FACULTY
OF MATHEMATICS
AND PHYSICS
Charles University

RIGOROUS THESIS

Filip Seitl

Modeling and statistics of random tessellations with applications to the study of microstructure of polycrystallic materials

Department of Probability and Mathematical Statistics

Supervisor of the doctoral thesis: prof. RNDr. Viktor Beneš, DrSc.
Study programme: Probability and statistics, econometrics and financial mathematics
Study branch: Probability

Prague 2023

I declare that I carried out this thesis independently, and only with the cited sources, literature and other professional sources. It has not been used to obtain another or the same degree.

I understand that my work relates to the rights and obligations under the Act No. 121/2000 Sb., the Copyright Act, as amended, in particular the fact that the Charles University has the right to conclude a license agreement on the use of this work as a school work pursuant to Section 60 subsection 1 of the Copyright Act.

In Prague, 27.10.2023 Filip Seitl

I would like to thank to everybody who believed I can finish this astonishing work. To my colleagues, friends, family, especially to my parents, and finally to my supervisor.

Title: Modeling and statistics of random tessellations with applications to the study of microstructure of polycrystallic materials

Author: Filip Seitl

Department: Department of Probability and Mathematical Statistics

Supervisor: prof. RNDr. Viktor Beneš, DrSc., Department of Probability and Mathematical Statistics

Abstract: We present a parametric statistical methodology for analysing a Laguerre tessellation data set viewed as a realization of a marked point process of generators. We study the dependence of the marks and the point process in detail. Further, we suggest two general models. The first one is based on marked Gibbs point processes and leads to a broad class of Gibbs-Laguerre tessellations. Under mild assumptions we prove the existence of the infinite-volume Gibbs measure. Then the choice of energy function for applications is discussed in detail. The second model is hierarchical, where in the first step the point pattern is modelled and in the second step the marks are modelled conditionally on points, using exponential family models based on geometrical characteristics of the tessellation. Statistical tools for the parameter estimation, model selection and fitting are suggested and implemented. We apply our methodology for a 3D Laguerre tessellation data set representing the microstructure of a polycrystalline metallic material, where simulations under a fitted model may substitute expensive laboratory experiment.

Keywords: Random tessellation models, simulation algorithms, statistical inference, grain microstructure, 3D image analysis

Contents

Introduction **3**

1 Motivation **9**
- 1.1 Laguerre tessellation and periodic condition . . . 9
- 1.2 Data acquisition . . . 9
- 1.3 Introduction of datasets . . . 10
  - 1.3.1 NiTi alloy . . . 10
  - 1.3.2 Aluminium alloy . . . 14

2 Theoretical background **19**
- 2.1 Tessellations . . . 20
  - 2.1.1 Tessellation characteristics . . . 24
  - 2.1.2 Periodic boundary conditions . . . 25
  - 2.1.3 Point positions in Laguerre geometry . . . 25
  - 2.1.4 Hypergraphs . . . 28
- 2.2 Point processes and marked point processes . . . 30
  - 2.2.1 Poisson point process . . . 31
  - 2.2.2 Gibbs point process . . . 32
  - 2.2.3 Marked point processes . . . 36
  - 2.2.4 Summary characteristics . . . 38
  - 2.2.5 Estimation of summary characteristics and boundary effects 40
  - 2.2.6 Connection to tessellations . . . 42
- 2.3 Simulation algorithms . . . 42
  - 2.3.1 Metropolis-Hastings birth-death-move algorithm . . . 43
  - 2.3.2 Metropolis within Gibbs algorithm . . . 44
- 2.4 Statistical methods . . . 45
  - 2.4.1 Maximum pseudolikelihood . . . 45
  - 2.4.2 Global envelopes . . . 48
  - 2.4.3 Permutation tests . . . 48

3 Gibbs-Laguerre tessellations **51**
- 3.1 Energy function . . . 51
  - 3.1.1 Periodic energy function . . . 51
  - 3.1.2 Examples of potential functions . . . 52
- 3.2 Existence of Gibbs-Laguerre tessellations . . . 54
  - 3.2.1 Energy over hypergraph structure . . . 55
  - 3.2.2 Assumptions and existence theorem . . . 57
  - 3.2.3 Auxiliary lemmas . . . 59
  - 3.2.4 Existence of Gibbs-Laguerre models . . . 59
- 3.3 Parameter estimation . . . 62
  - 3.3.1 Pseudolikelihood for Gibbs-Laguerre tessellations . . . 63
  - 3.3.2 Estimation of hardcore parameters . . . 64
- 3.4 Gibbs-Laguerre tessellations . . . 64
  - 3.4.1 Modelling and reconstruction . . . 64
  - 3.4.2 Moment reconstruction . . . 66

3.4.3 Histogram reconstruction 69
3.4.4 Summary of numerical results 73
3.5 Statistical reconstruction 75
3.6 Comparison of reconstruction approaches 76
3.7 Non-data based Gibbs-Laguerre tessellation models 78
3.7.1 Number of faces per cell 78
3.7.2 Neighbour-volume ratio 82

4 Hierarchical model 85
4.1 Independent modelling 85
4.1.1 Summary point process statistics 85
4.1.2 Pair correlation function analysis 87
4.1.3 Mark correlation 88
4.1.4 Independent marking 90
4.2 Hierarchical approach 92
4.2.1 Point process models for the points 93
4.2.2 Exponential tessellation models for the radii given the points 94
4.2.3 Joint model 95
4.3 Parameter estimation 95
4.3.1 Profile pseudolikelihood for the points 95
4.3.2 Pseudolikelihood for the radii given the points 96
4.3.3 Properties of MPLE 96
4.4 Hierarchical modelling 97
4.4.1 The fitted model for the points 97
4.4.2 The fitted model for the radii conditioned on the points . . 99

Conclusion 103

Bibliography 105

List of Figures 111

List of Tables 117

List of Abbreviations 119

List of publications 123

A Attachments 125
A.1 Pseudo-periodic configurations 125
A.1.1 Configuration for $\mathbb{R}^3$ tetrahedrization 125
A.2 Sphericity 129
A.3 Comparison of Gibbs and Poisson Laguerre tessellations 132
A.4 Profile maximum pseudolikelihood for multiscale point process . . 136
A.5 Software 138
A.5.1 Overview of the existing software used for computations . 138
A.5.2 Library `LagMod` 138

Introduction

The study of random tessellations presents an important sub-field of stochastic geometry and it finds many applications in natural and technical sciences. This thesis focuses on both probabilistic and statistical modelling of datasets represented by 3D Laguerre tessellations, including such tools as stochastic simulation and data reconstruction. We develop statistical methodology for analyzing the datasets and illustrate our approaches on two particular examples of Laguerre tessellations representing microstructure of polycrystalline materials.

Some background and motivation

Laguerre tessellations generalise the well-known Voronoi tessellations. While the latter are generated by point patterns, Laguerre tessellations are generated by marked point patterns; and besides spatial locations so-called radius marks also play a role. They control the size and shape of the tessellation cells. In brief, a 3D Laguerre tessellation (also called a power diagram or a generalised Voronoi tessellation) is a flexible way of modelling a subdivision of 3D space In practice, the tessellation is described by a data set $(\mathbf{x}_n, \mathbf{r}_n)$ where $\mathbf{x}_n = \{x_1, \dots, x_n\} \subset \mathbb{R}^3$ is a finite point pattern and $\mathbf{r}_n = (r_1, \dots, r_n) \in (0, \infty)^n$ is an associated vector of positive numbers called marks, and we identify $(\mathbf{x}_n, \mathbf{r}_n)$ with the marked point pattern $\{(x_1, r_1), \dots, (x_n, r_n)\}$ (see Sections 2.1 and 2.2.5 for the details). Viewing the marked point pattern as a realization of a marked point process, statistical analysis of random Laguerre tessellations can be reduced to statistical analysis of marked point processes, and also modelling is simplified to modelling marked point processes.

In materials science, discovering and quantifying relationships between microstructure and bulk properties of materials is one of the most important research goals, see Grimvall [1999]. The traditional approach is to analyze samples of real materials. Whereas this method arguably returns the most realistic results, it is time consuming and demanding to produce, image and investigate the specimens. With the increase in readily available computing power, it is possible to support such investigations today with in silico experiments, which drastically reduce the time spent in the laboratory, see, e.g., Redenbach et al. [2012], Stenzel et al. [2016]. An effective approach for this is building parametric stochastic models of the microstructure that provide realistic virtual samples whose physical properties can be computed numerically. Based on these results, it is then possible to study relationships between geometrical characteristics and descriptors of macroscopic physical properties, and, as a consequence, reduce the laboratory experiments needed to validate these relationships. On the other hand, when experimental datasets are available, one can generate further samples using the ideas of statistical reconstruction, see Illian et al. [2008].

In particular, spatial (three-dimensional) Laguerre tessellations, cf. Chiu et al. [2013], have become a valuable tool in the statistical analysis of polycrystalline microstructures of present materials. Such analyses are an important modern research topic in materials sciences, see Petrich et al. [2019], where the polycrystalline aspect is challenging. The use of Laguerre tessellations in this context is natural, since the members of a large class of spatial tessellations in the three-dimensional space can be considered as Laguerre tessellations, see Lautensack and Zuyev [2008].

Data sets of the form $(\mathbf{x}_n, \mathbf{r}_n)$ appear in various contexts of material science. For example, in 3D X-ray diffraction microscopy of polycrystalline materials the volumes and centroids of grains are obtained, and representation of such measurements by a 3D Laguerre tessellation can be produced by various mathematical optimization methods, cf. Lyckegaard et al. [2011], Spettl et al. [2016], Quey and Renversade [2018], Kuhn et al. [2020]. Development of stochastic models for the tessellation follows, however, only a few papers including Spettl et al. [2015] have been dealing with statistical methodology in the context of polycrystalline materials.

Inference for statistical models of Laguerre tessellation data sets is a difficult task unless $(\mathbf{x}_n, \mathbf{r}_n)$ follows a simple model such as a (marked) Poisson process. In this context the question arises whether the radius marks are independent or spatially correlated. In the former case modelling would be greatly simplified. Indeed, analytical calculations for Laguerre tessellations are rather complicated and even the pioneers in the field, Lautensack and Zuyev [2008], limited themselves to the case of independent marks. Therefore little is known about the influence of radius mark correlations in Laguerre tessellations. Often the points in $\mathbf{x}_n$ exhibit regularity and different marked point process models for $(\mathbf{x}_n, \mathbf{r}_n)$ have been suggested: simple random models for packings of hard balls, cf. Chiu et al. [2013], obtained by iterative procedures such as random sequential adsorption in Lautensack [2008] (with applications to foam structures) or variants of collective-rearrangement algorithms in Spettl et al. [2015] (with application to a polycrystalline microstructure).

Our contribution

On examples of two polycrystalline microstructures represented by Laguerre tessellations we develop various stochastic models and in detail discuss their properties. The datasets consist of marked point patterns, where points and marks are obviously not independent, see Section 4.1 and Stoyan et al. [2021]. Therefore we need to consider more complex models.

Inspired by the work of Dereudre and Lavancier [2011] on Gibbsian models for 2D Voronoi tessellations, we introduce in Seitl et al. [2021] Gibbsian models for random 3D Laguerre tessellations. We prove the existence of particular infinite-volume Gibbs-Laguerre tessellations in Jahn and Seitl [2020]. In the Gibbs setting we can built a variety of models with prescribed geometrical properties. In particular, we suggest two ways of so called ‘reconstruction’ in order to produce samples statistically similar to the data, see Section 3.4. We use the notion ‘reconstruction’ since the approach is up to some extent inspired by statistical reconstruction introduced by Tscheschel and Stoyan [2006] in order to overcome the problem of parameter estimation. These Gibbs models, which are described in Chapter 3 of this thesis, were first thought to be tempting to use because they may incorporate various properties of the Laguerre tessellation and an interaction between its cells. See Section 3.7 for extensive simulation studies and Attachment A.3 where we compare Gibbs-Laguerre to Poisson-Laguerre tessellations in order to show that much more variability in terms of cell shapes can be obtained using Gibbs models. However, the models are complicated to use for statistical inference and they are time-consuming to simulate.

In order to come up with an simpler model, we explore in detail spatial-mark dependencies within the marked point patterns of Laguerre generators. In Stoyan et al. [2021] we show that radius mark correlations play an important role in the context of random Laguerre tessellations; dispelling any hope that these correlations can be ignored (and hence in practice a marked Poisson process is rarely a reasonable assumption). We demonstrate this on both an empirical data set, see Section 4.1 and our own simulations based on marked Gibbs point processes, see Section 3.7.1. En passant, we show that the choice of bandwidths for pair correlation function estimation in the three-dimensional space is more complicated than described in the literature, for details see Section 4.1.2.

Finally, in Seitl et al. [2022] we introduce hierarchical statistical models consisting of first a parametric Gibbs point process model for $\mathbf{x}_n$ and second a parametric model for $\mathbf{r}_n$ conditioned on $\mathbf{x}_n$, see Section 4.2. Specifically, we use in the first case a nested sequence of flexible pairwise interaction points processes called multiscale processes Penttinen [1984] and in the second case various exponential family models where the canonical sufficient statistic is based on tessellation characteristics such as surface area or volume of cells or absolute difference in volumes of neighbouring cells. Apart from reducing the dimension from 4 (when viewing $(\mathbf{x}_n, \mathbf{r}_n)$ as a 4-dimensional point pattern) to 3 (when considering $\mathbf{x}_n$), an advantage is that we specify two much simpler cases of models with parameters which do not depend on each other. Hence we can separate between how to simulate and estimate unknown parameters for $\mathbf{x}_n$ and $\mathbf{r}_n \mid \mathbf{x}_n$, respectively. The parameters are simply estimated by maximum pseudolikelihood methods and well-known MCMC algorithms are used for simulations. Thereby estimation and simulation become much faster than in Seitl et al. [2021] when fitting specific models given in Section 4.2 to the data from Section 1.3.1. A further advantage is that the model construction makes it possible to develop a rather straightforward model selection procedure, see Section 4.4: For $\mathbf{x}_n$, the procedure starts with the simplest case of a Poisson process and continues with constructing more and more complex multiscale processes until a satisfactory fit is obtained when considering global envelopes and tests from Myllymäki et al. [2017] based on various functional summary statistics. For $\mathbf{r}_n$ conditioned on $\mathbf{x}_n$, more and more complex exponential models are developed, where we demonstrate how to compare fitted models of the same parameter dimension by considering maximized log pseudolikelihood functions. Further, we evaluate selected fitted models by comparing moment properties of tessellation characteristics under simulations from the model with empirical moments, by considering plots of global envelopes, and by evaluating values of global envelope tests. This comparison is not only done by looking at those tessellation characteristics used for specifying the canonical sufficient statistic of the exponential model but also for various other tessellation characteristics. To the best of our knowledge, this is the first time that such a model selection procedure has been used when analysing polycrystalline materials, cf. Šedivý et al. [2018] and the references therein.

Besides the already existing software we implemented a code for pseudolikelihood estimation and the MCMC algorithms used for simulation of Gibbs-

Laguerre models and hierarchical models. This code is available at `https://github.com/VigoFierry/Lag_mod`.

Outline

The main body of the thesis is organized in four chapters.

Chapter 1 in brief explains what is Laguerre tessellation and introduces two datasets represented by Laguerre tessellation, which motivate a development of stochastic models – the main goal of our work. It describes the process of data acquisition in Section 1.2 and provides a detail statistical description of the two datasets in Section 1.3.

Chapter 2 contains a theoretical background regarding: tessellations in Section 2.1 – a general theory, Voronoi and Laguerrre models, Section 2.1.1 introduces important geometric characteristics of tessellations, Section 2.1.2 presents concept of periodic boundary, Section 2.1.3 deals with mutual positions of Laguerre generators, Section 2.1.4 introduces a theoretical concept of hypergraphs and hyperedge potentials; and (marked) point processes in Section 2.2 – a general theory, Poisson point process is defined in Section 2.2.1, Gibbs point processes are introduced in Section 2.2.2, Section 2.2.3 describes marked point processes in detail, Section 2.2.4 states some important summary characteristics of (marked) point processes, Section 2.2.5 deals with their estimation, Section 2.2.6 connects (marked) point patterns with tessellation models. Simulation algorithms used in this thesis are stated in Section 2.3. Section 2.4 describes three statistical concepts, namely maximum pseudolikelihood estimation method is described in Section 2.4.1, global envelopes are mentioned in Section 2.4.2 and permutation test in Section 2.4.3.

Chapter 3 is named 'Gibbs-Laguerre tessellations' and focuses on simultaneous Gibbs model for generators of Laguerre tessellation $(\mathbf{x}_n, \mathbf{r}_n)$. Section 3.1 is devoted to the energy function of marked Gibbs point process – it shows a periodic adjustment in Section 3.1.1 and several examples in Section 3.1.2. Section 3.2 deals with the existence of infinite-volume Gibbs models – the energy function is rewritten in terms of hypergraphs in Section 3.2.1, general existence theorem is formulated in Section 3.2.2 (the existence result relies on three key assumptions **(R)**, **(S)** and **(U)** where details concerning the third assumption **(U)** are available in Attachment A.1), two auxiliary lemmas are stated in Section 3.2.3, Section 3.2.4 proves existence of two particular Gibbs-Laguerre tessellation models. Section 3.3 shows how parameters of Gibbs-Laguerre can be estimated in two-step procedure – estimates of possible hard-core parameters defined in Section 3.3.2 are plugged into equations obtained from maximizing the pseudolikelihood function in Section 3.3.1. Section 3.4 presents simulations of Gibbs-Laguerre tessellation whose goal is to resemble the data – a general modelling approach is sketched in Section 3.4.1, two different methods are presented in Sections 3.4.2 and 3.4.3. Basic statistical reconstruction described in Section 3.5 is compared to our Gibbs-Laguerre simulations in Section 3.6. Section 3.7 presents two more non-data based simulation studies (they are not meant to produce samples resembling the data).

Chapter 4 is called 'Hierarchical model' and aims at finding a more simpler model than the Gibbs-Laguerre model from the previous chapter. Instead of simultaneous modelling of $(\mathbf{x}_n, \mathbf{r}_n)$, the idea is split the task to modelling of $\mathbf{x}_n$ and $\mathbf{r}_n$ separately. Obviously, the data are too complex than $\mathbf{x}_n$ and $\mathbf{r}_n$ could be modelled independently, dependencies between points and marks are in detail studied in Section 4.1. Point process summary characteristics are studied in Section 4.1.1; among them a prominent place takes the pair correlation function, whose analysis is presented in Section 4.1.2; correlations of marks and points are studied in Section 4.1.3; a formal test of independence is performed in Section 4.1.4. When the independent approach fails, a hierarchical model from Section 4.2 can be adopted – first, we model the point pattern $\mathbf{x}_n$ in Section 4.2.1, and second, the marks/radii $\mathbf{r}_n$ are modelled conditionally on $\mathbf{x}_n$ in Section 4.2.2. Section 4.3 deals with parameter estimation in the hierarchical model. The simulations are presented in Section 4.4 – a model for $\mathbf{x}_n$ is selected in Section 4.4.1 and a suitable joint model is chosen in Section 4.4.2.

The thesis is accompanied by several attachments, four of them supply an additional information to the text of the thesis. The last fifth Attachment A.5 describes software, both already existing and self-implemented, used for the computations.

1. Motivation

A widely used mathematical model for capturing the microstructure of polycrystalline materials is the Laguerre tessellation. It is the most general model for materials for which the assumption of convexity of their grains (structure elements) is reasonable. This chapter introduces the concept of Laguerre tessellation in Section 1.1, while the proper definition is postponed into Section 2.1. Further, in Section 1.2 we will look at how Laguerre tessellation samples can be obtained from microscopical measurements of microstructures. Finally, in Section 1.3 we introduce two data samples from materials science and focus on their elementary statistical analyses. The aim of the thesis is to develop suitable statistical models for such kind of data.

1.1 Laguerre tessellation and periodic condition

Consider $\mathbf{x}$ to be a finite or locally finite set of points (called a point pattern) in $\mathbb{R}^3$ and $\mathbf{r}$ a set of positive marks (called radii) attached to these points. The points can be numbered, i.e., $(\mathbf{x}, \mathbf{r}) = (x_j, r_j)_{j \in \mathbb{N}}$. Then we can define 3D Laguerre tessellation generated by $(x_j, r_j)_{j \in \mathbb{N}}$ as the collection of nonempty sets (called cells) given by

$$C_i = \{z \in \mathbb{R}^d : ||z - x_i||^2 - r_i^2 \leq ||z - x_j||^2 - r_j^2 \ \forall (x_j, r_j) \in (\mathbf{x}, \mathbf{r}), i \neq j\},$$

$(x_i, r_i) \in (\mathbf{x}, \mathbf{r})$, where $|| \cdot ||$ is the Euclidean norm. We denote $L(\mathbf{x}, \mathbf{r})$ the Laguerre tessellation generated by the marked point pattern $(\mathbf{x}, \mathbf{r})$. For the proper definition and more details see Definition 9.

Note that Laguerre tessellation is defined as a division of the whole Euclidean space. In applications, the set of points $\mathbf{x}$ together with their marks $\mathbf{r}$ is finite and observed only on a finite observation window W. Then such finite marked point pattern $(\mathbf{x}, \mathbf{r})$ can be extended to the whole space $\mathbb{R}^3$ by periodic repetition, for details see Definition 11. Note that the periodic extension may cause skipping some generators out of $(\mathbf{x}, \mathbf{r})$ as their cells may become empty and empty cells are excluded by definition of Laguerre tessellation, for precise description of this phenomenon see the paragraph below Definition 11.

1.2 Data acquisition

In materials science the microstructure (set of grains) of a polycrystalline material is measured by some desctructive or undestructive microscopy technique. In the following, we will meet with synchrotron X-ray tomographic imaging and 3D X-ray diffraction microscopy (3D-XRD). The measurement then needs to be processed in order to obtain the experimental data (input data for our statistical models). Whereas the first microscopy technique gives us a 3D image of the microstructure (this still needs to be postprocessed by segmentation techniques), using 3D-XRD microscopy only the volumes and centroids of grains are obtained.

Having such kind of measurements (3D image or centroids with volumes), some optimization method is used to obtain an approximation of the microstructure by a deterministic Laguerre tessellation, i.e., our experimental data. Frequently used optimization methods are cross-entropy [Spettl et al., 2016], simulated annealing [Šedivý et al., 2016], gradient descent algorithms [Kuhn et al., 2020], etc.

1.3 Introduction of datasets

This section introduces two Laguerre tessellation datasets coming from materials research. The methods of acquisition of such datasets were briefly described in Section 1.2, here we focus on visualization and elementary statistical description of the two Laguerre tessellations. The visual inspection helps us to judge the homogeneity of marked point patterns of generators. The statistical analysis of the geometric tessellation characteristics gives us the first insight into microstructural properties of the materials. Some more advanced statistical properties of the marked point patterns of generators will be investigated (not necessarily the same ones in case of the both datasets) in Chapters 3 and 4 which are devoted to building up statistical models for the Laguerre tessellations.

1.3.1 NiTi alloy

The first Laguerre tessellation data set comes from the study Sedmák et al. [2016] of a polycrystalline microstructure of a nickel titanium alloy. We deal with a finite marked point pattern $(\mathbf{x}, \mathbf{r})$ extracted from a larger data set collected in Petrich et al. [2019] by the so-called cross-entropy method applied to 3D-XRD measurements. Specifically, $\mathbf{x}$ consists of the 2009 points which are contained in a 3D rectangular observation window W of size $40 \times 40 \times 85$ μm^3. We visualize the Laguerre tessellation data set by a series of 2D slices – Figure 1.1 shows four equidistant slices perpendicular to the z axis.

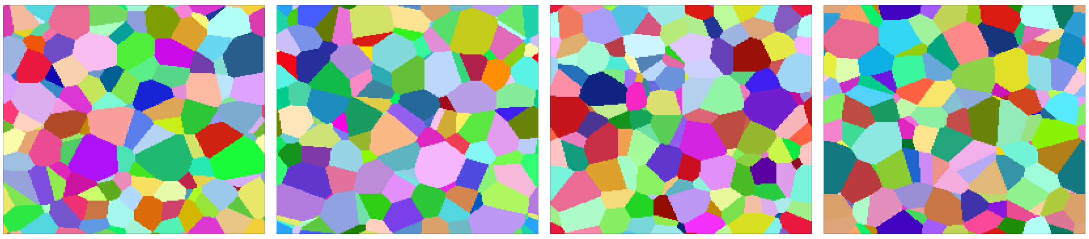


Figure 1.1: Four slices in xy plane of the observed Laguerre tessellation. The cells are coloured randomly.

To account for that W is a cut-off from the entire material specimen (causing undesirable boundary effects) and for computational reasons (i.e., the use of software `Voro++`, Rycroft [2009]), we choose to apply a periodic extension thereby obtaining our Laguerre tessellation data set $(\mathbf{x}_n, \mathbf{r}_n)$, see Definition 11 and notation at the end of Section 2.2.5. Here $n = 1965$, so by applying the periodic extension at most 44 cells are 'lost'. From now, all NiTi tessellation-related calculations are carried out in the periodic setup.

Figure 1.2 depicts $\mathbf{x}_n$ together with W and its projections onto the xy, xz and yz planes, respectively. None of these indicate spatial inhomogeneity of $\mathbf{x}_n$. So

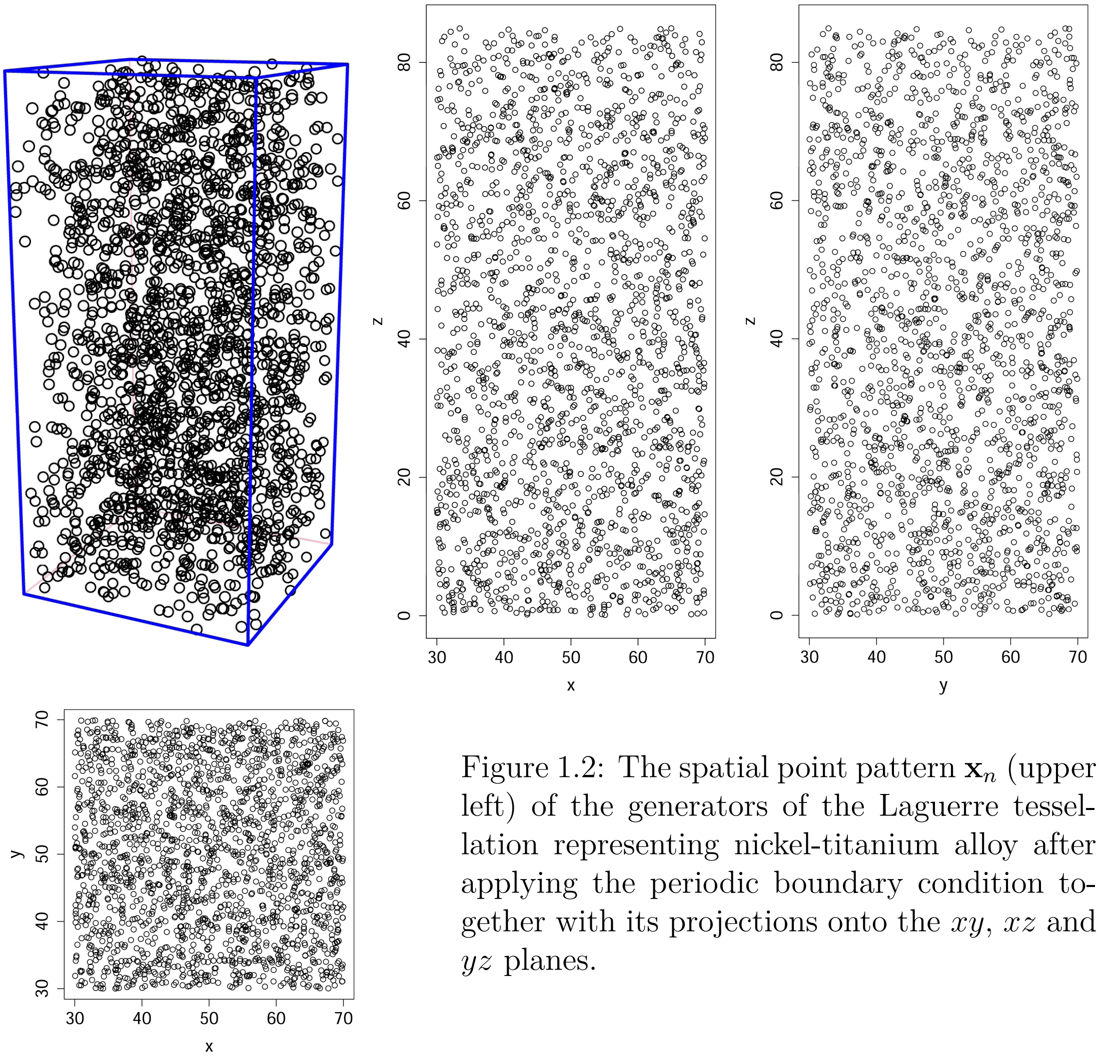
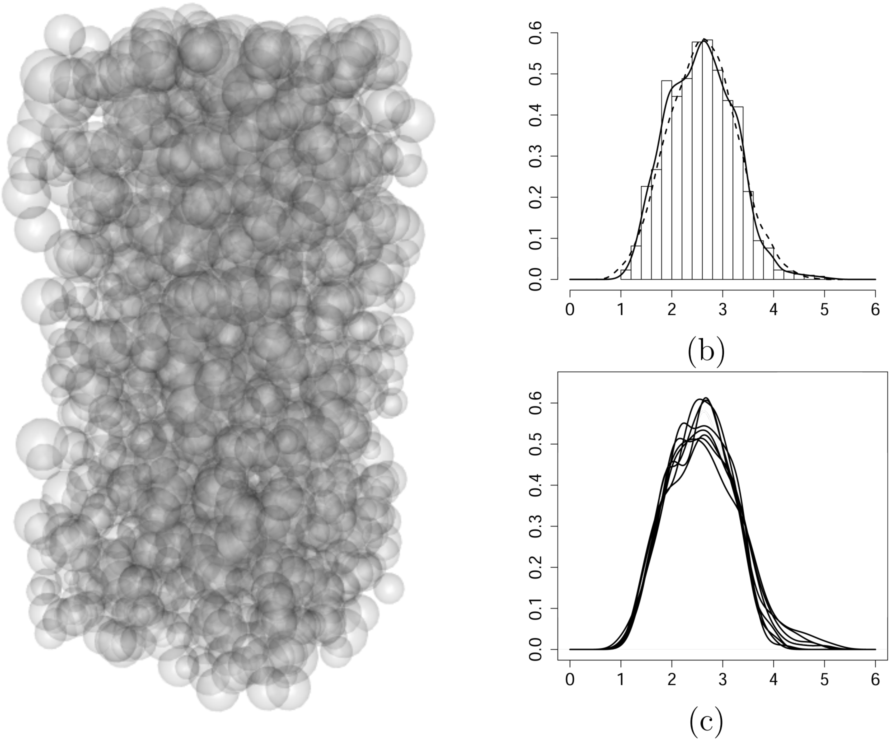


Figure 1.2: The spatial point pattern $\mathbf{x}_n$ (upper left) of the generators of the Laguerre tessellation representing nickel-titanium alloy after applying the periodic boundary condition together with its projections onto the xy, xz and yz planes.

considering the periodic extension $\mathbf{x}_n^*$ (defined as in (2.5) but with $(\mathbf{x}, \mathbf{r})$ replaced by $(\mathbf{x}_n, \mathbf{r}_n)$), we find it reasonable to assume that the distribution underlying $\mathbf{x}_n^*$ is invariant under translations and rotations in 3D space. Equivalently, we assume that the distribution underlying $\mathbf{x}_n$ is invariant under shifts and rotations when W is wrapped on a 3D torus.

The panel on the left in Figure 1.3 shows the balls defined by $(\mathbf{x}_n, \mathbf{r}_n)$. This indicates that the distribution underlying the radii $\mathbf{r}_n$ conditioned on $\mathbf{x}_n$ can be assumed to be homogeneous. As a consequence we have invariance of this conditional distribution when making shifts and rotations of $\mathbf{x}_n$ on the torus. The panel on the right shows: b) histogram of the radii; c) eight kernel density estimates of the radii distributions corresponding to a subdivision of W into the eight subsets obtained by dividing the three sides of the rectangular region W into halves. For the number of points per subset, the mean is 232, the minimum is 219 and the maximum is 270. The similarity of the eight density estimates is in accordance with the assumption that the distribution of $\mathbf{r}_n$ conditioned on $\mathbf{x}_n$ is homogeneous.

Figure 1.4 depicts histograms of some cell characteristics, cf. Section 2.1.1. It shows that many cells are small, the distribution of number of faces is rather symmetric and ranges from 4 (the smallest possible value) to 35, many cells are



(a)

(b)

(c)

Figure 1.3: The radii $\mathbf{r}_n$ of the generators of the Laguerre tessellation representing nickel-titanium alloy after applying the periodic boundary condition: a) the balls defined by $(\mathbf{x}_n, \mathbf{r}_n)$; b) histogram of the radii together with a kernel density estimate (solid line) and a fitted density of a beta distribution (dashed line) obtained by maximum likelihood estimation; c) eight kernel density estimates based on the radii associated to the points in the eight sets for the subdivision of W obtained by dividing its sides into halves.

rather spherical (i.e., with sphericity, see 2.3 for definition, close to 1) and many pairs of neighbouring cells have a range of differences in cell sizes of the same order as the cell size itself. Table 1.1 summarizes the data sample in terms of means and standard deviations of the chosen cell characteristics – besides those already considered in Figure 1.4, i.e., cell volume, surface area, number of faces, sphericity and absolute difference in volume for two neighbouring cells, it adds radius of the generator corresponding to the cell, number of vertices, total edge length and neighbour-volume ratio.

The upper right triangle in Table 1.2 shows the sampling correlation coefficients for the cell characteristics volume, surface area, number of faces, sphericity and total edge length. The lower left triangle in the table shows the empirical correlations for absolute difference of volumes of two neighbouring cells and three other face characteristics: face area, face perimeter and number of edges in a face. We see that all correlations are high except for dvol which is nearly uncorrelated to any other face characteristic. In particular volume, surface area and total edge length are highly correlated, and face area and face perimeter are highly correlated (correlations > 0.9).

Table 1.3 shows the empirical correlations of the cell characteristics with radii.

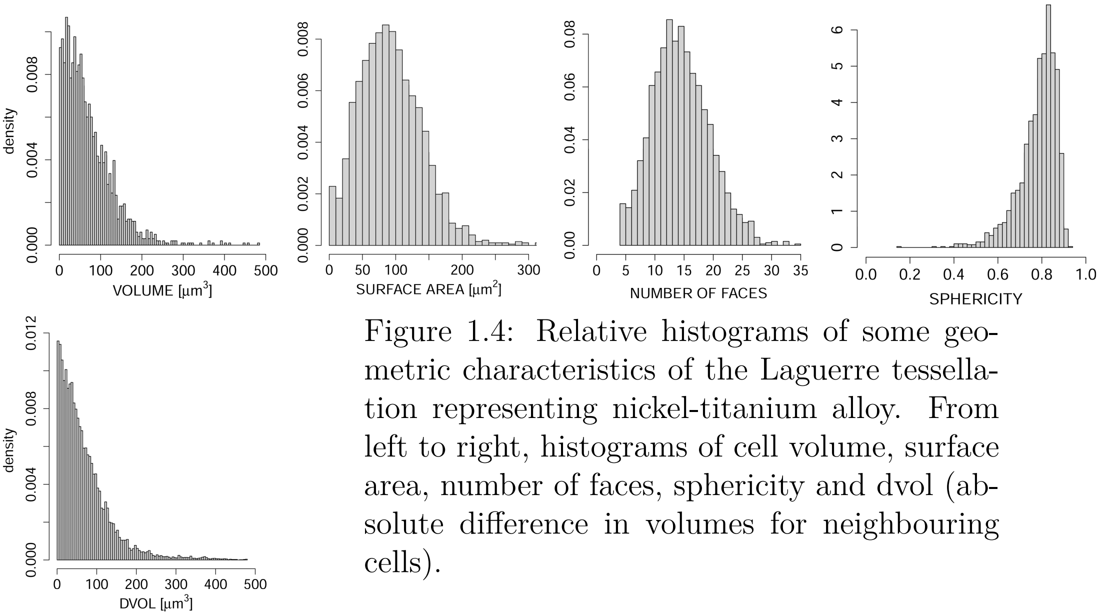


Figure 1.4: Relative histograms of some geometric characteristics of the Laguerre tessellation representing nickel-titanium alloy. From left to right, histograms of cell volume, surface area, number of faces, sphericity and dvol (absolute difference in volumes for neighbouring cells).

Table 1.1: Laguerre tessellation characteristics for the nickel-titanium alloy data. In the upper part of the table there are numbers of cells and faces per sample. In the lower part of the table there are mean values and standard deviations (sd) of cell characteristics evaluated from the data.

per sample:		
number of cells	1965	
number of faces	14721	
per cell:	mean	sd
radius [μm]	2.5719	0.64686
nof	14.9832	4.9193
nov	25.9664	9.8387
vol [μm^3]	69.2112	58.8862
surf [μm^2]	92.4555	47.8593
tel [μm]	67.9204	27.3815
spher	0.78232	0.087038
dvol [μm^3]	68.8902	65.0457
NVR	15.8262	862.9134

Table 1.2: Correlations of tessellation characteristics.

vol	0.971	0.938	0.841	0.680
farea	surf	0.974	0.874	0.737
0.923	fper	tel	0.941	0.793
0.751	0.751	fnoe	nof	0.754
0.074	0.062	0.025	dvol	spher

We see that among the cell characteristics surface area, volume and total edge length have the highest correlations with radii (around 0.8).

In all tables and even sometimes in the text we abbreviate the names of cell and face characteristics, for explanation of the shortcuts see Section 2.1.1 or List of Abbreviations.

Table 1.3: Correlations of tessellation characteristics with radii.

	vol	surf	tel	nof	spher
radius	0.795	0.818	0.768	0.647	0.541

1.3.2 Aluminium alloy

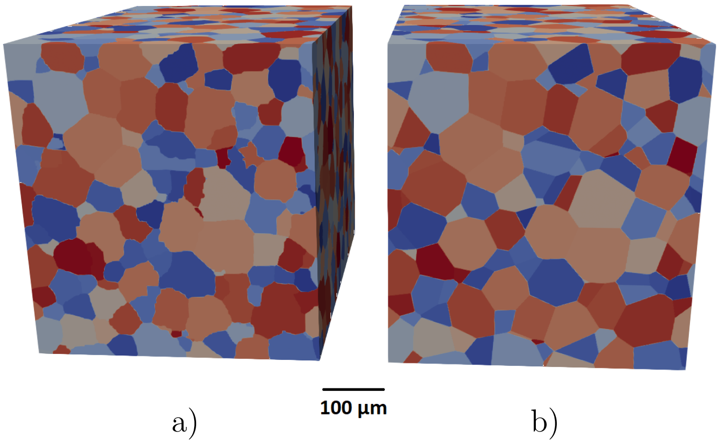
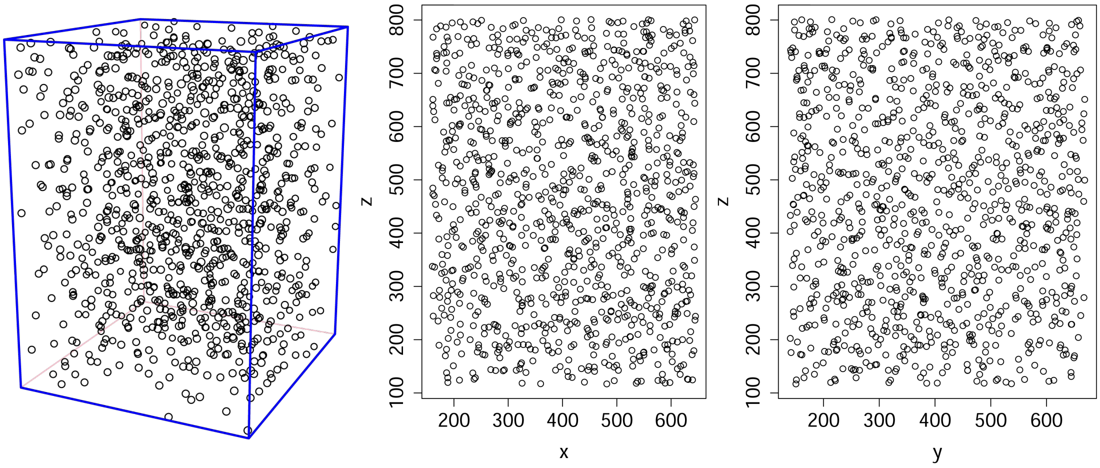


Figure 1.5: Aluminum alloy specimen measured by synchrotron X-ray tomography (cf. Spettl et al. [2016]) : a) original voxelized image; b) approximation by a Laguerre tessellation. The grains are coloured randomly.

The second Laguerre tessellation data sample, which belongs to a sample of polycrystalline material, namely an aluminium alloy, is described in this section. The image data, cf. Figure 1.5a, are obtained by synchrotron X-ray tomographic imaging together with watershed segmentation and present a cutout of the polycrystalline microstructure of an Al-5 wt% Cu sample, which is described in Spettl et al. [2016]. Figure 1.5b shows their approximation by a deterministic Laguerre tessellation obtained by cross-entropy method.

The corresponding marked point pattern of generators $(\mathbf{x}, \mathbf{r})$ contains 1057 marked points in the parallelepipedal window W of size $486 \times 529 \times 685$ μm^3. The generators create 1057 nonempty tessellation cells if we consider an empty outside configuration, i.e., there are no further generators outside the window. However, with periodic boundary conditions (periodic continuation outside the window, Definition 11) these 1057 generators create only 1049 nonempty cells, i.e., we obtain Laguerre tessellation data $(\mathbf{x}_n, \mathbf{r}_n)$ with $n = 1049$, cf. the end of Section 2.2.5. From now, all Al-5 wt% Cu tessellation-related calculations are carried out in the periodic setup. The key assumption of homogeneity of $\mathbf{x}_n$ and $\mathbf{r}_n$ conditioned on $\mathbf{x}_n$, which is necessary to justify the periodic boundary conditions, is investigated in Figures 1.6 and 1.7. Figure 1.6 depicts $\mathbf{x}_n$ together with the observation window and the projections of $\mathbf{x}_n$ onto xy, xz and yz planes. The figure indicates spatial homogeneity of $\mathbf{x}_n$. In Figure 1.7, the left panel shows the balls defined by $(\mathbf{x}_n, \mathbf{r}_n)$ and the right panel shows: b) histogram of the radii; c) the radii distributions corresponding to a subdivision of W into eight sets of equal size and shape. None of these contradict the assumption of homogeneity of $\mathbf{r}_n$ conditioned on $\mathbf{x}_n$.



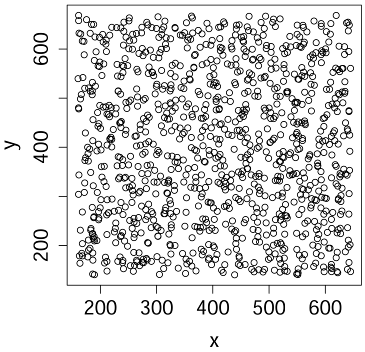
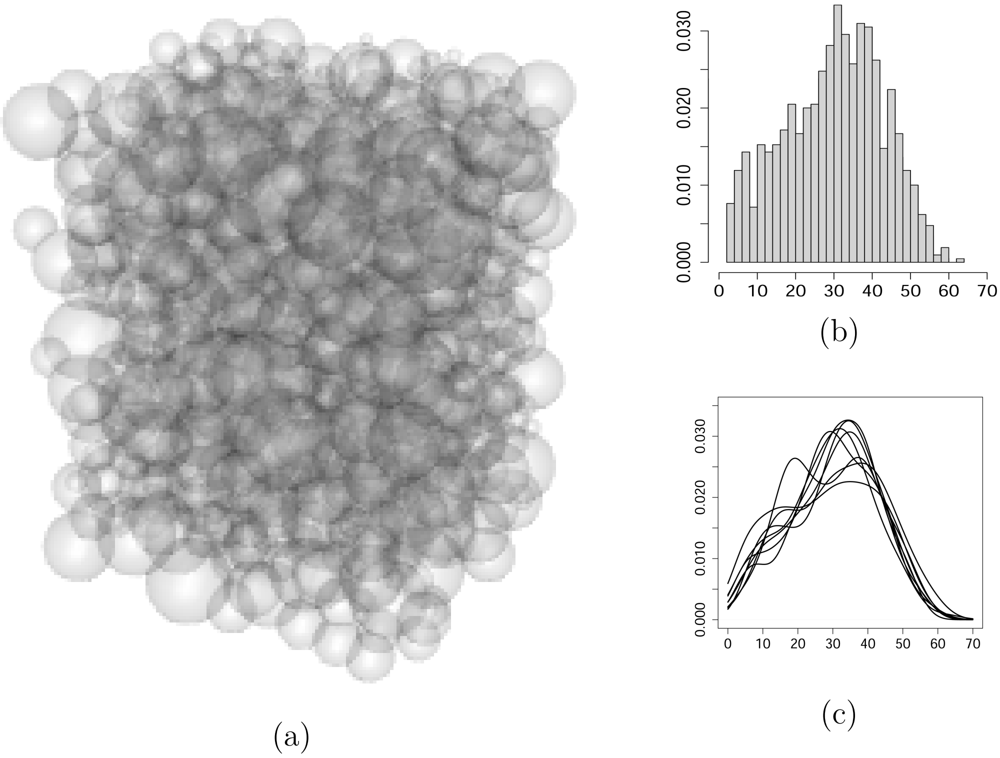


Figure 1.6: The spatial point pattern $\mathbf{x}_n$ (upper left) of the generators of the Laguerre tessellation representing aluminium alloy after applying the periodic boundary condition together with its projections onto the xy, xz and yz planes.

Figure 1.8 and Table 1.4 deal with chosen cells characteristics, cf. Section 2.1.1. Namely, histograms of volume, surface area, number of faces, sphericity and dvol are shown in Figure 1.8. Table 1.4 presents mean values and standard deviations of the same characteristics and of radii, number of vertices, total edge length and neighbour-volume ratio. When compared to Figure 1.4 and Table 1.1, we see that again many cells are small. The distribution of number of faces is more skewed to the left, its mean is 14.21 instead of 14.98. The cell are more spherical: the mean icreased from 0.78 to 0.83 and standard deviation decreased from 0.087 to 0.069. Histogram of dvol suggests that many pairs of neighbouring cells have similar volumes. This is confirmed by NVR, where both the mean value and standard deviation are very small. Comparing both dvol and NVR to NiTi alloy, we conclude that the number of neighbours with a bigger difference in volumes increased, on the other hand there are no extreme differences in volumes of neighbouring cells as mean and standard deviation of NVR is much smaller.

Finally, the empirical correlations of the cell and face characteristics are very similar as in the case of the first dataset, see Tables 1.2 and 1.3 – highly correlated characteristics in case of NiTi alloy remain highly correlated in case of aluminium alloy.



(a) (b) (c)

Figure 1.7: The radii $\mathbf{r}_n$ of the generators of the Laguerre tessellation representing aluminium alloy after applying the periodic boundary condition: a) the balls defined by $(\mathbf{x}_n, \mathbf{r}_n)$; b) histogram of the radii together with a kernel density estimate (solid line) and a fitted density of a beta distribution (dashed line) obtained by maximum pseudolikelihood estimation; c) eight kernel density estimates based on the radii associated to the points in the eight sets for the subdivision of W obtained by dividing its sides into half parts.

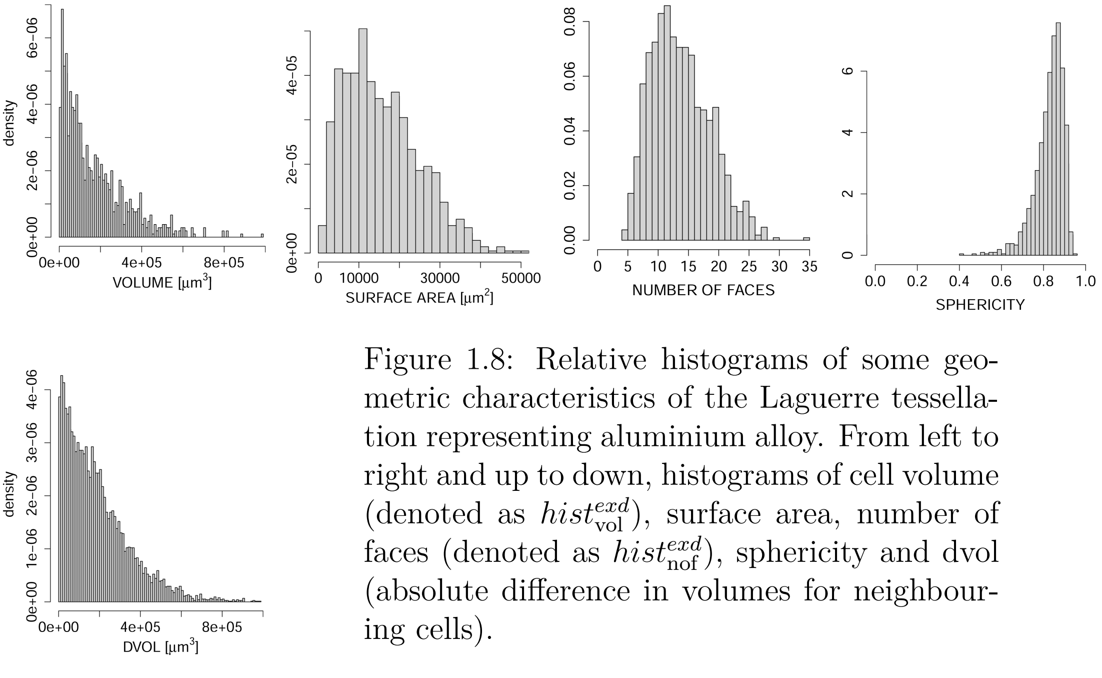


Figure 1.8: Relative histograms of some geometric characteristics of the Laguerre tessellation representing aluminium alloy. From left to right and up to down, histograms of cell volume (denoted as $hist^{exd}_{\mathrm{vol}}$), surface area, number of faces (denoted as $hist^{exd}_{\mathrm{nof}}$), sphericity and dvol (absolute difference in volumes for neighbouring cells).

Table 1.4: Laguerre tessellation characteristics for the aluminium alloy data. In the upper part of the table there are numbers of cells and faces per sample. In the lower part of the table there are mean values of characteristics per cell and standard deviations (sd) for data.

per sample:		
number of cells	1049	
number of faces	7453	
per cell:	mean	sd
radius [μm]	29.7693	18.9564
nof	14.21	4.84
nov	24.42	9.68
vol [μm^3]	$1.68{\cdot}10^5$	$1.53{\cdot}10^5$
surf [μm^2]	$1.56{\cdot}10^4$	$9.21{\cdot}10^3$
tel [μm]	$8.59{\cdot}10^2$	$3.84{\cdot}10^2$
spher	$8.28{\cdot}10^{-1}$	$6.90{\cdot}10^{-2}$
dvol [μm^3]	$1.885{\cdot}10^5$	$1.57{\cdot}10^5$
NVR	1.6995	1.8754

2. Theoretical background

In this chapter we introduce tessellations and point processes, some methods of spatial statistics and stochastic simulation. Basic references are Chiu et al. [2013], Schneider and Weil [2008], Møller and Waagepetersen [2003]. Throughout this thesis, if not stated otherwise, we will work on a complete separable metric space $\mathbb{X}$ equipped with metric ρ and Borel σ-field $\mathcal{B}(\mathbb{X})$ with reference measure γ. $\mathcal{B}_b(\mathbb{X})$ is the system of bounded Borel sets on $\mathbb{X}$. When speaking about randomness, we will always consider a probability space $(\Omega, \mathcal{A}, \mathbb{P})$. We start with some concepts and notations mostly from measure theory.

Definition 1. *For any subset $A \subset \mathbb{X}$ let* $\operatorname{diam} A$ *denote its diameter, that is*

$$\operatorname{diam} A = sup_{x,y \in A} \rho(x,y), \ \operatorname{diam}\ \emptyset = 0.$$

Let $A \subset S$ and $\epsilon > 0$. Define

$$H^d_\epsilon(A) = \inf \left(\sum_{i=1}^{\infty} (\operatorname{diam} U_i)^d : \bigcup_{i=1}^{\infty} U_i \supseteq A, \operatorname{diam} U_i < \epsilon \ \forall i \right).$$

Then the d-dimensional Hausdorff measure of A is

$$H^d(A) = \sup_{\epsilon > 0} H^d_\epsilon(A) = \lim_{\epsilon \to 0} H^d_\epsilon(A).$$

Finally we define the Hausdorff dimension of A by

$$dH(A) = \inf\{d \geq 0 : H^d(A) = 0\}.$$

Definition 2 (Locally finite measure)**.** *A measure μ on $(\mathbb{X}, \mathcal{B}(\mathbb{X}))$ is locally finite if $\mu(K) < \infty$ for every K compact on $\mathbb{X}$.*

Let $\mathbf{M}$ denote the space of all locally finite measures on $\mathbb{X}$ and $\mathbf{N} \subset \mathbf{M}$ the space of measures with values in $\mathbb{N} \cup \{0, \infty\}$.

Definition 3. *A point $x \in \mathbb{X}$ is an atom of a measure $\mu \in \mathbf{N}$ if $\mu(\{x\}) > 0$.*

From local finiteness it follows that each measure $\mu \in \mathbf{N}$ has at most countably many atoms. The following Lemma 1 says, that these atoms can be enumerated in a measurable way.

Lemma 1. *For $\mu \in \mathbf{N}$ there exist measurable mappings $\zeta_i : \mathbf{N} \to \mathbb{X}$ such that*

$$\mu = \sum_{i=1}^{\mu(\mathbb{X})} \delta_{\zeta_i(\mu)}.$$

Proof. See Definition 9.1.XI and Lemma 9.1.XIII in Daley and Vere-Jones [2008]. □

We identify any measure from $\mathbf{N}$ with a locally finite point configuration (also called point pattern - the set of atoms of the measure) and thus we write $\mathbf{N} = \{\mathbf{x} \subset \mathbb{X} : \mathrm{card}(\mathbf{x} \cap B) < \infty, B \in \mathcal{B}_b(\mathbb{X})\}$. Then $\mathcal{N} = \sigma(\{\mathbf{x} \in \mathbf{N} : \mathrm{card}(\mathbf{x} \cap B) = m\}, B \in \mathcal{B}(\mathbb{X}), m \in \mathbb{N} \cup \{0\})$ is the appropriate σ-algebra. Further, $\mathbf{N}^f$ is the family of all finite point configurations in $\mathbb{X}$ equipped with the trace σ-algebra $\mathcal{N}^f$ of $\mathcal{N}$, i.e., $\mathcal{N}^f = \{\mathbf{N}^f \cap N : N \in \mathcal{N}\}$. Thus we have measurable spaces $(\mathbf{N}, \mathcal{N})$ and $(\mathbf{N}^f, \mathcal{N}^f)$. Specially, $\mathbf{N}^{f,k} \subset \mathbf{N}^f$ is the family of configurations with exactly k points, and $\mathbf{N}_W$ is the family of all finite point configurations in $W \in \mathcal{B}_b(\mathbb{X})$.

Definition 4. *A mark space* $\mathbb{M}$ *is a separable locally compact metric space with Borel* σ*-algebra* $\mathcal{B}(\mathbb{M})$*. A pair* (x, m)*,* $x \in \mathbb{X}, m \in \mathbb{M}$ *is called a marked point. Locally finite marked point configuration* $(\mathbf{x}, \mathbf{m})$ *is such that* $\mathrm{card}(\mathbf{x} \cap B) < \infty$ *for all* $B \in \mathcal{B}_b(X)$*.*

We will denote the measurable space of marked point configurations (equivalently marked point patterns) by the same symbol $(\mathbf{N}, \mathcal{N})$ as before. It will be clear whether $\mathbf{N}$ stands for locally finite point patterns or locally finite marked point patterns from the context. Analogously we will use the symbols $\mathbf{N}^f$, $\mathbf{N}_W$, $\mathcal{N}$, ... in the marked case as well.

2.1 Tessellations

We will restrict ourselves to Euclidean spaces, i.e., consider $\mathbb{X} = \mathbb{R}^d$ and ρ to be Euclidean distance. We shorten $\mathcal{B}^d = \mathcal{B}(\mathbb{R}^d)$ and denote by $\mathbf{p}^\top \mathbf{q}$ a scalar product of two column vectors $\mathbf{p}, \mathbf{q} \in \mathbb{R}^d$. A tessellation in $\mathbb{R}^d$ is a locally finite system of space-filling closed sets, which are nonempty and have mutually disjoint interiors, more precisely:

Definition 5 (Tessellation)**.** *A tessellation of* $\mathbb{R}^d$ *is a countable system of sets* $T = \{C_i : i \in \mathbb{N}\}$ *such that*

- $\mathrm{int}(C_i) \cap \mathrm{int}(C_j) = \emptyset$*,* $i \neq j$*,*
- $\bigcup_i C_i = \mathbb{R}^d$*,*
- T *is locally finite (i.e.* $\mathrm{card}(\{C_i \in T : C_i \cap B \neq \emptyset\}) < \infty$ *for all* $B \in \mathcal{B}_b^d$*).*

The tessellation is convex if moreover holds that

- *each* $C_i \in T$*,* $i \in \mathbb{N}$*, is a compact convex set.*

Sets C_i of a tessellation are commonly called cells or in engineering applications grains. More details about tessellation models in $\mathbb{R}^d$ can be found in Møller [1989] and Okabe et al. [2000]. Before exploration of further aspects of tessellations we define a supporting hyperplane of a convex set in Euclidean space.

Definition 6. *A supporting hyperplane of a convex set* $B \in \mathbb{R}^d$ *is a* $(d-1)$*-dimensional hyperplane* $S(\mathbf{p}, b) = \{\mathbf{x} \in \mathbb{R}^d : \mathbf{p}^\top \mathbf{x} = b\}$*,* $\mathbf{p} \in \mathbb{R}^d \setminus \{\mathbf{0}\}$*,* $b \in \mathbb{R}$*, such that either* $B \subseteq \{\mathbf{x} \in \mathbb{R}^d : \mathbf{p}^\top \mathbf{x} \leq b\}$ *or* $B \subseteq \{\mathbf{x} \in \mathbb{R}^d : \mathbf{p}^\top \mathbf{x} \geq b\}$*, and* $\partial B \cap S(\mathbf{p}, b) \neq \emptyset$*.*

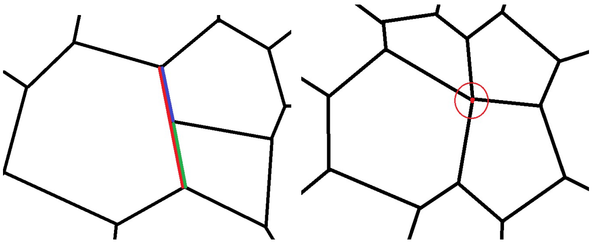


Figure 2.1: Violation of normality in 2D. Left: the tessellation is not face-to-face. Right: the vertex (0-face) in the red circle is contained in the boundary of 4 cells instead of 3.

Definition 7. *Suppose T is a tessellation of $\mathbb{R}^d$ with nonempty compact convex cells.*

1. *Given a cell C, $C \in T$, we consider the intersections of C with its supporting hyperplanes. For $s \in \{0, \dots, d-1\}$ we call an intersection of Hausdorff dimension s an s-face of C. We call the 0-faces the vertices, the 1-faces the edges, and specially in $\mathbb{R}^3$ the 2-faces the faces. For convenience the cell C is considered as a d-face.*

2. *Write $\Delta_s(C)$ for the set of s-faces of a cell C. Further denote*
$$F(y) = \bigcap_{C \in T:\ y \in C} C,\ y \in \mathbb{R}^d$$
and $S_s(T) = \{F(y) : \dim F(y) = s, y \in \mathbb{R}^d\}$, $s \in \{0, \dots, d\}$, the set of s-faces of the tessellation T.

3. *T is called face-to-face if the s-faces of the cells and the s-faces of the tessellation coincide, i.e., if $\Delta_s(T) = S_s(T)$ for $s = 0, \dots, d$.*

4. *T is called normal if it is face-to-face and every s-face of T is contained in the boundary of exactly $d - s + 1$ cells for $s = 0, \dots, d-1$.*

In Figure 2.1 we present examples of tessellations in $\mathbb{R}^2$ that violate conditions of being normal.

We will look at several deterministic tessellation models now, all of them being determined by a locally finite (possibly marked) point pattern in sense that each (marked) point determines a cell. We call generators the elements x_i of such point pattern $\mathbf{x}$ (or (x_i, m_i) elements of marked point pattern $(\mathbf{x}, \mathbf{m})$). The tessellation is then said to be generated by the (marked) point pattern. Since the definition of a tessellation allows only nonempty cells, we assume that all generators of $\mathbf{x}$ or $(\mathbf{x}, \mathbf{m})$ create nonempty cells. Otherwise, we consider only the subset of $\mathbf{x}$ or $(\mathbf{x}, \mathbf{m})$ that generates nonempty cells, i.e., generators creating empty cells are excluded. Definitions of the two most common tessellations follow.

Definition 8 (Voronoi tessellation)**.** *Let $\mathbf{x}$ be a locally finite system of points in $\mathbb{R}^d$. The Voronoi tessellation generated by the point pattern $\mathbf{x}$ is the collection of cells given by*
$$C_i = \{z \in \mathbb{R}^d : ||z - x_i|| \leq ||z - x_j||\ \forall x_j \in \mathbf{x}, i \neq j\}, x_i \in \mathbf{x}.$$
We denote $Vor(\mathbf{x})$ the Voronoi tessellation generated by the point pattern $\mathbf{x}$.

The Voronoi tessellation can be also interpreted as a result of a growth process. The points x_i of $\mathbf{x}$ play the role of nuclei, in which the growth begins at the same instant. The speed of the growth is uniform in all directions. In this case no empty cells can arise.

A generalization of Voronoi tessellation is Laguerre tessellation. It was first introduced by Aurenhammer [1987].

Definition 9 (Laguerre tessellation). *Let* $(\mathbf{x}, \mathbf{m}) = \{(x_i, m_i)\}$*,* $x_i \in \mathbb{R}^d$*,* $m_i \in \mathbb{M} = \mathbb{R}$*, be a countable set such that* $min_{(x,m)\in(\mathbf{x},\mathbf{m})} ||x-y||^2 - m$ *exists for each* $y \in \mathbb{R}^d$*. The Laguerre tessellation generated by the marked point pattern* $\{(x_i, m_i)\}$ *(marks* m_i *are called weights) is the collection of nonempty cells given by*

$$C_i = \{z \in \mathbb{R}^d : ||z - x_i||^2 - m_i \leq ||z - x_j||^2 - m_j \ \forall (x_j, m_j) \in (\mathbf{x}, \mathbf{m}), i \neq j\},$$

$(x_i, m_i) \in (\mathbf{x}, \mathbf{m})$*. We denote* $L(\mathbf{x}, \mathbf{m})$ *the Laguerre tessellation generated by the marked point pattern* $(\mathbf{x}, \mathbf{m})$*.*

We will restrict only to positive weigths, i.e., $m_i > 0$ and $\mathbb{M} = \mathbb{R}_+$, set $m_i = r_i^2$ and call

$$\rho(y, b(x, r)) = ||x - y||^2 - r^2 \tag{2.1}$$

the power distance of a point $y \in \mathbb{R}^d$ to a ball $b(x, r)$, $x \in \mathbb{R}^d$, $r > 0$. The values $\{r_i\}$ are called radii of Laguerre tessellation. We will write $(\mathbf{x}, \mathbf{r})$ for the corresponding marked point pattern and $L(\mathbf{x}, \mathbf{r})$ for the corresponding Laguerre tessellation. In the similar manner, $\mathbb{M}$ will now denote the space of radii (instead of weights) and we will consider $\mathbb{M} \subseteq \mathbb{R}_+$.

The interpretation of the power distance is as follows: for each $z \in \mathbb{R}^d$ outside the sphere $b(x, r)$, the value $\rho(z, b(x, r))$ equals the squared length of the tangent line segment from z to the sphere, cf. Fig. 2.2. The power distance $\rho(z, b(x, r))$ equals 0 if z lies on the boundary of the sphere, and it is smaller than 0 if z is inside the sphere. If a further generator $(y, q) \in \mathbf{x}$ overlaps with (x, r) over the center (i.e., $y \in b(x, r)$), it can happen that either the cell corresponding to marked point (y, q) does not cover y or even that there is no cell at all (in this case the generator can be omitted, it holds that $L(\mathbf{x}, \mathbf{r}) = L(\mathbf{x} \setminus \{y\}, \mathbf{r} \setminus \{q\})$ and we say that (y, q) is redundant in $(\mathbf{x}, \mathbf{r})$.). Note that the Laguerre tessellations are invariant under transformations of radii of the form $r \mapsto \sqrt{r^2 + t}$, where $t \in \mathbb{R}$ is fixed such that all radii remain positive. More details about Laguerre tessellations can be found in Lautensack [2007].

The Voronoi and Laguerre tessellations are two examples of tessellation models generated by a (marked) point pattern. All these models have for a given set of generators $\mathbf{g}$ the cells defined as

$$C_i = \{z \in \mathbb{R}^d : d(z, g_i) \leq d(z, g_j) \ \forall g_j \in \mathbf{g}, i \neq j\}. \tag{2.2}$$

by means of an appropriate distance d in $\mathbb{R}^d$. The distance d is the Euclidean distance in case of Voronoi tessellation and the power distance, (2.1), in case of Laguerre tessellation.

Both Voronoi and Laguerre tessellations have planar cell boundaries and are created by convex cells. However, Laguerre tessellation can produce tessellations with much more variation in terms of both cell size and aspect ratio. In the

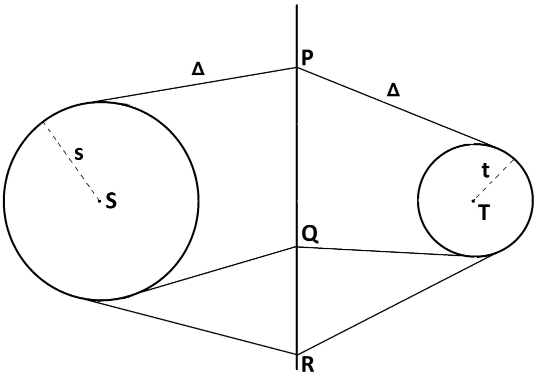


Figure 2.2: Illustration of the power distance given in (2.1). All three points P, Q, R have the same power distance with respect to circles with centers S, T and radii s, t, respectively. Δ is the square root of the power distance

case that all marks are equal, the Laguerre tessellation reduces to a Voronoi tessellation.

A further generalization of Laguerre tessellation is possible if we instead of Euclidean norm use an elliptic norm. This generalization is called generalized balanced power diagram (GBPD). The non-convex GBPD is beyond the scope of this thesis, as we restrict to convex tessellations only. More details about GBPD can be found in Alpers et al. [2015].

Definition 10. *We say that* $(\mathbf{x}, \mathbf{r}) \subset \mathbb{R}^d \times \mathbb{R}_+$ *fulfills regularity condition if*

(R1) for every $z \in \mathbb{R}^d$ *and every* $t \in \mathbb{R}$ *only finitely many elements* $(x, r) \in (\mathbf{x}, \mathbf{r})$ *satisfy* $||x - z||^2 - r^2 \leq t$*, and*

(R2) $\operatorname{conv}\{x : (x, r) \in (\mathbf{x}, \mathbf{r})\} = \mathbb{R}^d$.

Further, we say that the points of $(\mathbf{x}, \mathbf{r})$ *are in general position if the following conditions hold:*

(GP1) no $k + 1$ *points* x, $(x, r) \in (\mathbf{x}, \mathbf{r})$*, are contained in a* $(k - 1)$*-dimensional affine subspace of* $\mathbb{R}^d$ *for* $k = 2, \ldots, d$*, and*

(GP2) no $d + 2$ *points have the equal power distance with respect to some point in* $\mathbb{R}^d$.

If the set of radii is bounded, the first part of regularity condition implies the local finiteness of the set of marked points $\{x : (x, r) \in (\mathbf{x}, \mathbf{r})\}$.

Theorem 2. *Let* $L(\mathbf{x}, \mathbf{r})$ *be the Laguerre tessellation generated by marked point pattern* $(\mathbf{x}, \mathbf{r})$*. Let* $(\mathbf{x}, \mathbf{r})$ *satisfy regularity condition. Then*

1. *every cell of* $L(\mathbf{x}, \mathbf{r})$ *is compact;*
2. *the set of cells of* $L(\mathbf{x}, \mathbf{r})$ *is locally finite and space filling;*
3. *the* $L(\mathbf{x}, \mathbf{r})$ *is face-to-face;*
4. *moreover if* $(\mathbf{x}, \mathbf{r})$ *is in general position, then* $L(\mathbf{x}, \mathbf{r})$ *is normal.*

Proof. Proofs can be found in Lautensack [2007], Propositions 2.2.2, 2.2.4., 2.2.5 and Corollary 2.2.7. ☐

Theorem 3. *Every normal tessellation of* $\mathbb{R}^d$ *for* $d \geq 3$ *is a Laguerre tessellation.*

Proof. Proof can be found in Lautensack [2007], Theorem 2.4.3. ☐

2.1.1 Tessellation characteristics

Given a tessellation T, we are interested how to quantify its properties. In Definition 7 we have introduced so called s-faces of the cells, $s = 0, \ldots, d$. In case $d = 3$ we call the s-faces, $s = 0, \ldots, 3$, vertices, edges, faces and cells, respectively. These objects can be described by a pleiad of geometrical characteristics. We restrict ourselves to the case $d = 3$ and provide a reader with a list of the most common characteristics together with their abbreviations which are used throughout all formulae and tables within this thesis.

Cell characteristics: volume (abbreviated as 'vol'), surface area ('surf'), total edge length ('tel'), number of faces ('nof'), number of edges ('noe'), number of vertices ('nov'), sphericity ('spher').

Sphericity is defined by

$$\text{spher} = \frac{\pi^{1/3}(6 \cdot \text{vol})^{2/3}}{\text{surf}}. \tag{2.3}$$

Note that $0 < \text{spher} \leq 1$ with the value 1 corresponding to a sphere. Further, in 3D it holds that the minimal number of faces is 4 and

$$\text{nov} = \text{nof} + \text{noe - 2}.$$

Further geometrical characteristics can be invented for groups of cells in order to describe cell interactions. We call the characteristics describing a pairwise interaction of two neighbouring cells as face characteristics.

Face characteristics: face area ('farea'), face perimeter ('fper'), number of edges per face ('fnoe'), number of vertices per face ('fnov').

Another examples of such characteristics are absolute difference in volumes of two neighbouring cells ('dvol') and neighbour-volume ratio ('NVR') defined by

$$\text{NVR}(C_1, C_2) = \begin{cases} \left(\frac{\max\{|C_1|,|C_2|\}}{\min\{|C_1|,|C_2|\}} - 1\right)^{1/2} & \text{if } C_1 \sim C_2, \\ 0 & \text{otherwise,} \end{cases} \tag{2.4}$$

where $\sim$ stands for a neighbourhood relation. We say that two cells are neighbours if they share a common face. In general, for $k \geq 2$, if the k-tuple $C_1, \ldots, C_k$ can be arranged such that $C_{i_1} \sim C_{i_2} \sim \ldots \sim C_{i_k}$, we write $C_1, \ldots, C_k \sim_k$ and call it a k-neighbourhood relation. Specially for $k = 3$ and 4 we define a proper neighbourhood relation if the 3-tuple and 4-tuple of cells have a common edge and vertex (we can speak about edge and vertex characteristics), respectively.

For all above mentioned geometrical characteristics means and standard deviations over the tessellation are usually of interest. Alternatively to the moment properties the whole distribution can be examined, e.g., by means of histogram. Let us consider some geometrical characteristic taking values in an interval $[a, b]$. For some integer J, let $D = \{t_i\}_{i=0}^J$ with $t_i < t_{i+1}$ for all i be a decomposition of the interval $[a, b]$ into J subintervals such that $t_0 = a$ and $t_J = b$ (D does not need to be equidistant). Each histogram H can then be represented by some numbers $h_1, \dots, h_J$ interpreted as frequencies of the classes $1, \dots, J$ (i.e., h_i is the number of facets for which the value of the considered geometrical characteristic belongs to the interval $[t_{i-1}, t_i)$). Except of the histogram, one could deal with the cumulative histogram or empirical distribution function.

2.1.2 Periodic boundary conditions

Estimation of tessellation characteristics requires a special treatment of cells touching the boundary of the observation window W. These cells, cutted by the boundary, are often assumed to be not complete. Due to homogeneity, a convenient way how to deal with this undesirable effect is introduction of so called periodic boundary condition. Because of applications, the periodic boundary condition is formulated in case $d = 3$. We give a formal definition in case of a marked point pattern $(\mathbf{x}, \mathbf{r})$. Equivalently, application of the periodic boundary condition can be viewed as wrapping the observation window W on a torus.

Definition 11 (Periodic conditions)**.** *Let* $W = [0, a] \times [0, b] \times [0, c]$ *be a rectangular parallelepiped and* $(\mathbf{x}, \mathbf{r})$ *a finite marked point pattern such that* $\mathbf{x} \subset W$. *Define a periodic extension of* $(\mathbf{x}, \mathbf{r})$ *which is in accordance to* W, *i.e., the infinite marked point pattern*

$$(\mathbf{x}^*, \mathbf{r}^*) = \bigcup_{((x,y,z),r) \in (\mathbf{x},\mathbf{r})} \bigcup_{(i,j,k) \in \mathbb{Z}^3} \{((ia + x, jb + y, kc + z), r)\} \tag{2.5}$$

where $\mathbb{Z}^3$ *is the 3D integer lattice. Note that* $\mathbf{x} = \mathbf{x}^* \cap W$.

Application of the periodic boundary condition to a marked point pattern $(\mathbf{x}, \mathbf{r})$ of generators of Laguerre tessellation may cause that some generators of nonempty cells in $L(\mathbf{x}, \mathbf{r})$ might generate empty cells in $L(\mathbf{x}^*, \mathbf{r}^*)$. Then we keep only those Laguerre cells $C(x, r \mid \mathbf{x}^*, \mathbf{r}^*)$ which are non-empty and have $x \in \mathbf{x}$. Let n be the number of such cells and $(\mathbf{x}_n, \mathbf{r}_n) \subseteq (\mathbf{x}, \mathbf{r})$ be the marked point pattern specifying the generators of these cells. Note that $L(\mathbf{x}^*, \mathbf{r}^*) = L(\mathbf{x}_n^*, \mathbf{r}_n^*)$.

In the thesis, all simulations are carried out in the periodic setting and parameter estimation in Chapter 4 relies on the periodic treatment of boundary effects as well.

2.1.3 Point positions in Laguerre geometry

In this and the following section we deal with arrangement of marked points in the space $\mathbb{R}^3 \times \mathbb{M}$, where $\mathbb{M} \subseteq \mathbb{R}_+$. Recall that we use the notation $(\mathbf{x}, \mathbf{r})$ for a marked point pattern and write $(x, r) \in (\mathbf{x}, \mathbf{r})$ when considering a single point of the point pattern. We use the symbols $\mathbf{x}$ and $\mathbf{r}$ separately when reffering to points in $\mathbb{R}^3$ and radii in $\mathbb{M}$, respectively. By η we will always denote a finite subset of $(\mathbf{x}, \mathbf{r})$, i.e., $\eta = \{(x_j, r_j)\}_{j \in J}$ for some J finite subset of $\mathbb{N}$. Then η_x and η_r will denote the points and radii of η, respectively, i.e., $\eta_x = \{x_j\}_{j \in J}$ and $\eta_r = \{r_j\}_{j \in J}$. Recall that $(\mathbf{x}, \mathbf{r})_W$ denotes a restriction of $(\mathbf{x}, \mathbf{r})$ to $W \in \mathcal{B}_b^3$. The mutual arrangement of marked points is described in terms of power distance (2.1) (this is what we call a Laguerre geometry in the section title).

Definition 12. *Let* $(\mathbf{x}, \mathbf{r})$ *be a marked point pattern in* $\mathbb{R}^3 \times \mathbb{M}$ *and* $\rho(z, (x, r)) = ||z - x||^2 - r^2$ *denote the power distance of* $z \in \mathbb{R}^3$ *to a marked point* $(x, r) \in (\mathbf{x}, \mathbf{r})$.

1. *We say that two marked points* $(x, r), (y, s)$ *are orthogonal if* $\rho(x, (y, s)) = r^2$.
2. *We define the characteristic point of* $\eta \subset (\mathbf{x}, \mathbf{r})$ *as a marked point* $(x_\eta, r_\eta) \in \mathbb{R}^3 \times \mathbb{R}_+$ *which is orthogonal to every* $(y, s) \in \eta$.
3. $\eta \subset (\mathbf{x}, \mathbf{r})$ *is said to be Laguerre-cospherical, if there exists its characteristic point* (x_η, r_η).
4. *We say that* $\eta \subset (\mathbf{x}, \mathbf{r})$ *is regular in* $(\mathbf{x}, \mathbf{r})$ *if* $\rho(x_\eta, (y, s)) \geq r_\eta^2$ *for all* $(y, s) \in (\mathbf{x}, \mathbf{r})$.

It follows, that for arbitrary $\eta \subset (\mathbf{x}, \mathbf{r})$ the characteristic point of η needs not to exist or it may be non-unique. The statement $\eta \subset (\mathbf{x}, \mathbf{r})$ is Laguerre-cospherical can be paraphrased as 'for η there exists a point $x_\eta \in \mathbb{R}^3$ possesing the same power distance ρ to all points $(y, s) \in \eta$.

Using the terminology introduced in this section and Definition 12, the general position of $(\mathbf{x}, \mathbf{r})$ stated in Definition 10 can be reformulated in $\mathbb{R}^3$ as follows:

(GP1) $\eta \subset (\mathbf{x}, \mathbf{r}), 3 \leq \mathrm{card}(\eta) \leq 4 \Rightarrow \eta_x$ is an affinely independent set of points in $\mathbb{R}^3$,

(GP2) $\eta \subset (\mathbf{x}, \mathbf{r}), \mathrm{card}(\eta) > 4 \Rightarrow \eta$ is not Laguerre cospherical.

Note that the points $x_0, \ldots, x_k \in \mathbb{R}^3$, $k \leq 3$, are affinely independent if the vectors $x_1 - x_0, \ldots, x_k - x_0$ are linearly independent. Let $\mathbf{N}_{gp}$ denote the set of all locally finite marked point configurations $(\mathbf{x}, \mathbf{r})$ being in general position.

Definition 13. *For* $(\mathbf{x}, \mathbf{r}) \in \mathbf{N}_{gp}$ *we define*

$$\mathcal{LD}(\mathbf{x}, \mathbf{r}) := \{\eta \subset (\mathbf{x}, \mathbf{r}) : \eta \textit{ is regular}\},$$

$$\mathcal{LD}_k(\mathbf{x}, \mathbf{r}) = \{\eta \in \mathcal{LD}(\mathbf{x}, \mathbf{r}) : \mathrm{card}(\eta_x) = k\}, \ \textit{for } k = 1, \ldots, 4.$$

Claim 4. *Let* $(\mathbf{x}, \mathbf{r}) \in \mathbf{N}_{gp}$. *It holds that characteristic point* (x_η, r_η) *exists for all* $\eta \subset (\mathbf{x}, \mathbf{r})$ *with* card $(\eta) \leq 4$ *and moreover* (x_η, r_η) *is unique whenever* card $(\eta) = 4$.

Proof. Assume $\eta = \{(x_1, r_1), \ldots, (x_4, r_4)\} \subset (\mathbf{x}, \mathbf{r}) \in \mathbf{N}_{gp}$ to be arbitrary. The characteristic point has to satisfy $||x_\eta - x_i||^2 - {r_i}^2 = r_\eta^2$ for all $i = 1, \ldots, 4$. Denote coordinates of x_i as a_i, b_i, c_i and analogically coordinates of x_η as a_η, b_η, c_η. Then the system of equations becomes

$$a_\eta^2 + b_\eta^2 + c_\eta^2 - r_\eta^2 - 2a_i a_\eta - 2b_i b_\eta - 2c_i c_\eta = {r_i}^2 - a_i^2 - b_i^2 - c_i^2, \ i = 1, \ldots, 4.$$

It is linear with respect to $(\alpha_\eta, a_\eta, b_\eta, c_\eta)$, where $\alpha_\eta = a_\eta^2 + b_\eta^2 + c_\eta^2 - r_\eta^2$. The space locations $x_1, \dots, x_4$ are affinely independent, since $\eta \in \mathbf{N}_{gp}$, and hence the matrix formulation of the system is of the full rank. Therefore, the system has a unique solution. In case of $\operatorname{card} \eta < 4$, the system is underdetermined and of the full row rank as $\eta \in \mathbf{N}_{gp}$. That implies there are infinitely many solutions. □

Definition 14 (Delaunay tessellation)**.** *Let* $(\mathbf{x}, \mathbf{r}) \in \mathbf{N}_{gp}$. *Let The Delaunay tessellation determined by* $(\mathbf{x}, \mathbf{r})$ *is the collection of cells given by*

$$D_\eta = \{z \in \mathbb{R}^3 : z \in \operatorname{conv}(\eta_x)\}, \ \eta \textit{ regular in } (\mathbf{x}, \mathbf{r}), \operatorname{card}(\eta) = 4.$$

We denote $D(\mathbf{x}, \mathbf{r})$ *the Delaunay tessellation determined by the marked point pattern* $(\mathbf{x}, \mathbf{r})$*.*

In words, the Delaunay tessellation $D(\mathbf{x}, \mathbf{r})$ in $\mathbb{R}^3$ is a collection of tetrahedrons whose set of vertices is regular, i.e., collection of cells

$$\{D_\eta\}_{\eta \in \mathcal{LD}_4(\mathbf{x})} = \{\operatorname{conv}(\eta_x)\}_{\eta \in \mathcal{LD}_4(\mathbf{x})}.$$

The Delaunay tessellations of $\mathbb{R}^3$ are often called Delaunay tetrahedrizations, since their cells form tetrahedra.

As both Laguerre and Delaunay tessellations are determined by a marked point pattern $(\mathbf{x}, \mathbf{r})$, it is a question whether there is a relationship between them. Indeed, the Laguerre and Delaunay tessellations are dual, i.e., we can construct one if we know the other. Let $C(x, r)$ be a Laguerre cell corresponding to the generator (x, r) and $\mathcal{F}_0(C(x, r))$ denote the set of vertices of a nonempty cell $C(x, r)$, $(x, r) \in (\mathbf{x}, \mathbf{r})$. Then

$$\tilde{D}_v = \operatorname{conv}\{x : (x, r) \in (\mathbf{x}, \mathbf{r}), v \in \mathcal{F}_0(C(x, r))\}$$

is a tetrahedron corresponding to the vertex v (each vertex is an intersection of exactly four Laguerre cells). Let $\mathcal{F}_{0,(\mathbf{x},\mathbf{r})} = \cup_{(x,r) \in (\mathbf{x},\mathbf{r})} \mathcal{F}_0(C(x, r))$ denote the set of all vertices of the tessellation $L(\mathbf{x}, \mathbf{r})$. The Delaunay tessellation $D(\mathbf{x}, \mathbf{r})$ can then be alternatively defined as a collection of cells $\tilde{D}_v$, $v \in \mathcal{F}_{0,(\mathbf{x},\mathbf{r})}$. For $(x_1, r_1), (x_2, r_2), (x_3, r_3), (x_4, r_4) \in (\mathbf{x}, \mathbf{r})$ the set $\tilde{D}_v = \operatorname{conv}\{x_1, x_2, x_3, x_4\}$ is a Delaunay cell if and only if $C(x_1, r_1) \cap C(x_2, r_2) \cap C(x_3, r_3) \cap C(x_4, r_4) \neq \emptyset$. It holds that $D_\eta = \tilde{D}_v$ if $\eta = \{(x, r) \in (\mathbf{x}, \mathbf{r}) : v \in \mathcal{F}_0(C(x, r))\}$.

All terms defined in Definition 12 for marked point pattern $(\mathbf{x}, \mathbf{r})$ using power distance have their analogies for unmarked point pattern $\mathbf{x}$ using Euclidean distance. The characteristic point becomes a circumball circumscribed to a given subset of points of $\mathbf{x}$, Laguerre-cosphericity becomes cosphericity with respect to Euclidean distance and regularity becames so called empty sphere property, i.e., the property meaning that the circumball circumscribed to a given subset of points of $\mathbf{x}$ contains no other points of $\mathbf{x}$. The Delaunay tessellation in $\mathbb{R}^3$ given by $\mathbf{x}$ is a collection of sets

$$\{\operatorname{conv}(\eta_x) : \ \operatorname{card}(\eta_x) = 4, \ \eta_x \text{ satisfies empty sphere property in } \mathbf{x}\}.$$

Sometimes, we use term Delaunay tessellation only when reffering to an unmarked point pattern and instead the term Laguerre Delaunay tessellation when reffereing to a marked point pattern. Recall, that in the case that all marks are equal, the Laguerre tessellation reduces to a Voronoi tessellation and in the same maner we say that Laguerre Delaunay tessellation reduces to Delaunay tessellation.

2.1.4 Hypergraphs

In this section, we follow the usage of hypergraphs from Dereudre et al. [2012]. The concept of hypergraph is mainly the theoretical tool how to effectively describe the tessellation geometry.

Comment on notation: To keep the formulations concise we will write $\mathbf{g}$ instead of both unmarked locally finite point pattern $\mathbf{x}$ and marked locally finite point pattern $(\mathbf{x}, \mathbf{r})$. By $\eta \subset \mathbf{g}$ we always mention a finite (marked) point configuration. Writing $\eta \in \mathbf{N}^f$ and $\mathbf{g} \in \mathbf{N}$ is correct as we use symbols $\mathbf{N}, \mathbf{N}^f$, etc. in both unmarked and marked case.

Definition 15 (Hypergraph)**.** *A hypergraph structure is a measurable subset* $\mathcal{E}$ *of* $(\mathbf{N}^f \times \mathbf{N}, \mathcal{N}^f \otimes \mathcal{N})$ *such that* $\eta \subset \mathbf{g}$ *for all* $(\eta, \mathbf{g}) \in \mathcal{E}$*. We call* η *a hyperedge of* $\mathbf{g}$ *and write* $\eta \in \mathcal{E}(\mathbf{g})$*, where* $\mathcal{E}(\mathbf{g}) = \{\eta : (\eta, \mathbf{g}) \in \mathcal{E}\}$*. For a given* $\mathbf{g} \in \mathbf{N}$*, the pair* $(\mathbf{g}, \mathcal{E}(\mathbf{g}))$ *is called a hypergraph.*

Notice that we have already seen five hypergraph structures. Indeed, for $\mathbf{g} = (\mathbf{x}, \mathbf{r}) \in \mathbf{N}$, the pairs $((\mathbf{x}, \mathbf{r}), \mathcal{LD}(\mathbf{x}, \mathbf{r}))$ and $((\mathbf{x}, \mathbf{r}), \mathcal{LD}_k(\mathbf{x}, \mathbf{r}))$, $k = 1, \ldots, 4$, are hypergraphs.

The finite point configuration η, $\operatorname{card}(\eta) = k$, is said to be connected in $\mathcal{LD}(\mathbf{x}, \mathbf{r})$ if for any two points $(x, r), (y, s) \in \eta$ there exists a path between (x, r) and (y, s) in the graph $\mathcal{LD}_2(\mathbf{x}, \mathbf{r})$, that is, there exists $m \in \{1, \ldots, k-1\}$ and a sequence $(x_0, r_0), \ldots, (x_m, r_m) \in \eta$ such that $(x, r) = (x_0, r_0)$, $(y, s) = (x_m, r_m)$ and

$$\{(x_i, r_i), (x_{i+1}, r_{i+1})\} \in \mathcal{LD}_2(\mathbf{x}, \mathbf{r})$$

for all $i \in \{0, \ldots, m-1\}$. We define the graph of connected k-tuples

$$\mathcal{CG}_k = \{(\eta, (\mathbf{x}, \mathbf{r})) : \eta \subset (\mathbf{x}, \mathbf{r}), \operatorname{card}(\eta) = k, \eta \text{ is connected in } \mathcal{LD}(\mathbf{x}, \mathbf{r})\}$$

and

$$\mathcal{CG}_{k,b} = \{(\eta, (\mathbf{x}, \mathbf{r})) \in \mathcal{CG}_k : \forall (x, r) \in \eta : C(x, r) \text{ is bounded}\}.$$

Both $\mathcal{CG}_k$ and $\mathcal{CG}_{k,b}$ are hypergraph structures.

Definition 16 (Sublinearity)**.** *A hypergraph* $\mathcal{E}$ *is sublinear if there exists a constant* $c_s < \infty$ *such that for all* $(\mathbf{x}, \mathbf{r}) \in \mathbf{N}^f$ *we have* $\operatorname{card}(\mathcal{E}(\mathbf{x}, \mathbf{r})) \leq c_s \operatorname{card}(\mathbf{x})$.

Definition 17 (Hyperedge potential)**.** *A hyperedge potential is a measurable function* $\varphi : \mathcal{E} \to \mathbb{R} \cup \{+\infty\}$*. The hyperedge potential is said to be hereditary if* $\varphi(\eta, \mathbf{g}) < \infty$ *for some* $(\eta, \mathbf{g}) \in \mathcal{E}$ *implies that* $\varphi(\eta, \tilde{\mathbf{g}}) < \infty$ *for all* $\tilde{\mathbf{g}} \subset \mathbf{g}$ *such that* $\eta \in \mathcal{E}(\tilde{\mathbf{g}})$.

For notational convenience, we set $\varphi = 0$ on $\mathcal{E}^c$. The function φ introduces interactions on hyperedges that need not be hereditary since φ is allowed to take the value ∞ (hard-core case).

Definition 18. *Let* φ *be a hyperedge potential for a hypergraph structure* $\mathcal{E}$*. Then* φ *is*

(i) shift-invariant if

$$(\vartheta_t \eta, \vartheta_t \mathbf{g}) \in \mathcal{E} \text{ and } \varphi(\vartheta_t \eta, \vartheta_t \mathbf{g}) = \varphi(\eta, \mathbf{g}) \text{ for all } (\eta, \mathbf{g}) \in \mathcal{E} \text{ and } t \in \mathbb{R}.$$

(ii) unary if there exists a measurable function $\hat{\varphi} : \mathbf{N} \to \mathbb{R} \cup \{+\infty\}$ such that

$$\varphi(\eta, \mathbf{g}) = \hat{\varphi}(\eta) \text{ for } \eta \in \mathcal{E}(\mathbf{g}).$$

If $\mathbf{g} = (\mathbf{x}, \mathbf{r})$, then $\vartheta_t(\mathbf{x}, \mathbf{r}) = \{(x - t, r) : (x, r) \in (\mathbf{x}, \mathbf{r})\}$ is the translation of the positional part of the configurations by the vector $-t \in \mathbb{R}^3$. All hyperedge potentials mentioned in this thesis are assumed to be shift-invariant.

Definition 19 (Finite horizon)**.** *A set $\Delta \in \mathcal{B}_b^3$ is a finite horizon for the pair $(\eta, (\mathbf{x}, \mathbf{r})))\in \mathcal{E}$ and the hyperedge potential φ if for all $(\tilde{\mathbf{x}}, \tilde{\mathbf{r}}) \in \mathbf{N}, (\tilde{\mathbf{x}}, \tilde{\mathbf{r}}) = (\mathbf{x}, \mathbf{r})$ on $\Delta \times \mathbb{M}$*

$$(\eta, (\tilde{\mathbf{x}}, \tilde{\mathbf{r}})) \in \mathcal{E} \text{ and } \varphi(\eta, (\tilde{\mathbf{x}}, \tilde{\mathbf{r}})) = \varphi(\eta, (\mathbf{x}, \mathbf{r})).$$

Claim 5 (Finite horizon in case of $\mathcal{LD}$ and unary potential)**.** *Let $\mathbb{M} = [0, R_{\max}]$ for some $R_{\max} > 0$. Then the set $\Delta = b(x_\eta, \sqrt{r_\eta^2 + R_{\max}^2})$ is a finite horizon for $(\eta, (\mathbf{x}, \mathbf{r})) \in \mathcal{LD}$ and arbitrary unary potential $\varphi(\eta, (\mathbf{x}, \mathbf{r})) = \hat{\varphi}(\eta)$. Moreover the finite horizon Δ can be decomposed on a spherical shell with a bounded thickness and a ball containing no points of $\mathbf{x}$.*

Proof. Let $\eta \in \mathcal{LD}(\mathbf{x}, \mathbf{r})$ and (x_η, r_η) be its characteristic point. Taking an arbitrary point $(y, s) \in (\mathbf{x}, \mathbf{r})$ we obtain from regularity of η in $(\mathbf{x}, \mathbf{r})$ that $r_\eta^2 \leq \rho(x_\eta, (y, s)) = ||x_\eta - y||^2 - s^2 \leq ||x_\eta - y||^2$. It follows that the ball $b(x_\eta, r_\eta)$ does not contain the points of $\mathbf{x}$. The goal is to construct a finite horizon for a given pair $(\eta, (\mathbf{x}, \mathbf{r})) \in \mathcal{LD}$ and a given unary potential. Notice that the finite horizon does not depend on the unary potential, since for arbitraty $(\tilde{\mathbf{x}}, \tilde{\mathbf{r}})$ such that $\eta \subset (\tilde{\mathbf{x}}, \tilde{\mathbf{r}})$ we have $\varphi(\eta, (\tilde{\mathbf{x}}, \tilde{\mathbf{r}})) = \hat{\varphi}(\eta) = \varphi(\eta, (\mathbf{x}, \mathbf{r}))$, where $\hat{\varphi}$ is from the definition of the unary potential. Therefore, all we need to fulfill is that $(\eta, (\tilde{\mathbf{x}}, \tilde{\mathbf{r}})) \in \mathcal{E}$ for every $(\tilde{\mathbf{x}}, \tilde{\mathbf{r}}) \in \mathbf{N} : (\tilde{\mathbf{x}}, \tilde{\mathbf{r}}) = (\mathbf{x}, \mathbf{r})$ on $\Delta \times \mathbb{M}$. To do that we use the fact that the mark space is bounded, $\mathbb{M} = [0, R_{\max}]$. Then $\Delta = b(x_\eta, \sqrt{r_\eta^2 + R_{\max}^2})$ is sufficient as a horizon, since any point $(y, s) \in (\mathbb{R}^3 \setminus \Delta) \times \mathbb{M}$ satisfies $\|x_\eta - y\| \geq \sqrt{r_\eta^2 + R_{\max}^2}$ and therefore cannot violate the regularity of η. Indeed, since $s \leq R_{\max}$, $\|x_\eta - y\| \geq \sqrt{r_\eta^2 + s^2}$, what means that η is regular in arbitrary $(\tilde{\mathbf{x}}, \tilde{\mathbf{r}}) \in \mathbf{N} : (\tilde{\mathbf{x}}, \tilde{\mathbf{r}}) = (\mathbf{x}, \mathbf{r})$ on $\Delta \times \mathbb{M}$. Moreover, all points of η_x are contained in the set Δ, since for every $(y, s) \in \eta$ we have $\|x_\eta - y\| = \sqrt{r_\eta^2 + s^2} \leq \sqrt{r_\eta^2 + R_{\max}^2}$. The set $\Delta \setminus b(x_\eta, r_\eta)$ is a spherical shell (a generalization of an annulus to three dimensions) with thickness $\sqrt{r_\eta^2 + R_{\max}^2} - r_\eta = \frac{R_{\max}^2}{\sqrt{r_\eta^2 + R_{\max}^2} + r_\eta} \leq R_{\max}$. Although the number of points in the spherical shell is not bounded, we have the bound on its thickness. ☐

Definition 20 (Range confinement)**.** *Let $W \in \mathcal{B}_b^3$. Define the set*

$$\mathcal{E}_W(\mathbf{x}, \mathbf{r}) := \{\eta \in \mathcal{E}(\mathbf{x}, \mathbf{r}) : \varphi(\eta, \zeta \cup (\mathbf{x}, \mathbf{r})_{W^c}) \neq \varphi(\eta, (\mathbf{x}, \mathbf{r})) \text{ for some } \zeta \in \mathbf{N}_W\}.$$

We say a configuration $(\mathbf{x}, \mathbf{r}) \in \mathbf{N}$ confines the range of φ from W if there exists a set $\partial W(\mathbf{x}, \mathbf{r}) \in \mathcal{B}_b^3$ such that $\varphi(\eta, \zeta \cup (\tilde{\mathbf{x}}, \tilde{\mathbf{r}})_{W^c}) = \varphi(\eta, \zeta \cup (\mathbf{x}, \mathbf{r})_{W^c})$ whenever $(\tilde{\mathbf{x}}, \tilde{\mathbf{r}}) = (\mathbf{x}, \mathbf{r})$ on $\partial W(\mathbf{x}, \mathbf{r}) \times \mathbb{M}$, $\zeta \in \mathbf{N}_W$ and $\eta \in \mathcal{E}_W(\zeta \cup (\mathbf{x}, \mathbf{r})_{W^c})$. In this case we write $(\mathbf{x}, \mathbf{r}) \in \mathbf{N}_{\mathrm{cr}}^W$. We denote $r_{W,(\mathbf{x},\mathbf{r})}$ the smallest possible r such that $(W + b(0, r)) \setminus W$ satisfies the definition of $\partial W(\mathbf{x}, \mathbf{r})$. We will use the abbreviation $\partial_W(\mathbf{x}, \mathbf{r}) = (\mathbf{x}, \mathbf{r})_{\partial W((\mathbf{x},\mathbf{r}))}$.

2.2 Point processes and marked point processes

The theory regarding the point processes can be found for example in Daley and Vere-Jones [2003], Daley and Vere-Jones [2008] and Møller and Waagepetersen [2003].

Definition 21 (Point process). *A point process on* $\mathbb{X}$ *is a measurable mapping* $\Phi : (\Omega, \mathcal{A}, \mathbb{P}) \to (\mathbf{N}, \mathcal{N})$. *The distribution of the point process* Φ *is a probability measure* Q *on* $(\mathbf{N}, \mathcal{N})$ *given by* $Q(\cdot) = \mathbb{P}(\{\omega \in \Omega : \Phi(\omega) \in \cdot\})$. *The measure on* $\mathbb{X}$ *defined by* $\Lambda(B) = \mathsf{E}\, \Phi(B)$, $B \in \mathcal{B}(\mathbb{X})$ *is called the intensity measure of the point process* Φ. *We say that the point process* Φ *is simple if* $\mathbb{P}(\Phi(\{x\}) \leq 1, \forall x \in \mathbb{X}) = 1$ *and it is finite if* $\mathbb{P}(\Phi(\mathbb{X}) < \infty) = 1$.

In the rest of the thesis we will consider only simple point processes.

Definition 22 (Intensity function). *Let* Λ *be the intensity measure of a point process* Φ *on* $\mathbb{X}$ *satisfying for some non-negative measurable function* λ

$$\Lambda(B) = \int_B \lambda(x) dx, B \in \mathcal{B}(\mathbb{X}),$$

then λ *is called the intensity function of point process* Φ.

Theorem 6. *Let* Φ *be a point process on* $\mathbb{X}$ *with distribution* Q *and intensity measure* $\Lambda \in \mathbf{M}$. *Then exists a Markov kernel* P *from* $(\mathbb{X}, \mathcal{B}(\mathbb{X}))$ *to* $(\mathbf{M}, \mathcal{M})$ *such that*

$$\int_{\mathbf{M}} \int_{\mathbb{X}} f(x, \nu) \nu(dx) Q(d\nu) = \int_{\mathbb{X}} \int_{\mathbf{M}} f(x, \nu) P(x, d\nu) \Lambda(dx)$$

for arbitrary measurable function f *on* $\mathbb{X} \times \mathbf{M}$. *If there is different Markov kernel* P' *with this property, then* $\Lambda(\{x \in \mathbb{X} : P(x, \cdot) \neq P'(x, \cdot)\}) = 0$.

Proof. Theorem 4.2 in Chiu et al. [2013]. ☐

Definition 23 (Palm distribution). *Let* P *be a markov kernel from Theorem 6, the distribution* $P_x(\cdot) = P(x, \cdot)$ *is called a Palm distribution of the point process* Φ *in point* $x \in \mathbb{X}$.

Throughout the thesis we will be interested in point processes on Euclidean space, i.e., we set $\mathbb{X} = \mathbb{R}^d$, $d \in \mathbb{N}$. Remind that we shorten $\mathcal{B}(\mathbb{R}^d) = \mathcal{B}^d$.

Definition 24 (Stationary point process). *For each* $z \in \mathbb{R}^d$ *we denote* ϑ_z *the translation operator on* $\mathbf{N}$ *given by* $\vartheta_z \mu(B) = \mu(B - z)$, $\mu \in \mathbf{N}$, $B \in \mathcal{B}^d$. *A point process* Φ *on* $\mathbb{R}^d$ *is said to be stationary if* $\vartheta_z \Phi \sim^d \Phi$ *for all* $z \in \mathbb{R}^d$, *i.e., the distribution of the point process is invariant with respect to translations.*

Note that, the intensity function of a stationary point process is constant (in this case we speak just about intensity). A point process with constant intensity is usually called to be homogeneous.

Except of invariance with respect to translations, another important property of point processes on $\mathbb{R}^d$ is invariance with respect to rotations. By symbol $SO(d)$, $d \in \mathbb{N}$, we denote a d-dimensional rotation group. For details about the group and about various possible representations of rotations see Morawiec [2003]. Here we just write $\mathcal{O}$ for a rotation and we avoid to specify any particular representation of $\mathcal{O}$.

Definition 25 (Isotropic point process)**.** *Let* $\mathcal{O} \in SO(d)$ *we denote* $R_{\mathcal{O}}$ *the operator of rotation on* $\mathbf{N}$ *given by* $R_{\mathcal{O}}\mu(B) = \mu(\mathcal{O}B)$, $\mu \in \mathbf{N}$, $B \in \mathcal{B}^d$*. A point process* Φ *on* $\mathbb{R}^d$ *is said to be isotropic if* $R_{\mathcal{O}}\Phi \sim^d \Phi$ *for all rotations* $\mathcal{O}$*, i.e., the distribution of the point process is invariant with respect to rotations around origin.*

2.2.1 Poisson point process

Definition 26 (Poisson point process)**.** *Let* Λ *be a locally finite measure on* $\mathbb{X}$*. A point process* Φ *such that*

i. $\Phi(B)$ *has Poisson distribution with parameter* $\Lambda(B)$, $B \in \mathcal{B}_b(\mathbb{X})$,

ii. $\Phi(B_1), \dots, \Phi(B_n)$, $n \in \mathbb{N}$, *are independent for* $B_1, \dots, B_n \in \mathcal{B}_b(\mathbb{X})$ *pairwise disjoint,*

is called Poisson point process (PPP) with intensity measure Λ.

Let π_W^Λ denote the distribution of the restriction of a Poisson point process with intensity measure Λ on $W \in \mathcal{B}_b^d$. In homogeneous case we write π_W^z, where z is the intensity. If $z = 1$, then we write $\pi_W = \pi_W^1$ for short.

Definition 27 (Point process with density with respect to Poisson point process)**.** *Let* $p : \mathbf{N}^f \to \mathbb{R}^+$ *be a measurable function such that* $\int_{\mathbf{N}^f} p(\nu)\pi_W^\Lambda(d\nu) = 1$*. Then a point process* Φ *on* W *with distribution*

$$\mathbb{P}(\Phi \in A) = \int_A p(\nu)\pi_W^\Lambda(d\nu), A \in \mathcal{N}^f$$

is said be the (finite) point process with density p *with respect to Poisson point process with distribution* π_W^Λ.

To have a density with respect to Poisson point process with intensity measure Λ on W means to have a density w.r.t. Poisson measure given by

$$\pi_W^\Lambda(F) = e^{-\Lambda(W)}\left[\mathbb{I}_{\{F=\emptyset\}} + \sum_{n=1}^{\infty}\frac{1}{n!}\int\dots\int \mathbb{I}_{\{F=\{x_1,\dots,x_n\}\}}\Lambda(dx_1)\dots\Lambda(dx_n)\right], \tag{2.6}$$

where $F \in \mathcal{N}^f$.

Definition 28 (Papangelou conditional intensity)**.** *For a point process* Φ *with a density* p *we define Papangelou conditional intensity by*

$$\lambda^*(x,\nu) = \frac{p(\nu+\delta_x)}{p(\nu)}, x \in \mathbb{R}^d, \nu \in \mathbf{N}^f,$$

where we set $c/0 = 0$, $c \in \mathbb{R}_+$.

We can interpret the Papangelou conditional intensity of the point process Φ as the conditional probability that in an infinitesimal neighbourhood of some fixed point $x \in \mathbb{X}$, there will be a point of Φ, given we know the location of all points of Φ outside this neighbourhood.

2.2.2 Gibbs point process

A Gibbs point process (GPP) is a point process with a special form of the density with respect to finite Poisson point process. It allows to introduce interactions between the points. The interaction is described by an energy function on the space of configurations. In this section we will define Gibbs point process in two steps. In the first step we will state the finite volume case, i.e., the theory of Gibbs point processes on a bounded set $W \in \mathcal{B}_b^d$. Then the theory will be generalized to the whole space $W = \mathbb{R}^d$ and we will talk about infinite volume Gibbs point processes. Details regarding the theory of GPP can be found in Dereudre [2019].

Finite volume Gibbs point processes

Definition 29. *An energy function is a measurable function* $E : \boldsymbol{N}^f \to \mathbb{R} \cup \{\infty\}$*. The support of the energy function is* $\{\mathbf{x} \in \boldsymbol{N}^f : E(\mathbf{x}) < \infty\}$*. For a set* $W \in \mathcal{B}_b^d$ *we define*

$$E_W(\mathbf{x}) = E(\mathbf{x}) - E(\mathbf{x}_{W^c}),\ \mathbf{x}_W = \mathbf{x} \cap W,\ \mathbf{x} \in \boldsymbol{N}^f,$$

with the convention $\infty - \infty = 0$*.*

Roughly speaking the quantity $E_W(\mathbf{x})$ gives the energetic contribution of points $\mathbf{x}_W$ in the computation of the energy of $\mathbf{x}$.

Definition 30. *An energy function* E *is called*

(i) non-degenerate if $E(\emptyset) < \infty$*,*

(ii) hereditary if

$$\forall \mathbf{x} \in \boldsymbol{N}^f\ \forall y \in \mathbb{R}^d : E(\mathbf{x}) = \infty \Rightarrow E(\mathbf{x} + \delta_y) = \infty,$$

(iii) stable if there exists a constant K *such that for every* $\mathbf{x} \in \boldsymbol{N}^f$ *it holds* $E(\mathbf{x}) \geq K card(\mathbf{x})$*,*

(iv) stationary if for any $y \in \mathbb{R}^d$ *and* $\mathbf{x} \in \boldsymbol{N}^f$ *we have* $E(\tau_y \mathbf{x}) = E(\mathbf{x})$*.*

Definition 31. *An energy function* E *has a finite range* $R > 0$ *if for every* $W \in \mathcal{B}_b^d$ *and any* $\mathbf{x} \in \boldsymbol{N}^f$ *it holds*

$$E_W(\mathbf{x}) = E(\mathbf{x}_{W \oplus B(0,R)}) - E(\mathbf{x}_{W \oplus B(0,R) \setminus W}).$$

From now when talking about the energy function, we will assume that it is non-degenerate and stable. If E is nonnegative, it is stable with $K = 0$.

In many practical cases we assume only a bounded observation window $W \in \mathcal{B}_b^d$. Therefore we define Gibbs point process in the finite volume case. Moreover we will consider only the stationary reference Poisson point process π_W^z, $z > 0$. Recall that we write $\pi_W = \pi_W^1$.

Definition 32 (Gibbs point process)**.** *Let* W *be in* $\mathcal{B}_b^d$*. The finite volume Gibbs point process on* W *with activity* $z > 0$ *and the energy function* E *is a finite point process* Φ *with density with respect to* π_W *of the form*

$$p_W(\mathbf{x}) = \frac{1}{Z_W} z^{card(\mathbf{x}_W)} \exp\left(-E(\mathbf{x})\right),\ \mathbf{x} \in \boldsymbol{N}^f, \tag{2.7}$$

where

$$Z_W = \int z^{card(\mathbf{x}_W)} \exp\left(-E(\mathbf{x})\right)\pi_W(d\mathbf{x})$$

is a normalizing constant. Distribution P_W of Φ is called a (finite volume) Gibbs measure.

Note that the Definition 32 is correct - thanks to the non-degeneracy and stability of the energy function E the distribution P_W is well defined since the normalizing constant is positive and finite.

The local conditional distributions of Gibbs point process in any bounded window Δ given the configuration outside Δ can be described by DLR (Dobrushin, Lanford and Ruelle) equations:

Theorem 7. *For $W, \Delta \in \mathcal{B}_b^d$, $\Delta \subset W$, $\mathcal{L}(\Delta) > 0$ we have for P_W-a.a. $\mathbf{x}_{\Delta^c}$*

$$P_W(d\mathbf{x}_\Delta|\mathbf{x}_{\Delta^c}) = \frac{1}{Z_\Delta(\mathbf{x}_{\Delta^c})} z^{card(\mathbf{x}_\Delta)} \exp\left(-E_\Delta(\mathbf{x})\right)\pi_\Delta(d\mathbf{x}_\Delta),$$

where

$$Z_\Delta(\mathbf{x}_{\Delta^c}) = \int z^{card(\mathbf{x}_\Delta)} \exp\left(-E_\Delta(\mathbf{x})\right)\pi_\Delta(d\mathbf{x}_\Delta)$$

is a normalizing constant.

Proof. Proof can be found in [Dereudre, 2019], Proposition 3. ☐

A simple example of a Gibbs point process is the so called Strauss point process which is given by a density

$$p(\mathbf{x}) \propto \beta^{\mathrm{card}(\mathbf{x})}\gamma^{S_\delta(\mathbf{x})}, \mathbf{x} \in \mathbf{N}^f,$$

where $\beta > 0$, $0 \leq \gamma \leq 1$, $\delta > 0$ are parameters and $S_\delta(\mathbf{x}) = \sum_{j<k} \mathbb{I}_{[||x_j - x_k|| \leq \delta]}$. The limit case with $\gamma = 1$ reduces to Poisson point process with intensity measure $\beta\Lambda$. The case $\gamma = 0$ results in so called hard-core point process, where no two points of the process are closer than δ.

Strauss point process is a special case of a class of Gibbs point processes called multiscale point process [Penttinen, 1984]:

Definition 33 (Multiscale point process)**.** *Multiscale point process $\mathcal{M}_q$, $q \in \mathbb{N}$*

$$p(\mathbf{x}) \propto \beta^{\mathrm{card}(\mathbf{x})}\Pi_{i=1}^{q-1}\gamma_i^{\sum_{j<k}\mathbb{I}_{[\delta_{i-1}<\|x_j-x_k\|\leq\delta_i]}}, \mathbf{x} \in \mathbf{N}^f, \tag{2.8}$$

where $\beta > 0$, $0 \leq \gamma_1 \leq 1$, $\ldots$, $0 \leq \gamma_{q-1} \leq 1$ and $0 < \delta_1 < \ldots < \delta_{q-1}$ are unknown parameters and we set $\delta_0 = 0$ and $0^0 = 1$.

If $q = 1$ we interpret the right hand side in (2.8) as $\beta^{\mathrm{card}(\mathbf{x})}$. Note that $\mathcal{M}_1$ is just a Poisson point process with intensity measure $\beta\Lambda$ and $\mathcal{M}_2$ is just a Strauss point process.

The multiscale point process can be viewed as a special case of a class of pairwise interaction point processes. These processes constitute a flexible class of models for regularity. A point process Φ is called to be a pairwise interaction point process with homogeneous interaction functions ϕ (a function which is invariant with respect to shifts and rotations) if its distribution is given by a density with respect to Poisson point process, where the density is of the form

$$p(\mathbf{x}) \propto \beta^{\mathrm{card}(\mathbf{x})} \prod_{i<j} \phi(\|y_i - y_j\|), \mathbf{x} \in \mathbf{N}^f,$$

where $\beta > 0$ is a parameter and $\phi \geq 0$. In order to obtain a well-defined density, a further condition on the interaction function is needed; it suffices to assume that $0 \leq \phi \leq 1$ in which case normalizing constant is positive and finite.

Infinite volume Gibbs point processes

Under additional assumptions we can define an infinite volume Gibbs point process on the whole space $\mathbb{R}^d$.

Definition 34. *Let E be a stationary energy function. An infinite volume Gibbs point process with activity $z > 0$ and energy function E is a point process with (stationary) distribution P such that for any $W \in \mathcal{B}_b^d$ with $\mathcal{L}(W) > 0$ and for P-a.a. $\mathbf{x}_{W^c}$*

$$P(d\mathbf{x}_W|\mathbf{x}_{W^c}) = \frac{1}{Z_W(\mathbf{x}_{W^c})} z^{card(\mathbf{x}_W)} \exp\left(-E_W(\mathbf{x})\right)\pi_W(d\mathbf{x}_W), \tag{2.9}$$

where

$$Z_W(\mathbf{x}_{W^c}) = \int z^{\mathrm{card}(\mathbf{x}_W)} \exp\left(-E_W(\mathbf{x})\right)\pi_W(d\mathbf{x}_W)$$

is a normalizing constant. P is called a Gibbs measure.

The definition is correct if we can extend the definition of E_W from $\mathbf{N}^f$ to $\mathbf{N}$. This can be done for example when E has a finite range R, since then we obtain $E_W(\mathbf{x}) = E_W(\mathbf{x}_{W \oplus B(0,R)}) = E(\mathbf{x}_{W \oplus B(0,R)}) - E(\mathbf{x}_{W \oplus B(0,R) \setminus W})$, $\mathbf{x} \in \mathbf{N}$.

Definition 34 gives us the local conditional distributions of Gibbs point process in any bounded set W. The equations (2.9) are known as DLR equations. In terms of conditional densities $p(\mathbf{x}_W|\mathbf{x}_{W^c})$ with respect to the Poisson point process π_W the equations (2.9) are expressed as

$$p(\mathbf{x}_W|\mathbf{x}_{W^c}) = \frac{1}{Z_W(\mathbf{x}_{W^c})} z^{card(\mathbf{x}_W)} \exp\left(-E_W(\mathbf{x})\right). \tag{2.10}$$

Gibbs point process is a point process with Gibbs measure as its distribution. Gibbs measure is generally defined as a solution of DLR equations, which can be described as (2.9), but the existence and uniqueness of a such solution are non trivial questions.

The following theorem ensures the existence of infinite volume Gibbs measure.

Theorem 8. *Let E be a stationary energy function which is non-degenerate, hereditary, stable and having a finite range (see Definitions 30 and 31). Let $z > 0$. Then there exists infinite volume Gibbs measure with intensity z.*

Proof. Proof can be found in [Dereudre, 2019], Theorem 1. □

Some results about uniqueness and non-uniqueness of Gibbs measures can be found in [Dereudre, 2019], Sections 2.7 and 2.8.

Georgii-Nguyen-Zessin equation for GPP

In the previous section we defined infinite volume Gibbs point process using Dobrushin-Landford-Ruelle (DLR) equations, which describe complete conditional properties of the Gibbs measure. Alternatively, infinite volume Gibbs point process can be defined using Georgii-Nguyen-Zessin (GNZ) equations. The main advantage of GNZ equations is that they do not contain the normalizing constant which is in many cases difficult to evaluate. Further we interpret $E_W(\mathbf{x})$, $W \in \mathcal{B}_b^d$, $\mathbf{x} \in \mathbf{N}$, as the energy of $\mathbf{x}$ inside W given the configuration x_{W^c} outside W.

We say that energy function contains a hardcore part if $E_W(\mathbf{x}) = \infty$ for some $\mathbf{x} \in \mathbf{N}$ and some $W \in \mathcal{B}_b^d$. We denote by $\mathbf{N}_\infty$ the set of admissible configurations:

$$\mathbf{N}_\infty = \{\mathbf{x} \in \mathbf{N} : \forall W \in \mathcal{B}_b^d : E_W(\mathbf{x}) < \infty\}. \tag{2.11}$$

Definition 35 (Local energy)**.** *Let $\mathbf{x} \in \boldsymbol{N}$, $y \in \mathbb{R}^d$, $W \in \mathcal{B}_b^d$ containing y such that $E_W(\mathbf{x}) < \infty$. Then the local energy of y in $\mathbf{x}$ is defined by*

$$h(y, \mathbf{x}) = E_W(\mathbf{x} \cup \{y\}) - E_W(\mathbf{x}). \tag{2.12}$$

The definition of local energy of a point in a point pattern does not depend on the choice of W. The local energy function for Gibbs point process is the logarithm of the reciprocal conditional intensity λ^*,

$$\frac{1}{\lambda^*} = \frac{p_W(\mathbf{x}_W | \mathbf{x}_{W^c})}{p_W(\mathbf{x}_W \cup \{y\} | \mathbf{x}_{W^c})} = \frac{\exp\left(-E_W(\mathbf{x})\right)}{\exp\left(-E_W(\mathbf{x} \cup \{y\})\right)} = \exp\left(h(y, \mathbf{x})\right).$$

We will distinguish the hereditary and the non-hereditary case.

Hereditary case Assume now that we have energy function E fulfilling assumptions (i)-(iii) from definition 30, specially that it is hereditary.

Theorem 9. *Let P be a probability measure on $\boldsymbol{N}$. Let E be a finite range energy function and $z > 0$. Then P is an infinite volume Gibbs measure with energy function E and activity z if and only if for any positive measurable function $g : \mathbb{R}^d \times \boldsymbol{N} \to \mathbb{R}$ holds*

$$\int_{\boldsymbol{N}} \sum_{y \in \mathbf{x}} g(y, \mathbf{x} \setminus \{y\}) P(d\mathbf{x}) = z \int_{\boldsymbol{N}} \int_{\mathbb{R}^d} g(y, \mathbf{x}) e^{-h(y, \mathbf{x})} \mathcal{L}(dy) P(d\mathbf{x}) \tag{2.13}$$

Proof. Proof can be found in [Dereudre, 2019], Theorem 2. □

Non-hereditary case The formulation of Georgii-Nguyen-Zessin equation in Theorem 9 is unfortunately not valid in general for a non-hereditary Gibbs point processes (i.e., for GPP with non-hereditary energy function involving some hardcore part). Let us now explore the non-hereditary case and state a generalization of Theorem 9.

The main problem, in non-hereditary case, is that the energy of a point y in a configuration $\mathbf{x}$ is not always defined. Indeed, the energy of $\mathbf{x} - \delta_y$ may be locally infinite, even if the energy of $\mathbf{x}$ is locally finite. In this case, the energy of y in $\mathbf{x}$ would be minus infinity, which makes no sense. Therefore the concept of removable points is introduced.

Definition 36. *Let $\mathbf{x} \in \boldsymbol{N}$ and y be a point of $\mathbf{x}$. We say, that y is removable from $\mathbf{x}$ if*

$$\exists W \in \mathcal{B}_b^d \text{ such that } y \in W \text{ and } E_W(\mathbf{x} - \delta_y) < \infty. \tag{2.14}$$

We denote the set of removable points in $\mathbf{x}$ by $\mathcal{R}(\mathbf{x})$.

Theorem 10. *If $\mathbf{x}$ is in $\boldsymbol{N}_\infty$ and y is a point of $\mathbf{x}$, then y is removable from $\mathbf{x}$ if and only if $\mathbf{x} - \delta_y$ is in $\boldsymbol{N}_\infty$.*

Proof. Proof can be found in [Dereudre and Lavancier, 2009], Proposition 1. □

Now, the definition of the local energy should be adjusted for removable points.

Definition 37. *Let y be a removable point in a configuration $\mathbf{x} \in \boldsymbol{N}$. The local energy of y in $\mathbf{x} - \delta_y$ is defined as*

$$h(y, \mathbf{x} - \delta_y) = E_W(\mathbf{x}) - E_W(\mathbf{x} - \delta_y),$$

where W is a bounded set containing y such that $E_W(\mathbf{x} - \delta_y)$ is finite.

Note that, the definition of local energy of removable point does not depend on the choice of W.

Finally, the GNZ equation can be generalized to non-hereditary Gibbs point processes.

Theorem 11. *Let P be a Gibbs measure with intensity $z > 0$. Then for every bounded non-negative measurable function $g : \mathbb{R}^d \times \boldsymbol{N} \to \mathbb{R}$ it holds*

$$\int_{\boldsymbol{N}} \sum_{y \in \mathcal{R}(\mathbf{x})} g(y, \mathbf{x} \setminus \{y\}) P(d\mathbf{x}) = z \int_{\boldsymbol{N}} \int_{\mathbb{R}^d} g(y, \mathbf{x}) e^{-h(y,\mathbf{x})} \mathcal{L}(dy) P(d\mathbf{x}). \tag{2.15}$$

Proof. Proof can be found in [Dereudre and Lavancier, 2009], Proposition 2. □

The converse implication is not true, i.e., (2.15) does not characterize the measure P. Consider, for example, $\mathbf{x}$ P-almost surely not containing any removable points, then (2.15) becomes the trivial equation $0 = 0$. The equation is interesting only in case that $\mathbf{x}$ contains, P-almost surely, some removable points. Equation (2.15) in the hereditary case becomes the classical Georgii-Nguyen-Zessin equation (2.13).

2.2.3 Marked point processes

Since generators of Laguerre tessellations are marked point patterns we would like to add the concept of marks into the theory of point processes.

Definition 38 (Marked point process)**.** *Let $\mathbb{M}$ be a mark space. A marked point process is a simple point process Φ^m on space $\mathbb{X} \times \mathbb{M}$ such that its intensity measure Λ_m fulfills $\Lambda_m(B \times \mathbb{M}) < \infty$ for every $B \in \mathcal{B}_b(\mathbb{X})$. For every marked point process we consider unmarked point process Φ given by $\Phi(B) = \Phi^m(B \times \mathbb{M})$, $B \in \mathcal{B}(\mathbb{X})$.*

Definition 39. *A marked point process Φ^m on $\mathbb{R}^d$ with mark space $\mathbb{M}$ is stationary if its distribution is invariant under translations $\vartheta_y(\mu^m) = \mu^m - y := \{(x-y,m) : (x,m) \in \mu^m\}$, $\mu^m \in \mathbf{N}$, $y \in \mathbb{R}^d$, and isotropic if its distribution is invariant under rotations around origin $(R_{\mathcal{O}}\mu^m)(B \times L) = \mu^m(\mathcal{O}B \times L)$, $\mu^m \in \mathbf{N}$, $B \in \mathcal{B}^d$, $L \in \mathcal{B}(\mathbb{M})$.*

Using Lemma 1, a marked point process can be expressed as a countable sum of Dirac measures where the atoms are numbered in a measurable way - for $\mu^m \in \mathbf{N}$ there exist measurable mappings $\zeta_i : \mathbf{N} \to \mathbb{X} \times \mathbb{M}, \mu^m \mapsto (x_i, m_i)$ such that

$$\mu^m = \sum_{i=1}^{\mu(\mathbb{X})} \delta_{(x_i,m_i)}.$$

Theorem 12. *Let Φ^m be a stationary marked point process with a finite intensity $\lambda > 0$. Then there exists a unique probability measure $\mathbb{Q}$ on $\mathbb{M}$ such that the intensity measure of Φ^m is of the form*

$$\Lambda_m(B \times L) = \lambda \mathcal{L}(B)\mathbb{Q}(L), \; B \in \mathcal{B}^d, \; L \in \mathcal{B}(\mathbb{M}).$$

Proof. See Section 4.2.2 in Chiu et al. [2013]. □

Definition 40 (Mark distribution)**.** *A probability measure $\mathbb{Q}$ from Theorem 12 is called mark distribution.*

Definition 41. *We say that the intensity measure Λ of a point process Φ is diffuse, if $\Lambda(\{y\}) = 0$ holds for any $y \in \mathbb{X}$.*

Definition 42 (Poisson marked point process)**.** *Let Φ^m be Poisson point process on $\mathbb{X} \times \mathbb{M}$ with diffuse (non-atomic) intensity measure Λ_m such that $\Lambda_m(B \times \mathbb{M}) < \infty$ for every $B \in \mathcal{B}_b(\mathbb{X})$. Then Φ^m is called Poisson marked point process with intensity measure Λ_m.*

Definition 43 (Independent marking)**.** *A marked point process Φ^m is called independently marked if the random marks $\{m_i\}$ are independent, identically distributed and independent of the unmarked point process Φ.*

Let us consider a marked point process Φ^m which is independently marked and let the appropriate unmarked point process Φ be a Poisson point process. Then Φ^m is a Poisson marked point process.

Definition 44 (Geostatistical marking)**.** *Let Φ be a simple point process on $\mathbb{R}^d$ and $\{Z(x) : x \in \mathbb{R}^d\}$ be a random field with values in the space of marks $\mathbb{M}$ independent of Φ. Then $\Phi^m = \sum_{x \in supp\Phi} \delta_{x,Z(x)}$ is geostatistically marked point process.*

Geostatistical marking is one of possible approaches admitting dependences between marks, but it still assumes independence of marks and locations.

In the rest of the thesis, we will consider only marked point processes with simple unmarked counterparts on Euclidean space, i.e., $\mathbb{X} = \mathbb{R}^d$, $d \in \mathbb{N}$, and real marks, i.e., $\mathbb{M} \subset \mathbb{R}$. We introduce a formal definition of finite marked Gibbs point process in a similar way to Definition 32. Further definitions and statements of Section 2.2.2 can be extended to the marked case analogically.

Definition 45. *Let $W \in \mathcal{B}_b^d$. Let $\pi_W^{\Lambda_m}$ be distribution of the marked Poisson point process with intensity measure $\Lambda_m(B \times L) = z\mathcal{L}(B)\mathbb{Q}(L)$, $B \in \mathcal{B}^d$, $L \in \mathcal{B}(\mathbb{M})$, which is restricted on W. The finite volume marked Gibbs point process on W with activity $z > 0$, the energy function E and mark distribution $\mathbb{Q}$ is a finite point process Φ with density with respect to π_W (distribution of marked PPP with intensity* 1*) of the form*

$$p_W(\mathbf{x}, \boldsymbol{m}) = \frac{1}{Z_W} z^{card(\mathbf{x}_W)} \exp\left(-E(\mathbf{x}, \boldsymbol{m})\right), (\mathbf{x}, \boldsymbol{m}) \in \boldsymbol{N}^f, \tag{2.16}$$

where

$$Z_W = \int z^{card(\mathbf{x}_W)} \exp\left(-E(\mathbf{x}, \boldsymbol{m})\right)\pi_W(d(\mathbf{x}, \boldsymbol{m}))$$

is a normalizing constant.

Proceeding to an infinite volume marked Gibbs measure P, the GNZ equation (2.13) becomes

$$\begin{aligned} \int_{\mathbf{N}} \sum_{(y,m)\in(\mathbf{x},\mathbf{m})} g((y, m), (\mathbf{x}, \mathbf{m}) \setminus \{(y, m)\})P(d(\mathbf{x}, \mathbf{m})) = \\ = z \int_{\mathbf{N}} \int_{\mathbb{R}^d \times \mathbb{M}} g((y, m), (\mathbf{x}, \mathbf{m}))e^{-h((y,m),(\mathbf{x},\mathbf{m}))}\mathcal{L}(dy)\mathbb{Q}(dm)P(d(\mathbf{x}, \mathbf{m})). \end{aligned} \tag{2.17}$$

2.2.4 Summary characteristics

In this part we focus on various summary characteristics of (marked) homogeneous point processes on $\mathbb{R}^d$. For a broader overview of point process and mark characteristics see [Baddeley et al., 2015] and [Illian et al., 2008].

Point process characteristics

First, consider unmarked point process Φ on $\mathbb{R}^d$. If Φ is a homogeneous point process, Palm distributions $P_x(\cdot)$, $x \in \mathbb{R}^d$, from Definition 23 are determined by Palm distribution at origin $\mathbf{o}$ and by relation $P_x(\cdot) = P_{\mathbf{o}}(\vartheta_x^{-1}(\cdot))$. To describe mutual positions of the points of Φ the following summary characteristics can be used.

Definition 46 (F function)**.** *Let $D_{\mathbf{o}}$ be distance from the origin $\mathbf{o}$ to the nearest point of Φ. Spherical contact distribution function (F function) is given by*

$$F(t) = \mathbb{P}(\Phi(b(\mathbf{o}, t)) > 0) = \mathbb{P}(D_{\mathbf{o}} \leq t), t > 0.$$

Definition 47 (G function)**.** *Nearest-neighbour distance distribution function (G function) is given by*

$$G(t) = P_{\mathbf{o}}(\{\nu \in \mathbf{N} : \nu(b(\mathbf{o}, t)) > 1\}), t > 0.$$

Definition 48 (K function)**.** *Let* $0 < \lambda < \infty$*. Reduced second-order moment function (Ripley's K function) is given by*

$$\lambda K(t) = \mathsf{E}^!_{\mathbf{o}}\, \Phi(b(\mathbf{o}, t)) = \int_{\mathbf{N}} \nu(b(\mathbf{o}, t) \setminus \{\mathbf{o}\}) P_{\mathbf{o}}(d\nu), t > 0.$$

L function is a tranformation of K function given by

$$L(t) = \left(\frac{K(t)}{\omega_d}\right)^{1/d}, t > 0,$$

where $\omega_d = \frac{\pi^{d/2}}{\Gamma(1+d/2)}$ *is volume of the d-dimensional unit sphere in* $\mathbb{R}^d$*.*

Definition 49 (Pair correlation function)**.** *Pair correlation function is given by*

$$g(t) = \frac{K'(t)}{\sigma_d t^{d-1}} \text{ for } t > 0,$$

where K' *is the derivative of the K function and* $\sigma_d = d\omega_d = \frac{2\pi^{d/2}}{\Gamma(d/2)}$ *is surface of the* $(d-1)$*-dimensional unit sphere in* $\mathbb{R}^d$*.*

Theorem 13. *In case of homogeneous Poisson point process with intensity* λ *we obtain* $F(t) = G(t) = 1 - \exp(-\lambda\omega_d t^d)$*,* $L(t) = t$ *and* $g(t) = 1$ *for all* $t > 0$*.*

Proof. See Section 2.3 in Chiu et al. [2013]. ☐

Mark characteristics

Given a marked point process Φ^m, we are now interested in a mutual relation between points and marks.

Definition 50 (Mark correlation function)**.** *Mark correlation function is given by*

$$k_{mm}(t) = \frac{\mathsf{E}_{\mathbf{o},\mathbf{t}}[m(\mathbf{o})m(\mathbf{t})]}{\mathsf{E}[m(\mathbf{o})]\,\mathsf{E}[m(\mathbf{t})]}, t > 0,$$

where $m(\mathbf{o}), m(\mathbf{t})$ *are marks of the points* $\mathbf{o}, \mathbf{t} \in \mathbb{R}^d$*,* $||\mathbf{o} - \mathbf{t}|| = t$ *and* $\mathsf{E}_{\mathbf{o},\mathbf{t}}$ *is the conditional expectation, where we condition on the presence of points of the marked point process at* $\mathbf{o}$ *and* $\mathbf{t}$*.*

The mark correlation function $k_{mm}(t)$ characterizes aspects of the correlation of marks at distance t. Values of $k_{mm}(t)$ larger than 1 indicate that the product of marks of point pairs of a distance t tends to be larger than the squared mean mark.

Definition 51 (Mark variogram)**.** *Mark variogram is given by*

$$\gamma_m(t) = \frac{1}{2}\, var_{\mathbf{o},\mathbf{t}}[m(\mathbf{o}) - m(\mathbf{t})] = \frac{1}{2}\, \mathsf{E}_{\mathbf{o},\mathbf{t}}[m(\mathbf{o}) - m(\mathbf{t})]^2, t > 0,$$

where $\mathsf{E}_{\mathbf{o},\mathbf{t}}$ *and* $var_{\mathbf{o},\mathbf{t}}$ *are the conditional expectation and variance, where we condition on the presence of points of the marked point process at* $\mathbf{o}$ *and* $\mathbf{t}$*,* $||\mathbf{o} - \mathbf{t}|| = t$*,* $\mathbf{o}, \mathbf{t} \in \mathbb{R}^d$*.*

The mark variogram $\gamma_m(t)$ tends for large t towards the mark variance σ_μ^2. Small values for small t indicate similarity of marks of generators close together, but not that the marks are small.

The simplest form of marking a point process is independent marking, where the points get their marks by independent sampling from a mark distribution. In the case of independent marking it is $k_{mm}(t) \equiv 1$ and $\gamma_m(t) \equiv \sigma_\mu^2$. A marking a bit more complicated is geostatistical marking. There are two independent objects, the non-marked starting point process and a stationary and isotropic random field $\{Z(x)\}$. The mark of point x_i is then $Z(x_i)$. In geostatistically marked point processes the marks of points close together can be similar, but a joint tendency that the marks of points close together are similar and small is impossible. The corresponding mark correlation function and the mark variogram satisfy the equation

$$k_{mm}(t) = 1 + \frac{\sigma_\mu^2}{\mu^2} - \frac{\gamma_m(t)}{\mu^2}, \tag{2.18}$$

where μ and σ_μ^2 are mean and variance of the mark distribution.

2.2.5 Estimation of summary characteristics and boundary effects

When estimating the summary characteristics of (marked) point patterns and tessellations, it is neccessary to treat so called edge or boundary effects. These effects arise as we observe the marked point pattern/tessellation $(\mathbf{x}, \mathbf{r})$ only in a window $W \in \mathcal{B}_b^d$, whereas we do not know what is outside. Here, we show how to estimate the summary characteristics mentioned in previous sections together with the most frequent ways how to overcome the undesired boundary effects, for details see for example Baddeley et al. [2015] and Illian et al. [2008].

Estimate of F function with Kaplan-Meier correction is

$$\widehat{F(t)} = 1 - \prod_{s \le t} \left(1 - \frac{\operatorname{card}\{x \in I_a \cap W : d(x) = s, d(x) \le c(x)\}}{\operatorname{card}\{x \in I_a \cap W : d(x) \ge s, c(x) \ge s\}} \right),$$

where $I_a = y + a\mathbb{Z}^d = \{(y_1 + a_1 z_1, \dots, y_d + a_d z_d) \in \mathbb{R}^d : z_i \in \mathbb{Z}\}$, $y \in \mathbb{R}^d$, $a \in \mathbb{R}_+^d$, i.e., $a_i > 0$ for $i = 1, \dots, d$, is a regular lattice in $\mathbb{R}^d$. Further $d(x) = d(x, \mathbf{x})$ is a distance of x to its nearest point in $\mathbf{x}$ and $c(x) = d(x, \partial W)$ is a distance of x to boundary of the window W.

Estimate of G function with Kaplan-Meier correction is

$$\widehat{G(t)} = 1 - \prod_{s \le t} \left(1 - \frac{\operatorname{card}\{x \in \mathbf{x} \cap W : e(x) = s, e(x) \le c(x)\}}{\operatorname{card}\{x \in \mathbf{x} \cap W : e(x) \ge s, c(x) \ge s\}} \right),$$

where $e(x) = d(x, \mathbf{x} \setminus \{x\})$ is a distance of x to its nearest neighbour in $\mathbf{x}$ and $c(x) = d(x, \partial W)$ is a distance of x to boundary of the window W.

Estimate of K function with isotropic correction is obtained from

$$\widehat{\lambda^2 K(t)} = \sum_{x, y \in \mathbf{x} \cap W}^{\neq} \frac{\mathbb{I}_{[||x-y|| \le t]}}{|W|} \frac{|\partial b(x, ||x-y||)|}{|\partial b(x, ||x-y||) \cap W|},$$

as $\widehat{K(t)} = \frac{\widehat{\lambda^2 K(t)}}{\widehat{\lambda^2}}$, where the estimate of λ^2 is

$$\widehat{\lambda^2} = \frac{\mathbf{x}(W)(\mathbf{x}(W)-1)}{|W|^2},$$

with $\mathbf{x}(W) = \mathrm{card}(\mathbf{x}_W)$ being the number of points of $\mathbf{x}$ in W.

Estimate of pair correlation function with isotropic correction is obtained in a similar way,

$$\widehat{g(t)} = \frac{1}{\widehat{\lambda^2}} \sum_{x,y \in \mathbf{x} \cap W}^{\neq} \frac{k_b(t - ||x-y||)}{\sigma_d t^{d-1} |W|} \frac{|\partial b(x, ||x-y||)|}{|\partial b(x, ||x-y||) \cap W|}, \tag{2.19}$$

where $k_b(\cdot)$ is a kernel function with bandwidth b.

Note that Kaplan-Meier estimates are not unbiased. The estimate of K function with isotropic correction is ratio-unbiased for $t < t_0 = \inf\{u > 0 : |W^{(u)}| < |W|\}$, where $W^{(t)} = \{x \in W : \partial b(x,t) \cap W \neq \emptyset\}$, since $\widehat{\lambda^2}$ is unbiased estimate of λ^2 and $\widehat{\lambda^2 K(t)}$ is unbiased estimate of $\lambda^2 K(t)$ for $t < t_0$. The estimate of pair correlation function is not unbiased, the choice of bandwidth influences the bias and variability of the estimate – for excessively small bandwidth, the estimates $\widehat{g(t)}$ will have a high variance.

When estimating mark correlation function and mark variogram with isotropic correction one first computes

$$\widehat{\lambda_f^{(2)}(t)} = \sum_{x,y \in \mathbf{x} \cap W}^{\neq} \frac{f(m(x), m(y)) k_b(t - ||x-y||)}{\sigma_d t^{d-1} |W|} \frac{|\partial b(x, ||x-y||)|}{|\partial b(x, ||x-y||) \cap W|}$$

and

$$\widehat{\lambda^{(2)}(t)} = \sum_{x,y \in \mathbf{x} \cap W}^{\neq} \frac{k_b(t - ||x-y||)}{\sigma_d t^{d-1} |W|} \frac{|\partial b(x, ||x-y||)|}{|\partial b(x, ||x-y||) \cap W|}.$$

Estimate of mark correlation function is then

$$\widehat{k_{mm}(t)} = \frac{\widehat{\lambda_f^{(2)}(t)}}{\left(\frac{1}{\mathbf{x}(W)} \sum_{x \in \mathbf{x} \cap W} m(x)\right)^2 \widehat{\lambda^{(2)}(t)}}$$

with $f(m_1, m_2) = m_1 m_2$.

Estimate of mark variogram is

$$\widehat{\gamma_m(t)} = \frac{\widehat{\lambda_f^{(2)}(t)}}{\widehat{\lambda^{(2)}(t)}}$$

with $f(m_1, m_2) = \frac{1}{2}(m_1 - m_2)^2$.

The estimates of mark correlation function and mark variogram are not unbiased, again the choice of bandwidth influences the bias and variability of the estimate.

2.2.6 Connection to tessellations

In Section 2.1 we have introduced several deterministic tessellation models, all of them being determined by a locally finite (possibly marked) point pattern in sense that each (marked) point determines a cell. Now, we define a random tessellation and show how it can be identified with a (marked) point process.

Definition 52. *Let us denote* $\boldsymbol{T}$ *the set of all tessellations in* $\mathbb{R}^d$ *and equip it with* σ*-algebra* $\mathcal{T} = \sigma(\{A \subset \boldsymbol{T} : \partial A \cap K \neq \emptyset\} : K \in \mathcal{K})$*, where* ∂A *means the union of the boundaries of the cells* $C \in A$ *A random tessellation in* $\mathbb{R}^d$ *is a random variable on* $(\Omega, \mathcal{A}, \mathbb{P})$ *with values in* $(\boldsymbol{T}, \mathcal{T})$*. It is called normal or face-to-face if its realizations are almost surely normal or face-to-face.*

Note that detailed introduction of σ-algebra $\mathcal{T}$ on $\mathbf{T}$ can be found in [Møller, 1989].

According to this definition a random tessellation is a random variable with value in the measurable space of all tessellations $(\mathbf{T}, \mathcal{T})$. Considering that a tessellation is generated by a locally finite (marked) point pattern, then its random counterpart can be viewed as a random variable with value in measurable space of locally finite (marked) point patterns, i.e., as a marked point process. E.g., combining stationary Poisson point processes with Voronoi tessellation we obtain Poisson-Voronoi tessellations. In this simple case, closed analytical formulas are available for the moments of geometrical characteristics of cells, see Section (2.2.4), such as volume, number of faces, surface area, etc., cf. Okabe et al. [2000].

Theorem 14. *Let* Φ *be a stationary marked Poisson process on* $\mathbb{R}^d$ *with intensity* $\lambda > 0$ *and mark distribution* $\mathbb{Q}$*. Suppose* R *is a random variable with distribution* $\mathbb{Q}$*. Then the following statements are equivalent:*

- *(i) The Laguerre tessellation of* Φ *exists, i.e.,* $min_{(x,r)\in\Phi}||x-z||^2 - r^2$ *almost surely exists for all* $z \in \mathbb{R}^d$*.*
- *(ii) We have* $\mathsf{E}[R^d] < \infty$*.*

Proof. The proof can be found in Lautensack [2007], Proposition 3.1.4. □

The random tessellation from Theorem 14 is reffered to as Poisson-Laguerre tessellation, more details about this class of models can be found in Lautensack [2007] and Lautensack and Zuyev [2008]. Similarly one can obtain Gibbs-Laguerre tessellations, but the existence is not so straightforward anymore, since on top of mark distribution we have to care about existence of the underlying Gibbs point process itself. In Dereudre et al. [2012] they deal with the existence of Gibbs measures with energy functions based on a tessellation geometry. In particular, the Gibbs-Voronoi tessellations in $\mathbb{R}^2$ are studied in Dereudre and Lavancier [2011].

2.3 Simulation algorithms

In this section we describe three algorithms which will be used for simulation of Laguerre tessellations later.

2.3.1 Metropolis-Hastings birth-death-move algorithm

Metropolis-Hastings birth-death-move algorithm can be used to simulate realizations of (marked) point processes with density with respect to Poisson point process. As we focus on Laguerre tessellations and Gibbs models in this thesis, we formulate the algorithm in the context of marked point patterns $(\mathbf{x}, \mathbf{r})$ with $\mathbf{x} \in \mathbf{N}_W$, $W \in \mathcal{B}_b^3$, and density $p(\mathbf{x}, \mathbf{r}) \propto z^{\operatorname{card}(\mathbf{x})} \exp(-E(\mathbf{x}, \mathbf{r}))$. Except Laguerre generators $(\mathbf{x}, \mathbf{r})$, the algorithm will be used for simulation of unmarked point patterns $\mathbf{x}$, the adaption of the algorithm to the unmarked point pattern $\mathbf{x}$ is straightforward. The general form of the MCMC Metropolis-Hastings birth-death-move algorithm is described in Møller and Waagepetersen [2003] and also in Geyer and Møller [1994].

Recall that admissibility, Definition 58, of a finite point configuration $(\mathbf{x}, \mathbf{r})$ means that its energy $E(\mathbf{x}, \mathbf{r})$ is finite. Let p be defined as in equation (3.9). The evolution step of the MHBDM algorithm for an admissible $(\mathbf{x}_0, \mathbf{r}_0) \in W \times \mathbb{M}$ with $n = \operatorname{card}(\mathbf{x}_0)$ can be written as follows.

Algorithm 1 (evolution step of MHBDM).
do one of the following (with probability $\frac{1}{3}$ each):

(a) "birth": generate a point y uniformly in W (i.e., $y \sim Unif(W)$) and a radius $s \sim Unif(\mathbb{M})$ and set

$$(\mathbf{x}_1, \mathbf{r}_1) = \begin{cases} (\mathbf{x}_0, \mathbf{r}_0) \cup \{(y, s)\} & \text{with probability } \min\left(1, \frac{p((\mathbf{x}_0, \mathbf{r}_0) \cup \{(y,s)\})}{(n+1) p(\mathbf{x}_0, \mathbf{r}_0)}\right), \\ (\mathbf{x}_0, \mathbf{r}_0) & \text{otherwise;} \end{cases}$$

(b) "death": choose a point (x, r) from $\mathbf{x}_0$ at random and set

$$(\mathbf{x}_1, \mathbf{r}_1) = \begin{cases} (\mathbf{x}_0, \mathbf{r}_0) \setminus \{(x, r)\} & \text{with probability } \min\left(1, \frac{n p((\mathbf{x}_0, \mathbf{r}_0) \setminus \{(x,r)\})}{p(\mathbf{x}_0, \mathbf{r}_0)}\right), \\ (\mathbf{x}_0, \mathbf{r}_0) & \text{otherwise;} \end{cases}$$

(c) "move": choose a point (x, r) from $\mathbf{x}_0$ at random and generate $y \sim N_3(x, \Sigma)$ with the covariance matrix Σ, $s \sim N^T(r, \sigma_R^2)$ and set

$$(\mathbf{x}_1, \mathbf{r}_1) = \begin{cases} ((\mathbf{x}_0, \mathbf{r}_0) \setminus \{(x, r)\}) \cup \{(y, s)\} & \text{with probability} \\ & \min\left(1, \frac{p(((\mathbf{x}_0, \mathbf{r}_0) \setminus \{(x,r)\}) \cup \{(y,s)\})}{p(\mathbf{x}_0, \mathbf{r}_0)}\right), \\ (\mathbf{x}_0, \mathbf{r}_0) & \text{otherwise.} \end{cases}$$

Here,

$$\Sigma = diag\{\sigma^2, \sigma^2, \sigma^2\}, \ \sigma > 0, \tag{2.20}$$

$Unif$ denotes the uniform distribution, N_3 denotes the trivariate Gaussian distribution and N^T is the truncated Gaussian distribution on $\mathbb{M}$. All proposals are sampled independently of each other. Using the Gaussian distribution for the move proposal distribution is a common choice, cf. Dereudre and Lavancier [2011]. The constants $\sigma, \sigma_R > 0$ of the proposal distributions need to be chosen carefully. First, note that in "move" step of the algorithm, the point y can always be considered to belong to W; if $y \notin W$ then the periodic image of y in W is taken. The second observation is that the acceptance ratios H_k, $k =$ "birth", "death", "move", in the steps 1(a),(b),(c), respectively, are of the form

$$H_k = c_k \exp(E_b - E_a), \tag{2.21}$$

where E_a is the energy of the proposal (e.g., $\mathbf{x}_0 \cup \{(y, s)\}$ in step 1(a)), E_b is the energy of $\mathbf{x}_0$ and the constant c_k is equal to $\frac{z}{n+1}$ in step 1a, $\frac{n}{z}$ in step 1(b) and 1 in step 1(c). Finally, note that the uniform and Gaussian distributions in Algorithm 1 can be replaced by any other reasonable distribution.

The Metropolis-Hastings birth-death-move (MHBDM) algorithm is summarized in Algorithm 2.

Algorithm 2 (MHBDM).

1. construct an admissible marked point configuration $(\mathbf{x}_0, \mathbf{r}_0)$,
2. $n \leftarrow \text{card}(\mathbf{x}_0)$,
3. run the *Algorithm 1* (taking $(\mathbf{x}_0, \mathbf{r}_0)$ and yielding $(\mathbf{x}_1, \mathbf{r}_1)$),
4. $(\mathbf{x}_0, \mathbf{r}_0) \leftarrow (\mathbf{x}_1, \mathbf{r}_1)$,
5. repeat steps 2. to 4. $(S - 1)$ times,
6. return $(\mathbf{x}_0, \mathbf{r}_0)$.

Clearly, the number of iterations S depends on the considered model; more complex models tend to require more iterations in *Algorithm 2* to approach the target distribution. The convergence of the basic MHBDM algorithm was proved under mild conditions in Møller and Waagepetersen [2003], Section 7.3. As claimed in Dereudre and Lavancier [2011], Section 3.2, for models with hardcore potentials the convergence of Algorithm 2 is difficult to prove in cases when the tessellation model becomes too rigid. Further discussion concerning the hardcore case can be found in Dereudre and Lavancier [2011].

2.3.2 Metropolis within Gibbs algorithm

The Metropolis within Gibbs algorithm, a composition of Gibbs sampler and Metropolis-Hastings algorithm, will be used for simulation of radii $\mathbf{r}$ from fully conditional distributions $R_i \mid \mathbf{r}_{-i}$, where $\mathbf{r}_{-i} = (r_1, \ldots, r_{i-1}, r_{i+1}, \ldots, r_n)$, $i = 1, \ldots, n$. The cardinality $\text{card}(\mathbf{r}) = n$ is fixed. For a detailed description of both algorithms see Møller and Waagepetersen [2003]. The outer algorithm – Gibbs sampler, assumes that we are able to simulate from all fully conditional distributions.

Algorithm 3 (Gibbs sampler).

1. choose initial state $\mathbf{r}_0 = (r_{0,1}, \ldots, r_{0,n})$, set $t = 0$,
2. simulate $r_{t+1,n}$ from conditional distribution $R_n \mid r_{t,1}, \ldots, r_{t,n-1}$,
   simulate $r_{t+1,n-1}$ from conditional distribution $R_{n-1} \mid r_{t,1}, \ldots, r_{t,n-2}, r_{t+1,n}$,
   $\ldots$
   simulate $r_{t+1,1}$ from conditional distribution $R_1 \mid r_{t+1,2}, \ldots, r_{t+1,n}$,

3. if $t+1 < T$ then increase t and go to 2., else return $\mathbf{r}_t$.

Algorithm 3 describes the so called reversed order Gibbs sampler. In contrast to sampling in increasing order, the reversed order sampling gives us a reversible Markov chain.

If we are not able to simulate directly from fully conditional distributions in Algorithm 3, we can use Metropolis-Hastings algorithm to do so. Such combination of both algorithms is known as Metropolis within Gibbs algorithm. A description of the inner algorithm - Metropolis-Hastings, related to the case of radii follows.

The distributions $R_i \mid \mathbf{r}_{-i}$, $i = 1, \dots, n$, are target distributions for Metropolis-Hastings algorithm. For a given i, $i = 1, \dots, n$, we will show how to simulate from the appropriate target distribution, which we assume to have a density p with respect to some σ-finite reference measure ν on the measurable space $(\mathbb{M}, \mathcal{B}(\mathbb{M}))$ (in our case we can set $\nu = \mathcal{L}$). Let Q be a Markov kernel on $\mathbb{M}$, $Q(r, ds) = q(r, s)\nu(ds)$ for some q called a proposal density. It determines a proposal probability of a transition from state r to s, where $r, s \in \mathbb{M}$. Set

$$\alpha(r, s) = \begin{cases} \min\left(\frac{p(s)q(s,r)}{p(r)q(r,s)}, 1\right) & \text{for } p(r)q(r,s) > 0, \\ 1 & \text{otherwise.} \end{cases}$$

For $r, s \in \mathbb{M}$, $\alpha(r, s)$ is the proposal acceptance probability of the transition from r to s.

Algorithm 4 (Metropolis-Hastings algorithm).

1. choose r_0, set $t = 0$,
2. generate s from $Q(r_t, \cdot)$ and set

$$r_{t+1} = \begin{cases} s & \text{with probability } \alpha(r_t, s), \\ r_t & \text{otherwise;} \end{cases}$$

3. if $t+1 < T$ then increase t and go to 2., else return r_t.

If $q(r, s) = q(s, r)$, i.e., symmetric proposal density, we do not need to evaluate the proposal density and the algorithm is called shortly Metropolis algorithm. In general, the proposal density needs to be carefully specified – first, we need to be able to visit the whole space of radii, second, if the proposal density is poorly chosen, either the acceptance rate is low, or the Markov chain moves throughout the support of the invariant distribution too slow (even could be stuck around one place). It is recommended to achieve roughly 30% acceptance rate.

2.4 Statistical methods

2.4.1 Maximum pseudolikelihood

Inaccessability of the normalizing constant, a common problem for models with a complex density, causes difficulties in maximum likelihood estimation as the normalizing constant needs to be accurately estimated first. In case of Gibbs point processes, this can be done using Monte Carlo methods. Unfortunately, such approaches can be computationally very demanding. Therefore alternative ways to avoid calculating the intractable normalizing constants were developed. Among them the idea of maximum pseudolikelihood is considered in [Besag, 1974, 1977, Besag et al., 1982].

Definition 53 (Pseudolikelihood)**.** *Given random variables* $X_1, \dots, X_n$ *with the distribution depending on an unknown parameter* $\theta \in \Theta \subseteq \mathbb{R}^q$, $q \in \mathbb{N}$, *the pseudolikelihood of a realization* $\mathbf{x} = \{x_1, \dots, x_n\}$ *is*

$$\mathcal{PL}(\theta; \mathbf{x}) = \prod_{j=1}^{n} p_\theta(x_j \mid \mathbf{x}_{-j}). \tag{2.22}$$

Next, we introduce the psudolikelihood function in the context of unmarked point processes (the marked case can be treated analogically).

Definition 54 (Pseudolikelihood for a point process)**.** *Consider a finite point process on* $W \in \mathcal{B}_b^d$ *with Papangelou conditional intensity* $\lambda_\theta^\star$, *where* $\theta \in \Theta \subseteq \mathbb{R}^q$ *is the vector of unknown parameters. Then the pseudolikelihood function for a point realization* $\mathbf{x} \in \mathbf{N}_W$ *is*

$$\mathcal{PL}(\theta; \mathbf{x}) = \mathcal{PL}_W(\theta; \mathbf{x}) = \exp\left(|W| - \int_W \lambda_\theta^\star(x, \mathbf{x})\,\mathrm{d}x\right) \prod_{x \in \mathbf{x}} \lambda_\theta^\star(x, \mathbf{x} \setminus \{x\}). \tag{2.23}$$

The equation (2.23) is in accordance with the equation (2.22). This is easy to see, when we realize that the point process density is a density with respect to Poisson measure given by (2.6).

The maximum pseudolikelihood estimate (MPLE) is then found by maximizing the pseudolikelihood function $\mathcal{PL}$ in θ, i.e., if the parameter space $\Theta \subseteq \mathbb{R}^q$ is an open set, the MPLE is a solution to the pseudolikelihood estimating equation $\frac{\partial}{\partial\theta} \log \mathcal{PL}(\theta; \mathbf{x}) = 0$. More details about the pseudolikehood and maximum pseudolikelihood estimates in the context of point processes can be found in Møller and Waagepetersen [2003] and Jensen and Møller [1991].

Note that (2.23) agrees with the maximum likelihood function in case of unmarked Poisson point process, i.e., when $\lambda_\theta^*(x, \mathbf{x})$ depends only on x. So for point processes with weak interaction, the MPLE and the MLE may be expected to be close.

In practice the integrals arising in (2.22) and (2.23) are approximated by numerical methods, e.g., in the context of unmarked point processes:

Suppose we partition W into a finite number of cells C_i, and let $c_i \in C_i$ denote a given 'centre point'. Let $\{u_j; j = 1, \dots, m\}$ denote a list of these centre points and the points in $\mathbf{x}_W$. Then the integral in (2.23) is approximated by

$$\int_W \lambda_\theta^*(x, \mathbf{x}) dx \approx \sum_{j=1}^{m} \lambda_\theta^*(u_j, \mathbf{x} \setminus \{u_j\}) w_j,$$

where $w_j = \frac{|C_i|}{(1+\mathrm{card}(\mathbf{x}_{C_i}))}$ if $u_j \in C_i$. Then

$$\log \mathcal{PL}(\theta; \mathbf{x}) \approx \sum_{j=1}^{m} (y_j \log \lambda_j^* - \lambda_j^*) w_j,$$

where $y_j = \mathbb{I}_{\{u_j \in \mathbf{x}\}}/w_j$ and $\lambda_j^* = \lambda_\theta^*(u_j, \mathbf{x} \setminus \{u_j\})$.

Note that approximation involves a 'discontinuity error'. Namely the cost of the discretization using a grid of values is biasedness of the MPLE. The bias can be removed in several ways

- using random u_i instead of a grid of values,
- Richardson extrapolation.

For details see Baddeley and Turner [2014].

Exponential family models

The exponential family, Barndorff-Nielsen [2014], is an important class of probabilistic densities. Most of densities of Gibbs point processes (3.9) belong to this family.

Definition 55. *A density p_θ belongs to an exponential family, if it can be written as*

$$p_\theta(\cdot) = \frac{1}{Z(\theta)} b(\cdot) \exp(\theta t(\cdot)), \theta \in \Theta \subseteq \mathbb{R}^q,$$

where b is nonnegative function and t is a function taking values in $\mathbb{R}^q$.

The vector $t(\cdot)$ is called the canonical sufficient statistic. In case of point processes, $\lambda_\theta^*(x, \mathbf{x}) = b(x, \mathbf{x}) \exp(\theta t(x, \mathbf{x}))$, where $b(x, \mathbf{x}) = \frac{b(\mathbf{x} \cup \{x\})}{b(\mathbf{x})}$ and $t(x, \mathbf{x}) = t(\mathbf{x} \cup \{x\}) - t(\mathbf{x})$. A particular example of an exponential family model is Strauss point process. Indeed, if we fix interaction range $\delta > 0$ and exclude the hard-core case $\gamma = 0$ we obtain an exponential family model with $b = 1$, $\theta = (\theta_1, \theta_2) = (\log \beta, \log \gamma)$, $t(\mathbf{x}) = (\mathrm{card}(\mathbf{x}), s_\delta(\mathbf{x}))$ and $\Theta = \mathbb{R} \times (-\infty, 0)$.

Theorem 15 (Pseudolikelihood of an exponential model)**.** *Let p_θ belong to the exponential family. Then $\mathcal{PL}(\theta; \mathbf{x})$ is a log-concave function in θ.*

Proof. Proposition 2.3 in Jensen and Møller [1991]. □

The same proposition gives a condition for strict concavity.

Provided a model with a density belonging to the exponential family, the $\mathcal{PL}(\theta; \mathbf{x})$ can be easily maximized using (multidimensional) Newton-Raphson method, Magrenan and Argyros [2018].

Consistency and asymptotic normality of maximum pseudolikelihood estimates are briefly discussed in Møller and Waagepetersen [2003], Section 9.2.3.

Takacs-Fiksel method

From the integral representation (2.13) – GNZ equation, we obtain for any Gibbs point process that

$$\sum_{y \in \mathbf{x}} g(y, \mathbf{x} \setminus \{y\}) = \int_{\mathbb{R}^d} g(y, \mathbf{x}) \lambda_\theta^*(y, \mathbf{x}) dy, \tag{2.24}$$

is an unbiased estimating equation for any real function $g : \mathbb{R}^d \times \mathbf{N} \to \mathbb{R}$ (assuming that the expectations exist). A Takacs-Fiksel estimate is a solution of (2.24) for a number $k \geq q$ of functions g_i, $i = 1, \dots, k$, i.e., k unbiased estimating equations are obtained, cf. Møller and Waagepetersen [2003].

The i-th component, $i = 1, \dots, q$, of the pseudolikelihood estimating equation $\frac{\partial}{\partial\theta} \log \mathcal{PL}(\theta; \mathbf{x}) = 0$ emerges as the special case with g in (2.24) replaced by

$$g_i(x, \mathbf{x}) = \mathbb{I}_{\{x \in W\}} \frac{d}{d\theta_i} \log \lambda^*_\theta(x, \mathbf{x}_W),$$

provided $\lambda^*_\theta(x, \mathbf{x})$ only depends on $\mathbf{x}$ through $\mathbf{x}_W$, $W \in \mathcal{B}^d_b$.

The extension to the marked case is rather straightforward; the analogy of GNZ in the marked case is (2.17) and many details concerning the Takacs-Fiksel estimating procedure can be found in Coeurjolly et al. [2012].

2.4.2 Global envelopes

Global envelopes present a useful tool in spatial statistics for the graphical interpretation of results from tests based on functional or multivariate statistics, for determining central regions of functional or multivariate data, and also for determining confidence or prediction bands.

The word "global" means that the envelope is given with the prescribed coverage $100(1-\alpha)\%$ simultaneously for all the elements of the multivariate or functional statistic. The functional data first have to be discretized (the discretization can be arbitrary, as long as it is the same for all functions). Then for d-dimensional vectors $T_1, T_2, \dots, T_s$, $T_i = (T_{i,1}, \dots, T_{i,d})$, $i = 1, \dots, s$, a $100(1-\alpha)\%$ global envelope is considered to be a band bounded by the vectors $T^\alpha_{low} = (T^\alpha_{low,1}, \dots, T^\alpha_{low,d})$ and $T^\alpha_{upp} = (T^\alpha_{upp,1}, \dots, T^\alpha_{upp,d})$ such that the probability that T_i falls outside this envelope in any of the d points is equal to α, for $\alpha \in (0, 1)$, i.e.,

$$P\left(T_{i,j} \notin [T^\alpha_{low,j}, T^\alpha_{upp,j}] \text{ for any } j \in \{1, \dots, d\}\right) = \alpha.$$

When constructing the global envelope, the d-dimensional vectors T_i need to be ordered from the most extreme to the least extreme. For this purpose, many different measures exist. The `R`-package `GET` [Myllymäki and Mrkvička, 2019, Myllymäki et al., 2017] focuses on such measures for which it is possible to construct the global envelope with a practically interesting graphical interpretation. A prominent place occupy completely non-parametric envelopes which are based on some kind of 'rank' measure. In order to avoid possible ties special rank measures such as extreme rank length, continuous and area ranks are used. For definitions of mentioned rank measures see Myllymäki and Mrkvička [2019].

Among many other ranks, the envelopes based on area rank are the most robust according to recommendation in Myllymäki and Mrkvička [2020] and hence used throughout this thesis.

2.4.3 Permutation tests

A permutation test, see, e.g., Edgington and Onghena [2007], is an exact nonparametric statistical test that allows a hypothesis testing that assumes very little about the distribution of the data. Instead of assuming a distribution for the data sample and using it to derive the distribution of the test statistic, the permutation test works directly with the data. The idea is to permute (rearrange) the data by shuffling their labels of treatments, and then calculate a test statistic on each permutation. The collection of test statistics from the permuted data constructs the distribution under the null hypothesis H_0. The hypothesis H_0 assumes that data are exchangeable, i.e., the joint distribution is independent of positions (does not change when the positions are altered). In practise, the amount of all possible permutations is huge and only a fixed number of random permutations is used – so called Monte Carlo permutation test. The test statistic computed from the data is compared to those computed from permutations. Its extremity then determines the p-value.

In our setting, the permutation test can be used for testing the independence of points $\mathbf{x}_n$ and marks $\mathbf{r}_n$ – the data sample is the vector of marks/radii $\mathbf{r}_n$ and positions are given by point pattern $\mathbf{x}_n$. By permutation of the positions and keeping the marks we mimic the independent setting, i.e., the independency of points and marks. We can use a suitable functional characteristic of marked point processes as our test statistic. The extremity of the functional statistic computed from data among those computed from permutations can be then measured by rank measures as extreme rank length, continuous rank and area rank, cf. Myllymäki and Mrkvička [2019]. This leads to global envelope testing from Section 2.4.2.

3. Gibbs-Laguerre tessellations

Throughout this chapter we will consider $(\mathbf{x}_n, \mathbf{r}_n)$ to be a realization of a marked Gibbs point process. When $(\mathbf{x}_n, \mathbf{r}_n)$ are generators of a Laguerre tessellation, we speak about Gibbs-Laguerre tessellation models. The majority of the results presented in this chapter was published in Seitl et al. [2021] – introduction of particular Gibbs-Laguerre tessellation models, maximum pseudolikelihood estimation and reconstructions; and in Jahn and Seitl [2020] – existence of the particular Gibbs-Laguerre tessellation models.

3.1 Energy function

Definition 56. *A potential function* $V : \mathbf{N}^f \to \mathbb{R} \cup \{\infty\}$ *is a measurable symmetric function, where symmetry means that the value of the function remains the same when we permute its arguments. In particular,* $V_k : \mathbf{N}^{f,k} \to \mathbb{R} \cup \{\infty\}$ *is a potential function of order* k*, where* $k = 1, 2, \dots, n$*.*

The energy function E, given in Definition 29, can be built as a sum of potential functions (shortly potentials), see Baddeley [2007], Def. 4.2. We distinguish two types of potential functions. We speak about soft-core potentials if they are finite. On the other hand, hard-core potentials take on only one of the values 0 or $+\infty$. In the rest of the thesis, when writing arguments of a potential function, we identify cells of tessellation with their generators, i.e., $V(\mathbf{x}, \mathbf{r}) = V(\{C(x, r) : (x, r) \in (\mathbf{x}, \mathbf{r})\})$.

3.1.1 Periodic energy function

Because of the bounded sampling window we possibly need to resolve the boundary effects. This requires knowledge of the process outside the window, either we can observe (marked) points beyond the window or we add the outer (marked) point configuration artificially. An elegant way how to circumvent this issue is the employment of periodic boundary conditions, see Definition 11. In the periodic setup a potential function of k-th order is summed over k-tuples of k-neighbouring cells in the periodic domain $C_1, \dots, C_k \sim_k$ (to recall the notation see Section 2.1.1), $C_1, \dots, C_k \in L(\mathbf{x}^*, \mathbf{r}^*)$, such that each periodic k-tuple makes a unique contribution. In other words, there is only one contribution to the potential from the periodical extension $(\mathbf{x}_k^*, \mathbf{r}_k^*)$ of the k-tuple $(\mathbf{x}_k, \mathbf{r}_k) = \{(x_1, r_1), \dots, (x_k, r_k)\}$ of generators of given cells $C_1, \dots, C_k \sim_k$.

In general, a periodic energy function $\tilde{E} : \mathbf{N}^f \to \mathbb{R} \cup \{\infty\}$ can combine different potentials of different orders and can be written in a parametric form (the case of different orders)

$$\begin{aligned}\tilde{E}(\mathbf{x}, \mathbf{r}) = V_{hard} &+ \theta_1 \sum_{\substack{C \in L(\mathbf{x}^*, \mathbf{r}^*)\\ \mathrm{bar}(C) \in W}} V_1(C) + \theta_2 \sum_{\substack{C_1, C_2 \in L(\mathbf{x}^*, \mathbf{r}^*); C_1, C_2 \sim_2\\ \text{unique contribution}}} V_2(C_1, C_2) + \dots \\ &+ \theta_{n-1} \sum_{\substack{C_1, \dots, C_{n-1} \in L(\mathbf{x}^*, \mathbf{r}^*)\\ C_1, \dots, C_{n-1} \sim_{n-1}\\ \text{unique contribution}}} V_{n-1}(C_1, C_2, \dots, C_{n-1}) + \theta_n V_n(C_1, C_2, \dots, C_n),\end{aligned} \tag{3.1}$$

where all hard-core potentials are included in the term V_{hard}, $\theta_1, \dots, \theta_n$ are real-valued parameters, $\mathrm{bar}(\cdot)$ denotes the barycenter of a given set and $\sim_k$ is the k-neighbourhood relation, see Section 2.1.1. Recall that for $k = 2$, neighbouring cells are those which share a common face. For $k = 3, 4$ we distinguish the general form of k-neighbourhood relation $\sim_k$ and its special form – the proper neighbourhood relation, when the cells share an edge if $k = 3$ or a vertex if $k = 4$. In this case, i.e., $k = 3, 4$, the potential determines whether we sum over cells in the general k-neighbourhood or in the proper neighbourhood relation – e.g., $V_3(C_1, C_2, C_3) = \mathrm{el}(C_1 \cap C_2 \cap C_3)$, where 'el' stands for the edge length, makes sense only for C_1, C_2, C_3 in proper neighbourhood relation because otherwise $C_1 \cap C_2 \cap C_3 = \emptyset$. For $k = n$, the entire tessellation is considered to be neighbouring in the sense that all cells can be arranged such that $C_{i_1} \sim C_{i_2} \sim \dots \sim C_{i_n}$, where $\{i_1, \dots, i_n\}$ is a permutation of $\{1, \dots, n\}$. It is important that each subset of cells makes a unique contribution to the energy function. Note that there can be several potentials of the same order. Furthermore, the potential V_{hard} can be written as a sum of hard-core potentials, i.e., potentials that can either be equal to zero or $+\infty$, i.e.,

$$V_{hard} = \sum_{\substack{C \in L(\mathbf{x}^*, \mathbf{r}^*) \\ \mathrm{bar}(C) \in W}} V_{1,hard}(C) + \dots + V_{n,hard}(C_1, C_2, \dots, C_n).$$

Both energy function and periodic energy function are defined for $(\mathbf{x}, \mathbf{r}) \in \mathbf{N}^f$ only. If $E(\mathbf{x}, \mathbf{r}) < +\infty$ or $\tilde{E}(\mathbf{x}, \mathbf{r}) < +\infty$ in the periodic setup, we say that the configuration $(\mathbf{x}, \mathbf{r})$ is admissible.

3.1.2 Examples of potential functions

We will deal with the following choices of potential functions. First, we consider the hard-core potential of first order forbidding empty Laguerre cells:

$$V_{1,hard}(C) = \begin{cases} +\infty & \text{if } C \text{ is empty,} \\ 0 & \text{else.} \end{cases} \tag{3.2}$$

Similarly as in the hierachical approach, we will avoid empty Laguerre cells in our models. Therefore from now we assume, that the hard-core potential (3.2) is part of all following Gibbs-Laguerre models.

Further, hard-core potentials may influence the shape and size of cells, e.g.,

$$V_{1,hard}(C) = \begin{cases} +\infty & \text{if } h_{min}(C) \leq \alpha, \\ +\infty & \text{if } h_{max}(C) \geq \beta, \\ +\infty & \text{if } h_{max}^3(C) \geq B|C|, \\ 0 & \text{else,} \end{cases} \tag{3.3}$$

where $h_{min}(C)$, $h_{max}(C)$ denotes the minimum, maximum distance between the cell barycenter and a face of C, respectively, with $0 < \alpha < \beta$, $B > 0$. The parameter α forces the cells to be not too small, while β forces them to be not too large. The parameter B controls the shape of the cells—the smaller the value of B, the more regular are the shapes of the cells.

A soft-core potential of k-th order $V_k(C_1, \dots, C_k)$ is a symmetric function of a k-tuple of k-neighbouring cells. In practice, these potentials are often assumed to be nonnegative and bounded. In 2D these two properties ensure the stability property, Definition 30(iii), of the energy function, cf. Dereudre et al. [2012] (unfortunately, this implication does not seem to be generally preserved in higher dimensions). In case there is no upper bound, an artificial bound $K > 0$ can be used (K is some large constant depending on the particular potential). A pair potential function studied later on is given by

$$V_{2,\mathrm{NVR}}(C_1, C_2) = \mathrm{NVR}(C_1, C_2) \wedge K, \tag{3.4}$$

where NVR is defined in (2.4). The potential given in (3.4) may be multiplied by a real parameter θ. In such case, the sign of θ is crucial. In the case when $\theta > 0$, the neighboring cells tend to have a similar volume; on the other hand $\theta < 0$ forces the neighboring cells to have substantially different volumes. The choice of the power $1/2$, in (3.4) and later in (3.5), is not compulsory, in general, the power can be chosen arbitrarily.

To introduce higher order potentials, we set the following formal notation: Let $s_G : \mathcal{C}^k \to \mathbb{R}^l$ and $s_G(C_1, \dots C_k)$ denote a sample of l values of the geometric characteristic G assigned to a collection of cells $C_1, \dots C_k$, $k, l \in \mathbb{N}$. In particular, $s_{\mathrm{nof}}(C_1, \dots, C_k)$ and $s_{\mathrm{dvol}}(C_1, \dots, C_k)$ are samples of $l = k$ values of number of faces computed from k cells and $l > k$ values of dvol computed from k cells, respectively. In the latter case, l is the total number of faces among the cells $C_1, \dots, C_k$. Let $T : \mathbb{R}^l \to \mathbb{R}$ be a function defined on the sample $s_G(C_1, \dots, C_k)$, namely $T(s_G(C_1, \dots, C_k)) = \bar{s}_G(C_1, \dots, C_k)$ and $T(s_G(C_1, \dots, C_k)) = S^2_{s_G}$ stand for the sample mean and sample variance computed over the sample $s_G(C_1, \dots, C_k)$, respectively. $s_0 \in \mathbb{R}$ is the value we want $T(s_G(\cdot))$ to take. The meaning of the notation is demonstrated on an example: $T(s_{\mathrm{nof}}(C_1, \dots, C_k)) = \bar{s}_{\mathrm{nof}}(C_1, \dots, C_k) = \frac{1}{k}\sum_{i=1}^{k} \mathrm{nof}(C_i) = 12$ means that the mean number of faces per cells $C_1, \dots, C_k$ is 12, i.e., $G = \mathrm{nof}$, $l = k$.

The potential of n-th order has a very special meaning. Recall that n is the cardinality of the observed marked point pattern $(\mathbf{x}_n, \mathbf{r}_n)$ (i.e., total number of cells) on the bounded sampling window W. During simulations/reconstructions carried out in Section 3.4, this marked point pattern on $W \times \mathbb{M}$ is allowed to change its cardinality; thus, n is not constant in time. An example of a potential function of n-th order is

$$V^G_{n,T}(C_1, \dots, C_n) = \left(|T(s_G(C_1, \dots, C_n)) - s_0|\right)^{1/2}. \tag{3.5}$$

Later on, potential functions of the form (3.5) will be referred to as reconstructing potentials. A special case of this potential of n-th order is

$$V^G_{n,\mathrm{dsc}}(C_1, \dots, C_n) = \left(\mathrm{dsc}(H_{s_G(C_1,\dots,C_n)}, H'_s)\right)^{1/2}, \tag{3.6}$$

where $T(s_G(C_1, \dots, C_n)) = \mathrm{dsc}(H_{s_G(C_1,\dots,C_n)}, H'_{s_G})$, $s_0 = 0$, dsc is so called discrepancy of histogram $H_{s_G(C_1,\dots,C_n)}$ of the chosen geometrical characteristic computed from all cells and prescribed targetting histogram H'_{s_G} of the geometrical characteristic s that we want to approach (this can be typically obtained from data). The discrepancy is defined in (3.7).

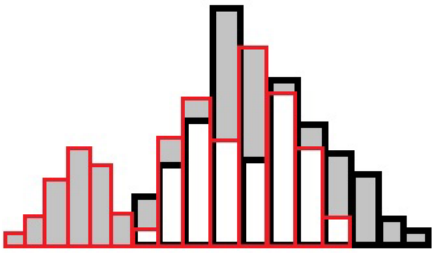


Figure 3.1: For the two histograms coloured in this figure by red and black, their discrepancy is a constant proportional to a sum of frequencies corresponding to the regions shaded in gray.

Recall that the histogram H is viewed as a collection of frequencies $h_1, \ldots, h_J$ in $J \in \mathbb{N}$ bins covering a given interval, for details see Section 2.1.1. Using the notation $S = \sum_{i=1}^{J} h_i$, the discrepancy between a pair of histograms (H, H') defined over the same interval and having the same bins (this implies the same number of classes) can be written as

$$\mathrm{dsc}(H, H') = \sum_{i=1}^{J} \left| \frac{h_i}{S} - \frac{h_i'}{S'} \right|. \tag{3.7}$$

An illustration of dsc can be seen in Fig. 3.1. The discrepancy measures the difference between two histograms and it is minimized when they are identical up to some positive multiplicative constant (i.e., there exists a constant $M > 0$ such that $h_i = M h_i'$ for every $i = 1, \ldots, J$). If we omit the normalizations S, S' in the definition of discrepancy, given in (3.7), then the discrepancy is minimized if the two histograms are identical. Alternatively, we could deal with the cumulative histogram or empirical distribution function instead of the histogram.

The potentials in (3.5) are called reconstructing potentials since they allow us to control the distributional aspects of the cells generated by $(\mathbf{x}, \mathbf{r})$. E.g., we may influence the moments (using $T(s_G(\cdot)) = \bar{s}(\cdot)$ for the first moment) or the entire distribution (using $T(s_G(\cdot)) = \mathrm{dsc}(H_{s_G(\cdot)}, H'_{s_G}))$ of some geometric characteristic G.

3.2 Existence of Gibbs-Laguerre tessellations

In this section we present one of the main theoretical achievements of this thesis, namely the existence of Gibbs-Laguerre tessellations in $\mathbb{R}^3$. The idea comes from the existence theorem published in Dereudre et al. [2012], we describe in detail its assumptions. As the theorem is formulated in the context of hypergraphs we have to precise what is the energy function. But first, we start with two key assumptions and their justification:

Assumption 1. *Assume* $(\mathbf{x}, \mathbf{r}) \in \mathbf{N}_{gp}$.

The assumption $(\mathbf{x},\mathbf{r}) \in \mathbf{N}_{gp}$ is not restrictive, since $\mathbf{N}_{gp}$ is measurable and for any $W \in \mathcal{B}_b^3$ it holds that $\pi_W(\mathbf{N} \setminus \mathbf{N}_{gp}) = 0$, cf. Section 4 in Zessin [2008].

Assumption 2. $\mathbb{M} = [0, R_{\max}]$ *for some* $R_{\max} > 0$.

In practise we observe $(\mathbf{x},\mathbf{r})$ in a bounded observation window W, hence $(\mathbf{x},\mathbf{r})$ is finite and we can choose an upper bound on radii marks.

3.2.1 Energy over hypergraph structure

Definition 29 says that energy function is a measurable function $E : \mathbf{N}^f \longrightarrow \mathbb{R} \cup \{+\infty\}$. In general, it can be expressed (resembling the definition of Hamiltonian below) as a sum of values of the hyperedge potential φ over hyperedges of some hypergraph structure $\mathcal{E}$, i.e.,

$$E(\mathbf{x},\mathbf{r}) = \sum_{\eta \in \mathcal{E}(\mathbf{x},\mathbf{r})} \varphi(\eta, (\mathbf{x},\mathbf{r})) \text{ for } (\mathbf{x},\mathbf{r}) \in \mathbf{N}_f.$$

We assume that energy function is non-degenerate, as the non-degeneracy ensures that it is not identically infinite.

Next, we would like to be able to say what is an energetic contribution of $(\mathbf{x},\mathbf{r})_W$, $W \in \mathcal{B}_b^3$, in the computation of the energy of $(\mathbf{x},\mathbf{r})$, i.e., to precise what is $E_W(\mathbf{x},\mathbf{r})$ from Definition 29, in the context of a hypergraph structure and a hyperedge potential. We will do it now directly for $(\mathbf{x},\mathbf{r}) \in \mathbf{N}$ with justification of the extension from $\mathbf{N}^f$ to $\mathbf{N}$ provided later.

Definition 57. *Considering* $(\mathbf{x},\mathbf{r}) \in \mathbf{N}$ *and taking* $\xi \in \mathbf{N}_W$ *we define the energy (Hamiltonian) of* ξ *in* W *with boundary condition* $(\mathbf{x},\mathbf{r})$ *by the formula*

$$E_{W,(\mathbf{x},\mathbf{r})}(\xi) = \sum_{\eta \in \mathcal{E}_W(\xi \cup (\mathbf{x},\mathbf{r})_{W^c})} \varphi(\eta, \xi \cup (\mathbf{x},\mathbf{r})_{W^c}), \tag{3.8}$$

provided the sum is well-defined (i.e., the negative part is finite, $E^-_{W,(\mathbf{x},\mathbf{r})}(\xi) < \infty$*). In the case* $\xi = (\mathbf{x},\mathbf{r})_W$ *we write* $E_{W,(\mathbf{x},\mathbf{r})}((\mathbf{x},\mathbf{r})_W) = E_W(\mathbf{x},\mathbf{r})$ *and speak about the local energy of* $(\mathbf{x},\mathbf{r})$ *in* W.

Recall that according to Definition 34, the infinite-volume Gibbs point process with activity z and energy function E is a point process with distribution P (called Gibbs measure) having conditional densities with respect to π_W of the form

$$p_W\left((\mathbf{x},\mathbf{r})_W | (\mathbf{x},\mathbf{r})_{W^c}\right) = \frac{1}{Z_W((\mathbf{x},\mathbf{r})_{W^c})} z^{\operatorname{card}(\mathbf{x}_W)} \exp\left(-E_W(\mathbf{x},\mathbf{r})\right) \tag{3.9}$$

for every $W \in \mathcal{B}_b^3$ and for $P - a.a.$ $(\mathbf{x},\mathbf{r})_{W^c}$, where

$$Z_W((\mathbf{x},\mathbf{r})_{W^c}) = \int z^{\operatorname{card}(\mathbf{x}_W)} \exp\left(-E_W(\mathbf{x},\mathbf{r})\right) \pi_W(d(\mathbf{x},\mathbf{r})_W),$$

is the normalizing constant called partition function and the argument indicates its dependence on $(\mathbf{x},\mathbf{r})_{W^c}$.

Definition 58 (Admissibility)**.** *A configuration* $(\mathbf{x},\mathbf{r}) \in \mathbf{N}$ *is called admissible for a region* W *and an activity* $z > 0$ *if* $E^-_{W,(\mathbf{x},\mathbf{r})}(\xi) < \infty$ *for* π_W*-a.a.* $\xi \in \mathbf{N}_W$ *(i.e.,* $E_{W,(\mathbf{x},\mathbf{r})}$ *is almost surely well-defined) and* $0 < Z_W((\mathbf{x},\mathbf{r})_{W^c}) < \infty$.

It is clear that a Gibbs measure is concentrated on the set of admissible configurations for all $W \in \mathcal{B}_b^3$. The extension of the energy function from finite point configurations to all locally finite configurations is required in (3.8) and (3.9). This can be achieved if a configuration $(\mathbf{x}, \mathbf{r})$ confines the range of φ from W, i.e., $(\mathbf{x}, \mathbf{r}) \in \mathbf{N}_{cr}^W$, Definition 20, then

$$E_{W,(\mathbf{x},\mathbf{r})}(\xi) = \sum_{\eta \in \mathcal{E}_W(\xi \cup \partial_W(\mathbf{x},\mathbf{r}))} \varphi(\eta, \xi \cup (\mathbf{x}, \mathbf{r})_{W^c}).$$

Even in the case when conditional densities are well-defined the existence of Gibbs point process is not obvious and further conditions on the energy function E need to be stated.

Examples of potential functions

All potentials mentioned in Section 3.1.2 can be reformulated in the context of hypergraphs, for example:

1. Soft-core pair potential defined for two Laguerre cells having a common face

$$\varphi_{2,\mathrm{NVR}}(\eta, (\mathbf{x}, \mathbf{r})) = \mathrm{NVR}(C(x_1, r_1), C(x_2, r_2)) \wedge K, \tag{3.10}$$

   with $\eta = \{(x_1, r_1), (x_2, r_2)\} \in \mathcal{LD}_2(\mathbf{x}, \mathbf{r})$ and $K > 0$.

2. Soft-core reconstructing potential

$$\varphi_{k,T}^G(\eta, (\mathbf{x}, \mathbf{r})) = \left(|T(s_G(\{C(x, r) : (x, r) \in \eta\})) - s_0|\right)^{1/2} \wedge K, \tag{3.11}$$

   where $\eta \in \mathcal{CG}_{k,b}(\mathbf{x}, \mathbf{r})$, $K > 0$ and $k \in \mathbb{N}$.

In order to satisfy the range condition **(R)**, see Section 3.2.2, we need to fix the number of cells entering the reconstructing potential (3.5). Therefore we introduce $V_{k,T}^G$, a k-th order version, $k \in \mathbb{N}$, of the reconstructing potential $V_{n,T}^G$. The hypergraph counterpart of $V_{k,T}^G$ is the hyperedge potential $\varphi_{k,T}^G$ from (3.11). If k is close to the expected number of generators inside the observation window, the $V_{k,T}^G$ is a good approximation of $V_{n,T}^G$. Although we need a fixed number of cells because of theoretical reasons, we keep the potential $V_{n,T}^G$ for simulations. If we use $V_{k,T}^G$ for simulations, we need to consider all connected k-tuples of cells where at least one cell was changed in order to evaluate the change of the energy in each iteration of the MHBDM algorithm, Algorithm 2. The number of such k-tuples can be enormous causing that the computational time becomes excessively long. The potential $V_{n,T}^G$ leads to a significant improvement in computational time, since the distributional characteristics (sample mean, sample variance or histogram) of the entire tessellation can be computed only once for the initial configuration. Then they are easily modified in each iteration since only a small number of cells differ in the proposal. Moreover, the drawback that the range condition **(R)** no longer holds in the case of (3.5) since it admits a variable number of cells in its argument is irrelevant for finite volume Gibbs point processes.

3.2.2 Assumptions and existence theorem

Firstly we state three assumptions from Dereudre et al. [2012] which are sufficient for the existence of a Gibbs measure, cf. Theorem 16.

(R) *The range condition.* There exist constants $\ell_R, n_R \in \mathbb{N}$ and $\delta_R < \infty$ such that for all $(\eta, (\mathbf{x}, \mathbf{r})) \in \mathcal{E}$ one can find a finite horizon Δ satisfying: For every $x, y \in \Delta$ there exist ℓ open balls $B_1, \ldots, B_\ell$ (with $\ell \leq \ell_R$) such that

- the set $\cup_{i=1}^{\ell} \bar{B}_i$ is connected and contains x and y, and
- for each i, either $\operatorname{diam}(B_i) \leq \delta_R$ or $\operatorname{card}(\mathbf{x} \cap B_i) \leq n_R$.

(S) *Stability.* The energy function E is called *stable* if there exists a constant $c_S \geq 0$ such that

$$E_{W,(\mathbf{x},\mathbf{r})}(\zeta) \geq -c_S \cdot \operatorname{card}(\zeta \cup \partial_W(\mathbf{x}, \mathbf{r}))$$

for all $W \in \mathcal{B}_b^3$, $\zeta \in \mathbf{N}_W$ and $(\mathbf{x}, \mathbf{r}) \in \mathbf{N}_{\mathrm{cr}}^W$.

(U) *Upper regularity.* M and Γ, cf. Subsection (A.1.1), can be chosen so that the following holds.

(U1) *Uniform confinement*: $\bar{\Gamma} \subset \mathbf{N}_{\mathrm{cr}}^W$ for all $W \in \mathcal{B}_b^3$ and

$$r_\Gamma := \sup_{W \in \mathcal{B}_b^3} \sup_{\mathbf{x} \in \bar{\Gamma}} r_{W,(\mathbf{x},\mathbf{r})} < \infty, \tag{3.12}$$

where $r_{W,(\mathbf{x},\mathbf{r})}$ was defined in Definition 20.

(U2) *Uniform summability*:

$$c_\Gamma^+ := \sup_{(\mathbf{x},\mathbf{r}) \in \bar{\Gamma}} \sum_{\eta \in \mathcal{E}(\mathbf{x},\mathbf{r}) : \eta_x \cap C \neq \emptyset} \frac{\varphi^+(\eta, (\mathbf{x}, \mathbf{r}))}{\operatorname{card}(\hat{\eta})} < \infty,$$

where $\hat{\eta} := \{k \in \mathbb{Z}^3 : \eta_x \cap C(k) \neq \emptyset\}$ and φ^+ is the positive part of φ.

(U3) *Strong non-rigidity*: $e^{z|C|}\pi_C^z(\Gamma) > e^{c_\Gamma}$, where c_Γ is defined as in (U2) with φ in place of φ^+.

($\hat{\mathbf{U}}$) *Alternative upper regularity.* M and Γ can be chosen so that the following holds.

($\hat{\text{U}}$1) *Lower density bound*: There exist constants $c, d > 0$ such that

$$\operatorname{card}(\zeta) \geq c|W| - d$$

whenever $\zeta \in \mathbf{N}_f \cap \mathbf{N}_W$ is such that $E_{W,(\mathbf{x},\mathbf{r})}(\zeta) < \infty$ for some $W \in \mathcal{B}_b^3$ and some $(\mathbf{x}, \mathbf{r}) \in \bar{\Gamma}$.

($\hat{\text{U}}$2) = (U2) *Uniform summability.*

($\hat{\text{U}}$3) *Weak non-rigidity*: $\pi_C^z(\Gamma) > 0$.

Before stating the existence theorem, we describe the conditions **(R)**, **(S)** and **(U)** in more detail and put them into context with conditions from Definition 30.

The fact that the hypergraph structure possesses a type of locality property is crucial for the existence of Gibbs measures. The first assumption **(R)** reflects this requirement. It says that hyperedges with a large horizon require the existence of a large ball with only a few points. It implies that the energy function $E_{W,(\mathbf{x},\mathbf{r})}$ depends only on the points of $(\mathbf{x},\mathbf{r})$ in a bounded set $\partial W(\mathbf{x},\mathbf{r})$, i.e., that $r_{W,(\mathbf{x},\mathbf{r})}$ from Definition 20 is a finite range from Definition 30(v) for the given marked point pattern $(\mathbf{x},\mathbf{r})$. If in addition all horizon sets can be chosen to have uniformly bounded diameters, the finite ranges $r_{W,(\mathbf{x},\mathbf{r})}$ do not depend on a particular choice of $(\mathbf{x},\mathbf{r})$ and the range condition from Definition 30(v) is trivially satisfied. Further, the assumption **(R)** justifies the restriction on $\mathbf{N}_{\mathrm{cr}}^{W}$ since any translation-invariant locally finite counting measure is concentrated on the set $\mathbf{N}_{\mathrm{cr}}^{W}$, see Proposition 3.1. in Dereudre et al. [2012].

The stability assumption **(S)** resembles the stability condition from Definition 30(iii) and in the same way ensures the finiteness of all normalizing constants $Z_W((\mathbf{x},\mathbf{r})_{W^c})$ in (3.9). Stability is trivially satisfied if φ is non-negative, then **(S)** holds with $c_S = 0$. In case that the hyperedge potential is bounded below, i.e., $\varphi(\eta,(\mathbf{x},\mathbf{r})) \geq -c_\varphi$ for some $c_\varphi < \infty$, stability is ensured if the hypergraph $\mathcal{E}$ is sublinear, see Definition 16.

Finally, when verifying the upper regularity conditions **(U)**, we can restrict ourselves only on the so called pseudo-periodic configurations defined in Appendix A.1. The assumption (U1) states that the pseudo-periodic configurations in $\bar{\Gamma}$ confine the range of φ in a uniform way. Condition (U2) provides a uniform upper bound for the local energy $E_{W,\cdot}$ on $\bar{\Gamma}$. The last condition (U3) is satisfied for all $z \geq z_0$ for some $z_0 \geq 0$, provided that (U2) holds and $\pi_C^z(\Gamma) > 0$. The alteration **($\hat{\mathbf{U}}$)** can help us when it is difficult to satisfy (U3) for small values of z.

Remark (Simplification of the upper regularity condition). Using the set Γ^A, defined in Subsection A.1.1, the conditions (U2) and (U3) can be simplified. In (U2), $\mathrm{card}(\hat{\eta}) = \mathrm{card}(\eta)$ since each point of η is in a different set $C(k)$. In (U3), $\pi_C^z(\Gamma)$ can be directly calculated:

$$\pi_C^z(\Gamma^A) = \pi_C^z(\{\xi \in \mathbf{N}_C : \xi = \{p\}, p \in A\}) = e^{-z|A|}z|A|e^{-z|C\setminus A|} = e^{-z|C|}z|A|.$$

(U3) is then of the form $z|A| > e^{C_\Gamma}$. By taking A as in (A.1) one obtains

$$|A| = \frac{4}{3}\pi\rho^3 a^3 \cdot \left(\frac{a}{2}(1-2\rho)\right)^2 = \frac{1}{3}\pi a^5\rho^3(1-2\rho)^2.$$

Theorem 16. *For every hypergraph structure $\mathcal{E}$, hyperedge potential φ and activity $z > 0$ satisfying **(S)**, **(R)** and **(U)** there exists at least one Gibbs measure.*

Alternative version of the theorem is obtained by replacing the condition **(U)** by **($\hat{\mathbf{U}}$)**.

Theorem 17. *A Gibbs measure exists also under the assumptions **(S)**, **(R)** and **($\hat{\mathbf{U}}$)**.*

The proof of both theorems can be found in Dereudre et al. [2012], see Theorems 3.2 and 3.3 together with Remark 3.7.

3.2.3 Auxiliary lemmas

The following two lemmas show two particular situations when either the range condition or the uniform confinement condition is satisfied. Lemma 18 states that the range condition is fulfilled in case of hypergraph structures $\mathcal{LD}$ and $\mathcal{LD}_k$, $k = 1, \dots, 4$, and an arbitrary unary potential.

Lemma 18. *For models with the hypergraph structure $\mathcal{LD}$ or $\mathcal{LD}_k$, $k = 1, \dots, 4$, with a unary potential, the range condition **(R)** is satisfied with parameters $\ell_R = 3, n_R = 0, \delta_R = 2R_{\max}$.*

Proof. The lemma was published as Lemma 4.6 in Jahn and Seitl [2020]. Take the horizon set $\Delta = b(x_\eta, \sqrt{r_\eta^2 + R_{\max}^2})$. As described in Claim 5, Δ can be decomposed into the ball $b(x_\eta, r_\eta)$ and $\Delta \backslash b(x_\eta, r_\eta)$, a spherical shell with thickness

$$\sqrt{r_\eta^2 + R_{\max}^2} - r_\eta = R_{\max}^2 / \left(\sqrt{r_\eta^2 + R_{\max}^2} + r_\eta \right).$$

The ball $b(x_\eta, r_\eta)$ determined by the characteristic point (x_η, r_η) cannot contain any points of $(\mathbf{x}, \mathbf{r})$. Although the spherical shell $\Delta \setminus b(x_\eta, r_\eta)$ does not have any bound on the number of points, its thickness is bounded by $R_{\max}$. This means that any $x, y \in \Delta$ can be connected by the spheres $b(x, R_{\max}), b(x_\eta, r_\eta), b(y, R_{\max})$, yielding the parameters $\ell_R = 3, n_R = 0, \delta_R = 2R_{\max}$. □

Lemma 19 states that uniformly bounded finite horizons imply uniform confinement (U1).

Lemma 19. *Let $\Gamma \subset \mathbf{N}$ be a class of configurations. If there exists $d_{max} > 0$ such that $\operatorname{diam}\Delta < d_{max}$ for the horizon Δ of any $(\eta, (\mathbf{x}, \mathbf{r})), \eta \in \mathcal{E}(\mathbf{x}, \mathbf{r}), (\mathbf{x}, \mathbf{r}) \in \Gamma$, then*

$$r_\Gamma < d_{max},$$

where r_Γ is defined as in the condition $(U1)$.

Proof. The lemma was published as Lemma 4.5 in Jahn and Seitl [2020]. Choose $W \in \mathcal{B}_b^3$ and $\mathbf{x} \in \Gamma$. Let $\zeta \in \mathbf{N}_W$, $\eta \in \mathcal{E}_W(\zeta \cup (\mathbf{x}, \mathbf{r})_{W^c})$ and denote Δ the finite horizon of $(\eta, (\mathbf{x}, \mathbf{r}))$. Then $\Delta \cap W \neq \emptyset$, since

$$\eta \in \mathcal{E}_W(\zeta \cup (\mathbf{x}, \mathbf{r})_{W^c}) \Leftrightarrow \exists \xi \in \mathbf{N}_W : \varphi(\eta, \zeta \cup (\mathbf{x}, \mathbf{r})_{W^c}) \neq \varphi(\eta, \xi \cup (\mathbf{x}, \mathbf{r})_{W^c})$$

$$\Rightarrow \exists \xi \in \mathbf{N}_W : \xi_x \cap \Delta \neq \emptyset \Rightarrow W \cap \Delta \neq \emptyset.$$

Therefore $\Delta \subset W + B(0, d_{max})$. If we take $(\tilde{\mathbf{x}}, \tilde{\mathbf{r}}) \in \Gamma$ such that $(\tilde{\mathbf{x}}, \tilde{\mathbf{r}}) = (\mathbf{x}, \mathbf{r})$ on $\partial W(\mathbf{x}, \mathbf{r})$ then $\varphi(\eta, \zeta \cup (\mathbf{x}, \mathbf{r})_{W^c}) = \varphi(\eta, \zeta \cup (\tilde{\mathbf{x}}, \tilde{\mathbf{r}})_{W^c})$ since $\zeta \cup (\mathbf{x}, \mathbf{r})_{W^c}$ and $\zeta \cup (\tilde{\mathbf{x}}, \tilde{\mathbf{r}})_{W^c}$ differ only on Δ^c. □

3.2.4 Existence of Gibbs-Laguerre models

Laguerre cell face interaction

In 3D we examine the Laguerre tessellation where every face contributes to the energy function. Since the potential depends on pairs of neighbouring cells having a common face, we talk about pair interaction.

Theorem 20. *There exists at least one Gibbs measure for the hypergraph structure $\mathcal{LD}_2$, hyperedge potential $\varphi_{2,NVR}$ and every activity $z > 0$.*

Proof. The theorem was published as Proposition 4.12 in Jahn and Seitl [2020].

(R) Claim 5 tells us that $\Delta = B(\xi_x, \sqrt{\xi_r^2 + R_{\max}^2})$ is the finite horizon for $(\xi, (\mathbf{x}, \mathbf{r})) \in \mathcal{LD}$ and unary hyperedge potential. Moreover it shows that the horizon can be decomposed to a ball with no interior points and a spherical shell with thickness bounded by $R_{\max}$. Since the pair potential depends on two cells of $\eta = \{(x_1, r_1), (x_2, r_2)\} \subset (\mathbf{x}, \mathbf{r})$. We define so-called Laguerre doubleflower, see Fig. 3.2, $LF_2 := \bigcup_{\xi \in \mathcal{LD}_4, \xi \cap \eta \neq \emptyset} \bar{b}(\xi_x, \sqrt{\xi_r^2 + R_{\max}^2})$. For $y_1, y_2 \in LF_2$ one can find $\xi_1, \xi_2 \in \mathcal{LD}_4$ such that $y_1 \in \bar{b}(\xi_{1x}, \sqrt{{\xi_1}_r^2 + R_{\max}^2})$ and $y_2 \in \bar{b}(\xi_{2x}, \sqrt{{\xi_2}_r^2 + R_{\max}^2})$. Then either $\bar{b}(\xi_x, \sqrt{\xi_r^2 + R_{\max}^2}) \cap \bar{b}(\xi_{2x}, \sqrt{{\xi_2}_r^2 + R_{\max}^2}) \neq \emptyset$ or $\bar{b}(\xi_x, \sqrt{\xi_r^2 + R_{\max}^2}) \cup \bar{b}(\zeta_x, \sqrt{\zeta_r^2 + R_{\max}^2}) \cup \bar{b}(\xi_{2x}, \sqrt{{\xi_2}_r^2 + R_{\max}^2})$, $\eta \subset \zeta \in \mathcal{LD}_4$, is connected. Therefore the range condition holds with $l_R = 7$, $n_r = 0$, $\delta_R = 2R_{\max}$ and Laguerre doubleflower as the horizon for every $(\eta, (\mathbf{x}, \mathbf{r})) \in \mathcal{LD}_2$ and the hyperedge potential $\varphi_{2,NVR}$.

(S) Stability is satisfied because φ is non-negative.

(U) We choose M and Γ as in Subsubsection A.1.1.

(U1) By Remark A.1.1 there exists $R_0 > 0$ such that $\eta_r \leq R_0$ for all $\eta \in \mathcal{LD}_4(\mathbf{x}, \mathbf{r}), (\mathbf{x}, \mathbf{r}) \in \bar{\Gamma}^A$. For every $(\eta, (\mathbf{x}, \mathbf{r})) \in \mathcal{LD}$ the diameter of the set $\Delta = b(\eta_x, \sqrt{\eta_r^2 + R_{\max}^2})$ can be bounded, cf. Claim 5: $\operatorname{diam} \Delta = 2\sqrt{\eta_r^2 + R_{\max}^2} \leq 2\sqrt{R_0^2 + R_{\max}^2}$. In the case of Laguerre doubleflower LF_2 we obtain the bound for its radius $3\sqrt{R_0^2 + R_{\max}^2}$. Since LF_2 is the horizon for every $(\eta, (\mathbf{x}, \mathbf{r})) \in \mathcal{LD}_2$ and the hyperedge potential $\varphi_{2,NVR}$, we obtain $r_\Gamma \leq 6\sqrt{R_0^2 + R_{\max}^2}$ using Lemma 19.

(U2) $c_A^+ = sup_{(\mathbf{x},\mathbf{r}) \in \bar{\Gamma}^A} \sum_{\eta \in \mathcal{LD}_2(\mathbf{x},\mathbf{r}), \eta_x \cap C \neq \emptyset} \frac{\varphi_{2,VNR}^+(\eta,(\mathbf{x},\mathbf{r}))}{card(\eta_x)} \leq \sum_{\eta \in \mathcal{LD}_2(\mathbf{x},\mathbf{r}), \eta_x \cap C \neq \emptyset} \frac{K}{2}$, where the right side is finite since the sum is finite. Indeed the number of incident tetrahedra to each vertex is at most 32, cf. Remark A.1.1. Each incident tetrahedron contains 3 edges ending in the considered vertex. Therefore the number of summands is for sure ≤ 96.

(U3) The lower bound on z is gained from the condition $z|A| > e^{C_\Gamma}$, cf. Remark 3.2.2. Since $|A| = \frac{1}{3}\pi a^5 \rho^3 (1 - 2\rho)^2$ the inequality results in $z > C_0 \frac{1}{a^5} e^{C_1}$, where $C_0, C_1 > 0$ are constants not depending on a. Minimizing the right side over $a > 0$ we obtain $z > 0$ as $a \to \infty$.

$\square$

Laguerre cell distribution interaction

The goal of the potential $\varphi_{k,T}^G$ is to influence the distribution of a geometric characteristic s computed from the sample of size k, $k \in \mathbb{N}$. In the stationary case the distribution on $(\mathbf{x}, \mathbf{r})$ is well approximated by the distribution on a finite η possessing a sample of size k large enough.

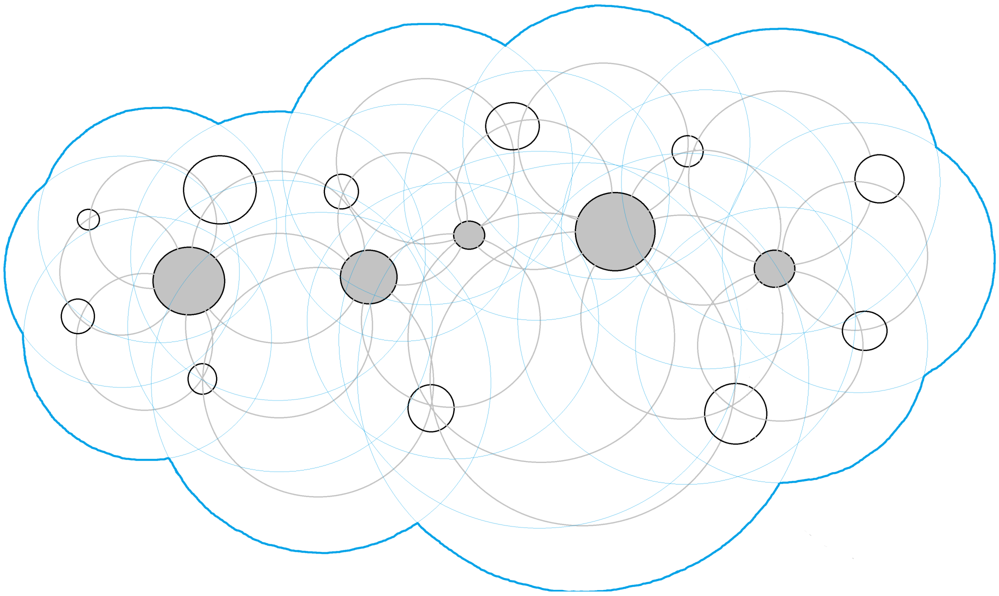


Figure 3.2: Laguerre 5-flower LF_5 of grey generators. Black spheres depicture generators. Grey and blue spheres are their characteristic points and finite horizons, respectively.

Theorem 21. *There exists at least one Gibbs measure for the hypergraph structure $\mathcal{CG}_{k,b}$, hyperedge potential $\varphi^G_{k,T}$, $k \in \mathbb{N}$, and every activity $z > 0$.*

Proof. The theorem was published as Proposition 4.13 in Jahn and Seitl [2020].

(R) Claim 5 tells us that $\Delta = b(\xi_x, \sqrt{\xi_r^2 + R_{\max}^2})$ is the finite horizon for $(\xi, (\mathbf{x}, \mathbf{r})) \in \mathcal{LD}$ and a unary hyperedge potential. Moreover it shows that the horizon can be decomposed to the ball with no interior points and annulus with width bounded by $R_{\max}$. The hyperedge η contains k points. Similarly as in Theorem 20 we define Lagurre k-flower, see Fig. 3.2, $LF_k := \bigcup_{\xi \in \mathcal{LD}_4, \xi \cap \eta \neq \emptyset} \bar{b}(\xi_x, \sqrt{\xi_r^2 + R_{\max}^2})$. Then for $y_1, y_2 \in LF_k$ one can find $\xi_1, \xi_2 \in \mathcal{LD}_4$ such that $y_1 \in \bar{b}(\xi_{1x}, \sqrt{{\xi_1}_r^2 + R_{\max}^2})$ and $y_2 \in \bar{b}(\xi_{2x}, \sqrt{{\xi_2}_r^2 + R_{\max}^2})$. These two balls can be connected in the worst case by $k-1$ balls $\bar{b}(\xi_x, \sqrt{\xi_r^2 + R_{\max}^2})$, where $\mathrm{card}(\mathbf{x}_\xi \cap \mathbf{x}_\eta) \geq 2$. The range condition **(R)** is then satisfied with $l_R = 2k + 3$, $n_r = 0$, $\delta_R = 2R_{\max}$ and Laguerre k-flower as the horizon for every $(\eta, (\mathbf{x}, \mathbf{r})) \in \mathcal{CG}_{k,b}$ and the hyperedge potential $\varphi^G_{k,T}$.

(S) Stability is satisfied since φ is non-negative.

(U) We choose M and Γ as in Subsubsection A.1.1.

(U1) Since the radius of the set Δ of any $(\eta, (\mathbf{x}, \mathbf{r}))$, $\eta \in \mathcal{LD}_4$, $(\mathbf{x}, \mathbf{r}) \in \bar{\Gamma}^A$, can be bounded uniformly by $\sqrt{R_0^2 + R_{\max}^2}$ the radius of the Laguerre k-flower LF_k is bounded by $(k+1)\sqrt{R_0^2 + R_{\max}^2}$. Further by Lemma 19 the bound for r_Γ can be taken as $2(k+1)\sqrt{R_0^2 + R_{\max}^2}$.

(U2) $c_A^+ = \sup_{(\mathbf{x},\mathbf{r}) \in \bar{\Gamma}^A} \sum_{\eta \in \mathcal{CG}_{k,b}(\mathbf{x},\mathbf{r}), \eta_x \cap C \neq \emptyset} \frac{\varphi^G_{k,T}(\eta,(\mathbf{x},\mathbf{r}))}{card(\eta_x)} \leq \sum_{\eta \in \mathcal{CG}_{k,b}(\mathbf{x},\mathbf{r}), \eta_x \cap C \neq \emptyset} \frac{K}{k}$, where the right hand side is finite since the sum is finite. Indeed the number of incident edges (vertices) to each vertex is at most 96, cf. proof of the previous Theorem 20. To each of these at most 96 vertices we can take again all his incident vertices. Repeating this $k-2$ times we obtain k-neighbourhood which is finite and we have finitely many possibilities how to choose k-tuple of vertices from it.

(U3) The lower bound on z is gained from the condition $z|A| > e^{c\mathrm{r}}$, cf. Remark 3.2.2. Since $|A| = \frac{1}{3}\pi a^5 \rho^3 (1-2\rho)^2$ the inequality results in $z > C_0 \frac{1}{a^5} e^{C_1}$, where $C_0, C_1 > 0$ are constants not depending on a. Minimizing the right side over $a > 0$ we obtain $z > 0$ as $a \to \infty$.

□

Corollary. The statement of the Theorem 21 holds for $T = \mathrm{dsc}(H, H')$ and $K = \infty$.

Proof. Note that in the case of $T = \mathrm{dsc}(H, H')$, it holds that $\varphi^G_{k,T} \leq K \wedge 2$. □

Remark (On sublinearity in Delaunay tessellation). Note that, in 2D we have sublinearity, see Definition 16, of the cardinalities of the Delaunay edges and triangles thanks to the Euler's formula. This would allow the parametrization of hyperedge potentials in 2D analogues of Theorems 20 and 21 by a negative parameter. Unfortunately, the similar statement in the higher dimensions is not so obvious and no result about its validity is known.

3.3 Parameter estimation

Let $\tilde{W} \subset W \subset \mathbb{R}^3$ be bounded Borel sets and assume that we have observed a tessellation on the larger set (observation window) W. We consider a parametric model for generators, cf. (3.1), where the energy function $E(\mathbf{x}, \mathbf{r})$ depends on a parameter vector $\theta = (\theta_1, \ldots, \theta_n)$ and possibly on some hardcore parameters ε, e.g., $\varepsilon = (\alpha, \beta, B)$, see (3.3). Thus we write $E(\mathbf{x}, \mathbf{r}) = E((\mathbf{x}, \mathbf{r}); \varepsilon, \theta)$ to emphasize the dependency on parameters. The activity $z > 0$ is considered as an unknown parameter, too. Methods for parameter estimation were suggested in Dereudre and Lavancier [2011] for Gibbs-Voronoi tessellations in $\mathbb{R}^2$. They can be extended straightforwardly to Gibbs-Laguerre tessellations in $\mathbb{R}^3$, at least when the values of the parameters are in a certain range.

In order to estimate the parameters z and θ we need to distinguish whether the energy function of Gibbs point process is hereditary, cf. Definition 30, or not. In Gibbs-type tessellations the heredity is frequently violated because of hard-core conditions. For example, when we delete a generator of an admissible configuration, it may happen that one of the new cells is too big to satisfy the hard-core condition given by the parameter β. Therefore, we introduce the notion of removable points $\mathcal{R}(\mathbf{x}, \mathbf{r})$ from the marked point configuration $(\mathbf{x}, \mathbf{r})$, cf. Definition 36. To emphasize the dependency on hardcore parameters ε we write $\mathcal{R}^\varepsilon$.

3.3.1 Pseudolikelihood for Gibbs-Laguerre tessellations

First, focus on estimation of parameters z and θ provided that the hardcore parameters are known or estimated in advance. The parameters z, θ are then estimated by means of the maximum pseudolikelihood method, cf. Møller and Waagepetersen [2003], where the (estimated) values of the hard-core parameters are plugged-in.

A derivation of the log pseudolikelihood contrast function (recall Definition 54)

$$cl\mathcal{PL}_{\tilde{W}} := -\log \mathcal{PL}_{\tilde{W}}$$

is in hereditary setting briefly described in Møller and Waagepetersen [2003]. A generalization of the maximum pseudolikelihood method to the non-hereditary case is described in Dereudre and Lavancier [2009] – he presence of hardcore parameters means that the log pseudolikelihood contrast function $cl\mathcal{PL}_{\tilde{W}}$ involves summing up over the set of removable points only. We state the log pseudolikelihood contrast function for a marked point realization $(\mathbf{x}, \mathbf{r}) \in \mathbf{N}_{\tilde{W}}$, compare this with Definition 54:

$$\begin{aligned} cl\mathcal{PL}_{\tilde{W}}(z, \varepsilon, \theta; (\mathbf{x}, \mathbf{r})) = &\int_{\tilde{W} \times \mathbb{M}} z \exp(-h((y,s), (\mathbf{x}, \mathbf{r}); \varepsilon, \theta)) d(y,s) \\ &+ \sum_{(y,s) \in \mathcal{R}^{\varepsilon}(\mathbf{x}, \mathbf{r}) \cap (\tilde{W} \times \mathbb{M})} (h((y,s), (\mathbf{x}, \mathbf{r}) \setminus \{(y,s)\}; \varepsilon, \theta) - \ln(z)). \end{aligned} \tag{3.13}$$

The maximum pseudolikelihood estimates of θ and z are then obtained as $(\hat{z}, \hat{\theta}) = argmin_{z,\theta}\ cl\mathcal{PL}_{\tilde{W}}(z, \hat{\varepsilon}, \theta; (\mathbf{x}, \mathbf{r}))$. This optimization requires the computation of the local energy h, cf. Definition 35, that requires the knowledge of the configuration outside $\tilde{W}$. This can be solved by taking $\tilde{W}$ to be a strict subset of the observation window W.

Differentiation of the log pseudolikelihood contrast function $cl\mathcal{PL}_{\tilde{W}}$ in (3.13) yields the representation of z given by

$$\hat{z} = \frac{\mathrm{card}(\mathcal{R}^{\hat{\varepsilon}}(\mathbf{x}, \mathbf{r}) \cap (\tilde{W} \times \mathbb{M}))}{\int_{\tilde{W} \times \mathbb{M}} \exp(-h((y,s), (\mathbf{x}, \mathbf{r}); \hat{\varepsilon}, \theta)) d(y,s)}, \tag{3.14}$$

and the equations

$$\begin{aligned} z \int_{(\tilde{W} \times \mathbb{M})'} &\frac{\partial h}{\partial \theta_j}((y,s), (\mathbf{x}, \mathbf{r}); \hat{\varepsilon}, \theta) \exp\left(-h((y,s), (\mathbf{x}, \mathbf{r}); \hat{\varepsilon}, \theta)\right) d(y,s) = \\ &= \sum_{\mathcal{R}^{\hat{\varepsilon}}(\mathbf{x}, \mathbf{r}) \cap (\tilde{W} \times \mathbb{M})} \frac{\partial h}{\partial \theta_j}((y,s), (\mathbf{x}, \mathbf{r}) \setminus \{(y,s)\}; \hat{\varepsilon}, \theta), j = 1, \dots, q, \end{aligned} \tag{3.15}$$

where $(\tilde{W} \times \mathbb{M})' = \{(y,s) \in \tilde{W} \times \mathbb{M} : (\mathbf{x}, \mathbf{r}) \cup \{(y,s)\}$ is admissible$\}$. In practice, we replace z in (3.15) by the right-hand side of (3.14) and obtain q equations whose solution $\hat{\theta}$ gives an estimate of θ. Then we compute $\hat{z}$ by plugging $\hat{\theta}$ into (3.14). The integrals in (3.14) and (3.15) have to be evaluated numerically, using, e.g., Monte Carlo techniques. The equations (3.14) and (3.15) are then solved numerically by the Newton-Raphson method, Magrenan and Argyros [2018].

3.3.2 Estimation of hardcore parameters

Let $\tilde{W} \subset W \subset \mathbb{R}^3$ be bounded Borel sets and similarly as in (3.3) let $h_{min}(C)$, $h_{max}(C)$ be the minimum, maximum distance between the barycenter of the cell C and any of its faces, respectively. Assume that we observe the Laguerre tessellation $L(\mathbf{x}, \mathbf{r})$ given by the generators $(\mathbf{x}, \mathbf{r}) \in \mathbf{N}$ on a window $W \in \mathcal{B}_b^3$ and that all cells intersecting $\tilde{W}$ are fully contained in W. The estimate of the hardcore parameter $\varepsilon = (\alpha, \beta, B)$ is $\hat{\varepsilon} = (\hat{\alpha}, \hat{\beta}, \hat{B})$, where

$$\begin{aligned}
\hat{\alpha} &= \min \left\{ w_{min}(C) : C \in L(\mathbf{x}, \mathbf{r}) \text{ intersecting } \tilde{W} \right\}, \\
\hat{\beta} &= \max \left\{ w_{max}(C) : C \in L(\mathbf{x}, \mathbf{r}) \text{ intersecting } \tilde{W} \right\}, \\
\hat{B} &= \max \left\{ \frac{w_{max}^3(C)}{|C|} : C \in L(\mathbf{x}, \mathbf{r}) \text{ intersecting } \tilde{W} \right\}.
\end{aligned}$$

Their consistency w.r.t. an unboundedly expanding observation window $\tilde{W}$ can be shown analogously to the proof given in Dereudre and Lavancier [2011].

3.4 Gibbs-Laguerre tessellations

The Gibbs-Laguerre tessellation model (3.5) introduced in Section 3.1.2 is used in case of Laguerre tessellation data representing the Aluminium alloy microstructure from Section 1.3.2. Results of this section are published in Seitl et al. [2021].

3.4.1 Modelling and reconstruction

The aim is to create models of random tessellations whose realizations are similar to the experimental data. The first approach to do so is to estimate the parameter values of various Gibbs-Laguerre tessellations chosen a priori, using standard techniques like the maximum pseudolikelihood method as described in Section 3.3.1. We tested this approach first on simulated data, see Section 3.7.2, Table 3.8. The results of pseudolikelihood method are not very accurate, especially for the activity z. Some practical aspects regarding the problem of estimation by the maximum pseudolikelihood method are mentioned in [Dereudre and Lavancier, 2011, Sections 4 and A.3].

An alternative approach inspired by the statistical reconstruction method (Section 3.5) is preferred in this section. The key element of this approach is so called reconstructing potential (3.5). We rely on distributional properties of geometric characteristics of the grains within the microstructure and distinguish two types of ‘reconstruction’. In Section 3.4.2 we focus on moments, especially on the mean and variance. We demonstrate that we can simulate tessellations with pre-specified moment properties on an example of two grain characteristics. In Section 3.4.3 the aim is not only to ‘reconstruct’ moment properties of chosen grain characteristics but their entire distribution.

In both cases, the reconstruction is done using the birth-death-move algorithm, Algorithm 2, but instead of a priori fixed number of iterations S it uses a ‘stopping criterion’. In brief, the recontruction stops when there is no change in the energy value greater than δ for some time t. More details about the stopping criterion can be found in Section 3.4.4.

The models depend on two parameters z and θ. The activity parameter z is fixed in advance such that its value corresponds approximately to the desired number of Laguerre cells within simulated tessellations. Further, its importance drops when either the first moment or whole distribution of the cell volume is about to be reconstructed – this is because of constraint that cell volumes always sum up to the volume of the observation window W. In this section, we set $z = 2000$. The choice of the parameters θ is done heuristically: a sequence of increasing parameter values is considered and the lowest values giving satisfactory results are chosen in the final model; this will be demonstrated in Section 3.4.3.

Table 3.1: Means and standard deviations of various tessellation characteristics of the aluminium alloy data, the third column, are normalized to the unit volume, the fourth column. Normalization means that volume-based, area-based and length-based characteristics are divided by the volume, the square root of the volume and the cube root of the volume of the window W, respectively.

	W	$486 \times 529 \times 685$	$[0,1]^3$
radius $[\mu m]$	mean	29.77	$5.31{\cdot}10^{-2}$
	sd	18.96	$3.38{\cdot}10^{-2}$
	$\mathbb{M}$	[0,70]	[0, 0.125]
vol $[\mu m^3]$	mean	$1.68 \cdot 10^5$	$9.6712 \cdot 10^{-4}$
	sd	$1.53 \cdot 10^5$	$1.0782 \cdot 10^{-4}$
dvol $[\mu m^3]$	mean	$1.89{\cdot}10^5$	$1.07{\cdot}10^{-3}$
	sd	$1.57{\cdot}10^5$	$8.92{\cdot}10^{-4}$
surf $[\mu m^2]$	mean	$1.56{\cdot}10^4$	1.18
	sd	$9.21{\cdot}10^3$	$6.94{\cdot}10^{-1}$
tel $[\mu m]$	mean	$8.59{\cdot}10^2$	1.53
	sd	$3.84{\cdot}10^2$	$6.85{\cdot}10^{-1}$

The last comment concerns the observation window W. For the purposes of the reconstructions, we normalize the window W to the unit window $[0,1]^3$. This means that the volume based characteriristics, as vol and dvol, are divided by the volume of the cuboidal domain of the experimental data, i.e., by $|W|$. Further, areas and lengths are normalized by the square and cube root of the volume the cuboidal domain, respectively. Dimensionless characteristictics as nof, spher, etc. remain unchanged. Those characteristics which value has changed by normalization are in Table 3.1 - the new values are compared to the original values from Table 1.4. Specially, the radii are divided by the cube root of $|W|$, and thus we obtain that radii are smaller than 0.124 (which corresponds to 70, cf. Fig. 1.8(b), before normalization). Hence, the space of marks $\mathbb{M}$ can be set to interval $[0, 0.125]$. The symbols $hist_{\text{nof}}^{exd}$ and $hist_{\text{vol}}^{exd}$ denote the relative histogram of the number of faces and the relative histogram of the cell volume from Figure 1.8, respectively. The upper index "exd" means that the histogram corresponds to experimental data. We use the same notation even for their normalized counterparts, which are assumed in the rest of this section.

3.4.2 Moment reconstruction

First, our goal is to reconstruct the Laguerre tessellation data using moments. We will show that realizations with prescribed empirical moments of various cell characteristics can be sampled. In particular, we investigate the first and the second moment together. We specify three models, where the first two consider either solely the number of faces per cell, (3.16), or solely the cell volume, (3.17), whereas in the third model both characteristics are involved, (3.18). The models are based on the reconstructing potentials $V_{n,T}^{G}$, (3.5). The energy function of the first two Gibbs-Laguerre models combines two reconstructing potentials, namely (together with the parameter specification)

$$\begin{aligned} G_1 &= \text{nof}, T_1(s_{G_1}) = \bar{s}_{G_1}, s_{0,1} = 14.1608, \theta_n^1 = 10\ 000, \\ G_2 &= \text{nof}, T_2(s_{G_2}) = S^2_{s_{G_2}}, s_{0,2} = 23.5783, \theta_n^2 = 1\ 000, \end{aligned} \tag{3.16}$$

in the case of nof characteristic and

$$\begin{aligned} G_1 &= vol, T_1(s_{G_1}) = \bar{s}_{G_1}, s_{0,1} = 9.6712 \cdot 10^{-4} [\mu m^3], \theta_n^1 = 100\ 000, \\ G_2 &= vol, T_2(s_{G_2}) = S^2_{s_{G_2}}, s_{0,2} = 1.1626 \cdot 10^{-8} [\mu m^3], \theta_n^2 = 500\ 000, \end{aligned} \tag{3.17}$$

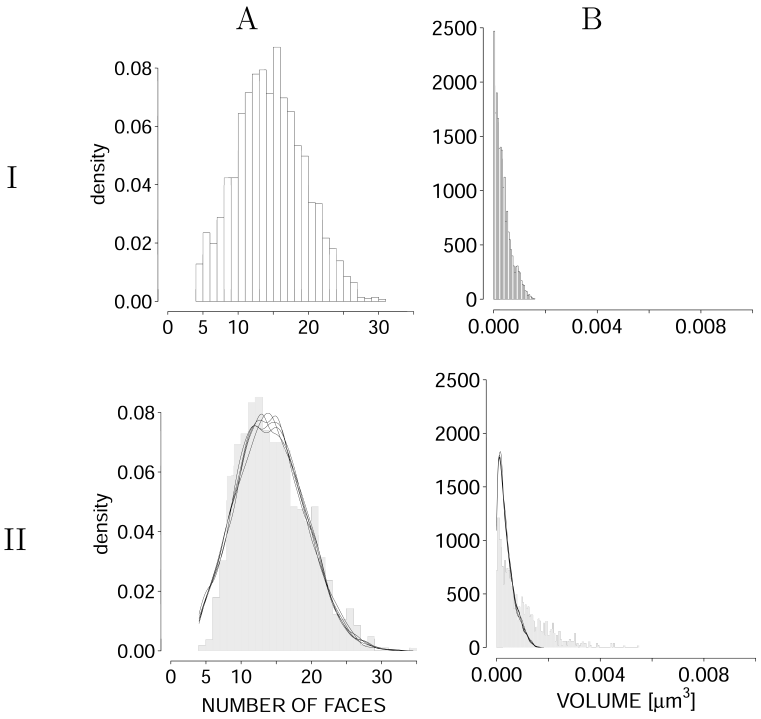
in the case of vol characteristic. The symbol $S^2_{s_G}$ denotes the sample variance $\frac{1}{n-1}\sum_C (s_G(C) - \bar{s}_G)^2$. The stopping criterion which we use in both cases is $(\delta, t) = (0.002, 500\ 000)$.

Figures 3.3a) and 3.3b) show results of the reconstructions (3.16) and (3.17), respectively. Both figures show histograms of the number of faces per cell and histograms of the cell volume in columns A nad B, respectively. In row I, there are histograms computed from one reconstructed tessellation. In row II, there are gray histograms computed from the data together with kernel estimates of the densities evaluated from ten reconstructions. Table 3.2 depicts the numerical values obtained. In case of reconstructions (3.16), we see (compare values in Tables 3.2 and Table 1.4) that the mean and standard deviation of number of faces are similar to those obtained from data. On the other hand, the mean and standard deviation of volume do not match perfectly. In case of reconstructions (3.17), we see conversely (as expected, the model is based on vol instead of nof) that the mean and standard deviation of number of faces differ and the mean and standard deviation of volume are very similar.

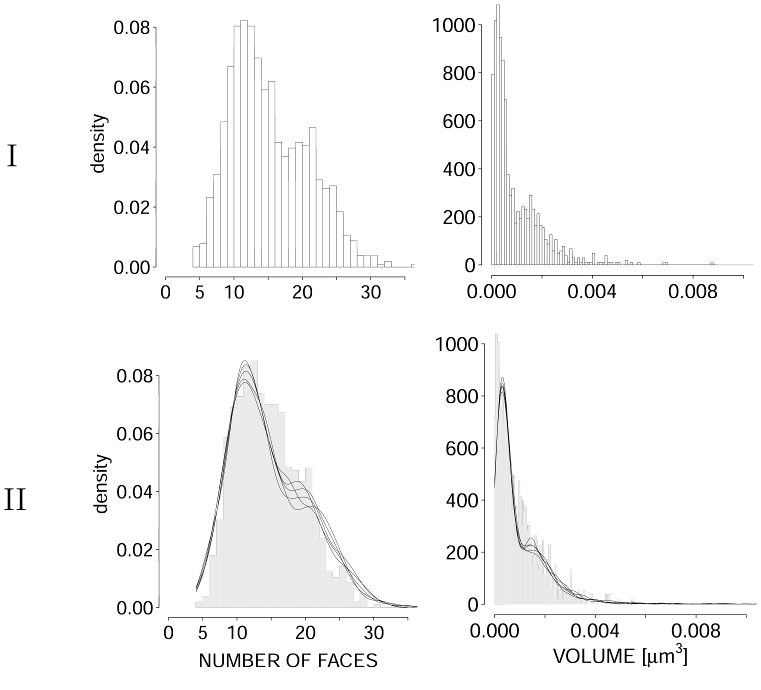
Table 3.2: Moment reconstruction - numerical values. Two parameter specifications (3.16) and (3.17) controlling the first two moments of nof and the first two moments of vol, respectively, were used to obtain realizations described in this table.

	number of cells	nof		vol $[\mu m^3]$	
		mean	sd	mean	sd
spec. (3.16)	2810	14.1609	4.8557	$3.5587 \cdot 10^{-4}$	$3.0310 \cdot 10^{-4}$
spec. (3.17)	1033	14.6583	5.6277	$9.6805 \cdot 10^{-4}$	$1.0782 \cdot 10^{-4}$

Until now random tessellations considering either solely the number of faces per cell or solely the cell volume were investigated. The next goal is to reconstruct



a) The reconstruction using the first two moments of nof with parameter specification given in (3.16).



b) The reconstruction using the first two moments of vol with parameter specification given in (3.17).

Figure 3.3: The columns A and B show number of faces and volume [μm^3], respectively. Relative histograms on row I are computed from a single realization. To show the variability of the model row II shows kernel estimates of the densities computed from ten realizations together with histograms coming from the experimental data—in gray.

the first two moments of the number of faces per cell and the cell volume at once. This means that the energy function of the third Gibbs-Laguerre model consists of four potentials. The objective is met by random tessellation having the specification:

$$\begin{array}{llll} G_1 = \text{nof}, & T_1(s_{G_1}) = \bar{s}_{G_1}, & s_{0,1} = 14.1608, & \theta_n^1 = 100\,000, \\ G_2 = \text{nof}, & T_2(s_{G_2}) = S^2_{s_{G_2}}, & s_{0,2} = 23.5783, & \theta_n^2 = 500\,000, \\ G_3 = \text{vol}, & T_3(s_{G_3}) = \bar{s}_{G_3}, & s_{0,3} = 9.6712 \cdot 10^{-4}[\mu m^3], & \theta_n^3 = 10\,000, \\ G_4 = \text{vol}, & T_4(s_{G_4}) = S^2_{s_{G_4}}, & s_{0,4} = 1.1626 \cdot 10^{-8}[\mu m^3], & \theta_n^4 = 8\,000\,000. \end{array} \tag{3.18}$$

The stopping criterion which we use is $(\delta, t) = (0.002, 500\,000)$.

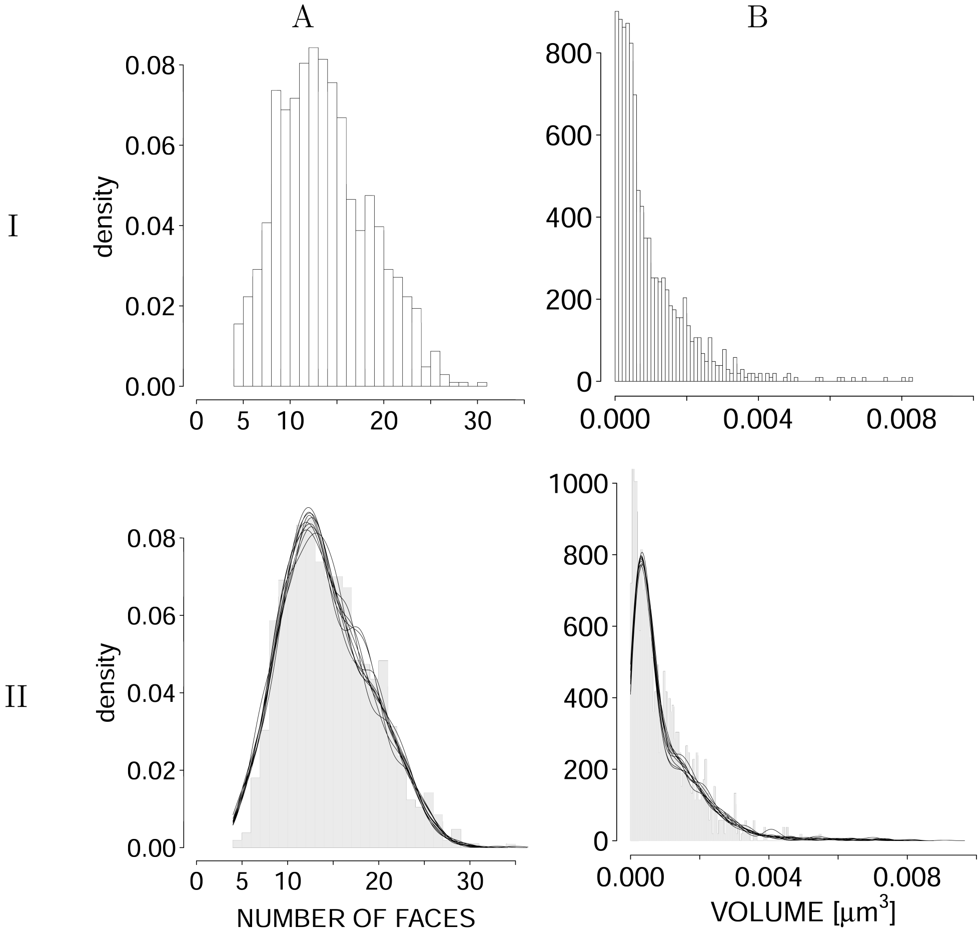


Figure 3.4: The reconstruction using the first and the second moments of both, nof and vol, model (3.18) – column A and B show relative histograms and densities of number of faces and volume $[\mu m^3]$, respectively; the histograms of relative frequencies in row I are computed for one realization; to illustrate the variability of the random tessellation, row II shows densities computed for a bunch of ten realizations together with histograms coming from the experimental data—in gray.

Figure 3.4 shows relative histograms of both the numbers of faces per cells and the cell volumes. Moreover, it contains densities of both characteristics per ten realizations to illustrate the variability of the characteristics. The numerical

Table 3.3: Moment reconstruction of the experimental data – model (3.18) (values from the experimental data are given in brackets).

number of cells	nof		vol [μm^3]	
	mean	sd	mean	sd
1032	14.1609	4.8558	$9.6899 \cdot 10^{-4}$	$1.0499 \cdot 10^{-4}$
(1034)	(14.1608)	(4.8558)	$(9.6712 \cdot 10^{-4})$	$(1.0782 \cdot 10^{-4})$

values obtained are presented in Table 3.3. With an increasing number of potentials combined in the energy function it is more difficult to get close to the prescribed values, but even for the potential of the third model (3.18) the results are satisfactory. We can compare the first two empirical moments of both, number of faces and volume, as reached in reconstructions with those obtained from data, cf. Table 1.4, and conclude that they are quite similar.

3.4.3 Histogram reconstruction

A more sophisticated approach to the reconstruction of tessellations is to control not only a few moments but the entire distribution of a geometrical characteristic. The easiest way to accomplish this is to measure the discrepancy between histograms, see Fig. 3.1. We will consider three different models. The first two controls the distribution of a single cell characteristic. The energy function of both is given by the reconstructing potential $V^G_{n,T}$, (3.5), and we set

$$G = \text{nof}, \quad T(s_G) = \text{dsc}(H_{s_G}, H'_{s_G}), \quad H'_{s_{\text{nof}}} = hist^{exd}_{\text{nof}}, \quad \theta_n, \tag{3.19}$$

in the case of the first model, and

$$G = \text{vol}, \quad T(s_G) = \text{dsc}(H_{s_G}, H'_{s_G}), \quad H'_{s_{\text{vol}}} = hist^{exd}_{vol}, \quad \theta_n, \tag{3.20}$$

in the case of the second model. In both models, H_{s_G} stands for the empirical histogram of the characteristic G computed within simulations. In addition to providing the results, we examine the choice of parameter θ_n. The stopping criterion which we use here is $(\delta, t) = (0.002, 500\ 000)$.

Table 3.4 and Figure 3.5 show how the value of θ_n influences the distribution of number of faces within the simulated realizations under model (3.19). The table shows that with increasing θ_n the histogram discrepancy of the number of faces per cell $\text{dsc}(H_{s_{\text{nof}}}, hist^{exd}_{\text{nof}})$ decreases. The reconstructing potential considers only the number of faces per cell (nof); therefore, the discrepancy of the histograms of cell volume $\text{dsc}(H_{s_{\text{vol}}}, hist^{exd}_{\text{vol}})$ is not controlled. In summary, a small value of θ_n results in a large discrepancy $\text{dsc}(H_{s_{\text{nof}}}, hist^{exd}_{\text{nof}})$. On the other hand, increasing θ_n beyond a certain level leads to no further improvement, because the probabilities of acceptance in Algorithm 1 tend to zero. The figure compares empirical kernel density estimates computed from simulations with histogram $hist^{exd}_{\text{nof}}$. We see, that with increasing value of θ_n the curves of the density estimates resemble the shape of histogram $hist^{exd}_{\text{nof}}$ more and more. The variability of the realizations is higher for small values of θ_n and it decreases when θ_n grows. The results should be compared to the experimental data, see Figure 1.8.

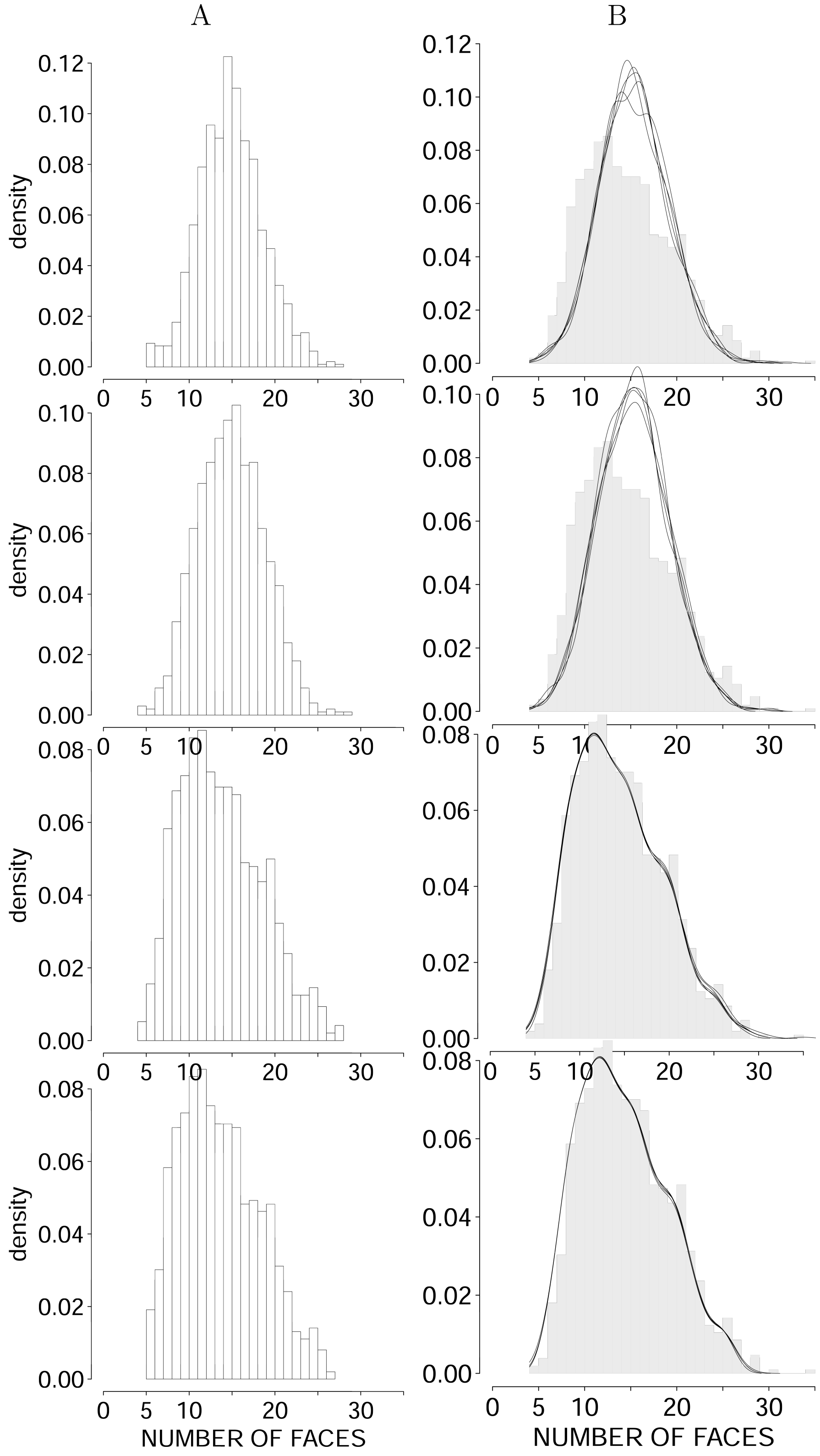


Figure 3.5: Reconstruction of experimental data, controlling the distribution of the numbers of faces per cell, model (3.19): from top to bottom the parameter θ_n takes on the values 10, 100, 1 000 and 10 000; column A shows the histograms of relative frequency computed from a single realization, and column B shows kernel density estimates based on ten realizations together with the histogram $hist_{\mathrm{nof}}^{exd}$ coming from the experimental data (cf. Figure 1.8)—in gray.

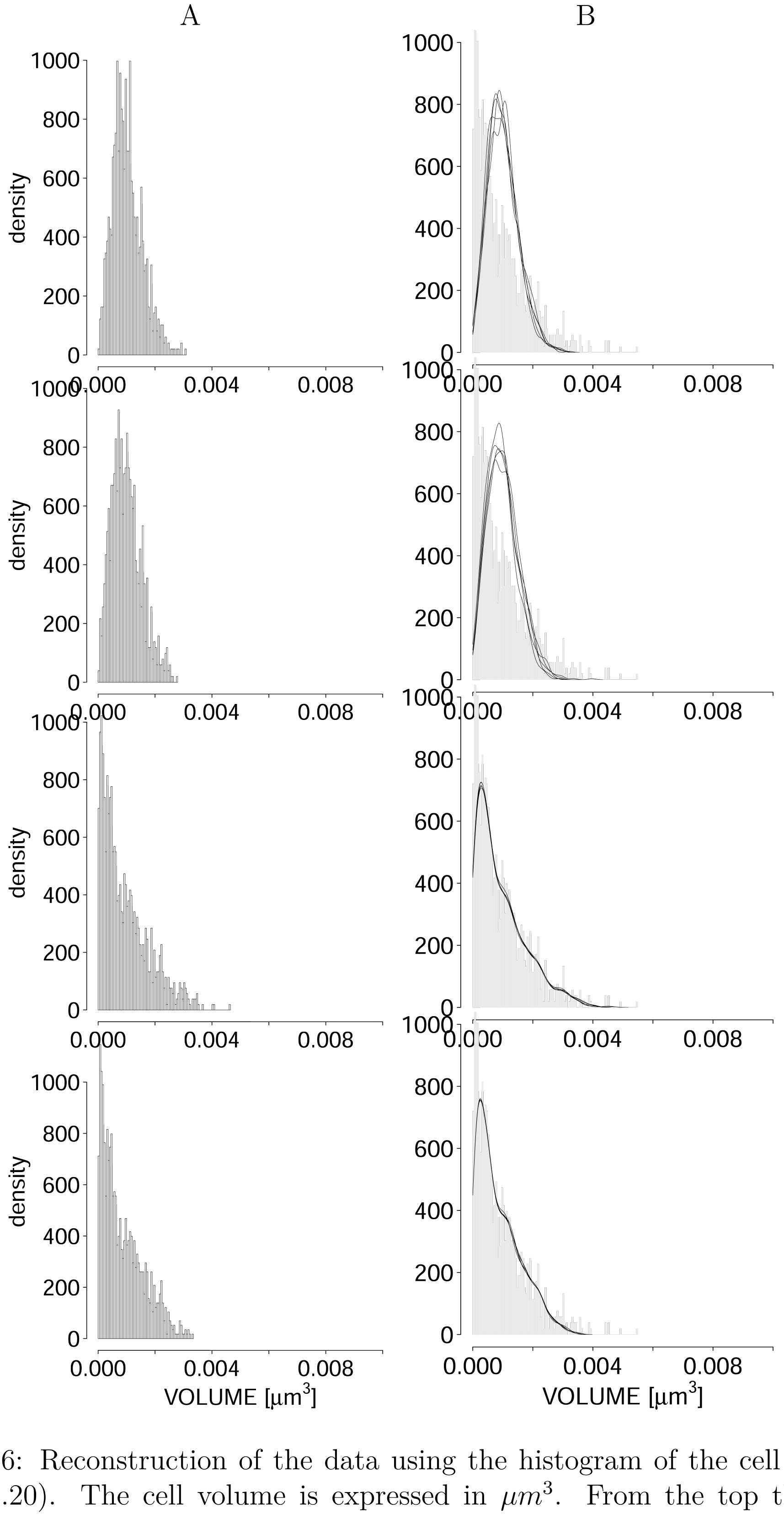


Figure 3.6: Reconstruction of the data using the histogram of the cell volume, model (3.20). The cell volume is expressed in μm^3. From the top the value of parameter θ_n increases, the values are 10, 100, 1000 and 10000. Column A shows the evolution of the histogram of relative frequencies (computed from one realization) and column B shows the densities of ten realizations together with the targeting histogram $hist_{\text{vol}}^{exd}$, that comes from the experimental data, see Figure 1.8. The corresponding discrepancies are stated in Table 3.5.

Table 3.4: Dependence on θ_n of the discrepancy between histograms for the number of faces per cell and histograms for the cell volumes in model (3.19).

θ_n	discrepancy nof	 vol
10	0.46822	0.70841
100	0.41481	0.68919
1 000	0.02964	0.45739
10 000	0.02529	1.32095
100 000	0.02356	1.28102
1 000 000	0.02102	1.32543

A similar behaviour can be observed in case of model (3.20). Table 3.5 demonstrates what happens with the histogram discrepancy when the parameter θ_n is increasing. Figure 3.6 shows how the value of parameter θ_n influences the empirical density and variability of the cell volume within simulated realizations.

Table 3.5: Dependence on θ_n of the discrepancy between histograms for the number of faces per cell and histograms for the cell volumes in model (3.20).

θ_n	discrepancy nof	 vol
10	0.49742	0.69294
100	0.46457	0.64480
1000	0.23373	0.03997
10000	0.20450	0.07205
100000	0.22514	0.08221
1000000	0.20899	0.06267

The third model combines the two previous models, i.e., it aims at reconstructing both cell characteristics:

$$\begin{aligned} G_1 &= \text{nof}, \quad T_1(s_{G_1}) = \text{dsc}(H_{s_{G_1}}, H'_{s_{G_1}}), \quad H'_{s_{\text{nof}}} = hist^{exd}_{\text{nof}}, \quad \theta_n^1, \\ G_2 &= \text{vol}, \quad T_2(s_{G_2}) = \text{dsc}(H_{s_{G_2}}, H'_{s_{G_2}}), \quad H'_{s_{\text{vol}}} = hist^{exd}_{\text{vol}}, \quad \theta_n^2, \end{aligned} \tag{3.21}$$

In addition to providing the results, we examine the choice of parameters θ_n^1, θ_n^2 in detail. Once again, the stopping criterion employed is $(\delta, t) = (0.002, 500\,000)$.

The conclusions made based on models (3.19) and (3.20) remain valid even for model (3.21), which combines two reconstructing potentials based on the histogram discrepancy. Combining more than one potential causes some difficulties. The values of both parameters have to be in a reasonable proportion as described in Table 3.6. Moreover, it is easy to see that the value of the parameter corresponding to the histogram of cell volumes must be the larger of the two. Figure 3.7 shows the reconstruction results in case of $\theta_n^1 = 1\,000$ and $\theta_n^2 = 10\,000$. The results should be compared to the experimental data, see Figure 1.8, in order to verify the success of the reconstruction visually.

Table 3.6: Dependence on the parameters θ_n^1 and θ_n^2 of the discrepancy of histograms for the number of faces per cell and for the cell volume in model (3.21).

$\theta_n^1 \backslash \theta_n^2$		10^3	10^4	10^5	10^6
10^2	nof	0.28966	0.21418	0.20543	-
	vol	0.36386	0.06484	0.05571	
10^3	nof	0.05294	0.07903	0.14432	0.13548
	vol	1.21265	0.08136	0.06971	0.06634
10^4	nof	-	0.01671	0.06802	0.09268
	vol		1.14779	0.09701	0.06514
10^5	nof	-	-	0.01327	0.05756
	vol			1.02774	0.09387

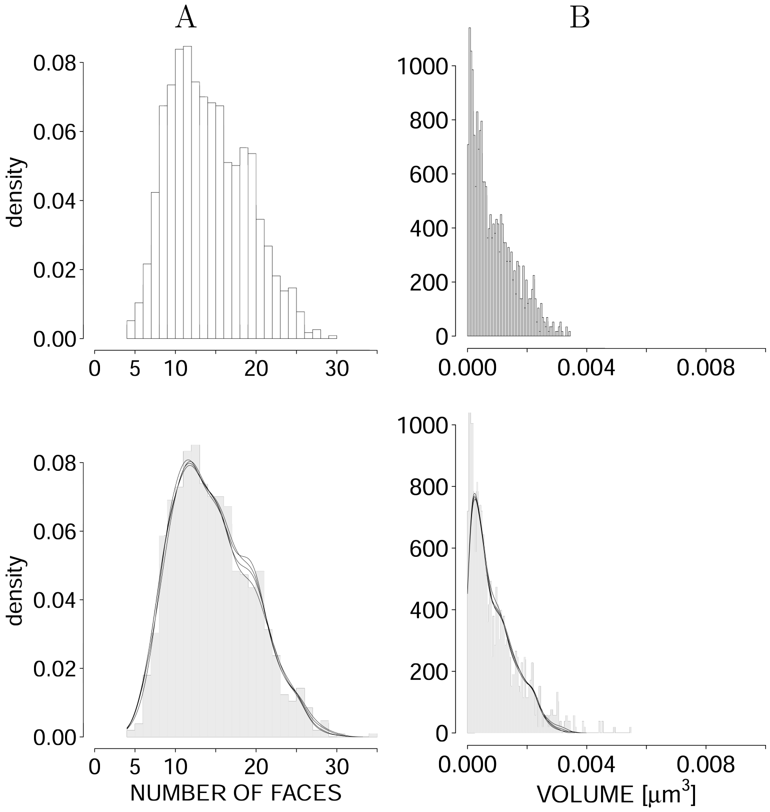


Figure 3.7: Reconstruction of experimental data, controlling distributions of the number of faces per cell and of the cell volume, model (3.21) with $\theta_n^1 = 1\,000$ and $\theta_n^2 = 10\,000$: column A shows the number of faces per cell, the histogram of relative frequencies computed from one realization and kernel density estimates based on ten realizations; column B shows the same plots for the cell volume. The gray histograms are those of the experimental data, i.e., $hist_{\mathrm{nof}}^{exd}$ and $hist_{\mathrm{vol}}^{exd}$, cf. Figure 1.8.

3.4.4 Summary of numerical results

In this section we have introduced two methods how to simulate Gibbs-Laguerre tessellations statistically similar to the experimental data. Recall, that the methods are inspired by statistical reconstruction of (marked) point pattern as in-

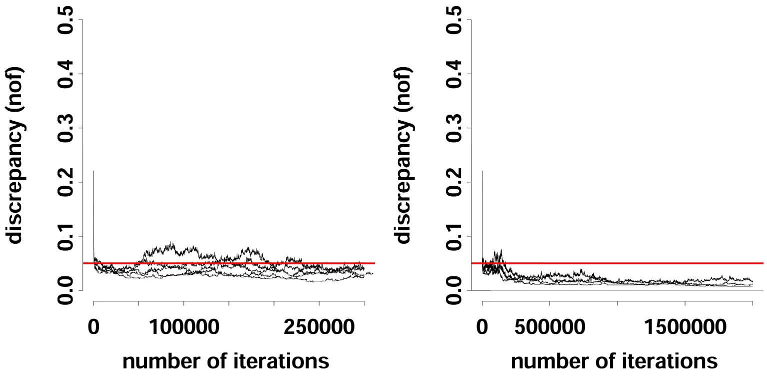


Figure 3.8: The comparison of the reconstruction stopped by stopping criterion $(\delta, t) = (0.02, 100000)$ and the simulation performing 2 millions steps. Both are based on model (3.19) with control parameter $\theta_n = 1000$. Two subfigures show the time evolution of the histogram discrepancy (energy) for 5 simulations. The stopping criterion (δ, t) results in shorter runs - all 5 reconstructions (on the left) were stopped after less than 320 thousands of iterations.

troduced in Tscheschel and Stoyan [2006] and described in Section 3.5. Both reconstruction methods have proved to be successful in sense that Laguerre tessellations similar to the experimetal data are simulated. The moment reconstruction method, cf. Section 3.4.2, produced samples with the first two empirical central moments of two cell characteristics being similar to the prescribed values (i.e., five-parameter model – activity and four control parameters). The number of parameters increases not only with the number of tessellation characteristic but with the number of considered moments as well. Therefore, the second method wants to keep the number of control parameters to be equal to the number of tessellation characteristics. The histogram reconstruction method, cf. Section 3.4.3, produced samples with empirical distribution of two cell characteristics being similar to the prescribed shape, i.e., only three-parameter model outperforming the moment reconstruction (the whole distribution is mimicked instead of the first two moments). It is clear that the first method cannot be so accurate as the second method, which can be viewed as a ‘limiting’ case in sense that we would need to consider all moments in the first method in order to obtain the same effect as the second method has. On the other hand, for small number of moments, the first method is usually less time demanding.

The time of reconstructions can be reduced in the other way – instead of using a priori fixed number of iterations in *Algorithm 2*, a stopping criterion (δ, t) can be used: the *Algorithm 2* is terminated when the energy remains in a band of the width δ for a time period t. A suitable choice of δ and t enables to shorten the runtime of the reconstructions significantly. In Figure 3.8 the shorter reconstruction terminated by the stopping criterion (δ, t) is compared with a much more longer simulation. Although the decrease of energy (in this case equivalent to a single discrepancy between histograms of the number of faces per cell) continues even after the event when the stopping criterion could be applied, it is very gradual. For these reasons, shorter and thus less time demanding reconstructions ended by the stopping criterion are preferred.

3.5 Statistical reconstruction

The concept of statistical reconstruction of (marked) point patterns was first introduced in Tscheschel and Stoyan [2006] (see Illian et al. [2008] for a textbook version). It uses no a priori stochastic model. The aim of the stochastic reconstruction is to generate point patterns with a priori chosen distributional characteristics close to those of a given point pattern (data). This is usually carried out by choosing an admissible entry point configuration (note that there can be some hard-core restrictions on the considered point configurations) and by running some iterative optimization method, e.g., greedy algorithm, Cormen et al. [2009], Chapter 16, or simulated annealing, van Laarhoven and Aarts [1987]. Depending on the particular method, the number of points can be fixed. The goal is to minimize the deviation of the distributional characteristics computed from the point pattern created by the iterative method and the point pattern representing the data. In each step of the optimization method the actual point configuration is proposed to be modified. The proposal is accepted according to the rules of the optimization method, e.g., in case of greedy algorithm only proposals leading to a decrease of the deviation are accepted. The resulting point pattern is taken as a reconstruction of the data pattern.

An example of the statistical reconstruction using greedy algorithm in case of marked point pattern follows. Choose distributional characteristics of the marked point pattern which we want to liken to the data $(\mathbf{x}_n, \mathbf{r}_n)$ and decide how to compute the deviation. We will denote the deviation of the distributional characteristics computed from $(\mathbf{x}_0, \mathbf{r}_0)$ and from the data $(\mathbf{x}_n, \mathbf{r}_n)$ by $E(\mathbf{x}_0, \mathbf{r}_0)$, $E : \mathbf{N}^f \to \mathbb{R}_+ \cup \{\infty\}$ (we use the same symbol as for the energy function of Gibbs point processes since the same functions can often be used in both cases). In the case when the periodic boundary conditions are applied we will write $\tilde{E}(\mathbf{x}_0, \mathbf{r}_0)$. Set $L \in \mathbb{N}$.

Algorithm 5 (Reconstruction via greedy algorithm).

1. construct an admissible marked point configuration $(\mathbf{x}_0, \mathbf{r}_0)$ of n marked points that generate only nonempty Laguerre cells,
2. choose a point (x, r) from $(\mathbf{x}_0, \mathbf{r}_0)$ at random, generate $(y, s) \sim Q(x, r)$ and set
$$(\mathbf{x}_1, \mathbf{r}_1) = (\mathbf{x}_0, \mathbf{r}_0) \setminus \{(x, r)\} \cup \{(y, s)\},$$
3. $(\mathbf{x}_0, \mathbf{r}_0) \leftarrow (\mathbf{x}_1, \mathbf{r}_1)$ if $\tilde{E}(\mathbf{x}_1, \mathbf{r}_1) < \tilde{E}(\mathbf{x}_0, \mathbf{r}_0)$,
4. if the point configuration $(\mathbf{x}_0, \mathbf{r}_0)$ has not changed over the last L iterations, then return $(\mathbf{x}_0, \mathbf{r}_0)$, else goto 3.

The function $\tilde{E}$ can be the periodic energy function of the form (3.1). The reconstruction starts by fixing the total number n of nonempty cells in the sampling window W. Then an admissible marked point pattern $(\mathbf{x}_0, \mathbf{r}_0)$ of n generators is sampled in $W \times \mathbb{M}$ such that it generates only nonempty Laguerre cells. In each iteration of *Algorithm 5* a random marked point $(x, r) \in (\mathbf{x}_0, \mathbf{r}_0)$ is chosen and proposed to be replaced by a new marked point (y, s). Here, the marked point (y, s) is drawn from a probability distribution $Q(x, r)$ possibly depending on (x, r), e.g., $Q(x, r) = Unif(W \times \mathbb{M})$. The replacement is carried out if the periodic energy of the proposal $(\mathbf{x}_1, \mathbf{r}_1)$ is smaller than the periodic energy of $(\mathbf{x}_0, \mathbf{r}_0)$. The non-emptiness of Laguerre cells is preserved by the hard-core potential (3.2). The reconstruction ends if there is no replacement in $L \in \mathbb{N}$ consecutive iterations.

3.6 Comparison of reconstruction approaches

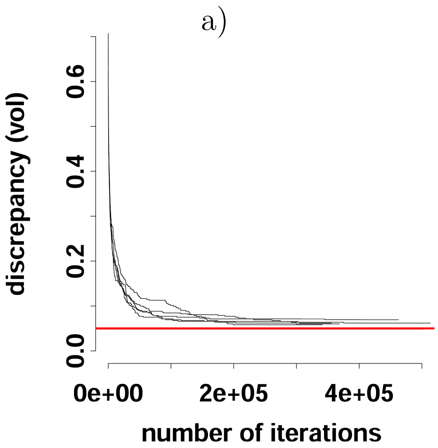
In this section, a short comparative study of reconstruction via MHBDM (*Algorithm 2* with stopping criterion (δ, t)) and the classical approach using the greedy algorithm described in Cormen et al. [2009], Chapter 16, is presented. The pseudocode briefly describing the greedy approach to statistical reconstruction is provided by Algorithm 5.

The reconstruction was carried out on the experimental data set introduced in Section 1.3.2 and encompassing 1049 nonempty cells. An important decision is which potentials will be incorporated in (3.1). We provide two comparisons, both defined by a single soft-core potential based on the histogram discrepancy.

The first comparison concerns the volumes of cells. Both reconstruction approaches consider the discrepancy (3.7) between the histogram of cell volumes of each generated tessellation and the corresponding histogram of the experimental data $hist^{exd}_{\mathrm{vol}}$ shown in Figure 1.8. In both algorithms there are some auxiliary parameters that need to be specified: namely, in *Algorithm 2* we set $\theta_n = 1000$ and $(\delta, t) = (0.01, 100000)$, and in *Algorithm 5* we set $n = 1049$, $Q(x, r) = Unif(W \times \mathbb{M})$ and $L = 50000$. In Figure 3.9 we observe that the discrepancy stops decreasing after 200 000 iterations in the case of the greedy reconstruction and after 50 000 iterations in the case of the MHBDM reconstruction. The computational time to arrive at this point is roughly the same for both approaches. The MHBDM reconstruction yields smaller discrepancies, as can be seen in Figure 3.9.

On the other hand, when dealing with histograms of the number of faces per cell (the second comparison), the results, cf. Figure 3.10, are better for the greedy reconstruction.

In the literature, cf. e.g., Illian et al. [2008], statistical reconstruction of point patterns is considered to be a non-parametric method. Our method interconnects statistical reconstruction with the simulation of Gibbs point processes and uses auxiliary parameters to control the precision of the fit. Altogether, there is a common step, *Algorithm 1*, that can be used in the simulation of marked Gibbs point processes and the reconstruction of marked point patterns. Note that, in contrast to the classical reconstruction as in Illian et al. [2008], the number of points in the reconstructed pattern does not need to be fixed. In summary, the benefit of the MHBDM reconstruction compared to the greedy algorithm introduced in Cormen et al. [2009] is that the former allows more flexibility in how close the reconstructed tessellation tracks the data. Moreover, since the MHBDM algorithm is not prone to getting stuck in local minima, a better fit can be achieved.



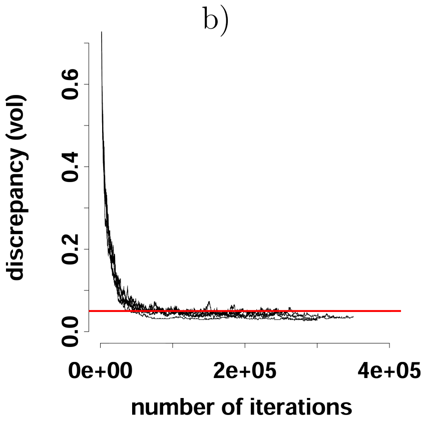


Figure 3.9: Evolution of the discrepancy of histograms of cell volume for (a) the greedy algorthm and (b) the MHBDM algorithm. The red line represents the discrepancy $y = 0.05$

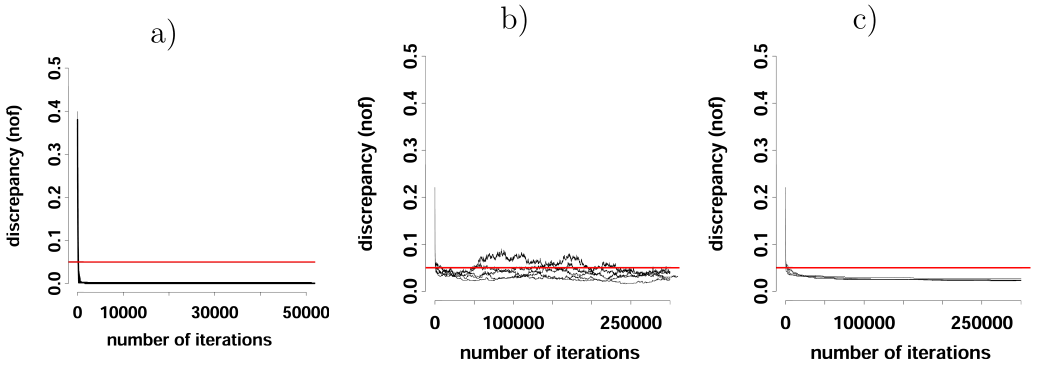


Figure 3.10: Evolution of the discrepancy of histograms of number of faces per cell for (a) the greedy algorithm, (b) the MHBDM algorithm with $\theta_n = 1\ 000$ and (c) the MHBDM algorithm with $\theta_n = 10\ 000$. The red line represents the discrepancy $y = 0.05$

3.7 Non-data based Gibbs-Laguerre tessellation models

Until now in this chapter we developed statistical models of Laguerre tessellations based on real datasets introduced in Section 1.3. In this section we introduce several non-data based models and examine somehow 'extreme' parameter specifications. In particular, we examine models based on the following geometrical characteristics: number of faces, NVR and sphericity. Section 3.7.1 shows that tessellations with rather extreme mean of number of faces can be simulated, this simulation study was published in Stoyan et al. [2021]. Section 3.7.2 examines the so called regular and irregular model based on NVR, where tessellations with similar and different volumes of neighbouring cells are obtained, respectively. Besides simulations it considers parameter estimation using Takacs-Fiksel method, the results were published in Seitl et al. [2021]. More simulation studies are presented within the attachments. Attachment A.2 shows histogram reconstruction of sphericity and Attachment A.3 compares chosen Gibbs-Laguerre tessellations to Poisson-Laguerre tessellations with the same intensity. If not stated otherwise, the activity parameter z of Gibbs-Laguerre tessellation models is set to be 1000.

3.7.1 Number of faces per cell

In this section tessellations with rather extreme mean number of nof are simulated in the observation window $W = [0, 560.524]^3$, units in μm. The size of the observation window is chosen to have the same volume as the observation window in case of alluminium alloy dataset, cf. Section 1.3.2. The energy function $E(\mathbf{x}, \mathbf{r})$ used here is of the form (3.1) and consists of three terms, which each aims to produce special geometric effects. We adapt the energy function to periodic boundary conditions, cf. Section 2.5, and write

$$\tilde{E}(\mathbf{x}, \mathbf{r}) = V_{hard} + \theta_1 \sum V_1(C) + \theta_n V_n(C_1, \dots, C_n).$$

The hardcore energy part V_{hard}, (3.2), is simply infinite if one or some marked points of $(\mathbf{x}, \mathbf{r})$ create empty Laguerre cells. That means that we avoid generators creating empty cells.

The repulsive potential $V_1(C)$ creates 'soft' distances between the generators. It has the form

$$V_1(C) = \begin{cases} \log(d/d_{\min}(C)) & \text{if } d_{min}(C) < d, \\ 0 & \text{otherwise,} \end{cases}$$

where $d_{min}(C)$ is the Euclidean distance from the generator of C to the nearest generator in $L(\mathbf{x}^*, \mathbf{r}^*) \setminus \{C\}$ and d is a parameter, we used $d = 22.42$.

The last energy term $V_n(C_1, \dots, C_n)$ is a kind of reconstructing potential (3.5) and helps to generate tessellations with a mean nof close to the prescribed value s_0. We set $\theta_1 = 1$ and $\theta_n = 1000$. We stopped the simulations after $200,000$ iterations, the target value s_0 was usually reached within $140,000$ iterations.

We present here results for two cases with the different parameters $s_0 = 12$ and $s_0 = 18$; we call the corresponding tessellations T12 and T18. These choices of s_0 aim to generate tessellations with mean values of nof 12 and 18, which so differ clearly in an important characteristic from the Poisson-Voronoi tessellation (where the mean nof is 15.54, see Okabe et al. [2000]). Our choice is not the only possible and similar results can be obtained for other cases.

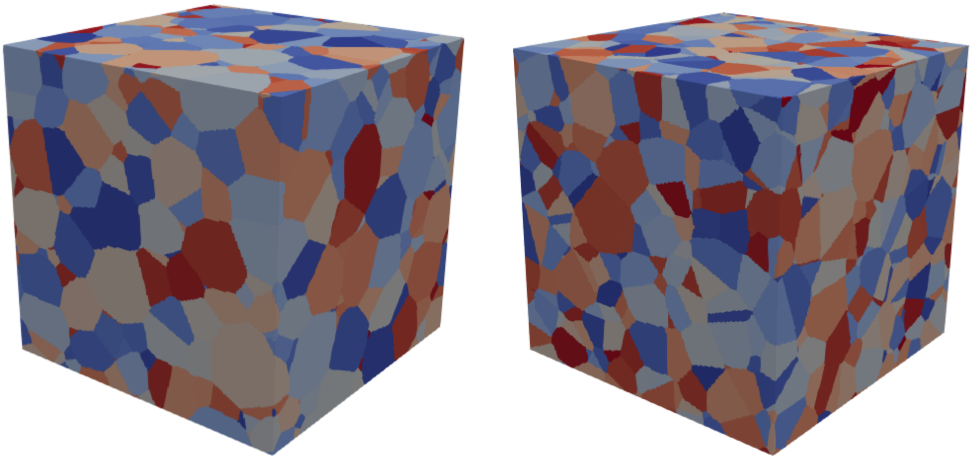


Figure 3.11: Samples of simulated Laguerre tessellations T12, left, and T18, right. The corresponding numbers of cells are 1040 and 1033, respectively. The cells are coloured randomly.

The tessellations T12 and T18 differ greatly in their geometrical structure, see Figure 3.11 for visual comparison, while the generating programs are identical up to the target parameter s_0. Our choice of the potential V_n and parameter θ_n made the simulations to 'brute force' simulations that really hit the target values $s_0 = 12$ and 18, see Table 3.7. This table evaluates five independent realizations of both T12 and T18 and moreover ten samples obtained by permutation of the radius marks from the first simulated sample. It shows in the columns 'data' and 'permuted' a large mean variability of the tessellation within the samples (rows II in Table 3.7), while the variability of the means of the characteristics for the whole tessellations (row III) is much smaller. We conclude that the simulated samples yield reasonably stable results for the mean characteristics of T12 and T18.

The averaged pair correlation functions of the simulated samples shown in Figure 3.12 are similar for both tessellations, but perhaps the range of correlation of the T18 pattern is a bit larger than that of T12. Also the mark variograms show only little difference in their course, see Figure 3.13, the corresponding ranges of correlation nearly coincide. The mark variance of T18 is smaller than that of T12. Both summary characteristics $g(r)$ and $\gamma_m(r)$ are similar to their counterparts for the data in Section 1.3.2. However, for the radius-mark correlation functions different results are obtained in case of T12 and T18. While Figure 3.14 shows that the mark correlation functions for T12 have a form comparable with those for the data introduced in Section 1.3.2 (point pairs close together tend to have small mark products), it has for T18 a form quite different. This indicates that in the patterns there is a tendency that generator pairs close together have mark products that tend to be larger than the square of the mean mark, which imply the values of $k_{mm}(r)$ larger than one. Probably, this is the price to pay for having the large mean nof of 18.

According to the aim to investigate mark correllation structure we compared

Table 3.7: Laguerre tessellation characteristics for samples T12 and T18. Five independently simulated samples and ten permuted samples from the first simulated sample are evaluated. In the upper part of the table there are the means and below the standard deviations (sd) of numbers of cells and faces per sample. In the lower part of the table there are for each characteristic three numbers. Firstly, I, the mean over all cells in all five samples (columns ‘data’) or ten samples (columns ‘permuted’). Further, II, the square root of the mean of variances within the samples and III, the square root of the variance of mean values within the samples (again with respect to five samples in columns ‘data’ and to ten samples in columns ‘permuted’).

per sample:		T12		T18	
		data	permuted	data	permuted
number of cells	mean	1019.6	789.90	992.41	884.50
	sd	32.01	16.36	30.96	5.244
number of faces	mean	6146.4	5544.6	8724.6	6479.5
	sd	580.3	100.9	320.6	64.45
per cell:					
nof	I	12.00	14.04	18.00	14.65
	II	8.59	6.84	4.49	5.97
	III	0	$8.87 \cdot 10^{-2}$	0	$6.80 \cdot 10^{-2}$
nov	I	20.00	24.08	32.00	25.30
	II	17.19	13.68	8.98	11.94
	III	0	$1.77 \cdot 10^{-1}$	0	$1.36 \cdot 10^{-1}$
vol	I	$1.74 \cdot 10^{5}$	$2.23 \cdot 10^{5}$	$1.82 \cdot 10^{5}$	$1.99 \cdot 10^{5}$
	II	$3.03 \cdot 10^{5}$	$2.99 \cdot 10^{5}$	$1.64 \cdot 10^{5}$	$2.29 \cdot 10^{5}$
	III	$1.59 \cdot 10^{4}$	$3.77 \cdot 10^{3}$	$6.77 \cdot 10^{3}$	$1.18 \cdot 10^{3}$
surf	I	$1.30 \cdot 10^{4}$	$1.78 \cdot 10^{4}$	$1.84 \cdot 10^{4}$	$1.77 \cdot 10^{4}$
	II	$1.68 \cdot 10^{4}$	$1.63 \cdot 10^{4}$	$9.77 \cdot 10^{3}$	$1.23 \cdot 10^{4}$
	III	$1.06 \cdot 10^{3}$	$2.46 \cdot 10^{2}$	$4.89 \cdot 10^{2}$	$9.17 \cdot 10^{2}$
spher	I	$6.86 \cdot 10^{-1}$	$7.37 \cdot 10^{-1}$	$7.53 \cdot 10^{-1}$	$7.62 \cdot 10^{-1}$
	II	$1.56 \cdot 10^{-1}$	$1.24 \cdot 10^{-1}$	$8.04 \cdot 10^{-2}$	$1.12 \cdot 10^{-1}$
	III	$3.20 \cdot 10^{-3}$	$3.55 \cdot 10^{-3}$	$1.65 \cdot 10^{-3}$	$1.52 \cdot 10^{-3}$

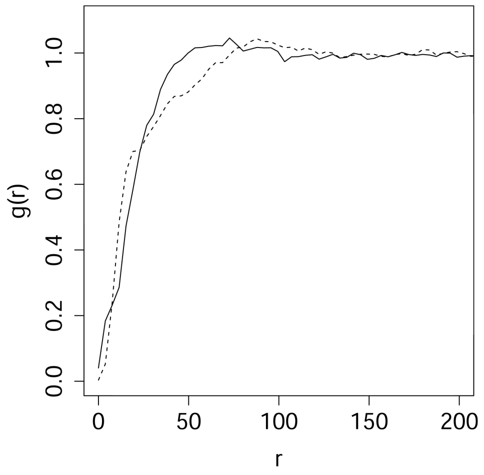


Figure 3.12: The empirical pair correlation functions averaged for the five simulated patterns. Solid line = T12, dashed line = T18.

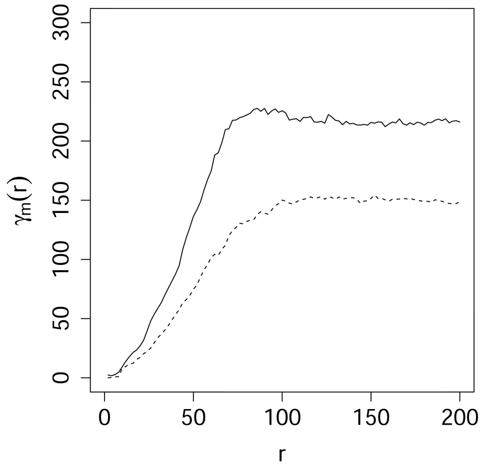


Figure 3.13: The empirical mark variograms $\gamma_m(r)$ averaged for the five simulated patterns. Solid line = T12, dashed line = T18.

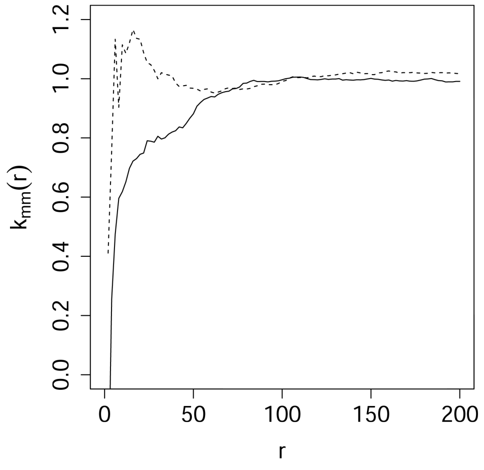


Figure 3.14: The mark correlation functions $k_{mm}(r)$ averaged for the five simulated patterns. Solid line = T12, dashed line = T18.

the values of mean nof for T12 and T18 with those for each ten tessellations with permuted radius marks (obtained for the first simulation). As shown in Table 3.7, for the permuted tessellations, we got the values of 14.04 (instead of 12) and 14.65 (instead of 18). We conclude that spatially correlated radius marks enable extreme values of mean nof such as 12 and 18, while independent marks with the same mark distribution and the same generators lead to tessellations with moderate values of number of faces, which may be seen as a form of 'regression to mediocrity'.

In order to understand the effect on the mark correlation function in the case of T18, we studied the spatial distributions of radius marks and of face numbers. We found for these geometrical characteristics a close correlation (large radius marks tend to generate large cells with many faces) and observed that pairs of generators close together with similarly large radius marks appear frequently, in the whole simulation window. The face numbers of large cells turned out to be so large that the small face numbers of their small neighbours were more than compensated.

3.7.2 Neighbour-volume ratio

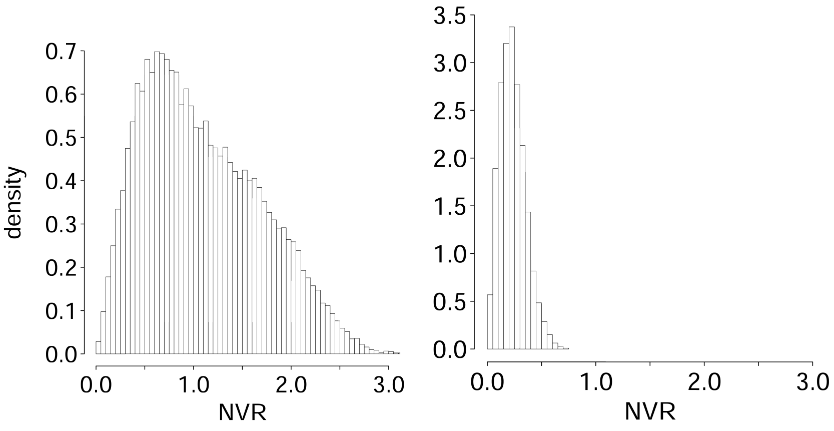
In this simulation study we consider unit observation window, i.e., $W = [0,1]^3$, and the energy function of the form (3.1) consisting of two terms:

$$\tilde{E}(\mathbf{x}, \mathbf{r}) = V_{hard} + \theta \sum V_2(C_1, C_2), \tag{3.22}$$

where the hard-core potential V_{hard} combines (3.2) and (3.3), i.e., it forbids empty Laguerre cells and impose some restrictions on cell sizes and shapes via three parameters α, β and B. The pair potential V_2 is minimum of neighbour-volume ratio (NVR) and a fixed constant $K > 0$, see (3.4). The constant K is important due to theoretical results of existence, in practise it is neglected as it is set to a really large positive number. Similar model is studied in Dereudre and Lavancier [2011] in case of 2D Voronoi tessellations. Here, we will use two parameter specifications differing in the value of θ, namely the regular model with $\theta = 1$ and the irregular model with $\theta = -1$, the choice $z = 2000$ is common to both models. Whether θ is positive or negative has a strong effect on the number of cells in W – in the irregular case, there are often many more cells than in the regular case, given a fixed value of z. Hardcore parameters in (3.3) can be used to reduce the range for number of cells in W significantly (in particular by using the bounds α and β). We set $\alpha = 0.02$, $\beta = 0.095$ and $B = \infty$ (i.e., no restriction imposed by parameter B).

Figure 3.15 compares the histogram of NVR characteristic for a realization under irregular model and a realization under regular model. The sign of the parameter θ, which influences the strength of interactions, is crucial: in case of positive θ, the neighbouring cells tend to have similar volumes and hence all cells tend to have similar volumes, whereas these volumes tend to be substantially different if $\theta < 0$.

Note that comparing Figures 1.4 and 1.8 with Figure 3.15, we observe that the NVR of our datasets is closer to an irregular model rather than to a regular one.



a) b)

Figure 3.15: Histogram of relative frequencies of the NVR statistic for the simulated tessellation based on (a) irregular model (4497 cells) and (b) regular model (752 cells).

Table 3.8: Estimated parameters based on tessellations obtained from regular and irregular model - the table shows estimates of hardcore parameters α, β, soft-core parameter θ and activity z; the means and standard deviations of estimates are obtained from 20 realizations.

		irregular	true value	regular	true value
$\hat{\alpha}$	mean	$2 \cdot 10^{-2}$	$2 \cdot 10^{-2}$	$3.4 \cdot 10^{-2}$	$2 \cdot 10^{-2}$
	sd	$3.58 \cdot 10^{-6}$		$6.22 \cdot 10^{-4}$	
$\hat{\beta}$	mean	$9.5 \cdot 10^{-2}$	$9.5 \cdot 10^{-2}$	$9.5 \cdot 10^{-2}$	$9.5 \cdot 10^{-2}$
	sd	$8.73 \cdot 10^{-7}$		$1.59 \cdot 10^{-6}$	
$\hat{\theta}$	mean	$-9.13 \cdot 10^{-1}$	-1	1.087	1
	sd	$1.64 \cdot 10^{-1}$		$2.18 \cdot 10^{-1}$	
$\hat{z}$	mean	1787.21	2000	1911.66	2000
	sd	194.95		249.23	

Parameter estimation in NVR model

The two-step estimation method described in Sections 3.3.1 and 3.3.2 is used for parameter estimation in case of both regular and irregular model. Table 3.8 shows estimates of the parameters α, β, θ and z computed from simulated tessellations. The results of pseudolikelihood method are not very accurate, especially for parameter z in the irregular case. This can be caused by an insufficient number of removable points, Definition 36. More details about practical aspects regarding the estimation of the parameters can be found in Dereudre and Lavancier [2011], Section 4.

4. Hierarchical model

A hierarchical approach splitting a model for $(\mathbf{x}_n, \mathbf{r}_n)$ into two steps, first the point pattern $\mathbf{x}_n$ is modelled and consequently the marks/radii $\mathbf{r}_n$ are modelled conditioned on $\mathbf{x}_n$, can make the modelling easier. Instead of one complex model of dimensionality 4, in case of Gibbs-Laguerre tessellations in Section 3, we obtain two simpler models of dimensionality 3, in the first step, and 1, in the second step. Before building up a particular hierarchical model we focus on dependencies between (unmarked) point process and marks/radii – by a permutation test we show that these cannot be thougt to be independent. The results presented in this chapter were published in Stoyan et al. [2021] – dependencies between point process and marks, and Seitl et al. [2022] – the hierarchical modelling approach.

4.1 Independent modelling

The simplest model of a random marked point process is obtained by independent marking. It can be realized in two steps. First, an unmarked point process is generated and afterwards, the points are labeled with independent random marks following some common distribution. Let $\Phi = \sum_{i=1}^{\mu(\mathbb{X})} \delta_{x_i}$ be an unmarked point process and let $\{m_i\}$ be i.i.d. sequence that is independent of Φ. The corresponding random marked point process $\Phi^m = \sum_{i=1}^{\mu(\mathbb{X})} \delta_{(x_i, m_i)}$ is said to be independently marked, cf. Definition 43. Another model where the point process and marks are independent can be obtained by geostatistical marking, cf. Definition 44. In this model, the marks are determined by some random field. In both cases we can model separately an unmarked point process and a mark distribution. Otherwise, we have to take into account interactions between the points and marks.

Possible spatial dependencies in a point process of spatial locations of generators and dependencies between point process and marks need to be explored in detail, as we did in Stoyan et al. [2021], in order to suggest a reasonable statistical model for the generators of Laguerre tessellations. We study dependencies within a point process in Sections 4.1.1 and 4.1.2 and dependencies between point process and marks in Section 4.1.3. Formally, to see if there is dependence between point process and marks/radii, we consider the empirical mark correlation function computed from $(\mathbf{x}_n, \mathbf{r}_n)$ (see Section 2.2.5 or Illian et al. [2008]) and perform a permutation test, cf. Section 2.4.3, based on the global area rank envelopes, permuting 1000 times the marks/radii $\mathbf{r}_n$ when $\mathbf{x}_n$ is fixed and calculating the empirical mark correlation function each time, see Section 4.1.4.

4.1.1 Summary point process statistics

A first insight into an unmarked point pattern gives us a histogram of nearest neighbour distance, cf. Figure 4.1 A broader overview of the geometry of an unmarked point process of spatial locations gives us further summary characteristics introduced in Section 2.2.4. For their empirical counterparts see Section 2.2.5. Figures 4.2 and 4.3 show empirical F, G, K and $L(r) - r$ functions for NiTi alloy and alluminium alloy data samples, respectively. The estimates of F and G functions use Kaplan-Meier correction, the estimates of K and L functions

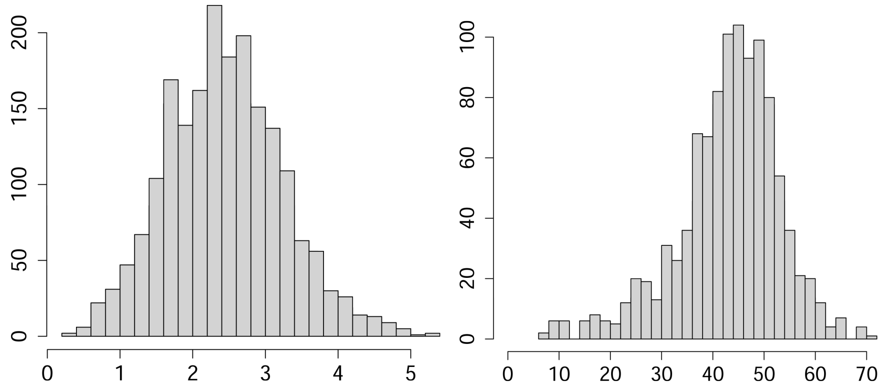


Figure 4.1: Nearest neighbour distance. Left: NiTi alloy, the minimal realized distance is $0.3824\mu m$. Right: alluminium alloy, the minimal realized distance is $7.737\mu m$.

use isotropic correction. The empirical functions in black are compared to the theoretical values under Poisson point process (model of complete spatial randomness) in red. There are visible differences in case of F and G functions. The difference at the lower end of K function curve can be clearly observed when the function is transformed to the $L(r) - r$ function.

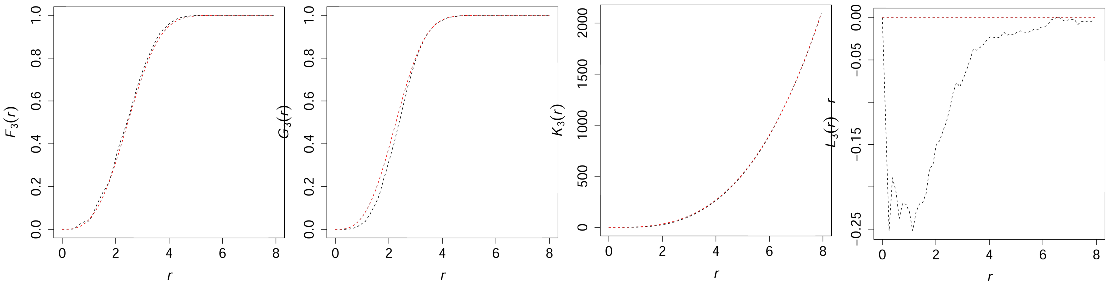


Figure 4.2: Empirical F, G, K and $L(r) - r$ functions for generators' locations of Laguerre tessellation representing NiTi alloy (in black) compared to the theoretical curves under Poisson point process (in red).

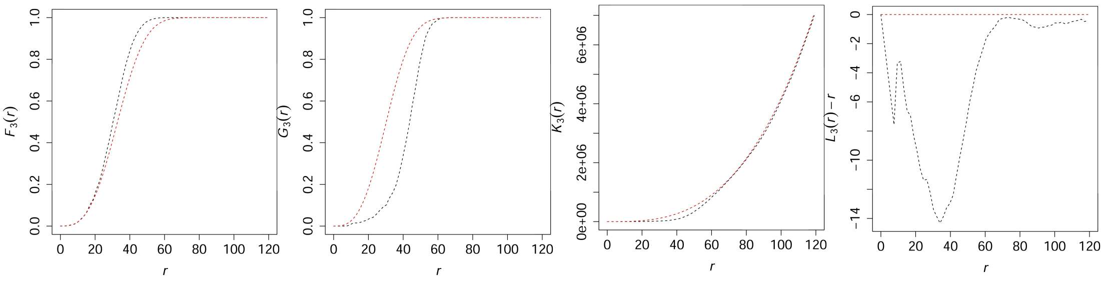


Figure 4.3: Empirical F, G, K and $L(r) - r$ functions for generators' locations of Laguerre tessellation representing alluminium alloy (in black) compared to the theoretical curves under Poisson point process (in red).

The last functional characteristic concerning the spatial arrangement of the generators' locations mentioned here is pair correlation function (shortly pcf).

Figure 4.4 shows estimated pcf with isotropic correction for the generators' locations of Laguerre tessellation representing NiTi alloy. Since its estimate is based on some propability kernel, the bandwidth h needs to be specified. In the literature there is a lot of advices and rules how to choose it, but none of these cannot be considered as a universal choice. Thanks to this and the fact that its misspecification may have a significant impact on the result, a detailed pair correlation function analysis is carried out in the next section in the case of aluminium alloy data sample.

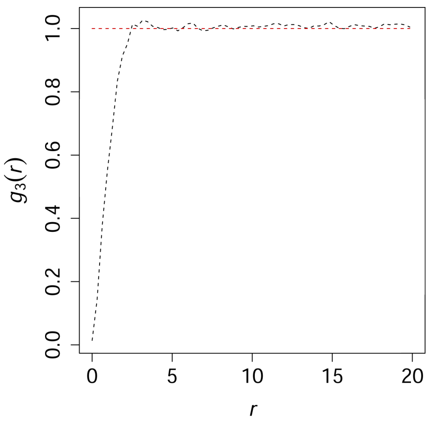


Figure 4.4: Empirical pair correlation function for generators' locations of Laguerre tessellation representing NiTi alloy, $h = 1.4$.

4.1.2 Pair correlation function analysis

In this section we deal with estimating the pair correlation function, Definition 49, of the point process of generators with kernel estimators. The samples are large enough for a reliable estimation of this function and due to homogeneity they permit usage of the isotropic estimator based on Ripley's edge correction, cf. (2.19). To do so, the choice of the bandwidth h has to be made with care. To demonstrate how to choose a reasonable bandwidth we proceed only with the alluminium alloy dataset.

The recommendation in [Illian et al., 2008, p. 236] is

$$h = 0.05/\lambda^{1/3}, \tag{4.1}$$

where λ denotes the intensity. With the estimated intensity 5.96×10^{-6} this led to $h = 2.76$. This value turned out to be much too small and we obtained a curve rather jagged. Therefore we changed to $h = 5$ and obtained the dashed curve in Figure 4.5. It has two clear peaks, one at $r = 10$ and one at $r = 55$. We considered the first peak as unusual and detected (only!) seven point pairs of an inter-point distance smaller than 15 (compare with the right pannel of Figure 4.1) that are responsible for the first peak.

In a second choice we followed R-package `spatstat`, cf. Baddeley et al. [2015], with a bandwidth $h = 14.34$ as recommended there, using

$$h = 0.26/\lambda^{1/3}. \tag{4.2}$$

We obtained a pair correlation function too smooth as is the solid curve in Figure 4.5 with only a soft shoulder at $r = 20$. Therefore, we decided to work with $h = 5$, which corresponds to a numerator of 0.09 in (4.1) and (4.2).

The seven pairs of generator points mentioned above we consider as unusual due to their small inter-point distances and very small radius-marks. Almost all 14 generators have small radius marks (only 3 of them have it bigger than 11 μm, the maximum is 31.14 μm), and small cell volume as well (5 smaller than 5000 μm^3, 6 between 5000 μm^3 and 15000 μm^3 and the volume of last 3 is 23153.3 μm^3, 41460.9 μm^3, 107717 μm^3); compare these numbers with Figures 1.7(b) and 1.8. Interestingly, none of these generators leads to an empty cell.

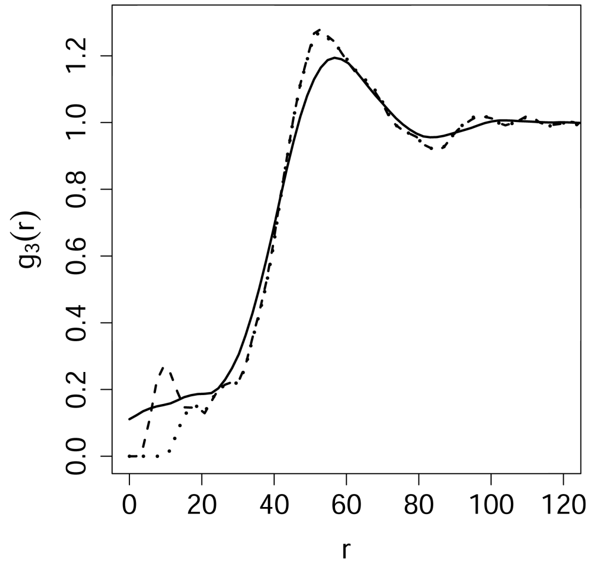
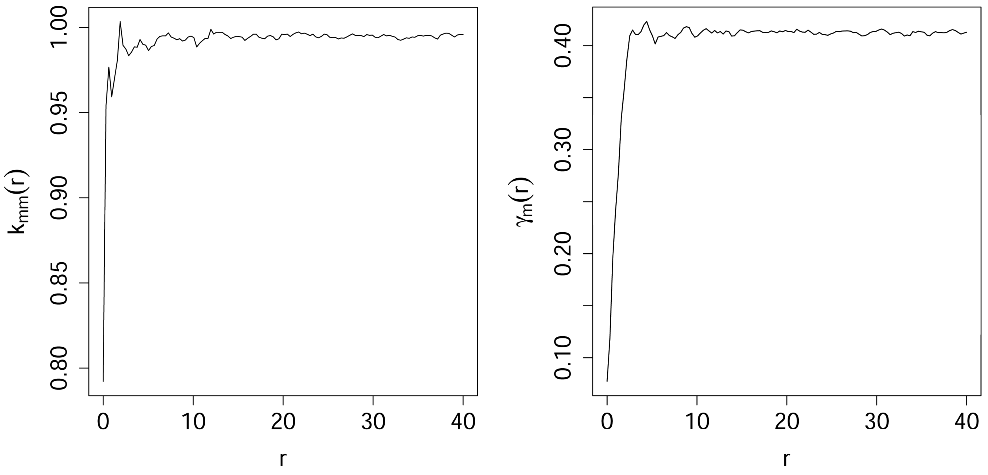


Figure 4.5: Empirical pair correlation functions for the Laguerre tessellation data representing alluminium alloy, all estimated with isotropic edge correction using spatstat, Baddeley et al. [2015]. Solid line: all data, $h = 14.34$; dashed line: all data, $h = 5$; dotted line: reduced data (without 7 close pairs), $h = 5$. For $r > 20$, the dashed and dotted lines almost coincide.

When we omitted the corresponding 14 points, we obtained the dotted curve in Figure 4.5, which has a form one may expect for a Gibbs point process with soft interaction. The Laguerre tessellation corresponding to the reduced sample is quite similar to the original tessellation. Table 4.1, column ‘subdata’, shows the characteristics corresponding to the subsample. We concluded that the omitted 14 points could be ignored as unessential. By the way, alternatively only 7 points (just one per pair) can be omitted. Both approaches lead to very similar results with respect to the form of empirical pair correlation function and tessellation statistics.

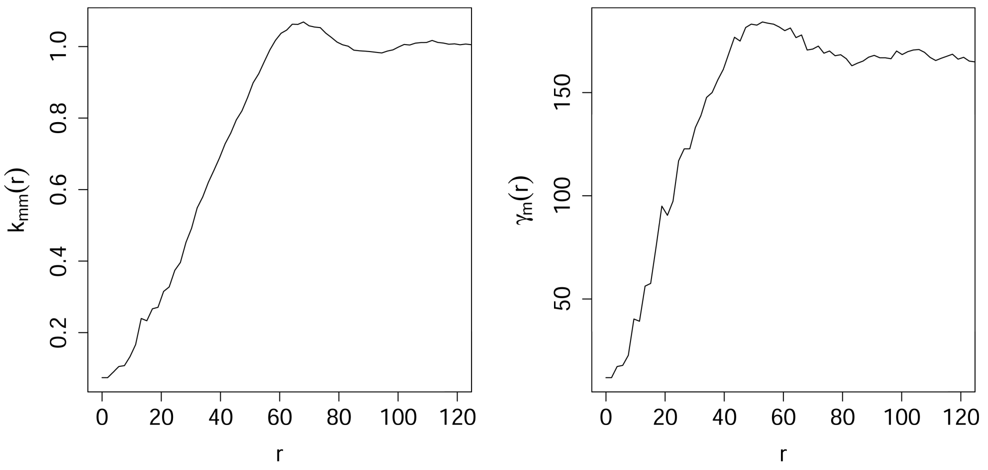
4.1.3 Mark correlation

Here we present the results of mark correlation statistics using mark correlation function, Definition 50, and mark variogram, Definition 51. The empirical counterparts, see Section 2.2.5, requires similarly to the pair correlation function a selection of the bandwidth h.



a) b)

Figure 4.6: Empirical mark correlation function, a), and mark variogram, b), for the Laguerre tessellation data representing NiTi alloy, $h = 1.4$.



a) b)

Figure 4.7: Empirical mark correlation function, a), and mark variogram, b), for the Laguerre tessellation data representing alluminium alloy, $h = 5$.

In Figures 4.6a) and 4.7a) there is the empirical mark correlation function for the two Laguerre tessellation data samples. It has a form often observed for such functions. The usual interpretation, which is applicable also here, is: point pairs which want to be close together have to pay the price that the product of their marks must be small, relative to the square of the mean mark. For the Laguerre tessellation this means that point pairs located close together tend to have also small radius marks, which means pairs of small tessellation cells.

In Figures 4.6b) and 4.7b) there is the empirical mark variogram for the two Laguerre tessellation data samples. Also its form is a form often observed; it looks like a variogram that is typical in geostatistics, first increasing and then constant. It says that points close together tend to have similar marks.

Both functions, $k_{mm}(r)$ and $\gamma_m(r)$, present a clear message: points close together tend to have small and similar marks. This excludes that the marks of the sample can be considered as coming from geostatistical marking. Equation (2.18) does not hold for the functions represented by the solid curves in Figures 4.6 and 4.7. In order to assure the reader, in next section we consider also independent marking, which we mimic by random permutation of the marks.

4.1.4 Independent marking

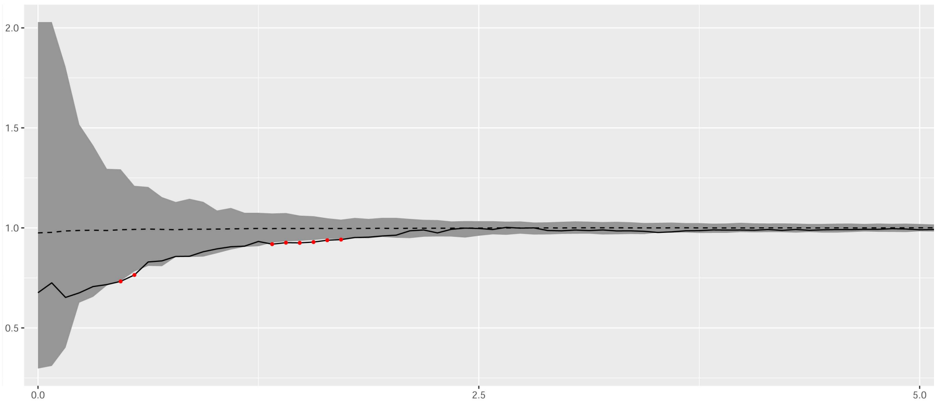
The case of independent marks, with the same mark distribution as the original data, is mimicked by random permutations of the marks. Formally, to see if there is dependence between point process and marks, we take the empirical mark correlation function and perform a permutation test based on the global area rank envelopes, cf. Section 2.4.2 and Myllymäki and Mrkvička [2019], permuting 1000 times the marks/radii $\mathbf{r}_n$ when $\mathbf{x}_n$ is fixed and calculating the empirical mark correlation function each time.

Figure 4.8 shows the functions based on the data (solid lines): both of them fall outside the 95%-global envelope (in case a) just on two small intervals, in case b) on the majority of the plotted x-axis range). The corresponding p-values obtained by the global area rank envelope test are 1.4% and $< 0.01\%$ for NiTi, Section 1.3.1, and aluminium, Section 1.3.2, alloy samples, respectively. Therefore we reject the null hypothesis of independence between point process and marks in both cases.

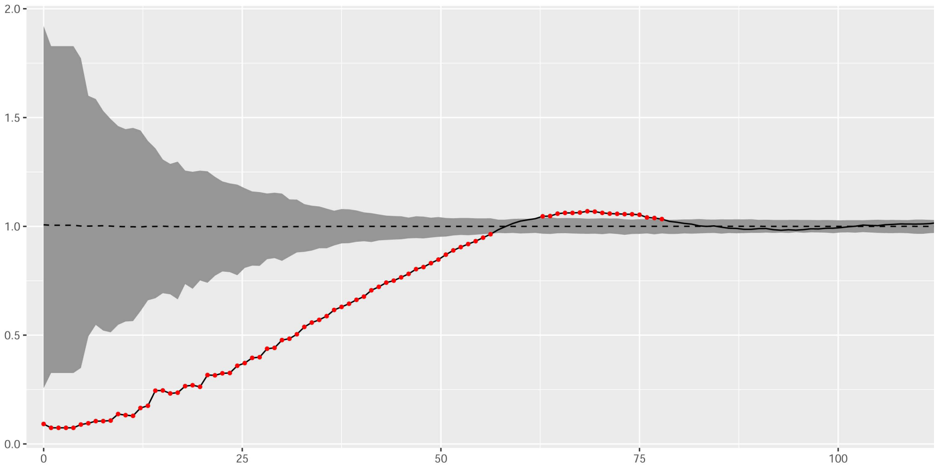
On example of aluminium alloy data sample we can see how geometric characteristics of the tessellation change when permuting the marks. The last column of Table 4.1 shows mean values and standard deviations of chosen characteristics resulting from 10 marked point patterns with permuted marks. We made these calculations since we suspected that there is an influence of the marking type on the tessellation properties. Indeed, the mean number of faces increased significantly in the permuted samples, a difference as that between 14.21 and 14.78 would matter in engineering applications. The standard deviation of the same characteristic decreased from 4.84 to 4.43 in average. Note that, with randomly permuted marks it happens that some generators create empty cells, but as you can see in the last column of Table 4.1 – average decrease by 9.1 cells, this effect is negligible.

To conclude, this section warns that modelling of three-dimensional tessellations by Laguerre tessellations may significantly fail if the assumptions made on distributional properties are too simple, e.g., in the case that correlations of the marks are ignored. If the generator points are given, the marks play an important role. Along with the mark distribution, the spatial correlation of the marks also counts. We have shown that important tessellation characteristics differ for correlated and independent marks in case of fixed mark distribution.

This fact makes modelling of Laguerre tessellations difficult, since it is not sufficient to model only the point process of generators and to take independent radius marks: points and marks should be modelled simultaneously. As the simultaneous modelling of points and marks in a marked point process, see



a) NiTi alloy



b) Alluminium alloy

Figure 4.8: The empirical mark correlation function (the solid line) and a 95%-global envelope (the grey region) obtained by permuting the radii when fixing the points. The dashed line is the average of the simulated mark correlation functions and the red dots indicate when the empirical function is outside the envelope.

Chapter 2, may be too difficult task, one alternative is to adapt the approach in Christoffersen et al. [2021], where locations of points in 3D are modelled hierarchically in two steps – first they model a point pattern in xy-plane by a suitable two-dimensional point process model, and second the values along the z-axis conditionally on the points in xy-plane by a Markov random field model. We propose the hierarchical approach to model a marked point pattern representing generators of a 3D Laguerre tessellation – first the point pattern is modelled and then conditionally on the pattern the marks.

Consequently, if the aim of modelling is only to obtain a basis for generating samples of tessellations similar to an empirically given one, we recommend that one does not use parametric models but instead generates samples by some kind of reconstruction (either nonparametric or with parameters chosen heuristically), see Sections 3.4 and 3.5.

Table 4.1: Laguerre tessellation characteristics for the aluminium alloy data, subdata without 14 generators with close neighbours and average values for ten samples with randomly permuted radius marks. In the upper part of the table there are numbers of cells and faces per sample, for permuted samples mean values are given (with standard deviation in brackets). In the lower part of the table there are mean values of characteristics per cell and standard deviation (below the means) for data and subdata. For the permuted samples overall means and square roots of the mean of variances within samples are given.

per sample:	data	subdata	permuted
number of cells	1049	1035	1039.90 (3.11)
number of faces	7453	7369	7683.80 (50.40)
per cell:			
nof	14.21	14.24	14.78
	4.84	4.79	4.43
nov	24.42	24.48	25.56
	9.68	9.59	8.86
vol	$1.68{\cdot}10^5$	$1.70{\cdot}10^5$	$1.69{\cdot}10^5$
	$1.53{\cdot}10^5$	$1.54{\cdot}10^5$	$1.40{\cdot}10^5$
surf	$1.56{\cdot}10^4$	$1.58{\cdot}10^4$	$1.63{\cdot}10^4$
	$9.21{\cdot}10^3$	$9.17{\cdot}10^3$	$8.52{\cdot}10^3$
spher	$8.28{\cdot}10^{-1}$	$8.28{\cdot}10^{-1}$	$8.10{\cdot}10^{-1}$
	$6.90{\cdot}10^{-2}$	$6.87{\cdot}10^{-2}$	$7.98{\cdot}10^{-2}$

4.2 Hierarchical approach

In this section we introduce a hierarchical model for $(\mathbf{x}_n, \mathbf{r}_n)$ consisting of first a parametric Gibbs point process model for $\mathbf{x}_n$ and second a parametric model for $\mathbf{r}_n$ conditioned on $\mathbf{x}_n$, Seitl et al. [2022]. Specifically, we use in the first step a nested sequence of flexible pairwise interaction points processes called multiscale processes, cf. Definition 33 and [Penttinen, 1984], and in the second step various exponential family models, Definition 55, where the canonical sufficient statistic is based on tessellation characteristics such as surface area or volume of cells or absolute difference in volumes of neighbouring cells, see Section 2.1.1. Apart from reducing the dimension from 4 (when viewing $(\mathbf{x}_n, \mathbf{r}_n)$ as a 4-dimensional point pattern) to 3 (when considering $\mathbf{x}_n$), an advantage is that we specify two much simpler models with parameters which do not depend on each other. Hence we can separate between how to simulate and estimate unknown parameters for $\mathbf{x}_n$ and $\mathbf{r}_n \mid \mathbf{x}_n$. The parameters are simply estimated by maximum pseudolikelihood methods, Section 2.4.1, and well-known MCMC algorithms, Section 2.3, are used for simulations.

A further advantage is that the model construction makes it possible to develop a rather straightforward model selection procedure: For $\mathbf{x}_n$, the procedure starts with the simplest case of a Poisson process and continues with constructing more and more complex multiscale processes until a satisfactory fit is obtained when considering global envelopes and tests [Myllymäki et al., 2017] based on various functional summary statistics. For $\mathbf{r}_n$ conditioned on $\mathbf{x}_n$, more and more complex exponential models are developed, where we demonstrate how to compare fitted models of the same dimension by considering maximized log pseudolikelihood functions. Further, we evaluate selected fitted models by comparing moment properties of tessellation characteristics under simulations from the model with empirical moments, by considering plots of global envelopes, and by evaluating values of global envelope tests. This comparison is not only done by looking at those tessellation characteristics used for specifying the canonical sufficient statistic of the exponential model but also for various other tessellation characteristics. Since both point process models for $\mathbf{x}_n$ and exponential models for $\mathbf{r}_n \mid \mathbf{x}_n$ require estimation of parameters, we are dealing with the composite hypotheses that both observed point pattern and radii conditioned on point pattern derives from a parametric model with unknown parameters. The plugging-in of the estimated parameter values can lead to conservative tests. Myllymäki et al. [2017] in Section 7 suggest to solve this conservativeness problem by method developed by Dao and Genton [2014]. Unfortunately, this requires a large number of additional simulations, hence not applicable in our case.

To avoid confusion, we reserve the notation $(\mathbf{x}_n, \mathbf{r}_n)$ for the data and use the notation $(\mathbf{y}_m, \mathbf{t}_m)$ when we consider arguments of densities where $\mathbf{y}_m = \{y_1, \dots, y_m\} \subset W$ is a finite point configuration and $\mathbf{t}_m = (t_1, \dots, t_m) \in \mathbb{M}^m$ are associated marks (if $m = 0$ then $\mathbf{y}_0$ is the empty point configuration and $\mathbf{t}_0$ can be ignored).

4.2.1 Point process models for the points

When modelling the point pattern data set $\mathbf{x}_n$ as a realization of a spatial point process Φ on W, we assume that the distribution of Φ is invariant under shifts and rotations when wrapping W on a 3D torus, i.e., when applying periodic boundary conditions, Definition 11. Further we rectrict ourselves only on models for Φ exhibiting regularity. Pairwise interaction point processes constitute a flexible class of models for regularity.

As a parametric approximation of such a pairwise interaction point process we consider a multiscale process, Definition 33 and [Penttinen, 1984]: for $q = 1, 2, \dots$, $\mathcal{M}_q$ denote the model class given by densities of the form

$$p(\mathbf{y}_m) \propto \beta^m \Pi_{i=1}^{q-1} \gamma_i^{\sum_{j<k} \mathbb{I}[\delta_{i-1} < \|y_j - y_k\|_W \leq \delta_i]} \tag{4.3}$$

where $\beta > 0$, $0 \leq \gamma_1 \leq 1$, …, $0 \leq \gamma_{q-1} \leq 1$ and $0 < \delta_1 < \ldots < \delta_{q-1}$ are unknown parameters. Note that $\mathcal{M}_1$ is just a homogeneous Poisson process on W with intensity β and $\mathcal{M}_2$ is just a Strauss process.

Simulation under the Poisson model $\mathcal{M}_1$ is well-known (see e.g. Møller and Waagepetersen [2003]) and for simulation of the models $\mathcal{M}_q$, $q \geq 2$, we used the birth-death-move Metropolis-Hastings algorithm, Algorithm 2 and [Geyer and Møller, 1994, Møller and Waagepetersen, 2003]. The parameter estimation is carried out by maximum pseudolikelihood method, Section 2.4.1.

4.2.2 Exponential tessellation models for the radii given the points

When modelling the observed radii $\mathbf{r}_n$ as a realisation of an n-dimensional vector, we condition on $\Phi = \mathbf{x}_n$ and consider a conditional probability density function (pdf) $p(r_1, \ldots, r_n \mid x_1, \ldots, x_n)$ on $(0, \infty)^n$. More precisely we assume $\mathbf{R}$ is a random vector (of random length) which conditioned on $\Phi = \mathbf{y}_m$ has a conditional pdf $p(\mathbf{t}_m \mid \mathbf{y}_m)$ which is zero whenever $C(y_j, t_j \mid \mathbf{y}^*_m, \mathbf{t}^*_m) = \emptyset$ for some $j \in \{1, \ldots, m\}$ (here $(\mathbf{y}^*_m, \mathbf{t}^*_m)$ is defined in a similar way as $(\mathbf{x}^*, \mathbf{r}^*)$, i.e., in the right hand side of (2.5) $(\mathbf{x}, \mathbf{r})$ is replaced by $(\mathbf{y}_m, \mathbf{t}_m)$). Furthermore, in order to work with a well-defined conditional pdf in (4.4) below we assume a mark space $\mathbb{M} = [0, R_{\max}]$ so that $p(\mathbf{t}_m \mid \mathbf{y}_m) = 0$ if $\mathbf{t}_m \notin \mathbb{M}^m$.

We now give a general exponential family form of the conditional pdf where $s \in \{1, 2, \ldots\}$ is the dimension, $\theta = (\theta_1, \ldots, \theta_s)$ denotes the canonical parameter and $H = (H_1, \ldots, H_s)$ the canonical sufficient statistic: for $\mathbf{t}_m \in \mathbb{M}^m$,

$$p(\mathbf{t}_m \mid \mathbf{y}_m) \propto \mathbb{I}_{[C(y_j,t_j|\mathbf{y}^*_m,\mathbf{t}^*_m)\neq\emptyset,\, j=1,\ldots,m]} \exp\left(\sum_{i=1}^{s} \theta_i H_i(\mathbf{y}_m, \mathbf{t}_m)\right). \qquad (4.4)$$

The idea is to let each $H_i(\mathbf{y}_m, \mathbf{t}_m)$ depend on either the radii, or tessellation characteristics of the cells $C(y_j, t_j \mid \mathbf{y}^*_m, \mathbf{t}^*_m)$ with $j = 1, \ldots, m$ or interactions between these cells. Specifically, in Section 4.4.2 we consider the following cases (a)–(e), using the abbreviations from Section 2.1.1, for short writing H_i for $H_i(\mathbf{y}_m, \mathbf{t}_m)$ and considering in (b)–(d) a sum over the cells $C(y_j, t_j \mid \mathbf{y}^*_m, \mathbf{t}^*_m)$, $j = 1, \ldots, m$, and in (e) a sum over all unordered pairs of cells sharing a face:

(a) including both $H_i = \sum_{j=1}^{m} \log \frac{t_j}{6}$ and $H_{i'} = \sum_{j=1}^{m} \log\left(1 - \frac{t_j}{6}\right)$ (with $i \neq i'$) somehow corresponds to a scaled beta distribution for the radii if no other terms are included in (4.4) – ‘somehow’ because it is not exactly a beta distribution since $p(\mathbf{t}_m \mid \mathbf{y}_m) = 0$ if $C(y_j, t_j \mid \mathbf{y}^*_m, \mathbf{t}^*_m) = \emptyset$ for some $j \in \{1, \ldots, m\}$;

(b) $H_i = \sum \text{nof}$ is twice the total number of faces;

(c) $H_i = \sum \text{surf}$ is twice the total surface area of faces;

(d) $H_i = \sum \text{vol}^2$ is the sum of squared volumes of cells;

(e) $H_i = \sum \text{dvol}$ is the sum of difference in volumes of two cells which share a face.

For simulation under (4.4) we use a Metropolis within Gibbs algorithm, Section 2.3.2, where we alternate between updating from the conditional densities

$$p(t_j \mid t_k,\, k \neq j,\, y_1, \ldots, y_m) \propto p(t_1, \ldots, t_m \mid y_1, \ldots, y_m), \quad j = 1, \ldots, m, \qquad (4.5)$$

using a Metropolis algorithm with a normal proposal.

4.2.3 Joint model

The next proposition summarizes the whole hierarchical model and justifies its existence.

Claim 22. *The model for* $(\mathbf{x}_n, \mathbf{r}_n)$ *given by density*

$$p(\mathbf{y}_m, \mathbf{t}_m) \propto$$
$$\beta^m \Pi_{i=1}^{q-1} \gamma_i^{\sum_{j<k} \mathbb{I}_{[\delta_{i-1} < \|y_j - y_k\|_W \leq \delta_i]}} \mathbb{I}_{[C(y_j, t_j | \mathbf{y}_m^*, \mathbf{t}_m^*) \neq \emptyset,\, j=1,\ldots,m]} \exp\left(\sum_{i=1}^{s} \theta_i H_i(\mathbf{y}_m, \mathbf{t}_m) \right)$$

with $q, s \in \mathbb{N}$ *is well-defined for* $\beta > 0$, $0 \leq \gamma_1 \leq 1$, $\ldots$, $0 \leq \gamma_{q-1} \leq 1$, $0 < \delta_1 < \ldots < \delta_{q-1}$ *and either* $\theta_i > -1$ *and* $\theta_{i'} > -1$ *in case of (a) or all* $\theta \in \mathbb{R}^s$ *in cases of (b)-(e). Moreover, for fixed* $(\delta_1, \ldots, \delta_{q-1})$ *the joint model is an exponential family model with canonical parameter* $(\log\beta, \log\gamma_1, \ldots, \log\gamma_{q-1}, \theta_1, \ldots, \theta_s)$.

Proof. This follows from the definition of multiscale point process (Definition 33) and from the fact that the mark space $\mathbb{M}$ is bounded. ☐

4.3 Parameter estimation

In this chapter we have introduced several parametric models for generators of Laguerre tessellations. Now, we will look at how to estimate the parameters of these models.

4.3.1 Profile pseudolikelihood for the points

Consider a point process with a given realization $\mathbf{y}_m$ in the observation window W and modelled by a parametric density p_θ (with respect to the unit rate Poisson process) where θ is an unknown real parameter vector. In particular, we consider p_θ to be a multiscale process $\mathcal{M}_q$, Definition 33, with $q > 1$. Then using Definition 54 the log pseudolikelihood function becomes

$$\begin{aligned} \log \mathcal{PL}(\beta, \gamma; \mathbf{y}_m) = & |W| - \int_W \beta \prod_{i=1}^{q-1} \gamma_i^{t_{\delta_i}(u, \mathbf{y}_m) - t_{\delta_{i-1}}(u, \mathbf{y}_m)} \, \mathrm{d}u \\ & + m \log\beta + 2 \sum_{i=1}^{q-1} \Big(S_{\delta_i}(\mathbf{y}_m) - S_{\delta_{i-1}}(\mathbf{y}_m) \Big) \log\gamma_i, \end{aligned}$$

where $t_\delta(u, \mathbf{y}_m) = \sum_{y_j \in \mathbf{y}_m} \mathbb{I}_{\{0 < \|u, y_j\| \leq \delta\}}$ and $S_\delta(\mathbf{y}_m) = \sum_{u \in \mathbf{y}_m} t_\delta(u, \mathbf{y}_m)/2$, $\gamma = (\gamma_1, \ldots, \gamma_{q-1})$ and $\delta = (\delta_1, \ldots, \delta_{q-1})$. Please note, that the pseudolikelihood function depends only on parameters β and γ. The parameter δ (interaction radii) is supressed here. Indeed, the maximum pseudolikehood method can be used only for estimation of parameters β and γ when the parameter δ is fixed. Since $\log \mathcal{PL}(\beta, \gamma; \mathbf{y}_m)$ is concave function of $\log\beta$ and $\log\gamma = (\log\gamma_1, \ldots, \log\gamma_{q-1})$, cf. Claim 23, we can find MPLE of $\log\beta$ and $\log\gamma$ using the Newton-Raphson algorithm. Since this MPLE depends on the parameter δ, we obtain a profile log pseudolikelihood which is maximized with respect to $(\delta_1, \ldots, \delta_{q-1})$ defined over a $(q-1)$-dimensional grid, thereby providing the final MPLE. The integral appearing in $\mathcal{PL}(\beta, \gamma; \mathbf{y}_m)$ is estimated numerically using a grid of values as descibed in Section 2.4.1.

4.3.2 Pseudolikelihood for the radii given the points

Pseudolikelihood function for the radii $\mathbf{t}_m$ given the points $\mathbf{y}_m$ is defined by the product of the conditional densities of each t_j given the t_k with $k \neq j$, see Definition 53,

$$\mathcal{PL}(\theta; \mathbf{t}_m) = \prod_{j=1}^{n} p_\theta(t_j \mid y_j, (y_k, t_k) \text{ with } k \neq j) = \prod_{j=1}^{n} \frac{p_\theta(\mathbf{t}_m \mid \mathbf{y}_m)}{\int_{\mathbb{M}} p_\theta(\mathbf{t}_m \mid \mathbf{y}_m)\, \mathrm{d}t_j},$$

where $p_\theta(\mathbf{t}_m \mid \mathbf{y}_m)$ is of the form (4.4) and θ is an uknown real parameter vector. Assume that $(\mathbf{y}_m, \mathbf{t}_m)$ is feasible, i.e., all Laguerre cells $C(y_j, t_j \mid \mathbf{y}_m^*, \mathbf{t}_m^*)$ are nonempty; we denote this property by $(\mathbf{y}_m, \mathbf{t}_m) \notin L(\emptyset)$. Then the log pseudolikelihood function for the radii distribution given the points is

$$\begin{aligned}\log \mathcal{PL}(\theta; \mathbf{t}_m) =& n \sum_{i=1}^{s} \theta_i H_i(\mathbf{y}_m, \mathbf{t}_m) \\ & - \sum_{j=1}^{m} \log \int_{\mathbb{M}} \mathbb{I}_{\{(\mathbf{y}_m, \mathbf{t}_m^{j,u}) \notin L(\emptyset)\}} \exp\left(\sum_{i=1}^{s} \theta_i H_i(\mathbf{y}_m, \mathbf{t}_m^{j,u}) \right) \mathrm{d}u,\end{aligned}$$

where $\mathbf{t}_m^{j,u} = (t_1, \ldots, t_{j-1}, u, t_{j+1}, \ldots, t_m)$.

Since $\log \mathcal{PL}(\theta; \mathbf{t}_m)$ is concave function of θ, cf. Claim 23, Newton-Raphson method can be used for finding the MPLE. Finally, we calculate various integrals which appear in the pseudolikelihood by numerical methods.

4.3.3 Properties of MPLE

In the following proposition we state some non-trivial properties of the maximum pseudolikelihood estimates from the previous two sections.

Claim 23.

1. *$\mathcal{PL}(\beta, \gamma; \mathbf{y}_m)$ is log-concave function of* $(\log \beta, \log \gamma_1, \ldots, \log \gamma_{q-1})$
2. *$\mathcal{PL}(\theta; \mathbf{t}_m)$ is log-concave function of θ*
3. *MPLEs of (β, γ) are consistent estimates of (β, γ)*
4. *MPLEs of (β, γ) are asymptotically normal (with respect to enlarging window)*

Proof. 1.,2. Both (4.3) and (4.4) are exponential family models with canonical parameters $(\log \beta, \log \gamma_1, \ldots, \log \gamma_{q-1})$ and θ, respectively. According to Theorem 15 $\log \mathcal{PL}(\beta, \gamma; \mathbf{y}_m)$ and $\log \mathcal{PL}(\theta; \mathbf{t}_m)$ are concave.

3. In case of point process, maximum pseudolikelihood estimates $\hat{\beta}$ and $\hat{\gamma}$ are (weakly) consistent since the potential of multiscale point process has a finite range – δ_{q-1}, cf. Theorem 3.2 and Example 2 in Jensen and Møller [1991]. Strong consistency can be proved as well, cf. Propositions 2 and 3 in Mase [1995].

4. Thanks to the finite range of the potential of multiscale point process, maximum pseudolikelihood estimates $\hat{\beta}$ and $\hat{\gamma}$ are asymptotically normal with respect to enlarging observation window, cf. Theorem 1.1 in Jensen and Künsch [1994].

□

4.4 Hierarchical modelling

The hierarchical approach introduced in Section 4.2 is demonstrated here on Laguerre tessellation data representing the NiTi alloy microstructure from Section 1.3.1. The section details the model selection procedure briefly described in Section 4.2 and which leads to our final joint model for $\mathbf{x}_n$ and $\mathbf{r}_n$. The classical goodness of fit procedures are based on measures given by maximizing likelihood functions and accounting for model complexity by including a penalty in terms of the number of parameters (e.g., AIC and BIC criteria). In Šedivý et al. [2018] the Vapnik-Chervonenkis theory is used instead and a criterion based on residual sum of squares is developed for partially ordered sets of deterministic tessellation models. Since we deal with stochastic tessellation models but our likelihood functions are intractable to maximize, we present another approach that corresponds to modern trends in spatial statistics. It is based on how well geometrical tessellation characteristics are described when comparing fitted models, where we account for mutual correlations among the characteristics and we consider the maximum of pseudolikelihood functions when comparing models with the same number of parameters.

4.4.1 The fitted model for the points

The first model to be taken into account is a homogeneous Poisson point process (with the intensity estimated by n divided by the volume of W), i.e., the model of complete spatial randomness. The way how functional summary statistics L, F and G, see Definitions 48, 46 and 47, based on the data differ from those computed from simulations under Poisson point process, see the envelope in Figure 4.9, indicates regularity in the point pattern $\mathbf{x}_n$. This observation justifies the choice of class of multiscale point processes, Defintion 33, for $\mathbf{x}_n$ as these point processes constitute a flexible class of models for regularity.

Then, for the point pattern $\mathbf{x}_n$ and $q = 1, 2, \ldots$, we select the first model $\mathcal{M}_q$ which provides a satisfactory fit when considering global envelopes and global area rank envelope test provided by `R`-package `GET`, see Section 2.4.2. The success of fit is assessed using empirical L, F and G functional summary statistics which are concatenated for the purposes of the test.

Figures 4.9, 4.10 and 4.11 show estimated/empirical functions $\hat{L}(t) - t$, $\hat{F}(t)$ and $\hat{G}(t)$ of the point pattern $\mathbf{x}_n$ together with concatenated 95%-global envelopes/confidence regions (the grey areas) obtained by simulations under the multiscale point process models $\mathcal{M}_1$, $\mathcal{M}_2$ and $\mathcal{M}_3$, respectively. The models $\mathcal{M}_2$ and $\mathcal{M}_3$ are fitted by profile maximum pseudolikelihood (for details concerning the estimation procedure see Section 4.3.1 and Attachment A.4). For definitions and interpretations of the empirical functions using edge correction factors see Section 2.2.4 or Baddeley et al. [2015] (for $\hat{L}$ we used Ripley's isotropic edge correction factor and for $\hat{F}$ and $\hat{G}$ we used Kaplan-Meier edge correction factors). The fact that the 95%-global envelopes are concatenated means that all three empirical functions are expected to be within the envelope with probability 0.95 in order to not reject the null hypothesis of the validity of a particular model $\mathcal{M}_d$, $d = 1, 2, 3$. The global envelopes were obtained using the `R`-package `GET` [Myllymäki and Mrkvička, 2019] with 1999 simulations for each process (increas-

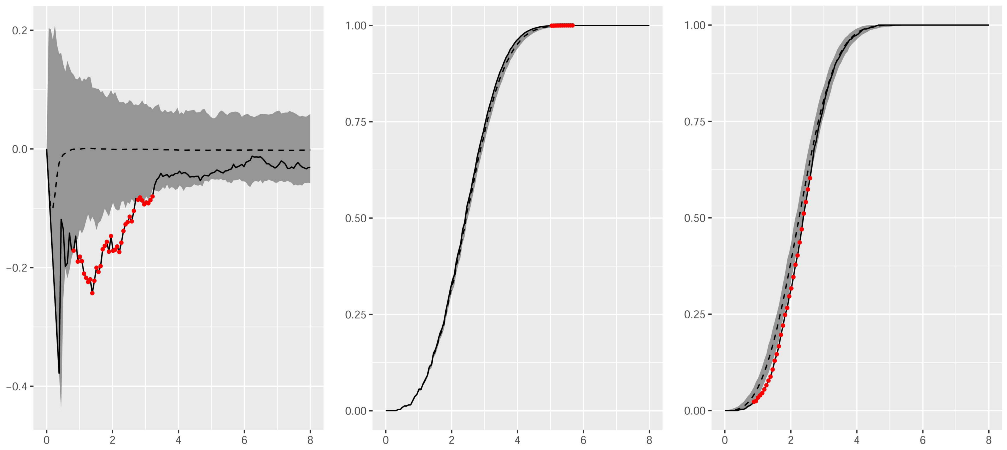


Figure 4.9: From left to right, empirical functional summary statistics $\hat{L}(t) - t$, $\hat{F}(t)$ and $\hat{G}(t)$ (solid lines) and simulated 95%-global envelope (grey regions) obtained under a fitted homogeneous Poisson process. Dashed lines are averages of the simulated functional summary statistics and the dots indicate when the empirical functions are outside the envelope.

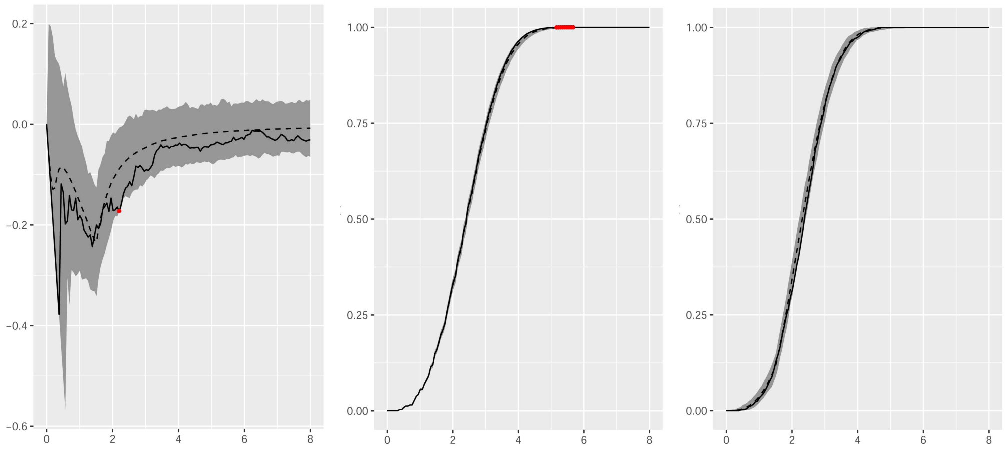


Figure 4.10: From left to right, empirical functional summary statistics $\hat{L}(t) - t$, $\hat{F}(t)$ and $\hat{G}(t)$ (solid lines) and simulated 95%-global envelope (grey region) obtained under a fitted Strauss process. Dashed lines are averages of the simulated functional summary statistics and the dots indicate when the empirical functions are outside the envelope.

ing this to 9999 simulations did not change the results). Note that because of plugging-in of the estimated parameter values, the tests can be conservative.

Neither Poisson model $\mathcal{M}_1$ nor Strauss model $\mathcal{M}_2$ is providing a satisfactory fit. Figures 4.9 and 4.10 show that the empirical functions are outside the envelope, and the corresponding p-values obtained by the global area rank envelope test are below 0.1% in case of $\mathcal{M}_1$ and 4.4% in case of $\mathcal{M}_2$.

The first model which is not significant at level 5% is $\mathcal{M}_3$: Figure 4.11 shows that the empirical functional summary statistics are within the 95%-global envelope. The corresponding p-value obtained by the global area rank envelope test is 18.8%, and the maximum pseudolikelihood estimates, see Section 4.3.1, are $\hat{\beta} = 0.0168$, $\hat{\gamma}_1 = 0.5328$, $\hat{\gamma}_2 = 0.8432$, $\hat{\delta}_1 = 1.25$ and $\hat{\delta}_2 = 2.25$.

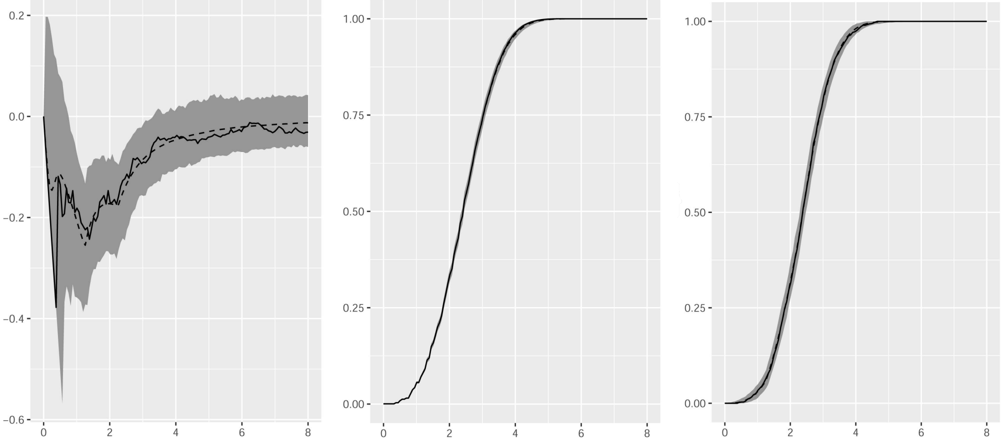


Figure 4.11: From left to right, empirical functional summary statistics $\hat{L}(t) - t$, $\hat{F}(t)$ and $\hat{G}(t)$ (solid lines) and simulated 95%-global envelope (grey regions) obtained under a fitted multiscale process with $d = 3$. Dashed lines are averages of the simulated functional summary statistics.

4.4.2 The fitted model for the radii conditioned on the points

Now, having fitted the model $\mathcal{M}_3$ for $\mathbf{x}_n$ it remains to obtain a model for $\mathbf{r}_n$ conditioned on $\mathbf{x}_n$ where we use a model selection procedure as follows. We consider a list of tessellation characteristics

$$\mathcal{L} = \{\text{nof}, \text{vol}, \text{surf}, \text{tel}, \text{spher}, \text{dvol}\} \tag{4.6}$$

when creating and evaluating more and more complex models as given by (4.4) and (a)–(e) in Section 4.2.2. When comparing fitted models of the same dimension q, we select the one with the highest value of the maximized log pseudolikelihood function. The selected model is then evaluated by a global area rank envelope test, see Section 2.4.2, based on kernel smoothed densities of the empirical distributions for the six tessellation characteristics in $\mathcal{L}$, using 499 simulations of the joint model for $\mathbf{x}_n$ and $\mathbf{r}_n$. Here we concatenate the six densities and use the `R`-package `GET` [Myllymäki and Mrkvička, 2019], and we simulate from the joint model rather than from the conditional model of $\mathbf{r}_n$ given $\mathbf{x}_n$, since our aim is to replace expensive laboratory experiments with simulations from the joint model. Therefore, in Table 4.2, we also evaluate our fitted models by comparing empirical (column 'data') and simulated means and standard deviations of the tessellation characteristics in $\mathcal{L}$, using 100 simulations of first $\mathbf{y}_m$ and second under the different fitted models of $\mathbf{t}_m$ conditioned on $\mathbf{y}_m$. The table also shows the (signed) difference between the empirical and the simulated values divided by the empirical values (the values given in percentages) – we refer to such a value as a deviation. Again note that the tests can be conservative.

First, we consider the empirical distribution of the radii: As the histogram in Figure 1.3(b) looks like a beta distribution, we first propose in (4.4) only to include the terms in (a) so that $s = 2$, $H_1 = \sum_{j=1}^{m} \log \frac{t_j}{6}$ and $H_2 = \sum_{j=1}^{m} \log\left(1 - \frac{t_j}{6}\right)$ – we refer to this model as 'beta'. The column 'beta' in Table 4.2 shows that except for dvol the means match the data well, and the deviations of standard deviations vary by 12 to 24 percent. However, the fitted 'beta' model was highly significant when evaluated by the global area rank envelope test.

Table 4.2: Means and standard deviations of the tessellation characteristics given in (4.6) and as obtained by simulations under various joint models and from the data. Deviations are given in percentages. See the text for details.

		models			data
		beta	beta + dvol	beta + nof + dvol	
nof	mean	14.82	14.87	14.95	14.98
		-1%	-1%	<-1%	
	sd	5.58	5.38	5.20	4.92
		+13%	+9%	+6%	
vol	mean	70.65	70.65	70.65	69.21
		+2%	+2%	+2%	
	sd	65.65	63.42	62.01	58.89
		+12%	+8%	+5%	
surf	mean	91.14	91.53	93.58	92.46
		-1%	-1%	+1%	
	sd	57.26	54.07	51.50	47.86
		+20%	+13%	+8%	
tel	mean	66.16	66.12	67.69	67.92
		-3%	-3%	<-1%	
	sd	31.99	30.86	29.52	27.38
		+17%	+13%	+8%	
spher	mean	0.76	0.76	0.77	0.78
		-3%	-3%	-1%	
	sd	0.11	0.10	0.10	0.087
		+24%	+19%	+16%	
dvol	mean	79.87	73.72	69.82	68.89
		+16%	+7%	+1%	
	sd	72.75	68.57	62.78	65.04
		+12%	+5%	-4%	

Table 4.3: Maximized log pseudolikelihood functions for radii models conditioned on the points. The value for the 'beta' model is given as a reference.

model	$\log \mathcal{PL}(\hat{\theta})$
beta	-2532.45
beta + vol^2	-2866.39
beta + nof	-2594.72
beta + surf	-2483.98
beta + dvol	-2468.76
beta + vol^2 + nof	-2513.89
beta + vol^2 + surf	-2823.71
beta + vol^2 + dvol	-2714.10
beta + nof + surf	-2498.58
beta + nof + dvol	-2465.12
beta + surf + dvol	-2477.63

Second, we expand the model by including one of the terms $H_3 = \sum \mathrm{vol}^2$, $\sum \mathrm{nof}$, $\sum \mathrm{surf}$ or $\sum \mathrm{dvol}$ so that $s = 3$. Here, we do not include $\sum \mathrm{vol}$, since this is a constant; or $\sum \mathrm{tel}$, since the correlation coefficient between surf and tel for each cell is close to 1 (the highest in Table 1.2); or $\sum \mathrm{spher}$, since by definition spher is given by vol and surf for each cell. Table 4.3 shows that the model with $H_3 = \sum \text{dvol}$ has the largest maximized log pseudolikelihood function. For this model, comparing the columns ‘beta’ and ‘beta+dvol’ in Table 4.2 we obtain now better results for dvol. Moreover, all standard deviations are now reduced, and the p-value based on global area rank envelope test is 9.8%.

Third, we investigate the effect of expanding the model ‘beta’ with any two of the four terms $\sum \text{vol}^2$, $\sum \text{nof}$, $\sum \text{surf}$ and $\sum \text{dvol}$ so that $s = 4$. Table 4.3 shows that ‘beta+nof+dvol’ provides the best fit according to the maximized log pseudolikelihood function. For this model, the p-value based on global area rank envelope test is 10.6% which is slightly larger than the p-value of 9.8% for the model ‘beta+dvol’. Table 4.2 shows an improvement for both the mean values and the standard deviations when comparing ‘beta+nof+dvol’ with ‘beta+dvol’. Figure 4.12 shows empirical kernel estimates of the densities for the six characteristics in $\mathcal{L}$ together with a 95%-global envelope obtained when concatenating all six empirical densities. The empirical functions are completely covered by the envelope. Moreover, the maximum pseudolikehood estimates are $\hat{\theta}_1 = 4.709$, $\hat{\theta}_2 = 5.982$, $\hat{\theta}_3 = -2.376 \times 10^{-1}$ and $\hat{\theta}_4 = 3.021 \times 10^{-2}$. Thus under this fitted model realizations become more likely as the total number of faces decreases or the sum of differences in volumes between neighbouring cells increases (when all other terms H_i in (4.4) are fixed).

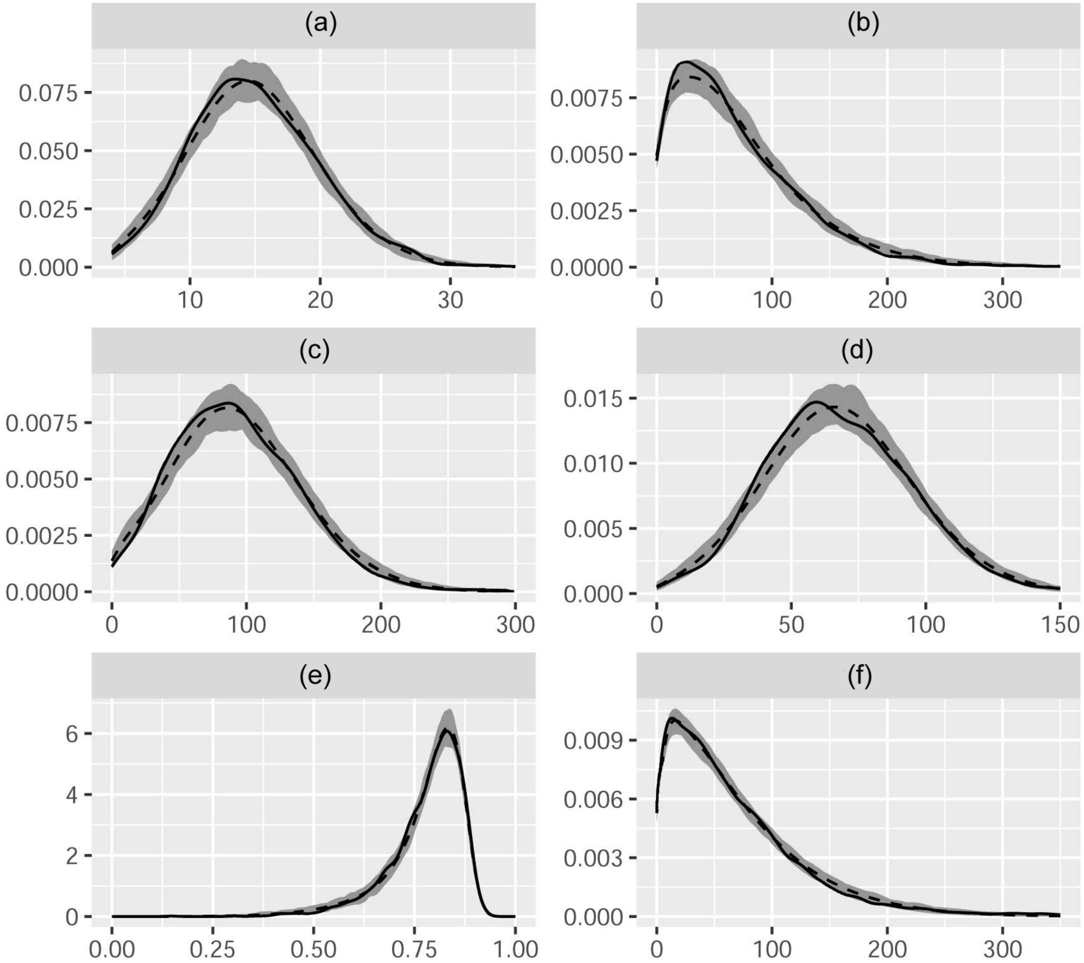


Figure 4.12: Estimated densities of tessellation characteristics, namely nof (a), vol (b), surf (c), tel (d), spher (e) and dvol (f). The solid lines are the functions based on the data $(\mathbf{x}_n, \mathbf{r}_n)$ and the grey regions are 95%-global envelopes under the fitted model 'beta+nof+dvol'. Dashed lines are averages of the simulated densities.

Conclusion

The thesis is written on the base of results obtained by the author when being a member of the team solving the international research project of the Czech Science Foundation ‘Parametric representation and stochastic 3D modelling of the grain microstructure of polycrystalline materials using the random marked tessellations’. The aim of the thesis is to develop stochastic geometry models of given complex spatial systems and corresponding methods of spatial statistics, including their implementation an discussion of numerical results.

Laguerre tessellation presents a suitable geometrical object for the approximation of polycrystalline microstructures investigated in the thesis and it brings a considerable compression of voxelized three-dimensional data since it is determined by a marked point pattern with one-dimensional marks. Stochastic modelling and statistical inference of Laguerre tessellations is developed in two different ways.

First in Chapter 3 the Gibbs-Laguerre tessellation is introduced and thoroughly investigated. There are interesting theoretical problems such as the existence of an infinite-volume Gibbs measure (solved in Section 3.2). When dealing with real data, statistical inference and reconstruction of Gibbs models (Sections 3.4- 3.6) can hardly be realized without intensive computing, supported by Markov chain Monte Carlo simulations. Thank to this new class of random tessellations we are able to simulate realizations with prescribed geometrical properties and their variability, at the cost of computationally demanding algorithms.

On the contrary to an apriori probabilistic model in Chapter 3 we proceed by a direct data-based modeling of a Laguerre tessellation as a marked point process in Chapter 4. We carefully examine relationship of points and radii in the datasets (Section 4.1) and proceed in a hierarchical way, as in the basic formula for conditional probability density. Tayloring statistical methods (maximum pseudo-likelihood, global envelopes) to the new environment yields a respectful approach to model selection and leads to a good fit (Section 4.4) between a model and data.

The results achieved in the thesis have a general validity and will be used in the continuing research collaboration between mathematicians and physicists.

Bibliography

Andreas Alpers, Andreas Brieden, Peter Gritzmann, Allan Lyckegaard, and Henning Friis Poulsen. Generalized balanced power diagrams for 3D representations of polycrystals. Philosophical Magazine, 95(9):1016–1028, 2015.

Franz Aurenhammer. Power diagrams: properties, algorithms and applications. SIAM Journal on Computing, 16(1):78–96, 1987.

Adrian J. Baddeley. Spatial point processes and their applications. In Wolfgang Weil, editor, Stochastic Geometry: Lectures given at the C.I.M.E. Summer School held in Martina Franca, Italy, September 13–18, 2004, pages 1–75. Springer, 2007.

Adrian J. Baddeley and Rolf Turner. Bias correction for parameter estimates of spatial point process models. Journal of Statistical Computation and Simulation, 84(8):1621–1643, 2014.

Adrian J. Baddeley, Ege Rubak, and Rolf Turner. Spatial Point Patterns: Methodology and Applications with R. Chapman and Hall/CRC, Boca Raton, 2015.

Ole Eiler Barndorff-Nielsen. Information and Exponential Families: In Statistical Theory. John Wiley & Sons, 2014.

Julian E. Besag. Spatial interaction and the statistical analysis of lattice systems (with discussion). Journal of the Royal Statistical Society Series B, 36:192–236, 1974.

Julian E. Besag. Some methods of statistical analysis for spatial data. Bulletin of the International Statistical Institute, 47:77–92, 1977.

Julian E. Besag, Robin K. Milne, and Stan Zachary. Point process limits of lattice processes. Journal of Applied Probability, 19:210–216, 1982.

Sung Nok Chiu, Dietrich Stoyan, Wilfrid S. Kendall, and Joseph Mecke. Stochastic Geometry and its Applications. John Wiley & Sons, Chichester, 3rd edition, 2013.

Andreas Dyreborg Christoffersen, Jesper Møller, and Heidi Søgaard Christensen. Modelling columnarity of pyramidal cells in the human cerebral cortex. Australian & New Zealand Journal of Statistics, 63(1):33–54, 2021.

Jean-François Coeurjolly, David Dereudre, Rémy Drouilhet, and Frédéric Lavancier. Takacs–Fiksel method for stationary marked Gibbs point processes. Scandinavian Journal of Statistics, 39(3):416–443, 2012.

Thomas Cormen, Charles Leiserson, Ronald Rivest, and Clifford Stein. Introduction to Algorithms. MIT Press, 3rd edition, 2009.

Daryl J. Daley and David Vere-Jones. An Introduction to the Theory of Point Processes: Volume I: Elementary Theory and Methods. Springer, 2003.

Daryl J Daley and David Vere-Jones. *An Introduction to the Theory of Point Processes. Volume II: General Theory and Structure*. Springer, 2008.

Ngoc Anh Dao and Marc G Genton. A Monte Carlo-Adjusted Goodness-of-Fit Test for Parametric Models Describing Spatial Point Patterns. *Journal of Computational and Graphical Statistics*, 23(2):497–517, 2014.

David Dereudre. Introduction to the theory of Gibbs point processes. In David Coupier, editor, *Stochastic Geometry*, pages 181–229. Springer, 2019.

David Dereudre and Frédéric Lavancier. Campbell equilibrium equation and pseudo-likelihood estimation for non-hereditary Gibbs point processes. *Bernoulli*, 15(4):1368–1396, 2009.

David Dereudre and Frédéric Lavancier. Practical simulation and estimation for Gibbs Delaunay–Voronoi tessellations with geometric hardcore interaction. *Computational Statistics & Data Analysis*, 55(1):498–519, 2011.

David Dereudre, Rémy Drouilhet, and Hans-Otto Georgii. Existence of Gibbsian point processes with geometry-dependent interactions. *Probability Theory and Related Fields*, 153(3-4):643–670, 2012.

Eugene Edgington and Patrick Onghena. *Randomization Tests*. Chapman and Hall/CRC, 2007.

Fropuff. The vertex configuration of a tetrahedral-octahedral honeycomb, 2006. URL `https://en.wikipedia.org/wiki/File:TetraOctaHoneycomb-VertexConfig.svg`.

Charles J. Geyer and Jesper Møller. Simulation procedures and likelihood inference for spatial point processes. *Scandinavian Journal of Statistics*, 21:359–373, 1994.

Göran Grimvall. *Thermophysical Properties of Materials*. Elsevier, 1999.

Janine Illian, Antti Penttinen, Helga Stoyan, and Dietrich Stoyan. *Statistical Analysis and Modelling of Spatial Point Patterns*, volume 70. John Wiley & Sons, Chichester, 2008.

Daniel Jahn and Filip Seitl. Existence and simulation of Gibbs-Delaunay-Laguerre tessellations. *Kybernetika*, 56(4):617–645, 2020.

Jens Ledet Jensen and Hans R. Künsch. On asymptotic normality of pseudo likelihood estimates for pairwise interaction processes. *Annals of the Institute of Statistical Mathematics*, 46(3):475–486, 1994.

Jens Ledet Jensen and Jesper Møller. Pseudolikelihood for exponential family models of spatial point processes. *The Annals of Applied Probability*, 1(3): 445–461, 1991.

Jannick Kuhn, Matti Schneider, Petra Sonnweber-Ribic, and Thomas Boelke. Fast methods for computing centroidal Laguerre tessellations for prescribed volume fractions with applications to microstructure generation to polycrystalline materials. Computer Methods in Applied Mechanics and Engineering, 369:113–175, 2020.

Claudia Lautensack. Random Laguerre tessellations. PhD thesis, University of Karlsruhe, 2007.

Claudia Lautensack. Fitting three-dimensional Laguerre tessellations to foam structures. Journal of Applied Statistics, 35(9):985–995, 2008.

Claudia Lautensack and Sergei Zuyev. Random Laguerre tessellations. Advances in Applied Probability, 40(3):630–650, 2008.

Allan Lyckegaard, Erik Mejdal Lauridsen, Wolfgang Ludwig, Richard Warren Fonda, and Henning Friis Poulsen. On the use of Laguerre tessellations for representations of 3D grain structures. Advanced Engineering Materials, 13: 165–170, 2011.

Alberto A. Magrenan and Ioannis Argyros. A Contemporary Study of Iterative Methods. Elsevier, 1st edition, 2018.

Shigeru Mase. Consistency of the maximum pseudo-likelihood estimator of continuous state space gibbsian processes. The Annals of Applied Probability, 5 (3):603–612, 1995.

Jesper Møller. Random tessellations in $\mathbb{R}^d$. Advances in Applied Probability, 21 (1):37–73, 1989.

Jesper Møller and Rasmus Plenge Waagepetersen. Statistical Inference and Simulation for Spatial Point Processes. Chapman and Hall/CRC, Boca Raton, 2003.

Adam Morawiec. Orientations and Rotations. Springer, 2003.

Mari Myllymäki and Tomáš Mrkvička. GET: Global Envelopes in R, 2019. Available at arXiv:1911.06583.

Mari Myllymäki and Tomáš Mrkvička. Comparison of non-parametric global envelopes, 2020. Avaliable at arXiv:2008.09650.

Mari Myllymäki, Tomáš Mrkvička, Henri Seijo, Pavel Grabarnik, and Ute Hahn. Global envelope tests for spatial processes. Journal of the Royal Society of Statistics Series B, 79(2):381–404, 2017.

Atsuyuki Okabe, Barry Boots, Kokichi Sugihara, and Sung Nok Chiu. Spatial Tessellations: Concepts and Applications of Voronoi Diagrams. 2nd edition, 2000.

Antti Penttinen. Modelling Interactions in Spatial Point Patterns: Parameter Estimation by the Maximum Likelihood Method. Number 7 in Jyväskylä Studies in Computer Science, Economics and Statistics. University of Jyväskylä, 1984.

Lukas Petrich, Jakub Staněk, Mingyan Wang, Daniel Westhoff, Luděk Heller, Petr Šittner, Carl Emil Krill III, Viktor Beneš, and Volker Schmidt. Reconstruction of grains in polycrystalline materials from incomplete data using Laguerre tessellations. Microscopy and Microanalysis, 25(3):743–752, 2019.

Romain Quey and Loïc Renversade. Optimal polyhedral description of 3D polycrystals: Method and application to statistical and synchrotron X-ray diffraction data. Computational Methods in Applied Mechanical and Enginering, 330:308–333, 2018.

Claudia Redenbach, Inga Shklyar, and Heiko Andrä. Laguerre tessellations for elastic stiffness simulations of closed foams with strongly varying cell sizes. International Journal of Engineering Science, 50(1):70–78, 2012.

Chris H. Rycroft. Voro++: A three-dimensional Voronoi cell library in C++. Chaos, 19:041111, 2009.

Rolf Schneider and Wolfgang Weil. Stochastic and Integral Geometry, volume 1. Springer, 2008.

Ondřej Šedivý, Tim Brereton, Daniel Westhoff, Leoš Polívka, Viktor Beneš, Volker Schmidt, and Aleš Jäger. 3D reconstruction of grains in polycrystalline materials using a tessellation model with curved grain boundaries. Philosophical Magazine, 96(18):1926–1949, 2016.

Pavel Sedmák, Jan Pilch, L udě Heller, Jaromír Kopeček, Jon P. Wright, Petr Sedlák, Miroslav Frost, and Petr Šittner. Grain-resolved analysis of localized deformation in nickel-titanium wire under tensile load. Science, 353:559–562, 2016.

Filip Seitl, Lukas Petrich, Jakub Staněk, Carl Emil Krill III, Volker Schmidt, and Viktor Beneš. Exploration of Gibbs-Laguerre tessellations for three-dimensional stochastic modeling. Methodology and Computing in Applied Probability,, 23(2):669–693, 2021.

Filip Seitl, Jesper Møller, and Viktor Beneš. Fitting three-dimensional laguerre tessellations by hierarchical marked point process models. Spatial Statistics, 51:100658, 2022.

Aaron Spettl, Raphael Wimmer, Thomas Werz, Michael Heinze, Stefan Odenbach, Carl Emil Krill III, and Volker Schmidt. Stochastic 3D modeling of Ostwald ripening at ultra-high volume fractions of the coarsening phase. Modelling and Simulation in Materials Science and Engineering, 23(6):065001, 2015.

Aaron Spettl, Tim Brereton, Qibin Duan, Thomas Werz, Carl Emil Krill III, Dirk P. Kroese, and Volker Schmidt. Fitting Laguerre tessellation approximations to tomographic image data. Philosophical Magazine, 96(2):166–189, 2016.

Ole Stenzel, Omar Pecho, Lorenz Holzer, Matthias Neumann, and Volker Schmidt. Predicting effective conductivities based on geometric microstructure characteristics. AIChE Journal, 62(5):1834–1843, 2016.

Dietrich Stoyan, Viktor Beneš, and Filip Seitl. Dependent radius marks of laguerre tessellations: a case study. Australian & New Zealand Journal of Statistics, 63(1):19–32, 2021.

André Tscheschel and Dietrich Stoyan. Statistical reconstruction of random point patterns. Computational Statistics & Data Analysis, 51(2):859–871, 2006.

Peter J. M. van Laarhoven and Emile H. L. Aarts. Simulated Annealing: Theory and Applications. Springer, 1987.

Ondřej Šedivý, Daniel Westhoff, Jaromír Kopeček, Carl Emil Krill III, and Volker Schmidt. Data-driven selection of tessellation models describing polycrystalline microstructures. Journal of Statistical Physics, 172:1223–1246, 2018.

Hans Zessin. Point processes in general position. Journal of Contemporary Mathematical Analysis, 43(1):59–65, 2008.

List of Figures

1.1 Four slices in xy plane of the observed Laguerre tessellation. The cells are coloured randomly. 10

1.2 The spatial point pattern $\mathbf{x}_n$ (upper left) of the generators of the Laguerre tessellation representing nickel-titanium alloy after applying the periodic boundary condition together with its projections onto the xy, xz and yz planes. 11

1.3 The radii $\mathbf{r}_n$ of the generators of the Laguerre tessellation representing nickel-titanium alloy after applying the periodic boundary condition: a) the balls defined by $(\mathbf{x}_n, \mathbf{r}_n)$; b) histogram of the radii together with a kernel density estimate (solid line) and a fitted density of a beta distribution (dashed line) obtained by maximum likelihood estimation; c) eight kernel density estimates based on the radii associated to the points in the eight sets for the subdivision of W obtained by dividing its sides into halves. . . . 12

1.4 Relative histograms of some geometric characteristics of the Laguerre tessellation representing nickel-titanium alloy. From left to right, histograms of cell volume, surface area, number of faces, sphericity and dvol (absolute difference in volumes for neighbouring cells). 13

1.5 Aluminum alloy specimen measured by synchrotron X-ray tomography (cf. Spettl et al. [2016]) : a) original voxelized image; b) approximation by a Laguerre tessellation. The grains are coloured randomly. 14

1.6 The spatial point pattern $\mathbf{x}_n$ (upper left) of the generators of the Laguerre tessellation representing aluminium alloy after applying the periodic boundary condition together with its projections onto the xy, xz and yz planes. 15

1.7 The radii $\mathbf{r}_n$ of the generators of the Laguerre tessellation representing aluminium alloy after applying the periodic boundary condition: a) the balls defined by $(\mathbf{x}_n, \mathbf{r}_n)$; b) histogram of the radii together with a kernel density estimate (solid line) and a fitted density of a beta distribution (dashed line) obtained by maximum pseudolikelihood estimation; c) eight kernel density estimates based on the radii associated to the points in the eight sets for the subdivision of W obtained by dividing its sides into half parts. . . 16

1.8 Relative histograms of some geometric characteristics of the Laguerre tessellation representing aluminium alloy. From left to right and up to down, histograms of cell volume (denoted as $hist^{exd}_{\mathrm{vol}}$), surface area, number of faces (denoted as $hist^{exd}_{\mathrm{nof}}$), sphericity and dvol (absolute difference in volumes for neighbouring cells). . . . 16

2.1 Violation of normality in 2D. Left: the tessellation is not face-to-face. Right: the vertex (0-face) in the red circle is contained in the boundary of 4 cells instead of 3. 21

2.2 Illustration of the power distance given in (2.1). All three points P, Q, R have the same power distance with respect to circles with centers S, T and radii s, t, respectively. Δ is the square root of the power distance 23

3.1 For the two histograms coloured in this figure by red and black, their discrepancy is a constant proportional to a sum of frequencies corresponding to the regions shaded in gray. 54

3.2 Laguerre 5-flower LF_5 of grey generators. Black spheres depicture generators. Grey and blue spheres are their characteristic points and finite horizons, respectively. 61

3.3 The columns A and B show number of faces and volume [μm^3], respectively. Relative histograms on row I are computed from a single realization. To show the variability of the model row II shows kernel estimates of the densities computed from ten realizations together with histograms coming from the experimental data—in gray. 67

3.4 The reconstruction using the first and the second moments of both, nof and vol, model (3.18) – column A and B show relative histograms and densities of number of faces and volume [μm^3], respectively; the histograms of relative frequencies in row I are computed for one realization; to illustrate the variability of the random tessellation, row II shows densities computed for a bunch of ten realizations together with histograms coming from the experimental data—in gray. 68

3.5 Reconstruction of experimental data, controlling the distribution of the numbers of faces per cell, model (3.19): from top to bottom the parameter θ_n takes on the values 10, 100, 1 000 and 10 000; column A shows the histograms of relative frequency computed from a single realization, and column B shows kernel density estimates based on ten realizations together with the histogram $hist^{exd}_{\text{nof}}$ coming from the experimental data (cf. Figure 1.8)—in gray. 70

3.6 Reconstruction of the data using the histogram of the cell volume, model (3.20). The cell volume is expressed in μm^3. From the top the value of parameter θ_n increases, the values are 10, 100, 1000 and 10000. Column A shows the evolution of the histogram of relative frequencies (computed from one realization) and column B shows the densities of ten realizations together with the targeting histogram $hist^{exd}_{\text{vol}}$, that comes from the experimental data, see Figure 1.8. The corresponding discrepancies are stated in Table 3.5. 71

3.7 Reconstruction of experimental data, controlling distributions of the number of faces per cell and of the cell volume, model (3.21) with $\theta^1_n = 1\,000$ and $\theta^2_n = 10\,000$: column A shows the number of faces per cell, the histogram of relative frequencies computed from one realization and kernel density estimates based on ten realizations; column B shows the same plots for the cell volume. The gray histograms are those of the experimental data, i.e., $hist^{exd}_{\text{nof}}$ and $hist^{exd}_{\text{vol}}$, cf. Figure 1.8. 73

3.8 The comparison of the reconstruction stopped by stopping criterion $(\delta, t) = (0.02, 100000)$ and the simulation performing 2 millions steps. Both are based on model (3.19) with control parameter $\theta_n = 1000$. Two subfigures show the time evolution of the histogram discrepancy (energy) for 5 simulations. The stopping criterion (δ, t) results in shorter runs - all 5 reconstructions (on the left) were stopped after less than 320 thousands of iterations. 74

3.9 Evolution of the discrepancy of histograms of cell volume for (a) the greedy algorthm and (b) the MHBDM algorithm. The red line represents the discrepancy $y = 0.05$ 77

3.10 Evolution of the discrepancy of histograms of number of faces per cell for (a) the greedy algorithm, (b) the MHBDM algorithm with $\theta_n = 1\ 000$ and (c) the MHBDM algorithm with $\theta_n = 10\ 000$. The red line represents the discrepancy $y = 0.05$ 77

3.11 Samples of simulated Laguerre tessellations T12, left, and T18, right. The corresponding numbers of cells are 1040 and 1033, respectively. The cells are coloured randomly. 79

3.12 The empirical pair correlation functions averaged for the five simulated patterns. Solid line = T12, dashed line = T18. 81

3.13 The empirical mark variograms $\gamma_m(r)$ averaged for the five simulated patterns. Solid line = T12, dashed line = T18. 81

3.14 The mark correlation functions $k_{mm}(r)$ averaged for the five simulated patterns. Solid line = T12, dashed line = T18. 81

3.15 Histogram of relative frequencies of the NVR statistic for the simulated tessellation based on (a) irregular model (4497 cells) and (b) regular model (752 cells). 83

4.1 Nearest neighbour distance. Left: NiTi alloy, the minimal realized distance is $0.3824 \mu m$. Right: alluminium alloy, the minimal realized distance is $7.737 \mu m$. 86

4.2 Empirical F, G, K and $L(r) - r$ functions for generators' locations of Laguerre tessellation representing NiTi alloy (in black) compared to the theoretical curves under Poisson point process (in red). 86

4.3 Empirical F, G, K and $L(r) - r$ functions for generators' locations of Laguerre tessellation representing alluminium alloy (in black) compared to the theoretical curves under Poisson point process (in red). 86

4.4 Empirical pair correlation function for generators' locations of Laguerre tessellation representing NiTi alloy, $h = 1.4$. 87

4.5 Empirical pair correlation functions for the Laguerre tessellation data representing alluminium alloy, all estimated with isotropic edge correction using `spatstat`, Baddeley et al. [2015]. Solid line: all data, $h = 14.34$; dashed line: all data, $h = 5$; dotted line: reduced data (without 7 close pairs), $h = 5$. For $r > 20$, the dashed and dotted lines almost coincide. 88

4.6 Empirical mark correlation function, a), and mark variogram, b), for the Laguerre tessellation data representing NiTi alloy, $h = 1.4$. 89

4.7 Empirical mark correlation function, a), and mark variogram, b), for the Laguerre tessellation data representing alluminium alloy, $h = 5$. 89

4.8 The empirical mark correlation function (the solid line) and a 95%-global envelope (the grey region) obtained by permuting the radii when fixing the points. The dashed line is the average of the simulated mark correlation functions and the red dots indicate when the empirical function is outside the envelope. 91

4.9 From left to right, empirical functional summary statistics $\hat{L}(t)-t$, $\hat{F}(t)$ and $\hat{G}(t)$ (solid lines) and simulated 95%-global envelope (grey regions) obtained under a fitted homogeneous Poisson process. Dashed lines are averages of the simulated functional summary statistics and the dots indicate when the empirical functions are outside the envelope. 98

4.10 From left to right, empirical functional summary statistics $\hat{L}(t)-t$, $\hat{F}(t)$ and $\hat{G}(t)$ (solid lines) and simulated 95%-global envelope (grey region) obtained under a fitted Strauss process. Dashed lines are averages of the simulated functional summary statistics and the dots indicate when the empirical functions are outside the envelope. 98

4.11 From left to right, empirical functional summary statistics $\hat{L}(t)-t$, $\hat{F}(t)$ and $\hat{G}(t)$ (solid lines) and simulated 95%-global envelope (grey regions) obtained under a fitted multiscale process with $d = 3$. Dashed lines are averages of the simulated functional summary statistics. 99

4.12 Estimated densities of tessellation characteristics, namely nof (a), vol (b), surf (c), tel (d), spher (e) and dvol (f). The solid lines are the functions based on the data $(\mathbf{x}_n, \mathbf{r}_n)$ and the grey regions are 95%-global envelopes under the fitted model 'beta+nof+dvol'. Dashed lines are averages of the simulated densities. 102

A.1 An example of pseudo-periodic configuration for unmarked points in 2D. 126

A.2 Nonredundancy of points in 2D. Spatial parts of sets A, i.e., $B(0, \rho a)$, are gray. Black circles are marked points with the maximal radii represented by bold interrupted line of the length $\frac{a}{2}(1-2\rho)$ and $a(1-2\rho)$ in the upper and lower part of the picture, respectively. Those in the upper part do not overlap and those in the lower part do not overlap over center. 126

A.3 Tetrahedral-octahedral honeycomb. 127

A.4 Fropuff [2006] Tetrahedral-octahedral honeycomb tessellate in 3D. Each vertex is incident to eight regular tetrahedra (yellow) and six regular octahedra (blue), which are shown in an exploded view. . 128

A.5 Prescribed sphericity distribution for the simulations of Gibbs Laguerre tessellation. 129

A.6 Densities of the sphericity distribution in MCMC simulations of Gibbs Laguerre tessellations. The dashed line represents the prescribed sphericity distribution, cf. Figure A.5. Black and grey full lines are densities coming from simulations. In diagram a) densities for θ equal to $-10000, -2000, -1000, -1, 1, 1000, 1500$ have their peak magnitudes in the descending order. In diagram b) the detail of diagram a) is present for θ equal to $-1, 1, 1000, 1500$. . . 130
A.7 The 3D simulations of the sphericity model for $\theta = -10000$, a), and $\theta = 1500$, b). The colouring is random. 130
A.8 Discrepancy between simulated and prescribed histogram (up) and number of cells (down) for varying values of the parameter θ in MCMC simulations of Gibbs Laguerre tessellations. In column A and B the negative and positive values of parameter θ are involved, respectively. Namely in column A, $\theta = -1$ corresponds to black full line, $\theta = -1000$ to grey full line, $\theta = -2000$ to black dashed line and $\theta = -10000$ to grey dashed line. In column B, $\theta = 1$ corresponds to black full line, $\theta = 1000$ to grey full line and $\theta = 1500$ to black dashed line. We draw each 10^5-th iteration. 131
A.9 Comparison of Gibbs-Laguerre tessellations obtained from model (A.3) and model (3.22)-irregular (red) with Poisson-Laguerre tessellations (black) – estimated distributions of Laguerre radii, cell volumes, number of faces per cell and NVR are presented. 133
A.10 Comparison of the Gibbs-Laguerre tessellations obtained from model model (3.22)-regular and model (A.4) (red) with Poisson-Laguerre tessellations (black) – estimated distributions of Laguerre radii, cell volumes, number of faces per cell and NVR are presented. . . 134

List of Tables

1.1 Laguerre tessellation characteristics for the nickel-titanium alloy data. In the upper part of the table there are numbers of cells and faces per sample. In the lower part of the table there are mean values and standard deviations (sd) of cell characteristics evaluated from the data. 13
1.2 Correlations of tessellation characteristics. 13
1.3 Correlations of tessellation characteristics with radii. 14
1.4 Laguerre tessellation characteristics for the aluminium alloy data. In the upper part of the table there are numbers of cells and faces per sample. In the lower part of the table there are mean values of characteristics per cell and standard deviations (sd) for data. . 17

3.1 Means and standard deviations of various tessellation characteristics of the aluminium alloy data, the third column, are normalized to the unit volume, the fourth column. Normalization means that volume-based, area-based and length-based characteristics are divided by the volume, the square root of the volume and the cube root of the volume of the window W, respectively. 65
3.2 Moment reconstruction - numerical values. Two parameter specifications (3.16) and (3.17) controlling the first two moments of nof and the first two moments of vol, respectively, were used to obtain realizations described in this table. 66
3.3 Moment reconstruction of the experimental data – model (3.18) (values from the experimental data are given in brackets). 69
3.4 Dependence on θ_n of the discrepancy between histograms for the number of faces per cell and histograms for the cell volumes in model (3.19). 72
3.5 Dependence on θ_n of the discrepancy between histograms for the number of faces per cell and histograms for the cell volumes in model (3.20). 72
3.6 Dependence on the parameters θ_n^1 and θ_n^2 of the discrepancy of histograms for the number of faces per cell and for the cell volume in model (3.21). 73
3.7 Laguerre tessellation characteristics for samples T12 and T18. Five independently simulated samples and ten permuted samples from the first simulated sample are evaluated. In the upper part of the table there are the means and below the standard deviations (sd) of numbers of cells and faces per sample. In the lower part of the table there are for each characteristic three numbers. Firstly, I, the mean over all cells in all five samples (columns 'data') or ten samples (columns 'permuted'). Further, II, the square root of the mean of variances within the samples and III, the square root of the variance of mean values within the samples (again with respect to five samples in columns 'data' and to ten samples in columns 'permuted'). 80

3.8 Estimated parameters based on tessellations obtained from regular and irregular model - the table shows estimates of hardcore parameters α, β, soft-core parameter θ and activity z; the means and standard deviations of estimates are obtained from 20 realizations. 83

4.1 Laguerre tessellation characteristics for the aluminium alloy data, subdata without 14 generators with close neighbours and average values for ten samples with randomly permuted radius marks. In the upper part of the table there are numbers of cells and faces per sample, for permuted samples mean values are given (with standard deviation in brackets). In the lower part of the table there are mean values of characteristics per cell and standard deviation (below the means) for data and subdata. For the permuted samples overall means and square roots of the mean of variances within samples are given. 92
4.2 Means and standard deviations of the tessellation characteristics given in (4.6) and as obtained by simulations under various joint models and from the data. Deviations are given in percentages. See the text for details. 100
4.3 Maximized log pseudolikelihood functions for radii models conditioned on the points. The value for the 'beta' model is given as a reference. 100

A.1 Dependency of numerical characteristics on the value of parameter θ. 129
A.2 $\mathcal{M}_2$: Values of contrast log pseudolikelihood function increased by observation window volume (additive constant) and the estimates of parameters β and γ_1 for the prescribed grid values of δ_1. 136
A.3 $\mathcal{M}_3$: Values of contrast log pseudolikelihood function increased by observation window volume (additive constant) and the estimates of parameters β, γ_1 and γ_2 for the prescribed grid values of δ_1 and δ_2. 137

List of Abbreviations

$\mathbb{R}$	real numbers	
$\mathbb{R}_+$	non-negative real numbers	
$\mathbb{N}$	natural numbers	
$\mathbb{Z}$	integer numbers	
$\mathbb{X}$	locally compact Polish space	19
$\mathbb{M}$	mark space	20
∂D	boundary of $D \subset \mathbb{R}^d$	
$\bar{D}$	closure of $D \subset \mathbb{R}^d$	
$\mathrm{int}(D)$	interior of $D \subset \mathbb{R}^d$	
card	cardinality	
conv	convex hull	
diam	diameter	
dim	dimension	
bar	barycenter	
$\oplus$	Minkowski sum operator	
$\mathcal{L}$	Lebesgue measure	
δ_x	Dirac measure concentrated in a point x	
$\sigma(\cdot)$	σ-algebra generated by collection of sets	
$\mathcal{B}$	Borel sets	
$\mathcal{B}_b$	bounded Borel sets	
$\mathcal{B}^d$	Borel sets on $\mathbb{R}^d$	
$\mathcal{B}_b^d$	bounded Borel sets on $\mathbb{R}^d$	
$\mathcal{C}$	system of all convex polygons in $\mathbb{R}^3$	
3D	three-dimensional	
$SO(d)$	d-dimensional rotation group	30
$\|\cdot\|$	Euclidean distance	
$\mathbf{o}$	origin	
$b(x,r)$	open ball in $\mathbb{R}^d$ with center $x \in \mathbb{R}^d$ and radius $r > 0$	
$\angle(u,v)$	angle between vectors u and v	
$\vartheta_t(\cdot)$	translation by $t \in \mathbb{R}^d$	30
$R_{\mathcal{O}}(\cdot)$	rotation by $\mathcal{O} \in SO(d)$	31
ω_d	volume of the d-dimensional unit sphere in $\mathbb{R}^d$	39
σ_d	surface of the $d-1$-dimensional unit sphere in $\mathbb{R}^d$	39
$\Gamma(\cdot)$	Gamma function	
$\mathbb{I}_{[\cdot]}$	indicator function	
f^+, f^-	positive, negative part of function f	
$\mathrm{dsc}(H,H')$	histogram discrepancy between a pair of histograms (H,H')	54
$\mathbf{p}^\top$	transposition of vector $\mathbf{p} \in \mathbb{R}^d$	
$a \wedge b$	maximum of a nad b	

$\mathbf{N}$	set of locally finite (possibly marked) counting measures/point patterns	19
$\mathbf{x}$	point pattern/ configuration	
$\mathbf{m}$	vector of marks	
$(\mathbf{x}, \mathbf{r})$	marked point pattern/ configuration with marks interpreted as radii	
$(\mathbf{x}^*, \mathbf{r}^*)$	periodic configuration	25
$\mathcal{N}$	σ-algebra on $\mathbf{N}$	20
$\mathbf{N}^f$	set of finite (marked) point patterns	20
$\mathcal{N}^f$	σ-algebra on $\mathbf{N}^f$	20
$\mathbf{N}^{f,k}$	set of finite (marked) point patterns with exactly k points	20
$\mathbf{N}_W$	set of all finite (marked) point configurations in $W \in \mathcal{B}_b(\mathbb{X})$	20
$\mathbf{N}_{gp}$	set of all locally finite (marked) point configurations being in general position	26
$\mathbf{N}_\infty$	set of admissible (marked) point configurations	35

$(\Omega, \mathcal{A}, \mathbb{P})$	probability space	
$\sim_d$	equality in distribution	
$\mathsf{E}\,(X)$	expected value of random variable X	
var	variance	
sd	standard deviation	
$\bar{Y}$	sample mean of the sample $Y = (Y_1, \dots, Y_l)$	
S_Y^2	sample variance of the sample $Y = (Y_1, \dots, Y_l)$	
$Unif$	uniform distribution	
$\mathbb{Q}$	mark distribution	37

Λ	intensity measure	30
λ	intensity function	30
λ^*	Papangelou conditional intensity	31
P_x	Palm distribution in point x	30
π_W^Λ	distribution of Poisson point process	31

PPP	Poisson point process	31
GPP	Gibbs point process	32
$\mathcal{M}_q$	Multiscale point process of order $q \in \mathbb{N}$	33

DLR	Dobrushin-Lanford-Ruelle equations	33
GNZ	Georgii-Nguyen-Zessin equations	35
E	energy function	32
Z_W	partition function on $W \in \mathcal{B}_b^d$	32
V_k	potential function of order k	51
$\bar{E}$	periodic energy function	51
$h(y, \mathbf{x} - \delta_y)$	local energy of y in $\mathbf{x} - \delta_y$	35
$\mathcal{R}(\mathbf{x})$	set of removable points in $\mathbf{x} \in \mathbf{N}$	36
$V_{k,T}^G$	reconstructing potential	53
$\varphi_{k,T}^G$	hypergraph version of reconstructing potential $V_{k,T}^G$	56

$F(t)$	F function	38
$G(t)$	G function	39
$K(t)$	K function	39
$L(t)$	L function	39
$g(t)$	pair correlation function	39
$k_{mm}(t)$	mark correlation function	39
$\gamma_m(t)$	mark variogram	39

$\mathbf{T}$	set of all tessellations on $\mathbb{R}^d$	42
$\mathcal{T}$	σ-algebra on $\mathbf{T}$	42
C_i, $C(x,r)$	the i-th tessellation cell, tessellation cell generated by marked point (x,r)	20
$\sim_k$	k-neighbourhood relation	24
$\Delta_s(C)$	set of s-faces of a cell C	21
$S_s(T)$	set of s-faces of a tessellation T	21

$Vor(\mathbf{x})$	Voronoi tessellation	21
$L(\mathbf{x},\mathbf{r})$	Laguerre tessellation	22
$D(\mathbf{x},\mathbf{r})$	Delaunay tessellation	27
GBPD	generalized balanced power diagram	23

vol	cell volume	24
nof	number of faces (per cell)	24
noe	number of edges (per cell)	24
nov	number of vertices (per cell)	24
surf	surface area	24
tel	total edge length	24
spher	sphericity	24
dvol	absolute difference in volume for two neighbouring cells	24
NVR	neighbour-volume ratio	24
farea	face area	24
fper	face perimeter	24
fnoe	number of edges in a face	24

MLE	maximum likelihood estimate	
MPLE	maximum pseudolikelihood estimate	46
$\mathcal{PL}$	pseudolikelihood function	46
$cl\mathcal{PL}$	log pseudolikelihood contrast function	63

(x_η, r_η)	characteristic point	26
$\mathcal{E}$	hypergraph structure	28
$\mathcal{LD}(\mathbf{x},\mathbf{r})$	set of regular subsets of $(\mathbf{x},\mathbf{r})$	26
$\mathcal{LD}_k(\mathbf{x},\mathbf{r})$	set of regular subsets of $(\mathbf{x},\mathbf{r})$ with cardinality k	26
$\mathcal{CG}_k(\mathbf{x},\mathbf{r})$	graph of connected k-tuples	28
$\mathcal{CG}_{k,b}(\mathbf{x},\mathbf{r})$	graph of connected k-tuples with bounded cells	28

$\varphi(\eta,\mathbf{g})$	hyperedge potential	28
$\mathbf{N}_{cr}^{W}$	set of locally finite (marked) point configurations confining the range of φ from $W \in \mathcal{B}_b$	29

List of publications

Petr Král, Jakub Staněk, Lenka Kunčická, Filip Seitl, Lukas Petrich, Volker Schmidt, Viktor Beneš, Václav Sklenička. Microstructure changes in HPT - processed copper occurring at room temperature. Materials Characterization, 151: 602–611, 2019.

Jaromír Kopeček, Jakub Staněk, Stanislav Habr, Filip Seitl, Lukas Petrich, Volker Schmidt, Viktor Beneš. Analysis of polycrystalline microstructure of AlMgSc alloy observed by 3D EBSD. Image Analysis & Stereology, 39(1): 1–11, 2020.

Jakub Staněk, Jaromír Kopeček, Petr Král, Iva Karafiátová, Filip Seitl, Viktor Beneš. Comparison of segmentation of 2D and 3D EBSD measurements in polycrystalline materials. Metallic Materials, 58: 301–319, 2020.

Daniel Jahn, Filip Seitl. Existence and simulation of Gibbs-Delaunay-Laguerre tessellations. Kybernetika, 56(4): 617–645, 2020.

Filip Seitl, Lukas Petrich, Jakub Staněk, Carl E. Krill III, Volker Schmidt and Viktor Beneš. Exploration of Gibbs-Laguerre tessellations for three-dimensional stochastic modeling. Methodology and Computing in Applied Probability, 23: 669–693, 2021.

Dietrich Stoyan, Viktor Beneš, Filip Seitl. Dependent radius marks of Laguerre tessellations: a case study. Australian & New Zealand Journal of Statistics, 63(1): 19–32, 2021.

Filip Seitl, Viktor Beneš, Jesper Møller. Fitting three-dimensional Laguerre tessellations by hierarchical marked point process models. Spatial Statistics, 51:100658, 2022.

A. Attachments

A.1 Pseudo-periodic configurations

When dealing with the existence problem of Gibbs models, we can restrict ourselves to so-called pseudo-periodic configurations, a key concept in the verification of the upper regularity condition **(U)**, see Subsection 3.2.2 and Dereudre et al. [2012].

Let $M \in \mathbb{R}^{3\times 3}$ be an invertible 3×3 matrix with column vectors (M_1, M_2, M_3). For each $k \in \mathbb{Z}^3$ define the cell

$$C(k) = \{Mx \in \mathbb{R}^3 : x - k \in [-1/2, 1/2)^3\}.$$

These cells partition $\mathbb{R}^3$ into parallelepipeds, i.e., solids whose six faces are all parallelograms in $\mathbb{R}^2$. We write $C = C(0)$. Let $\Gamma \subset \mathbf{N}_C$ be measurable and non-empty. Then we define the pseudo-periodic configurations $\bar{\Gamma}$ as

$$\bar{\Gamma} = \{\mathbf{x} \in \mathbf{N} : \vartheta_{Mk}\mathbf{x}_{C(k)} \in \Gamma \text{ for all } k \in \mathbb{Z}^3\},$$

the set of all configurations whose restriction to $C(k)$, when shifted back to C, belongs to Γ. The prefix pseudo- refers to the fact that the configuration itself does not need to be identical in all $C(k)$, it merely needs to belong to the same class of configurations.

A.1.1 Configuration for $\mathbb{R}^3$ tetrahedrization

Here we introduce and analyze the pseudo-periodic configuration used in the proofs of existence of our tessellation models. Fix some $A \subset C \times \mathbb{M}$, where $\mathbb{M} = [0, R_{max}]$, $R_{max} > 0$, and define

$$\Gamma^A = \{\zeta \in \mathbf{N}_C : \zeta = \{p\}, p \in A\},$$

the set of configurations consisting of exactly one point in the set A. The set of pseudo-periodic configurations $\bar{\Gamma}$ thus contains only one point in each $C(k), k \in \mathbb{Z}^3$.

Let M be such that $|M_i| = a > 0$ for $i = 1, 2, 3$ and $\angle(M_i, M_j) = \pi/3$ for $i \neq j$. We choose

$$A = B(0, \rho a) \times \left[0, \left(\frac{a}{2}(1 - 2\rho)\right)^2\right] \tag{A.1}$$

in order for balls to never overlap. Hence no problems with point redundancy occur. We choose $\rho < 1/4$ in order for points to remain in a general position.

Remark (Nonredundancy of points). The interval for radii in (A.1) is too conservative since we can admit balls to partially overlap and instead of $\left(\frac{a}{2}(1 - 2\rho)\right)^2$ use $(a(1 - 2\rho))^2$ as an upper bound. Then the balls may overlap but not over its centers. Therefore the nonredundancy of points is still preserved.

A useful mental model of how to think of the class $\bar{\Gamma}$ is to start from a configuration

$$\mathbf{x}_0 = \{(M_a k, 0) \in \mathbb{R}^3 \times \mathbb{M} : k \in \mathbb{Z}^3\} \in \bar{\Gamma},$$

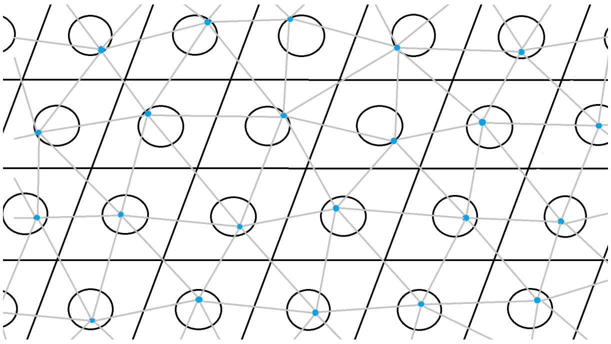


Figure A.1: An example of pseudo-periodic configuration for unmarked points in 2D.

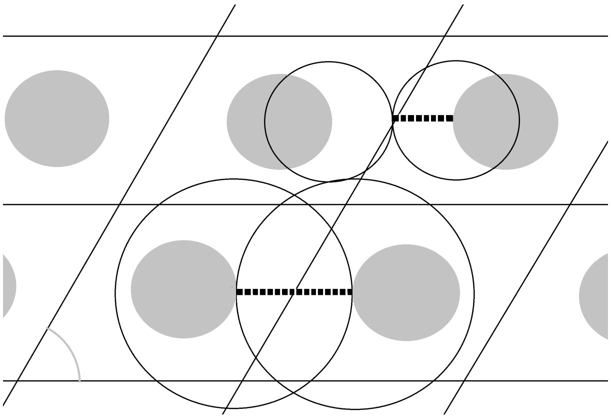


Figure A.2: Nonredundancy of points in 2D. Spatial parts of sets A, i.e., $B(0, \rho a)$, are gray. Black circles are marked points with the maximal radii represented by bold interrupted line of the length $\frac{a}{2}(1-2\rho)$ and $a(1-2\rho)$ in the upper and lower part of the picture, respectively. Those in the upper part do not overlap and those in the lower part do not overlap over center.

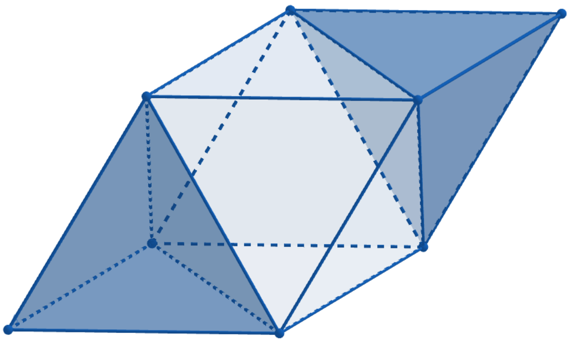


Figure A.3: Tetrahedral-octahedral honeycomb.

with points at the centers of the set A and then imagine any configuration $\mathbf{x} \in \bar{\Gamma}$ as a perturbed version of $\mathbf{x}_0$. In the following remark we describe how the tetrahedrizations formed by $\mathbf{x}_0$ look like. The tetrahedrization of any $\mathbf{x} \in \bar{\Gamma}$ is considered as a perturbed version of tetrahedrization of $\mathbf{x}_0$. When speaking about these perturbed tetrahedrizations we will shortly say: "up to a perturbation".

Remark (Pseudo-periodic tessellation). While in the 2D case Dereudre et al. [2012] the point configuration forms a tessellation out of equilateral triangles (up to a perturbation), the three-dimensional case results into the so-called tetrahedral-octahedral honeycomb (up to a perturbation), cf. Fig. A.3. This tessellation, if tetrahedrized, contains two different types of tetrahedra : a regular tetrahedron with side length a and an irregular tetrahedron with side lengths $(a, a, a, a, a, \sqrt{2}a)$, again, up to a perturbation. We will refer to a tetrahedron that is a perturbed version for the regular tetrahedron as T_1 and similarly T_2 for the irregular tetrahedron. In the tetrahedron-octahedron tessellation, each vertex is incident to eight regular tetrahedra and six regular octahedra, cf. Fig. A.4. Since each octahedron contains four tetrahedra, we obtain the bound for n_T, the number of incident tetrahedra of each vertex, $n_T \leq 8 + 6 \cdot 4 = 32$.

To show that the range of interactions is limited for the configurations in $\bar{\Gamma}$ two quantities need to be shown to be uniformly bounded. The first one is the circumdiameter of the tetrahedra described in Remark A.1.1.

Remark (Bounding the circumdiameter of tetrahedra). As noted in Remark A.1.1, there are two types of tetrahedra in the tetrahedrization, T_1 and T_2. The following argument describes how to obtain the bound for the regular tetrahedron (T_1), but the same procedure applies to T_2 as well. The optimization problem to be solved is

$$\begin{aligned} \underset{x_1,x_2,x_3,x_4 \in \mathbb{R}^3}{\text{maximize}} \quad & \chi(\{x_1, x_2, x_3, x_4\}) \\ \text{subject to} \quad & \|x_i - t_i\| \leq \rho a, \quad t_i \in \mathbb{R}^3, i = 1, 2, 3, 4, \\ & \|t_i - t_j\| = a, \quad i = 1, 2, 3, 4. \end{aligned} \tag{A.2}$$

An essential finding is that the points $x_1, \ldots, x_4$ which maximize the circumdiameter χ form a sphere tangent[1] to the spheres $\mathbb{S}_i := \partial B(t_i, \rho a), i = 1, \ldots, 4$. This reduces the number of possible solutions to 2^4 (even less because of symmetry) and all that remains is to check the largest solution. If $t_i = (t_{i,1}, t_{i,2}, t_{i,3})$,

---

[1] Spheres are said to be tangent if they intersect at a point.

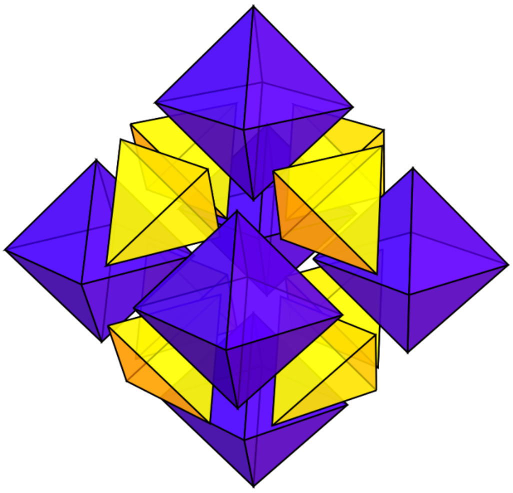


Figure A.4: Fropuff [2006] Tetrahedral-octahedral honeycomb tessellate in 3D. Each vertex is incident to eight regular tetrahedra (yellow) and six regular octahedra (blue), which are shown in an exploded view.

$i = 1, 2, 3, 4$, then we are looking for a sphere $\mathbb{S} = \partial B(y, r)$, $y = (y_1, y_2, y_3) \in \mathbb{R}^3$, such that

$$(y_1 - t_{i,1})^2 + (y_2 - t_{i,2})^2 + (y_3 - t_{i,3})^2 = (r - e_i \rho a)^2, \quad i = 1, 2, 3, 4,$$

where $e_i = \pm 1$, creating the said 2^4 possible solutions. It is possible to linearize these equations. Solving linear equations and choosing the solution yielding the largest circumradius give the following bounds. For $T1$ tetrahedra, we obtained the bound

$$\chi_1(\rho) := 2(\sqrt{6}/4 + \rho)$$

and the maximum in (A.2) is $a\chi_1(\rho)$. For T2, we obtain the bound

$$\chi_2(\rho) := 2\frac{2\rho + \sqrt{2 - 32\rho^2 + 64\rho^4}}{2 - 32\rho^2},$$

and the maximum is $a\chi_2(\rho)$. Both cases are valid for $\rho < 1/4$.

The second quantity to be bounded is the weight of a characteristic point.

Remark (Bounding the weight of the characteristic point). Since the perturbation happens on a bounded window and the points' weights are bounded, this amounts to proving that the points cannot come arbitrarily close to, or even attain, a coplanar position. However, this is equivalent to the boundedness of the circumdiameter of the tetrahedron, which we have already proven.

A.2 Sphericity

In order to give a vivid description of a model based on reconstructing potential (3.5), we would like to simulate Gibbs-Laguerre tessellation whose energy consists of a single reconstructing potential $\varphi_{k,T}^{G}$ multiplied by a real parameter θ:

$$\tilde{E}(\mathbf{x}, \mathbf{r}) = V_{hard} + \theta V_{n,T}^{G}(C_1, \ldots, C_n),$$

where V_{hard}, (3.2), forbids the empty cells. The reconstructing potential is important in practice as it allows simulation of tessellations that are similar to real data in terms of the distribution of a chosen geometric characteristic G.

In particular we set the functional T to be the histogram discrepancy, i.e., $T(s_G(\cdot)) = \mathrm{dsc}(H_{s_G(\cdot)}, H'_{s_G})$, and the cell characteristic G to be sphericity. The constant s_0 is set to be 0. The targetting histogram of sphericity H'_{s_G} is plotted in Figure A.5. The activity z is fixed to be 1000.

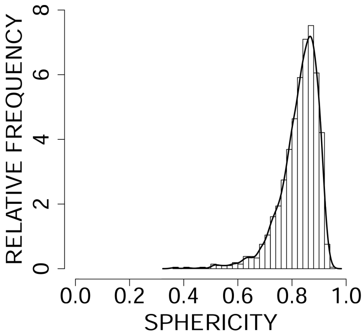


Figure A.5: Prescribed sphericity distribution for the simulations of Gibbs Laguerre tessellation.

Table A.1: Dependency of numerical characteristics on the value of parameter θ.

θ	total number of cells	discrepancy
-10000	349	1.57805
-5000	374	1.578
-2000	468	1.441
-1500	805	1.076
-1000	846	0.9386
-1	950	0.573
1	1018	0.566
1000	1020	0.187
1500	985	0.0175
10000	1000	0.0169

The choice of the parameter θ determines the properties of the model. The presented simulation study shows the influence of θ on the sphericity distribution, the discrepancy, and the number of non-empty cells in Λ. Numerical results are summarized in Table A.1 and Figures A.6 and A.7. Finally, Figure A.8 shows how

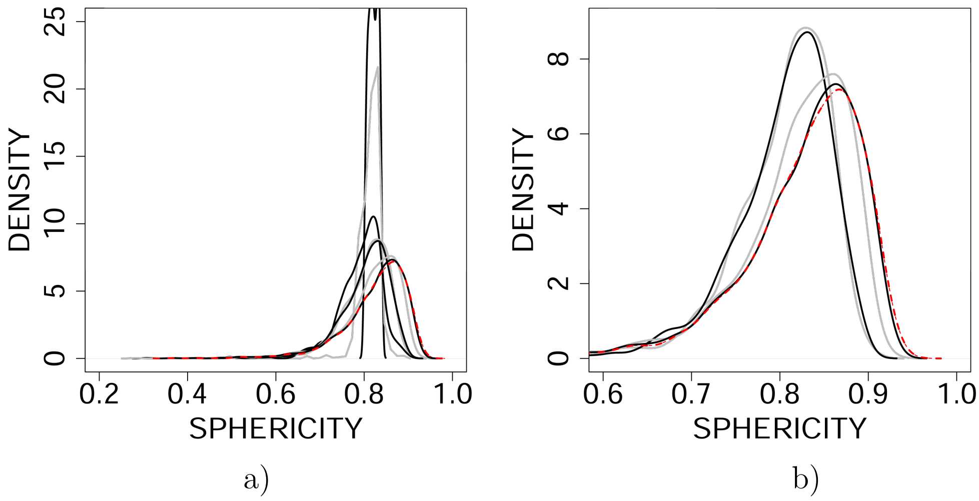


Figure A.6: Densities of the sphericity distribution in MCMC simulations of Gibbs Laguerre tessellations. The dashed line represents the prescribed sphericity distribution, cf. Figure A.5. Black and grey full lines are densities coming from simulations. In diagram a) densities for θ equal to $-10000, -2000, -1000, -1, 1, 1000, 1500$ have their peak magnitudes in the descending order. In diagram b) the detail of diagram a) is present for θ equal to $-1, 1, 1000, 1500$.

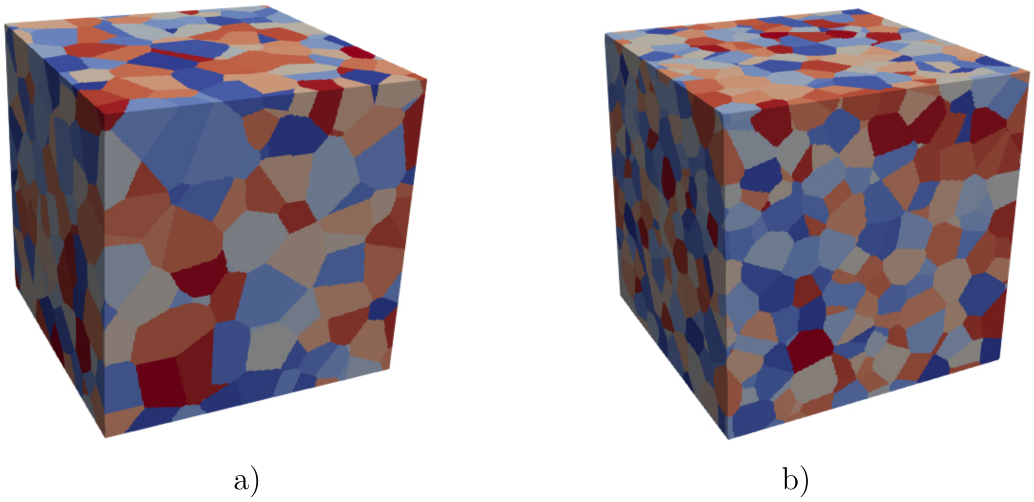


Figure A.7: The 3D simulations of the sphericity model for $\theta = -10000$, a), and $\theta = 1500$, b). The colouring is random.

the histogram discrepancy and number of non-empty Laguerre cells evolve during MCMC simulation, *Algorithm 2*, in the case of different values of parameter θ.

With an increasing value of the parameter θ the discrepancy decreases, i.e., the sphericity distribution of simulated tessellations is getting closer to the prescribed one. As θ decreases, the tessellations become quite different. Instead of moving horizontally, corresponding to less spherical cells occurring, the peak of the density curve increases, i.e., sphericities in the narrow range are favoured. This causes a dramatic increase of discrepancy which suddenly stops below the value 1.58. This behaviour is observable for $\theta = -5000$ and smaller. Narrowness of the density means that all cells in the tessellation have very similar sphericity and are in this sense rigid. The densities for values $-1, 1$ are quite similar. For $\theta = 1500$, the density is quite similar to the prescribed one and any further increase of the parameter value has only a negligible effect on the decay of the discrepancy. The discrepancies fluctuate more for smaller (in absolute value) values of θ. The fluctuations are considerably smaller for values ≤ -5000 and ≥ 1500. The same phenomenon is observable for the number of cells. Moreover the number of cells rapidly decreases with decreasing negative values of θ. 3D visualizations of the simulated realizations for $\theta = -10000$ and $\theta = 1500$ are available in Figure A.7 (note that the numbers of cells differ significantly).

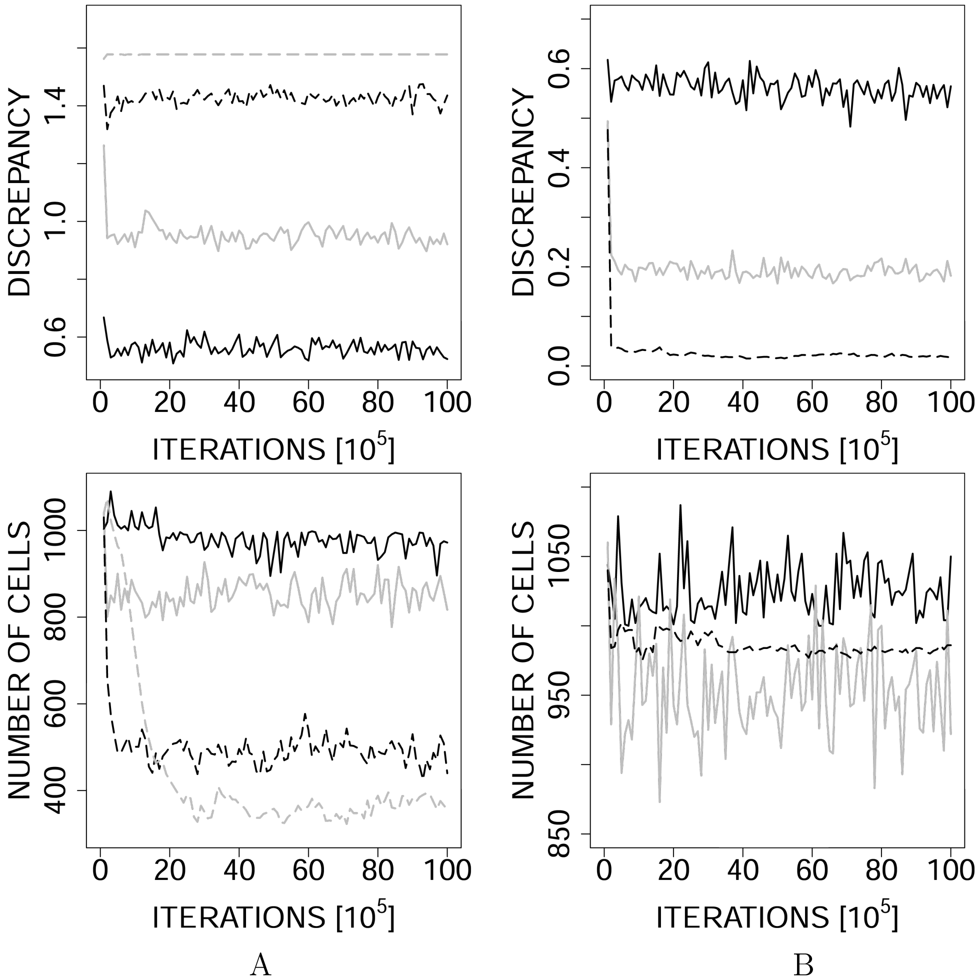


Figure A.8: Discrepancy between simulated and prescribed histogram (up) and number of cells (down) for varying values of the parameter θ in MCMC simulations of Gibbs Laguerre tessellations. In column A and B the negative and positive values of parameter θ are involved, respectively. Namely in column A, $\theta = -1$ corresponds to black full line, $\theta = -1000$ to grey full line, $\theta = -2000$ to black dashed line and $\theta = -10000$ to grey dashed line. In column B, $\theta = 1$ corresponds to black full line, $\theta = 1000$ to grey full line and $\theta = 1500$ to black dashed line. We draw each 10^5-th iteration.

A.3 Comparison of Gibbs and Poisson Laguerre tessellations

In contrast with Poisson type tessellations, the Gibbs point process allows for quick modification of various geometrical characteristics of the cells using the potentials introduced in Section 3.1. In the Poisson case, the theoretical values (depending on the intensity) for the first two moments of various characteristics can be determined, cf. Okabe et al. [2000] and Lautensack [2007]. A simulation study demonstrating that Gibbs-Laguerre tessellations outperform Poisson-Laguerre tessellations in terms of the variety of possible cell shapes and numerical values of cell characteristics is presented in this section.

Using the potentials introduced in Section 3.1 one can easily modify various geometric characteristics of the cells. Using Gibbs-type tessellations instead of Poisson-type ones one can obtain different cell shapes preserving the intensity of cells. This will be demonstrated on four Gibbs-Laguerre models:
a model based on a single reconstructing potential $V^G_{n,T}$, (3.5), given by

$$G = \mathrm{nof}, T(s_G) = \bar{s}_G, \quad s_0 = 20, \quad \theta_n = 1\,000; \tag{A.3}$$

a model based on a combination of two reconstructing potentials given by

$$\begin{aligned} &G_1 = \mathrm{nof}, T_1(s_{G_1}) = \bar{s}_{G_1}, \quad s_{0,1} = 16.5, \quad \theta^1_n = 10\,000, \\ &G_2 = \mathrm{nof}, T_2(s_{G_2}) = S^2_{s_{G_2}}, \quad s_{0,2} = 25, \quad \theta^2_n = 1\,000; \end{aligned} \tag{A.4}$$

and regular and irregular model from Section 3.7.2.

Figures A.9 and A.10 show how the realizations drawn from various Gibbs-Laguerre tessellation models differ from those of the Poisson-Laguerre tessellations (PLT) with the same intensity and distribution of radii (marked independently). Instead of histograms, we are plotting kernel density estimates of tessellation characteristics. The red line corresponds to the empirical density of one particular realization of the Gibbs-Laguerre tessellation and the black lines correspond to the empirical densities of PLTs. For every Gibbs-Laguerre tessellation only a single realization is compared with a bunch of twenty appropriate PLTs in order to show how it extends the PLT model. In Figure A.9 the tessellation obtained from the model (A.3) and the model (3.22)-irregular are compared with their Poisson-Laguerre counterparts. The model (A.3) results in tessellations with an average number of faces per cell equal to 20, the variance of the same characteristic is not controlled by the potential and is comparable with the value obtained for PLT. Medium size cells are favored. The model (3.22)-irregular creates tessellations differing from the PLT significantly in both the distribution of the number of faces per cell and the distribution of the cell volume. The difference is influenced by the presence of hardcore parameters, which in this case allow a smaller range for the cell sizes. In Fig. A.10 the tessellations obtained from the model (3.22)-regular and the model (A.4) are compared with their Poisson-Laguerre counterparts. As for the irregular tessellation, significant differences are visible even for the tessellations obtained from model (3.22)-regular. Here, the range of cell sizes given by hardcore limits is even smaller since the model favors neighbouring cells to have similar volumes. The model (A.4), controlling the

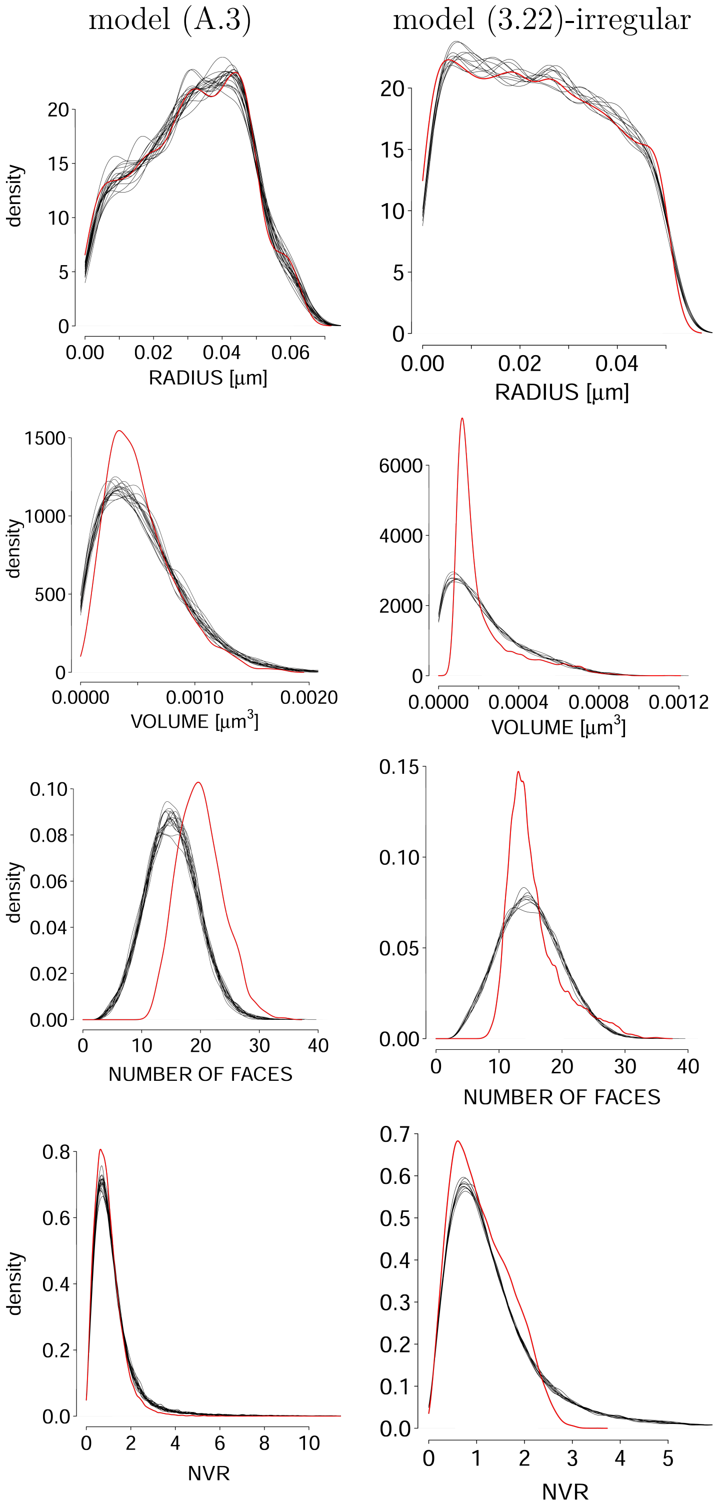
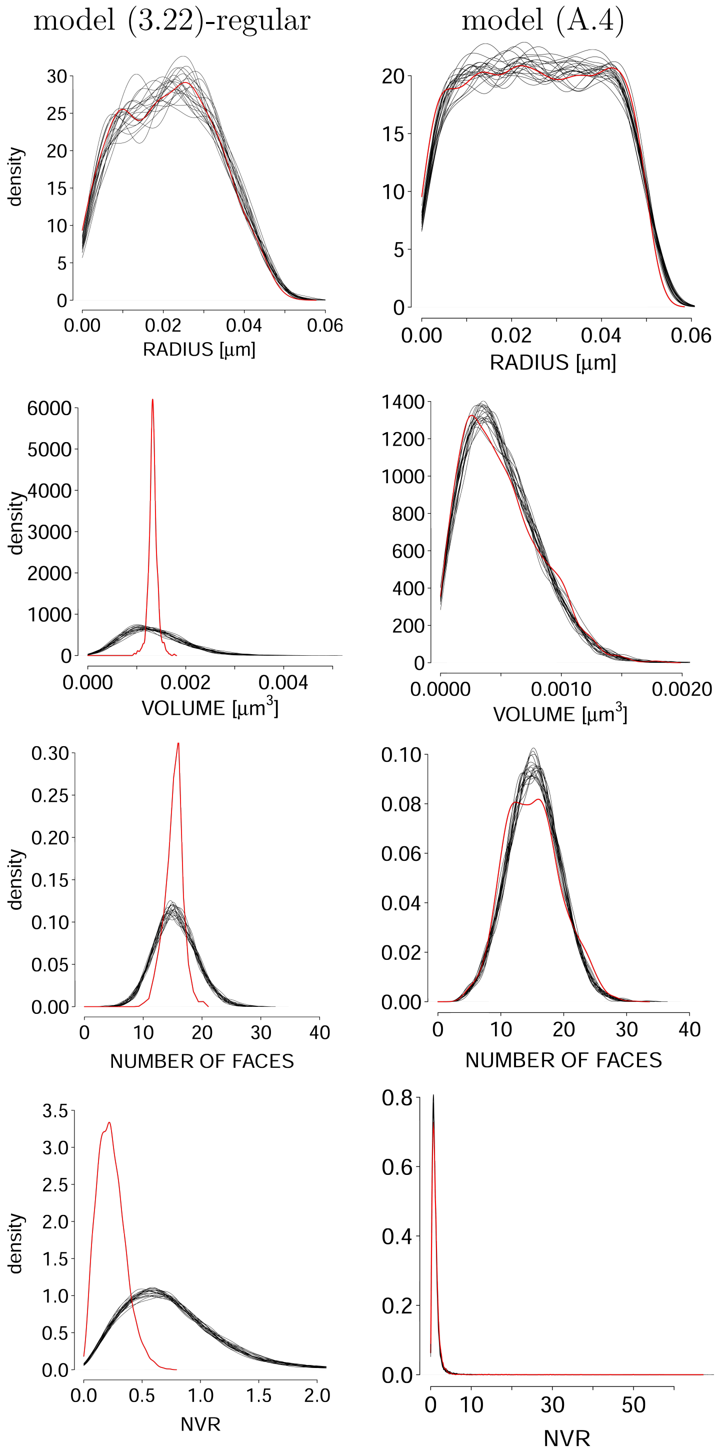


Figure A.9: Comparison of Gibbs-Laguerre tessellations obtained from model (A.3) and model (3.22)-irregular (red) with Poisson-Laguerre tessellations (black) – estimated distributions of Laguerre radii, cell volumes, number of faces per cell and NVR are presented.


model (3.22)-regular

model (A.4)

Figure A.10: Comparison of the Gibbs-Laguerre tessellations obtained from model model (3.22)-regular and model (A.4) (red) with Poisson-Laguerre tessellations (black) – estimated distributions of Laguerre radii, cell volumes, number of faces per cell and NVR are presented.

mean and variance of the number of faces per cell, performs similarly as PLT with respect to the distribution of cell volumes, but the distribution of the number of faces per cell has larger mean and variance than the PLT model.

A.4 Profile maximum pseudolikelihood for multiscale point process

Consider the class of multiscale point processes, Definition 33. Then the parameters β, γ and δ of $\mathcal{M}_d$, $d > 1$, are estimated using profile maximum pseudolikehood procedure sketched in Section 4.3.1. This method uses a grid of values of δ and estimates β and γ at each grid point. These estimates are shown in Tables A.2 and A.3 for models $\mathcal{M}_2$ and $\mathcal{M}_3$, respectively. Moreover, the tables show for each grid point the value *clPL* which is the shortcut for value of contrast log pseudolikelihood function evaluated in $(\hat{\beta}, \hat{\gamma})$ and increased by observation window volume, i.e., the value $-\log \mathcal{PL}(\hat{\beta}, \hat{\gamma}) + |W|$. The grid point and appropriate estimates corresponding to the largest value of contrast log pseudolikelihood function is highlighted by red color and is taken as the final estimate of the triplet (β, γ, δ).

Table A.2: $\mathcal{M}_2$: Values of contrast log pseudolikelihood function increased by observation window volume (additive constant) and the estimates of parameters β and γ_1 for the prescribed grid values of δ_1.

δ_1	cl$\mathcal{PL}$	$\hat{\gamma}_1$	$\hat{\beta}$
0.25	10310.3	$9.94287 \cdot 10^{-4}$	$1.4499 \cdot 10^{-2}$
0.5	10308.7	0.400712	$1.45513 \cdot 10^{-2}$
0.75	10304.9	0.51253	$1.46677 \cdot 10^{-2}$
1	10296.3	0.529527	$1.49071 \cdot 10^{-2}$
1.25	10282.8	0.539163	$1.53054 \cdot 10^{-2}$
1.5	10281.7	0.633133	$1.56259 \cdot 10^{-2}$
1.75	10288.3	0.734503	$1.58022 \cdot 10^{-2}$
2	10283.4	0.759278	$1.62963 \cdot 10^{-2}$
2.25	10281.9	0.790985	$1.67555 \cdot 10^{-2}$
2.5	10289.7	0.843616	$1.68163 \cdot 10^{-2}$
2.75	10298.9	0.894212	$1.65644 \cdot 10^{-2}$
3	10294.9	0.893557	$1.72605 \cdot 10^{-2}$
3.25	10299.2	0.917589	$1.72107 \cdot 10^{-2}$
3.5	10305.2	0.945237	$1.67104 \cdot 10^{-2}$
3.75	10305.7	0.95217	$1.69583 \cdot 10^{-2}$
4	10306.1	0.957812	$1.70712 \cdot 10^{-2}$
4.25	10306.9	0.963817	$1.71511 \cdot 10^{-2}$
4.5	10305.4	0.962259	$1.78561 \cdot 10^{-2}$
4.75	10305.1	0.964197	$1.82936 \cdot 10^{-2}$
5	10305.5	0.967718	$1.85147 \cdot 10^{-2}$
5.25	10304.7	0.968132	$1.91756 \cdot 10^{-2}$
5.5	10304.6	0.969892	$1.96449 \cdot 10^{-2}$
5.75	10304.9	0.972358	$1.99401 \cdot 10^{-2}$
6	10306.4	0.976875	$1.96253 \cdot 10^{-2}$
6.25	10310.3	0.987499	$1.74356 \cdot 10^{-2}$
6.5	10309.9	0.98706	$1.79756 \cdot 10^{-2}$

Table A.3: $\mathcal{M}_3$: Values of contrast log pseudolikelihood function increased by observation window volume (additive constant) and the estimates of parameters β, γ_1 and γ_2 for the prescribed grid values of δ_1 and δ_2.

$\delta_1 \backslash \delta_2$	0.75	1	1.25	1.5	1.75	2	2.25	2.5	2.75
0.5	10304.6 0.39923 0.560209 0.0146677	10296 0.396478 0.548518 0.0149072	10282.4 0.392742 0.549157 0.0153055	10280.9 0.391433 0.642366 0.0156263	10286.8 0.392011 0.74255 0.0158032	10281.7 0.39088 0.76498 0.0162979	10280 0.390715 0.795272 0.0167578	10287.4 0.392775 0.847105 0.0168191	10296.2 0.395129 0.897097 0.0165675
0.75	10304.9	10296.3 0.509239 0.544327 0.0149071	10282.7 0.505074 0.548545 0.0153055	10280.9 0.503034 0.651618 0.0156266	10286 0.503111 0.753983 0.0158044	10280.7 0.501668 0.7733 0.0162998	10278.7 0.50155 0.801791 0.0167605	10285.5 0.503405 0.8527 0.0168233	10293.7 0.506126 0.901913 0.0165725
1		10296.3	10282.7 0.525475 0.553512 0.0153054	10280 0.523171 0.679221 0.0156271	10283.3 0.522795 0.782457 0.0158064	10277.6 0.521038 0.792657 0.0163033	10275 0.520428 0.816343 0.0167659	10280.6 0.521966 0.864887 0.0168305	10287.6 0.524277 0.912222 0.016581
1.25			10282.8	10277.7 0.536761 0.765003 0.0156276	10277.8 0.53596 0.847201 0.0158087	10272.1 0.533815 0.830887 0.0163079	10268.7 0.532765 0.843202 0.0167731	10272.4 0.533775 0.886635 0.0168404	10277.5 0.535529 0.930246 0.0165928
1.5				10281.7	10280.7 0.631993 0.907179 0.0158074	10275.5 0.629527 0.852482 0.0163068	10271.9 0.628122 0.858111 0.016773	10274.7 0.628764 0.901252 0.0168431	10278.6 0.630272 0.943848 0.0165987
1.75					10288.3	10282.4 0.732009 0.813804 0.0162994	10279.1 0.730418 0.843722 0.0167648	10282.3 0.73079 0.901009 0.0168372	10285.9 0.732105 0.948844 0.0165966
2						10283.4	10279.7 0.757536 0.868466 0.0167624	10281.3 0.757525 0.931828 0.0168378	10283 0.758419 0.976758 0.0166023
2.25	legend: $\mathrm{cl}\mathcal{PL}$ $\hat{\gamma}_1$ $\hat{\gamma}_2$ $\hat{\beta}$						10281.9		

A.5 Software

For the development of the models introduced in this thesis we use various software either already existing or self implemented.

A.5.1 Overview of the existing software used for computations

Particular tasks are solved using already existing software, namely we use

- `C++` library `Voro++` [Rycroft, 2009] for computations of Laguerre tessellations,
- `R`-package `spatstat` [Baddeley et al., 2015] for computations of point process functional characteristics,
- `R`-package `GET` [Myllymäki and Mrkvička, 2019] to perform global envelope tests.

In more detail, the `C++` library `Voro++`, [Rycroft, 2009], is open source library which enables effective computations of a single Laguerre cell. This is an advantage for the estimation and MCMC simulation procedures we use, since we need to recompute only a small number of cells in each iteration (not the whole tessellation). The library computes the cell, its neighbours and its basic geometric characteristics. In case that window W is rectangular parallelepiped, it allows periodic boundary conditions.

For the calculation of estimates of functional summary statistics (specifically $\hat{F}(t)$, $\hat{G}(t)$, $\hat{L}(t)$ and $\hat{g}(t)$), we use the `R`-package `spatstat` [Baddeley et al., 2015]. This package was used for visualization of points as well.

Global envelopes and the corresponding p-values are computed using the `R`-package `GET`, [Myllymäki and Mrkvička, 2019].

A.5.2 Library `LagMod`

Besides the functionality of above mentioned libraries we implemented all other computations, most of them in `C++`.

`LagMod` is `C++` library which encompasses the computations connected with all models introduced in this thesis. In brief, we extended capabilities of `Voro++` library and implemented a code for pseudolikelihood estimation and the MCMC algorithms used for simulation of the (marked) point process models and the models for the radii given the points.

Most of data analysis was carried out in `R`. Moreover, we implemeted empirical mark correlation function and empirical mark variogram in `R`.

The code is available at `https://github.com/VigoFierry/Lag_mod` together with several examples of usage.