General framework for testing Poisson-Voronoi assumption for real microstructures

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

DOI
10.1002/asmb.2517

Publication date
2020

Document Version
Final published version

Published in
Applied Stochastic Models in Business and Industry

Received: 25 March 2019 | Revised: 9 January 2020 | Accepted: 14 January 2020

DOI: 10.1002/asmb.2517

RESEARCH ARTICLE

WILEY

General framework for testing Poisson-Voronoi assumption for real microstructures

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

[1]Department of Applied Mathematics, Delft University of Technology, XE Delft, the Netherlands

[2]Materials Innovation Institute (M2i), XE Delft, the Netherlands

[3]Tata Steel, IJmuiden Technology Centre, CA IJmuiden, the Netherlands

[4]Department of Materials Science and Engineering, Delft University of Technology, CD Delft, the Netherlands

Correspondence
Martina Vittorietti, Department of Applied Mathematics, Delft University of Technology, Mourik Broekmanweg 6, XE Delft 2628, the Netherlands.
Email: m.vittorietti@tudelft.nl

Funding information
Materials innovation institute, S41.5.14547b; Nederlandse Organisatie voor Wetenschappelijk Onderzoek

Abstract

Modeling microstructures is an interesting problem not just in materials science, but also in mathematics and statistics. The most basic model for steel microstructure is the Poisson-Voronoi diagram. It has mathematically attractive properties and it has been used in the approximation of single-phase steel microstructures. The aim of this article is to develop methods that can be used to test whether a real steel microstructure can be approximated by such a model. Therefore, a general framework for testing the Poisson-Voronoi assumption based on images of two-dimension sections of real metals is set out. Following two different approaches, according to the use or not of periodic boundary conditions, three different model tests are proposed. The first two are based on the coefficient of variation and the cumulative distribution function of the cells area. The third exploits tools from to topological data analysis, such as persistence landscapes.

KEYWORDS

cumulative distribution function, hypothesis testing, persistence landscape, Poisson-Voronoi diagrams, real microstructures, scaling

1 | INTRODUCTION

The problem of quantifying complicated and fascinating microstructures of materials like metals has been around for many years.[1] It is an important issue in materials science because modeling 3D microstructures and relating these models to specific properties of the metals can give rise to new kinds of metals with desired performance. Indeed, having a good model for the microstructure, simulations can be performed to generate "digital versions" of the microstructure and testing its properties, for instance, mechanical properties, using yet other models that establish the relation between microstructural and mechanical properties. These simulations, approximating reality, allow the researcher to test material at relatively low cost and relatively fast, compared with real physical experiments. It is clear that an important and challenging statistical question to be answered is whether a specific model for a microstructure is adequate, given measured data.

In the tentative answering of this last question, several points need to be touched upon. The first point concerns the choice of a model. There exists a vast choice of models and among them, Voronoi diagrams have been extensively studied and used.[2] In particular, Poisson-Voronoi diagrams, only involving one nonnegative intensity parameter λ, represent the most basic case for modeling microstructures. In fact, they are often used in applications involving single-phase steel.[2-4] More sophisticated models have been proposed, but in this article we will concentrate on the Poisson-Voronoi model.

A second point concerns the available data. While the microstructure of a material is the arrangement of grains and phases in a three-dimensional (3D) space, the material is typically observed in two dimensions (2D). Usually, a small sample from inside the material is obtained and the exposed surface is examined in a microscope. Therefore, the work involves the study of 2D sections, from which 3D microstructure information has to be extracted.

Under the 3D Poisson-Voronoi model, the observable 2D section is a realization of a so-called *2D sectional Poisson-Voronoi diagram*, often denoted by $\mathcal{V}_\Phi(2,3)$. It is the result of the intersection of a fixed plane and a 3D Poisson-Voronoi diagram. Only limited results about the geometrical characteristics of its grains have been obtained analytically, but for most of them numerical results have been obtained through Monte Carlo simulations.[5] If a Poisson-Voronoi diagram is a good model, using 2D sections for the estimate of the intensity parameter λ, it is possible to infer distributions of almost all 3D microstructural properties, such as grain volume, grain surface area and grain number of faces.[6]

The last point is about the model validation. The question that this article wants to answer is "Given a real 2D materials section, could a Poisson-Voronoi diagram be a good model for approximating the 3D materials microstructure?" We propose several tests for the Poisson-Voronoi hypothesis. These are all based on contrasts between features of the observed 2D picture and the features one would expect if the data were generated according to the Poisson-Voronoi model.

The article structure is as follows. After having reviewed the basic concepts of Voronoi diagrams (Section 2), we recall the main stereological relations, which can be used to estimate λ based on a 2D sectional Poisson-Voronoi diagram and the most used intensity estimators introduced in Reference 7 (Section 3).

Then, we move to the testing framework (Section 4). We distinguish periodic and nonperiodic boundary conditions. The former case is very popular in material science practice and it allows to approximate "infinite structures," giving nice scaling properties and avoiding so-called "edge effects." The latter more closely resembles real situations. Assuming periodic boundary conditions, in Section 4.1 the distributions of the main geometrical characteristics of the 2D sectional cells are numerically obtained and two model tests are proposed. The first, already introduced in Reference 8, is based on the coefficient of variation of the cell (or grain) areas; the second is a Kolmogorov-Smirnov type test based on the cumulative distribution function (CDF) of the cell areas. In Section 4.2 the two tests previously mentioned are adapted to the nonperiodic boundaries setting. An additional test is defined, using tools from the emergent area of topological data analysis (TDA), that combines the two disciplines of statistics and topology. The focus is on *persistent homology*, the branch of TDA that summarizes the 2D picture using various functions. After having briefly and intuitively explained the basic concepts of persistent homology and the common ways of representing it (*persistence diagram*), a test based on the squared distances between *persistence landscapes* is presented (Section 4.2).

In Section 5, we carry out a computer simulation for estimating the quantiles for the proposed model test statistics. We consider null distributions for the test statistics conditional on the number of visible cells in 2D. For a general test statistic, the conditional distribution is expressed in terms of quantities that involve the (unknown) intensity parameter λ of the 3D Poisson process and quantities independent of λ. Therefore, a bootstrap approach for computing a 90% confidence interval for λ is proposed.

In Section 6, the power of the tests is discussed with respect to a specific parametric alternative hypothesis: 2D Poisson-Voronoi diagram. The new tests proposed in Section 4 result to be more powerful than the one already proposed in literature. Finally, in Section 7, we show an application of our work based on scanned images by Hahn and Lorz[8] of alumina ceramics. The different tests belonging to the different approaches are performed and the results are compared. A brief discussion on future developments follows in Section 8.

2 | VORONOI DIAGRAMS

We begin reviewing the generic definition and the basic properties of the 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

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

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

TABLE 1 Estimated moments of the geometrical features of 36 480 600 two-dimension sectional cells, λ=1

(a) Area		(b) Perimeter		(c) Number of Edges	
μ_1	0.68524	μ_1	3.13345	μ_1	6.00000
σ	0.47342	σ	1.60552	σ	1.69195
μ_2	0.69367	μ_2	12.39622	μ_2	38.86268
μ_3	30.37169	μ_3	2072.73503	μ_3	9818.30810
μ_4	40.94590	μ_4	10 695.17596	μ_4	72 107.17324

If $\mathbf{X} = \Phi = \{x_i\}$ is the realization of a homogeneous Poisson point process, then we will refer to the resulting structure as the *Poisson-Voronoi diagram* and denote it by $\mathcal{V}_\Phi$. This model is characterized by one single intensity parameter λ, the mean number of points generated according to the Poisson point process per unit volume.

Okabe et al[2] synthesize previous research activity on the properties of Poisson-Voronoi diagrams. Despite the fact that the moments of several geometrical characteristics are known, the distributions of the main features, such as grain volume, grains surface area, and grain number of faces, especially in 3D, are not. In Reference 6 a simulation study is conducted for finding accurate approximations for these distributions. A generalized gamma distribution is found to be the best approximating distribution among the well-known parametric densities frequently used in this framework. Exploiting the scaling property of the Poisson process, one obtains the distribution of the main geometrical characteristics for every value of the intensity parameter λ. In real experiments, it is often not possible to deal directly with 3D structures. Instead, one has to base inference on pictures of 2D sections of the 3D structure. In Reference 9, Chiu et al answer a fundamental question: "For integers $2 \leq t \leq d-1$, is the intersection between an arbitrary, but fixed t-dimensional linear affine subspace of $\mathbb{R}^d$ and the d-dimensional Voronoi tessellation generated by a point process Φ a t-dimensional Voronoi tessellation?" The answer to this question is negative when Φ is a Poisson point process.[9,10] Moreover, each cell in a sectional Poisson-Voronoi tessellation is almost surely a *non-Voronoi* cell.[9] For 2D and 3D Poisson-Voronoi diagrams, also for 2D sectional Poisson-Voronoi diagrams, much information about moments and scaling for the main geometrical characteristics is known, but little information and no analytic expressions for their distributions of them are available so far. In this article, we focus on the distribution of the area, the perimeter, and the number of edges of cells in 2D sectional Voronoi diagrams. The major results are summarized in Table 1.

In the following section, we will see how stereological relations can be used to obtain estimates of the intensity parameter λ of the 3D generating Poisson process based on the 2D sections.

3 | Stereological estimators for the intensity parameter λ

Basic stereological relationships exist, which are independent of any underlying tessellation model. Moreover, in the literature explicit (scaling) relations are known expressing the expected number of vertices per unit area, P_A, the expected number of cells per unit area, N_A, and the mean total edge length per unit area, L_A, in terms of the intensity parameter λ for a generating 3D Poisson process. Combining stereological and scaling relationships, the following expressions hold.[7]

$$
\begin{aligned}
P_A &= \frac{8}{15} \cdot \left(\frac{3}{4}\right)^{1/3} \cdot \pi^{5/3} \Gamma\left(\frac{4}{3}\right) \cdot \lambda^{2/3} = c_1 \cdot \lambda^{2/3} \\
N_A &= \frac{4}{15} \cdot \left(\frac{3}{4}\right)^{1/3} \cdot \pi^{5/3} \Gamma\left(\frac{4}{3}\right) \cdot \lambda^{2/3} = \frac{c_1}{2} \cdot \lambda^{2/3}, \text{and} \\
L_A &= \pi \cdot \left(\frac{\pi}{6}\right)^{1/3} \cdot \Gamma\left(\frac{5}{3}\right) \cdot \lambda^{1/3} = c_2 \cdot \lambda^{1/3}.
\end{aligned}
$$

Furthermore, exploiting the simple relation between N_A and the expected area of the cell profiles, $\mathbb{E}(a)$, $N_A = \frac{1}{\mathbb{E}(a)}$, four estimators for λ can be obtained:

$$
\begin{aligned}
\hat{\lambda}_P &= \left(\frac{\hat{P}_A}{c_1}\right)^{3/2} \approx 0.2008 \cdot \hat{P}_A^{3/2}, \hat{\lambda}_N = \left(\frac{2\hat{N}_A}{c_1}\right)^{3/2} \approx 0.5680 \cdot \hat{N}_A^{3/2} \\
\hat{\lambda}_L &= \left(\frac{\hat{L}_A}{c_2}\right)^{3} \approx 0.0837 \cdot \hat{L}_A^{3}, \hat{\lambda}_a = \left(\frac{2}{c_1 \bar{a}}\right)^{3/2} \approx 0.5680 \cdot \bar{a}^{-3/2}.
\end{aligned} \tag{1}
$$

Here, the hats indicate natural estimates for the mean quantities based on the data (like "number of cells divided by observed area," $\hat{N}_A$). In Reference 7, the behavior of the estimators is investigated by means of a computer simulation. The authors state that the estimators show hardly any difference concerning bias and variance and that the biases are less than 1% for sample size $n = 50$ and that they decrease rapidly with increasing sample size.

Once we have an estimate of the intensity parameter $\hat{\lambda}$, it can be used for estimating the distribution of the main geometrical 3D features of the grains.[6] An additional important issue of interest is whether the Poisson-Voronoi diagrams assumption is suitable in view of the observed 2D picture. We will consider this problem in the following section.

4 | MODEL TESTS FOR VALIDITY OF THE POISSON-VORONOI ASSUMPTION

In References 4,8,11 several model tests based on the distribution of geometrical features of the grains in random plane sections of a spatial tessellation are proposed. More precisely in Reference 4, the authors propose five stereological model tests based on the distribution of the number of cell vertices. The power of the model tests is investigated under some special parametric alternative hypotheses: a Matérn cluster point process (CVT), a Matérn hard-core point process (HVT) and a simple sequential inhibition point process (SVT). Second, in Reference 8 three different model tests are considered: the first two are based on the variability of the section cells area, the third is motivated by a well-known relationship between specific edge length L_A and point process intensities λ and P_A. In line with their previous work, the authors propose one-sided and two-sided tests for distinguishing Poisson-Voronoi tessellation from more regular tessellations (HVT and SVT) or irregular tessellations (CVT). The null distributions of the test statistics are approximated using simulation. Simulations also show that the model tests are quite powerful in discriminating the different kind of plane sections. It is interesting to note that all their tests are based on summarizing indices such as the coefficient of variation, skewness index, and so on, and that the best behavior among them is reported to be the one based on the coefficient of variation of the cells area (Equation (2)), also used by the authors in Reference 11.

In this article, we introduce test statistics that use more information contained in the data than only summarizing indices. To this end, we use tools belonging to different branches in statistics. Moreover, we describe a partly simulation-based framework to approximate null distributions of the test statistics considered.

Before going more deeply into the testing problem, it is necessary to make a distinction between periodic and nonperiodic boundary conditions. On the one hand, periodic boundary conditions are mathematically convenient as these provide a natural way to deal with edge effects. Moreover, for large volumes and large values of λ, the construction really mimics the infinite volume situation where the convenient scaling results as mentioned in Section 3. For real materials, the periodic boundary constraint is not realistic. The approach without periodic boundary conditions is more realistic. It will be seen that determining null distributions of test statistics, the approach will be slightly more simulation based, but also more tailored to the data and 3D object at hand.

4.1 | Periodic boundary conditions

The first simulation study involves a Monte Carlo procedure. The following results are obtained by randomly generating approximately 1000 points in a box of dimension $10 \times 10 \times 10$ and using Equation 1 for creating 3D Poisson-Voronoi cells. This is equivalent to saying that the generator points of the Poisson-Voronoi diagram are generated according to a Poisson process with intensity parameter $\lambda = 1$. Then, one section with dimensions 10×10 (parallel to the cube face for reducing boundary effect) for every 3D structure is randomly taken. On average, the number of 2D cells in a section turns out to be approximately 146. The simulation is conducted using the software provided by TATA Steel. The algorithm that the software exploits is described in Reference 6. The procedure consists of three main steps:

Repeat 1000000 times:

Step 1: Generate a 3D Poisson-Voronoi diagram with intensity parameter $\lambda = 1$ applying periodic boundary conditions;
Step 2: Take a random 2D section of the 3D structure;
Step 3: Determine the geometrical characteristics of all cells in the 2D section.

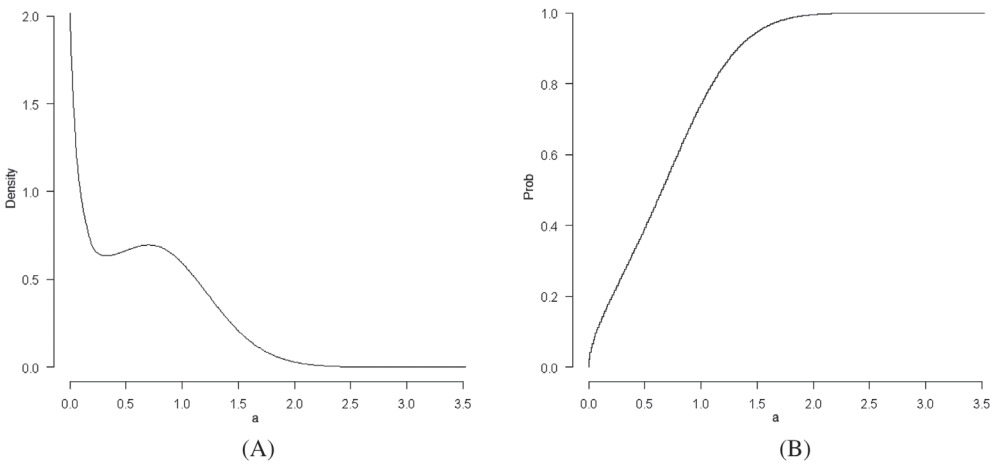


FIGURE 1 (A) Boundary corrected Kernel density estimate (Epanechnikov kernel, linear combination correction, $h = 0.2^{12}$) and (B) empirical cumulative distribution function of the area of 36 480 600 (originating from the 1 000 000 slices) two-dimension sectional cells, $\lambda = 1$

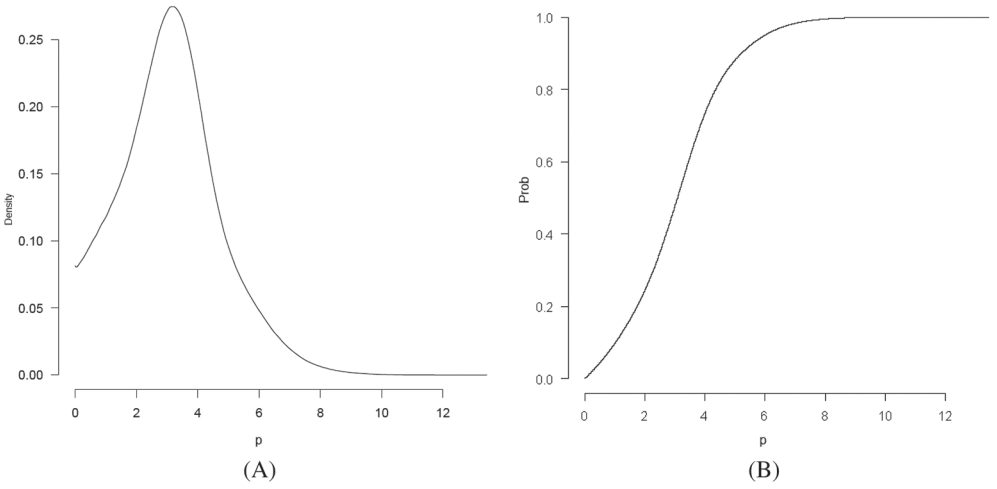


FIGURE 2 (A) Boundary corrected Kernel density estimate (Epanechnikov kernel, linear combination correction, $h = 0.1^{12}$) and (B) empirical cumulative distribution function of the perimeter of 36 480 600 two-dimension sectional cells, $\lambda = 1$

Graphical representations of the results are shown in Figures 1 to 3. For the grain area and the grain perimeter distributions estimation a simple boundary correction for kernel density estimation is adopted.[12] Quantities such as area and perimeter are by definition nonnegative. This means that in general the probability density functions of these quantities have a discontinuity at zero. Kernel estimators are biased at such points of discontinuity and can be corrected for bias using boundary corrected estimate described in Reference 12. The linear correction approach, as proposed in Reference 12 Formula 3.4, prevents the estimate to assign mass outside $[0, \infty)$.

The values in Table 1 are the estimated values of the main geometrical features for a 2D sectional Poisson-Voronoi diagram. They are in agreement with both theoretical and simulation results known in the literature (cf[2]).

If it comes to the study of mechanical properties of metal, the grain size is known to be an important parameter. In 2D, grain area therefore represents one of the most interesting features for real materials sections, especially for single-phase materials.[13] Therefore, in this article we restrict ourselves to tests based on observed cell areas. The first one, mentioned before and already used in Reference 8,11, is based on the coefficient of variation of the observed cell areas:

$$C = \frac{\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(a_i - \bar{a})^2}}{\bar{a}}. \tag{2}$$

Here, a_i is the area of the ith sectional cell and $\bar{a}$ is the mean cell area in the section. As the coefficient of variation is scale invariant, one just needs to compute the coefficient of variation of the area of the cells of a real section applying periodic boundary conditions and compare it with the quantile of the distribution of this test statistic. In fact, the information contained in the 2D section is clearly related to the number of cells observed (n) and comparing the observed value of C

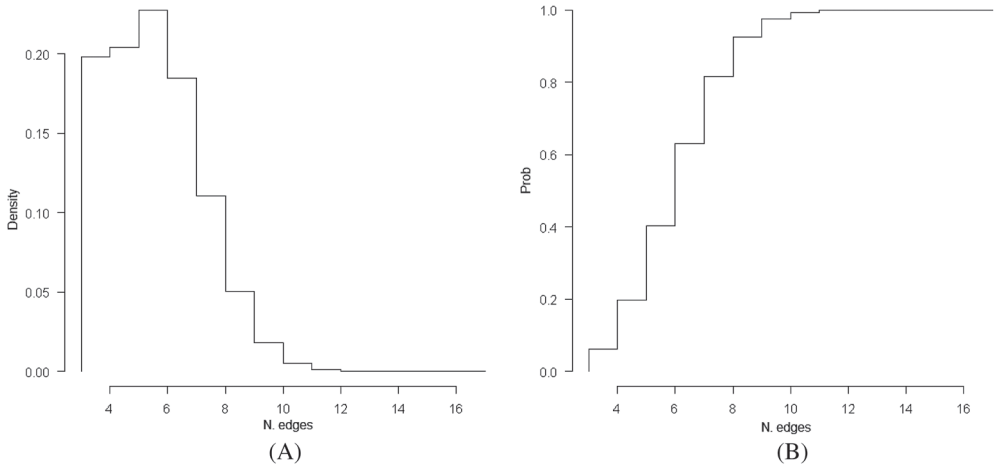


FIGURE 3 (A) Relative frequencies and (B) empirical cumulative distribution function of the number of edges of 36 480 600 two-dimension sectional cells, $\lambda = 1$

with a quantile of the conditional distribution of C given n will only depend on the number of cells observed in the 2D section.

The second test is based on the CDF of the area of the 2D sectional cells. More precisely, it is a Kolmogorov-Smirnov type test given by the supremum distance between the CDF of the area of the cells of the section, for which one wants to test the Poisson-Voronoi hypothesis and a function that reflects our expectation of the empirical distribution function under the Poisson-Voronoi assumption. For the latter, we choose a very accurate simulation-based approximation of the CDF of the area of 36480600 sectional Poisson-Voronoi cells. Let F_1 be the CDF of the areas of the 2D sectional cells with intensity parameter $\lambda = 1$ approximated via simulation as described above and let $\hat{G}$ be the empirical distribution function of the area of n cells of a 2D section from a 3D structure with intensity parameter λ. First, we use Equation (1) for estimating the intensity based on the considered section, $\hat{\lambda}_a$. Furthermore, inspired by lemma 3 in Reference 6, we define the next test statistic as the supremum distance between the two functions:

$$D(F, \hat{G}) = \sup_{x \geq 0} |F_1(x) - \hat{G}(\hat{\lambda}^{\frac{2}{3}} x)|. \tag{3}$$

We will return to the issue of approximating the null distribution of this test statistic in Section 5.

4.2 | Nonperiodic boundary conditions

In most real situations, the data available are relative to a material section with completely visible as well as partially visible grains. In such situations, it is not realistic to use periodic boundary conditions in the model. We fix the geometry of the 3D volume and 2D slice as in the periodic boundaries case (Section 4.1). Then the procedure can be summarized in three main steps:

Repeat 1000000 times:

Step 1: Generate a 3D Poisson-Voronoi diagram with intensity parameter λ not applying periodic boundary conditions. In this article, for reasons that will become clear later, $\lambda = 0.2$ is chosen;
Step 2: Take a random 2D section of the 3D structure;
Step 3: Determine the geometrical characteristics of the completely visible and the partially visible cells in the 2D section.

In this setting, we consider three different tests.

The first recall exactly the one shown in the setting of Section 4.1 (Equation (2)) and it is based on the coefficient of variation of the area of the totally and partially visible cells. Obviously, the referring quantile of the distribution of the statistical test are different with respect to the previous case. The second statistic is in line with the test based on the CDF of the cell areas seen in Section 4.1, but the formulation is slightly different. It is expressed by

$$D(\overline{F}_{\lambda\ n_{2D}}, \hat{G}_{n_{2D}}) = \sup_{x \geq 0} |\overline{F}_{\lambda\ n_{2D}}(x) - \hat{G}_{n_{2D}}(x)|, \tag{4}$$

where $\overline{F}_{\lambda\ n_{2D}}$ is the expected CDF conditioned to the event of observing exactly n_{2D} sectional cells with estimated parameter λ. In Section 5, it will be explained in more detail how this can be computed. $\hat{G}_{n_{2D}}$ is the CDF of the areas of the totally and partially visible cells of the section under study.

The last test exploits tools coming from the emergent field of TDA. We will now explain the main concepts of persistence homology necessary for using our model test.

Test based on persistence plots

Instead of giving rigid mathematical and topological definitions, the aim of this section is to guide the reader via intuitive concepts in the construction of *persistence diagrams* and *persistence landscapes* used for the last model test. Looking at one 2D image, it is hard to identify the really "important" features that univocally characterize it. TDA is a relatively new discipline that has provided new insight into the study of qualitative features of data. In particular, persistent homology is the branch of TDA that provides tools both for identifying qualitative features of data and to give a measure of the importance of those features. Key topological features of a set include connected components, holes, voids, and so on. The main aim of persistent homology is to record the evolution of those characteristics with respect to a scale parameter r that usually can be interpreted as time.

For avoiding too long digressions that can drift away from the real scope of the article, most of the main concepts belonging to homology and persistent homology field are just roughly mentioned. For readers that aim to come to a formal definition of the following procedure, in Reference 14 more details are provided.

For illustrative reasons and because in this study 2D images are used, the 2D case is considered but generalization to higher dimensions is not complicated. The input of the analysis typically takes the form of a point cloud $\mathbf{X}$ (Figure 4A). Based on that, a special structure is built. It provides information about the qualitative features discussed above. This structure is based on so-called *simplices*. A *geometric k-simplex* is the convex hull of $k+1$ affinely independent points $v_0, v_1, \dots, v_k$. More precisely the 0-simplex identifies vertices, the 1-simplex line segments, and the 2-simplex triangles.

One way for building this structure starts off with the so-called *Delaunay triangulation*, DT($\mathbf{X}$) of $\mathbf{X}$. Basically, this is a graph consisting of vertices in $\mathbf{X}$ and edges between two points if and only if they share a Voronoi edge (Figure 4B). Then, circles are grown with increasing radius r, centered at the points in $\mathbf{X}$. The *Alpha complex** at radius r, $\alpha_r(\mathbf{X})$, is a subcomplex of DT($\mathbf{X}$). In fact, for r very small, the Alpha complex is nothing but the set $\mathbf{X}$ of the generator points. Then r grows and once two circles intersect, the edge of the underlying DT between the two circle centers is added to $\alpha_r(\mathbf{X})$. Eventually, for r very big, the Alpha complex is the DT itself (Figure 4C-I).

Now, rather than considering this structure for some fixed value of r, its evolution for growing $r > 0$ is registered. In particular, we keep track of the birth time b and a death time d of connected components and holes†, where the "time" is given by the radius of the circles corresponding to those events. One can think of the circles radii growing at constant rate. At time zero, the Alpha complex equals $\mathbf{X}$. All individual points are separate connected components. These are born at time zero. After some time, when the first two points get connected because their circles touch, one can say two connected components merge or one connected component "dies" (in a more general case, in which the birth times might be different the edge that causes two components to merge is paired with the component that was born most recently, that is, the one with the later birth "time"). In Figure 4, this happens for $r = 0.47$; see subplot (C). For one connected component, we therefore have $(b, d) = (0, 0.47)$. Increasing r further, more connected components will "die" until only one remains for all r large enough because all points are covered by the union of all large circles. During the same process, it is also possible that holes appear. This happens when a polygon (eg, triangle) appears in the picture, such that the r-circles around the corner points of this polygon do not cover the whole polygon. At this time a hole is "born," yielding a birth time b for this feature. It will also "die" again, when r is further increased and the circles centered at the corners do cover the whole polygon. Note that not all polygons that appear correspond to the birth of a hole. For instance, in the case of triangles, acute triangles always generate a hole, while obtuse triangles not. In Figure 4G a triangle appears but the circles centered at the three corners immediately cover the whole triangle.

*Other common choices of simplicial complexes are Čech and Rips complexes.[14-16]

†In 3D one could also consider other topological features, such as voids or cavities (holes in more than 2D).

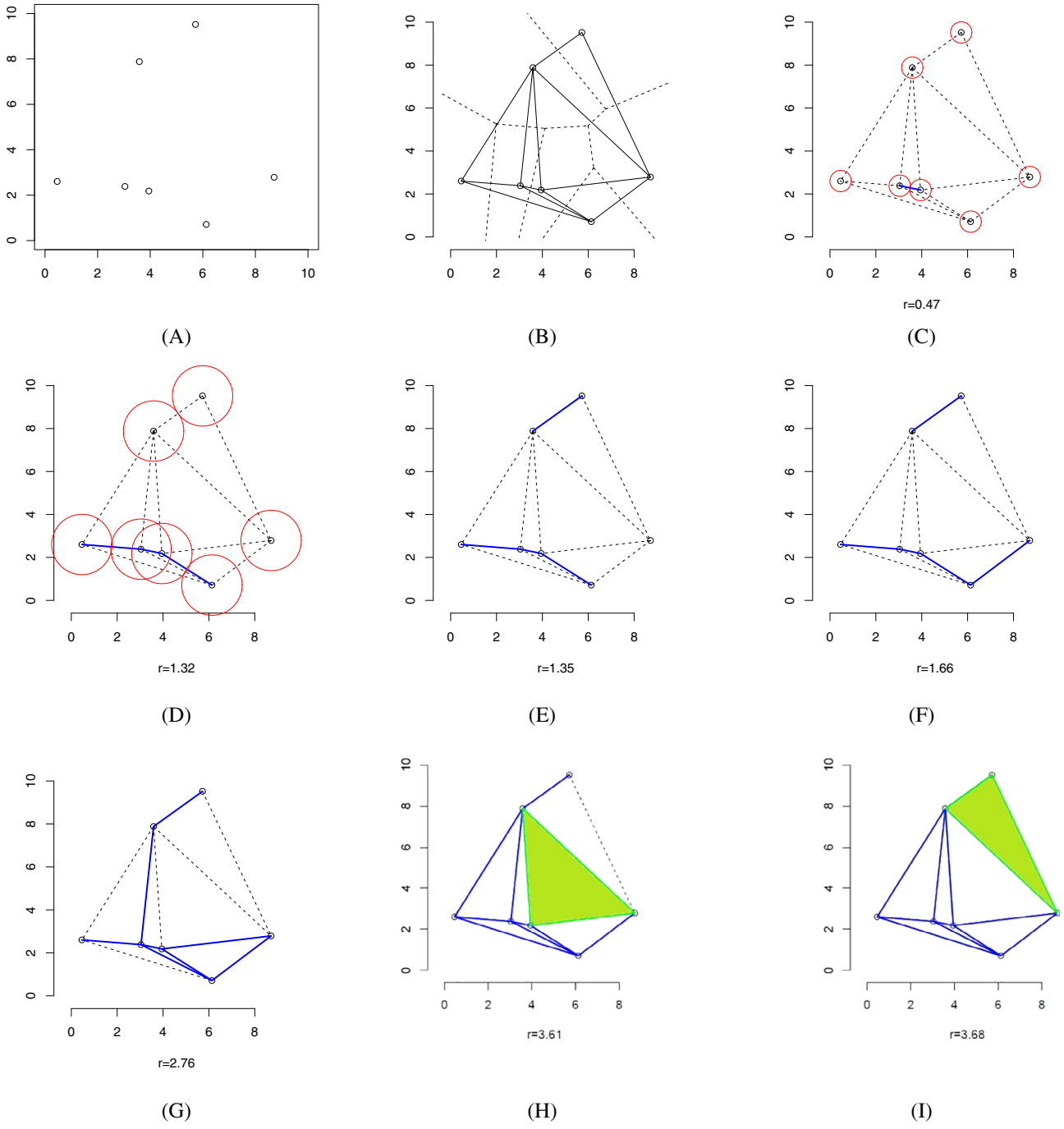


FIGURE 4 (A) Set of points $\mathbf{X}$, (B) Voronoi diagram (dashed) and Delaunay triangulation (solid), (C) Circles with radius 0.47 around the points of $\mathbf{X}$; the Alpha complex $\alpha_r(\mathbf{X})$ consists of the individual points of $\mathbf{X}$ and the one edge corresponding to the two touching circles, (D) Alpha complex for $r = 1.32$, (E) Alpha complex, $r = 1.35$, (F) Alpha complex, $r = 1.66$, (G) Alpha complex, $r = 2.76$, (H) Alpha complex, $r = 3.61$, and (I) Alpha complex, $r = 3.68$ [Colour figure can be viewed at wileyonlinelibrary.com]

The points (b, d) thus obtained can be used as coordinates and plotted on a plane, resulting in the so-called *persistence diagram*. (see 17 for a detailed explanation or alternatively 18 for barcode plot). Since the topological features (connected components, holes) can only die *after* they are born ($d \geq b$), necessarily each point appears on or above the diagonal line $y = x$. The persistence diagram corresponding to the data in Figure 4 is shown in Figure 5. The black dots, D_{0i}, on the vertical axis represent the "deaths" of connected components; the lowest being the aforementioned $(b, d) = (0, 0.47)$, the highest, $(b, d) = (0, 2.67)$, corresponding to Figure 4G. The red triangles D_{1i}, represent the birth- and death times of the holes.

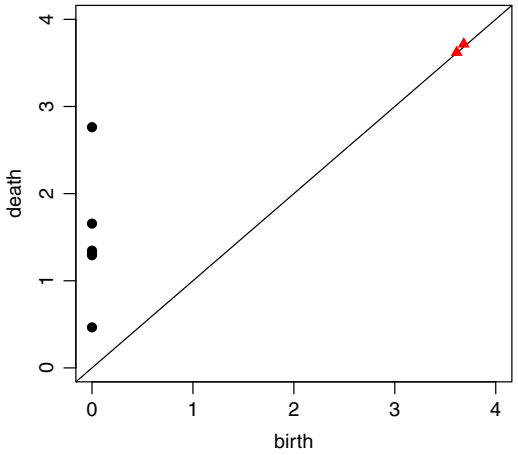


FIGURE 5 Persistence diagram. The black dots indicate the birth- and death time of connected components and the red triangles the birth- and death times of the holes. The data are the same as those used for Figure 4 [Colour figure can be viewed at wileyonlinelibrary.com]

Based on persistence diagrams, several descriptive summarizing functions have been proposed in the literature. For example, rank functions,[19] landscapes and silhouettes,[20,21] and accumulated persistence functions.[22] In this article, we follow the persistence landscapes approach, but any other summary statistic could also be used for testing.

We first describe in words how to construct a landscape from a persistence diagram. Then, the formal definition follows. For each point (b, d) in the persistence diagram, count the number of points to its left top (north-west). This is the *rank* of the point (b, d) and it can be interpreted as the number of features that are alive at time b and that are still alive at time d. Then, draw horizontal and vertical lines from each point (b, d) in the persistence diagram to the diagonal and "tip the diagram on its side." Then take the contour of the projection of the points with the same rank. This results in the so-called landscape. This is done for connected components and holes separately, see Figure 6.

More formally, a persistence landscape is a sequence of continuous, piecewise linear functions $\lambda(k, \cdot) : \mathbb{R}^+ \to \mathbb{R}^+, k = 1, 2, \ldots$. Denote the set of "persistence points" in the persistence diagram by D. Then for each $p = (b, d) \in D$ define the triangular functions

$$\Lambda_p(t) = \begin{cases} t - b & t \in \left[b, \frac{b+d}{2}\right] \\ d - t & t \in \left(\frac{b+d}{2}, d\right] \\ 0 & \text{otherwise.} \end{cases}$$

Then, the persistence landscape of the persistence diagram is defined by

$$\lambda_D(k, t) = \text{kmax}_{p \in D} \Lambda_p(t), \quad t \geq 0, k \in \mathbb{N}. \tag{5}$$

Here, kmax selects the kth largest value in the set, so for a particular k at each t, $\lambda_D(1, t)$ is the largest value in the set $\{\Lambda_1(t), \Lambda_2(t), \ldots, \Lambda_p(t)\}, p \in D$, $\lambda_D(2, t)$ the second largest value, and so on.

Our test will be the contrast between the observed landscape and a landscape one would expect under the null hypothesis that the 3D structure is Poisson-Voronoi. For this *mean landscape*, we use the conditional expectation of the landscape given that $N_{2D} = n_{2D}$ and approximate this using the simulation procedure described in Section 4.1. To be more specific,

$$\overline{\lambda}_{D_j}(k, t) = \frac{1}{n} \sum_{i=1}^{n} \lambda_{D_j(i)}(k, t) \quad j = 0, 1, \quad t \geq 0, \tag{6}$$

where n is the number of 2D Poisson-Voronoi sections generated with $N_{2D} = n_{2D}$. Inspired by the approach proposed in Reference 19, the test statistics are then given by the distance between persistence landscapes and mean persistence landscapes using L^2 norm,

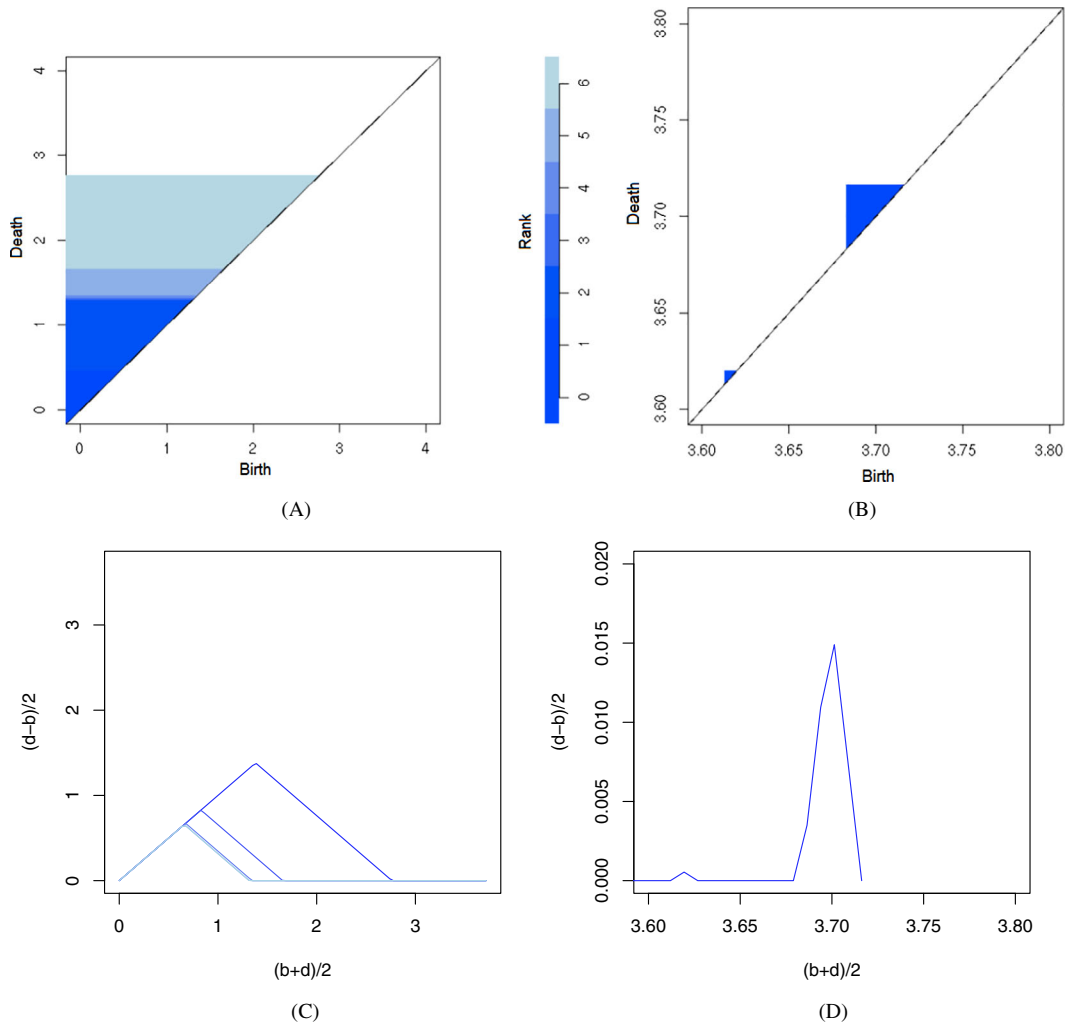


FIGURE 6 Rank function for connected components (A) and holes (B) persistence landscapes for connected components (C) and holes (D) [Colour figure can be viewed at wileyonlinelibrary.com]

$$L_0 = \|\hat{\lambda}_{D_0} - \overline{\lambda}_{D_0}\|_2 = \left[\sum_{k=1}^{n_{2D}-1} \int_0^T (\hat{\lambda}_{D_0}(k,t) - \overline{\lambda}_{D_0}(k,t))^2 \mathrm{dt}\right]^{\frac{1}{2}}$$
$$L_1 = \|\hat{\lambda}_{D_1} - \overline{\lambda}_{D_1}\|_2 = \left[\sum_{k=1}^{\infty} \int_0^T (\hat{\lambda}_{D_1}(k,t) - \overline{\lambda}_{D_1}(k,t))^2 \mathrm{dt}\right]^{\frac{1}{2}}. \quad (7)$$

Here, $\hat{\lambda}_{D_j}(k,\cdot), j = 0, 1$ is the kth landscape for the connected components ($j = 0$) and for the holes ($j = 1$) for the 2D section under study. If both L_0 and L_1 are less than the threshold quantiles, the Poisson-Voronoi hypothesis is not rejected.

5 | Bootstrap Confidence Interval for λ and Quantiles of the model tests

In Reference 8, the authors carry out a simulation for estimating the quantiles of the test statistics proposed there. Cells of 3D spatial Poisson-Voronoi diagrams are generated with $\lambda = 1$. Then, a random planar section of the 3D structure is taken and square observation windows are drawn in the section planes with an expected number of 50, 100, 150, and 200 cells, respectively.

We provide an expression for the distribution of any test statistic given the number of observed cells in the section, separating a part that depends on the parameter λ and a part that does not. We consider the situation where we see a

window (with known shape and size) of a 2D planar section of a 3D Poisson-Voronoi diagram in a 3D object of known geometry. As before, denote by N_{3D} the number of cells in the 3D object and N_{2D} the number of 2D cells visible in the 2D window. Lemma 1 below gives an expression of the null distribution of a test statistic T, given n_{2D} cells are observed in the section. It separates a part that depends on the intensity parameter λ and a part that does not.

Lemma 1. *Let T denote a general model test for the Poisson-Voronoi assumption validation. The conditional probability $P_\lambda(T \geq t|N_{2D} = n_{2D})$ can be expressed as*

$$P_\lambda(T \geq t|N_{2D} = n_{2D}) = \frac{\sum_{k=n_{2D}}^{\infty} P(T \geq t|N_{3D} = k, N_{2D} = n_{2D})\, P(N_{2D} = n_{2D}|N_{3D} = k)\frac{(\lambda\mathcal{V})^k}{k!}}{\sum_{j=n_{2D}}^{\infty} P(N_{2D} = n_{2D}|N_{3D} = j)\frac{(\lambda\mathcal{V})^j}{j!}}. \tag{8}$$

Proof.

$$\begin{aligned} P_\lambda(T \geq t|N_{2D} = n_{2D}) &= \sum_{k=0}^{\infty} P_\lambda(T \geq t, N_{3D} = k|N_{2D} = n_{2D}) \\ &= \sum_{k=n_{2D}}^{\infty} P_\lambda(T \geq t, N_{3D} = k|N_{2D} = n_{2D}) \\ &= \sum_{k=n_{2D}}^{\infty} P_\lambda(T \geq t|N_{3D} = k, N_{2D} = n_{2D})\, P_\lambda(N_{3D} = k|N_{2D} = n_{2D}) \\ &= \sum_{k=n_{2D}}^{\infty} P(T \geq t|N_{3D} = k, N_{2D} = n_{2D})\, P_\lambda(N_{3D} = k|N_{2D} = n_{2D}). \end{aligned} \tag{9}$$

In the last equality the λ-dependence disappears from the first factor, because, conditionally on N_{3D}, the distribution of T does not depend on λ. The λ-dependent part in Equation (9) can be made more explicit also using that conditionally on N_{3D}, the distribution of N_{2D} does not depend on λ:

$$\begin{aligned} P_\lambda(N_{3D} = k|N_{2D} = n_{2D}) &= \frac{P(N_{2D} = n_{2D}|N_{3D} = k)P_\lambda(N_{3D} = k)}{P_\lambda(N_{2D} = n_{2D})} \\ &= \frac{P(N_{2D} = n_{2D}|N_{3D} = k)P_\lambda(N_{3D} = k)}{\sum_{j=n_{2D}}^{\infty} P(N_{2D} = n_{2D}|N_{3D} = j)P_\lambda(N_{3D} = j)} \\ &= \frac{P(N_{2D} = n_{2D}|N_{3D} = k)\frac{(\lambda\mathcal{V})^k}{k!}}{\sum_{j=n_{2D}}^{\infty} P(N_{2D} = n_{2D}|N_{3D} = j)\frac{(\lambda\mathcal{V})^j}{j!}}. \end{aligned} \tag{10}$$

Combining Equations (9) and (10) yields Equation (8). ∎

For computing P-values in practice, the value of λ is needed. In order to take into account the uncertainty in the estimate of λ, while computing P-values for model tests, we compute a 90% confidence interval for λ. To do this, as this value is not known, we propose a bootstrap approach. More precisely, we want to compute

$$\begin{aligned} P_\lambda(\sqrt{\hat{\lambda}} - \sqrt{\lambda} \leq u) &= \sum_{k=0}^{\infty} P_\lambda(\sqrt{\hat{\lambda}} - \sqrt{\lambda} \leq u, N_{3D} = k) \\ &= \sum_{k=0}^{\infty} P_\lambda(\sqrt{\hat{\lambda}} - \sqrt{\lambda} \leq u|N_{3D} = k)P_\lambda(N_{3D} = k). \end{aligned} \tag{11}$$

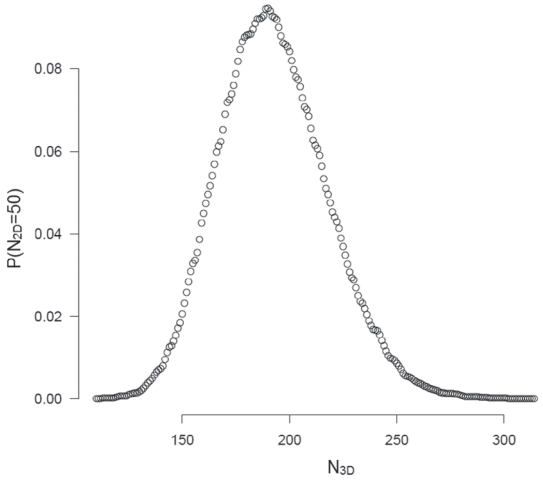


FIGURE 7 Monte Carlo approximation of $P(N_{2D} = 50|N_{3D} = k)$

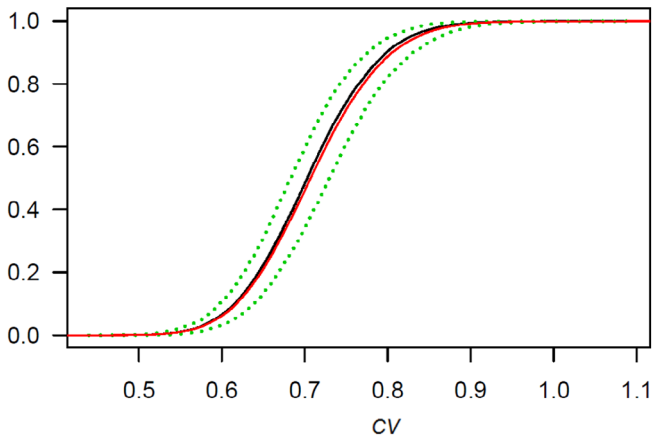


FIGURE 8 Cumulative distribution function of the coefficient of variation of the two-dimension (2D) sectional cells area conditioned on $N_{2D} = 50$ (black line; green dotted lines are obtained using the upper and lower limit of the confidence set for $\hat{\lambda} = 0.2$) and unconditioned (red line) [Colour figure can be viewed at wileyonlinelibrary.com]

The procedure can be summarized as follows: first we estimate λ from a real 2D image, using $\hat{\lambda}_a$ (Equation (1)). For computing $P_\lambda(N_{3D} = k)$, $P_{\hat{\lambda}_a}(N_{3D} = k)$ is then used. Second, for computing $P_\lambda(\sqrt{\hat{\lambda}} - \sqrt{\lambda} \leq u|N_{3D} = k)$, 10000 Poisson-Voronoi diagrams for each realization of a Poisson process with $\hat{\lambda}_a$ in a cube are generated. Then a 2D section from each 3D diagram is randomly taken and the number of cells in the section is used for estimating λ. Next, the probability of having exactly k cells in 3D, $P(N_{3D} = k)$, is used as weight for computing a weighted mean CDF. Finally, a square root transformation for normalizing and stabilizing the variance is used for computing the confidence set:[23]

$$P\left[\sqrt{\hat{\lambda}} - l_{0.95} \leq \sqrt{\lambda} \leq \sqrt{\hat{\lambda}} - l_{0.05}\right] \approx 0.90. \tag{12}$$

As an example, if $\hat{\lambda}_a = 0.2$ (as in the application shown in Section 7 $n_{2D} = 50$ and window size$= 10 \times 10$), the resulting 90%-confidence set is given by:

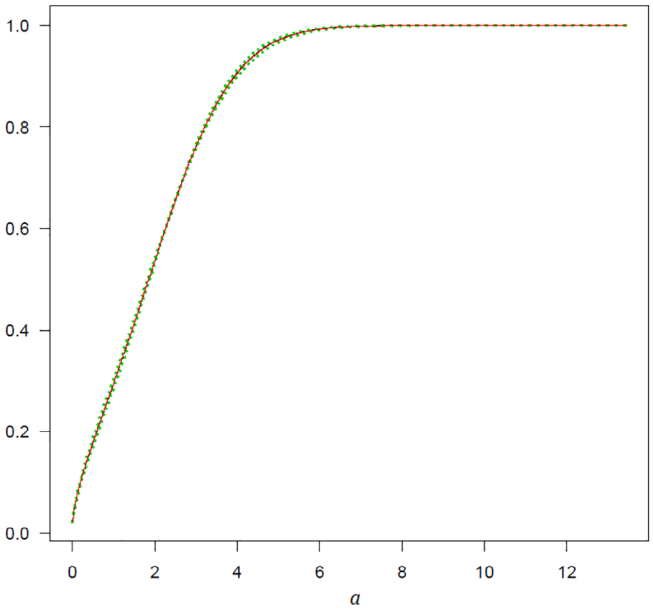
$$[0.1498; 0.2439]. \tag{13}$$

Having a confidence set for λ at hand, the next step is to compute the null distribution described in Lemma 1 for the various test statistics. A visualization of the probability of observing exactly 50 cells given $P(N_{3D} = k)$ is shown in (Figure 7). The resulting P-values depend on λ, but we can consider these for all λ in the confidence set constructed. We start with the coefficient of variation as test statistic (Equation (2)). In Figure 8 it is possible to see the difference between the CDFs of the coefficient of variation of the 2D sectional cells area unconditioned and conditioned on seeing exactly 50 cells in the 2D section. Moreover, the green dotted lines represent the CDF of the coefficient of variation for the lower and upper bounds of the λ confidence set. Note that the distance between the two cdfs is small, showing that the approach of Reference 8 to use an unconditional distribution in this particular setting leads to comparable results. In Table 2, quantiles for the conditional distribution of the CV of the cells area are shown ($\hat{\lambda} = 0.2$).

TABLE 2 Quantiles of the conditional distribution of the coefficient of variation of the two-dimension (2D) sectional cells area given that $N_{2D} = 50$, ($\lambda = 0.2$)

α	0.005	0.01	0.0125	0.025	0.05	0.1	0.9	0.95	0.975	0.9875	0.99	0.995
c_α	0.531	0.547	0.553	0.571	0.591	0.615	0.798	0.826	0.853	0.875	0.883	0.903

FIGURE 9 Cumulative distribution function of the two-dimension (2D) sectional cells area conditioned on $N_{2D} = 50$ (black line; green dotted lines are obtained using the upper and lower limit of the confidence set for $\hat{\lambda} = 0.2$) and unconditioned (red line) [Colour figure can be viewed at wileyonlinelibrary.com]



In Figure 9, the conditional weighted mean CDF for cells area (black line), its confidence bands (green dotted lines) and the unconditional mean are shown. More precisely, we define

$$\begin{aligned}\overline{F}_{\lambda\ n_{2D}}(x) &= \mathbb{E}_\lambda\left\{F_{N_{2D}}(x)|N_{2D} = n_{2D}\right\} = \mathbb{E}_\lambda\left\{\mathbb{E}(F_{N_{2D}}(x)|N_{2D} = n_{2D}, N_{3D})\right\} \\ &= \sum_{k=n_{2D}}^{\infty} P_\lambda(N_{3D} = k|N_{2D} = n_{2D}) \cdot \mathbb{E}(F_{N_{2D}=n_{2D},N_{3D}=k}(x)). \end{aligned} \tag{14}$$

where $F_{N_{2D}=n_{2D},N_{3D}=k}(x)$ is the empirical distribution function of the areas given k cells in 3D structure and n_{2D} visible on the slice. The same type of expression is used also for $\overline{\lambda}_{D_0}(1,t)$ and $\overline{\lambda}_{D_1}(1,t)$.

In Figure 10 the CDF of the test based on the supremum distance between ecdfs of the 2D sectional cells area is shown. As for the test based on coefficient of variation, the difference between conditional and unconditional approach is relatively small.

Switching to the test based on persistence landscapes, Figures 11 and 12 are visualizations of k mean persistence landscapes conditioned on $N_{2D} = 50$ for connected components and holes, respectively, when $\hat{\lambda} = 0.2$. Instead, Figures 13 and 14 are the conditional maximum weighted means (black lines) and their confidence bands (green dotted lines).

In Figures 15 and 16 the CDF of the test based on the L_2 distance between persistence landscapes (connected components and holes) are shown. In addition in this case, the difference between conditional and unconditional approach seems to be irrelevant.

For computing the quantiles of the distribution of the model tests based on CDF and on persistence landscape (Tables 3, 4 and 5), we use a "leave one out" procedure.[24] Here, we use the B 2D slices generated as follows:

- For the test based on CDFs difference Equation (4):

$$d_i = \sup_{x\in\mathbb{R}} |\overline{F}_{\lambda\ n_{2D}(-i)}(x) - \hat{F}_{n_{2D}(i)}(x)|, 1 \le i \le B. \tag{15}$$

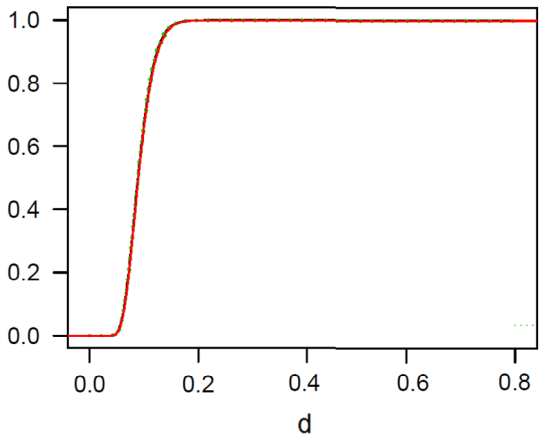


FIGURE 10 Cumulative distribution function of the ecdf test of the two-dimension (2D) sectional cells area conditioned on $N_{2D} = 50$ (black line; green dotted lines are obtained using the upper and lower limit of the confidence set for $\hat{\lambda} = 0.2$) and unconditioned (red line) [Colour figure can be viewed at wileyonlinelibrary.com]

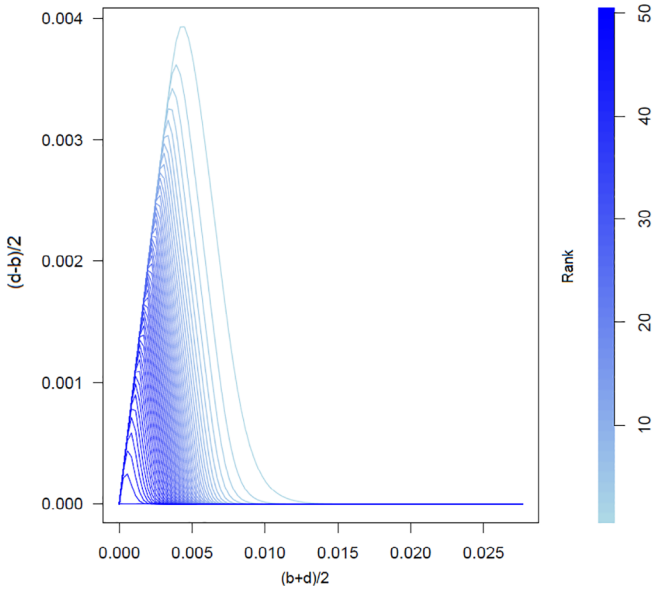


FIGURE 11 *k* Weighted mean landscapes (connected components) for sections with exactly 50 two-dimension sectional cells, ($\hat{\lambda} = 0.2$) [Colour figure can be viewed at wileyonlinelibrary.com]

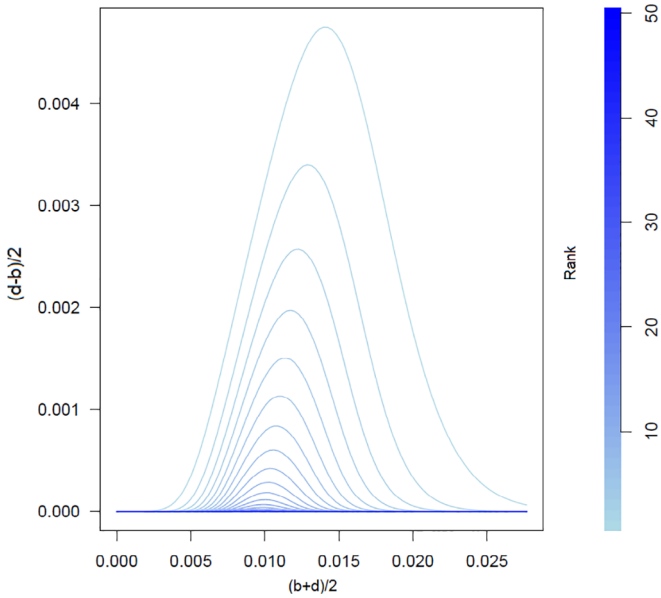


FIGURE 12 *k* Weighted mean landscapes (holes) for sections with exactly 50 2D sectional cells, ($\hat{\lambda} = 0.2$) [Colour figure can be viewed at wileyonlinelibrary.com]

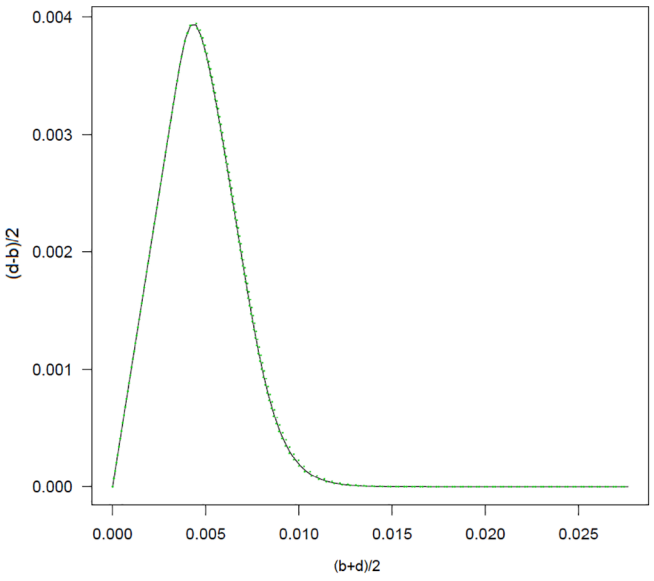


FIGURE 13 Max weighted mean landscape (connected components) for sections with exactly 50 two-dimension sectional cells (black line; green dotted lines are obtained using the upper and lower limit of the confidence set for $\hat{\lambda} = 0.2$) [Colour figure can be viewed at wileyonlinelibrary.com]

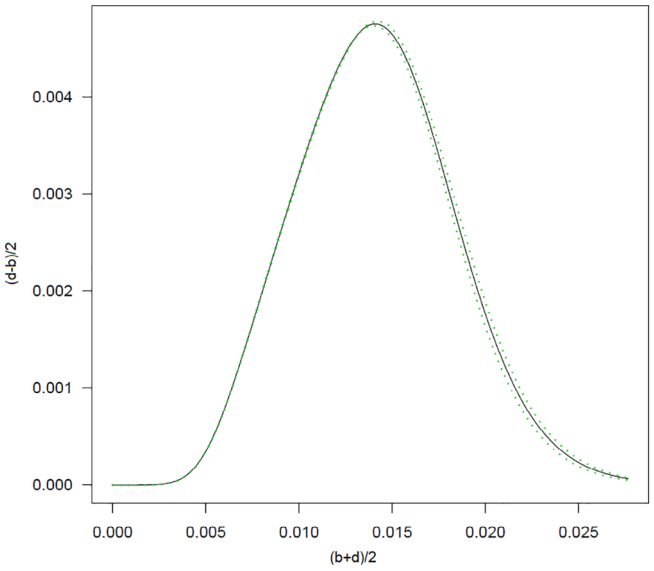


FIGURE 14 Max weighted mean landscape (holes) for sections with exactly 50 two-dimension sectional cells (black line; green dotted lines are obtained using the upper and lower limit of the confidence set for $\hat{\lambda} = 0.2$) [Colour figure can be viewed at wileyonlinelibrary.com]

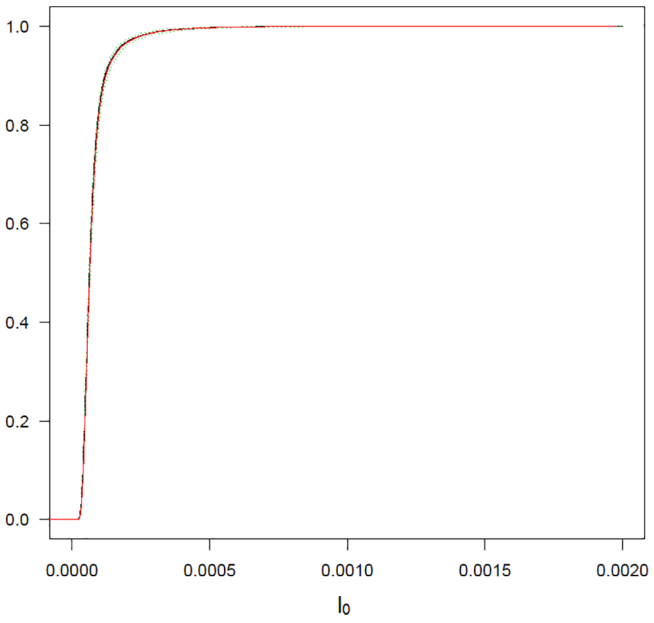


FIGURE 15 Cumulative distribution function of the test based on the L_2 distance between persistence landscapes L_0, (7), of the two-dimension sectional cells area conditioned on $N_{2D} = 50$ (black line; green dotted lines are obtained using the upper and lower limit of the confidence set for $\hat{\lambda} = 0.2$) and unconditioned (red line) [Colour figure can be viewed at wileyonlinelibrary.com]

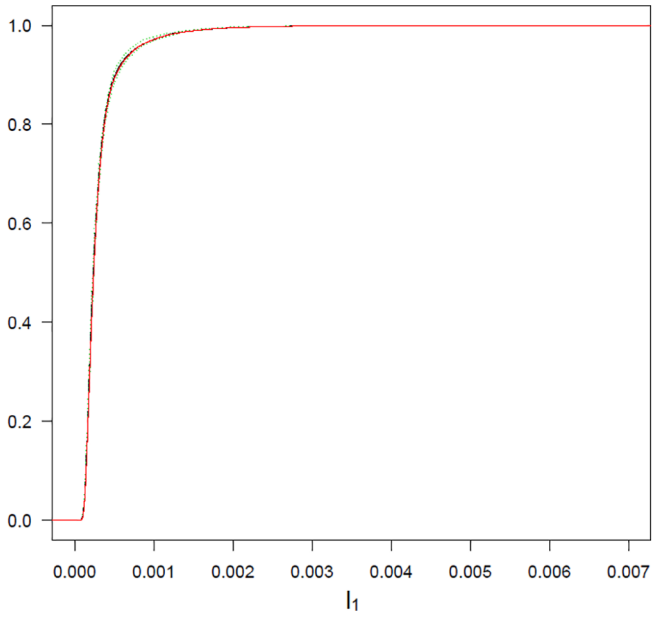


FIGURE 16 Cumulative distribution function of the test based on the L_2 distance between persistence landscapes L_1, (7), of the two-dimension sectional cells area conditioned on $N_{2D} = 50$ (black line; green dotted lines are obtained using the upper and lower limit of the confidence set for $\hat{\lambda} = 0.2$) and unconditioned (red line) [Colour figure can be viewed at wileyonlinelibrary.com]

TABLE 3 Quantiles of the conditional distribution of the test based on the difference between cumulative distribution functions of the two-dimension sectional cells area given that N_{2D}=50, ($\hat{\lambda} = 0.2$)

α	0.005	0.01	0.0125	0.025	0.05	0.1	0.9	0.95	0.975	0.9875	0.99	0.995
d_α	0.047	0.050	0.051	0.054	0.058	0.064	0.123	0.135	0.146	0.155	0.159	0.168

TABLE 4 Quantiles of the conditional distribution of the test based on the difference between the observed landscapes and the conditional mean landscapes (connected components) of the two-dimension sectional cells area given that $N_{2D} = 50$, ($\hat{\lambda} = 0.2$)

α	0.005	0.01	0.0125	0.025	0.05	0.1	0.9	0.95	0.975	0.9875	0.99	0.995
$l_{0\alpha} \times 10^{-5}$	2.402	2.602	2.802	3.003	3.403	3.803	10	20	20	30	30	40

TABLE 5 Quantiles of the conditional distribution of the test based on the difference between the observed landscapes and the conditional mean landscapes (holes) of the two-dimension sectional cells area given that $N_{2D} = 50$, ($\hat{\lambda} = 0.2$)

α	0.005	0.01	0.0125	0.025	0.05	0.1	0.9	0.95	0.975	0.9875	0.99	0.995
$l_{1\alpha} \times 10^{-5}$	9.109	9.810	9.810	10	10	10	50	70	100	140	150	190

- For the test based on persistence landscapes difference Equation (7):

$$l_{0(i)} = \left[\sum_{k=1}^{n_{2D}-1} \int_0^T (\hat{\lambda}_{D_0(i)}(k,t) - \overline{\lambda}_{D_0(-i)}(k,t))^2 \mathrm{dt} \right]^{\frac{1}{2}} \quad 1 \le i \le B,$$

$$l_{1(i)} = \left[\sum_{k=1}^{\infty} \int_0^T (\hat{\lambda}_{D_1(i)}(k,t) - \overline{\lambda}_{D_1(-i)}(k,t))^2 \mathrm{dt} \right]^{\frac{1}{2}} \quad 1 \le i \le B. \tag{16}$$

Here, $\hat{F}_{n_{2D}(i)}$, $\hat{\lambda}_{D_0(i)}$, and $\hat{\lambda}_{D_1(i)}$ are the empirical results for the section i and $\overline{F}_{n_{2D}(-i)}$, $\overline{\lambda}_{D_0(-i)}$, and $\overline{\lambda}_{D_1(-i)}$ are the mean result computed for all the B sections leaving out the ith.

6 | POWER OF THE MODEL TESTS

To assess the power of the three different tests proposed in Section 4, the values of their power functions are estimated under a specific alternative hypothesis: 2D Poisson-Voronoi diagram. As stated in Section 2, the authors in References 9,10

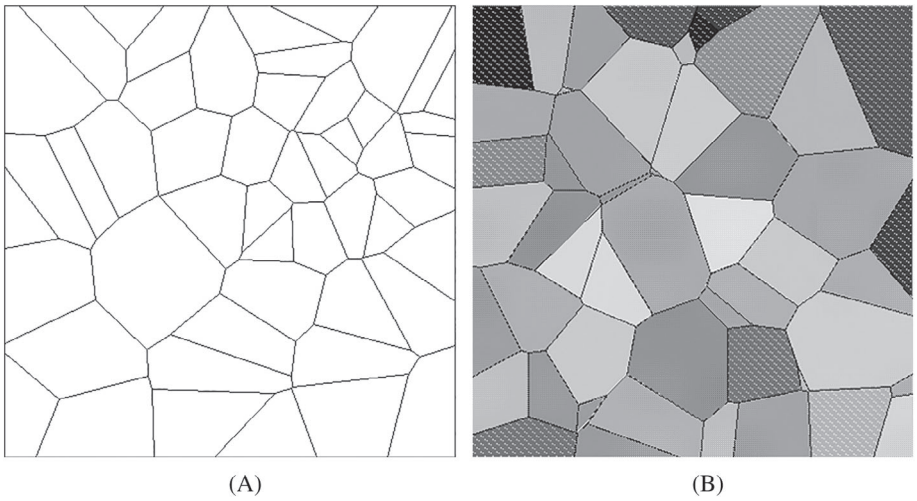


FIGURE 17 (A) 2D Poisson-Voronoi diagram and (B) 2D sectional Poisson-Voronoi diagram, $N_{2D} = 50$. 2D, two dimensions

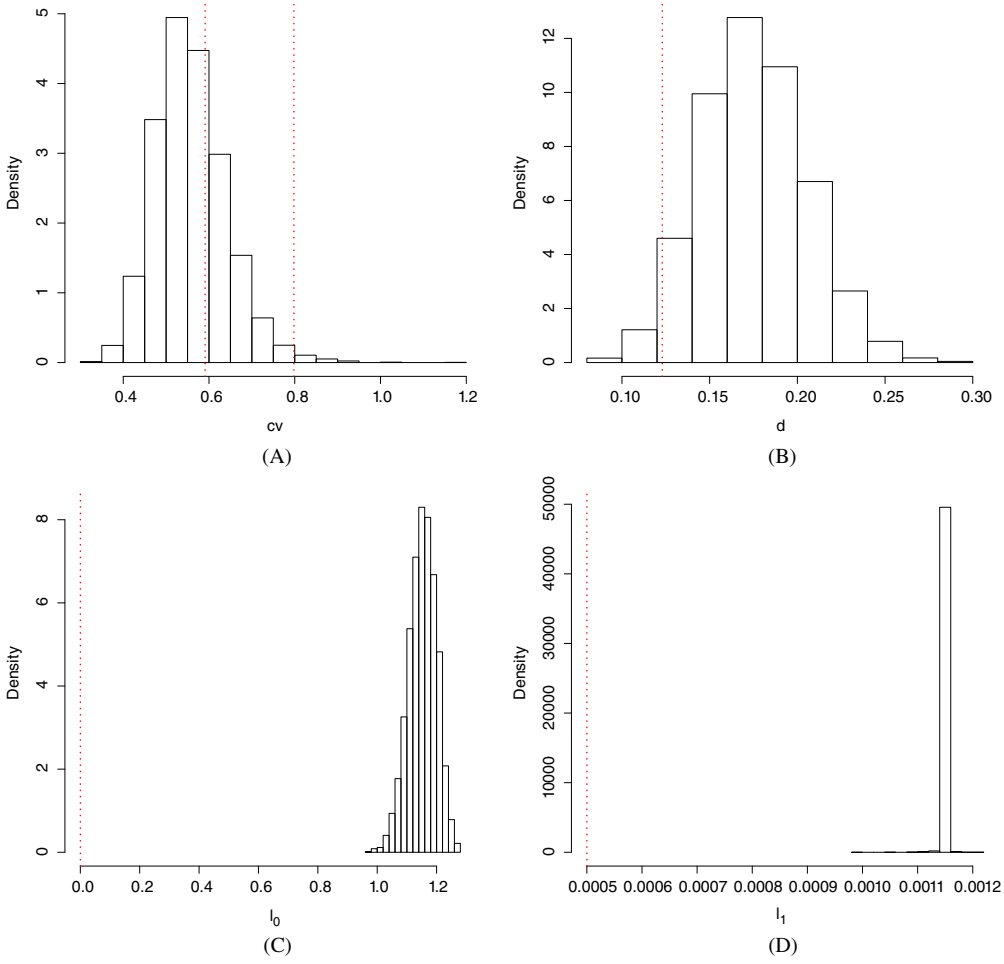


FIGURE 18 Histograms of the results of the coefficient of variation test (A), of the ecdf test (B), of the test based on the L_2 distance between persistence landscape L_0 (C) and L_1 (D) for 7000 two-dimensions Poisson-Voronoi diagram. Red dotted lines identify the quantile of the conditional distribution of the test for $N_{2D} = 50$ [Colour figure can be viewed at wileyonlinelibrary.com]

showed that a 2D sectional Poisson-Voronoi diagram cannot be a planar Poisson-Voronoi tessellation. Some mean values of the geometrical characteristics of the sectional diagram clearly deviate from the tessellation resulting from a 2D Poisson point process. Nevertheless, at a first sight the difference between them is not immediate (Figure 17).

We generate 7000 2D Poisson-Voronoi diagrams with exactly 50 cells and intensity parameter $\lambda = 0.2$ under nonperiodic boundary conditions and the main graphical results are shown in Figure 18. As the simulations results show, the most powerful test is the one based on the L_2 distance between persistence landscapes. Good results are obtained also for the test based on the ecdf difference: its power value is around 90%. The coefficient of variation reaches, however, a considerably lower power 78.6%.

7 | APPLICATION TO SINGLE-PHASE ALUMINA CERAMICS

In Reference 8, it is stated that single-phase microstructures, for example, alumina ceramics, can be well approximated by Poisson-Voronoi diagrams. Using the same images shown in Reference 8, the tests proposed in the previous section (Section 4) are performed.

First all the cells in the images (Figure 19A) are involved in tests computation. Hereafter, for illustrative purposes and for a better comparison with the theoretical results shown in the previous section, we decide to consider just part of the images used in Reference 8. In fact, the original window size is reduced until exactly 50 cells are visible or partially visible (Figure 19B). In Tables 6 and 7, the test statistics and the *P*-values (values in brackets) are shown for the four model tests and following the two different approaches. Figures 20 to 29) are graphical representations of the CDF test and the persistence approach steps. In particular, for applying the test based on the difference between persistence landscapes, we take the center of mass of the cells in the images (Figures 21 and 26), then compute the persistence diagrams (Figures 22 and 27), and finally the persistence landscapes (Figures 23, 24, 28, and 29) as explained in Section 4.2.

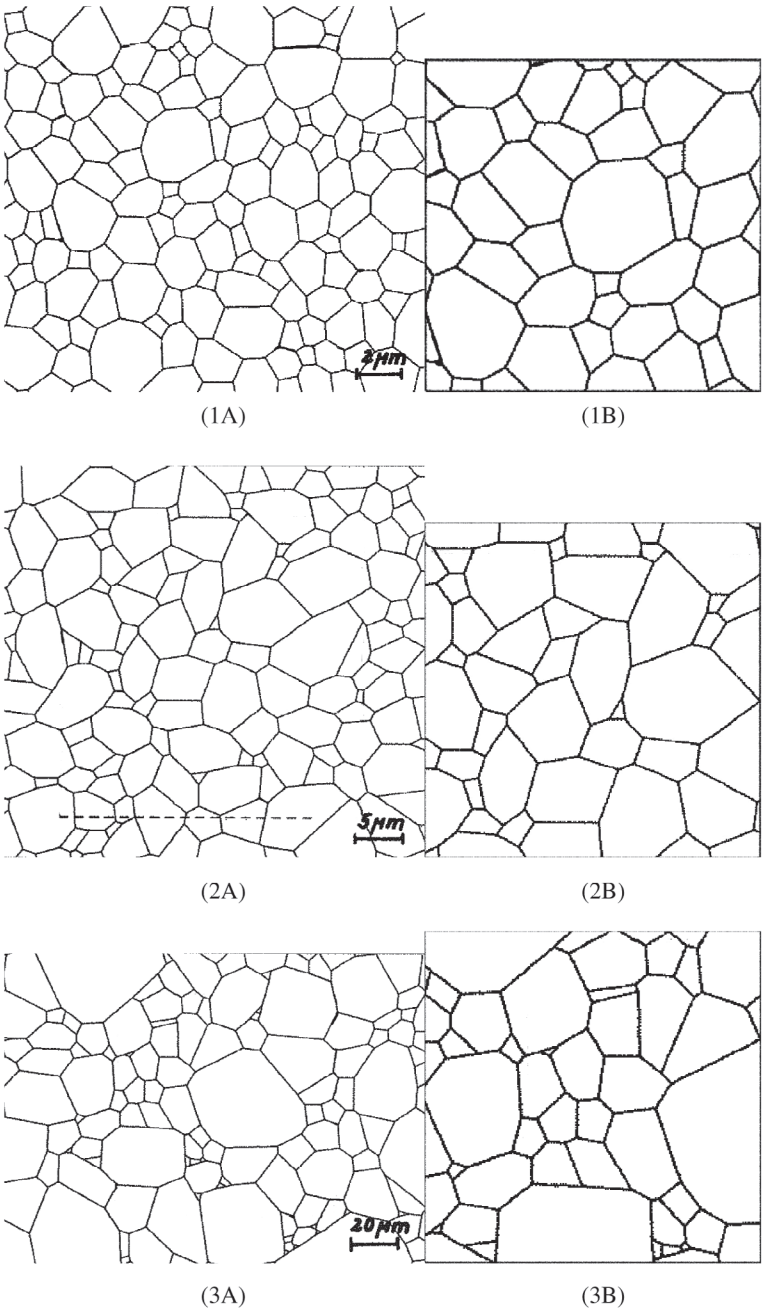


FIGURE 19 Schemes as planar tessellations of plane sections of alumina ceramics: Preprocessing (A) Hahn and Lorz[8] and (B) cut of the plane sections with exactly 50 cells

TABLE 6 Values of the different model tests for the schemes of plane sections of alumina ceramics (Figure 19 1a, 2a, 3a)

(a) 1b		(b) 2b		(c) 3b	
c	0.848 (0.073)	c	0.959 (0.002)	c	1.492 (0)
d	0.078 (0.710)	d	0.121 (0.172)	d	0.168 (0.006)
l_0	0.058 (0)	l_0	0.057 (0)	l_0	0.062 (0)
l_1	0.019 (0)	l_1	0.028 (0)	l_1	0.018 (0)

TABLE 7 Values of the different model tests for the cuts of the sections of alumina ceramics with exactly 50 cells (Figure 19 1b, 2b, 3b)

(a) 1b		(b) 2b		(c) 3b	
c	0.931 (0.004)	c	1.002 (0.0002)	c	1.328 (0)
d	0.154 (0.014)	d	0.172 (0.004)	d	0.248 (0)
l_0	0.077 (0)	l_0	0.116 (0)	l_0	0.137 (0)
l_1	0.041 (0)	l_1	0.024 (0)	l_1	0.009 (0)

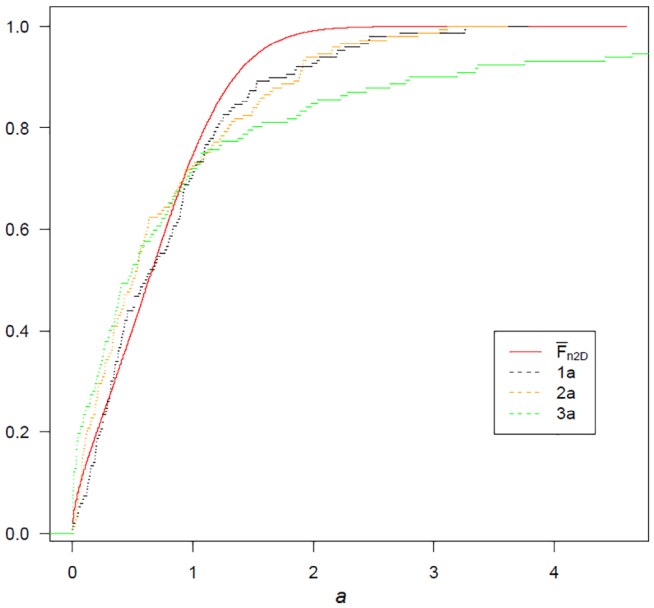


FIGURE 20 Cumulative distribution function comparison of the cells area of the schemes of plane sections of alumina ceramics (Figure 19 1a black line, 2a yellow line, 3a green line) and of the two-dimension sectional Poisson-Voronoi cells area (red line) [Colour figure can be viewed at wileyonlinelibrary.com]

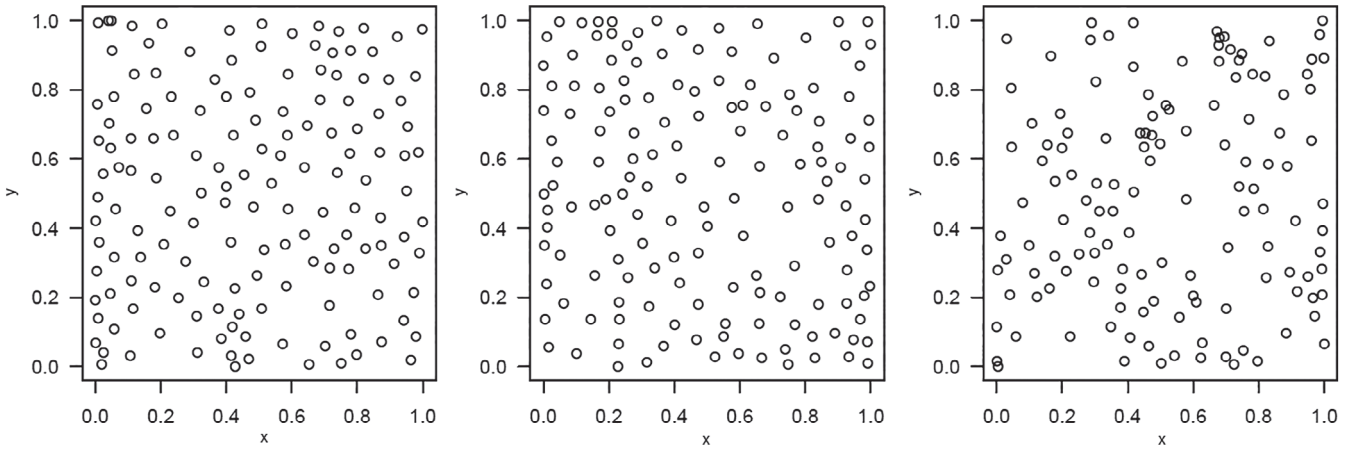


FIGURE 21 From left to right centers of mass of the schemes of plane sections of alumina ceramics (Figure 19 1a, 2a, 3a)

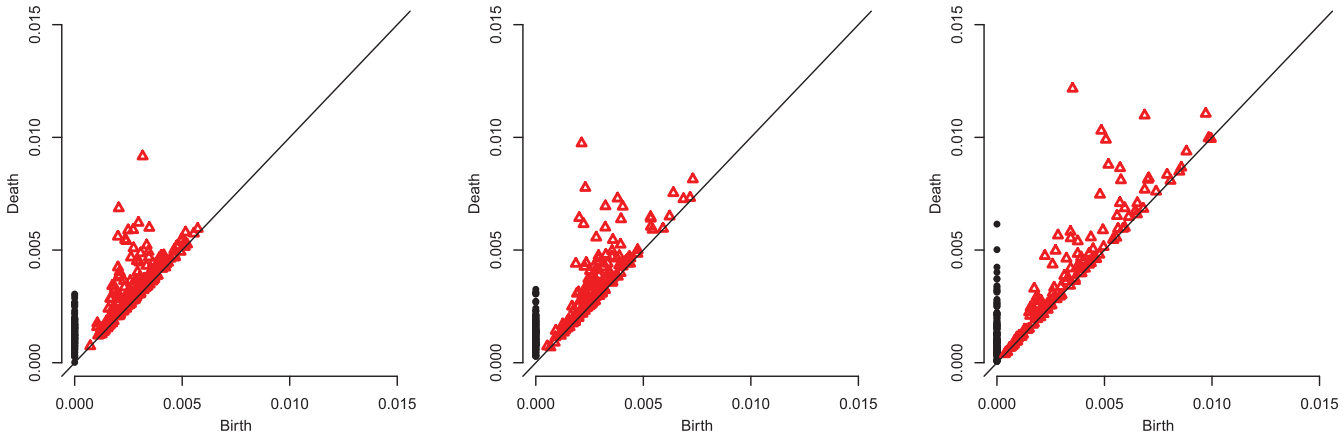


FIGURE 22 From left to right persistence diagrams of the centers of mass of the schemes of plane sections of alumina ceramics (Figure 19 1a, 2a, 3a) [Colour figure can be viewed at wileyonlinelibrary.com]

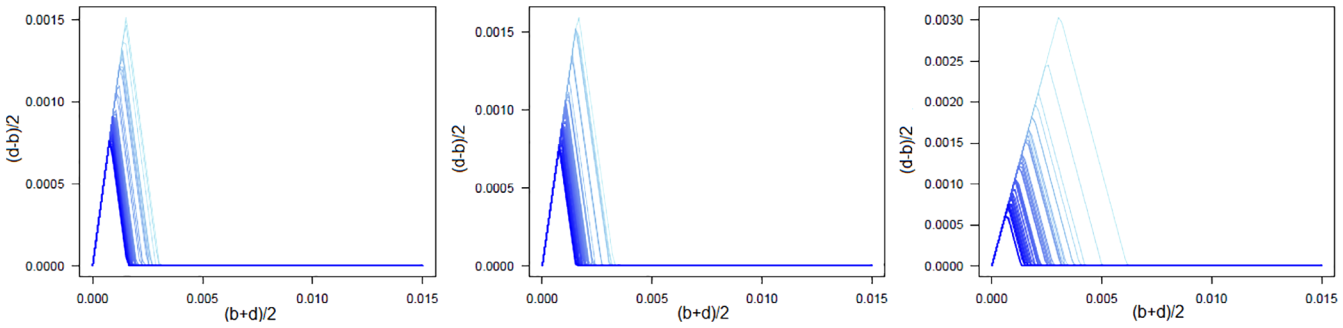


FIGURE 23 From left to right persistence landscapes (connected components) of the schemes of plane sections of alumina ceramics (Figure 19 1a, 2a, 3a) [Colour figure can be viewed at wileyonlinelibrary.com]

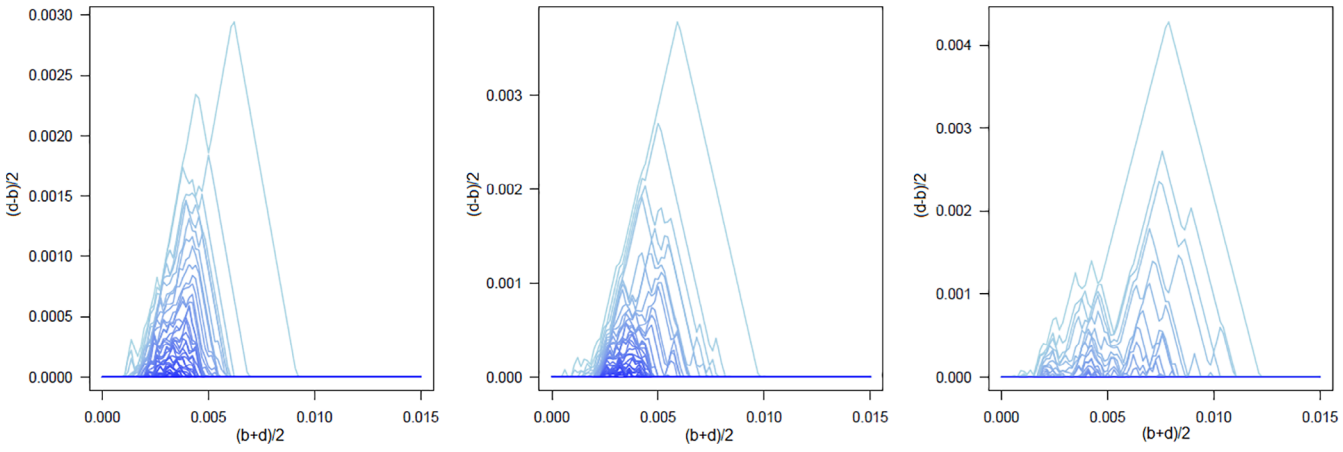


FIGURE 24 From left to right persistence landscapes (holes) of the schemes of plane sections of alumina ceramics (Figure 19 1a, 2a, 3a) [Colour figure can be viewed at wileyonlinelibrary.com]

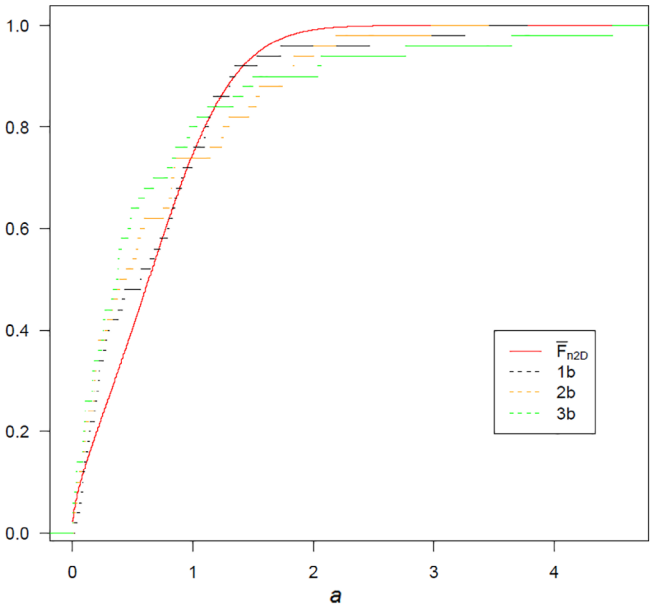


FIGURE 25 Cumulative distribution function comparison of the cuts of the sections of alumina ceramics with exactly 50 cells (Figure 19 1b black line, 2b yellow line, 3b green line) and of the two-dimension sectional Poisson-Voronoi cells area conditioned on $N_{2D} = 50$ (red line) [Colour figure can be viewed at wileyonlinelibrary.com]

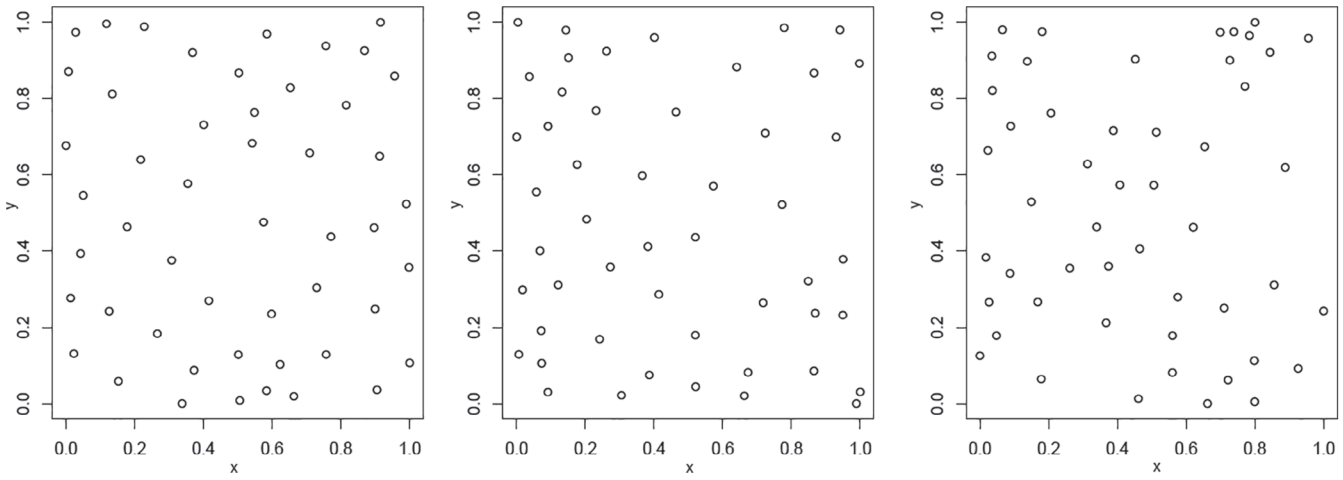


FIGURE 26 From left to right centers of mass of the cuts of the sections of alumina ceramics with exactly 50 cells (Figure 19 1b, 2b, 3b)

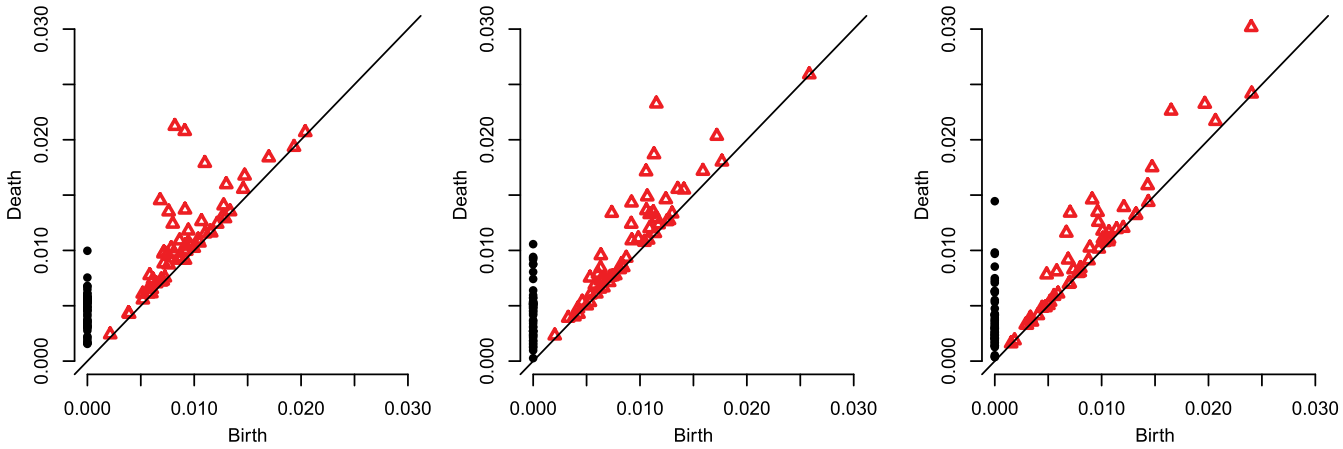


FIGURE 27 From left to right persistence diagrams of the centers of mass of the cuts of the sections of alumina ceramics with exactly 50 cells (Figure 19 1b, 2b, 3b) [Colour figure can be viewed at wileyonlinelibrary.com]

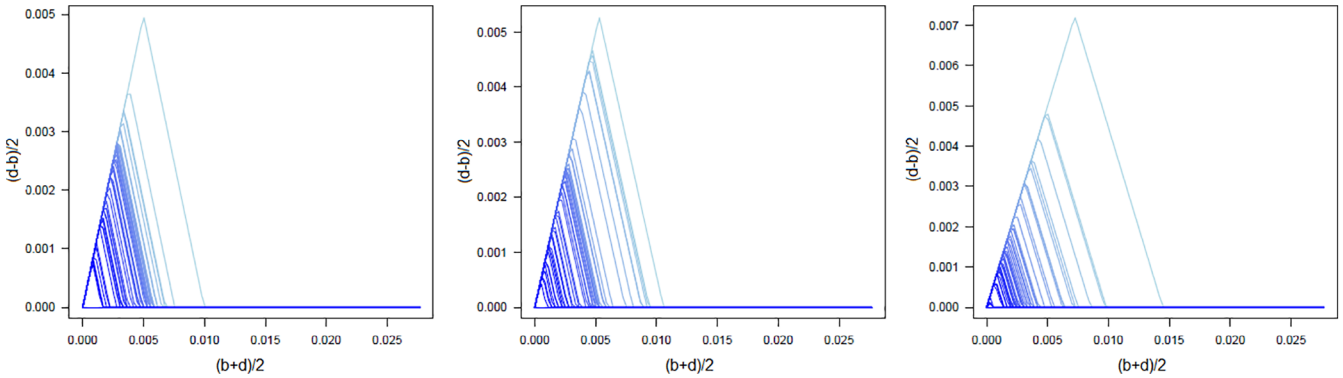


FIGURE 28 From left to right persistence landscapes (connected components) of the cuts of the sections of alumina ceramics with exactly 50 cells (Figure 19 1b, 2b, 3b) [Colour figure can be viewed at wileyonlinelibrary.com]

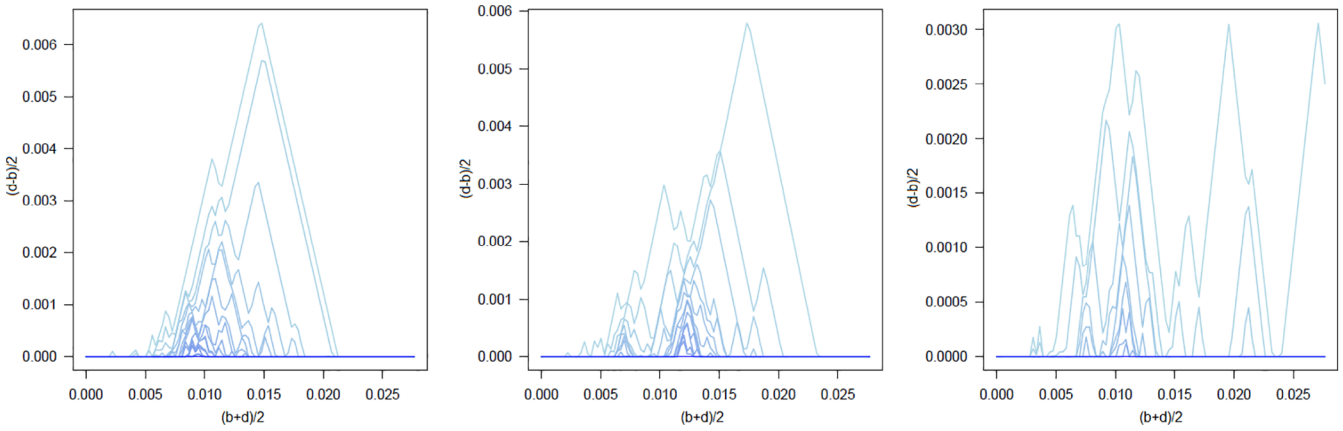


FIGURE 29 From left to right persistence landscapes (holes) of the cuts of the sections of alumina ceramics with exactly 50 cells (Figure 19 1b, 2b, 3b) [Colour figure can be viewed at wileyonlinelibrary.com]

Results using the two different approaches lead to slightly different results regarding the first two images (Figures 19 1a, 1b, 2a, 2b). For the first image, considering all the cells, the coefficient of variation test and the test based on the CDF of cells area suggest that the Poisson-Voronoi model could be reasonably used for approximating alumina ceramics; instead, looking at the cuts, the hypothesis is rejected by both tests. For the second image, the coefficient of variation test based on all the cells is in agreement with the results obtained for the reduced sections; just the test based on the CDF considering all the cells does not reject the Poisson-Voronoi hypothesis. Using tests from persistence approach instead, the use of Poisson-Voronoi model is discouraged in both cases.

8 | RESULTS AND DISCUSSION

This work provides a general setting for testing whether a microstructure is generated by a Poisson-Voronoi diagram, based on a cross-section of the microstructure. Taking inspiration from previous work in this field,[4,8] we widen the testing framework proposing new model tests. In particular, we introduce test statistics using tools coming from different statistical branches like goodness of fit and TDA, that show to be more powerful under the specific alternative hypothesis, 2D Poisson-Voronoi diagram. We consider the situation with periodic boundary conditions, which is popular in materials science applications and without these conditions. Our approach is very general and can also be extended to test hypotheses for more complicated models describing the 3D structure based on a 2D section.

Being able to accept the Poisson-Voronoi model on the basis of 2D real metal sections means having complete probabilistic information on the underlying 3D structure. Furthermore, testing more general model assumption on metal microstructure and accepting the underlying hypothesis will allow to perform mechanical experiments using virtual microstructures avoiding waste of material and discovering new interesting relations between microstructure features and mechanical properties much faster than possible using physical experiments.

Moreover, since as showed in Section 7, Poisson-Voronoi model can be a too simple model for metal microstructure as alumina ceramics, alternative models are required. Hence, future developments involve testing of more general and less understood Voronoi structures for more complicated microstructures, such as multilevel Voronoi diagrams. Another interesting direction is to consider fully data-based approaches for analyzing 2D sections. For instance, analyzing such a section using a persistence landscape does not need the rigid restrictions on the geometry of the cells as present in the Poisson-Voronoi model.

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 Vanessa Robins and Jeroen Spandaw for the inspiring discussion about Persistent Homology and the anonymous reviewers for the useful comments that lead to a new section of the paper.

ORCID


Martina Vittorietti https://orcid.org/0000-0002-7695-3209

REFERENCES

1. Ohser J, Mücklich F. *Statistical Analysis of Microstructures in Materials Science*. Chichester, NH: John Wiley & Sons Ltd; 2000.
2. Okabe A, Boots B, Sugihara K, Chiu SN. *Spatial Tessellations: Concepts and Applications of Voronoi Diagrams*. Vol 501. Oxford, UK: John Wiley & Sons; 2009.
3. Kumar S, Kurtz SK. Simulation of material microstructure using a 3D Voronoi tesselation: calculation of effective thermal expansion coefficient of polycrystalline materials. *Acta Metall Mater*. 1994;42(12):3917-3927.
4. Lorz U, Krawietz R. Stereological model tests for the spatial Poisson-Voronoi tessellation. *Acta Stereologica*. 1991;42(12):3917-3927.
5. Lorz U. Cell-area distributions of planar sections of spatial Voronoi mosaics. *Mater Charact*. 1990;25:297-309.
6. Vittorietti M, Kok PJJ, Sietsma J, Jongbloed G. Accurate representation of the distributions of the 3D Poisson-Voronoi typical cell geometrical features. *Computational Materials Science*. 2019;166:111-118.
7. Hahn U, Lorz U. On the precision of some stereological estimators for the model parameter of the spatial Poisson-Voronoi tessellation. *Acta Stereologica*. 1994;13(1):245-250.
8. Hahn U, Lorz U. Stereological model tests for the spatial Poisson-Voronoi tessellation II. *Acta Stereologica*. 1993;12:131-131.
9. Chiu SN, Van de Weygaert R, Stoyan D. The sectional Poisson Voronoi tessellation is not a Voronoi tessellation. *Adv Appl Probab*. 1996;28(2):356-376.
10. Mecke J. Parametric representation of mean values for stationary random mosaics. *Ser Stat*. 1984;15.3:437-442.
11. Schwertel J, Stamm H. Analysis and modelling of tessellations by means of image analysis methods. *J Microsc*. 1997;186(2):198-209.
12. Jones MC. Simple boundary correction for kernel density estimation. *Stat Comput*. 1993;3(3):135-146.
13. Hermann H, Wendrock H, Stoyan D. Cell-area distributions of planar Voronoi mosaics. *Metallography*. 1989;23(3):189-200.
14. Hatcher A. *Algebraic topology*. Cambridge, MA: Cambridge University Press; 2002.
15. Edelsbrunner H, Harer J. *Computational Topology: An Introduction*. Providence, RI: American Mathematical Society; 2010.
16. Edelsbrunner H, Mücke EP. Three-dimensional alpha shapes. *ACM Trans Graphic (TOG)*. 1994;13.1:43-72.
17. Fasy BT, Lecci F, Rinaldo A, Wasserman L, Balakrishnan S, Singh A. Confidence sets for persistence diagrams. *Ann Stat*. 2014;42(6):2301-2339.
18. Ghrist R. Barcodes: the persistent topology of data. *Bull Am Math Soc*. 2008;45(1):61-75.
19. Robins V, Turner K. Principal component analysis of persistent homology rank functions with case studies of spatial point patterns, sphere packing and colloids. *Phys D: Nonlinear Phenomena*. 2016;334:99-117.
20. Bubenik P. Statistical topological data analysis using persistence landscapes. *J Mach Learn Res*. 2015;16(1):77-102.
21. Chazal F, Fasy BT, Lecci F, Rinaldo A, Wasserman L. Stochastic convergence of persistence landscapes and silhouettes. Paper presented at: Proceedings of the 13th Annual Symposium on Computational Geometry; 2014:474; ACM.
22. Biscio C, Møller J. The accumulated persistence function, a new useful functional summary statistic for topological data analysis, with a view to brain artery trees and spatial point process applications. 2016; arXiv preprint arXiv:1611.00630.

23. Sahai H, Khurshid A. Confidence intervals for the mean of a Poisson distribution: a review. *Biom J*. 1993;35(7):857-867.
24. Loosmore NB, Ford ED. Statistical inference using the G or K point pattern spatial statistics. *Ecology*. 2006;87(8):1925-1931.

How to cite this article: Vittorietti M, Kok PJJ, Sietsma J, Li W, Jongbloed G. General framework for testing Poisson-Voronoi assumption for real microstructures. *Appl Stochastic Models Bus Ind*. 2020;1–24.
https://doi.org/10.1002/asmb.2517