Statistical study of the size and spatial distribution of defects in a cast aluminium alloy for the low fatigue life assessment

Pablo Wilson, Nicolas Saintier, Thierry Palin-Luc, Bruno Sudret, Sebastien Bergamo

Statistical study of the size and spatial distribution of defects in a cast aluminium alloy for the low fatigue life assessment

Pablo Wilson [a,c], Nicolas Saintier [a,*], Thierry Palin-Luc [a], Bruno Sudret [b], Sebastien Bergamo [c]

[a] Arts et Metiers Science and Technology, CNRS, University of Bordeaux, Bordeaux INP, I2M Bordeaux, Esplanade des Arts et Metiers, 33405 Talence, France
[b] Chair of Risk, Safety and Uncertainty Quantification, ETH Zurich, Stefano-Franscini-Platz 5, 8093 Zurich, Switzerland
[c] Renault S.A.S, 1 avenue du Golf, 78280 Guyancourt, France

A B S T R A C T

Cast aluminium alloys, and more widely cast materials, are frequently used in industry. The casting process allows for complex geometries of parts, but, on the downside, often causes materials voids. It is well known these material defects are harmful for material fatigue performances, but the nature of these defects, in a statistical manner, are more seldom studied. This paper aims at proposing a methodology for finding the underlying characteristics of the defect population (size and spatial distribution) and determine their implication on fatigue behaviour in the presence of stress/strain gradients (notched specimens). To do so, various statistical tools are brought from different fields, such as point processes, and applied to experimentally observed defect distributions (by μCT tomography on virgin test specimens). The population of defects is clearly identified, and it is shown these defects are not randomly distributed, but rather in cluster. It is also shown there is no strong link between the defect size an it's location. Knowing the statistics of the defect population, it is then possible to confront the result of fatigue tests (and the observed initiating defects) with the simulated defect population: the fatigue crack initiation mechanisms, which favour (sub-) surface rather than core initiating defects, reduce the size of the active zone and therefore artificially shift the defect size distribution (by reducing their number).

1. Introduction

Cast aluminium alloys are widely used in the automotive industry [1]. Due to their good specific mechanical properties, their good machinability, their high thermal conductivity, combined with a reasonable pricing, they are often chosen for highly mechanical and thermal loaded parts [2]. Amongst these aluminium alloy, Al–Si based alloy, combined with smaller fraction of Cu and Mg, are often favoured for engine components. While the balance of the alloying elements plays a key role in the material behaviour, the casting defects also are of paramount importance, especially for fatigue behaviour (see for example [3]). In the past ten years a great effort has been made by the fatigue community in order to propose microstructure sensitive numerical methods [4–6] that allow to consider more precisely the role of key microstructural features. In the literature, many works related to the defect sensitivity of the fatigue behaviour of aluminium alloys focus on the role of the size and the shape of defects [7,8], but very little work are dedicated to the precise description and modelling of the spatial distribution of defects, especially in 3D [9,10].

Particle distribution description has been a subjects of interest for many years in a very wild range of scientific fields [11–13], such as in the study of forests seeds dissemination or astronomy [14]. In recent years, it has gained much interest when applied to microstructural elements: for reinforcement of composites, metallic grains or, as studied in this paper, voids. This paper is a proposal to better address the description of 3D spatial distribution of defects for cast aluminium alloys and to linking this distribution to the high temperature low cycle fatigue behaviour of the considered alloys.

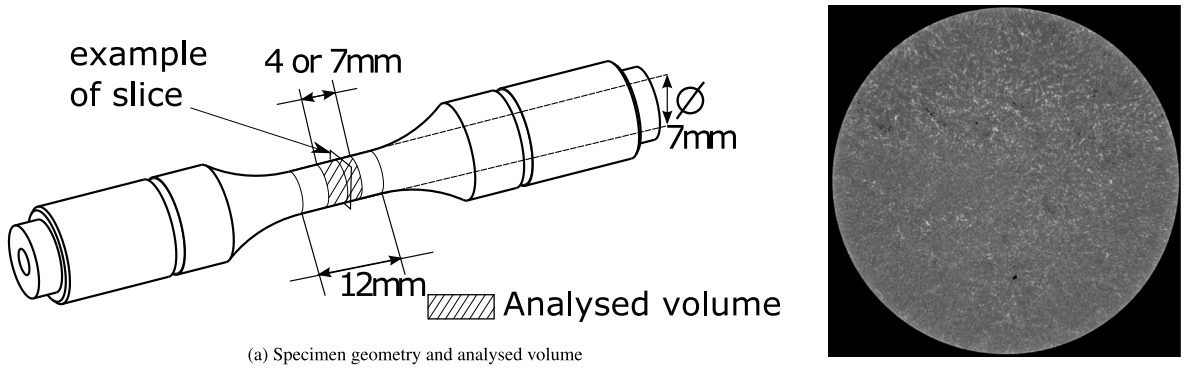
2. Experimental conditions

2.1. Material and specimen

The material used for this study is an Aluminium–Silicon–Copper based cast aluminium alloy. Its measured chemical composition is shown in Table 1 and the secondary dendritic arm spacing (SDAS) has a value of 24 μm, close to what is observed for similar casting processes

* Corresponding author.
E-mail address: nicolas.saintier@ensam.eu (N. Saintier).

[1] Except for sample 202F, which is as cast.



(a) Specimen geometry and analysed volume

(b) Example of slice

Fig. 1. Specimen used for fatigue testing (The 12 mm zone being the fatigue tested zone) and tomography analysed volume.

Table 1
Cast alloy measured composition.

	Si	Cu	Mg	Ti	Fe	Mn	Zn	Ni	Al
Measured(w%)	7.86	3.18	0.33	0.14	0.34	0.23	0.35	0.04	Balance

in literature (see [15]). Castings were made in cylindrical bars of approximately 20 mm diameter and 200 mm long. They underwent a T5 heat treatment.[1] From the bars were machined fatigue test specimens following description on Fig. 1.

2.2. Tomography

X-ray Tomographic studies for material science have been increasingly used in the past years ([3,16] for similar studies, see [17,18] for more broad reviews of X-ray tomography possibilities. In our case, the analysis was used to observe the voids in the material. These voids can be of 2 types: gas pores, which are the result of entrapped gas during the casting process, and shrinkage pores, which are caused by heterogeneous cooling rates during the cooling process. While it is possible to separate the pore type based on its sphericity, the spatial resolution used here was unsufficient to clearly separate the two populations. Therefore, in our study, the type of pore will not be considered. In order to have the specimen representativeness, 3 of them were analysed by X-ray tomography (see Table 3).

2.3. Fatigue tests

A more detailed description of the fatigue tests and the defect size analysis used in this study can be found in [19]. Only the main characteristics of the tests and the results needed are presented hereafter.

Fatigue tests are performed on cylindrical low-cycle fatigue specimens shown in Fig. 1(a)). The active zone is a 7 mm diameter and 14 mm height cylinder. A 12 mm gauge extensometer is used to perform strain-controlled tests. Fatigue tests are performed on an Instron® 8500 servo-hydraulic machine equipped with a high-temperature MTS® furnace. The strain imposed signal waveform is triangular as to maintain a constant strain-rate ($\dot{\varepsilon}$) of 1.10^{-3} s^{-1} at strain amplitudes ($\Delta\varepsilon/2$) ranging from $\pm 0.2\%$ to 0.4%, and strain ratio $R_\varepsilon = -1$. Surface roughness of all the specimens (arithmetic average of absolute values) is lower than $R_a = 0.8$ µm. Fatigue tests are conducted under 4 different controlled temperatures: 20 (ambient), 150, 200 and 250 °C. The number of cycles to failure was defined by a 10% drop from the trend line of the maximum stress per cycle. In order to measure the critical defect size, the fracture surface was systemically observed by Secondary Electron Microscopy and the 2D Feret Diameter was measured. An example of defect measurement is shown in Fig. 2. For each fatigue test, the critical defect was identified and two parameters were measured (see Table 2):

Table 2
Defect measurement.

Specimen no. (–)	Feret diameter (µm)	Distance to surface (µm)
88T5	236	0
68T5	351	0
87T5	356	0
56T5	178	0
14T5	241	159
72T5	320	0
130T5	144	90
28T5	223	30
70T5	189	0
170T5	325	0
49T5	152	0
59T5	371	263
67T5	128	0
25T5	273	35
71T5	195	195
62T5	130	180
105T5	164	0

the defect size (2D Feret diameter) and the distance to surface (which is the smallest distance between the defect frontier and the specimen outside surface).

3. Statistical description of defect size

3.1. Segmentation of tomographic data for porosity identification

In order to ensure comparison between specimens, the first aim of this work was to propose a reproducible method for porosity detection of laboratory X-ray tomographic data. Indeed, the highly non-linear aspect of the measurement and reconstruction method makes data thresholding highly dependent on the raw data itself (in the present case the Avizo software was used). It is often observed that only data measured consecutively on the same machine and the same operator can be compared. Even in this case, automatic range scaling parameters could render the comparison difficult. To overcome this and to compare specimens from different set of measurements, two specimens are used (specimens 12T5 and 125T5, see Table 3) to set up a proper methodology. These two specimens are made of the same material and the same batch, but specimen 12T5 was analysed using a Versa 500 Zeiss, while specimen 125T5 was analysed by a General Electrics X-ray nanotom.

Being obtained from two different X-ray systems, the raw grey histograms of the specimens shown Fig. 3 appear different. A first step using linear scaling of the data intensity is therefore performed on specimen 125T5 in order to have the same general aspect as the 12T5 specimen:

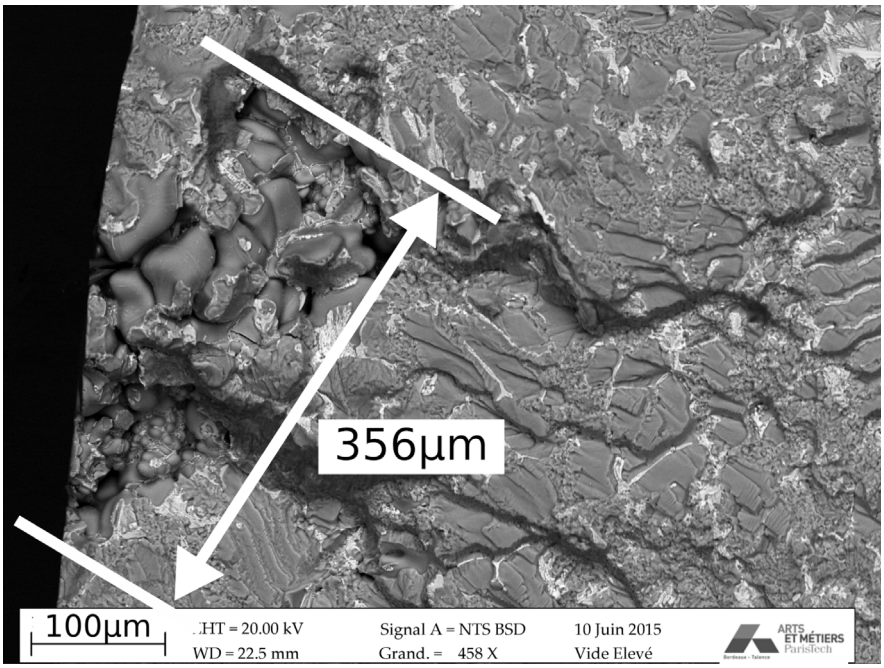


Fig. 2. Measurement of defect size (Specimen 87T5, $\Delta\varepsilon/2 = 0.30\%$, 20 °C, $R_\varepsilon = -1$, $N_f = 2904$ cycles).

Table 3
Specimens analysed by tomography.

Specimen	Volume	Size	Resolution
12T5	7 × 7 × 6.5 mm	1920 × 1920 × 1750 pixels	3.7 µm
202F	7 × 7 × 6.5 mm	1920 × 1920 × 1750 pixels	3.7 µm
125T5	7 × 7 × 4 mm	1450 × 1450 × 800 pixels	4.8 µm

- the first peak around 12 000 corresponds to the void value: after reconstruction, the voxel intensity of void space is not strictly zero. This peak is slightly higher for specimen 125T5.
- the second peak is the mean material intensity. Both specimens (12T5 and 125T5) have the same peak value, around 28 000.

Applying this method requires the two tomographic analysis to be similar, i.e. they must contain the same proportion of material and void, otherwise the peaks could be artificially shifted. After linear shifting of the data, both histograms have the same trend.

After this image correction, phase separation can be made by data thresholding. An example is given on a slice of specimen 12T5 shown Figs. 4(a), 4(b), 4(c) and 4(d), showing the effect of different thresholding levels of (15 000, 22 000 and 20 000), respectively. Based on the histogram shown Fig. 3, three remarkable values of thresholding could be chosen:

- the mean void intensity value (x-value labelled (1) in Fig. 3, around 12 000): correspond to the maximum of the first peak. However, visual comparison shows this value underestimates the void size. Even for a higher thresholding of 15 000 (used in Fig. 4(b)), visual identification of voids does not match the binary thresholding.
- the local minima between the two peaks (x-value labelled (3) in Fig. 3, around 22 000): given the minimum curvature, this value is highly imprecise and cannot be used as a robust indicator. The visual comparison (see Fig. 4(c)) shows that void size might be slightly overestimated.
- the middle distance between the two peaks (x-value labelled (2) in Fig. 3, here 20 000), can be used to provide a robust thresholding value. Fig. 4(d) shows that large porosities edges are correctly detected, while smaller voids (of only a few voxels) might go undetected.

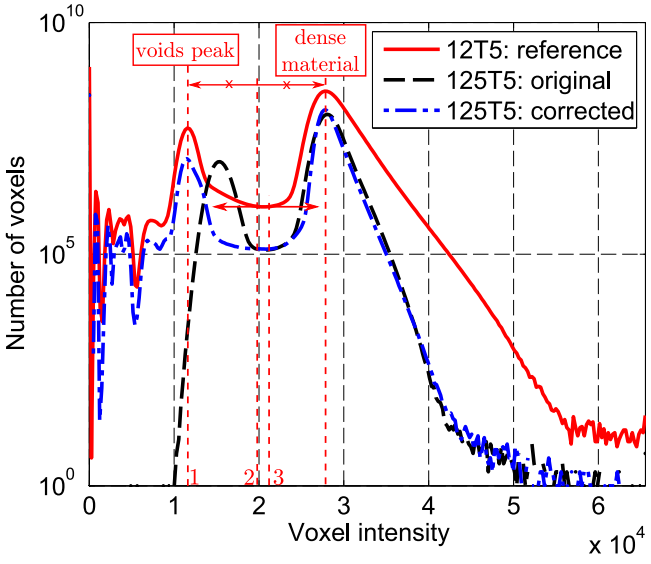


Fig. 3. Histogram equalization: specimen 125T5 is from a different source than specimen 12T5, the intensity is manually corrected (linear correction) for the two local maximums to fit. Voxel intensity is between 0 and $2^{16} = 65\,536$ (16 bits).

The threshold value of 20 000 was finally considered as the reference by visual comparison. The threshold problematic arises not only to detect the edges (an error of measurement of 1 or 2 voxels on total defect size has little impact), but first and foremost for void separation: because of the highly tortuous shape of shrinkage defects, numerical thresholding can separate or coalesce two or more close voids.

3.2. Feret diameter analysis

Once the thresholding and segmentation are done, a measurand must be chosen to quantify porosity size. In this study, the Feret

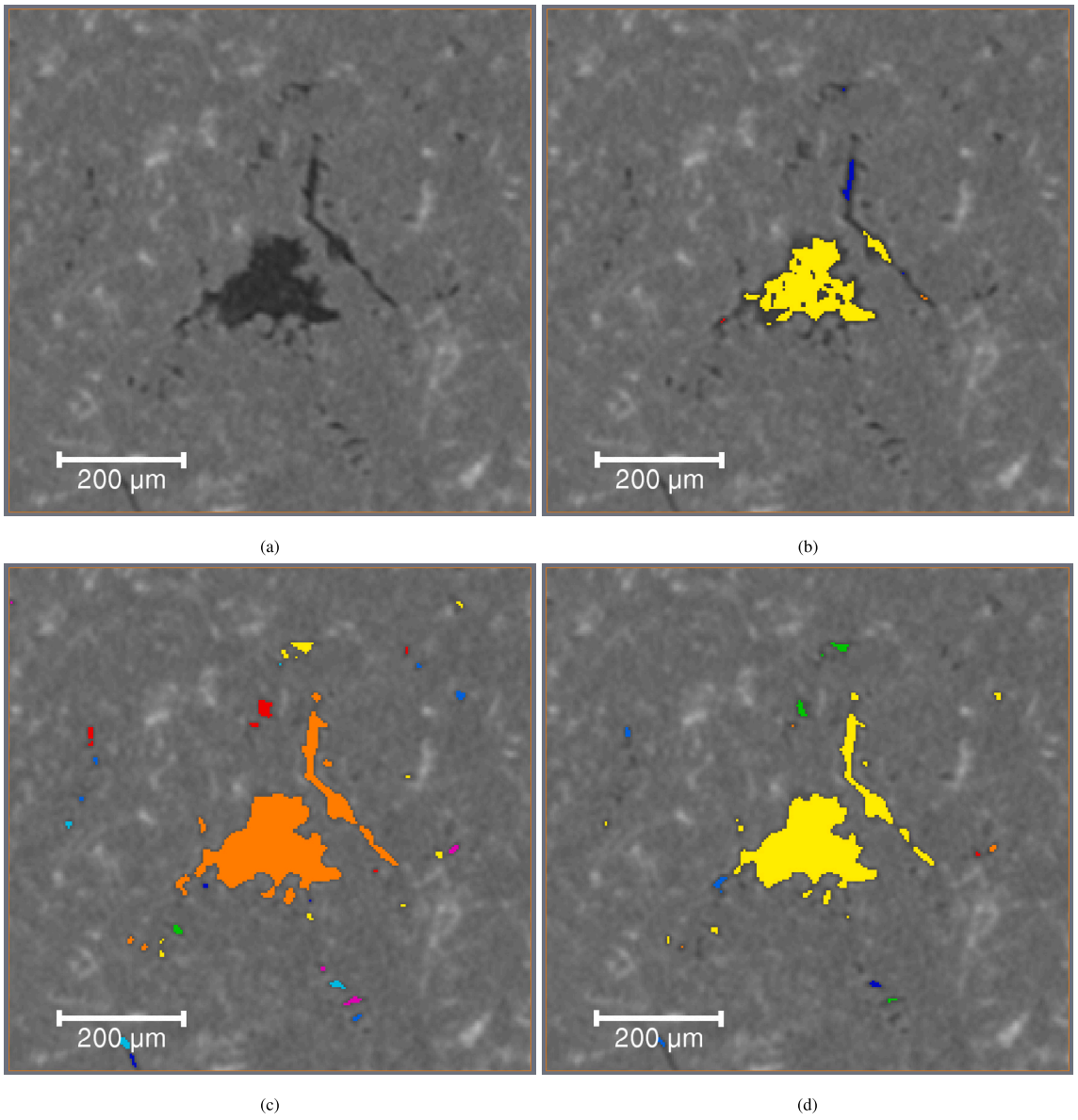


Fig. 4. Effect of thresholding value on void detection: (a) raw data, (b) threshold value of 15 000, (c) threshold value of 22 000, (d) threshold value of 20 000.

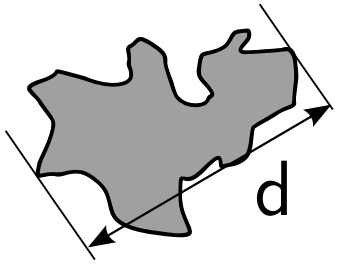


Fig. 5. Two dimensional measurement of Feret diameter.

diameter is chosen. This measure corresponds to the largest distance between two parallel tangents (line in 2D and plane in 3D) on opposite sides. An illustration of a 2D measure of the Feret diameter is given Fig. 5.

The histogram of the Feret diameter for the different specimens, and thresholding values, are shown Fig. 6. A few remarks can be made:

- for specimen 12T5, the effect of thresholding is also observed. Concurrently with the visual observation made previously, thresholding values of 20 000 and 22 000 yield similar results, especially for the larger defects. For example, the largest defect detected is the same and only increases by 34 µm (from 498 µm to 532 µm for a thresholding value of 20 000 to 22 000). For the thresholding value of 15 000 however, the larger defects are clearly less numerous. The thresholds of 15 000 and 22 000 are no further discussed and only the 20 000 is used for all specimens.
- Specimen 125T5, which was analysed by a different tomograph, has a slightly higher defect number than specimen 12T5 but tail values show similar quantities of larger defects.
- Specimen 202F has a lesser number of defects and the largest defect measures 251 µm. This could be the result of the different heat treatment and/or the material variability.

In order to analyse the defect distribution by statistical means, a censoring of smaller defects was applied: assuming only the larger defects are critical for fatigue behaviour, only defects larger than the threshold $\mu_e = 40\,\mu m$ are considered. By applying this censoring, a more important weight is given to the larger defects in the statistical

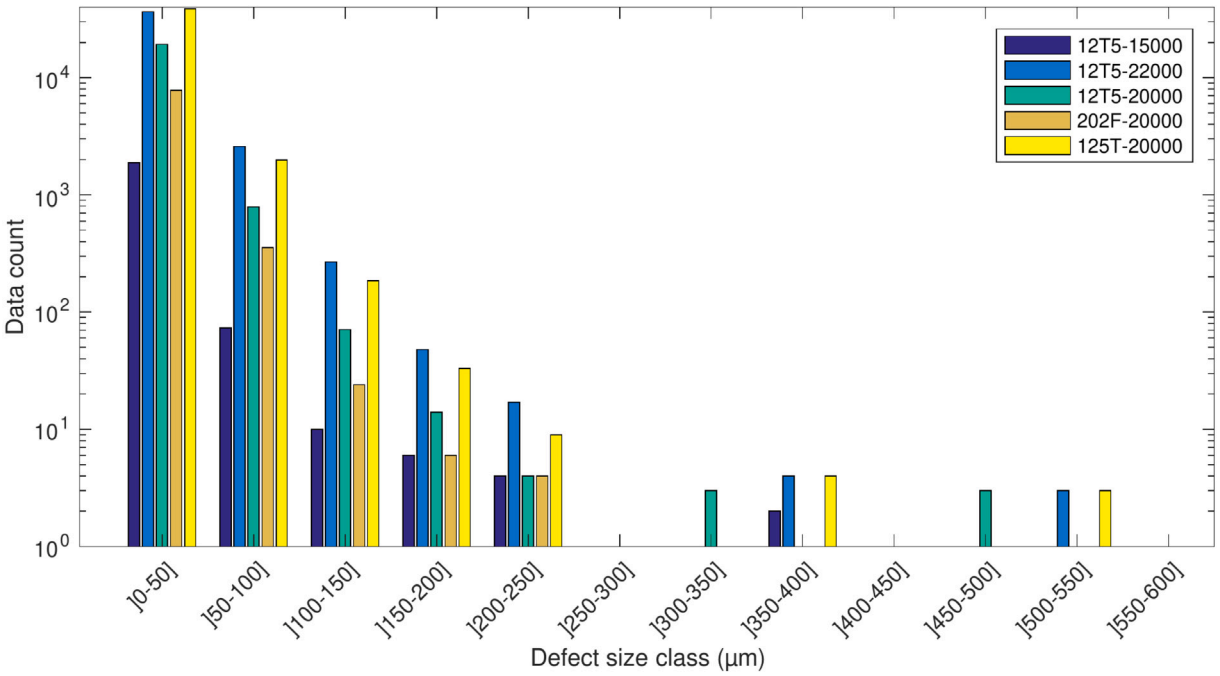


Fig. 6. Histogram of Feret diameter for 3 specimens. The legend corresponds to "specimen name - thresholding value".

inference of the defect distribution, which are consequently better depicted. The numerical value chosen here reflects the compromise which arises between the number of defects of the statistical specimen (which becomes lower with the increase of μ_e) and the weight of the larger defects in the statistical inference (which increases as μ_e approaches the largest defect). Further analysis would however be necessary to clearly define this numerical censoring value.

3.3. Statistical inference of Feret diameter size

The defects being identified, the following step establishes the best statistical distribution fitting the experimental data. The specimen 12T5 is used as the identification specimen, while specimen 125T5 and 202F are used to validate the distribution inferred.

Many distributions were tested (exponential, extreme value, gamma distribution, Rayleigh, Weibull) but only the three best fits are presented: the log-normal distribution (LOGN), the generalized extreme value distribution (GEV) and the generalized Pareto distribution (GPD).

The probability density function (PDF) of the LOGN distribution f_{logn} is:

$$f_{logn}(x) = \frac{1}{\sigma_{logn}\sqrt{2\pi}} \exp\left[-\frac{(\ln x - \mu_{logn})^2}{2\sigma_{logn}^2}\right], x > 0 \tag{1}$$

where μ_{logn} is the location parameter and σ_{logn} the scale parameter. The LOGN distribution is defined only for positive values.

The PDF of the GEV distribution, f_{gev} is given by:

$$f_{gev}(x) = \frac{1}{\sigma_{gev}}\left[1+\xi_{gev}\left(\frac{x-\mu_{gev}}{\sigma_{gev}}\right)\right]^{(-1/\xi_{gev})-1} \times \exp\{-\left[1+\xi_{gev}\left(\frac{x-\mu_{gev}}{\sigma_{gev}}\right)\right]^{-1/\xi_{gev}}\} \tag{2}$$

where μ_{gev} is the location parameter, $\sigma_{gev} > 0$ the scale parameter and ξ_{gev} the shape parameter. For the data sample analysed, ξ_{gev} is always positive, meaning the PDF is defined for $x_{gev} > \mu_{gev} - \sigma_{gev}/\xi_{gev}$.

Finally, the GPD probability density function f_{gpd} is defined by:

$$f_{gpd}(x) = \frac{1}{\sigma_{gpd}}\left(1+\frac{\xi_{gpd}(x-\mu_{gpd})}{\sigma_{gpd}}\right)^{-\frac{1}{\xi_{gpd}}-1} \tag{3}$$

Table 4
Log-likelihood of the different distributions used for inference of experimental data.

Distribution	log-likelihood
Log-normal	−6939
Generalized extreme value	−6480
Generalized pareto	−6400

where ξ_{gpd} is the shape parameter, μ_{gpd} the location parameter and σ_{gpd} the scale parameter. The GPD was introduced by Pickands [20] to approximate conditional excess distributions. Given a random variable X of cumulative distribution function F_X, the conditional excess distribution F_{μ_e} for a threshold μ_e is defined by:

$$F_{\mu_e}(y) = P(x - \mu_e \leq y | X > \mu_e) \tag{4}$$

and $y = x - \mu_e$ are the excesses. For numerous underlying cumulative distributions F_X, and for large threshold μ_e, the GPD approximates well F_{μ_e}.

The three distributions are fitted using maximum likelihood estimation. Table 4 shows the log-likelihood corresponding to the inference of the different distributions: the higher the value the better. The graphical results of the identification of the three best distributions are shown Fig. 7 and confirm the ranking obtained by comparing the log-likelihood values: the GPD is slightly better than the GEV, while the LOGN-normal does not seem appropriate for the data. Indeed, the ranking of the log-normal distribution was to be expected since it is not meant for censored data. This is of prior importance since all tomographic experimental data are censored.

In order to ascertain the representativeness of the inferred distribution, the results are compared with two other samples. While Fig. 8 shows the graphical comparison of the kernel density estimates, Table 5 shows the parameters of the fitting of the GPD on the three samples separately. Statistically, the three samples appear very close and it is reasonable to use the single distribution identified on the reference specimen (12T5).

The statistical analysis conducted here shows it is therefore possible to "prepare" the sample to fit particular needs: by removing the smaller defects, which are of less importance for fatigue behaviour, and by

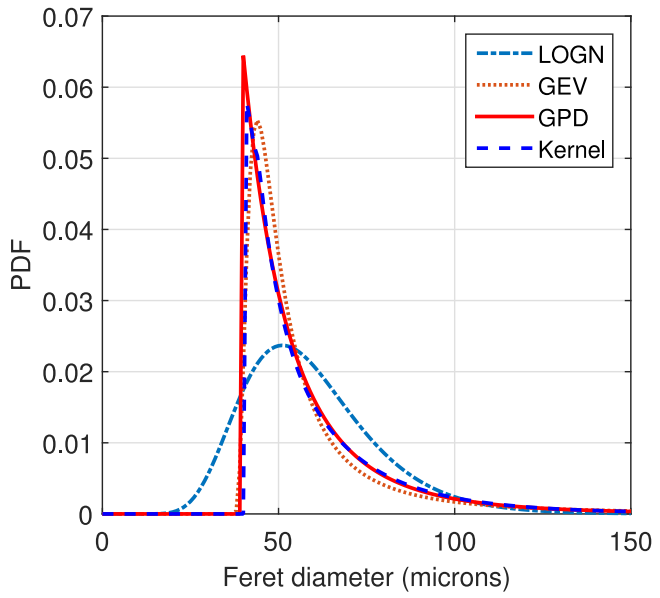


Fig. 7. Identification of the best distribution fit on specimen 12T5. LOGN is the log-normal distribution, GEV the generalized extreme value distribution and GPD the generalized Pareto distribution, while kernel is the kernel estimate.

Table 5
Parameter inference of the different specimens. The inference made on specimen 12T5 is considered as the reference. The resolution indicates the voxel size of the tomography analysis.

Specimen	Resolution	GPD parameters		
	(µm)	σ_{gpd}	ξ_{gpd}	μ_{gpd}
12T5 (ref.)	3.7	15.5	0.216	40.0
202F	3.7	15.7	0.168	40.0
125T5	5.6	16.3	0.243	40.0

choosing an appropriate distribution, which takes the censoring into account, it is possible to correctly represent the size of the larger defects.

4. Statistical description of the defect location and correlation with the defect size

To completely define a defect population, it is not only necessary to identify the size of the defects, as seen in Section 3, but also their position. In this section, the theory of point processes is first described, as well as the point processes used herein. Then, the Ripley K-function is introduced. This function helps describing and identifying a population of points. Using this theoretic background, the defect population identified by tomography is analysed with this framework. Finally, marked point processes are introduced: these processes can help better understand if there is a link, in a statistical manner, between the size and position of the defects.

4.1. Theory of point process: mathematical background and basic processes

Given a set of n points $\mathbb{X} = \{x_i \in W \subset \mathbb{R}^{d_R}, i = 1, \dots, n\}$, point process theory is used to characterize the spatial distribution of the points (usually $d_R = 2, 3$). In many cases, the process occurs in a large subset $S_{sub} \subset \mathbb{R}^{d_R}$, but can only be observed in a smaller window $W \subset S_{sub}$. Specific tools were developed to infer properties of the underlying process given a limited observation window W.

A point process $\mathbb{X}$ is a given set of $n_{\mathbb{X}}$ points, $n_{\mathbb{X}}$ also being a random variable, defined on a subset $S_{sub} \subset \mathbb{R}^{d_R}$:

$$\mathbb{X} = \{x_i, i = 1, \dots, n_{\mathbb{X}}\} \tag{5}$$

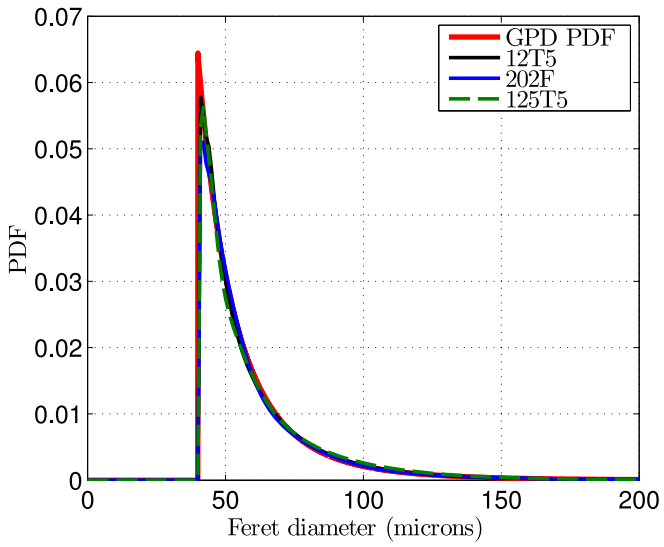


Fig. 8. Comparison of the generalized pareto probability density function (GPD PDF), identified on specimen 12T5, with the kernel density estimates of specimen 12T5, 202F and 125T5.

where the x_i are also random vectors. For ease of calculations, the considered observation windows are all rectangles (2D) or rectangular cuboids (3D):

$$W = [x_{min}, x_{max}] \times [y_{min}, y_{max}] \quad \text{(for 2D case)} \tag{6}$$

$$W = [x_{min}, x_{max}] \times [y_{min}, y_{max}] \times [z_{min}, z_{max}] \quad \text{(for 3D case)} \tag{7}$$

Hypotheses are made concerning the studied point processes:

- Simple point process: there are not repeated points, i.e. $x_i \neq x_j$ if $i \neq j$;
- Homogeneous (stationary) point process: all properties of the point process are not modified by a translation.

The Poisson point process. A stationary point process is a Poisson point process if:

- the number of points in a subset $B \subset W$ follows a Poisson distribution of parameter $\lambda|B|$, $|B|$ being the volume of B:

$$\mathbb{P}(N(B) = k) = \frac{[\lambda|B|]^k}{k!} e^{-\lambda|B|} \tag{8}$$

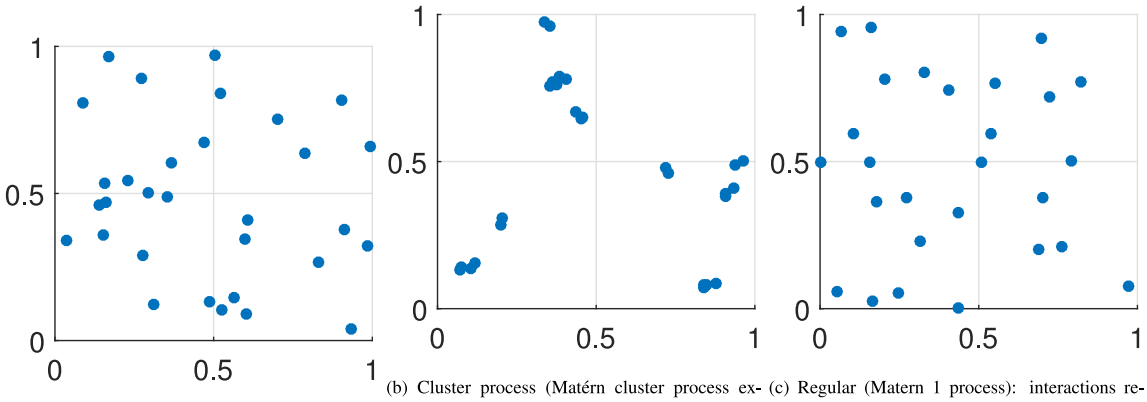
- λ is the point process intensity, which corresponds to the average number of points per volume unit ($\mathbb{E}$ being the statistical expected value):

$$\lambda = \frac{\mathbb{E}[\mathbb{N}(\mathbb{B})]}{|B|} \tag{9}$$

- For m disjoints sets $B_1, \dots, B_m$, the random variables $N(B)_1, \dots, N(B)_m$ are independent.

The Poisson point process is used to represent Complete Spatial Randomness (CSR): each point is stochastically independent, and there is absolutely no interaction between them (see Fig. 9(a)). Therefore, the Poisson point process serves as reference to evaluate if a point process is clustered (the points are attracted to each other, see Fig. 9(b)) or regular (the points repulse each other, see Fig. 9(c)).

Cluster process. Cluster point processes occur when groups of points, or "clusters", form: in each cluster, the distance between points is smaller than the overall average smallest distance between points. In nature, these appear when parents (for example a tree), produces children (for example seeds), that stay in close proximity of their parent.



(a) Poisson: the points are randomly distributed.

(b) Cluster process (Matérn cluster process example): points are attracted and appear in small groups called clusters.

(c) Regular (Matern 1 process): interactions repulses points, there is a low probability of having two points nearby.

Fig. 9. Different examples of point processes.

Mathematically, a cluster point process is defined by:

- A parent point process $\mathbb{X}_p$;
- For each parent $c \in \mathbb{X}_p$, a child point process $\mathbb{Z}_c$.

The cluster point process $\mathbb{X}$ is therefore:

$$\mathbb{X} = \bigcup_{c \in \mathbb{X}_p} \mathbb{Z}_c \tag{10}$$

The parent process $\mathbb{X}_p$ is used solely in the construction process and is not included in the final process, and parent points are not considered in the final sample.

Neyman–scott processes: Matérn-cluster and modified Thomas processes. A Neyman–Scott process [21] is defined by 3 steps (see [22], p. 384):

1. The parent process $\mathbb{X}_p$ is a homogeneous Poisson point process of intensity λ_p;
2. Each parent $c \in \mathbb{X}_p$, produces N_{ch} children points given by a discrete probability distribution. For the considered cases, the distribution is a Poisson distribution with parameter $\bar{c}$:

$$\mathbb{P}(N_{ch} = k) = e^{-\bar{c}} \frac{\bar{c}^k}{k!} \tag{11}$$

3. The N_{ch} children are independently drawn around the cluster centre c (rigorously, the children probability distribution is defined conditionally by the cluster centre).

A particular Neyman–Scott process is the Matérn cluster. The children are randomly distributed in a R_0 radius ball centred on the parent point c:

$$z \in \mathbb{Z}_c \sim \mathcal{U}(b(c, R_0)) \tag{12}$$

where $\mathcal{U}$ is the uniform distribution, and $b(c, R_0)$ the ball of radius R_0 centred on c. Fig. 10 illustrates such a point process.

A second specific Neyman–Scott process considered here is the Modified Thomas process. The children are distributed around the parent point with a multivariate ($d_R = 2, 3$ following the space dimension) Gaussian distribution centred on the parent point c, and of covariant matrix $\sigma_T^2 \mathbf{I}_{d_R}$ (the d_R-dimension identity matrix $\mathbf{I}_{d_R}$ implies there is no correlation between the different dimensions):

$$z \in \mathbb{Z}_c \sim \mathcal{N}(c, \sigma_T^2 \mathbf{I}_{d_R}) \tag{13}$$

where $\mathcal{N}$ is the multivariate Gaussian distribution. An example of such a point process is given in Fig. 11.

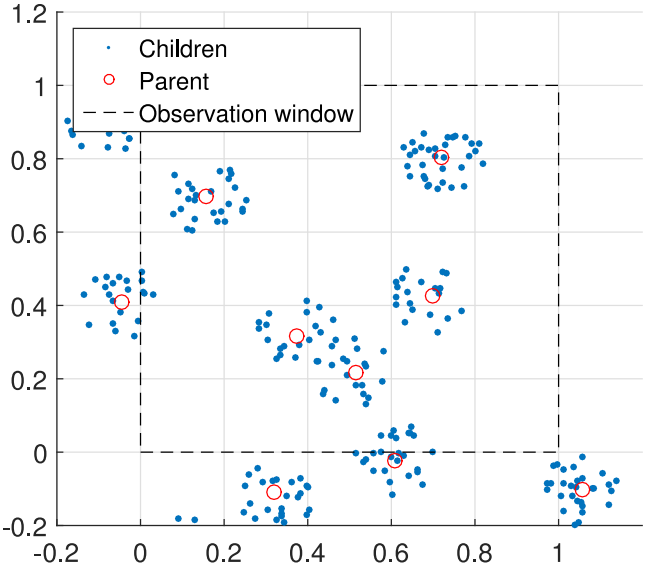


Fig. 10. Example of Matérn cluster point process: $\lambda_p = 35, \bar{c} = 25, R_0 = 0.1$. The unity observation window shows children points can be observed, despite the parent point being outside the window.

4.2. The Ripley K-function

After the presentation of the point process of interest, it is necessary to find means to choose between them, as well as to identify the different parameters. The K Ripley function is often chosen to do so. Given a point process $\mathbb{X}$, the Ripley's K function is defined by:

$$K(r) = \frac{1}{\lambda} \mathbb{E}_{\boldsymbol{x}} \left[N\Big(b(\boldsymbol{x}, r) \backslash \{\boldsymbol{x}\} \Big) \right], \quad r > 0 \tag{14}$$

where λ is the intensity of the point process. The function is the average number of points in a r-radius sphere, where each sphere center is a point of the process itself but is not counted.

For example, given a stationary Poisson point process (PPP) in a d_R dimension, the Ripley's K function is exactly the volume of the sphere $V_{d_R}(r)$ (since the K function is normalized by $1/\lambda$):

$$K^{PPP}(r) = \pi r^2 \qquad \text{(for 2D case)} \tag{15}$$

$$K^{PPP}(r) = \frac{4}{3} \pi r^3 \qquad \text{(for 3D case)} \tag{16}$$

For a clustered process, and a small radius r_1, there is a locally higher number of points (given the center of the counting sphere is also

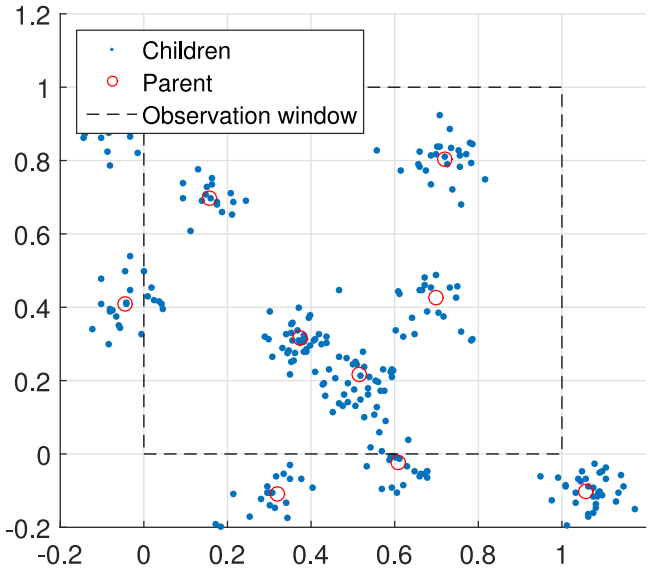


Fig. 11. Example of modified Thomas cluster point process: $\lambda_p = 35, \bar{c} = 25, \sigma_T = 0.05$. the infinite support of the Gaussian distribution can create children who appear isolated.

part of the daughter point process), than for a Poisson point process which represents Complete Spatial Randomness. Therefore, locally, the clustered $K^{cluster}$ Ripley function follows:

$$K^{cluster}(r_1) > K^{PPP}(r) = V_{d_R}(r) \tag{17}$$

Conversely, for a regular point process, the $K^{regular}$ Ripley function follows:

$$K^{regular}(r) < K^{PPP}(r) = V_{d_R}(r) \tag{18}$$

In order to provide robust parameter inference, analytical formulas of the different Ripley K functions considered are preferred. Of concern here are the Matérn and modified Thomas (see [22], p. 376–377). The modified Thomas point process could not be found in 3 dimensions in the scientific literature and had to be derived.

For the Neyman–Scott process, the K^{NS} Ripley function can be written as:

$$K^{NS}(r) = K^{PPP}(r) + \frac{\mathbb{E}[N(N-1)]}{\lambda_p \bar{c}^2} F_d(r) \tag{19}$$

where λ_p and $\bar{c}$ are the parent point process intensity and the average number of children, respectively. F_d is the cumulative distribution function of the distance between two children points of a same cluster.

For Matérn and modified Thomas point process, the expected value $\mathbb{E}[N(N-1)]$ can be simplified: both processes parent points generate N children following a Poisson of expected value $\bar{c}$ (which is also equal to the variance). Therefore, $\mathbb{E}[N(N-1)] = \bar{c}^2$ and specifically for Matérn and modified Thomas point processes:

$$K^{NS}(r) = K^{PPP}(r) + \frac{1}{\lambda_p} F_d^{NS}(r) \tag{20}$$

while some cumulative distribution functions can be found in [22], others had to be derived for some cases. The results can be found in Table 6.

4.3. Estimation of point process parameters on a dataset: application on observed tomographic data

The sample used for the estimation of spatial characteristics is the same as the one used for the defect size distribution: the specimen 12T5 is studied with a grey level thresholding of 20 000 and censoring of

Table 6
Analytical probability density function (PDF) and cumulative distribution function (CDF) of the distance between two children points of a same cluster (see [22]).

Matérn 3D	PDF	$f_d(r) = \frac{3r^2}{2R_0^6}\left(R_0 - \frac{r}{2}\right)^2\left(2R_0 + \frac{r}{2}\right)$
	CDF	$F_d(r) = \frac{3}{16R_0^6}\left(\frac{r^6}{6} - 3R_0^2 r^4 + \frac{16}{3}R_0^3 r^3\right)$
Modified Thomas 3D	PDF	$f_d(r) = \frac{r^2}{\sqrt{4\pi}\sigma_T^3}\exp\left(-\frac{r^2}{4\sigma_T^2}\right)$
	CDF	$F_d(r) = 2\Phi\left(\frac{r}{\sqrt{2}\sigma_T}\right) - \frac{\sqrt{2}r}{\sigma_T}\varphi\left(\frac{r}{\sqrt{2}\sigma_T}\right) - 1$

With $\varphi(x) = 1/\sqrt{2\pi}\exp(-x^2/2)$ and $\Phi(x) = \int_{-\infty}^{x}\varphi(t)dt$.

defects smaller than 40 µm. Each defect is represented by a point. The point coordinates being the barycentre of the voxels which belong to the defect.

In order to correctly analyse the sample, the estimation methods require cuboid samples. Therefore, the largest possible cuboid is extracted from the initially cylindrical sample: the remain cuboid is 159.25 µm^3 and 1323 data points remain.

The first and most important parameter to estimate, is the process intensity λ. For an observation window W, the average number of points in this window is:

$$\mathbb{E}[N(W)] = \lambda|W| \tag{21}$$

where $|W|$ is the window volume (or surface in the 2D case). For a particular process realization of n points, an unbiased intensity estimate $\hat{\lambda}$ is:

$$\hat{\lambda} = \frac{n}{|W|} \tag{22}$$

The estimated intensity of the observed sample is therefore:

$$\hat{\lambda} = \frac{1323}{159.29} = 8.31 \text{ points/mm}^3 \tag{23}$$

Once the intensity has been estimated, it is necessary to estimate the K Ripley function. For this, it is first necessary to calculate the symmetric distance matrix (d_{ij}). For n observed points $(\boldsymbol{x}_i)_{\{i=1\ldots n\}}$, the $n \times n$ distance matrix is:

$$d_{ij} = \|\boldsymbol{x}_i - \boldsymbol{x}_j\| \tag{24}$$

where $\|\bullet\|$ is the usual Euclidean distance. For a finite observation window W, a biased estimate of λK can then be derived:

$$\lambda\hat{K}_{bias}(r) = \frac{1}{n}\sum_{i=1}^{n}\sum_{i\neq j}\#\{d_{ij} \leq r\} \tag{25}$$

where $\#(\bullet)$ counts the number of data that verifies the condition. This estimate places n radius balls centred on all the $\boldsymbol{x}_i$ (which belong to the point process), and counts the average number or points in the balls. In Fig. 12, the counting centred on $\boldsymbol{x}_1$ provides an accurate estimate of the local intensity. For the ball centred on $\boldsymbol{x}_2$, the number of process points is artificially lower due to the unobserved points outside the W window, even if $r_1 = r_2$, thus causing a strong bias in the estimator.

Different methods can be used to avoid this strong edge effect and provide an unbiased estimate (see [22], Section 4). For this study, the translation correction is chosen. For a homogeneous point process, the edge effect can be corrected by using a translated window $W_{\boldsymbol{x}^*}$:

$$W_{\boldsymbol{x}^*} = \{\boldsymbol{u} + \boldsymbol{x}^*, \boldsymbol{u} \in W\} \tag{26}$$

which is used to calculate the correction factor w^{trans}:

$$w^{trans}(\boldsymbol{x}_i, \boldsymbol{x}_j) = \frac{|W \cap W_{\boldsymbol{x}_j - \boldsymbol{x}_i}|}{|W|} \tag{27}$$

An example is shown in Fig. 13. The area for two points $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$, $|W \cap W_{\boldsymbol{x}_2 - \boldsymbol{x}_1}|$, corresponds to the shaded area. The K function estimate

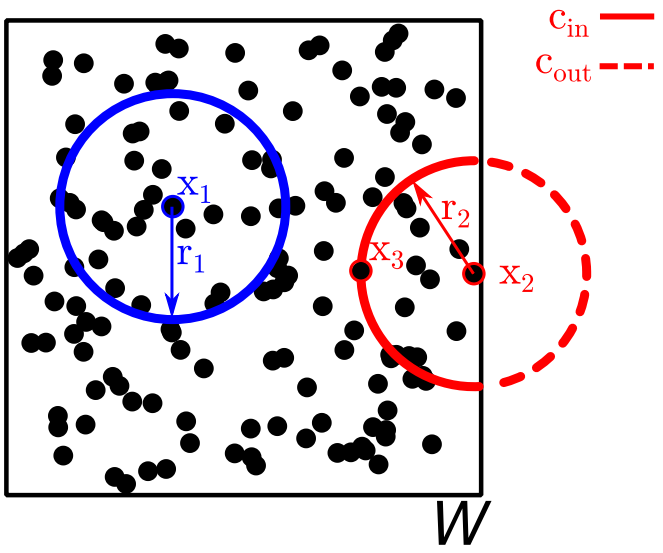


Fig. 12. Ripley's K function estimate: the ball centred on x_1 correctly estimate the local intensity (number of points divided by circumference), whereas it is clearly underestimated by the ball centred on x_2.

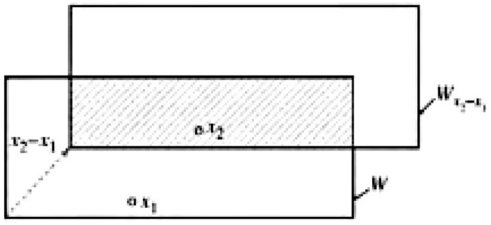


Fig. 13. Translation corrected estimation of the K Ripley function, see [22] chapter 4.

is then:

$$\lambda \hat{K}_{trans}(r) = \frac{1}{n}\sum_{i=1}^{n}\sum_{i\neq j}\frac{1}{w^{trans}(\boldsymbol{x}_i,\boldsymbol{x}_j)}\#\{d_{ij}\leq r\} \tag{28}$$

Underlying this correction is the probability of observing two points far from each other. For very close points, the window bias is rather small and only a small number of pair of points are unobserved. The further apart the points are (i.e. for large r), the larger number of pair of points are not counted.

Now the function estimate has been derived, it is necessary to estimate the best set of parameters of the selected models (Poisson, Matérn and modified Thomas). For an experimental set of points $\mathbb{X} = \{\boldsymbol{x}_i, i = 1,\dots,n_{\mathbb{X}}\}$, the experimental estimate $\hat{K}_{exp}(r)$ function is calculated (in this study, $\hat{K}_{exp} = \hat{K}_{trans}$) . For a set of parameters $\boldsymbol{\theta}$, the experimental function is compared with the analytical Ripley function $K_{\boldsymbol{\theta}}(r)$ (see Eqs. (16) and (19)) by using the contrast function $D(\boldsymbol{\theta})$:

$$D(\boldsymbol{\theta}) = \int_{r_{min}}^{r_{max}} \left[\hat{K}_{exp}(r)^{q_1} - K_{\boldsymbol{\theta}}(r)^{q_1}\right]^{q_2} dr, \tag{29}$$

where $(q_1, q_2) \in \mathbb{R}^2$. This method consist in minimizing a certain q_1-norm of the Ripley functions (to the power of q_2) between two radii r_{min} and r_{max}.

For a clustered process, the different values suggested in [23] are used:

- $r_{min} = 0$ and r_{max} is half the minimum observation window size;
- $q_2 = 2$ as in the least square method;
- $q_1 = \frac{1}{4}$

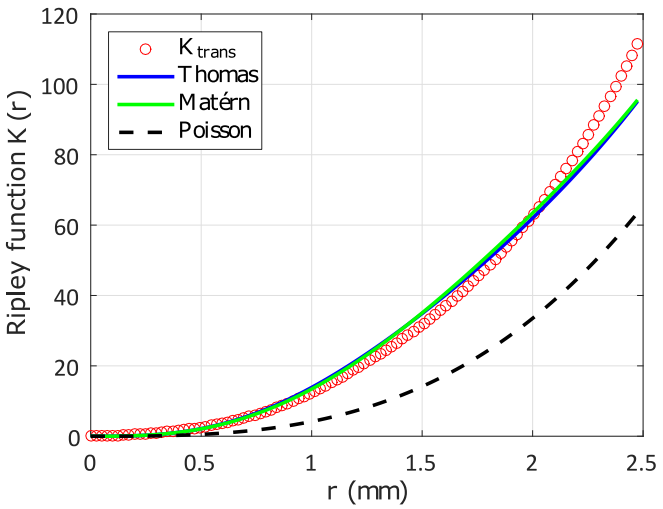


Fig. 14. Spatial statistics inference: the experimental Ripley's function K_{trans} is close to the optimized Matérn and Thomas point process ones. The Poisson point process distribution is unable to capture the spatial data.

Table 7
Inferred parameters for the T5 specimen with the data thresholded at 20 000, and the length filtered $x_f > 40$ µm. D parameter Eq. (29) given for information.

	$\hat{\lambda}$ (points/mm^3)	λ_p (points/mm^3)	σ_T (mm)	R_0 (mm)	$\bar{c} = \hat{\lambda}/\lambda_p$	D
Poisson	8.31	–	–	–	–	0.38
Thomas	–	0.0305	0.5973	–	272.8	0.017
Matérn	–	0.0312	–	1.2534	266.6	0.019

The method is then used to find the best parameters $\boldsymbol{\theta}^*$ in the admissible space $D_{\boldsymbol{\theta}}$:

$$\boldsymbol{\theta}^* = \underset{\boldsymbol{\theta}\in D_{\boldsymbol{\theta}}}{\arg\min}\, D(\boldsymbol{\theta}) \tag{30}$$

This optimal parameter set is achieved by minimizing the $D(\boldsymbol{\theta})$ function (using a hybrid genetic algorithm). This methodology is applied for the selected models and the results shown in Table 7 and Fig. 14.

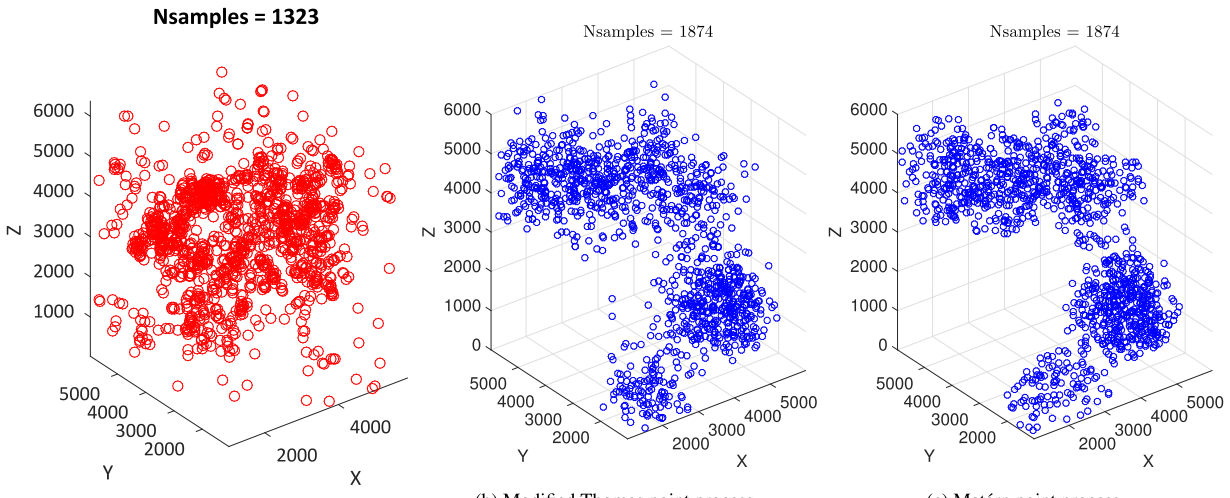
The Poisson point process is clearly unable to represent the observed data, whereas both the modified Thomas and the Matérn point process give a better and similar fit.

However, the clustered simulated processes slightly overestimates the Ripley's K function for radii r between 1 and 2 mm, and underestimate for $r > 2$ mm. As seen in Fig. 15, the clustered process is unable to capture the seemingly isolated data points outside the clusters shown on real data (see Fig. 15(a)). To maintain the global intensity of the process, the number of children $\bar{c}$ has to be higher to compensate, thereby increasing Ripley's K function for short distances ($r < 2$ mm). For the longer ranges ($r > 2$ mm), there are no isolated data points and the simulated Ripley's K function is lower than observed. The isolated points could be linked to isolated gas porosities, and an added Poisson point process could be superimposed in order to simulate these isolated data points. However, the parameter inference robustness would suffer from the added complexity.

4.4. Is there a link between defect size and their location? use of the marked point process

In order to study the correlation between the defects position and size (hereafter considered as points), marked point processes can be used. For a point, the mark can be a qualitative or a quantitative one, and represents a characteristic. For the scope of this study, the mark of interest is therefore the defect size.

Several theoretical models can be applied to study marked point processes:



(a) 3d representation of the data points used.

(b) Modified Thomas point process.

(c) Matérn point process

Fig. 15. Comparison of experimental and simulated samples.

- *Random marks*: for this case, spatial and mark distribution are considered independent. The spatial distribution is characterized and for each point, the mark is considered as independent of its position (and consequently, independent of its neighbours marks).
- *Random superposition*: for quantitative marks m defined in $\{m_1, \dots, m_Q\}$ with $Q \in \mathbb{N}$, Q sub process are defined for each mark class. The final process is defined by superposition of the Q sub processes $\{\mathbb{X}_i, i = 1, \dots, Q\}$:

$$\mathbb{X}_i = \{\boldsymbol{x}_j^{(i)}, j = 1, \dots, n_i\}, i = 1, \dots, Q \tag{31}$$

$$\mathbb{X} = \bigcup_{i=1}^{Q} (\mathbb{X}_i \times \{m_i\}) = \left\{ (\boldsymbol{x}_j^{(i)}, m_i), j = 1, \dots, n_i; i = 1, \dots, Q \right\} \tag{32}$$

- *Geostatic marks*: here, the marks are defined based on a random continuous field $Z(\boldsymbol{x}, \omega)$. Once the point position $\boldsymbol{x}_j$ is known, the corresponding mark is $Z(\boldsymbol{x}_j)$. Hence the process $\mathbb{W}$:

$$\mathbb{X} = \left\{ \left(\boldsymbol{x}_j, Z(\boldsymbol{x}_j)\right), j = 1, \dots, n \right\} \tag{33}$$

For this section, sample "12T5-20 000" is used (i.e. sample 12T5 with a 20 000 grey threshold,). However, since the largest defects are no longer the main concern (i.e. the right-tail of the distribution), a compromise is set to increase the number of defects while excluding the smaller ones (which are mainly noise related). Therefore, only defects smaller than 10 µm are excluded (rather than the 40 µm threshold used for the defect size distribution study). The defect size measure remains the 3D Feret diameter which is considered as the point mark.

First order characteristics. For qualitative marks $m_1, \dots, m_Q$, the probability distribution is defined by:

$$p_i = \mathbb{P}(M = m_i) \tag{34}$$

The probability p_i is easily obtained with the sub process intensity λ_i containing solely the marks m_i:

$$p_i = \frac{\lambda_i}{\lambda} \tag{35}$$

where λ is the intensity of the spatial point process (i.e. all point considered, with no mark differentiation).

For quantitative marks $M \in D_M$, let us consider a subspace $B_{sub} \in W$ and a subset $]-\infty, m_0]$. For a given m_0, a sub process can be defined by considering all marks such as $m \leq m_0$. The number of points in this subspace is therefore:

$$\mathbb{E}[N(B_{sub}; m(\boldsymbol{x}) \leq m_0)] \overset{\text{def}}{=} \lambda(m_0)|B| \tag{36}$$

where $|B_{sub}|$ is the volume of B_{sub} (in 3 dimensions). When m_0 tends to infinity, all the marks are considered, and therefore:

$$\lim_{m_0 \to \infty} \lambda(m_0) = \lambda \tag{37}$$

Consequently, the cumulative distribution of marks F_M is defined by:

$$F_M(m) = \frac{\lambda(m)}{\lambda} \tag{38}$$

from which the probability density function (pdf) can be obtained $f_M(m) = F'_M(m)$. The mean and variance are defined by:

$$\mu_m = \int_{D_M} m f_M(m) dm \tag{39}$$

$$\sigma_m^2 = \int_{D_M} (m - \mu_m)^2 m f_M(m) dm \tag{40}$$

These different quantities can be easily estimated. For an observed volume W, the mark mean estimate is:

$$\hat{\mu_m} = \frac{1}{n} \sum_{\boldsymbol{x} \in W} m(\boldsymbol{x}) \tag{41}$$

and the mark variance:

$$\hat{\sigma_m}^2 = \frac{1}{n} \sum_{\boldsymbol{x} \in W} (m(\boldsymbol{x}) - \hat{\mu_m})^2 \tag{42}$$

Nearest neighbour correlation index. The nearest neighbour correlation index aims at pointing out if there is a correlation between the mark of a point, and the mark of the nearest point. For $\boldsymbol{x} \in \mathbb{X}$, the nearest neighbour is referred to as nei($\boldsymbol{x}$):

$$\text{nei}(\boldsymbol{x}) = \underset{\boldsymbol{y} \in \mathbb{X} \setminus \{x\}}{\arg\min} \|\boldsymbol{y} - \boldsymbol{x}\| \tag{43}$$

Using the nearest neighbour, three indexes can be defined [22]:

- The nearest neighbour mean:

$$\boldsymbol{n}_1 = \mathbb{E}_{\boldsymbol{x}}[m(\text{nei}(\boldsymbol{x}))]/\mu_m \tag{44}$$

- The mean mark product:

$$\boldsymbol{n}_{mm} = \mathbb{E}_{\boldsymbol{x}}[m(\boldsymbol{x})m(\text{nei}(\boldsymbol{x}))]/\mu_m^2 \tag{45}$$

- The variographic mark index:

$$\boldsymbol{n}_\gamma = \mathbb{E}_{\boldsymbol{x}}\left[(m(\boldsymbol{x}) - m(\text{nei}(\boldsymbol{x})))^2\right]/\sigma_m^2 \tag{46}$$

Table 8
Nearest neighbour correlation index for specimen 12T5.

Index	n_1	n_{mm}	n_γ
Value	0.9985	1.0001	0.9509

These indexes are to be distinguished from correlation indexes, for they can be greater than 1. Due to the normalizations, they are to be compared to one: a value close to one indicating a lack of correlation. These index estimates are readily obtained, for edge effect introduces only a small bias and no edge correction method are necessary [24].

These indexes are calculated for the tomographic observed data and shown in Table 8. All the values are all close to one, which is a first step indicating the lack of correlation between defect size and position.

K-Ripley and L-Besag functions. Detailed study of marked point process can be achieved by using the weighted Ripley K function. For an unmarked point process, the Ripley K function (in Eq. (14)) can be rewritten:

$$K(r) = \frac{1}{\lambda}\mathbb{E}_{\boldsymbol{x}}\left[\sum_{\boldsymbol{y}\in\mathbb{X}\setminus\boldsymbol{x}} \mathbf{1}_{b(\boldsymbol{x},r)}(\boldsymbol{y})\right] \tag{47}$$

where $\mathbf{1}_{b(\boldsymbol{x},r)}$ is the indicator function centred on $\boldsymbol{x}$ of radius r. The weighted Ripley K-function K_{mm} is:

$$K_{mm}(r) = \frac{1}{\lambda\mu_m^2}\mathbb{E}_{\boldsymbol{x}}\left[\sum_{\boldsymbol{y}\in\mathbb{X}\setminus\boldsymbol{x}} m(\boldsymbol{x})m(\boldsymbol{y})\mathbf{1}_{b(\boldsymbol{x},r)}(\boldsymbol{y})\right]. \tag{48}$$

As for unmarked point process, an L Besag function can be defined using the unit volume sphere in d dimension V_{d_R}:

$$L_{mm}(r) = \left(\frac{K_{mm}(r)}{V_{d_R}(1)}\right)^{1/d}, \tag{49}$$

which translates in 3 dimensions to:

$$L_{mm}(r) = \left(\frac{3K_{mm}(r)}{4\pi}\right)^{1/3}. \tag{50}$$

As for unmarked point process, the K_{mm} Ripley weighted function estimation requires an edge correction method. The same method of translation correction is applied (see Eq. (28)):

$$\lambda\mu_m^2\hat{K}_{mm,trans}(r) = \frac{1}{n}\sum_{i=1}^{n}\sum_{i\neq j}\frac{m(\boldsymbol{x}_i)m(\boldsymbol{x}_j)}{w^{trans}(\boldsymbol{x}_i,\boldsymbol{x}_j)}\#\{(\boldsymbol{x}_i,\boldsymbol{x}_j)\in W : d_{ij}\leq r\} \tag{51}$$

The weighted K_{mm} Ripley function is compared to the K Ripley function, which is calculated without the marks. If spatial and mark distribution are uncorrelated, these two functions are equal.

The L Besag function can also be estimated by:

$$\hat{L}_{mm,trans}(r) = \left(\frac{3}{4\pi\hat{\lambda}\hat{\mu_m}^2}\frac{1}{n}\sum_{i=1}^{n}\sum_{i\neq j}\frac{m(\boldsymbol{x}_i)m(\boldsymbol{x}_j)}{w^{trans}(\boldsymbol{x}_i,\boldsymbol{x}_j)}\#\{(\boldsymbol{x}_i,\boldsymbol{x}_j)\in W : d_{ij}\leq r\}\right)^{1/3} \tag{52}$$

For reading easiness, the estimate functions using translation correction are hereafter simply referred to as $\hat{K}_{mm}$ and $\hat{L}_{mm}$.

Lastly, a mark independence test can be used to test spatial and mark distributions independence. Given a sample, marks of this sample are randomly permuted creating B_{sub} artificial samples. For each artificially generated sample k, the L Besag function $\hat{L}_{mm}^{(k)}$ is calculated creating a sheaf of curves. The extreme curves are then defined for each abscissa r by:

$$L_{mm,min}(r) = \min_{k=1,\dots,B}\hat{L}_{mm}^{(k)}(r) \tag{53}$$

$$L_{mm,max}(r) = \max_{k=1,\dots,B}\hat{L}_{mm}^{(k)}(r) \tag{54}$$

Independence between spatial and mark distribution is then checked by plotting $L_{mm,min} - \hat{L}_{mm}$ and $L_{mm,max} - \hat{L}_{mm}$ and comparing with the null hypothesis, represented by the abscissa axis.

The K_{mm} weighted Ripley function is shown in Fig. 16(a). Both the weighted and the standard Ripley function appear perfectly equal, confirming the independence of point mark and position indicated by the different indexes (see Table 8). However, the mark independence test shown in Fig. 16(b) shows there is in fact a small interaction of marks for shorter distances.

In conclusion, the study of marked point process shows there is only a small link between size and position of the defects. It is therefore reasonable to simulate defect position (may it be clustered or completely random) and to subsequently affect a size, following a specific probability distribution, to each defect independently of its position in space.

5. Extreme value statistics and critical defects

Now that populations of defects can be simulated, this section aims at comparing the critical defects observed in fatigue tests to the defect distribution identified on virgin specimens by tomography.

5.1. Mathematical background and analytical study using extreme value statistics

Let $X_1, X_2, \dots, X_n$ be n independent and identically distributed random variables of probability distribution function (pdf) f_X and cumulative distribution function (cdf) F_X.

Let $M_n = \max\{X_1, \dots, X_n\}$ be the maximum of the n random variables, the exact maximum distribution can be derived:

$$\Pr(M_n \leq y) = \Pr(X_1 \leq y, \dots, X_n \leq y) = F_X(y)^n = G_X(y) \tag{55}$$

For the study of defects, f_{gpd} is the probability density function of the defect size distribution (identified on specimen 12T5 and shown in Table 5):

$$f_{gpd}(x) = \frac{1}{\sigma_{gpd}}\left(1 + \frac{\xi_{gpd}(x-\mu_{gpd})}{\sigma_{gpd}}\right)^{-\frac{1}{\xi_{gpd}}-1} \tag{56}$$

and F the corresponding cdf:

$$F_{gpd}(x) = 1 - \left(1 + \frac{\xi_{gpd}(x-\mu_{gpd})}{\sigma_{gpd}}\right)^{-\frac{1}{\xi_{gpd}}} \tag{57}$$

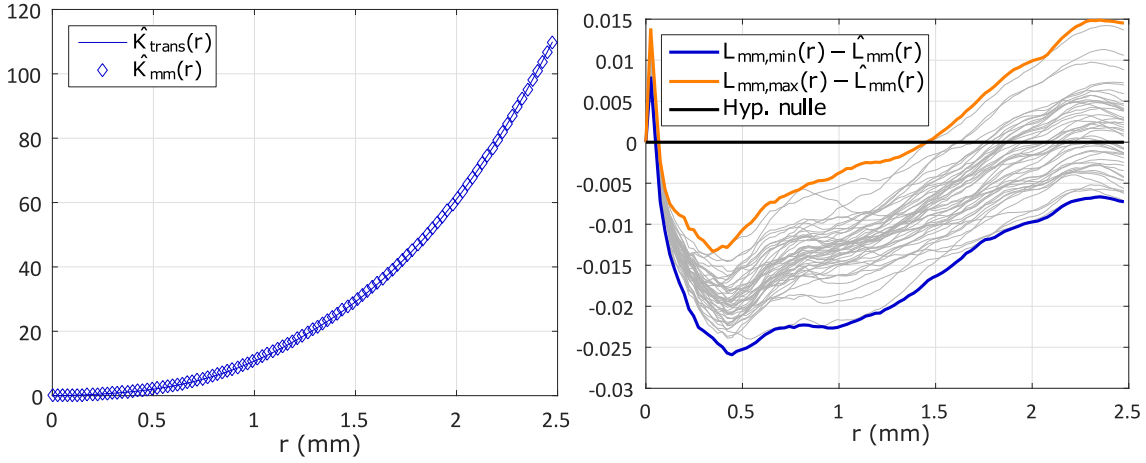
Supposing there is an identical number n of defects per specimen, the cdf of maximum defect size G_{gpd} is:

$$G_{gpd}(y) = F_{gpd}(y)^n = \left(1 - \left(1 + \frac{\xi_{gpd}(y-\mu_{gpd})}{\sigma_{gpd}}\right)^{-\frac{1}{\xi_{gpd}}}\right)^n \tag{58}$$

Following Eq. (58), the number of defects per specimen n appears critical to derive the maximum size distribution G. This can be seen in Fig. 17. For $n = 1000$ defects to $n = 10000$, the median goes from 315 µm to 540 µm, and the 95% percentile (i.e. the 5% of larger defects) from 575 µm to 960 µm.

5.2. Specimen choice and statistical representative volume

Cylindrical specimen. Given a specimen geometry of volume $|W|$ (corresponding to the cylindrical fatigue tested zone, $|W| = 461\ \text{mm}^3$), the number of defects per specimen depends on the chosen spatial point process:



(a) K Ripley function and weighted K_{mm} Ripley function

(b) Mark independence test.

Fig. 16. Study of spatial and mark distributions correlation by marked point processes.

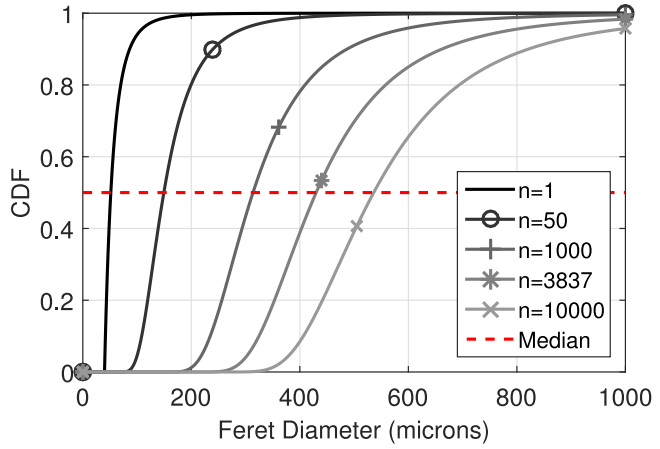


Fig. 17. Cumulative distribution function of maximum defect size G_{gpd} for different number of defects per specimen n. 3837 is the number of defects of the fatigue tested sample (a 7 mm diameter and 12 mm high cylinder with a $8.31\,\mathrm{mm}^{-3}$ density of defects) if the number of defects is considered deterministically proportional to the volume.

- Homogeneous distribution (see Fig. 18(a)): the deterministic number of defects n is proportional to the volume of the sample. The defect concentration is the same as the intensity λ identified in the point process statistics:

$$\lambda = 8.31\,\mathrm{mm}^{-3} \tag{59}$$

and the number of defects per sample is therefore:

$$n = \lambda |W| = 3837 \tag{60}$$

- Poisson distribution (see Fig. 18(b)): the number of defects in a sample is the realization of a Poisson distribution. The variable n is therefore a random one:

$$n \sim \mathcal{P}(\lambda |W|) \tag{61}$$

where $\mathcal{P}$ is the Poison distribution, λ the process intensity ($\lambda = 8.31\,\mathrm{mm}^{-3}$) and $|W|$ the window volume

- Clustered point process (see Fig. 18(c)): the number of defects per sample depends on the parameters of the point process and the geometrical shape of the sample. Each sample has to be numerically simulated, and the generated defects counted.

The different empirical cumulative distributions of number of defects are shown in Fig. 19, and some characteristic statistical values in Table 9. The homogeneous and the Poisson point process provide analytical values, whereas for the clustered processes, the different distributions and values were obtained by simulating 10 000 samples.

The deterministic homogeneous method and the Poisson process provide similar results, with only a slight variance for the Poisson process. Using the clustered point processes (Matérn or Thomas), the variance increases drastically. This increase can be explained by the variable number of parent points. For the Thomas point process, the parent point process intensity λ_p is:

$$\lambda_p = 0.0305\,\mathrm{mm}^{-3} \tag{62}$$

The Poisson distribution of the number of parents n_p is therefore of parameter $\bar{p}$:

$$\bar{p} = \lambda_p |W| = 14.1 \tag{63}$$

i.e. there are 14.1 parents per sample in average. The standard deviation σ_{n_p} of the Poisson distribution can also be calculated:

$$\sigma_{n_p} = \sqrt{\lambda_p |W|} = 3.75 \tag{64}$$

The number of children per parent has a mean value $\bar{c} = 272.8$ and a standard deviation σ_c:

$$\sigma_c = \sqrt{\bar{c}} = 16.52 \tag{65}$$

The total number of points being the product of the number of parent points by the number of children (both random variables being independent), the standard deviation of the total number of points σ_n can be calculated by:

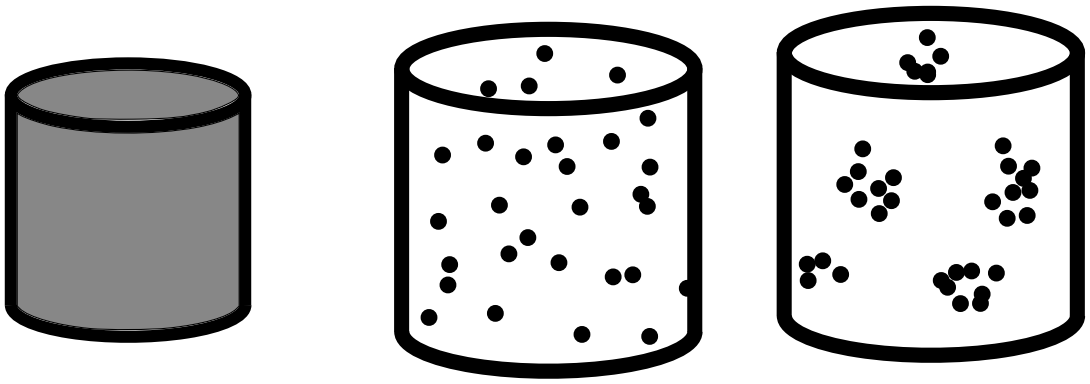
$$\sigma_n^2 = \sigma_{n_p}^2 \sigma_c^2 + \bar{p}^2 \sigma_c^2 + \bar{c}^2 \sigma_{n_p}^2 \tag{66}$$

Which numerically is:

$$\sigma_n = 1048.1 \tag{67}$$

The discrepancy with the observed value (950.3) can be explained by two factors. Firstly, for the simulations, parent points outside the sample could generate children points inside the sample (this is overcome by simulating a larger sample from which the sample of interest is taken). Secondly, some parent points generate children points outside the sample of interest which are not counted.

Similarly, for the Matérn point process, the calculated approximate variance 1020.8 is close to the descriptive statistic value of 894.3.



(a) Homogenous defect distribution: the number of defects per sample is deterministically proportional to the specimen volume.

(b) Poisson defect distribution: the number of defects per sample is a realization of a Poisson distribution.

(c) Clustered defect distribution: each sample has to be numerically simulated to account for the geometrical effect on the number of defects per sample.

Fig. 18. Methods of simulation of defect population.

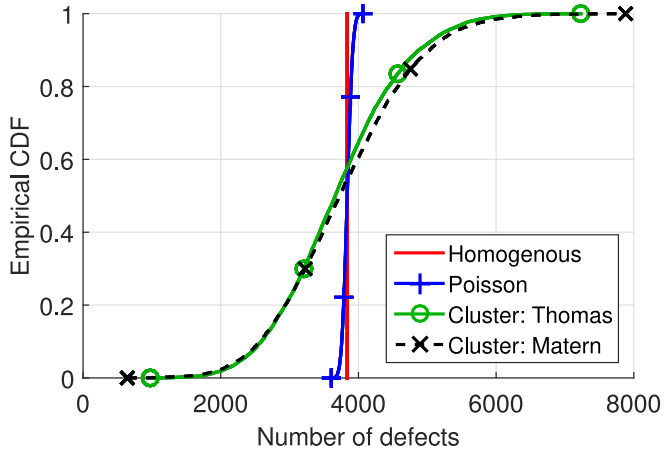


Fig. 19. Cylindrical sample: empirical cumulative distribution function (CDF) of number of defects per sample for different simulation methods. The estimates are based on 10 000 simulated samples.

Table 9
Cylindrical sample: descriptive statistics of the number of defects per sample.

Method	Mean $\bar{n}$	Standard deviation σ_n
Homogeneous	3837	0
Poisson	3836.5	62.1
Clustered: Matérn	3695	894.3
Clustered: Thomas	3735.7	950.3

Ring sample. Going back to real life observations, Table 2 shows the initiating defects are all located at a distance lower than 263 µm from the specimen surface. To investigate this effect, ring numerical specimen are created (see Fig. 20). In this thought experiment, the defects in the center cylinder (of radius R_{min}) are omitted, and only the defects of the outside rim ($R_{min} < r < R$, with $R = 3.5$ mm the specimen radius) are considered.

For the numerical value of $R - R_{min}$, a value of 300 µm is taken. The difference between the "distance to surface" parameter and $R - R_{min}$ is noteworthy: the distance to the surface is defined as the shortest distance from the defect contour to the specimen surface, while the distance $R - R_{min}$ encompasses the maximal distance from the specimen surface to the numerical defect center.

Table 10
Ring sample: descriptive statistics of the number of defects per sample.

Method	Mean $\bar{n}$	Standard deviation σ_n
Homogeneous	629	0
Poisson	629.3	25.6
Clustered: Matérn	617.6	220.7
Clustered: Thomas	607.1	153.1

The sample volume $|W|$ is therefore:

$$|W| = \pi(R^2 - R_{min}^2).h = 75.8\,\text{mm}^3 \tag{68}$$

h being the height of the sample. The ring sample is approximately six times smaller than the cylindrical sample.

Four different models are used: homogeneous, Poisson, Matérn and Thomas. For the clustered processes, 10 000 samples are simulated to derive the empirical CDF (see Fig. 21), and the statistical values (see Table 10).

Similar conclusions to the ones observed for the cylindrical samples can be drawn: the clustered processes drastically increase the standard deviation while the mean value is only slightly modified.

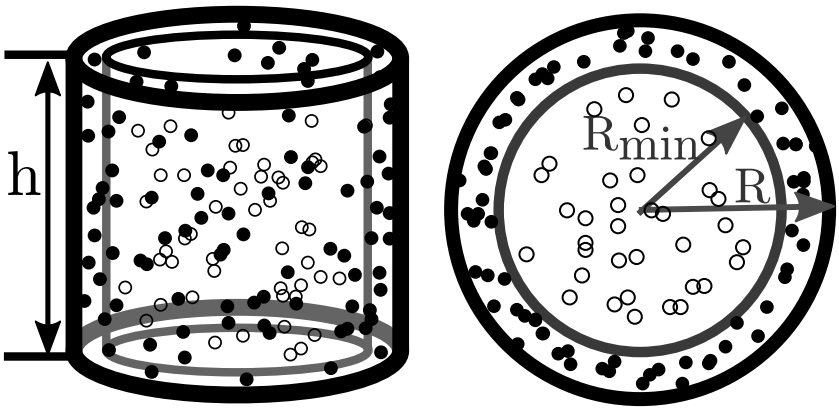
5.3. Comparison with experimental critical defects

Given the size distribution of defects per sample and the independence of the spatial distribution, the numerical method enables to simulate a maximum defect size for each sample. The goal being the comparison with the observed data, a conversion of the Feret diameter from 3D to 2D is done. Indeed, the simulated data is generated according to the tomographic data, in 3 dimensions. However, on the observed fractured specimens, only a 2D Feret diameter can be measured, and 3D defects are approximated by 2D defects:

Fig. 22 shows an example of this conversion is done:

1. A 3D Feret diameter d_{3D} is numerically generated following the size distribution identified in Section 3 and corresponding to what could be measured by tomography.
2. Given the randomness of the 3D Feret diameter angle α, for each 3D Feret diameter, such a random angle is generated with respect to the plane (x, y), where z is the loading direction.
3. For each 3D Feret diameter d_{3D}, a corresponding 2D Feret diameter d_{2D} is generated:

$$d_{2D} = d_{3D}.\cos\alpha \tag{69}$$



(a) 3D representation of the specimen

(b) Top view of the specimen

Fig. 20. Ring sample: numerical defects are created in the cylindrical sample, but only the ones on the outside ring (•), and not in the centeral cylinder (o), are used.

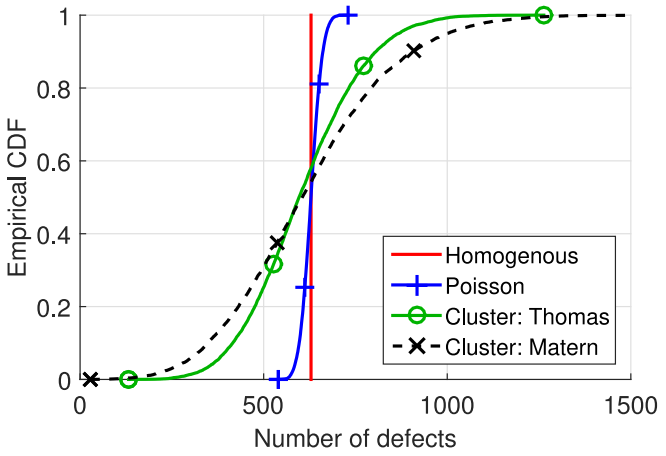


Fig. 21. Ring sample: empirical cumulative distribution function (CDF) of number of defects per sample for different simulation methods. The estimates are based on 10 000 simulated samples.

This method is applied to the cylindrical sample and the ring sample.

Cylindrical sample. Using the cylindrical sample, the sample maxima distribution is generated for the different point processes (homogeneous, Poisson, Thomas and Matérn). The results are shown in Fig. 23. For the 4 different point processes used, the sample maxima distribution is almost identical. The higher standard deviation of the number of defects per numerical sample of the clustered processes has very little effect on the sample maxima distribution.

Ring sample. The sample maxima method is now applied to the ring sample for the different point processes (homogeneous, Poisson, Thomas and Matérn). As for the cylindrical sample, the different point processes all provide the same results.

5.4. Results

Using the results of the previous Section, the results on the cylindrical and ring samples are compared to the experimental results of the fatigue test results obtained in Section 2.3 (i.e. the critical defects measured on the fracture surfaces). The comparison is shown in Fig. 25. The differences between the simulation methods being very small, the different simulation methods are not differentiated (the reader can refer to the previous Figs. 23 and 24).

The cylindrical sample clearly overestimates the maximum defect size, meaning the simulated volume is too large, while the ring sample provides a good fit.

In order to compare the different fittings, the different data sets are fitted with an Generalized Extreme Value distribution (GEV). An example of fitting for the homogeneous process is shown in Fig. 26. The GEV provides an excellent fit for both numerically generated samples (cylinder and ring). Given the small number of experimental critical defects measured (17), it is difficult to discriminate and fit the best distribution. For comparison purposes with the numerical samples, the GEV has been chosen. This allows the comparison of the distribution parameters (location, scale and shape).

The GEV obtained by fitting on the cylinder sample does not capture the experimental results, while the one obtained by the ring sample does.

For the higher values of maximum defect size (over 600 µm), further studies should be conducted in order to ascertain the tails of these extreme value distributions.

The same distribution fitting method is applied to the other point processes. Table 11 shows the results of this distribution fitting. For each identified distribution, a Kolmogorov–Smirnov test hypothesis (KS-Test) is used with a 5% significance level.[2]

For the ring sample, and all the point process methods, the location parameter μ is well captured, but the scale parameter σ is underestimated and the shape parameter ξ overestimated. However, given the small number of data, these parameters are estimated with a high uncertainty. For all the methods, the KS-Test fails to reject the null hypothesis, confirming the goodness of fit of the distributions.

For the cylindrical sample, all the point processes methods are rejected by the KS-Test. This demonstrates the cylindrical sample is not representative of the defect size population with respect to the crack initiation defect. While for some mechanisms, the larger the volume studied, the better the representativeness, the description of the largest defect in a sample requires an accurate estimate of the volume likely to contain a critical defect. Here, all the critical defects are close to the specimen surface, and the representative volume is the ring sample. This study also shows that the clustered aspect of defects plays only a small role in the size of the largest defect. This is due to the important size of the samples: even

[2] The test result is 1 if the test rejects the null hypothesis at the 5% significance level. The null hypothesis being that the Experimental Data comes from the distribution identified by simulation.

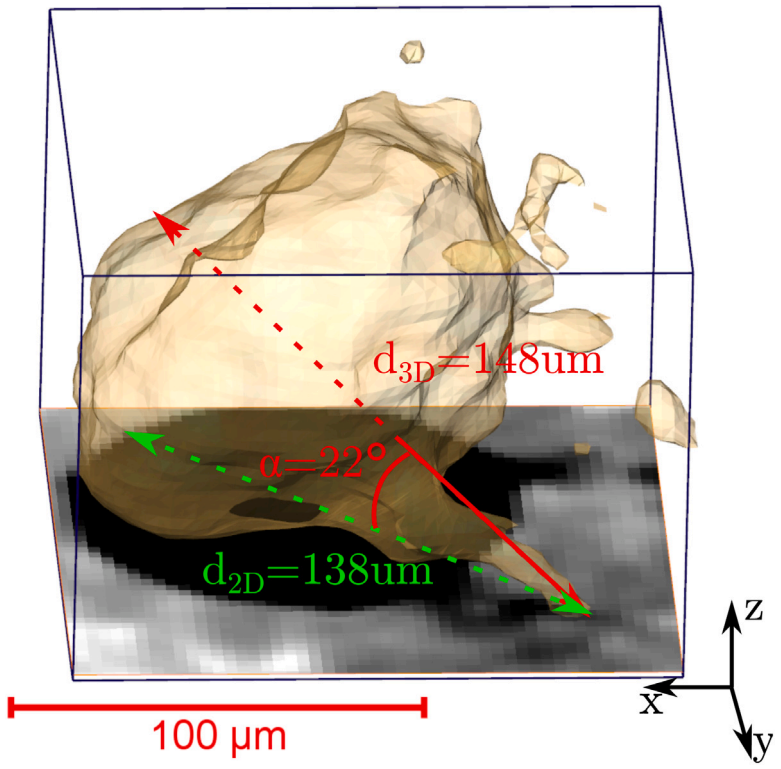


Fig. 22. Feret diameter conversion from 3D to 2D on real defect.

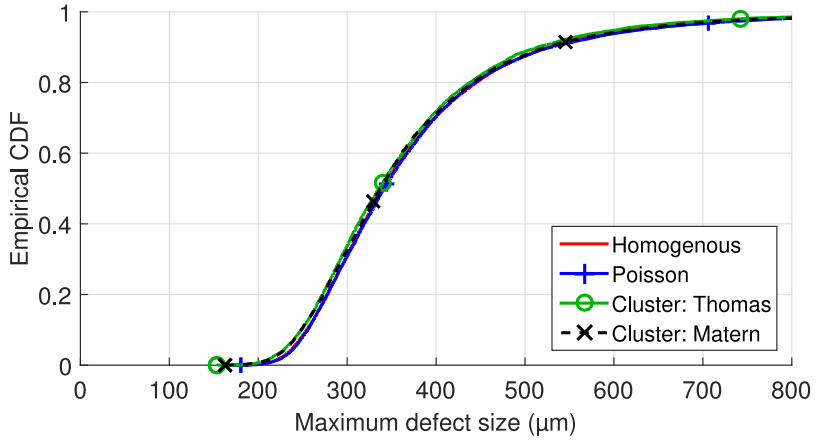


Fig. 23. Cylindrical sample: empirical cumulative distribution function (CDF) of the largest defect per sample for different simulation methods. The estimates are based on 10 000 simulated samples.

for the smaller ring sample, it contains in average around 600 defects. This high number of defects combined with the independence of the defect location within the sample weakens the impact of the clustering of defects compared to the random Poisson point process.

6. Application to notch specimens

A fatigue campaign on notched specimens was carried out to study the impact of defect distribution on in the highly loaded volume. Indeed, for these specimens, only a small portion of the total volume is highly mechanically loaded, stressing out the importance of proper spatial defect distribution description on the fatigue life. Finite element simulations are used to asses the stress and strain field around the notch and a previously developed fatigue model is used to estimate the LCF life estimation and the impact of defect clustering.

6.1. Test conditions and setup

Notched specimen were designed so that the stress gradient in the specimen match that of the studied industrial components. The component and numerical results on the component are not shown for confidentiality reasons. Circular specimens with U-shape groove inducing a $K_t = 1.78$ were designed. The dimensions are indicated in Fig. 27.

For the fatigue tests, an extensometer with a gauge length $l_0 = 12\,\text{mm}$ was used to perform strain controlled LCF tests. The notch is

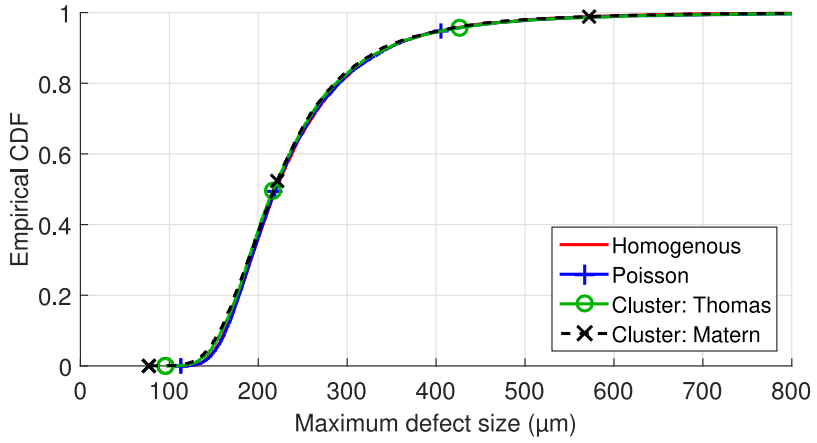


Fig. 24. Ring sample: empirical cumulative distribution function (CDF) of the largest defect per sample for different simulation methods. The estimates are based on 10 000 simulated samples.

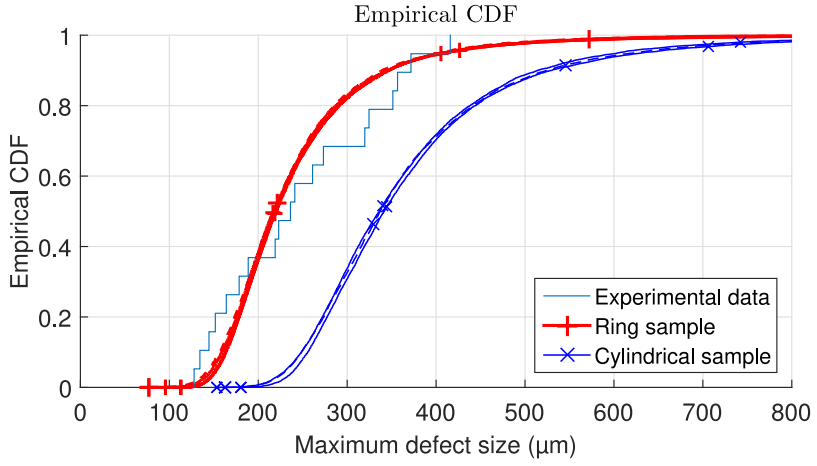


Fig. 25. Comparison of the ring and cylindrical samples sample maxima distribution to the experimentally observed defects.

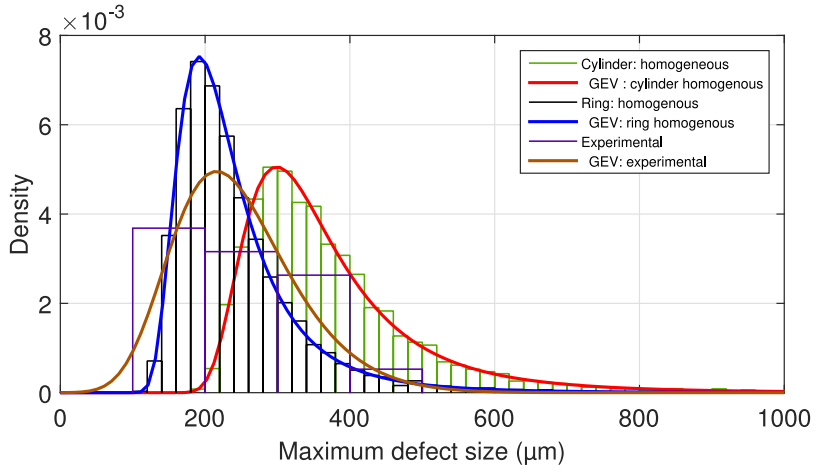


Fig. 26. GEV distribution fitting.

Table 11
Generalized extreme value distribution fitting of the numerical samples (cylindrical and ring specimen) and the experimentally observed defects. In parenthesis are the confidence intervals.

Data	Method	μ_{gev}	σ_{gev}	ξ_{gev}	KS-Test
Experimental		209 (±22.0)	74.5 (±17.5)	−0.0937 (±0.319)	
Ring sample					
	Homogeneous	201.2	49.9	0.216	0
	Poisson	200.8	49.4	0.234	0
	Matern	198.3	51.8	0.203	0
	Thomas	198.0	53.8	0.156	0
Cylinder sample					
	Homogeneous	312.9	74.3	0.213	1
	Poisson	312.3	74.6	0.224	1
	Matern	306.5	76.2	0.188	1
	Thomas	308.2	77.6	0.187	1

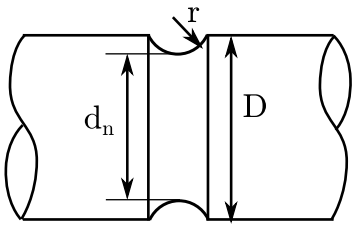


Fig. 27. Notch shape: $D = 7$ mm, $d_n = 5.6$ mm, $r = 1.2$ mm.

Table 12
Experimental matrix for notched specimens tests: number of tests per test condition. The tests are run at $L_0 \pm d$ (i.e. a d amplitude, or equivalently a $2d$ range), l_0 being the initial extensometer gauge length (12 mm).

T (°C)	$\Delta\epsilon_{macro}/2$	20	150	200	250
$d = 0.016$ mm	±0.13%	1	1	1	1
$d = 0.020$ mm	±0.17%	2	0	3	1
$d = 0.025$ mm	±0.21%	1	0	0	0

placed in the center of the extensometer gauge length. The loading is applied by an Instron 8500 servo-hydraulic fatigue testing machine equipped with a furnace (see [19]). Unlike for smooth specimens, the local strain cannot be directly imposed, and the controlled parameter is solely the total extensometer gauge length l. The displacement d is further used in the graphics (the displacement d is the relative displacement $d = l - L_0$). In order to define the test frequency, a macroscopic strain ϵ_{macro} is defined:

$$\epsilon_{macro} = \frac{l - L_0}{L_0} \tag{70}$$

and the frequency is chosen so that $\dot{\epsilon}_{macro} = 10^{-3}$ (which is the same strain rate used for the smooth specimens). Fatigue tests were performed for 3 imposed displacements at 20 °C and 2 for other investigated temperatures (see Table 12). The imposed signal waveform was triangular and symmetrical ($R = -1$). Tests were conducted under 3 different controlled temperatures: 20 (ambient), 150, 200 and 250 °C. For all tests, the specimen lifetime N_f correspond to a 10% load drop.

6.2. Fatigue test results

Fatigue test results on notched specimens are presented in Table 13 and in Fig. 28. Seemingly to the fatigue tests results on smooth specimens (see [19]), increasing temperature decreases the fatigue lifetime for a given imposed displacement. For all the specimens, the fatigue crack initiated at the notch root. As for crack mechanisms observed on smooth specimens in (see [19]), the main crack initiates on a shrinkage defect close to the specimen free surface.

Table 13
Fatigue test results on notched specimens ($K_t = 1.78$).

Specimen	d (mm)	Temperature (°C)	N_f (cycles)
41T5	0.016	20	11 790
16T5	0.020	20	2376
75T5	0.020	20	1698
162T5	0.025	20	753
83T5	0.016	150	9924
17T5	0.016	200	6159
64T5	0.020	200	1880
10T5	0.020	200	1650
53T5	0.020	200	1456
38T5	0.016	250	3302
174T5	0.020	250	1358

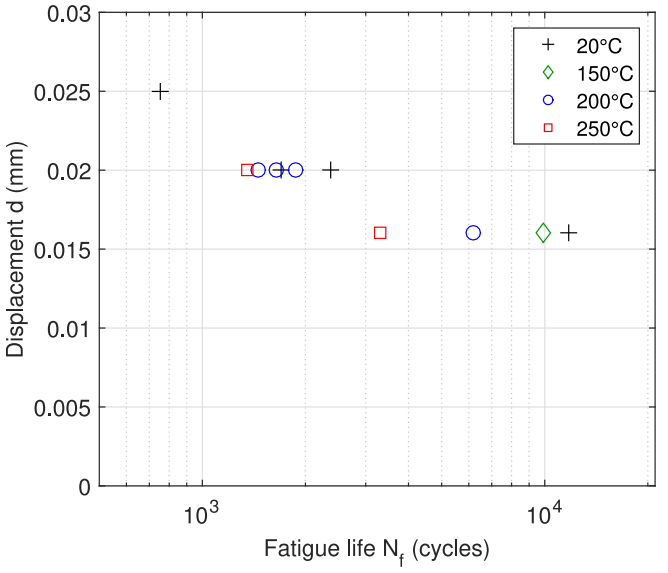


Fig. 28. Fatigue life for notched specimens ($K_t = 1.78$).

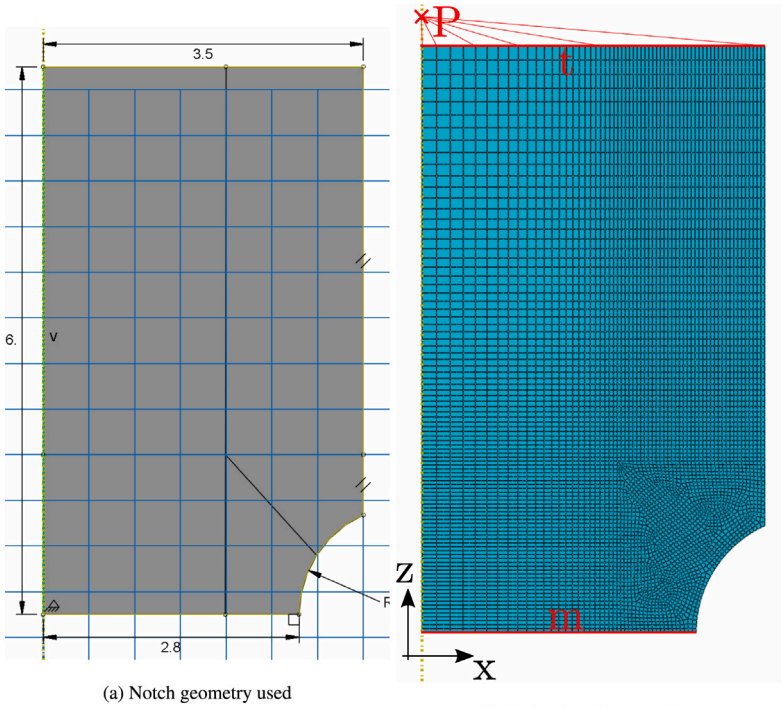
6.3. Numerical simulation of notch specimens

Given the test set-up, only macroscopic values can be experimentally measured (extensometer displacement and applied load). Numerical simulation were used to assess the evolution of local mechanical quantities.

6.3.1. Numerical and mesh generation

In order to simulate notched specimens, Abaqus© finite element (FEM) code was used. Given the problem symmetry, a quarter of the specimen was simulated using axi-symmetric 2D CAX8R elements (see Fig. 29).

The boundary conditions used are:



(a) Notch geometry used

(b) Optimal mesh generation.

Fig. 29. Mesh of the notched specimen.

- Middle nodes (on line m): the vertical displacement (z) is set to 0, as to impose the model symmetry.
- Top nodes (on line t): the vertical displacement d_{sim} (along z) is imposed according to experimental conditions.
- node P: the z degree of freedom of all nodes on line t are linked to the z degree of freedom of node P.

Using this method, twice the z displacement of node P matches the measured displacement of the experimental extensometer, and the reaction force on node P corresponds to the experimental load. These are the values used in the following numerical study.

A short convergence study allowed to optimize the mesh size for proper strain and stress field description on the notch area (the optimized mesh was reduced to 40 µm close to the notch root). Given the cyclic nature of the simulation model used, numerical cyclic stabilization must also be ensured. A combined kinematic and isotropic hardening model was identified on cyclic hardening tests performed on smooth specimens at the different investigated temperatures. For confidentiality reasons the model description and model parameters cannot be provided but the identified model could describe the experimental stabilized behaviour for all investigated temperatures (example is given Fig. 30 for 200 °C).

Numerically, the stabilized behaviour at the notch root was reached after 15 cycles. Fig. 31 compares the simulated and experimental value of the macroscopic load for all the temperatures and imposed displacement. The simulations and experimental results are all shown for cycle $N = 15$ and show good agreement between simulation and experimental results.

6.4. SEM and tomography analyses of fracture surfaces

Notched specimen fatigue mechanisms are described in this Section. For all the specimens, the fatigue crack initiated at the notch root. Fig. 32 shows an example of a general fracture surface (at $T = 250$ °C and defect distance to surface d = 0.016 mm). The main crack initiates on a shrinkage defect close to the specimen free surface. The steady crack propagation region, however, transforms in a complex shape induced by the notch stress concentration and do not show the standard semi elliptical shape observed on smooth samples. Fig. 32 also shows a secondary crack with characteristic features, which has less propagated and exhibits a crack shape similar to smooth specimen cracks.

To investigate the crack shape before failure, specimen 53T5 was analysed using laboratory X-ray tomography. The fatigue test was stopped after crack initiation was detected and the cracked sample scanned. A 5.1 × 5.1 × 5.1 µm voxel was used. The general 3D shape of the sample is shown in Fig. 33. The main crack appears at or very close to the notch root along the direction: the crack is contained in a ±400 µm interval from the notch middle showing successive and numerous crack bifurcations. The segmentation of the crack shown in Fig. 34 confirm the complex 3d shape of the crack and the limited crack propagation towards the center of the notched specimen (400 µm) due to the strong gradient effect. Finally Fig. 35 displays what is identified as the main crack initiation. The initiating defect is highly spherical (see Fig. 35(c)), suggesting a gas porosity defect of 80 µm. The initial shape of the crack corresponds to a classical semi circular crack (see Fig. 35(b)) centred on the identified defect. The steady crack appears roughly 1200 µm wide before it changes aspect (see Fig. 35(a)).

6.5. Fatigue life assessment of notched specimen with numerical defect generation

Numerical defects are generated according to specific spatial distribution (homogeneous or clustered process), as well as each defect size. For each defect, a life prediction can be made using the crack propagation model described hereafter. The number of cycles to failure of a given configuration is then defined as the lowest simulated value of all defects. This critical value depends on the defect size, considered as the initial crack length, and the defect location, which modifies the driving mechanical forces of the model (plastic dissipated and elastic energies). The following describes these different steps.

6.5.1. Defect population generation

A defect population is generated according to the method described in Section 4: defect location and the 2D equivalent Feret diameter are

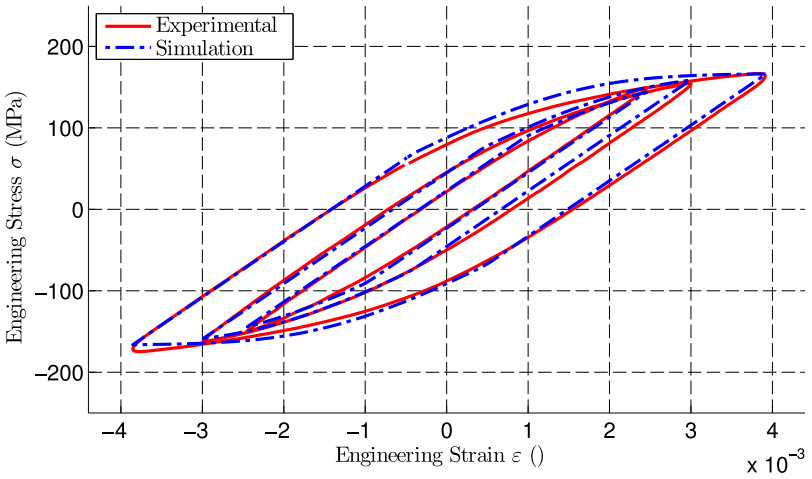


Fig. 30. Numerical simulation of 3 strain controlled tests ($\Delta\varepsilon/2 = \pm$ 0.25; 0.3; 0.4%) at 200 °C, 10th cycle.

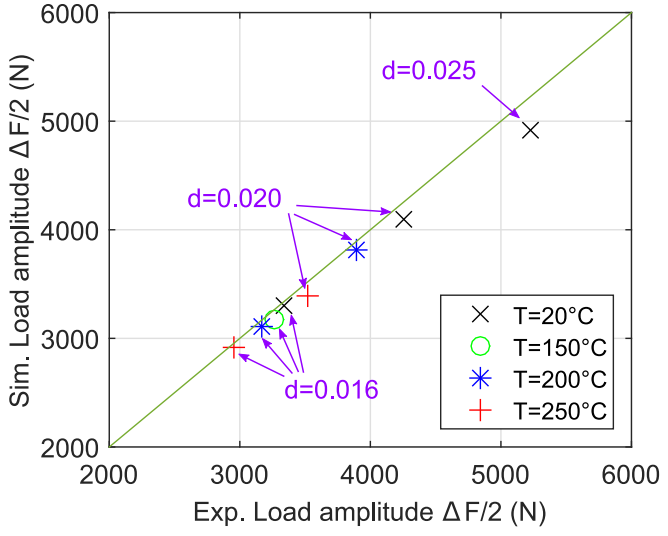


Fig. 31. Comparison of experimental and simulated load amplitude for different temperatures and imposed displacements d (indicated in mm) for notched specimen ($K_t = 1.78$).

generated in a large window defined on $(x, y, z) \in [-5; 5]^3$ englobing the notch area. The defect location was generated using a Poisson distribution, or a clustered one (as the one illustrated in Fig. 36(a)). For this study, the clustered distribution is limited to the Thomas Point process. Each defect i is located by its cylindrical coordinates (ρ_i, θ_i, z_i). For subsequent illustrations, all the defects are projected on the $\theta = 0$ cross section (see Fig. 36(b)) and defects outside the virtual notched specimen are omitted.

6.6. FEM mapping and calculation method

Using the results of the FEM calculation without considering any defect, each defect can be associated with a local value of elastic energy (see Fig. 37(a)) and plastic strain energy (see Fig. 37(b)). To obtain the mapping, FEM results are firstly extracted at calculation nodes and results linearly interpolated at defect location. Although the stress values are only approximate values at nodes, the fine mesh used in our FEM model allows neglecting the approximation.

After obtaining the energy values at each defect center, the crack propagation model (defined in [19]) is used to simulate the predicted life $N^i_{f,sim}$ of each defect:

$$N^i_{f,sim}(d_i) = \int_{d_i}^{a_f} \frac{da}{\frac{1}{\lambda}\frac{da}{dN}(\Delta W_e(\rho_i, z_i), \Delta W_p(\rho_i, z_i), a)} \tag{71}$$

where d_i is the ith defect 2D feret diameter of cylindrical coordinate (ρ_i, z_i). The angle coordinate θ is omitted given the revolution symmetry. The variable $\Delta W_e(\rho, z)$ (resp. $\Delta W_p(\rho, z)$) is the elastic energy field (resp. the plastic dissipation energy field). The final crack length considered a_f is equal to 3 mm. This method aims at quantifying the influence of clustering on fatigue life considering the small volume subjected to high loading induced by the notch, thereby creating a smaller fatigue active volume.

For a rigorous calculation, the β_e and β_p parameters should account for the notch geometry and the inhomogeneous mechanical field compared to the uniform stress/strain field for smooth specimens. However, in a first instance both parameters will be taken from results obtained in [19] on smooth specimens, the results presented hereafter mostly aiming at emphasizing the effect of defect clustering on fatigue life.

Fig. 38 shows the result of the fatigue life simulation for one specimen. Given the theoretical stress concentration at the notch root, all the shorter lives are found close to the root, even though some larger defects slightly correspond to fatigue life in the same range.

6.7. Simulation of defect mapping and life estimation: results

Using this simulation method, it is possible to generate a high number of specimens, all with a specific defect population, and calculate for each specimen a number of cycles to failure. Combined with kernel density estimation, the model provides an estimate probability density function of the fatigue life. Confronting the simulation results with experimental ones (see Fig. 39(a)) shows both homogeneous and clustered process underestimate the fatigue life of notch specimens. Compared with smooth specimens (see [19]), which presented much better correlation results, the discrepancy could be caused by the hypothesis taken. Indeed, the crack propagation model, used here in a high stress gradient environment, does not account for the heterogeneous stress field. The SEM and tomography observation suggest some complex sequences for crack initiation and propagation in the millimetre range with a crack arrest at about 400 μm from the notch root. A more local detection of the crack initiation may lead to better predictions. Similarly, the interaction of the crack propagation with the neighbouring defects in an heterogeneous stress field is quite different

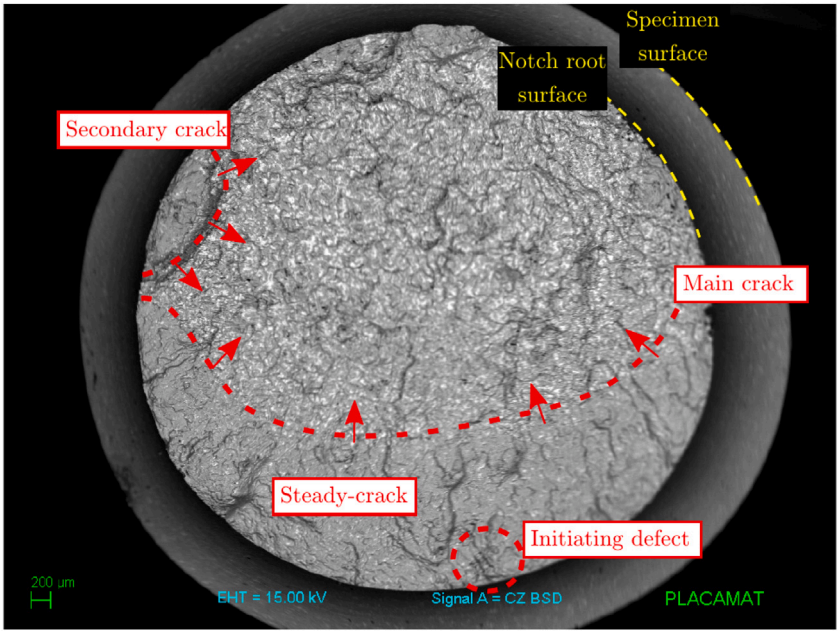


Fig. 32. General view of the fractured surface of a notched fatigue specimen (specimen 38T5, d = 0.016 mm, T = 250, Nf = 3302 cycles.

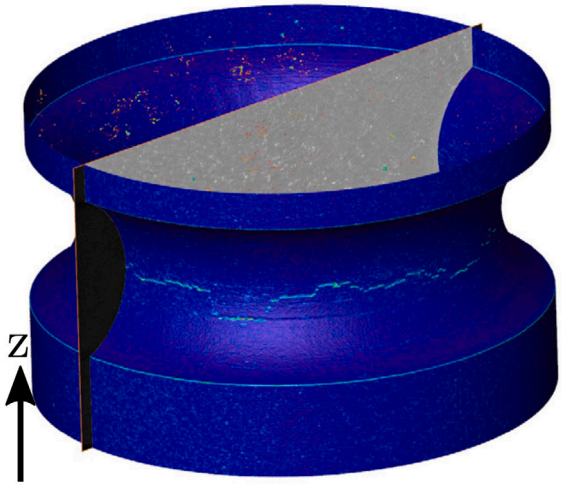


Fig. 33. 3D isosurface view of fatigue notched specimen analysed by tomography. The colour indicates surface curvature, allowing to reveal the fatigue crack. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

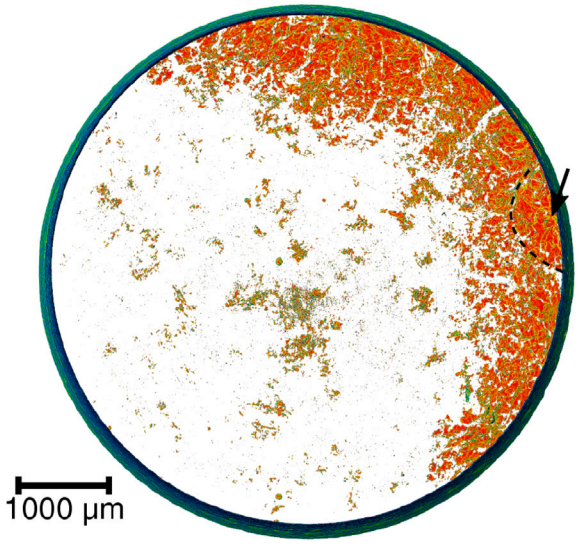


Fig. 34. 3D crack reconstruction by tomography before total failure.

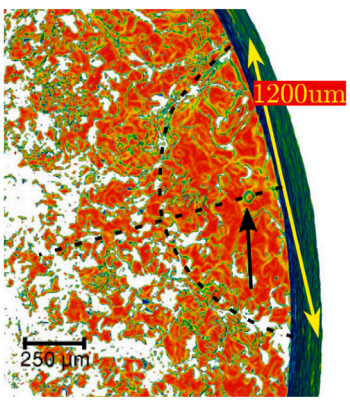
to the one acting in an homogeneous (uniaxial) stress/strain field. This could cause higher experimental fatigue life and is not straightforward to capture explicitly in a model. Finally, the tomography analyses suggest some tortuous paths at the notch root that may be a sign of multiple crack initiation and possible crack shielding effect. This should be further investigated from both the experimental and numerical side.

All things considered, the model proposed provides interesting results for the study of clustering defects on highly loaded volumes. This can be seen on Fig. 39(b): the mean number of cycles to failure is higher when using clustered point process generated defects. While counter-intuitive, these conclusions have a simple explanation: in some cases, the clustered point process generates very few defects in the highly stressed volume. An example of such a numerical specimen is given Fig. 39(c): the parents points of the clustered point process are all far from the notch root, causing a local low density of defects (52 in the restricted area r_c). Due to the low number of defects, the critical one is comparatively smaller (23 µm, compared to the mean values over all the specimens of 70 µm) and the estimated fatigue life higher (811 cycles, as opposed to the mean value 423).

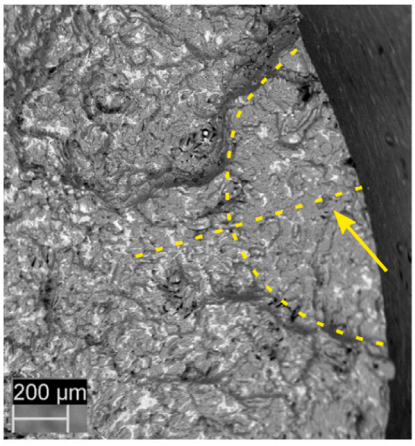
7. Conclusions and prospects

The thorough study of the statistics of defects can help better understand the fatigue mechanisms and the representativeness, from a statistical point of view, of a sample. In a more detailed way, these other conclusions can be drawn:

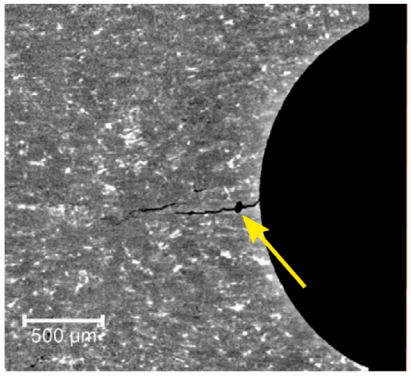
- Properly thresholding a tomography sample can be a difficult task. Working directly on the global histogram can help having



(a) Tomography reconstruction crack initiation

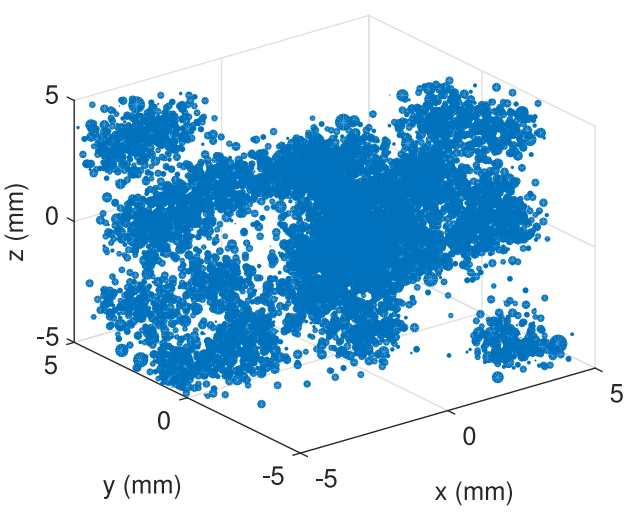


(b) corresponding SEM observation

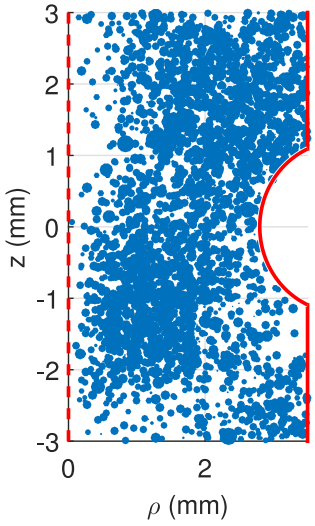


(c) Gas spherical defect : ferret diameter of 80μm, plane of observation perpendicular to image a) and b).

Fig. 35. Analysis of first stages of crack initiation and propagation on microstructural defect identified by the arrow (specimen 53T5, d = 0.020 mm, T = 200 , Nf = 1456 cycle).



(a) 3D defect generation.



(b) 2D projection of defect position

Fig. 36. Defect population generation in a notched specimen. The diameter of each symbol is proportional to the simulated 2D Feret diameter.

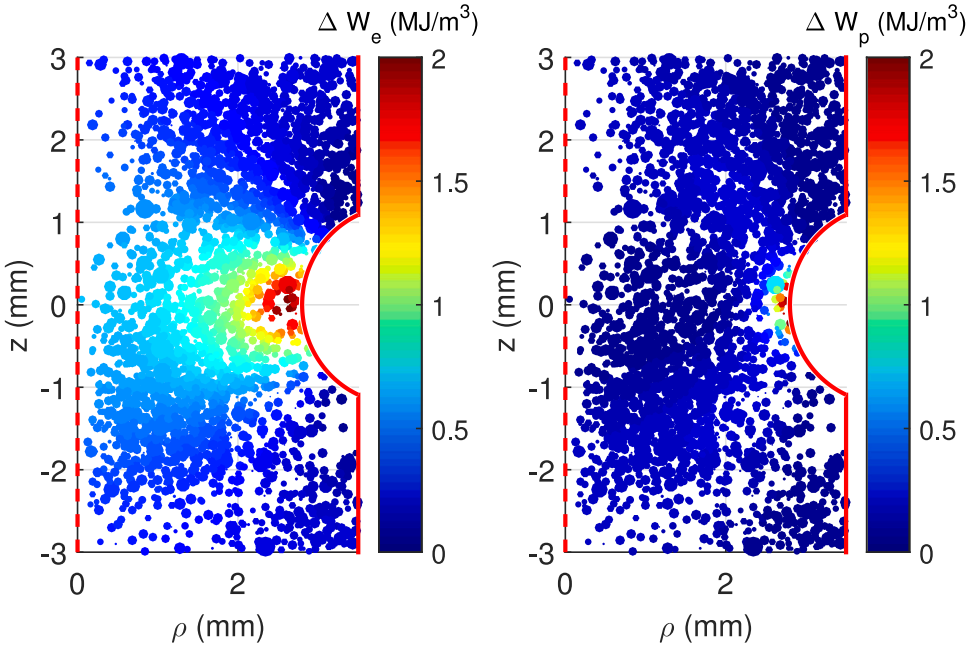
comparable image processing, especially for samples analysed by different means.

- Since the larger defects are here of concern, censoring the smaller defects can help better grasp the defect size distribution in a selected area, here the right hand tail distribution. Combined with an appropriate statistical distribution (the generalized pareto distribution), the statistical inference shows good agreement.
- Studying the nature of the spatial distribution of defects by use of the theory of point process ascertains what can be visually observed : the defects aggregate in clusters. By combining with the study of marked point processes, the correlation between the location and the size of defects can also be better understood: for this specific case, there is no link between location and size (or very little).
- Using the proper size and location distribution, combined with the observation of critical defects for low cycle fatigue, the statistical representative volume is reduced to only the subsurface volume.
- Despite the higher scatter in number of defects, the extreme defects (i.e. the largest defect per sample) is only slightly affected by this clustering.
- However, when the size of clusters is comparable to the sample size, the scatter in the number of defects per sample drastically increases. For small diameter specimens with internal defects, the consequences may be important.
- Similarly, when the characteristic length of the cluster point process is similar to that of the highly stressed volume of a notched specimen, critical defects size can be affected.

In future with the progress of X-ray tomography resolution or 3D atom probe, the proposed methodology could be used to identify the statistical distribution of non-metallic inclusion since their drastic effect on LCF resistance is well-known [10]. This could be useful for high fidelity microstructure modelling and simulation.

Data availability

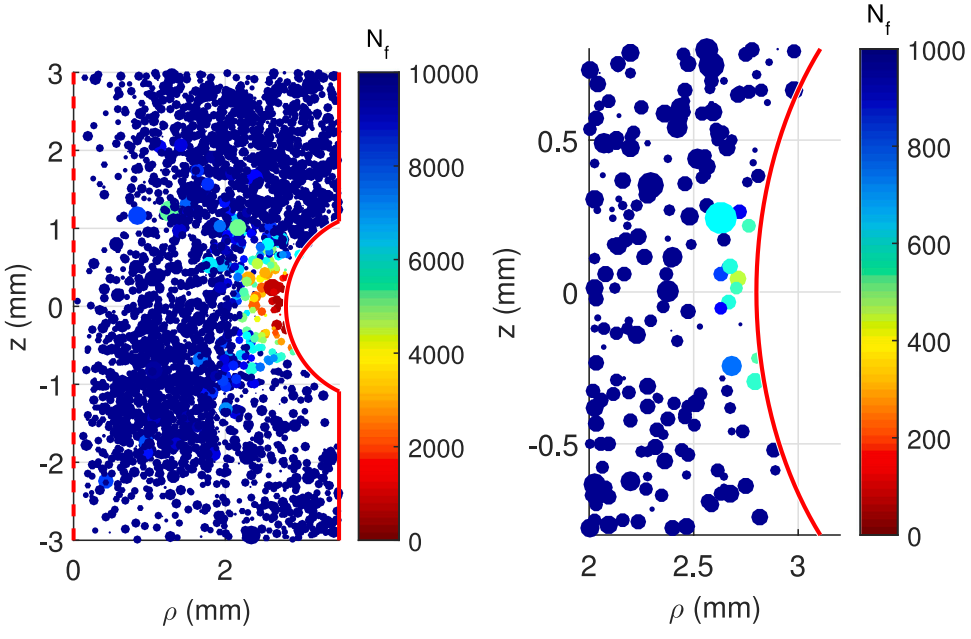
The data that has been used is confidential.



(a) Elastic energy ΔW_e

(b) Plastic dissipated energy ΔW_p

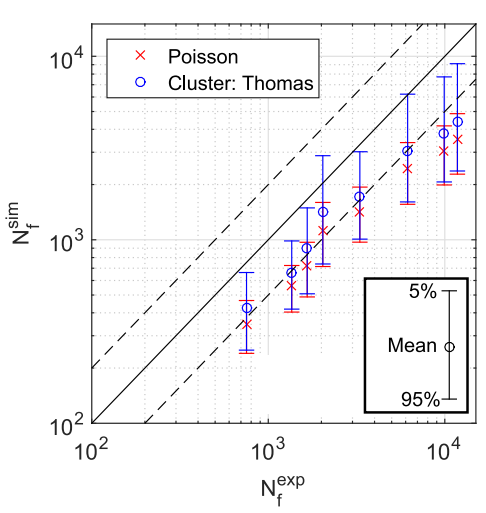
Fig. 37. Defect population generation in a notched specimen and FEM calculation mapping ($d = 0.025$ mm and $T = 20\,^{\circ}$C).



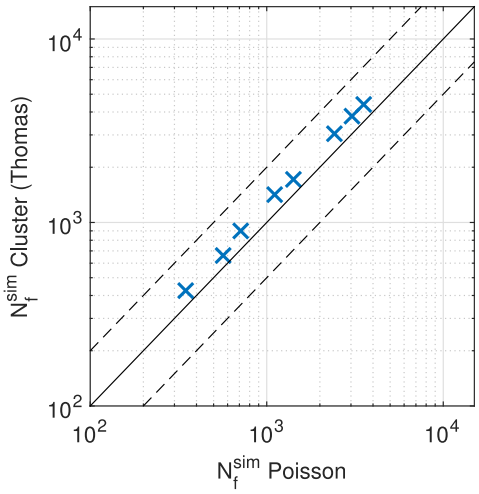
(a) Global view

(b) Zoom of notch region, notice the different in the scale of the color bar for N_f.

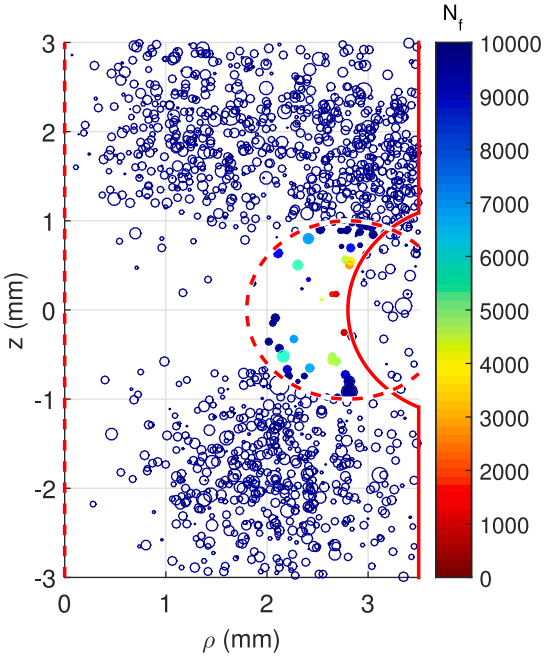
Fig. 38. Fatigue life simulation of number of cycles to failure (N_f) of notched specimen ($d = 0.025$ mm and $T = 20\,^{\circ}$C, $N_f = 753$). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)



(a) Effect of clustering on the dispersion of fatigue life prediction, 5% and 95% percentiles are shown.



(b) Simulation of the mean number of cycles to failure: Poisson or Clustered (modified Thomas) point processes generated defects.



(c) Fatigue life simulation of the number of cycles to failure (N_f) of notched specimen with defects (K_t = 1.78, d = 0.025mm and T = 20 °C, N_f = 753). Specific case of low number of defects close to the notch root. Only the fatigue lives of defects close to the notch root are calculated (full markers)

Fig. 39. Influence of point process on fatigue life prediction.

References

[1] Thomas JJ, Verger L L, Bignonnet A, Charkaluk E. Thermomechanical design in the automotive industry. Fatigue Fract Eng Mater Struct 2004;27(10):887–95. http://dx.doi.org/10.1111/j.1460-2695.2004.00746.x.

[2] Mondolfo LF. Aluminium alloys, structure and propreties. Butterworths; 1976.

[3] Buffière J-Y, Savelli S, Jouneau PH, Maire E, Fougères R. Experimental study of porosity and its relation to fatigue mechanisms of model Al-Si7-Mg0.3 cast Al alloys. Mater Sci Eng 2002;A316:115–26.

[4] McDowell D, Dunne F. Microstructure-sensitive computational modeling of fatigue crack formation. Int J Fatigue 2010;32(9):1521–42. http://dx.doi.org/10.1016/j.ijfatigue.2010.01.003, Emerging Frontiers in Fatigue. URL https://www.sciencedirect.com/science/article/pii/S0142112310000162.

[5] Przybyla CP, Musinski WD, Castelluccio GM, McDowell DL. Microstructure-sensitive HCF and VHCF simulations. Int J Fatigue 2013;57:9–27. http://dx.doi.org/10.1016/j.ijfatigue.2012.09.014, Fatigue and Microstructure: A special issue on recent advances. URL https://www.sciencedirect.com/science/article/pii/S0142112312002782.

[6] Hor A, Saintier N, Robert C, Palin-Luc T, Morel F. Statistical assessment of multiaxial HCF criteria at the grain scale. Int J Fatigue 2014;67:151–8. http://dx.doi.org/10.1016/j.ijfatigue.2014.01.024, Multiaxial Fatigue 2013. URL https://www.sciencedirect.com/science/article/pii/S0142112314000358.

[7] Ben Ahmed A, Houria MI, Fathallah R, Sidhom H. The effect of interacting defects on the HCF behavior of Al-Si-Mg aluminum alloys. J Alloys Compd 2019;779:618–29. http://dx.doi.org/10.1016/j.jallcom.2018.11.282, URL https://www.sciencedirect.com/science/article/pii/S0925838818344141.

[8] El Khoukhi D, Morel F, Saintier N, Bellett D, Osmond P, Le V-D. 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. Int J Fatigue 2021;147:106177. http://dx.doi.org/10.1016/j.ijfatigue.2021.106177, URL https://www.sciencedirect.com/science/article/pii/S0142112321000372.

[9] El Khoukhi D, Saintier N, Morel F, Bellett D, Osmond P, Le V-D. Spatial point pattern methodology for the study of pores 3D patterning in two casting aluminium alloys. Mater Charact 2021;177:111165. http://dx.doi.org/10.1016/j.matchar.2021.111165, URL https://www.sciencedirect.com/science/article/pii/S1044580321002953.

[10] Chen B, Hashimoto T, Vergeer F, Burgess A, Thompson G, Robinson I. Three-dimensional analysis of the spatial distribution of iron oxide particles in a decorative coating by electron microscopic imaging. Prog Org Coat 2014;77(6):1069–72. http://dx.doi.org/10.1016/j.porgcoat.2014.03.005, URL https://www.sciencedirect.com/science/article/pii/S0300944014000939.

[11] Baddeley AJ, van Lieshout MNM. Area-interaction point processes. Ann Inst Stat Math 1995;47:601–19.

[12] Baddeley A, Rubak E, Turner R. Spatial point patterns: Methodology and applications with R. CRC Press; 2015.

[13] Bailey TC, Gatrell AC. Interactive spatial data analysis. Longman Scientific Technica; 1995.

[14] Babu GJ, Feigelson ED. Spatial point processes in astronomy. J Statist Plann Inference 1996;50(3):311–26. http://dx.doi.org/10.1016/0378-3758(95)00060-7, Spatial Statistics, Part I. URL https://www.sciencedirect.com/science/article/pii/0378375895000607.

[15] Shabestari SG, Moemeni H. Effect of copper and solidification conditions on the microstructure and mechanical properties of Al-Si-Mg alloys. J Mater Process Technol 2004;153–154:193–8.

[16] Salvo L, Cloetens P, Maire E, Zabler S, Blandin JJ, Buffiére JY, Ludwig W, Boller E, Bellet D, Josserond C. X-ray micro-tomography an attractive characterisation technique in materials science. Nucl Instrum Methods Phys Res B 2003;200:273–86. http://dx.doi.org/10.1016/s0168-583x(02)01689-0.

[17] Buffiere JY, Maire E, Adrien J, Masse JP, Boller E. In situ experiments with X ray tomography: An attractive tool for experimental mechanics. In: Proceedings of the society for experimental mechanics, Inc., Vol. 67. 2010, p. 289–305. http://dx.doi.org/10.1007/s11340-010-9333-7.

[18] Stock S. Microcomputed tomography: Methodology and applications. 2011, p. 364. http://dx.doi.org/10.1201/9781420058772.fmatt.

[19] Wilson P, Saintier N, Palin-Luc T, Bergamo S. Isothermal fatigue damage mechanisms at ambient and elevated temperature of a cast Al-Si-Cu aluminium alloy. Int J Fatigue 2019;121:112–23. http://dx.doi.org/10.1016/J.IJFATIGUE.2018.11.016.

[20] Pickands J. Statistical inference using extreme order statistics. Ann Statist 1975;3(1):119–31. http://dx.doi.org/10.1214/aos/1176343003.

[21] Neyman J, Scott EL. A theory of the spatial distribution of galaxies. Astrophys J 1952;116:144. http://dx.doi.org/10.1086/145599.

[22] Illian J, Penttinen A, Stoyan H, Stoyan D. Statistical analysis and modelling of spatial point patterns. Wiley-Blackwell; 2007, http://dx.doi.org/10.1002/9780470725160.

[23] Waagepetersen R. An estimating function approach to inference for inhomogeneous Neyman-Scott Processes. Biometrics 2006;63:252–8.

[24] Pommerening A, Stoyan D. Edge-corrrection needs in estimating indices of spatial forest structure. Can J For Res 2006;36:1723–39.