb{R}^2$ sampled on a grid of points $\{a\boldsymbol{i} + \boldsymbol{c} | \boldsymbol{i} \in \mathbb{Z}^2,\ (1,1) \leq \boldsymbol{i} \leq \boldsymbol{n}\}$ for some $a > 0$, $c \in \mathbb{R}^2$ and $\boldsymbol{n} = (n_1, n_2) \in \mathbb{N}^2$ such that the values of the indicator $\mathbb{1}_X$ on the grid define a matrix $\boldsymbol{M} \in \{0,1\}^{n_1,n_2}$,

$$(\boldsymbol{M})_{\boldsymbol{i}} = \mathbb{1}_X\big(a\boldsymbol{i} + \boldsymbol{c}\big), \quad \boldsymbol{i} \in \{1,\ldots,n_1\}\times\{1,\ldots,n_2\},$$

Let $\boldsymbol{I}$ be a matrix of the same size as $\boldsymbol{M}$ with all elements equal to one.

The natural unbiased estimator $\hat{p}_p$ of the volume fraction is defined by (4.18). Concerning second order properties, our first aim is to analyse the properties of the discrete estimator

$\hat{\kappa}_p(\boldsymbol{r})$ defined by (4.17) and its intrinsically balanced modification $\hat{\kappa}_p^\bullet(\boldsymbol{r})$ defined by (4.20). All of those estimators can be calculated with the help of the fast Fourier transform using relations (4.19), (4.22) and (4.23).

Except direct estimators we also study isotropic adapted estimators $\hat{\kappa}_{p;h}^{AI}(r)$ defined by (4.24) and $\hat{\kappa}_p^{\bullet AI}(r)$ defined analogously, with K being the Gaussian kernel and $h(r)$ being the variable bandwidth that is throughout the analysis taken to be

$$h(r) = 30 \cdot (1 - e^{-0.001r}) + 0.5.$$

Such a function increases from 0.5 to 30.5 with fastest increase around zero and represents a reasonable choice balancing the need for a small bandwidth when r is small and a sufficiently large bandwidth when r is medium or large.

In order to quantify the performance of analysed estimators the following basic characteristics are used. For the analysed estimator $\hat{f}$ of a model characteristic f with known expectation $\mathbb{E}\,\hat{f}$ its variance $\operatorname{var}\hat{f}$ is estimated by the sample variance $s_N^2(\hat{f})$ and the standard deviation $\sqrt{\operatorname{var}\hat{f}}$ by the sample standard deviation $s_N(\hat{f})$. Estimators $s_N^2(\hat{f})$ and $s_N(\hat{f})$ are defined by

$$s_N^2(\hat{f}) = \frac{1}{N}\sum_{i=1}^{N}(\hat{f}_i - \mathbb{E}\,\hat{f})^2 \quad \text{and} \quad s_N(\hat{f}) = \sqrt{s_N^2(\hat{f})}, \tag{4.34}$$

respectively, where $\hat{f}_i$ is the value of $\hat{f}$ on i-th realisation of the model. If the true value of the expectation $\mathbb{E}\,\hat{f}$ of the analysed estimator $\hat{f}$ is not known, then the sample mean $\overline{\hat{f}}$ as the estimator of $\mathbb{E}\,\hat{f}$ is used. It is given by

$$\overline{\hat{f}} = \frac{1}{N}\sum_{i=1}^{N}\hat{f}_i. \tag{4.35}$$

In that case the sample variance $s_N^2(\hat{f})$ and the sample standard deviation $s_N(\hat{f})$ are defined by

$$s_N^2(\hat{f}) = \frac{1}{N-1}\sum_{i=1}^{N}\left(\hat{f}_i - \overline{\hat{f}}\right)^2 \quad \text{and} \quad s_N(\hat{f}) = \sqrt{s_N^2(\hat{f})}, \tag{4.36}$$

respectively. Furthermore, the mean squared error $\mathbb{E}(\hat{f} - f)^2$ from the true value f is estimated by $d_N^2(\hat{f}; f)$ and the deviation from the true value given by $\sqrt{\mathbb{E}(\hat{f} - f)^2}$ is estimated by $d_N(\hat{f}; f)$, where $d_N^2(\hat{f}; f)$ and $d_N(\hat{f}; f)$ are defined by

$$d_N^2(\hat{f}; f) = \frac{1}{N}\sum_{i=1}^{N}(\hat{f}_i - f)^2 \quad \text{and} \quad d_N(\hat{f}; f) = \sqrt{d_N^2(\hat{f}; f)}, \tag{4.37}$$

respectively. Sometimes we also estimate the two sided 90% confidence interval (confidence bounds) $I_{0.9}(\hat{f})$ of $\hat{f}$ defined via the property $\mathbb{P}\left(\hat{f} \in I_{0.9}(\hat{f})\right) = 0.9$ together with

$$\mathbb{P}\left(\hat{f} \text{ on the left of } I_{0.9}(\hat{f})\right) = \mathbb{P}\left(\hat{f} \text{ on the right of } I_{0.9}(\hat{f})\right) = 0.05.$$

If f is a function (e.g. the correlation function), we define the confidence bounds at every point of its domain independently. Denoting $I_{0.9}(\hat{f}) = [a(\hat{f}), b(\hat{f})]$, the bound $a(\hat{f})$ is estimated by $\hat{a}(\hat{f}) = f_{i_{0.05}}$, where $i_{0.05} \in 1, 2, \ldots, N$ is the index of a realization such that there are exactly $\lfloor 0.05 \cdot N \rfloor$ realizations with values of $\hat{f}$ smaller than $\hat{f}_{i_{0.05}}$. Note that $\lfloor 0.05 \cdot N \rfloor$ means the integral part of $0.05 \cdot N$. The estimator of $b(\hat{f})$ is defined analogously.

4.4.1 The Boolean model

The analysis is based on two different models, the Boolean model and the level excursion set. The Boolean model was introduced in Section 2.7 as a germ-grain model with a Poisson ground point process. It represents a short-range dependent structure. We analyse realizations of a Boolean model X with ground stationary Poisson point process X^0 having intensity λ and with typical grain X_0 being a disc of fixed radius R centred at the origin, i.e $X_0(\omega) = B_R(\mathbf{0})$ for all $\omega \in \Omega$. From (2.37) follows that the volume fraction p of X is

$$p = 1 - e^{-\lambda\, \mathbb{E}\, \nu_d(\check{X}_0)} = 1 - e^{-\lambda\pi R^2},$$

since $\check{X}_0 = B_R(\mathbf{0})$. In order to determine the correlation function we need to express the set covariance $\gamma_{X_0}(\boldsymbol{r}) = \nu_d\big(X_0 \cap (X_0 - \boldsymbol{r})\big)$ of X_0. It is a simple matter to check that

$$\gamma_{X_0}(\boldsymbol{r}) = 2R^2 \arccos\left(\frac{r}{2R}\right) - rR\sqrt{1 - \left(\frac{r}{2R}\right)^2}$$

for $r \equiv \|\boldsymbol{r}\| \in [0, 2R)$ and $\gamma_{X_0}(\boldsymbol{r}) = 0$ otherwise. Hence by (2.38),

$$\kappa(\boldsymbol{r}) = \kappa(r) = \begin{cases} \frac{1-p}{p}(e^{\lambda\gamma_{X_0}(\boldsymbol{r})} - 1) & \text{for } r \in [0, 2R), \\ 0 & \text{for } r \geq 2R. \end{cases} \tag{4.38}$$

The fact that $\kappa(\boldsymbol{r})$ is a function of $\|\boldsymbol{r}\|$ also follows from the isotropy of X. To show the isotropy of X one can use Theorem 2.2.2, relation (2.36) and the fact that $\Lambda(\theta C + \check{X}_0) = \lambda\nu(\theta C + \theta B_R(\mathbf{0})) = \lambda\nu(\theta(C + B_R(\mathbf{0})) = \Lambda(C + \check{X}_0)$ for every rotation $\theta \in SO_2$.

The parameters of the model used in simulations are $R = 10$ and $p = 0.3$ which corresponds to

$$\lambda = \frac{1}{\pi R^2} \log\left((1-p)^{-1}\right) \doteq 11.35 \cdot 10^{-4}.$$

The random closed set X is analysed on the grid sampled with $a = 1$ and $\boldsymbol{c} = -1(1, 1)$ inside two square windows $W_{3000} = [0, 3000] \times [0, 3000]$ and $W_{5000} = [0, 5000] \times [0, 5000]$ in order to analyse the influence of window size. To construct a realization of X in W_{3000} we generate points of the Poisson process X^0 in a larger window $[-10, 3010] \times [-10, 3010]$ so that the final Boolean model in W_{3000} is without simulation based edge effects. Analogously we obtain realizations in W_{5000}. An example of a realization is shown in Figure 4.1. The results of estimations of the correlation function for the same realization are shown in Figure 4.2. We can see that they are close to the true value $\kappa(r)$.

The study was performed by generating $N = 1000$ realizations. The results for the volume fraction are summarized in Table 4.1. We can see the influence of a larger window in the decrease of the standard deviation and narrowing the confidence interval. The decrease in the variance should be consistent with

$$s^2_N(\hat{p}_{p;W_{5000}}) = \frac{\nu(W_{3000})}{\nu(W_{5000})} s^2_N(\hat{p}_{p;W_{3000}}) = \frac{9}{25} s^2_N(\hat{p}_{p;W_{3000}}),$$

which follows from relation (4.5) that approximately holds even in the discrete case with short range dependence and large W. Since $9/25 \cdot 6.8 \cdot 10^{-6} = 2.46 \cdot 10^{-6}$, we see that it is approximately true.

Deviations of correlation function estimators from true values are shown in Figure 4.3. We can see the better performance of intrinsically balanced versions of estimators compared to ordinary ones, especially at larger distances. Also the isotropic adapted estimators are generally better, but only for large r. This improvement is due to more information used by the isotropic estimator which is a result of averaging. For small r the influence of larger bias, yielding larger deviations then usual estimators, is clearly visible. The bias is a result of the

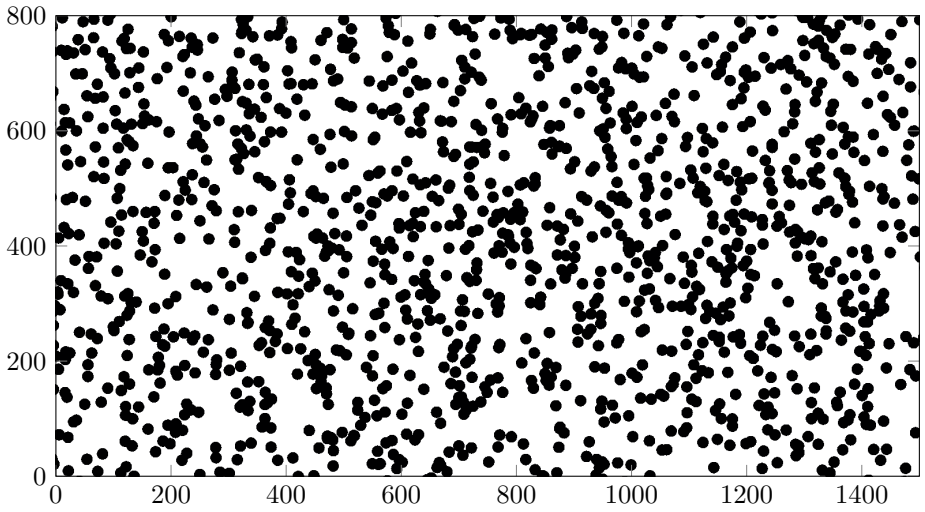


Figure 4.1: A part of one concrete realization of the Boolean model X in window W_{3000}.

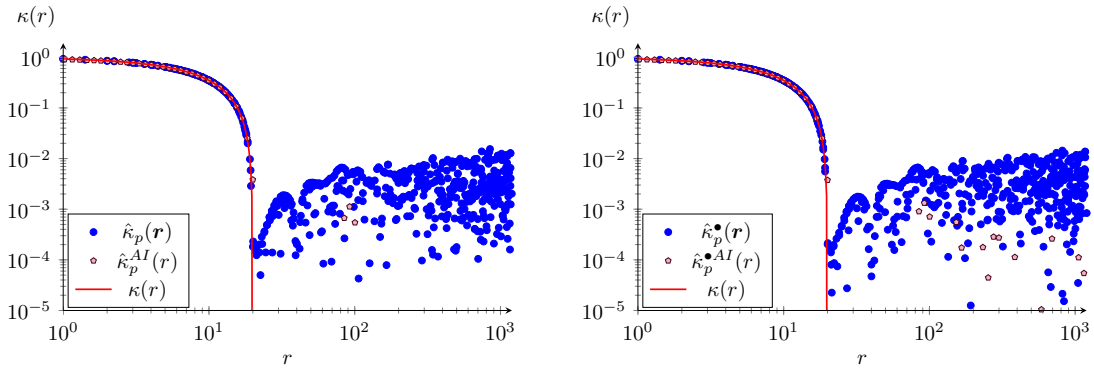


Figure 4.2: A result of correlation function estimators for one realization in window W_{3000}. On the left: ordinary estimators, on the right: intrinsically balanced estimators. The true value $\kappa(r)$ is superimposed in both cases.

Table 4.1: Basic characteristics of $\hat{p}_p$.

Window	$I_{0.9}(\hat{p}_p)$	$s_N^2(\hat{p}_p)$	$s_N(\hat{p}_p)$
W_{3000}	$[0.2958, 0.3043]$	$6.8 \cdot 10^{-6}$	0.0026
W_{5000}	$[0.2975, 0.3025]$	$2.3 \cdot 10^{-6}$	0.0015

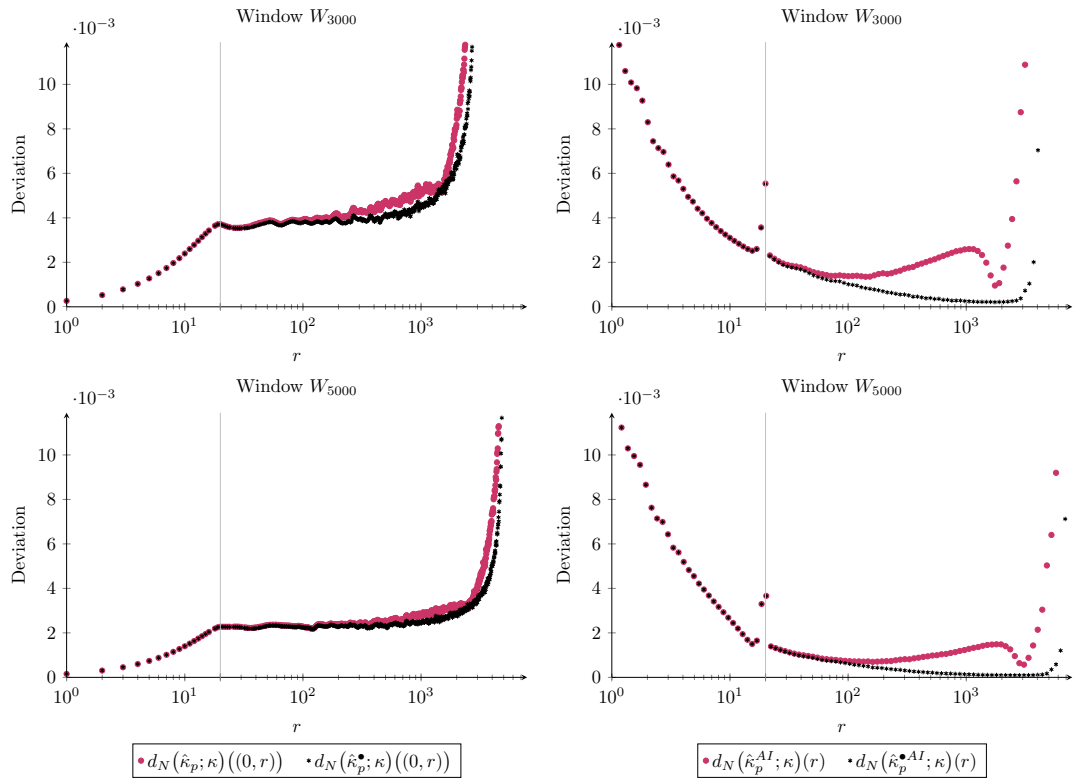


Figure 4.3: Estimated deviations of $\hat{\kappa}_p(\boldsymbol{r})$, $\hat{\kappa}_p^{\bullet}(\boldsymbol{r})$ from the true value $\kappa(\boldsymbol{r}) \equiv \kappa(r)$ along $\boldsymbol{r} = (0, r)$ on the left and of $\hat{\kappa}_p^{AI}(r)$, $\hat{\kappa}_p^{\bullet AI}(r)$ on the right.

convolution and of the particular choice of a bandwidth function $h(r)$ as was mentioned in the end of Subsection 4.2.3. The estimated bias of adapted isotropic estimators is shown in Figure 4.4. As can be expected, the bias of estimators obtained for W_{3000} and W_{5000} is almost the same. One can see the clear difference between regions $r < 2R$ and $r > 2R$. The first is the region with strongly decreasing κ and thus the bias produced by the convolution is significant. On the other side, in the second region κ is constantly equal to 0 and hence the bias should be zero.

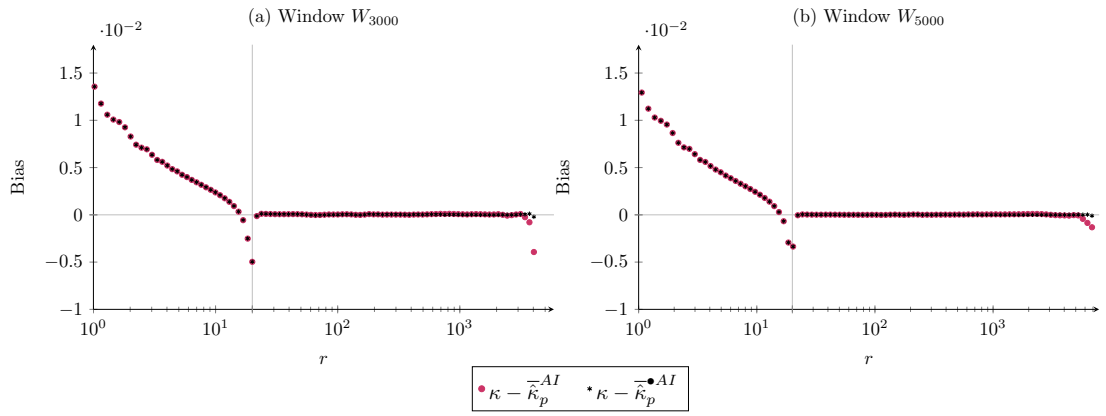


Figure 4.4: Estimated bias of $\hat{\kappa}_p^{AI}(r)$ and of $\hat{\kappa}_p^{\bullet AI}(r)$.

4.4.2 The level excursion set

The second model is given by the level excursion set introduced in Section 2.4. The analysis is based on realizations from a 0-level excursion set $X_0(Z)$ of a Gaussian random field Z determined by the mean $\mu_Z = 0$ and the covariance function

$$\mathrm{cov}_Z(\boldsymbol{x}, \boldsymbol{y}) = \kappa_C(\|\boldsymbol{x} - \boldsymbol{y}\|),$$

where $\kappa_C(r)$ is the **Cauchy covariance function** defined by

$$\kappa_C(r) = \left(1 + \left(\frac{r}{\theta}\right)^{\alpha}\right)^{-\frac{\beta}{\alpha}} \quad \text{for} \quad r \geq 0. \tag{4.39}$$

The Cauchy covariance function, or one may say Cauchy correlation function since $\kappa_C(0) = 1$, depends on shape parameters $\alpha \in (0, 2]$, determining behaviour for small arguments, $\beta > 0$, determining behaviour for large arguments, and the scale parameter $\theta > 0$. The class formed by all Cauchy covariance functions, called the Cauchy class, was introduced in [96]. See also [97], where it is argued that it corresponds to a valid covariance function, i.e. positive semi-definite, for all combinations of the parameters.

It is important to note that Z determined by μ_Z and cov_Z has almost surely continuous sample functions, since it satisfies Corollary 2.4.1. To check this it is enough to see that $\big(1 - C(r)\big)\,|\log(r)|^{1+\varepsilon}$ is bounded as $r \to 0_+$. Using the Taylor expansion one can by standard methods of calculus show that the limit equals 0. Hence Z has almost surely continuous sample functions. Therefore $X_0(Z) = \{\boldsymbol{x} | Z(\boldsymbol{x}) \geq 0\}$ is by Theorem 2.4.2 a stationary and isotropic random closed set.

From (2.14) follows that the volume fraction p of $X_0(Z)$ is $p = 0.5$. The covariance function $\mathrm{cov}(r)$ of $X_0(Z)$ is given by (2.15) and the correlation function by (2.16) as

$$\kappa_L(r) = \frac{2}{\pi} \arcsin\left(\left(1 + \left(\frac{r}{\theta}\right)^{\alpha}\right)^{-\frac{\beta}{\alpha}}\right). \tag{4.40}$$

Since $\arcsin(x)/x \to 1$ as $x \to 0$, it follows that

$$\kappa_L(r) \sim \frac{2}{\pi}\left(\frac{r}{\theta}\right)^{-\beta} \qquad (r \to \infty).$$

Clearly, $\mathrm{cov}(r)$ has the same asymptotic with different constant factor. Let assume that $\beta \in (0, 2)$. Then the random closed set $X_0(Z)$ satisfies Definition 2.8.2 and hence exhibits

Table 4.2: Settings used in simulations. The volume fraction is $p = 0.5$ and the covariance is given by (4.40) with parameters α, β, θ. The simulation is performed in window W and the sampling is based on a regular grid scheme with a and $\boldsymbol{c} = -a(1,1)$.

Setting	W	a	α	β	θ
1–a	W_{3000}	1	1.8	0.8	3
1–b	W_{3000}	1	2	1.05	4
2–a	W_{5000}	1	1.8	0.8	3
2–b	W_{5000}	1	2	1.05	4
3–a	W_{10000}	2	1.8	0.8	3
3–b	W_{10000}	2	2	1.05	4

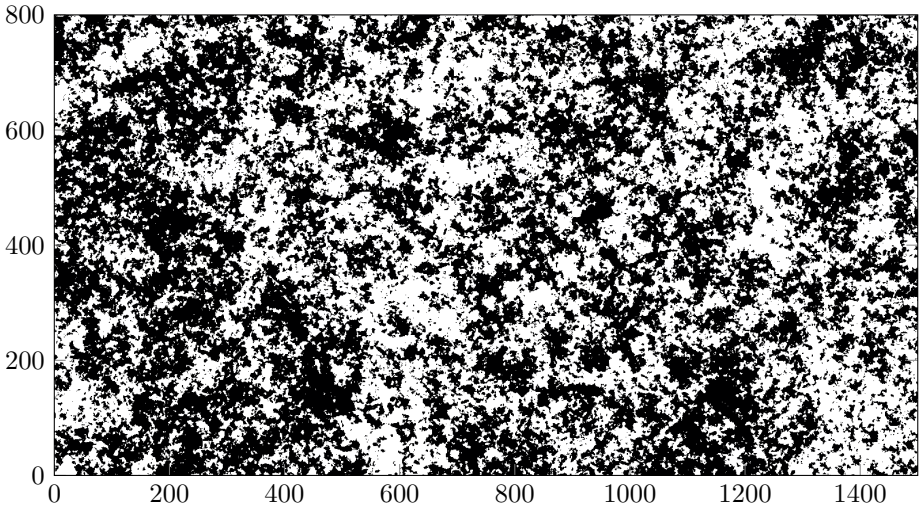


Figure 4.5: A part of one realization of the 0-level excursion set $X_0(Z)$ in the setting 2–a.

isotropic long-range dependence, which means, as a result of Corollary 2.8.2, that it is long-range dependent. Therefore we may assume poor statistical properties compared to those observed for the Boolean model in the previous subsection.

The values of Gaussian random field Z and the corresponding 0-level excursion set $X_0(Z)$ on a regular grid may be simulated using the method of circular embedding with the help of fast Fourier transform, see [98]. Note that the method can be used only if the circular embedding results in a non-negative definite covariance matrix which is for the Cauchy class not true in general. However, for all settings we have analysed the method works well. The concrete `MATLAB` codes used in simulations were taken from [99].

The numerical analysis was performed for 6 settings that are combined from 3 sampling schemes and 2 correlation schemes that follow. The sampling schemes are regular grids with $a = 1$ inside $W_{3000} = [0, 3000] \times [0, 3000]$, $a = 1$ inside $W_{5000} = [0, 5000] \times [0, 5000]$, and $a = 2$ inside $W_{10000} = [0, 10000] \times [0, 10000]$. In all schemes we take $\boldsymbol{c} = -a(1,1)$. The correlation schemes are determined by $\alpha = 1.8$, $\beta = 0.8$, $\theta = 3$ and $\alpha = 2$, $\beta = 1.05$, $\theta = 4$. Thus the first case has more slowly decaying correlations than the second. The summary of parameters of all settings can be found in Table 4.2. An example of one realization is shown in Figure 4.5.

Table 4.3: Basic characteristics of $\hat{p}_p$.

Setting	$I_{0.9}(\hat{p}_p)$	$s_N^2(\hat{p}_p)$	$s_N(\hat{p}_p)$
1–a	$[0.437, 0.563]$	$1.43 \cdot 10^{-3}$	0.038
2–a	$[0.453, 0.549]$	$8.55 \cdot 10^{-4}$	0.029
3–a	$[0.463, 0.539]$	$5.50 \cdot 10^{-4}$	0.023
1–b	$[0.462, 0.536]$	$4.87 \cdot 10^{-4}$	0.022
2–b	$[0.471, 0.528]$	$3.00 \cdot 10^{-4}$	0.017
3–b	$[0.482, 0.521]$	$1.45 \cdot 10^{-4}$	0.012

During the analysis we have generated $N = 500$ realizations for each setting. Again, as in the previous subsection, we estimate variances, standard deviations, and deviations from true values of studied estimators according to (4.34), (4.37) and sometimes also the sample mean according to (4.35). When it is instructive we also determine the 90% confidence interval of the estimator.

The results for the volume fraction are summarized in Table 4.3. On can clearly see the strong influence of the correlation strength. The (b) settings that correspond to weaker correlations have much smaller variances then (a) settings corresponding to stronger correlations. By comparing to Table 4.1 it is clearly visible that variances in all settings are much larger than for the short-range Boolean model. Similarly, there is again the influence of the sample window size. The larger window corresponds to smaller standard deviation, variance, and to more narrow confidence interval. However, here the variance is not linearly dependent on the window volume as it was for the Boolean model. This holds because random closed set $X_0(Z)$ is now long-range dependent. Relation (4.5) is no longer true since the integral range A is infinity. Instead, relation (4.6) should approximately hold, yielding

$$s_N^2(\hat{p}_{p;W_{5000}}) \approx \left(\frac{\nu(W_{3000})}{\nu(W_{5000})}\right)^{\frac{\beta}{2}} s_N^2(\hat{p}_{p;W_{3000}}) = \left(\frac{3}{5}\right)^{\beta} s_N^2(\hat{p}_{p;W_{3000}})$$

and analogously for other combinations of W_{3000}, W_{5000} and W_{10000}. Since

$$1.43 \cdot 10^{-3} \cdot \left(\frac{3}{5}\right)^{0.8} = 9.50 \cdot 10^{-4}, \qquad 4.87 \cdot 10^{-4} \cdot \left(\frac{3}{5}\right)^{1.05} = 2.85 \cdot 10^{-4},$$
$$1.43 \cdot 10^{-3} \cdot \left(\frac{3}{10}\right)^{0.8} = 5.46 \cdot 10^{-4}, \qquad 4.87 \cdot 10^{-4} \cdot \left(\frac{3}{10}\right)^{1.05} = 1.38 \cdot 10^{-4},$$
$$8.55 \cdot 10^{-4} \cdot \left(\frac{5}{10}\right)^{0.8} = 4.91 \cdot 10^{-4}, \qquad 3.00 \cdot 10^{-3} \cdot \left(\frac{5}{10}\right)^{1.05} = 1.45 \cdot 10^{-4},$$

we see that it works with less then 15% error in all cases. The better result is again achieved for weaker correlations with $\beta = 1.05$ in (b) cases. This is because relation (4.6) holds asymptotically and hence systems with faster decrease of correlations achieve the asymptotic region faster.

Deviations from the true value κ of correlation function estimators $\hat{\kappa}_p$, $\hat{\kappa}_p^{\bullet}$ are shown in Figure 4.6 and of isotropically adapted estimators $\hat{\kappa}_p^{AI}$, $\hat{\kappa}_p^{\bullet AI}$ in Figure 4.7. Again one can see a clear influence of the window size and the strength of correlations. Thus the deviation decreases with increasing window and weaker correlations. In all cases the intrinsically balanced estimators are better then their ordinary counterparts, especially at larger distances. Actually, the effect of intrinsically balancing is much stronger here than it was for the Boolean

model, see Figure 4.3 for comparison. This is in contradiction to dismissal of such performance improvement of covariance function estimators pointed out by Picka in [90, p. 694], where the intrinsically balanced estimators were introduced. He argued that: "the intrinsically balanced modification has no obvious beneficial or detrimental effect when the dependence structure is present".

The isotropic adapted estimators are generally better only for large values of r. The reason is the same as in the previous subsection. For small r there is an influence of larger bias caused by the convolution that results in large deviations. The biases of adapted isotropic estimators are shown in Figure 4.8. The domain of a bias can be divided into two parts. For r smaller then approximately 5 the main part of the bias is due to convolution and it is not sensitive to the window size. The messy beginnings in settings 3–a and 3–b are because of the large grid step $a = 2$ in sampling which disables the precise estimation of the correlation function at very small distances. On the other hand, for larger values of r there is a decrease in the bias when the window is enlarged. This is because of the presence of systematic bias of underlying estimators $\hat{\kappa}_p(\boldsymbol{r})$ and $\hat{\kappa}_p^\bullet(\boldsymbol{r})$ in that range. To visualise the situation we show the estimated mean values of $\hat{\kappa}_p^\bullet(\boldsymbol{r})$ together with 90% confidence bounds in Figure 4.9. One can see the clear underestimation of correlation function values at the right half of every case. Again the situation is better for (b) cases and for larger windows.

Finally, we estimate the values of α, β and θ that determine the correlation function. This is done by fitting of estimated values of the correlation function with parametric function $\kappa_L(r, \boldsymbol{\theta})$ defined by (4.40), where $\boldsymbol{\theta} = (\alpha, \beta, \theta)$ is the vector of unknown parameters. The fit is obtained by minimizing the weighted least squares. Hence as an estimate we take the vector of parameters $\hat{\boldsymbol{\theta}} = (\hat{\alpha}, \hat{\beta}, \hat{\theta})$, $0 < \hat{\beta}, \hat{\theta} < \infty$, $0 < \hat{\alpha} \leq 2$, that minimizes

$$Q(\boldsymbol{\theta}) = \sum_{\boldsymbol{k}=-\boldsymbol{n}+1}^{\boldsymbol{n}-1} w_{\boldsymbol{k}} \big(\hat{\kappa}_p(a\boldsymbol{k}) - \kappa_C(\|a\boldsymbol{k}\|\,; \boldsymbol{\theta})\big)^2,$$

where $\boldsymbol{n} = (n_1, n_2)$ and the weights $w_{\boldsymbol{k}}$ are chosen accordingly to [49, §2.6.2] as $w_{\boldsymbol{k}} = \frac{N_{\boldsymbol{k}}}{a\|\boldsymbol{k}\|+1}$. The coefficient $N_{\boldsymbol{k}} = (n_1 - |k_1|) \cdot (n_2 - |k_2|)$ gives the number of points that were used for the calculation of $\hat{\kappa}_p(a\boldsymbol{k})$. Analogously we define the estimator $\hat{\boldsymbol{\theta}}^\bullet$ based on the intrinsically balanced estimator $\hat{\kappa}_p^\bullet$. Note that we do not use isotropic adapted estimators since they bring additional bias into the fitting mechanism leading to poor performance of resulting estimators of the parameters. In Table 4.4 we summarized estimated statistical properties of $\hat{\boldsymbol{\theta}}$ and $\hat{\boldsymbol{\theta}}^\bullet$ in all analysed settings.

Clearly, as before, the estimation is better for larger windows and for weaker correlations. It is also much better for intrinsically balanced estimators. For the individual parameters we may say that α is underestimated whereas β and θ are overestimated. The parameter of a special importance is β since it corresponds to the power law coefficient of the long-range correlation function (or equally covariance function) decay. We can see that the bias in the estimation of β can be up to 24% in the setting 1–a and the best achieved value is in 3–b with bias slightly less then 4%. Even in this case 5% of realizations lead to values larger then 1.15 which is more than 10% overestimation. This is also demonstrated in Figure 4.9 with estimated mean values of estimator $\hat{\kappa}_p^\bullet$ and superimposed $\kappa_L\big(r, \overline{\boldsymbol{\theta}}^\bullet\big)$, where $\overline{\boldsymbol{\theta}}^\bullet = \big(\overline{\hat{\alpha}}^\bullet, \overline{\hat{\beta}}^\bullet, \overline{\hat{\theta}}^\bullet\big)$ is the vector of determined mean parameters. We may see that the estimation is mainly influenced by the negative bias of $\hat{\beta}^\bullet$. This means that the green dashed line has always a faster decay then $\kappa(r)$ plotted as a red line.

It is also interesting to see how the situation looks like for one concrete realization. The example of a result of intrinsically balanced estimators together with the fit of two realizations is shown in Figure 4.10. We see that for the first realization the parameter fit $\hat{\boldsymbol{\theta}}^\bullet$ is more precise that in the mean situation. On the other hand the result of the estimation on second realization is worse then in the mean. However, in both cases the general qualitative features of estimations are similar and also similar to the mean situation depicted in Figure 4.9. They are

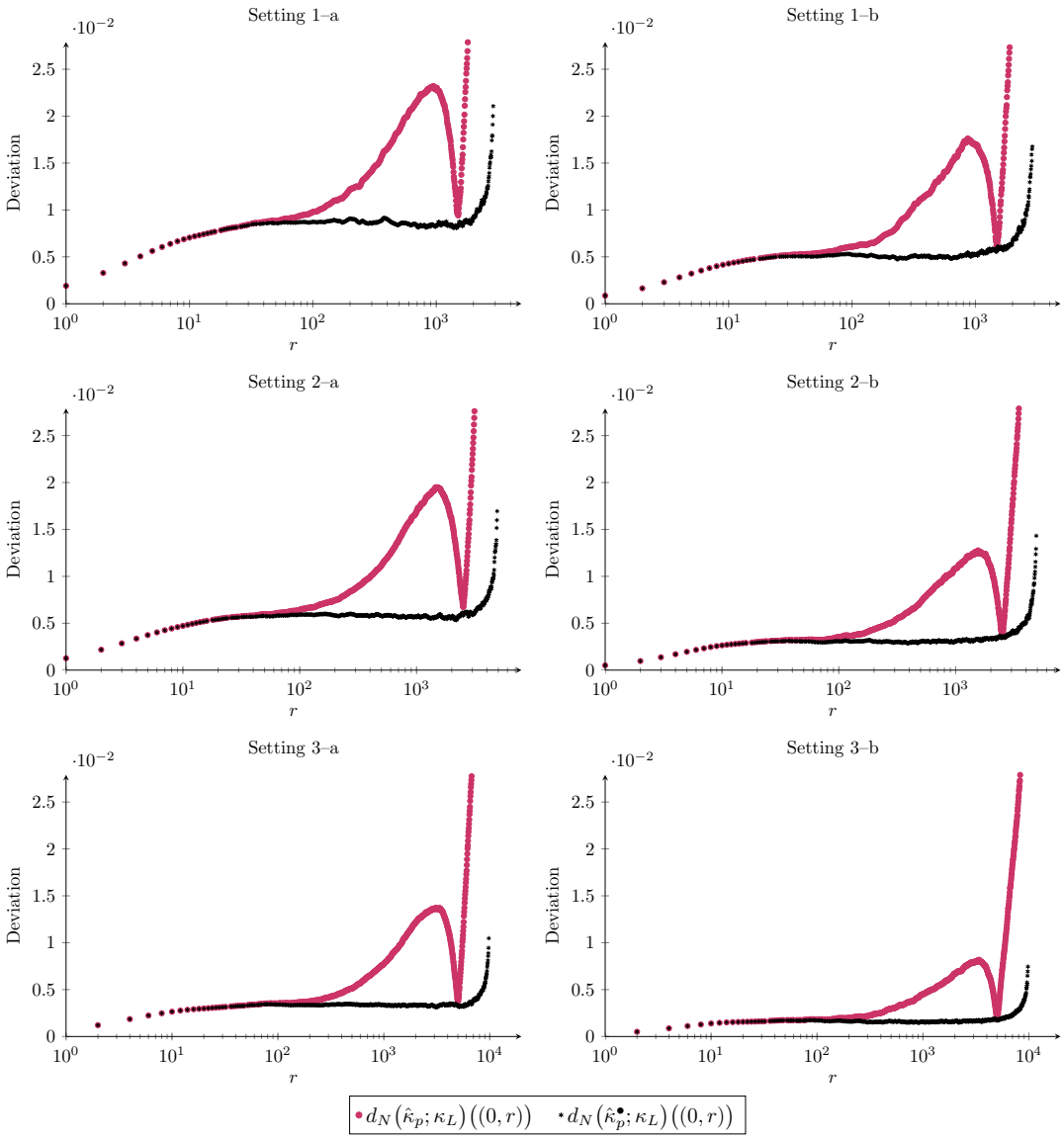


Figure 4.6: Estimated deviations of $\hat{\kappa}_p(\boldsymbol{r})$ and $\hat{\kappa}_p^{\bullet}(\boldsymbol{r})$ from the true value $\kappa_L(\boldsymbol{r})$ along $\boldsymbol{r} = (0, r)$.

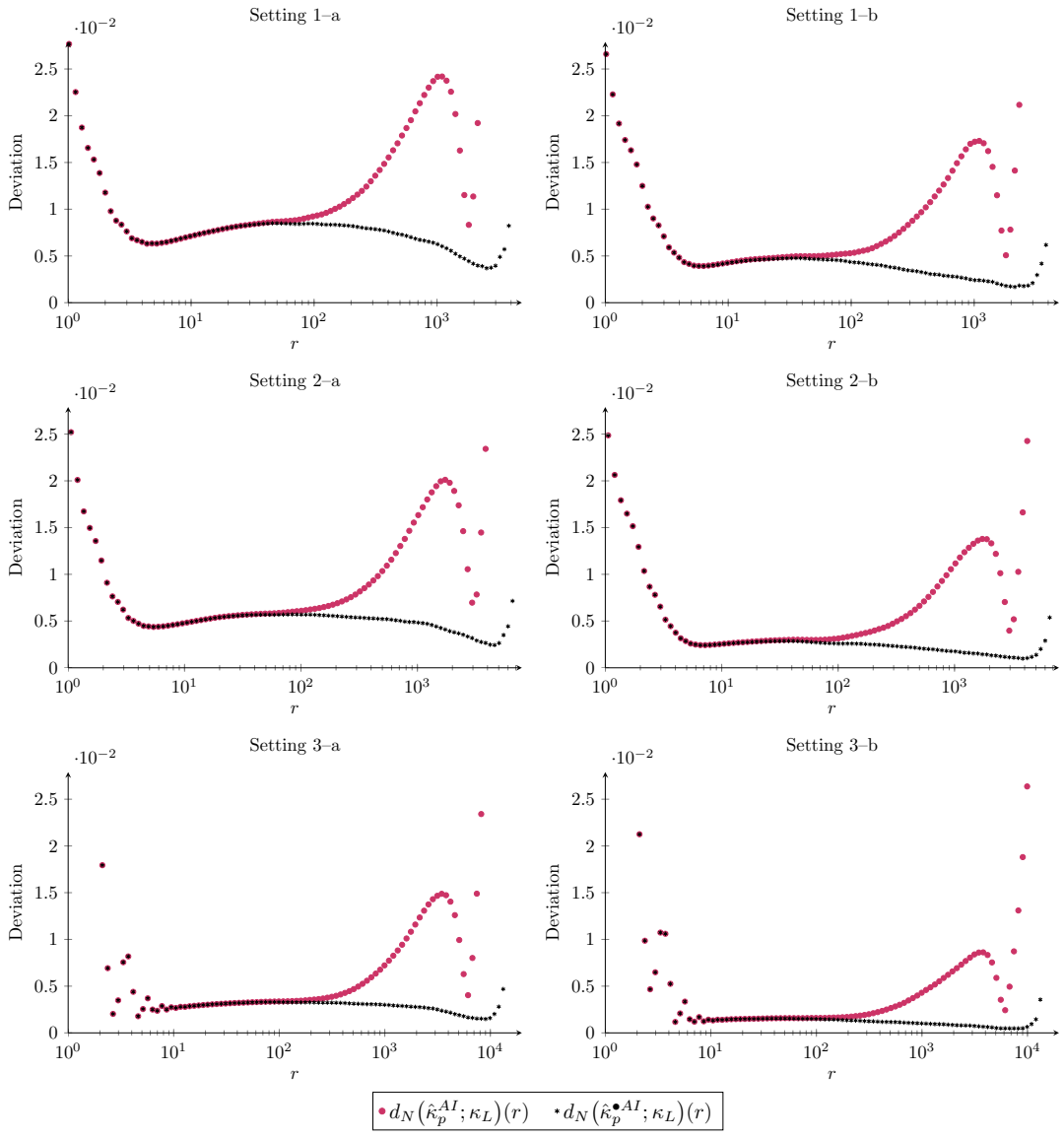


Figure 4.7: Estimated deviations of $\hat{\kappa}_p^{AI}(r)$ and $\hat{\kappa}_p^{\bullet AI}(r)$ from the true value $\kappa_L(r)$.

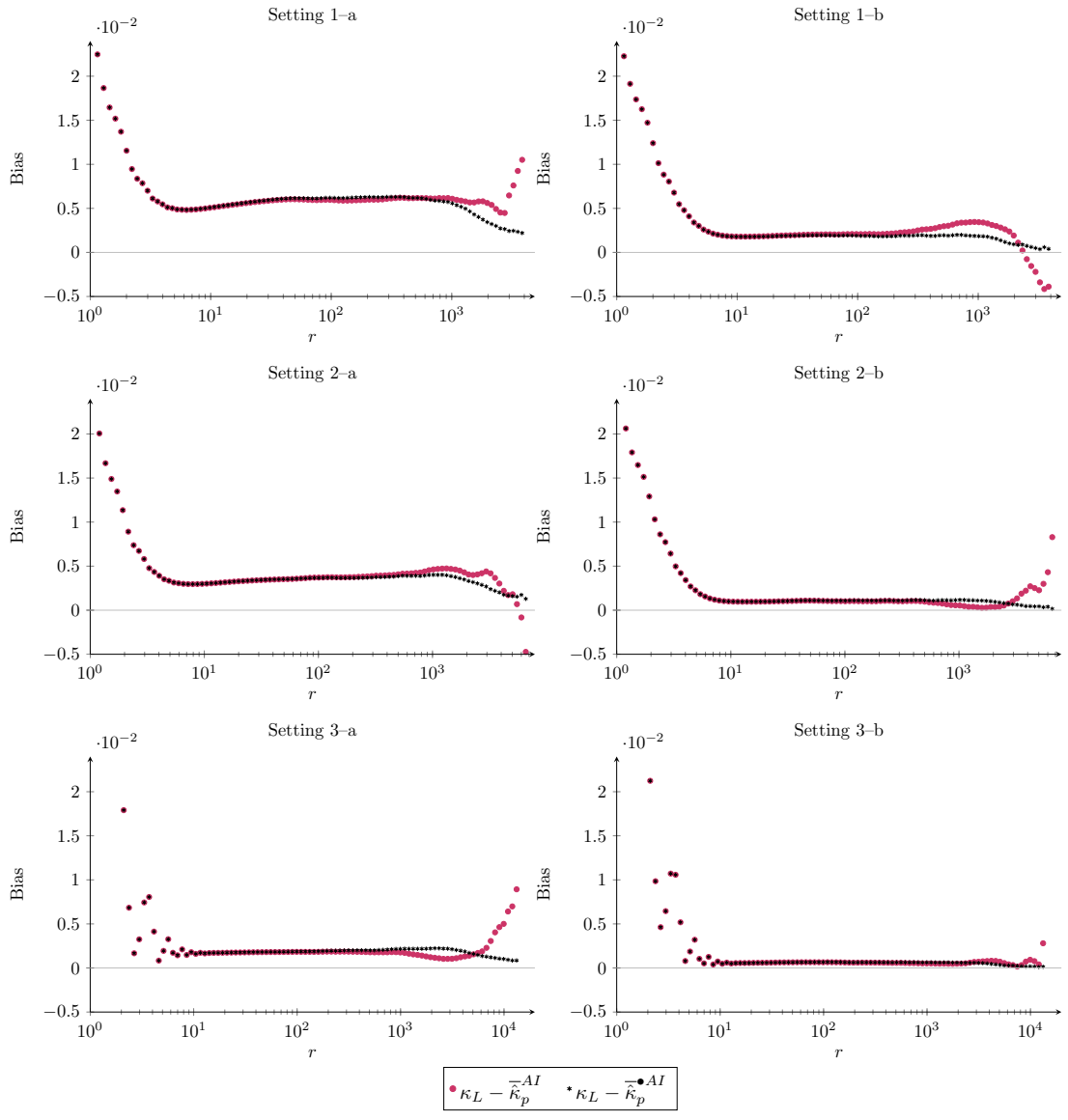


Figure 4.8: Estimated bias of $\hat{\kappa}_p^{AI}(r)$ and of $\hat{\kappa}_p^{\bullet AI}(r)$.

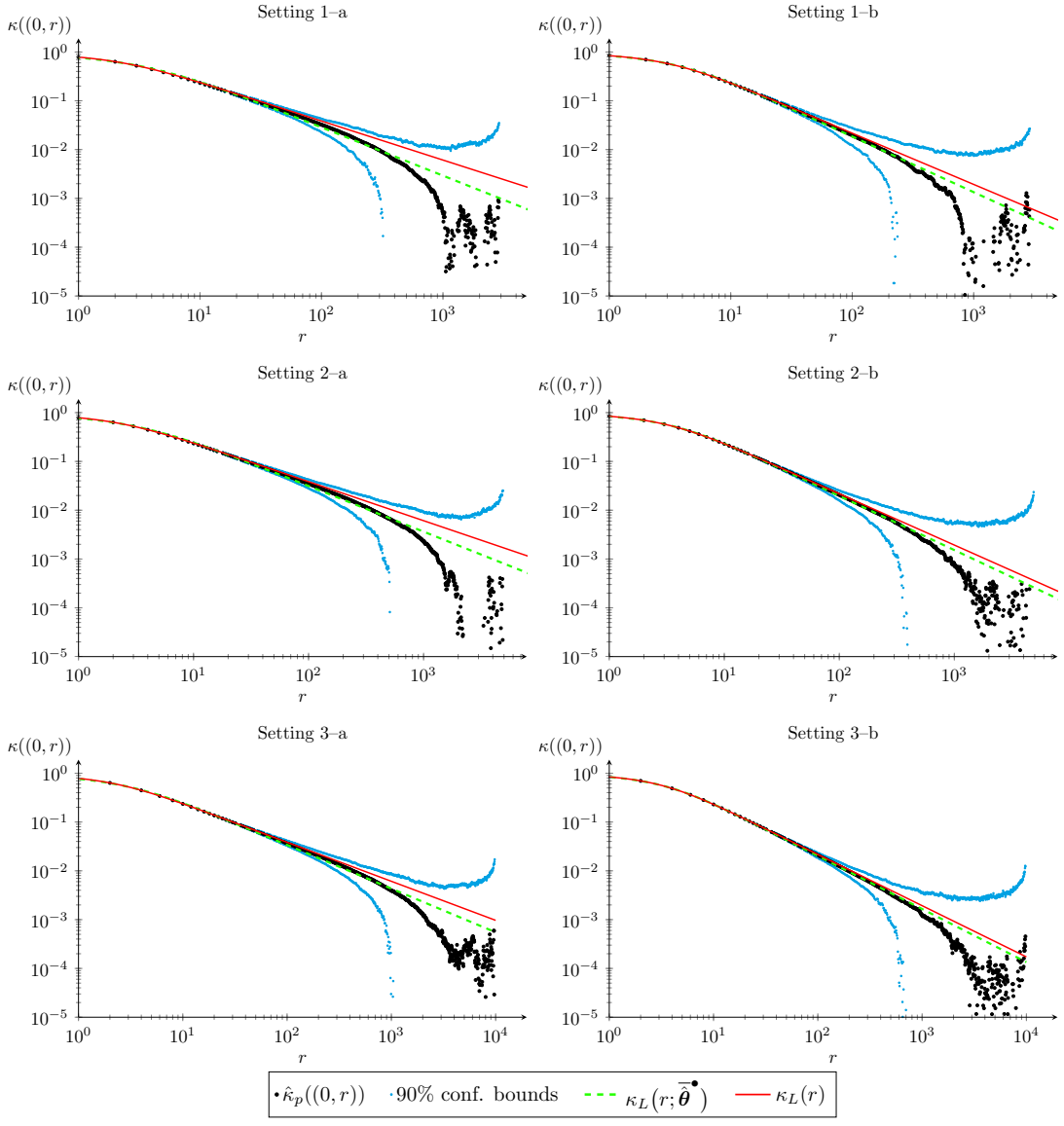


Figure 4.9: The sample mean of $\hat{\kappa}_p^{\bullet}(\boldsymbol{r})$ along $\boldsymbol{r} = (0, r)$ together with estimated two-sided 90% confidence bounds. We also plot $\kappa_L\big(r; \overline{\hat{\boldsymbol{\theta}}}^{\bullet}\big)$ where $\overline{\hat{\boldsymbol{\theta}}}^{\bullet}$ is the sample mean of fitted vector of parameters. The true value $\kappa_L(r)$ is superimposed.

Table 4.4: The summary of estimated properties of $\hat{\boldsymbol{\theta}} = (\hat{\alpha}, \hat{\beta}, \hat{\theta})$ based on $\hat{\kappa}_p(\boldsymbol{r})$ and of $\hat{\boldsymbol{\theta}}^\bullet = (\hat{\alpha}^\bullet, \hat{\beta}^\bullet, \hat{\theta}^\bullet)$ based on $\hat{\kappa}_p^\bullet(\boldsymbol{r})$. $\overline{\hat{\alpha}}$ is the sample mean, $I_{0.9}(\hat{\alpha})$ the 90% confidence interval, $d_N(\hat{\alpha}, \alpha)$ the deviation from the true value given by (4.37), $s_N(\hat{\alpha})$ the sample standard deviation given by (4.36), and analogously for β, θ and $\hat{\boldsymbol{\theta}}^\bullet$.

Setting	α	$\overline{\hat{\alpha}}$	$I_{0.9}(\hat{\alpha})$	$d_N(\hat{\alpha}, \alpha)$	$s_N(\hat{\alpha})$	$\overline{\hat{\alpha}}^\bullet$	$I_{0.9}(\hat{\alpha}^\bullet)$	$d_N(\hat{\alpha}^\bullet, \alpha)$	$s_N(\hat{\alpha}^\bullet)$
1–a	1.8	1.46	$[0.89, 2.00]$	0.52	0.40	1.49	$[1.31, 1.74]$	0.34	0.14
2–a	1.8	1.48	$[0.89, 2.00]$	0.51	0.40	1.53	$[1.38, 1.80]$	0.30	0.13
3–a	1.8	1.54	$[0.90, 2.00]$	0.49	0.42	1.55	$[1.39, 1.85]$	0.29	0.14
1–b	2.0	1.77	$[1.39, 2.00]$	0.32	0.22	1.87	$[1.76, 2.00]$	0.15	0.08
2–b	2.0	1.82	$[1.46, 2.00]$	0.27	0.20	1.90	$[1.80, 2.00]$	0.12	0.07
3–b	2.0	1.84	$[1.55, 2.00]$	0.23	0.17	1.93	$[1.85, 2.00]$	0.09	0.05

	β	$\overline{\hat{\beta}}$	$I_{0.9}(\hat{\beta})$	$d_N(\hat{\beta}, \beta)$	$s_N(\hat{\beta})$	$\overline{\hat{\beta}}^\bullet$	$I_{0.9}(\hat{\beta}^\bullet)$	$d_N(\hat{\beta}^\bullet, \beta)$	$s_N(\hat{\beta}^\bullet)$
1–a	0.8	1.18	$[0.62, 2.22]$	0.65	0.52	0.99	$[0.81, 1.16]$	0.22	0.11
2–a	0.8	1.08	$[0.60, 1.88]$	0.49	0.40	0.94	$[0.79, 1.07]$	0.16	0.09
3–a	0.8	0.97	$[0.59, 1.60]$	0.35	0.31	0.90	$[0.78, 0.99]$	0.12	0.07
1–b	1.05	1.22	$[0.82, 1.74]$	0.34	0.29	1.14	$[0.99, 1.28]$	0.13	0.09
2–b	1.05	1.14	$[0.76, 1.56]$	0.25	0.24	1.11	$[1.01, 1.21]$	0.09	0.06
3–b	1.05	1.11	$[0.85, 1.37]$	0.17	0.16	1.09	$[1.02, 1.15]$	0.06	0.04

	θ	$\overline{\hat{\theta}}$	$I_{0.9}(\hat{\theta})$	$d_N(\hat{\theta}, \theta)$	$s_N(\hat{\theta})$	$\overline{\hat{\theta}}^\bullet$	$I_{0.9}(\hat{\theta}^\bullet)$	$d_N(\hat{\theta}^\bullet, \theta)$	$s_N(\hat{\theta}^\bullet)$
1–a	3.0	7.90	$[1.91, 24.37]$	9.84	8.54	4.38	$[3.16, 5.64]$	1.58	0.78
2–a	3.0	6.70	$[1.87, 19.97]$	7.18	6.16	4.04	$[2.98, 4.93]$	1.20	0.60
3–a	3.0	5.52	$[1.72, 15.34]$	5.34	4.71	3.80	$[2.90, 4.59]$	0.95	0.52
1–b	4.0	5.20	$[2.96, 8.95]$	2.23	1.89	4.49	$[3.78, 5.13]$	0.65	0.42
2–b	4.0	4.69	$[2.63, 7.53]$	1.64	1.49	4.34	$[3.87, 4.83]$	0.46	0.31
3–b	4.0	4.49	$[3.07, 6.21]$	1.13	1.02	4.24	$[3.91, 4.55]$	0.31	0.20

namely: for small values of r the values of the estimator and the fit follows the true correlation function well. For medium values of r there is an almost straight plateau of values still very close to the true function that drops down to negative values with possible small bump before that decrease. Then there is a region, where the values are predominantly negative, which can be seen as the absence of $\hat{\kappa}^{\bullet AI}(r)$. Finally, the estimators again reach the positive values with usually overcoming even the true values of κ.

As a conclusion we may recommend to use the intrinsically balanced estimators and to be aware of the possible resulting bias that can be large especially for small windows. One should also note, that obtained results often suggest weaker correlation structure than is actually presented.

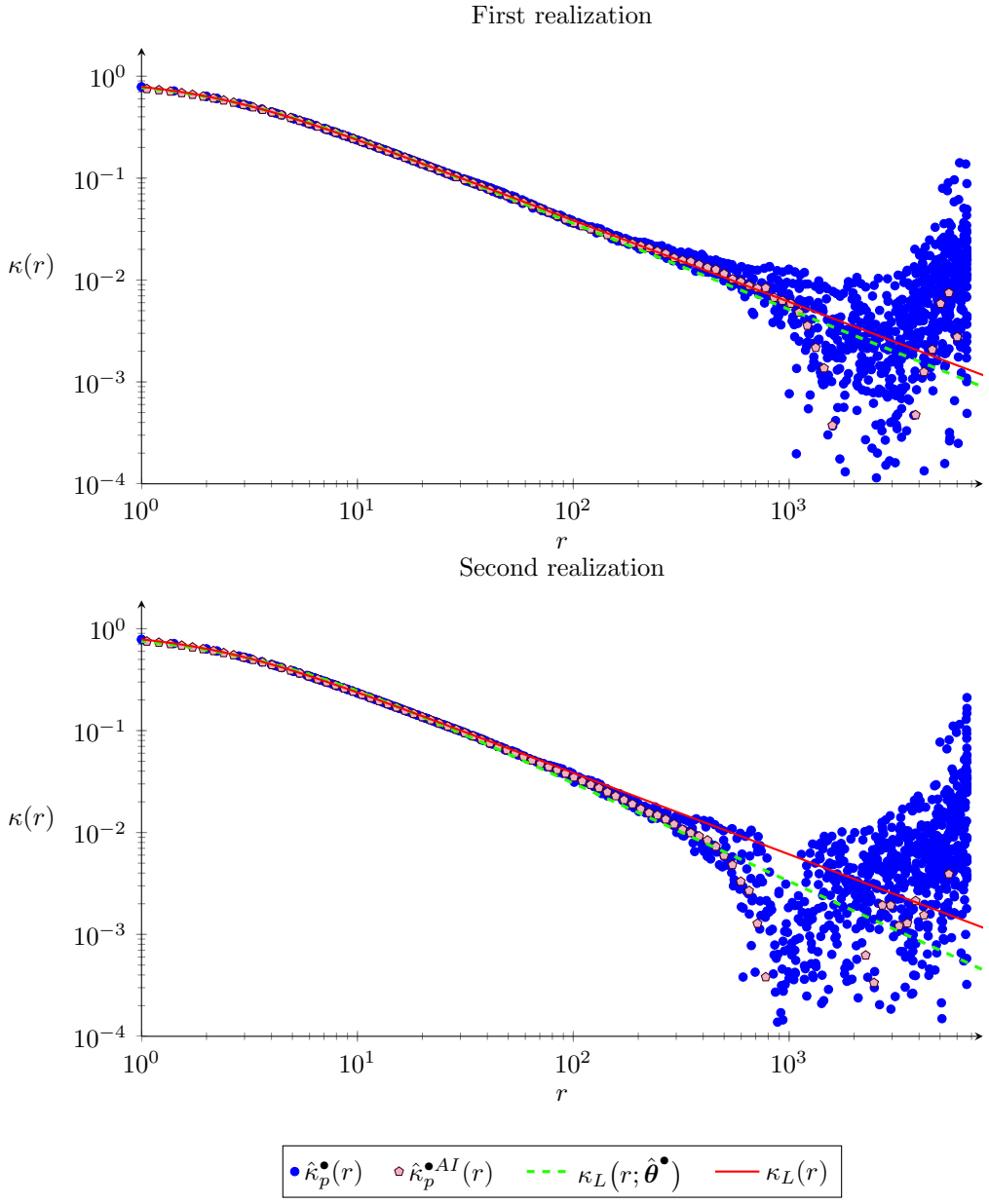


Figure 4.10: Correlation function estimators based on $\hat{\kappa}_p^{\bullet}(\boldsymbol{r})$ calculated for two different realization of 0-level excursion set $X_0(Z)$ in the setting 2–a. Fitted parameters are $\hat{\alpha}^{\bullet} = 1.66$, $\hat{\beta}^{\bullet} = 0.85$, $\hat{\theta}^{\bullet} = 3.36$ for the first realization and $\hat{\alpha}^{\bullet} = 1.44$, $\hat{\beta}^{\bullet} = 0.97$, $\hat{\theta}^{\bullet} = 4.34$. The true covariance function κ_L is also shown.

Chapter 5

Built-up structure properties

This chapter is devoted to the introduction and analysis of a built-up structure in cities. After introducing a built-up structure as a collection of buildings that are represented by polygons, formulas for the centroid and area of a polygon are derived. Then we focus on centres of cities. The built-up area is there taken as a realization of a stationary random closed set. We especially study the second order properties using the direct correlation function estimators and using the analysis of the variance in balls. In the next section a wider city area is analysed. Here the built-up area is taken as a realisation of non-stationary random closed set. As a basic characteristic we estimate the volume fraction using non-parametric kernel method introduced in the previous chapter. The optimal bandwidth is determined by an iterative procedure based on the correlation function estimation and an assumption that it depends only on the distance between two points. In subsequent sections also the radial cumulative volume fraction dependence and the distribution of building sizes are analysed. Finally, the connection between the fractality and observed long-range dependence is discussed.

5.1 Data description

The detailed spatial information about buildings is available in various countries. In some of them, like the Czech Republic, it is based on exact geodetic measurements and is collected mainly for tax purposes. Here, buildings can be recognised easily as a unique type of land. In other countries, especially in the USA, the spatial information about buildings is based on automatic extraction through very high resolution aerial or satellite imagery. Those data are usually collected by local government authorities as a part of the Geographic Information System (GIS) and are available for download through respective web pages. The next general source is the `OpenStreetMap` (www.openstreetmap.org). In our study we use the GIS based data for buildings in USA cities and `OpenStreetMap` in the rest of the world. Complete details about analysed datasets and their sources are given in Appendix B. In the following analysis we sometimes use also the information about water areas which give a natural constrain to building construction. Those data are solely from the `OpenStreetMap`. In Figure 5.1 an example of a part of Berlin is shown.

All used datafiles were in ESRI Shapefile format. For a complete specification of this format see [100]. Regardless of the source, the data were transformed to a representation, where individual buildings are given by polygons in the generalized sense explained below. These polygons representing buildings are called **building footprints**. The coordinates of points defining them are always given in optimal local coordinate reference system. The summary of used coordinate systems can be found in Table B.3.

The axes of each coordinate system are always treated as perpendicular and the unit of length is one metre. The points can therefore be considered as elements of $\mathbb{R}^2$. A polygon P is a compact subset of $\mathbb{R}^2$ such that its boundary ∂P consists of one or more rings: $\partial P =$

Figure 5.1: Polygon representation of the built-up structure in the centre of Berlin. Shaded polygons are building footprints and blue polygons correspond to water areas (Spree river).

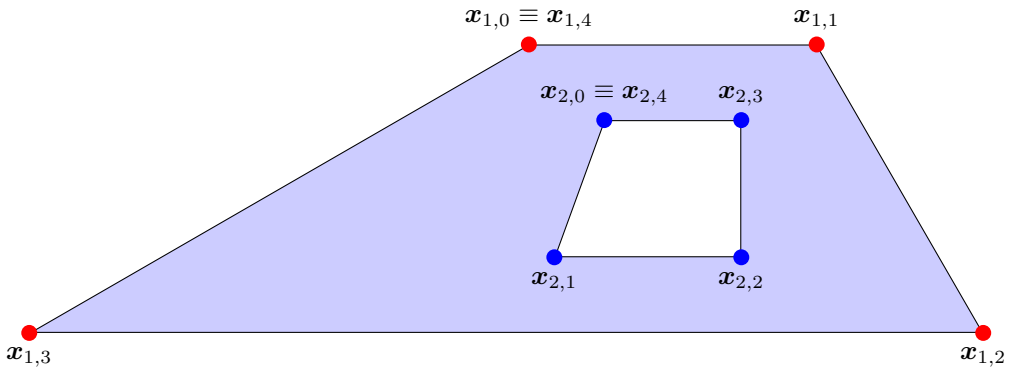


Figure 5.2: A polygon P with $\partial P = \{C_1, C_2\}$, $C_1 = \{\boldsymbol{x}_{1,0}, \boldsymbol{x}_{1,1}, \boldsymbol{x}_{1,2}, \boldsymbol{x}_{1,3}, \boldsymbol{x}_{1,4}\}$, and $C_2 = \{\boldsymbol{x}_{2,0}, \boldsymbol{x}_{2,1}, \boldsymbol{x}_{2,2}, \boldsymbol{x}_{2,3}, \boldsymbol{x}_{2,4}\}$. The outer ring C_1 has clockwise ordered vertices and the inner ring C_2 has counter-clockwise ordered vertices.

$\{C_1, \ldots, C_k\}$. The ith ring C_i is a closed non-self-intersecting curve specified by a sequence of four or more points, $C_i = \{\boldsymbol{x}_{i,0}, \boldsymbol{x}_{i,1}, \ldots, \boldsymbol{x}_{i,n_i}\}$, called vertices so that the curve consists of line segments connecting the consecutive vertices. Clearly, the last and the first point are identical, $\boldsymbol{x}_{i,0} \equiv \boldsymbol{x}_{i,n_i}$. The boundary of a polygon may contain multiple outer and inner rings that do not intersect. The order of vertices indicates which side of the ring is the interior of the polygon. The neighbourhood on the right hand side of an observer walking along the ring in vertex order is the neighbourhood inside the polygon. Vertices for a single-ringed polygon are therefore always in the clockwise order. Rings defining holes have the counter-clockwise orientation. An example of a polygon with a hole is presented in Figure 5.2.

5.2 Formulae for the area and centroid of a polygon

For a given polygon P its area $\nu_2(P)$ and the centroid $\boldsymbol{z}(P) \in \mathbb{R}^2$ can be easily calculated. To derive the respective formulas let us begin with a simple case of single ringed polygon P, $\partial P = \Big\{\{\boldsymbol{x}_0, \boldsymbol{x}_1, \ldots, \boldsymbol{x}_n \equiv \boldsymbol{x}_0\}\Big\}$, where the vertices are clockwise ordered. Let the origin of a given coordinate system be denoted by $\mathbf{0}$ and let $\boldsymbol{x}_i = \big(x_{i;1}, x_{i;2}\big)$ for all $i = 0, \ldots, n$.

It is well known that the norm of a vector product $\boldsymbol{u} \times \boldsymbol{v}$ in $\mathbb{R}^3$ equals to the area of parallelogram with sides $\boldsymbol{u}, \boldsymbol{v}$ and that the direction follows the right hand rule. If we therefore take the half of the third component of

$$\big(x_{0;1}, x_{0;2}, 0\big) \times \big(x_{1;1}, x_{1;2}, 0\big) = \big(0, 0, x_{0;1}x_{1;2} - x_{1;1}x_{0;2}\big),$$

it gives us both the area of a triangle $\mathbf{0}\boldsymbol{x}_0\boldsymbol{x}_1$ and the information about the ordering. If the third component is positive, then the points $\boldsymbol{x}_0, \boldsymbol{x}_1$ are counter-clockwise ordered around the origin. For the clockwise ordered single ringed polygon P, it is easy to see that the area of P is given by minus the sum off all oriented areas of $\mathbf{0}\boldsymbol{x}_i\boldsymbol{x}_{i+1}$ for all i. Hence the area of the whole clockwise ordered single ringed polygon P, $\partial P = \{\{\boldsymbol{x}_0, \ldots, \boldsymbol{x}_n\}\}$, can be calculated by

$$\nu_2(P) = -\frac{1}{2}\sum_{j=0}^{n-1}\big(x_{j;1}x_{j+1;2} - x_{j+1;1}x_{j;2}\big).$$

The previous formula is also a consequence of the divergence theorem well known from geometry (e.g. [65, 101, 67]). We can use the version in $\mathbb{R}^2$ saying that if $V \subset \mathbb{R}^2$ is compact with a piecewise smooth boundary ∂V and $\vec{F}$ is a continuously differentiable vector field on a neighbourhood of V, then

$$\int_V (\nabla \cdot \vec{F})\,\mathrm{d}\boldsymbol{x} = \oint_{\partial V} (\vec{F} \cdot \vec{n})\,\mathrm{d}l,$$

where $\vec{n}$ is the outward pointing unit normal field of the boundary ∂V which is traversed counter-clockwise. Taking $V = P, \vec{F}(\boldsymbol{x}) = \boldsymbol{x}/2 = (x_1/2, x_2/2)$ together with the fact that ∂P is traversed clockwise, one obtains the desired result.

If we use the same formula for an inner ring of a general polygon, it gives us the negative value because the vertices of that ring are counter-clockwise ordered. Moreover, since the rings $C_1, \ldots, C_k$ of a polygon P with $\partial P = \{C_1, \ldots, C_k\}$ do not intersect each other, it is easy to see that the general formula for the area of P is given by

$$\nu_2(P) = -\frac{1}{2}\sum_{i=1}^{k}\sum_{j=0}^{n_i-1}\big(x_{i,j;1}x_{i,j+1;2} - x_{i,j+1;1}x_{i,j;2}\big), \tag{5.1}$$

where $C_i = \{\boldsymbol{x}_{i,0}, \boldsymbol{x}_{i,2}, \ldots, \boldsymbol{x}_{i,n_i}\}$ for all $i = 1, \ldots, k$. Note that the minus sign is necessary as the outer ring is always clockwise ordered.

For the centroid a polygon P is taken as an object with uniform density. Under these circumstances, its centroid can be defined by

$$\boldsymbol{z}(P) = \frac{1}{\nu_2(P)}\int_P \boldsymbol{x}\,\mathrm{d}\boldsymbol{x} = \frac{1}{\nu_2(P)}\left(\int_P x_1\,\mathrm{d}\boldsymbol{x}, \int_P x_2\,\mathrm{d}\boldsymbol{x}\right),$$

because we assume that the area $\nu_2(P)$ of P is always positive. The centroid of a triangle $O\boldsymbol{x}_0\boldsymbol{x}_1$ is known to be at $2/3$ of the median. That gives us

$$\boldsymbol{z}(\mathbf{0}\boldsymbol{x}_0\boldsymbol{x}_1) = \frac{2}{3}\frac{(\boldsymbol{x}_0 + \boldsymbol{x}_1)}{2} = \frac{\boldsymbol{x}_0 + \boldsymbol{x}_1}{3}.$$

Generally for a triangle $\boldsymbol{y}\boldsymbol{x}_0\boldsymbol{x}_1$ we have $\boldsymbol{z}(\boldsymbol{y}\boldsymbol{x}_0\boldsymbol{x}_1) = \boldsymbol{y} + ((\boldsymbol{x}_0 - \boldsymbol{y}) + (\boldsymbol{x}_1 - \boldsymbol{y}))/3$.

Similarly as for the area, the centroid of a single ringed polygon P can be calculated by

$$\boldsymbol{z}(P) = -\frac{1}{6\nu_2(P)} \sum_{j=0}^{n-1} \big(x_{j;1}x_{j+1;2} - x_{j+1;1}x_{j;1}\big)\big(\boldsymbol{x}_j + \boldsymbol{x}_{j+1}\big).$$

One can prove this formula directly or use the divergence theorem with $V = P, \vec{F}(\boldsymbol{x}) = (x_1^2, x_1x_2)$ for the first component and analogously the second one. Finally, it easily follows that the centroid $\boldsymbol{z}(P)$ of a general polygon P with $\partial P = \{C_1, \dots, C_k\}$, $C_i = \{x_{i,0}, \dots, \boldsymbol{x}_{i,n_i}\}$, $\boldsymbol{x}_{i,j} = (x_{i,j;1}, x_{i,j;2})$, can be calculated by

$$\boldsymbol{z}(P) = -\frac{1}{6\nu_2(P)} \sum_{i=1}^{k} \sum_{j=0}^{n_i-1} \big(x_{i,j;1}x_{i,j+1;2} - x_{i,j+1;1}x_{i,j;2}\big)\big(\boldsymbol{x}_{i,j} + \boldsymbol{x}_{i,j+1}\big),$$

where $\nu_2(P)$ can be calculated using (5.1).

5.3 Built-up area as a stationary random closed set

In this part we analyse the built-up area from the perspective of theory of random closed sets introduced in Section 2.2. Thus we take the built-up area in a particular city as a realization of a random closed set X in $\mathbb{R}^2$ observed in a sample window $W \subset \mathbb{R}^2$. Moreover, we focus only on a central part of the city where, as was argued in [1, 23], the volume fraction is roughly constant. This assumptions is reasonable since throughout the historical development every location in this area had the same theoretical chance of being occupied by a building.

In the following it is therefore assumed that the random closed set X is stationary. It is often argued (see e.g. discussions in [82, 102]) that this approach is pragmatically justified when one studies a small subregion, where the pattern appear stationary. The stationarity implies that the volume fraction is constant and the analysis could then concentrate on investigating of the second order properties. Moreover, we assume that X is ergodic, an assumption not susceptible to statistical analysis if there is only one sample, but one, that is necessary if any statistical analysis is sensible.

For the stationarity assumption it is necessary to choose the observation window W properly. Since we do not want to rely on methods from the following section that is devoted to non-stationary approach, we choose W subjectively based on a visualisation of building footprints in a city centre. In the next section we justify the choice for each city by a comparison with volume fraction estimates. When choosing W it is appropriate to exclude natural constrains of the urbanization. Those are particularly represented by large water areas like seas, rivers, and lakes and it is reasonable to exclude them from the analysis. In practice, we choose W as some rectangular region D from which the polygons corresponding to water areas are subtracted. An example of this choice for Pittsburgh is shown in Figure 5.3. The GPS coordinates of rectangular parts of chosen windows are summarized in Table 5.1.

Having W, the observed part $X \cap W$ of X can be obtained by

$$X \cap W = \bigcup_{i=1}^{n} (P_i \cap W),$$

where $\{P_i\}_{i=1}^n$ are the building footprints in a given city. The volume fraction p of X is estimated using the empirical volume fraction $\hat{p}_v$ defined by (4.2),

$$\hat{p}_v = \frac{\nu_2(X \cap W)}{\nu_2(W)} = \frac{1}{\nu_2(W)} \sum_{i=1}^{n} \nu_2(P_i \cap W).$$

Since W is a polygon and an intersection of two polygons is again a polygon, possibly of zero volume, we may easily calculate all terms in the previous formula using (5.1). The results for analysed cities are shown in Table 5.2.

Figure 5.3: Observation window selection for Pittsburgh.

Table 5.1: GPS coordinates of rectangular bounding box D of W.

	Lower left corner		Upper right corner		
City	Latitude	Longitude	Latitude	Longitude	Size of D [m]
Boston	N 42° 20.275′	W 71° 05.813′	N 42° 22.803′	W 71° 03.101′	3700 × 4700
Chicago	N 41° 54.599′	W 87° 44.203′	N 41° 57.053′	W 87° 38.608′	7700 × 4600
Los Angeles	N 33° 58.910′	W 118° 17.280′	N 34° 01.893′	W 118° 13.782′	5400 × 5500
Pittsburgh	N 40° 26.464′	W 79° 58.166′	N 40° 28.142′	W 79° 55.178′	4300 × 3000
Seattle	N 47° 36.214′	W 122° 21.562′	N 47° 37.755′	W 122° 19.291′	2900 × 2800
Berlin	N 52° 29.814′	E 13° 21.379′	N 52° 32.186′	E 13° 26.050′	5200 × 4500
Birmingham	N 52° 28.238′	W 1° 54.788′	N 52° 29.692′	W 1° 52.134′	3000 × 2700
Milan	N 45° 27.512′	E 9° 10.897′	N 45° 28.659′	E 9° 12.705′	2350 × 2130
Minsk	N 53° 53.387′	E 27° 33.764′	N 53° 55.044′	E 27° 36.617′	3100 × 3100
Moscow	N 55° 44.767′	E 37° 35.633′	N 55° 46.415′	E 37° 38.446′	3000 × 3000
Oslo	N 59° 54.605′	E 10° 43.375′	N 59° 55.493′	E 10° 45.038′	1550 × 1650
Paris	N 48° 50.878′	E 2° 18.711′	N 48° 53.055′	E 2° 22.200′	4300 × 4000
Prague	N 50° 04.296′	E 14° 25.055′	N 50° 05.512′	E 14° 26.488′	2000 × 2000
St. Petersburg	N 59° 55.355′	E 30° 17.378′	N 59° 56.640′	E 30° 20.610′	3000 × 2400

5.3.1 Correlation function estimation

To study second order properties of X we use the correlation function $\kappa(\boldsymbol{r})$. Since it is normalized it enables us to compare the results for individual cities. The best available estimator of $\kappa(\boldsymbol{r})$ is the intrinsically balanced estimator $\hat{\kappa}_v^\bullet(\boldsymbol{r})$ given by (4.14). However, it is computationally demanding because one has to calculate the volume of the intersection $X \cap (X-\boldsymbol{r}) \cap W \cap (W-\boldsymbol{r})$ for every $\boldsymbol{r}$ for which $\kappa(\boldsymbol{r})$ is to be known. Since there are typically thousands of buildings in a city core (and in W), this approach is practically unusable.

It is advantageous to use a digitalized version of the estimator. This approach was introduced in Subsection 4.2.3. Hence we first digitalize $W \cap X$ by sampling the values of the indicator $\mathbb{1}_{X\cap W}$ on some fixed grid of points given by the intersection of a regular point lattice $L^2 = a\mathbb{Z}^2 + \boldsymbol{c}$ with a rectangular bounding box D of W. The shift $\boldsymbol{c}$ of the lattice is always chosen such that the lower left corner of D corresponds to $a(1,1)+\boldsymbol{c}$. Let $D = [\boldsymbol{a}, \boldsymbol{b}] \equiv [a_1, b_1] \times [a_2, b_2]$.

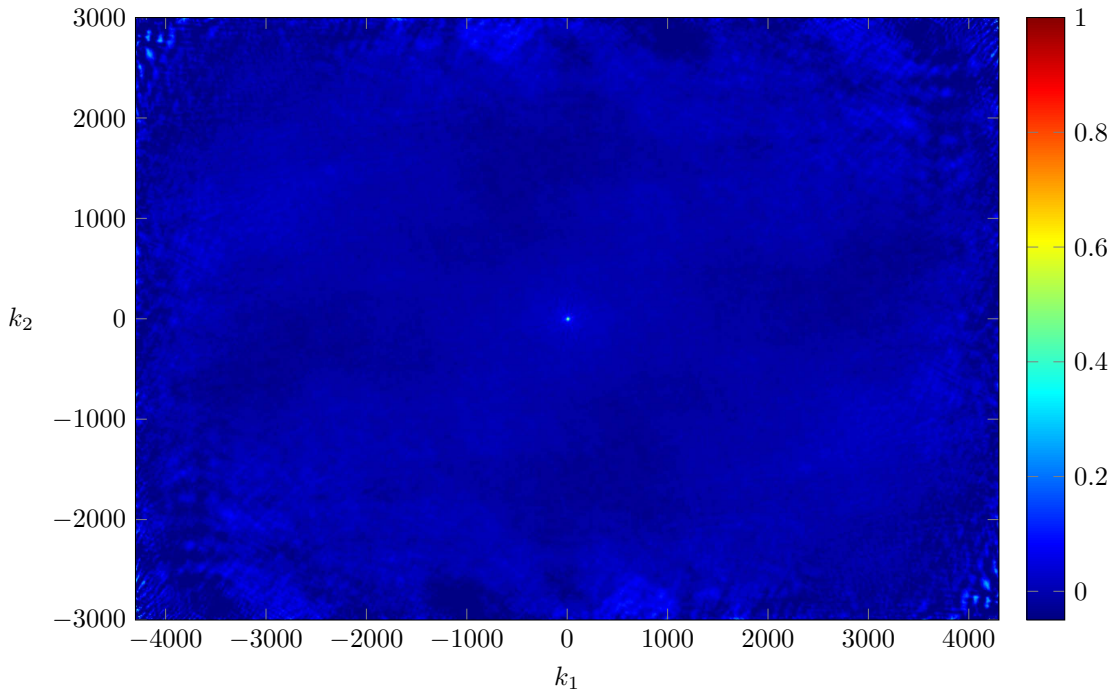


Figure 5.4: The estimation $\hat{\kappa}_p^{\bullet}(\boldsymbol{k})$ of the correlation function for Pittsburgh. The built-up structure is sampled with $a = 1\,\mathrm{m}$.

The values of the indicator $\mathbb{1}_{X\cap W}$ on the grid define a matrix $\boldsymbol{M} \in \{0,1\}^{n_1,n_2}$ by

$$(\boldsymbol{M})_{\boldsymbol{i}} = \mathbb{1}_{X\cap W}\big(a\boldsymbol{i}+\boldsymbol{c}\big), \quad \boldsymbol{i} \in \{1,\dots,n_1\}\times\{1,\dots,n_2\},$$

where $n_j \in \mathbb{N}, (b_j - a_j)/a - 1 \le n_j \le (b_j - a_j)/a$ for $j = 1, 2$. Similarly we set

$$(\boldsymbol{I})_{\boldsymbol{i}} = \mathbb{1}_{W}\big(a\boldsymbol{i}+\boldsymbol{c}\big), \quad \boldsymbol{i} \in \{1,\dots,n_1\}\times\{1,\dots,n_2\},$$

for the observation window only.

The discrete version $\hat{\kappa}_p^{\bullet}(\boldsymbol{r})$ of the estimator $\hat{\kappa}_v^{\bullet}(\boldsymbol{r})$ is according to (4.20),

$$\hat{\kappa}_p^{\bullet}(a\boldsymbol{k}) = \frac{\hat{C}_p(a\boldsymbol{k}) - \hat{p}_p[\check{\mathbf{0}}, a\boldsymbol{k}]\hat{p}_p[\mathbf{0}, a\check{\boldsymbol{k}}]}{\sqrt{\hat{p}_p[\check{\mathbf{0}}, a\boldsymbol{k}]\big(1-\hat{p}_p[\check{\mathbf{0}}, a\boldsymbol{k}]\big)}\sqrt{\hat{p}_p[\mathbf{0}, a\check{\boldsymbol{k}}]\big(1-\hat{p}_p[\mathbf{0}, a\check{\boldsymbol{k}}]\big)}}$$

for all $\boldsymbol{k}$ from a subset of $\{-n_1+1,\dots,n_1-1\}\times\{-n_2+1,\dots,n_2-1\}$ such that the denominator is positive and the estimator $\hat{C}_p(a\boldsymbol{k})$ given by (4.16) is well defined, and therefore also $\hat{p}_p[\check{\mathbf{0}}, a\boldsymbol{k}], \hat{p}_p[\mathbf{0}, a\check{\boldsymbol{k}}]$ given by (4.21) are well defined. See the discussion following formula (4.16) for more details.

As we know from Subsection 4.2.3 the estimators $\hat{C}_p(a\boldsymbol{k}), \hat{p}_p[\check{\mathbf{0}}, a\boldsymbol{k}]$, and $\hat{p}_p[\mathbf{0}, a\check{\boldsymbol{k}}]$ can be calculated with help of the discrete fast Fourier transform using relations (4.19), (4.22), and (4.23), respectively. Note that matrices $\boldsymbol{M}$ and $\boldsymbol{I}$ have to be padded with zeros before calculating the discrete Fourier transform because of its periodicity assumption.

Throughout the analysis, the data were sampled with $a = 1\,\mathrm{m}$. A typical result of the correlation function estimation is given in Figure 5.4. The pattern seems to correspond to the isotropic situation. This assumption is reasonable also from the theoretical point of view since there are so many different directional trends in a city centre, influenced e.g. by roads,

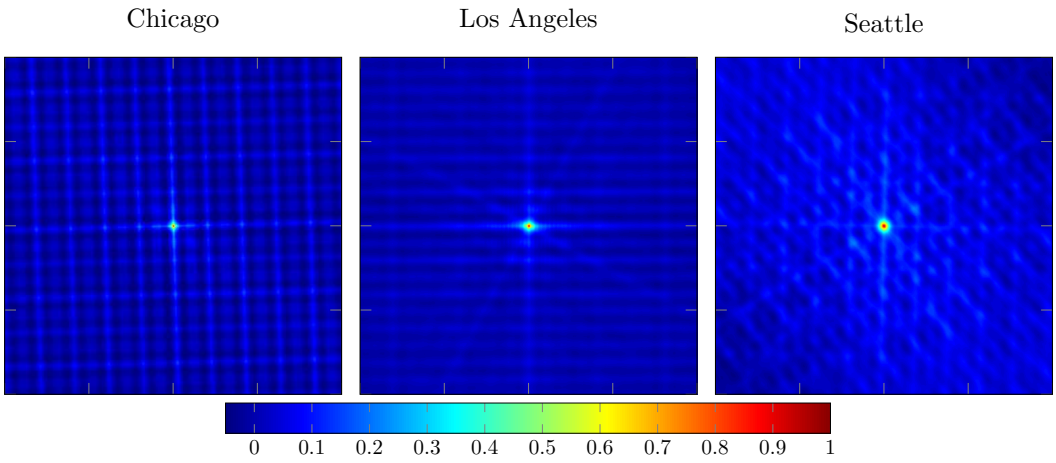


Figure 5.5: The correlation function $\hat{\kappa}_p^{\bullet}(r)$ in the range $r \in (-500, 500) \times (-500, 500)$ for North American cities with gridiron urban structure.

water sides, terrain slopes, and cardinal directions, such that we may assume that they mutually neglect. In the following it is therefore assumed that X is isotropic. Hence we may use the adapted isotropic intrinsically balanced covariance function estimator $\hat{\kappa}_p^{\bullet AI}(r)$ defined analogously to (4.24) by

$$\hat{\kappa}_p^{\bullet AI}(r) = \frac{\sum_{\boldsymbol{k}} K_{h(r)}\big(\|a\boldsymbol{k}\| - r\big)\hat{\kappa}_p^{\bullet}(a\boldsymbol{k})}{\sum_{\boldsymbol{k}} K_{h(r)}\big(\|a\boldsymbol{k}\| - r\big)},$$

where K is the Gaussian kernel, so that $K_h(x) = \frac{1}{h\sqrt{2\pi}} e^{-\frac{x^2}{2h^2}}$ and the bandwidth function $h(r)$ is chosen to be

$$h(r) = 30 \cdot (1 - e^{-0.001 r/\mathrm{m}})\,\mathrm{m} + 0.5\,\mathrm{m}.$$

Such a function represents a reasonable choice balancing the need for a small bandwidth when r is small and a sufficiently large bandwidth when r is medium or large.

Obtained estimates for analysed cities are sown in Figure 5.6. In Chicago, Los Angeles, and Seattle one can see small oscillations of the correlation function for r approximately in $(20, 1000)\,\mathrm{m}$. This is a consequence of the grid urban planning of many North American cities (see [103]). In those cities, especially in their older parts, the rigid regular shaping of the of the built-up structure is presented. Consequently, also the correlation function partially follows this regular structure as can bee seen in Figure 5.5 for Los Angeles, Chicago, and Seattle. As a result, the adapted isotropic estimator $\hat{\kappa}_p^{\bullet AI}(r)$ shows oscillations in the above mentioned range of r. Those oscillations however does not influence the overall asymptotic behaviour of the correlation function as the waves blur and eventually disappear for large values of r. Thus even in those cities we do not drop the isotropy assumption that leads to use of $\hat{\kappa}_p^{\bullet AI}(r)$ and to the following parametric fit. In the next subsection we confirm, with help of a special estimator sensitive to periodic structures, that the oscillatory background of the correlation function is of marginal importance and does not influence the overall characteristics of the correlation function decay for large r.

An example of a result of both estimators $\hat{\kappa}_p^{\bullet}(r)$ and $\hat{\kappa}_p^{\bullet AI}(r)$ for Pittsburgh is shown in Figure 5.7. We see the similar pattern as in Figure 4.10 corresponding to the 0-excursion set of a Gaussian random field with Cauchy covariance function.

Therefore it is instructive to fit estimated values of the correlation function with a function from some simple parametric class of correlation functions. Among commonly used parametric

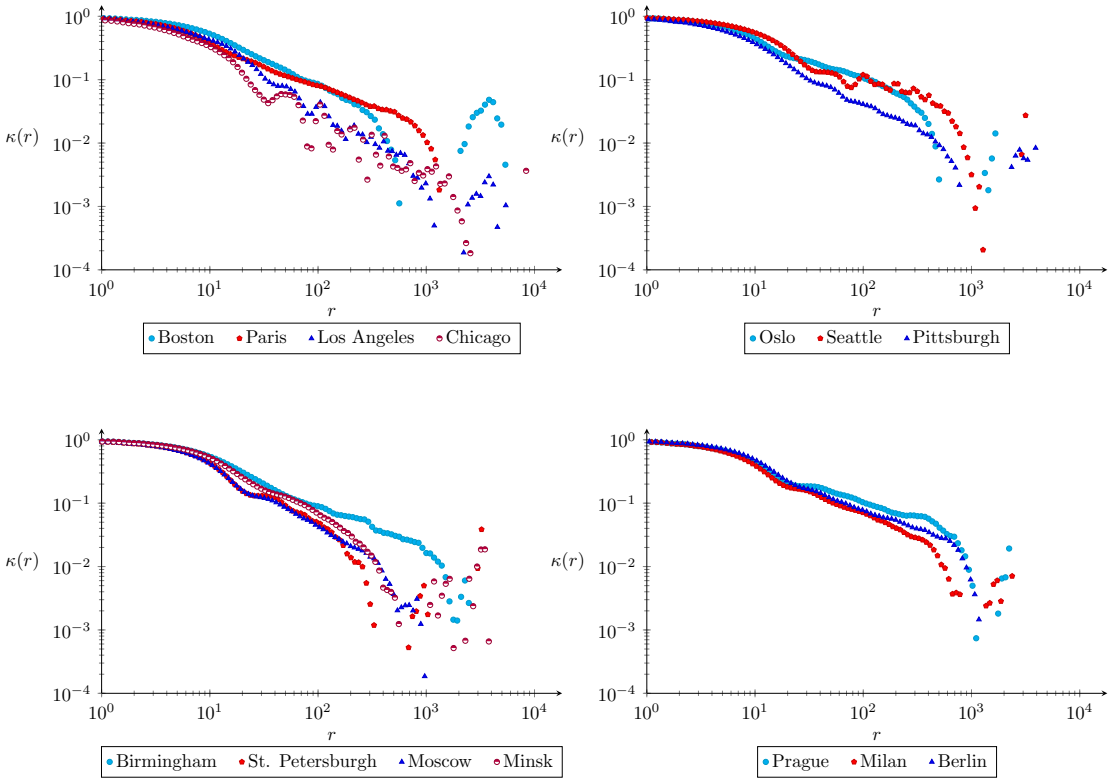


Figure 5.6: The results of the adapted estimator $\hat{\kappa}_p^{\bullet AI}(r)$ for all analysed cities.

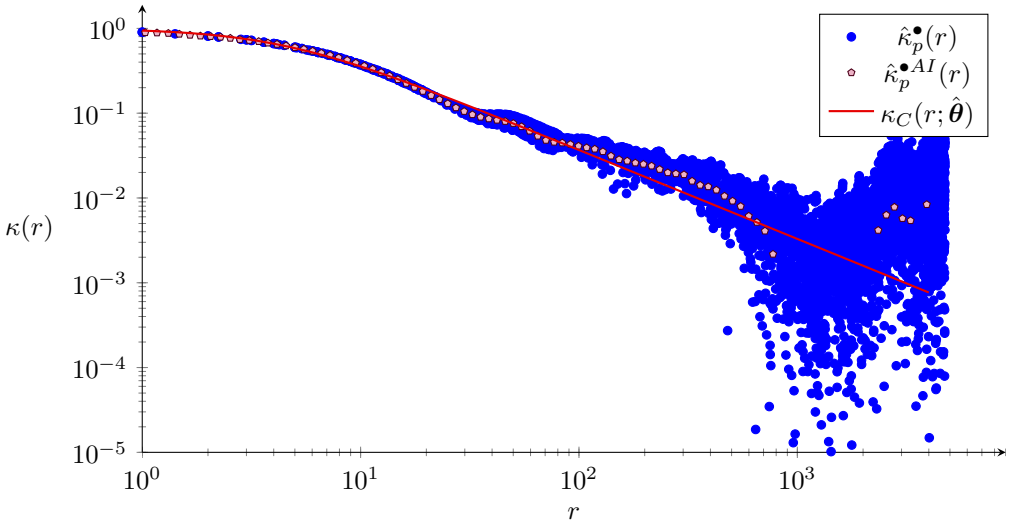


Figure 5.7: An example of Pittsburgh. The result of the intrinsically balanced estimator $\hat{\kappa}_p^{\bullet}(r)$ is shown together with result of its adapted isotropic modification $\hat{\kappa}_p^{\bullet AI}(r)$. The values of $\hat{\kappa}_p^{\bullet}(r)$ are fitted by the Cauchy correlation function $\kappa_C(r;\hat{\boldsymbol{\theta}})$ with $\hat{\alpha} = 1.60$, $\hat{\beta} = 1.05$, and $\hat{\theta} = 4.32\,\mathrm{m}$.

Table 5.2: The volume fraction and fitted values of parameters of the Cauchy correlation function for analysed cities.

City	$\hat{p}_v$	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\theta}$ [m]
Boston	0.30	1.32	1.01	7.86
Chicago	0.32	1.39	1.32	5.05
Los Angeles	0.32	1.27	1.33	7.90
Pittsburgh	0.22	1.60	1.05	4.32
Seattle	0.34	2.00	0.78	4.49
Berlin	0.26	2.00	0.84	4.23
Birmingham	0.32	2.00	0.78	4.59
Milan	0.47	2.00	0.84	3.44
Minsk	0.20	1.58	1.10	7.16
Moscow	0.35	2.00	1.07	4.60
Oslo	0.36	1.20	0.89	5.95
Paris	0.49	1.98	0.77	3.04
Prague	0.42	2.00	0.65	2.77
St. Petersburg	0.41	2.00	1.00	4.39

classes, for their inventory we refer the reader to [49, Section 2.5], the best fit is achieved by the Cauchy class that was introduced in Section 4.4. The general form of a correlation function from the Cauchy class is

$$\kappa_C(r;\boldsymbol{\theta}) = \left(1 + \left(\frac{r}{\theta}\right)^{\alpha}\right)^{-\frac{\beta}{\alpha}}, \tag{5.2}$$

where $\boldsymbol{\theta} = (\alpha, \beta, \theta)$ with $\theta > 0$, $0 < \alpha \leq 2$, and $\beta > 0$.

The fit is performed by minimizing the least squares with predefined weights. Our aim is to choose the vector $\hat{\boldsymbol{\theta}} = (\hat{\alpha}, \hat{\beta}, \hat{\theta})$ of parameters constrained to $0 < \hat{\beta}, \hat{\theta} < \infty$, $0 < \hat{\alpha} \leq 2$, such that it minimizes

$$Q(\boldsymbol{\theta}) = \sum_{\boldsymbol{k}=-\boldsymbol{n}+1}^{\boldsymbol{n}-1} w_{\boldsymbol{k}} \left(\hat{\kappa}_p^{\bullet}(a\boldsymbol{k}) - \kappa_C(\|a\boldsymbol{k}\|;\boldsymbol{\theta})\right)^2,$$

where $\boldsymbol{n} = (n_1, n_2)$ and the weights $w_{\boldsymbol{k}}$ are chosen accordingly to [49] as

$$w_{\boldsymbol{k}} = \frac{N_{\boldsymbol{k}}}{a\,\|\boldsymbol{k}\| + 1}.$$

The coefficient $N_{\boldsymbol{k}} = \sum_{\boldsymbol{i} \in A(\boldsymbol{k})} (\boldsymbol{I})_{\boldsymbol{i}} (\boldsymbol{I})_{\boldsymbol{i}+\boldsymbol{k}}$, where $A(\boldsymbol{k})$ is defined by $A(\boldsymbol{k}) = \{\boldsymbol{i} \in \mathbb{Z}^2 | 1 \leq \boldsymbol{i} \leq \boldsymbol{n}, 1 \leq \boldsymbol{i} + \boldsymbol{k} \leq \boldsymbol{n}\}$, gives the number of points that were used for the calculation of $\hat{\kappa}_p^{\bullet}(a\boldsymbol{k})$. Note that the least squares fitting is based on the estimator $\hat{\kappa}_p^{\bullet}(a\boldsymbol{k})$ and not on the adapted isotropic modification $\hat{\kappa}_p^{\bullet AI}(r)$, because it yields the better performance as was discussed in Section 4.4. An example of a fit for Pittsburgh is shown in Figure 5.7. The fitted coefficients for all cities are summarized in Table 5.2.

From the fact that the correlation function may be well fitted by the Cauchy covariance function follows that the built-up area viewed as a stationary random closed set posses a long-range dependence. To see this let note that

$$\left(1 + \left(\frac{r}{c}\right)^{\alpha}\right)^{-\frac{\beta}{\alpha}} \sim \left(\frac{r}{c}\right)^{-\beta} \qquad (r \to \infty).$$

Furthermore, from Table 5.2 follows that $\beta \in (0, d)$, where $d = 2$. Hence X satisfies Definition 2.8.2 and it is isotropic long-range dependent. One must be careful here since we were able

to study the correlation function at largest distances only around $5 \cdot 10^3$, which is surely not close to the infinity. On the other hand by comparing Figures 5.6 and 5.7 with Figures 4.9 and 4.10, the qualitative behaviour of the correlation function seems to be similar to the case of long-range dependence observed in Section 4.4 for level excursion sets. Thus we may argue that the long-range dependence of a built-up structure is at least strongly supported by our results.

5.3.2 Variance in balls: theory and simulations

As an additional approach to the analysis of the possible long-range dependence of a built-up structure we use the method based on studying variances in balls. This approach was introduced in [104] for analysis of the spatial distribution of the matter in cosmology. In connection to a built-up structure it was introduced by the author and co-authors in [105, 106]. In the following the method is systematically developed and finally some numerical experiments to confirm the approach are also performed. Then it is used to study the properties of a built-up area.

Let X be a stationary random closed set in $\mathbb{R}^2$ and let ν_X be its volume measure defined by $\nu_X(B) = \nu_2(B \cap X) \equiv \nu(B \cap X)$ for all Borel $B \subset \mathbb{R}^2$. Note that we may work in a more general space $\mathbb{R}^d$. However, since our aim is to study the planar situation only, we restrict ourselves on $\mathbb{R}^2$. In what follows the 2-dimensional Lebesgue measure is denoted by ν and a closed ball with radius r centred at $\boldsymbol{x}$ by $B_r(\boldsymbol{x})$, using a special notation $B_r(\mathbf{0}) \equiv B_r$.

The idea of the method is to analyse the dependence of the variance $\mathrm{var}\left(\nu_X(B_r)\right)$ in a closed ball B_r on the ball radius r. In the following we use the term **variance in balls** for $\mathrm{var}\left(\nu_X(B_r)\right)$ thought as a function of r. The variance $\mathrm{var}\left(\nu_X(B_r)\right)$ can be similarly to (4.3) expressed using the covariance function cov of X as

$$\mathrm{var}\left(\nu_X(B_r)\right) = \int_{B_r}\int_{B_r} \mathrm{cov}(\boldsymbol{x}-\boldsymbol{y})\,\mathrm{d}\boldsymbol{x}\,\mathrm{d}\boldsymbol{y}.$$

This can be further rewritten using the correlation function κ and volume fraction p of X as

$$\mathrm{var}\left(\nu_X(B_r)\right) = p(1-p)\int_{B_r}\int_{B_r} \kappa(\boldsymbol{x}-\boldsymbol{y})\,\mathrm{d}\boldsymbol{x}\,\mathrm{d}\boldsymbol{y} = p(1-p)\int_{B_r}\int_{B_r(\boldsymbol{v})} \kappa(\boldsymbol{u})\,\mathrm{d}\boldsymbol{u}\,\mathrm{d}\boldsymbol{v}.$$

Now the behaviour of $\mathrm{var}\left(\nu_X(B_r)\right)$ for very large r depends on the value of integral

$$A = \int_{\mathbb{R}^d} \kappa(\boldsymbol{r})\,\mathrm{d}\boldsymbol{r},$$

which is called the **integral range** whenever the integral exists. See subsection 4.2.1 for more details and references on the integral range.

There are three interesting different situations depending on the value of A. First is when $0 < A < \infty$. Then we clearly obtain the following asymptotic,

$$\mathrm{var}\left(\nu_X(B_r)\right) \sim p(1-p)A\pi r^2 \qquad (r \to \infty).$$

Thus $\mathrm{var}\left(\nu_X(B_r)\right)$ is for large r proportional to r^2.

Next case corresponds to $A = \infty$. This is usually connected with long-range behaviour that was discussed in Section 2.8. The variance in this situation is generally larger and also of faster increase than in the previous case. To see this let us take X to be an isotropically long-range dependent random closed set according to Definition 2.8.2. Hence its covariance function, and also correlation function, is of the form (2.40), where the long-range exponent is $\beta \in (0,2)$. We may use Proposition 2.8.6 and obtain

$$\mathrm{var}\left(\nu_X(B_r)\right) \sim r^{4-\beta}\, F_{\beta;B_1}\ell(r) \qquad (r \to \infty),$$

since $\lambda_D = 0$ for random closed sets as was discussed in Subsection 2.5.4. The constant F_{β,B_1} is independent of r. This means that $\operatorname{var}\big(\nu_X(B_r)\big)$ is for large r proportional to $r^{4-\beta}$, where $4-\beta > 2$.

Now we focus on $A = 0$. In this case one may expect the variance to be smaller and increase more slowly than r^2. The coefficient of a power law decay is governed by the behaviour of the Bartlett spectrum Γ of X in the vicinity of $\mathbf{0}$, which was introduced in Subsections 2.3.3 and 2.5.5. Since the covariance function is integrable, Γ is absolutely continuous with respect to ν, with corresponding density denoted by f_Γ. By Proposition 2.5.8 (d),

$$\operatorname{var}\big(\nu_X(B_r)\big) = 2\pi \int_{\mathbb{R}^2} \big|\check{\mathbb{1}}_{B_r}(\boldsymbol{\omega})\big|^2 f_\Gamma(\boldsymbol{\omega})\,\mathrm{d}\boldsymbol{\omega},$$

where $\check{\mathbb{1}}_{B_r} \equiv \mathscr{F}^{-1}\mathbb{1}_{B_r}$ is the inverse Fourier transform of the ball indicator $\mathbb{1}_{B_r}$ defined by (A.8). It can be shown that

$$\check{\mathbb{1}}_{B_r}(\boldsymbol{\omega}) = \check{\mathbb{1}}_{B_r}(\omega) = r\frac{J_1(r\omega)}{\omega}, \tag{5.3}$$

where $\omega = \|\boldsymbol{\omega}\|$ and J_1 is a Bessel function of the first kind (see e.g. [95, Chapter 9] for definition and properties). Hence by transferring to polar coordinates:

$$\operatorname{var}\big(\nu_X(B_r)\big) = (2\pi)r^2 \int_0^\infty \int_0^{2\pi} \frac{J_1^2(r\omega)}{\omega} f_\Gamma(\omega,\phi)\,\mathrm{d}\omega\,\mathrm{d}\phi$$

and after substitution $r\omega = \xi$,

$$\operatorname{var}\big(\nu_X(B_r)\big) = (2\pi)r^2 \int_0^\infty \int_0^{2\pi} \frac{J_1^2(\xi)}{\xi} f_\Gamma(\xi/r,\phi)\,\mathrm{d}\xi\,\mathrm{d}\phi.$$

Since $A = 0$, from (2.26) follows $f_\Gamma(\mathbf{0}) = 0$. Now if f_Γ behaves as $C\,\|r\|^\sigma$ with $\sigma > 0$ as $r \to 0$ and $C > 0$,then

$$\operatorname{var}\big(\nu_X(B_r)\big) \sim (2\pi)^2 r^{2-\sigma} C \int_0^\infty \frac{J_1^2(\xi)}{\xi^{1-\sigma}}\,\mathrm{d}\xi \qquad (r \to \infty). \tag{5.4}$$

Thus $\operatorname{var}\big(\nu_X(B_r)\big)$ is for large r proportional to $r^{2-\sigma}$. Note that this derivation works also for $\sigma = \beta - 2$, $0 < \beta < 2$, which by Theorem 2.8.2 corresponds to a certain subclass of isotropic long-range dependent random closed sets with power law coefficient β, that were already discussed.

Finally, the last case occurs when A is not defined. This may for example hold for simple situations of randomly shifted periodic patterns. Let assume that $X = X + \boldsymbol{b}$ for some $\boldsymbol{b} \neq \mathbf{0}$. Then clearly $\kappa(\boldsymbol{x}+\boldsymbol{b}) = \kappa(\boldsymbol{x})$. If $\kappa(\boldsymbol{x}) < 0$ for some $\boldsymbol{x}$ and its neighbourhood, we clearly obtain that A is undefined since it has infinitely large negative part. As a particular example let us take a random closed set with correlation function given by $\kappa(\boldsymbol{r}) = \operatorname{cov}(\boldsymbol{r}) = \cos(r_1)\cos(r_2)$. It is easy to check from (2.25) that the corresponding Bartlett spectrum is

$$\Gamma = \frac{2\pi}{4}\big(\delta_{(1,1)} + \delta_{(-1,1)} + \delta_{(1,-1)} + \delta_{(-1,-1)}\big),$$

where $\delta_{\boldsymbol{x}}$ is a Dirac measure at $\boldsymbol{x}$ defined by (A.1). Hence by Proposition 2.5.8 (d) and same arguments as in the previous case we obtain

$$\operatorname{var}\big(\nu_X(B_r)\big) = 2\pi^2 r^2 J_1^2\big(r\sqrt{2}\big).$$

Since in [107, Lemma IV.3.11] was shown that $J_m(r) = \sqrt{2/\pi r}\cos(r - \pi m/2 - \pi/4) + \mathcal{O}(r^{-3/2})$ as $r \to \infty$ we finally obtain

$$\operatorname{var}\big(\nu_X(B_r)\big) = \frac{4\pi}{\sqrt{2}} r \cos^2(r - \pi 3/4) + \mathcal{O}(1) \qquad (r \to \infty).$$

Hence in this case, $\operatorname{var}\big(\nu_X(B_r)\big)$ is $\mathcal{O}(r)$, i.e. of order not exceeding r, and it has a periodical component that reaches 0 every multiple of π shifted by $\pi/4$.

The previous particular case can be generalized to every periodic correlation function κ. Since it is also an even function it can be expanded to the Fourier cosine series. Hence by the similar approach as before one may show that its Bartlett spectrum is again a sum, possibly infinite, of Dirac measures. Now the situation depends on a presence or absence of $\mathbf{0}$ in the support of Γ, which is equivalent to the presence or absence of the absolute term in the Fourier series. If $\Gamma(\{0\}) > 0$, then it represents a dominating term in the variance. Since $J_1(a)/a \to 1/2$ as $a \to 0_+$ ([95, Eq. 9.1.7]) we get $\left|\check{\mathbb{1}}_{B_r}(\mathbf{0})\right|^2 = r^4/4$ and thus $\operatorname{var}\big(\nu_X(B_r)\big)$ is for large r proportional to r^4. Note that in this case the random closed set is also long-range dependent. The second possibility is when $\Gamma(\{0\}) = 0$. Since for a fixed period $\boldsymbol{b}$, the norms of frequencies appearing in Γ are bounded from below by $(2\pi)\,\|\boldsymbol{b}\|^{-1}$, we obtain, similarly to the situation with cosine alone, that $\operatorname{var}\big(\nu_X(B_r)\big)$ is $\mathcal{O}(r)$. Random closed sets with $\operatorname{var}\big(\nu_X(B_r)\big)$ being proportional to $r^{2-\sigma}$ for some $\sigma > 0$ are generally called **super-homogeneous**.

From the previous discussion follows that it is interesting to study the dependence of the variance $\operatorname{var}\big(\nu_X(B_r)\big)$ in balls. If one obtains

$$\operatorname{var}\big(\nu_X(B_r)\big) \sim C r^{\eta} \qquad (r \to \infty),$$

then for $\eta > 2$ the random closed set X is isotropically long-range dependent with coefficient $\beta = 4 - \eta$, for $\eta = 2$ it is an ordinary short range dependent, and for $\eta < 2$ it is super-homogeneous with possible periodicity for $\eta = 1$.

The non-trivial task, however, is the estimation of the variance in a bounded window W. We propose the following mechanism. For a selected radius r we choose locations $\{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n\}$ of n points in W such that $B_r(\boldsymbol{x}_i) \subset W$ for all $i = 1, \ldots, n$. For those points we determine values of $\nu_X\big(B_r(\boldsymbol{x}_i)\big) = \nu\big(X \cap B_r(\boldsymbol{x}_i)\big)$. The estimator $\hat{\operatorname{var}}_v\big(\nu_X(B_r)\big)$ of the variance is then defined by

$$\hat{\operatorname{var}}_v\big(\nu_X(B_r)\big) = \frac{1}{n-1}\sum_{i=1}^{n}\Big(\nu_X\big(B_r(\boldsymbol{x}_i)\big) - \overline{\nu_X(B_r)}\Big)^2, \tag{5.5}$$

where

$$\overline{\nu_X(B_r)} = \frac{1}{n}\sum_{i=1}^{n}\nu_X\big(B_r(\boldsymbol{x}_i)\big).$$

Note that $\mathbb{E}\,\overline{\nu_X(B_r)} = p2\pi r^2$ and hence $\overline{\nu_X(B_r)}$ is an unbiased estimator of $p\nu(B_r)$. The estimator $\hat{\operatorname{var}}_v\big(\nu_X(B_r)\big)$ can be further rewritten as

$$\begin{aligned}\hat{\operatorname{var}}_v\big(\nu_X(B_r)\big) &= \frac{1}{n-1}\sum_{i=1}^{n}\Big(\nu_X^2\big(B_r(\boldsymbol{x}_i)\big) - 2\nu_X\big(B_r(\boldsymbol{x}_i)\big)\overline{\nu_X(B_r)} + \overline{\nu_X(B_r)}^2\Big)\\ &= \frac{1}{n-1}\left(\sum_{i=1}^{n}\nu_X^2\big(B_r(\boldsymbol{x}_i)\big) - n\overline{\nu_X(B_r)}^2\right).\end{aligned}$$

Using the definition of $\overline{\nu_X(B_r)}$,

$$\begin{aligned}\hat{\mathrm{var}}_v\big(\nu_X(B_r)\big) &= \frac{1}{n-1}\left(\sum_{i=1}^{n}\nu_X^2\big(B_r(\boldsymbol{x}_i)\big) - \frac{n}{n^2}\sum_{i,j=1}^{n}\nu_X\big(B_r(\boldsymbol{x}_i)\big)\nu_X\big(B_r(\boldsymbol{x}_j)\big)\right)\\ &= \frac{1}{n(n-1)}\left((n-1)\sum_{i=1}^{n}\nu_X^2\big(B_r(\boldsymbol{x}_i)\big) - \sum_{i\neq j}^{n}\nu_X\big(B_r(\boldsymbol{x}_i)\big)\nu_X\big(B_r(\boldsymbol{x}_j)\big)\right).\end{aligned}$$

Applying the expectation yields

$$\begin{aligned}\mathbb{E}\,\hat{\mathrm{var}}_v\big(\nu_X(B_r)\big) &= \frac{1}{n(n-1)}\left((n-1)\sum_{i=1}^{n}\mathbb{E}\,\nu_X^2\big(B_r(\boldsymbol{x}_i)\big) - \sum_{i\neq j}^{n}\mathbb{E}\,\nu_X\big(B_r(\boldsymbol{x}_i)\big)\nu_X\big(B_r(\boldsymbol{x}_j)\big)\right)\\ &= \frac{1}{n(n-1)}\Bigg((n-1)\sum_{i=1}^{n}\mathbb{E}\,\nu_X^2\big(B_r(\boldsymbol{x}_i)\big) - n(n-1)\big(\mathbb{E}\,\nu_X(B_r)\big)^2\\ &\qquad + n(n-1)\big(\mathbb{E}\,\nu_X(B_r)\big)^2 - \sum_{i\neq j}^{n}\mathbb{E}\,\nu_X\big(B_r(\boldsymbol{x}_i)\big)\nu_X\big(B_r(\boldsymbol{x}_j)\big)\Bigg)\\ &= \mathrm{var}\,\big(\nu_X(B_r)\big) - \frac{1}{n(n-1)}\sum_{i\neq j}^{n}\mathrm{cov}\,\Big(\nu_X\big(B_r(\boldsymbol{x}_i)\big),\nu_X\big(B_r(\boldsymbol{x}_j)\big)\Big). \qquad (5.6)\end{aligned}$$

The covariance term can be calculated using the covariance function cov of X as

$$\mathrm{cov}\,\Big(\nu_X\big(B_r(\boldsymbol{x}_i)\big),\nu_X\big(B_r(\boldsymbol{x}_j)\big)\Big) = \int_{B_r(\boldsymbol{x}_i)}\int_{B_r(\boldsymbol{x}_j)}\mathrm{cov}(\boldsymbol{x}-\boldsymbol{y})\,\mathrm{d}\boldsymbol{x}\mathrm{d}\boldsymbol{y}.$$

If the covariance function is positive, then the estimator $\hat{\mathrm{var}}_v\big(\nu_X(B_r)\big)$ is negatively biassed. This particularly holds for many isotropic long-range dependent random closed sets, e.g. for sets with the Cauchy correlation function.

In order to analyse the asymptotic properties of $\mathrm{cov}\,\Big(\nu_X\big(B_r(\boldsymbol{x}_i)\big),\nu_X\big(B_r(\boldsymbol{x}_j)\big)\Big)$ we continue by

$$\begin{aligned}\mathrm{cov}\,\Big(\nu_X\big(B_r(\boldsymbol{x}_i)\big),\nu_X\big(B_r(\boldsymbol{x}_j)\big)\Big) &= \int_{\mathbb{R}^2}\int_{\mathbb{R}^2}\mathbb{1}_{B_r(\boldsymbol{x}_i)}(\boldsymbol{x})\mathbb{1}_{B_r(\boldsymbol{x}_j)}(\boldsymbol{y})\,\mathrm{cov}(\boldsymbol{x}-\boldsymbol{y})\,\mathrm{d}\boldsymbol{x}\mathrm{d}\boldsymbol{y}\\ &= \int_{\mathbb{R}^2}\int_{\mathbb{R}^2}\mathbb{1}_{B_r}(\boldsymbol{v}+\boldsymbol{x}_i)\mathbb{1}_{B_r}(\boldsymbol{v}+\boldsymbol{u}+\boldsymbol{x}_j)\,\mathrm{cov}(\boldsymbol{u})\,\mathrm{d}\boldsymbol{u}\mathrm{d}\boldsymbol{v}\\ &= \int_{\mathbb{R}^2}(\mathbb{1}_{B_r}*\mathbb{1}^*_{B_r})(\boldsymbol{u}+\boldsymbol{x}_j-\boldsymbol{x}_i)\,\mathrm{cov}(\boldsymbol{u})\,\mathrm{d}\boldsymbol{u}.\end{aligned}$$

Since for the inverse Fourier transform $\check{g}$ of $g(\boldsymbol{x}) = f(\boldsymbol{x}-\boldsymbol{a})$ follows $\check{g}(\boldsymbol{\omega}) = e^{i\boldsymbol{a}\cdot\boldsymbol{\omega}}\check{f}(\boldsymbol{\omega})$, then by Proposition A.2.8,

$$\mathrm{cov}\,\Big(\nu_X\big(B_r(\boldsymbol{x}_i)\big),\nu_X\big(B_r(\boldsymbol{x}_j)\big)\Big) = (2\pi)\int_{\mathbb{R}^2}e^{-i(\boldsymbol{x}_j-\boldsymbol{x}_i)\cdot\boldsymbol{\omega}}\left|\check{\mathbb{1}}_{B_r}(\boldsymbol{\omega})\right|\,\Gamma(\mathrm{d}\boldsymbol{\omega})$$

and using (5.3) we get

$$\mathrm{cov}\,\Big(\nu_X\big(B_r(\boldsymbol{x}_i)\big),\nu_X\big(B_r(\boldsymbol{x}_j)\big)\Big) = (2\pi)r^2\int_{\mathbb{R}^2}\cos\big((\boldsymbol{x}_j-\boldsymbol{x}_i)\cdot\boldsymbol{\omega}\big)\frac{J_1^2(r\,\|\omega\|)}{\boldsymbol{\omega}^2}\,\Gamma(\mathrm{d}\boldsymbol{\omega}),$$

because the integral is a real number and Γ is a positive measure. Substitution $r\boldsymbol{\omega} = \boldsymbol{\xi}$ leads to

$$\operatorname{cov}\Big(\nu_X\big(B_r(\boldsymbol{x}_i)\big), \nu_X\big(B_r(\boldsymbol{x}_j)\big)\Big) = (2\pi) r^4 \int_{\mathbb{R}^2} \cos\left(\frac{1}{r}(\boldsymbol{x}_j - \boldsymbol{x}_i)\cdot\boldsymbol{\xi}\right) \frac{J_1^2(\|\boldsymbol{\xi}\|)}{\boldsymbol{\xi}^2}\, \Gamma\left(\mathrm{d}\frac{\boldsymbol{\xi}}{r}\right).$$

The covariance therefore depends on the behaviour of the Bartlett spectrum Γ of X in the vicinity of $\mathbf{0}$. Since $\cos\big((\boldsymbol{x}_j - \boldsymbol{x}_i)\cdot\boldsymbol{\xi}/r\big) \to 1$ as $r \to \infty$, we see that $\operatorname{cov}\big(\nu_X\big(B_r(\boldsymbol{x}_i)\big), \nu_X\big(B_r(\boldsymbol{x}_j)\big)\big)$ can be arbitrary close to $\operatorname{var}\big(\nu_X(B_r)\big)$. In particular it has the same asymptotic behaviour.

This pointed out the key problem in the estimation procedure. It is that the estimator $\hat{\operatorname{var}}_v\big(\nu_X(B_r)\big)$ may, with increasing r, return values that are in the mean much smaller than true values of $\operatorname{var}\big(\nu_X(B_r)\big)$. Particularly it may lead to different (if any) power law behaviour. Since we are primarily focused on the estimation of the asymptotic behaviour of the variance and especially on the estimation of the power law coefficient, we need an estimator that follows the same asymptotic. For this, it is perfectly sufficient to construct an estimator of the constant multiple of the true variance.

Such an estimator is $\hat{\operatorname{var}}_v\big(\nu_X(B_r)\big)$, defined by (5.5), for the proper choice of sampling points $\{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n\}$. If we choose them to be dependent on r in such a way that inter-point distances are always equal to some multiple of r, i.e. $\boldsymbol{x}_j - \boldsymbol{x}_i = \boldsymbol{a}_{i,j} r$ for some vectors $\{\boldsymbol{a}_{i,j}\}_{i\neq j}$, we get $\cos\big((\boldsymbol{x}_j - \boldsymbol{x}_i)\cdot\boldsymbol{\xi}/r\big) = \cos(\boldsymbol{a}_{i,j}\cdot\boldsymbol{\xi}) < 1$ for all $\boldsymbol{\xi}$ close to $\mathbf{0}$ and not perpendicular to $\boldsymbol{a}_{i,j}$. Let us now observe how it helps in certain cases. If Γ has an atom at $\mathbf{0}$, then there is clearly no benefit and $\hat{\operatorname{var}}_v\big(\nu_X(B_r)\big)$ still does not follow r^4 asymptotic of $\operatorname{var}\big(\nu_X(B_r)\big)$. In that case the asymptotic behaviour is be governed by other parts of Γ.

If Γ has atom at point $\boldsymbol{x} \neq \mathbf{0}$ that is not perpendicular to $\boldsymbol{a}_{i,j}$, we obtain the same formula as for the variance but with multiplicative factor $\cos(r\boldsymbol{a}_{i,j}\cdot\boldsymbol{x})$. Hence the covariance $\operatorname{cov}\big(\nu_X\big(B_r(\boldsymbol{x}_i)\big), \nu_X\big(B_r(\boldsymbol{x}_j)\big)\big)$ is again $\mathcal{O}(r)$ as $r \to \infty$. Now lets us inspect the contribution of the continuous part of Γ corresponding to the density $f_\Gamma(\boldsymbol{\omega})$. Whenever it behaves around $\mathbf{0}$ as $K\,\|\boldsymbol{\omega}\|^\sigma$ with $\sigma > -2$ and $K > 0$, we obtain

$$\operatorname{cov}\big(\nu_X\big(B_r(\boldsymbol{x}_i)\big), \nu_X\big(B_r(\boldsymbol{x}_j)\big)\big) \sim (2\pi)^2 r^{2-\sigma} K \int_0^\infty J_0\big(\,\|\boldsymbol{a}_{i,j}\|\,\xi\big)\, \frac{J_1^2(\xi)}{\xi^{1-\sigma}}\,\mathrm{d}\xi \qquad (r \to \infty),$$

because $\int_0^{2\pi} \cos(a_1\cos\phi + a_2\sin\phi)\,\mathrm{d}\phi = 2\pi J_0\big(\,\|\boldsymbol{a}\|\,\big)$ as follows from the integral representation of J_0 ([95, Eq. 9.1.18]). Since $J_0(x) < 1$ for all $x > 0$, by comparing to (5.4) we finally get that $\operatorname{cov}\big(\nu_X\big(B_r(\boldsymbol{x}_i)\big), \nu_X\big(B_r(\boldsymbol{x}_j)\big)\big)$ is for large r proportional to $r^{2-\sigma}$. Note that the range $-2 < \sigma$ covers also the long-range dependence, where $\beta = \sigma + 2$ is the power law decay coefficient.

From the previous considerations and from (5.6) follow that whenever the asymptotic behaviour of the variance $\operatorname{var}\big(\nu_X(B_r)\big)$ is governed by atomic components located outside of the origin or by a power law behaviour of the absolutely continuous part of the Bartlett spectrum Γ around the origin, the expectation of the estimator $\hat{\operatorname{var}}_v\big(\nu_X(B_r)\big)$, based on proper scale invariant sampling points $\{\boldsymbol{x}_i\}$ introduced above, is just the constant multiple of the variance, i.e.

$$\mathbb{E}\,\hat{\operatorname{var}}_v\big(\nu_X(B_r)\big) = C\operatorname{var}\big(\nu_X(B_r)\big),$$

where C is independent on r. Hence it has the same asymptotic behaviour. Moreover, in other cases the power law exponent of $\mathbb{E}\,\hat{\operatorname{var}}_v\big(\nu_X(B_r)\big)$, if the power law is presented, will be smaller or equal than the power law coefficient of the variance $\operatorname{var}\big(\nu_X(B_r)\big)$. Note that $\hat{\operatorname{var}}_v\big(\nu_X(B_r)\big)$ in this setting is not a good estimator of true values of $\operatorname{var}\big(\nu_X(B_r)\big)$, but rather of its asymptotic behaviour.

In the following we precisely construct the estimator in a rectangular window W. Let us without loss of generality assume that $W = [0, v_1] \times [0, v_2]$, where $\boldsymbol{v} = (v_1, v_2)$ is the upper right corner of W. Let further $\boldsymbol{n} = (n_1, n_2) \in \mathbb{N}^2$ and $\delta > 0$. The sample points $\{\boldsymbol{x_i}\}$ indexed

by two dimensional integer vector $\mathbf{0} \leq \boldsymbol{i} = (i_1, i_2) \leq \boldsymbol{n}$ are chosen by

$$\boldsymbol{x_i} = r\delta\boldsymbol{i} + r. \tag{5.7}$$

We thus have $n = (n_1+1)\cdot(n_2+1)$ points that satisfy the scaling property $\boldsymbol{x_i} - \boldsymbol{x_j} = r\delta(\boldsymbol{i}-\boldsymbol{j}) = r\boldsymbol{a_{i,j}}$. The particular choices of $\boldsymbol{n}$ and δ determine the maximal possible value of r accordingly to condition $B_r(\boldsymbol{x_i}) \subset W$, which is equivalent to

$$r\delta n_1 + 2r \leq v_1 \quad \text{and} \quad r\delta n_2 + 2r \leq v_2.$$

The maximal radius r_M is thus $r_M = \min\{v_1/(r\delta n_1+2r), v_2/(r\delta n_1+2r)\}$. It should be always large enough such that the behaviour of the variance approaches the asymptotic regime.

The estimator $\hat{\mathrm{var}}_v\big(\nu_X(B_r)\big)$ defined by (5.5) with previous choice of $\{\boldsymbol{x_i}\}$ clearly uses only limited information from the window W. It uses a part $W_{r;\mathbf{0}} = [0, r\delta n_1+2r]\times[0, r\delta n_2+2r] \subset W$, which can be only a small fraction of W, when r is small comparing to r_M. This can be enhanced by repeating the estimation with suitably shifted points $\{\boldsymbol{x_i}\} + \boldsymbol{y}$ for several shifts $\boldsymbol{y}$, and then take the estimate as the sample mean of these estimates. Hence we choose $m = (m_1+1)\cdot(m_2+1)$ points $\{\boldsymbol{y_k}\}_{\boldsymbol{k}=0}^{\boldsymbol{m}}$ with $\boldsymbol{m} = (m_1, m_2)$ such that shifted windows $W_{r;\boldsymbol{y_k}} = W_{r;\mathbf{0}} + \boldsymbol{y_k}$ for all $\boldsymbol{y_k}$ cover the whole W. In particular we take

$$\boldsymbol{y_k} = (b_1 \cdot k_1, b_2 \cdot k_2)$$

for a proper choice of $\boldsymbol{b} = (b_1, b_2)$ such that

$$b_1 m_1 = v_1 - r\delta n_1 - 2r \quad \text{and} \quad b_2 m_2 = v_2 - r\delta n_2 - 2r, \tag{5.8}$$

i.e. the upper right corner of the last sub window $W_{r;\boldsymbol{y_m}}$ exactly equals the upper right corner (v_1, v_2) of W.

In this way we use the information inside W more homogeneously. Note that both $\boldsymbol{m}$ and $\boldsymbol{b}$ generally depend on r. It brings no benefit to have large number m of sub windows when r is close to r_M, since then even the first sub window $W_{r;\mathbf{0}}$ uses large fraction of W. In such a case it makes sense to have small $\boldsymbol{m}$. On the other hand, for small values of r there is a clear benefit of having $\boldsymbol{n}$ large. Here, the natural choice is to take $\boldsymbol{b} = r\delta\boldsymbol{n} + 2r$ and $\boldsymbol{m}$ accordingly to (5.8).

The final estimator of the variance is

$$\hat{\mathrm{var}}_v^M\big(\nu_X(B_r)\big) = \frac{1}{(m_1+1)(m_2+1)} \sum_{\boldsymbol{k}=\mathbf{0}}^{\boldsymbol{m}} \hat{\mathrm{var}}_v\big(\nu_X(B_r); \boldsymbol{y_k}\big), \tag{5.9}$$

where $\hat{\mathrm{var}}_v\big(\nu_X(B_r); \boldsymbol{y_k}\big)$ is the estimator given by (5.5) for sample points $\{\boldsymbol{x_i} + \boldsymbol{y_k}\}_{\boldsymbol{i}=\mathbf{0}}^{\boldsymbol{n}}$, i.e.

$$\hat{\mathrm{var}}_v\big(\nu_X(B_r); \boldsymbol{y_k}\big) = \frac{1}{(n_1+1)(n_2+1)-1} \sum_{\boldsymbol{i}=\mathbf{0}}^{\boldsymbol{n}} \Big(\nu_X\big(B_r(\boldsymbol{x_i}+\boldsymbol{y_k})\big) - \overline{\nu_X\big(B_r(\boldsymbol{y_k})\big)}\Big)^2,$$

with

$$\overline{\nu_X\big(B_r(\boldsymbol{y_k})\big)} = \frac{1}{(n_1+1)(n_2+1)} \sum_{\boldsymbol{i}=\mathbf{0}}^{\boldsymbol{n}} \nu_X\big(B_r(\boldsymbol{x}_i+\boldsymbol{y_k})\big).$$

Clearly

$$\mathbb{E}\,\hat{\mathrm{var}}_v^M\big(\nu_X(B_r)\big) = \mathbb{E}\,\hat{\mathrm{var}}_v\big(\nu_X(B_r); \mathbf{0}\big).$$

Hence they share the same asymptotic and in particular the same relation to the asymptotic of $\mathrm{var}\big(\nu_X(B_r)\big)$ as discussed above.

In the rest of this part we present some numerical analysis of the proposed estimator. We use the same random closed set models as in Section 4.4, i.e. the Boolean model with $R = 10$,

$p = 0.3$ in window W_{3000} sampled with $a = 1$, and the 0-level excursion set based on the Cauchy covariance function with 6 different settings summarized in Table 4.2. We again work with the discrete realizations of those models on the grid $\{a\boldsymbol{i} + \boldsymbol{c}\}$ of points in W. Therefore we work with the discrete estimator $\hat{\mathrm{var}}_p^M\big(\nu_X(B_r)\big)$ that is defined by (5.9) with the only difference in approximation of the volume $\nu_X\big(B_r(\boldsymbol{z})\big)$ by the sum of values of the indicator $\mathbb{1}_X$ at points of the grid that are inside $B_r(\boldsymbol{z})$.

In both models the variance in balls has asymptotic power law behaviour. For the Boolean model it is

$$\mathrm{var}\,\big(\nu_X(B_r)\big) \sim Cr^2 \qquad (r \to \infty)$$

and for the 0-level excursion set with covariance function (4.40) determined by parameters α, β and θ it is

$$\mathrm{var}\,\big(\nu_X(B_r)\big) \sim Cr^{4-\beta} \qquad (r \to \infty).$$

As we have discussed before in both cases the same asymptotic behaviour holds also for the expectation $\mathbb{E}\,\hat{\mathrm{var}}_v^M\big(\nu_X(B_r)\big)$ and hence also for its discrete version $\hat{\mathrm{var}}_p^M\big(\nu_X(B_r)\big)$.

In simulations, sample points $\{\boldsymbol{x}_i\}$ were given by (5.7) with $\delta = 3$ and $\boldsymbol{n} = (4, 4)$, yielding 25 points. Every two points $\boldsymbol{x_i}, \boldsymbol{x_j}, \boldsymbol{i} \neq \boldsymbol{j}$ are separated by the distance larger than $\delta r = 3r$ and thus every two balls $B_r(\boldsymbol{x_i}), B_r(\boldsymbol{x_j})$ are separated by at least r. The maximal possible radius r_M is $r_M = \min\{v_1/14, v_2/14\}$, i.e. approximately $r_M = 210$ for W_{3000}, $r_M = 350$ for W_{5000}, and $r_M = 710$ for W_{10000}.

For each model and each setting we have generated $N = 1000$ realizations. For each realization the values of $\hat{\mathrm{var}}_v^M\big(\nu_X(B_r)\big)$ were calculated for radii r, equidistantly distributed in their logarithm from $r = 30$ to r_M with l values, where $l = 10$ for $r_M = 210$, $l = 12$ for $r_M = 350$, and $l = 15$ for $r_M = 750$.

Mean values together with 90% confidence bands, defined below equation (4.35), and true expectations obtained by numerical integration are shown in Figure 5.9 (a) for the Boolean model and in Figure 5.8 for all settings of the 0-level excursion set. Deviations from the true value are plotted in Figures 5.9 (b) and 5.10. We can see that mean values of the discrete estimator $\hat{\mathrm{var}}_p^M\big(\nu_X(B_r)\big)$ are in all cases very close to the theoretical values of $\mathbb{E}\,\hat{\mathrm{var}}_v^M\big(\nu_X(B_r)\big)$. The 90% confidence bounds are wider for larger values of r as follows from the boundedness of the observation window W. Analogously, the deviation from true values increases with increasing r. The effect of the decrease of deviations for larger windows is also visible. Note that for the Boolean model it follows from (4.38) that the covariance function is zero for $r > 2R = 20$ and hence the covariance $\mathrm{cov}\,\big(\nu_X\big(B_r(\boldsymbol{x_i})\big), \nu_X\big(B_r(\boldsymbol{x_j})\big)\big)$ is zero for all $\boldsymbol{i} \neq \boldsymbol{j}$ and all r in the analysed range $[30, r_M]$. Therefore $\mathbb{E}\,\hat{\mathrm{var}}_v^M\big(\nu_X(B_r)\big) = \mathrm{var}\,\big(\nu_X(B_r)\big)$ and the estimator $\hat{\mathrm{var}}_v^M\big(\nu_X(B_r)\big)$ is in the analysed range of r unbiased.

For the estimation of the power law exponent β we use the least squares log-linear fit according to the relation

$$\log \hat{\mathrm{var}}_p^M\big(\nu_X(B_r)\big) = (4 - \hat{\beta}_v) \log r + c. \tag{5.10}$$

In this way we obtain the estimator $\hat{\beta}_v$ of β. It is a good idea not to use all values of r in the analysed range. For small values of r the variance is still not sufficiently close to the asymptotic regime and for r close to r_M there is a poor statistic as demonstrated by wide confidence bounds in Figures 5.9 (a) and 5.8. It turns out that best performance is achieved when the values of r are restricted onto the range $r \in (46, 136)$ for W_{3000} corresponding to $r_M = 210$, or $r \in (47, 224)$ for W_{5000} corresponding to $r_M = 350$, or $r \in (47, 452)$ for W_{10000} corresponding to $r_M = 750$. An example of such a fit for two realizations is shown in Figure 5.11.

The summary of statistical results of $\hat{\beta}_v$ is given in Table 5.3. By comparing to Table 4.4 we see that in contrary to the correlation function based estimator $\hat{\beta}$ from Section 4.4, this method has tendency to underestimate the values of the exponent β. The underestimation is clearly stronger for larger values of β. The overall bias is generally much smaller for the

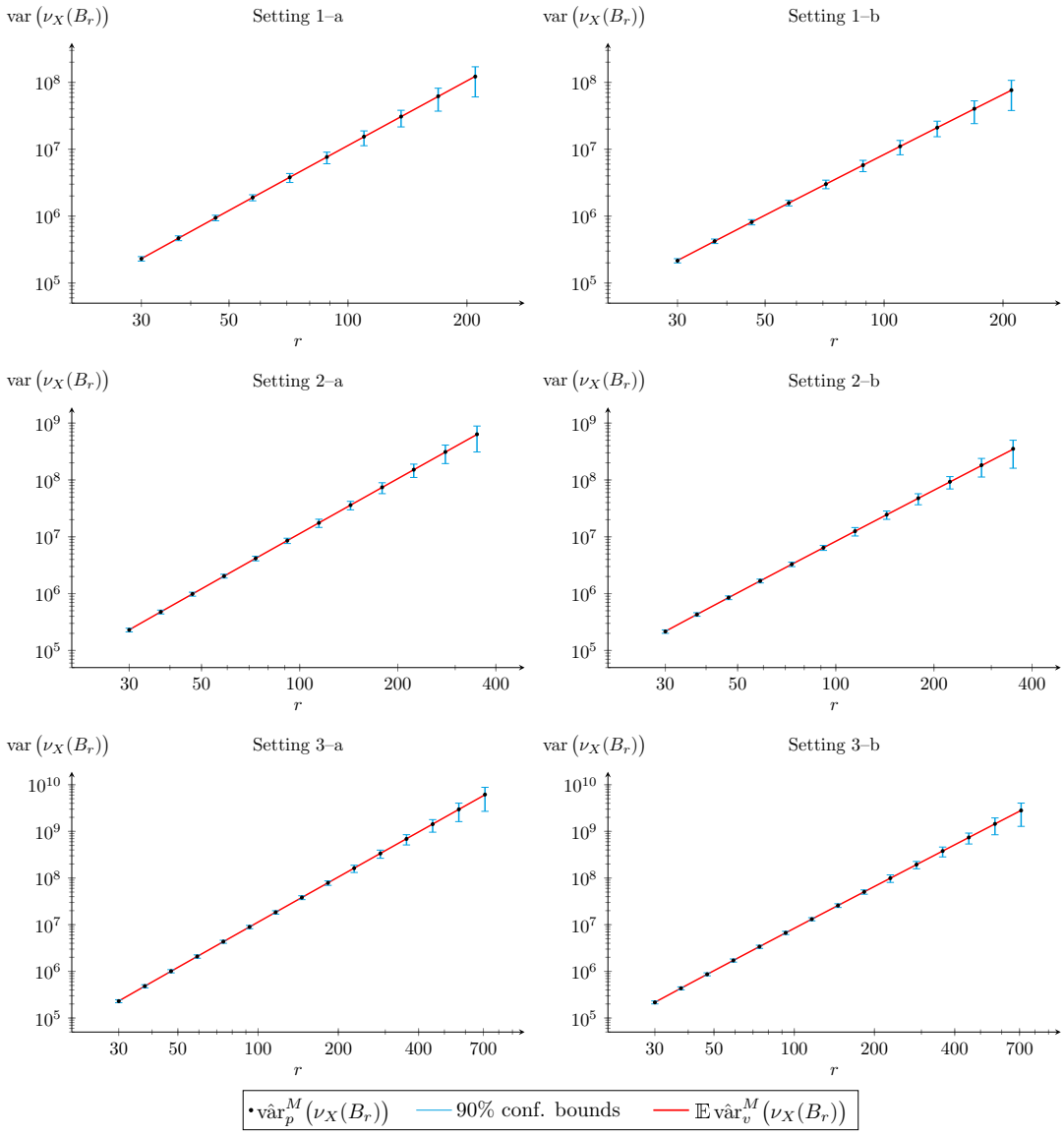


Figure 5.8: Estimated expectations and two-sided 90% confidence bounds of $\hat{\mathrm{var}}_p^M\big(\nu_X(B_r)\big)$ for the 0-level excursion set. The true value of $\mathbb{E}\,\hat{\mathrm{var}}_v^M\big(\nu_X(B_r)\big)$ is also shown.

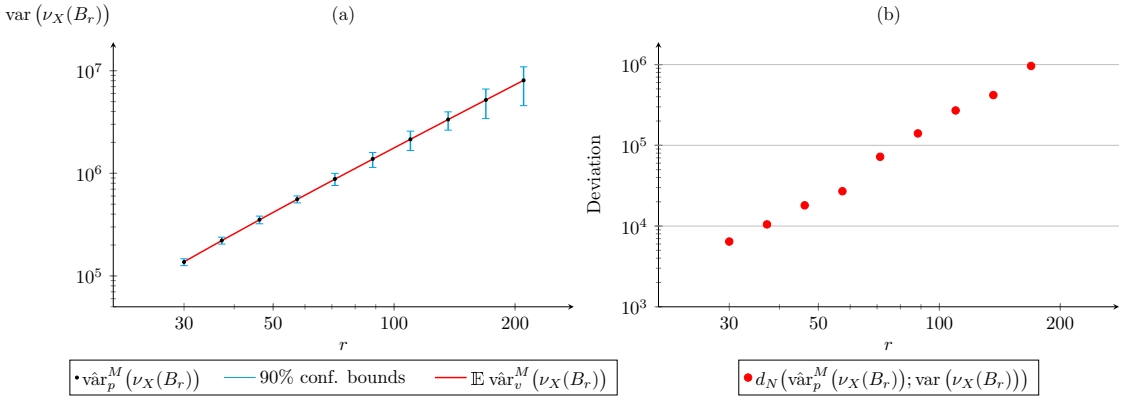


Figure 5.9: The results for the Boolean model. (a): Estimated expectations and two-sided 90% confidence bounds of $\hat{\text{var}}_p^M(\nu_X(B_r))$. The true value $\mathbb{E}\,\hat{\text{var}}_v^M(\nu_X(B_r))$ of the expectation is also shown. (b): Estimated deviations of $\hat{\text{var}}_p^M(\nu_X(B_r))$ from true values of $\text{var}(\nu_X(B_r))$.

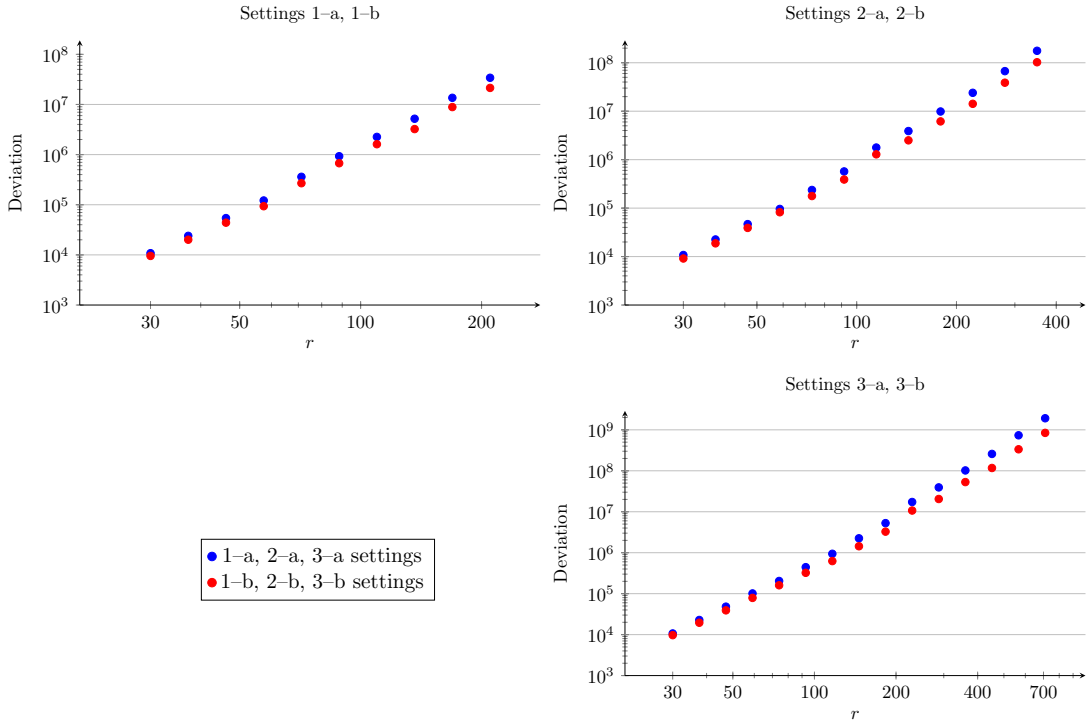


Figure 5.10: Estimated deviations of $\hat{\text{var}}_p^M(\nu_X(B_r))$ from true values of $\text{var}(\nu_X(B_r))$ for level excursion set models.

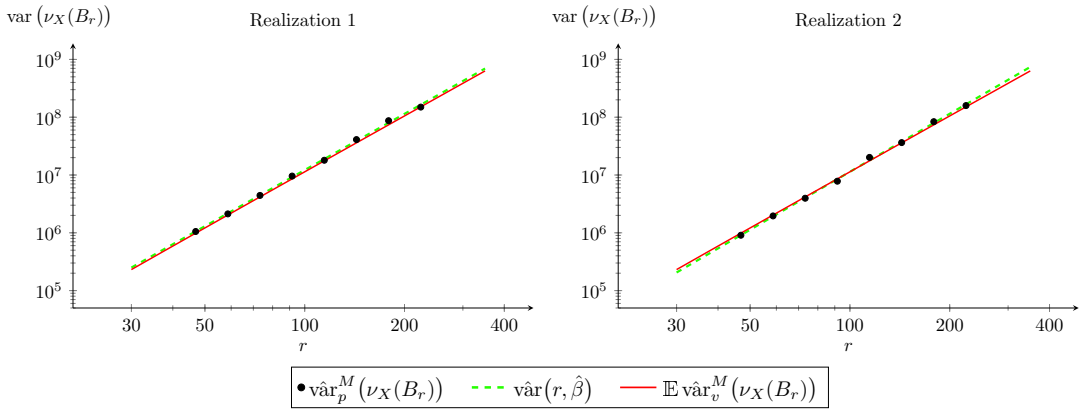


Figure 5.11: Example of the log linear fit for two realizations of the 0-level excursion set in the setting 2–a. The values of $\hat{\text{var}}_p^M\big(\nu_X(B_r)\big)$ in the narrowed range $r \in (47, 223)$ are fitted accordingly to (5.10) by $\hat{\text{var}}\big(r, \hat{\beta}\big)$. The values of the estimation of β are $\hat{\beta} = 0.77$ for realization 1 and $\hat{\beta} = 0.67$ for realization 2. The expectation $\mathbb{E}\,\hat{\text{var}}_v^M\big(\nu_X(B_r)\big)$ is superimposed in both plots.

Table 5.3: The summary of estimated properties of $\hat{\beta}$ based on the variance in balls: β is the true value of the power law decay coefficient; β^* is the value obtained by the fit of true values of $\mathbb{E}\,\hat{\text{var}}_v^M\big(\nu_X(B_r)\big)$; $\overline{\hat{\beta}}$ is the sample mean of $\hat{\beta}$ defined by (4.35); $I_{0.9}(\hat{\beta})$ the 90% confidence interval of $\hat{\beta}$; $d_N(\hat{\beta}, \beta)$ the deviation from the true value defined by (4.37). The estimator $\hat{\beta}$ is based on the log-linear fit of $\hat{\text{var}}_p^M\big(\nu_X(B_r)\big)$ for k values of $r \in J$ according to (5.10).

Setting	β	β^*	$\overline{\hat{\beta}}$	$I_{0.9}(\hat{\beta})$	$d_N(\hat{\beta}, \beta)$	k	J
Boolean model	2.00	1.92	1.93	$[1.75, 2.09]$	0.13	6	$[46, 136]$
1–a	0.80	0.78	0.79	$[0.56, 1.00]$	0.13	6	$[46, 136]$
2–a	0.80	0.78	0.79	$[0.63, 0.94]$	0.10	8	$[47, 224]$
3–a	0.80	0.79	0.80	$[0.67, 0.91]$	0.07	11	$[47, 452]$
1–b	1.05	0.99	1.00	$[0.78, 1.21]$	0.14	6	$[46, 136]$
2–b	1.05	1.00	1.01	$[0.86, 1.15]$	0.10	8	$[47, 224]$
3–b	1.05	1.01	1.02	$[0.91, 1.13]$	0.07	11	$[47, 452]$

variance based estimator. This holds especially for strong correlations represented by settings 1–a, 2–a, and 3–a, where it has also smaller deviation. On the other hand, the variance based estimator is generally less stable since it has similar or larger deviations and much smaller bias that contributes significantly to deviations of the correlation function based estimator. This can be also observed by the fact that confidence intervals of $\hat{\beta}_v$ are wider then confidence bounds of $\hat{\beta}$. In contrary to $\hat{\beta}$ we see that deviations and the width of the confidence interval of $\hat{\beta}_v$ are independent of the concrete model (setting) and depend only on the window size.

Finally, we show the result of the variance estimation for the model of a randomly shifted chessboard. That is a random closed set $X = X_s + \boldsymbol{x}$, where X_s is a fixed set defined by $\mathbb{1}_{X_s}(\boldsymbol{x}) = \text{sgn}\big(\sin(x_1\pi/s)\sin(x_2\pi/s)\big)$, with $s > 0$ being a side of the small rectangle, and $\boldsymbol{x}$ is a random variable in $[0, 2s]^2$ with uniform distribution. The random closed set X is clearly periodic. Hence from the considerations above the variance var $\big(\nu_X(B_r)\big)$ is asymptotically of order $\mathcal{O}(r)$. In Figure 5.12 we show the basic realization for $\boldsymbol{x} = 0$ of the chessboard model of side $s = 100$ and estimation of the corresponding variances in balls. We see that there is no

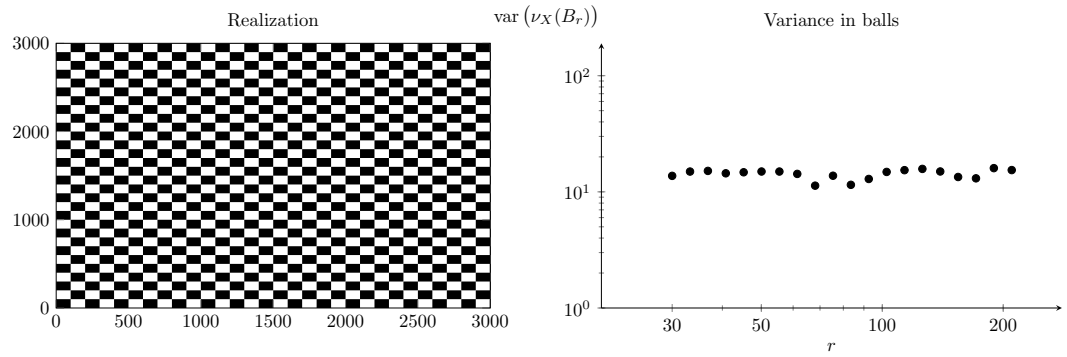


Figure 5.12: The base realization and result of the variance in balls estimation given by $\hat{\mathrm{var}}_p^M\big(\nu_X(B_r)\big)$ for the chessboard model with small rectangle side $s = 100$ in the window W_{3000}.

clear power law behaviour of the variance. In particular, there is no increase of order larger than $\mathcal{O}(r)$.

5.3.3 Variance in balls: results

For the analysis of the built-up area represented by a stationary random closed set X in an observation window W we perform the estimation of the variance using the estimator $\hat{\mathrm{var}}_v^M\big(\nu_X(B_r)\big)$ defined by (5.9). For all cities we take sample points $\{\boldsymbol{x}_i\}$ given by (5.7) with $\delta = 3$ and $\boldsymbol{n} = (4, 4)$. Hence we have 25 points and all balls of radius r are separated by at least $3r$. We use only balls that are fully included in W, i.e. $B_r(\boldsymbol{x}_i + \boldsymbol{y}_{\boldsymbol{k}}) \subset W$. The maximal possible radius r_M is determined by the window size. If $D = [u_1, u_2] \times [v_1, v_2]$ is the rectangular bounding box of the window W, then $r_M = \min\{(v_1 - u_1)/14, (v_2 - u_2)/14\}$. For each city we calculate the values of $\hat{\mathrm{var}}_v^M\big(\nu_X(B_r)\big)$ for k different radii r that are equidistantly distributed in their logarithm in a certain interval I such that smallest values are around $r = 48\,\mathrm{m}$ and largest up to $r = 332\,\mathrm{m}$.

The results of the estimator $\hat{\mathrm{var}}_v^M\big(\nu_X(B_r)\big)$ and the corresponding log-linear fit according to equation

$$\log \hat{\mathrm{var}}_v^M\big(\nu_X(B_r)\big) = (4 - \hat{\beta}_v) \log r + c \tag{5.11}$$

are plotted in Figure 5.13. The estimated values $\hat{\beta}_v$ of the power law coefficient β together with other parameters of the estimation are summarized in Table 5.4. We can see that values of $\hat{\beta}_v$ are generally smaller than values of $\hat{\beta}$ from Table 5.2. The only exceptions are Paris and Minsk, where $\hat{\beta}_v$ are larger then $\hat{\beta}$. For Minsk, Moscow, and St. Petersburgh the values are similar. The largest differences are for Oslo, Pittsburgh, and Seattle. In Oslo, Seattle, and also in Milan, and Prague, however, the observation window W is small and thus also the range I of values of r for which we estimate the variance in balls is narrow. Hence the results for those cities are highly unreliable.

Finally let us discuss the results for Chicago, Los Angeles, and Seattle that are of gridiron structure as discussed on page 111. We have seen that the correlation function $\kappa(\boldsymbol{r})$ partially follows a regular structure which is reflected by small oscillations of the isotropic adapted estimator of the correlation function. The question is whether the structure is not actually periodic as discussed in the previous subsection. In such a case the variance in balls should not follow a power law and moreover it should be of asymptotic order not exceeding r as $r \to \infty$. On the example in Figure 5.12 we have seen that the estimator $\hat{\mathrm{var}}_v^M\big(\nu_X(B_r)\big)$ is capable of recognizing this type of behaviour.

Table 5.4: The results of the estimation of the power law dependence coefficient β of the variance in balls $\operatorname{var}\big(\nu_X(B_r)\big)$ for all analysed cities. The estimator $\hat{\beta}_v$ is based on the log-linear fit of k values of $\hat{\operatorname{var}}_v^M\big(\nu_X(B_r)\big)$ for $r \in I$ according to (5.11). The values denoted by * are unreliable since they are determined from a range of r with length only around 100 m.

City	$\hat{\beta}_v$	k	I [m]
Boston	0.85	13	$[51, 260]$
Chicago	1.11	15	$[51, 320]$
Los Angeles	1.07	15	$[51, 332]$
Pittsburgh	0.61	12	$[50, 210]$
Seattle	0.40^*	11	$[44, 155]$
Berlin	0.67	15	$[51, 320]$
Birmingham	0.56	12	$[49, 190]$
Milan	0.70^*	11	$[48, 150]$
Minsk	1.16	13	$[45, 220]$
Moscow	1.02	11	$[50, 184]$
Oslo	0.28^*	10	$[45, 110]$
Paris	1.05	12	$[51, 215]$
Prague	0.42^*	11	$[47, 140]$
St. Petersburg	0.95	12	$[48, 170]$

In Figure 5.13 the power law behaviour is clearly visible in the whole analysed range. Moreover as follows from Table 5.4 the exponent β of the power law decay is far smaller then $\beta = 3$ that should hold (as the upper bound) in the periodic situation and it also does not correspond to super-homogeneous situations that holds for $\beta > 2$. Hence we may conclude that the gridiron structure produces only small perturbations of the correlation function and the overall behaviour on large distances is fully dominated by the power law decay.

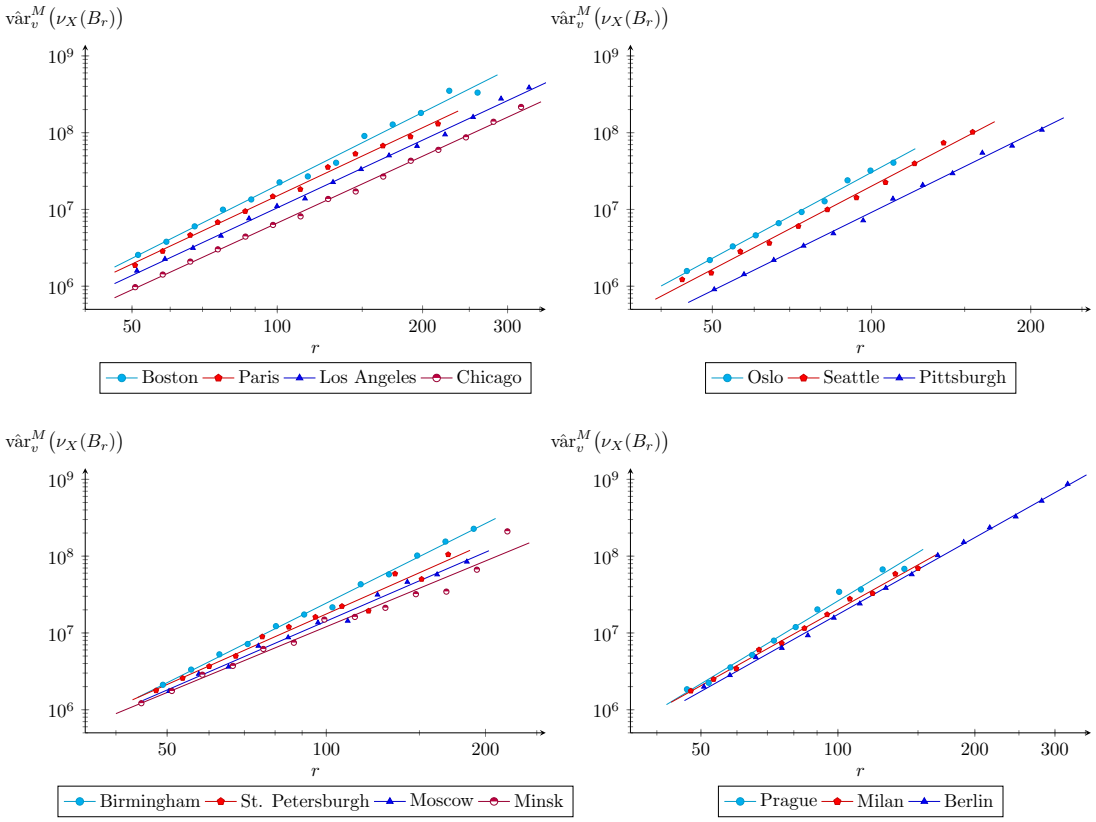


Figure 5.13: The log-linear fits according to (5.11) of the power law behaviour of the variance in balls estimator $\hat{\mathrm{var}}_v^M\big(\nu_X(B_r)\big)$ for analysed cities.

5.4 Built-up area as a non-stationary random closed set

In the previous section the built-up area in a city centre was analysed when taken as a part of a stationary random closed set. In this section the stationarity assumption is removed and we analyse the built-up area using non-stationary methods introduced in Section 4.3. Hence we take the built-up area in a particular city as a realization of a random closed set X in $\mathbb{R}^2$. Without stationarity there is no ergodicity, however, we still assume that the estimation from one sample makes sense. Moreover, we always assume P-continuity of X which implies continuity of the (non-constant) volume fraction and covariance as was discussed in Section 2.3.

The random closed set X is observed in a sample window $W \subset \mathbb{R}^2$ which is, in opposite to the previous section, chosen as large as possible in order to study the city as a whole. In practice we obtain W by choosing a rectangular region D and subtracting water areas inside it. The choice of D is usually limited by the availability of the data. In the USA cities building footprints are available only in a certain range around the city. For European cities the most determining factor is the quality of the `Openstreetmap` data. This means that for many cities the building footprints are usually available only in the close vicinity of the city centre. In more remote areas the building information is incomplete or fully missing. The GPS coordinates of rectangular parts of chosen windows are summarized in Table 5.5.

Having the window W the observed part of a random closed set X in W can be obtained

Table 5.5: The selections of the rectangular region D for analysed cities in the non-stationary approach. The coordinates are given in the GPS format.

City	Lower left corner Latitude	Lower left corner Longitude	Upper right corner Latitude	Upper right corner Longitude	Size of D [m]
Boston	N 42° 16.615′	W 71° 10.273′	N 42° 24.147′	W 71° 01.488′	12000 × 14000
Chicago	N 41° 46.508′	W 87° 46.480′	N 41° 59.741′	W 87° 36.479′	13650 × 24600
Los Angeles	N 33° 46.058′	W 118° 27.791′	N 34° 07.205′	W 118° 05.860′	33900 × 39000
Pittsburgh	N 40° 21.470′	W 80° 04.714′	N 40° 30.371′	W 79° 51.567′	19000 × 16000
Seattle	N 47° 29.149′	W 122° 25.018′	N 47° 46.537′	W 122° 16.340′	11500 × 32000
Berlin	N 52° 27.223′	E 13° 16.796′	N 52° 33.558′	E 13° 28.484′	13000 × 12000
Birmingham	N 52° 23.978′	W 1° 56.913′	N 52° 33.673′	W 1° 45.129′	13000 × 18000
Milan	N 45° 26.436′	E 9° 07.671′	N 45° 30.745′	E 9° 14.594′	9000 × 8000
Minsk	N 53° 50.181′	E 27° 27.068′	N 53° 56.474′	E 27° 41.390′	15600 × 11800
Moscow	N 55° 38.543′	E 37° 24.710′	N 55° 52.283′	E 37° 48.114′	25000 × 25000
Oslo	N 59° 53.123′	E 10° 41.982′	N 59° 57.430′	E 10° 49.604′	7100 × 8000
Paris	N 48° 44.839′	E 2° 11.038′	N 48° 57.903′	E 2° 32.139′	26000 × 24000
Prague	N 50° 00.569′	E 14° 21.118′	N 50° 08.546′	E 14° 31.262′	14000 × 13000
St. Petersburg	N 59° 49.925′	E 30° 12.838′	N 60° 03.072′	E 30° 30.154′	16000 × 24500

as

$$X \cap W = \bigcup_{i=1}^{n} (P_i \cap W),$$

where $\{P_i\}_{i=1}^n$ are the building footprints in a given city. First task in the non-stationary analysis is to estimate the volume fraction. Since we do not assume any specific parametrizable result we use the general kernel based non-parametric estimation method. Hence we use the estimator (4.25),

$$\hat{m}_h(\boldsymbol{x}) = \frac{1}{e_h(\boldsymbol{x})} \int_W \mathbb{1}_X(\boldsymbol{u}) K_h(\boldsymbol{x} - \boldsymbol{u}) \, \mathrm{d}\boldsymbol{u},$$

where the kernel function K_h taken in the analysis is the bivariate Gaussian kernel

$$K_h(\boldsymbol{x}) = \frac{1}{2\pi h^2} e^{-\frac{\boldsymbol{x}^2}{2h^2}}$$

and

$$e_h(\boldsymbol{x}) = \int_W K_h(\boldsymbol{x} - \boldsymbol{u}) \, \mathrm{d}\boldsymbol{u}.$$

Since all polygons are disjoint, it follows

$$\hat{m}_h(\boldsymbol{x}) = \frac{1}{e_h(\boldsymbol{x})} \sum_{i=1}^{n} \int_{P_i \cap W} K_h(\boldsymbol{x} - \boldsymbol{u}) \, \mathrm{d}\boldsymbol{u}. \tag{5.12}$$

Thus all items in the previous relation are given by the convolution of the Gaussian kernel with some polygon. In the following subsection we develop an approximate method how to determine those integrals in practice.

5.4.1 Kernel estimate approximation for polygons

The main task of this part is to calculate the integral

$$\int_P K_h(\boldsymbol{y} - \boldsymbol{x}) \, \mathrm{d}\boldsymbol{x}, \tag{5.13}$$

where P is a polygon. For the Gaussian kernel the closed formula, however, cannot be found. We may only found an approximate relation. Let us first consider the case when the polygon is just a rectangle B. The kernel can be expressed in a special form

$$K_h(\boldsymbol{y}-\boldsymbol{x}) = \frac{1}{2\pi h^2} e^{-\frac{(\boldsymbol{x}-\boldsymbol{y})^2}{2h^2}} = \frac{1}{\sqrt{2\pi}h} e^{-\frac{(x_1-y_1)^2}{2h^2}} \frac{1}{\sqrt{2\pi}h} e^{-\frac{(x_2-y_2)^2}{2h^2}} = \phi_h(x_1-y_1)\phi_h(x_2-y_2),$$

where $\phi_h(x)$ denotes the one-dimensional probability distribution function of the normal distribution $N(0,h)$ with zero mean and variance h^2. Let

$$\Phi_h(x) = \frac{1}{\sqrt{2\pi}h} \int\limits_{-\infty}^{x} e^{-\frac{t^2}{2h^2}}\,\mathrm{d}t \tag{5.14}$$

be the cumulative distribution function of $N(0,h)$. The integral of $K_h(\boldsymbol{y}-\boldsymbol{x})$ over the rectangle B is therefore given by

$$\begin{aligned}\int\limits_B K_h(\boldsymbol{y}-\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} &= \int\limits_B K_h(\boldsymbol{x}-\boldsymbol{y})\,\mathrm{d}\boldsymbol{x} = \int\limits_{B+\boldsymbol{y}} K_h(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} = \\ &= \int\limits_{v_2+y_2}^{u_2+y_2}\int\limits_{v_1+y_1}^{u_1+y_1} \phi_h(x_1)\phi_h(x_2)\,\mathrm{d}x_1\,\mathrm{d}x_2 = \\ &= \big(\Phi_h(u_1+y_1)-\Phi_h(v_1+y_1)\big)\big(\Phi_h(u_2+y_2)-\Phi_h(v_2+y_2)\big), \end{aligned} \tag{5.15}$$

where $\boldsymbol{v}=(v_1,v_2)$ is the lower left corner of B and $\boldsymbol{u}=(u_1,u_2)$ is the upper right corner of B.

When P is not a rectangle one may approximate the value of the integral by the value calculated for its bounding box B_P multiplied by the ratio of the polygon area and the area of the bounding box. If we denote the lower left corner of B_P by $\boldsymbol{v}_P=(v_{P;1},v_{P;2})$ and the upper right corner by $\boldsymbol{u}_P=(u_{P;1},u_{P;2})$ we can approximately take

$$\begin{aligned}&\int\limits_P K_h(\boldsymbol{y}-\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} \approx \frac{\nu(P)}{\nu(B_P)} \int\limits_{B_P} K_h(\boldsymbol{y}-\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} = \\ &= \frac{\nu(P)}{(u_{P;1}-v_{P;1})(u_{P;2}-v_{P;2})}\big(\Phi_h(u_{P;1}+y_1)-\Phi_h(v_{P;1}+y_1)\big)\big(\Phi_h(u_{P;2}+y_2)-\Phi_h(v_{P;2}+y_2)\big),\end{aligned}$$

where $\nu(P)\equiv\nu_2(P)$ is the area of P.

In order to estimate the error of such approximation we prove the following assertion.

Proposition 5.4.1. *Let P be a polygon and B be a rectangle, both in $\mathbb{R}^2$, such that $P\subset B$. Let further K_h be the Gaussian kernel, given for all $\boldsymbol{x}\in\mathbb{R}^2$ by*

$$K_h(\boldsymbol{x}) = \frac{1}{2\pi h^2} e^{-\frac{\boldsymbol{x}^2}{2h^2}}.$$

Then for all $\boldsymbol{y}\in\mathbb{R}^2$

$$\left|\int\limits_P K_h(\boldsymbol{y}-\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} - \frac{\nu(P)}{\nu(B)}\int\limits_B K_h(\boldsymbol{y}-\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}\right| \le p(1-p)\frac{\nu(B)}{2\pi h^2}\min\left\{1,\frac{\operatorname{diam}(B)}{2h}\right\},$$

where $\operatorname{diam}(B)=\max_{\boldsymbol{x},\boldsymbol{y}\in B}\|\boldsymbol{x}-\boldsymbol{y}\|$ *is the largest distance of two points inside B and* $p=\frac{\nu(P)}{\nu(B)}$.

Proof. Let $\boldsymbol{y} \in \mathbb{R}^2$ be fixed and let $m_{\boldsymbol{y}} = \inf_{\boldsymbol{x}\in B} K_h(\boldsymbol{y}-\boldsymbol{x})$ and $M_{\boldsymbol{y}} = \inf_{\boldsymbol{x}\in B} K_h(\boldsymbol{y}-\boldsymbol{x})$. Then we have

$$m_{\boldsymbol{y}}\nu(P) = m_{\boldsymbol{y}} \int\limits_B \mathbb{1}_P(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} \le \int\limits_B K_h(\boldsymbol{y}-\boldsymbol{x})\mathbb{1}_P(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} \le M_{\boldsymbol{y}} \int\limits_B \mathbb{1}_P(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} = M_{\boldsymbol{y}}\nu(P).$$

The rectangle B is a convex compact set and $K_h(\boldsymbol{y}-\boldsymbol{x})$ is continuous in $\boldsymbol{x}$. Therefore $K_h(\boldsymbol{y}-\boldsymbol{x})$ as a function of $\boldsymbol{x}$ attains every value of the interval $[m_{\boldsymbol{y}}, M_{\boldsymbol{y}}]$ and in particular there exists $\boldsymbol{u}_{\boldsymbol{y}} \in B$ such that

$$K_h(\boldsymbol{y}-\boldsymbol{u}_{\boldsymbol{y}})\nu(P) = \int\limits_B K_h(\boldsymbol{y}-\boldsymbol{x})\mathbb{1}_P(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} = \int\limits_P K_h(\boldsymbol{y}-\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}.$$

Analogously there is $\boldsymbol{v}_{\boldsymbol{y}} \in B$ such that

$$K_h(\boldsymbol{y}-\boldsymbol{v}_{\boldsymbol{y}})\nu(B\setminus P) = \int\limits_{B\setminus P} K_h(\boldsymbol{y}-\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}.$$

Using $\nu(B\setminus P) = \nu(B) - \nu(P)$ we have

$$\begin{aligned}
&\left|\int\limits_P K_h(\boldsymbol{y}-\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} - \frac{\nu(P)}{\nu(B)}\int\limits_B K_h(\boldsymbol{y}-\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}\right| = \\
&\qquad = \left|\int\limits_P K_h(\boldsymbol{y}-\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} - \frac{\nu(P)}{\nu(B)}\left(\int\limits_P K_h(\boldsymbol{y}-\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} + \int\limits_{B\setminus P} K_h(\boldsymbol{y}-\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}\right)\right| = \\
&\quad = \left|\nu(P)K_h(\boldsymbol{y}-\boldsymbol{u}_{\boldsymbol{y}}) - \frac{\nu(P)}{\nu(B)}\Big(\nu(P)K_h(\boldsymbol{y}-\boldsymbol{u}_{\boldsymbol{y}}) + \big(\nu(B)-\nu(P)\big)K_h(\boldsymbol{y}-\boldsymbol{v}_{\boldsymbol{y}})\Big)\right| = \\
&\qquad\qquad = p(1-p)\nu(B)\left|K_h(\boldsymbol{y}-\boldsymbol{u}_{\boldsymbol{y}}) - K_h(\boldsymbol{y}-\boldsymbol{v}_{\boldsymbol{y}})\right|,
\end{aligned}$$

where $p = \frac{\nu(P)}{\nu(B)} \le 1$. Introducing new variable $\boldsymbol{x}_{\boldsymbol{y}} = \boldsymbol{y} - \frac{\boldsymbol{u}_{\boldsymbol{y}}+\boldsymbol{v}_{\boldsymbol{y}}}{2}$ gives us

$$|K_h(\boldsymbol{y}-\boldsymbol{u}_{\boldsymbol{y}}) - K_h(\boldsymbol{y}-\boldsymbol{v}_{\boldsymbol{y}})| = \left|K_h\left(\boldsymbol{x}_{\boldsymbol{y}} - \frac{\boldsymbol{u}_{\boldsymbol{y}}-\boldsymbol{v}_{\boldsymbol{y}}}{2}\right) - K_h\left(\boldsymbol{x}_{\boldsymbol{y}} + \frac{\boldsymbol{u}_{\boldsymbol{y}}-\boldsymbol{v}_{\boldsymbol{y}}}{2}\right)\right|.$$

Using the explicit formula for K_h one obtains

$$|K_h(\boldsymbol{y}-\boldsymbol{u}_{\boldsymbol{y}}) - K_h(\boldsymbol{y}-\boldsymbol{v}_{\boldsymbol{y}})| = \frac{1}{2\pi h^2}\left|e^{-\frac{1}{2h^2}\left(\boldsymbol{x}_{\boldsymbol{y}} - \frac{\boldsymbol{u}_{\boldsymbol{y}}-\boldsymbol{v}_{\boldsymbol{y}}}{2}\right)^2} - e^{-\frac{1}{2h^2}\left(\boldsymbol{x}_{\boldsymbol{y}} + \frac{\boldsymbol{u}_{\boldsymbol{y}}-\boldsymbol{v}_{\boldsymbol{y}}}{2}\right)^2}\right|. \tag{5.16}$$

Let us now inspect the function $f(\boldsymbol{x}) = \left|e^{-(\boldsymbol{x}-\boldsymbol{a})^2} - e^{-(\boldsymbol{x}+\boldsymbol{a})^2}\right|$. It is clearly even in the sense that $f(\boldsymbol{x}) = f(-\boldsymbol{x})$ for all $\boldsymbol{x} \in \mathbb{R}^2$. With the decomposition $\boldsymbol{x} = \boldsymbol{x}_\perp + \boldsymbol{x}_\parallel$, where $\boldsymbol{x}_\perp$ is perpendicular to $\boldsymbol{a}$ and $\boldsymbol{x}_\parallel$ is parallel to $\boldsymbol{a}$, we have

$$f(\boldsymbol{x}) = \left|e^{-(\boldsymbol{x}_\perp+\boldsymbol{x}_\parallel-\boldsymbol{a})^2} - e^{-(\boldsymbol{x}_\perp+\boldsymbol{x}_\parallel+\boldsymbol{a})^2}\right| = e^{-\boldsymbol{x}_\perp^2}\left|e^{-(\boldsymbol{x}_\parallel-\boldsymbol{a})^2} - e^{-(\boldsymbol{x}_\parallel+\boldsymbol{a})^2}\right| \le f(\boldsymbol{x}_\parallel).$$

The maximum of f is therefore always at $\boldsymbol{x}$ located on the line given by the vector $\boldsymbol{a}$. Thus we can without loss of generality restrict ourselves to one dimensional case and study the maximum of $f(x) = e^{-(x-a)^2} - e^{-(x+a)^2}$ where $a > 0$ and $x > 0$. Taking the derivative and setting it to 0,

$$f'(x) = -2(x-a)e^{-(x-a)^2} + (x+a)e^{-(x+a)^2} = 2e^{-x^2-a^2}\left((x+a)e^{-xa} - (x-a)e^{xa}\right) = 0,$$

leads to equation

$$1 - \frac{2a}{x+a} = e^{-4xa}. \tag{5.17}$$

The left side is positive only when $x > a$, so for the solution $x_{\max}$ we have $x_{\max} > a$. The expansion of the right side gives us

$$1 - \frac{2a}{x+a} = 1 - 4xa + \sum_{k=2}^{\infty} \frac{(-4xa)^k}{k!}.$$

In the limit $a \to 0$ we thus have $x_{\max} = \frac{1}{\sqrt{2}}$. Now we prove that for all $a > 0$ holds $x_{\max} > \frac{1}{\sqrt{2}}$. In order to see this it is useful to rewrite equation (5.17) with $x = \frac{u}{\sqrt{2}} + \frac{1}{\sqrt{2}}$:

$$1 - \frac{2\sqrt{2}a}{1+u+\sqrt{2}a} - e^{-2\sqrt{2}a(1+u)} = 0.$$

Note that the denominator of the left side is positive since we know that in the neighbourhood of the solution is $x > a$. The function on the left side is obviously increasing in u. By the standard methods of analysis it is easy to show that for $u = 0$ and all $a > 0$,

$$1 - \frac{2\sqrt{2}a}{1+\sqrt{2}a} - e^{-2\sqrt{2}a} < 0.$$

On the other side, the limit of the left side equals 1 when $u \to \infty$. The solution is therefore positive, $u_{\max} > 0$, which proves the fact that $x_{\max} > \frac{1}{\sqrt{2}}$.

Let us now inspect the value of $f(x_{\max})$:

$$f(x_{\max}) = e^{-(x_{\max}-a)^2} - e^{-(x_{\max}+a)^2} = e^{-(x_{\max}-a)^2}\left(1 - e^{-4x_{\max}a}\right).$$

With the use of the relation (5.17) we obtain

$$f(x_{\max}) = e^{-(x_{\max}-a)^2}\left(\frac{2a}{x_{\max}+a}\right) \leq \frac{2a}{x_{\max}+a}.$$

Since we know that $x_{\max} > a$ and $x_{\max} > \frac{1}{\sqrt{2}}$, we have

$$\frac{2a}{x_{\max}+a} < \frac{2a}{a+a} = 1$$

and for $a < \frac{1}{\sqrt{2}}$ also

$$\frac{2a}{x_{\max}+a} < \frac{2a}{\frac{1}{\sqrt{2}}+a} < \sqrt{2}a.$$

This can be finally written together as

$$f(x_{\max}) < \min\{1, \sqrt{2}a\}.$$

Now we can return back to equation (5.16). The maximal error can be bounded from above by removing the dependence of $\boldsymbol{u}, \boldsymbol{v}$, and $\boldsymbol{x}$ on $\boldsymbol{y}$

$$\sup_{\boldsymbol{y}\in\mathbb{R}^2} \left| e^{-\frac{1}{2h^2}\left(\boldsymbol{x}_{\boldsymbol{y}} - \frac{\boldsymbol{u}_{\boldsymbol{y}}-\boldsymbol{v}_{\boldsymbol{y}}}{2}\right)^2} - e^{-\frac{1}{2h^2}\left(\boldsymbol{x}_{\boldsymbol{y}} + \frac{\boldsymbol{u}_{\boldsymbol{y}}-\boldsymbol{v}_{\boldsymbol{y}}}{2}\right)^2} \right| \leq \sup_{\boldsymbol{x}\in\mathbb{R}^2, \boldsymbol{u},\boldsymbol{v}\in B} \left| e^{-\frac{1}{2h^2}\left(\boldsymbol{x} - \frac{\boldsymbol{u}-\boldsymbol{v}}{2}\right)^2} - e^{-\frac{1}{2h^2}\left(\boldsymbol{x} + \frac{\boldsymbol{u}-\boldsymbol{v}}{2}\right)^2} \right|.$$

From the previous analysis we know that the maximal error is when $\left\|\frac{\boldsymbol{u}-\boldsymbol{v}}{2}\right\|$ is maximal. The maximum is at $\boldsymbol{x}_{\max}$, which is the multiplication of the vector $\frac{\boldsymbol{u}-\boldsymbol{v}}{2}$ with the lower bound of its norm, $\|\boldsymbol{x}_{\max}\| > \left\|\frac{\boldsymbol{u}-\boldsymbol{v}}{2}\right\|$. Since $\boldsymbol{u}, \boldsymbol{v}$ are points from B the norm $\left\|\frac{\boldsymbol{u}-\boldsymbol{v}}{2}\right\|$ in its maximal value equals $\frac{\operatorname{diam}(B)}{2}$. Using the previous estimate of $f(x_{\max})$ with $a = \frac{\operatorname{diam}(B)}{2\sqrt{2}h}$ yields the final assertion. □

It is obvious that the error is smaller for smaller polygons. If the bounding box B of a polygon P has sides $a, b < h$ than the error

$$D(P,B) = \left| \int_P K_h(\boldsymbol{y}-\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} - \frac{\nu(P)}{\nu(B)} \int_B K_h(\boldsymbol{y}-\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} \right|$$

is, as a consequence of the previous proposition, bounded by

$$D(P,B) \le p(1-p)\frac{ab\sqrt{a^2+b^2}}{4\pi h^3} \le \frac{ab\sqrt{a^2+b^2}}{16\pi h^3}$$

since $p(1-p) \le \frac{1}{4}$. Let divide the bounding box B of P into $n \times n$ equal pieces $B_1, \dots, B_{n^2}$. Clearly all B_i have sides $\frac{a}{n}, \frac{b}{n}$. Let further $P_i = B_i \cap P$ be the part of P in the i-th small rectangle B_i. The error in estimating the convolution for P_i is bounded by

$$D(P_i,B_i) \le p_i(1-p_i)\frac{ab\sqrt{a^2+b^2}}{4\pi h^3 n^3} \le \frac{ab\sqrt{a^2+b^2}}{16\pi h^3 n^3}.$$

In the overall sum it leads to

$$D(P,B) \le \sum_{i=1}^{n^2} D(P_i,B_i) \le \sum_{i=1}^{n^2} p_i(1-p_i)\frac{ab\sqrt{a^2+b^2}}{4\pi h^3 n^3} \le \frac{ab\sqrt{a^2+b^2}}{16\pi h^3 n} = \mathcal{O}(n^{-1})$$

as $n \to \infty$. If n is large then many of small rectangles B_i are either fully contained in P ($p_i = 1$) or contain zero volume from P ($p_i = 0$). Thus in the limit $n \to \infty$ the number of small rectangles with $p_i(1-p_i) > 0$ is proportional to n with a constant L. Hence

$$\sum_{i=1}^{n^2} p_i(1-p_i)\frac{ab\sqrt{a^2+b^2}}{4\pi h^3 n^3} \approx \sum_{i=1}^{Ln} p_{k_i}(1-p_{k_i})\frac{ab\sqrt{a^2+b^2}}{4\pi h^3 n^3} \le L\frac{ab\sqrt{a^2+b^2}}{16\pi h^3 n^2} = \mathcal{O}(n^{-2})$$

as $n \to \infty$. In other words, the overall error is of order not exceeding n^{-2}. From this follow that it is reasonable to approximate the correct value of (5.13) by setting n big enough and using the formula

$$\int_P K_h(\boldsymbol{y}-\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} \approx \sum_{i=1}^{n^2} \frac{\nu(P_i)}{\nu(B_i)} \int_{B_i} K_h(\boldsymbol{y}-\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}, \tag{5.18}$$

where B_i is the i-th small rectangle given by the division of the bounding box B into $n \times n$ pieces and $P_i = P \cap B_i$. In figure 5.14 an example of the dependence of the error on n is shown.

In the numerical analysis of the built-up structure we chose n for each building footprint P individually in such a way that $\frac{a}{n}, \frac{b}{n} < 5\,\mathrm{m}$, where a, b are the sides of the bounding box B of P.

5.4.2 Choosing the bandwidth and estimating the correlation function

The estimation result of the kernel estimator $\hat{m}_h(\boldsymbol{x})$ depends strongly on the choice of bandwidth h. As was discussed in Subsection 4.3.1, choosing the large bandwidth may introduce a large bias whereas too small bandwidth may cause a large variance of $\hat{m}_h(\boldsymbol{x})$. Thus as the best theoretical choice the value minimizing the mean integrated square error (MISE)

$$\mathrm{MISE}\big(\hat{m}_h\big) = \int_W \big(\,\mathbb{E}\,\hat{m}_h(\boldsymbol{x}) - m(\boldsymbol{x})\big)^2\,\mathrm{d}\boldsymbol{x} + \int_W \operatorname{var}\hat{m}_h(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}$$

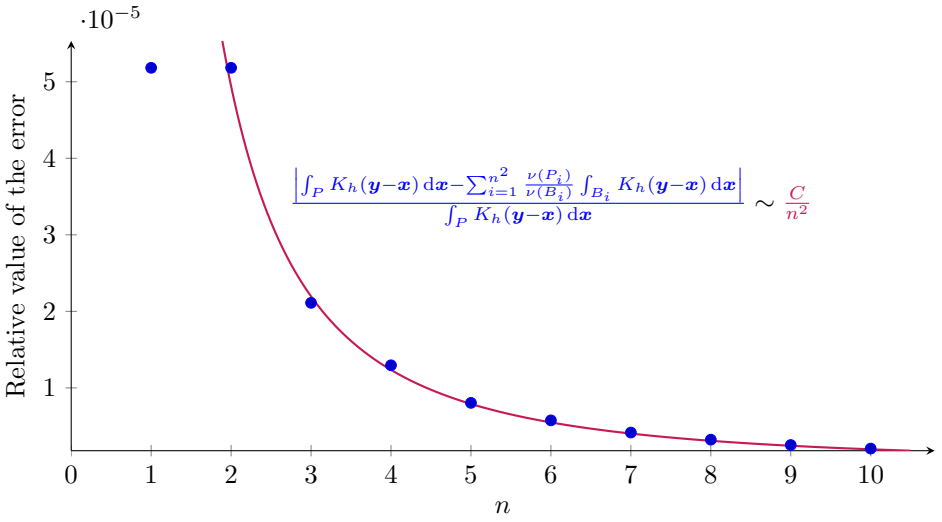


Figure 5.14: An example of the decrease of a relative error given by the ratio between the absolute value of the error and the correct value. The notation is the same as in (5.18). The relative error is fitted with curve Cn^{-2}. Polygon P is chosen to be a square with side $a = 25$ centred at the origin and rotated by $\frac{\pi}{4}$. The bandwidth is $h = 1000$ and values are evaluated for $\boldsymbol{y} = (100, 0)$. The correct value was calculated by rotating P and $\boldsymbol{y}$ by $\frac{\pi}{4}$ and using formula (5.15).

is usually considered. Such a value is called the **optimal bandwidth** and one may assume a reasonable performance of the estimator based on this value.

The aim is therefore to estimate the optimal bandwidth and then use it in the estimation of the volume fraction. In the classical problem of kernel univariate or bivariate density estimation several methods of choosing the optimal bandwidth are known, see e.g. [76, 80], however, for our task the situation is much more complicated and we can take them as a motivation rather than usable methods. The situation in our case is different since we have only one sample and values of the indicator $\mathbb{1}_X(\boldsymbol{x})$ are spatially correlated. Thus, except several specific cases like in Proposition 4.3.3, the choice of optimal bandwidth is analytically intractable.

In the following we introduce a heuristic approach to choosing the approximate optimal bandwidth based solely on one realization. Some numerical arguments in favour of this approach are presented further in Subsection 5.4.4. First note that according to (4.28) we have

$$\begin{aligned}\mathrm{MISE}\big(\hat{m}_h\big) = &\int_W\int_W\int_W \mathrm{cov}(\boldsymbol{u},\boldsymbol{v})\frac{K_h(\boldsymbol{x}-\boldsymbol{u})K_h(\boldsymbol{x}-\boldsymbol{v})}{e_h^2(\boldsymbol{x})}\,\mathrm{d}\boldsymbol{u}\mathrm{d}\boldsymbol{v}\mathrm{d}\boldsymbol{x}\\ &+\int_W\int_W\int_W m(\boldsymbol{u})m(\boldsymbol{v})\frac{K_h(\boldsymbol{x}-\boldsymbol{u})K_h(\boldsymbol{x}-\boldsymbol{v})}{e_h^2(\boldsymbol{x})}\,\mathrm{d}\boldsymbol{u}\mathrm{d}\boldsymbol{v}\mathrm{d}\boldsymbol{x}\\ &-2\int_W\int_W m(\boldsymbol{u})m(\boldsymbol{x})\frac{K_h(\boldsymbol{x}-\boldsymbol{u})}{e_h(\boldsymbol{x})}\,\mathrm{d}\boldsymbol{u}\mathrm{d}\boldsymbol{x}+\int_W m^2(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}. \qquad (5.19)\end{aligned}$$

Hence in order to minimize the MISE with respect to h one must have some preliminary estimates of $m(\boldsymbol{x})$ and $\mathrm{cov}(\boldsymbol{x},\boldsymbol{y})$ for all $\boldsymbol{x},\boldsymbol{y}\in W$. The optimal bandwidth is then determined by minimising $\mathrm{MISE}\big(\hat{m}_h\big)$, where those preliminary estimates are used instead of the true volume fraction and covariance function.

A natural choice of a volume fraction estimator is the kernel estimator $\hat{m}_h$ for some initial a priori chosen bandwidth h_0. In practice we use the approximated version $\hat{m}_{A;h}$ of $\hat{m}_h$ ob-

tained using the approximation of the kernel integral for polygons introduced in the previous subsection. Hence combining (5.12), (5.18), and (5.15) we get

$$\begin{aligned}\hat{m}_{A;h}(\boldsymbol{x}) &= \frac{1}{e_h(\boldsymbol{x})}\sum_{i=1}^{n}\sum_{j=1}^{n_i^2}\frac{\nu(P_i\cap W\cap B_{i;j})}{\nu(B_{i;j})}\int\limits_{B_{i;j}} K_h(\boldsymbol{x}-\boldsymbol{u})\,\mathrm{d}\boldsymbol{u}\\ &= \frac{1}{e_h(\boldsymbol{x})}\sum_{i=1}^{n}\sum_{j=1}^{n_i^2}\frac{\nu(P_i\cap W\cap B_{i;j})}{\nu(B_{i;j})}\prod_{\ell=1}^{2}\big(\Phi_h(u_{B_{i;j};\ell}+x_\ell)-\Phi_h(v_{B_{i;j};\ell}+x_\ell)\big),\end{aligned}$$

where $B_{i;j}$ is the j-th small rectangle given by the division of bounding box B_i of $P_i\cap W$ into $n_i\times n_i$ pieces, $\boldsymbol{v}_{B_{i;j}}=(v_{B_{i;j};1},v_{B_{i;j};2})$ is the lower left corner of $B_{i;j}$, $\boldsymbol{u}_{B_{i;j}}=(u_{B_{i;j};1},u_{B_{i;j};2})$ is the upper right corner of $B_{i;j}$, and $\Phi_h(x)$ denotes the cumulative distribution function of $N(0,h)$ defined by (5.14). The division of each polygon $P_i\cap W$ is taken such that $u_{B_{i;j};\ell}-v_{B_{i;j};\ell}<5\,\mathrm{m}$ for $\ell=1,2$. An example of $\hat{m}_{A;h_0}$ calculated for Prague with $h_0=2000$ is in Figure 5.15.

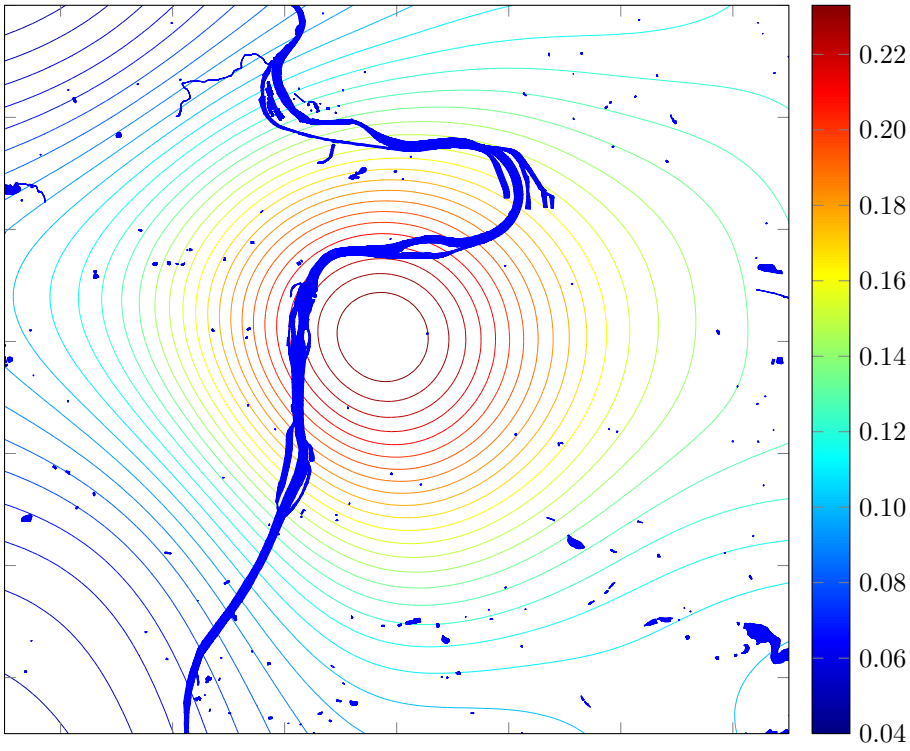


Figure 5.15: The volume fraction in Prague obtained using the kernel estimator $\hat{m}_{A;h_0}$ with $h_0=2000\,\mathrm{m}$.

Having $\hat{m}_{h_0}$ the $\mathrm{cov}(\boldsymbol{x},\boldsymbol{y})$ may be further estimated. This is done in the following way. Let us assume that the correlation function $\kappa(\boldsymbol{x},\boldsymbol{y})$ approximately depends only on $\|\boldsymbol{x}-\boldsymbol{y}\|$, which yields

$$\mathrm{cov}(\boldsymbol{x},\boldsymbol{y})\approx\kappa\big(\,\|\boldsymbol{x}-\boldsymbol{y}\|\,\big)\sqrt{m(\boldsymbol{x})-m(\boldsymbol{x})^2}\sqrt{m(\boldsymbol{y})-m(\boldsymbol{y})^2}. \tag{5.20}$$

The approximate estimator of cov based on $\hat{m}_h$ is therefore given by

$$\hat{\mathrm{cov}}_h(\boldsymbol{x},\boldsymbol{y})=\hat{\kappa}\big(\,\|\boldsymbol{x}-\boldsymbol{y}\|\,\big)\sqrt{\hat{m}_{A;h}(\boldsymbol{x})-\hat{m}_{A;h}(\boldsymbol{x})^2}\sqrt{\hat{m}_{A;h}(\boldsymbol{y})-\hat{m}_{A;h}(\boldsymbol{y})^2}, \tag{5.21}$$

where $\hat{\kappa}(\boldsymbol{r})$ is a suitable estimator of the correlation function. In order to be able to use $\hat{\mathrm{cov}}_h(\boldsymbol{x},\boldsymbol{y})$ in formula (5.19) for the MISE one needs to have an estimator that enables fast calculation of its values. Since $\hat{\mathrm{cov}}_h(\boldsymbol{x},\boldsymbol{y})$ consists of two parts, $\hat{m}_{A;h}$ and $\hat{\kappa}$, this requirement should be satisfied for both of them. In the case of $\hat{m}_{A;h}(\boldsymbol{x})$ can assume that it is, for h large enough, only slowly varying with $\boldsymbol{x}$. Thus its values may be calculated on a dense grid of points (we use inter-point distances from 50 m to 150 m) in W and then use the cubic spline interpolation. As a result the values of $\hat{m}_{A;h}(\boldsymbol{x})$ can be quickly obtained in arbitrary points of W with a sufficient precision.

For $\hat{\kappa}$ a different approach is used. It is approximated by a correlation function from a suitable parametric family of correlation functions fitted to the values of the estimator $\hat{\kappa}_{p;h}(\boldsymbol{r})$. The estimator $\hat{\kappa}_{p;h}(\boldsymbol{r})$ is defined by (4.32) and represents a digitalized counterpart of $\hat{\kappa}_{v;h}(r)$ defined by (4.30) that is too computationally demanding to be of a practical use.

The digitalization process was for non-stationary random closed sets introduced in Subsection 4.3.2. Here it is performed exactly in the same way. We begin with standardizing the values of the indicator $\mathbb{1}_{X\cap W}$ using $\hat{m}_{A;h}$, which yields

$$g_h(\boldsymbol{u}) = \mathbb{1}_W(\boldsymbol{u})\frac{\mathbb{1}_X(\boldsymbol{u}) - \hat{m}_{A;h}(\boldsymbol{u})}{\sqrt{\hat{m}_{A;h}(\boldsymbol{u}) - \hat{m}^2_{A;h}(\boldsymbol{u})}}.$$

The digitalization is now performed by sampling the values of g_h on some fixed grid of points given by the intersection of a regular point lattice $L^2 = a\mathbb{Z}^2 + \boldsymbol{c}$ with a rectangular box $B \subset W$. The shift $\boldsymbol{c}$ is chosen such that the lower left corner of B corresponds to $a(1,1) + \boldsymbol{c}$. Let $B = [\boldsymbol{a},\boldsymbol{b}] \equiv [a_1,b_1]\times[a_2,b_2]$. The values of g_h on the grid define a matrix $\boldsymbol{G}_h \in \{0,1\}^{n_1,n_2}$ with elements

$$(\boldsymbol{G}_h)_{\boldsymbol{i}} = g_h\big(a\boldsymbol{i} + \boldsymbol{c}\big), \quad \boldsymbol{i} \in \{1,\dots,n_1\}\times\{1,\dots,n_2\},$$

where $n_j \in \mathbb{N}, (b_j - a_j)/a - 1 \le n_j \le (b_j - a_j)/a$ for $j = 1,2$. Similarly we set

$$(\boldsymbol{I})_{\boldsymbol{i}} = \mathbb{1}_W\big(a\boldsymbol{i} + \boldsymbol{c}\big), \quad \boldsymbol{i} \in \{1,\dots,n_1\}\times\{1,\dots,n_2\},$$

for the observation window only. The discrete version $\hat{\kappa}_{p;h}(\boldsymbol{r})$ of the estimator $\hat{\kappa}_{v;h}(\boldsymbol{r})$ is given by (4.32) as

$$\hat{\kappa}_{p;h}(a\boldsymbol{k}) = \frac{\sum_{\boldsymbol{i}\in A(\boldsymbol{k})}(\boldsymbol{G}_h)_{\boldsymbol{i}}(\boldsymbol{G}_h)_{\boldsymbol{i}+\boldsymbol{k}}}{\sum_{\boldsymbol{i}\in A(\boldsymbol{k})}(\boldsymbol{I})_{\boldsymbol{i}}(\boldsymbol{I})_{\boldsymbol{i}+\boldsymbol{k}}}$$

for all $\boldsymbol{k}$ from a subset of $\{-n_1+1,\dots,n_1-1\}\times\{-n_2+1,\dots,n_2-1\}$ such that the denominator is positive, and $A(\boldsymbol{k}) = \{\boldsymbol{i}\in\mathbb{Z}^2 | 1 \le \boldsymbol{i} \le \boldsymbol{n}, 1 \le \boldsymbol{i}+\boldsymbol{k} \le \boldsymbol{n}\}$. Again $\hat{\kappa}_{p;h}(a\boldsymbol{k})$ can be calculated with the help of the discrete fast Fourier transform using relation (4.33), where matrices $\boldsymbol{G}$ and $\boldsymbol{I}$ must be padded with zeros.

In practice the data for different cities are sampled with different values of a in the range $[0.7, 3]$ m, which is chosen so that the sizes of obtained matrices $\boldsymbol{G}$ and $\boldsymbol{I}$ are around 10000×10000. This is a limit, determined by the computer memory, for which we are able to perform the fast Fourier transform of padded matrices that are of double size. In contrary to the stationary approach from the previous section, the sampling region B is not taken as the bounding box D of the observation window W, but it is taken smaller by 1 km in each direction. This is to avoid largest edge effects in $\hat{m}_h$ arising near the boundaries.

The result of the calculation of $\hat{\kappa}_{p;h}(\boldsymbol{r})$ for Prague is shown in Figure 5.16. One can see that $\hat{\kappa}_{p;h}$ seems to be direction independent. Thus in the following it is assumed that $\kappa(\boldsymbol{r})$ depends only on the norm $\|\boldsymbol{r}\|$ of $\boldsymbol{r}$. For a visualisation of the dependence of $\hat{\kappa}_{p;h}(\boldsymbol{r})$ on $\|\boldsymbol{r}\|$ the isotropic adapted estimator $\hat{\kappa}^{AI}_{p;h}(r)$ is used. It is defined by (4.24) as

$$\hat{\kappa}^{AI}_{p;h}(r) = \frac{\sum_{\boldsymbol{k}} L_{h(r)}\big(\|a\boldsymbol{k}\| - r\big)\hat{\kappa}_{p;h}(a\boldsymbol{k})}{\sum_{\boldsymbol{k}} L_{h(r)}\big(\|a\boldsymbol{k}\| - r\big)},$$

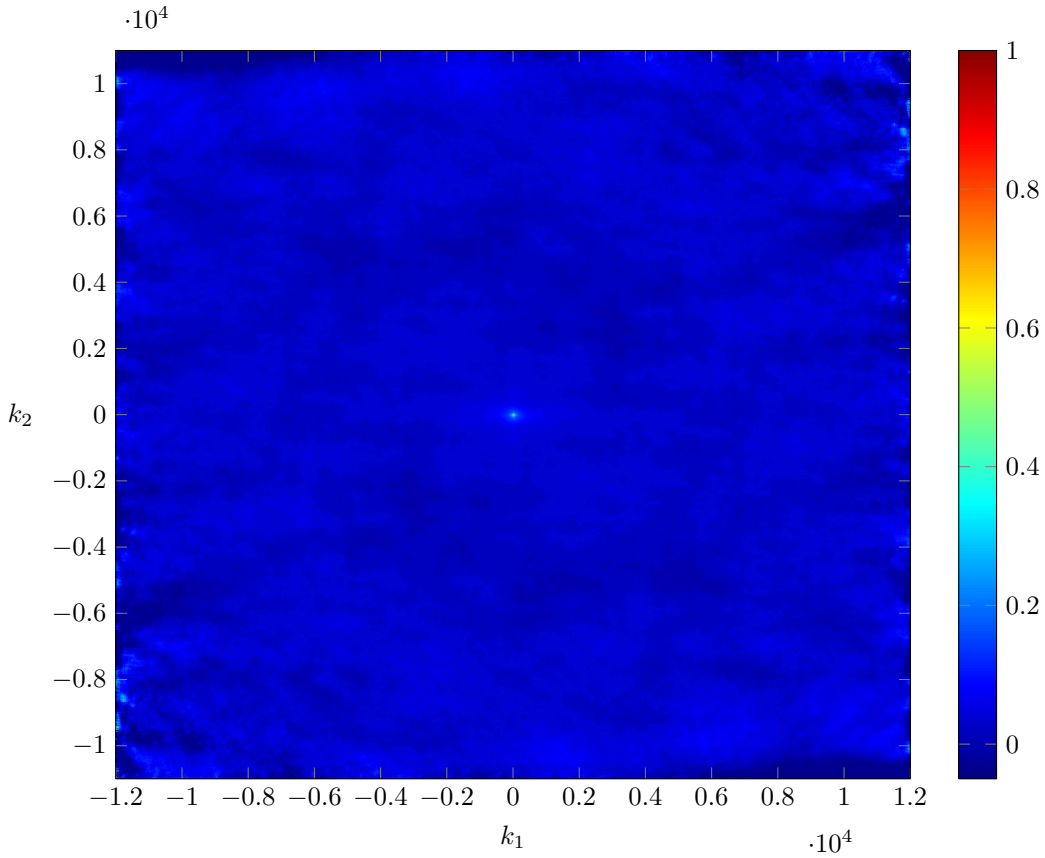


Figure 5.16: The estimator $\hat{\kappa}_{p;h_0}(\boldsymbol{k})$ with $h_0 = 2000\,\mathrm{m}$ for Prague. The built-up area is sampled with $a = 1\,\mathrm{m}$.

where L is the Gaussian kernel so that $L_h(x) = \frac{1}{h\sqrt{2\pi}} e^{-\frac{x^2}{2h^2}}$ and the bandwidth function $h(r)$ is chosen to be

$$h(r) = 30 \cdot (1 - e^{-0.001 r/\mathrm{m}})\,\mathrm{m} + 0.5\,\mathrm{m}$$

exactly as in Section 5.3. The result obtained for Prague is shown in Figure 5.17.

Similarly to the stationary case the obtained values of the correlation function are well fitted by the Cauchy correlation function of the form (5.2). The fit by the Cauchy correlation function is also the best among other commonly used covariance classes. The fit is performed by minimizing the least squares with predefined weights. As the estimator $\hat{\boldsymbol{\theta}}_h = (\hat{\alpha}_h, \hat{\beta}_h, \hat{\theta}_h)$ of the vector $\boldsymbol{\theta} = (\alpha, \beta, \theta)$ of parameters we choose $\boldsymbol{\theta}$ constrained to $0 < \beta, \theta < \infty$, $0 < \alpha \leq 2$, that minimizes

$$Q(\boldsymbol{\theta}) = \sum_{\boldsymbol{k}=-\boldsymbol{n}+1}^{\boldsymbol{n}-1} w_{\boldsymbol{k}} \big(\hat{\kappa}_{p;h}(a\boldsymbol{k}) - \kappa_C(\|a\boldsymbol{k}\|\,;\boldsymbol{\theta})\big)^2,$$

where $\boldsymbol{n} = (n_1, n_2)$ and the weights $w_{\boldsymbol{k}}$ are chosen as $w_{\boldsymbol{k}} = \frac{N_{\boldsymbol{k}}}{a\|\boldsymbol{k}\|+1}$. The coefficient $N_{\boldsymbol{k}} = \sum_{\boldsymbol{i}\in A(\boldsymbol{k})} (\boldsymbol{I})_{\boldsymbol{i}} (\boldsymbol{I})_{\boldsymbol{i}+\boldsymbol{k}}$ gives the number of points that were used for the calculation of $\hat{\kappa}_{p;h}(a\boldsymbol{k})$. An example of such a fit for Prague is presented in Figure 5.17.

Taking all together, given h we are able to construct the approximate estimator $\hat{\mathrm{cov}}_h(\boldsymbol{x}, \boldsymbol{y})$ according to (5.21) as

$$\hat{\mathrm{cov}}_h(\boldsymbol{x}, \boldsymbol{y}) = \kappa_C\big(\,\|\boldsymbol{x} - \boldsymbol{y}\|\,;\hat{\boldsymbol{\theta}}_h\big) \sqrt{\hat{m}_{A;h}(\boldsymbol{x}) - \hat{m}_{A;h}(\boldsymbol{x})^2} \sqrt{\hat{m}_{A;h}(\boldsymbol{y}) - \hat{m}_{A;h}(\boldsymbol{y})^2}.$$

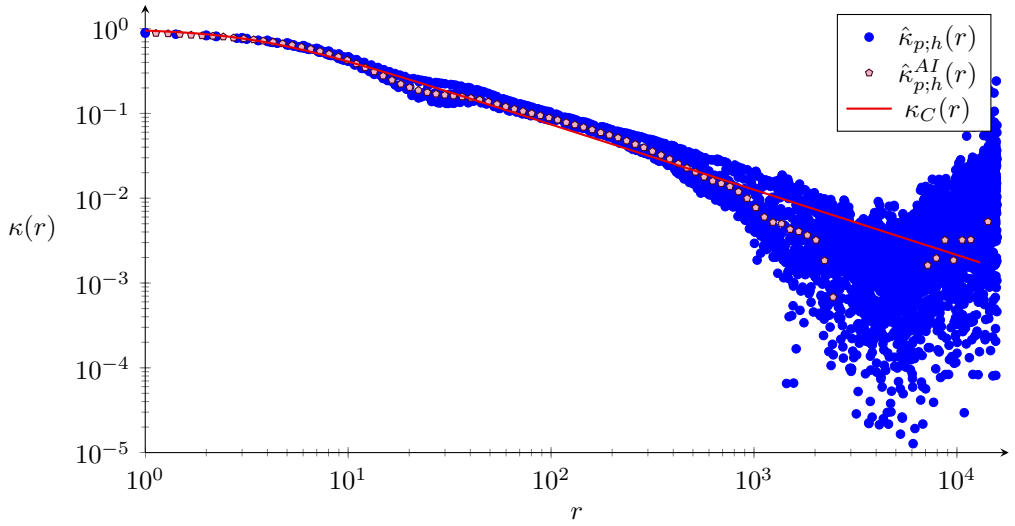


Figure 5.17: An example of Prague. The result of the adapted estimator $\hat{\kappa}^{AI}_{p;h_0}(r)$ is shown together with $\hat{\kappa}_{p;h_0}$ for $h_0 = 2000\,\mathrm{m}$. The values of $\hat{\kappa}_{p;h_0}$ are fitted by the Cauchy correlation function $\kappa_C(r;\hat{\boldsymbol{\theta}})$ with $\hat{\alpha} = 1.71$, $\hat{\beta} = 0.77$, and $\hat{\theta} = 3.41\,\mathrm{m}$.

Based a priory chosen initial value h_0 we obtain the volume fraction estimate $\hat{m}_{h_0}$, correlation function estimate $\hat{\kappa}_{p;h_0}(\boldsymbol{r})$, parameter estimate $\hat{\boldsymbol{\theta}}_{h_0}$ of the Cauchy fit $\kappa_C(r;\hat{\boldsymbol{\theta}}_{h_0})$ of the correlation function, and consequently the covariance function estimate $\hat{\mathrm{cov}}_h(\boldsymbol{x},\boldsymbol{y})$. Let $\mathrm{MISE}_{h_0}(\hat{m}_h)$ denote the mean integrated square error defined by (5.19) with m replaced by $\hat{m}_{h_0}$ and cov by $\hat{\mathrm{cov}}_{h_0}$. As a further step the numerical minimization of $\mathrm{MISE}_{h_0}(\hat{m}_h)$ is performed yielding the minimizing value h_1 that is taken as the optimal bandwidth. It is worth noting that the numerical minimization of $\mathrm{MISE}_{h_0}(\hat{m}_h)$ with some reasonable accuracy is computationally very demanding task since one has to calculate sixfold integrals in $\mathbb{R}$. Finally, having the optimal bandwidth h_1 the estimation procedure is repeated yielding the volume fraction estimate $\hat{m}_{h_1}$, correlation function estimate $\hat{\kappa}_{p;h_1}(\boldsymbol{r})$, and parameter estimate $\hat{\boldsymbol{\theta}}_{h_1}$ of the Cauchy fit $\kappa_C(r;\hat{\boldsymbol{\theta}}_{h_1})$

In the analysis of the built-up structure $h_0 = 2000\,\mathrm{m}$ is used as the initial value for all cities. Note that this choice corresponds to the optimal bandwidth from Proposition 4.3.3 for Gaussian case with parameters $B = 0.3$, $a = 10\,\mathrm{km}$ and exponential correlation function with $\alpha = 1/50\,\mathrm{m}^{-1}$. Those parameters belongs to theoretical city with maximal volume fraction 0.3 and with approximately 40% of all build-up land within the a distance form the core and approximately 86% within the $2a = 20\,\mathrm{km}$ distance from the core. The value $\alpha = 1/50\,\mathrm{m}^{-1}$ corresponds to situation when the correlation function at $x = 50\,\mathrm{m}$ equals $e^{-1} \approx 0.37$.

The estimated values of $\hat{\boldsymbol{\theta}}_{h_0}$, h_1, and $\hat{\boldsymbol{\theta}}_{h_1}$ for all cities are summarized in Table 5.6. The volume fraction in Prague is shown in Figure 5.15, the correlation function in Figure 5.16, and its fit by the Cauchy correlation function in Figure 5.17. In Figure 5.18 the difference between volume fraction estimates $\hat{m}_{h_1}$ and $\hat{m}_{h_0}$ for Prague is shown. Since in Prague $h_1 < h_0$ the estimate $\hat{m}_{h_1}$ is sharper leading to higher values in the city core and lower values in the periphery. In Figure 5.19 we show the difference between the correlation function estimates in Prague obtained for h_0 and h_1. The results of the isotropic adapted estimator $\hat{\kappa}^{AI}_{p;h_1}(r)$ based on the optimal bandwidth h_1 are for all cities shown in Figure 5.20.

It is also interesting to see how correlations depend on different locations of a city. This means how plausible is the assumption $\kappa(\boldsymbol{x},\boldsymbol{y}) \approx \kappa(\|\boldsymbol{x}-\boldsymbol{y}\|)$. In order to check this in Prague, Boston, Paris, and Los Angeles we choose sampling regions $B_1,\ldots,B_4 \subset W$ and estimate the correlation function in each region separately. For each $i = 1,\ldots,4$ we denote by $\hat{\kappa}^{AI}_{p;h_1;B_i}$ the

Table 5.6: Estimated parameters for all analysed cities: a is the sampling lattice constant; $\hat{\alpha}_{h_0}$, $\hat{\beta}_{h_0}$, and $\hat{\theta}_{h_0}$ are parameters of the best fit of $\hat{\kappa}_{p;h_0}$ by the Cauchy correlation function $\kappa_C(r;\hat{\boldsymbol{\theta}}_{h_0})$; h_1 is the optimal bandwidth obtained by the minimisation of $\mathrm{MISE}_{h_0}(\hat{m}_h)$; $\hat{\alpha}_{h_1}$, $\hat{\beta}_{h_1}$, and $\hat{\theta}_{h_1}$ are parameters of the best fit of $\hat{\kappa}_{p;h_1}$ by the Cauchy correlation function $\kappa_C(r;\hat{\boldsymbol{\theta}}_{h_1})$.

City	a [m]	$\hat{\alpha}_{h_0}$	$\hat{\beta}_{h_0}$	$\hat{\theta}_{h_0}$ [m]	h_1 [m]	$\hat{\alpha}_{h_1}$	$\hat{\beta}_{h_1}$	$\hat{\theta}_{h_1}$ [m]
Boston	1	2.00	0.68	2.57	2578	2.00	0.67	2.62
Chicago	1.5	2.00	0.62	2.47	5376	2.00	0.58	2.29
Los Angeles	3	2.00	0.67	2.23	3312	2.00	0.64	2.13
Pittsburgh	1.5	1.97	0.72	2.14	2439	2.00	0.69	2.06
Seattle	2	2.00	0.53	1.56	6070	2.00	0.52	1.58
Berlin	1	2.00	0.81	4.13	2636	2.00	0.79	4.09
Birmingham	1.5	2.00	0.80	2.88	2217	2.00	0.79	2.85
Milan	0.7	2.00	0.77	3.87	1907	2.00	0.78	3.91
Minsk	1	1.61	0.82	5.05	2727	1.76	0.78	4.70
Moscow	2	1.39	0.93	6.63	2220	1.41	0.91	6.45
Oslo	0.7	2.00	0.65	2.59	2723	2.00	0.61	2.50
Paris	2	1.12	0.88	4.73	2035	1.12	0.88	4.70
Prague	1	1.71	0.77	3.41	1775	1.65	0.80	3.63
St. Petersburg	1.5	1.85	0.83	4.96	2105	1.86	0.82	4.91

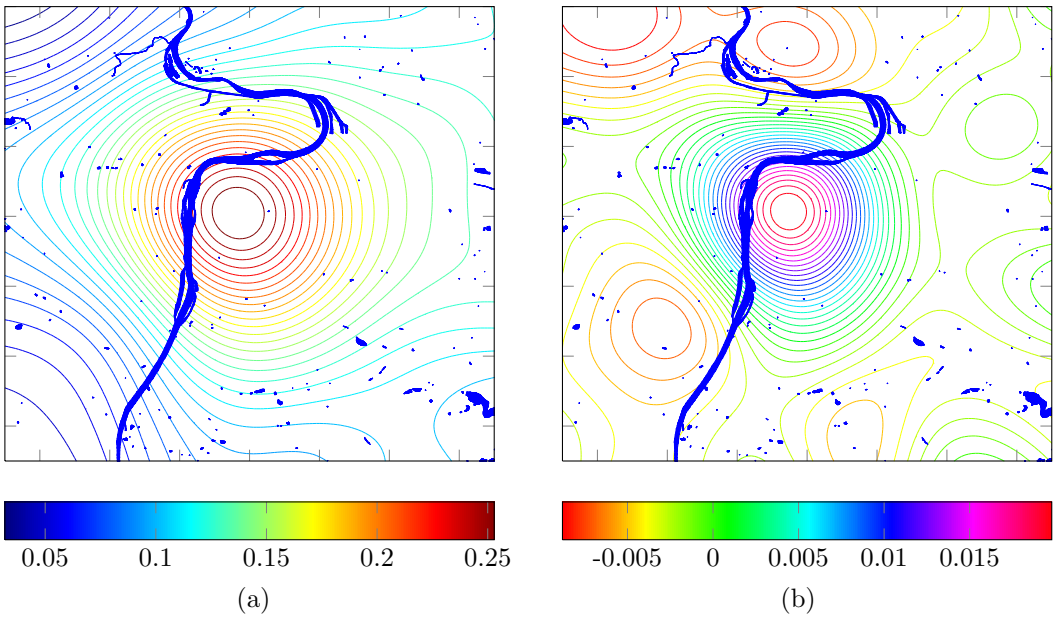


Figure 5.18: (a): The volume fraction in Prague obtained using the kernel estimator $\hat{m}_{A;h_1}$ with $h_1 = 1775\,\mathrm{m}$. (b): The difference $\hat{m}_{A;h_1}(\boldsymbol{x}) - \hat{m}_{A;h_0}(\boldsymbol{x})$ between volume fractions obtained for $h_0 = 2000\,\mathrm{m}$ (shown in Figure 5.16) and for h_1.

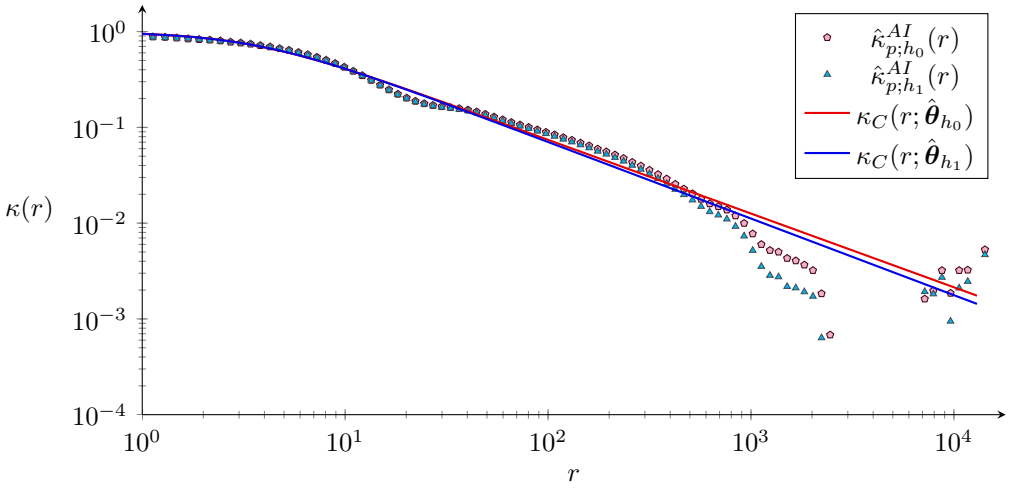


Figure 5.19: A comparison of correlation function estimates based on $h_0 = 2000\,\mathrm{m}$ and $h_1 = 1775\,\mathrm{m}$ in Prague.

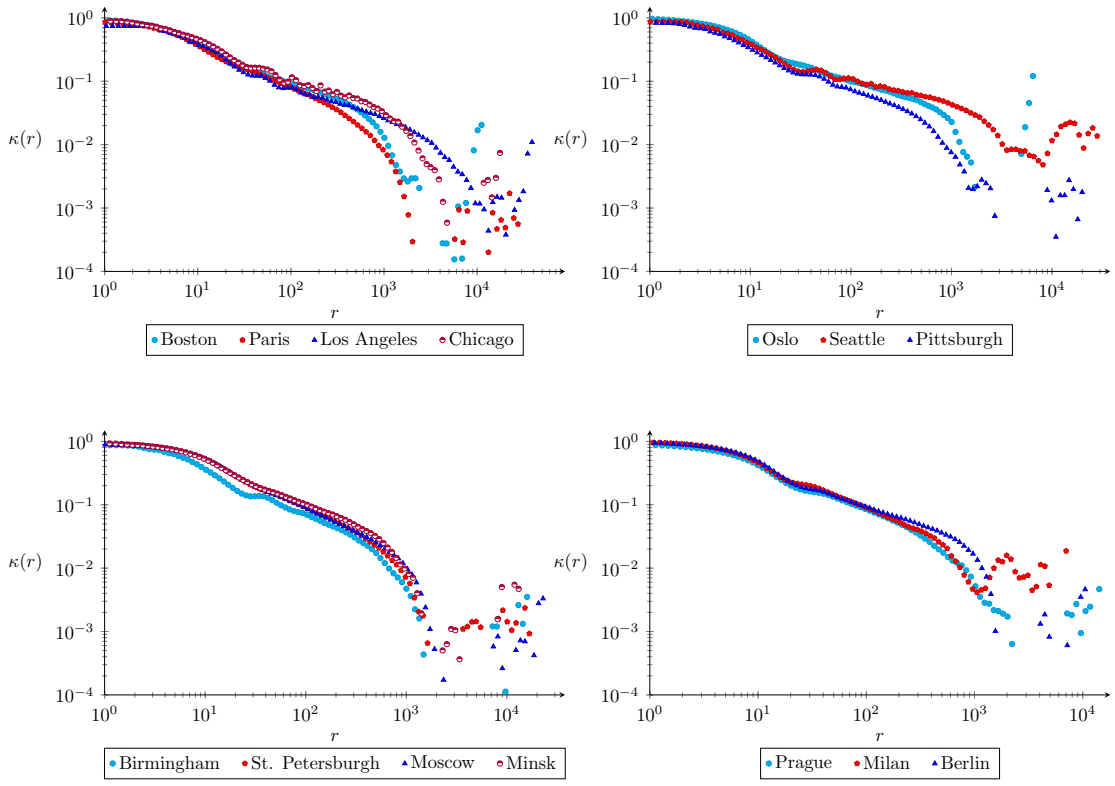


Figure 5.20: The results of the isotropic adapted estimator $\hat{\kappa}^{AI}_{p;h_1}(r)$ based on optimal bandwidths h_1 for all analysed cites.

isotropic adapted estimator based on h_1 that is obtained in B_i. The positions of sampling regions in W and the obtained estimators are shown in Figure 5.21. One can see that the behaviour is relatively similar in all subregions. The only larger deteriorations are visible in Prague and Paris where results corresponding to regions B_2 in the range $r \in (200, 2000)\,\mathrm{m}$ are larger then for other regions. This sampling regions partially cover areas with the highest volume fraction.

In Boston and especially in Los Angeles one can further see small oscillations of the correlation function in the range $r \in (20, 400)\,\mathrm{m}$. The reason is the same as it was discussed in the stationary approach in Section 5.3 on pages 111 and 124. Thus it is a result of the gridiron urban plans that are common in many North American cities. Those oscillations, however, do not influence the asymptotic behaviour of the correlation function as the waves blur and eventually disappear for large values of r. This effect is also visible in Figure 5.20 for Los Angeles, Chicago, and Seattle.

5.4.3 Justification of the window choice for the stationary approach

In Section 5.3 the built-up area was analysed using the stationary random closed set approach. The stationarity is assumed to hold in a particular part of a city around the city core determined by the observation window W_s. In order to justify the specific choice of W_s for a given city (see Table 5.1 for coordinates of bounding boxes of windows used in the analysis) it is reasonable to observe how the volume fraction $\hat{m}_{A;h_1}$ based on the optimal bandwidth h_1 changes inside W_s. This is done by determining $\max_{\boldsymbol{x}\in W_s} \hat{m}_{A;h_1}(\boldsymbol{x})$, $\min_{\boldsymbol{x}\in W_s} \hat{m}_{A;h_1}(\boldsymbol{x})$, and the relative maximal difference of those values

$$\Delta_r \hat{m}_{A;h_1} = \frac{\max_{\boldsymbol{x}\in W_s} \hat{m}_{A;h_1}(\boldsymbol{x}) - \min_{\boldsymbol{x}\in W_s} \hat{m}_{A;h_1}(\boldsymbol{x})}{\max_{\boldsymbol{x}\in W_s} \hat{m}_{A;h_1}(\boldsymbol{x})}.$$

The results for analysed cities are shown in Table 5.7. It is interesting to compare values of the volume fraction $\hat{m}_{A;h_1}$ in W_s with values of $\hat{p}_v$ summarized in Table 5.2 obtained in the stationary approach. Since W_s is located around a city core, where the volume fraction is highest, the values obtained from this small window are generally larger then values obtained by the kernel estimator. The reason is in the fact that for large bandwidth the kernel estimator is based on a large area meaning that it is influenced by changes of the volume fraction in large distances. If this is combined with W_s of sides not much larger then h, it leads to generally lower values of the volume fraction estimates obtained using $\hat{m}_{A;h}$. Thus for a volume fraction that forms a peak with a sufficiently fast decrease on distances compared to the value of h and with W_s having dimensions not much larger then h, the values of the stationary estimator $\hat{p}_v$ are larger then those of the kernel estimator $\hat{m}_{A,h}$. The smaller is h with respect to the size of W_s the smaller is the difference between the maximal volume fraction estimated by the kernel estimator and by the stationary estimator. This is clearly visible for Berlin and Los Angeles, where the values are close together. On the other hand for Seattle and Oslo, where h is larger then dimensions of W_s, there is a much bigger difference between $\hat{p}_v$ and $\hat{m}_{A,h}$. A depiction of the concrete situation for Prague and Pittsburgh is given in Figure 5.22.

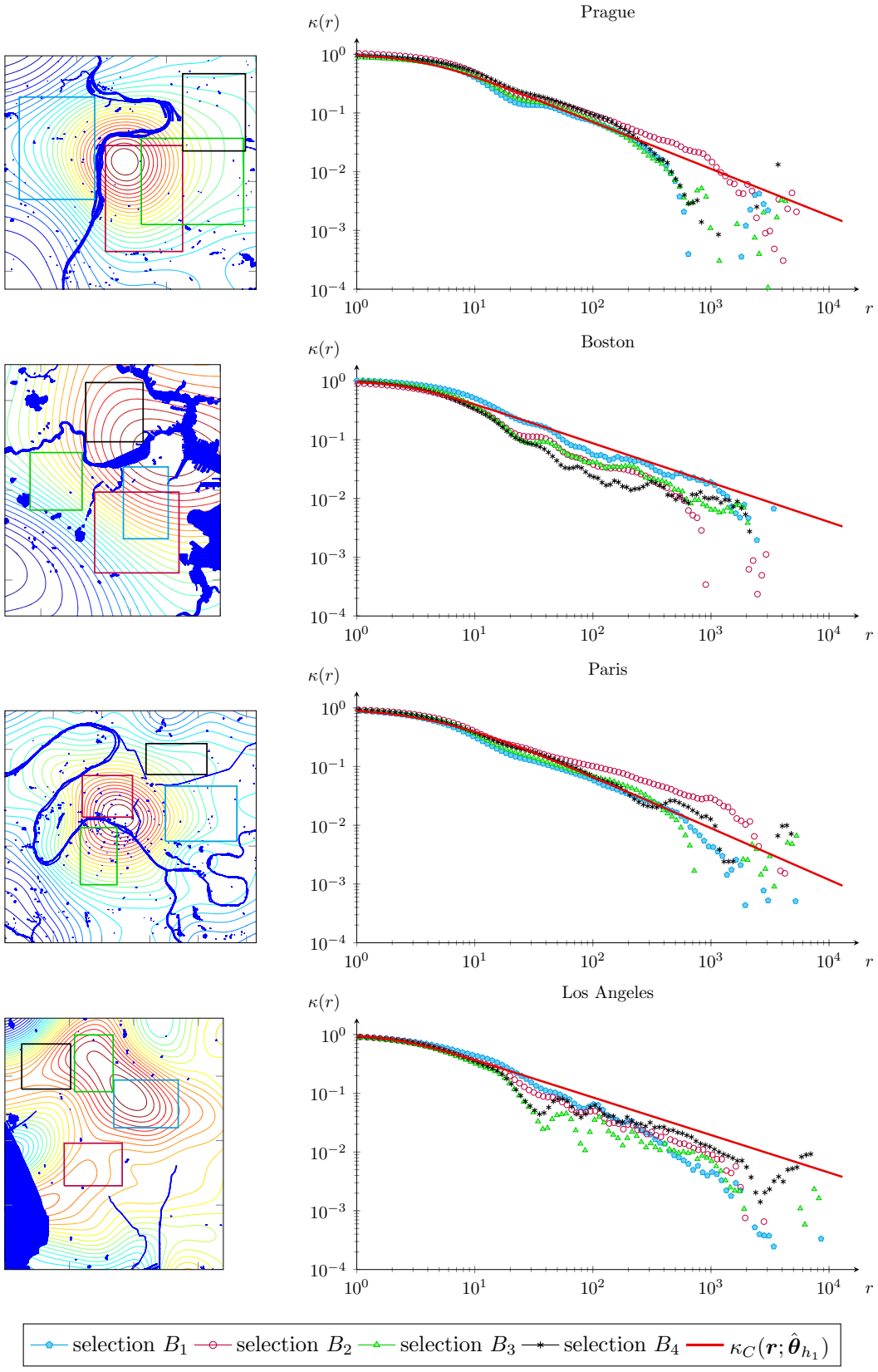
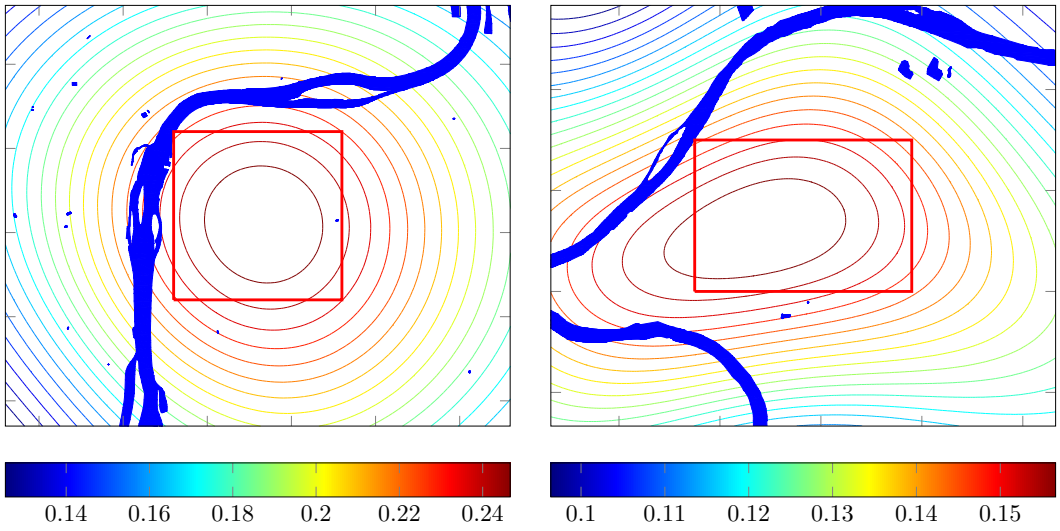


Figure 5.21: On the left: Selections $B_1, \ldots, B_4$ together with the volume fraction $\hat{m}_{h_1}$ in a given city. On the right: Estimates $\hat{\kappa}^{AI}_{p;h_1;B_i}$ of the correlation function based on h_1 and $B_1, \ldots, B_4$. The Cauchy correlation function with parameters $\hat{\alpha}_{h_1}$, $\hat{\beta}_{h_1}$ and $\hat{\theta}_{h_1}$ from Table 5.6 is also shown.

Table 5.7: Maximal, minimal, and relative differences of values of the volume fraction estimator $\hat{m}_{A;h_1}$ inside the stationary selection window W_s used in Section 5.3.

City	h_1 [m]	$\max \hat{m}_{A;h_1}$	$\min \hat{m}_{A;h_1}$	$\Delta_r \hat{m}_{A;h_1}$
Boston	2578	0.26	0.23	13.3%
Chicago	5376	0.28	0.27	4.1%
Los Angeles	3312	0.31	0.28	8.2%
Pittsburgh	2439	0.16	0.14	11.7%
Seattle	6070	0.20	0.19	2.9%
Berlin	2636	0.25	0.23	7.0%
Birmingham	2217	0.23	0.21	8.3%
Milan	1907	0.36	0.34	6.7%
Minsk	2727	0.16	0.15	6.5%
Moscow	2220	0.26	0.23	10.2%
Oslo	2723	0.21	0.20	6.5%
Paris	2035	0.44	0.38	14.5%
Prague	1775	0.25	0.23	11.0%
St. Petersburgh	2105	0.33	0.30	9.8%



(a) Prague (b) Pittsburgh

Figure 5.22: (a): The volume fraction $\hat{m}_{A;h_1}$ with $h_1 = 1775\,\mathrm{m}$ and the window W_s used in the stationary analysis in Prague. (b): The volume fraction $\hat{m}_{A;h_1}$ with $h_1 = 2439\,\mathrm{m}$ and W_s in Pittsburgh.

5.4.4 Numerical simulations

In the following we check several numerical properties of the estimators used in first parts of this section. As a suitable model of a built-up area we choose the level excursion set introduced in Section 2.4 by Definition 2.4.5. In particular we take the 0-level excursion set $X_0(Z)$ of a Gaussian random field Z determined by the mean function $\mu_Z(\boldsymbol{x})$ and covariance function $\mathrm{cov}_Z(\boldsymbol{x}, \boldsymbol{y})$ given by

$$\mu_Z(\boldsymbol{x}) = \Phi^{-1}\left(Ae^{-\frac{\boldsymbol{x}^2}{2\sigma^2}} + B\right)$$

and

$$\mathrm{cov}_Z(\boldsymbol{x},\boldsymbol{y}) = \kappa_C(\|\boldsymbol{x}-\boldsymbol{y}\|;\boldsymbol{\theta}) = \left(1+\left(\frac{\|\boldsymbol{x}-\boldsymbol{y}\|}{\theta}\right)^{\alpha}\right)^{-\frac{\beta}{\alpha}}, \tag{5.22}$$

where Φ is a distribution function of a standard normal random variable, $A,B,\beta,\theta > 0$, $A+B \le 1$, and $\alpha \in (0,2]$. By tuning parameters A,B,α,β,θ qualitative properties of $X_0(Z)$ may be changed.

Again as in Section 4.4 Z has almost surely continuous sample functions since the covariance function satisfies Corollary 2.4.1. Therefore $X_0(Z) = \{\boldsymbol{x}|Z(\boldsymbol{x}) \ge 0\}$ is a non-stationary random closed set. For the volume fraction $m(\boldsymbol{x})$ of $X_0(Z)$ we use (2.12) and obtain

$$m(\boldsymbol{x}) = 1-\Phi\left(-\mu_Z(\boldsymbol{x})\right) = \Phi\left(\mu_Z(\boldsymbol{x})\right) = Ae^{-\frac{\boldsymbol{x}^2}{2\sigma^2}} + B. \tag{5.23}$$

The covariance function $\mathrm{cov}(\boldsymbol{x},\boldsymbol{y})$ of $X_0(Z)$ is from (2.13) given by

$$\mathrm{cov}(\boldsymbol{x},\boldsymbol{y}) = \int_0^{\kappa_C(\|\boldsymbol{x}-\boldsymbol{y}\|)} \frac{1}{2\pi\sqrt{1-z^2}} e^{-\frac{\tilde{x}^2+\tilde{y}^2-2z\tilde{x}\tilde{y}}{2(1-z^2)}}\,\mathrm{d}z,$$

where $\tilde{x} = -\mu_Z(\boldsymbol{x})$, $\tilde{y} = -\mu_Z(\boldsymbol{y})$. In particular

$$\mathrm{cov}(\boldsymbol{x},\boldsymbol{x}) = \int_0^1 \frac{1}{2\pi\sqrt{1-z^2}} e^{-\frac{\mu_Z(\boldsymbol{x})}{2(1+z)}}\,\mathrm{d}z = m(\boldsymbol{x})\big(1-m(\boldsymbol{x})\big).$$

The correlation function is thus

$$\kappa(\boldsymbol{x},\boldsymbol{y}) = \frac{1}{\sqrt{\mathrm{cov}(\boldsymbol{x},\boldsymbol{x})\,\mathrm{cov}(\boldsymbol{y},\boldsymbol{y})}} \int_0^{\kappa_C(\|\boldsymbol{x}-\boldsymbol{y}\|)} \frac{1}{2\pi\sqrt{1-z^2}} e^{-\frac{\mu_Z(\boldsymbol{x})^2+\mu_Z(\boldsymbol{y})^2-2z\mu_Z(\boldsymbol{x})\mu_Z(\boldsymbol{y})}{2(1-z^2)}}\,\mathrm{d}z.$$

Since $\mu_Z(\boldsymbol{x})$ depends on $\boldsymbol{x}$ we see that in general $\kappa(\boldsymbol{x},\boldsymbol{y}) \ne \kappa(\boldsymbol{x}-\boldsymbol{y},\boldsymbol{0})$. On a concrete example in Figure 5.25 it is shown how large the difference may be.

For simulations of the Gaussian random field Z and corresponding excursion set $X_0(Z)$ the circular embedding is used similarly as in Section 4.4. Thus we are able to simulate values of Z on the regular grid of points $\{a\boldsymbol{k}+\boldsymbol{c}|k_1 = 1,\dots,n_1,\ k_2 \in 1,\dots,n_2\}$ for some $a > 0, \boldsymbol{c} \in \mathbb{R}^2$, and $\boldsymbol{n} = (n_1,n_2) \in \mathbb{N}^2$. Hence we obtain a matrix $\boldsymbol{Z} \in \mathbb{R}^{n_1,n_2}$ with elements corresponding to the values of Z at the grid points,

$$(\boldsymbol{Z})_{\boldsymbol{i}} = Z(a\boldsymbol{i}+\boldsymbol{c}).$$

The values of the indicator of $X_0(Z)$ at the grid points form a matrix $\boldsymbol{M} \in \{0,1\}^{n_1,n_2}$ with elements

$$(\boldsymbol{M})_{\boldsymbol{i}} = \mathbb{1}_{X_0(Z)}(a\boldsymbol{i}+\boldsymbol{c}) = \mathbb{1}\big((\boldsymbol{Z})_{\boldsymbol{i}} \ge 0\big).$$

To estimate the volume fraction the discrete kernel estimator

$$\hat{m}_{p;h}(a\boldsymbol{k}+\boldsymbol{c}) = \frac{\sum_{\boldsymbol{i}=(1,1)}^{\boldsymbol{n}} (\boldsymbol{M})_{\boldsymbol{i}} K_h(a\boldsymbol{k}-a\boldsymbol{i})}{\sum_{\boldsymbol{i}=(1,1)}^{\boldsymbol{n}} K_h(a\boldsymbol{k}-a\boldsymbol{i})}$$

is used, where K is the Gaussian kernel and $1 \le \boldsymbol{k} \le \boldsymbol{n}$.

The correlation function is estimated by $\hat{\kappa}_{p;h}$, which is defined, analogously to the previous part, by (4.32), where the matrix $\boldsymbol{G}$ is now

$$(\boldsymbol{G}_h)_{\boldsymbol{i}} = \frac{(\boldsymbol{M})_{\boldsymbol{i}} - \hat{m}_{p;h}(a\boldsymbol{i}+\boldsymbol{c})}{\sqrt{\hat{m}_{p;h}(a\boldsymbol{i}+\boldsymbol{c}) - \hat{m}^2_{p;h}(a\boldsymbol{i}+\boldsymbol{c})}}.$$

Table 5.8: Settings used in simulations. The volume fraction is defined through (5.23) with parameters A, B, σ and the covariance through (5.22) with parameters α, β, θ. The simulations are performed in the window W and discrete estimators are based on a regular grid scheme determined by parameters a and $\boldsymbol{c}$.

	Volume fraction			Window and sampling			Correlation		
Setting	A	B	σ	W	a	$\boldsymbol{c}$	α	β	θ
1–a	0.2	0.1	1800	$[-5000, 5000]^2$	2	$(-5002, -5002)$	1.8	0.8	3
1–b	0.2	0.1	1800	$[-5000, 5000]^2$	2	$(-5002, -5002)$	2	1.05	4
2–a	0.25	0.05	3000	$[-2500, 2500]^2$	1	$(-2501, -2501)$	1.8	0.8	3
2–b	0.25	0.05	3000	$[-2500, 2500]^2$	1	$(-2501, -2501)$	2	1.05	4

The rest of the estimation process remains the same as before with the only difference. Instead of the Cauchy class of correlation functions, which is used to fit the values of $\hat{\kappa}_{p;h}$ we take the class of correlation functions of the 0-level excursion set based on a stationary standard Gaussian random field with Cauchy covariance. Hence by (2.16), the general form of the correlation function is

$$\kappa_L(r) \equiv \kappa_L(r; \boldsymbol{\theta}) = \frac{2}{\pi} \arcsin\left(\left(1 + \left(\frac{r}{\theta} \right)^{\alpha} \right)^{-\frac{\beta}{\alpha}} \right) \tag{5.24}$$

for $0 < \alpha \leq 2$ and $\beta, \theta > 0$ with $\boldsymbol{\theta} = (\alpha, \beta, \theta)$. Note that this class was already used in Section 4.4. Thus,

$$\hat{\mathrm{cov}}_h(\boldsymbol{x}, \boldsymbol{y}) = \kappa_L\big(\|\boldsymbol{x} - \boldsymbol{y}\| ; \hat{\boldsymbol{\theta}}_h \big) \sqrt{\hat{m}_{p;h}(\boldsymbol{x}) - \hat{m}_{p;h}(\boldsymbol{x})^2} \sqrt{\hat{m}_{p;h}(\boldsymbol{y}) - \hat{m}_{p;h}(\boldsymbol{y})^2} \tag{5.25}$$

is taken as the approximate estimator (5.21) of the covariance function cov of $X_0(Z)$.

Since the extensive analysis of our estimation procedure is computationally demanding we focus on four particular settings, 1–a, 1–b, 2–a, and 2–b, determined by specific settings of $\mu_Z(\boldsymbol{x})$ and $\mathrm{cov}_Z(\boldsymbol{x}, \boldsymbol{y})$. The first two, 1–a and 1–b, have a sharp volume fraction profile determined by coefficients $A = 0.2$, $B = 0.1$, $\sigma = 1800$ in (5.23). They are analysed in the window $W = [-5000, 5000] \times [-5000, 5000]$ and sampled with $a = 2$ and $\boldsymbol{c} = (-5002, -5002)$. The last two, 2–a and 2–b, have a more flat volume fraction profile determined by $A = 0.25$, $B = 0.05$, $\sigma = 3000$ in the window $W = [-2500, 2500] \times [-2500, 2500]$ and sampled with $a = 1$ and $\boldsymbol{c} = (-2501, -2501)$. In the correlation structure we combine two different settings. Settings 1–a and 2–a have slowly decaying correlations given by $\alpha = 1.8$, $\beta = 0.8$, $\theta = 3$ in (5.22). Settings 1–b and 2–b have faster decaying correlations given by $\alpha = 2$, $\beta = 1.05$, $\theta = 4$. The parameters of all four combinations are summarized in Table 5.8 and corresponding volume fractions are shown in Figure 5.24. An example of one realization in the setting 1–a is shown in Figure 5.23.

The dependence of the correlation function $\kappa(\boldsymbol{x}, \boldsymbol{x} + \boldsymbol{r})$ on the location $\boldsymbol{x}$ inside the window W for the first setting 1–a is shown in Figure 5.25. The figure suggests that the largest values occurs when $\boldsymbol{x} = \boldsymbol{x}_4 = (0, 0)$, which is the point with highest volume fraction and the lowest values occur when $\boldsymbol{x} = \boldsymbol{x}_1 = (5000, 5000)$ is on the boundary of W, i.e. for points with low values of the volume fraction. The relative difference

$$\Delta(r) = \frac{\max_{\boldsymbol{x} \in W}(\kappa(\boldsymbol{x}, \boldsymbol{x} + \boldsymbol{r})) - \min_{\boldsymbol{x} \in W}(\kappa(\boldsymbol{x}, \boldsymbol{x} + \boldsymbol{r}))}{\max_{\boldsymbol{x} \in W}(\kappa(\boldsymbol{x}, \boldsymbol{x} + \boldsymbol{r}))} \approx \Delta_{\kappa_4, \kappa_1}(r)$$

has a maximum that approximately equals 0.4. For the setting 1–b, which differs only in the strength of correlations, the result is analogous. The overall slope of the correlation function

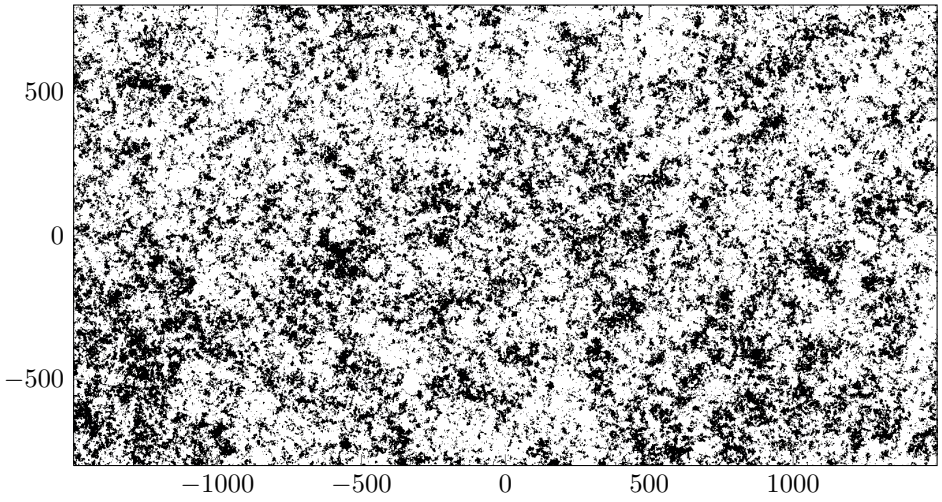


Figure 5.23: A part of one realization of $X_0(Z)$ for the setting 1–a.

Table 5.9: Minimal and maximal values of the volume fraction $m(\boldsymbol{x})$ and of the expectation $\mathbb{E}\,\hat{m}_{p;h_0}(\boldsymbol{x})$ of its kernel estimator $\hat{m}_{p;h_0}(\boldsymbol{x})$ for $h_0 = 2000$ in W. For a visualisation see Figure 5.24.

Setting	$\min_{\boldsymbol{x}\in W} m(\boldsymbol{x})$	$\max_{\boldsymbol{x}\in W} m(\boldsymbol{x})$	$\min_{\boldsymbol{x}\in W} \mathbb{E}\,\hat{m}_{p;h_0}(\boldsymbol{x})$	$\max_{\boldsymbol{x}\in W} \mathbb{E}\,\hat{m}_{p;h_0}(\boldsymbol{x})$
1–a, 1–b	0.100	0.300	0.111	0.192
2–a, 2–b	0.175	0.300	0.249	0.259

has a power law decay with coefficient 1.05 instead of 0.8 but the relative error remains the same. For settings 2–a and 2–b the result is similar with lower maximal value of the relative difference since the volume fraction profile is more flat inside W. The maximal obtained relative difference is 0.3 in both 2–a and 2–b. If one allows $\boldsymbol{x}$ to take arbitrary values then the maximal relative difference in settings 1–a and 1–b remains to be approximately 0.4, whereas in settings 2–a and 2–b it increases up to 0.6. This is a consequence of the fact that the volume fraction peak in settings 1–a, 1–b is relatively narrow and thus mostly contained in W. Hence, relaxing the condition $\boldsymbol{x} \in W$ does not lead to the possibility of reaching significantly smaller values of the volume fraction. On the other hand in settings 2–a, 2–b the volume fraction outside W still rapidly decreases to 0.05 as $\|\boldsymbol{x}\| \to \infty$. This leads to the increase of the relative difference between the correlation function $\kappa(\boldsymbol{x}, \boldsymbol{x}+\boldsymbol{r})$ with $\boldsymbol{x}$ at the centre of the peak and with $\boldsymbol{x}$ far from from it.

For each of the four settings we generated $N = 300$ realizations. For each realization the whole estimation process with initial bandwidth $h_0 = 2000$ was performed. Thus the estimates by $\hat{m}_{p;h_0}$, $\hat{\kappa}_{p;h_0}$, $\hat{\kappa}^{AI}_{p;h_0}$, $\hat{\boldsymbol{\theta}}_{h_0}$, and $\hat{\mathrm{cov}}_{h_0}$ were calculated. Then the minimisation of $\mathrm{MISE}_{h_0}(\hat{m}_h)$ was performed yielding the optimal bandwidth h_1. Finally, h_1 was used to calculate estimates by $\hat{m}_{p;h_1}$, $\hat{\kappa}_{p;h_1}$, $\hat{\kappa}^{AI}_{p;h_1}$, and $\hat{\boldsymbol{\theta}}_{h_1}$.

The expectation of the volume fraction estimator $\hat{m}_h$ is according to (4.26) given by the convolution of the true volume fraction m with the Gaussian kernel K_h. The expectation of

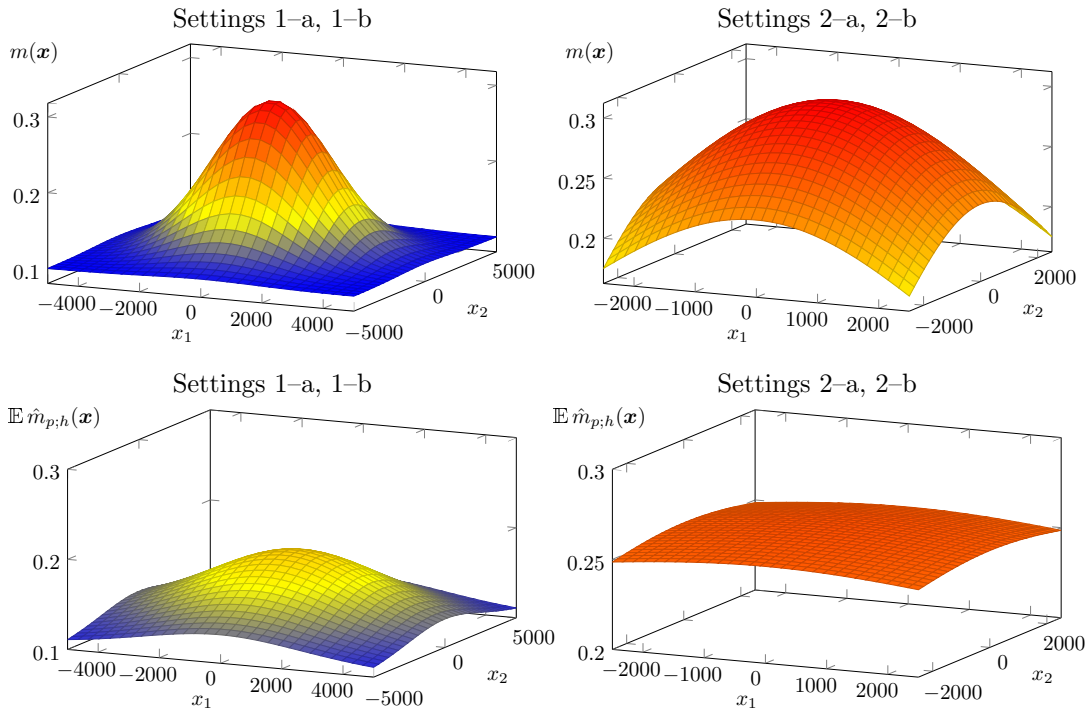


Figure 5.24: A depiction of the true volume fraction $m(\boldsymbol{x})$ in simulated settings. (a) corresponds to (5.23) with $A = 0.2,\ B = 0.1,\ \sigma = 1800$ and (b) with $A = 0.25,\ B = 0.05,\ \sigma = 3000$. The expectation $\mathbb{E}\,\hat{m}_{p;h_0}$ of the kernel estimator $\hat{m}_{p;h_0}$ given by (5.26) with $h_0 = 2000$ is plotted in (c) for the first case and in (d) for the second case. For extremal values in all four settings see Table 5.9.

the discrete version $\hat{m}_{p;h}$ equals

$$\mathbb{E}\,\hat{m}_{p;h}(\boldsymbol{x}) = \frac{\sum_{\boldsymbol{i}=(1,1)}^{\boldsymbol{n}} m(a\boldsymbol{i}+\boldsymbol{c})K_h(a\boldsymbol{k}-a\boldsymbol{i})}{\sum_{\boldsymbol{i}=(1,1)}^{\boldsymbol{n}} K_h(a\boldsymbol{k}-a\boldsymbol{i})} \approx \frac{1}{e_h(\boldsymbol{x})}\int_W m(\boldsymbol{u})K_h(\boldsymbol{x}-\boldsymbol{u})\,\mathrm{d}\boldsymbol{u} = \mathbb{E}\,\hat{m}_h. \quad (5.26)$$

Since W is a rectangular window then by (5.15)

$$e_h(\boldsymbol{x}) = \big(\Phi_h(u_1+x_1) - \Phi_h(v_1+x_1)\big)\big(\Phi_h(u_2+x_2) - \Phi_h(v_2+x_2)\big),$$

where $\boldsymbol{v} = (v_1, v_2)$ is the lower left corner of W, $\boldsymbol{u} = (u_1, u_2)$ is the upper right corner of W and Φ_h is the cumulative distribution function of $N(0, h)$. The expectation $\mathbb{E}\,\hat{m}_{p;h_0}$ for all settings is shown in Figure 5.24. The extreme values of $m(\boldsymbol{x})$ and $\mathbb{E}\,\hat{m}_{p;h_0}(\boldsymbol{x})$ inside W are summarized in Table 5.9. The flattening effect of the convolution is clearly visible. The variance of $\hat{m}_{p;h_0}$ depends on the second order structure of $X_0(Z)$ similarly to (4.28), which holds for the variance of $\hat{m}_{h_0}$. We estimate its value by calculating the sample variance, which is according to (4.34) defined by

$$s_N^2(\hat{m}_{p;h_0})(\boldsymbol{x}) = \frac{1}{N}\sum_{i=1}^{N}\big(\hat{m}_{i;p;h_0}(\boldsymbol{x}) - \mathbb{E}\,\hat{m}_{p;h_0}(\boldsymbol{x})\big)^2,$$

where $\hat{m}_{i;p;h_0}(\boldsymbol{x})$ denotes the result of the estimation on the i-th realisation. The estimator of the standard deviation is then defined by $s_N(\hat{m}_{p;h_0})(\boldsymbol{x}) = \sqrt{s_N^2(\hat{m}_{p;h_0})(\boldsymbol{x})}$. The estimated values of the standard deviation for our settings and $h_0 = 2000$ are shown in Figure 5.26. High

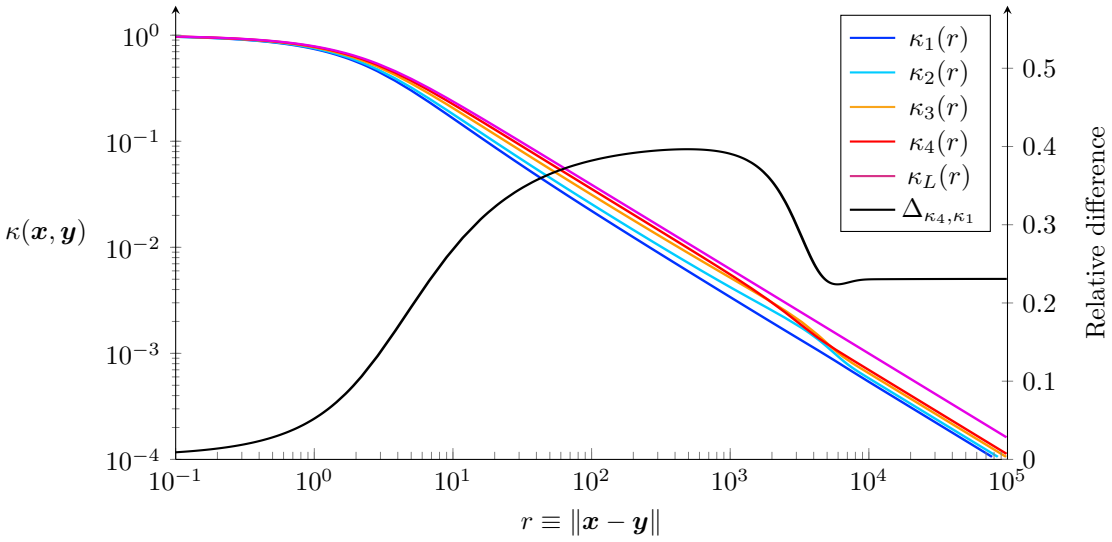


Figure 5.25: The dependence of the correlation function $\kappa(\boldsymbol{x},\boldsymbol{x}+r\boldsymbol{e}_y)$ of $X_0(Z)$ on r for different values of $\boldsymbol{x} \in W$ in the first setting 1–a, where $\boldsymbol{e}_y = (0,1)$ and $\kappa_i(r) = \kappa(\boldsymbol{x}_i, \boldsymbol{x}_i + r\boldsymbol{e}_y)$. The points are $\boldsymbol{x}_1 = (5000, 5000)$, $\boldsymbol{x}_2 = (-2000, -3000)$, $\boldsymbol{x}_3 = (-2000, -2000)$, and $\boldsymbol{x}_4 = (0,0)$. The correlation function $\kappa_L(r)$ of a stationary standard Gaussian 0-level excursion set given by (4.40) with $\alpha = 1.8$, $\beta = 0.8$, $\theta = 3$ is plotted for the comparison. Finally, the relative difference $\Delta_{\kappa_4,\kappa_1}(r) = \frac{\kappa_4(r)-\kappa_1(r)}{\kappa_4(r)}$ is superimposed with axis on the right side of the figure.

values occur at locations where the individual estimators $\hat{m}_{i;p;h_0}(\boldsymbol{x})$ differ significantly from the expectation $\mathbb{E}\,\hat{m}_{p;h_0}(\boldsymbol{x})$. We may observe several properties. First, as can be expected, the deviations are larger for larger values of $\mathbb{E}\,\hat{m}_{p;h_0}(\boldsymbol{x})$. Second, the deviations are larger for stronger correlation dependence. Therefore settings 1–b and 2–b have larger deviations then 1–a and 2–a, respectively. Third, the deviations are generally larger for smaller observation window. Finally, the deviations are larger for $\boldsymbol{x}$ close to the boundary of the observation window.

Now we focus on the analysis of correlation function estimators $\hat{\kappa}_{p;h_0}$, based on the initial bandwidth h_0, and $\hat{\kappa}_{p;h_1}$, based on the optimal bandwidth h_1 obtained by minimising $\mathrm{MISE}_{h_0}(\hat{m}_h)$. As was already discussed, the correlation function $\kappa(\boldsymbol{x},\boldsymbol{y})$ of $X_0(Z)$ depends on both $\boldsymbol{x}$ and $\boldsymbol{r} = \boldsymbol{y} - \boldsymbol{x}$. However, this is not reflected by estimators $\hat{\kappa}_{p;h_0}(\boldsymbol{r})$ and $\hat{\kappa}_{p;h_1}(\boldsymbol{r})$ because they are actually estimating the spatial average $\overline{\kappa}_W(\boldsymbol{r})$ of $\kappa(\boldsymbol{x},\boldsymbol{x}+\boldsymbol{r})$ over the window $W \cap (W-\boldsymbol{r})$ defined by

$$\overline{\kappa}_W(\boldsymbol{r}) = \frac{1}{\nu_2\big(W \cap (W-\boldsymbol{r})\big)} \int\limits_{W\cap(W-\boldsymbol{r})} \kappa(\boldsymbol{x},\boldsymbol{x}+\boldsymbol{r})\,\mathrm{d}\boldsymbol{x}$$

as follows from relation (4.30) and discussion preceding it. The sample means defined by (4.35) and estimates of 90% confidence bounds of $\hat{\kappa}_{p;h_0}$ and of $\hat{\kappa}_{p;h_1}$ along one concrete direction of $\boldsymbol{r}$ together with superimposed spatially averaged values of $\overline{\kappa}_W(\boldsymbol{r})$ are shown in Figure 5.27 for h_0 and in Figure 5.29 for h_1. The estimates $s_N(\hat{\kappa}_{p;h_0})$ and $s_N(\hat{\kappa}_{p;h_1})$ of standard deviations of $\hat{\kappa}_{p;h_0}$ and $\hat{\kappa}_{p;h_1}$ defined by (4.36) are shown in Figures 5.31 and 5.32, respectively.

One can see that the estimators show typical deviations, which were observed already in the stationary situation in Section 4.4 and also for the real built-up structure in this section. The values of the spatially averaged correlation function $\overline{\kappa}_W(\boldsymbol{r})$ are almost always inside estimated confidence bounds. For small and medium values of r they are also close to values of the sample

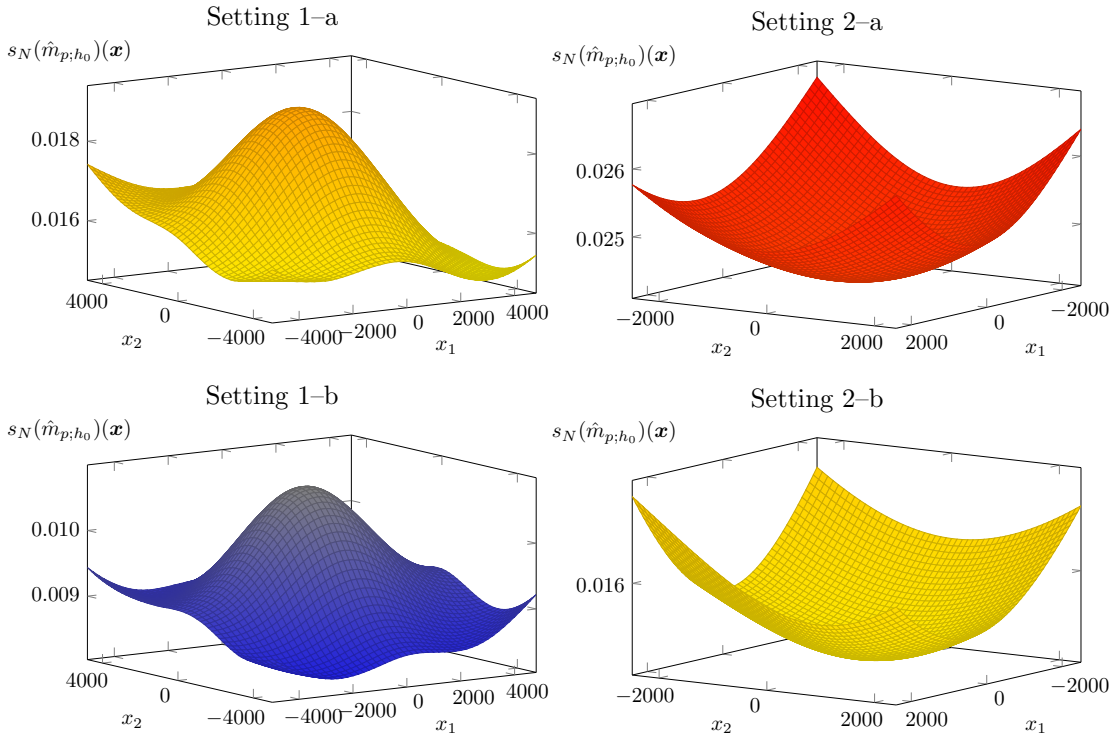


Figure 5.26: A depiction of estimated standard deviations $s_N(\hat{m}_{p;h_0})$ for all settings.

mean. This is especially true for $\hat{\kappa}_{p;h_1}$, where the improvement against $\hat{\kappa}_{p;h_0}$ is clearly visible in settings 1–b and 2–a.

For the isotropic adapted estimators $\hat{\kappa}^{AI}_{p;h_0}$ and $\hat{\kappa}^{AI}_{p;h_1}$ the sample mean and sample standard deviations are again analysed. Both estimators should estimate the isotropised spatially averaged correlation function $\overline{\kappa}^I_W$ given by averaging $\kappa(\boldsymbol{x}, \boldsymbol{x}+\boldsymbol{r})$ over the window $W \cap (W - \boldsymbol{r})$ and over all directions, i.e.

$$\overline{\kappa}^I_W(r) = \frac{1}{2\pi} \int_0^{2\pi} \overline{\kappa}_W\big(r \cdot (\cos\phi, \sin\phi)\big)\, \mathrm{d}\phi.$$

The sample mean together with 90% confidence bounds and $\overline{\kappa}^I_W$ are shown in Figure 5.28 for $\hat{\kappa}^{AI}_{p;h_0}$ and in Figure 5.30 for $\hat{\kappa}^{AI}_{p;h_1}$. Again there is a significant improvement of the estimator based on h_1 instead of h_0, especially in midscale ranges of settings 1–b and 2–a. The estimates of the standard deviations are plotted in Figures 5.31 and 5.32.

The result of estimations by $\hat{\kappa}_{p;h_1}$ and $\hat{\kappa}^{AI}_{p;h_1}$ based on h_1 for one realization in the setting 1–a is shown in Figure 5.33. By comparison to Figures 5.27 (a) and 5.28 (a) one can see the clear qualitative similarity to the expected situation.

Finally, we study the performance of $\hat{\boldsymbol{\theta}}_h = (\hat{\alpha}_h, \hat{\beta}_h, \hat{\theta}_h)$, which is an estimator of parameters of the fit of the values of $\hat{\kappa}_{p;h}$ by the parametric correlation function $\kappa_L(r;\boldsymbol{\theta})$. Based on Figure 5.25, we may assume that the estimator should be able to find the true value β and possibly also α of parameters of the underlying Gaussian random field Z. We may also expect underestimation of θ comparing to the true value for Z, since the lower value of θ pushes the corresponding curve $\kappa_L(r;\boldsymbol{\theta})$ in Figure 5.25 down leading to better intermediate approximation. In Table 5.10 the estimated statistical properties of $\hat{\boldsymbol{\theta}}_{h_0}$ and $\hat{\boldsymbol{\theta}}_{h_1}$ in all four settings are summarized.

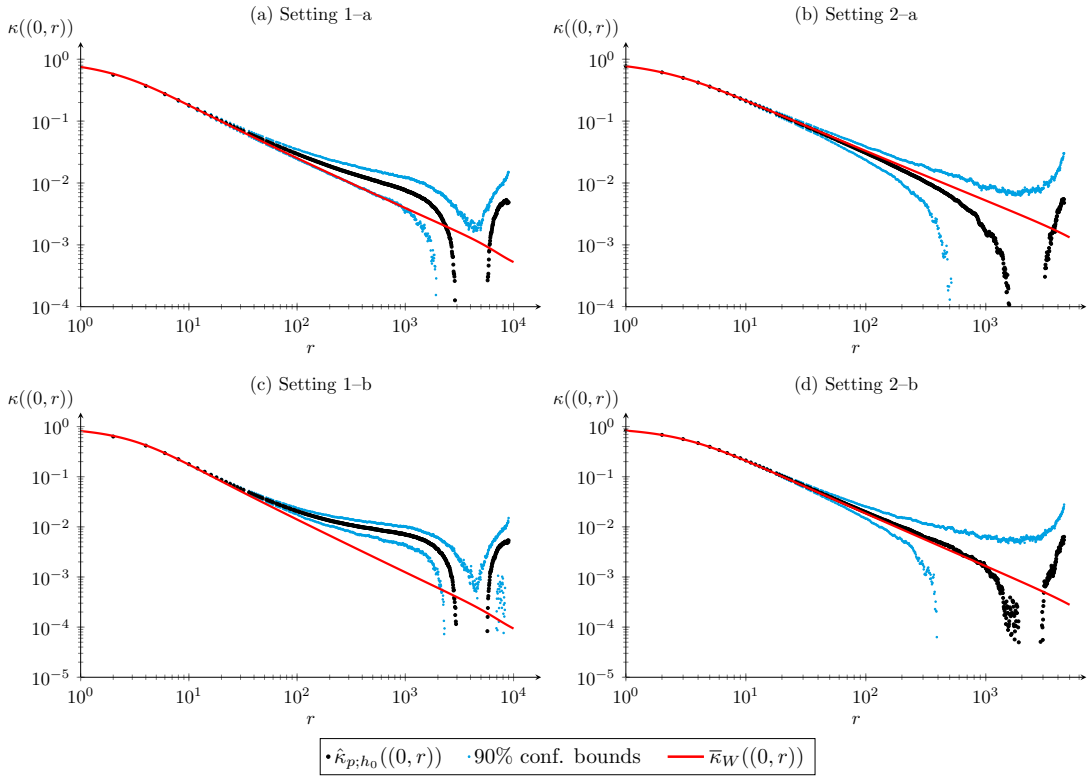


Figure 5.27: The sample mean of $\hat{\kappa}_{p;h_0}(\boldsymbol{r})$ along $\boldsymbol{r} = (0,r)$ with estimated two-sided 90% confidence bounds and the theoretical value $\overline{\kappa}_W(\boldsymbol{r})$.

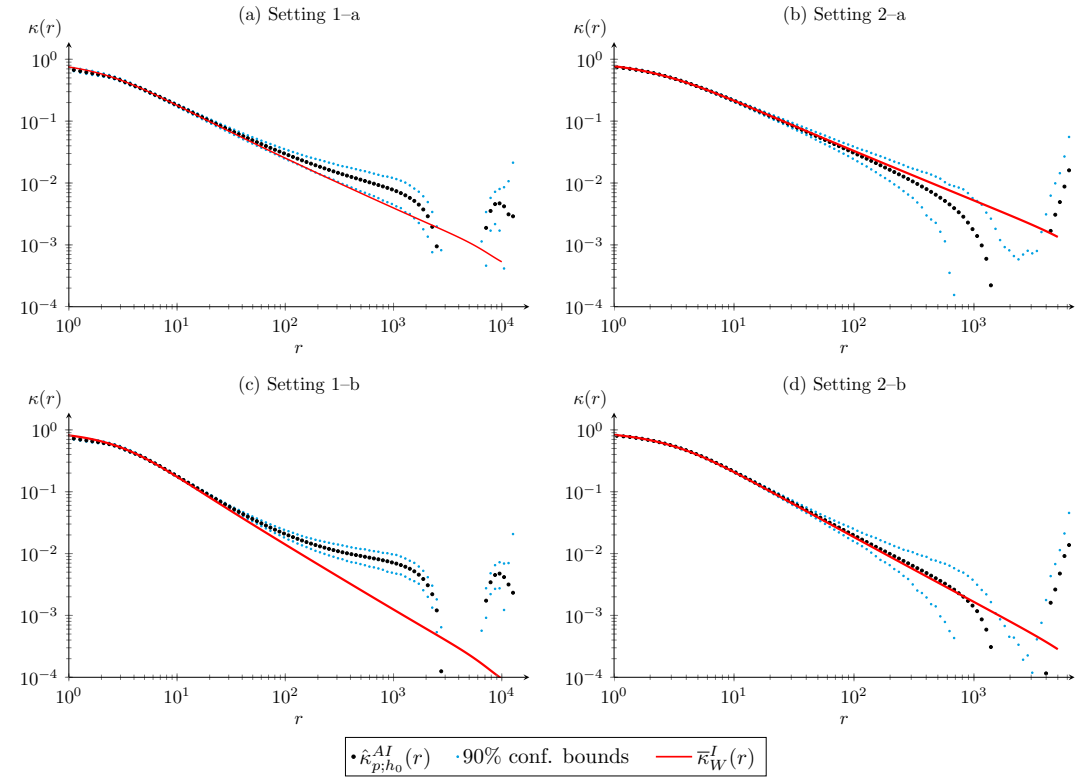


Figure 5.28: The sample mean of $\hat{\kappa}^{AI}_{p;h_0}(r)$ with estimated two-sided 90% confidence bounds and the theoretical value $\overline{\kappa}^{I}_W(r)$.

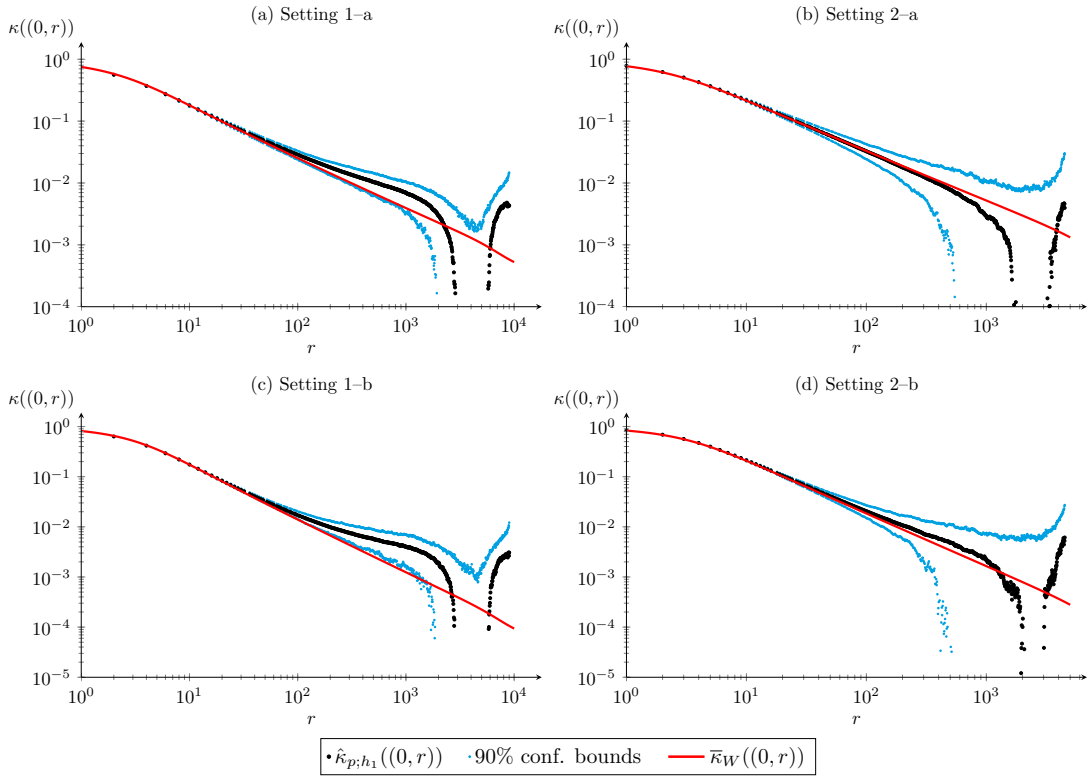


Figure 5.29: The sample mean of $\hat{\kappa}_{p;h_1}(\boldsymbol{r})$ along $\boldsymbol{r} = (0,r)$ with estimated two-sided 90% confidence bounds and the theoretical value $\overline{\kappa}_W(\boldsymbol{r})$.

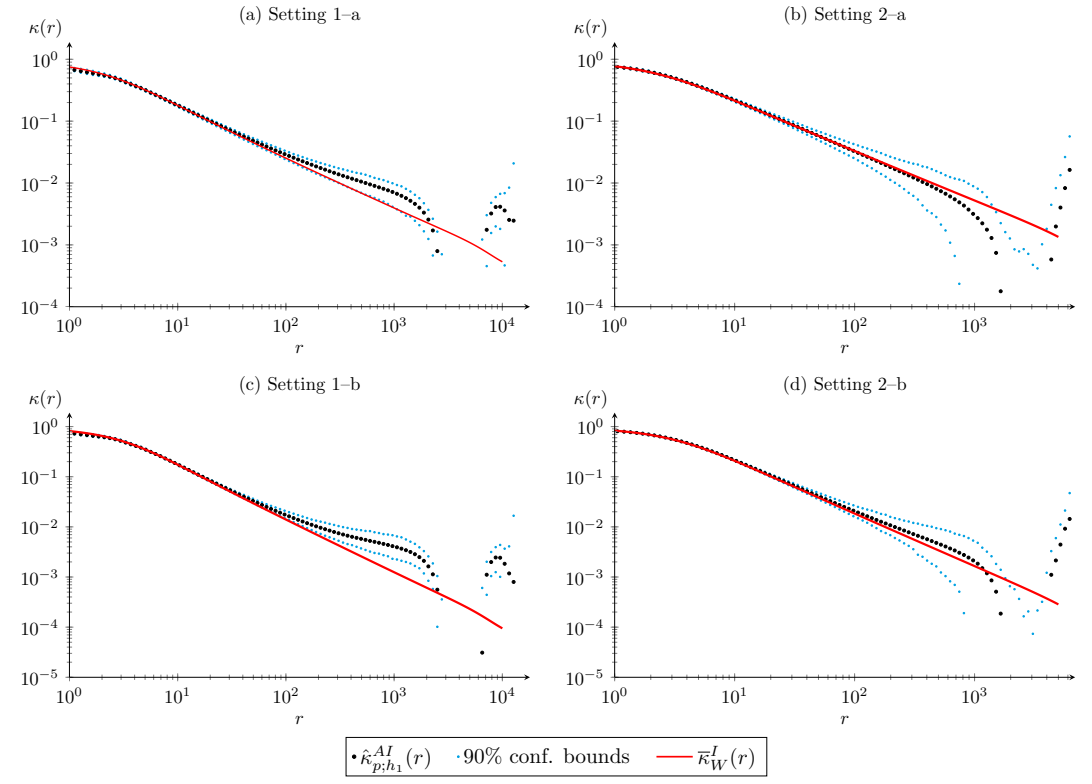


Figure 5.30: The sample mean of $\hat{\kappa}^{AI}_{p;h_1}(r)$ with estimated two-sided 90% confidence bounds and the theoretical value $\overline{\kappa}^{I}_W(r)$.

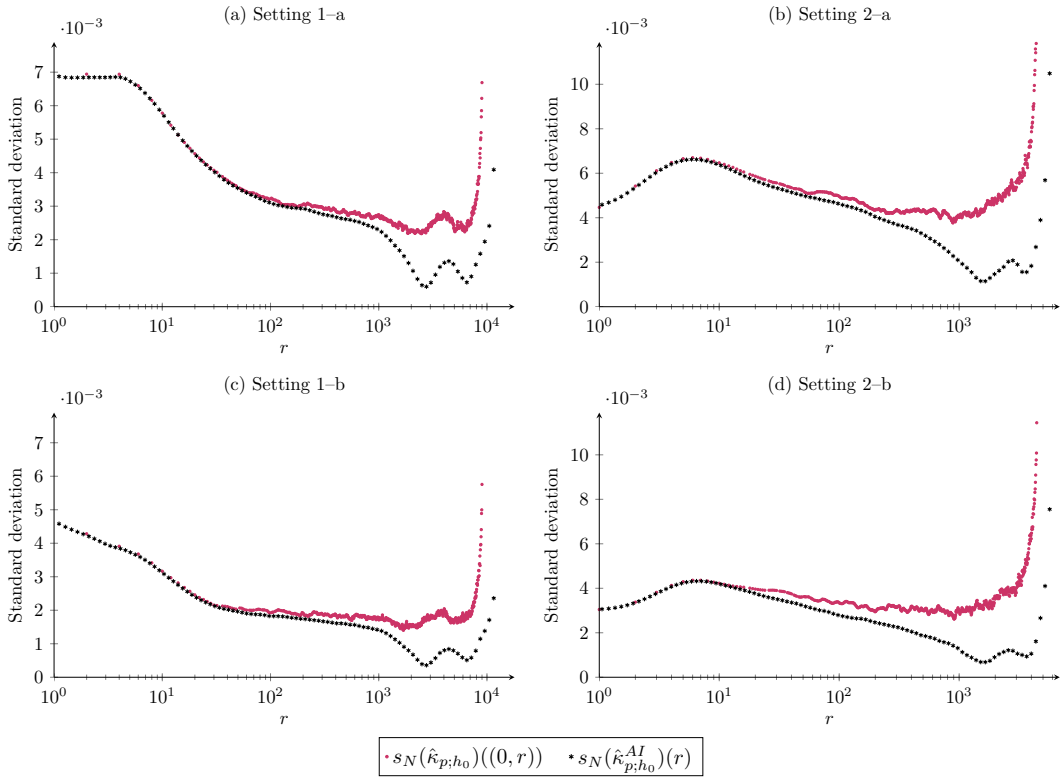


Figure 5.31: Estimated standard deviations of $\hat{\kappa}_{p;h_0}(\boldsymbol{r})$ along $\boldsymbol{r} = (0, r)$ and of $\hat{\kappa}^{AI}_{p;h_0}(r)$.

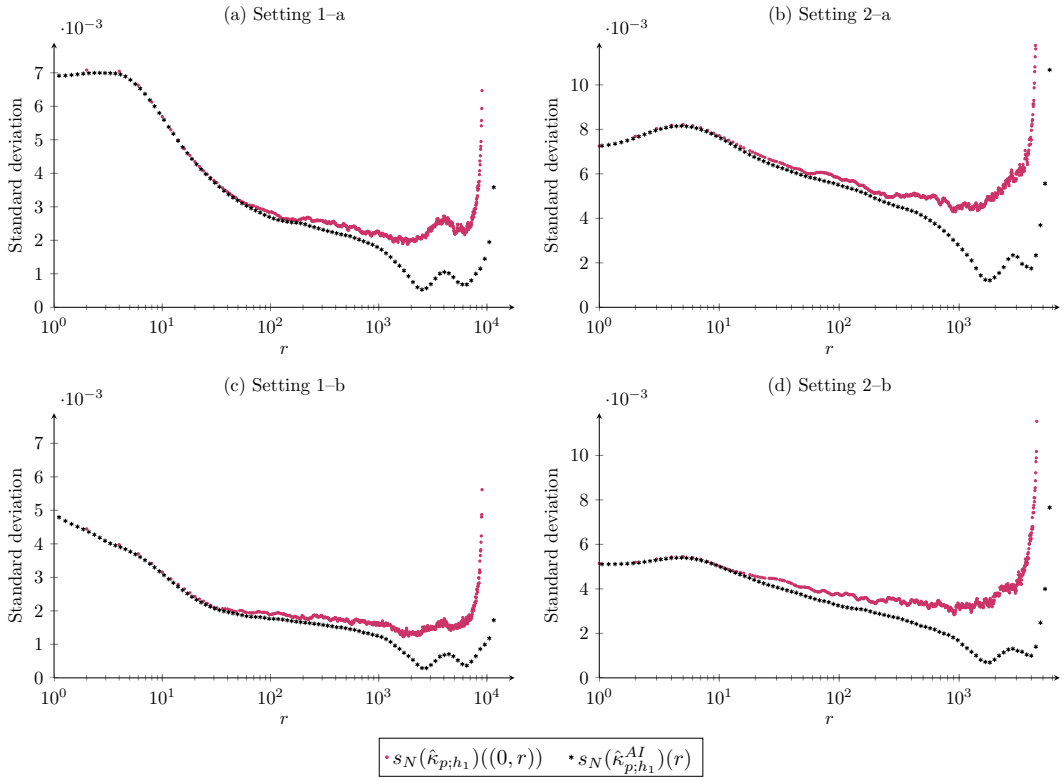


Figure 5.32: Estimated standard deviations of $\hat{\kappa}_{p;h_1}(\boldsymbol{r})$ along $\boldsymbol{r} = (0, r)$ and of $\hat{\kappa}^{AI}_{p;h_1}(r)$.

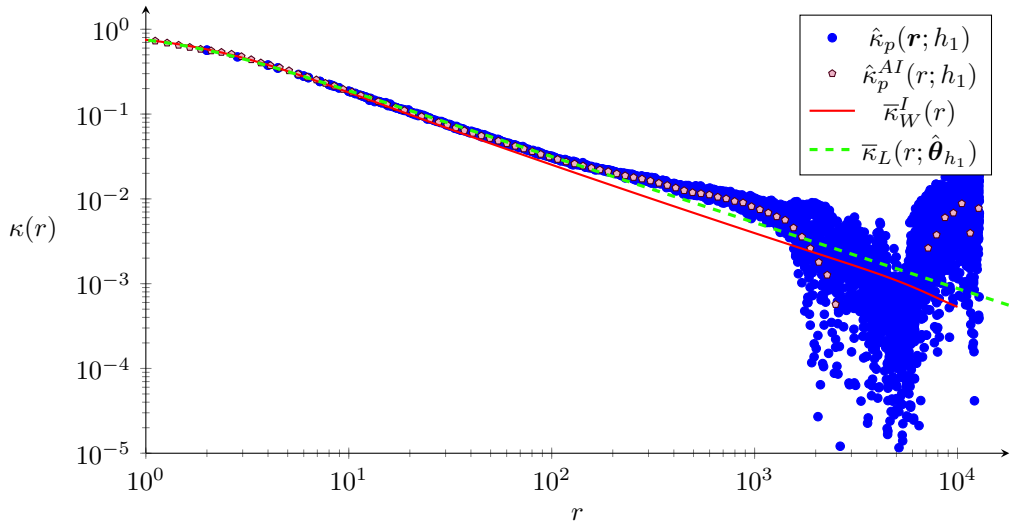


Figure 5.33: A result of correlation function estimators based on h_1 calculated for one realization in the setting 1–a. Here $h_1 = 2014$ and $\hat{\alpha}_{h_1} = 2.00, \hat{\beta}_{h_1} = 0.78, \hat{\theta}_{h_1} = 2.10$.

Table 5.10: The summary of estimated properties of $\hat{\boldsymbol{\theta}}_{h_0} = (\hat{\alpha}_{h_0}, \hat{\beta}_{h_0}, \hat{\theta}_{h_0})$ based on h_0 and of $\hat{\boldsymbol{\theta}}_{h_1}$ based on h_1: $\overline{\hat{\alpha}}_{h_0}$ is the sample mean; $I_{0.9}(\hat{\alpha}_{h_0})$ the 90% confidence interval; $d_N(\hat{\alpha}_{h_0}, \alpha)$ the deviation from the true value defined by (4.37); $s_N(\hat{\alpha}_{h_0})$ the sample standard deviation defined by (4.36); and analogously for β, θ, and h_1.

Setting	α	$\overline{\hat{\alpha}}_{h_0}$	$I_{0.9}(\hat{\alpha}_{h_0})$	$d_N(\hat{\alpha}_{h_0}, \alpha)$	$s_N(\hat{\alpha}_{h_0})$	$\overline{\hat{\alpha}}_{h_1}$	$I_{0.9}(\hat{\alpha}_{h_1})$	$d_N(\hat{\alpha}_{h_1}, \alpha)$	$s_N(\hat{\alpha}_{h_1})$
1–a	1.80	1.99	[1.91, 2.00]	0.19	0.04	1.98	[1.87, 2.00]	0.18	0.05
2–a	1.80	1.58	[1.47, 1.73]	0.23	0.08	1.68	[1.50, 1.90]	0.17	0.12
1–b	2.00	2.00	[2.00, 2.00]	0.00	0.00	2.00	[2.00, 2.00]	0.00	0.00
2–b	2.00	1.99	[1.92, 2.00]	0.03	0.03	1.99	[1.95, 2.00]	0.02	0.02

	β	$\overline{\hat{\beta}}_{h_0}$	$I_{0.9}(\hat{\beta}_{h_0})$	$d_N(\hat{\beta}_{h_0}, \beta)$	$s_N(\hat{\beta}_{h_0})$	$\overline{\hat{\beta}}_{h_1}$	$I_{0.9}(\hat{\beta}_{h_1})$	$d_N(\hat{\beta}_{h_1}, \beta)$	$s_N(\hat{\beta}_{h_1})$
1–a	0.80	0.79	[0.72, 0.86]	0.04	0.04	0.81	[0.76, 0.88]	0.04	0.04
2–a	0.80	0.94	[0.83, 1.05]	0.16	0.07	0.90	[0.78, 1.04]	0.13	0.08
1–b	1.05	0.93	[0.86, 0.99]	0.13	0.04	1.03	[0.94, 1.10]	0.05	0.05
2–b	1.05	1.08	[0.98, 1.17]	0.06	0.06	1.06	[0.95, 1.16]	0.06	0.06

	θ	$\overline{\hat{\theta}}_{h_0}$	$I_{0.9}(\hat{\theta}_{h_0})$	$d_N(\hat{\theta}_{h_0}, \theta)$	$s_N(\hat{\theta}_{h_0})$	$\overline{\hat{\theta}}_{h_1}$	$I_{0.9}(\hat{\theta}_{h_1})$	$d_N(\hat{\theta}_{h_1}, \theta)$	$s_N(\hat{\theta}_{h_1})$
1–a	3.00	2.09	[1.86, 2.33]	0.92	0.14	2.13	[1.94, 2.40]	0.88	0.14
2–a	3.00	3.55	[2.94, 4.18]	0.67	0.38	3.35	[2.72, 4.03]	0.53	0.41
1–b	4.00	2.71	[2.48, 2.95]	1.30	0.14	3.02	[2.74, 3.24]	1.00	0.15
2–b	4.00	3.82	[3.49, 4.15]	0.27	0.20	3.79	[3.45, 4.10]	0.29	0.20

Table 5.11: Basic characteristics of h_1: $\overline{h}_1$ is the sample mean; $I_{0.9}(h_1)$ is the 90% confidence interval; $s_N(h_1)$ is the sample standard deviation.

Setting	$\overline{h}_1$	$I_{0.9}(h_1)$	$s_N(h_1)$
1–a	1934	$[1731, 2125]$	116
1–b	1596	$[1512, 1705]$	60
2–a	7667	$[2288, 30416]$	8059
2–b	3606	$[2124, 5778]$	2985

One can see the clear benefit of using h_1 instead of just h_0. In almost all analysed settings deviations from true values of estimators based on h_1 are smaller then deviations of estimators based on h_0. The only difference is for θ in the setting 2–b where $d_N(\hat{\theta}_{h_1}, \theta)$ is slightly larger then $d_N(\hat{\theta}_{h_0}, \theta)$. The benefit may be also visualised by comparing mean based correlation functions $\kappa_L(r; \hat{\boldsymbol{\theta}}_{h_0})$ and $\kappa_L(r; \hat{\boldsymbol{\theta}}_{h_1})$ with theoretical values of the correlation function $\kappa(\boldsymbol{x}, \boldsymbol{x}+\boldsymbol{r})$ for different choices of $\boldsymbol{x}$ and $\boldsymbol{r}$. This is presented in Figure 5.34. One can see that the fit based on h_1 is generally much closer to the strip of real values then the fit based on h_0.

Since h_1 is by construction a random variable representing an estimator of the true optimal bandwidth we may analyse its statistical properties. In Table 5.11 the basic characteristics of h_1 are summarized. Based on the results the following observation may be pointed out. The variability of h_1 is much larger for smaller window in settings 2–a, 2–b and especially in the setting 2–a, where the correlation is stronger, which is consistent with larger standard deviations of $\hat{\kappa}_{p;h_1}$, $\hat{\kappa}^{AI}_{p;h_1}$ in those settings as can be seen by comparing Figures 5.31 and 5.32.

As a final conclusion we may say that generally, there is a significant improvement in the performance of estimators based on the optimal bandwidth h_1 comparing to estimators based on the initial bandwidth h_0. This particularly holds for the estimator $\hat{\beta}$ of β which is capable of successful estimation of the value of the power law decay coefficient of the correlation function. Such a result may be viewed as a partial confirmation of the validity of the proposed method of obtaining the optimal bandwidth and estimators based on it. Moreover it also shows that the approximation of the covariance function given by (5.20) leading to the estimator (5.25) is useful.

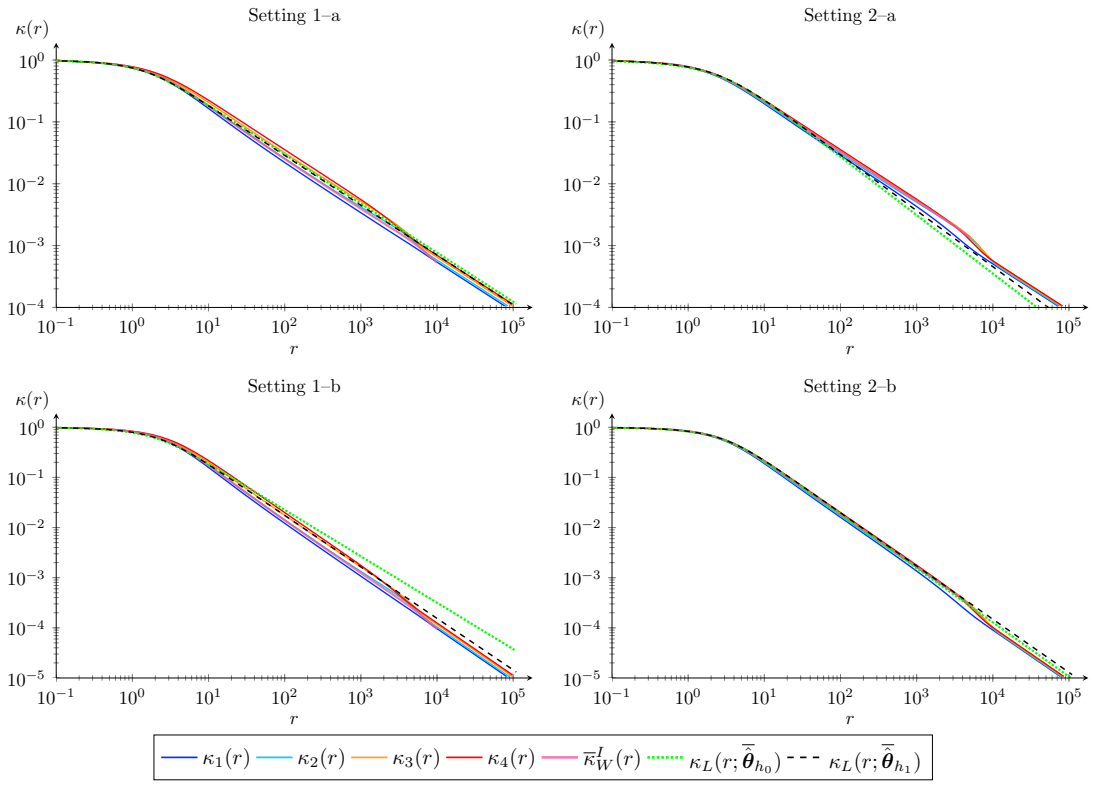


Figure 5.34: Correlation functions $\kappa_L\big(r;\overline{\hat{\boldsymbol{\theta}}}_{h_0}\big)$ and $\kappa_L\big(r;\overline{\hat{\boldsymbol{\theta}}}_{h_1}\big)$ defined by (5.24) and determined by sample mean parameters $\overline{\hat{\boldsymbol{\theta}}}_{h_0}$ and $\overline{\hat{\boldsymbol{\theta}}}_{h_1}$, respectively. For comparison, $\overline{\kappa}^I_W(r)$ and $\kappa_i(r) = \kappa(\boldsymbol{x}_i, \boldsymbol{x}_i + r\boldsymbol{e}_y)$ for $i = 1, \ldots, 4$, where $\boldsymbol{e}_y = (0, 1)$, are also shown. The points $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_4$ in settings 1–a and 1–b are chosen to be $\boldsymbol{x}_1 = (5000, 5000)$, $\boldsymbol{x}_2 = (-2000, -3000)$, $\boldsymbol{x}_3 = (-2000, -2000)$, $\boldsymbol{x}_4 = (0, 0)$ and in settings 2–a and 2–b they are chosen to be $\boldsymbol{x}_1 = (2500, 2500)$, $\boldsymbol{x}_2 = (-1700, -2500)$, $\boldsymbol{x}_3 = (0, -2000)$, $\boldsymbol{x}_4 = (0, 0)$.

5.5 Discussion of random closed set approach results

In the previous two sections the built-up area of certain cities was analysed by studying second order properties of its representation by random closed sets. In Section 5.3 the built-up area was taken as a realization of a stationary random closed set. The second order properties were analysed using two methods. The first is the direct estimation of the correlation function $\kappa(r)$. We found that under the assumption of isotropy the correlation function can be well fitted inside the parametric class of Cauchy correlation functions $\kappa_C(r, \boldsymbol{\theta})$, where $\boldsymbol{\theta} = (\alpha, \beta, \theta)$ is the vector of parameters. The Cauchy correlation function has the power law asymptotic decay determined by the coefficient β. Thus, by fitting of the estimated values of the correlation function we were able to obtain the estimator $\hat{\boldsymbol{\theta}}$ of the parameters and in particular the estimator $\hat{\beta}$ of β, which is from now on denoted by $\hat{\beta}_s$.

The second method used in the stationary scenario was the estimation of the variance in balls $\operatorname{var}\big(\nu_X(B_r)\big)$. We derived that if the correlation function has a power law decay with coefficient β then the dependence of $\operatorname{var}\big(\nu_X(B_r)\big)$ on the ball radius r follows a power law with coefficient $4 - \beta$. To obtain an estimator of β we thus analyse the power law behaviour of the estimator $\hat{\operatorname{var}}_v\big(\nu_X(B_r)\big)$ of the variance in balls on r using least squares log-linear fitting which yields the estimator $\hat{\beta}$, which is in the following denoted by $\hat{\beta}_l$.

Table 5.12: Results of estimations of the coefficient β.

City	$\hat{\beta}_s$	$\hat{\beta}_v$	$\hat{\beta}_{h_0}$	$\hat{\beta}_{h_1}$
Boston	1.01	0.85	0.68	0.67
Chicago	1.32	1.11	0.62	0.58
Los Angeles	1.33	1.07	0.67	0.64
Pittsburgh	1.05	0.61	0.72	0.69
Seattle	0.78	0.40	0.53	0.52
Berlin	0.84	0.67	0.81	0.79
Birmingham	0.78	0.56	0.80	0.79
Milan	0.84	0.70	0.77	0.78
Minsk	1.10	1.16	0.82	0.78
Moscow	1.07	1.02	0.93	0.91
Oslo	0.89	0.28	0.65	0.61
Paris	0.77	1.05	0.88	0.88
Prague	0.65	0.42	0.77	0.80
St. Petersburg	1.00	0.95	0.83	0.82

In Section 5.4 the built-up area was studied as a realization of a non-stationary random closed set. Here the non-constant volume fraction was first estimated using the kernel estimator $\hat{m}_{h_0}$ for the initial bandwidth h_0. Then, second order properties were analyse by the direct estimation of the correlation function $\kappa(\boldsymbol{x}, \boldsymbol{y})$ under the assumption that it depends only on the difference $\boldsymbol{x} - \boldsymbol{y}$. We saw that under the additional assumption of isotropy the best fit of the estimated correlation function is again given by the Cauchy correlation function. Thus, we obtain the estimator $\hat{\boldsymbol{\theta}}_{h_0}$ of its parameters and in particular the estimator $\hat{\beta}_{h_0}$ of the power law decay coefficient β. Next, both the volume fraction estimate and correlation function estimate were used to minimise the mean integrated square error $\mathrm{MISE}_{h_0}(\hat{m}_h)$ and consequently to obtain the bandwidth h_1 that was taken as the optimal bandwidth. Based on h_1 the estimation of the volume fraction and of the correlation function was again performed. This after fitting by the Cauchy correlation function leads to the estimator $\hat{\boldsymbol{\theta}}_{h_1}$ and hence to $\hat{\beta}_{h_1}$ as the estimator of β.

Based on results of all three previously recapitulated approaches we conjecture that the correlation function is asymptotically governed by the power law behaviour, i.e.

$$\kappa(\boldsymbol{x}, \boldsymbol{y}) \sim C \, \|\boldsymbol{x} - \boldsymbol{y}\|^{-\beta} \qquad (r \to \infty).$$

Hence the built up structure as a random closed set exhibits isotropic long-range dependence according to Definition 2.8.2. The overall summary of estimated values of β is given in Table 5.12. We see that the values are for some cities rather different. From the numerical analysis performed in Section 4.4 follows that the estimator $\hat{\beta}_s$ has a tendency to overestimate the true values of β. On the other hand the numerical analysis in Subsection 5.3.2 suggests that $\hat{\beta}_l$ has a tendency to underestimate the true values. Both tendencies are observable in Table 5.12 for all cities except Minsk and Paris. It is also interesting to compare estimators $\hat{\beta}_{h_0}$ and $\hat{\beta}_{h_1}$. From numerical simulations in Subsection 5.4.4 follows that $\hat{\beta}_h$ returns smaller values for larger h, which is consistent with results in Table 5.6 (repeated in the above table) corresponding to the built-up area. However, as follows from numerical simulations the more precise estimate is always given by $\hat{\beta}_{h_1}$.

In the overall comparison the least reliable values are given by $\hat{\beta}_v$ since the variance in balls was estimated for only a limited range of radii. On the other hand $\hat{\beta}_{h_1}$ should be preferable since much larger window than in the other methods was used for its determination. The only

possible weakness of the estimator $\hat{\beta}_{h_1}$ is the fact that it is based on the approximation in which the correlation function $\kappa(\boldsymbol{x}, \boldsymbol{y})$ depends only on $\|\boldsymbol{x} - \boldsymbol{y}\|$. This property was not observed in any non-stationary random closed set model in Chapter 2. However, from simulations in Subsection 5.4.4 follows that this may not be a problem since the general underlying behaviour of the correlation function, and particularly its power law decay, is well reflected by all used correlation function estimators based on that approximation.

It is worth to discuss the possible confounding between the density and interaction which is a well known phenomenon in the point process analysis. It was shown (see [82] and references therein) that for point processes it is sometimes not possible to recognize the difference of spatial inhomogeneities from clustering in a single realization of a point process. In the theory of point processes it represents a fundamental limitation of the scope of statistical inference. The natural question is whether this kind of misinterpretation may not be presented in our analysis. In that case, when applied in the same observation window, the stationarity based estimators compared to non-stationarity ones should produce larger correlations due to the influence of the non-constant volume fraction and consequently of smaller values of the power law coefficient β. In our analysis the non-stationarity based estimators are based on much larger windows than the stationary ones, with more rapid volume fraction changes so that the effect may be partially neglected. However, as can be seen in Table 5.12, the values of $\hat{\beta}_{h_1}$ estimated under the non-stationarity assumption are comparable and often even smaller than values of $\hat{\beta}_s$ and $\hat{\beta}_v$ obtained under stationarity assumption. From this we conclude that the confounding is not presented in the built-up structure and thus the long-range correlations are a vital part of the structure and not any misinterpretation of the influence of the non-constant density. Moreover we conclude that the non-stationary approach is more suitable for the analysis of large built-up areas whereas the stationary approach should be limited only to neighbourhoods of true city centres.

Finally note that all long-range dependent models used for numerical simulations in Subsections 4.4.2, 5.3.2, and 5.4.4 were Gaussian level excursion sets. It should be mentioned that the use this particular model is not new in connection to urban structures. Makse et al in [32, 33] introduced a correlated percolation model which exactly corresponds to the digitalized level excursion set on a regular grid of points. They used it for simulation of the urban structure on large scales. However, they obtain the best results for strongly correlated case with $\beta = 0.05$, which is a rather different value from our results.

5.6 Radial density analysis

In this section the radial dependence of the volume fraction is studied. We again take a built-up area in a given city as a realization of the non-stationary random closed set X observed in some observation window W. Let further ν_X be the volume measure of X given by $\nu_X(B) = \nu(X \cap B)$ for all Borel $B \subset \mathbb{R}^2$, where ν is the Lebesgue measure on $\mathbb{R}^2$. In the following we are going to analyse the dependence of

$$\hat{m}_{r;\boldsymbol{x}_0} = \frac{\nu_X(B_r(\boldsymbol{x}_0) \cap W)}{\nu(B_r(\boldsymbol{x}_0) \cap W)}$$

on r, where $\boldsymbol{x}_0$ is a suitably chosen centre point of the city. The values of r are taken such that $0 < \nu(B_r \cap W) < \nu(W)$, i.e. $B_r(\boldsymbol{x}_0)$ is included in W but does not fully contain W. The expectation is given by

$$\mathbb{E}\,\hat{m}_{r;\boldsymbol{x}_0} = \frac{1}{\nu(B_r(\boldsymbol{x}_0) \cap W)} \int\limits_{B_r(\boldsymbol{x}_0)} 1_W(\boldsymbol{x}) m(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} = \frac{\Lambda(B_r(\boldsymbol{x}_0) \cap W)}{\nu(B_r(\boldsymbol{x}_0) \cap W)},$$

where $m(\boldsymbol{x})$ is the volume fraction of X and Λ is the intensity measure of ν_X.

As is mentioned in the introduction of this thesis, it was observed (see e.g. [1, 8, 9, 10]) that the intensity measure on a ball of radius r centred at the city centre $\boldsymbol{x}_0$ behaves for large

Table 5.13: GPS coordinates of the centre point $\boldsymbol{x}_0$ and coefficient γ of a power law decay of $\hat{m}_{r;\boldsymbol{x}_0}$.

City	Latitude	Longitude	$\hat{\gamma}$
Boston	N 42° 21.466′	W 71° 03.759′	0.25
Chicago	N 41° 55.915′	W 87° 38.042′	$\sim$
Los Angeles	N 34° 00.781′	W 118° 15.330′	0.24
Pittsburgh	N 40° 27.136′	W 79° 57.232′	0.49
Seattle	N 47° 37.837′	W 122° 20.012′	0.23
Berlin	N 52° 31.233′	E 13° 23.623′	0.23
Birmingham	N 52° 28.960′	W 1° 53.594′	0.66
Milan	N 45° 28.099′	E 9° 11.766′	0.46
Minsk	N 53° 54.400′	E 27° 35.093′	0.25
Moscow	N 55° 45.544′	E 37° 37.100′	0.61
Oslo	N 59° 55.019′	E 10° 41.659′	0.45
Paris	N 48° 51.984′	E 2° 20.684′	0.59
Prague	N 50° 04.861′	E 14° 25.831′	0.78
St. Petersburg	N 59° 56.073′	E 30° 19.081′	0.61

r according to a power law, i.e.

$$\Lambda(B_r(\boldsymbol{x}_0)) \approx Cr^{2-\gamma}$$

for large r and some $C > 0$, possibly with different values of γ in different ranges of r. Furthermore, according to [1, 23], for certain initial range $r \in (0, r_0)$, there should be a plateau with $\gamma = 0$ and then for $r > r_0$ the power law behaviour continues with $\gamma > 0$.

Under this hypothesis one should obtain

$$\mathbb{E}\,\hat{m}_{r;\boldsymbol{x}_0} \sim Kr^{-\gamma} \tag{5.27}$$

as $r \to \infty$ and $\mathbb{E}\,\hat{m}_{r;\boldsymbol{x}_0} \approx K$ for values of r in some initial range $(0, r_0)$ and some $K > 0$.

To be able to perform the analysis, the city centre point $\boldsymbol{x}_0$ should be first determined. We decided to choose a point with the highest volume fraction estimated by the kernel estimator $\hat{m}_{h_1}$, which corresponds to the optimal bandwidth h_1 from the previous Section. In papers mentioned in the beginning of this section the centroid is usually used as a centre point. However, since our observation window W often does not cover the whole city with its neighbourhood, the centroid based only on information from W may be significantly shifted from the true centroid of the whole city. Moreover, we think that the centroid is actually not suitable in et all, since for many geographically constrained cities, like e.g. Chicago or Seattle, the downtown that should contain the centre point is located close to the boundary of the excluded area (usually lake or sea). In this case the centroid is significantly shifted towards the inland. On the other side, our kernel estimator uses edge corrections and as a result its highest value is correctly located inside the downtown. The GPS coordinates of centre points in analysed cities selected using our method are given in Table 5.13.

The resulting dependence of $\hat{m}_{r;\boldsymbol{x}_0}$ on r is shown in Figure 5.35. The range of values of r is for each city chosen such that $\nu(B_r(\boldsymbol{x}_0) \cap W) \approx Cr^2$ for some $C > 0$, meaning that we avoid the systematic changes of $\hat{m}_{r;\boldsymbol{x}_0}$ due to the complicated geometry of the window W. We can see that the expected behaviour is roughly followed. The plateau of constant volume fraction seems to be presented in the range $r \in (0, r_0)$ for r_0 being usually between 1 km and 5 km. The fluctuations for small values of r are given by the large influence of individual buildings in the vicinity of the centre point $\boldsymbol{x}_0$ and should not be misinterpreted as fluctuations of the expected volume fraction. Note that based on previous sections we assume that the built-up

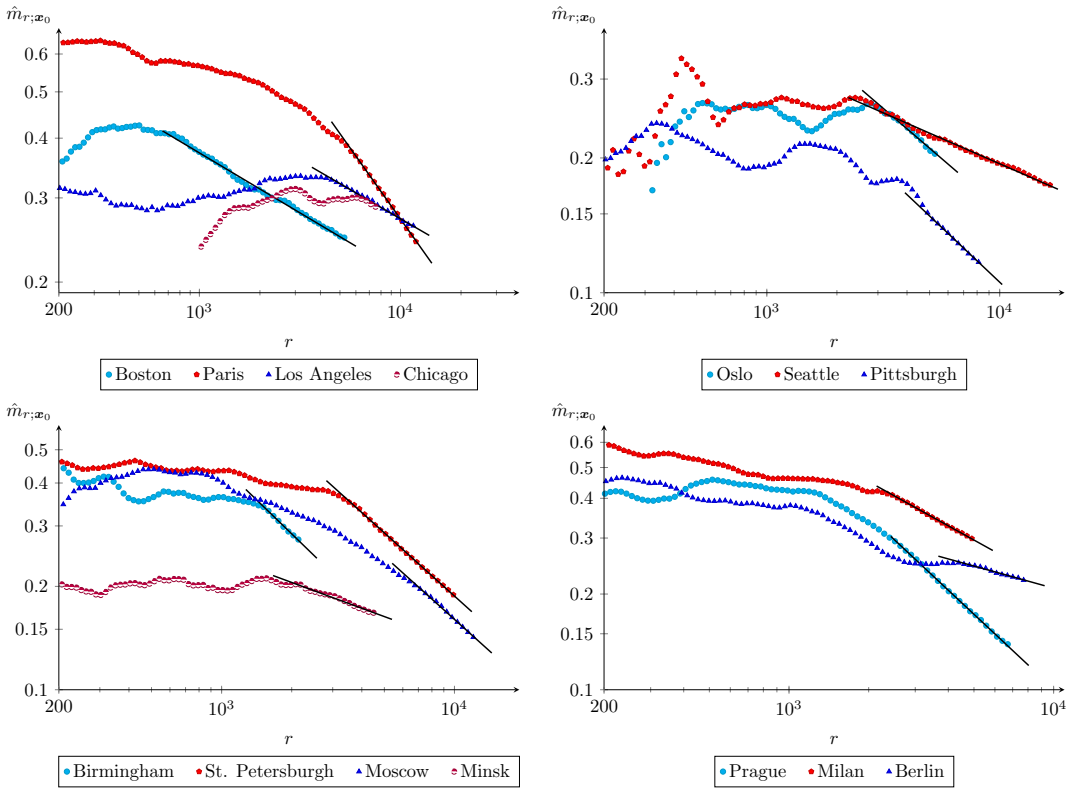


Figure 5.35: The dependence of $\hat{m}_{r;\boldsymbol{x}_0}$ on r for all analysed cities.

area is long-range dependent. Hence the estimator $\hat{m}_{r;\boldsymbol{x}_0}$ has a large variance for small and medium values of r. For large values of r the power law decay of $\hat{m}_{r;\boldsymbol{x}_0}$ is visible for all cities. For some of them, e.g. Oslo, Birmingham, and Milan, the power law part is not very long. For Chicago it is actually very short. Therefore, we are not able to surely confirm the hypothesis of the power law decay. However, the suggestion is that it is true. Finally, we provide a least squares log-linear fit of the power law dependence according to a relation

$$\log(\hat{m}_{r;\boldsymbol{x}_0}) = \hat{\gamma}\log(r) + c$$

that holds for large values of r. Hence we obtain the estimator $\hat{\gamma}$ of the coefficient γ from (5.27). In Table 5.13 we show estimated values of γ for all analysed cities except Chicago. The fits are also shown in Figure 5.35. The values of $\hat{\gamma}$ are between 0.23 and 0.78 which corresponds to those obtained in previous works. However, the values for individual cities are different (see [8] for concrete values). This is probably a consequence of the scope of our analysis. Whereas the previous works focused on large urban areas around analysed cities we deal with much smaller neighbourhoods of city cores.

5.7 Distribution of building sizes

In Sections 5.3 and 5.4 the built-up area was analysed as a random closed set. This random closed set was obtained as a union set of individual buildings that are represented by polygons. Instead of the union set we may study the original collection of buildings as a particle process defined in Section 2.7. Hence, the buildings in some area W are taken as the realization of

some particle process Y in $\mathbb{R}^2$ observed in window W. In the following we restrict ourselves to the assumption that Y is stationary.

Since all buildings have positive volume, Y may be restricted on the collection $\mathcal{W}$ of all compact sets with positive Lebesgue measure, which is according to Theorem 3.1.1 measurable set in C'. Hence Y may be taken as a particle process in $\mathcal{W}$. Furthermore, according to Theorem 3.1.2 one may take the centroid defined by (3.1) as a suitable centre function $\boldsymbol{z}$ on $\mathcal{W}$. Given the centre function $\boldsymbol{z}$, the grain space $\mathcal{W}_0$ is defined by (2.29). Finally, the stationarity of Y implies the existence of the typical grain Y_0 with grain distribution $\mathbb{Q}$ on $\mathcal{W}_0$.

In this section we focus only on the distribution of sizes of individual buildings, i.e. on a random variable $X = \nu(Y_0)$, where ν is the Lebesgue measure on $\mathbb{R}^2$. Our aim is to estimate the distribution of X, which is connected to $\mathbb{Q}$ by

$$\mathbb{P}(X \le x) = \mathbb{Q}(\nu(Y_0) \le x).$$

Since Y is assumed to be stationary, we use buildings only inside the same observation window as was used in the stationary random closed set approach in Section 5.3. It turns out that the crucial role in estimating the distribution of X is played by the generalized beta distribution of the second kind (GB2).

The GB2 distribution depends on four positive parameters, a, b, p, q, and its distribution function $F(x)$ is given by

$$F(x) = \begin{cases} I_z(p,q) & \text{for } x \ge 0, \\ 0 & \text{for } x < 0, \end{cases} \quad \text{with } z = \frac{\left(\frac{x}{b}\right)^a}{1 + \left(\frac{x}{b}\right)^a},$$

where $I_z(p,q)$ is the regularized incomplete beta function that is for all p, q with $\Re p, \Re q > 0$ and all $z \in [0,1]$ defined by

$$I_z(p,q) = \frac{1}{B(p,q)} \int_0^z u^{p-1}(1-u)^{q-1}\,\mathrm{d}u$$

and $B(p,q)$ is the beta function defined by

$$B(p,q) = \int_0^1 u^{p-1}(1-u)^{q-1}\,\mathrm{d}u, \quad \Re p, \Re q > 0.$$

The probability density function f of GB2(a,b,p,q) is

$$f(x) = \frac{1}{B(p,q)} \frac{a}{b^{ap}} \frac{x^{ap-1}}{\left(1 + \left(\frac{x}{b}\right)^a\right)^{p+q}} \quad \text{for} \quad x > 0 \tag{5.28}$$

and $f(x) = 0$ otherwise. One can easily see that b is a scale parameter and a, p, q are shape parameters. The density for $x \gg 1$ behaves like x^{-aq-1} and for $x \ll 1$ like x^{ap-1}. The distribution is therefore capable of fitting different power law behaviours for large and small values of x. Moreover it follows that moments exist for $-ap < k < aq$ and are

$$\mathbb{E}(X^k) = \frac{b^k B(p + k/a, q - k/a)}{B(p,q)}, \tag{5.29}$$

where $X \sim$ GB2(a,b,p,q) is a random variable. In particular the second moment exists when $aq > 2$.

The important property of the GB2 distribution is the invariance with respect to scaling and raising to a positive power. To see this let us take $X \sim$ GB2(a,b,p,q) and try to find the distribution of

$$Y = cX^\alpha,$$

Table 5.14: Parameter estimates in the case $R \sim \mathrm{GB2}(a, b, p, q)$ obtained by maximum likelihood.

City	$\hat{a}$	$\hat{b}$	$\hat{p}$	$\hat{q}$
Birmingham	1.47	8.46	2.97	2.36
Milan	2.07	497.79	0.67	174.10
Minsk	4.99	15.18	0.48	0.66
Moscow	6.01	14.33	0.35	0.60
Oslo	9.99	10.83	0.22	0.33
St. Petersburgh	12.64	20.07	0.13	0.29

where $c > 0$ and $\alpha > 0$. For the distribution function of Y we have

$$F_Y(y) = \mathbb{P}(Y \le y) = \mathbb{P}(cX^\alpha \le y) = \mathbb{P}\left(X \le (c^{-1}y)^{\frac{1}{\alpha}}\right) = F_X((c^{-1}y)^{\frac{1}{\alpha}}) = I_{\tilde{z}}(p, q),$$

where

$$\tilde{z} = \frac{\left(\frac{(c^{-1}y)^{\frac{1}{\alpha}}}{b}\right)^a}{1 + \left(\frac{(c^{-1}y)^{\frac{1}{\alpha}}}{b}\right)^a} = \frac{\left(\frac{y}{cb^\alpha}\right)^{\frac{a}{\alpha}}}{1 + \left(\frac{y}{cb^\alpha}\right)^{\frac{a}{\alpha}}}.$$

Hence Y has again the GB2 distribution with parameters $a/\alpha, cb^\alpha, p, q$, i.e.

$$X \sim \mathrm{GB2}(a, b, p, q) \quad \Longleftrightarrow \quad cX^\alpha \sim \mathrm{GB2}(a/\alpha, cb^\alpha, p, q). \tag{5.30}$$

The GB2 distribution is often used in economics, where it was introduced as an income distribution in [108]. Therefore, it is not a surprise, that it may be appear in connection to the distribution of building sizes. As a full four parameter family, it contains many other important distributions as its sub-families. For details and further discussion see [109].

Now let us focus on the analysis of the distribution of building sizes. There are two scenarios. The first holds for all analysed European cities except Paris, Berlin, and Prague. Here the distribution of X is well, and best among other standard families of distributions, approximated by the GB2 distribution. Since in the following part we would like to approximate buildings by balls and use the distribution of their radii, we do not directly estimate GB2 parameters of X, but instead we introduce a new random variable R given by

$$R = \frac{\sqrt{X}}{\sqrt{\pi}}.$$

Thus R corresponds to the equivalent radius of a ball with area X. As was discussed previously, it has again the GB2 distribution with the relation (5.30) between parameters. Hence in the following results of the fit of the distribution of R are presented. The estimation was performed using the maximum likelihood and estimated values of parameters are summarized in Table 5.14. The histograms and quantile-quantile (Q–Q) plots of fitted distributions for selected cities are shown in Figure 5.36.

The second scenario holds for analysed cities in the USA and for Berlin, Paris, and Prague. Here the distribution of X corresponds to a mixture of two random variables with GB2 distributions. Such a distribution denoted by $\mathrm{MGB2}(a_1, b_1, p_1, q_1; a_2, b_2, p_2, q_2; p)$ is determined by the probability density function

$$f_{\mathrm{MGB2}}(r) = p f_{\mathrm{GB2}}(r; a_1, b_1, p_1, q_1) + (1 - p) f_{\mathrm{GB2}}(r; a_1, b_1, p_1, q_1),$$

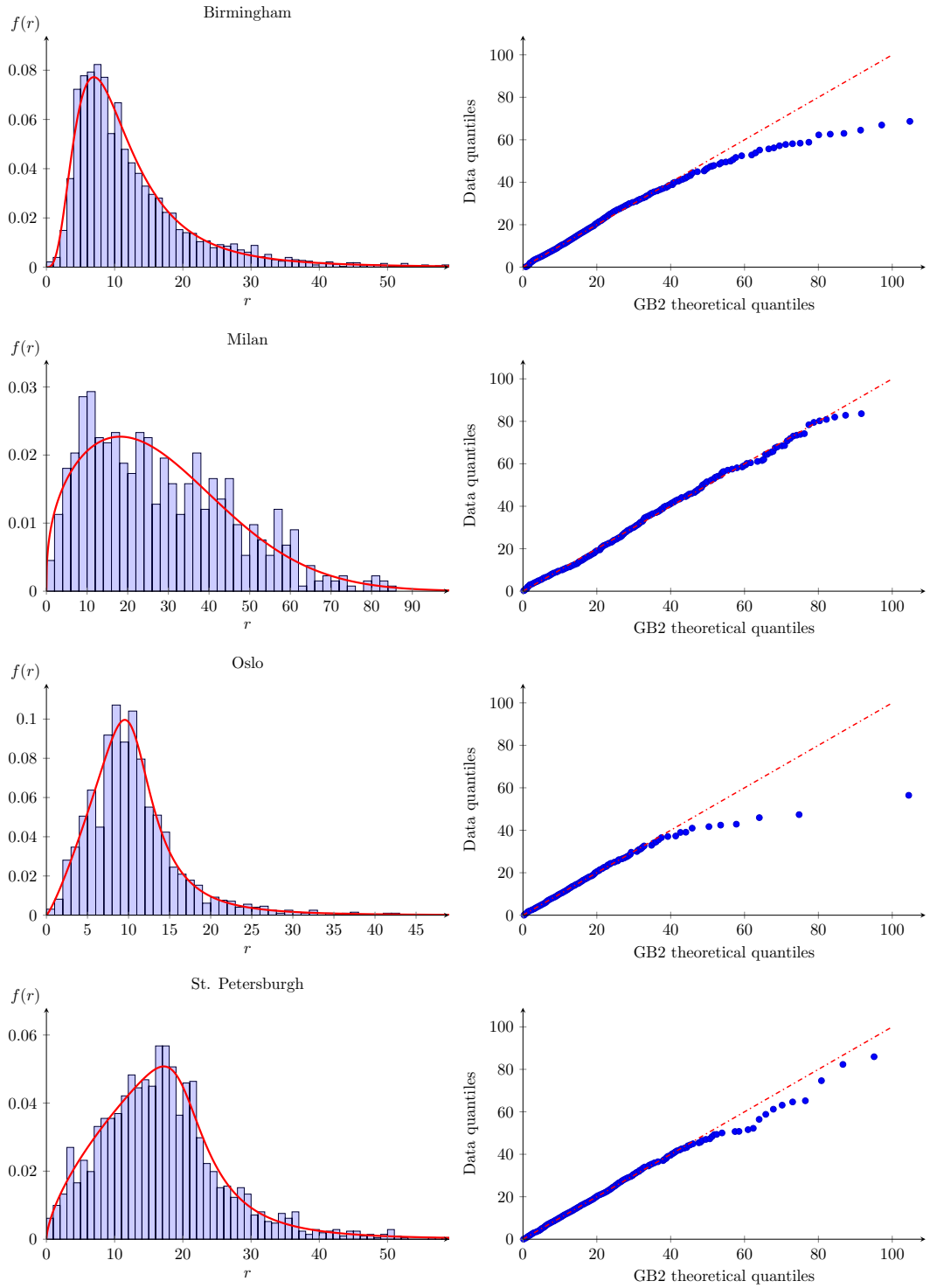


Figure 5.36: The visualisation of distributions of R compared to their GB2 fits obtained by maximum likelihood. On the left: the histogram together with fitted theoretical probability density function. On the right: the Q–Q plot comparing the empirical and fitted distributions.

Table 5.15: Parameter estimates in the case $R \sim \mathrm{MGB2}(a_1, b_1, p_1, q_1; a_2, b_2, p_2, q_2; p)$ obtained by maximum likelihood.

City	$\hat{a}_1$	$\hat{b}_1$	$\hat{p}_1$	$\hat{q}_1$	$\hat{a}_2$	$\hat{b}_2$	$\hat{p}_2$	$\hat{q}_2$	$\hat{p}$
Berlin	10.23	8.10	0.76	0.27	51.14	14.23	0.035	0.055	0.55
Boston	43.24	5.74	0.085	0.049	2.99	40.39	0.47	2.41	0.67
Chicago	512	3.50	0.017	0.004	13.06	5.75	3.14	1.55	0.66
Los Angeles	0.46	0.04	78.45	8.99	27.53	7.53	0.14	0.42	0.54
Paris	2.82	1.29	1.19	0.67	6.04	8.61	0.63	0.60	0.12
Pittsburgh	4.70	17.96	0.33	0.65	0.53	3.61	63.44	51.72	0.088
Prague	0.70	26.83	6.86	16.54	7.54	5.73	19.94	0.47	0.49
Seattle	20.60	5.93	1.89	0.36	0.78	204	4.65	37.19	0.11

Table 5.16: Comparison of basic characteristics of the distribution of R and of the corresponding fit by GB2 or MGB2 distribution: N is the number of buildings; α, β are power law coefficients of the fitted probability density function according to (5.31); μ and $\overline{R}_N$ are the expectation of the fitted distribution and the sample mean, respectively; σ^2 and s_N^2 are the variance of the fitted distribution and the sample variance, respectively.

City	N	α	β	μ	$\overline{R}_N$	σ^2	s_N^2
Berlin	8431	0.81	3.75	12.93	12.79	101	58
Boston	5047	0.42	3.13	12.56	12.51	283	137
Chicago	75297	7.91	3.24	5.85	5.83	41	13
Los Angeles	40747	2.81	5.16	6.83	6.80	35	28
Pittsburgh	13339	0.57	4.04	6.58	6.57	28	24
Seattle	2435	2.62	8.39	14.33	14.33	94	93
Birmingham	3280	3.35	4.47	12.79	12.68	117	87
Milan	665	0.39	360.52	28.37	28.27	319	323
Minsk	1776	1.38	4.28	15.43	15.44	109	101
Moscow	4244	1.08	4.62	13.18	13.15	66	59
Oslo	1960	1.21	4.29	10.71	10.64	46	35
Paris	23705	2.35	2.89	8.88	8.88	∞	30
Prague	3760	3.83	4.54	10.47	10.46	36	34
St. Petersburgh	2113	0.66	4.62	17.17	17.15	114	96

where $f_{\mathrm{GB2}}(r; a_i, b_i, p_i, q_i)$ is a probability density function of $\mathrm{GB2}(a_i, b_i, p_i, q_i)$ given by (5.28). For the MGB2 distribution the condition analogous to (5.30) is clearly valid. Hence we may again analyse the equivalent radius R instead of X, because it has also the MGB2 distribution. The results of the estimation performed using the maximum likelihood are summarized in Table 5.15. The histograms and quantile-quantile (Q–Q) plots of fitted distributions are shown in Figures 5.37 and 5.38.

Now let us discuss the performance and properties of the fitting. As follows from presented figures the estimated distributions in all cities except Chicago reasonably fit to empirical histograms. It means that they perform well in regions with largest probability. The worse results are usually obtained when one focus on the tail of the distribution, which can be seen in Q–Q plots. Here the empirical quantiles of the data are for large probability generally smaller than the theoretical quantiles. This means that there is a less number of buildings with very large

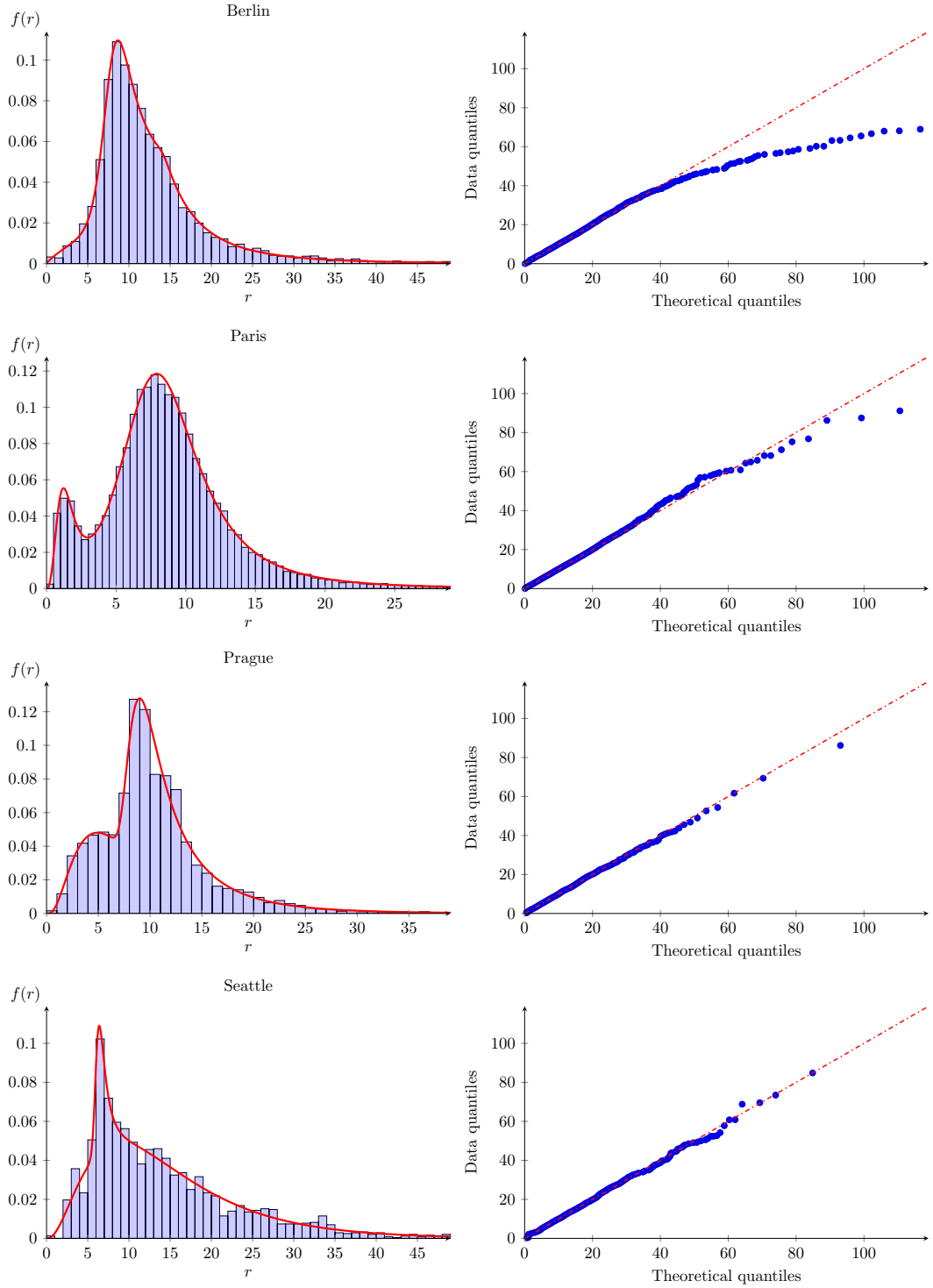


Figure 5.37: The visualisation of distributions of R compared to their MGB2 fits obtained by maximum likelihood. On the left: the histogram together with fitted theoretical probability density function. On the right: the Q–Q plot comparing the empirical and fitted distributions.

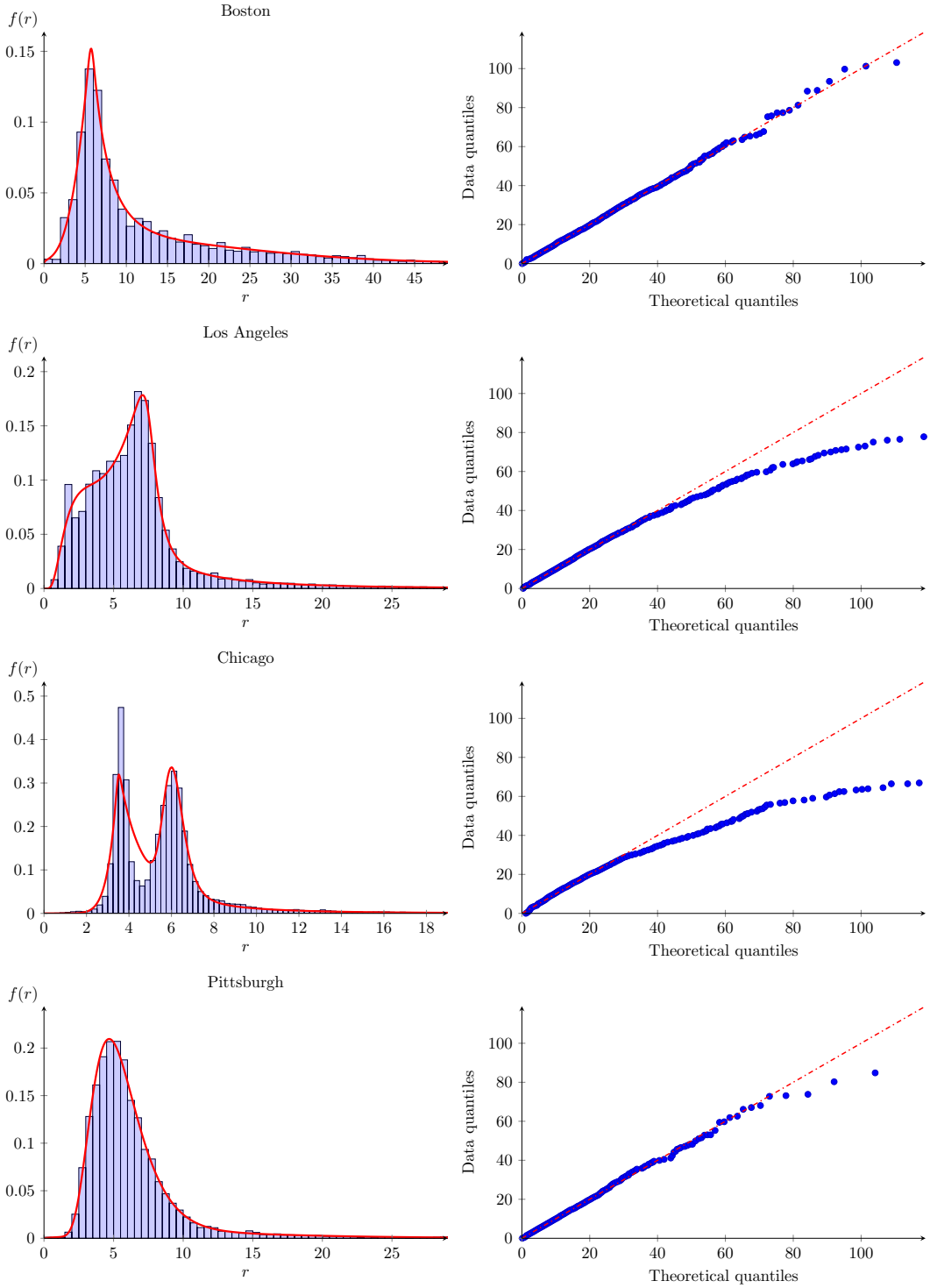


Figure 5.38: The visualisation of distributions of R compared to their MGB2 fits obtained by maximum likelihood. On the left: the histogram together with fitted theoretical probability density function. On the right: the Q–Q plot comparing the empirical and fitted distributions.

values of R, than predicted by the estimated theoretical distribution. In the case of Chicago the fit by the MGB2 is not precise even for small values of R, where it is unable to follow the probability density function decay of the first peak. The tail of the distribution is again overestimated.

As a simple verification of the performance of the fitted distribution we use the expectation μ and the variance σ^2. Let us denote the number of buildings by N. The expectation can be directly estimated by the sample mean $\overline{R}_N$ and the variance by the sample variance s_N^2 defined by

$$\overline{R}_N = \frac{1}{N}\sum_{i=1}^{N} R_i, \qquad s_N^2 = \frac{1}{N-1}\sum_{i=1}^{N}(R_i - \overline{R}_N)^2.$$

Those estimators should lead to similar values as the theoretical expectation μ and variance σ^2 of the fitted distribution. The theoretical vales for the GB2 distribution are from (5.29) given by

$$\mu_{\mathrm{GB2}} = \frac{bB(p+1/a, q-1/a)}{B(p,q)}, \quad \sigma^2_{\mathrm{GB2}} = \frac{b^2 B(p+2/a, q-2/a)}{B(p,q)} - \mu^2_{\mathrm{GB2}},$$

and for the MGB2 distribution by

$$\mu_{\mathrm{MGB2}} = p\frac{b_1 B(p_1+1/a_1, q_1-1/a_1)}{B(p_1,q_1)} + (1-p)\frac{b_2 B(p_2+1/a_2, q_2-1/a_2)}{B(p_2,q_2)},$$
$$\sigma^2_{\mathrm{MGB2}} = p\frac{b_1 B(p_1+2/a_1, q_1-2/a_1)}{B(p_1,q_1)} + (1-p)\frac{b_2 B(p_2+2/a_2, q_2-2/a_2)}{B(p_2,q_2)} - \mu^2_{\mathrm{MGB2}}.$$

The results are presented in Table 5.16, where we also show the power law coefficients α and β of the estimated theoretical probability density function f in the vicinity of 0 and ∞ according to

$$f(r) \sim r^{\alpha} \quad (r \to 0_+) \qquad \text{and} \qquad f(r) \sim r^{-\beta} \quad (r \to \infty), \tag{5.31}$$

respectively. We can see that $\overline{R}_N$ and μ are always very close to each other. The worse situation is when comparing the variances s_N^2 and σ^2. This is caused by the strong impact of the tail of the distribution. Hence when the value of β is small, indicating the slow power law decay, then the difference in the variance is large. This holds for Boston, Chicago, and Berlin. Moreover, in Paris the fitted values correspond to $\beta = 2.89$. Here the power law decay is dominated by the $r^{-a_1 q_1 - 1}$ term of the first component of the mixture. Since $a_1 q_1 = 1.89$ then the second moment and the variance are infinite. On the other side, in Milan the large value of β, indicating strong decrease of the probability density function, is probably caused by the small size of the sample.

It is also interesting to discuss the influence of number N of buildings involved in the estimation. In Table 5.16 we see that all values of N are relatively large. However, the buildings in a city are correlated due to urbanisation plans and various social forces. Hence the observed values of R are not independent, which limits the overall performance of the estimation process, which is then lower compared to the estimation based on independent samples.

Finally, it is a question if one needs the full four parameter family GB2 or even the nine parameter family MGB2 of distributions to describe the true distribution of R. During the analysis we were not able to find any simplifying relations between parameters. The need for the mixture MGB2 seems to be important, since it may indicate two different and partially independent components involved in the built-up process. However, here the more detailed analysis is needed before drawing any conclusions.

5.7.1 Covariance for small arguments

The knowledge of the distribution of building sizes can help us to determine the covariance of the corresponding built-up area. To do this let again take buildings as particles of a stationary

particle process Y. Let further Y^0 denote the stationary ground point process of Y and Y_0 the typical grain of Y, corresponding to the centroid centre function $\boldsymbol{z}$, with distribution $\mathbb{Q}$. The built-up area taken as a stationary random closed set Z according to Section 5.3 is precisely the union set of Y.

When the particles of Y do not overlap, which holds for buildings by principle, then the volume fraction p of Z is given by

$$p = \lambda\, \mathbb{E}\, \nu(Y_0) = \lambda\, \mathbb{E}\, X,$$

where λ is the intensity of Y^0 and $X = \nu(Y_0)$ is the area of the typical grain. To show this let us take $\tilde{Y}$ as the stationary marked point process corresponding to Y by Theorem 2.7.3 and use its representation by

$$\tilde{Y} = \sum_{i=1}^{\tau} \delta_{(\boldsymbol{y}_i, C_i)}, \quad \tau = \tilde{Y}(\mathbb{R}^d \times \mathcal{W}_0)$$

according to (2.27). If the particles of Y do not overlap then

$$(\boldsymbol{y}_i + C_i) \cap (\boldsymbol{y}_j + C_j) = \emptyset \quad \text{for all } i \neq j$$

with probability one. This implies that

$$p = \mathbb{E}\, \mathbb{1}_Z(\boldsymbol{0}) = \mathbb{E} \sum_{i=1}^{\tau} \mathbb{1}_{\boldsymbol{y}_i + C_i}(\boldsymbol{0}) = \mathbb{E} \sum_{i=1}^{\tau} \mathbb{1}_{C_i}(-\boldsymbol{y}_i).$$

Using the Campbell theorem 2.5.2 and decomposition of the intensity measure Θ of $\tilde{Y}$ from Theorem 2.6.1, we finally obtain

$$p = \lambda \int_{\mathcal{W}_0} \int_{\mathbb{R}^2} \mathbb{1}_C(-\boldsymbol{y})\, \mathrm{d}\boldsymbol{y}\, \mathbb{Q}(\mathrm{d}C) = \lambda \int_{\mathcal{W}_0} \nu(-C)\, \mathbb{Q}(\mathrm{d}C) = \lambda\, \mathbb{E}\, \nu(Y_0).$$

Now we focus on the covariance C of Z. For non-overlapping particles it may be decomposed into two terms:

$$C(\boldsymbol{r}) = C_1(\boldsymbol{r}) + C_2(\boldsymbol{r}),$$

where C_1 corresponds to the interaction of particles with themselves and C_2 to the interaction between different particles. Clearly $C_2(\boldsymbol{r})$ is small for small $\boldsymbol{r}$ and it is even zero, when all particles are separated by some minimal positive distance. Hence the behaviour of the covariance in the vicinity of 0 is primarily influenced by C_1. The covariance C_1 of particles with itself is by analogous reasons as before given as the expectation of the intersection of the typical grain Y_0 with shifted copy of itself multiplied by the intensity λ, i.e.

$$C_1(\boldsymbol{r}) = \lambda\, \mathbb{E}\, \nu\big(Y_0 \cap (Y_0 - \boldsymbol{r})\big) = \lambda\, \mathbb{E}\, \gamma_{Y_0}(\boldsymbol{r}),$$

where $\gamma_{Y_0}(\boldsymbol{r}) = \nu\big(Y_0 \cap (Y_0 - \boldsymbol{r})\big)$ is the set covariance of Y_0.

Our aim is to determine the covariance C_1 from the distribution of X. We introduce the following simplification. From now on let us assume that the buildings are balls. Therefore, the distribution $\mathbb{Q}$ of the typical grain is fully determined by the distribution of $R = \sqrt{X/\pi}$. The covariance $C_1(\boldsymbol{y})$ then depends only on the norm $y \equiv \|\boldsymbol{y}\|$ and is given by

$$C_1(y) = \lambda\, \mathbb{E}\, \gamma_R(y) = p \frac{\mathbb{E}\, \gamma_R(y)}{\pi\, \mathbb{E}\, R^2}, \tag{5.32}$$

where $\gamma_R(y) = \nu\big(B_R \cap (B_R - y\boldsymbol{u})\big)$ is the set covariance of a ball of radius R and $\boldsymbol{u}$ is an arbitrary unit vector. The intersection area $\gamma_r(y)$ is non-zero for $y \in [0, 2r)$ and can be calculated as

$$\gamma_r(y) = 2r^2 \arccos\left(\frac{y}{2r}\right) - yr\sqrt{1 - \left(\frac{y}{2r}\right)^2}.$$

Thus, in order to calculate the expectation, it is necessary to expand $\gamma_r(y)$ into series in powers of $y/2r$. We use the following Taylor expansions (see for instance [95]):

$$\arccos(x) = \frac{\pi}{2} - \sum_{n=0}^{\infty} \frac{(2n)!}{4^n (n!)^2} \frac{1}{(2n+1)} x^{2n+1}, \quad |x| < 1,$$
$$(1+x)^\alpha = \sum_{n=0}^{\infty} \binom{\alpha}{n} x^n, \quad |x| < 1.$$

Using the identity

$$\binom{\frac{1}{2}}{n} = \frac{(-1)^{n+1}(2n)!}{4^n (n!)^2} \frac{1}{(2n-1)}, \quad n = 0, 1, 2, \ldots,$$

that can be easily proven by induction, one finally obtains

$$\gamma_r(y) = \pi r^2 + 4 \sum_{n=0}^{\infty} \frac{(2n)!}{2^{2n}(n!)^2} \frac{1}{(4n^2-1)} \frac{y^{2n+1}}{2^{2n+1} r^{2n-1}}. \tag{5.33}$$

Up to the first four terms it reads

$$\gamma_r(y) = \pi r^2 - 2ry + \frac{y^3}{12r} + \frac{y^5}{320 r^3} + \mathcal{O}\left(\frac{y^7}{r^5}\right).$$

As we have seen in the previous part, the distribution of R is given by the generalized beta distribution of the second kind (GB2) or by the mixture of two GB2 distributions. Hence we need to calculate both expectations in (5.32) with respect to the GB2(a, b, p, q) distribution. Since $\gamma_r(y)$ is 0 for $y \geq 2r$, the expectation of $\gamma_R(y)$ is given by

$$\mathbb{E}\,\gamma_R(y) = \int_0^\infty \gamma_r(y)\, \mathbb{P}(\mathrm{d}r) = \int_{\frac{y}{2}}^\infty \gamma_r(y)\, \mathbb{P}(\mathrm{d}r) \equiv \mathbb{E}\left(\gamma_R(y); 2R > y\right).$$

Using expansion (5.33) of $\gamma_r(y)$ we obtain

$$\mathbb{E}\,\gamma_R(y) = \pi\, \mathbb{E}(R^2; 2R > y) + 4 \sum_{n=0}^{\infty} \frac{(2n)!}{2^{2n}(n!)^2} \frac{1}{(4n^2-1)} \left(\frac{y}{2}\right)^{2n+1} \mathbb{E}(R^{1-2n}; 2R > y). \tag{5.34}$$

Hence we need to calculate

$$\mathbb{E}(R^k; 2R > y) = \int_{\frac{y}{2}}^\infty r^k\, \mathbb{P}(\mathrm{d}r) = \int_{\frac{y}{2}}^\infty r^k f(r)\, \mathrm{d}r,$$

where $f(r)$ is the probability density function of the GB2 distribution and $k = 2, 1, -1, -3, \ldots$. Using expression (5.28) for the density, the integral becomes

$$\mathbb{E}(R^k; 2R > y) = \int_{\frac{y}{2}}^\infty r^k \frac{1}{B(p,q)} \frac{a}{b^{ap}} \frac{r^{ap-1}}{\left(1 + \left(\frac{r}{b}\right)^a\right)^{p+q}} \mathrm{d}r.$$

After substitution $\left(\frac{r}{b}\right)^a = t$ we get

$$\mathbb{E}(R^k; 2R > y) = \frac{b^k}{B(p,q)} \int_{\left(\frac{y}{2b}\right)^a}^\infty \frac{t^{p-1+\frac{k}{a}}}{(1+t)^{p+q}}\, \mathrm{d}t$$

and further substitution $u = \frac{1}{1+t}$ brings it to

$$\mathbb{E}(R^k; 2R > y) = \frac{b^k}{B(p,q)} \int_0^x u^{q-1-\frac{k}{a}} (1-u)^{p-1+\frac{k}{a}} \,\mathrm{d}u,$$

where

$$x = \frac{1}{1 + \left(\frac{y}{2b}\right)^a}.$$

Using the incomplete beta function $B_x(a,b)$, that is defined by

$$B_x(a,b) = \int_0^x u^{a-1}(1-u)^{b-1} \,\mathrm{d}u, \quad 0 < x < 1, \quad \Re a > 0, \tag{5.35}$$

the expectation can be finally rewritten as

$$\mathbb{E}(R^k; 2R > y) = \frac{b^k}{B(p,q)} B_x\left(q - \frac{k}{a}, p + \frac{k}{a}\right) \tag{5.36}$$

with the same x as before. We assume $aq > 2$, which means that the second moment of the GB2(a,b,p,q) distribution exists. The expression is therefore correct for all $k = 2, 1, -1, -3, \ldots$. It may, however, diverge as $x \to 1$ and thus as $y \to 0_+$. The divergence occurs when $k < -pa$. In order to extract the behaviour for x close to 0 and 1 we use the following relations between the incomplete beta function and the hypergeometric function (see e.g. [110]):

$$B_x(a,b) = \frac{x^a}{a} F(a, 1-b; a+1; x), \tag{5.37}$$

$$= \frac{x^a (1-x)^b}{a} F(a+b, 1; a+1; x). \tag{5.38}$$

Here $F(a,b;c;x)$ is the hypergeometric function defined by the following series

$$F(a,b;c;z) = \sum_{i=0}^{\infty} \frac{(a)_n (b)_n}{(c)_n} \frac{z^i}{n!}, \quad |z| < 1,$$

where

$$(a)_n = \begin{cases} 1 & \text{for } n = 0, \\ a(a+1)\cdots(a+n-1) & \text{for } n > 0. \end{cases}$$

Now let us inspect individual terms in $\mathbb{E}\,\gamma_R(y)$. The first term

$$\pi\,\mathbb{E}(R^2; 2R > y) = \frac{\pi b^2}{B(p,q)} B_x\left(q - \frac{2}{a}, p + \frac{2}{a}\right)$$

can be further rewritten as

$$\pi\,\mathbb{E}(R^2; 2R > y) = \frac{\pi b^2}{B(p,q)} \left(B\left(p + \frac{2}{a}, q - \frac{2}{a}\right) - B_{1-x}\left(p + \frac{2}{a}, q - \frac{2}{a}\right)\right),$$

which follows from the property

$$B(a,b) = B_x(b,a) + B_{1-x}(a,b)$$

of the incomplete beta function that can be easily checked. Using (5.37) we finally obtain

$$\pi\,\mathbb{E}(R^2; 2R > y) =$$
$$\frac{\pi b^2 B\left(p+\frac{2}{a}, q-\frac{2}{a}\right)}{B(p,q)} - \frac{\pi b^2}{B(p,q)} \frac{\left(\frac{y}{2b}\right)^{ap+2}}{\left(1+\left(\frac{y}{2b}\right)^a\right)^{p+\frac{2}{a}}} F\left(p+\frac{2}{a}, 1-q+\frac{2}{a}; p+1+\frac{2}{a}; \frac{\left(\frac{y}{2b}\right)^a}{1+\left(\frac{y}{2b}\right)^a}\right).$$

Remaining terms in $\mathbb{E}\,\gamma_R(y)$, that depend on n, can be divided into two groups by the condition $2n-1 < ap$. The usefulness of this division will be seen later and it is based on the fact that the range of the hypergeometric function $F(a,b;c;x)$ is bounded and the limit as $x \to 1$ can be evaluated when $\Re c > \Re(a+b)$. The limit is then given by Gauss's theorem (see e.g. [110]) as

$$\lim_{x\to 0_+} F(a,b;c;x) = F(a,b;c;1) = \frac{\Gamma(c)\Gamma(c-a-b)}{\Gamma(c-a)\Gamma(c-b)}.$$

If $n < (ap+1)/2$, then we use relation (5.37) and obtain the n-th term of (5.34) as

$$4\frac{(2n)!}{2^{2n}(n!)^2}\frac{1}{(4n^2-1)}\frac{b^2\left(\frac{y}{2b}\right)^{2n+1}}{\left(1+\left(\frac{y}{2b}\right)^a\right)^{q+\frac{2n-1}{a}}}\frac{F\left(q+\frac{2n-1}{a}, 1-p+\frac{2n-1}{a}; q+1+\frac{2n-1}{a}; \frac{1}{1+\left(\frac{y}{2b}\right)^a}\right)}{B(p,q)\left(q+\frac{2n-1}{a}\right)},$$

with bounded range of the hypergeometric function. If $n > (ap+1)/2$ then we use relation (5.38) and obtain the n-th term of (5.34) as

$$4\frac{(2n)!}{2^{2n}(n!)^2}\frac{1}{(4n^2-1)}\frac{b^2\left(\frac{y}{2b}\right)^{ap+2}}{\left(1+\left(\frac{y}{2b}\right)^a\right)^{p+q}}\frac{F\left(p+q, 1; q+1+\frac{2n-1}{a}; \frac{1}{1+\left(\frac{y}{2b}\right)^a}\right)}{B(p,q)\left(q+\frac{2n-1}{a}\right)}.$$

Again the values of the hypergeometric function are bounded.

In order to obtain the expansion of the covariance $C_1(y)$ we need to calculate the denominator in (5.32). It is just the second moment of the GB2 distribution, which is by (5.29) equal to

$$\mathbb{E}\,R^2 = b^2\frac{B\left(p+\frac{2}{a}, q-\frac{2}{a}\right)}{B(p,q)}.$$

Now let n_0 be an integer such that $(ap+1)/2 - 1 < n_0 < (ap+1)/2$. Taking all together we finally obtain

$$C_1(y) = p - p\frac{\left(\frac{y}{2b}\right)^{ap+2}}{\left(1+\left(\frac{y}{2b}\right)^a\right)^{p+\frac{2}{a}}}\frac{F\left(p+\frac{2}{a}, 1-q+\frac{2}{a}; p+1+\frac{2}{a}; \frac{\left(\frac{y}{2b}\right)^a}{1+\left(\frac{y}{2b}\right)^a}\right)}{B\left(p+\frac{2}{a}, q-\frac{2}{a}\right)}$$
$$+\frac{4p}{\pi}\sum_{n=0}^{n_0}\frac{(2n)!}{2^{2n}(n!)^2}\frac{1}{(4n^2-1)}\frac{\left(\frac{y}{2b}\right)^{2n+1}}{\left(1+\left(\frac{y}{2b}\right)^a\right)^{q+\frac{2n-1}{a}}}\frac{F\left(q+\frac{2n-1}{a}, 1-p+\frac{2n-1}{a}; q+1+\frac{2n-1}{a}; \frac{1}{1+\left(\frac{y}{2b}\right)^a}\right)}{B\left(p+\frac{2}{a}, q-\frac{2}{a}\right)\left(q+\frac{2n-1}{a}\right)}$$
$$+\frac{4p}{\pi}\frac{\left(\frac{y}{2b}\right)^{ap+2}}{\left(1+\left(\frac{y}{2b}\right)^a\right)^{p+q}}\sum_{n=n_0}^{\infty}\frac{(2n)!}{2^{2n}(n!)^2}\frac{1}{(4n^2-1)}\frac{F\left(p+q, 1; q+1+\frac{2n-1}{a}; \frac{1}{1+\left(\frac{y}{2b}\right)^a}\right)}{B\left(p+\frac{2}{a}, q-\frac{2}{a}\right)\left(q+\frac{2n-1}{a}\right)}. \quad (5.39)$$

The convergence is quite fast with increasing n. Hence the sum may be truncated to only first few terms and it still approximates the result well.

From the covariance C_1 we may obtain the correlation function κ_1 by

$$\kappa_1(r) = \frac{C_1(r) - p^2}{p(1-p)}.$$

It is interesting to compare the values of $\kappa_1(r)$ with correlation function $\kappa(r)$ estimated directly in Section 5.3. In further we use notation $\kappa_1 \equiv \kappa_1(r; a, b, p, q)$ for the correlation function κ_1 based on the GB2 distribution with parameters a, b, p, q. For the cities with GB2 distribution of R we use the estimated values of parameters that are given in Table 5.14. For cities with R distributed according to the mixture MGB2$(a_1, b_1, p_1, q_1; a_2, b_2, p_2, q_2; q)$ it can be easily shown that the correlation function $\kappa_1(r)$ is given by

$$\kappa_1(r) = C\kappa_1(r; a_1, b_1, p_1, q_1) + (1 - C)\kappa_1(r; a_2, b_2, p_2, q_2),$$

where

$$C = \frac{qb_1^2 \frac{B\left(p_1+\frac{2}{a_1}, q_1-\frac{2}{a_1}\right)}{B(p_1,q_1)}}{qb_1^2 \frac{B\left(p_1+\frac{2}{a_1}, q_1-\frac{2}{a_1}\right)}{B(p_1,q_1)} + (1-q)b_2^2 \frac{B\left(p_2+\frac{2}{a_2}, q_2-\frac{2}{a_2}\right)}{B(p_2,q_2)}}.$$

For those cities we use estimated values of parameters from Table 5.15. The comparison is performed visually by plotting both κ_1 and κ together. The results for all cities except Paris, where the distribution of R do not have finite second moment, are shown in Figures 5.39 and 5.39. The quality of the interpolation of κ by κ_1 seems to be mainly influenced by the validity of the assumption that the buildings are balls. For many cities the assumption is more or less broken since there are many buildings given by large, complicated polygons often with holes. In that case the inter particle covariance decrease faster than for circular grains of equivalent sizes. This phenomenon is clearly observable for the majority of analysed cities.

Finally, let us look on the asymptotic behaviour as $y \to 0_+$. Clearly the argument of the hypergeometric function in the second term of $C_1(y)$ is close to 0 leading the hypergeometric function to be close to 1. On the other hands arguments of hypergeometric functions in the remaining terms are close to 1 leading them close to the limiting value mentioned before. The smallest coefficient of the power law terms is y^1 and it appears in the first term of the sum corresponding to $n = 0$. Taking just this term we get

$$C_1(y) = p - \frac{4p}{\pi} \frac{\Gamma\left(q + 1 - \frac{1}{a}\right)\Gamma\left(p + \frac{1}{a}\right)}{\left(q - \frac{1}{a}\right) B\left(p + \frac{2}{a}, q - \frac{2}{a}\right)\Gamma(p+q)} \left(\frac{y}{2b}\right) + \mathcal{O}(y^2) \quad (y \to 0_+).$$

Using $\Gamma(x+1) = x\Gamma(x)$ it can be further simplified to

$$C_1(y) = p - \frac{4p}{\pi} \frac{\Gamma\left(q - \frac{1}{a}\right)\Gamma\left(p + \frac{1}{a}\right)}{\Gamma\left(p + \frac{2}{a}\right)\Gamma\left(q - \frac{2}{a}\right)} \left(\frac{y}{2b}\right) + \mathcal{O}(y^2) \quad (y \to 0_+).$$

For the correlation function $\kappa_1(r)$ one easily obtains

$$\kappa_1(y; a, b, p, q) = 1 - \frac{4}{\pi(1-p)} \frac{\Gamma\left(q - \frac{1}{a}\right)\Gamma\left(p + \frac{1}{a}\right)}{\Gamma\left(p + \frac{2}{a}\right)\Gamma\left(q - \frac{2}{a}\right)} \left(\frac{y}{2b}\right) + \mathcal{O}(y^2) \quad (y \to 0_+).$$

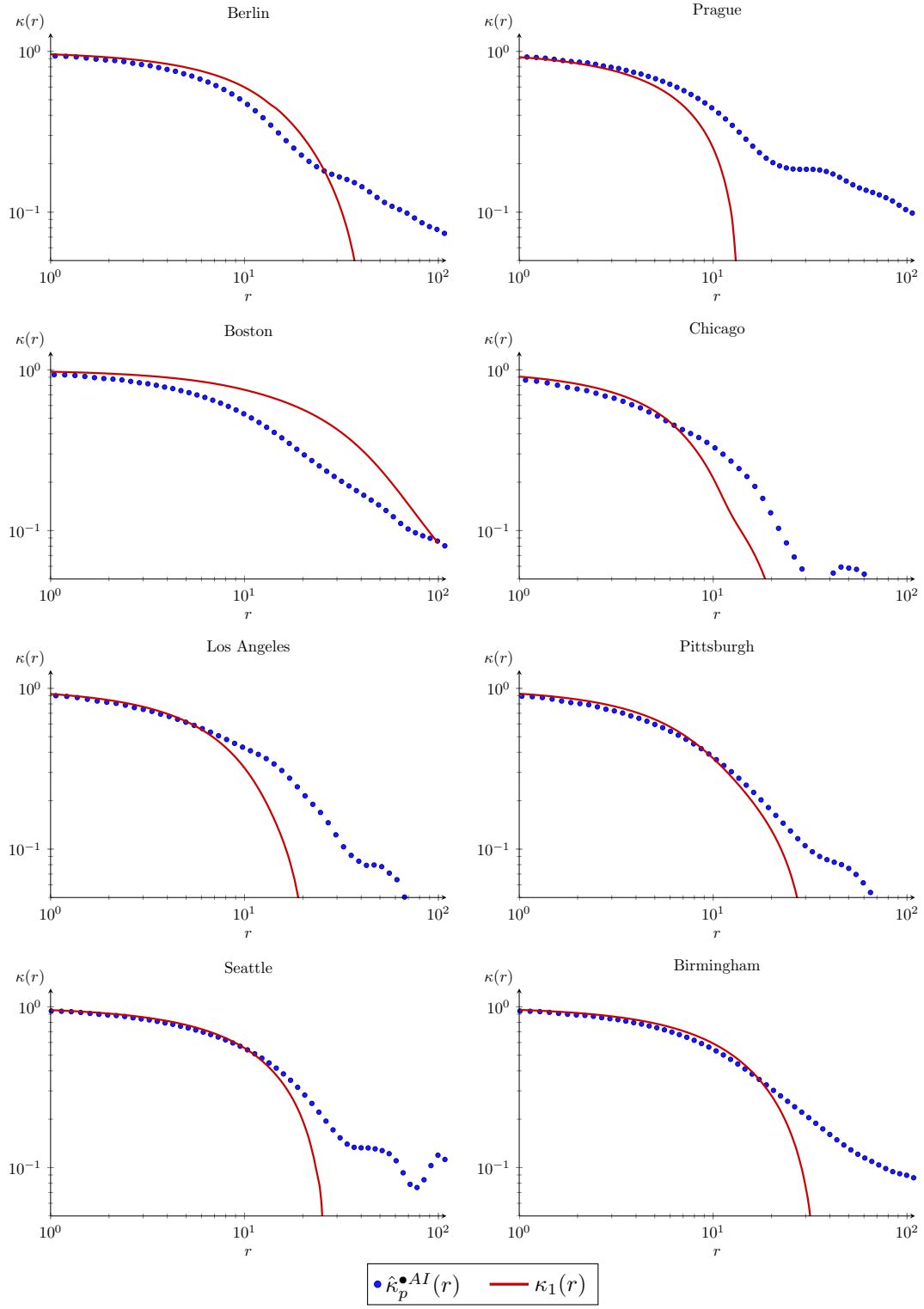


Figure 5.39: The comparison of the correlation function estimated by $\hat{\kappa}_p^{\bullet AI}$ and of the inter-particle correlation function κ_1.

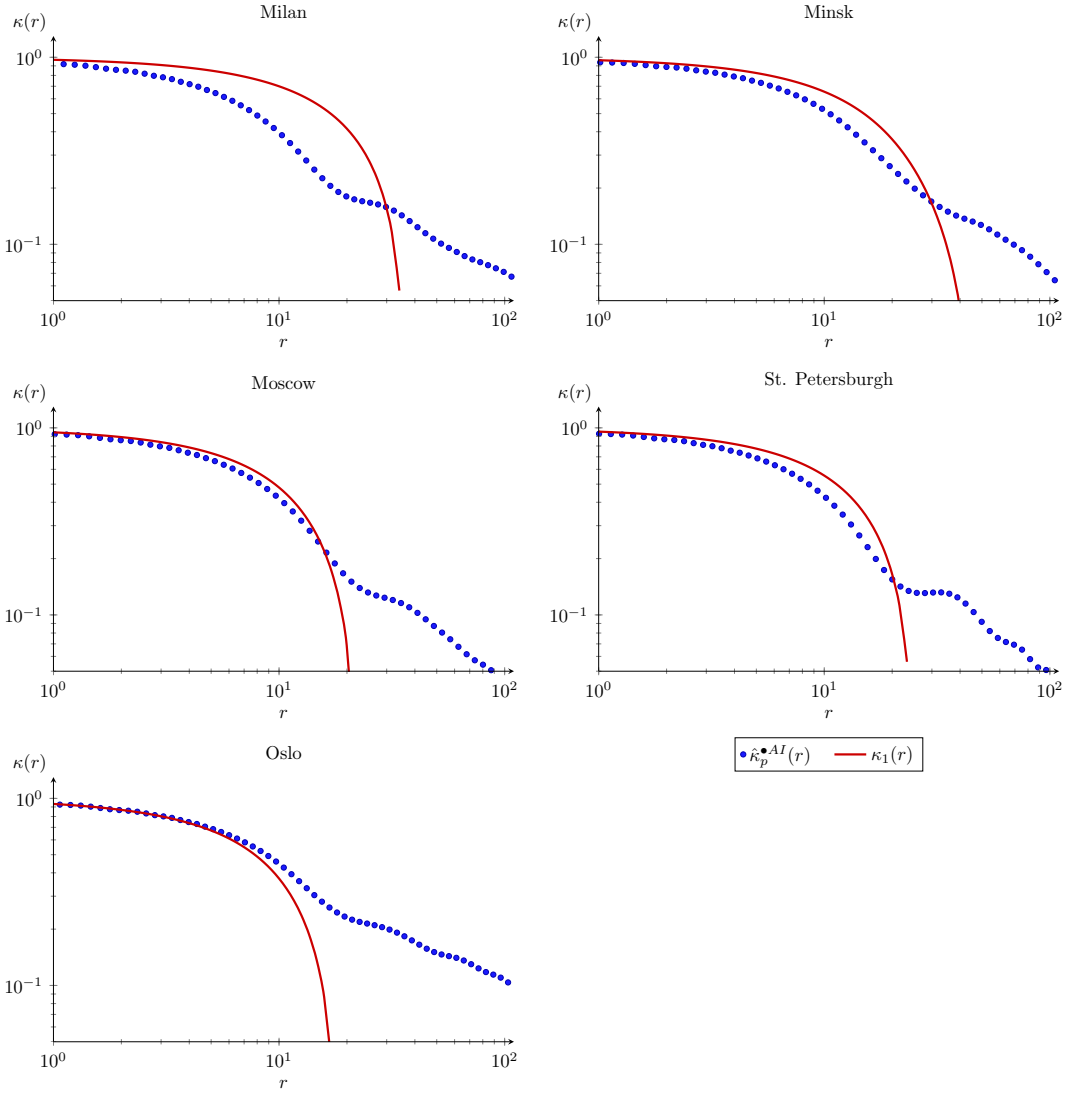


Figure 5.40: The comparison of the correlation function estimated by $\hat{\kappa}_p^{\bullet AI}$ and of the inter-particle correlation function κ_1.

5.8 Fractality versus long-range dependence

As was discussed in the introduction, the structural properties of the built-up area are often described in context of fractal geometry. On scales of large urban areas there appear several scaling relations that are usually shared by fractal objects. Generally, the fractality is a local rather than global concept. It is often connected with certain properties that are namely: the fine structure, i.e. details on an arbitrary scale, irregularity to be described by traditional language of geometry, some form of the self-similarity, and a fractal dimension, usually Hausdorff, greater than the topological dimension.

In the analysis of the built-up area, however, the fractality observed on medium-large scales is not reflected on small scales. When decreasing the scale then individual buildings appear and finally they fully dominate the overall geometry. Since the buildings are standard geometrical objects, i.e. a connected closed set of the positive Lebesgue measure with piecewise smooth boundary of finite length, they clearly do not follow any of the basic characteristics of fractals. Hence on small scales the built-up area is not fractal. This fact is usually overlooked by urban scientists.

As we have seen in previous sections there is, however, a scaling relation that persists even under the full detailed analysis. It is the power law decay of the correlation function that represents a manifestation of the so called long-range dependence. For a built-up area it starts at distances of several tens of meters and is clearly observed in the whole range of our analysis which is usually up to 10-30 kilometres.

In principle the long-range dependence and fractality mean something completely different. Fractal behaviour is very much a local property, whereas the power law decay of the correlation function is a global characteristic. However, in many systems those two properties are closely related as the result of the self-similarity (see e.g. [25, 58] for details). On the other hand it is possible to construct models, where the fractal dimension can be varied independently on the power law coefficient. Such a model is for example given by a Gaussian random field in $\mathbb{R}^d$ with the Cauchy covariance function introduced by Gneiting and Schlather in [96]. Here the fractal behaviour means the Hausdorff dimension of the graph of a realization. It was shown in [51] that the Hausdorff dimension D of the graph is given by the power law behaviour of the covariance function in the vicinity of 0. In particular if $C(0) - C(r) \sim c\,|r|^{\alpha}$ as $r \to 0$ for some $\alpha \in (0, 2]$, then $D = d + 1 - \alpha/2$ with probability one. This is exactly the case of the Cauchy covariance function given by (4.39). Clearly it is independent of the coefficient β of the power law decay when $r \to \infty$.

It also interesting to look at the 0-level excursion set based on the previous Gaussian random field. From the fact that sample paths are with probability one continuous, as was discussed in Section 4.4, and the volume fraction is positive clearly follows that with probability one every non-empty realization of the 0-level excursion set has positive Lebesgue measure. Hence its Hausdorff dimension is d. On the other side the power law decay remains the same as for the original Gaussian random field as follows from (4.40).

From our analysis of second order properties of the built-up area follows that it should be viewed as a long-range dependent structure. This indicates that one should be aware of a generally lower performance of various estimators. It may also provide an important insight into the built-up structure, where the long-range dependence may help to explain observed fluctuations of the built-up area and other phenomena. On the other hand, the non-stationarity on large scales is not a result of fluctuations and it represents a complication in the estimation procedure. In Section 5.4 proposed a method how to overcome this problem with the help of kernel density estimators and with assuming the special form of the correlation function. Numerical simulations indicate that this approach is appropriate.

Conclusion and final remarks

In this thesis we study the built-up structure in cities using the framework of stochastic geometry. Let us now briefly recapitulate the content and major findings of our analysis. The introductory chapter provides a short summary of standard approaches to the urban analysis and especially to the analysis of the built-up structure in cities.

Chapter 2 is devoted to the recapitulation of essential parts of stochastic geometry. We particularly focus on random closed sets and random measures as their extension. For both of of them the first and second order characteristics, ergodicity, and Bartlett spectrum are introduced. As an important model of random closed sets the level excursion set of a random (Gaussian) field is correctly defined by showing its measurability. Furthermore, we derived several formulas for the basic first and second order characteristics that are extensively used in subsequent parts. With regards to random measures we focus on the correct definition of the (reduced) second order (factorial) covariance measure and the Bartlett spectrum as its Fourier transform. For the convenience of the reader the relevant parts of the theory of Fourier analysis of positive-semi definite measures are recapitulated in Section A.2, where some less known results are proven. In the following sections marked processes and particle processes together with their relations are recapitulated and a second important random closed set model given by the union set of a particle process, and in particular by the Boolean model as its special case, is defined. The last section deals with the long-range dependence. Here we developed an extension of the usually treated long-range dependence of random measures in $\mathbb{R}$ to $\mathbb{R}^d$ that was indicated in [43]. Thus, the general definition for both random measures and random closed sets is given and the important connection to the Bartlett spectrum is derived. As a special case we also defined the isotropic long-range dependence, which is characterized by the asymptotic power law decay of the correlation function $\mathrm{cov}(r)$ as $r \to \infty$. The equivalent formulation given through the power law behaviour of the Bartlett spectrum in the vicinity of $\mathbf{0}$ was also shown.

Chapter 3 deals with the construction of a specific centre function based on the centroid of a set. The aim of the construction is to generalize the classical notion of the centroid from sets with positive Lebesgue measure to certain sets with zero Lebesgue measure. The generalization is developed in two ways. The first is given by the limit of centroids of ε-neighbourhoods of the given set as $\varepsilon \to 0_+$. By explicit counterexample it was shown that this generalization does not exist for an arbitrary set. Further, we proved several sufficient conditions of its existence. The most important one, given by Corrolary 3.3.1, is based on the Minkowski measurability. In particular it is possible to prove the existence of the generalized centroid for compact subsets of m-dimensional C^1 (or Lipschitzian) submanifolds of $\mathbb{R}^d$ and their finite unions, which was done in Section 3.4. The measurability on many important subclasses of sets was also proven. The second generalization of the centroid is based on the Hausdorff measure. Its existence and measurability on certain subclasses of sets was also proven and its relation to the previous generalized centroid was discussed. It should be noted that the first generalization is much more natural in connection to the numerical estimation than the second.

In Chapter 4 some statistical issues concerning the estimation of basic characteristics of random closed sets are treated. After a brief recapitulation of statistics for point processes in the first section we focused on stationary random closed sets. The detailed discussion of

the volume fraction and second order characteristics estimators was provided. In particular the intrinsically balanced estimator of the correlation function was introduced. In Subsection 4.2.3 the use of the fast Fourier transform that may significantly decrease the computational time of discrete second order estimators was discussed. Then we proceeded with statistical issues for non-stationary random closed sets. Here the kernel volume fraction estimator was proposed and its basic asymptotic properties and the problem of the bandwidth choice were discussed. For a specific example of a random closed set with the Gaussian volume fraction and exponential correlation function, the approximate explicit formula for the optimal bandwidth was derived. Finally, in the last section several numerical simulations were performed to analyse basic properties of previously proposed estimators in the stationary case. Those simulations were based on the Boolean model and on the 0-level excursion set of a Gaussian random field with the Cauchy correlation function, which is a suitable model of a long-range dependent random closed set.

Chapter 5 provides a detailed analysis of the built-up structure in cities. In the beginning the built-up structure as a collection of buildings represented by polygons was introduced and then the formulae for the area and centroid of polygons were derived. Section 5.3 is devoted to the analysis of the built-up area taken as the realization of a stationary random closed set. The stress is put on second order properties that were analysed by two methods. The first was the direct estimation of the correlation function $\kappa(r)$. We observed that under the additional assumption of isotropy, the estimation of the correlation function may be well fitted by the Cauchy correlation function, which asymptotically follows a power law decay. This observation leads to the preliminary assumption of the long-range dependence of the built-up area. The second method was the analysis of the variance in balls. The theoretical behaviour of the variance in balls as a function of the ball radius was discussed in various cases depending on the value of the integral range and on the behaviour of the Bartlett spectrum in the vicinity of $\mathbf{0}$. We particularly derived that if the correlation function asymptotically follows a power law with coefficient β, then the dependence of the variance in balls on the ball radius follows a power law with coefficient $4 - \beta$. Estimating the variance in balls for the built-up area in different cities the clear power law behaviour was observed with coefficients β smaller than 2, that is, with coefficients typical for long-range dependence.

Next, in Section 5.4, the built-up area was analysed as a realization of a non-stationary random closed set. We introduced a one step iterative estimation method that enables us to approximately determine the optimal bandwidth needed for the kernel volume fraction estimator and consequently for other second order estimators based on the volume fraction. The method uses a priory chosen bandwidth h_0 to calculate the preliminary estimates of the volume fraction and correlation function. The volume fraction is estimated by the kernel volume fraction estimator $\hat{m}_{A;h_0}$. The correlation function $\kappa(\boldsymbol{x}, \boldsymbol{y})$ is further assumed to depend only on the norm $\|\boldsymbol{x} - \boldsymbol{y}\|$, where the validity of this assumption is also discussed later in the same section. Under this approximate assumption the observed results of the discrete correlation function estimator $\hat{\kappa}_{p;h_0}$ may be well fitted by the Cauchy correlation function, which is then taken as the preliminary estimator of the true correlation function. Those preliminary estimates enables us to approximate the mean integrated square error function as a function of the bandwidth h and consequently to numerically find its minimizing value h_1. Based on h_1, which is taken as the optimal bandwidth, we finally performed the analogous estimation of the volume fraction and of the correlation function. The results of the discrete correlation function estimator $\hat{\kappa}_{p;h_1}$ based on the optimal bandwidth h_1 were again well fitted by the Cauchy correlation function. This may be interpreted as another evidence for the long-range dependence of the built-up structure. Finally in this section, numerical simulations based on non-stationary level excursion sets were performed to confirm the previously introduced iterative method. The results clearly reflected the benefit of performing such a procedure.

Section 5.5 contains a brief summary of both stationary and non-stationary random closed set approaches to the analysis of the built-up area. We argued that based on the results the built-up structure should be viewed as a long-range dependent structure.

In Section 5.6 the radial density dependence was analysed. The power law decay mentioned in previous works (e.g. [9, 10, 1, 18]) was visible on large distances from the city centre. The next section is devoted to the study of the distribution of individual building sizes. The privileged position is given by the generalized beta distribution of the second kind (GB2). We observed two different situations. For the first part of cities the building sizes are approximately distributed according to the GB2 distribution, whereas for the rest of the cities the best approximation is given by the mixture of two GB2 distributions. Using this result and under the additional assumption that individual buildings are balls we further derived an approximate analytic expression for the covariance $C(r)$ and consequently for the correlation function $\kappa(r)$ for small values of r. This expression was then compared with observed values of the correlation function.

In the last section we discussed the relation of the observed long-range dependence on the fractal behaviour that is probably the most often mentioned morphological characteristic in previous studies (see the introduction for references). Based on our observations we can conclude that the built-up structure in cities is a long range dependent meaning that the correlation function of its random closed set representation has a power law decay. From the physical point of view this corresponds to the very important similarity with critical systems, i.e. systems at the critical point of the second order phase transition ([27, 28]). This fact supports the hypothesis of the self-organized criticality of urban systems mentioned by some earlier works (see e.g. [9, 1, 10, 26, 18, 16]).

A possible further extension of the analysis of the built-up structure may be given by using the theory of Gibbs point processes. Here, it may be interesting to determine the energy and thus the potential of the interaction between individual buildings.

Appendix A

Selected parts from Topology and Measure theory

For the convenience of the reader we repeat in this chapter the relevant material from various sources with references, thus making our exposition self-contained. The proofs are given only when we present a new extension of known results or when they are important for later use.

In the following we shall use the set of **extended real numbers**

$$[-\infty, +\infty] = \mathbb{R} \cup \{-\infty, +\infty\}$$

with obvious ordering and with algebraic operations defined in an usual way. The undefined expressions are $\pm\infty + \mp\infty$ and $\pm\infty \cdot 0$.

A.1 General measure theory

Here we mention only the vital part of measure theory. For more details we refer the reader to [46, 111]. Measures are constructed on σ-algebras.

Definition A.1.1. A collection $\mathcal{A}$ of subsets of some space Ω is a **σ-algebra** if

(a) $\Omega \in \mathcal{A}$,

(b) $A \in \mathcal{A}$ implies $A^c \in \mathcal{A}$ (complement),

(c) if $A_1, A_2, \ldots \in \mathcal{A}$, then $A_1 \cup A_2 \cup \ldots \in \mathcal{A}$.

We also recall the definition of a measure.

Definition A.1.2. Let Ω be a space and $\mathcal{A}$ a σ-algebra of subsets of Ω. A **measure** μ on Ω or on $(\Omega, \mathcal{A})$ is a map $\mu : \mathcal{A} \to [0, \infty]$ that satisfies the two conditions:

(a) $\mu(\emptyset) = 0$,

(b) if $A_1, A_2, \ldots \in \mathcal{A}$ is a disjoint sequence, then

$$\mu\left(\bigcup_{k=1}^{\infty} A_k\right) = \sum_{k=1}^{\infty} \mu(A_k).$$

A measure μ is **finite** if $\mu(\Omega) < \infty$ and it is a **probability** measure if $\mu(\Omega) = 1$. If $\Omega = A_1 \cup A_2 \cup \ldots$ for some finite or countable sequence of sets from $\mathcal{A}$ satisfying $\mu(A_k) < \infty$, then μ is called **σ-finite**.

Let E be a topological space. The smallest σ-algebra containing all open sets of E is called a **Borel σ-algebra** and we denoted it by $\mathcal{B}(E)$. The sets from $\mathcal{B}(E)$ are called **Borel sets**. A measure μ defined on the σ-algebra that contains all Borel sets is called a **Borel measure**. A Borel measure μ is **locally finite** if for every $x \in E$ there is an open neighbourhood U of x such that $\mu(U) < \infty$. Thus, a locally finite Borel measure is finite on compact sets.

It is useful to recall the basic continuity properties of measures. If $\{A_n\}$ is a sequence of sets increasing in the sense of inclusion (i.e. $A_n \subset A_{n+1}$ for all n) and $A = \bigcup_n A_n$, we use the notation $A_i \nearrow A$. Similarly, if $\{A_n\}$ is a sequence of sets decreasing in the sense of inclusion (i.e. $A_n \supset A_{n+1}$ for all n) and $A = \bigcap_n A_n$, we use the notation $A_i \searrow A$.

Theorem A.1.1. *Let μ be a measure on a measurable space $(\Omega, \mathcal{A})$.*

(a) (monotonicity) If $A, B \in \mathcal{A}$ and $A \subset B$, then $\mu(A) \leq \mu(B)$.

(b) (continuity form below) If $A, A_1, A_2, \ldots \in \mathcal{A}$ and $A_i \nearrow A$, then $\mu(A_i) \nearrow \mu(A)$.

(c) (continuity form above) If $A, A_1, A_2, \ldots \in \mathcal{A}$ and $A_i \searrow A$ and if $\mu(A_1) < +\infty$, then $\mu(A_i) \searrow \mu(A)$.

(d) If μ is σ-finite on $\mathcal{A}$, then $\mathcal{A}$ cannot contain an uncountable, disjoint collection of sets of positive μ-measure.

Proof. Theorem 10.2 in [46]. □

An important example of a finite Borel measure is a **Dirac measure** at x denoted by δ_x, which is defined for every Borel set B by

$$\delta_x(B) = \begin{cases} 1 & \text{if } x \in B, \\ 0 & \text{otherwise.} \end{cases} \tag{A.1}$$

More generally, we say that a Borel measure μ on a separable metric space E has an atom at $x \in \mathbb{E}$ if $\mu(\{x\}) > 0$. If μ is σ-finite, then it can have at most countably many atoms $\{x_j | j \in \mathbb{N}\}$. Writing $b_i = \mu(\{x_i\})$, the measure μ can be uniquelly decomposed into

$$\mu = \mu_a + \mu_d,$$

where $\mu_a = \sum_{i=1}^{\infty} b_i \delta_{x_i}$ is the **atomic** component of μ and ν_d is the **diffuse** component of μ.

Now let us introduce the concept of an outer measure, which is necessary for the geometric measure theory, discussed in the following section.

Definition A.1.3. A set function $\eta : \{A | A \subset \Omega\} \to [0, \infty]$ is called an **outer measure** on Ω if

(a) $\eta(\emptyset) = 0$,

(b) $\eta(A) \leq \eta(B)$ whenever $A \subset B \subset \Omega$,

(c) $\mu\left(\bigcup_{k=1}^{\infty} A_k\right) \leq \sum_{k=1}^{\infty} \mu(A_k)$ whenever $A_1, A_2, \ldots \subset \Omega$.

Note that every measure μ defined on $(\Omega, \mathcal{A})$ may be extended to an outer measure μ^* on Ω by

$$\mu^*(A) = \inf\{\mu(B) | A \subset B \in \mathcal{A}\}.$$

On the other side, an outer measure gives an ordinary measure when restricted to a collection of the so called measurable sets. Let μ be the outer measure on Ω. A set $A \subset \Omega$ is μ **measurable** if

$$\mu(B) = \mu(B \cap A) + \mu(B \setminus A) \quad \text{for all } B \subset \Omega.$$

Theorem A.1.2. *Let μ be an outer measure and let $\mathcal{M}$ be the collection of all μ measurable subsets of Ω. Then $\mathcal{M}$ is a σ-algebra. Moreover if $\mu(A) = 0$, then $A \in \mathcal{M}$.*

Proof. Theorem 11.1 in [46]. □

Let E be a topological space. An outer measure μ on E is called **Borel** if all Borel sets of E are measurable. Furthermore, μ is **Borel regular** if μ is Borel and for every $A \subset E$ there is a Borel set $B \subset E$ such that $A \subset B$ and $\mu(A) = \mu(B)$. By a **Radon measure** we mean a Borel regular measure μ on a locally compact second countable topological space E such that $\mu(C) < \infty$ for compact sets $C \subset E$, $\mu(U) = \sup\{\mu(C) | C \subset U \text{ is compact}\}$ for every open set $V \subset E$, and $\mu(A) = \sup\{\mu(U) | U \subset A \text{ is open}\}$ for every $A \subset E$.

The map $f : \Omega \to G$ from some space with σ-algebra $\mathcal{A}$ to a topological space G is **measurable** if $f^{-1}(A) \in \mathcal{A}$ for every open set A in G. The map $f : E \to G$ from some topological space E to a topological space G is called **Borel measurable** if $f^{-1}(A)$ is a Borel set in E for every open set A in G. When checking the Borel measurability of some mapping the following theorem is often useful.

Theorem A.1.3. *Let $f : E \to G$ be a map from a topological space E to a topological space G. Let further Γ be the system of sets from G that generates the Borel σ-algebra $\mathcal{B}(G)$ on G, i.e. $\mathcal{B}(G)$ is the minimal σ-algebra containing Γ. Then f is Borel measurable if and only if $f^{-1}(A) \in \mathcal{B}(E)$ for every $A \in \Gamma$.*

Proof. See 2.3.2 in [65]. □

Note that one can particularly take $\Gamma = \mathcal{B}(G)$ and conclude that f is Borel measurable if and only if $f^{-1}(A) \in \mathcal{B}(E)$ for all Borel sets A in G.

We often use the concept of the weak convergence in metric spaces. Let S be a metric space and $\mathcal{B}(S)$ be a Borel σ-algebra generated by the open sets of S. From now on, all measures are taken on $(S, \mathcal{B}(S))$.

The sequence of finite measures $\{\mu_n\}_{n\in\mathbb{N}}$ **converges weakly** to a measure μ , $\mu_n \xrightarrow{\text{w}} \mu$, if

$$\int_S f \,\mathrm{d}\mu_n \to \int_S f \,\mathrm{d}\mu \quad \text{for every } f \text{ in } C_b(S), \tag{A.2}$$

where $C_b(S)$ denote all bounded, continuous real functions on S.

The following theorem provides useful conditions equivalent to the weak convergence. Usually it is mentioned only for the weak convergence of probability measures but we need it in the more general form.

Theorem A.1.4 (Portmanteau theorem)**.** *Let $\mu, \mu_1, \mu_2, \ldots$ be finite measures on $(S, \mathcal{B}(S))$. Then the following conditions are equivalent:*

(a) $\mu_n \xrightarrow{\text{w}} \mu$,

(b) $\int f \,\mathrm{d}\mu_n \to \int f \,\mathrm{d}\mu$ *for all bounded, uniformly continuous real* f,

(c) $\limsup_{n\to\infty} \mu_n(F) \le \mu(F)$ *for every closed* F *and* $\mu_n(S) \to \mu(S)$,

(d) $\liminf_{n\to\infty} \mu_n(G) \ge \mu(G)$ *for all open* G *and* $\mu_n(S) \to \mu(S)$,

(e) $\lim_n \mu_n(A) = \mu(A)$ *for every Borel* A *such that* $\mu(\partial A) = 0$.

Proof. The proof in the case of probability measures is given in [111] as a proof of Theorem 2.1. Here we give arguments only when there is a difference from that proof. In (b)⇒(c): Taking $f \equiv 1$ in (b) gives $\mu_n(S) \to \mu(S)$. In (c)⇔(d): From complementation $G = S \setminus F$ we have

$$\begin{aligned}\liminf_{n\to\infty} \mu_n(G) = \liminf_{n\to\infty} \big(\mu_n(S) - \mu_n(F)\big) &\geq \liminf_{n\to\infty} \mu_n(S) + \liminf_{n\to\infty} \big(-\mu_n(F)\big) \\ &\geq \liminf_{n\to\infty} \mu_n(S) - \limsup_{n\to\infty} \mu_n(F) = \mu(S) - \mu(F) = \mu(G).\end{aligned}$$

The opposite implication is analogous. Proofs of (a)⇒(b), (c)&(d)⇒(e), (e)⇒(a) and the rest of (b)⇒(c) can be done exactly in the same way as for probability measures. □

The following theorem ensures the existence of a continuous bounded function that equals 1 on a given set and vanishes outside of the ε-neighbourhood of it. If $d(x, y)$ is a distance of points in S then the **distance from** x **to** F is given by $d(x, F) = \inf\{d(x, y)|y \in F\}$.

Theorem A.1.5. *Let F be a closed set in S with metric $\rho(x, y)$ and $\varepsilon > 0$. Then there is a uniformly continuous function $f \in C_b(S)$ such that $f(x) = 1$ for $x \in F$, $f(x) = 0$ for $d(x, F) \geq \varepsilon$, and $0 \leq f(x) \leq 1$ for all $x \in S$.*

Proof. From the triangle inequality for distance d follows that $d(x, A) \leq d(x, y) + d(y, A)$ and $d(x, A)$ is therefore uniformly continuous. If we define

$$f(x) = \begin{cases} 1 - \frac{d(x,F)}{\varepsilon}, & \text{if } d(x, F) \leq \varepsilon, \\ 0, & \text{if } \varepsilon < d(x, F), \end{cases}$$

then f has the required properties. □

In the special case $S = \mathbb{R}^d$ there is another equivalent criterion of the weak continuity for probability measures. Before presenting the statement let us recall one important object. The **distribution function** F of a probability measure $\mathbb{P}$ is given by

$$F(\boldsymbol{x}) = \mathbb{P}\left((-\infty, \boldsymbol{x}]\right) = \mathbb{P}\left((-\infty, x_1] \times \ldots \times (-\infty, x_d]\right)$$

for all $x = (x_1, \ldots, x_d) \in \mathbb{R}^d$. The properties of distribution functions can be formulated as follows.

Lemma A.1.1. *A distribution function $F(\boldsymbol{x})$ have the following properties*

(a) F is continuous from above (i.e. $\lim_{\boldsymbol{x}\to\boldsymbol{y}_+} F(\boldsymbol{x}) = F(\boldsymbol{y})$ for all $\boldsymbol{y}$),

(b) $0 \leq F(\boldsymbol{x}) \leq 1$ for all $\boldsymbol{x} \in \mathbb{R}^d$, F is non-decreasing in each coordinate, and for each d-dimensional rectangle $(\boldsymbol{a}, \boldsymbol{b}]$ holds

$$\sum_{(\theta_1,\ldots,\theta_d)\in\{0,1\}^d} (-1)^{\sum \theta_i} F\big(a_1 + \theta_1(b_1 - a_1), \ldots, a_d + \theta_d(b_d - a_d)\big) \geq 0.$$

(c) $F(\boldsymbol{x}) \to 0$ as any one coordinate of $\boldsymbol{x}$ goes to $-\infty$, and $F(\boldsymbol{x}) \to 1$ as all coordinates of $\boldsymbol{x}$ goes to $+\infty$.

Moreover, if F is a real function satisfying those properties, there exists a unique probability measure $\mathbb{P}$ on $\mathbb{R}^d$ such that F is a distribution function of $\mathbb{P}$.

Proof. Theorem 12.5 in [46]. □

Theorem A.1.6. *A sequence of probability measures $\{\mathbb{P}_i\}_{i\in\mathbb{N}}$ on $\mathbb{R}^d$ converges weakly to a probability measure $\mathbb{P}$ on $\mathbb{R}^d$ if and only if*

$$\lim_{n\to+\infty} F_n(\boldsymbol{x}) = F(\boldsymbol{x}) \quad \text{for all continuity points } \boldsymbol{x} \in \mathbb{R}^d \text{ of } F, \tag{A.3}$$

where F_i, F for all $i \in \mathbb{N}$ are distribution functions of probability measures $\mathbb{P}_i, \mathbb{P}$, respectively.

Proof. Theorem 29.1 in [46]. □

A.2 Fourier analysis of positive semi-definite measures

In order to be able to correctly define the covariance measure and the Bartlett spectrum of a random measure, the generalized theory of signed measures have to be used. It is based on the view of a measure as a continuous liner functional on the space $C_c(E)$ of continuous complex-valued functions on a locally compact, Hausdorff space E having compact support, see e.g. [112]. The continuity of a linear functional μ means the following condition: for every compact subset K of E, there is a positive number M_K such that, for every function $f \in C_c(E)$ whose support is contained in K,

$$|\mu(f)| \leq M_k \sup_{x \in E} |f(x)| . \tag{A.4}$$

A **complex measure** μ on E is a continuous linear functional on $C_c(E)$. A **signed measure** μ on E is a complex measure taking only real values on real functions, i.e. $\mu(f) \in \mathbb{R}$ for all $f \in C_c(E)$ such that $f(x) \in \mathbb{R}$ for every $x \in E$. A **positive measure** μ on E is a signed measure taking only non-negative values on non-negative functions, i.e. $\mu(f) \geq 0$ for all $f \in C_c(E)$ such that $f \geq 0$. We say that a complex measure μ is **bounded** if there exists a finite number $M \geq 0$ such that, for every function $f \in C_c(E)$,

$$|\mu(f)| \leq M \sup_{x \in E} |f(x)| .$$

Bounded measures are continuous in the topology of uniform convergence in E. For details see [112, Chapter III §1.8]. Note that the previously defined terminology is not mixed with the terminology of ordinary measure theory from previous section.

Given a locally finite Borel measure $\tilde{\mu}$ (in the sense of the previous section) on E, it follows from the linearity of the Lebesgue integral that a functional μ defined for every $f \in C_c(E)$ by

$$\mu(f) = \int_E f \, \mathrm{d}\tilde{\mu}$$

is a positive linear functional and thus a positive measure. The converse statement that a positive measure corresponds, in the sense of previous formula, to a unique locally finite Borel (Radon) measure is a result called the Riesz representation theorem, see e.g. [113, Theorem 2.14]. Hence every positive measure is an ordinary locally finite Borel measure in the sense of the previous section. The same holds for complex measures ([113, Theorem 6.19]). From this reason, given a signed (positive) measure μ we use the notation $\int f \, \mathrm{d}\mu$, $\int f(x)\mu(\mathrm{d}x)$ for $\mu(f)$. It is clear that a positive measure μ is bounded if and only if $\tilde{\mu}$ is bounded, i.e. $\tilde{\mu}(E) < \infty$.

For a complex measure μ its **total variation** $|\mu|$ is defined by

$$|\mu|(f) = \sup_{|g| \leq f,\ g \in C_c(E)} |\mu(g)|$$

for all $f \in C_c(E)$. It can be shown (e.g. [112, Chapter III §1.6]) that $|\mu|$ is a positive measure. The complex measure μ is bounded if and only if $|\mu|$ is bounded. From the definition follows

$$|\mu(f)| \leq |\mu|(|f|)$$

for every $f \in C_c(E)$. The **upper variation** μ^+ and **lower variation** μ^- of a signed measure μ are positive measures given by

$$\mu^+ = \frac{1}{2}(|\mu| + \mu) \quad \text{and} \quad \mu^- = \frac{1}{2}(|\mu| - \mu),$$

respectively. It is clear that

$$\mu = \mu^+ - \mu^- \quad \text{and} \quad |\mu| = \mu^+ + \mu^-.$$

If μ is a positive measure, then $\mu = |\mu| = \mu^+$.

In order order to extend the integration outside of $C_c(E)$ for a signed measure μ we use the notation $f \in L^1(E, \mathrm{d}\mu)$ if $f \in L^1(E, \mathrm{d}\,|\mu|)$ (in the usual sense for the Borel measure $|\mu|$) meaning

$$\int |f| \ \mathrm{d}\,|\mu| < \infty.$$

Thus by definition $L^1(E, \mathrm{d}\mu) = L^1(E, \mathrm{d}\,|\mu|)$. For every $f \in L^1(\mathrm{d}\mu)$ we set

$$\int f \,\mathrm{d}\mu = \int f \,\mathrm{d}\mu^+ - \int f \,\mathrm{d}\mu^-.$$

Both integrals are finite so their difference is well defined.

The theory of the Fourier transform of complex measures in the previous sense is developed on locally compact abelian groups. The basic references are [114] and [115]. In the following we restrict ourselves only on $\mathbb{R}^d$.

Note that on $\mathbb{R}^d$ every measurable function g that is locally integrable with respect to the Lebesgue measure ν_d defines a complex measure $g\nu_d$ by

$$g\nu_d(f) = \int\limits_{\mathbb{R}^d} g(\boldsymbol{x}) f(\boldsymbol{x}) \,\mathrm{d}\boldsymbol{x} \tag{A.5}$$

for every $f \in C_c(\mathbb{R}^d)$. The **involution** f^* of a complex function f on $\mathbb{R}^d$ is

$$f^*(x) = \overline{f(-x)} \quad \text{for all} \quad \boldsymbol{x} \in \mathbb{R}^d. \tag{A.6}$$

The **convolution** of functions $f, g \in L^1(\mathbb{R}^d, \nu_d)$ is

$$(f * g)(\boldsymbol{x}) = \int\limits_{\mathbb{R}^d} f(\boldsymbol{y}) g(\boldsymbol{x} - \boldsymbol{y}) \,\mathrm{d}\boldsymbol{y} \quad \text{for all} \quad \boldsymbol{x} \in \mathbb{R}^d.$$

If one of functions f, g is continuous and one, possibly the same, has compact support, then $f * g$ is continuous and if $f, g \in C_c(\mathbb{R}^d)$, then $f * g \in C_c(\mathbb{R}^d)$, see [116, Chapter VIII §4.5 Proposition 11].

Proposition A.2.1. *Let $f \in L^1(\mathbb{R}^d, \nu_d)$ and $g \in L^\infty(\mathbb{R}^d, \nu_d)$. Then $f * g$ exists at every point $\boldsymbol{x}$ and is uniformly continuous.*

Proof. Proposition 14 in Chapter VIII §4.5 in [116]. □

Now the positive semi-definiteness may be defined. Let us begin with a notion for functions. We say that a complex valued function f defined on $\mathbb{R}^d$ is **positive semi-definite** if

$$\sum_i \sum_j \lambda_i \overline{\lambda_j} f(\boldsymbol{x}_i - \boldsymbol{x}_j) \geq 0$$

for all $n \in \mathbb{N}$, $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n \in \mathbb{R}^d$ and $\lambda_1, \ldots, \lambda_n \in \mathbb{C}$.

Proposition A.2.2. *A continuous function $g : \mathbb{R}^d \to C$ is positive semi-definite if and only if*

$$\int\limits_{\mathbb{R}^d} g(\boldsymbol{x})(f * f^*)(\boldsymbol{x}) \,\mathrm{d}\boldsymbol{x} \geq 0$$

for all $f \in C_c(\mathbb{R}^d)$.

Proof. Proposition 4.1 in [115]. □

Definition A.2.1. A complex measure μ is **positive semi-definite** if for every $f \in C_c(\mathbb{R}^d)$

$$\mu(f * f^*) = \int_{\mathbb{R}^d} (f * f^*)(\boldsymbol{x})\ \mu(\mathrm{d}\boldsymbol{x}) \geq 0.$$

It is clear that a complex measure $g\nu_d$ which corresponds to a continuous function g according to (A.5) is positive semi-definite if and only if g is positive semi-definite.

Proposition A.2.3. *A complex measure μ is positive semi-definite if and only if for every bounded measurable function f with bounded support*

$$\mu(f * f^*) = \int_{\mathbb{R}^d} (f * f^*)(\boldsymbol{x})\ \mu(\mathrm{d}\boldsymbol{x}) \geq 0.$$

Proof. One implication is obvious. We prove that if μ is positive semi-definite, then the integral is positive for every bounded measurable function with bounded support. Since $C_c(\mathbb{R}^d)$ is dense in $L^1(\mathbb{R}^d, \nu_d)$, there exist a sequence $\{f_n\}$ with $f_n \in C_c(\mathbb{R}^d)$, such that $f_n \to f$ in $L^1(\mathbb{R}^d, \nu_d)$ as $n \to \infty$. Moreover, as follows from Luzin's theorem [113, Theorem 2.23], $\{f_n\}$ can be chosen such that there is a positive number $K > 0$, a compact set $C \subset \mathbb{R}^d$, and for all n holds: $f_n(\boldsymbol{x}) = 0$ for all $\boldsymbol{x} \in \mathbb{R}^d \setminus C$ and $|f_n(\boldsymbol{x})| \leq K$ for all $\boldsymbol{x} \in C$. Without loss of generality we assume that the same holds for f, i.e. $f(\boldsymbol{x}) = 0$ for all $\boldsymbol{x} \in \mathbb{R}^d \setminus C$ and $|f(\boldsymbol{x})| \leq K$ for all $\boldsymbol{x} \in C$.

Therefore

$$\begin{aligned}
|(f_n * f_n^*)(\boldsymbol{x}) - (f * f^*)(\boldsymbol{x})| \leq & |(f_n * f_n^*)(\boldsymbol{x}) - (f_n * f^*)(\boldsymbol{x})| + |(f_n * f^*)(\boldsymbol{x}) - (f * f^*)(\boldsymbol{x})| \\
\leq & \|f_n^*\|_\infty \int |f_n^*(\boldsymbol{x} - \boldsymbol{y}) - f^*(\boldsymbol{x} - \boldsymbol{y})|\ \mathrm{d}\boldsymbol{y} \\
& + \|f^*\|_\infty \int |f_n^*(\boldsymbol{x} - \boldsymbol{y}) - f^*(\boldsymbol{x} - \boldsymbol{y})|\ \mathrm{d}\boldsymbol{y} \\
\leq & 2K \|f_n - f\|_1 ,
\end{aligned}$$

where $\|\cdot\|_1$ is the norm in $L^1(\mathbb{R}^d, \nu_d)$ and $\|f\|_\infty = \sup_{\boldsymbol{x} \in \mathbb{R}^d} |f(\boldsymbol{x})|$. We have used the obvious property of isometry of the involution $f \mapsto f^*$ on L^1. The sequence $\{f_n * f_n^*\}$ thus converges to $f * f^*$ uniformly. Proposition A.2.1 implies that $f_n * f_n^*$ for all n and $f * f^*$ are uniformly continuous. Moreover they all have bounded support that is included in the support of $\mathbb{1}_C * \mathbb{1}_C^*$. The continuity (A.4) of μ yields

$$\int_{\mathbb{R}^d} (f_n * f_n^*)(\boldsymbol{x})\ \mu(\mathrm{d}\boldsymbol{x}) \to \int_{\mathbb{R}^d} (f * f^*)(\boldsymbol{x})\ \mu(\mathrm{d}\boldsymbol{x}) \quad \text{as } n \to \infty.$$

Hence the assertion follows. □

Before the definition of the Fourier transform of a measure we present some know results from the Fourier analysis on $\mathbb{R}^d$. The Fourier transform $\hat{f} \equiv \mathscr{F} f$ and the inverse Fourier transform $\check{f} \equiv \mathscr{F}^{-1} f$ of a function $f \in L^1(\mathbb{R}^d, \nu_d)$ are given by

$$\hat{f}(\boldsymbol{\xi}) = (2\pi)^{-d/2} \int_{\mathbb{R}^d} e^{-i\boldsymbol{\xi}\cdot\boldsymbol{x}} f(\boldsymbol{x})\, \mathrm{d}\boldsymbol{x}, \tag{A.7}$$

$$\check{f}(\boldsymbol{x}) = (2\pi)^{-d/2} \int_{\mathbb{R}^d} e^{i\boldsymbol{\xi}\cdot\boldsymbol{x}} f(\boldsymbol{\xi})\, \mathrm{d}\boldsymbol{\xi}. \tag{A.8}$$

Note that $\hat{f}(-\boldsymbol{\xi}) = \check{f}(\boldsymbol{\xi})$, $\mathscr{F}^{-1} f^* = \overline{\hat{f}}$, and $\mathscr{F}^{-1}(f * f^*) = (2\pi)^{d/2} \left|\check{f}\right|^2$, which is a well known property (see e.g. [92, Theorem IX.3]).

The Fourier transform can be naturally extended to any bounded complex measure μ by setting

$$\hat{f}_\mu(\boldsymbol{\xi}) = (2\pi)^{-d/2} \int_{\mathbb{R}^d} e^{-i\boldsymbol{\xi}\cdot\boldsymbol{x}} \mu(\mathrm{d}\boldsymbol{x}) \tag{A.9}$$

for all $\boldsymbol{\xi} \in \mathbb{R}^d$. It can be shown that $\hat{f}_\mu$ is a uniformly continuous, bounded function. Thus $\hat{f}_\mu \nu_d$ is a complex measure.

Theorem A.2.1 (The inversion formula)**.** *Let μ be a bounded, complex measure on $\mathbb{R}^d$. If the Fourier transform $\hat{f}_\mu \in L^1(\mathbb{R}^d, \nu_d)$, then the measure μ is absolutely continuous with respect to ν_d, and its density f_μ is given by*

$$f_\mu(\boldsymbol{x}) = (2\pi)^{-d/2} \int_{\mathbb{R}^d} e^{i\boldsymbol{\xi}\cdot\boldsymbol{x}} \hat{f}_\mu(\boldsymbol{\xi})\, \mathrm{d}\boldsymbol{\xi}.$$

Proof. Theorem 2.6 in [115]. □

If we restrict ourselves to positive bounded measures, the Fourier transform is positive semi-definite function. Moreover, the following equivalence holds.

Theorem A.2.2 (Bochner)**.** *A continuous function f is positive semi-definite if and only if there exists a positive bounded measure σ on $\mathbb{R}^d$ such that*

$$f(\boldsymbol{x}) = (2\pi)^{-d/2} \int_{\mathbb{R}^d} e^{i\boldsymbol{\xi}\cdot\boldsymbol{x}}\ \sigma(\mathrm{d}\boldsymbol{\xi}).$$

Proof. Theorem IX.9 in [92]. □

Now we may define the Fourier transform of a generally unbounded, positive semi-definite, complex measure μ. The base of the definition is stated by the following theorem.

Theorem A.2.3. *Let μ be a positive semi-definite measure on $\mathbb{R}^d$. Then there exists a unique positive measure $\hat{\mu}$ on $\mathbb{R}^d$ such that*

$$\mu(f * f^*) = (2\pi)^{d/2} \hat{\mu}\big(|\check{f}|^2\big)$$

for all $f \in C_c(\mathbb{R}^d)$.

Proof. Theorems 4.5 and 4.7 in [115]. □

Definition A.2.2. The **Fourier transform** $\mathscr{F}\mu$ of a positive semi-definite measure μ on $\mathbb{R}^d$ is the positive measure $\hat{\mu}$ on $\mathbb{R}^d$ associated with μ by the previous theorem.

It can be shown that if μ is absolutely continuous with respect to the continuous positive semi-definite function f_μ, then the Fourier transform $\hat{\mu}$ is the same as the measure σ corresponding to f_μ from the Bochner theorem, see [115, Proposition 4.3]. Thus

$$f_\mu(\boldsymbol{x}) = (2\pi)^{-d/2} \int_{\mathbb{R}^d} e^{i\boldsymbol{\xi}\cdot\boldsymbol{x}}\ \hat{\mu}(\mathrm{d}\boldsymbol{\xi}).$$

The connection between the Fourier transform and the extension of the Fourier transform on bounded measures defined by (A.9) is given by the following proposition.

Proposition A.2.4. *The bounded complex measure μ is positive semi-definite if and only if $\hat{f}_\mu(\boldsymbol{\xi}) \geq 0$ for all $\boldsymbol{\xi} \in \mathbb{R}^d$, where $\hat{f}_\mu$ is defined by (A.9). If μ is a bounded positive semi-definite complex measure, the Fourier transform $\hat{\mu}$ of μ is equal to $\hat{f}_\mu \nu_d$.*

Proof. Proposition 4.14 in [115]. □

The Fourier transform of a bounded positive semi-definite complex measure is therefore absolutely continuous with respect to ν_d, and its density is given by (A.9). By the Inversion formula A.2.1 follows that if $\hat{f}_\mu \in L^1(\mathbb{R}^d, \nu_d)$, then μ is also absolutely continuous with density $f_\mu = \mathscr{F}^{-1}\hat{f}_\mu$.

Proposition A.2.5. *Let $\hat{\mu}$ be the Fourier transform of a positive semi-definite measure μ on $\mathbb{R}^d$ and f be a bounded measurable function with bounded support. Then $\check{f} \in L^2(\mathbb{R}^d, \hat{\mu})$ and*

$$\mu(f * f^*) = (2\pi)^{d/2}\hat{\mu}\big(|\check{f}|^2\big).$$

Proof. Analogously as in the proof of Proposition A.2.3 there exist a sequence $\{f_n\}$ with $f_n \in C_c(\mathbb{R}^d)$ such that $f_n \to f$ in $L^1(\mathbb{R}^d, \nu_d)$ as $n \to \infty$, $f_n * f_n^*$ converges to $f * f^*$ uniformly on $\mathbb{R}^d$, and

$$\mu(f_n * f_n^*) \to \mu(f * f^*) \quad \text{as } n \to \infty. \tag{A.10}$$

For every n we have $\sup_{\boldsymbol{x}\in\mathbb{R}^d} |\check{f}_n(\boldsymbol{x})| \le (2\pi)^{-d/2} \|f_n\|_1$, where $\|\cdot\|_1$ is the norm in $L^1(\mathbb{R}^d, \nu_d)$. Hence $\check{f}_n \to \check{f}$ uniformly on $\mathbb{R}^d$ as $n \to \infty$, which yields $|\check{f}_n|^2 \to |\check{f}|^2$ uniformly as $n \to \infty$. From Theorem A.2.3 follows that for every n is $\check{f}_n \in L^2(\mathbb{R}^d, \check{\mu})$ and

$$\mu(f_n * f_n^*) = (2\pi)^{d/2}\hat{\mu}\big(|\check{f}_n|^2\big).$$

From (A.10) follows that $\{|\check{f}_n|^2\}$ is a Cauchy sequence in $L^1(\mathbb{R}^d, \hat{\mu})$. Finally, the statement is a consequence of the completeness of $L^1(\mathbb{R}^d, \hat{\mu})$, see e.g. [113, Theorem 3.11]. □

This particularly gives that the Fourier transform of an indicator of a bounded Borel set A is $\hat{\mu}$ squared integrable, the result presented in [117]. It is easy to see that both the Dirac measure $\delta_{\mathbf{0}}$ at $\mathbf{0}$ and the Lebesgue measure ν_d are positive semi-definite.

Proposition A.2.6. *The Fourier transform of $\delta_{\mathbf{0}}$ is ν_d and vice versa.*

Proof. This easily follows from Theorem A.2.3 and Parseval's identity implied by the Plancherel theorem (see e.g. [92, Theorem IX.6]). □

The complex measure μ is said to be **translation bounded** if for every compact set $C \subset \mathbb{R}^d$

$$\sup_{\boldsymbol{x}\in\mathbb{R}^d} |\mu|\,(C + \boldsymbol{x}) < \infty.$$

Proposition A.2.7. *The Fourier transform $\hat{\mu}$ of a positive semi-definite measure μ on $\mathbb{R}^d$ is translation bounded.*

Proof. Proposition 4.9 in [115]. □

Finally, we introduce the concept of the approximation of the identity. Let ϕ be an integrable function on $\mathbb{R}^d$ such that $\int_{\mathbb{R}^d} \phi(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} = 1$ and for $a > 0$ let take $\phi_a(\boldsymbol{x}) = a^{-d}\phi(a^{-1}\boldsymbol{x})$. The set $\{\phi_a | a > 0\}$ is called an **approximation of the identity**. The well known result of approximations of the identity is their convergence to delta function.

Theorem A.2.4. *Let μ be a complex measure and $\{\phi_a | a > 0\}$ be an approximation of the identity. If f is bounded, uniformly continuous, then*

$$\lim_{a\to 0_+} (\phi_a * f)(\boldsymbol{x}) = f(\boldsymbol{x})$$

uniformly and if $f \in C_c(\mathbb{R}^d)$, then also in $L^p(\mathbb{R}^d, \mu)$.

Proof. We follow the proof of Theorem 2.1 in [118]. Since ϕ has integral 1,

$$(\phi_a * f)(\boldsymbol{x}) - f(\boldsymbol{x}) = \int_{\mathbb{R}^d} \phi(\boldsymbol{y})\big(f(\boldsymbol{x} - a\boldsymbol{y}) - f(\boldsymbol{x})\big)\,\mathrm{d}\boldsymbol{y}.$$

By the uniform continuity of f, given $\varepsilon > 0$ one can choose $\delta > 0$ such that if $\|h\| < \delta$,

$$|f(\boldsymbol{x} + h) - f(\boldsymbol{x})| \leq \frac{\varepsilon}{2\,\|\phi\|_1}$$

for all $\boldsymbol{x} \in \mathbb{R}^d$, where $\|\phi\|_1$ is the $L^1(\mathbb{R}^d, \nu_d)$ norm of ϕ. Let $K = \sup_{\boldsymbol{x}\in\mathbb{R}^d} f(\boldsymbol{x}) < \infty$. For now fixed δ, by integrability of ϕ, we can take sufficiently small $t > 0$ such that

$$\int_{\|\boldsymbol{y}\|\geq\delta/t} |\phi(\boldsymbol{y})|\,\mathrm{d}\boldsymbol{y} \leq \frac{\varepsilon}{4K}.$$

Finally by the triangular inequality for the absolute value,

$$|(\phi_a * f)(\boldsymbol{x}) - f(\boldsymbol{x})| \leq \int_{\|\boldsymbol{y}\|<\delta/t} |\phi(\boldsymbol{y})|\,|f(\boldsymbol{x} + h) - f(\boldsymbol{x})|\,\mathrm{d}\boldsymbol{y} + 2K \int_{\|\boldsymbol{y}\|\geq\delta/t} |\phi(\boldsymbol{y})|\,\mathrm{d}\boldsymbol{y} < \varepsilon.$$

The proof for the $L^p(\mathbb{R}^d, \mu)$ norm $\|\cdot\|_{p;\mu}$ is analogous. We just use the first bound

$$\|f(\boldsymbol{x} + h) - f(\boldsymbol{x})\|_{p;\mu} \leq \frac{\varepsilon}{2\,\|\phi\|_1},$$

the second bound with $K = \|f\|_{p;\mu}$, and the Minkowski inequality in the last relation. □

Corollary A.2.1. *Let μ be a complex measure, $\{\phi_a | a > 0\}$ be an approximation of the identity, and $f \in C_c(\mathbb{R}^d)$. Then there is $a_0 > 0$ such that $(\phi_a * f) \in L^1(\mathbb{R}^d, \mu)$ for all $0 < a < a_0$.*

Now we may state the important proposition for translation bounded measures.

Proposition A.2.8. *Let $\hat{\mu}$ be the Fourier transform of a translation bounded positive semi-definite measure μ on $\mathbb{R}^d$. If f is a bounded measurable function with bounded support in $\mathbb{R}^d$, then $\check{f} \in L^1(\mathbb{R}^d, \hat{\mu})$ and*

$$\mu(f) = \hat{\mu}(\check{f}).$$

Proof. First note that

$$\hat{\mathbb{1}}_{[-1,1]^d}(\boldsymbol{y}) = (2\pi)^{-d/2} \int_{[-1,1]^d} e^{-i\boldsymbol{x}\cdot\boldsymbol{y}}\,\mathrm{d}\boldsymbol{x} = (2\pi)^{-d/2} 2^d \prod_{j=1}^{d} \frac{\sin(y_j)}{y_j}.$$

Let take the approximation of the identity (sometimes called the **Fejér kernel**) given by

$$\phi(\boldsymbol{y}) = \frac{\big|\hat{\mathbb{1}}_{[-1,1]^d}\big|^2(\boldsymbol{y})}{2^d} = \prod_{j=1}^{d} \frac{\sin^2(y_j)}{\pi y_j^2}.$$

Clearly $\phi^*(\boldsymbol{y}) = \phi(\boldsymbol{y})$ for all $\boldsymbol{y} \in \mathbb{R}^d$. Theorem A.2.4 yields $\int f * \phi_a^* \mu \to \int f\mu$ as $a \to 0_+$ for every $f \in C_c(\mathbb{R}^d)$. The inverse Fourier transform $\check{\phi}_a$ of ϕ_a is, with $b = 1/a$,

$$\check{\phi}_a(\boldsymbol{\xi}) = \frac{(2\pi)^{-d/2}}{2^d b^d}\,\mathbb{1}_{[-b,b]^d} * \mathbb{1}^*_{[-b,b]^d}(\boldsymbol{\xi}) = \frac{\nu_d\big([-b,b]^d \cap ([-b,b]^d - \boldsymbol{\xi})\big)}{(2\pi)^{d/2}\nu_d([-b,b]^d)},$$

because $\hat{\mathbb{1}}_{[-1,1]^d}(b\boldsymbol{y}) = b^{-d}\hat{\mathbb{1}}_{[-b,b]^d}(\boldsymbol{y})$. Clearly $\check{\phi}_a(\boldsymbol{\xi}) \in C_c(\mathbb{R}^d)$ and $\check{\phi}_a(\boldsymbol{\xi}) \to (2\pi)^{-d/2}$ as $a \to 0_+$ for every $\boldsymbol{\xi} \in \mathbb{R}^d$.

Using polarization identities on the relation $\mu(f * f^*) = (2\pi)^{d/2}\hat{\mu}\big(|\check{f}|^2\big)$ for $f \in C_c(\mathbb{R}^d)$ leads to

$$\mu(f * g^*) = (2\pi)^{d/2}\hat{\mu}\big(\check{f}\overline{\check{g}}\big)$$

for $f, g \in C_c(\mathbb{R}^d)$. Let fix $a > 0$ small enough such that Corollary A.2.1 yields $(f * \phi_a^*) \in L^1(\mathbb{R}^d, \mu)$. Since $\phi_a \in L^1(\mathbb{R}^d)$ and $C_c(\mathbb{R}^d)$ is dense in $L^1(\mathbb{R}^d)$, [113, Theorem 3.14], there exists a sequence $\{g_n\}$, $g_n \in C_c(\mathbb{R}^d)$ for all n, such that $g_n^* \to \phi_a^*$ in $L^1(\mathbb{R}^d)$ as $n \to \infty$ and as a consequence $\overline{g_n} \to \overline{\check{\phi}_a}$ pointwise (uniformly) as $n \to \infty$. For the left side we have

$$|(f * g_n^*)(\boldsymbol{x}) - (f * \phi_a^*)(\boldsymbol{x})| \leq \int_{\mathbb{R}^d} |f(\boldsymbol{x} - \boldsymbol{y})|\, |g_n^*(\boldsymbol{y}) - \phi_a^*(\boldsymbol{y})|\, \mathrm{d}\boldsymbol{y}.$$

From the translation boundedness of μ follows that there exists K such that $\|f(\cdot - \boldsymbol{y})\|_{\mu;1} \leq K\|f\|_{\mu;1}$, where $\|f\|_{\mu;1}$ is the $L^1(\mathbb{R}^d, \mu)$ norm of f. From this follows

$$\begin{aligned}\|(f * g_n^*)(\boldsymbol{x}) - (f * \phi_a^*)(\boldsymbol{x})\|_{\mu;1} &= \int_{\mathbb{R}^d} |(f * g_n^*)(\boldsymbol{x}) - (f * \phi_a^*)(\boldsymbol{x})|\, |\mu|\, (\mathrm{d}\boldsymbol{x}) \\ &\leq K\|f\|_{\mu;1}\, \|g_n^*(\boldsymbol{y}) - \phi_a^*(\boldsymbol{y})\|_1\,.\end{aligned}$$

Hence $\mu(f * g_n^*) \to \mu(f * \phi_a^*)$ as $n \to \infty$ and therefore $\check{f}\overline{\check{g}_n}$ is a Cauchy sequence in $L^1(\mathbb{R}^d, \hat{\mu})$ that, by previous considerations, converges pointwise to $\check{f}\overline{\check{\phi}_a}$. From a completeness of $L^1(\mathbb{R}^d, \hat{\mu})$, [113, Theorem 3.11], follows $\hat{\mu}\big(\check{f}\overline{\check{g}_n}\big) \to \hat{\mu}\big(\check{f}\overline{\check{\phi}_a}\big)$ as $n \to \infty$. Thus we finally obtain

$$\mu(f * \phi_a^*) = (2\pi)^{d/2}\hat{\mu}\big(\check{f}\overline{\check{\phi}_a}\big)$$

for every $f \in C_c(\mathbb{R}^d)$. Now we use Theorem A.2.4 and obtain $\mu(f * \phi_a^*) \to \mu(f)$ as $a \to 0_+$ on the left side. Consequently, the right side also converges to the point-wise limit $\check{f}$, which is in $L^1(\mathbb{R}^d, \hat{\mu})$ as a consequence of the completeness of $L^1(\mathbb{R}^d, \hat{\mu})$. Letting $a \to 0_+$ leads to

$$\int_{\mathbb{R}^d} f(\boldsymbol{x})\mu(\mathrm{d}\boldsymbol{x}) = \int_{\mathbb{R}^d} \check{f}(\boldsymbol{\xi})\hat{\mu}(\mathrm{d}\boldsymbol{\xi})$$

for every $f \in C_c(\mathbb{R}^d)$. The statement for f bounded with bounded support follows by the similar arguments as in Proposition A.2.5. □

A.3 Geometric measure theory

Here we recapitulate some important results from the Geometric measure theory in $\mathbb{R}^d$. For a thorough treatment we refer the reader to [65, 101, 119]. First let us introduce the Hausdorff measure and Hausdorff dimension. We follow the usual Carathéodory's construction.

The **diameter** of a set $A \subset \mathbb{R}^d$ is defined to be

$$\operatorname{diam} A = \sup\{\|\boldsymbol{x} - \boldsymbol{y}\|\,|\boldsymbol{x}, \boldsymbol{y} \in A\}.$$

The countable (or finite) collection of sets $\{U_i\}$ is a **δ-cover** of A if $\operatorname{diam} U_i \leq \delta$ for all i and $A \subset \bigcup_i U_i$. Suppose that α is a non-negative number and B is a subset of $\mathbb{R}^d$. Then the **α-dimensional Hausdorff measure** $\mathcal{H}^\alpha(B)$ of B is given by

$$\mathcal{H}^\alpha(B) = \lim_{\delta \to 0_+} \left(\inf \left\{ \sum_{i=1}^{\infty} c_\alpha 2^{-\alpha} (\operatorname{diam} U_i)^\alpha \middle| \{U_i\} \text{ is a } \delta\text{-cover of } B \right\} \right),$$

where

$$c_\alpha = \frac{\pi^{\frac{\alpha}{2}}}{\Gamma\left(1+\frac{\alpha}{2}\right)} \text{ for } \alpha > 0, \quad c_0 = 1.$$

The formula for c_α is the generalization of the expression for the α-dimensional volume of a unit ball in $\mathbb{R}^\alpha$ when α is a natural number. It can be shown that $\mathcal{H}^\alpha$ is an outer Radon measure.

The following theorem gives the connection between the d-dimensional Hausdorff measure and d-dimensional Lebesgue measure.

Theorem A.3.1. *Let B be a Borel set in $\mathbb{R}^d$. Then $\nu_d(B) = \mathcal{H}^d(B)$.*

Proof. Theorem 2.10.35 in [65]. □

Next important property of the Hausdorff measure is its invariance with respect to Lipschitz functions.

Theorem A.3.2. *Suppose $\alpha \geq 0$ and n, m are positive integers. If $f : \mathbb{R}^m \to \mathbb{R}^n$ satisfies*

$$\|f(\boldsymbol{x}) - f(\boldsymbol{y})\| \leq L \|\boldsymbol{x} - \boldsymbol{y}\| \quad \text{for all } \boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^m,$$

i.e. if f is a Lipschitz function with Lipschitz constant not exceeding L, then for any $A \subset \mathbb{R}^n$

$$\mathcal{H}^\alpha(f(A)) \leq L^\alpha \mathcal{H}^\alpha(A).$$

Proof. Theorem 3.1.16 in [101]. □

The **Hausdorff dimension** of a set $A \subset \mathbb{R}^d$ is defined to be

$$\dim_H A = \inf\{\alpha \geq 0 | \mathcal{H}^\alpha(A) = 0\} = \sup\{\alpha | \mathcal{H}^\alpha(A) = \infty\}.$$

A simple analysis (see for instance [25]) shows that $\mathcal{H}^\alpha(A) = 0$ for $\alpha > \dim_H A$ and $\mathcal{H}^\alpha(A) = \infty$ for $\alpha < \dim_H A$. If $\alpha = \dim_H A$, then $\mathcal{H}^\alpha(A)$ may be zero or infinite, or may satisfy $0 < \mathcal{H}^\alpha(A) < \infty$. In the last case A is called an **α-set**.

For latter use it is important to introduce the Minkowski content of a set. The ε-neighbourhood of a set A is given by

$$A_\varepsilon = \{\boldsymbol{x} \mid d(\boldsymbol{x}, A) < \varepsilon\},$$

where $d(\boldsymbol{x}, A)$ is the distance from $\boldsymbol{x}$ to A defined in the previous section.

Definition A.3.1. Let assume $A \subset \mathbb{R}^d$ and $0 \leq \alpha \leq d$. The **α-dimensional upper** and **lower Minkowski contents of** A, denoted by $\mathcal{M}^{*\alpha}(A)$ and $\mathcal{M}^\alpha_*(A)$, are defined by

$$\mathcal{M}^{*\alpha}(A) = \limsup_{\varepsilon \to 0_+} \frac{\nu_d(A_\varepsilon)}{c_{d-\alpha}\varepsilon^{d-\alpha}} \quad \text{and} \quad \mathcal{M}^\alpha_*(A) = \liminf_{\varepsilon \to 0_+} \frac{\nu_d(A_\varepsilon)}{c_{d-\alpha}\varepsilon^{d-\alpha}},$$

respectively. If $\mathcal{M}^{*\alpha}(A) = \mathcal{M}^\alpha_*(A)$, their common value is called the **α-dimensional Minkowski content of** A and denoted by $\mathcal{M}^\alpha(A)$.

It is easy to see that $\mathcal{M}^{*\alpha}(A) = \mathcal{M}^{*\alpha}(\bar{A})$, where $\bar{A}$ is the topological closure of A, and similarly for the lower content. Therefore the Minkowski contents do not distinguish between a set and its closure. If B is the Borel set of positive volume $\nu_d(B) > 0$, then obviously $\mathcal{M}^d(B) = \nu_d(B)$.

Analogously to the Hausdorff dimension one can define the **upper** and **lower Minkowski dimensions** by

$$\overline{\dim}_M A = \inf\{\alpha \geq 0 | \mathcal{M}^{*\alpha}(A) = 0\} \quad \text{and} \quad \underline{\dim}_M A = \inf\{\alpha \geq 0 | \mathcal{M}^\alpha_*(A) = 0\},$$

respectively. If both dimensions coincide we call their common value the **Minkowski dimension of** A and denote it by $\dim_M(A)$. We say that A is **Minkowski measurable** if $0 < \mathcal{M}_*^\alpha(A) = \mathcal{M}^{*\alpha}(A) < \infty$, where α is the Minkowski dimension of A. It turns out that upper and lower Minkowski dimensions equal to upper and lower box dimensions that are used quite often in fractal analysis. For a further discussion of fractal dimensions and their relations we refer the reader to [25].

It is worth noting that the upper and lower Minkowski contents are not outer measures, and not Borel measures when restricted to Borel sets, as they are not countably sub-additive. Actually only the upper Minkowski content is finitely sub-additive. The finite additivity can be shown for Minkowski content and Minkowski measurable sets that are separated by a positive distance. A deeper discussion of problems related to Minkowski content can be found in [120].

There is no general relation between the Minkowski content and Hausdorff measure. The best result is the following lower bound.

Proposition A.3.1. *For $0 \leq \alpha \leq d$ and any set $A \subset \mathbb{R}^d$, we have*

$$\mathcal{H}^\alpha(A) \leq 3^\alpha \frac{c_{d-\alpha} c_\alpha}{c_d} \mathcal{M}_*^\alpha(A).$$

Proof. Proposition 3.3.3 in [101]. □

From this follows that for every set A

$$\dim_H A \leq \underline{\dim}_M A \leq \overline{\dim}_M A.$$

Even though one generally cannot link the Hausdorff measure and Minkowski content, for nice sets both are the same. First, let us recall the concept of rectifiability. We follow the construction in [65, 101].

Definition A.3.2. Suppose $B \subset \mathbb{R}^d$, ϕ is an outer measure on $\mathbb{R}^d$, and m is a positive integer:

(a) B is **m-rectifiable** if B is the image of some bounded subset of $\mathbb{R}^m$ under a Lipschitz function.

(b) B is **countably m-rectifiable** if B equals the countable union of m-rectifiable sets.

(c) B is **countably (ϕ, m)-rectifiable** if there is some countably m-rectifiable set containing ϕ almost all of B.

(d) B is **(ϕ, m)-rectifiable** if B is countably (ϕ, m)-rectifiable and $\phi(B) < \infty$.

Familiar examples of rectifiable sets are rectifiable curves and C^1 manifolds in $\mathbb{R}^d$. We are particularly interested in $(\mathcal{H}^m, m)$-rectifiable sets since there is a nice connection to C^1 submanifolds in $\mathbb{R}^d$. First recall the definition of C^1 submanifolds of $\mathbb{R}^d$.

Definition A.3.3. Let m be a non-negative integer. By an m-dimensional C^1 submanifold of $\mathbb{R}^d$ we mean a subset M of $\mathbb{R}^d$ that satisfies: For each $\boldsymbol{x} \in M$ there exists a neighbourhood T of $\boldsymbol{x}$ in $\mathbb{R}^d$, a C^1 diffeomorphism $\varphi : T \to \mathbb{R}^d$, and an m-dimensional vectorsubspace Z of $\mathbb{R}^d$ such that

$$\varphi(M \cap T) = Z \cap \varphi(T).$$

Clearly, a subset of $\mathbb{R}^d$ is a 0-dimensional submanifold if and only if all of its points are isolated. The relation between submanifolds and rectifiability is given by the following theorem.

Theorem A.3.3. *Let m be a positive integer. A subset of $\mathbb{R}^d$ is countably $(\mathcal{H}^m, m)$-rectifiable if and only if, except for a set of $\mathcal{H}^m$ measure zero, it is contained in a countable union of m-dimensional C^1 submanifolds of $\mathbb{R}^d$.*

Proof. Theorem 3.2.29 in [65] or Lemma 5.4.2 in [67]. □

The 1-rectifiability is often much easier to obtain than the higher dimensional rectifiability because of the following result.

Theorem A.3.4. *Every compact connected set $C \subset \mathbb{R}^n$ with $\mathcal{H}^1(C) < \infty$ is a Lipschitz image of a subinterval of $\mathbb{R}$.*

Proof. Theorem 1.1.8 in [121]. □

Besides the global connection to C^1 submanifolds one can formulate rectifiability in terms of the local structure.

Theorem A.3.5 (Besicovitch-Marstrand-Matilla)**.** *Let $M \subset \mathbb{R}^d$ be an $\mathcal{H}^m$ measurable set with $\mathcal{H}^m(M) < \infty$. Then, M is $(\mathcal{H}^m, m)$-rectifiable if and only if*

$$\lim_{\varepsilon \to 0_+} \frac{\mathcal{H}^m\left(M \cap B_r(\boldsymbol{x})\right)}{(2\varepsilon)^m} = 1$$

for $\mathcal{H}^m$-almost all $\boldsymbol{x} \in M$.

Proof. See [122] or Theorem 17.6 in [119]. □

A.4 Fell topology

In the following two sections some topological and measurability aspects connected with the theory of random closed sets from Section 2.2 are recapitulated. For a fuller treatment we refer the reader to [39, 40, 42].

Let E be a locally compact second countable topological space. From now on the Hausdorff separation property of topological spaces is always assumed. By $\mathcal{F} = \mathcal{F}(E), \mathcal{C} = \mathcal{C}(E)$, and $\mathcal{G} = \mathcal{G}(E)$ we mean the system of closed, compact, and open subsets of E, respectively. Empty set is always included: $\emptyset \in \mathcal{F}, \mathcal{C}, \mathcal{G}$. Corresponding systems of non-empty sets are denoted by $\mathcal{F}', \mathcal{C}'$, and $\mathcal{G}'$.

The most important topological properties of E are summarized in the following theorem. We use the abbreviation $\bar{A}$ for the topological closure of a set $A \subset E$.

Theorem A.4.1. *Let E be a locally compact topological space with a countable base. Then:*

(a) For every point $x \in E$ and every open neighbourhood G of x there exists an open, relatively compact neighbourhood D of x such that $\bar{D} \subset G$.

(b) The topology of E has a countable base $\mathcal{D}$ consisting of open, relatively compact sets such that every open set $G \subset E$ is the union of sets $D \in \mathcal{D}$ satisfying $\bar{D} \subset G$.

(c) There is a sequence $\{G_i\}_{i\in\mathbb{N}}$ of open, relatively compact sets in E satisfying $\bar{G}_i \subset G_{i+1}$ for all i and $\bigcup_i G_i = E$.

(d) For every compact set $C \subset E$ there exists a decreasing sequence $\{G_i\}_{i\in\mathbb{N}}$ of open, relatively compact neighbourhoods of C such that to every open set $G \subset E$ with $C \subset G$ there is an i with $G_i \subset G$.

Further, there is a decreasing sequence $\{H_i\}_{i\in\mathbb{N}}$ of open, relatively compact sets with $\bar{H}_{i+1} \subset H_i$ and $\bigcap_i H_i = C$.

(e) If $C \subset E$ is compact and $G_1, G_2 \subset E$ are open sets with $C \subset G_1 \cup G_2$, then there are compact sets $C_1 \subset G_1$ and $C_2 \subset G_2$ with $C = C_1 \cup C_2$.

Proof. Theorem 12.1.1 in [40]. □

The theory of random closed sets is based on a special topology on the system $\mathcal{F} = \mathcal{F}(E)$ of closed subsets of E. It is useful to present here the most important properties of this topology. For each $A \subset E$ let us define

$$\mathcal{F}_A = \{F \in \mathcal{F} | F \cap A \neq \emptyset\},$$

as the system of closed sets with non-empty intersection with A (sets that hit A) and

$$\mathcal{F}^A = \{F \in \mathcal{F} | F \cap A = \emptyset\},$$

as the system of closed sets with empty intersection (sets that miss A). It is easy to see that

$$\mathcal{F}^A \cap \mathcal{F}^B = \mathcal{F}^{A \cup B}, \quad \mathcal{F}_A \cup \mathcal{F}_B = \mathcal{F}_{A \cup B}, \quad \text{and} \quad (\mathcal{F}_A)^c = \mathcal{F}^A.$$

It is further convenient to define the system

$$\mathbb{B} = \{\mathcal{F}^C_{G_1, G_2, \ldots, G_n} | C \in \mathcal{C}, G_i \in \mathcal{G}, n \in \mathbb{N}_0\},$$

where

$$\mathcal{F}^C_{G_1, G_2, \ldots, G_n} = \mathcal{F}^C \cap \mathcal{F}_{G_1} \cap \ldots \cap \mathcal{F}_{G_n}, \ n > 0$$

and

$$\mathcal{F}^C_{G_1, G_2, \ldots, G_n} = \mathcal{F}^C, \ n = 0.$$

$\mathcal{F}^C_{G_1, G_2, \ldots, G_n}$ is the system of closed sets that miss compact C and hits all the open G_i. In particular $\mathcal{F}^\emptyset_G = \mathcal{F}_G$ and

$$\mathcal{F}^C_{G_1, \ldots, G_n} \cap \mathcal{F}^{C'}_{G'_1, \ldots, G'_m} = \mathcal{F}^{C \cup C'}_{G_1, \ldots, G_n, G'_1, \ldots, G'_m}.$$

The family $\mathbb{B}$ of such closed sets contains $\mathcal{F} = \mathcal{F}^\emptyset$ and is closed under finite intersections. Therefore it constitutes a base of the so called **Fell** or **hit-or-miss** topology on $\mathcal{F}$. Recall that a base of the topology means a collection of open sets such that every open set from topology can be written as a union of elements from the base. Throughout this thesis we always assume that $\mathcal{F}$ is equipped with this topology.

Theorem A.4.2. *$\mathcal{F}$ is a Hausdorff separable compact space with a countable base. $\mathcal{F}'$ is a Hausdorff separable locally compact space with a countable base.*

Proof. Theorem 12.2.1 in [40] or Theorem 1-2-1 in [39] for the first part and a trivial consequence in Remark (b) after Theorem 12.2.1. in [40] for a second part. □

It is useful to characterize the convergence in the space $\mathcal{F}$.

Theorem A.4.3. *Let $\{F_i\}_{i \in \mathbb{N}}$ be a sequence in $\mathcal{F}$ and let $F \in \mathcal{F}$. Then $F_i \to F$ as $i \to +\infty$ if and only if it satisfies the following two conditions:*

(a) If $x \in F$, then for almost all i (except for a finite number) there is $x_i \in F_i$ so that $x_i \to x$ as $i \to +\infty$ in E,

(b) if $\{F_{j_k}\}_{k \in \mathbb{N}}$ is a subsequence and the points $x_{j_k} \in F_{j_k}$ are such that $x = \lim_{k \to +\infty} x_{j_k}$ exists, then $x \in F$.

Proof. See Theorem 12.2.2 in [40] or Theorem 1-2-2 in [39]. □

The important corollary of this theorem is that every closed set $F \in \mathcal{F}$ can be obtained as a limit of finite subsets in E.

Corollary A.4.1. *Let $\mathcal{T}$ be the class of the finite subsets in E and $\mathcal{T}' = \mathcal{T} \setminus \{\emptyset\}$ the class of the non-empty finite subsets in E. Then $\mathcal{T}$ is dense in $\mathcal{F}$, $\mathcal{T}'$ is dense in $\mathcal{F}' = \mathcal{F} \setminus \{\emptyset\}$, and, if E is not compact, $\mathcal{T}'$ is dense in $\mathcal{F}$.*

Proof. See Corollary 2 after Theorem 1-2-2 in [39]. □

In the next corollary we show one special type of convergence in $\mathcal{F}$ that is used later.

Corollary A.4.2. *Let $\{F_n\}$ be a sequence in $\mathcal{F}$ and $F_n \searrow F$ (i.e. F_n is decreasing in the sense of inclusion and $F = \bigcap_n F_n$). Then $F_n \to F$ in $\mathcal{F}$.*

Proof. The set F is given by $F = \bigcap_n F_n$ and therefore it is closed, $F \in \mathcal{F}$. To show the convergence, it is sufficient to check both conditions of Theorem A.4.3. In (a) one can for each $x \in F$ take the constant sequence $x_i = x$ since $x \in F_i$ for all i.

Let now assume that $\{F_{j_k}\}_{k\in\mathbb{N}}$ is a subsequence and that points $x_{j_k} \in F_{j_k}$ are chosen such that $x = \lim_{k\to+\infty} x_{j_k}$ exists. For each n there exists k_0 such that $j_k > n$ for all $k > k_0$. Therefore $x_{j_k} \in F_n$ for all $k > k_0$. Since F_n is closed, we have $x \in F_n$. Hence $x \in \bigcap_n F_n = F$, which proves (b). □

When working with mappings related to random closed sets it is necessary to prove their measurability. For many of them the measurability follows from their continuity or semi-continuity.

A map $\varphi : T \to \mathcal{F}$ from a topological space T into $\mathcal{F}$ is called **upper semi-continuous** if for any $C \in \mathcal{C}$, the set $\varphi^{-1}(\mathcal{F}^C)$ is open in T, and φ is **lower semi-continuous** if for any $G \in \mathcal{G}$, the set $\varphi^{-1}(\mathcal{F}_G)$ is open in T. In the following we summarize the continuity and semi-continuity properties of the most important mappings.

Lemma A.4.1. *(a) The union mapping $(F, F') \mapsto F \cup F'$ from the product space $\mathcal{F} \times \mathcal{F}$ onto $\mathcal{F}$ is continuous.*

(b) The intersection mapping $(F, F') \mapsto F \cap F'$ from $\mathcal{F} \times \mathcal{F}$ onto $\mathcal{F}$ is upper semi-continuous (and not continuous).

(c) The mapping $F \mapsto \overline{F^c}$ from $\mathcal{F}$ to itself is upper semi-continuous.

(d) If E is locally connected, the boundary mapping $F \mapsto \partial F$ from $\mathcal{F}$ to itself is lower semi-continuous.

(e) If the topological group G acts continuously on E, then the map $(g, F) \mapsto gF$ from $G \times \mathcal{F}$ to $\mathcal{F}$ is continuous.

Proof. (a) Theorem 12.2.3 in [40] or Corollary 1 after Theorem 1-2-2 in [39].
(b) - (d) Theorem 12.2.6 in [40] or Corollaries 1-3 after Proposition 1-2-4 in [39].
(e) Theorem 13.1.1 in [40]. □

For the next assertion let us recall the upper (lower) semi-continuity of a real function. A function $f : E \to [-\infty, +\infty]$ from a topological space E into the extended real line $[-\infty, +\infty]$ is said to be **lower semi-continuous** (resp. **upper semi-continuous**), if for each finite real number h, $f^{-1}((h, +\infty])$ (resp. $f^{-1}([-\infty, h))$) is an open set in E.

It is a standard result (see e.g. [113]) that every lower (upper) semi-continuous function is Borel measurable, i.e. $f^{-1}(V)$ is a Borel set (with respect to the topology of E) for every open V.

Proposition A.4.1. *A real-valued function $f : E \to [-\infty, +\infty]$, defined on a topological space E, is lower semi-continuous (resp. upper semi-continuous) if and only if $\liminf_{x\to a} f(x) \geq f(a)$ (resp. $\limsup_{x\to a} f(x) \leq f(a)$) for all $a \in E$.*

Proof. Proposition 3 in [123]. □

A useful characterization of the upper semi-continuity of a real function on $\mathcal{F}(E)$ is given by the following lemma.

Lemma A.4.2. *Let φ be an increasing mapping (i.e. $F_1 \subset F_2$ in $\mathcal{F}$ implies $\varphi(F_1) \leq \varphi(F_2)$) from $\mathcal{F}$ into $[-\infty, +\infty]$. Then φ is upper semi-continuous if and only if $F_n \searrow F$ (decreasing in the sense of inclusion and $F = \bigcap_n F_n$) in $\mathcal{F}$ implies $\varphi(F_n) \searrow \varphi(F)$ (decreasing and $\varphi(F_n) \to \varphi(F)$).*

Proof. See Corollary 5 after Proposition 1-2-4 in [39]. □

For a finite measure μ defined on the Borel σ-algebra $\mathcal{B}(\mathcal{F})$ the previous lemma can be used to show upper semi-continuity of the function $F \mapsto \mu(F)$.

Theorem A.4.4. *Any finite measure μ on E is upper semi-continuous on $\mathcal{F}$.*

Proof. Let $\{F_i\}_{i\in\mathbb{N}}$ be a sequence in $\mathcal{F}$ that is decreasing in the sense of inclusion,

$$F_i \supset F_{i+1} \quad \text{for all } i \in \mathbb{N},$$

such that $F = \bigcap_i F_i \in \mathcal{F}$. From the σ-additivity of μ follows that μ is increasing, i.e. $F_1 \supset F_2$ implies $\mu(F_1) \geq \mu(F_2)$. Since $\mu(F_1) < +\infty$, Theorem A.1.1 (a) implies

$$\mu(F_i) \searrow \mu(F).$$

Finally, Lemma A.4.2 yields the result. □

This result does not hold for general measures (see [40] for more details and references), e.g. the Lebesgue measure on $E = \mathbb{R}^d$ is neither upper nor lower semi-continuous and therefore not continuous. The non-continuity of the Lebesgue measure also directly follows from Corollary A.4.1.

An important example of an upper semi-continuous function is the set indicator.

Theorem A.4.5. *The indicator function map $\mathbb{1} : \mathcal{F} \times E \to \mathbb{R}$, given by $(F, x) \mapsto \mathbb{1}_F(x)$, is upper semi-continuous.*

Proof. Theorem 12.2.7 in [40]. □

The previous theorem particularly says that for each fixed $x \in E$, the indicator function $\mathbb{1}_F(x)$ is an upper semi-continuous function from $\mathcal{F}(E)$ to $\mathbb{R}$. In the special case $E = \mathbb{R}^d$ there are other natural mappings that are also continuous or semi-continuous.

Theorem A.4.6. *Let $E = \mathbb{R}^d$. Then the mappings*

(a) $F \mapsto -F$ from $\mathcal{F}$ to itself,

(b) $(\alpha, F) \mapsto \alpha F$ from $\mathbb{R}^+ \times \mathcal{F}$ onto $\mathcal{F}$

are continuous and the mapping

(c) $(F, F') \mapsto \overline{F + F'}$ from $\mathcal{F} \times \mathcal{F}$ to $\mathcal{F}$

is lower semi-continuous.

Proof. See Theorem 12.3.1 in [40] or Proposition 1-5-1 in [39]. □

Let us now discuss the connection between measurability and semi-continuity. First, it is appropriate to mention the following lemma.

Lemma A.4.3. *The σ-algebra $\mathcal{B}(\mathcal{F})$ of Borel sets of $\mathcal{F}$ is generated by either of the systems*

$$\{\mathcal{F}^C | C \in \mathcal{C}\} \quad \text{and} \quad \{\mathcal{F}_G | G \in \mathcal{G}\}.$$

Proof. Lemma 2.1.1 in [40]. □

From this lemma and Theorem A.1.3 follows that if $\varphi : T \to \mathcal{F}$ is a lower (upper) semi-continuous map from a topological space T into $\mathcal{F}$, then φ is Borel measurable. If $f : T \to [-\infty, +\infty]$ is a lower (upper) semi-continuous real function from a topological space T into the extended real line $[-\infty, +\infty]$, then it is also Borel measurable. This follows from the fact that the collection of all intervals of the form $(h, \infty]$ (or $[-\infty, h)$) generates the Borel σ-algebra of $[-\infty, +\infty]$.

The next assertion is that $\mathcal{C}$, a collection of all compact sets in $\mathcal{F}$, is a Borel set in $\mathcal{F}$.

Lemma A.4.4. *$\mathcal{C}$ is a Borel set in $\mathcal{F}$.*

Proof. Lemma 2.1.2 in [40]. □

The same holds for a collection $\mathcal{X}_m \subset \mathcal{F}$ of $(\mathcal{H}^m, m)$-rectifiable closed sets in $\mathbb{R}^d$.

Theorem A.4.7. *For each integer m, $0 \leq m \leq d$, $\mathcal{X}_m$ is a Borel set in $\mathcal{F}$.*

Proof. Theorem 2.2.1 in [124]. □

A.5 Hausdorff metric

In this section we assume $E = \mathbb{R}^d$, the d-dimensional Euclidean space. Let $d(\cdot, \cdot)$ denote the Euclidean metric on $\mathbb{R}^d$ which is for $x, y \in \mathbb{R}^d$ defined using the standard scalar product and the corresponding norm $\|\cdot\|$ as $d(x, y) = \|x - y\|$. We use it to define the Hausdorff metric on the collection $\mathcal{C}' = \mathcal{C} \setminus \{\emptyset\}$ of non-empty compact subsets of $\mathbb{R}^d$.

Definition A.5.1. The **Hausdorff distance** $\delta(C_1, C_2)$ of $C_1, C_2 \in \mathcal{C}'$ is defined by

$$\delta(C_1, C_2) = \max\left\{\max_{x \in C_1} \min_{y \in C_2} d(x, y), \max_{x \in C_1} \min_{y \in C_2} d(x, y)\right\}.$$

This can be equivalently written as

$$\delta(C_1, C_2) = \min\left\{\varepsilon \geq 0 \,\middle|\, C_1 \subset C_2 + \varepsilon B^d, C_2 \subset C_1 + \varepsilon B^d\right\},$$

where B^d is the closed d-dimensional unit ball centred at the origin.

The Hausdorff distance constitutes a metric, called the **Hausdorff metric**, on $\mathcal{C}'$ and can be extended on $\mathcal{C}$ by putting $\delta(C_1, C_2) = +\infty$ if exactly one of the sets C_1, C_2 is empty set and $\delta(\emptyset, \emptyset) = 0$. Empty set $\emptyset$ is therefore an isolated point. In the following we mean by $(\mathcal{C}, \delta)$ the topological space with topology induced by the Hausdorff metric δ on $\mathcal{C}$.

Theorem A.5.1. *The space $\mathcal{C}'$ of non-empty compact sets is complete and locally compact in the Hausdorff metric.*

Proof. Theorem 1.8.2 and Theorem 1.8.4 in [125]. □

Since we deal with the theory of random closed sets, it is important to present the connection between the topology on $\mathcal{C}$ induced by the Hausdorff metric and the trace topology on $\mathcal{C}$ induced by the Fell topology of $\mathcal{F}$.

Theorem A.5.2. *The topology of the Hausdorff metric on $\mathcal{C}$ is strictly finer than the trace topology induced by $\mathcal{F}$.*

Proof. Theorem 12.3.2 in [40] or Proposition 1-4-1 and 1-4-4 in [39]. □

Therefore every sequence in $\mathcal{C}$ that is converging in the Hausdorff metric is also converging in the trace topology of $\mathcal{F}$. The following theorem gives the strict equivalence.

Theorem A.5.3. *The convergence of a sequence $\{C_i\}_{i\in\mathbb{N}}$ in $(\mathcal{C},\delta)$ is equivalent to the following conditions taken together:*

(a) $\{C_i\}_{i\in\mathbb{N}}$ converges in $\mathcal{F}$,

(b) $\{C_i\}_{i\in\mathbb{N}}$ is uniformly bounded, that is, there exists a set $K\in\mathcal{C}$ with $C_i\subset K$ for all i.

Proof. Theorem 12.3.3 in [40] or Theorem 1-4-1 in [39]. □

The same assertion as in Corollary A.4.1 holds in $(\mathcal{C},\delta)$.

Corollary A.5.1. *The set $\mathcal{T}'=\mathcal{T}\setminus\{\emptyset\}$ (the class of the non-empty finite subsets of $\mathbb{R}^d$) is dense in $\mathcal{C}'=\mathcal{C}\setminus\{\emptyset\}$.*

Proof. Corollary 4 after Theorem 1-4-1 in [39]. □

It follows from the definition of lower (upper) semi-continuity of a real function that every real function on $\mathcal{C}$ lower (upper) semi-continuous with respect to the trace topology of $\mathcal{F}$ is also lower (upper) semi-continuous on $(\mathcal{C},\delta)$. One can prove the useful analogue of Lemma A.4.2.

Lemma A.5.1. *Let φ be an increasing real-valued function from $\mathcal{C}$ into $[-\infty,+\infty]$. Then φ is upper semi-continuous if and only if $C_n\searrow C$ (decreasing in the sense of inclusion and $C=\bigcap_n C_n$) implies $\varphi(C_n)\searrow\varphi(C)$.*

Proof. Every sequence $C_n\searrow C$ of compact sets converges in the trace topology induced by $\mathcal{F}$ as a consequence of Corollary A.4.2 and as a consequence of Theorem A.5.3, also in topology induced by the Hausdorff metric since $\{C_n\}$ is uniformly bounded. The assertion is then an immediate consequence of Lemma A.4.2. □

Now we can obtain the extended version of Theorem A.4.4.

Theorem A.5.4. *Any Borel σ-finite measure μ on $\mathbb{R}^d$ is upper semi-continuous on $(\mathcal{C},\delta)$.*

Proof. The proof using a decreasing sequence $\{C_i\}_{i\in\mathbb{N}}$ is analogous as the proof of Theorem A.4.4. The crucial condition $\mu(C_1)<+\infty$ follows from σ-finiteness of μ. □

The continuity properties of most important mappings with respect to the Hausdorff metric are summarized in the following theorem.

Theorem A.5.5. *If $\mathcal{C}$ is equipped with the topology of the Hausdorff metric, then the following maps are continuous:*

(a) $\begin{array}{ccc}\mathcal{C}\times\mathcal{F} & \to & \mathcal{F}\\ (C,F) & \mapsto & C\cup F\end{array}$ *and* $\begin{array}{ccc}\mathcal{C}\times\mathcal{C} & \to & \mathcal{C}\\ (C,C') & \mapsto & C\cup C',\end{array}$

(b) $\begin{array}{ccc}\mathcal{C}\times\mathcal{F} & \to & \mathcal{F}\\ (C,F) & \mapsto & C+F\end{array}$ *and* $\begin{array}{ccc}\mathcal{C}\times\mathcal{C} & \to & \mathcal{C}\\ (C,C') & \mapsto & C+C',\end{array}$

(c) the symmetrical reflection mapping $C\mapsto -C$ from $\mathcal{C}$ to itself,

(d) the multiplication mapping $(\alpha,C)\mapsto\alpha C$ from $\mathbb{R}^+\times\mathcal{C}$ to $\mathcal{C}$.

Proof. Theorem 12.3.5 in [40]. □

From Theorem A.5.2 we know that the topology on $\mathcal{C}$ induced by the Hausdorff measure is strictly finer than the trace topology on $\mathcal{C}$ induced by the topology on $\mathcal{F}$. The question is whether the difference holds also for the induced Borel σ-algebras.

Theorem A.5.6. *Let $\mathcal{B}_{\mathcal{F}}(\mathcal{C})$ be the Borel σ-algebra on $\mathcal{C}$, when $\mathcal{C}$ is equipped with trace topology induced by the topology of $\mathcal{F}$, and let $\mathcal{B}_{\delta}(\mathcal{C})$ be the Borel σ-algebra on $\mathcal{C}$, when $\mathcal{C}$ is equipped with topology induced by the Hausdorff metric. Then $\mathcal{B}_{\mathcal{F}}(\mathcal{C}) = \mathcal{B}_{\delta}(\mathcal{C})$.*

Proof. Theorem 2.4.1 in [40]. □

From now on we write $\mathcal{B}(\mathcal{C})$ for the σ-algebra from the previous theorem. The previous theorem holds in the same way for $\mathcal{C}'$.

Corollary A.5.2. *$\mathcal{C}'$ is a Borel set in both $\mathcal{F}', \mathcal{F}$, and $\mathcal{B}_{\mathcal{F}'}(\mathcal{C}') = \mathcal{B}_{\delta}(\mathcal{C}')$.*

Proof. Follows from Lemma A.4.4, the previous theorem, and the fact that $\emptyset$ is an isolated point in the topology of the Hausdorff metric. □

We may use the same argumentation as after Lemma A.4.3 to see that every continuous or upper (lower) semi-continuous real function is Borel measurable. Now lets look at the Hausdorff measure $\mathcal{H}^{\alpha}$. Here, Theorem A.5.4 cannot be used since for $\alpha < d$ the Hausdorff measure $\mathcal{H}^{\alpha}$ is not σ-finite. One can hovewer still show measurability.

Theorem A.5.7. *The functions $\mathcal{H}^{\alpha} : \mathcal{C} \to [0, \infty]$ and $\dim_H : \mathcal{C}' \to [0, d]$ are Borel measurable. Moreover, for any closed $B \in \mathcal{F}$, $\mathcal{H}^{\alpha}(C \cap B)$ as a function of C is Borel measurable.*

Proof. The first part follows from Theorem 2.1 in [126], which shows that they are Baire's class 2. From 2.2.15 in [65] we know that every Baire function is Borel measurable. The second part is Theorem 2.1.3 in [124]. □

The following proposition shows the measurability of the measures that are absolutely continuous with respect to $\mathcal{H}^{\alpha}$. Note that from the previous theorem follows that the class $\mathcal{C}'_{\mathcal{H}^{\alpha}}$ of non-empty compact α-sets is a Borel set in $\mathcal{C}'$ since $\mathcal{C}'_{\mathcal{H}^{\alpha}} = (\mathcal{H}^{\alpha})^{-1}\big((0, \infty)\big)$.

Proposition A.5.1. *Let $0 \leq \alpha \leq d$ be fixed, f be a non-negative Borel measurable function on $\mathbb{R}^d$, and μ be a Borel measure on $\mathbb{R}^d$ defined by $\mu(B) = \int_B f(\boldsymbol{x})\, \mathcal{H}^{\alpha}(\mathrm{d}\boldsymbol{x})$ for all Borel $B \subset \mathbb{R}^d$. Let further $\mathcal{C}'_{\mathcal{H}^{\alpha}}$ be equipped with trace topology induced by the Hausdorff metric on $\mathcal{C}'$. Then $\mu : \mathcal{C}'_{\mathcal{H}^{\alpha}} \to [0, \infty]$ is Borel measurable.*

Proof. Let us denote $\mathcal{F}_{\mathcal{H}^{\alpha}}$ the class of all closed subsets F of $\mathbb{R}^d$ for which $\mathcal{H}^{\alpha}(F \cap \cdot)$ is a locally finite measure on $\mathbb{R}^d$. Clearly $\mathcal{C}'_{\mathcal{H}^{\alpha}} \subset \mathcal{F}_{\mathcal{H}^{\alpha}}$. By Corollary 2.1.4 in [124], $\mathcal{F}_{\mathcal{H}^{\alpha}}$ is a Borel set in $\mathcal{F}$ and for any Borel $B \in \mathbb{R}^d$ is $\mathcal{H}^{\alpha}(\cdot \cap B) : \mathcal{F}_{\mathcal{H}^{\alpha}} \to [0, \infty]$ measurable with respect to the trace σ-algebra of $\mathcal{F}$. The Borel σ-algebra $\mathcal{B}_{\delta}(\mathcal{C}'_{\mathcal{H}^{\alpha}})$ of the trace topology is clearly the trace σ-algebra $\mathcal{B}_{\delta}(\mathcal{C}) \cap \mathcal{C}'_{\mathcal{H}^{\alpha}}$ and by Theorem A.5.6 also the trace σ-algebra $\mathcal{B}_{\mathcal{F}}(\mathcal{C}) \cap \mathcal{C}'_{\mathcal{H}^{\alpha}}$ which further equals to the trace σ-algebra $\mathcal{B}(\mathcal{F}) \cap \mathcal{C}'_{\mathcal{H}^{\alpha}}$. Since $\mathcal{C}'_{\mathcal{H}^{\alpha}}$ is a Borel set in $\mathcal{C}'$ by Theorem A.5.2 it is a Borel set in $\mathcal{F}$. Hence for any Borel $B \in \mathbb{R}^d$ is $\mathcal{H}^{\alpha}(\cdot \cap B) : \mathcal{C}'_{\mathcal{H}^{\alpha}} \to [0, \infty]$ Borel measurable with respect to $\mathcal{B}_{\delta}(\mathcal{C}'_{\mathcal{H}^{\alpha}})$. From the basic construction of the Lebesgue integral (e.g. [113, Theorem 1.17]) we know that there is a pointwise increasing sequence $\{s_n\}_{n \in \mathbb{N}}$ of simple measurable functions on $\mathbb{R}^d$ such that $s_n(\boldsymbol{x}) \to f(\boldsymbol{x})$ as $n \to \infty$ for every $\boldsymbol{x} \in \mathbb{R}^d$. Moreover, for a simple function $s = \sum_{i=1}^{k} a_i \mathbb{1}_{A_i}$ with $a_i \geq 0$ and $A_1, \ldots, A_k$ Borel, we have $\int_C s(\boldsymbol{x})\, \mathcal{H}^{\alpha}(\mathrm{d}\boldsymbol{x}) = \sum_{i=1}^{k} a_i\, \mathcal{H}^{\alpha}(C \cap A_i)$ which is by previous considerations measurable on $\mathcal{C}'_{\mathcal{H}^{\alpha}}$ as a function of C. Since by the Lebesgue monotone convergence theorem $\int_C s_n(\boldsymbol{x})\, \mathcal{H}^{\alpha}(\mathrm{d}\boldsymbol{x}) \to \mu(C)$ as $n \to \infty$ for every $C \in \mathcal{C}'_{\mathcal{H}^{\alpha}}$ the measurability of μ on $\mathcal{C}'_{\mathcal{H}^{\alpha}}$ follows. □

The prominent role in $(\mathcal{C}, \delta)$ is played by the subset of convex bodies $\mathcal{K}$, i.e. compact convex sets.

Theorem A.5.8. *On the set $\mathcal{K}'$ of non-empty compact convex sets, the topology of the Hausdorff metric and the trace topology induced by $\mathcal{F}$ coincide.*

Proof. Theorem 12.3.4 in [40]. □

From the topological point of view it may be shown that convex sets are a closed subset of compact sets.

Theorem A.5.9. *$\mathcal{K}$ is a closed subset of $\mathcal{C}$.*

Proof. Theorem 1.8.5 in [125] or Corollary after Proposition 1-5-4 in [39]. □

The d-dimensional Lebesgue measure was upper semi-continuous on $\mathcal{C}$ as a consequence of Theorem A.5.4. The following theorem tells us that on $\mathcal{K}'$ the Lebesgue measure is actually continuous.

Theorem A.5.10. *The d-dimensional Lebesgue measure ν_d is continuous on $\mathcal{K}'$.*

Proof. Lemma 3-5-2 in [39] or Theorem 1.8.16 in [125]. □

Finally we look at the **convex ring** $\mathcal{R}$ given by the finite unions of sets from $\mathcal{K}$. The non-empty convex ring $\mathcal{R}$ is defined by $\mathcal{R}' = \mathcal{R} \setminus \{\emptyset\}$. Obviously $\mathcal{R} \subset \mathcal{C}$.

Theorem A.5.11. *$\mathcal{R}$ is a Borel set in $\mathcal{F}$ and in $\mathcal{C}$. $\mathcal{R}'$ is a Borel set in $\mathcal{F}'$ and in $\mathcal{C}'$.*

Proof. Theorem 2.4.2 in [40] and Corollary A.4.4 resp. A.5.2. □

Appendix B

Dataset details

The `OpenStreetMap` data in ESRI shapefile format can be downloaded from http://osmdata.thinkgeo.com. The GIS sources for the USA cities are given in Table B.1. The unit conversion factor is in Table B.2.

Table B.1: Building-footprints GIS dataset sources and length units.

Location	URL	Original Unit
Boston	http://www.mass.gov	Meter
Chicago	https://data.cityofchicago.org	Foot_US
Los Angeles	http://egis3.lacounty.gov	Foot_US
Pittsburgh	http://www.alleghenycounty.us	Foot_US
Seattle	https://data.seattle.gov	Foot_US

Table B.2: Units conversion table.

Unit label	meter [m]
Meter	1 m
Foot_US	0.3048006096012192 m

Table B.3: Coordinate reference systems (CRS) used for different countries and cities.

Country	City	Code	Name
Belarus	Minsk	2525	Pulkovo 1942 / 3-degree Gauss-Kruger zone 9
Czech Republic	Prague	2065	S-JTSK (Ferro) / Krovak
France	Paris	2154	RGF93 / Lambert-93
Germany	Berlin	31468	DHDN / 3-degree Gauss-Kruger zone 4
Great Britain	Birmingham	27700	OSGB 1936 / British National Grid
Italy	Milan	3003	Monte Mario / Italy zone 1
Norway	Oslo	27393	NGO 1948 (Oslo) / NGO zone III
Russia	Moscow	2705	Pulkovo 1995 / 3-degree Gauss-Kruger CM 39E
	Saint Petersburg	2702	Pulkovo 1995 / 3-degree Gauss-Kruger CM 30E
USA	Boston	2894	NAD 1983 HARN StatePlane Massachusetts Mainland FIPS 2001
	Chicago	3443	NAD 1983 HARN StatePlane Illinois East FIPS 1201
	Los Angeles	2874	NAD 1983 HARN StatePlane California V FIPS 0405
	Pittsburgh	3365	NAD 1983 HARN StatePlane Pennsylvania South FIPS 3702
	Seattle	2926	NAD 1983 HARN StatePlane Washington North FIPS 4601

Bibliography

[1] F. Schweitzer, Brownian Agents and Active Particles, 1st Edition, Springer, 2003.

[2] G. K. Zipf, Human Behavior and the Principle of Least Effort, Addison-Wesley, 1949.

[3] H. D. Rozenfeld, D. Rybski, X. Gabaix, H. A. Makse, The Area and Population of Cities: New Insights from a Different Perspective on Cities, NBER Working Papers 15409, National Bureau of Economic Research, Inc (Oct. 2009).
URL http://ideas.repec.org/p/nbr/nberwo/15409.html

[4] M. Batty, Cities and Complexity: Understanding Cities with Cellular Automata, Agent-Based Models, and Fractals, The MIT press, Cambridge, 2007.

[5] S. L. Arlinghaus, Fractals Take a Central Place, Geografiska Annaler. Series B 67 (2) (1985) 83–88.

[6] M. Batty, P. Longley, The fractal simulation of urban structure, Environment and Planning A 18 (1986) 1143–1179.

[7] S. L. Arlinghaus, W. C. Arlinghaus, The Fractal Theory of Central Place Geometry: A Diophantine Analysis of Fractal Generators for Arbitrary Loschian Numbers, Geographical Analysis 21 (2) (1989) 103–121.

[8] M. Batty, P. Longley, Fractal Cities: A Geometry of Form and Function, 1st Edition, Academic Press, 1994.

[9] F. Schweitzer, J. Steinbink, Urban cluster growth: Analysis and computer simulations of urban aggregations, in: F. Schweitzer (Ed.), Self-Organization of Complex Structures: From Individual to Collective Dynamics, CRC Press, 1997, pp. 501 – 518.

[10] P. Frankhauser, The fractal approach: A new tool for the spatial analysis of urban agglomerations, Population: An English Selection (1998) 205–240.

[11] G. Shen, Fractal dimension and fractal growth of urbanized areas, International Journal of Geographical Information Science 16 (5) (2002) 419–437. doi:10.1080/13658810210137013.

[12] M.-L. D. Keermaecker, P. Frankhauser, I. Thomas, Using fractal dimensions for characterizing intra-urban diversity: The example of brussels, Geographical Analysis 35 (4).

[13] R. Carvalho, A. Penn, Scaling and universality in the micro-structure of urban space, Physica A 332 (2004) 539–547.

[14] I. Thomas, P. Frankhauser, M.-L. D. Keermaecker, Fractal dimension versus density of built-up surfaces in the periphery of brussels, Papers in Regional Science 86 (2).

[15] A. Lagarias, Fractal analysis of the urbanization at the outskirts of the city: Models, measurement and explanation, Cybergeo : European Journal of Geography 391.

[16] Y. Chen, Y. Zhou, Scaling laws and indications of self-organized criticality in urban systems, Chaos, Solitons & Fractals 35 (2008) 85–98.

[17] I. Thomas, P. Frankhauser, C. Biernacki, The morphology of built-up landscapes in wallonia (belgium): A classification using fractal indices, Landscape and Urban Planning 84 (2008) 99–115.

[18] P. Frankhauser, Fractal Geometry for Measuring and Modeling Urban Patterns, in: S. Albeverio, D. Andrey, P. Giordano, A. Vancheri (Eds.), The Dynamics of Complex Urban Systems: An Interdisciplinary Approach, Physica-Verlag Heidelberg, New York, 2007, pp. 213–243.

[19] Y. Chen, Fractal dimension evolution and spatial replacement dynamics of urban growth, Chaos, Solitons & Fractals 45 (2012) 115–124.

[20] I. Thomas, P. Frankhauser, D. Badariotti, Comparing the fractality of european urban neighbourhoods: do national contexts matter?, J Geogr Syst 14 (2012) 189–208.

[21] Y. Chen, Fractal analytical approach of urban form based on spatial correlation function, Chaos, Solitons & Fractals 49 (2013) 47–60.

[22] W. Kinsner, A Unified Approach To Fractal Dimensions, Journal of Information Technology Research 1 (2008) 62–85. doi:10.4018/jitr.2008100105.

[23] M. Guérois, D. Pumain, Built-up encroachment and the urban field: a comparison of forty European cities, Environment and Planning A 40 (2008) 2186 – 2203.

[24] B. B. Mandelbrot, The Fractal Geometry of Nature, Freeman, 1982.

[25] K. Falconer, Fractal Geometry: Mathematical Foundations and Applications, 2nd Edition, Wiley, 2003.

[26] M. Batty, Y. Xie, Self-organized criticality and urban development, Discrete Dynamics in Nature and Society 3 (1999) 109–124. doi:10.1155/S1026022699000151.

[27] K. Huang, Statistical Mechanics, 2nd Edition, Springer, 1987.

[28] I. Herbut, A modern approach to Critical Phenomena, Cambridge University Press, 2007.

[29] M. Batty, P. Longley, S. Fotheringham, Urban growth and form: scaling, fractal geometry, and diffusion-limited aggregation, Environment and Planning A 21 (1989) 1447–1472.

[30] M. Batty, Generating urban forms from diffusive growth, Environment and Planning A 23 (1991) 511–544. doi:10.1068/a230511.

[31] M. Batty, Cities as Fractals: Simulating Growth and Form, in: A. Crilly, R. Earnshow, H. Jones (Eds.), Fractals and Chaos, Springer New York, 1991, pp. 43–69. doi:10.1007/978-1-4612-3034-2_4.

[32] H. A. Makse, S. Havlin, H. E. Stanley, Modelling urban growth patterns, Nature 377 (1995) 608–612. doi:10.1038/377608a0.

[33] H. A. Makse, J. S. Andrade, M. Batty, S. Havlin, H. E. Stanley, Modeling urban growth patterns with correlated percolation, Physical Review E 58 (6) (1998) 7054–7062. doi:10.1103/PhysRevE.58.7054.

[34] R. White, G. Engelen, Cellular automata and fractal urban form: a cellular modelling approach to the evolution of urban land-use patterns, Environment and Planning A 25 (1993) 1175–1199. doi:10.1068/a251175.

[35] R. White, Cities and Cellular Automata, Discrete Dynamics in Nature and Society 2 (1998) 111–125. doi:10.1155/S1026022698000090.

[36] R. White, G. Engelen, High-resolution integrated modelling of the spatial dynamics of urban and regional systems, Computers, Environment and Urban Systems 24 (2000) 383–400. doi:0.1016/S0198-9715(00)00012-0.

[37] M. Fialkowski, A. Bitner, Universal rules for fragmentation of land by humans, Landscape Ecology 23 (9) (2008) 1013–1022. doi:10.1007/s10980-008-9268-x.

[38] P. Exner, P. Šeba, D. Vašata, The distribution of landed property, Physica A: Statistical Mechanics and its Applications 388 (21) (2009) 4619–4623. doi:10.1016/j.physa.2009.07.019.

[39] G. Matheron, Random Sets and Integral Geometry, John Wiley & Sons, 1975.

[40] R. Schneider, W. Weil, Stochastic and Integral Geometry, Springer, 2008. doi:10.1007/978-3-540-78859-1.

[41] S. N. Chiu, D. Stoyan, W. S. Kendall, J. Mecke, Stochastic Geometry and Its Applications, 3rd Edition, Wiley, 2013.

[42] I. Molchanov, Theory of Random Sets, Springer, 2005.

[43] D. Daley, D. Vere-Jones, An Introduction to the Theory of Point Processes, Volume II, 2nd Edition, Springer, 2008.

[44] L. Heinrich, On existence and mixing properties of germ-grain models, Statistics 23 (3) (1992) 271–286. doi:10.1080/02331889208802375.

[45] X. X. Nguyen, H. Zessin, Ergodic theorems for spatial processes, Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete 48 (2) (1979) 133–158. doi:10.1007/BF01886869.

[46] P. Billingsley, Probability and Measure, 3rd Edition, John Wiley & Sons, 1995.

[47] S. Böhm, L. Heinrich, V. Schmid, Kernel Estimation of the Spectral Density of Stationary Random Closed Sets, Australian & New Zealand Journal of Statistics 46 (2004) 41 – 51. doi:10.1111/j.1467-842X.2004.00310.x.

[48] K. Koch, J. Ohser, K. Schladitz, Spectral theory for random closed sets and estimating the covariance via frequency space, Advances in Applied Probability 35 (3) (2003) 603–613. doi:10.1239/aap/1059486820.

[49] J.-P. Chillès, P. Delfiner, Geostatistics: Modeling Spatial Uncertainty, 2nd Edition, Wiley, 2012.

[50] J. Ohser, K. Schladitz, 3D Images of Materials Structures, Wiley-VCH, 2010. doi:10.1002/9783527628308.

[51] R. J. Adler, The Geometry of Random Fields, John Wiley & Sons, 1981.

[52] J. L. Doob, Stochastic Processes, John Wiley & Sons, 1953.

[53] J. Potthoff, Sample properties of random fields i: Separability and measurability, Communications on Stochastic Analysis 3 (1) (2009) 143–153.

[54] J. T. Kent, Continuity properties for random fields, The Annals of Probability 17 (4) (1989) 1432–1440. doi:10.1214/aop/1176991163.

[55] H. Cramér, M. R. Leadbetter, Stationary and Related Stochastic Processes, John Wiley & Sons, New York, 1967.

[56] D. Daley, D. Vere-Jones, An Introduction to the Theory of Point Processes, Volume I, 2nd Edition, Springer, 2003.

[57] A. A. Tepel'man, Ergodic theorems for general dynamical systems, Tr. Mosk. Mat. Obs. 26 (1972) 95–132.

[58] J. Beran, Y. Feng, S. Ghosh, R. Kulik, Long-Memory Processes: Probabilistic Properties and Statistical Methods, Springer, 2013. doi:10.1007/978-3-642-35512-7.

[59] F. Lavancier, Long memory random fields, in: P. Bertail, P. Soulier, P. Doukhan (Eds.), Dependence in Probability and Statistics, Vol. 187 of Lecture Notes in Statistics, Springer New York, 2006, pp. 195–220. doi:10.1007/0-387-36062-X_9.

[60] R. Dobrushin, P. Major, Non-central limit theorems for non-linear functional of gaussian fields, Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete 50 (1) (1979) 27–52. doi:10.1007/BF00535673.

[61] N. H. Bingham, C. M. Goldie, J. L. Teugels, Regular variation, Cambridge University Press, 1989.

[62] V. S. Vladimirov, Y. N. Drozzinov, B. I. Zavialov, Tauberian Theorems for Generalized Functions, Kluwer Academic Publishing, 1988. doi:10.1007/978-94-009-2831-2.

[63] M. Kabanava, Tempered radon measures, Revista Matemática Complutense 21 (2) (2008) 553–564.

[64] V. M. Shelkovich, Tauberian theorems for distributions in the lizorkin spaces (2008). URL http://arxiv.org/abs/math/0512579v1

[65] H. Federer, Geometric Measure Theory, Springer, 1969.

[66] L. Ambrosio, N. Fusco, D. Pallara, Functions of Bounded Variation and Free Discontinuity Problems, Oxford University Press, 2000.

[67] S. G. Krantz, H. R. Parks, Geometric Integration Theory, Birkhäuser, 2008.

[68] J. Illian, A. Penttinen, H. Stoyan, D. Stoyan, Statistical Analysis and Modelling of Spatial Point Patterns (Statistics in Practice), John Wiley & Sons, 2008.

[69] G. R. Grimmett, D. R. Stirzaker, Probability and Random Processes, 3rd Edition, Oxford University Press, 2001.

[70] D. G. Horvitz, D. J. Thompson, A generalization of sampling without replacement from a finite universe, Journal of the American Statistical Association 47 (260) (1952) 663–685.

[71] A. Baddeley, Spatial sampling and censoring, in: O. E. Barndorff-Nielsen, W. S. Kendall, M. N. M. V. Lieshout (Eds.), Stochastic geometry: likelihood and computation, Chapman & Hall, 1998, Ch. 2, pp. 37–78.

[72] D. Stoyan, H. Stoyan, Fractals, Random Shapes and Point Fields: Methods of Geometrical Statistics, Wiley, 1994.

[73] D. Stoyan, H. Stoyan, Improving ratio estimators of second order point process characteristics, Scandinavian Journal of Statistics 27 (4) (2000) 641–656. doi:10.1111/1467-9469.00213.

[74] P. Diggle, A kernel method for smoothing point process data, Applied Statistics 34 (2) (1985) 138–147.

[75] M. Wand, M. Jones, Kernel Smoothing, Chapman & Hall, London, 1995.

[76] B. W. Silverman, Density Estimation for Statistics and Data Analysis, Chapman & Hall, London, 1986.

[77] J. S. Simonoff, Smoothing Methods in Statistics, Springer, 1996.

[78] P. J. Diggle, Nonparametric methods, in: A. E. Gelfand, P. Diggle, P. Guttorp, M. Fuentes (Eds.), Handbook of Spatial Statistics, CRC Press, 2010, pp. 299 – 316.

[79] A. J. Baddeley, J. Moller, R. Waagepetersen, Non- and semi-parametric estimation of interaction in inhomogeneous point patterns, Statistica Neerlandica 54 (3) (2000) 329–350. doi:10.1111/1467-9574.00144.

[80] A. W. Bowman, A. Azzalini, Applied Smoothing Techniques for Data Analysis, Oxford University Press, 1997.

[81] U. Hahn, E. B. V. Jensen, M.-C. van Lieshout, L. S. Nielsen, Inhomogeneous spatial point processes by location-dependent scaling, Advances in Applied Probability 35 (2) (2003) 319–336. doi:10.1239/aap/1051201648.

[82] A. Baddeley, Modeling strategies, in: A. E. Gelfand, P. Diggle, P. Guttorp, M. Fuentes (Eds.), Handbook of Spatial Statistics, CRC Press, 2010, pp. 339 – 369.

[83] A. M. Yaglom, Correlation theory of stationary and related random functions I, Springer, 1987.

[84] A. M. Yaglom, Correlation theory of stationary and related random functions II, Springer, 1987.

[85] N. A. C. Cressie, Statistics for Spatial Data, Wiley, 1991.

[86] B. D. Ripley, Spatial Statistics, John Wiley & Sons, 1981.

[87] A. J. Baddeley, L. M. Cruz-Orive, The Rao-Blackwell Theorem in Stereology and Some Counterexamples, Advances in Applied Probability 27 (1) (1995) 2–19.

[88] D. B. Percival, A. T. Walden., Spectral Analysis for Physical Applications, Cambridge University Press, 1993.

[89] E. J. Hannan, The uniform convergence of autocovariances, The Annals of Statistics 2 (4) (1974) 803–806. doi:10.1214/aos/1176342767.

[90] J. D. Picka, Variance reducing modifications for estimators of standardized moments of random sets, Advances in Applied Probability 32 (3) (2000) 682–700. doi:10.1239/aap/1013540239.

[91] T. Mattfeldt, D. Stoyan, Improved estimation of the pair correlation function of random sets, Journal of Microscopy 200 (2) (2000) 158–173. doi:10.1046/j.1365-2818.2000.00748.x.

[92] M. Reed, B. Simon, Methods of Mathematical Physics, vol. II: Fourier analysis, self-adjointness, Academic Press, 1975.

[93] E. O. Brigham, The Fast Fourier Transform and Its Applications, Prentice Hall, 1988.

[94] I. Molchanov, S. N. Chiu, Smoothing techniques and estimation methods for nonstationary Boolean models with applications to coverage processes, Biometrika 87 (2) (2000) 265–283. doi:10.1093/biomet/87.2.265.

[95] M. Abramowitz, I. A. Stegun, Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, Dover, New York, 1964.

[96] T. Gneiting, M. Schlather, Stochastic models that separate fractal dimension and the hurst effect, SIAM Review 46 (2) (2004) 269–282. doi:10.1137/S0036144501394387.

[97] T. Gneiting, Power-law correlations, related models for long-range dependence and their simulation, Journal of Applied Probability 37 (4) (2000) 1104–1109. doi:10.1239/jap/1014843088.

[98] T. Gneiting, H. Ševčíková, D. B. Percival, M. Schlather, Y. Jiang, Fast and Exact Simulation of Large Gaussian Lattice Systems in $\mathbb{R}^2$: Exploring the Limits, Journal of Computational and Graphical Statistics 15 (3) (2006) 483–501. doi:10.1198/106186006X128551.

[99] D. P. Kroese, Z. I. Botev, Spatial process generation, in: V. Schmidt (Ed.), Stochastic Geometry, Spatial Statistics and Random Fields, Volume II: Analysis, Modeling and Simulation of Complex Structures, Springer, 2014, to appear.

[100] ESRI Shapefile Technical Description, ESRI White Paper Edition (July 1998).
URL http://www.esri.com/library/whitepapers/pdfs/shapefile.pdf

[101] S. G. Krantz, H. R. Parks, The Geometry of Domains in Space, Birkhäuser, 1999.

[102] G. Matheron, Estimating and Choosing, Springer-Verlag, 1989. doi:10.1007/978-3-642-48817-7.

[103] P. Kivell, Land and the City: Patterns and Processes of Urban Change, Geography and environment series, Routledge, 1993.

[104] A. Gabrielli, M. Joyce, F. Sylos Labini, Glass-like universe: Real-space correlation properties of standard cosmological models, Phys. Rev. D 65 (2002) 083523. doi:10.1103/PhysRevD.65.083523.

[105] D. Vašata, P. Exner, P. Šeba, Built-up structure criticality, Physica A: Statistical Mechanics and its Applications 390 (21–22) (2011) 3922–3931. doi:10.1016/j.physa.2011.06.009.

[106] D. Vašata, P. Exner, P. Šeba, Urban structure analysis, in: P. Exner (Ed.), Mathematical Results in Quantum Physics, 2011, pp. 260–264.

[107] E. M. Stein, G. Weiss, Introduction to Fourier Analysis on Euclidean Spaces, Princeton University Press, 1971.

[108] J. B. McDonald, Some Generalized Functions for the Size Distribution of Income, Econometrica 52 (3) (1984) 647–63.

[109] C. Kleiber, S. Kotz, Statistical Size Distributions in Economics and Actuarial Sciences, John Wiley & Sons, Hoboken, 2003.

[110] N. M. Temme, Special Functions, An Introduction to the Classical Functions of Mathematical Physics, John Wiley & Sons, 1996.

[111] P. Billingsley, Convergence of Probability Measures, 2nd Edition, John Wiley & Sons, 1999.

[112] N. Bourbaki, Elements of Mathematics: Integration I, Springer, 2004.

[113] W. Rudin, Real and Complex Analysis, McGraw-Hill, New York, 1970.

[114] L. Argabright, J. G. de Lamadrid, Fourier Analysis of Unbounded Measures on Locally Compact Abelian Groups, no. 145 in Mem. Amer. Math. Soc., AMS, 1974.

[115] C. Berg, G. Forst, Potential Theory on Locally Compact Abelian Groups, Springer, 1975. doi:10.1007/978-3-642-66128-0.

[116] N. Bourbaki, Elements of Mathematics: Integration II, Springer, 2004.

[117] A. P. Robertson, M. L. Thornett, On translation-bounded measures, Journal of the Australian Mathematical Society (Series A) 37 (1984) 139–142. doi:10.1017/S1446788700021820.

[118] J. Duoandikoetxea, Fourier Analysis, AMS, 2001.

[119] P. Mattila, Geometry of Sets and Measures in Euclidean Spaces, Cambridge University Press, 1995.

[120] F. Mendivil, J. C. Saunders, On minkowski measurability, Fractals 19 (04) (2011) 455–467. doi:10.1142/S0218348X11005506.

[121] G. David, S. Semmes, Analysis of and on Uniformly Rectifiable Sets, American Mathematical Society, 1993.

[122] P. Mattila, Hausdorff m regular and rectifiable sets in n-space, Transactions of the American Mathematical Society 205 (1975) 263–274.

[123] N. Bourbaki, Elements of Mathematics: General Topology, Addison Wesley, 1967.

[124] M. Zähle, Random processes of hausdorff rectifiable closed sets, Mathematische Nachrichten 108 (1) (1982) 49–72. doi:10.1002/mana.19821080105.

[125] R. Schneider, Convex Bodies: The Brunn-Minkowski Theory, Cambridge University Press, Cambridge, U.K., 1993.

[126] P. Mattila, D. R. Mauldin, Measure and dimension functions: measurability and densities, Mathematical Proceedings of the Cambridge Philosophical Society 121 (1997) 81–100. doi:10.1017/S0305004196001089.