Spatial point pattern methodology for the study of pores 3D patterning in two casting aluminium alloys

Driss El Khoukhi, Nicolas Saintier, Franck Morel, Daniel Bellett, Pierre Osmond, Viet-Duc Le

Version of Record: https://www.sciencedirect.com/science/article/pii/S1044580321002953
Manuscript_2c1b404f7d56db17f4d3b58c54f39b11

Spatial Point Pattern Methodology for the study of pores 3D patterning in two casting aluminium alloys

Driss EL KHOUKHI[1, 2, 3,*], Nicolas SAINTIER[2], Franck MOREL[1], Daniel BELLETT[1], Pierre OSMOND[3], Viet-Duc Le[1]

[1]LAMPA, Arts et Métiers Institute of Technology, 49 035 Angers, Cedex, France
[2]I2M, Arts et Métiers Institute of Technology, 33170 Talence, Cedex, France
[3]Groupe PSA, 78955 Carrières-sous-Poissy, Cedex, France

Abstract

Two cast aluminium alloys fabricated by different casting processes (gravity die-casting and lost foam casting) and showing different degrees of porosity were characterized with X-ray Computed tomography. Information concerning the pore distribution inside the investigated materials is obtained in terms of pore size and the pore positions in 3D space. Subsequently, a spatial point pattern analysis is undertaken to investigate the pore distributions. Different methods, including a nearest neighbor analysis, Ripley's K-function and Clark-Evans tests developed for 3D applications, are used to analyse the observed patterns. The results show that the Homogeneous Poisson process, which provides the Complete Spatial Randomness (CSR) is suitable to approximate the spatial distribution of the pores present in the investigated alloys. Synthetic microstructures that mimic key macroscale features of the materials in terms of pore size and the 3D spatial distribution of the pores were generated. These microstructures can be used in the probabilistic modelling of fatigue behaviour.

Keywords: Casting process, X-ray Computed Tomography, Porosity, Point pattern, spatial statistics, Ripley's K-function, nearest neighbor function, Poisson process, Fatigue.

1. Introduction, context and objectives

In terms of fatigue failure, many authors have shown that surface porosity is the most important casting defect that affects fatigue life. (Ammar et al 2008) showed that porosity acts as the most likely site for crack initiation as 92% of all tested specimens fractured from surface or near-surface porosity site (Figure 1).

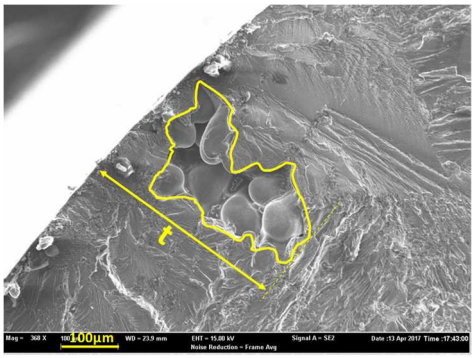

Figure 1: A critical pore located at the surface of a specimen (El Khoukhi et al. 2019)

Moreover, in the literature, many studies (Ammar et al. 2008 ; Le et al. 2016 ; El Khoukhi et al. 2019 ; Murakami 1991) have emphasized the importance of the distance between the critical defect and the free surface. Figure 2 highlights the importance of this parameter. This figure shows a specimen loaded in the High Cycle Fatigue regime, in which a very large defect can be seen in the very middle of the specimen (√Area=575 µm) and a smaller one is visible near the surface (√Area=367 µm). SEM observations clearly show that the smaller pore close to the surface is responsible for fatigue crack initiation and failure.

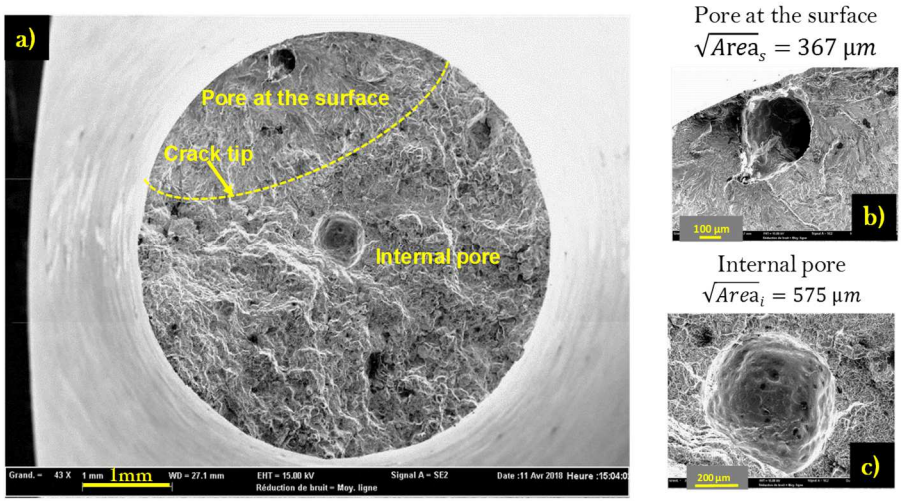

Figure 2: Fatigue failure surface for a specimen taken from cast aluminum alloy (AlSi7Mg03 - T7), has failed at Sa= 55 MPa and Nf=919402 cycles for a stress ratio R=0.1. (a) Macroscopic view, detailed view of (b) surface pore and (c) internal pore. It shows the importance of the position of a defect relative to the free surface, in terms of fatigue failure (El Khoukhi et al. 2019)

Furthermore, experimental investigations have shown that the characteristics of the defect distribution have a major impact on the sensitivity of the fatigue behaviour to the size effect and the stress gradient effect as well as the scatter in the fatigue strength (El Khoukhi et al. 2019).

(Zhu et al. 2007) have developed a probabilistic fatigue model to establish the relationship between the porosity population and the resulting fatigue strength of a cast aluminum alloy. The predictions show that, when the fatigue life is controlled by the porosity population, the mean and standard deviation of the fatigue strength decrease with increasing mean pore size, pore size standard deviation, and porosity density. In addition, the specimen size and shape are shown to influence the fatigue strength by affecting the number of pores and the probability of pores intersecting with the specimen surface within the stressed volume.

The same authors, in (Yi et al. 2007) have extended this approach to examine how specific changes to the nature of the porosity population, such as mean pore size, pore size standard deviation, porosity density, and the volume of the specimen gauge section, affect the predicted fatigue behavior of the alloy. In their model, the authors have supposed that casting pores are randomly distributed in each specimen, and that the pore sizes follow a lognormal distribution. However, in real materials the population of defects can exhibit 3D pattering and may not follow the lognormal distribution in terms of their size.

From these studies, it can be concluded that the variability in the fatigue strength is linked to the variability in microstructural heterogeneities (i.e. size, spatial distribution, shape, density…). Therefore, in the development of probabilistic models and for an accurate prediction of the resulting variability in the fatigue strength, the precise characterization of the defect population is needed in order to create numerical microstructures that ideally mimic key macroscale features of the material in terms of the size distribution and the 3D spatial distribution of defect.

3D characterisation of defects is also important for Additive Manufacturing emerging technologies and other casting process such as high-pressure die-casting. This type of characterization can help understand the links between the process parameters and the porosity of the resulting parts (Chen et al. 2014).

In the literature, many authors have investigated the characterisation of the size and the shape of defect populations (Guerchais et al. 2017; Ben Ahmed et al. 2019; Billaudeau, et al. 2004; Murakami

1991), but relatively little work has focused on characterizing the spatial distribution of defects, especially in 3D space (Wilson 2017). Typically, to take in account the spatial distribution of defects, the density, defined as the number of defects per unit volume, is used. This is a first order parameter which is not capable of describing the defect pattern in the material (i.e. clustering, randomness, etc.). Figure 3 and Figure 4 highlight the fact that two alloys with the same defect density but with different spatial arrangements: (a) clustering (b) random, can lead to different behaviour in terms of high cycle fatigue.

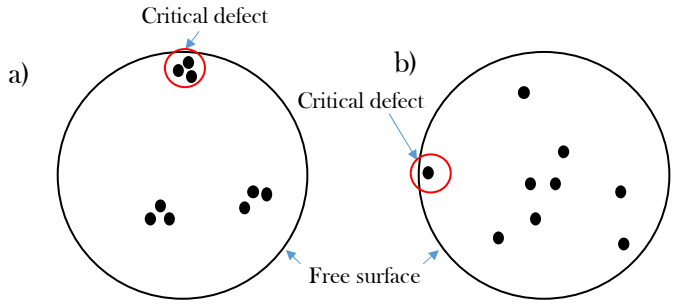

Figure 3: Two samples with the same defect density but with different defect patterns (a) clustering, and (b) random.

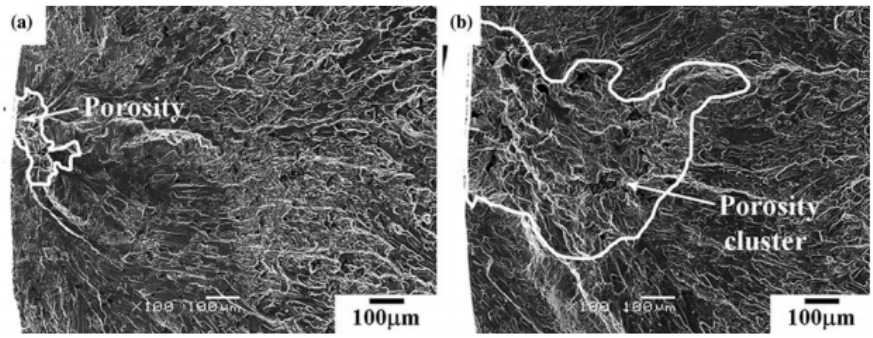

Figure 4: SEM fractography of 319-T7 specimens fatigued at the same maximum stress of 100 MPa but with different fatigue lives of (a) 413 495 and (b) 189 221 cycles (Jang et al. 2009)

For these reasons, it is concluded that the 3D characterisation of the spatial defect distribution inside the material is particularly important in terms of fatigue response. The rapid development of X-ray Computed Tomography makes it possible to access to this information and more, such as the size and shape of individual defects.

As discussed by (Ben Ahmed et al. 2019) the presence of micro-shrinkage porosity in cast materials is directly related to the solidification process and in particular the change in volume of the material which occurs as the material changes state and cools. This is a physical process that is highly dependent on the cooling rate, which in a complex real component is not constant in all zones of the part.

On the other hand, different approaches have been proposed in the literature to predict the appearance of defects in the casting process (Stefanescu 2005). For example, a multiphase (solid, liquid, and porosity) model that predicts melt pressure, feeding flow, and porosity formation and growth during solidification has been developed by (Blair et al. 2005). This model can predict the location in a casting, the defect volume fraction, and the size (diameter) of individual pores. It could be interesting

to characterize their spatial arrangement. This information can be useful to enhance the parameters of the process to avoid the appearance of the clusters of pores in some locations in the parts.

For those reasons, in the present work, we will present the application of certain techniques that can be used to characterize the 3D spatial distribution of defects contained in cast alloys. These techniques have been developed in the framework of the Point Process Theory (Diggle 1983) and are often used in other disciplines such as forestry (Perrin et al. 2006), plant ecology (Illian et al. 2007) as well as astronomy (Babu and Feigelson 1996). However, very few researchers have applied it to the characterization of the spatial distribution of material defects (Wilson 2017; Wilson et al. 2019). The ultimate goal of this work is to propose a high cycle fatigue model that takes into account the defect size and spatial distributions within a loaded volume.

2. Scope of the present work

The main objective of this paper is to characterize the spatial distribution of micro-shrinkage and gas porosity in two cast aluminium alloys using the Point Pattern Process theory. This characterization tells us whether the defects are aggregated (i.e. clustered) or dispersed in comparison to a Complete State of Randomness described by a Homogeneous Poisson process and over which spatial scales do patterns exist. With this information it will be possible to create numerical microstructures that ideally mimic the features of the material with respect to the size and spatial distribution of the defects present in the alloys. Note that in this work it is assumed that the defect size distribution and the 3D spatial distributions are not coupled. This hypothesis was verified by (Wilson 2017; Wilson et al. 2019) in an alloy of the same kind.

The characterization of the defect size has been presented in a previous publication (El Khoukhi et al. 2018), and it is generally accepted that the effect of the shape of a defect does not have a strong influence on its fatigue behaviour (Murakami 1991). Hence, the focus of this paper is the characterisation of the 3D spatial defect distribution.

This step is mandatory for future modelling work to predict the fatigue resistance of components made from these alloys by coupling the defect size distribution, the spatial distribution as well as the loaded volume. This modelling work will help understand how the scatter in the fatigue strength is linked to the microstructural variability of the material, such as the variability in the defect size and the way the defects are arranged in 3D space (clustering, regularity…).

As already mentioned, to characterize the 3D spatial defect distribution, the techniques developed in the framework of Point Process Theory are used. In the following sections, these techniques are presented, as well as their application to the investigated alloys.

3. Stat of Art, theory and definitions

According to (Diggle 1983), a Spatial Point Pattern is a set of locations, or events, within a specified region. The events are irregularly placed and are modelled as the result of an unknown underlying stochastic Process, referred to as a Spatial Point Process. We can think of the spatial distribution of pores as the result of one such process. Point processes are a well-studied domain in probability theory and the subject of powerful tools in statistics for modelling and analysing spatial data, which is of interest in such diverse disciplines as forestry, plant ecology, astronomy and many more. As discussed by (Diggle 1983), spatial point patterns can be divided into three main categories (Figure 5):

- **Random**: or Complete Spatial Randomness (CSR) where point or events are distributed randomly. This state is provided by the Poisson Process.

- **Clustering**: if the points or events attract each other and appear in small groups called clusters.
- **Regularity**: if the points or events repulse each other; no two points are close to each other.

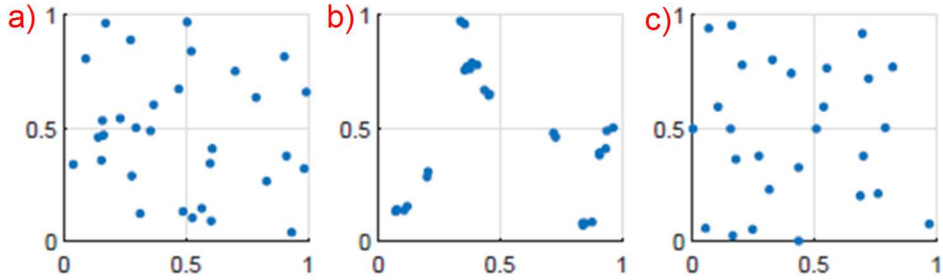

Figure 5: Spatial Pattern categories (a) Complete Spatial Randomness, (b) clustering (d) regular process.

Point Pattern Analysis (PPA) is the study of the spatial arrangements of points in space (3 dimensions in our case). The simplest formulation is a set $S = \{x \in D\}$ where D, which can be called the 'study volume', is a subset of $\mathbf{R}^3$, a 3-dimensional Euclidean space.

When exploring the properties of an unknown spatial distribution of defects contained in a material, we can look at the density ρ, which can be estimated as the average number of pores per unit volume. A spatial distribution also can be characterized by its second-order properties, which describe how events or point are placed in relation to each other.

(Bailey and Gatrell 1995) have classified these techniques as Exploratory Analyses of the first order properties and the second order properties of a process.

The exploratory analyses of **first order properties** of a process aim to analyse how the density, of defects in a material for example, varies over an area or a volume. This category includes Quadrat Analyses and Kernel Estimation.

On the other hand, exploratory analyses of **second order properties** of a process estimate the presence of spatial dependence among events (pores) based on the distances from one another in comparison to the Poisson process. In this category, different techniques can be used:

- The Nearest-Neighbor distance distribution function (G-Function)
- The Empty space function of the point process (F-Function)
- The K-Ripley Function (Ripley 1977)

Furthermore, statistical test can be conducted to obtain quantitative information on how the studied process differs from the Poisson process. The Clark-Evans (CE) test is usually used (Clark and Evans 1954). This test aims to calculate the statistical significance level through comparison with the null hypothesis of complete spatial randomness (CSR) provided by the homogeneous Poisson process. In other words, it tends to measure the distance of the investigated process from the Poisson process.

We used the exploratory analyses of second order properties of the patterns to determine how random the defect population is. Then the Clark-Evans test was used to measure how far it is from the Poisson process.

In the following sections, the detailed definitions of the different techniques used to analyse the spatial distribution of defects contained in two casting aluminium alloys, will be presented. At the simplest level, the data being analysed consist only of the coordinate locations of the events (i.e. the pores). Each event or pore could also have attributes or variables associated with them like their size or shape.

Since the analyses of second order properties of a process aims to compare it to the Poisson process, firstly we start by defining the Poisson process.

3.1. **The Poisson Process**

A homogeneous Poisson process (or Poisson point process) is characterized by:

i. the number of points in a subset V from R^3 follows a Poisson distribution of parameter $\rho|V|$, where $|V|$ is the volume of V:

$$P\{N(V) = k\} = \frac{(\rho|V|)^k}{k!} e^{-\rho|V|} \quad \text{(eq. 1)}$$

ii. $\rho > 0$ is the point process density, which corresponds to the average number of points per volume unit.

iii. For m disjoints sets $V_1, \ldots V_m$, the random variables $N(V)_1, ..., N(V)_m$ are independent.

The Poisson point process is used to produce Complete Spatial Randomness (CSR): each point is stochastically independent, and there is absolutely no interaction between them (see Figure 5). Therefore, the Poisson process serves as a reference to evaluate if a point process is clustered (i.e. the points are attracted to each other (see Figure 5 (b)) or regular (i.e. the points repulse each other, see Figure 5 (c)).

3.2. **Techniques based on distances to the Nearest-Neighbor (NN-distances)**

This category includes the G, F and J functions that are commonly used to compare two sets of spatial patterns, or one pattern against the null hypothesis of complete spatial randomness (CSR). These functions are defined as follows.

G Function: nearest neighbor distance distribution function

The nearest neighbor distance distribution function of a point process S is the cumulative distribution function G of the distance from a typical random point of S to its nearest neighboring point (also of S). It is estimated by:

$$G(r) = \frac{1}{N}\sum_{i=1}^{N} I(d_i \leq r) \quad \text{(eq. 2)}$$

where d_i is the Euclidian distance between pore i and its nearest neighbor, and N is the number of pores. $I(\cdot)$ is the indicator function, which is one if the argument is true, else zero.

The estimate of G is then compared to the true value of G for the Poisson Process in 3D, defined by:

$$G(r) = 1 - e^{\frac{-4}{3}\rho\pi r^3} \quad \text{(eq. 3)}$$

where ρ is the density (expected number of points per unit area).

Deviations between the empirical and theoretical G curves may suggest spatial clustering or spatial regularity.

F Function: empty space function of the point process

The F function, also called the "spherical contact distribution" of a stationary point process, in contrast to the G function plots the cumulative frequency distribution for the distance of each grid point to its nearest pore when a regularly spaced grid of g points is superimposed upon the population. It is estimated by:

$$F(r) = \frac{1}{g}\sum_{i=1}^{g} I(z_i \leq r) \quad \text{(eq. 4)}$$

where zi is the distance of grid point i to the nearest pore. The F function is useful for detecting non-homogeneities in the distribution of pores within the field. In addition, the estimate of F is a useful statistic summarizing the gap sizes in the pattern. Like for the G function, the estimate of F is compared to the true value of F for a Poisson process in 3D, which is:

$$F(r) = 1 - e^{\frac{-4}{3}\rho\pi r^3} \quad \text{(eq. 5)}$$

Deviations between the empirical and theoretical F curves may suggest spatial clustering or spatial regularity.

J-Function

The J-function (Baddeley et Lieshout 1995) of a stationary point process is defined as:

$$J(r) = \frac{1-G(r)}{1-F(r)} \quad \text{(eq. 6)}$$

This function is used to compare a given distribution to a Poisson process. If :

- J(r) > 1: the point process tends to be regular
- J(r) = 1: the process is a Poisson process
- J(r) < 1: the point process tends to be clustered

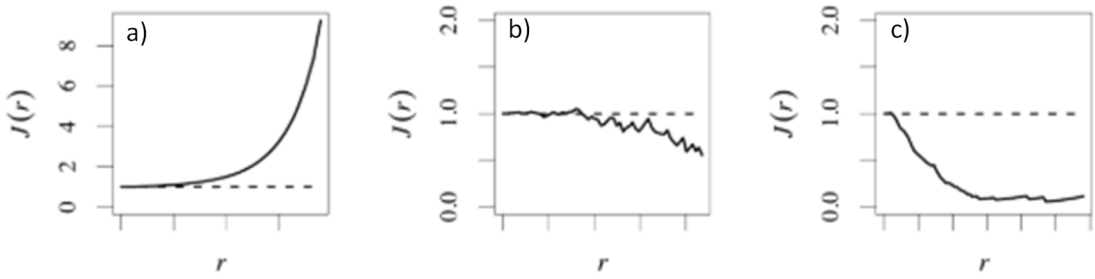

Figure 6**:** schematization des three cases of J-Function, (a) regularity, (b) Poisson process and (c) clustering (Baddeley et al. 2015).

3.3. **The K-Ripley Function**

While these three functions (G, F and J) provide qualitative information about the spatial distribution, the K-Ripley function (Ripley 1977) is often chosen to infer statistical properties. It is a more descriptive point pattern measure. Whereas the G-function and the F-function rely only on the nearest neighbor distances to describe a point process, the K-Function (also called “Ripley's K-function” or the “K-Ripley function” or the “reduced second moment function”) uses all distances between point pairs in the pattern [Figure 7]. The K-Function of point process S is determined by the second order moment properties of S. It is determined as follows:

1. Construct a circle of radius *d* around each point (event) *i*
2. Count the number of other events, labeled *j*, that fall inside this circle
3. Repeat these first two steps for all points *i*, and
4. Sum the results.

These steps equate to:

$$K(d) = \frac{|V|}{N}\sum_{i=1}^{N}\sum_{i\neq j}^{N}\frac{I(r_{ij}\leq d)}{N} \quad \text{(eq. 8)}$$

where, N is the number of events (pores) in the studied region, |V| its volume and r_{ij} is the distance between pore i and pore j. I(.) is the indicator function. $I(r_{ij})=1$ if the distance, r_{ij}, from *i* to *j* is less then *d*. *O*therwise $I(r_{ij})=0$.

5. Increase *d* by a small fixed amount
6. Repeat the calculation, giving values of K(*d*) for a set of distances, *d*

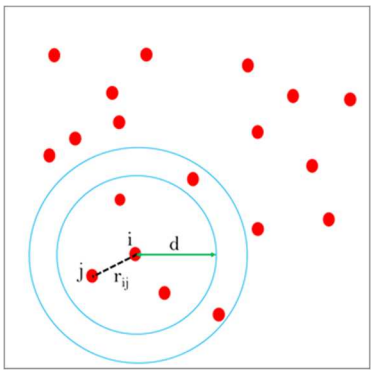

Figure 7: Representation of the steps used to construct Ripley's K function in 2D.

The estimate of K is a useful statistic, summarizing aspects of inter-point "dependence" and "clustering". It is usually compared to the theoretical value of K for a Poisson point process. In 3D, Ripley's K function for a Poisson process is exactly the volume of the sphere with radius d (since the K function is normalized by 1/ρ):

$$K_{CSR}(d) = \frac{4}{3}\pi d^3 \qquad \text{(eq. 10)}$$

Like for the precedent estimators, deviations between the empirical and theoretical K curves may suggest spatial clustering or spatial regularity.

Therefore, if:

- $K(d) > K_{Poisson}(d)$: the point process is clustered
- $K(d) < K_{Poisson}(d)$: the point process is regular

The K-Function describes the characteristics of a point processes at many distance scales. The G and F nearest-neighbor distance functions do not have this property.

All of these methods give qualitative information about the pattern; if it is clustered, regular or random. However, as most real patterns will almost never be perfectly random, even if the population is random, a way is needed to determine how far a process really is from the Poisson process. In order to evaluate the degree of clustering or randomness compared to the Poisson process with a certain level of confidence (ex. 1%, 5% or 10%) a statistical test should be used. The Clark-Evans test is one possibility. There are two versions of this test: (i) The one-tailed version used if the pattern is known to be clustered (or regular) and is compared to the Poisson process. (ii) The two-tailed version used when it is unknown if the process is clustered or regular.

In general, after using either the 1st or the 2nd order techniques, we can tell if the pattern is clustered or regular. Therefore, in the following only the one-tailed version of the Clark-Evans test is used.

3.4. **Clark-Evans test**

The "Clark and Evans" test (1954) calculates for each pore the distances from the nearest neighbor. The average value of these distances $\overline{D}_m$ can then be compared to the theoretical value μ calculated for a null hypothesis (i.e. the Poisson process) in order to characterize the spatial structure: clustering, regularity, etc. The standardized sample mean under the CSR Hypothesis is defined by:

$$Z_m = \frac{\overline{D}_m - \mu}{\sigma} \quad \text{(eq. 11)}$$

It follows from the Central Limit Theorem that independent sums of identically distributed random variables are approximately normally distributed. Hence, the most common test of the CSR Hypothesis based on nearest neighbors involves a normal approximation to the sample mean of D, as defined by:

$$\overline{D}_m = \frac{1}{m}\sum_{i=1}^{m} D_i \quad \text{(eq. 12)}$$

To construct this normal approximation, the mean and variance of the Poisson distribution in 3D are respectively given by:

$$\mu = E(X) = \left(\frac{4\rho\pi}{3}\right)^{\frac{-1}{3}} \Gamma(4/3) \quad \text{(eq. 13)}$$

$$\sigma = \sqrt{\frac{\Gamma(5/3) - \left(\Gamma(4/3)\right)^2}{\sqrt[3]{\left(\frac{4\rho\pi}{3}\right)^2}}} \quad \text{(eq. 14)}$$

where Γ is the gamma function.
But from the Central Limit Theorem it then follows that for sufficiently large sample sizes (m>30), D_m must be approximately normally distributed under the CSR Hypothesis with the mean and variance of the Poisson Process, so:

$$\overline{D}_m \sim N\left[\left(\frac{4\rho\pi}{3}\right)^{\frac{-1}{3}} \Gamma(4/3), \frac{\Gamma(5/3) - \left(\Gamma(4/3)\right)^2}{\sqrt[3]{\left(\frac{4\rho\pi}{3}\right)^2}}\right] \quad \text{(eq. 15)}$$

Hence, if we now denote the standardized sample mean under the CSR Hypothesis by:

$$Z_m = \frac{\overline{D}_m - \mu}{\sigma} = \frac{\left(\frac{4\rho\pi}{3}\right)^{\frac{-1}{3}} \overline{D}_m - \Gamma(4/3)}{\sqrt{\Gamma(5/3) - \left(\Gamma(4/3)\right)^2}} \quad \text{(eq. 16)}$$

Then Z_m follows the standard normal distribution,

$$Z_m \sim N(0,1) \quad \text{(eq. 17)}$$

If the CSR Hypothesis is true, then Z_m should be a sample from N(0,1).

One-tailed version

By definition, the level of significance of a test is the probability, α, that the null hypothesis is rejected when it is actually true. And the *upper-tail points*, z_α, for the standard normal distribution is defined by:

$$Pr(Z \geq Z_\alpha) = \alpha \quad \text{For} \quad Z \sim N(0, 1) \qquad \text{(eq. 18)}$$

It is important to highlight that there is no "best" choice for α. The typical values given by most statistical tests (Smith 2020) from the literature are listed in Table 1 below:

Table 1: One-tailed significance

Significance	**α**	$\mathbf{Z_\alpha}$
Strong	0.01	2.33
Standard	0.05	1.65
Weak	0.10	1.28

Using the one-tailed significance presented in Table 1:

- the non-randomness of a given pattern is considered "strongly" significant if the CSR Hypothesis can be rejected at the α = 0.01 level of significance.
- the non-randomness of a given pattern is considered "weakly" significant if the CSR Hypothesis can be rejected at the α = 0.10 level of significance.

The value α =0.05 is regarded as a standard (default) value indicating "significance". We mention that in the case of clustering we use ($-\mathbf{Z_\alpha}$) instead of ($\mathbf{Z_\alpha}$) in Table 1 (see Figure 8).

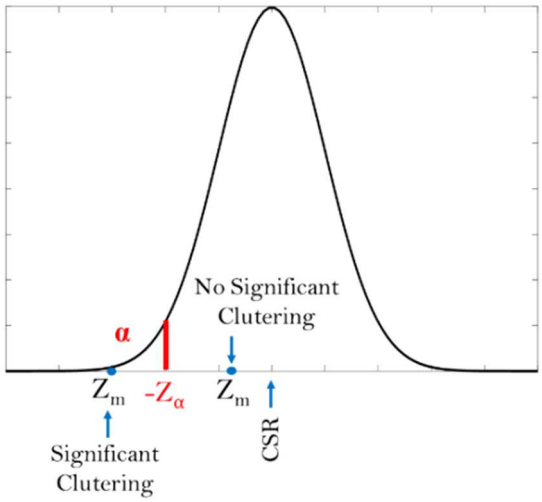

Figure 8**:** One-Tailed Test of Clustering.

P-Values for Test

Another approach is often adopted in evaluating test results. The idea is that, in the one-tailed test of clustering versus CSR discussed above, suppose that for the observed standardized mean value, Z_m, one simply asks how likely it would be to obtain a value this low if the CSR Hypothesis were true? This question is answered by calculating the probability of a sample value as low as Z_m for the standard normal distribution N(0,1). This probability, called the P-value of the test, is given by

$$P(Z \leq Z_m) = \phi(Z_m) \qquad \text{(eq. 19)}$$

where ϕ the probability density function for the normal distribution.

Taking in account the dependencies between nearest neighbor distances

However, one major difficulty with the precedent techniques is that we have used the entire point pattern (m=N) and have thus ignored the obviously dependencies between NN-distances. (Cressie 1993) calls this "intensive" sampling and shows with simulation analyses that this procedure tends to overestimate the significance of clustering (or regularity). The study of the dependencies between nearest neighbor distances is discussed in the following sections.

4. Materials – the investigated cast aluminum alloys

In this work, two cast aluminium alloys referred to as Alloy A (AlSi7Cu05Mg03) obtained using gravity die casting followed by a standard T7 heat treatment and Alloy B (AlSi7Mg03) obtained by lost foam casting followed by T7 heat treatment are used. These alloys have been largely studied and characterised in previous works by the present research group (El Khoukhi et al. 2019), (Le et al. 2016) and (Koutiri et al. 2013). The different manufacturing processes result in different porosity populations (volume fraction, defect size, defect spatial distribution…[see Figure 9].) and it should be kept in mind that alloy A contains an additional 0.5 wt% copper, compared to alloys B. The presence of copper results in a higher micro-hardness of the alpha phase.

Table 2, Table 3 and Table 4 summarize the mechanical and microstructural properties of these alloys.

Table 2: typical composition of the alloy A [AlSi7Cu0.5Mg0.3]

Element	Si	Cu	Mg	Zn	Mn	Ni	Ti	Pb	Fe	Sn
% in weight	6,5-7,5	0,4-0,6	0,28-0,4	<0,10	<0,10	<0,05	0,08-0,2	<0,05	<0,20	<0,05

Table 3: typical composition of the alloy B [AlSi7Mg0.3]

Element	Si	Cu	Mg	Zn	Mn	Ni	Ti	Pb	Fe	Sn
% in weight	6,5-7,5	<0,10	0,28-0,4	<0,10	<0,10	<0,05	0,08-0,2	<0,05	<0,20	<0,05

Table 4: Properties of the investigated cast Al-Si alloys

Grade	Alloy A	Alloy B
Designation	AlSi7Cu05Mg03 - T7	AlSi7Mg03 - T7
Casting Process	Gravity Die	Lost Foam
Young Modulus E (GPa)	77±6	68±5
Yield stress $\sigma_{Y0.2\%}$ (MPa)	260±2	240±5
Ultimate tensile strength σ_u (MPa)	304±4	251±6
Void fraction (%)	0.03	0.28

In a previous publication (El Khoukhi et al. 2019) presented the effect of microstructural heterogeneities (SDAS, pore size, oxides …) on the fatigue behavior of these materials in terms of fatigue failure mechanisms, the size effect and the stress gradient effect as well as the scatter in the fatigue strength. As the main objective of this paper is to analyse the 3D spatial distribution of pores that exists in these alloys, the following section shows the data concerning the micro-porosity obtained

from CT X-Ray tomography scans and the volume-rendering analyses done using the AVIZO software of samples extracted from the studied alloys.

In these analyses, only defect sizes larger than 18 μm^3 are taken into account. This decision was made due to experimental observations indicating that defects smaller than this value are harmless with regard to fatigue.

Porosity

In order to characterize the defect size and spatial distributions in these alloys, Computed-Tomography (CT) analyses were undertaken. The CT scans were done by the MATEIS laboratory at INSA Lyon with a resolution of 8 µm/voxel. The AVIZO software was then used in order to analyze the raw data. An inspection volume of 363 mm^3 was used. Note that the CT scan resolution can influence the precision of the defect size measurements, in particularly for small defects.

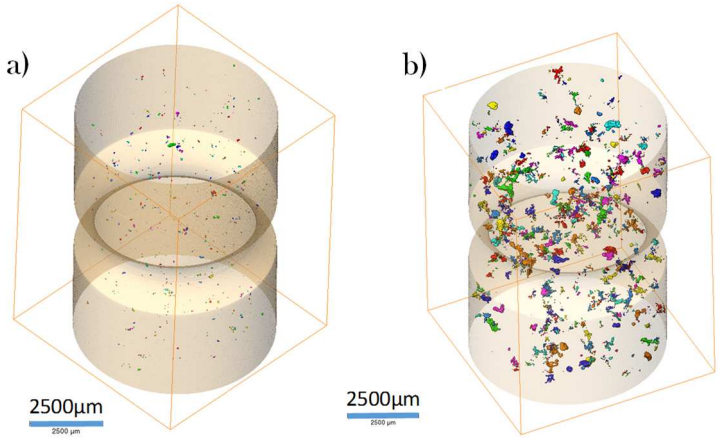

Figure 9: Micro-tomography scans of (a) Alloy A and (b) Alloy B

Figure 9 illustrates the defect populations within two notched specimens. For more detail on the use of the notched specimen see (El Khoukhi et al. 2019). It can be clearly seen that Alloy B has considerably larger pores when compared to Alloy A. This is to be expected as Alloy B was obtained by the lost foam casting process. From these two scans, only the scan corresponding to BVN1-05 is used in the Point Process analyses.

In addition, further measurements using the X-ray micro-tomography technique were conducted by (V. D. Le 2016) at the PLACAMAT (attached to the University of Bordeaux and CNRS), in France on a X GE V/TOME/SX micro-tomograph. A spatial resolution of 5 µm×5 µm×5 µm was used with a 160 kV (max) X-ray source. This data was also reconstructed and analyzed using the Avizo® software. In these measurements, 9 samples from Alloy B were scanned. Adding the sample BVN1-05, a total of 10 scanned samples for the alloy B are considered, which corresponds to a total scanned volume of 2.2×10^3 mm^3. For Alloy A, 4 samples were scanned corresponding to a total volume of 0.9×10^3 mm^3.

5. Statistical pattern analyses

The following sections show the application of the Point Pattern Process techniques discussed above to the data obtained from the CT X-Ray tomography for the 3D pore distributions in the cast alloys.

5.1. **Techniques based on the distance to the Nearest-Neighbor (NN-distances)**

Figure 10 shows the estimation of the G and F functions for both Alloy A and Alloy B. As previously discussed, the G function describes the nearest neighbor distance distribution of the pattern. And the F function measures the emptiness in the pattern.

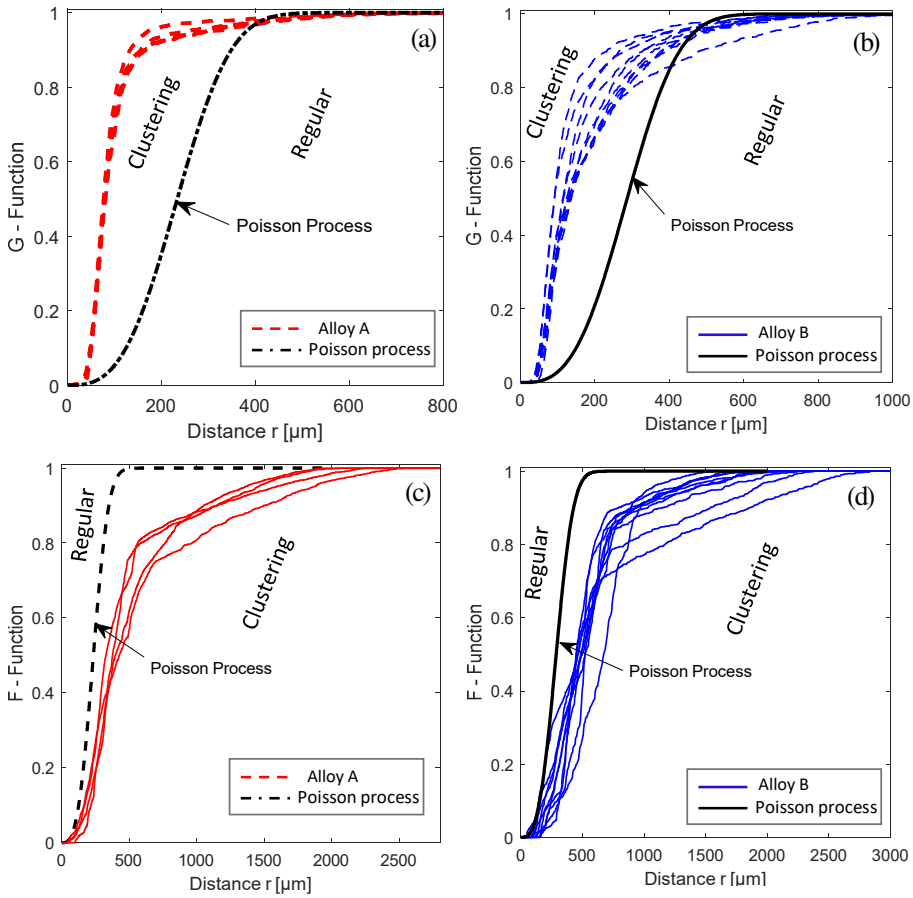

Figure 10: The G Function of the different curves (a) correspond to the 4 scanned samples for alloy A and (b) correspond to the 10 scanned samples for alloy. The F Function of the different curves in (c) correspond to the 4 scanned samples for alloy A and (b) correspond to the 10 scanned samples for alloy B.

According to the G function presented in Figures 10 (a) and (c), all the empirical curves for both alloys lie above the theoretical curve for the Poisson process. This indicates that the NN-distances in the data are shorter than those expected from a Completely Random pattern. This is consistent with clustering.

Figure 10 (b) and Figure 10 (d) show the results for the F-function for both alloys. It can be seen that the empirical curves lie under the theoretical curve for the Poisson process. This is an indication of the presence of clustering in both alloys.

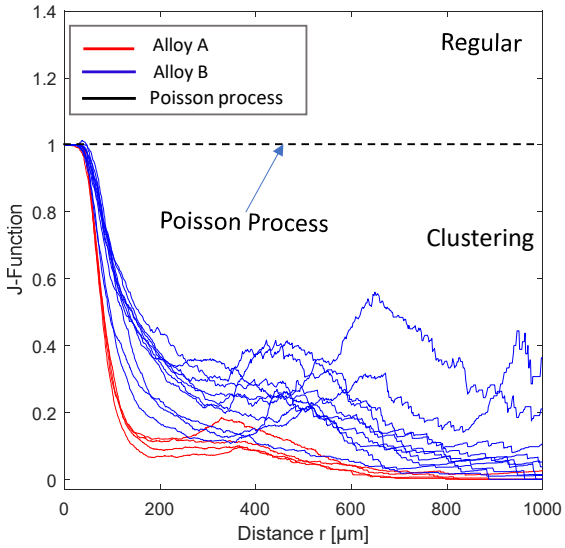

Figure 11: The J-Function for the 4 the scanned samples of Alloy A (in red), and for the 10 scanned samples of Alloy B (in blue) and the Poisson Process (in black).

Figure 11 shows the J-Function versus the NN-distance r, estimated for the two alloys. The results from the J-function confirm the presence of clusters in the studied patterns. This is because the curves are lower than the value of 1 corresponding to the Poisson Process. From Figure 11, it can also be concluded that there is more clustering in Alloy A compared to Alloy B.

In the next section, the results in terms of the K-Function are presented. This analysis gives greater detail as it does not only consider the Nearest Neighbor pores. The drawback of this technique is that it requires an edge effect correction. It is generally recommended that in order to minimize the edge effect, a maximum distance 'd' of approximately ¼ of the minimum distance should be used.

5.2. **The K-Ripley Function**

Figure 12 shows the estimation of the K-Ripley function for both alloys. The black curve is the K-Ripley function for the Poisson process. A curve that is above the Poisson process curve indicates clustering and a curve that is under the Poisson process curve indicates regularity.

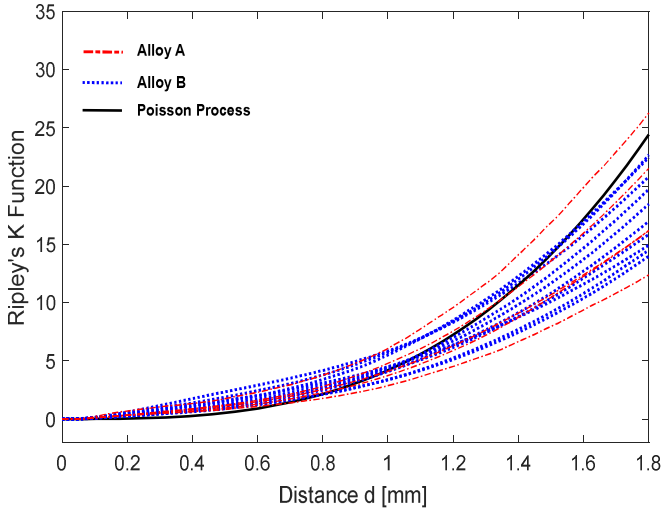

Figure 12: Ripley's K-function for Alloy A, Alloy B and the Poisson process

From Figure 12 it can be seen that the point patterns of both alloys show a combination of effects, clustering at small scales and regularity at large scales. This information was not detected by the techniques based only on the Nearest-Neighbor distance.

In order to better highlight the difference between a point pattern process and the Poisson process, the estimated K-function is often displayed as the difference between the estimated K-Ripley function and the expected $K_{CSR}(d) = \frac{4}{3}\pi d^3$.

When the estimated K-function value is similar to the expected value from the Poisson process the difference $\left(K(d) - \frac{4}{3}\pi d^3\right)$ is close to zero. This value is referred to as the "Residual K-Function". When the Residual K-Function is positive or the values of the estimated K-function are higher than Poisson process, this indicates clustering. When the difference is negative it indicates regularity. Moreover, the intersection of the K-Ripley function with the theoretical curve of the Poisson process determines a critical distance.

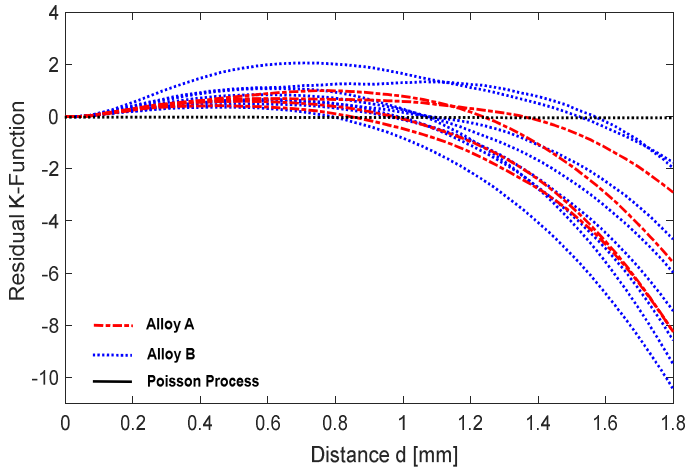

Figure 13: Residual K-Ripley functions of Alloy A, Alloy B and Poisson process.

The resulting plot is shown in Figure 13. The horizontal line indicates the "theoretical" values of $\left(K(d) - \frac{4}{3}\pi d^3\right)$ using the CSR hypothesis. Furthermore, the Residual K-Function (Figure 13) shows that the point patterns of both alloys have a combination of effects, clustering at small scales and regularity at large scales.

Moreover, the intersections between the experimental curves and the Poisson process curve provides information concerning the critical inter-pore distance below which there is risk of clustering. For both alloys, the critical distance varies between 800 µm to 1600 µm. This critical distance could be important in terms of the size of Representative Volume Element for the fatigue behavior of the material, although information concerning the defect size distribution would also be required.

Nevertheless, it would be beneficial to have quantitative evaluation that measures the degree of clustering in the studied point processes. The following section presents the results of the Clark-Evans test including a proposed methodology to adapt it to the 3D point pattern case. Note that only the One-Tailed version is used as it is concluded that both alloys show clustered patterns.

5.3. **Clark-Evans test in 3D**

The key to this statistical test is to distinguish between the degree of clustering that could easily occur haphazardly and those that could not (i.e. due to the physics of the process). Figure 14 shows a Poisson pattern that contains some clusters. Two examples are circled in red.

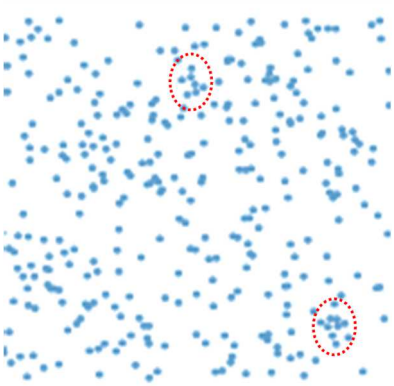

Figure 14: An example of a CSR pattern with a certain degree of clustering (i.e. the red circles) that occurred fortuitously.

Example for the case of intensive sampling (m=N) for a sample from Alloy A

To obtain insight into the Z_m variable, we first need an estimate of the Mean of the NN distance $\overline{D}_m$, then an estimate of the density. By using equation (eq. 16) the value of Z_m can be obtained. For instance, the batch A23 has a $\overline{D}_m = 101\ \mu m$ and a density ρ=13 mm^{-3}. Therefore:

$$Z_m = -1.58 > -Z_{0.1} = -1.65 \qquad \text{(eq. 20)}$$

Hence, from Table 1, this value of Z_m indicates "weak to standard significance" of clustering. Moreover, the P-value in this case is given by:

$$\text{P-value (A)} = \phi(Z_m = -1.58) = 0.12\ (\sim 12\%) \qquad \text{(eq. 21)}$$

This means that the chance of obtaining a mean Nearest-Neighbor distance this low with the Poisson process hypothesis is 12% for Alloy A.

The same steps have been done for the other batches; the results are summarized in Table 5 and the averages are show in Table 6. The results do not present a high dispersion for Z_m.

Table 5: Summary of the numerical results for intensive sampling (m=N).

Batch	ρ [mm^{-3}]	Z_m	P-Value
B10	4	-1.19	0.20
B19	8	-1.14	0.21
B20	10	-1.51	0.13
B30	7	-1.24	0.19
B31	8	-1.47	0.14
B32	6	-1.15	0.21
B43	6	-1.39	0.15
B48	6	-1.20	0.19
B49	6	-1.27	0.18
BVN1-05	4	-1.48	0.13
A01	13	-1.53	0.12
A23	13	-1.58	0.12
A24	13	-1.53	0.12
A30	12	-1.53	0.12

Table 6: Averages values for the alloys

Alloy	ρ [mm^{-3}]	Z_m	P-Value
Alloy A	13	-1.54	0.12
Alloy B	7	-1.30	0.17

However, As mentioned before, one major difficulty with this conclusion is that the entire point pattern (m = N) has been used and as discussed above the obvious dependencies between NN-distances have been ignored. With this in mind, a procedure for taking random subsamples of pattern points which tend to minimize this dependence problem is presented below.

For the case of non-intensive sampling m<N

This procedure makes it possible to take random subsamples, and thereby reduces the effect of dependencies among NN-distances. The objective is to take many subsamples of the same size (for example with m=N/10) and to examine the range of Z-values obtained. If almost all samples indicate significant clustering, then this yields a much stronger result that is clearly independent of the particular sample chosen. In addition, one might for example want to use the P-value obtained for the sample mean of Z as a more representative estimate of actual significance. Figure 15 shows the sampling procedure.

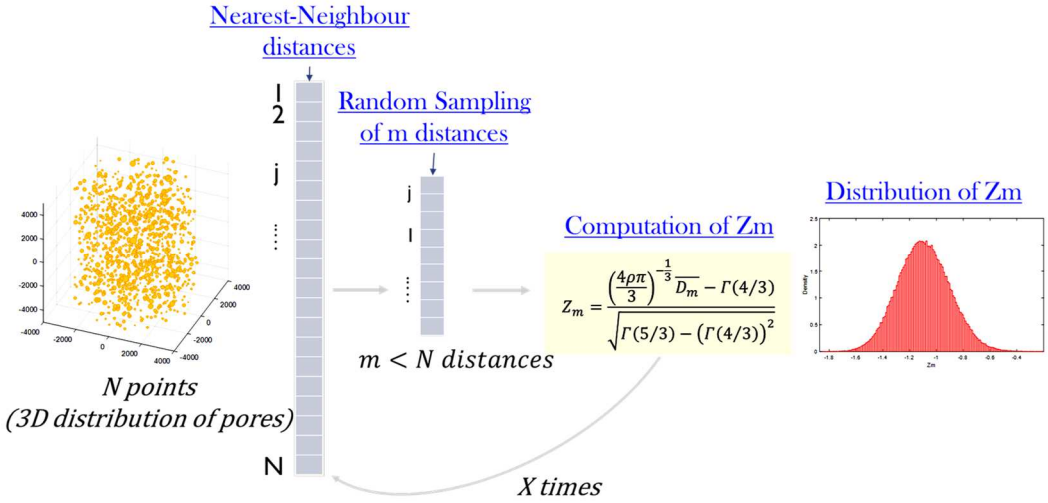

Figure 15: Principle of the sampling

This sampling procedure (Figure 15) was applied to both alloys and to a Poisson process generated numerically. The steps of this procedure are as follows :

1. Randomly choose m=N/10 nearest neighbor distances.
2. Average these m distances.
3. Calculate the expected mean distance given by the Poisson process.
4. Calculate the corresponding Z_m value (and calculate the P-value).
5. Repeat steps 1 to 4 'X' times.
6. Taking into account the average value of Z_m (and the average value of P-value) and considering the significance reference values given in Table 1, determine the degree of clustering.

These steps have been implemented using Matlab. The results are presented below.

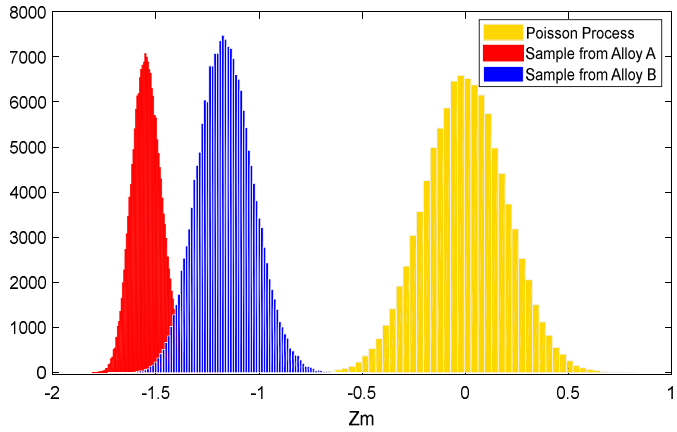

Figure 16: Sampling Distribution of Z-values for Alloy A (in red), Alloy B (in blue) and Poisson process (in yellow)

Figure 16 shows the results of the non-intensive sampling technique of the NN-distances for Alloy A (in red) and Alloy B (in blue) and the NN-distances from a numerical Poisson process (in yellow) used as a reference. The results are summarized in Table 7. It can be concluded that there is no strong evidence suggesting that a clustered spatial pattern is observed in the studied alloys (see Table 8). This implies that the generation of the spatial defect positions in a synthetic microstructure can be done using the Poisson distribution.

Table 7: Summary of the numerical results for intensive and Non-intensive sampling (m=N)

Batch	ρ [mm^{-3}]	Intensive sampling Zm	Non-intensive sampling Zm	P-Value
B10	4	-1.19	-1.16	0.20
B19	8	-1.14	-1.12	0.22
B20	10	-1.51	-1.49	0.13
B30	7	-1.24	-1.23	0.19
B31	8	-1.47	-1.46	0.14
B32	6	-1.15	-1.13	0.21
B43	6	-1.39	-1.38	0.15
B48	6	-1.20	-1.19	0.20
B49	6	-1.27	-1.26	0.18
BVN1-05	4	-1.48	-1.48	0.13
A01	13	-1.53	-1.53	0.12
A23	13	-1.58	-1.57	0.12
A24	13	-1.53	-1.53	0.12
A30	12	-1.53	-1.54	0.12

Table 8: Summary of the intensive and non-intensive sampling

Alloys	Intensive sampling Average (Zm)	Non-intensive sampling Average (Zm)	P-Value	Degree of clustering
Alloy A	-1.54	-1.54	0.12	Weak to standard
Alloy B	-1.30	-1.29	0.18	Weak to standard

From this quantitative evaluation of the degree of clustering in the studied alloys using the Clark-Evans test for both the intensive and non-intensive sampling cases, the following conclusions can be drawn.

i. The value of Z_m for Alloy A is -1.54 with the non-intensive sampling procedure, which is the same for the intensive sampling procedure.
ii. The value of Z_m for Alloy B is -1.19 with the non-intensive sampling procedure, which is very close to -1.30 for the intensive sampling.
iii. The scatter in the numerical distributions of Z_m in Figure 16 is dependent on the size of the sample (i.e. the number of pores in the pattern). The higher the number of pores, the lower the scatter in the variable Z_m.
iv. The values obtained with or without intensive sampling are almost the same. Therefore, it can be concluded that the effect of dependencies between the NN-distances is negligible.
v. From Figure 16, it confirms that the numerical methodology employed to generate pore locations using the Poisson process is quite good, as the average value of Z_m (in yellow) is equal to zero.

6. Generation of synthetic microstructures

In the previous section, it has been shown that for both alloys their point processes can be modelled using the Poisson Process. At the simplest level, the data are analysed only in terms of the coordinate locations of the events (pores). However as previously mentioned, the population of defects can have associated attributes such as the size or the shape of each defect. For the sake of simplicity, the point process will be modelled in association with the defect size that has been determined from the CT X-Ray tomography data. It is assumed that the size and the spatial distribution of the defects are independent of each other.

For the generation of synthetic microstructures, the initial step is to generate the defect or pore positions, and then to associate with each position a size with respect to the defect size distribution extracted using the CT X-Ray tomography data. For the sake of simplicity, the defect shape has not been taken into account. In addition, experimental work (Murakami 1991) has shown that the shape of the defect has a slight effect in the high cycle fatigue failure process. Hence, in this work all defects are assumed to be spherical in shape. The size of defects is expressed as the $\sqrt{Area}$ of this spherical volume.

6.1. Generation of the spatial positions of defects

In the following section, numerical processes are generated that have almost the same density as the natural processes with the homogeneous Poisson process. Firstly, a verification of the ability of the script that was developed to generate numerical Poisson processes is presented in Figure 17.

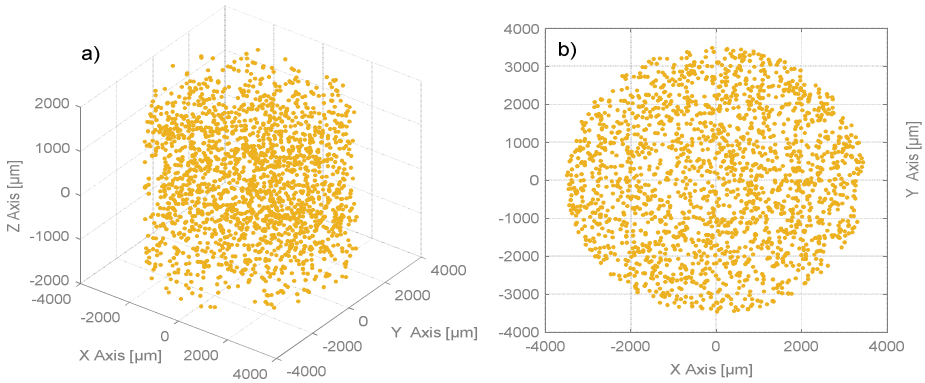

Figure 17: An example of defect positions generated following the Poisson process. (a) The 3D defect positions and (b) the projection of the defect positions onto the XY plane.

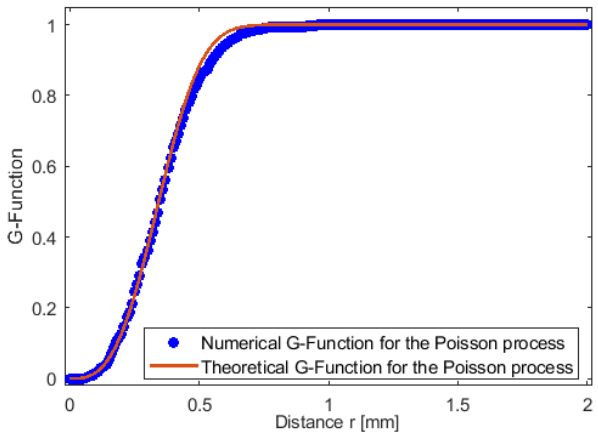

Figure 18: The G-function for the theoretical and the numerical Poisson process.

Figure 17 and Figure 18 confirm that the numerical procedure correctly generates a defect population that follows the Poisson process, as the numerical G-Function and the theoretical one are almost superimposed.

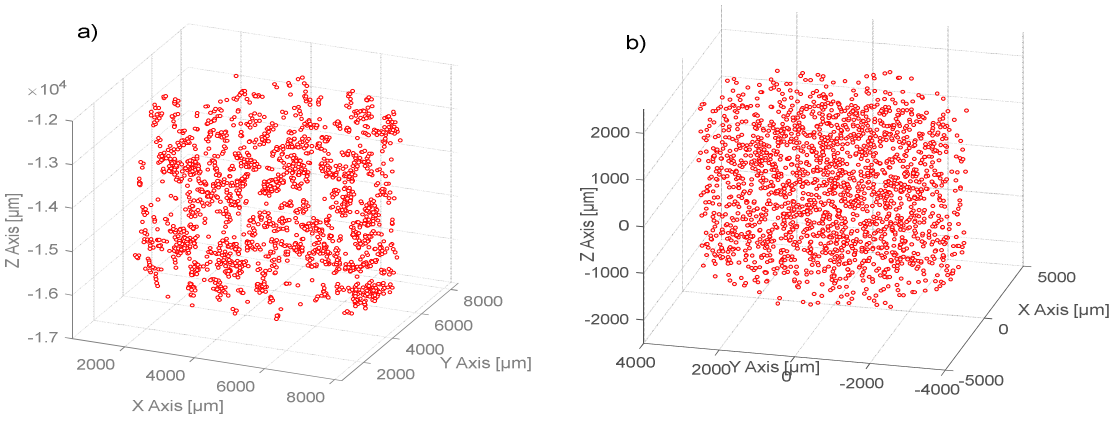

Figure 19: 3D defect positions for Alloy A. (a) The natural pattern obtained via a CT scan and (b) the numerical pattern generated using the Poisson process.

Figure 19 (a) and Figure 28 (a) show respectively an example of the natural positions in 3D space and its projection for a real sample from the alloy A. Figure 19 (b) and Figure 28 (b) show respectively an example of the numerical positions in 3D space and its projection for a numerical sample for the alloy A with the same density of 13 mm^{-3}. These figures give insight into the difference between the numerical and the natural patterns.

The same steps were used to generate a numerical process for Alloy B. The results are shown in Figure 20 and Figure 27. The numerical process shown in Figure 20 (b) and Figure 27 (b) was also generated using the homogeneous Poisson process with the same density of 7 mm^{-3}. Figure 20 (a) and Figure 27 (a) show an example of the natural pattern for Alloy B. These figures show that Alloy B has lower clustering than Alloy A.

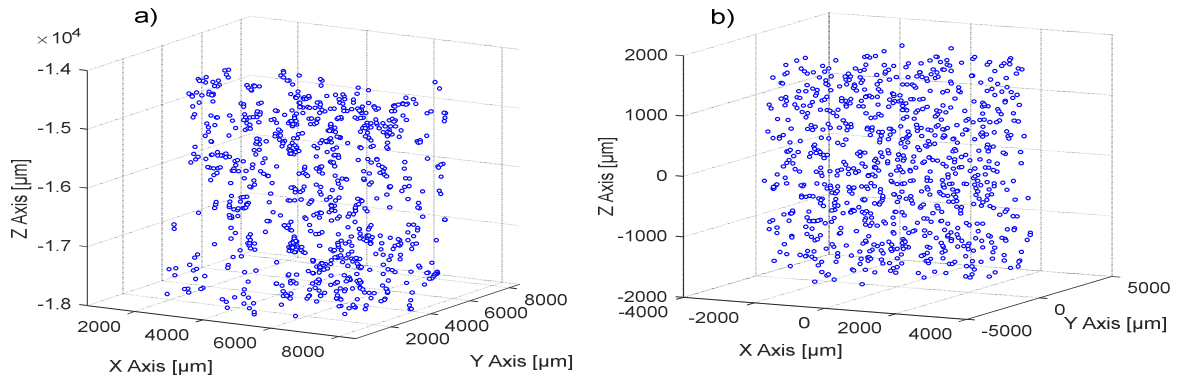

Figure 20: 3D defect positions for Alloy B. (a) The natural pattern obtained via a CT scan and (b) the numerical pattern generated using the Poisson process.

From these results, the presence of some clustering can be seen in the natural pattern, when compared to the numerical patterns corresponding to a homogenous Poisson process. Nevertheless, as previously estimated using the Clark-Evans test, the degree of cluttering is still less than the standard level. Therefore, the resulting process can be and will be used for future modelling work.

6.2. **Association of the defect size to the generated positions**

Figure 21 shows the defect size distributions for the two alloys in terms of the equivalent Murakami parameter (Murakami 1991), $\sqrt{Area}_{eq}$ of the defect. The relationship between the pore volume obtained by tomography and the equivalent square root value of the projected area is given in (eq. 24). This relationship is obtained by assuming the pores have a spherical shape. The data are obtained from different scans, with a total volume of almost 2.16×10^3 mm^3 for Alloy B and almost 0.86×10^3 mm^3 for Alloy A.

$$\sqrt{Area}_{eq} = \pi^{1/6}\,(\frac{3V}{4})^{1/3} \qquad \text{(eq. 24)}$$

This approximation makes it possible to compare the pore sizes obtained by the CT scans and those measured on the fatigue failure surfaces. It can be seen that for this data the maximum defect size for Alloy A is 285 µm and for Alloy B it is 1492 µm.

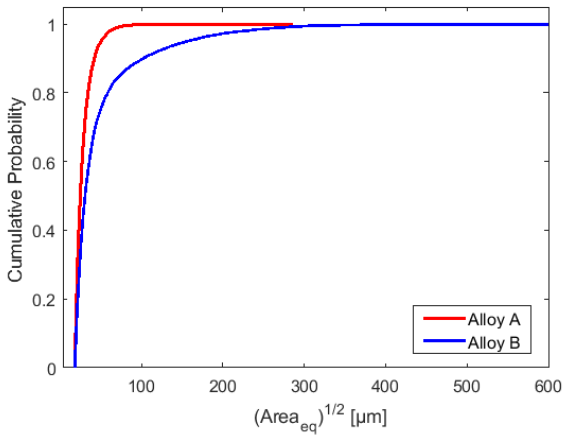

Figure 21: The cumulative probability functions of the defect size in the studied alloys in $\sqrt{Area}_{eq}$

Statistical tests were used to choose the best fit of the defect size distributions. The results of the comparison between the different distributions shows that the Generalized Extreme Value distribution has the maximum value of Ln-Likelihood for both alloys (Table 9). This distribution is therefore

chosen to model the defect size distributions for both alloys. The distribution parameters identified for each material are summarized in Table 10.

Table 9: The Goodness-Of-Fit of the defect size distribution for the studied alloys.

Alloy	Criteria	Generalized extreme value	Lognormal	Gamma	Weibull	Gumbel
A	Ln-Likelihood	**-57218**	-60145	-61363	-64348	-79376
B	Ln-Likelihood	**-64830**	-69005	-71664	-72662	-92007

Table 10: Parameters of the distribution chosen to model the defect size distributions.

	Alloy A	Alloy B
ρ : density of pores (mm^{-3})	13	7
Spatial distribution	Poisson process	Poisson process
Distribution model	Generalized Extreme Value (k=0.35, σ=4.4, μ =22)	Generalized Extreme Value (k=0.9, σ=9, μ =25)

where μ is the location parameter, σ is the scale parameter, and k is the shape parameter of the generalized extreme value distribution.

The maximum pore sizes expressed in terms of the $\sqrt{Area}_{eq}$ parameter, obtained via the CT scans, are approximately 900 µm for Alloy A and 1500 µm for Alloy B. Almost the entire population of defects in Alloy A have a $\sqrt{Area}_{eq}$ that is lower than 100 µm whereas 20% of the defects in Alloy B have a $\sqrt{Area}_{eq}$ greater than 100 µm. The defect size distributions were fitted using the generalized extreme value density probability function. The resulting theoretical distributions were then used to generate a defects size to be associated with each point in the modelled pattern.

6.3. **Examples of synthetic microstructures and comparison with the natural ones**

The synthetic microstructures are generated by coupling the defect size distribution and the position distributions. Figure 23 compares the generated 3D patterns and the natural ones. In terms of their spatial distributions, the natural processes are random with a certain degree of clustering as previously discussed. The synthetic ones are also random processes but with a lower degree of clustering. This small deviation in the 3D pattering between the natural and numerical processes for both alloys can be observed visually, nevertheless, this deviation has already been quantified using the Point Pattern Theory and it has been shown that it is negligible according to the Clark-Evans test. Furthermore, it can be noticed that the clustering is mainly due to the smaller defects.

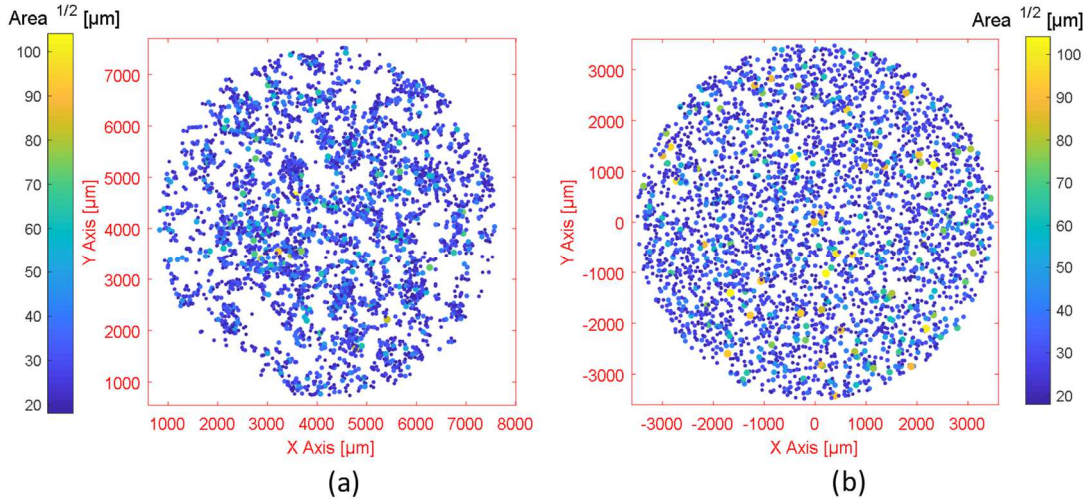

Figure 22: An example of size and spatial distributions for Alloy A projected onto the XY plane (a) from a CT scan and (b) a numerically generated patterns using the Poisson process. Both have height of h=8 mm and a radius of r=3.5 mm.

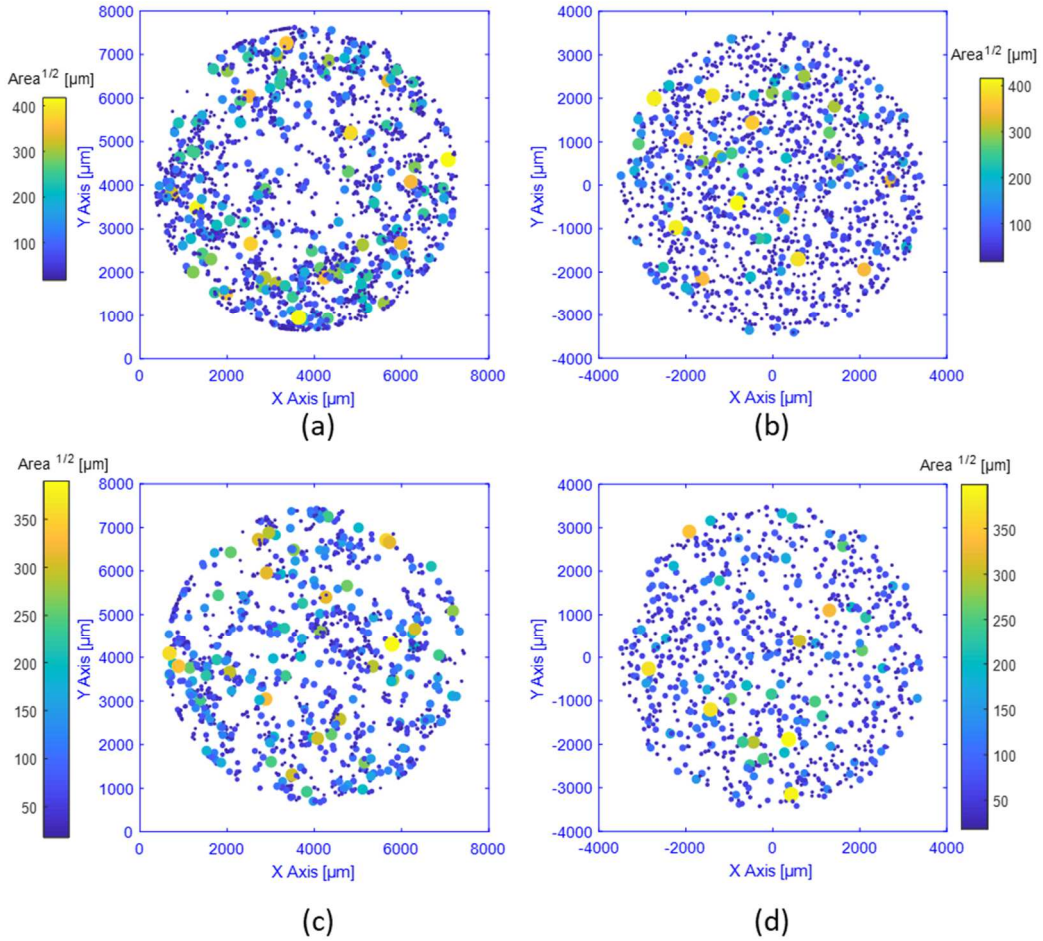

Figure 23: Spatial positions of defects projected onto the XY plane with associated defect size for Alloy B. (a) and (c) are examples of natural patterns, and (b) and (c) are examples of numerical patterns generated using the Poisson process.

From Figure 23, for Alloy B, it can be seen that clustering occurs predominately for the smallest defects. Furthermore, in terms of fatigue behavior, it is generally accepted that only the largest pores are involved in the fatigue process of failure. Therefore, in addition to the low level of clustering given by the Clark-Evans test for this alloy, clustering is only observed for the smallest pores. These results justify the choice of the Poisson process to model the spatial distribution of the defects in these alloys.

Figure 24 and Figure 25, visually demonstrate that if the smallest defects are not taken in account, less clustering would be observed in both alloys. In previously published work (El Khoukhi et al. 2019), the smallest critical defects measured on the fatigue failure surfaces were approximately 30 µm in size. Given this observation, in figures Figure 24 (b) and Figure 25 (b), only the defects with a size larger than 30 µm are considered.

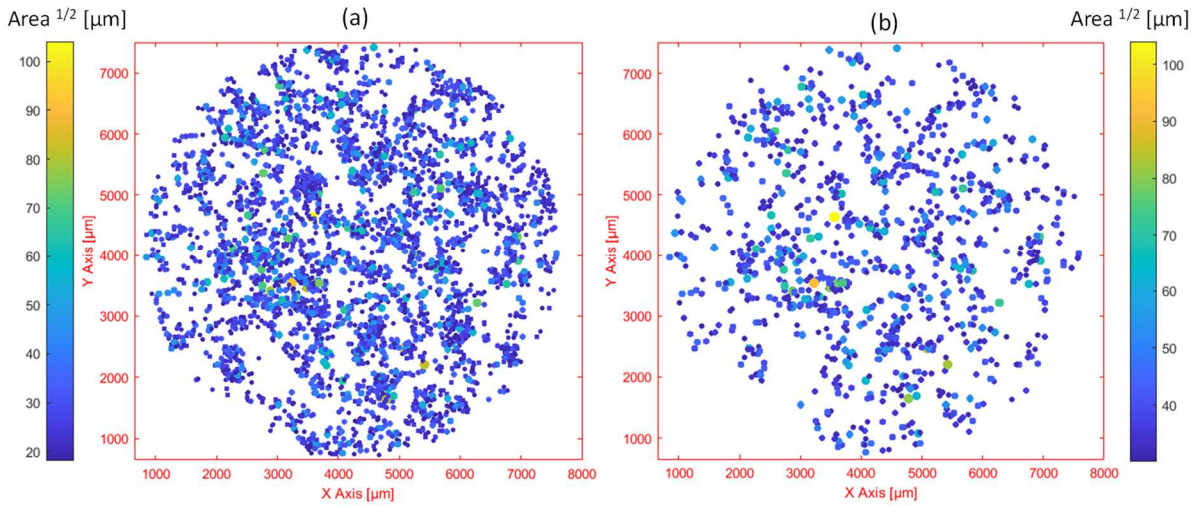

Figure 24: Spatial positions of defects projected onto the XY plane with their associated defect size for Alloy A, with (a) all scanned defects (larger than 18 µm), (b) only the defects larger than 30 µm.

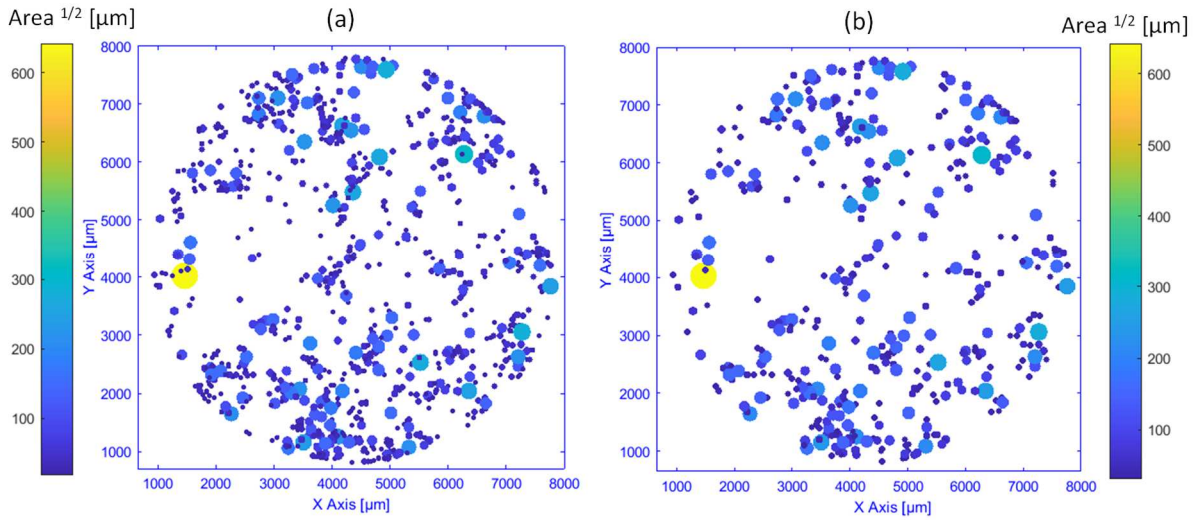

Figure 25: Spatial positions of defects projected onto the XY plane with their associated defect size for Alloy B, with (a) all scanned defects (larger than 18 µm), (b) with only the defects larger than 30 µm.

Figure 26 shows that the size distributions of the numerically generated defects are almost identical to the natural defect size distributions. Furthermore, on this figure, the GEV distributions giving the best Goodness-of-Fit are presented for both alloys.

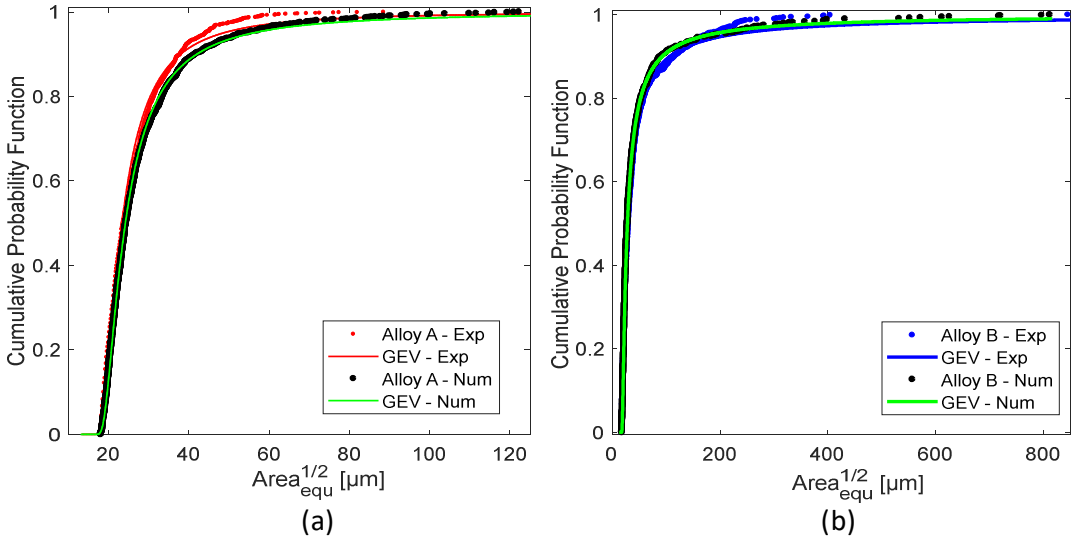

Figure 26: The defects sizes distributions for the natural and numerical process for (a) Alloy A and (b) Alloy B

To sum up, in Section 6.1 the objective was to deal with the arrangement of the defect positions in 3D space, independently of their size. In this section, for the experimental data (Figure 19 (a) and Figure 20 (a)), the position of the defect refers to the center of gravity of each defect obtained from the CT scans of the real samples. On the other hand, for the Numerical data (Figure 19 (b) and Figure 20 (b)), the defect positions are generated using the Poisson process. For this, the defect density must be known, which corresponds to the number of defects per unit volume obtained from the experimental CT data for each alloy.

Whereas, in Section 6.3, the objective is to discusses the generation of numerical microstructures that are as similar as possible to the experimental ones in terms of their defect spatial and size distributions. In this section, for the Experimental data (Figure 22 (a), Figure 23 (a) and (c)), the position of the defects and their sizes were extracted from the CT data. Here, the defect size is presented in terms of a sphere with an equivalent volume. Whereas, for the Numerical data (Figure 22 (b), Figure 23 (b) and (d))), the defect positions are generated using the Poisson process using the same defect density as that identified from the CT data. For each position, a defect size (with a spherical shape) has been attributed. This size was extracted randomly from the statistical distribution identified from the defect size data obtained by CT scans (Table 10).

7. Conclusions and perspectives

In this paper, it has been shown that Point Process Theory is an efficient way to characterize the 3D spatial distribution of defects in materials. The techniques developed in the framework of this theory have been applied to two cast aluminum alloys, referred to here as Alloy A and Alloy B. The statistical analyses have shown that the 3D patterns of the defects in the studied alloys do not present a high degree of clustering. Consequently, for the generation of the defect positions in synthetic microstructures the Homogeneous Poisson process can be used. We notice that in the case of presence of high level of clustering, the Neyman-Scott processes can be used. More information on those technics are provided in (Wilson 2017).

In order to generate numerical microstructures for the cast aluminum alloys, the defect size distributions from the CT-X ray tomography scans were used. The synthetic microstructures are generated by coupling the defect size distributions and the 3D position distributions. The Clark-Evans test indicates that the numerical processes are in agreement with the natural processes for both alloys. The resulting synthetic microstructures are used in modelling work (El Khoukhi et al. 2021) to predict

the average fatigue strength as well as the scatter for a given Highly Stressed Volume by providing information concerning the variability in the microstructural heterogeneities.

We believe that this paper will be with a great help to researchers studying:

- The structural integrity of load carrying component containing defects.
- The variability of the defect characteristics and their distributions within components obtained by certain processes such as additive manufacturing and casting processes.
- Defect based modeling of mechanical performance.

Acknowledgments

The authors gratefully acknowledge the financial support of the OPENLAB Materials and Processes, and the French National Agency for Research and Technology (ANRT).

References

Ammar, H. R., A. M. Samuel, and F. H. Samuel. 2008. "Effect of Casting Imperfections on the Fatigue Life of 319-F and A356-T6 Al–Si Casting Alloys." *Materials Science and Engineering: A* 473 (1): 65–75. https://doi.org/10.1016/j.msea.2007.03.112.

Babu, Gutti Jogesh, and Eric D. Feigelson. 1996. "Spatial Point Processes in Astronomy." *Journal of Statistical Planning and Inference*, Spatial Statistics, Part I, 50 (3): 311–26. https://doi.org/10.1016/0378-3758(95)00060-7.

Baddeley, A. J., and M. N. M. van Lieshout. 1995. "Area-Interaction Point Processes." *Annals of the Institute of Statistical Mathematics* 47 (4): 601–19. https://doi.org/10.1007/BF01856536.

Baddeley, Adrian, Ege Rubak, and Rolf Turner. 2015. *Spatial Point Patterns: Methodology and Applications with R*. CRC Press.

Bailey, Trevor C., and Anthony C. Gatrell. 1995. *Interactive Spatial Data Analysis*. Longman Scientific & Technical.

Ben Ahmed, Amal, Mohamed Iben Houria, Raouf Fathallah, and Habib Sidhom. 2019. "The Effect of Interacting Defects on the HCF Behavior of Al-Si-Mg Aluminum Alloys." *Journal of Alloys and Compounds* 779 (March): 618–29. https://doi.org/10.1016/j.jallcom.2018.11.282.

Billaudeau, T., Y. Nadot, and G. Bezine. 2004. "Multiaxial Fatigue Limit for Defective Materials: Mechanisms and Experiments." *Acta Materialia* 52 (13): 3911–20. https://doi.org/10.1016/j.actamat.2004.05.006.

Blair, Malcolm, Christoph Beckermann, Richard Hardin, Kent Carlson, and Charles Monroe. 2005. "Predicting the Occurrence and Effects of Defects in Castings." *JOM: The Journal of the Minerals, Metals & Materials Society* 57 (May): 29–34. https://doi.org/10.1007/s11837-005-0092-3.

Chen, Bo, Teruo Hashimoto, Frank Vergeer, Andrew Burgess, George Thompson, and Ian Robinson. 2014. "Three-Dimensional Analysis of the Spatial Distribution of Iron Oxide Particles in a Decorative Coating by Electron Microscopic Imaging." *Progress in Organic Coatings* 77 (6): 1069–72. https://doi.org/10.1016/j.porgcoat.2014.03.005.

Clark, P. J., and F. C. Evans. 1954. "Distances to Nearest Neighbor as a Measure of Spatial Relationships in Populations." *Ecology, Vol. 35, p. 445-453 (1954).* 35: 445–53.

Cressie, Noel A. C. 1993. *Statistics for Spatial Data*. J. Wiley.

Diggle, Peter. 1983. *Statistical Analysis of Spatial Point Patterns*. Academic Press.

El Khoukhi, Driss, Franck Morel, Nicolas Saintier, Daniel Bellett, and Pierre Osmond. 2018. "The Effect of Microstructural Heterogeneities on the High Cycle Fatigue Scatter of Cast Aluminium Alloys: From an Elementary Volume to the Structure." Edited by G. Hénaff. *MATEC Web of Conferences* 165: 14006. https://doi.org/10.1051/matecconf/201816514006.

El Khoukhi, Driss, Franck Morel, Nicolas Saintier, Daniel Bellett, Pierre Osmond, and Viet-Duc Le. 2021. "Probabilistic Modeling of the Size Effect and Scatter in High Cycle Fatigue Using a Monte-Carlo Approach: Role of the Defect Population in Cast Aluminum Alloys." *International Journal of Fatigue*, February, 106177. https://doi.org/10.1016/j.ijfatigue.2021.106177.

El Khoukhi, Driss, Franck Morel, Nicolas Saintier, Daniel Bellett, Pierre Osmond, Viet-Duc Le, and Jérôme Adrien. 2019. "Experimental Investigation of the Size Effect in High Cycle Fatigue: Role of the Defect Population in Cast Aluminium Alloys." *International Journal of Fatigue* 129 (December): 105222. https://doi.org/10.1016/j.ijfatigue.2019.105222.

Guerchais, R., F. Morel, and N. Saintier. 2017. "Effect of Defect Size and Shape on the High-Cycle Fatigue Behavior." *International Journal of Fatigue*, Multiaxial Fatigue 2016: Experiments and Modeling, 100 (July): 530–39. https://doi.org/10.1016/j.ijfatigue.2016.12.010.

Illian, Janine, Jesper Møller, and Rasmus Waagepetersen. 2007. "Spatial Point Process Analysis for a Plant Community with High Biodiversity." *Environmental and Ecological Statistics* 16 (September): 389–405. https://doi.org/10.1007/s10651-007-0070-8.

Jang, Younghwan, Youin Jeong, Chongho Yoon, and Sangshik Kim. 2009. "Fatigue Life Prediction for Porosity-Containing Cast 319-T7 Aluminum Alloy." *Metallurgical and Materials Transactions A* 40 (5): 1090–99. https://doi.org/10.1007/s11661-009-9795-3.

Koutiri, Imade, Daniel Bellett, Franck Morel, Louis Augustins, and Jérôme Adrien. 2013. "High Cycle Fatigue Damage Mechanisms in Cast Aluminium Subject to Complex Loads." *International Journal of Fatigue* 47 (February): 44–57. https://doi.org/10.1016/j.ijfatigue.2012.07.008.

Le, Viet Duc. 2016. "Etude de l'influence des hétérogénéités microstructurales sur la tenue en fatigue à grand nombre de cycles des alliages d'aluminium de fonderie," May. https://pastel.archives-ouvertes.fr/tel-01345228.

Le, Viet-Duc, Franck Morel, Daniel Bellett, Nicolas Saintier, and Pierre Osmond. 2016. "Multiaxial High Cycle Fatigue Damage Mechanisms Associated with the Different Microstructural Heterogeneities of Cast

Aluminium Alloys." *Materials Science and Engineering: A* 649 (January): 426–40. https://doi.org/10.1016/j.msea.2015.10.026.
Murakami, Yukitaka. 1991. "Effects of Small Defects and Nonmetallic Inclusions on the Fatigue Strength of Metals." Key Engineering Materials. Trans Tech Publications Ltd. 1991. https://doi.org/10.4028/www.scientific.net/KEM.51-52.37.
Perrin, Guillaume, Xavier Descombes, and Josiane Zerubia. 2006. "Point Processes in Forestry : An Application to Tree Crown Detection," January.
Ripley, B. D. 1977. "Modelling Spatial Patterns." *Journal of the Royal Statistical Society. Series B (Methodological)* 39 (2): 172–212.
Smith, Tony E. 2020. "Notebook on Spatial Data Analysis." http://www.seas.upenn.edu/~ese502/#notebook.
Stefanescu, Doru. 2005. "Computer Simulation of Shrinkage-Related Defects In Castings – A Review." In . Vol. 179.
Wilson, Pablo. 2017. "Inférence statistique de la population de défauts pour l'étude et la simulation de la fatigue oligocyclique d'un alliage d'aluminium de fonderie." Phdthesis, Ecole nationale supérieure d'arts et métiers - ENSAM. https://pastel.archives-ouvertes.fr/tel-01511338.
Wilson, Pablo, Nicolas Saintier, Thierry Palin-Luc, and Sebastien Bergamo. 2019. "Isothermal Fatigue Damage Mechanisms at Ambient and Elevated Temperature of a Cast Al-Si-Cu Aluminium Alloy." *International Journal of Fatigue* 121 (April): 112–23. https://doi.org/10.1016/j.ijfatigue.2018.11.016.
Yi, J. Z., X. Zhu, J. W. Jones, and J. E. Allison. 2007. "A Probabilistic Model of Fatigue Strength Controlled by Porosity Population in a 319-Type Cast Aluminum Alloy: Part II. Monte-Carlo Simulation." *Metallurgical and Materials Transactions A* 38 (5): 1123–35. https://doi.org/10.1007/s11661-006-9069-2.
Zhu, X., J. Z. Yi, J. W. Jones, and J. E. Allison. 2007. "A Probabilistic Model of Fatigue Strength Controlled by Porosity Population in a 319-Type Cast Aluminum Alloy: Part I. Model Development." *Metallurgical and Materials Transactions A* 38 (5): 1111–22. https://doi.org/10.1007/s11661-006-9070-9.

Annexes

Definition of the Ln-Likelihood Function

The ln-likelihood function is defined as:

$$l_N: \Theta \times R^N \to R$$

$$(\theta; x_1, x_2, \dots, x_n) \to l_N(\theta; x_1, x_2, \dots, x_n) = \sum_{i=1}^{N} \ln f_X(x_i; \theta)$$

Where:

- X is a continuous random variable with a probability density function denoted by $f_X(x_i; \theta)$, for $x \in R$. Here, five pdfs were evaluated: Generalized extreme value, Lognormal, Gamma, Weibull, Gumbel.
- $\theta = (\theta_1 \dots \theta_K)^T$ is a $K \times 1$ vector of unknown parameters, we assume that $\theta \in \Theta \subset R^K$
- Let us consider a sample $\{X_1, \dots, X_N\}$ of i.i.d. random variables with the same arbitrary distribution as X.
- The realization of $\{X_1, \dots, X_N\}$ (the data set…) is denoted $\{x_1, \dots, x_N\}$.
- And Ln is the Natural logarithm.

Complementary figures

Some complementary figures for the defects positions Figure 27 and Figure 28 and numerical samples in Figure 29 and Figure 30.

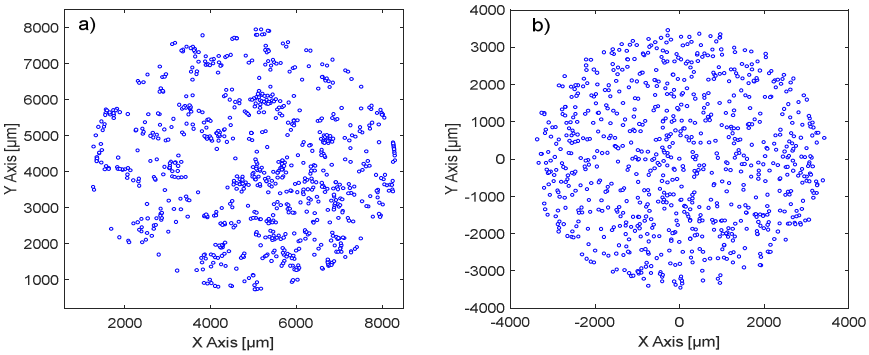

Figure 27: The projection of the defect positions onto the XY plane for (a) the natural pattern obtained via a CT scan and (b) the numerically generated pattern using Poisson process.

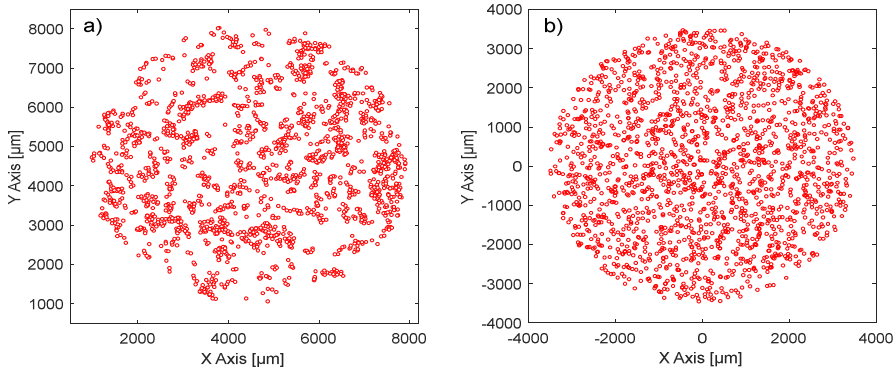

Figure 28: Projection of the defect positions onto the XY plane for Alloy A. (a) The natural pattern obtained via a CT scan, and (b) the numerically generated pattern using Poisson process.

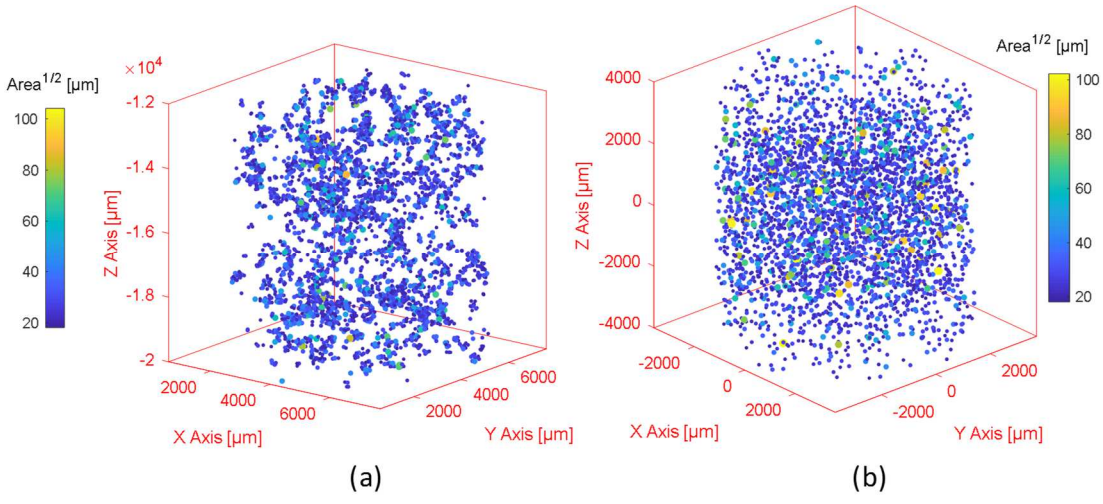

Figure 29: An example of size and 3D spatial distributions for Alloy A (a) from a CT scan and (b) a numerically generated pattern using Poisson process. Both have height of h= 8 mm and a radius of r= 3.5 mm.

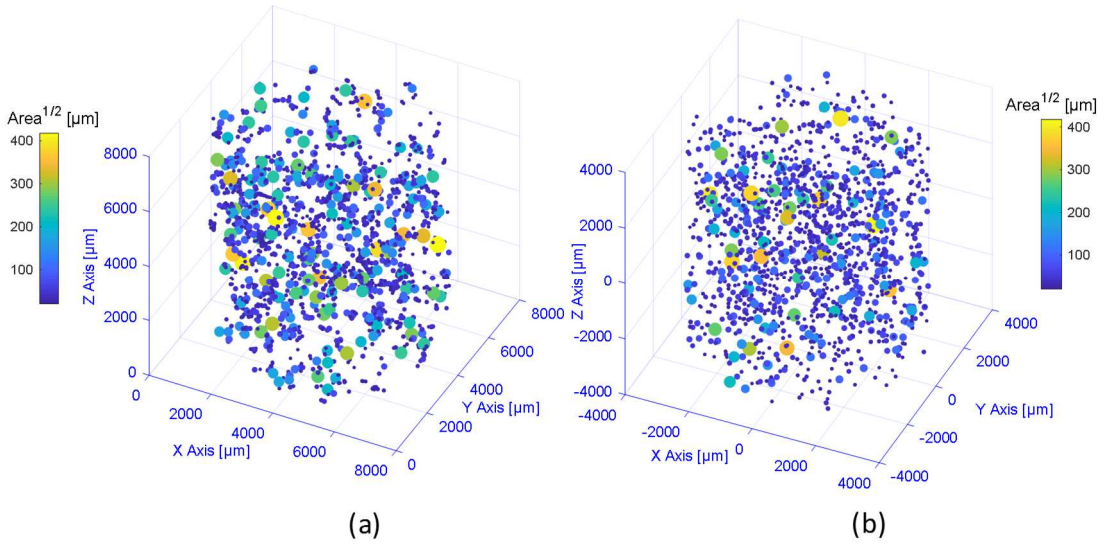

Figure 30: An example of size and 3D spatial positions of defects for Alloy B (a) from a CT scan and (b) numerically generated patterns using Poisson process. Both have height of h= 8 mm and a radius of r=3.5 mm.