



HAL
open science

Generation of synthetic microstructures containing casting defects: a machine learning approach

Arjun Kalkur Matpadi Raghavendra, Laurent Lacourt, Lionel Marcin, Vincent Maurel, Henry Proudhon

► **To cite this version:**

Arjun Kalkur Matpadi Raghavendra, Laurent Lacourt, Lionel Marcin, Vincent Maurel, Henry Proudhon. Generation of synthetic microstructures containing casting defects: a machine learning approach. *Scientific Reports*, 2023, 13 (1), pp.11852. <10.1038/s41598-023-38719-0>. <hal-04307940>

HAL Id: hal-04307940

<https://hal.science/hal-04307940v1>

Submitted on 26 Nov 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons CC BY-NC 4.0 - Attribution - Non-commercial use - International License

Generation of synthetic microstructures containing casting defects: a machine learning approach

Arjun Kalkur Matpadi Raghavendra^{1,2}, Laurent Lacourt¹, Lionel Marcin²,
Vincent Maurel¹, and Henry Proudhon^{1,*}

¹Centre des matériaux, Mines Paris, PSL University

²Safran Aircraft Engines, Etablissement de Villaroche

¹ *Corresponding author: Henry Proudhon, henry.proudhon@minesparis.psl.eu

Abstract

This paper presents a new strategy to generate synthetic samples containing casting defects. Four samples of Inconel 100 containing casting defects such as shrinkages and pores have been characterized using X-ray tomography and are used as reference for this application. Shrinkages are known to be tortuous in shape and more detrimental for the mechanical properties of materials, especially metal fatigue, whereas the pores can be of two types: broken shrinkage pores with arbitrary shapes and gaseous pores of spherical shape. For the generation of synthetic samples, an integrated module of Spatial Point Pattern (SPP) analysis and deep learning techniques such as Generative Adversarial Networks (GANs) and Convolutional Neural Networks (CNNs) are used. The SPP analysis describes the spatial distributions of casting defects in material space whereas GANs and CNNs generate a defect of arbitrary morphology very close to real defects. SPP analysis reveals the existence of two different void nucleation mechanisms during metal solidification associated to shrinkages and pores. Our deep learning model successfully generates casting defects with defect size ranging from 100 μm to 1.5 mm and of very realistic shapes. The whole synthetic microstructure generation process respects the global defect statistics of reference samples and the generated samples are validated by statistically comparing with real samples.

Keywords: Synthetic microstructures, Ni-based Superalloys, GAN, Machine learning, X-ray Computed Tomography

1 Introduction

Casted materials often carry defects formed during metal solidification. These defects can have a serious impact on the material properties whose magnitude depends on various microstructural and defect characteristics. Some of the defects that can appear in casted materials are shrinkages, pores, oxide films, etc [1–3]. Shrinkages are large tortuous cavities formed due to contraction of molten metal during solidification whereas pores and micro-voids are smaller in size and are generally formed due to trapped gases. These cavity defects can degrade material performance drastically by promoting initiation and propagation of crack driven by stress concentration [4–7]. The intensity of this degradation depends on various defect characteristics such as its size, position and morphology

33 [8]: the fatigue life is known to vary inversely with respect to defect size, a relationship demon-
34 strated by Kitagawa-Takahashi diagram [9, 10]. It is also known that defect location plays a very
35 prominent role in high cycle fatigue (HCF) [10, 11]. Cracks initiating from defects that are closer
36 to free surface propagate faster when compared to those initiating from internal defects given the
37 difference in their stress intensity factors (SIF) [1]. Furthermore, a tortuous morphology of defects
38 can drastically increase stress concentration facilitating crack initiations. Some of the independent
39 features that can characterize defect morphologies are sphericity, aspect ratio, etc [8]. **Whilst these**
40 **characteristics can induce a large scatter in fatigue life, the problem gets more complicated in ma-**
41 **terials containing high porosity levels which results in formation of defect clusters [12].** In clustered
42 defects, apart from the individual features of defects, they are also influenced by the stress gradients
43 of neighbouring defects. These defects can sometimes be found in aeronautical parts like turbine
44 disks and blades, and has received much less attention in mechanical domain. Analysing all the
45 features that might affect fatigue life requires a large number of samples to be tested which can be
46 extremely costly. Therefore, a plausible approach is to generate synthetic microstructures that are
47 very close to reality which can be simulated numerically to create a large database of mechanical
48 response to the presence of defects, their morphology and spatial distribution.
49 Present work focuses on analysing the effects of defect population in a naturally isotropic material
50 inconel 100 under cyclic loads where the granular characteristics of all tested samples are similar.
51 In such a case, synthetic microstructures can be generated by distributing the defects in a homoge-
52 neous material space according to a pattern similar to real defects. The spatial ordering of defects
53 can be analysed through spatial point pattern theory (SPP) with tools like Ripley’s K-function that
54 measures the second order properties of point distribution in space [13]. A similar approach was
55 applied by El Khoukhi et al where numerical microstructures were generated by placing spherical
56 defects in a homogeneous material space [14].

57
58 **Need for synthetic microstructures :** Samples containing clustered defects (see figure 1)
59 are known to produce very complex mechanical response under fatigue loading. Although, image
60 based finite element (FE) models can simulate this response and aid in locating the crack-initiation
61 site, it is still very difficult to simplify the process and predict its fatigue life with respect to defect
62 characteristics [15]. For an isolated defect, a Kitagawa – Takahashi diagram can be used but the
63 same Linear Elastic Fracture Mechanical (LEFM) approach cannot be applied to clustered defects.
64 **Therefore, a better estimation of the parameters influencing material’s fatigue life apart from just**
65 **the defect’s size is needed. The additional parameters or features could be the volume fraction of**
66 **defects, size of the cluster, sphericity, aspect ratio or other morphological parameters.** For such
67 an analysis, a large number of samples are needed and generating synthetic microstructures that
68 mimics the real specimens is seemingly an inexpensive approach. Furthermore, these generated
69 microstructures can then be converted to image based FE models and simulated numerically to
70 estimate prominent characteristics or to develop a probabilistic model with an approach similar to
71 Monte-Carlo.

72

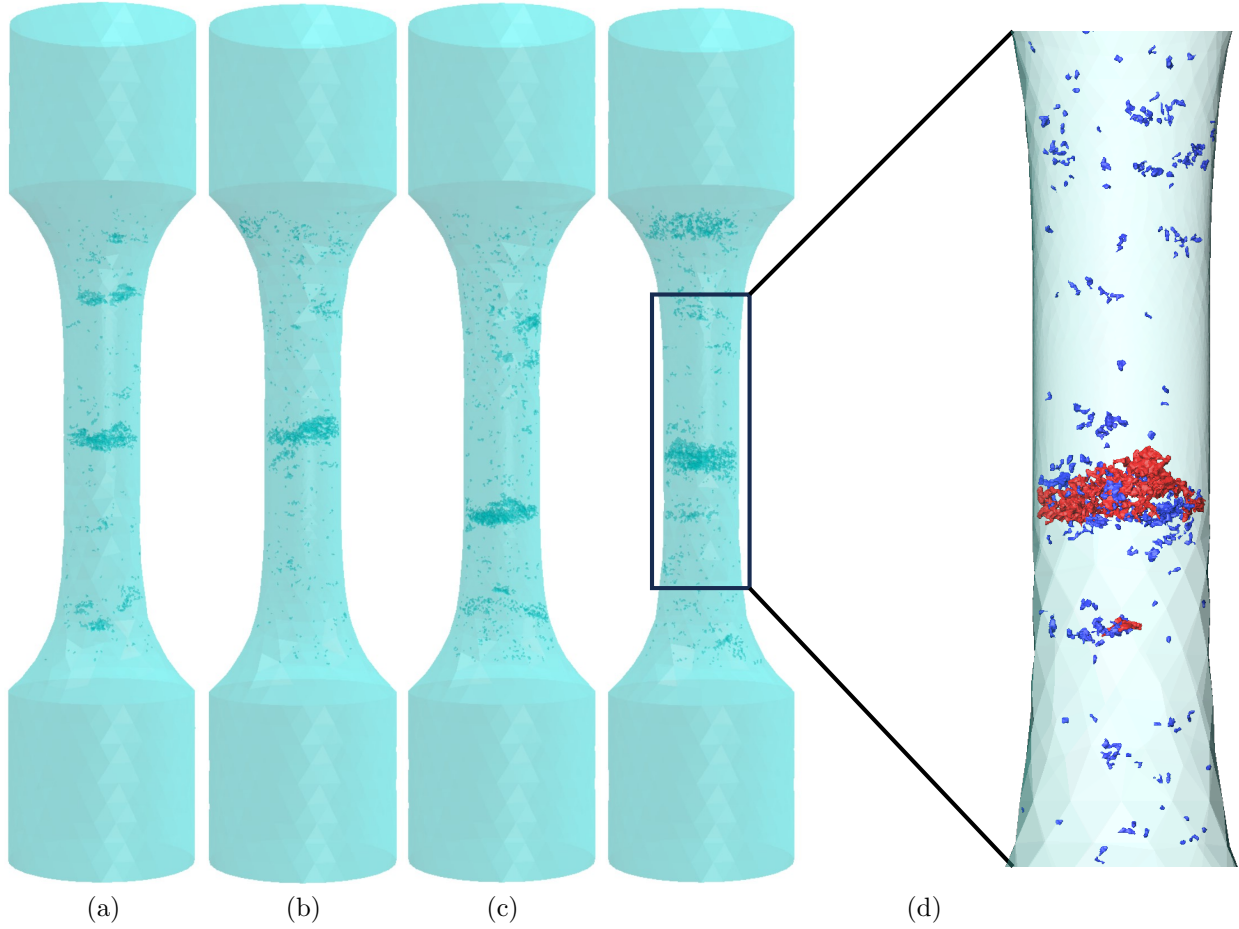


Figure 1: Image-based FE models from X-ray computed tomography (XCT) for reference sample a) A b) B c) C d) D along with a close up view of the clustered defects in sampled D colored as large defects (red) and small defects (blue). Samples are 40 mm long with gauge diameter of 3.7 mm.

73 **Reference samples :** Four *as-cast* Inconel 100 (IN100) cylindrical samples machined from
74 cast ingot bars are used to generate synthetic microstructures. The samples are 40 mm long with
75 a gauge diameter of 3.7 mm and contains clustered defects whose characteristics are assessed via
76 X-ray computed tomographic images (XCT) as seen in our previous work [15]. The XCT volumes
77 can be used to build image-based FE models of the same. Via numerical simulations, the critical
78 defect that could initiate primary crack during fatigue loads can be determined. *The volumes of*
79 *the studied XCT scans were 300 mm³.*

80 As seen in table 1, *defects occupy around 0.3 - 0.52 % of material volume*, among which many
81 of these defects are confined in a small thickness along the axis of the sample forming complicated
82 network of defects as seen in figure 1. The interaction of these clustered defects is too complicated

Table 1: Global defect statistics in IN100 samples

Sample	A	B	C	D
Total no. defects	215	295	346	434
Total defects volume (mm³)	0.935	0.84	1.36	1.58

83 and requires a profound analysis.

84

85 **Method to generate synthetic microstructure :** Synthetic microstructures can be gener-
86 ated by placing the defects in a fixed material space of a particular geometry similar to that of
87 real specimens. Placing the defects in material space is a stochastic process which requires a prior
88 understanding of spatial distribution of patterns, for eg: via SPP, which is significantly used in the
89 field of astronomy, forestry, cartography etc [16–20]. With tools like Ripley’s K-function, second
90 order properties of point pattern can be measured: the points in our context are centroids of defect
91 volumes [21]. Neither many researchers have considered SPP analysis to estimate the 3D spatial
92 characterization of defects nor to generate numerical microstructures using the same [14, 22].

93 In the case of clustered defects, regular shapes cannot be assumed for defects since contributions
94 of various features in degrading material’s performance is merely unknown. Hence, a deep learn-
95 ing strategy called Generative Adversarial Networks (GANs) and Convolutional Neural Networks
96 (CNNs) are integrated together to recreate realistic synthetic defects that can be placed via stochas-
97 tic process defined by SPP in material space. GANs are a very recent development in the field of
98 deep learning that can learn to create data that doesn’t exist in the database [23, 24]. Few re-
99 searchers have attempted to generate microstructures directly using different variants of GANs
100 [25–31]. Jangid et al developed a GAN that could generate random grain shapes [26] which were
101 validated by comparing with real grains. CNNs on the other hand are kernel based neural networks
102 which can learn various receptive kernels to be applied on the image data for classification and re-
103 gression purposes. Here, CNNs are used as post processing step to determine the size of generated
104 synthetic defects.

105 The generated defects are placed in material space respecting the global distributions of defect
106 features and also the spatial pattern. The uniqueness of the generated microstructures is main-
107 tained by applying a Poisson distribution over the mean number of defects while exploring different
108 K-functions similar to that of real specimens.

109

110 2 Methods

111

112 2.1 Spatial point pattern

113 Spatial point pattern (SPP) analysis is a branch of study in stochastics mainly used in the field
114 of astronomy, ecological survey etc. SPP is any point or location in a specified region (defects in
115 material space in our case). These events occur randomly and can be modelled with a specific
116 stochastic process. As discussed by [16], SPP can be divided into three main categories:

117 a **Random or Complete Spatial Randomness (CSR):** where the points or events are
118 randomly distributed and can be modelled via Poisson process.

119 b **Clustered:** where points or events attract to each other in space forming small groups called
120 clusters.

121 c **Regular:** where the points repel each other.

122 A point pattern analysis (PPA) is mainly concerned with describing and making sense of the
123 process that could have generated these random patterns, for eg:- the occurrence of defects in a

124 material is controlled by various parameters linked to material thermodynamical properties. A PPA
125 can be described with two properties [32], namely:

126 a **First order properties:** Are the descriptions on the basis of intensity functions, density of
127 defects at a particular material space for example.

128 b **Second order properties:** are the descriptions on the basis of interactions between each
129 event or points in their material space.

130 First order properties are studied over a sub-region for a large number of events or points. Variations
131 in these properties from each sub-region to another can make the point pattern inhomogeneous.
132 The first order properties are helpful for a global spatial distribution analysis but aren't efficient to
133 distribute the defects spatially in material space. Moreover, global parameters like volume fraction
134 of defects in a specimen has no strong relationship with fatigue life. Hence, patterns of defects in
135 space are studied via second order properties which include: Nearest Neighbor function (NND),
136 Ripley's K-function, etc. Second order properties are those where an occurrence of each event is
137 linked or dependent on one another characterized by the distance between them. Point patterns
138 are statistically compared with complete spatial randomness (CSR) of the null hypothesis for a
139 thorough analysis.

140

141 2.2 Poisson process

142 CSR is a state where events or points occur randomly in space with no interactions between
143 each other. CSR forms the continuum of the natural ordering or patterns of events. On the either
144 side of this continuum lies clustered and regular state of point patterns as explained by [33]. CSR
145 can be modelled with just one parameter such as the expected density of points in space. This can
146 be done via Poisson process since any random event follows a Poisson distribution with a mean
147 value or expected value (density of defects in this case) which is given by,

$$P\{N(V) = k\} = \frac{(\lambda)^k}{k!} \exp^{-\lambda} \quad (1)$$

148 Where λ is number of points per unit volume, sometimes also called as rate parameter, V is
149 the volume of material space in our case and N is the possible random variable. Equation 1
150 gives the probability of N being equal to k . For n disjoint sets V_1, \dots, V_n the random variables
151 $N(V_1), \dots, N(V_n)$ are independent of each other i.e., each point is stochastically independent and
152 there exists no interaction between them which defines CSR. This stochastically independent state
153 of point pattern is therefore often used as reference to evaluate if a point pattern is clustered or
154 dispersed (attracting or repulsing).

155

156 2.3 Univariate and Bi-variate Ripley's K-function

157 Ripley's K-function is an effective tool to quantify second order properties of a spatial point
158 pattern. With distance between each pair of events or points as the main parameter, Ripley's k-
159 function can estimate the probable number of events or points that can be found within a particular
160 distance. K-function can be expressed in multiple variants. When all the points in the study region
161 belongs to one type or class, it is said to be univariate K-function and when the points are divided

162 into two different types or classes, it can be called as bivariate K-function. In general form, K-
 163 function is given by,

$$K(d) = \lambda^{-1} E[\text{number of events within distance } d \text{ of a randomly chosen event}] \quad (2)$$

164 which is given as,

$$K(d) = \frac{V}{N} \sum_{i=1}^N \sum_{i \neq j}^N \frac{I(r_{ij} < d)}{N} \quad I \rightarrow \begin{cases} 1, & \text{if } r_{ij} < d \\ 0, & \text{otherwise} \end{cases} \quad (3)$$

165 Where V is the volume and N is the total number of points in material space. V/N is nothing but
 166 λ^{-1} which is the intensity of events (points) or the number of points in a unit volume and I is the
 167 identity operator which equals to one if the distance between point i and j is less than distance
 168 d and equals to zero otherwise. The value of K-function is usually compared to the theoretical
 169 value of K for CSR or homogenous poisson process. From the null hypothesis for CSR, Ripley's
 170 K-function reduces to the volume of the sphere with radius equal to distance d ,

$$K_{poisson}(d) = \frac{4}{3} \pi d^3 \quad (4)$$

171 The deviation of K from the theoretical value can estimate the nature of spatial distribution of
 172 events. If $K(d) > K_{poisson}(d)$, the pattern is said to be clustered and vice versa.

173 In bi-variate K-function, the events or points are classed into two types, for example, Orange trees
 174 and apples trees, stars and planets etc. Bivariate functions as a whole can be represented in matrix
 175 form where K_{11} and K_{22} are K-functions of type 1 and type 2 points. **The intensities of each type**
 176 **λ_1 and λ_2 are the two variables of the bi-variate K-functions.** The interaction between these two
 177 processes is measured with cross K-function K_{12} . The procedure to measure K_{12} remains the same
 178 as univariate K-function except that within a sphere or circle, number of other type points are
 179 counted. Cross K-function is given as,

$$K_{12}(d) = (\lambda_1 \lambda_2 V)^{-1} \sum_{i=1}^{N_1} \sum_{j=1}^{N_2} I(r_{ij} < d) \quad (5)$$

180 Where, λ_1 and λ_2 are the intensities of type 1 and type 2 points, N_1 and N_2 are the number of type
 181 1 and type 2 points or events.

182

183 **2.4 Estimation of parameters for Neyman-Scott process**

184 If a pattern is homogeneous, Poisson process can generate the pattern whilst for inhomogeneous
 185 pattern, strategies like Neyman-Scott process, Strauss process or Matern process needs to be used.
 186 The parameters for such processes needs to be estimated in prior [21]. In the current work, Neyman-
 187 Scott process is used where the points are classified into two types: parent and children. Parent
 188 forms the center around which children points are distributed with a known distribution whilst
 189 parent points are distributed homogeneously in space.

190 To classify the defects into types, a parameter θ was introduced in K-function. This parameter can
 191 be called as threshold size parameter with which defects are classified based on their sizes. With
 192 the introduction of θ , K and K_{12} functions can be expressed as,

$$K_{kk}(d, \theta) = \frac{V}{N_k(\theta)} \sum_{i=1}^{N_k(\theta)} \sum_{i \neq j}^{N_k(\theta)} \frac{I(r_{ij} < d)}{N_k(\theta)} \quad (6)$$

$$K_{12}(d, \theta) = (\lambda_1 \lambda_2 V)^{-1} \sum_{i=1}^{N_1(\theta)} \sum_{j=1}^{N_2(\theta)} I(r_{ij} < d) \quad (7)$$

194 where, $k = 1, 2$ depending on type of defect, $\lambda_1 = \frac{N_1(\theta)}{V}$, $\lambda_2 = \frac{N_2(\theta)}{V}$, $N_1(\theta)$ and $N_2(\theta)$ are the number
 195 of type 1 and type 2 defects.

196

197 2.5 Data acquisition

198 For the entire process, 4 reference samples of IN100 were used which were tomographed using
 199 Nikon XT H 450. XCT images of these specimens were processed and the defects were segmented
 200 to create binary masks. Each connected defect volumes were labelled separately such that they
 201 are identified and accessible in the segmented volumes. The resolution of each voxel was $(25 \mu\text{m})^3$.
 202 For the training of GAN and CNN, the defect volumes were cropped and rigorous augmentation
 203 techniques were applied to increase the size of database via random rotation, flip etc. Finally
 204 around 1200 pores and 1200 shrinkages were resized to a shape of $32 \times 32 \times 32$ pixels for pores and
 205 $64 \times 64 \times 32$ pixels for shrinkages. Pixel values of the dataset were normalized from $[0, 255]$ to $[0, 1]$.

206

207 2.6 Deep learning networks

208 Two Deep learning neural networks are integrated together in this work to generate defect: GAN
 209 and CNN. GANs are a generative model which usually contains two blocks of networks namely,
 210 generator and discriminator. The generator \mathcal{G} takes in random 1D vector z and generates a 3D
 211 image of defect volumes while discriminator \mathcal{D} is trained with both real $\mathcal{D}(x)$ and generated image
 212 $\mathcal{G}(z)$ to predict if the image is real or fake. The generator tries to minimize the value function of
 213 discriminator whilst the discriminator tries to maximize it. Hence, approach of GANs are sometimes
 214 also referred to as minmax game:

$$\min_{\mathcal{G}} \max_{\mathcal{D}} V(\mathcal{D}, \mathcal{G}) = \mathbb{E}_{x \sim p_{data}(x)} [\log(\mathcal{D}(x))] + \mathbb{E}_{z \sim p_{data}(z)} [\log(1 - \mathcal{D}(\mathcal{G}(z)))] \quad (8)$$

215 where p_{data} is the distributions pertaining to real images and z is the input distributions to generator
 216 \mathcal{G} . The convergence of the network is reached when the generator successfully fools the discriminator
 217 and discriminator fails to predict authenticity of the image. Theoretically, the value function at
 218 convergence is 0.5. In the current work, DCGAN inspired architecture has been used with binary
 219 cross entropy loss function. CNNs on the other hand are fairly simple to train. The network is
 220 trained to predict the actual width, height and depth of the real defects by training on the resized
 221 images [34–36]. The convergence is achieved by minimizing the mean squared error between the
 222 actual and predicted size via stochastic gradient descent.

223

224 **Network structure:** In our model, generator takes in a normally distributed random input
 225 vector of size 128. The input layer is connected to a fully dense layer followed by 3 transposed con-
 226 volutional layers and a convolution layer with kernel size of 4 and a stride of 2. Batch normalization
 227 and ReLU activation layers are added in between except in the last convolution layer and finally a
 228 sigmoid layer at the end. Discriminator on the other hand is an exact mirror of generator except
 229 for the last layer which is one single output. Furthermore, the ReLU layers are replaced with Leaky
 230 ReLU activate layers. A gaussian kernel initializer is used to assign initial values of weights and
 231 bias with a mean of 0 and standard deviation of 1.

232 The architecture of CNN contains 4 convolutional layers along with max pooling layers of size 2.

ReLU activation layers are added between each convolutional and max pooling layers followed by a dense fully connected layer and 3 linear output neurons at the end.

235

Training procedure: For the GANs, a batch size of 16 was used with adaptive moment estimation optimizer (ADAM) [37]. Learning rate for generator was set 2 times that of discriminator with a value of 0.0002. Generally in the Vanilla GAN, the generator is updated once per each update of discriminator. As a result of which the discriminator learns quicker when compared to generator. Therefore the generator is trained twice for each updated of discriminator. This helps to keep the balance in the training of generator and discriminator. Furthermore, hyperparameters such as learning rate of discriminator and generator, decay parameter of optimizer, number of filters of each layer etc. were tuned via a random search method. Model performance was seen to largely depend on learning rates and the number of filters associated to each layer. Initially, vanishing gradient problems were encountered during the training. However, adding batch normalization layers along with one sided noise smoothing of labels fixed the issue. One sided label smoothing is a method to add a small noise to the labels of discriminator. A random noise of $\pm 2\%$ was added to labels. CNNs also use ADAM optimizer with a learning rate of 0.001 and batch size of 32. CNN is converged by minimizing the mean squared error.

3 Results

3.1 Spatial point pattern

251

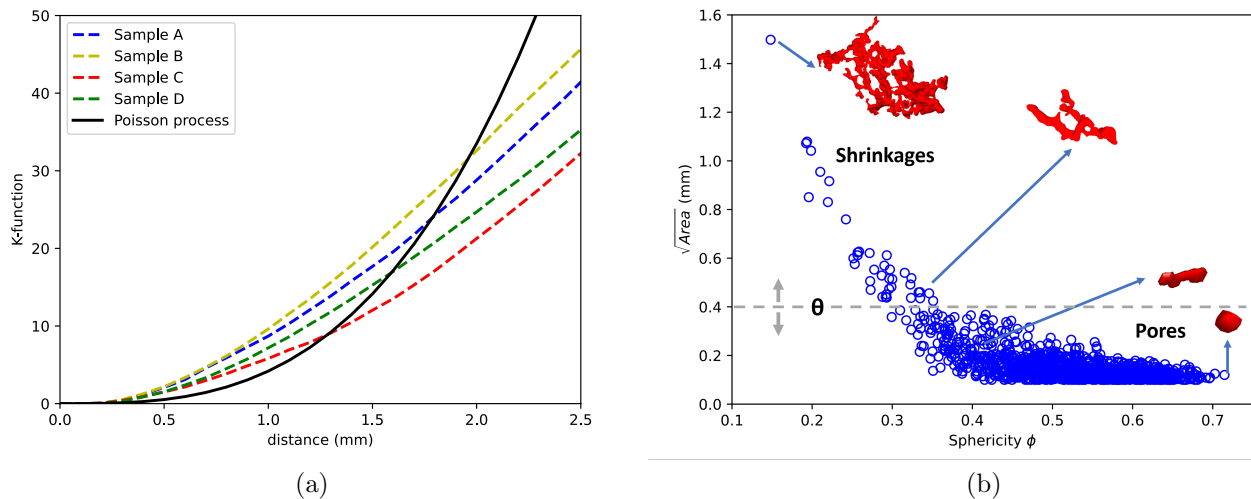


Figure 2: a) K-functions of samples comprising all defects showing the aggregation and dispersion effects (see text for more details) b) Defect size vs sphericity displaying the evolution of morphology along the defect size with shrinkages being very tortuous while pores more spherical in shape.

Spatial point patterns of defect distribution in 4 reference microstructures of IN100 were analysed via Ripley's K-function [14, 21]. K-functions are compared to the theoretical value of K for complete spatial randomness (CSR) or homogeneous Poisson process [38, 39]. Any deviations from Poisson process indicates the nature of spatial pattern i.e., if $K(d) > K_{poisson}(d)$, the points are said to be attracting or clustered and vice versa. From figure 2a, strong clustering effects are seen in

256

257 short distance ranges ($K(d) > K_{poisson}(d)$) and dispersion in large ranges ($K(d) < K_{poisson}(d)$). To
258 simulate such an in-homogeneous pattern, different strategies like Strauss process, Matern process
259 or Neyman-Scott process need to be employed [21].

260
261 The Defect size is plotted as a function of sphericity in figure 2b. Defect size is defined by
262 \sqrt{Area} where $Area$ is the projected area of defect on a plane perpendicular to loading direction [6]
263 while sphericity is a morphological parameter which measures how spherical a defect is: a value of 1
264 indicates a perfectly spherical defect [15]. Sphericity, ϕ is given by $\frac{\pi^{1/3}V_p}{A_p}$ where V_p is the volume of
265 defect and A_p is the surface area. Figure 2b shows a clear inverse relationship between defect size
266 and sphericity (defects get more and more spherical as the size reduces). Indeed, small pores are
267 mostly formed due to trapped gases while larger pores are shrinkages and tend to be much more
268 tortuous as the size increases. From figure 2b, it can be seen that defect size ranges from 100 μm to
269 1.5 mm. Due to this large variance in defect size, the K-function was modified to assess attraction
270 or repulsion among specific groups of defects (classified based on their size). A defect size threshold
271 θ which is a \sqrt{Area} value was introduced to classify defects into two groups (see section 2.4), the
272 two groups being shrinkages (larger defects) and pores (smaller defects). By varying θ , it is possible
273 to investigate the existence of two different processes in the formation of voids via bivariate
274 K-functions (see section 2.3). By splitting the defect into two groups as type 1 for defects of size
275 larger than θ and type 2 for defects smaller than θ , it is assumed that defects of type 1 and type 2
276 are two different processes for which K-functions and cross K-functions are analysed. Defects are
277 initially classified at $\theta = 1$ mm and varied upto $\theta = 0.1$ mm.

278 Cross K-function is a method to estimate interaction between two processes i.e., the spatial ordering
279 of type 2 defects around type 1 defects [21]. This kind of analysis helps to understand if the smaller
280 defects are aggregated with respect to each other or with the larger defects and furthermore aids
281 to simplify the simulation of in-homogeneous point process. Bivariate K-functions together can be
282 described in the form of a symmetrical matrix given that pattern is stationary where, K_{11} and K_{22}
283 are K-functions of type 1 and type 2 defects and K_{12} is the cross K-function between point process
284 of both type of defects. In other words, K_{11} is K-function of all defects larger than θ and K_{22} for
285 the defects smaller than θ . As θ reduces, defects from type 2 group are moved to type 1 group.

286

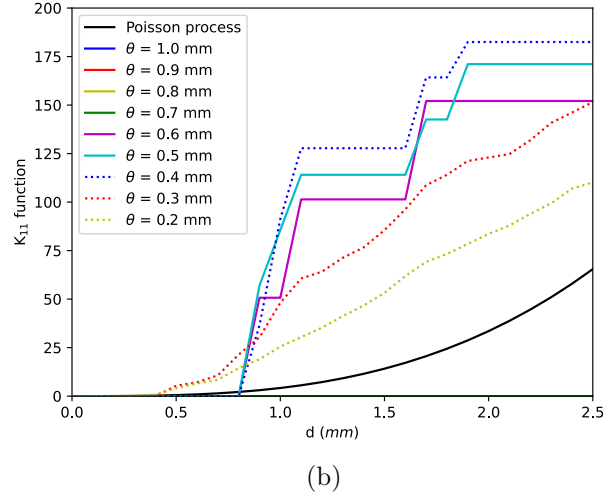
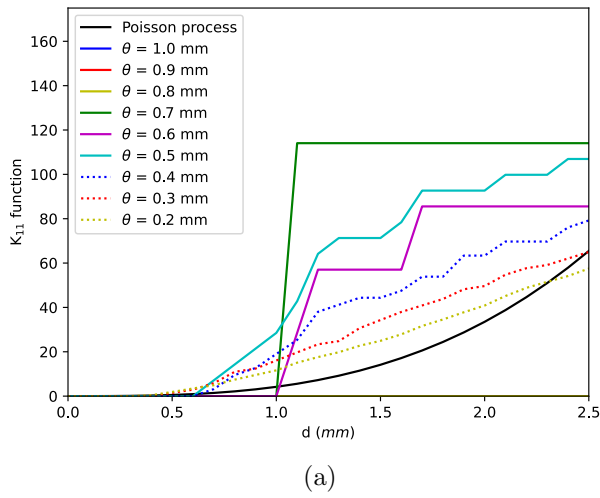


Figure 3: K_{11} functions for a) sample D b) sample B for different values of θ .

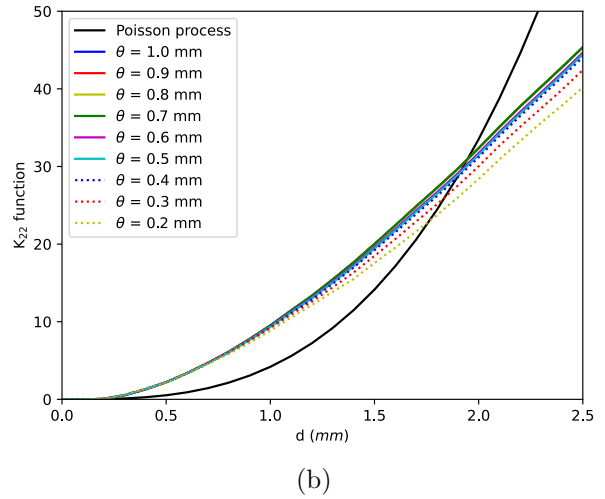
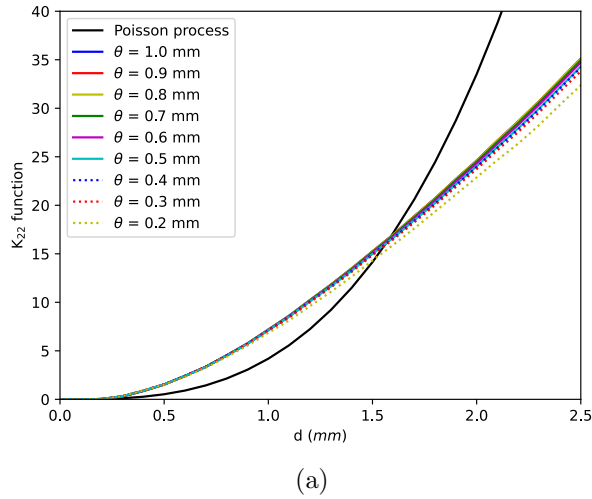


Figure 4: K_{22} functions for a) sample D b) sample B for different values of θ .

287 Results from K_{11} , K_{22} and K_{12} functions of two reference samples are shown in figures 3, 4 and
 288 5. From figure 3, strong aggregation can be seen among defects larger than [approximately 0.4 mm](#)
 289 (θ values above 0.4) but as the smaller defects are considered, the clustering effect reduces (with
 290 respect to θ). This reduction is due to the fact that smaller defects which are necessarily pores
 291 are spread across the length of the sample. Similar effect can be seen for K_{22} functions where the
 292 function remains more or less same for θ values between 1 - 0.4 mm and reduces thereafter signifying
 293 that clustering is driven by defects larger than 0.4 mm.

294

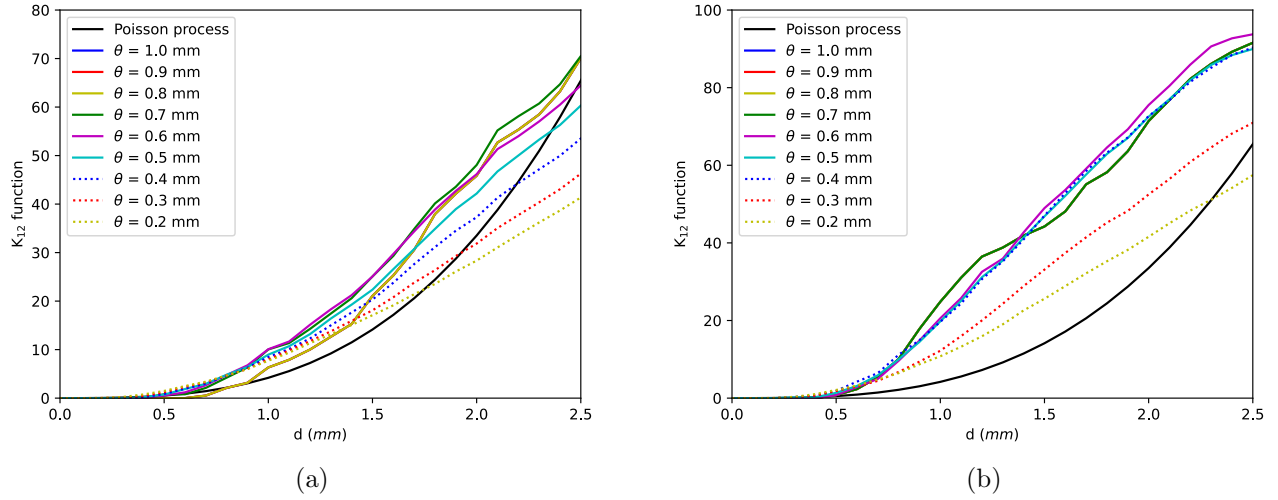


Figure 5: K_{12} functions for a) sample D b) sample B for different values of θ .

295 The interaction between two classes of defects with respect to parameter θ can be described
 296 with cross K-function K_{12} . At a given θ , K_{12} function measures if the defects smaller than θ are
 297 clustered or dispersed with larger defects. From figure 5b, it is seen that smaller defects are strongly
 298 clustered with larger defects upto a value of 0.4 mm and remains almost same upto this value similar
 299 to results of K_{11} and K_{22} functions. However, in some samples K_{12} function reduces marginally
 300 with respect to θ even for values greater than 0.4 mm. This reduction is attributed to the existence
 301 of secondary and tertiary clusters apart from one large primary cluster as depicted in figure 1d. The
 302 weak attraction of these subordinate clusters which contains defects larger than 0.4 mm reduces the
 303 cross K-function as seen in figure 5a. Similar effect can also be seen in K_{11} function of this sample,
 304 see figure 3a.

305
 306 Given the fact that in most scenarios K-functions remains nearly constant upto a θ of 0.4 mm,
 307 it is evident that defects above and below this size follow different processes of void formation
 308 mechanisms. One of the processes is where the smaller defects are nucleated randomly across the
 309 length of the sample while the other where the voids are localised to form clusters i.e., the larger
 310 defects whose K-functions shows strong aggregation. These two processes interact with each other
 311 causing the smaller defects to be attracted towards larger defects. This can also be seen in 1d where
 312 defects smaller than 0.4 mm (colored in blue) are spread across the sample but interacts with larger
 313 defects (colored in red) to form clusters. Those larger than 0.4 mm are certainly shrinkages which
 314 are tortuous in shape as seen in figure 2b whose formation is linked to thermodynamical processes
 315 of solidification whilst the rest are pores formed mostly due to trapped gases. Nevertheless, from
 316 figure 4 negligible attraction effect is seen even for pores ($K_{22} > K_{poisson}$) due to the interaction
 317 between two processes. It is however important to note that the clustering effects at all $\theta < 0.4$ mm
 318 for K_{11} and K_{12} functions are not caused by the same effect. In these functions, defects larger than
 319 θ are included in the calculations i.e., for example, at a θ of 0.1 mm, K_{11} function is measured for
 320 all defects larger than 0.1 mm. Therefore, in these functions the clustering effect for lower θ values
 321 is induced by the larger defects. Finally, with the knowledge of existence of two processes and the
 322 interaction between them as described by bivariate K-functions, Neyman-Scott process can be used
 323 to generate such an in-homogenous point pattern. In this process, the parent events or defects are
 324 distributed homogeneously in the material space and children defects are distributed around the

parent defects [40]. Shrinkages or defects larger than 0.4 mm typically found in the defect cluster are the parent defects whilst the pores are children defects.

However, nucleation of parent defects is in-homogeneous and occur at specific points along the length of the sample defined by a mixed Gaussian distributions as seen in figure 6. Furthermore, it is seen that children defects follow the same distribution along the axis of the sample due to the interaction between the two processes as already invoked. **More importantly, presence of multiple clusters is seen in the number of gaussians of this mixed gaussian distribution.** Mixed Gaussian distributions or Gaussian mixture models (GMM) are characterised by means μ_k , standard deviation σ_k and weights π_k where k is the number of Gaussians [41, 42]. Via expectation maximization algorithm, respective means, standard deviations and weights of each Gaussians can be found. Average standard deviation of the parent defects' Gaussians was found to be approximately 9 pixels or 225 μm whilst the means were found to be coherent with those of children defects as also seen in figure 6. Each Gaussian of parent defects acts as seeds for the nucleation of clustered defects in that zone of the sample. This preference of clustering along the length of the sample maybe due to solidification processes of cylindrical ingot bars which are used to machine the samples. Furthermore, it can also be due to the choice of location and orientation of samples to be machined from ingot bars: the axis of samples were placed parallel to the axis of the ingot bars during machining.

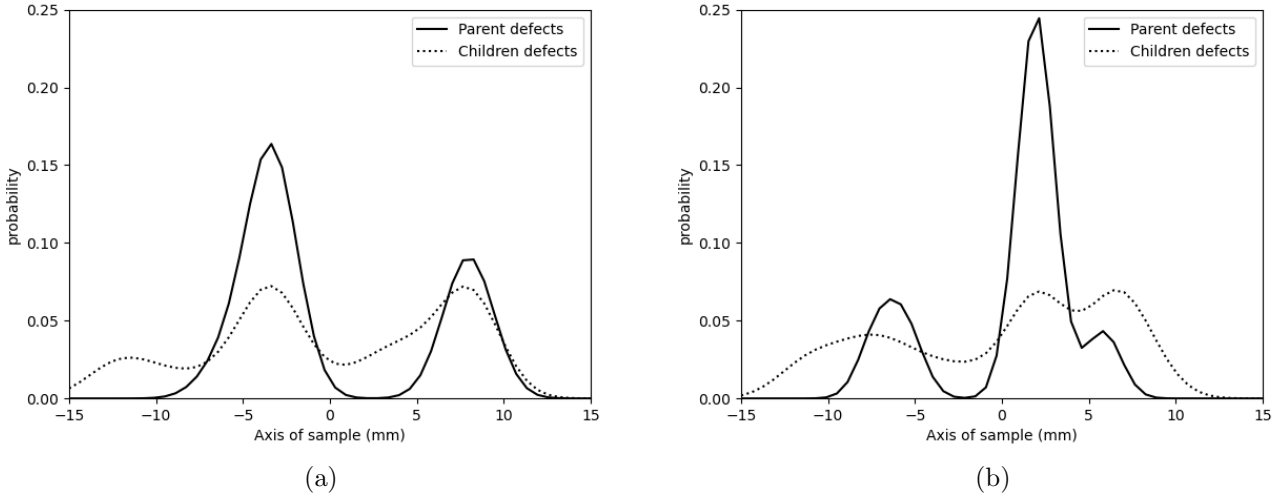


Figure 6: Distribution of defects along the length of sample a) D b) C showing the existence of multiple clusters.

3.2 Generation of synthetic defects

Morphology of the defects varies with respect to its size in an exponential pattern [15, 43]. It is difficult to train GANs to reproduce defects that can respect this relationship since all defects will be initially resized to a fixed size for training. Hence GANs were discretized into two parts i.e., two adversarial networks were trained to generate defects: one for shrinkages (defects > 0.4 mm) and the other for pores (defects < $\theta = 0.4$ mm). θ here is the threshold parameter as determined via SPP analysis. Since the number of shrinkages and pores were insufficient to train the network, a rigorous data augmentation step was carried out to increase the database size. The individual defect volumes were randomly rotated in 3D with angle bounds of -45 deg to $+45$ deg, flipped and inverted in the data augmentation step. All defects are then resized to a fixed size of $64 \times 64 \times 32$ voxels

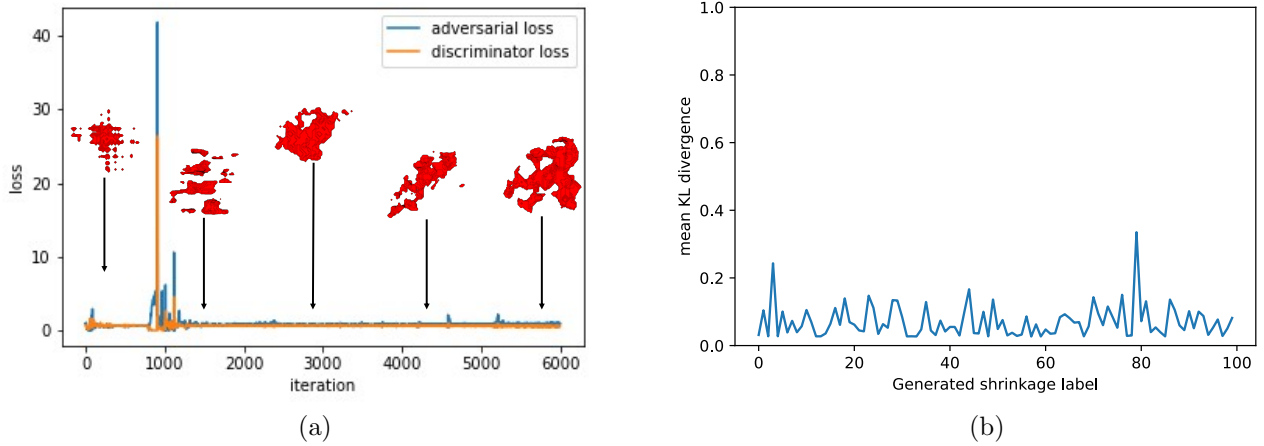


Figure 7: a) Evolution of loss of GAN and generated defect along the training period with iterations being each update of discriminator b) KL divergence value between gaussian curvature of generated shrinkage and mean gaussian curvature of real shrinkage showing that generated shrinkages are similar to real defects.

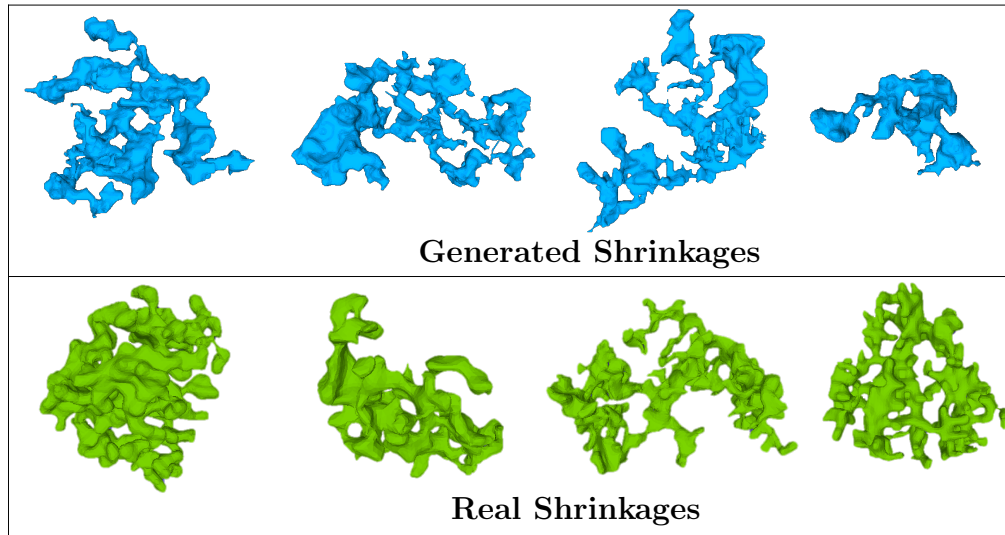


Figure 8: Examples of few generated shrinkages along with real shrinkages.

352 for shrinkages and $32 \times 32 \times 32$ voxels for the pores before training the adversarial networks. The
 353 resizing of images is done by applying a zero order interpolation function $Image_{resized} = D(Image)$
 354 where D is the interpolation function. To maintain the balance between generator and discriminator
 355 networks, the generator is updated twice per each update of discriminator. Furthermore, adding
 356 a small noise to the labels of discriminator has shown to improve the training of the adversarial
 357 network. The adversarial and discriminator loss balance out after as less as 5 epochs and the model
 358 would be trained within 60 epochs.

359 Since the generated defects are also of fixed size similar to training data, CNNs are used to learn
 360 the inverse of the interpolation function used and find the dimensions of defect's 3D slices. Since a
 361 relationship is assumed between defect size and morphology, it is fairly an easy and quick process to
 362 train the CNNs [34, 35]. The trained generators and CNNs can then be integrated to generate defects

363 of various sizes and morphologies for the synthetic microstructures. Trained generator generates
 364 the defect and lets trained CNN to predict its original size. The defect (3D image stack)
 365 is then upsampled and filtered to remove disconnected volumes as the final procedure in the generation of
 366 defects i.e. the largest volume is retained [44].

367 Evolution of the adversarial and discriminator loss of GAN is shown along with the evolution of
 generated defects with each discriminator update in figure 7a. The generated defects are validated

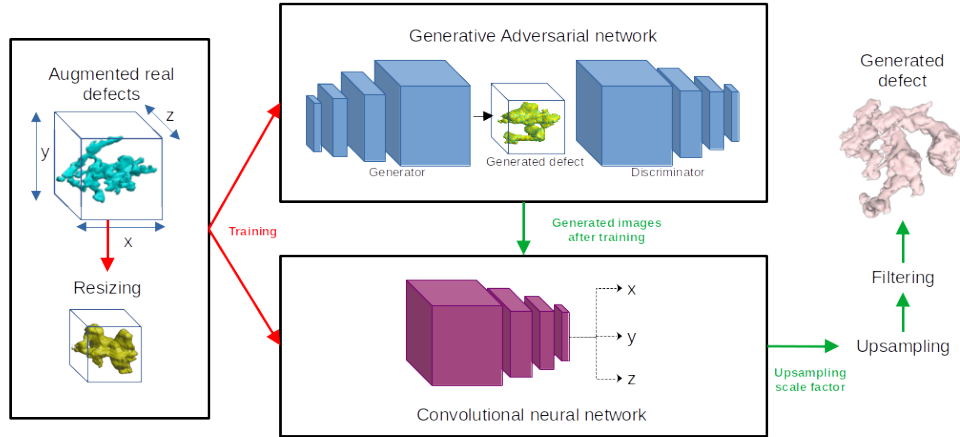


Figure 9: Representation of procedures for the generation of synthetic microstructures.

368
 369 by comparing local Gaussian curvatures [45] of generated and real defects. Gaussian curvatures are
 370 defined as product of principal curvatures (or eigen vectors of local curvatures) at each vertex of the
 371 surface mesh. In this work, the gaussian curvatures are measured as per the methods described by
 372 [46] using the python module trimesh [47]. Gaussian curvatures measured at all points on a given
 373 defect's surface forms a gaussian distribution. For each generated defect, the distance between its
 374 gaussian curvature distributions with the mean distribution of real defects is measured via Kullback-
 375 Leibler (KL) divergence distance [48] which is given by, $\log \frac{\sigma_R}{\sigma_G} + \frac{\sigma_G^2 + (\mu_G - \mu_R)^2}{2\sigma_R^2} - \frac{1}{2}$ where σ_G is the
 376 standard deviation of gaussian distribution of generated defect, σ_R average standard deviation for
 377 real defects and μ_G and μ_R are the respective means. Smaller the value of KL distance, more similar
 378 the two distributions are while for identical distributions, the value equals 0. As shown in figure
 379 7b, this distance metric remains low exhibiting similarity with real defects and at the same time,
 380 each generated defect is unique as seen in figure 7b and 8. The entire procedure of synthetic defect
 381 generation via GAN and CNN is summarized in figure 9.

382 3.3 Synthetic microstructures

383 During the generation of synthetic microstructure, firstly number of defects in the microstructure is
 384 defined by Poisson random number with $\lambda = E(N_{real})$ where N_{real} is the number of defects in real
 385 specimens. Through statistical estimations (ln-likelihood), generalised extreme value (GEV) was
 386 seen to best fit the defect size distribution in this material. Therefore, identified parameters of GEV
 387 distribution were used to estimate and generate number of defects for given size ranges discretely
 388 via our combined GAN and CNN model that produces unique synthetic defects as explained in

389 previous section. These generated defects are placed in material space using positions defined by K-
390 functions. The distribution of defects in real specimens is heterogeneous, therefore Neyman-Scott
391 process is adopted to replicate this heterogeneity [40] via bivariate K-functions. The shrinkages
392 (defects > 0.4 mm) act as parent defects and pores (defects < 0.4 mm) as children events. Contrary
393 to traditional method, a mixed Gaussian distribution defined on the axis of the sample is used to
394 distribute shrinkages (parent defects) in material space. Mixed gaussian distribution is randomly
395 generated with random number of clusters k , mean μ_k and variance σ_k equal to average variance
396 of reference samples whilst the weights π_k are randomly attributed to each Gaussian k such that
397 their sum equals unity. Each Gaussian of this GMM acts as seeds for the nucleation of primary
398 and subordinate clusters. Shrinkages are placed in the material where their planar co-ordinates
399 (radial positions) are randomly chosen whilst their position along the axis is extracted randomly
400 from the mixed Gaussian distribution. The process generates a random K_{11} function similar to
401 those of reference sample.
402 Furthermore, the children defects (pores) are added around the parent defects (shrinkages) con-
403 serving the interaction between the two processes via K_{12} function and interaction amongst the
404 pores via K_{22} functions. Since volume of material space is constant and the number of defects
405 are defined by Poisson random number, expected number of defects within any given distance d
406 can be computed using equations 6 and 7 with respect to maximum and minimum K-functions of
407 reference sample. It is ensured that K_{12} function of generated sample is always in between the
408 lowest and the highest value of reference samples. To not over-constrain the addition of gener-
409 ated defects as per K-functions in the material space, a small tolerance value is added to bivariate
410 K-functions such that K-functions of each generated microstructure is similar to real specimens
411 but unique. This way, the entire generation process is randomized and each generated microstruc-
412 ture is a Poisson random output with $\lambda =$ characteristics of real specimens. During this process,
413 attention is given to avoid overlapping of defects within themselves and with the material boundary.
414

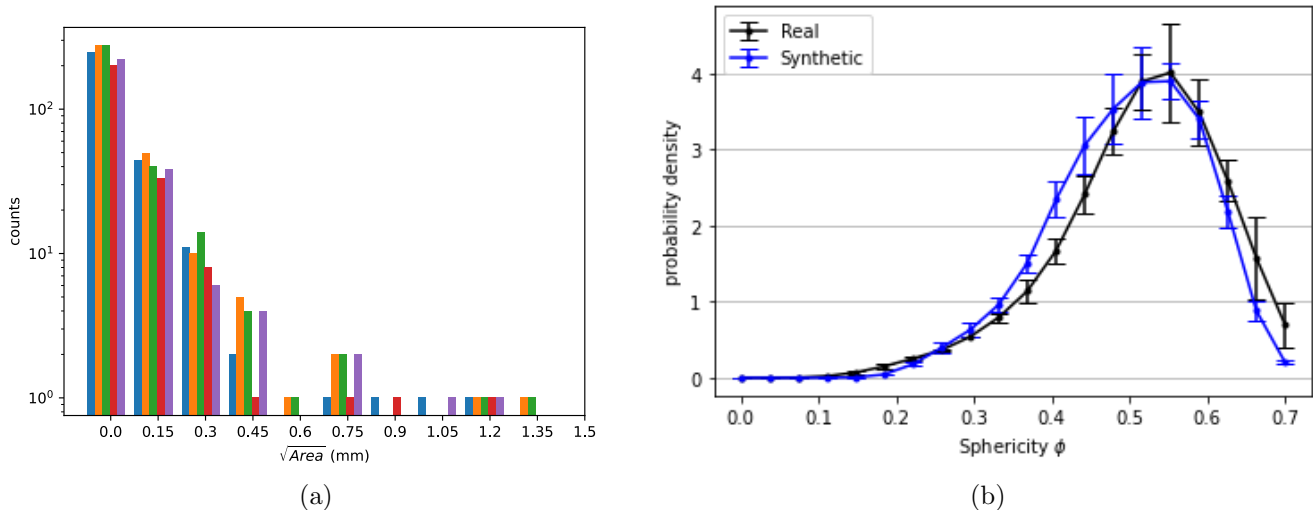


Figure 10: a) Defect size distribution of 5 generated samples showing that each generated microstructure is unique in terms of total number of defects, parent and children defects and maximum defect size b) comparison of probability densities of Sphericity displaying the morphological consistency of generated microstructures with real microstructures.

415 Figure 10a shows the defect size distribution of 5 synthetic microstructures. Total number of

416 defects of each synthetic microstructure is different since it is assumed that total number of defects
 417 follows Poisson distribution [49]. In other words, the number of defects, number of parent and
 418 children defects of synthetic microstructures are determined by the Poisson random number with
 419 metrics of reference specimens as rate parameter λ i.e., average number of defects in samples.
 420 Furthermore, sphericity of defects in the synthetic microstructures are compared in figure 10, the
 421 error bars represent 95 percent variances. Variance bars of generated microstructure's sphericity
 422 distribution lies within the distribution of reference samples at almost all instants describing the
 423 morphological consistency of generated samples.

424
 425 Some of the morphological features are usually correlated in real specimens for e.g., \sqrt{Area} and
 426 ϕ are negatively correlated [8] as also seen in figure 2b. Defect size on the other hand, can be
 427 expressed in various forms like cube root of volume, equivalent radius of a sphere assuming volume
 428 of defect is equal to that of this sphere etc. However, \sqrt{Area} is the one most used to describe fatigue
 429 since it allows to capture mode I crack propagation and is empirically linked to fatigue life, stress
 430 intensity factors of crack etc. [7–9, 50–52]. Inter-dependencies of these features are found to have a
 431 prominence in fatigue performance of a material. The correlations between each of such features can
 432 be measured via pearson correlation coefficient (PCC) [53–56]. Apart from these features, defect
 433 characteristics such as aspect ratio (AR) and distance from free surface (d) plays an important role
 434 in fatigue performance of the material too. AR is a ratio of major axis to minor axis of a defect
 435 projected on a particular plane (here, a plane perpendicular to loading axis which is the axis of
 436 specimen). Figure 11 shows the PCC between each of such features in the form of a matrix. It is
 437 seen that the generated microstructure preserves the inter-relationships between prominent defect
 438 features given the similarities between PCC of real and synthetic samples. Some of the generated
 439 synthetic microstructure are shown in figure 12.

440

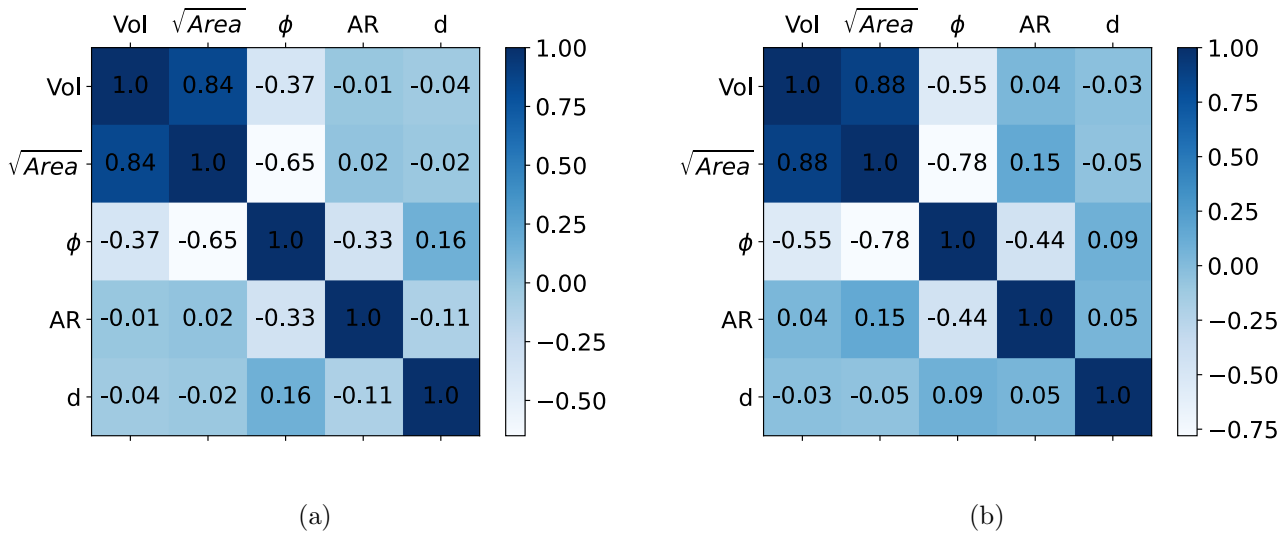


Figure 11: Comparison of global inter relationships between defect features via PCC a) Real microstructure b) Synthetic microstructure, displaying the statistical coherency between synthetic and real microstructure in terms of inter-relationships between the defect characteristics.

441 With all these global statistics being in good agreement with those of real specimens, it can be
 442 said that the method applied for the generation is efficient and can replicate the real specimens

443 in terms of spatial arrangement as well as the morphological and statistical aspects of defects.
444 Furthermore, the uniqueness of each generated microstructure is conserved by the random number
445 of defects that is defined by Poisson distribution, randomly explored K-functions along with new
446 unique defects generated by Deep Neural Networks (DNNs) for each microstructures.

447 4 Discussion

448 A novel strategy has been developed in this work to generate synthetic microstructure in a more
449 cost-friendly and efficient way by integrating SPP analysis and GAN as illustrated in figure 13. Gen-
450 erating greyscale XCT-like images directly would have been impractical and very computationally
451 demanding given the size of samples [57–60]. In addition, it would not have brought any additional
452 information as ultimately the images would have been thresholded to segment the defects. Also,
453 training such a model would require an enormous number of XCT images as input. In this regard,
454 combined use of SPP and DNNs is very effective.

455
456 Bi-variate ripley’s K-functions were used to analyse SPP of real reference microstructures and to
457 generate synthetic ones. Ripley’s K-function is usually affected by edge effects where the measure-
458 ment domain of distance d goes out of the study-region. One of the simplest methods to avoid this
459 error is to measure K-function only upto $1/3^{rd}$ of the largest possible distance. Hence, K-functions
460 were analysed only until a distance of 2 mm in the above cases. Moreover, mechanical interaction
461 of a pore with any shrinkage that is farther away than 2 mm from itself is nearly negligible given
462 a gauge section diameter of 3.7 mm and maximum possible defect size of approximately 1.5 mm.
463 Therefore, for all pores beyond 2 mm from any of the parent defects, a Poisson process was assumed
464 to distribute them in material space.

465 Furthermore, to assess if the process behind the point pattern is different for larger and smaller de-
466 fects, a defect size threshold parameter θ (which is \sqrt{Area} value) was introduced to classify defects
467 based on their sizes. K-functions were analysed at different θ values ranging from 1 mm to 0.1 mm.
468 From our previous work [15], defect population was classified into three groups: a) Shrinkages, b)
469 broken shrinkage pores and c) gaseous pores. Statistically, all defects larger than approximately 0.3
470 - 0.4 mm are shrinkages due to their tortuous morphology as seen in figure 2b and matches with
471 the findings of SPP analysis where a θ of 0.4 mm explicitly classifies the void nucleation mechanism
472 into two groups: Shrinkages and pores. The clustering of defects is driven by shrinkages which
473 nucleates at specific zones in material space and the gaseous porosity interacts with this process as
474 shown by cross K-functions.

475
476 The shrinkages of reference samples were resized to a cuboid size of $64 \times 64 \times 32(px^3)$ to train
477 the generator given the unsymmetrical size of shrinkages in three directions. The average radial
478 width of shrinkages (X and Y direction) was found to be around 50 px while the thickness along the
479 axis (Z direction) was found to be 29 pixels. This difference might be linked to gradient of cooling
480 rate along the radial axis of the ingot bars [61–63]. Furthermore, the current approach to generate
481 defects can be further developed by various means given the increasing popularity of deep learning
482 techniques in materials science [64, 65]. For example, GAN can be conditioned to generate a defect
483 of particular characteristics which would reduce the generation time for synthetic microstructure
484 and give further control to the user [66]. Furthermore, a Deep convolution GAN inspired architeture
485 was used in the current work which can be replaced with more advanced GAN networks such as
486 Wasserstein GANs, Style GAN, Spatial GANs etc [28, 67, 68]. It might also be possible to integrate
487 the generation of grains into this existing model to also capture the effects of grains, slip plane etc

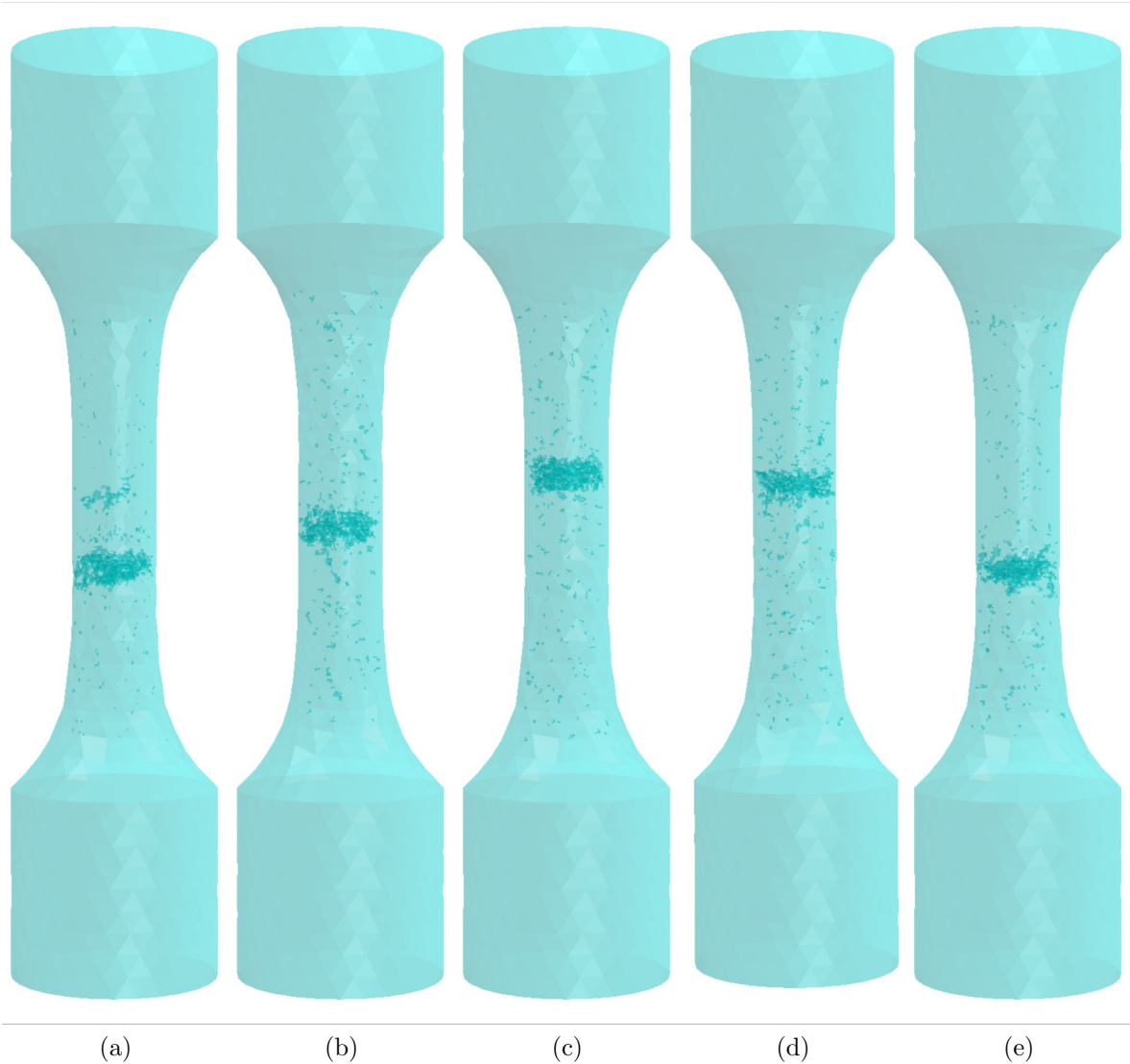


Figure 12: Examples of few generated samples.

488 on material's performance. In the current approach, CNN's predict the X, Y and Z dimension of
489 the generated defect which is later upscaled by interpolation. This can be replaced by bottle-neck
490 architectures like U-nets to directly upscale the defect which should probably remove the filtering
491 step in the model.

492

493 Assuming a similarity in granular characteristics in all samples, grains were not considered in
494 the current approach. However, the approach is also compatible if such granular microstructure
495 needs to be taken into account assuming that there is no correlation between grain size and defects.
496 This can be easily done by generating grains by Voronoi tessellation as demonstrated by Quey et al
497 [69] and performing a boolean operation between the granular microstructure and microstructure
498 containing defects for example.

499

500 With regards to further usage of this strategy, one immediate application would be to analyse the
501 influence of each feature of defect on the number of cycles to failure in fatigue loading via fracture
502 mechanics. Particularly in the case of clustered defects, such an analysis should aid in finding

503 an approximate function that can better predict the fatigue life of samples taking into account
 504 inter-defect interactions. Furthermore, similar to approach of El Khoukhi et al [70], a monte-carlo
 505 like approach can be implemented to estimate fatigue life in a probabilistic fashion. The results of
 506 all these extended works with the aid of synthetic microstructure will be presented in our future
 507 articles.

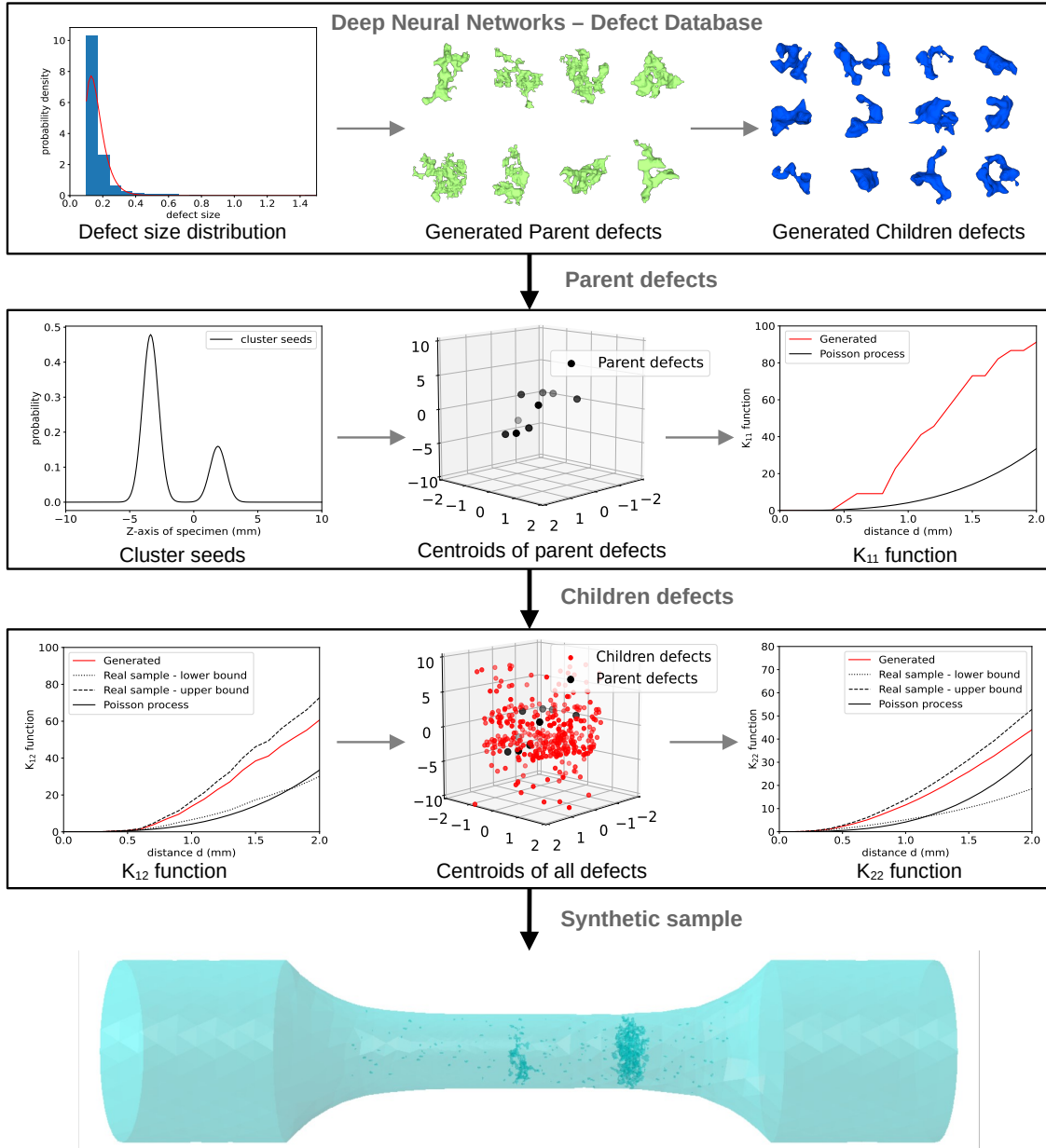


Figure 13: Illustration of the strategy to generate synthetic microstructures.

5 AUTHOR CONTRIBUTIONS

A.K.M.R developed the methods and trained the machine learning models presented in this paper. L.L contributed in the development of codes and methods related to statistical analysis. L.M, V.M and H.P designed the project. H.P supervised the whole project. A.K.M.R wrote the manuscript. All authors contributed in the refinement of manuscript and have given approval for the final version of manuscript.

6 COMPETING INTERESTS

The authors declare no Competing Financial or Non-Financial Interests.

7 DATA AVAILABILITY

X-ray Tomographic datasets of reference samples as well as the codes (machine learning and synthetic microstructure generation module) are available upon reasonable request from the corresponding author.

References

1. Rotella, A., Nadot, Y., Piellard, M., Augustin, R. & Fleuriot, M. Fatigue Limit of a Cast Al-Si-Mg Alloy (A357-T6) with Natural Casting Shrinkages Using ASTM Standard X-Ray Inspection. en. *International Journal of Fatigue* **114**, 177–188. ISSN: 01421123 (Sept. 2018).
2. Wang, Q. G., Apelian, D. & Lados, D. A. Fatigue Behavior of A356-T6 Aluminum Cast Alloys.Part I. Effect of Casting Defects. en. *Journal of Light Metals*, 12 (2001).
3. Kunz, L., Lukáš, P., Konečná, R. & Fintová, S. Casting defects and high temperature fatigue life of IN 713LC superalloy. *International Journal of Fatigue* **41**, 47–51. ISSN: 0142-1123. <https://www.sciencedirect.com/science/article/pii/S0142112311003173> (2012).
4. Wang, Q. G., Crepeau, P. N., Davidson, C. J. & Griffiths, J. R. Oxide Films, Pores and the Fatigue Lives of Cast Aluminum Alloys. en. *Metallurgical and Materials Transactions B* **37**, 887–895. ISSN: 1073-5615, 1543-1916 (Dec. 2006).
5. Kunz, L., Lukáš, P. & Konečná, R. High-cycle fatigue of Ni-base superalloy Inconel 713LC. *International Journal of Fatigue* **32**. Selected Papers of the 17th European Conference of Fracture (ECF 17), 908–913. ISSN: 0142-1123. <https://www.sciencedirect.com/science/article/pii/S0142112309000917> (2010).
6. Murakami, Y. & Endo, M. Effects of Defects, Inclusions and Inhomogeneities on Fatigue Strength. en. *International Journal of Fatigue* **16**, 163–182. ISSN: 01421123 (Apr. 1994).
7. Dezecot, S., Maurel, V., Buffiere, J.-Y., Szmytka, F. & Koster, A. 3D Characterization and Modeling of Low Cycle Fatigue Damage Mechanisms at High Temperature in a Cast Aluminum Alloy. en. *Acta Materialia* **123**, 24–34. ISSN: 13596454 (Jan. 2017).
8. Nadot, Y. Fatigue from Defect: Influence of Size, Type, Position, Morphology and Loading. *International Journal of Fatigue* **154**, 106531. ISSN: 0142-1123. <https://www.sciencedirect.com/science/article/pii/S0142112321003856> (2022).

- 544 9. Koutiri, I., Andreau, O. & Peyre, P. Multi-Scale Approach of HCF Taking into Account Plas-
545 ticity and Damage: Application to LPBF Materials. *Applied Mechanics* **3**, 544–559. ISSN:
546 2673-3161. <https://www.mdpi.com/2673-3161/3/2/32> (2022).
- 547 10. Le, V.-D., Morel, F., Bellett, D., Saintier, N. & Osmond, P. Simulation of the Kitagawa-
548 Takahashi Diagram Using a Probabilistic Approach for Cast Al-Si Alloys under Different
549 Multiaxial Loads. en. *International Journal of Fatigue* **93**, 109–121. ISSN: 01421123 (Dec.
550 2016).
- 551 11. El Khoukhi, D., Morel, F., Saintier, N., Bellett, D., Osmond, P., Le, V.-D. & Adrien, J. Exper-
552 imental Investigation of the Size Effect in High Cycle Fatigue: Role of the Defect Population
553 in Cast Aluminium Alloys. en. *International Journal of Fatigue* **129**, 105222. ISSN: 01421123
554 (Dec. 2019).
- 555 12. Bellomo, N. P., Öztürk, I., Günzel, M., Reed, R., Sundar, V., Ammar, A. & Schwalbe, C. Ident-
556 ifying Critical Defect Sizes From Pore Clusters in Nickel-based Superalloys Using Automated
557 Analysis and Casting Simulation. *Metallurgical and Materials Transactions A* **54**, 1699–1709
558 (2023).
- 559 13. Ripley, B. D. Modelling spatial patterns. *Journal of the Royal Statistical Society: Series B*
560 *(Methodological)* **39**, 172–192 (1977).
- 561 14. Driss, E. K., Nicolas, S., Franck, M., Daniel, B., Pierre, O. & Viet-Duc, L. Spatial point
562 pattern methodology for the study of pores 3D patterning in two casting aluminium alloys.
563 en. *Materials Characterization* **177** (May 2021).
- 564 15. Raghavendra, A. K. M., Armanni, T., Charles, S. & Marcin, L. Role of defects in fatigue
565 performance of IN100. *Engineering Fracture Mechanics* **261**, 108224 (2022).
- 566 16. Diggle, P. J. *Statistical analysis of spatial and spatio-temporal point patterns* (CRC press,
567 2013).
- 568 17. Gaines, K. F., Bryan, A. L. J. & Dixon, P. M. The Effects of Drought on Foraging Habitat
569 Selection of Breeding Wood Storks in Coastal Georgia. *Faculty Research & Creative Activity*
570 **44** (2000).
- 571 18. Pillay, T. & Ward, D. Spatial pattern analysis and competition between Acacia karroo trees
572 in humid savannas. *Plant Ecology* **213**, 1609–1619 (2012).
- 573 19. Veen, A. & Schoenberg, F. P. in *Case Studies in Spatial Point Process Modeling* (eds Baddeley,
574 A., Gregori, P., Mateu, J., Stoica, R. & Stoyan, D.) 293–306 (Springer New York, New York,
575 NY, 2006). ISBN: 978-0-387-31144-9. https://doi.org/10.1007/0-387-31144-0_16.
- 576 20. Neyman, J. & Scott, E. L. Spatial Distribution of GALAXIES—ANALYSIS of the Theory of
577 Fluctuations. *Proceedings of the National Academy of Sciences* **40**, 873–881 (1954).
- 578 21. Dixon, P. M. *Ripley's K Function in Encyclopedia of Environmetrics* ISBN: 9780470057339.
579 eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1002/9780470057339.var046>.
580 <https://onlinelibrary.wiley.com/doi/abs/10.1002/9780470057339.var046> (John
581 Wiley & Sons, Ltd, 2006).
- 582 22. Wilson, P., Saintier, N., Palin-Luc, T., Sudret, B. & Bergamo, S. Statistical study of the size
583 and spatial distribution of defects in a cast aluminium alloy for the low fatigue life assessment.
584 *International Journal of Fatigue* **166**, 107206. ISSN: 0142-1123. <https://www.sciencedirect.com/science/article/pii/S0142112322004595> (2023).
585

- 586 23. Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville,
587 A. & Bengio, Y. Generative adversarial networks. *Communications of the ACM* **63**, 139–144
588 (2020).
- 589 24. Creswell, A., White, T., Dumoulin, V., Arulkumaran, K., Sengupta, B. & Bharath, A. A.
590 Generative adversarial networks: An overview. *IEEE signal processing magazine* **35**, 53–65
591 (2018).
- 592 25. Chun, S., Roy, S., Nguyen, Y. T., Choi, J. B., Udaykumar, H. S. & Baek, S. S. Deep learning
593 for synthetic microstructure generation in a materials-by-design framework for heterogeneous
594 energetic materials. *Scientific Reports* **10** (2020).
- 595 26. Jangid, D. K., Brodnik, N. R., Khan, A., Goebel, M. G., Echlin, M. P., Pollock, T. M.,
596 Daly, S. H. & Manjunath, B. S. 3D Grain Shape Generation in Polycrystals Using Generative
597 Adversarial Networks. *Integrating Materials and Manufacturing Innovation* **11** (2022).
- 598 27. Laloy, E., Hérault, R., Jacques, D. & Linde, N. Training-Image Based Geostatistical Inversion
599 Using a Spatial Generative Adversarial Neural Network. *Water Resources Research* **54**, 381–
600 406 (2018).
- 601 28. Hsu, T., Epting, W. K., Kim, H., Abernathy, H. W., Hackett, G. A., Rollett, A. D., Sal-
602 vador, P. A. & Holm, E. A. Microstructure Generation via Generative Adversarial Network
603 for Heterogeneous, Topologically Complex 3D Materials. *JOM* **73**, 90–102 (2021).
- 604 29. Yang, Z., Li, X., Catherine Brinson, L., Choudhary, A. N., Chen, W. & Agrawal, A. Microstruc-
605 tural Materials Design Via Deep Adversarial Learning Methodology. *Journal of Mechanical*
606 *Design* **140**. 111416. ISSN: 1050-0472. eprint: [https://asmedigitalcollection.asme.org/
607 mechanicaldesign/article-pdf/140/11/111416/6375275/md\140\11\111416.pdf](https://asmedigitalcollection.asme.org/mechanicaldesign/article-pdf/140/11/111416/6375275/md\140\11\111416.pdf).
608 <https://doi.org/10.1115/1.4041371> (Oct. 2018).
- 609 30. Shen, S. C.-y. & Buehler, M. J. Nature-inspired architected materials using unsupervised deep
610 learning. *Communications Engineering* **1** (2022).
- 611 31. Nguyen, P. C. H., Vlassis, N. N., Bahmani, B., Sun, W., Udaykumar, H. S. & Baek, S. S.
612 Synthesizing controlled microstructures of porous media using generative adversarial networks
613 and reinforcement learning. *Scientific Reports* **12** (2022).
- 614 32. Bailey, T. C., Gatrell, A. C., *et al.* *Interactive spatial data analysis* **8** (Longman Scientific &
615 Technical Essex, 1995).
- 616 33. Smith, M. O., Ball, J., Holloway, B. B., Erdelyi, F., Szabo, G., Stone, E., Graham, J. &
617 Lawrence, J. J. Measuring aggregation of events about a mass using spatial point pattern
618 methods. *Spatial Statistics* **13**, 76–89. ISSN: 2211-6753. [https://www.sciencedirect.com/
619 science/article/pii/S2211675315000445](https://www.sciencedirect.com/science/article/pii/S2211675315000445) (2015).
- 620 34. Nibali, A., He, Z., Morgan, S. & Prendergast, L. *Numerical Coordinate Regression with Con-
621 volutional Neural Networks* 2018. <https://arxiv.org/abs/1801.07372>.
- 622 35. Shimobaba, T., Kakue, T. & Ito, T. *Convolutional Neural Network-Based Regression for Depth
623 Prediction in Digital Holography in 2018 IEEE 27th International Symposium on Industrial
624 Electronics (ISIE)* (2018), 1323–1326.
- 625 36. Sateesh Babu, G., Zhao, P. & Li, X.-L. *Deep Convolutional Neural Network Based Regression
626 Approach for Estimation of Remaining Useful Life in Database Systems for Advanced Appli-
627 cations* (eds Navathe, S. B., Wu, W., Shekhar, S., Du, X., Wang, X. S. & Xiong, H.) (Springer
628 International Publishing, Cham, 2016), 214–228. ISBN: 978-3-319-32025-0.

- 629 37. Kingma, D. P. & Ba, J. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*
630 (2014).
- 631 38. Ruan, Y., Yin, P., Li, F., Li, D., Lin, Q. & Li, K. The Accuracy of Determining Cluster Size by
632 Analyzing Ripley's K Function in Single Molecule Localization Microscopy. *Applied Sciences*
633 **9**. ISSN: 2076-3417. <https://www.mdpi.com/2076-3417/9/16/3271> (2019).
- 634 39. Marcon, E., Traissac, S. & Lang, G. The Effects of Drought on Foraging Habitat Selection of
635 Breeding Wood Storks in Coastal Georgia. *International Scholarly Research Network, ISRN*
636 *Ecology* **44** (2013).
- 637 40. Neyman, J. & Scott, E. A theory of the spatial distribution of galaxies. *The Astrophysical*
638 *Journal* **116**, 144 (1952).
- 639 41. Welsh, A. & Richardson, A. in *Robust Inference* 343–384 (Elsevier, 1997). <https://www.sciencedirect.com/science/article/pii/S0169716197150155>.
- 641 42. Hunt, L. & Jorgensen, M. Clustering mixed data. *WIREs Data Mining and Knowledge Dis-*
642 *covery* **1**, 352–361. eprint: <https://wires.onlinelibrary.wiley.com/doi/pdf/10.1002/widm.33>.
643 <https://wires.onlinelibrary.wiley.com/doi/abs/10.1002/widm.33> (2011).
- 644 43. Buffière, J.-Y., Savelli, S., Jouneau, P., Maire, E. & Fougères, R. Experimental Study of Porosity
645 and Its Relation to Fatigue Mechanisms of Model Al–Si7–Mg0.3 Cast Al Alloys. en. *Materials Science and Engineering: A* **316**, 115–126. ISSN: 09215093 (Nov. 2001).
- 647 44. Wu, J., Zhang, C., Xue, T., Freeman, B. & Tenenbaum, J. Learning a probabilistic latent
648 space of object shapes via 3d generative-adversarial modeling. *Advances in neural information*
649 *processing systems* **29** (2016).
- 650 45. Richard, J. L. Detection of zones of abnormal strains in structures using Gaussian curvature
651 analysis. *AAPG bulletin* **78**, 1811–1819 (1994).
- 652 46. Cohen-Steiner, D. & Morvan, J.-M. *Restricted delaunay triangulations and normal cycle in*
653 *Proceedings of the nineteenth annual symposium on Computational geometry* (2003), 312–321.
- 654 47. Dawson-Haggerty. *trimsh* version 3.2.0. <https://trimsh.org/>.
- 655 48. Joyce, J. M. in *International encyclopedia of statistical science* 720–722 (Springer, 2011).
- 656 49. Rényi, A. *Remarks on the Poisson process in Symposium on Probability Methods in Analysis*
657 (1967), 280–286.
- 658 50. Murakami, Y., Takagi, T., Wada, K. & Matsunaga, H. Essential structure of S-N curve: Pre-
659 diction of fatigue life and fatigue limit of defective materials and nature of scatter. *International Journal of Fatigue* **146**, 106138. ISSN: 0142-1123. <https://www.sciencedirect.com/science/article/pii/S0142112320306708> (2021).
- 662 51. Jiang, W., Li, P., Yao, W.-X., Rui, S.-S., Shi, H.-J. & Huang, J. The effect of porosity size on the
663 high cycle fatigue life of nickel-based single crystal superalloy at 980°C. *International Journal of Fatigue* **147**, 106191. ISSN: 0142-1123. <https://www.sciencedirect.com/science/article/pii/S0142112321000517> (2021).
- 666 52. Bortoluci Ormastroni, L. M., Mataveli Suave, L., Cervellon, A., Villechaise, P. & Cormier,
667 J. LCF, HCF and VHCF life sensitivity to solution heat treatment of a third-generation Ni-
668 based single crystal superalloy. *International Journal of Fatigue* **130**, 105247. ISSN: 0142-1123.
669 <https://www.sciencedirect.com/science/article/pii/S0142112319303512> (2020).
- 670 53. Sedgwick, P. Pearson's correlation coefficient. *Bmj* **345** (2012).

- 671 54. Benesty, J., Chen, J. & Huang, Y. On the Importance of the Pearson Correlation Coefficient in
672 Noise Reduction. *IEEE Transactions on Audio, Speech, and Language Processing* **16**, 757–765
673 (2008).
- 674 55. Benesty, J., Chen, J., Huang, Y. & Cohen, I. in *Noise Reduction in Speech Processing* 1–
675 4 (Springer Berlin Heidelberg, Berlin, Heidelberg, 2009). ISBN: 978-3-642-00296-0. https://doi.org/10.1007/978-3-642-00296-0_5.
676
- 677 56. Peng, X., Wu, S., Qian, W., Bao, J., Hu, Y., Zhan, Z., Guo, G. & Withers, P. J. The potency of
678 defects on fatigue of additively manufactured metals. *International Journal of Mechanical Sci-*
679 *ences* **221**, 107185. ISSN: 0020-7403. [https://www.sciencedirect.com/science/article/](https://www.sciencedirect.com/science/article/pii/S0020740322001102)
680 [pii/S0020740322001102](https://www.sciencedirect.com/science/article/pii/S0020740322001102) (2022).
- 681 57. Mangalagiri, J., Chapman, D., Gangopadhyay, A., Yesha, Y., Galita, J., Menon, S., Yesha,
682 Y., Saboury, B., Morris, M. & Nguyen, P. *Toward Generating Synthetic CT Volumes using a*
683 *3D-Conditional Generative Adversarial Network in 2020 International Conference on Compu-*
684 *tational Science and Computational Intelligence (CSCI)* (2020), 858–862.
- 685 58. Sun, L., Chen, J., Xu, Y., Gong, M., Yu, K. & Batmanghelich, K. Hierarchical Amortized GAN
686 for 3D High Resolution Medical Image Synthesis. *IEEE Journal of Biomedical and Health*
687 *Informatics* **26**, 3966–3975 (2022).
- 688 59. Uzunova, H., Ehrhardt, J. & Handels, H. Memory-efficient GAN-based domain translation of
689 high resolution 3D medical images. *Computerized Medical Imaging and Graphics* **86**, 101801.
690 ISSN: 0895-6111 (2020).
- 691 60. Cirillo, M. D., Abramian, D. & Eklund, A. *Vox2Vox: 3D-GAN for Brain Tumour Segmentation*
692 *in Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries* (eds Crimi,
693 A. & Bakas, S.) (Springer International Publishing, Cham, 2021), 274–284. ISBN: 978-3-030-
694 72084-1.
- 695 61. Ge, H., Ren, F., Li, J., Han, X., Xia, M. & Li, J. Four-Phase Dendritic Model for the Prediction
696 of Macroseggregation, Shrinkage Cavity, and Porosity in a 55-Ton Ingot. *Metallurgical and*
697 *Materials Transactions A*, 1139–1150 (2017).
- 698 62. Wu, M., Ludwig, A. & Kharicha, A. A four phase model for the macroseggregation and shrinkage
699 cavity during solidification of steel ingot. *Applied Mathematical Modelling* **41**, 102–120. ISSN:
700 0307-904X. <https://www.sciencedirect.com/science/article/pii/S0307904X16304474>
701 (2017).
- 702 63. Piwonka, T. & Flemings, M. Pore formation in solidification. *Aime Met Soc Trans* **236**, 1157–
703 1165 (1966).
- 704 64. Choudhary, K., DeCost, B., Chen, C., Jain, A., Tavazza, F., Cohn, R., Park, C. W., Choudhary,
705 A., Agrawal, A., Billinge, S. J. L., Holm, E., Ong, S. P. & Wolverton, C. Recent advances and
706 applications of deep learning methods in materials science. *npj Computational Materials* **8**
707 (2022).
- 708 65. Stuckner, J., Harder, B. & Smith, T. M. Microstructure segmentation with deep learning
709 encoders pre-trained on a large microscopy dataset. *npj Computational Materials* **8** (2022).
- 710 66. Han, C., Kitamura, Y., Kudo, A., Ichinose, A., Rundo, L., Furukawa, Y., Umemoto, K., Li, Y. &
711 Nakayama, H. *Synthesizing Diverse Lung Nodules Wherever Massively: 3D Multi-Conditional*
712 *GAN-Based CT Image Augmentation for Object Detection in 2019 International Conference*
713 *on 3D Vision (3DV)* (2019), 729–737.

- 714 67. Jetchev, N., Bergmann, U. & Vollgraf, R. Texture Synthesis with Spatial Generative Adver-
715 sarial Networks. arXiv: 1611.08207 [cs.CV] (2017).
- 716 68. Arjovsky, M., Chintala, S. & Bottou, L. Wasserstein GAN. arXiv: 1701.07875 [stat.ML]
717 (2017).
- 718 69. Quey, R., Dawson, P. & Barbe, F. Large-scale 3D random polycrystals for the finite element
719 method: Generation, meshing and remeshing. *Computer Methods in Applied Mechanics and*
720 *Engineering* **200**, 1729–1745. ISSN: 0045-7825. [https://www.sciencedirect.com/science/
721 article/pii/S004578251100003X](https://www.sciencedirect.com/science/article/pii/S004578251100003X) (2011).
- 722 70. El Khoukhi, D., Morel, F., Saintier, N., Bellett, D., Osmond, P. & Le, V.-D. Probabilistic
723 Modeling of the Size Effect and Scatter in High Cycle Fatigue Using a Monte-Carlo Approach:
724 Role of the Defect Population in Cast Aluminum Alloys. en. *International Journal of Fatigue*
725 **147**, 106177. ISSN: 01421123 (June 2021).