Accurate representation of the distributions of the 3D Poisson-Voronoi typical cell geometrical features

Vittorietti, Martina; Kok, Piet J.J.; Sietsma, Jilt; Jongbloed, Geurt

DOI
10.1016/j.commatsci.2019.04.054

Publication date
2019

Document Version
Accepted author manuscript

Published in
Computational Materials Science

Citation (APA)
Vittorietti, M., Kok, P. J. J., Sietsma, J., & Jongbloed, G. (2019). Accurate representation of the distributions of the 3D Poisson-Voronoi typical cell geometrical features. *Computational Materials Science*, *166*, 111-118. https://doi.org/10.1016/j.commatsci.2019.04.054

Accurate representation of the distributions of the 3D Poisson-Voronoi typical cell geometrical features

Martina Vittorietti*[1,2], Piet J.J. Kok[3], Jilt Sietsma[4], Geurt Jongbloed[1]

[1]Department of Applied Mathematics, Delft University of Technology, Mekelweg 4, 2628 CD Delft, The Netherlands

[2]Materials Innovation Institute (M2i), Mourik Broekmanweg 6, 2628 XE Delft, The Netherlands

[3]Tata Steel, IJmuiden Technology Centre, Postbus 10.000, 1970 CA, IJmuiden, The Netherlands

[4]Department of Materials Science and Engineering, Delft University of Technology, Mekelweg 2, 2628 CD Delft, The Netherlands

Abstract. Understanding the intricate and complex materials microstructure and how it is related to materials properties is an important problem in the Materials Science field. For a full comprehension of this relation, it is fundamental to be able to describe the main characteristics of the 3-dimensional microstructure. The most basic model used for approximating steel microstructure is the Poisson-Voronoi diagram. Poisson-Voronoi diagrams have interesting mathematical properties, and they are used as a good model for single-phase materials. In this paper we exploit the scaling property of the underlying Poisson process to derive the distribution of the main geometrical features of the grains for every value of the intensity parameter. Moreover, we use a sophisticated simulation program to construct a close Monte Carlo based approximation for the distributions of interest. Using this, we determine the closest approximating distributions within the mentioned frequently used parametric classes of distributions and conclude that these representations can be quite accurate. Finally we consider a 3D volume dataset and compare the real volume distribution to what is to be expected under the Poisson-Voronoi model.

1. Introduction

Investigating 3-dimensional structures is a fundamental aspect for many disciplines; especially for those related to materials study, but also for more abstract disciplines such as Mathematics and Statistics. For materials study, one of the most outstanding aims is to fully understand the intriguing relationship between microstructural features and mechanical properties of the materials. The very first step for achieving this objective is quantifying 3D microstructures. From Materials Science point of view this means examining and understanding the nature and variety of microstructures. From Statistics it means looking for the characterization of the 3D virtual microstructures under specific mathematical models.

In the past few years the use of Voronoi diagrams has rapidly increased. These diagrams represent an appealing structure, especially because they describe various natural processes quite well. In [1] an extensive list of fields in which Voronoi diagrams are adopted can be found. Among the many areas of application of this model, the field of materials science stands out. In fact, they are now among the most used mathematical models for microstructure characterization and depending on the specific kind of materials, it is possible to use a proper category of Voronoi diagrams.
In this paper, we discuss the most basic instance of the model: Poisson-Voronoi diagrams.

*Contact author: m.vittorietti@tudelft.nl

In this framework the nuclei or sites are generated by a homogeneous Poisson process with intensity parameter λ. The parameter λ represents the expected number of points (Voronoi cells) in a unit volume.

Although many interesting mathematical properties of Poisson-Voronoi diagrams are known, there is still much to discover about the distributions of the geometrical characteristics of its cells. Through simulations, many authors were able to obtain numerical approximations of the moments of the distribution of the volume, of the surface area, of the number of faces and many other geometrical characteristics of the grains. Nevertheless, analytic expressions of the distributions of many of these important features are not known, others are only obtained via complicated numerically intractable characterizations. Therefore, various proposals to obtain close approximations to the real distributions were put forward by several authors e.g. Lognormal- [14,15], Generalized Gamma- [7,17] and Rayleigh distributions [16]. But as far as we know, there is no theoretical support for preferring one of these distributions.

The aim of this paper is twofold. After explaining that λ, the intensity parameter of the Poisson process, is the only parameter determining all distributional properties of the geometrical structure of the grains, we show that if we have the distribution of a given geometrical characteristic for $\lambda = 1$, the distribution of the same quantity for every value of $\lambda > 0$ can be obtained by rescaling. More precisely, we consider volume, surface area and number of faces of the grains, but the approach can be extended to other characteristics. Secondly, we find a close Monte Carlo based approximation for the previously mentioned geometrical characteristics of the grains and using it we determine the most closely approximating distribution within the mentioned frequently used parametric classes of distributions. As said before, several well known probability distributions were used for approximating the grain geometrical characteristics distributions, but in this study we determine the most accurate of these.

After briefly reviewing the basic concepts of Voronoi diagrams in Section 2, in Section 3 we explain the scaling property of the Voronoi structure in terms of the intensity parameter and how it can be useful for studying distributional properties of the grain features. Section 4 describes our simulation approach and produces an accurate Monte Carlo approximation for the distribution of the grain volume and the grain surface area. In fact, we provide the approximate distributions of the grain volume and of the grain surface area for $\lambda = 1$ and we adapt it for the other values of λ using the aforementioned scaling properties. In Section 5, we study the approximation of the true distributions of the geometrical characteristics by some well-known and frequently used probability distributions in this context: the Gamma-, Generalized Gamma- and Lognormal distribution. Fitting these three distributions and comparing them through statistical measures, such as the supremum distance and the Total Variation distance between the Monte Carlo empirical distribution and its parametric approximations, we are not only able to identify the best approximation but also to give a measure of error if one of these parametric approximations is used.
In Section 6 we consider an experimental dataset of grain volumes and compare the distribution of these volumes to the distribution that would be expected in case of a Poisson-Voronoi generating process. Finally, we discuss the possibility to extend our approach according to different Voronoi Diagrams cases, such as Multi-level Voronoi and Laguerre Voronoi Diagrams. We want to remark that for the 3D Voronoi diagrams generation we use TATA Steel software and for data analysis the statistical software R.

2. Poisson-Voronoi Diagrams

We begin by reviewing the generic definition and the basic properties of the Poisson-Voronoi Diagram. Given a denumerable set of distinct points in $\mathbb{R}^d$, $\mathbf{X} = \{x_i : i \geq 1\}$, the Voronoi diagram of $\mathbb{R}^d$ with *nuclei* $\{x_i\}$ (also called *sites* or *generator points*) is a partition of $\mathbb{R}^d$ consisting of cells

$$C_i = \{y \in \mathbb{R}^d : \|x_i - y\| \leq \|x_j - y\| \; for \; j \neq i\}, \;\; i = 1, 2, \ldots$$

where $\|\cdot\|$ is the usual Euclidean distance. This means that given a set of two or more but finitely many distinct points, we associate all locations in that space with the closest member(s) of the point set with respect to the Euclidean distance [1].

If we assume that $\mathbf{X} = \Phi = \{x_i\}$ is the realization of a homogeneous Poisson point process, we will refer to the resulting structure as the *Poisson-Voronoi diagram*, $\mathcal{V}_\Phi$. In particular in this paper, we assume that the sites of the Poisson-Voronoi diagrams are generated according to a *uniform* or *homogeneous* Poisson process. This means that for any bounded set A in $\mathbb{R}$ with Lebesgue measure $|A|$ (length if $d = 1$, area if $d = 2$, volume if $d = 3$), the expected number of points falling in A, $\mu(A)$, is proportional to $|A|$. In formula

$$\mu(A) = \lambda |A|. \tag{2.1}$$

The parameter $\lambda > 0$ is referred to the intensity parameter of the Poisson process and for a homogeneous Poisson process is assumed to be constant.

For a formal definition of the Poisson process see [2].

As mentioned before, our aim is to find the distribution of the geometrical characteristics of the grains. In order to approximate these distributions, we generate a large sample of independent and identically distributed cells, more specifically *typical cells*. A typical Voronoi cell refers to a random polytope which loosely speaking has the same distribution as a randomly chosen cell from the diagram selected in such a way that every cell has the same chance of being sampled. Moreover, the distribution of the typical Poisson-Voronoi cell is by Slivnyak-Mecke formula [6] the same as the distribution of the Voronoi cell containing the origin, obtained when the origin is added to the point process Φ. This cell is formally given by

$$\mathcal{C} = \{y \in \mathbb{R}^d : \|y\| \leq \|y - x\| \; for \; all \; x \in \Phi\}.$$

Okabe et al. [1] synthesize previous research activity about the properties of Poisson Diagrams. Despite the fact that distributions of several geometrical characteristics are already known, the distributions of the main features, especially in 3D, are not. We describe a simulation approach to approximate these distributions in the next section.

3. Distribution of the geometrical properties of a typical cell

Given the complexity of finding explicit formulae for the distributions of the Poisson-Voronoi diagram geometrical characteristics, especially in 3D, many authors used Monte Carlo methods to approximate these. Among them Kiang [3], Kumar and Kurtz [4], Lorz and Hahn [5], Møller [6], Tanemura [7] obtained numerical results for the moments of the distribution of volume, surface area, and number of faces of the grains in 3D. They also give histogram estimates of these distributions and suggest approximations for them using various well known probability distributions. For instance, for the volume distribution, before 1990 most authors used the Lognormal distribution for approximating the grain size distribution in polycrystals. Nowadays, more flexible distributions such as Gamma or Generalized Gamma

are commonly used (e.g. [4,7]). Although these models fit the observed data rather well (as we will see in the next section) our approach allows to find an accurate representation of the true distribution and the parametric distribution that optimally fits the data.

The main idea is that, given a Poisson-Voronoi diagram generated by a Poisson point process Φ with intensity parameter λ, this λ is the only parameter determining the distributions of the geometrical features of the grains. Furthermore, the dependence of the distributions on the intensity parameter is via simple scaling of a 'parent distribution', due to the following important scaling property of the Poisson process.

Lemma 3.1 (Scaling Property)**.** *Let $\Phi = \{X_1, X_2, \dots\}$ be a Poisson process on $\mathbb{R}^d$ with intensity $\lambda = 1$. Choose $\lambda > 0$ and define $\Phi_\lambda = \{X_1/\lambda^{1/d}, X_2/\lambda^{1/d}, \dots\}$. Then Φ_λ is a Poisson process with intensity λ.*

The fact that Φ_λ is a Poisson process is a special instance of the 'Mapping theorem' [see 2, Section 2.3] for further details on the proof.

In the following subsections Lemma 3.1 will be used to study the dependence of the distributions of volume, surface area and number of faces of the grains on the intensity parameter λ.

3.1. **Grain Volume.** We first focus our attention on the grain volume distribution because of the direct relationship of this Poisson-Voronoi geometrical characteristic and the grain size distribution in microstructure characterization of materials.

Exploiting the properties of the Poisson process, the distribution for the normalized length of the Voronoi cell in 1D or size measure in 1D, can be shown to have density [8]

$$f_{1D}(y) = 4y \exp(-2y) \;\; 1_{[0,\infty)}(y)$$

In dimension $d > 1$, it was conjectured that the area (2D) and the volume (3D) of the typical cell in a Poisson-Voronoi diagram may be distributed as the sum of two and three gamma variables with shape and scale parameters equal to 2 [3], but [9] and [10] showed the conjecture to be false. In 2D an analytic, though computationally challenging result is provided by Calka [11], which gives an expression for the distribution of the area of the typical cell in 2D given the number of vertices. In 3D, as we know so far, no analytic expression for the volume distribution exists.

Lemma 3.2. *Denote by F_λ the distribution function of the volume (length if $d = 1$, area if $d = 2$) of the typical cell of the Poisson-Voronoi diagram based on a homogeneous Poisson process on $\mathbb{R}^d$ with intensity parameter $\lambda > 0$. Then, for all $x \geq 0$,*

$$F_\lambda(x) = F_1(\lambda x) \tag{3.1}$$

Proof. Let Φ be a homogeneous Poisson process on $\mathbb{R}^d$ with intensity 1. Denote by $\mathcal{C}$ the typical cell of the Voronoi diagram based on this process. Fix $\lambda > 0$ and consider the homogeneous Poisson process Φ_λ with intensity λ as introduced in the statement of **Lemma 3.1**. Then the typical cell in the Voronoi diagram based on Φ_λ is a scaled version of the typical cell of the Voronoi diagram based on Φ, in the sense that it is given by $\mathcal{C}_\lambda = \{x/\lambda^{1/d} : x \in \mathcal{C}\}$. This means that the volume V_λ of $\mathcal{C}_\lambda$ is exactly λ^{-1} times the volume V of $\mathcal{C}$. Therefore, for $x \geq 0$,

$$F_\lambda(x) = \mathrm{P}(V_\lambda \leq x) = \mathrm{P}\left(\frac{V}{\lambda} \leq x\right) = \mathrm{P}(V \leq \lambda x) = F_1(\lambda x)$$

□

3.2. **Grain Surface area.**

Lemma 3.3. *Denote by G_λ the distribution function of the surface area of the typical cell of the Poisson-Voronoi diagram based on a homogeneous Poisson process on $\mathbb{R}^3$ with intensity parameter $\lambda > 0$. Then, for all $x \geq 0$,*

$$G_\lambda(x) = G_1\left(\lambda^{\frac{2}{3}} x\right)$$

Proof. The argument follows the proof of **Lemma 3.2**. Denote by S_λ the surface area of $\mathcal{C}_\lambda$ and note that scaling of $\mathcal{C}_\lambda$ implies that S_λ is $\lambda^{-\frac{2}{3}}$ times the surface area of $\mathcal{C}$, S. Therefore

$$G_\lambda(x) = \mathrm{P}(S_\lambda \leq x) = \mathrm{P}\left(\frac{S}{\lambda^{\frac{2}{3}}} \leq x\right) = \mathrm{P}\left(S \leq \lambda^{\frac{2}{3}} x\right) = G_1\left(\lambda^{\frac{2}{3}} x\right)$$

□

3.3. **Number of grain faces.** Finally, another (discrete) property of interest regards the number of grain faces of the typical cell. It is clear that using either Φ or Φ_λ (from **Lemma 3.1**) as a basis for the Voronoi diagram, yields the same number of faces of the typical cell ($\mathcal{C}$ or $\mathcal{C}_\lambda$ respectively), leading to

Lemma 3.4. *Denote by N_λ the distribution function of the number of faces of the typical cell of the Poisson-Voronoi diagram based on a homogeneous Poisson process on $\mathbb{R}^d$ with intensity parameter $\lambda > 0$. Then, for all $x \geq 0$,*

$$N_\lambda(x) = N_1(x)$$

The same lemma holds for *number of corner points*, n_v. In fact, exploiting the Euler-Poincaré relation [1], it is possible to determine n_v when the number of faces is known.

4. Simulation Results

Now, we approximate the distribution function of the grain geometrical features, using the results obtained by a simulation based on 1 000 000 Voronoi diagrams. Our simulation approach also allows to determine which of the usual parametric models provides the best approximation to the true distribution and its deviation from it. We consider the volume, the surface area and the number of faces of a 3D Poisson Voronoi typical cell. Two possible approaches, well described in [1] are possible:

(1) generate a large number of points inside a bounded region B according to Φ, construct $\mathcal{V}_\Phi$ and measure the characteristics of all its cells.
(2) generate a sequence of independent typical Poisson Voronoi cells, measure the characteristics of each and then aggregate them to obtain the required distributions.

We follow the second approach. The reason for this choice derives from the convenience of having a sample of independent and identically distributed Voronoi cells such that we can quantify the agreement with the real distribution. Moreover, we are able to control and eliminate the boundary effect that is present because the structure is actually only constructed on a bounded region. For our objective it is important that only the distributions of the geometrical properties of the typical cell are needed, using $\lambda = 1$ in the simulations. By **Lemma 3.2, 3.3** and **3.4**, the distributions based on diagrams with different intensities can be obtained by scaling.

We conduct our simulation approach using the Voronoi software provided by TATA Steel. The implemented algorithm is based on the half plane intersection, which is closely related

to the original definition of a Voronoi tessellation [12] but a filter, which determines which neighboring points of a site are needed for the Voronoi cell construction of this site, is added for improving the computational speed, now comparable to the performance of the best algorithm [13].

A Monte-Carlo procedure is the adopted.
Repeat 1 000 000 times:

Step 1: : Generate a 3D Poisson-Voronoi diagram with added generator point $(0, 0, 0)$ with $\lambda = 1$;

Step 2: : Determine the geometrical characteristics of the realizations of the typical Voronoi cell, the cell that contains the point $(0, 0, 0)$, $\mathcal{C}(0)$;

Then, aggregate the 1 000 000 values.
The main graphical results are shown in Figures 1, 2 and 3.

The values of the estimated densities of the previously mentioned geometrical characteristics are given at http://dutiosb.twi.tudelft.nl/ martina.vittorietti/. An R-package is under construction. In Tables 1 and 2, we report the estimated moments of the main geometrical characteristics and the estimated probabilities for the number of faces. They are coherent with both the theoretical and numerical results obtained by other authors [4,7].

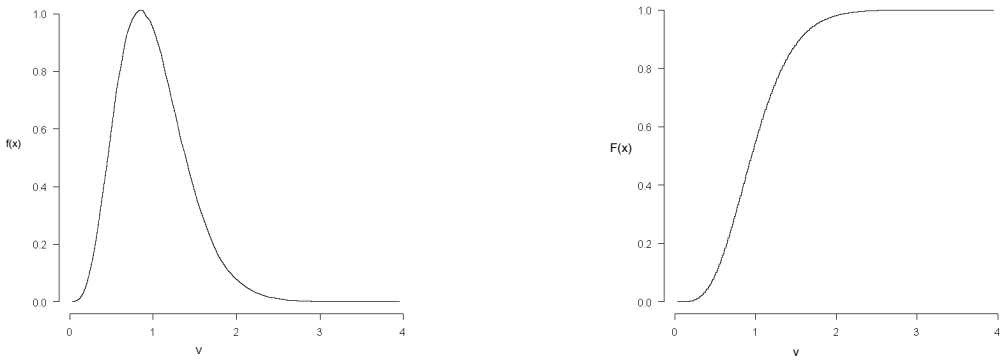
Table 1: Estimated moments of the geometrical features of 1 000 000 Poisson-Voronoi typical cells, λ=1

	(a) Volume		**(b)** Surface area		**(c)** Number of faces
μ_1	1.00008	μ_1	5.82670	μ_1	15.53071
σ	0.41189	σ	1.43821	σ	3.33896
μ_2	1.16981	μ_2	36.01888	μ_2	252.35173
μ_3	1.55900	μ_3	234.69091	μ_3	4277.80397
μ_4	2.32340	μ_4	1603.48468	μ_4	75464.60519

5. Parametric approximations to the distributions

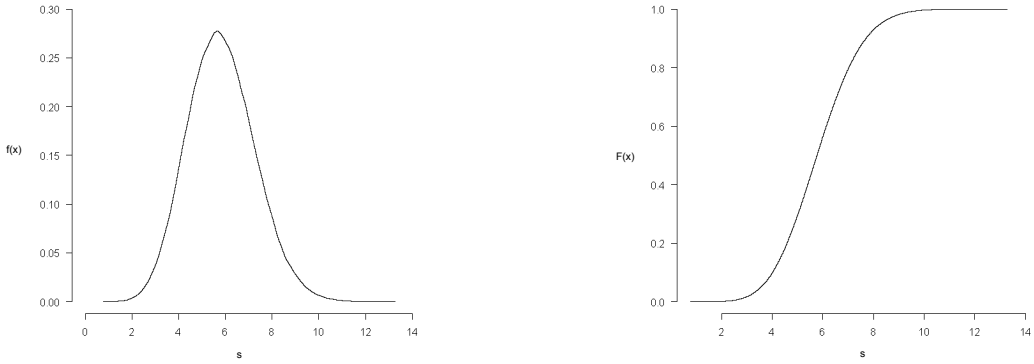
Various proposals to estimate the distributions of the geometrical properties of the typical cell were put forward by several authors such as the Lognormal distribution [14,15], Generalized Gamma distribution with 2 [17] or 3 parameters [7] and Rayleigh distribution [16]. Ferenc and Néda [10] propose their own function for the volume distribution. However, nobody was able to find an analytic expression for the grain geometrical characteristics distributions.

As noted in [17] the use of the Lognormal distribution function for representig grain size distribution lacks a solid physical basis and is not in general accurate. Nowadays, the debate regards mostly the Generalized Gamma Distribution with 2 or 3 parameters, but until now no physical explanation for using one preferential distribution exists. However, in view of the scaling properties described in the previous sections, it is natural to think that the distributions of the geometrical characteristics of the grain belong to a scale parametric family of



(a) **(b)**

Figure 1: (a) Kernel density estimate (Epanechnikov kernel, cross validation bandwidth $h = 0.05$) and (b) empirical cumulative distribution function of volume of 1 000 000 Poisson-Voronoi typical cells, $\lambda = 1$

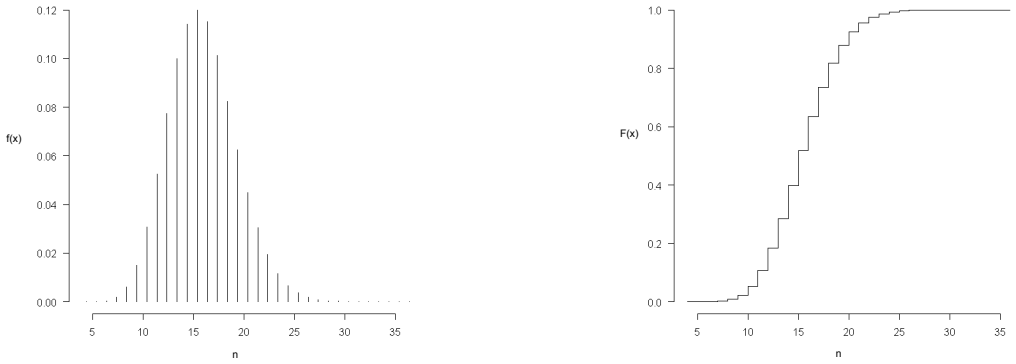


(a) **(b)**

Figure 2: (a) Kernel density estimate (Epanechnikov kernel, cross validation bandwidth $h = 0.25$) and (b) empirical cumulative distribution function of surface area of 1 000 000 Poisson-Voronoi typical cells, $\lambda = 1$

distributions. Only then the distributions of the quantities *for all* λ can belong to the class. One could, for instance, consider the Lognormal distribution. Its probability density is given by

$$f(x|\mu,\sigma) = \frac{1}{x\sigma\sqrt{2\pi}} \exp\left(-\frac{(\log(x)-\mu)^2}{2\sigma^2}\right).$$



(a) **(b)**

Figure 3: (a) Relative frequencies and (b) empirical cumulative distribution function of number of faces of 1 000 000 Poisson-Voronoi typical cells, $\lambda = 1$

Table 2: Distribution of the number of faces F of 1 000 000 Poisson-Voronoi typical cell, $\lambda = 1$

F	n_f	p_f	F	n_f	p_f	F	n_f	p_f
4	5	0.000005	16	115188	0.115188	28	435	0.000435
5	35	0.000035	17	101151	0.101151	29	224	0.000224
6	316	0.000316	18	82277	0.082277	30	95	0.000095
7	1822	0.001822	19	62408	0.062408	31	52	0.000052
8	6190	0.006190	20	44944	0.044944	32	18	0.000018
9	15051	0.015051	21	30477	0.030477	33	3	0.000003
10	30685	0.030685	22	19466	0.019466	34	1	0.000001
11	52528	0.052528	23	11682	0.011682	35	1	0.000001
12	77421	0.077421	24	6756	0.006756	36	1	0.000001
13	100094	0.100094	25	3631	0.003631			
14	114163	0.114163	26	1890	0.001890			
15	120015	0.120015	27	975	0.000975			

Let $\hat{\mu}_1$ and $\hat{\sigma}_1$ be the maximum likelihood estimates when $\lambda = 1$ (based on the 1,000,000 simulated values). Then, define a scale family based on that, $f_\lambda(x)$ as:

$$f_\lambda(x) = \lambda f(\lambda x|\hat{\mu}_1, \hat{\sigma}_1) = \frac{1}{x\hat{\sigma}_1\sqrt{2\pi}} \exp\left(-\frac{(\log(x) + \log(\lambda) - \hat{\mu}_1)^2}{2\hat{\sigma}_1^2}\right)$$

which corresponds to a Lognormal distribution with parameter vector $(\hat{\mu}_1 - log(\lambda), \hat{\sigma}_1^2)$. Therefore, we have a log-addition scaling on the first parameter, which is not consistent with the λ-scaling that is found for real distributions. Now let us consider the Generalized Gamma distribution. Its density function, parameterized according to [18], is given by

$$f(x|a, b, k) = \frac{bx^{bk-1}}{\Gamma(k)a^{bk}} e^{-\left(\frac{x}{a}\right)^b} \tag{5.1}$$

where a and b are the shape and the scale parameters, k the family parameter. Let $\hat{a}_1$, $\hat{b}_1$ and $\hat{k}_1$ be the maximum likelihood estimates for the parameters (based on the 1,000,000 simulated values) when $\lambda = 1$. Define $f_\lambda(x)$ as equal to:

$$f_\lambda(x) = \lambda f(\lambda x|\hat{a}_1, \hat{b}_1, \hat{k}_1) = \frac{\hat{b}_1 x^{\hat{b}_1\hat{k}_1 - 1}}{\Gamma(\hat{k}_1)} \left(\frac{\lambda}{\hat{a}_1}\right)^{\hat{b}_1\hat{k}_1} e^{-\left(\frac{\lambda}{\hat{a}_1}x\right)^{\hat{b}_1}}$$

which corresponds to a Generalized Gamma with parameters $(\frac{\hat{a}_1}{\lambda}, \hat{b}_1, \hat{k}_1)$. This suggests to look for a distribution that belongs to this scale family. Special cases of this family are the Gamma distribution with parameters a, k and $b = 1$ and the Weibull distribution with parameters a, b and $k = 1$. Beside the parametrization in *eq.* 5.1, another one is provided by Prentice [19]. This is in general more stable in the estimation of the parameters but both parametrizations lead to the same estimates. In the next subsections, we report the estimated parameters of the best Generalized Gamma approximations and we statistically compare the fits based on the Gamma distribution, the Generalized Gamma distribution and the Lognormal distribution. This comparison is based on two criteria:

- Supremum distance between two distribution functions:

$$D(F, G) = \sup_{x\in\mathbb{R}} |F(x) - G(x)| \tag{5.2}$$

  This distance is computed between the empirical distribution function F_n of the sample and the maximum likelihood fit within the respective parametric families. The data are the 1,000,000 simulated values from the distribution of interest, with $\lambda = 1$.
- Total Variation distance between distributions with distribution functions F and G and densities f and g respectively on $\mathbb{R}$:

$$TV(F, G) := \sup_{A\in\mathcal{B}} |\int_A dF(x) - \int_A dG(x)| = \frac{1}{2}\int |f(x) - g(x)|\, dx \tag{5.3}$$

  This distance is computed between the kernel estimate of the densities and maximum likelihood parametric fits based on the simulation results with $\lambda = 1$.

Note that information on these distances is based on the data obtained with $\lambda = 1$. If the estimates for more general values of λ are obtained via the rescaling, these distances do, however, not change under this scaling, so the distances also hold for the other values of λ.

Table 3: Estimated Generalized Gamma parameters for volume distribution approximation, λ=1

	$\hat{a}$	$\hat{b}$	$\hat{k}$
Estimate	0.380	1.287	3.583
Std. Error	0.005	0.006	0.0322

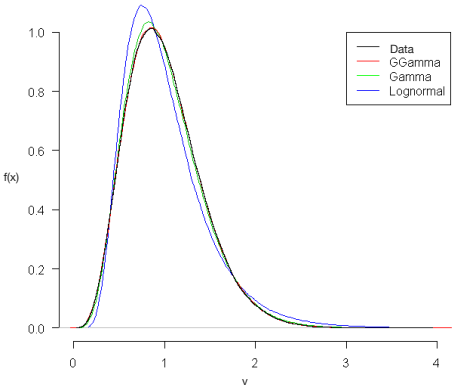
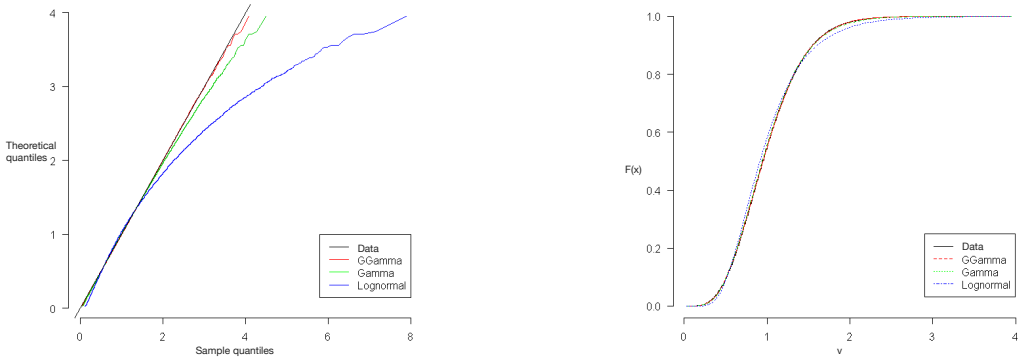


Figure 4: Comparison of parametric approximations to the volume distribution of 1 000 000 Poisson-Voronoi typical cells, $\lambda = 1$



(a) **(b)**

Figure 5: (a) QQplot and (b) cumulative distribution function comparison of parametric approximations to the volume distribution of 1 000 000 Poisson-Voronoi typical cells, $\lambda = 1$

Table 4: Comparison of Gamma-, Generalized Gamma- and Lognormal approximations for volume distribution in terms of Supremum- and Total Variation distance

	Gamma	Generalized Gamma	Lognormal
Supremum distance	0.013	0.005	0.041
TV distance	0.018	0.005	0.089

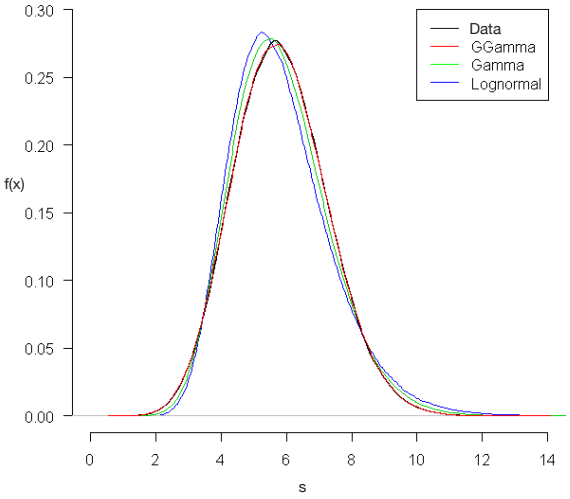


Figure 6: Comparison of parametric approximations to the surface area distribution of 1 000 000 Poisson-Voronoi typical cells, $\lambda = 1$

Table 5: Estimated Generalized Gamma parameters for surface area distribution approximation

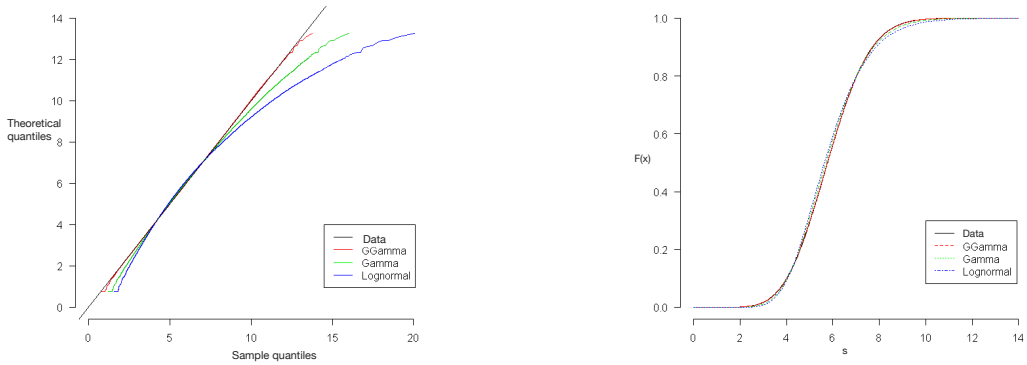
	$\hat{a}$	$\hat{b}$	$\hat{k}$
Estimate	3.174	2.102	3.839
Std. Error	0.025	0.011	0.036

Table 6: Comparison of Gamma-, Generalized Gamma- and Lognormal approximations for surface area distribution in terms of Supremum- and Total Variation distance

	Gamma	Generalized Gamma	Lognormal
Supremum distance	0.020	0.002	0.037
TV distance	0.035	0.003	0.082

6. Application

Data used for this application are kindly provided by Erik Offerman (TU Delft). They are relative to 882 grains in an actual steel microstructure in a volume of 950 μm $\times$ 1100 μm $\times$ 400 μm. The main information collected from the processed experimental data set is:



(a) **(b)**

Figure 7: (a) QQplot and (b) cumulative distribution function comparison of parametric approximations to the surface area of 1 000 000 Poisson-Voronoi typical cells, $\lambda = 1$

- center of mass of each grain;
- equivalent spherical radius;
- orientation as Miller indices.

More details about the 3D microstructure and the 3DXRD measurement can be found in [22,23].

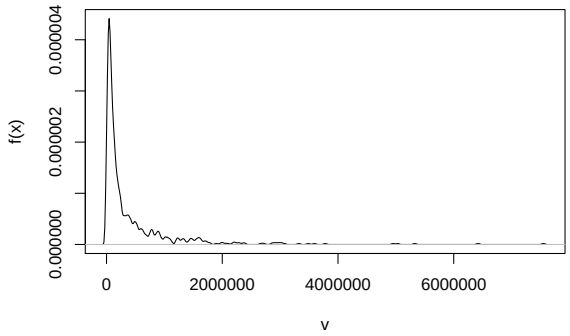


Figure 8: Kernel density estimate (Epanechnikov kernel, cross validation bandwidth $h = 21735.28$)

First, an estimate of λ is required for fitting. The most commonly used estimator is given by the ratio between the number of grains and the size of the domain [25]; in the specific case $\hat{\lambda} = 2.11 \times 10^{-6} \mu\text{m}^{-3}$. Another option is to use a moment estimator for λ, $\hat{\lambda} = \frac{\mu_1}{\bar{V}}$, where $\bar{V}$ is the empirical mean volume and $\mu_1 = \frac{a\Gamma(k+\frac{1}{b})}{\Gamma(k)}$ is known for the best

Generalized Gamma distribution approximating the volume distribution (Tab. 3). This leads to $\hat{\lambda} = 2.64 \times 10^{-6} \mu\text{m}^{-3}$.

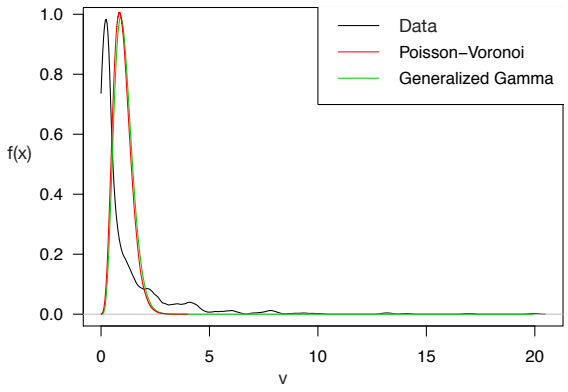
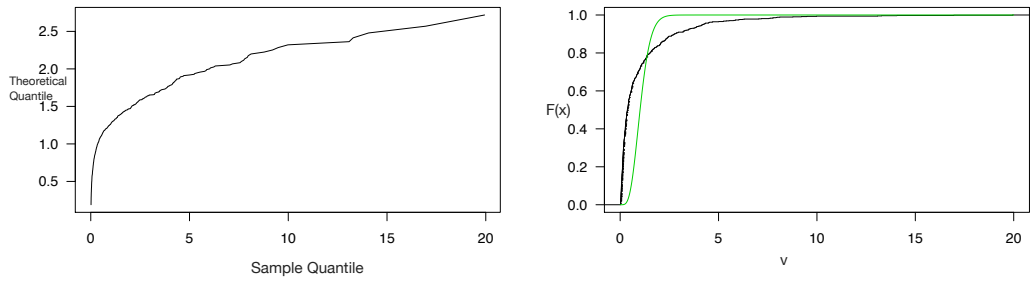


Figure 9: Comparison of parametric approximations to the volume distribution density(Epanechnikov kernel, cross validation bandwidth $h = 0.057$, black line), $\lambda = 1$

In figures 9, 10(a), 10(b) a graphical comparison in terms of probability density function, qq-plot and cumulative distribution function, respectively is performed. At first sight the result discourages the use of the Poisson-Voronoi model for this dataset. One major difference is related to the presence of many more small cells in the real microstructure than one would expect in the Poisson-Voronoi setting. One possible explanation could be the measurement inaccuracy of the smallest cells: it is, in fact, a very well known problem in the literature [24]. Moreover, volume values close to 0 may be attributed also to the non space filling reconstruction even in the center regions, as reported by data providers. This over-representation of small cells will also lead to an inaccurate estimate of λ and have a global impact on the comparison of both distributions. It is interesting to see whether the distributions in the higher volume regions fit better. One way to look at this is by conditioning on the (e.g. 10%) highest observed volumes. This conditional distribution then has to be compared to an appropriate Generalized Gamma distribution. We use a (conditional) method of moments estimate for λ to rescale the data on the same level as was done in Figure 10(b). This leads to Fig. 11. In



(a) **(b)**

Figure 10: (a) QQplot and (b) cumulative distribution function comparison of Generalized Gamma distribution to the real volume distribution, $\lambda = 1$

this graph we still see a difference between the two distribution functions as it comes to the thickness of the tails. However, the general shapes of both conditional distributions are quite similar.

All in all, we would not recommend the use of the Poisson-Voronoi model for the data at hand because of some complications (as the overrepresentation of small volumes, maybe due to premature stopping of the generating process) that could be addressed by using other models, such as Weighted Voronoi Diagrams. On the other hand, if especially the distributional behavior of the highest volumes is of interest and there are indications of inaccurate measurements of the smallest grains, the Poisson-Voronoi model may be used as a reasonable approximation and as a way of extrapolating distributional properties for cells with volume values close to 0.

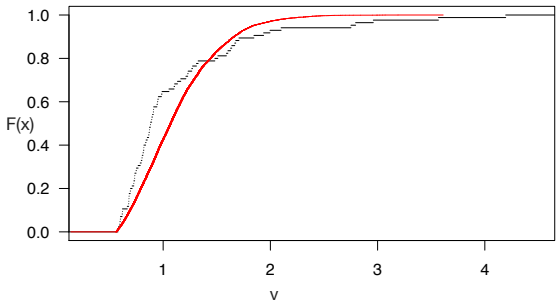


Figure 11: Cumulative distribution function comparison of Generalized Gamma distribution to the highest values of volume distribution $\lambda = 1$

7. Conclusions

This work provides a very accurate representation for the distributions of the main geometrical characteristics of the Poisson-Voronoi typical cell. It is now possible to exploit our approximated distribution for generating observations of approximately every geometrical characteristic that defines the grain size and for every possible value of λ. Moreover, we show that the Generalized Gamma distribution, with parameter $\hat{a} = 0.380$, $\hat{b} = 1.287$ $\hat{k} = 3.583$ for the volume and $\hat{b} = 3.174$, $\hat{a} = 2.102$, $\hat{k} = 3.839$ (Tab. 3 and 5) for the surface area is the best approximation in the class of the commonly used parametric distributions for grain size distributions. Nevertheless, it is not the true underlying distribution. In fact, the interpretation of the total variation distance as a measure of quality allows to say that using Generalized Gamma approximation for the grain volume distribution we could commit an error of about 0.5% and about 0.3% for the grain surface area (Tab. 4 and 6).

As noted in the introduction, Poisson-Voronoi diagrams are interesting and widely applied, but for modeling microstructures only constitute the most basic case. Their use in microstructure characterization is not fully evaluated yet. Therefore, we want to test their applicability and then extend our work to more general and less understood Voronoi structures, such as

Multi-level Voronoi diagrams [20] or Laguerre-Voronoi tessellations [21] in which the convenient scaling properties present in the Poisson-Voronoi diagrams are less obvious.

Acknowledgements

This research was carried out under project number S41.5.14547b in the framework of the Partnership Program of the Materials innovation institute M2i (www.m2i.nl) and (partly) financed by the Netherlands Organisation for Scientific Research (NWO). We also thank Erik Offerman, Gözde Dere and Markus Apel for permission to use the experimental data (processed and unprocessed) on the steel microstructure.

References

[1] Okabe A., Boots B., Sugihara K., Chiu S.N. Spatial tessellations: concepts and applications of Voronoi diagrams. John Wiley & Sons; 2009; vol. 501.
[2] Kingman J.F.C. Poisson processes. Wiley Online Library; 1993.
[3] Kiang T. Random fragmentation in two and three dimensions. Zeitschrift fur Astrophysik; 1966; vol.64; p. 433.
[4] Kumar S., Kurtz S.K. Monte-Carlo study of angular and edge length distributions in a three-dimensional Poisson-Voronoi tesselation. Elsevier; Materials characterization; 1995; vol.34; p. 15–27.
[5] Lorz U., Hahn U. Geometric characteristics of random spatial Voronoi tessellations and planar sections. Dekan des Fachbereiches Mathematik; 1993.
[6] Møller J. Lectures on random Voronoi tessellations. Ney York: Notes Statist. Springer; 1994; vol. 87.
[7] Tanemura M. Statistical distributions of Poisson Voronoi cells in two and three dimensions. SCIPRESS: FORMA-TOKYO; 2003; vol. 18; p. 221–247.
[8] Meijering J.L. Interface area, edge length, and number of vertices in crystal aggregates with random nucleation. Philips Res. Rep; 1953; vol. 8, p. 270–290.
[9] Zaninetti L. Poissonian and non-Poissonian Voronoi diagrams with application to the aggregation of molecules. Physics Letters A; 2009; vol. 373; p. 3223–3229.
[10] Ferenc J.S. Néda Z. On the size distribution of Poisson Voronoi cells. Elsevier; Physica A: Statistical Mechanics and its Applications; 2007; vol. 385; p. 518–526.
[11] Calka P. et al. Precise formulae for the distributions of the principal geometric characteristics of the typical cells of a two-dimensional Poisson-Voronoi tessellation and a Poisson line process. Advances in Applied Probability; 2003. p. 551–562
[12] O'Rourke J. Computational geometry in C. Cambridge university press; 1998.
[13] Fortune S. A sweepline algorithm for Voronoi diagrams. ACM: Proceedings of the second annual symposium on Computational geometry; 1986; p. 313–322.
[14] Smith J.E., Jordan M.L. Mathematical and graphical interpretation of the log-normal law for particle size distribution analysis. Elsevier; Journal of Colloid Science; 1964; vol. 19; p. 549–559.
[15] Söderlund J., Kiss L.B., Niklasson G.A., Granqvist C.G. Lognormal size distributions in particle growth processes without coagulation. APS; Physical review letters; 1998; vol. 80; p. 2386.
[16] Pande C.S. On a stochastic theory of grain growth. Elsevier; Acta Metallurgica; 1987; vol. 35; p. 2671–2678.
[17] Vaz M.F, Fortes M.A. Grain size distribution: The lognormal and the gamma distribution functions. Elsevier; Scripta metallurgica; 1988; vol. 22; p. 35–40.
[18] Stacy E.W., Mihram G.A. Parameter estimation for a generalized gamma distribution. Technometrics; 1965; p. 349–358.
[19] Prentice R.L. A log gamma model and its maximum likelihood estimation. Biometrika; 1974; p. 539–544.

[20] Kok P.J.J., Spanjer W., Vegter H. A microstructure based model for the mechanical behaviour of multiphase steels. Key Engineering Materials; 2015; vols. 651-653; p. 975–980.
[21] Lautensack C., Zuyev S. Random Laguerre Tessellations. Advances in Applied Probability; 2008; vol. 40; no. 3; p. 630–650.
[22] Sharma H., Huizenga R. M., & Offerman S. E. A fast methodology to determine the characteristics of thousands of grains using three dimensional X-ray diffraction. I. Overlapping diffraction peaks and parameters of the experimental setup. Journal of applied crystallography; 2012; vol. 45(4); p. 693-704.
[23] Sharma H., Huizenga R. M., & Offerman S. E. A fast methodology to determine the characteristics of thousands of grains using three dimensional X-ray diffraction. II. Volume, centre-of-mass position, crystallographic orientation and strain state of grains. Journal of Applied Crystallography; 2012; vol. 45(4); p. 705-718.
[24] Dere, E. G. Microstructure Control of Fire-resistant, Low-alloy Steel; An in-situ 3D X-ray Diffraction and A Small-angle X-ray Scattering Study. Ph.D thesis, Delft University of Technology. 2013
[25] Ohser J., & Mücklich F. Statistical analysis of microstructures in materials science. Wiley. 2000