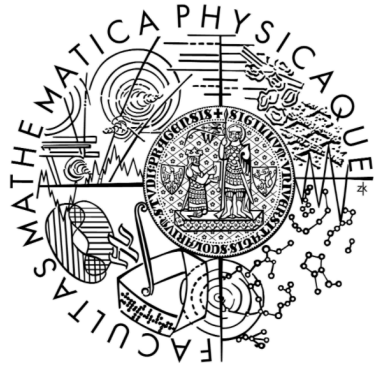
Charles University
Faculty of Mathematics and Physics

HABILITATION THESIS



Modelling and statistics of random sets in 2D and 3D

Jakub Staněk

Prague 2023

Acknowledgments

I would like to thank Kateřina Helisová and Viktor Beneš for their support, valuable advice and useful comments while writing my habilitation thesis as well as for final reading and corrections of the text. Further, I thank all my colleagues who consulted with me the results of my work, directed my research in the right way and created pleasant working conditions for me, which positively influenced my research activities.

Commented papers

This habilitation thesis is composed of the following collection of the author's papers published since 2016. The entries are in the order in which they are commented in the thesis, and their numbering corresponds to **Bibliography** section of this thesis.

[22] Helisová K., Staněk J. (2016): Quermass-interaction process with convex compact grains. *Applications of Mathematics* **61**, 463 – 487. DOI: 10.1007/s10492-016-0142-x

[10] Debayle J., Gotovac Đogaš V., Helisová K., Staněk J., Zikmundová M. (2021): Assessing similarity of random sets via skeletons. *Methodology and Computing in Applied Probability* **23**, 471 – 490. DOI: 10.1007/s11009-020-09785-y

[19] Gotovac Đogaš V, Helisová K., Radović B., Staněk J., Zikmundová M., Brejchová K. (2021): Two-step method for assessing similarity of random sets. *Image Analysis and Stereology* **40**, 127 – 140. DOI: 10.5566/ias.2600

[47] Neumann M., Hirsch Ch., Staněk J., Beneš V., Schmidt V. (2019): Estimation of geodesic tortuosity and constrictivity in stationary random closed sets. *Scandinavian Journal of Statistics* **46**, 848 – 884. DOI: 10.1111/sjos.12375

[46] Neumann M., Staněk J., Pecho M. O., Holzer L., Beneš V., Schmidt V. (2016): Stochastic 3D modeling of complex three-phase microstructures in SOFC-electrodes with completely connected phases. *Computational materials science* **118**, 353 – 364. DOI: 10.1016/j.commatsci.2016.03.013

[48] Petrich L., Staněk J., Wang M., Westhoff D., Heller L., Šittner P., Krill C. E., Beneš V., Schmidt V. (2019): Reconstruction of the grains in polycrystalline materials from incomplete data using Laguerre tessellations. *Microscopy and Microanalysis* **25**, 734 – 752. DOI: 10.1017/S1431927619000485

[33] Kopeček J., Staněk J., Habr S., Seitl F., Petrich L., Schmidt V., Beneš V. (2020): Analysis of the polycrystalline microstructure of AlMgSc alloy observed by 3D EBSD. *Image Analysis and Stereology* **39**, 1 – 11. DOI: 10.5566/ias.2224

[32] Král P., Staněk J., Kunčická L., Seitl F., Petrich L., Schmidt V, Beneš V., Sklenička V. (2019): Microstructure changes in HPT-processed copper occuring at room temperature. *Materials Characterization* **151**, 602 – 611. DOI: 10.1016/j.matchar.2019.03.046

[59] Staněk J., Kopeček J., Král P., Karafiátová I., Seitl F., Beneš V. (2020): Comparison of segmentation of 2D and 3D EBSD measurements in polycrystalline materials. *Kovové materiály - Metallic Materials* **58**, 301 – 319. DOI: 10.4149/km.2020.5.301

Contents

Notation

$\mathbb{N}$	natural numbers
$\mathbb{N}_0 = \mathbb{N} \cup \{0\}$	natural numbers including 0
$\mathbb{R}$	real numbers
$\mathbb{C}$	complex numbers
$x, y, \ldots \in \mathbb{R}^d$, $d \in \mathbb{N}$	points in $\mathbb{R}^d$
$o = (0, \ldots, 0) \in \mathbb{R}^d$	the origin in $\mathbb{R}^d$
$A, B, \ldots \subset \mathbb{R}^d$	subsets of $\mathbb{R}^d$
$A \cup B$	union of A and B
$A \cap B$	intersection of A and B
$\emptyset$	empty set
$A + x = \{a + x : a \in A\}$, $x \in \mathbb{R}^d$ fixed	the translation of the set A by $x \in \mathbb{R}^d$
$A \oplus B = \{a + b : a \in A, b \in B\}$	Minkowski addition of the sets A and B
$\check{A} = \{-a : a \in A\}$	symmetric set to A with respect to the origin
$\|.\|$	Euclidean norm
v_d	$d-$dimensional Lebesgue measure
$\mathcal{H}_d$	$d-$dimensional Hausdorff measure
$\mathrm{int}A$	interior of A
∂A	boundary of A
$\Phi(A) = \frac{\sqrt[3]{\pi}(6v_3(A))^{2/3}}{\mathcal{H}_2(\partial A)}$	sphericity of $A \subset \mathbb{R}^3$ with the volume $v_3(A) < \infty$ and the surface area $\mathcal{H}_2(\partial A) < \infty$
$b(x, r) = \{y \in \mathbb{R}^d : \|x - y\| \leq r, r > 0\}$	closed ball centred in $x \in \mathbb{R}^d$ with the radius r
$B(x, r) = \{y \in \mathbb{R}^d : \|x - y\| < r, r > 0\}$	open ball centred in $x \in \mathbb{R}^d$ with the radius r
$\mathcal{B}$	family of Borel sets
$\mathcal{K}$	family of compact subsets of $\mathbb{R}^d$
$\mathcal{F}$	family of closed subsets of $\mathbb{R}^d$
$\sigma_{\mathcal{F}} = \sigma(\{F \in \mathcal{F} : K \cap F \neq \emptyset\} : K \in \mathcal{K})$	σ-algebra generated by special subsets of $\mathcal{F}$
$\mathcal{N}$	system of locally finite subsets of $\mathbb{R}^d$
$\mathfrak{N} = \sigma(\{N \in \mathcal{N}\})$	σ-algebra generated by all subsets of $\mathcal{N}$
Ω	set of elementary events
$\mathcal{A} = \sigma(\{\omega \in \Omega\})$	σ-algebra generated by all subsets of Ω
P	probability
$\mathbb{E}$	expected value
$X, Y, \ldots$	point processes
$\mathbf{X}, \mathbf{Y}, \ldots$	random sets
$\mathbf{x}, \mathbf{y}, \ldots$	realisations of point processes / random sets
$X^*, Y^*, \ldots$	marked point processes
$\mathcal{T}$	tessellation
$C_i \in \mathcal{T}, i \in I$ (an index set)	cells of the tessellation $\mathcal{T}$

$D_A(M)$	dilatation of the set M by the set A
$E_A(M)$	erosion of the set M by the set A
$O_A(M)$	opening of the set M by the set A
$C_A(M)$	closing of the set M by the set A
$SK(M)$	morphological skeleton of the set M
$G(V, E)$	graph with vertices V and edges E
$\Gamma(z) = \int_0^\infty x^{z-1}e^{-x}\mathrm{d}x$	the Gamma function for $z \in \mathbb{C}, \mathrm{Real}(z) > 0$

Chapter 1

Introduction

Stochastic geometry and spatial statistics have been on the rise in the last few decades. Mostly investigated fields are point processes and random sets, see e.g. [7], [9], [11], [38] or [42]. The reasons for their attractiveness are, firstly, the development of computer technology, which allows faster simulations and hence derivation of statistical inferences based very often on these simulations, and secondly, a wide range of their applications, e.g. in biology, medicine or materials sciences.

The most common applications concern models in two-dimensional space (e.g. occurrences of illnesses on an area [3], shapes of plant undergrowths [41], cross-sections of cell tissue [43], etc.) and three-dimensional space (e.g. material structures or particles in them [55]). Therefore, also statistical methods of their analyses focus on these dimensions. The methods may be very various from model constructions through descriptive statistics to parameter estimates and testing several hypotheses about constructed models.

The 2D and 3D models and their statistical analyses are the main points of interest in the presented habilitation thesis. The thesis is intended as comments of nine papers published by the author since 2014 concerning the above-mentioned topics. The aim of the commented papers was to further develop the theory, simulations, methods for several statistical analyses and applications of random sets in 2D and 3D using already existing results for both random sets and point processes. Namely, a special planar model of random set called the Quermass-interaction process, see [27] or [40], is studied from an unexplored point of view so that the random disc process [40] is generalised to the process with convex compact grains and its the theoretical results from [27] are corrected and extended (this is done in the commented paper [22]). Further, the thesis concerns two methods for assessing similarity of planar random sets without knowledge of their models when having only two realisations of the random sets on the input. The first one takes into account positions of components in the realisations and assesses their similarity based on increment of mass around a given point (the commented paper [10]). The

second one focuses only on the geometrical characteristics of individual components in the realisations (the commented paper [19]). The last topic includes various 3D material analyses, where a mathematical definition of characteristics describing transmissivity through a material is given, their estimations are derived and properties are studied (the commented paper [47]), a stochastic model for a special material is constructed (the commented paper [46]), a method of processing incomplete dataset of crystalline material is suggested (the commented paper [48]), basic descriptive spatial statistics of data obtained from crystalline material using the 3D EBSD method are provided (the commented papers [32] and [33]) and a difference in segmentation of 2D and 3D EBSD data is briefly discussed (the commented paper [59]).

The thesis is organised as follows. In Chapter 2, basic definitions, notation and previous results are introduced. Chapter 3 consists of the comments on the papers concerning 2D analyses mentioned above, namely Section 3.1 concerns the paper on the Quermass-interaction process with convex compact grains and in Section 3.2, the methods for distinguishing realisations of random sets is described. Chapter 4 is related to 3D models, where Section 4.1 defines new characteristics of material properties, Section 4.2 concerns a construction of a model for a special material consisting of three phases, in Section 4.3, a method for reconstruction of material grain structure is suggested, Section 4.4 introduces descriptive statistics of two material specimen processed in special conditions and finally, Section 4.5 comments differences between 2D and 3D EBSD processing. At the end, the commented papers are enclosed.

Chapter 2

Theoretical background

In this habilitation thesis, there are used notations, definitions and propositions taken from or based mainly on the publications [5], [7], [11], [14], [27], [30], [31], [35], [36], [38], [40], [41], [42], [44], [54], [57] and [58]. This chapter provides a brief summary of that ones needed in the following two chapters dedicated to commented papers.

2.1 Geometry in Euclidean space

Denote $\mathbb{N}$ the set of natural numbers. For $d \in \mathbb{N}$, $\mathbb{R}^d$ denotes the $d-$dimensional Euclidean space, $|.|$ the Euclidean norm in $\mathbb{R}^d$, v_d the $d-$dimensional Lebesgue measure and $\mathcal{H}_k$ the $k-$dimensional Hausdorff measure for $1 \leq k \leq d$.

Let $A, B \subset \mathbb{R}^d$ and $x \in \mathbb{R}^d$. Then $A \cup B$ denotes the union of A and B, $A \cap B$ denotes their intersection, the interior of A is denoted as int$A$ and the boundary of A as ∂A, $\check{A} = \{-a : a \in A\}$ is the set symmetric to A with respect to the origin, $A + x = \{a + x : a \in A\}$ is the translation of the set A by $x \in \mathbb{R}^d$, $A \oplus B = \{x + y : x \in A, y \in B\}$ is the Minkowski addition of the sets A and B, $\emptyset$ is the empty set and $o = (0, ..., 0)$ denotes the origin in $\mathbb{R}^d$.

Open and closed balls with radius $r > 0$ centered at $x \in \mathbb{R}^d$ are denoted as $B(x, r)$ and $b(x, r)$, respectively, i.e. $B(x, r) = \{y \in \mathbb{R}^d : |x - y| < r\}$, $b(x, r) = \{y \in \mathbb{R}^d : |x - y| \leq r\}$.

Further, consider a smooth 2D curve c parameterised by a parameter $\varphi \in [0, \phi] \subset \mathbb{R}$, i.e. $c(\varphi) = (x(\varphi), y(\varphi))$. Then the curvature κ of c is defined as

$$\kappa(c(\varphi)) = \frac{x'(\varphi)y''(\varphi) - x''(\varphi)y'(\varphi)}{(x'^2(\varphi) + y'^2(\varphi))^{3/2}}.$$

It means that $\kappa(c(\varphi)) = \pm 1/R(\varphi)$, where $R(\varphi)$ is the radius of the osculating circle touching the curve in the point $[x(\varphi), y(\varphi)]$ and the choice between "+" and "−" is determined by the local convexity convention.

Finally, let $A \subset \mathbb{R}^3$, the volume $v_3(A) < \infty$ and the surface area $\mathcal{H}_2(\partial A) < \infty$. Then the sphericity of A is

$$\Phi = \frac{\sqrt[3]{\pi}(6v_3(A))^{2/3}}{\mathcal{H}_2(\partial A)}.$$

In Chapter 3 and Chapter 4, we work with special formations built in the Euclidean space, which are introduced in the following subsections.

2.1.1 Laguerre tessellation

Definition 2.1 *A tessellation in $\mathbb{R}^d$ is a countable collection of sets (called cells) $\mathcal{T} = \{C_i \subset \mathbb{R}^d : i \in \mathbb{N}\}$ such that*

- *$C_i \in \mathcal{K}$ for all $i \in \mathbb{N}$,*
- *$\mathrm{int}C_i \cap \mathrm{int}C_j = \emptyset$ for $i \neq j$,*
- *$\cup_{i\in\mathbb{N}} C_i = \mathbb{R}^d$,*
- *$\mathcal{T}$ is locally finite.*

Definition 2.2 *Consider a couple (x, r), $x \in \mathbb{R}^d$, $r \geq 0$, called a marked point, and $y \in \mathbb{R}^d$. Then the power distance (Laguerre distance measure) of the point y from the marked point (x, r) is defined by*

$$pow((x, r), y) = |y - x|^2 - r^2.$$

Definition 2.3 *Consider a set of marked points $\{(x_i, r_i)\}_{i\in I}$ ($I \subseteq \mathbb{N}$ is an index set). The Laguerre tessellation $\mathcal{T}$ is the system of the sets $\{C_i, i \in I\}$, where*

$$C_i = \{y \in \mathbb{R}^d : pow((x_i, r_i), y) \leq pow((x_j, r_j), y) \text{ for all } j \in I\}. \tag{2.1.1}$$

The set C_i is called the i-th cell of the Laguerre tessellation $\mathcal{T}$, the points x_i are called seeds, the values r_i are called weights and the marked points (x_i, r_i) are called weighted generating points.

The Laguerre tessellation, also called power diagram or additively weighted power Voronoi tessellation, is a generalization of the Voronoi tessellation, see [7] or [39], in the sense that in the Voronoi tessellation, the weights r_i are identical, i.e. each cell C_i is the set of points $y \in \mathbb{R}^d$ which are closer to the seed x_i than to any other seed x_j, $j \in I \setminus \{i\}$.

A special usage of the Laguerre tessellation is introduced in [40], where the authors work with the intersection of the Laguerre tessellation with a union of discs, called the power tessellation of the union of discs. It is a helpful tool for simulations of the Quermass-interaction process, see Section 2.2.3 below. It is constructed as follows. Consider a configuration of discs $\mathbf{b} = (b_1(x_1, r_1), \dots, b_n(x_n, r_n))$ and denote the union of the discs as $U_{\mathbf{b}} = \cup_{i=1}^n b_i(x_i, r_i)$. Let $\mathcal{T}$ is the Laguerre tessellation with weighted generating points (x_i, r_i), $i = 1, \dots, n$. Denote

$$\tilde{C}_i = C_i \cap b(x_i, r_i), \quad i = 1, \dots, n,$$

where C_i are the cells of the tessellation $\mathcal{T}$. The system $\{\tilde{C}_i, i = 1, \dots, n\}$ is the power tessellation of the union of discs $U_{\mathbf{b}}$.

2.1.2 Binary images, morphological operations and skeletons in $\mathbb{R}^2$

Definition 2.4 *Consider $A, M \subseteq \mathbb{R}^d$ arbitrary sets. The dilation, erosion, opening and closing of the set M by the structuring element A are defined, respectively, as*

$$\begin{aligned} D_A(M) &= M \oplus A = \{x \in \mathbb{R}^d : \check{A}_x \cap M \neq \emptyset\} = \bigcup_{a \in A} M_a, \\ E_A(M) &= M \ominus \check{A} = \{x \in \mathbb{R}^d : A_x \subseteq M\} = \bigcap_{a \in A} M_{-a}, \\ O_A(M) &= D_A(E_A(M)) = (M \ominus \check{A}) \oplus A, \\ C_A(M) &= E_A(D_A(M)) = (M \oplus A) \ominus \check{A}, \end{aligned}$$

where M_a is the set M translated by a, i.e. $M_a = \{x + a; x \in M\}$.

Definition 2.5 *Let $B(x, r)$ be the open disc with the radius r centred in $x \in \mathbb{R}^2$ and $M \subseteq \mathbb{R}^2$. The disc $B(x, r)$ is called maximal with respect to the set M if there exists no other disc $\tilde{B}(\tilde{x}, \tilde{r})$ included in M and containing $B(x, r)$, i.e.*

$$B(x, r) \subseteq \tilde{B}(\tilde{x}, \tilde{r}) \subseteq M \Rightarrow B(x, r) = \tilde{B}(\tilde{x}, \tilde{r}).$$

Definition 2.6 *Let I_M^{max} be the set of all maximal discs with respect to M. The morphological skeleton (called only skeleton in the sequel) $SK(M)$ of the set M is defined as the set of centres of all maximal discs, i.e.*

$$SK(M) = \{x; B(x, r) \in I_M^{max}, r > 0\}.$$

For $r > 0$, we can define the r-th skeleton subset $S_r(M)$ as

$$S_r(M) = \{x \in SK(M) : B(x, r) \in I_M^{max}\}.$$

Obviously, it holds that

$$SK(M) = \bigcup_{r>0} S_r(M).$$

As the r-th skeleton subset can be obtained by using morphological set transformations, namely

$$S_r(M) = \bigcap_{s>0} \left\{ E_{B(o,r)}(M) \backslash O_{B(o,s)}(E_{B(o,r)}(M)) \right\}, \tag{2.1.2}$$

where o denotes the origin, we can represent the skeleton of M as

$$SK(M) = \bigcup_{r>0} \bigcap_{s>0} \left\{ E_{B(o,r)}(M) \backslash O_{B(o,s)}(E_{B(o,r)}(M)) \right\}. \tag{2.1.3}$$

Binary image

In practice, we often work with planar data in the form of binary images. By a binary image, we mean the matrix of black and white pixels or, alternatively, of 1's and 0's. When we talk about a set A to be in the form of binary image, we mean that A is the set of black pixels (or the set of 1's) in the matrix, while the set of white pixels (or the set of 0's) corresponds to its complement A^c.

Skeleton of a binary image

Let M be in the form of binary image and nB be the binary disc with radius $n \in \mathbb{N}$ using the Manhattan distance, i.e. $nB = \{(x, y) \in \mathbb{Z}^2; |x| + |y| \leq n\}$. Let us find $N \in \mathbb{N}$ such that $E_{NB}(M) \neq \emptyset$ and $E_{(N+1)B}(M) = \emptyset$. Then, the skeleton $SK(M)$ is defined by

$$SK(M) = \bigcup_{n=0}^{N} S_n(M),$$

where $S_n(M) = E_{nB}(M) \backslash O_B(E_{nB}(M))$, cf. (2.1.2) and (2.1.3).

2.1.3 Beta-skeleton

Let X be an arbitrary locally finite set of points (called vertices) in $\mathbb{R}^d$. Consider $x, y \in X$ and $a \geq 1$. Denote

$$m_{x,y}^{(1)} = \frac{a}{2}x + \left(1 - \frac{a}{2}\right) y,$$
$$m_{x,y}^{(2)} = \frac{a}{2}y + \left(1 - \frac{a}{2}\right) x,$$
$$A_a(x, y) = b(m_{x,y}^{(1)}, |m_{x,y}^{(1)} - y|) \cap b(m_{x,y}^{(2)}, |m_{x,y}^{(2)} - x|).$$

Definition 2.7 *The beta-skeleton on X with the parameter a is defined as the graph $G_a(X) = (X, Edg_a)$, where X is the set of vertices and the set of edges is defined by*

$$Edg_a = \{(x, y) : x, y \in X, (X \setminus \{x, y\}) \cap A_a(x, y) = \emptyset\}.$$

In the case $a = 2$, the beta-skeleton coincides with the relative neighborhood graph, see [29]. It is easy to show that the beta-skeleton $G_a(X)$ is a connected graph for all $1 \leq a \leq 2$ if the beta skeleton $G_2(X)$ is a connected graph, see [46].

2.2 Basics of stochastic geometry

2.2.1 Point processes

Definition 2.8 *Let $(\Omega, \mathcal{A}, P)$ be a probability space. Consider $\mathcal{N}$ the system of locally finite subsets of $\mathbb{R}^d$ with the σ-algebra $\mathfrak{N} = \sigma(\{\mathbf{x} \in \mathcal{N} : n(\mathbf{x} \cap A) = m\} : A \in \mathcal{B}, m \in \mathbb{N}_0)$, where $\mathcal{B}$ denotes the system of bounded Borel sets and $n(\mathbf{x})$ denotes the number of points in the configuration $\mathbf{x}$. A point process X defined on $\mathbb{R}^d$ is a measurable mapping from $(\Omega, \mathcal{A})$ to $(\mathcal{N}, \mathfrak{N})$. The distribution P_X of the point process X is given by the relation $P_X(K) = P(\{\omega \in \Omega : X(\omega) \in K\})$ for $K \in \mathfrak{N}$.*

Definition 2.9 *A measure Λ on $\mathcal{B}$ satisfying $\Lambda(A) = \mathbb{E}X(A)$ for all $A \in \mathcal{B}$, where $X(A)$ denotes the number of points in A, is called the intensity measure. If there exists a function $\rho(x)$ for $x \in \mathbb{R}^d$ such that $\Lambda(A) = \int_A \rho(x)dx$, then $\rho(x)$ is called the intensity function. If the intensity function $\rho(x)$ is constant, $\rho(x) = \rho$, the point process is called homogeneous with the intensity ρ. Otherwise, it is said to be inhomogeneous.*

Definition 2.10 *A point process X is called to be*

- *stationary, if the distribution P_X is invariant under translation, i.e. $P_X = P_{X+v}$ for all $v \in \mathbb{R}^d$.*
- *isotropic, if the distribution P_X is invariant under rotation, i.e. $P_X = P_{\psi(X)}$ for each rotation ψ around the origin o in $\mathbb{R}^d$.*
- *motion invariant if it is both stationary and isotropic.*

Definition 2.11 *Consider a given locally finite measure Λ. Then the Poisson point process Y is the process which satisfies:*

- *for any finite collection $\{A_n\}$ of pairwise disjoint sets from $\mathcal{B}$, the numbers of points in these sets, $Y(A_n)$, are independent random variables,*
- *for each $A \in \mathcal{B}$, the number of points in A, $Y(A)$, has Poisson distribution with the parameter $\Lambda(A)$, i.e. $P[Y(A) = k] = \frac{\Lambda(A)^k}{k!} e^{-\Lambda(A)}$ for $k = 0, 1, 2, \ldots$*

Note that Λ in Definition 2.11 is the intensity measure of the Poisson point process.

Definition 2.12 *Let Y be the Poisson point process defined on $S \subseteq \mathbb{R}^d$. Denote $\mathcal{N}_f = \{\mathbf{x} \subset S : n(\mathbf{x}) < \infty\}$ and $\mathfrak{N}_f = \{F \in \mathfrak{N} : F \subseteq \mathcal{N}_f\}$. Then a point process X is given by a density $f : \mathcal{N}_f \to \mathbb{R}^+$ with respect to the Poisson point process Y if*

$$P(X \in F) = \int_F f(\mathbf{x}) P_Y(\mathrm{d}\mathbf{x}) \quad \text{for } F \in \mathfrak{N}_f.$$

Definition 2.13 *Let X be a point process given by a density f with respect to a Poisson process Y. For any configuration $\mathbf{x} \in \mathcal{N}_f$ and any point $v \in S \setminus \mathbf{x}$, the Papangelou conditional intensity is defined as*

$$\begin{aligned} \lambda(\mathbf{x}, v) &= f(\mathbf{x} \cup \{v\})/f(\mathbf{x}) && \text{if } f(\mathbf{x}) > 0, \\ &= 0 && \text{otherwise.} \end{aligned} \tag{2.2.1}$$

Definition 2.14 *A density $f : \mathcal{N}_f \to \mathbb{R}^+$ is called Ruelle stable if there exist positive constants α and β such that $f(\mathbf{x}) \leq \alpha \beta^{n(\mathbf{x})}$ for all $\mathbf{x} \in \mathcal{N}_f$.*

Definition 2.15 *A density $f : \mathcal{N}_f \to \mathbb{R}^+$ is called locally stable if there exists a constant β such that $\lambda(\mathbf{x}, v) \leq \beta$ for all $\mathbf{x} \in \mathcal{N}_f$ and all $v \in S \setminus \mathbf{x}$.*

Remark 2.1 *The local stability implies the Ruelle stability and moreover, both the stability properties imply the integrability of the density with respect to P_Y.*

Definition 2.16 *A point process X is said to be attractive if for the Papangelou conditional intensity, it holds that*

$$\lambda_\theta(\mathbf{x}, v) \geq \lambda_\theta(\mathbf{y}, v) \quad \textit{whenever } \mathbf{y} \subset \mathbf{x},\ \mathbf{x} \in \mathcal{N}_f, v \in S,$$

and repulsive if

$$\lambda_\theta(\mathbf{x}, v) \leq \lambda_\theta(\mathbf{y}, v) \quad \textit{whenever } \mathbf{y} \subset \mathbf{x},\ \mathbf{x} \in \mathcal{N}_f, v \in S.$$

Definition 2.17 *A marked point process on $\mathbb{R}^d$ is a random sequence $X^* = \{[x_i, m_i]\}$, where the points x_i form a point process in $\mathbb{R}^d$ and m_i are the marks (usually random variables) corresponding to each x_i. The marks belong to a given space $\mathbb{M}$ of marks which is assumed to be Polish.*

2.2.2 Random sets

Denote $\mathcal{K}$ the family of compact subsets of $\mathbb{R}^d$, $\mathcal{F}$ the family of closed subsets of $\mathbb{R}^d$ and $\sigma_\mathcal{F} = \sigma(\{F \in \mathcal{F} : K \cap F \neq \emptyset\} : K \in \mathcal{K})$.

Definition 2.18 *Let $(\Omega, \mathcal{A}, P)$ be a probability space. Then a measurable mapping $\mathbf{X}$ from $(\Omega, \mathcal{A})$ to $(\mathcal{F}, \sigma_\mathcal{F})$ is called a random closed set. The distribution $P_\mathbf{X}$ of the random set $\mathbf{X}$ is given by the relation $P_\mathbf{X}(F) = P(\{\omega \in \Omega : \mathbf{X}(\omega) \in F\})$ for $F \in \sigma_\mathcal{F}$.*

Definition 2.19 *The random set $\mathbf{X}$ is called to be*

- *stationary, if the distribution $P_\mathbf{X}$ is invariant under translation, i.e. $P_\mathbf{X} = P_{\mathbf{X}+v}$ for all $v \in \mathbb{R}^d$.*
- *isotropic, if the distribution $P_\mathbf{X}$ is invariant under rotations, i.e. $P_\mathbf{X} = P_{\psi(\mathbf{X})}$ for each rotation ψ around the origin o in $\mathbb{R}^d$.*
- *motion invariant, if it is both stationary and isotropic.*

Definition 2.20 *The volume fraction of a stationary random closed set $\mathbf{X}$ is defined by*

$$p = \mathbb{E}v_d(\mathbf{X} \cap [0,1]^d).$$

Definition 2.21 *Let $X^* = \{[x_i, \mathbf{A}_i]\}$ be a marked point process, where the points x_i form a point process X in $\mathbb{R}^d$ and $\mathbf{A}_1, \mathbf{A}_2, \ldots$ are random compact sets in $\mathbb{R}^d$. Then the union*

$$\mathbf{A} = \cup_{i=1}^\infty (x_i + \mathbf{A}_i) \tag{2.2.2}$$

is called a germ-grain model. The points $x_1, x_2, \ldots$ from (2.2.2) are called germs and the random sets $\mathbf{A}_1, \mathbf{A}_2, \ldots$ are called grains.

Definition 2.22 *Let $X = \{x_1, x_2, \ldots\}$ in Definition 2.21 be the Poisson point process in $\mathbb{R}^d$ and $\mathbf{A}_1, \mathbf{A}_2, \ldots$ independent identically distributed random compact sets in $\mathbb{R}^d$ that are independent on X. If $\mathbb{E}v_d(\mathbf{A}_1 \oplus K) < \infty$ for all compact sets K, then the union (2.2.2) is called the Boolean model.*

2.2.3 Random disc Quermass-interaction process

In general, the Quermass-interaction process, see e.g. [27], [40] or [41], is a special planar germ-grain model $\mathbf{X}$ given by a density with respect to a planar germ-grain model $\mathbf{Y}$, where $\mathbf{Y}$ is usually a Boolean model. The authors of [40] and [41] - the main motivation for study in the commented paper [22] summarised in section 3.1 - focus on a special case, when the grains are discs with random radii and an intensity function of the disc centers $\rho(x)$. The local integrability of this intensity function is assumed to ensure that with probability 1, the point process of the centres of the discs is finite for any bounded region $S \subset \mathbb{R}^2$ such that $\int_S \rho(x)\,\mathrm{d}x > 0$.

In the sequel, we denote $\mathbf{b} = \{b_1, \ldots, b_n\}$ a finite configuration of discs and $U_{\mathbf{b}}$ their union. We assume that the reference process $\mathbf{Y}$ has unit intensity of the disc centers in a given region S and zero intensity otherwise, i.e. $\rho(x) = 1$ for $x \in S$ and $\rho(x) = 0$ for $x \notin S$. Note that it means that the centres of the discs are contained in S but the discs themselves may extend outside S. Further, we assume that $\mathbf{X}$ is absolutely continuous with respect to the reference process $\mathbf{Y}$ and it is given by a density $f(\mathbf{b})$ with respect to $\mathbf{Y}$ in the form

$$f_\theta(\mathbf{b}) = \exp\{\theta \cdot T(U_{\mathbf{b}})\}/c_\theta, \tag{2.2.3}$$

where θ is a real parameter vector, $\cdot$ denotes the usual inner product, $T(U_{\mathbf{b}})$ is a vector of geometrical characteristics depending on the union $U_{\mathbf{b}}$ of the discs from the configuration $\mathbf{b}$ and c_θ is the normalising constant.

Definition 2.23 *The random disc Quermass-interaction process is the disc process $\mathbf{X}$ given by the density (2.2.3) with respect to a given reference random disc Boolean model $\mathbf{Y}$, in which we set $T(U_{\mathbf{b}}) = (A(U_{\mathbf{b}}), L(U_{\mathbf{b}}), \chi(U_{\mathbf{b}}))$, where $A(U_{\mathbf{b}})$ is the area, $L(U_{\mathbf{b}})$ is the perimeter and $\chi(U_{\mathbf{b}})$ is the Euler-Poincaré characteristic (i.e. the number of connected components minus the number of holes) of the set $U_{\mathbf{b}}$.*

Remark 2.2 *The authors of [40] and [41] work with so called extended random disc Quermass-interaction process, where more geometrical characteristics are included in the vector $T(U_{\mathbf{b}})$ (e.g. the number of individual discs, the number of connected components and the number of holes considered separately, etc.), but in the commented paper [22], the classical Quermass-interaction process using only $T(U_{\mathbf{b}}) = (A(U_{\mathbf{b}}), L(U_{\mathbf{b}}), \chi(U_{\mathbf{b}}))$ is studied.*

Simulation by Metropolis-Hastings algorithm

For simulation of the Quermass-interaction process (2.2.3), a simple version of the birth-death type Metropolis-Hastings algorithm studied in [14] is used.

Denote $\lambda_\theta(\mathbf{b}, b)$ Papangelou conditional intensity (2.2.1) of the Quermass-interaction process $\mathbf{X}$. Then the simulating algorithm works as follows.

1. Start from an arbitrary configuration $\mathbf{b_0}$.
2. Suppose that in iteration $t \in \{0, \dots, t_{\max}\}$, we have $\mathbf{b_t} = \{b_1, \dots, b_n\}$.
3. In the iteration $t + 1$:
   (a) with the probability $1/2$, we propose to add a disc b_{n+1}, i.e. to change the configuration to $\mathbf{b_t} \cup \{b_{n+1}\}$
      i. we accept the proposal with probability $\min\{1; H(\mathbf{b_t}, b_{n+1})\}$ and set $\mathbf{b_{t+1}} = \mathbf{b_t} \cup \{b_{n+1}\}$,
      ii. else we set $\mathbf{b_{t+1}} = \mathbf{b_t}$,
   (b) else, we propose to delete a randomly chosen disc b_i from $\mathbf{b_t}$, i.e. to change the configuration to $\mathbf{b_t} \backslash \{b_i\}$
      i. we accept it with probability $\min\{1; 1/H(\mathbf{b_t} \backslash \{b_i\}, b_i)\}$ and set $\mathbf{b_{t+1}} = \mathbf{b_t} \backslash \{b_i\}$,
      ii. else $\mathbf{b_{t+1}} = \mathbf{b_t}$,

where $H(\mathbf{b_t}, b_{n+1}) = \lambda_\theta(\mathbf{b_t}, b_{n+1}) \frac{|S|}{n+1}$ and $H(\mathbf{b_t} \backslash \{b_i\}, b_i) = \lambda_\theta(\mathbf{b_t} \backslash \{b_i\}, x = b_i) \frac{|S|}{n}$.

An advantage of the algorithm is that it requires only the value of Papangelou conditional intensity, so only local computations of the geometrical statistics are needed. Namely, denoting $G(\mathbf{b}, b) = G(U_\mathbf{b} \cup b) - G(U_\mathbf{b})$ the increment of an arbitrary geometrical characteristic $G \in \{A, L, \chi\}$ when adding a disc b to the configuration $\mathbf{b}$, we get the Papangelou conditional intensity in the form

$$\lambda_\theta(\mathbf{b}, b) = \exp\{\theta_1 A(\mathbf{b}, b) + \theta_2 L(\mathbf{b}, b) + \theta_3 \chi(\mathbf{b}, b)\}. \tag{2.2.4}$$

It means that $\lambda_\theta(\mathbf{b}, b)$ depends only on the increments of the geometrical characteristics and neither on the characteristics themselves nor on the normalising constant.

Moreover, the power tessellation of the union of discs $U_\mathbf{b}$ plays an important role. Its usefulness is such that it provides a division of the union of overlapping discs to the union of disjoint convex compact sets, thus instead of inclusion-exclusion formula

$$G(U_\mathbf{b}) = \sum_i G(b_i) - \sum_{\{i_1, i_2\}} G(b_{i_1} \cap b_{i_2}) + \ldots + (-1)^{n+1} \sum_{\{i_1, \dots, i_n\}} G(b_{i_1} \cap \cdots \cap b_{i_n}),$$

we can calculate

$$A(U_{\mathbf{b}}) = \sum_i A(\tilde{C}_i), \quad L(U_{\mathbf{b}}) = \sum_i L(\tilde{C}_i), \quad , \chi(U_{\mathbf{b}}) = N_1 - N_2 + N_3,$$

where $\tilde{C}_i$ are the cells of the power tessellation of the union $U_{\mathbf{b}}$, N_1 is the number of nonempty cells of the corresponding tessellation, N_2 the number of so called interior edges (the lines corresponding to the boundaries between two cells, i.e. to their intersections) and N_3 the number of so called interior vertices (the points where three cells meet, i.e. their intersections), since almost surely, more than three cells cannot meet in one point. It allows to make only local calculations in the sense that when we add or delete a disc, the geometrical characteristics are recalculated only for such cells of the tessellation which are intersected by the added or deleted disc, respectively, since for the other cells, the characteristics remain unchanged.

2.3 Equality of distributions of random functions

A part of this thesis, namely Section 3.2, works with a description of random sets by a group of functions, which are used for assessing dissimilarity of realisations of random sets. Therefore, we need methods for testing equality of distributions of random functions. In [10] and [19], which are summarised in Section 3.2, we use the envelope test from [44], which has become very popular in the field of spatial statistics in the last years, and the test from [18] based on N-distances of probability measures [30], since it appears to be a very promising procedure for this purpose.

2.3.1 Envelope test

Consider $s+1$ exchangeable random objects described by functional characteristics $T_i(u)$, $i = 1, ..., s+1$, $u \in I$ (index set). For each $u \in I$, let $R_i^{\uparrow}(u)$ and $R_i^{\downarrow}(u)$ denote the ranks of the values $T_i(u)$ from the smallest value with rank 1 to the largest one with rank $s+1$ and from the largest value with rank 1 to the smallest one with rank $s+1$, respectively. For each $u \in I$, we define u-wise ranks of $T_i(u)$ as $R_i(u) = \min\left(R_i^{\uparrow}(u), R_i^{\downarrow}(u)\right)$.

In practice, we observe $T_i(u)$ in a discrete index set $I = \{u_1, \ldots, u_n\}$, i.e. $T_i(u) = (T_i(u_1), \ldots, T_i(u_n))$. Denote $\mathbf{N}_i = (N_{i1}, \ldots N_{i\tilde{s}})$, where $N_{ik} = \sum_{j=1}^{n} \mathbf{1}(R_i(u_j) = k)$ and $\tilde{s} = \lfloor (s+2)/2 \rfloor$, and define

$$\mathbf{N}_i \prec \mathbf{N}_j \iff \exists m \leq \tilde{s}\ \forall k < m : N_{ik} = N_{jk}\ \&\ N_{im} > N_{jm}.$$

Then the p-value of the test is

$$p = \frac{1}{s+1}\left(1 + \sum_{i=1}^{s+1} \mathbf{1}(\mathbf{N}_i \prec \mathbf{N}_1)\right).$$

For our purposes, the aim is to test the equality of distributions of two random functions $t^{(1)}$ and $t^{(2)}$. Suppose we have m_1 samples of geometrical objects described by functions $t_1^{(1)}(u), \ldots, t_{m_1}^{(1)}(u)$ coming from a distribution and m_2 samples of geometrical objects described by functions $t_1^{(2)}(u), \ldots, t_{m_2}^{(2)}(u)$ coming from an other distribution. We use a Monte Carlo permutation test, see e.g. [12], for testing whether these two samples of functions come from the same distribution. It works as follows. We consider all functions $t_1^{(1)}(u), \ldots, t_{m_1}^{(1)}(u), t_1^{(2)}(u), \ldots, t_{m_1}^{(2)}(u)$ together and make s random permutations of them. Consequently, we split each permuted pooled sample into two groups of the lengths m_1 and m_2, and calculate the characteristics $T_i(u)$, $i = 2, \ldots s+1$, as the differences between the means of the functions from the i-th permutation.

2.3.2 Test based on N-distance

Definition 2.24 *Let $\mathcal{X}$ be a nonempty set. A map $\mathcal{L} : \mathcal{X} \times \mathcal{X} \to \mathbb{C}$ is called a negative definite kernel if for any $n \in \mathbb{N}$, arbitrary $c_1, ..., c_n \in \mathbb{C}$ such that $\sum_{i=1}^n c_i = 0$ and arbitrary $x_1, ..., x_n \in \mathcal{X}$, it holds that*

$$\sum_{i=1}^{n}\sum_{j=1}^{n} L(x_i, x_j) c_i \bar{c}_j \leq 0. \tag{2.3.1}$$

Definition 2.25 *The negative definite kernel $\mathcal{L}$ is called a strongly negative definite kernel if for an arbitrary probability measure μ and an arbitrary $f : \mathcal{X} \to \mathbb{R}$ such that $\int_{\mathcal{X}} f(x)d\mu(x) = 0$ and $\int_{\mathcal{X}}\int_{\mathcal{X}} \mathcal{L}(x,y)f(x)f(y)d\mu(x)d\mu(y)$ exists and is finite, the relation $\int_{\mathcal{X}}\int_{\mathcal{X}} \mathcal{L}(x,y)f(x)f(y)d\mu(x)d\mu(y) = 0$ implies $f(x) = 0$ μ-almost surely.*

Theorem (Klebanov [30]) Denote $\mathcal{I}_{\mathcal{L}}$ the set of all measures μ such that the integral $\int_{\mathcal{X}}\int_{\mathcal{X}} \mathcal{L}(x,y)d\mu(x)d\mu(y)$ exists. Let $\mathcal{L}(x,y) = \mathcal{L}(y,x)$. Then

$$\begin{aligned}\mathcal{N}(\mu,\nu) = 2\int_{\mathcal{X}}\int_{\mathcal{X}} \mathcal{L}(x,y)d\mu(x)d\nu(y) \\ - \int_{\mathcal{X}}\int_{\mathcal{X}} \mathcal{L}(x,y)d\mu(x)d\mu(y) - \int_{\mathcal{X}}\int_{\mathcal{X}} \mathcal{L}(x,y)d\nu(x)d\nu(y) \geq 0\end{aligned} \tag{2.3.2}$$

holds for all measures $\mu, \nu \in \mathcal{I}_{\mathcal{L}}$ with the equality in the case $\mu = \nu$ only, if and only if $\mathcal{L}$ is a strongly negative definite kernel.

In the sequel, the term $\mathcal{N}(\mu,\nu)$ from (2.3.2) is called the N-distance of the measures μ and ν. There are introduced some examples of strongly negative definite kernels $\mathcal{L}$ in [30] for $\mathcal{X}=\mathbb{R}^d$, e.g. the Euclidean distance used in [19], which is summarised in Section 3.2.2. Further, since we often focus on testing the equality in distribution of random functions $t^{(1)}$ and $t^{(2)}$, we use the kernel from [18] constructed especially for random functions. Namely, when we evaluate the testing functions $t^{(1)}$ and $t^{(2)}$ in discrete arguments from the index set $I=\{u_1,\dots,u_n\}$, the kernel is

$$\mathcal{L}(t^{(1)},t^{(2)})=\sum_{K\in 2^I}\left(\sum_{u_k\in K}\left(t^{(1)}(u_k)-t^{(2)}(u_k)\right)\right)^2\Bigg)^{1/2},$$

where 2^I denotes the set of all subsets of I.

The estimate of the N-distance of the (random) functions $t^{(1)}$ and $t^{(2)}$ based on the random samples $t_1^{(1)}(u),\dots,t_{m_1}^{(1)}(u)$ and $t_1^{(2)}(u),\dots,t_{m_2}^{(2)}(u)$, respectively, is evaluated as

$$\hat{\mathcal{N}}_1=\frac{2}{m_1m_2}\sum_{i=1}^{m_1}\sum_{j=1}^{m_2}\mathcal{L}(t_i^{(1)},t_j^{(2)})-\frac{1}{m_1^2}\sum_{i=1}^{m_1}\sum_{j=1}^{m_1}\mathcal{L}(t_i^{(1)},t_j^{(1)})-\frac{1}{m_2^2}\sum_{i=1}^{m_2}\sum_{j=1}^{m_2}\mathcal{L}(t_i^{(2)},t_j^{(2)}). \tag{2.3.3}$$

It plays the role of the testing characteristic so that we again use Monte-Carlo permutation test, i.e. we make s permutations of all functions $t_1^{(1)}(u),\dots,t_{m_1}^{(1)}(u),t_1^{(2)}(u),\dots t_{m_2}^{(2)}(u)$, split each permuted pooled sample into two groups of the lengths m_1 and m_2, respectively, and analogously to (2.3.3), we calculate $\hat{\mathcal{N}}_i$ for the i-th permutation, $i=2,\dots,s+1$. Then the p-value of the test is

$$p=\frac{\#\{i\in\{2,\dots,s+1\}:\hat{\mathcal{N}}_i\geq\hat{\mathcal{N}}_1\}+1}{s+1}.$$

2.4 Cross-entropy method

The cross-entropy method (denoted as CE in the sequel) is a stochastic optimisation method which is able to solve many continuous optimisation problems, see [31] and [54]. Consider a cost function c. The idea of the CE method is to find the global minimum $z^*=\arg\min_{z\in\mathbb{R}^d}c(z)$ by a stochastic searching from a d-dimensional distribution so that z^* is the global maximum with probability 1.

We assume that the distribution of the location of the global minimum z^* is given by a parametric density $f(z;\theta)$ with a parameter $\theta\in\mathbb{R}^{\tilde{d}}$. Consider $l,n\in\mathbb{N}$ such that $n>l$. Generate a sample $(z^{(1)},...,z^{(n)})$ from a given density $f(z;\theta)$ and calculate the corresponding values $c(z^{(1)}),...,c(z^{(n)})$. Let us call the set of l members from the sample

$(z^{(1)}, ..., z^{(n)})$ with the smallest values of the cost function $c(z)$ the elite set. Then, the parameter θ is updated by computing the maximal likelihood estimate from the samples in the elite set. This procedure is repeated until the distribution given by $f(z;\theta)$ is nearly determined.

The standard choice of the density $f(z;\theta)$ is

$$f(z;\theta) = \Pi_{i=1}^{d} \varphi(z_i; \mu_i, \sigma_i), \tag{2.4.1}$$

where $z = (z_1, ..., z_d)$, $\theta = (\mu_1, \sigma_1, ..., \mu_d, \sigma_d)$ and $\varphi(z; \mu, \sigma)$ is the density of the normal distribution with the mean $\mu \in \mathbb{R}$ and the standard deviation $\sigma > 0$. In this case, the CE algorithm works as follows:

1. **Initialization**: Choose an initial parameter $\theta^{(0)} = (\mu_1^{(0)}, \sigma_1^{(0)}, ..., \mu_d^{(0)}, \sigma_d^{(0)})$ and set $k = 0$.

2. **Sampling**: Generate a sample $(z^{(1)}, ..., z^{(n)})$ from the density $f(z;\theta^{(k)})$ and select the elite set $(z^{(e_1)}, ..., z^{(e_l)})$ from this sample with respect to the cost function c.

3. **Updating**: Calculate the MLE of the parameter θ from the elite sample, which corresponds to the sample means and the standard deviations for the density (2.4.1), i.e. calculate

$$\bar{z}_1 = \frac{1}{l}\sum_{i=1}^{l} z_1^{(e_i)}, ..., \bar{z}_d = \frac{1}{l}\sum_{i=1}^{l} z_d^{(e_i)},$$

$$z_1^{(sd)} = \sqrt{\frac{1}{l-1}\sum_{i=1}^{l}(z_1^{(e_i)} - z_1^{(sd)})^2}, ..., z_d^{(sd)} = \sqrt{\frac{1}{l-1}\sum_{i=1}^{l}(z_d^{(e_i)} - \bar{z}_d)^2},$$

   and set $\theta^{(k+1)} = (\mu_1^{(k+1)} = \bar{z}_1, \sigma_1^{(k+1)} = z_1^{(sd)}, ..., \mu_d^{(k+1)} = \bar{z}_d, \sigma_d^{(k+1)} = z_d^{(sd)})$.

4. **Iteration**: If the value of the cost function c does not decrease significantly for a certain number of steps, see (2.4.2) below, the algorithm is stopped. Otherwise, set $k = k + 1$ and go to the step 2.

Variance injection

When using the CE algorithm, the variance of the distribution $f(z;\theta)$ may decrease too quickly, hence the algorithm converges to a local minimum instead of the global minimum z^*. To avoid this situation, we can use variance injection introduced in [4]. The idea is to occasionally increase the variance of the distribution $f(z;\theta)$ to ensure that more realisations of this distribution leave the neighborhood of the local minimum. Usually,

the variance injection is applied when the cost function c does not decrease significantly for a chosen number of iterations τ, more precisely, when

$$\left| \frac{c_{\min}^{(k)} - \max_{t \in \{k-\tau,\dots,k-1\}} c_{\min}^{(t)}}{c_{\min}^{(k)}} \right| < \delta_{inject},$$

where $\delta_{inject} > 0$ is a chosen boundary and $c_{\min}^{(k)} = \min_{i=1,\dots,n} c(z^{(i)})$ for the sample $(z^{(1)}, ..., z^{(n)})$ of the k-th step. The size of the variance increment may depend on the current value of the cost function. If the cost function $c(z) = \sum_{i=1}^{m} c_{loc}(z_i)$, then the increment of the standard deviation σ_i can be proportional to the value $c_{loc}(z_i)$. If the effect of variance injection become negligible, i.e.

$$\left| \frac{c_{\min}^{(k)}}{c_{\min}^{(k')}} \right| > \gamma,$$

where $\gamma \in (0, 1)$ and $c_{\min}^{(k')}$ is the minimal cost function prior to the previous variance injection, we stop using variance injection.

The CE algorithm is terminated when

$$\left| \frac{c_{\min}^{(k)} - \max_{t \in \{k-\tau,\dots,k-1\}} c_{\min}^{(t)}}{c_{\min}^{(k)}} \right| < \delta_{term}, \tag{2.4.2}$$

where $\delta_{inject} > \delta_{term} > 0$.

Chapter 3

Statistical methods for random sets in 2D

3.1 Quermass-interaction process with convex compact grains

A generalisation of the random disc Quermass-interaction process from Section 2.2.3 is studied in [22]. Here, we focus on the Quermass-interaction process with convex compact grains. We follow up the papers [27] and [40], which provide some theoretical results, especially several stability properties. A simulation algorithm is moreover suggested in [40]. In [22], we add some new theoretical results, modify the simulation method and moreover, we correct some flaws which appear in [27]. The summary of [22] follows.

Definition 3.1 *The Quermass-interaction process with convex compact grains is a germ-grain model* $\mathbf{X}$*, which is absolutely continuous with respect to a reference germ-grain model* $\mathbf{Y}$ *and given by the density* $f(\mathbf{x})$ *with respect to* $\mathbf{Y}$*, where the density is of the form*

$$f_\theta(\mathbf{x}) = c_\theta^{-1} \exp\{\theta_1 A(U_\mathbf{x}) + \theta_2 L(U_\mathbf{x})\} + \theta_3 \chi(U_\mathbf{x})\}. \tag{3.1.1}$$

The germs lie in a bounded set $S \subset \mathbb{R}^d$ *and the grains of* $\mathbf{Y}$*, as well as the grains* $\{x_1, \ldots, x_n\}$ *of each configuration* $\mathbf{x}$*, are realisations of a random planar convex compact set.*

First, we wonder whether the Quermass-interaction process from Definition 3.1 is well defined, i.e. whether the density (3.1.1) is measurable and integrable. While the measurability is ensured by Lemma 2.2 from [27], we focus only on the Ruelle stability, which

is sufficient condition for integrability of the density (3.1.1), see [27]. The following two propositions are the first theoretical results in [22]. Note that it is a corrected version of the results from [27], in which we remove the mistakes as mentioned above.

Proposition 3.1 *1. Consider the (A, L)-interaction process, i.e. the process with*

$$f_\theta(\mathbf{x}) = c_\theta^{-1} \exp\{\theta_1 A(U_\mathbf{x}) + \theta_2 L(U_\mathbf{x})\}.$$

Then the density f_θ is Ruelle stable if one of the following conditions holds:

(a) $\theta_1 \leq 0$ and $\theta_2 \leq 0$,

(b) $\theta_1 > 0, \theta_2 > 0$ and $A(x)$ and $L(x)$ are bounded above for all grains x from the support of the distribution of the grains.

2. Consider the χ-interaction process, i.e. the process with

$$f_\theta(\mathbf{x}) = c_\theta^{-1} \exp\{\theta_3 \chi(U_\mathbf{x})\},$$

whose grains are random discs. Then the density f_θ is Ruelle stable for all $\theta_3 \in \mathbb{R}$.

3. Consider the χ-interaction process, i.e. the process with

$$f_\theta(\mathbf{x}) = c_\theta^{-1} \exp\{\theta_3 \chi(U_\mathbf{x})\},$$

whose grains are random polygons satisfying the following conditions:

(a) There exists an angle $\phi > 0$ and a radius $r > 0$ such that for all grains x (i.e. realisations of the random polygon forming the grains) and for all vertices v_x of each x, the intersection $b(v_x, r) \cap U_\mathbf{x}$ is a circular sector of angle at least ϕ.

(b) There exists a constant $K \in \mathbb{R}$ such that $x \subset [-K, K]^2$ for all realisations of the grains x.

Then the density f_θ is Ruelle stable.

The proof can be found in [22]. Note that the parts 1. and 2. are slight modifications of the proof in [27] while the point 3. requires more detailed analysis.

In the following proposition, we provide the conditions for the Ruelle stability of the χ-interaction process with a different type of grains, namely for grains with smooth boundary and bounded curvature (e.g. ellipses with bounded ratios of the main axes, several ovoid shapes etc.). The proof can be again found in [22].

Proposition 3.2 *Consider a convex compact grain x satisfying the following conditions:*

1. *There exists a constant $K > 0$ such that $x \subset [-K, K]^2$.*

2. *For each grain x, it holds that its boundary $\partial x \in \mathcal{C}^1$, i.e. there exists a tangent of the grain in each point on its boundary.*

3. *Denote T_v^x, T_w^x the tangents of ∂x in the points $v, w \in \partial x$ and $\alpha_{v,w}^x$ the smaller angle between T_v^x and T_w^x. Then there exists a constant L such that*

$$\alpha_{v,w}^x \leq L||v - w||.$$

   Just note that for a grain x such that $\partial x \in \mathcal{C}^2$, it means that the curvature of ∂x is bounded.

4. *There exists $\tilde{r} > 0$ such that for any two grains $x_i, x_j, i \neq j$, it holds that*

$$n(\partial x_i \cap \partial x_j \cap b(u, \tilde{r})) \leq 2, \ \forall u \in \mathbb{R}^2,$$

Then the χ-interaction process with convex compact grains satisfying the conditions (1)–(4) above is Ruelle stable.

Then, we study the attractiveness and the repulsiveness. These properties for the random disc Quermass-interaction process are introduced in [40]. In [22], they are generalised and formulated as follows.

Proposition 3.3 *For the Quermass-interaction process with convex compact grains, it holds:*

1. *The process with $\theta_2 = \theta_3 = 0$ and $\theta_1 \neq 0$, i.e. the A-interaction process, is attractive for $\theta_1 < 0$ and repulsive for $\theta_1 > 0$.*

2. *The process with $\theta_1 = \theta_3 = 0$ and $\theta_2 \neq 0$, i.e. the L-interaction process, is*

   (a) *both attractive and repulsive, if $L(U_{\mathbf{X}}) = 0$,*

   (b) *attractive for $\theta_2 < 0$ and repulsive for $\theta_2 > 0$, if $A(U_{\mathbf{X}}) = 0$ and $P(L(U_{\mathbf{X}}) > 0) > 0$,*

   (c) *neither attractive nor repulsive, if $P(A(U_{\mathbf{X}}) > 0) > 0$.*

3. *The process with $\theta_1 = \theta_2 = 0$ and $\theta_3 \neq 0$, i.e. the χ-interaction process, is*

   (a) *both attractive and repulsive, if $L(U_{\mathbf{X}}) = 0$,*

   (b) *neither attractive nor repulsive, if $P(L(U_{\mathbf{X}}) > 0) > 0$.*

Further contribution of the paper [22] is the suggestion of an easy algorithm for programming simulation of the process. The simulation procedure uses the basic Metropolis-Hastings birth-death algorithm, see [42], described above in Section 2.2.3. Several significant advantages of this procedure, described in [40] for the random disc Quermass-interaction process, are used here as well. Let us name e.g. the advantage of the Papangelou conditional intensity being independent on the normalising constant, which is explicitly inexpressible, or simplification of calculation of the Papangelou conditional intensity using local calculations of the increments of the characteristics in the density (3.1.1).

In [40], the authors uses the power tessellation of the union of discs to calculate the characteristics increments. It is enabled due to properties of intersections of discs.. Unfortunately, this is not possible in case of general convex compact grains. At least for the process with the grains being ellipses with random axes, we could use a generalisation of the Laguerre tessellation, see e.g. [2]. Nevertheless, in this case, we face to the problem that the cells of the tessellation are not convex, so it does not allow us to work with the tessellation in the same way as done in [40]. Therefore, we leave the idea of using tessellations and we base the calculation of the increments of the geometrical characteristics on discretisation, i.e. on a division of the observation window to a grid of pixels, which keep the needed values. Namely, each pixel has the information about

- the area, i.e. 1 or 0, when the pixel forms the set or not, respectively,
- the perimeter whereby the pixel contributes to the perimeter of the union,
- the information about the surround of the pixel in order to use the classical algorithm for calculation of the Euler-Poincaré characteristic based on local patterns,
- list of grains overlapping the pixel.

Then in case when a grain is added or deleted, the geometrical characteristics are recalculated only in the pixels overlapped by this grain.

In [22], we show several examples of simulated realisations, while we focus on the process with the grains being random rectangles or ellipses with random axes and rotations (for other grains, only slight modifications are needed). Moreover, we provide a comparison to the simulation method of the random disc process using the power tessellation from [40]. We obtain natural conclusion that our method depends on resolution of realisation images, while in case of higher resolution, the geometrical characteristics of the simulated realisations are close to the realisations obtained by the method from [40], but the time consumption grows quadratically as the grid refines, of course. Therefore, we recommend to use the method from [40] for simulations of the random disc Quermass-interaction process, while for the Quermass-interaction process with non circular grains, our method is suitable option since it works reliably as well.

3.2 Assessing similarity of realisations of random sets

Although it is beneficial to know a particular model of a given random set, it is not always necessary to find it. One can be interested only in distinguishing between two realisations in the sense to decide whether they come from the same or different models, or at least whether they are similar in some specific way. In case of parametric models, we can estimate the parameters and apply several statistical tests of their equality. For non-parametric models or when the model is unknown, different characteristics, e.g. the covariance function, the contact distribution function, see [7], or functions on morphological operations can be employed. Unfortunately, these methods are not sufficient in many cases. Moreover, we get only one estimation of the given characteristic from one realisation, so it can be difficult to formulate the task of comparing two random sets when only one realisation from each random set is available.

Some methods of comparing random sets, when only one realisation of each set is available, has been recently presented in [17], [18] and [20]. Although the methods have very satisfactory results, they have many disadvantages or are unusable in some situations. Some of them work only with approximations of the given realisations or require many choices of free parameters. All of them moreover focus only on shapes of components in the realisations but not on their mutual positions, which may also play an important role for assessing similarity in some practical cases. The papers [10] and [19], which are summarised in this section, improve some weaknesses of the above-mentioned papers.

3.2.1 Assessing similarity of random sets via morphological skeletons of their realisations

The method presented in [10] is based on using morphological skeletons and the corresponding maximal discs introduced in Section 2.1.2, which uniquely describe realisations of random sets. We consider each realisation as a binary image, so the skeleton together with the radii of the corresponding maximal discs can be identified with a realisation of a marked point process on a grid. For this process, we define a function, which is analogous to the mark-weighted K-function, see [25], with the difference that we evaluate the function in each point of the skeleton. Then, we sample a group of the functions from each realisation and compare their distributions using the envelope test and the N-distance test described in Section 2.3.

More precisely, we first define similarity of random sets in the following way. Consider a stationary random set $\mathbf{X}$ in the form of a random binary image observed in a bounded window $W \subset \mathbb{Z}^2$, we randomly choose a realisation $\mathbf{x}$ of $\mathbf{X}$ and from its skeleton $SK(\mathbf{x})$,

we randomly choose a point x. Then for $\mathbf{X}$, we define the random function

$$t^{\mathbf{X}}(u) = \sum_{r=1}^{\infty} \sum_{y \in W^{grid}} r\mathbf{1}(||x-y|| < u)\mathbf{1}(B(y,r) \in I_{\mathbf{x}}^{max}) \quad u \in \mathbb{N}, \tag{3.2.1}$$

where $I_{\mathbf{x}}^{max}$ is the set of all maximal discs of the realisation $\mathbf{x}$ and W^{grid} denotes the set of all pixels of the observation window W. The pixels play the role of units in the sequel.

Definition 3.2 *Random sets (binary images)* $\mathbf{X}$ *and* $\mathbf{Y}$ *are said to be similar if the distributions of the corresponding functions* $t^{\mathbf{X}}(u)$ *and* $t^{\mathbf{Y}}(u)$ *defined by* (3.2.1) *are equal.*

Note that the similarity of random sets is here generally defined through the probability distribution of a random function describing increases of mass around points of the skeletons corresponding to the realisations of the random sets. It means that we take into account mutual positions of the components.

As mentioned above, we focus on the situations when only two realisations of random sets available and our aim is to decide whether they come from similar random sets. We address the problem to testing the equality of distributions of the corresponding functions $t(u)$ based on samples of the functions obtained from the realisations (called testing functions in the sequel). The approach is as follows.

Consider a realisation $\mathbf{x}$ of a stationary random set as a binary image and let $SK(\mathbf{x})$ be its skeleton. For each $x_i \in SK(\mathbf{x})$, denote r_i the radius of the corresponding maximal disc and approximate the testing function (3.2.1) at the point x_i by

$$t_i(u) = \sum_{j \neq i} r_j \mathbf{1}(||x_i - x_j|| < u), \quad u = 1, 2, \ldots, U_{max}, \quad U_{max} \in \mathbb{N}. \tag{3.2.2}$$

Comparing to the pure estimate of the K-function for classical point processes, a complication occurs in our case. Since the skeletons basically form curves of the centres of the maximal discs, there are many pairs of the points x_i and x_j very close to each other, and the testing functions t_i and t_j are very similar when x_i is too close to x_j. This phenomenon can significantly effect the results, because the tests require exchangeability of the inputs, which is not satisfied for the functions whose reference points are too close. Therefore, we set a minimal distance D_{min} of the points in which the test functions are evaluated. The set M of such testing points is constructed using the second Matérn point process thinning method, see [7], applied to the set $SK(\mathbf{x})$. Then, we use only the test functions t_i corresponding to the points $x_i \in M$ (called testing points) for further statistical inference. A comprehensive discussion on appropriate choices of U_{max} and D_{min} is provided in [10].

After we set the above-mentioned parameters, we compare two realisations $\mathbf{x}^{(1)}$ and $\mathbf{x}^{(2)}$. We make a sample of the testing points, evaluate the corresponding testing functions $t_1^{(1)}(u), \ldots, t_{m_1}^{(1)}(u)$ and $t_1^{(2)}(u), \ldots, t_{m_2}^{(2)}(u)$, and apply the envelope test from Section 2.3.1 and the test based on N-distance from Section 2.3.2.

At the end of [10], we show a simulation study including three types of simulated realisations and compare both the realisations of the same models and the realisations of different models. From 100 pairs of realisations, we evaluate the p-values and study their distributions, expecting their values close to zero for pairs of realisations of different models and uniformly distributed on the interval $\langle 0, 1\rangle$ for pairs of realisations of the same model. The method gives excellent results in the simulation study both for assessing similarity of the same models and for distinguishing between different models.

3.2.2 Two-step method for assessing similarity of random sets

In [19], we come back to the similarity of two realisations of random sets based only on the shape of their components and not on their mutual positions as described in the subsection above. However, instead of the heuristic approximation of realisations used in [20] and [18], we describe geometrical characteristics of the components uniquely.

The characteristics we work with are the curvature of boundaries of the components and the ratio of the perimeter and the area of each component. The idea is that the boundary of a (connected) planar set X is given by a curve c whose theoretical curvature $\kappa(z)$ evaluated in a given point $z \in c$ can be approximated as follows. For a disc $b(z, r)$ with the center in z and a radius r small enough, it holds that

$$\kappa(z) \approx \frac{3A^*_{b(z,r)}}{r^3} - \frac{3\pi}{2r} = \frac{3\pi}{r}\left(\frac{A^*_{b(z,r),X}}{A_{b(z,r)}} - \frac{1}{2}\right),$$

where $A_{b(z,r)}$ is the area of the disc $b(z, r)$ and $A^*_{b(z,r),X}$ is the area of $b(z, r) \cap X$. Thus, considering a connected random set $\mathbf{X}$, we focus only on the ratio $A^*_{b(z,r),\mathbf{X}}/A_{b(z,r)}$. Denote $B_{\mathbf{X}}$ the boundary of $\mathbf{X}$, and define

$$\tilde{\kappa}_{\mathbf{X},r}(u) = |B_{\mathbf{X}}|^{-1} \int_{B_{\mathbf{X}}} \mathbf{1}\left\{\frac{A^*_{b(z,r),\mathbf{X}}}{A_{b(z,r)}} \leq u\right\} dz, \quad u \in [\,0, 1],$$

which is basically an analogy of the distribution function of the (shifted and rescaled) curvature at points on the boundary. The main difference is that it is evaluated for all boundary points, so it describes the distribution for strongly dependent values. The object of our interest is the function analogous to the corresponding density function, i.e.

$$t_{\mathbf{X},r}(u) = \tilde{\kappa}'_{\mathbf{X},r}(u).$$

Finally, denote $R_{\mathbf{X}}$ the random variable describing the ratio of the perimeter and the area of the random set $\mathbf{X}$, and define the similarity of random sets in the following way.

Definition 3.3 *Two connected random sets* $\mathbf{X}$ *and* $\mathbf{Y}$ *are considered to be similar if the distributions of* $\lim_{r\to 0} t_{\mathbf{X},r}$ *and* $\lim_{r\to 0} t_{\mathbf{Y},r}$ *as well as the distributions of* $R_{\mathbf{X}}$ *and* $R_{\mathbf{Y}}$ *are equal.*

Since in practice, we observe realisations of the random sets in the form of binary images, we adjust the task of assessing dissimilarity of the realisations consisting of black and white pixels as in the previous subsection. The pixels again play the role of units. The ratio of the perimeter and the area is simply given by the number of boundary pixels divided by the number of all pixels of the component. For evaluating the function describing the curvature, fix a radius $r \in \mathbb{N}$, denote P the set of all pixels of given realisation (binary image) $\mathbf{x}$, further denote $z_1, \ldots, z_n$ all boundary pixels, and for each boundary pixel z_i, define

$$K(z_i) = \frac{\sharp\{p \in P : p \in b(z_i, r) \cap \mathbf{x}\}}{\sharp\{p \in P : p \in b(z_i, r)\}}.$$

Then, the approximation of the function $t_{\mathbf{X},r}(u)$ is

$$t(u) = \frac{\sharp\{i \in \{1, \ldots, n\} : K(z_i) \in [u - 1/l, u)\}}{n}, \quad u = \frac{1}{l}, \frac{2}{l}, \ldots, 1.$$

Moreover, we suppose that in practice, the realisations consist of more than one component. If we can suppose that the components are independent and come from the same distribution, then we can define similarity of two random sets it the way that they are considered to be similar, if their components are similar in the meaning of Definition 3.3. In this way, we obtain two samples of components from two realisations $\mathbf{x}^{(1)}$ and $\mathbf{x}^{(2)}$, respectively, which are then used as the input samples $\mathbf{x}_1^{(1)}, \ldots, \mathbf{x}_{m_1}^{(1)}$ and $\mathbf{x}_1^{(2)}, \ldots, \mathbf{x}_{m_2}^{(2)}$, and apply the permutation test based on N-distance from Section 2.3.2 to both geometrical characteristics. The final p-value is then given by

$$p = \frac{\sharp\{i \in \{2, \ldots, s+1\} : \hat{\mathcal{N}}_i^R \geq \hat{\mathcal{N}}_1^R \wedge \hat{\mathcal{N}}_i^t \geq \hat{\mathcal{N}}_1^t\} + 1}{s+1},$$

where $\mathcal{N}^R$ and $\mathcal{N}^t$ are the N-distances of the corresponding perimeter/area ratios and t functions, respectively.

The procedure in [19] is justified by an extensive simulation study, similar to that one in [10], which is summarised above in Section 3.2.1. Except very good results, it also provides a discussion on the influence of the mutual dependence of individual components together with suggestions on how to solve complications caused by this dependence. Finally, the procedure is equipped by an application to real data from the medical environment, namely it is used to distinguish between mastopathic tissue and tissue affected by breast cancer.

Chapter 4

Stochastic modelling in 3D and statistics in materials sciences

4.1 Geodesic tortuosity and constrictivity

Geodesic tortuosity and constrictivity are structural characteristics, which describe transport properties in porous or composite materials. Briefly said, the tortuosity measures the windedness of paths through a given material and the constrictivity measures the bottleneck effects in the material. Although these characteristics are widely used, they were not analysed from a mathematical point of view before study in [47]. This section summarises the results published in [47], where we present tools for mathematical modelling and statistical analysis of these two characteristics.

An empirical estimator of the tortuosity τ, see [16], [49], [58] or [60], and the constrictivity β, see [24] or [60], has been already introduced, nevertheless, a precise definition of τ and β for random closed sets allows us not only to introduce the appropriate estimators of these two characteristics, but also to derive their properties which has not been studied yet.

Consider a stationary random closed set $\mathbf{X}$ in $\mathbb{R}^d$. Then, the mean geodesic tortuosity τ is defined as the expected length of the shortest path from the origin o to a predefined hyperplane through the interior of $\mathbf{X}$ under the condition that at least one such a path exists. For the precise definition of τ, we use the following notation. Let $F, F_0 \in \mathcal{F}$ (the family of closed subsets of $\mathbb{R}^d$), then

- $H_1 = \{x \in \mathbb{R}^d : x_d = 1\}$ denotes the hyperplane orthogonal to the d-th standard unit vector e_d) at the distance 1 to the origin o.

- $\mathcal{P}_{F_0}(x, F) = \{f : [0,1] \to \text{int}F_0 \text{ Lipschitz} : f(0) = x, f(1) \in F\}$ denotes the set of all paths from $x \in \mathbb{R}^d$ to F through the interior of the set F_0.
- $\mathcal{C}_{F_0}(F) = \{x \in \mathbb{R}^d : \mathcal{P}_{F_0}(x, F) \neq \emptyset\}$ denotes the set of all points connected to the set F through the interior of the set F_0.
- $\gamma_{F_o}(x, F) = \inf_{f \in \mathcal{P}_{F_0}(x,F)} \mathcal{H}_1(f[0,1])$ denotes the length of the shortest path from x to the set F through the interior of the set F_0.

The tortuosity τ is defined in [47] in the following way.

Definition 4.1 *The mean geodesic tortuosity of* $\mathbf{X}$ *is defined by*

$$\tau = \begin{cases} \mathbb{E}[\gamma_{\mathbf{X}}(o, H_1) | o \in \mathcal{C}_{\mathbf{X}}(H_1)], & \text{if } P(o \in \mathcal{C}_{\mathbf{X}}(H_1)) > 0, \\ -\infty, & \text{otherwise.} \end{cases}$$

Further in [47], we introduce a general definition of the constrictivity β. So far, the term constrictivity has been used mainly in 3-dimensional case. It has been defined as $\beta = \left(\frac{r_{min}}{r_{max}}\right)^2$, where r_{min} is the maximum radius r such that in expectation, at least half of the set $\mathbf{X} \cap [0,1]^3$ can be filled by an intrusion of balls with radius r in transport direction, and r_{max} is the maximum radius r such that in expectation at least half of the set $\mathbf{X} \cap [0,1]^3$ can be covered by balls of radius r, where these balls are completely contained in $\mathbf{X}$. For the precise definition of β, we use following notation:

- $M^{\ominus r} = E_{b(o,r)}(M)$ the erosion of the set $M \subset \mathbb{R}^d$ by the closed ball $b(o, r)$.
- $H_o = \{x \in \mathbb{R}^d : x_d = 0\}$ denotes the hyperplane orthogonal to the $d-$th standard unit vector e_d) containing the origin o.

The general and mathematically precise definition of the constrictivity from [47] follows.

Definition 4.2 *Let* $\mathbf{X}$ *be a stationary random closed set, and define*

$$r_{min} = \sup\{r \geq 0 : \mathbb{E}[v_d((\mathcal{C}_{\mathbf{X}^{\ominus r}}(H_o) \oplus b(0,r)) \cap [0,1]^d)] \geq \frac{1}{2}\mathbb{E}[v_d(\mathbf{X} \cap [0,1]^d)]\}$$
$$r_{max} = \sup\{r \geq 0 : \mathbb{E}[v_d(O_{b(o,r)}(\mathbf{X}) \cap [0,1]^d)] \geq \frac{1}{2}\mathbb{E}[v_d(\mathbf{X} \cap [0,1]^d)]\}.$$

Then, the constrictivity β *of* $\mathbf{X}$ *is defined by*

$$\beta = \left(\frac{r_{min}}{r_{max}}\right)^{d-1}.$$

From Definition 4.2 of r_{min} and r_{max}, we immediately have $r_{min} \leq r_{max}$, hence $0 \leq \beta \leq 1$. The interpretation of β is such that β close to 0 corresponds to many narrow constrictions in $\mathbf{X}$, while there are no constrictions at all in the case $\beta = 1$.

In the next part, we establish estimators of τ, r_{min} and r_{max} and show the sufficient conditions which ensure the strong consistence of these estimators.

Denote $W_N = [\frac{-N}{2}, \frac{N}{2}]^{d-1} \times [0,1]$ and $H_{o,N} = H_o \cap W_N$. The estimator of the mean geodesic tortuosity is defined by

$$\hat{\tau}_N = \frac{1}{\mathcal{H}_{d-1}(\mathcal{C}_{\mathbf{X}}(H_1) \cap H_{o,N})} \int_{\mathcal{C}_{\mathbf{X}}(H_1) \cap H_{o,N}} \gamma_{\mathbf{X}}(x, H_1) \mathcal{H}_{d-1}(dx).$$

Under some mild conditions on a random closed set $\mathbf{X}$ introduced in [47], the following theorem holds.

Theorem 4.1 *Let* $\mathbb{E}[\gamma_{\mathbf{X}}(o, H_1) \mathbb{I}_{o \in \mathbb{C}_{\mathbf{X}}(H_1)}] < \infty$. *Then* $\hat{\tau}_N$ *is a strongly consistent estimator of* τ.

To evaluate $\hat{\tau}_N$, we have to know the lengths of all shortest paths from $H_{o,N}$ to H_1. However, this information may not be available, because we usually observe the random set $\mathbf{X}$ only in a bounded sampling window. Therefore, a new estimator $\hat{\tau}_N^{\alpha}$ of the mean geodesic tortuosity based on a bounded window is establish in [47] and sufficient conditions for strong and weak consistency of the estimator are shown.

The estimators of r_{min} and r_{max} from data observed in the window W_N are defined by

$$\hat{r}_{min,N} = \sup\{r \geq 0 : 2v_d((\mathcal{C}_{\mathbf{X} \ominus r}(H_o) \oplus b(o,r)) \cap W_N) \geq v_d(\mathbf{X} \cap W_N)\},$$
$$\hat{r}_{max,N} = \sup\{r \geq 0 : 2v_d(O_{b(o,r)}(\mathbf{X}) \cap W_N) \geq v_d(\mathbf{X} \cap W_N)\}.$$

For a stationary random closed set $\mathbf{X}$, the following theorems are introduced and proved in [47].

Theorem 4.2 *If there exists at most one* $r_0 \geq 0$ *such that*

$$2\mathbb{E}[v_d((\mathcal{C}_{\mathbf{X} \ominus r_0}(H_o) \oplus b(o, r_0)) \cap W_1)] = \mathbb{E}[v_d(\mathbf{X} \cap W_1)],$$

then the estimator $\hat{r}_{min,N}$ *is strongly consistent as* $N \to \infty$.

Theorem 4.3 *If there exists at most one* $r_0 \geq 0$ *such that*

$$2\mathbb{E}[v_d(O_{b(o,r_0)}(\mathbf{X}) \cap W_1)] = \mathbb{E}[v_d(\mathbf{X} \cap W_1)],$$

then the estimator $\hat{r}_{max,N}$ *is strongly consistent as* $N \to \infty$.

Similarly as in case of the tortuosity τ, using the estimator $\hat{r}_{min,N}$ also requires information which may not be available from a bounded sampling window, therefore an estimator $\hat{r}^{\alpha}_{min,N}$ of r_{min} based on a bounded window is defined and the results about its consistency are presented in [47].

Further in [47], edge effects of the estimators of τ and r_{min} are studied for a model for multi-phase material that is incorporated in solid oxid fuel cells. This model is presented in [46], which is discussed in the next section.

Finally in [47], we show a simulation study of dependence of the estimates on N with focus on stabilisation of the estimates for the model introduced in the next section. As a part of this simulation study, we also introduce estimates from discretised data, i.e. from data in the form of voxels, which is common in practice.

4.2 Stochastic modelling of three-phase microstructures with completely connected phases

In this section, a stochastic 3D model for description of a three-phase materials is presented. The model is introduced in [46]. It is a parametric model of three phases in a material, where each phase forms a connected component. Such a material occurs e.g. in anodes of solid oxide fuel cells (SOFC). It consists of pores, nickel (Ni) and yttria-stabilized zirconia (YSZ). Although several models of SOFC electrode have been presented during the last decade, see [6], [8], [15], [28], [37] or [50], none of them can form connected phases, which can be observed in data, see e.g. [23]. The data from [23] has become our main motivation and also the experimental data for the study in [46], where three random connected graphs are used as a back-bone of three phases, which guarantees the connectivity of these phases.

The model is constructed in the following way. Let X_1, X_2 and X_3 be independent homogeneous Poisson point processes with intensities $\lambda_1, \lambda_2, \lambda_3 > 0$. Further $a_1, a_2, a_3 \geq 1$ and $G_i = G_{a_i}(X_i), i = 1, 2, 3$, are the beta-skeletons with the parameters a_i and set of vertices X_i. Let $\mathbf{X}_i$ be random sets defined as

$$\mathbf{X}_i = \{x \in \mathbb{R}^d : d(x, G_i) \leq \min_{j=1,2,3} d(x, G_j)\}, \tag{4.2.1}$$

where $d(x, G) = \min_{e \in E} \min_{y \in e} |x - y|$ is a minimal Euclidean distance from the point x to the graph $G = (X, E)$. The $i-$th phase is defined by the corresponding random set $\mathbf{X}_i$.

Proposition 4.4 *Let $d = 3$, $1 \leq a \leq 2$, X be a homogeneous Poisson point process with intensity λ and $G_a(X)$ the beta-skeleton on X with the parameter a. Denote $e_{\lambda,a}$ the*

expected total edge length of $G_a(X)$ *in* $[0,1]^3$. *Then*

$$e_{\lambda,a} = 8\Gamma\left(\frac{4}{3}\right)\sqrt[3]{\frac{12\lambda^2}{\pi(3a-1)^4}}, \tag{4.2.2}$$

where Γ *denotes the gamma function.*

Note that in [46], the proposition is formulated and proved for a general dimension d.

It is shown by a simulation study that there is a strong correlation between the total edge length of each graph and the empirical volume fraction of the corresponding phase. Namely, denote $V_i = \frac{e_{\lambda_i,a_i}}{e_{\lambda_1,a_1}+e_{\lambda_2,a_2}+e_{\lambda_3,a_3}}$, then the relationship is described by the linear regression model

$$\hat{p}_i = 0.9132V_i + 0.0292 + \varepsilon_i, \tag{4.2.3}$$

where $\varepsilon_i \sim N(0, 0.013)$. The relationship allows us to reduce the number of free parameters $\lambda_1, \lambda_2, \lambda_3, a_1, a_2, a_3$ of the model when the volume fractions p_1, p_2 and p_3 are given. Using the equations (4.2.2) and (4.2.3), for given volume fractions and parameters λ_1, a_1, a_2, a_3, we get the approximation

$$\lambda_i \approx \lambda_1\left(\frac{p_i}{p_1}\right)^{3/2}\left(\frac{3b_i-1}{3b_1-1}\right)^2, \quad i = 2, 3.$$

The simulation study also reveals the relationship between the volume fraction and the constrictivity, and between the volume fraction and the specific area of interfaces, where the specific area of the interfaces between the phases $\mathbf{X}_i$ and $\mathbf{X}_j$ is defined by $I_{ij} = \mathbb{E}\mathcal{H}_2(\mathbf{X}_i \cap \mathbf{X}_j \cap [0,1]^d)$. The relationships are described by

$$\frac{I_{ij}+I_{ik}}{I_{ij}+I_{jk}} \approx \frac{p_i(p_j+p_k)}{p_j(p_i+p_k)},$$
$$\hat{\beta}_i = 0.35\log\hat{p}_i + 0.8 + \varepsilon_i,$$

where $\hat{\beta}_i$ as an estimator of constrictivity of the phase $\mathbf{X}_i$, see Section 4.1, and $\varepsilon_i \sim N(0, 0.028)$.

The strong correlation between the volume fraction and the constrictivity shows us a lack of flexibility of the model. To increase the flexibility of the model, we replace the distance $d(x, G)$ from (4.2.1) by parametric distance measure $d'_\gamma(x, G) = \min\{\gamma d(x, G), d(x, X)\}$, where $\gamma \geq 1$ and $d(x, X) = \min_{y\in X}|x - y|$. This generalisation give us a more flexible model with respect to the constrictivity.

For an application of the model to anode of SOFC, the specific length of the triple phase boundary (TPB) defined by

$$\delta = \mathbb{E}\mathcal{H}_1(\mathbf{X}_1 \cap \mathbf{X}_2 \cap \mathbf{X}_3 \cap [0,1]^3)$$

is one of crucial characteristics, because in TPB the chemical reaction resulting the free electrons take place. The importance of this characteristic lead us to the last generalisation of the model, where the Gaussian smoothing is used. The possibility of changing the parameter during Gaussian smoothing gives us the required flexibility of the model with respect to the characteristic δ. On the other hand, after Gaussian smoothing, the connectivity of the components of $\mathbf{X}_i$ is not guaranteed.

Finally in [46], the generalised model is fitted to data obtained by FIB-tomography by the Nelden-Mead method [45] with the cost function

$$3\sum_{i=2}^{3}\frac{|\hat{p}_{i,sim}-p_i|}{p_i}+2\sum_{i=2}^{3}\frac{|\hat{\beta}_{i,sim}-\beta_i|}{\beta_i}+\sum_{i=2}^{3}\frac{|\hat{\tau}_{i,sim}-\tau_i|}{\tau_i}+\frac{|\hat{\delta}_{sim}-\delta|}{\delta},$$

where p_i, β_i and τ_i denote the volume fraction, the constrictivity and the tortuosity, respectively, for the $i-$th phase, δ is the length of TPB obtained from image data and $\hat{p}_{i,sim}, \hat{\beta}_{i,sim}, \hat{\tau}_{i,sim}$ and $\hat{\delta}_{sim}$ denote the estimators of the same characteristics obtained from model realisations. Note that $\mathbf{X}_2$ and $\mathbf{X}_3$ denote the YSZ- and the Ni-phase, respectively. The study in [46] is equipped by a verification of the model.

4.3 Reconstruction of grains in polycrystalline materials

Three-dimesional X-ray diffraction (3DXRD) microscopy [51] is one of the methods which allow us to investigate internal structure of polycrystalline materials in three dimensions. The method has been developed during the last two decades, see [1], [26] or [34]. It allows us to measure the centre of mass, volumes, crystal lattice orientations and internal stress of a large number (>10 000) of grains of a polycrystalline material. Unfortunately, this method is not able to determine grain boundary, hence the information about individual grain shapes in not available. However, knowledge of data boundaries can be key in many applications, e.g. in analysis of mechanic behaviour of polycrystalline microstructure by the mean of the finite-element (FE) method.

This section concerns a method of reconstruction grain boundaries based on incomplete 3DXRD data. The method is introduced and studied in details in [48] summarised in this section. The basic tool for the reconstruction is the Laguerre tessellation, see Section 2.1.1. Alternative ways of reconstruction of 3DXRD data based on Laguerre tessellation can be found e.g. in [35] and [52].

Consider a 3DXRD dataset $D = \{s_i, v_i\}_{i \in I}$, where $s_i = (s_{i,1}, s_{i,2}, s_{i,3}) \in \mathbb{R}^3$ is the location of the centre of mass and $v_i > 0$ is the volume of the i-th grain, I denotes an index set and N is the number of measured grains. To reconstruct the grain structure of the data by the Laguerre tessellation, we have to find a set of weighted generating points $G_{\mathcal{T}} = \{x_i = (x_{i,1}, x_{i,2}, x_{i,3}), r_i\}$ of the Laguerre tessellation $\mathcal{T}$ such that the $i-$th cell of $\mathcal{T}$ approximates the grain with the centre of mass s_i and the volume v_i. More precisely, we want to find the center of mass $\hat{s}_i$ and the volume $\hat{v}_i$ of the i-th cell $\mathcal{C}_i$ of $\mathcal{T}$ so that $(\hat{s}_i, \hat{v}_i)$ is close enough to (s_i, v_i).

To obtain the set of generating points $G_{\mathcal{T}}$, we use the CE method described in Section 2.4. Here $z = (x_{1,1}, x_{1,2}, x_{1,3}, r_1, ..., x_{N,1}, x_{N,2}, x_{N,3}, r_N)$ and the cost function

$$\begin{aligned} c(z) &= c(x_{1,1}, x_{1,2}, x_{1,3}, r_1, ..., x_{N,1}, x_{N,2}, x_{N,3}, r_N) \\ &= \frac{1}{N} \sum_{i=1}^{N} \left(\frac{3}{4} \min \left(\frac{|s_i - \hat{s}_i|}{r_i'}, 1 \right) + \frac{1}{4} \min \left(\frac{|v_i - \hat{v}_i|}{v_i}, 1 \right) \right), \end{aligned}$$

where $r_i' = \sqrt[3]{3v_i/4\pi}$ is the radius of a ball with volume v_i. As an initial tessellation (set of generating points) used by CE algorithm, we choose the Laguerre tessellation obtained by the procedure from [35], which is a heuristic two-step method providing quite a short running time of the procedure.

The conclusion in [48] is that the presented algorithm is highly parallel, so it can be used to a large dataset (>10 000 grains). Moreover, it is shown in a simulation study that it fits the simulated data well. The simulation study also shows good stability of the method. Finally, the reconstruction method is applied to two experimental datasets, namely to an aluminium-copper (AlCu) sample with known grain boundaries and to a nickle-titanium (NiTi) sample without the information about grain boundaries. We observe a good fit of the AlCu sample, while we get worse result for the NiTi sample, especially when looking at errors in the volume. Comparing our approach to other known methods, we get the following conclusions. The two-step method proposed in [35] is very fast, but our method gives a better fit of the data. The method from [52] fits the centers of mass better, but the relative errors of the volumes are again smaller in our method. Moreover, our procedure has a big advantage in computer solution. It can be well divided into multiple computational cores, which significantly reduces the time-consumption.

4.4 Analysis of polycrystalline microstructure

The three-dimensional electron backscattered diffraction method (3D-EBSD) is one of scanning techniques which allow us to measure crystal orientation of a crystalline or polycrystalline microstructure. The method provides data in the form of 2D slices. The slices can be subsequently joined into 3D image, hence by application of segmentation

methods to the 3D image, we get the information about the data grain structure. Tools for the processing are implemented e.g. in the software DREAM.3D [21]. The processed data can be further analysed by using methods of descriptive spatial statistic.

The papers [32] and [33], summarised in this section, show statistical analyses of samples of pollycrystalline material measured by 3D-EBSD and processed in the way described above.

4.4.1 AlMgSc alloy

The material investigated in [33] is the aluminium alloy, denoted here as Al-3Mg-0.2Sc. The 3D-EBSD data are processed in DREAM.3D, where the grains are detected by using the 'Segment Features (Misorientations)' filter with a misorientation threshold of 2°. After removing noise and very small grains, we get a specimen of 74 579 grains in the material.

The statistical analysis is divided into two parts. In the first part, the individual grain characteristics are statistically analysed. Namely, we analyse the equivalent diameter (i.e. diameter of a ball with the volume equal to the volume of the grain), the number of neighbours, the sphericity, the volume and the surface area. In the second part, the characteristics of individual grain faces are analysed. One of the analysed characteristic is the volume neighbour ratio (VNR), which is defined by

$$VNR = \sqrt{\frac{\max\{v_3(C_1), v_3(C_2)\}}{\min\{v_3(C_1), v_3(C_2)\}} - 1},$$

where C_1 and C_2 denote two neighbouring grains. For each face, its location is represented by the point

$$[x_1 + \frac{ed_1(x_2 - x_1)}{ed_1 + ed_2}, y_1 + \frac{ed_1(y_2 - y_1)}{ed_1 + ed_2}, z_1 + \frac{ed_1(z_2 - z_1)}{ed_1 + ed_2}],$$

where $[x_i, y_i, z_i]$ is the centroid of the grain C_i and ed_i is the volume-equivalent diameter of the i-th grain, $i = 1, 2$. The derived point pattern is then split into three classes, namely the class of low angle grain boundary (LAGB) including the points with the misorientation $10° \pm 5°$, high angle grain boundary (HAGB) consisting of the points with the misorientation $55° \pm 5°$) and the remaining points. The LAGB and HABG samples are further statistically analysed separately. For both samples, the empty space function, the nearest neighbor distribution function and the pair correlation function, see [7], are estimated and compared to the corresponding functions for the homogeneous Poisson process.

4.4.2 Ultrafine-grained copper

A statistical analysis of microstructure of ultrafine-grained copper specimen is presented in [32]. The analysed material was processed by 10 resolutions of high-pressure torsion and subsequently annealed at room temperature for 6 years. It is a similar processing way as that one applied to the specimen from the previous section. The processed specimen consists of 2265 grains and 11513 faces. The characteristic of the grains as well as the characteristics of the grain faces are analysed in a similar way as done in [33]. The spatial distribution of the special boundaries $\Sigma 3$ and $\Sigma 9$, see [13], are again investigated using the empty space function, the nearest neighbor distribution function and the pair correlation function, analogously as above.

4.5 Comparison of segmentation of 2D and 3D EBSD measurements in polycrystalline materials

The electron backscattered diffraction method (EBSD), already discussed in the previous section, is a powerful tool for studying microstructure of crystalline or polycrystalline materials. The EBSD method provides datasets in the form of a set of points with their positions and crystalline orientations (called measured points in the sequel). The crystalline orientation is described by Euler angles, see [13]. The reconstruction of the grain structure of the dataset is based on the following assumptions. Grain interiors have the same or very similar properties, namely the crystalline orientation, i.e. we construct grains by collecting the measured points with similar orientation. In other words, the EBSD method considers a grain to be the set of all points surrounded by a boundary of a given misorientation threshold (MT). It is obvious that the choice of the misorientation threshold may significantly affect the reconstructed grain structure and therefore, the whole analysis of the microstructure as well. Note that the choice of MT in 3D-EBSD usually comes from the experience with 2D-EBSD, because the 2D-EBSD method has been used since nineties, but the 3D-EBSD method is still quite new.

In this section, we summarise the study introduced in [59], which compares the results of segmentation procedure in 2D and 3D with the same choice of MT. The discrepancy between segmentation in 2D and 3D has been discussed theoretically and the problem is demonstrated by a simulation study, where a generalised Laguerre tessellation modelling of polycrystalline microstructure measured by EBSD, see [56], is used. Finally, the discrepancy between segmentation in 2D and 3D is shown for two real datasets. For this purpose, we use the real data discussed in the Section 4.4.

In [59], it is shown that the segmentation in 2D is more sensitive to a noise, especially when the low MT is chosen. The segmentation in 3D has stronger tendency to join

more real grains into one grain than the segmentation in 2D. It means that after 3D segmentation, we may observe large grains, which are in fact created by joining of more real grains, but in 2D case, we still observe these grains separately. It is quite a natural result, but it is often disregarded. This effect has significant impact when at least 10% - 15% of the boundaries of the real grains have the misorientation smaller than the chosen MT. It is also shown that the thickness of the specimen affect the discrepancy between segmentation in 2D and 3D. On the other hand, the effect of the morphology of the grains is negligible in this sense.

Bibliography

[1] Abdolvand H., Majkut M., Oddershede J., Schmidt S., Lienert U., Diak B. J., Withers P. J., Daymond M. R. (2015): On the deformation twinning of Mg AZ31B: A three-dimensional synchrotron X-ray diffraction experiment and crystal plasticity finite element model. *International Journal of Plasticity* **70**, 77 – 97.

[2] Altendorf H., Latourte F., Jeulin D., Faessel M., Saintyant L. (2014): 3D reconstruction of a multiscale microstructure by anisotropic tessellation models. *Image Analysis and Stereology* **33**, 121 – 130.

[3] Beneš V., Bodlák K., Møller J., Waagepetersen R. P. (2005): A case study on point process modelling in disease mapping. *Image Analysis and Stereology* **24**, 1 – 10.

[4] Botev Z., Kroese D. P. (2004): Global likelihood optimization via the cross-entropy method, with an application to mixture models. In: *Proceedings of the 2004 Winter Simulation Conference* **1** Washington DC, USA, 529 – 535.

[5] Bullard J. V., Garboczi E. J., Carter W. C., Fuller E. R. Jr. (1995): Numerical methods for computing interfacial mean curvature. *Computational Materials Science* **4**, 103 – 16.

[6] Cai Q., Adjiman C. S., Brandon N. P. (2011): Modelling the 3D microstructure and performance of solid oxide fuel cell electrodes: Computational parameters. *Electrochimica Acta* **56**, 5804 – 5814.

[7] Chiu S. N., Stoyan D., Kendall W. S., Mecke J. (2013): *Stochastic Geometry and Its Applications.* John Wiley & Sons, Chichester.

[8] Choi H. W., Berson A., Pharoah J. G., Beale S. B. (2011): Effective transport properties of the porous electrodes in solid oxide fuel cells. *Proceedings of the Institution of Mechanical Engineers, Part A: J. Power Energy* **225**, 183 – 197.

[9] Cressie N. A. C. (1993): *Statistics for spatial data.* John Wiley & Sons, New York.

[10] Debayle J., Gotovac Đogaš V., Helisová K., Staněk J., Zikmundová M. (2021): Assessing similarity of random sets via skeletons. *Methodology and Computing in Applied Probability* **23** , 471 – 490

[11] Daley D. J., Vere-Jones D. (1988): *An Introduction to the Theory of Point Processes.* Springer, New York.

[12] Davison A. C., Hinkley D. V. (1997): Bootstrap methods and their application (Cambridge Series in Statistical and Probabilistic Mathematics, Volume 1). Cambridge University Press.

[13] Engler O., Randle V. (2010): *Introduction to Texture Analysis: Macrotexture, Microtexture, and Orientation Mapping.* CRC Press, Boca Raton.

[14] Geyer C. J., Møller J. (1994): Simulation procedures and likelihood inference for spatial point processes. *Scandinavian Journal of Statistics* **21**, 359 – 373.

[15] Golbert J., Adjiman C. S., Brandon N. P. (2008): Microstructural Modeling of Solid Oxide Fuel Cell Anodes. *Industrial & Engineering Chemistry Research* **47**, 7693 – 7699.

[16] Gommes C. J., Bons A.-J., Blacher S., Dunsmuir J. H., Tsou A. H. (2009): Practical methods for measuring the tortuosity of porous materials from binary or gray-tone tomographic reconstructions. *AIChE Journal* **55**, 2000 – 2012.

[17] Gotovac V. (2019): Similarity between random sets consisting of many components. *Image Analysis and Stereology* **38**, 185 – 99.

[18] Gotovac Đogaš V., Helisová K. (2021): Testing equality of distributions of random convex compact sets via theory of N-distances. *Methodology and Computing in Applied Probability* **23**, 503 – 526.

[19] Gotovac Đogaš V, Helisová K., Radović B., Staněk J., Zikmundová M., Brejchová K. (2021): Two-step method for assessing similarity of random sets. *Image Analysis and Stereology* **40**, 127 – 140.

[20] Gotovac V., Helisová K., Ugrina I. (2016): Assessing dissimilarity of random sets through convex compact approximations, support functions and envelope tests. *Image Analysis and Stereology* **35**, 181 – 193.

[21] Groeber M. A., Jackson M. A. (2014): DREAM.3D: A digital representation environment for the analysis of microstructure in 3D. *Integrating Materials and Manufacturing Innovation* **3**, 5 – 17.

[22] Helisová K., Staněk J. (2016): Quermass-interaction process with convex compact grains. *Applications of Mathematics* **61**, 463 – 487.

[23] Holzer L., Münch B., Iwanschitz B., Cantoni M., Hocker T., Graule T., Power J. (2011): Quantitative relationships between composition, particle size, triple phase boundary length and surface area in nickel-cermet anodes for solid oxide fuel cells. *Journal of Power Sources* **196**, 7076 – 7089.

[24] Holzer L., Wiedenmann D., Münch B., Kelle L., Presta M., Gasse P., Robertson I., Grobéty B. (2013): The influence of constrictivity on the effective transport properties of porous layers in electrolysis and fuel cells. *Journal of Materials Science* **48**, 2934 – 2952.

[25] Illian J., Penttinen A., Stoyan H., Stoyan D. (2008): *Statistical analysis and modelling of spatial point patterns.* John Wiley & Sons, Chichester.

[26] Johnson G., King A., Honnicke M.G., Marrow J., Ludwig W. (2008): X-ray diffraction contrast tomography: A novel technique for three-dimensional grain mapping of polycrystals. II. The combined case. *J Appl Crystallogr* **41**, 310 – 318.

[27] Kendall W. S., van Lieshout M. N. M., Baddeley A. J. (1999): Quermass-interaction processes: conditions for stability. *Advances in Applied Probability* **31**, 315 – 342.

[28] Kenney B., Valdmanis M., Baker C., Pharoah J. G., Karan K. (2009): Computation of TPB length, surface area and pore size from numerical reconstruction of composite solid oxide fuel cell electrodes. *Journal of Power Sources* **189**, 1051 – 1059.

[29] Kirkpatrick D. R., Radke J. D. (1985): A framework for computational geometry. *Computational Geometry, Machine Intelligence and Pattern Recognition* **2**, 217 – 248.

[30] Klebanov L. (2006): *N-distances and Their Application.* Karolinum, Prague.

[31] Kroese D. P., Porotsky S., Rubinstein R. Y. (2006): The cross-entropy method for continuous multi-extremal optimization. *Methodology and Computing in Applied Probability* **8**, 383 – 407.

[32] Král P., Staněk J., Kunčická L., Seitl F., Petrich L., Schmidt V, Beneš V., Sklenička V. (2019): Microstructure changes in HPT-processed copper occuring at room temperature. *Materials Characterization* **151**, 602 – 611.

[33] Kopeček J., Staněk J., Habr S., Seitl F., Petrich L., Schmidt V., Beneš V. (2020): Analysis of the polycrystalline microstructure of AlMgSc alloy observed by 3D EBSD. *Image Analysis and Stereology* **39**, 1 – 11.

[34] Ludwig W., King A., Reischig P., Herbig M., Lauridsen E. M., Schmidt S., Proudhon H., Forest S., Cloetens P., Rolland du Roscoat S., Buffière J. Y., Marrow T. J., Poulsen H. F. (2009): New opportunities for 3D materials science of polycrystalline materials at the micrometre lengthscale by combined use of X-ray diffraction and X-ray imaging. *Materials Science and Engineering: A* **524**, 69 – 76.

[35] Lyckegaard A., Lauridsen E. M., Ludwig W., Fonda R. W., Poulsen H. F. (2011): On the use of Laguerre tessellations for representations of 3D grain structures. *Advanced Engineering Materials* **13**, 165 – 170.

[36] Matheron G. (1975): *Random Sets and Integral Geometry.* John Wiley & Sons, New-York.

[37] Metcalfe C., Kesler O., Rivard T., Gitzhofer F., Abatzoglou N. (2010): Connected three-phase boundary length evaluation in modeled sintered composite solid oxide fuel cell electrodes. *Journal of the Electrochemical Society* **157**, B1326 – B1335.

[38] Molchanov I. (2005): *Theory of Random Sets.* Springer, New York.

[39] Møller J. (1994): *Lectures on Random Voronoi Tessellations.* Springer, New York.

[40] Møller J., Helisová K. (2008): Power diagrams and interaction processes for unions of discs. *Advances in Applied Probability* **40**, 321 – 347.

[41] Møller J., Helisová K. (2010): Likelihood inference for unions of interacting discs. *Scandinavian Journal of Statistics* **37**, 365 – 381.

[42] Møller J., Waagepetersen R. P. (2004): *Statistical Inference and Simulation for Spatial Point Processes.* Chapman and Hall/CRC, Boca Raton.

[43] Mrkvička T., Mattfeldt T. (2011): Testing histological images of mammary tissues on compatibility with the Boolean model of random sets. *Image Analysis and Stereology* **30**, 11 – 18.

[44] Myllymäki M., Mrkvička T., Grabarnik P., Seijo H., Hahn U. (2017): Global envelope tests for spatial processes. *Journal of the Royal Statistical Society B* **79**, 381 – 404.

[45] Nelder J. A., Mead R. (1965): A simplex method for function minimization. *The Computer Journal* **7**, 308 – 313.

[46] Neumann M., Staněk J., Pecho M. O., Holzer L., Beneš V., Schmidt V. (2016): Stochastic 3D modeling of complex three-phase microstructures in SOFC-electrodes with completely connected phases. *Computational materials science* **118**, 353 – 364.

[47] Neumann M., Hirsch Ch., Staněk J., Beneš V., Schmidt V. (2019): Estimation of geodesic tortuosity and constrictivity in stationary random closed sets. *Scandinavian Journal of Statistics* **46**, 848 – 884.

[48] Petrich L., Staněk J., Wang M., Westhoff D., Heller L., Šittner P., Krill C. E., Beneš V., Schmidt V. (2019): Reconstruction of the grains in polycrystalline materials from incomplete data using Laguerre tessellations. *Microscopy and Microanalysis* **25**, 734 – 752.

[49] Peyrega C., Jeulin D. (2013): Estimation of tortuosity and reconstruction of geodesic paths in 3D. *Image Analysis and Stereology* **32**, 27 – 43.

[50] Pharoah J.G., Handl L., Schmidt V. (2015): The effect of electrode morphology on solid oxide fuel cell performance. *ECS Transactions* **68**, 2037 – 2045.

[51] Poulsen H.F. (2004). *Three-Dimensional X-Ray Diffraction Microscopy.* Springer, Berlin, Heidelberg.

[52] Quey R., Renversade L. (2018): Optimal polyhedral description of 3D polycrystals: Method and application to statistical and synchrotron X-ray diffraction data. *Computer Methods in Applied Mechanics and Engineering* **330**, 308 – 333.

[53] Rencher A. C. (2002): *Methods of Multivariate Analysis.* John Wiley & Sons, New York.

[54] Rubinstein R. Y., Kroese D. P. (2004): *The Cross-Entropy Method.* Springer, New York.

[55] Saxl I, Ponížil P., Löflerová M. (2009): Application of computer simulation in the stereology of materials. *International Journal of Microstructure and Materials Properties* **4**, 231 – 249.

[56] Seitl F., Petrich L., Staněk J., Krill C., Schmidt V., Beneš V. (2020): Exploration of Gibbs-Laguerre tessellations for three-dimensional stochastic modeling. *Methodology and Computing in Applied Probabability*, DOI: 10.1007/s11009-019-09757-x.

[57] Serra J. (1982): *Image Analysis and Mathematical Morphology. Volume 2: Theoretical Advances.* Academic Press.

[58] Soille P. (2003): *Morphological Image Analysis: Principles and Applications.* Springer, New York.

[59] Staněk J., Kopeček J., Král P., Karafiátová I., Seitl F., Beneš V. (2020): Comparison of segmentation of 2D and 3D EBSD measurements in polycrystalline materials. *Kovové materiály - Metallic Materials* **58**, 301 – 319.

[60] Stenzel O., Pecho O. M., Holzer L., Neumann M., Schmidt V. (2016): Predicting effective conductivities based on geometric microstructure characteristics. *AIChE Journal* **62**, 1834 – 1843.