ISSN 1330-3651 (Print), ISSN 1848-6339 (Online)
UDC/UDK 620.192.3:539.422.24]:519.23

# ESTIMATING A PDF OF PARAMETERS OF OXIDE INCLUSIONS ON FATIGUE FRACTURE SURFACE OF STANDARDIZED SPECIMEN

*Ana Bižal, Jernej Klemenc, Matija Fajdiga*

Original scientific paper

Studies show that, if present, inhomogenities should not be neglected when predicting the fatigue life of cast parts. Accounting for their presence presents a problem due to random nature of their location, distribution and size. Knowledge of statistical characteristics of inhomogenities is fundamental for describing their effect on fatigue life. This paper presents a statistical analysis of oxide inclusions which caused fatigue failure of AlSi9Cu3 specimens. Geometrical characteristics were described using geometry parameters $\sqrt{A}$ and $L$. Three different mixed PDFs were used for modelling the distribution of the selected geometry parameters and multivariate Gaussian function was found to be the most appropriate. For estimating unknown parameters of the mixture, modified EM algorithm and REBMIX were used and compared. Results identified modified EM algorithm as more suitable. Bootstrap analysis was carried out for the chosen joint PDF and algorithm, for assigning a measure of variation of the estimated PDF's parameters.

***Keywords:*** *EM algorithm, finite mixture distributions, inclusions, REBMIX algorithm, statistical distribution*

## Procjena utjecaja PDF parametara oksidnih umetaka na lom zbog zamora površine standardiziranih uzoraka

Izvorni znanstveni članak

Ispitivanja pokazuju da se pojave nehomogenosti ne bi trebale zanemariti kod procjene loma zbog zamora lijevanih dijelova. Objašnjavanje njihovog postojanja predstavlja problem zbog slučajnosti njihovih lokacija, raspodjele i dimenzija. Poznavanje statističkih karakteristika nehomogenosti od osnovnog je značenja za opis hjihovog učinka na lom zbog zamora. U ovom se radu daje statistička analiza oksidnih umetaka koji su doveli do loma AlSi9Cu3 uzoraka. Geometrijske karakteristike su opisane pomoću geometrijskih parametara $\sqrt{A}$ i $L$. Upotrebljena su tri različita miješana PDFa za modeliranje distribucije izabranih geometrijskih parametara, a Gaussian funkcija s više varijanti je odabrana kao najpogodnija. Za procjenu nepoznatih parametara mješavine upotrebljeni su i uspoređeni modificirani EM algoritam i REBMIX. Prikladnijim se, prema rezultatima, pokazao modificirani EM algoritam. Provedena je bootstrap analiza za PDF i algoritam odabranog spoja kako bi se odredila veličina promjene procijenjenih PDF parametara.

***Ključne riječi:*** *EM algoritam, raspodjele konačnih mješavina, REBMIX algoritam, statistička raspodjela, umetci*

## 1 Introduction

In attempt to lower production costs manufacturers in automotive industry commonly use die casting technologies to produce vehicle parts and components. Several defects can occur during manufacturing. While some defects affect only the appearance of the structural parts, others can have major effects on the structural integrity of the parts. There are seven basic categories of casting defects: metallic projections, cavities, discontinuities, defective surface, incomplete casting, incorrect dimensions or shape and inclusions. While some imperfections can be avoided by choosing optimal casting parameters, porosity and inclusions can only be reduced and not fully avoided. When designing structural components made from metal castings, reduction of materials' strength due to the presence of inhomogeneities is accounted for by increasing cross sections. By doing so the structures' weight increases and the product can become over designed. With a growing tendency to lean manufacturing and lightweight structures it becomes more and more important to identify location, size and effect of inhomogeneities on material properties of cast parts. The core problem of accounting for the presence of inhomogenities and their effect on endurance of cast parts is linked to the random nature of location, distribution and size of inhomogenities.

Many studies were carried out dealing with the effect of inhomogeneities on material properties of alloys. Most studies that were carried out for aluminum alloys have focused on porosity effects on material properties [1, 2, 3, 4]. They all showed detrimental effects on fatigue behavior. Linder et al. [5] determined a 15 % decrease in fatigue strength for an aluminum alloy when porosity level was increased from initial 0,7 % to 4,1 %. Wang [6] observed that the effect of porosity on fatigue life increases with pore size. For the fatigue life prediction researchers often calculate maximum expected pore size on fracture surface based on pore size distribution using extreme value statistics [5, 7, 8]. However, a scatter of fatigue life is sometimes also the result of shape, size and location variety of inhomogeneities present within the structure.

To account for various parameters describing inhomogeneities and their effect on the fatigue life of the products several models have been developed. Uhrus's study of ball bearings and oxide inclusions [9] showed that only inclusions larger than 30 μm have considerable effect on fatigue life. Duckworth and Ineson [10] discovered that the effect of same-sized inclusions on fatigue life vary depending on where they appear on fracture surface. Ramsey and Kedzie [11] linked the size of inclusions with fatigue strength. Size of inclusions was determined by geometric means of their length and width. De Kazincky [12] used diameter of circumscribing circle to describe the size of inclusion. New geometrical parameter $\sqrt{A}$ for describing inclusion size was suggested by Murakami and Endo [13]. $\sqrt{A}$ is a square root of inclusions' area projected onto the fracture surface of the specimen. Murakami and Endo model presumes that a mere size of inclusions is what effects fatigue life the most and not the inclusion shape. Murakami-Endo model is in accordance with the model of Duckworth and Ineson [10] who deliberately introduced inclusions of oval and

rectangular shape in material and noted negligible distinction between their effects on fatigue life.

On the basis of the studies presented it can be concluded that it is highly important to gain the knowledge of the statistical characteristics of inhomogeneities in castings prior to describing their effect on fatigue life of the product. This paper presents a statistical analysis of oxide inclusions which caused fatigue failure of standard specimens made from an aluminium alloy AlSi9Cu3. For determining E-N curve of AlSi9Cu3 alloy, 294 specimens were manufactured and subjected to cyclic testing until rupture. By visually inspecting fatigue fracture surface of every failed specimen we were able to isolate the specimens containing an oxide inclusion in the fractured cross section. Geometrical characteristics of each inclusion were recorded and the probability density function (PDF) for the selected geometrical parameters of inclusions was estimated. For modelling PDF various finite mixtures of multivariate PDFs were applied. Furthermore, for estimating the unknown parameters of the PDF mixtures two algorithms were used: a modification of an EM algorithm according to Figueiredo and Jain [14] and a REBMIX algorithm [15]. Based on results the most suitable joint PDF for modelling the statistical distribution of the geometrical parameters of inclusions was selected as well as the most suitable algorithm for PDF parameter estimation. For the most suitable combination of PDF and the PDF's training algorithm a bootstrap analysis was carried out for assigning a measure of variation of the estimated PDF's parameters.

Basic definitions of finite mixture models and the two training algorithms are given in Section 2. Experiments and specimens used are presented in Section 3. First the specimens and experiments are presented, followed by determining the geometrical parameters of inclusions on fatigue fracture cross sections of the specimens. The results of the statistical analysis are presented and discussed in Section 4. At the end of Section 4 measures of variation of estimated PDF's parameters are assessed. All the findings are summarized in concluding section.

## 2 Theoretical background

### 2.1 Finite mixture model

The basic finite mixture model for independent $\boldsymbol{d}$-dimensional vector of observations $\boldsymbol{x}$ can be written as [16, 17]:

$$f\left(\boldsymbol{x}\middle|m,\boldsymbol{w},\boldsymbol{\Theta}\right)=\sum_{j=1}^{m}w_j f\left(\boldsymbol{x}\middle|\boldsymbol{\theta}_j\right). \tag{1}$$

$f(\boldsymbol{x}\,|\,\boldsymbol{\theta}_j)$ is a multivariate PDF with a known mathematical form, $m$ is a number of components in the finite mixture, $w_j$ is a weight corresponding to the $j^{\text{th}}$ component in the mixture and $\boldsymbol{\theta}_j$ is a vector of parameters of the applied PDFs. The components $f(\boldsymbol{x}\,|\,\boldsymbol{\theta}_j)$ can generally belong to various parametric families. However, the finite mixture model is usually simplified in the way that all the components belong to the same parametric family, such as Gaussian function, $t$-function, Weibull function, etc. The mixture component can be either a true multivariate PDF (a joint PDF of the multivariate random variable $\boldsymbol{x}$) or a product of univariate PDF belonging to individual components of the multivariate random variable $\boldsymbol{x}$ [18]:

$$f\left(\boldsymbol{x}\middle|\boldsymbol{\theta}_j\right)=\prod_{i=1}^{d}f\left(x_i\middle|\boldsymbol{\theta}_{ij}\right). \tag{2}$$

In our research three different mixed PDFs were applied for modelling the distribution of geometry parameters of inclusions: a mixture of multivariate Gaussian functions, a mixture of multivariate $t$-functions and a mixture of products of univariate Gaussian functions and Weibull functions.

A multivariate Gaussian function is defined as:

$$f_j(\boldsymbol{x})=\frac{1}{(2\pi)^{d/2}\left|\boldsymbol{\Sigma}_j\right|^{1/2}}\exp\left\{-\frac{1}{2}\left(\boldsymbol{x}-\boldsymbol{\mu}_j\right)^{\mathrm{T}}\boldsymbol{\Sigma}_j^{-1}\left(\boldsymbol{x}-\boldsymbol{\mu}_j\right)\right\}, \tag{3}$$

where $\boldsymbol{\mu}_j$ is a $d$-dimensional mean vector, $\boldsymbol{\Sigma}_j$ is a $\boldsymbol{d}\times\boldsymbol{d}$ covariance matrix and $\left|\boldsymbol{\Sigma}_j\right|$ is its determinant and $d$ is the dimensionality of the vector random variables.

The mixture of multivariate Normal PDFs has the following form:

$$f(\boldsymbol{x})=\sum_{j=1}^{m}w_j f\left(\boldsymbol{x}\middle|\boldsymbol{\theta}_j\right)=\sum_{j=1}^{m}w_j\frac{1}{(2\pi)^{d/2}\left|\boldsymbol{\Sigma}_j\right|^{1/2}}\exp\left\{-\frac{1}{2}\left(\boldsymbol{x}-\boldsymbol{\mu}_j\right)^{\mathrm{T}}\boldsymbol{\Sigma}_j^{-1}\left(\boldsymbol{x}-\boldsymbol{\mu}_j\right)\right\}, \tag{4}$$

where $\boldsymbol{\mu}_j$ is the mean vector of the $j^{\text{th}}$ component of the normal mixture and $\boldsymbol{\Sigma}_j$ is the covariance matrix of the $j^{\text{th}}$ component of the normal mixture.

A multivariate $t$-function is defined as

$$f_j(\boldsymbol{x})=\frac{\Gamma\left[\left(v_j+d\right)/2\right]\cdot\sqrt{\det\left(\boldsymbol{\Sigma}_j^{-1}\right)}}{\left(\pi\cdot v_j\right)^{d/2}\cdot\Gamma\left(v_j/2\right)\cdot\left[1+\left(\boldsymbol{x}-\boldsymbol{\mu}_j\right)^{\mathrm{T}}\cdot\boldsymbol{\Sigma}_j^{-1}\cdot\left(\boldsymbol{x}-\boldsymbol{\mu}_j\right)/v_j\right]^{\left(v_j+d\right)/2}}, \tag{5}$$

where $\Gamma$ is the gamma function, $v_j$ represents the degrees of freedom of the $j^{\text{th}}$ component, $\boldsymbol{\mu}_j$ is the mean vector of the $j^{\text{th}}$ component and $\boldsymbol{\Sigma}_j$ is the covariance matrix.

The mixture of multivariate $t$ PDF has the following form

$$\begin{aligned}f(\boldsymbol{x})&=\sum_{j=1}^{m}w_j f\left(\boldsymbol{x}\middle|\boldsymbol{\theta}_j\right)=\\&=\sum_{j=1}^{m}w_j\frac{\Gamma\cdot\left[\left(v_j+d\right)/2\right]\cdot\sqrt{\det\left(\boldsymbol{\Sigma}_j^{-1}\right)}}{\left(\pi\cdot v_j\right)^{d/2}\cdot\Gamma\left(v_j/2\right)\cdot\left[1+\left(\boldsymbol{x}-\boldsymbol{\mu}_j\right)^{\mathrm{T}}\cdot\boldsymbol{\Sigma}_j^{-1}\cdot\left(\boldsymbol{x}-\boldsymbol{\mu}_j\right)/v_j\right]^{\left(v_j+d\right)/2}},\end{aligned} \tag{6}$$

$\Gamma$ is the gamma function, $v_j$ represents the degrees of freedom of the $j^{\text{th}}$ component of the mixture of $t$-functions, $\boldsymbol{\mu}_j$ is the mean vector of the $j^{\text{th}}$ component of the mixture of $t$-functions and $\boldsymbol{\Sigma}_j$ is the covariance matrix of $t$-functions.

In a mixture of products of univariate Gaussian functions and Weibull functions the individual components were represented by a product of the following functions:

$$f\left(\boldsymbol{x}\middle|\boldsymbol{\theta}_j\right)=\frac{1}{\sqrt{2\pi}\sigma_j}\cdot\exp\left\{-\frac{1\left(x-\mu_j\right)^2}{2\sigma_j^2}\right\}, \tag{7}$$

$$f\left(\boldsymbol{x}|\boldsymbol{\theta}_j\right)=\frac{\beta_j}{\eta_j}\left(\frac{x}{\eta_j}\right)^{\beta_j-1}\cdot\exp\left\{-\left(\frac{x}{\eta_j}\right)^{\beta_j}\right\}, \tag{8}$$

where $\mu_j$ and $\sigma_j$ are mean and standard deviation of the univariate Gaussian function – equation (7) and $\beta_j$ and $\theta_j$ are a shape and a scale factor of the univariate Weibull function.

### 2.2 REBMIX algorithm

Since its creation REBMIX has evolved and improved while the core of the algorithm remained unchanged [19, 20, 21, 22]. The basic idea of the algorithm lies in dividing independent observations into two sets, which is followed by a rough estimation of component parameters. Based on the set of random observations $\{\boldsymbol{x}_i;\ i = 1,\ldots,\ n\}$ empirical component densities $f_{j\max}$ are estimated. Then a global mode position is identified and its inherent maximum value of the variable $\boldsymbol{x}_{\max}$. If the number of independent observations is $n < 200$ a Parzen window is used for smoothening of the empirical PDF [23]. Once the global mode position and its empirical density are known, the rough component parameter estimation of the component density $f(\boldsymbol{x}|\ \boldsymbol{\theta}_j)$ is carried out. Algorithm enables user to select component densities from normal, lognormal or Weibull parametric families. The set of independent observations $\boldsymbol{x}$ is clustered into two subsets gradually. The first set, containing $n_j < n$ observations, is linked to the component density, while the second forms the residue that includes $r=n-n_j$ observations belonging to the rest of the components. After the number of observations $n_j$ belonging to the component density is estimated the mixing weight of predictive component density $w_j = n_j/n$ is calculated. The optimal number of components, weights and component parameters are gained iteratively. The process is repeated until the following criteria are met: $n_j > 1$, $w_j > 2jD_{\min}$ and $D_j > D_{\min}/w_j$ where $D_j$ stands for total positive relative deviation of the component density from the empirical density of the $j^{\text{th}}$ component. If any of the above criteria is not met, the enhanced component parameter estimation based on the MLE is produced. The REBMIX package has been published on the Comprehensive R Archive Network and is available at http://CRAN.R-project.org/package=rebmix.

### 2.3 Expectation Maximization Algorithm (EM)

EM algorithm is a general method of finding the maximum-likelihood estimate of the parameters of an underlying mixture distribution from a given data set $\{\boldsymbol{x}_i\ ;\ i = 1,\ldots,\ n\}$. For this purpose a logarithmic likelihood function $\ell$ is defined:

$$\lg \ell\left(\chi,Z|\theta\right)=\sum_{i=1}^{n}\sum_{j=1}^{m}z_{ij}\lg\left[w_j f\left(\boldsymbol{x}_i\middle|\boldsymbol{\theta}_j\right)\right]. \tag{9}$$

Standard EM algorithm considers the given data set to be incomplete or having some values missing. Since complete data is not available for sampling, expected value of logarithmic likelihood function $\ell$ of complete data is calculated using the measured values $\{\boldsymbol{x}_i\ ;\ i = 1,\ldots,\ n\}$ and the current estimates of the mixture parameters $\{\Theta^{(k)},\ q^{(k)}\}$. Calculation begins by selecting the number of components in the mixed PDF and defining the size of the solution space. From the range of possible PDF parameters, initial values of unknown parameters are randomly selected. By iteratively updating equations using parameter values from the previous step $\{\Theta^{(k)},\ q^{(k)}\}$, new PDF parameter values are calculated $\{\Theta^{(k+1)},\ q^{(k+1)}\}$. Therefore, the EM algorithm gives a range of approximate estimates of PDF parameters $\{\boldsymbol{\theta}(t),\ t=0,1,2,3,\ldots\}$. Calculation of the mixed PDF's parameters consists of two steps:

- E-step (expectation step): the conditional expectation of the complete log-likelihood function is calculated, given measured data $\chi$ and current estimate $\boldsymbol{\theta}(t)$. Since $\lg \ell\ (\chi,\ Z\,|\,\boldsymbol{\theta})$ is linear with respect to missing data Z, conditional expectation $W \equiv E[Z|\chi,\ \boldsymbol{\theta}(t)]$ can easily be computed and plug into $\lg \ell\ (\chi,\ Z\,|\,\boldsymbol{\theta})$. The result is the so-called $Q$-function:

$$Q\left(\theta,\boldsymbol{\theta}(t)\right)\equiv E\left[\lg f\left(\chi,Z|\boldsymbol{\theta}\right)\middle|\chi,\boldsymbol{\theta}(t)\right] \tag{10}$$

$$=\lg f\left(\chi,W|\boldsymbol{\theta}\right). \tag{11}$$

- M-step (maximization step): Update of parameter estimates according to

$$\boldsymbol{\theta}(t+1)=\operatorname{argmax}\left\{Q\left(\boldsymbol{\theta},\boldsymbol{\theta}(t)\right)+\lg f\left(\boldsymbol{\theta}\right)\right\}. \tag{12}$$

Iteration process continues until a defined convergence criterion is not met.

EM algorithm has three major drawbacks. Firstly, the final parameter values depend heavily on the chosen initial values. In attempt to overcome that, multiple random starts are used and the estimate with the highest likelihood is chosen. The second drawback is the possibility of algorithm to converge to the boundary of the parameter space. This tends to happen when the number of components assumed is larger than the optimal/true one. The third major drawback is that the number of components in the mixture must be determined beforehand. By implementing the MML criterion (Minimum Message Length) Figueiredo and Jain [14] eliminated this deficiency and significantly reduced EM's sensitivity to initialization. To some extent their modification also avoids convergence to the boundary of the parameter space. Their modification of the EM algorithm has already been successfully used for parameter estimation of mixtures of multivariate normal and $t$-functions by Klemenc and Fajdiga [24]. The modification of EM algorithm by Figueiredo and Jain [14] was also used in our research.

## 3 Experimental data

### 3.1 Specimens and testing

The specimens used in present study were produced with high-pressure die-casting. Their shape complies with ASTM 606 [25] standard. Specimens were produced from

aluminum alloy AlSi9Cu3. Shape and size of the specimens are presented in Fig. 1. To capture solely the phenomenon of inclusions being present within the structure, the specimens were not additionally handled before testing. Surface roughness, though a possible fatigue fracture initiation site, was not considered due to the fact that whenever present in the critical cross section-fatigue fracture always initiated in inclusions' vicinity.

All the fatigue testing was performed on the test stand for static and dynamic tests in the research department of a CIMOS d.d. company. A MTS 810.22 servo-hydraulic test stand was used. When the fatigue test was performed at higher temperatures the specimens were heated using a Hüttinger TIG 5/300 induction heater with a Eurotherm 2604. When measuring strain an MTS 632.53F-14 high-temperature extensometer with a measuring distance of 12 mm was used. The specimen's temperature was measured, following the manufacturers' advice, with a type-K thermocouple. The software packages "MTS FlexTest SE" and "2MTS TestWorks 4" were used for the test-stand control. The content of alloying elements, the loading rate, the surrounding medium and other influential parameters were kept constant during all the experiments in this study. For each combination of testing parameters at least 17 test runs were performed.

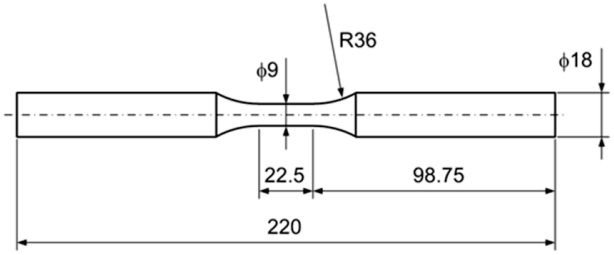


**Figure 1** Test specimens according to the ASTM 606 [26] standards

The fatigue tests resulted in an experimental database that included 294 fatigue-test results for specimens of 9 mm diameter and four testing temperatures (25 °C, 100 °C, 200 °C, and 300 °C). Each sample in the database included the following data: specimen diameter, testing temperature, strain amplitude, stress amplitude and mean value, number of cycles-to-failure.

### 3.2 Determining the inclusions geometry parameters

Fatigue fracture cross sections of all 294 specimens were inspected and photographed. In 34 cases oxide inclusions were present on fractured cross section. Given that the number of specimens with inclusion on the fractured cross section is relatively small, a model for describing the statistical characteristics of inclusion geometry should have had as few parameters as possible. Regarding that, the Murakami-Endo [13] model was chosen. According to this model the geometry and the position of the inclusion is defined with two parameters only. The geometry is described by the area of inclusions' projection on fractured cross section. The squared root of the area is then calculated and the geometry parameter is designated as $\sqrt{A}$ . The position of inclusion on the fractured cross section is defined by the shortest distance from the edge of inclusion to the outer surface of the specimen, $L$. For every specimen with the oxide inclusion both parameters were measured on the basis of a photograph of fractured cross section. The measurements were performed using ImageTool software [26]. Fig. 2 shows an example of fractured cross section with an oxide inclusion and the two measured parameters $\sqrt{A}$ and $L$. Geometric characteristics for all 34 specimens with inclusion are listed in Tab. 1.

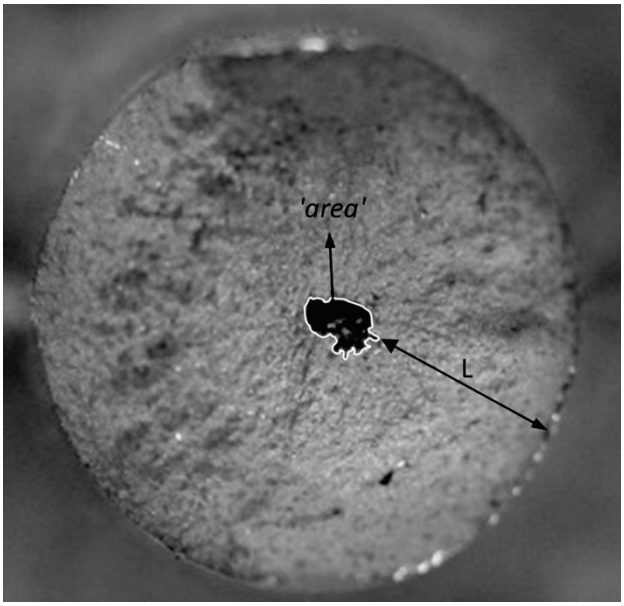


**Figure 2** Inclusion on fractured cross section of a specimen with measured parameters $\sqrt{A}$ and $L$

**Table 1** Measured values of variables $\sqrt{A}$ and $L$

| Spec. | $\sqrt{A}$ / mm | $L$ / mm | Spec. | $\sqrt{A}$ /mm | $L$ / mm |
|---|---|---|---|---|---|
| V 4 | 0,12649 | 2,80 | L 15 | 0,78740 | 0,32 |
| V 10 | 0,17321 | 1,35 | L 18 | 0,78740 | 2,60 |
| V 15 | 0,69282 | 0,60 | U14 | 0,69282 | 1,07 |
| V 16 | 0,88318 | 3,30 | U21 | 0,28284 | 3,70 |
| V 17 | 1,17047 | 2,10 | V 9 | 0,73485 | 2,70 |
| HS 5 | 0,36742 | 3,77 | L 10 | 1,69706 | 0,90 |
| L 7 | 1,26095 | 2,30 | L 12 | 1,50000 | 1,00 |
| U13 | 0,18439 | 1,60 | L 16 | 0,50000 | 0,72 |
| HS 1 | 0,81240 | 3,15 | L 21 | 1,27671 | 2,70 |
| HS 7 | 0,15811 | 0,25 | L 23 | 1,30384 | 0,97 |
| L 9 | 1,10454 | 0,90 | L 24 | 0,94868 | 2,40 |
| V 5 | 0,81240 | 3,80 | L 26 | 1,30384 | 2,28 |
| V 11 | 0,70711 | 3,00 | U2 | 0,53852 | 1,26 |
| U24 | 1,91311 | 2,60 | U5 | 0,40000 | 1,60 |
| U30 | 0,69282 | 4,20 | U38 | 1,32288 | 2,70 |
| L 4 | 1,36748 | 2,80 | U41 | 1,76918 | 2,60 |
| L 6 | 0,97980 | 0,91 | U52 | 1,47309 | 2,20 |

## 4 Results and discussion

Basic statistical parameters (mean and standard deviation), were estimated for parameters $\sqrt{A}$ and $L$ and are listed in Tab. 2. The minimal area of the inclusion on the fractured cross section was 0,016 mm$^2$ and the maximal area of inclusion measured was 3,66 mm$^2$. The inclusion which is the closest to the specimen outer surface lies 0,255 mm from it. The maximum distance from the specimen outer surface is 4,2 mm.

**Table 2** Basic statistics for variables $\sqrt{A}$ and $L$

| | $\sqrt{A}$ /mm | $L$ /mm |
|---|---|---|
| Minimum value | 0,126491 | 0,25 |
| Maximum value | 1,913113 | 4,20 |
| Mean, $\mu$ | 0,9037 | 2,0928 |
| Standard deviation, $\sigma$ | 0,487824 | 1,091577 |

The sample pairs of the variables $\sqrt{A}$ and $L$ that were used for estimating the PDF of the inclusion geometry parameters are displayed as a scatter plot in Fig. 3. In the figure also the line that limits the domain of $\sqrt{A}$ and $L$ parameters for the 9 mm specimen is drawn.

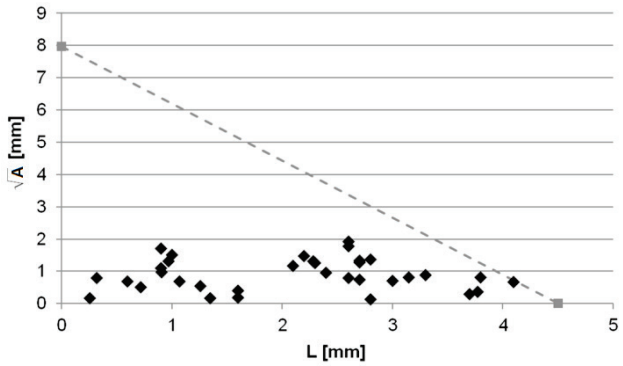


**Figure 3** Scatter plot displaying values for variables $\sqrt{A}$ and $L$

### 4.1 Estimating the mixed PDF for the inclusion parameters using the REBMIX algorithm

The multivariate PDF of $\sqrt{A}$ and $L$ parameters was modelled with a mixture PDF which was a product of univariate PDFs for $\sqrt{A}$ and $L$ variables. The mixture components were set-up by multiplying either univariate Gaussian functions (see Eq. (7)) or two-parametric Weibull functions (see Eq. (8)). In a case where Gaussian functions were used the REBMIX algorithm found one or five components to be the most suitable in the mixed PDF. Mixture parameters are given in Tab. 3 and both mixtures are plotted in Fig. 4. Considering the weights of components in five-component mixture we find that at least three out of five components occupy trivial weights (0,12; 0,17; 0,09), while only one component occupies a significant weight (0,41). In case where Weibull functions were used REBMIX algorithm again found one component to be the most suitable along with three component mixture. Weibull mixture parameters are given in Tab. 4 and both mixtures are plotted in Fig. 5. In three component mixture again one of the components has the weight of 0,4. The other two components have about the same weight of 0,3.

Observing Fig. 3, two separate clusters of points are clearly visible. In cases where the REBMIX identified only one component in the mixture the algorithm was clearly unable to detect the gap between the two clusters. In both cases where the REBMIX algorithm identified more than one component in the mixture the weights of components show that one component ($w \approx 0,4$) represents one cluster of points, while the sum of all other components ($w \approx 0,6$) the other cluster of points. Due to the inherent assumption that two parameters in each component are independent the algorithm is unable to account for skewness of cluster points. Therefore it compensates by identifying a larger number of smaller components in the mixture or a single large component.

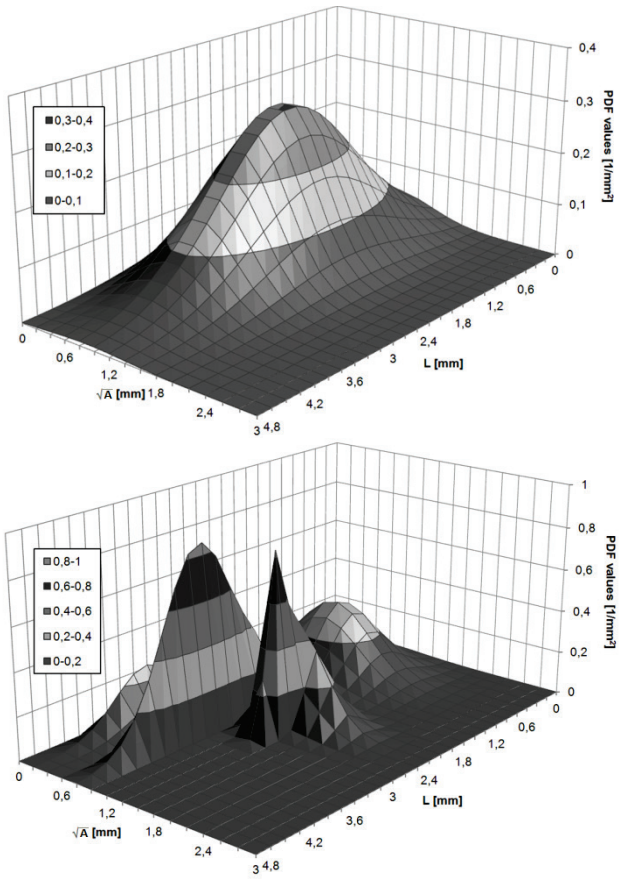


**Figure 4** Mixture of two-dimensional Gaussian functions, 1 component (above) and 5 components (below), REBMIX algorithm

**Table 3** Mixture parameters for Gaussian PDF, REBMIX algorithm

| $m$ | $j$ | $w$ | Variable / mm | $\mu$ / mm | $\sigma$ / mm |
|---|---|---|---|---|---|
| 1 | 1 | 1 | $\sqrt{A}$ | 0,904 | 0,481 |
| | | | $L$ | 2,090 | 1,080 |
| 5 | 1 | 0,12 | $\sqrt{A}$ | 1,320 | 0,049 |
| | | | $L$ | 2,780 | 0,144 |
| | 2 | 0,21 | $\sqrt{A}$ | 0,776 | 0,066 |
| | | | $L$ | 3,240 | 0,540 |
| | 3 | 0,17 | $\sqrt{A}$ | 1,390 | 0,351 |
| | | | $L$ | 2,390 | 0,182 |
| | 4 | 0,09 | $\sqrt{A}$ | 0,259 | 0,104 |
| | | | $L$ | 3,420 | 0,446 |
| | 5 | 0,41 | $\sqrt{A}$ | 0,765 | 0,477 |
| | | | $L$ | 0,961 | 0,394 |

Remarks: $m$ – number of component in mixture; $j$ – the number of components; $w$ – weight of component; $\mu$ – mean value; $\sigma$ – standard deviation

### 4.2 Estimating the mixed PDF for the inclusion parameters using the modified EM algorithm

One of fundamental differences between the EM and REBMIX algorithms is the assumption of the variable dependency. The EM algorithm, as opposed to the REBMIX algorithm, assumes that variables are statistically dependent within each component of the mixture. In this case the joint PDF of the parameters $\sqrt{A}$

was modelled either using the mixture of the multivariate Gaussian functions (see Eq. (4)) or the mixture of the multivariate *t*-functions (see Eq. (6)).

**Table 4** Mixture parameters for Weibull PDF, REBMIX algorithm

| $m$ | $j$ | $w$ | Variable / mm | $\beta$ / mm | $\eta$ / mm |
|---|---|---|---|---|---|
| 1 | 1 | 1 | $\sqrt{A}$ | 1,010 | 1,940 |
| | | | $L$ | 2,370 | 2,020 |
| 3 | 1 | 0,31 | $\sqrt{A}$ | 1,470 | 5,540 |
| | | | $L$ | 2,610 | 9,420 |
| | 2 | 0,29 | $\sqrt{A}$ | 0,705 | 2,580 |
| | | | $L$ | 3,350 | 6,760 |
| | 3 | 0,40 | $\sqrt{A}$ | 0,840 | 1,550 |
| | | | $L$ | 1,110 | 2,660 |

Remarks: $m$ – number of components in mixture; $j$ – the number of component; $w$ – weight of component; $\beta$ – shape parameter; $\eta$ – scale parameter

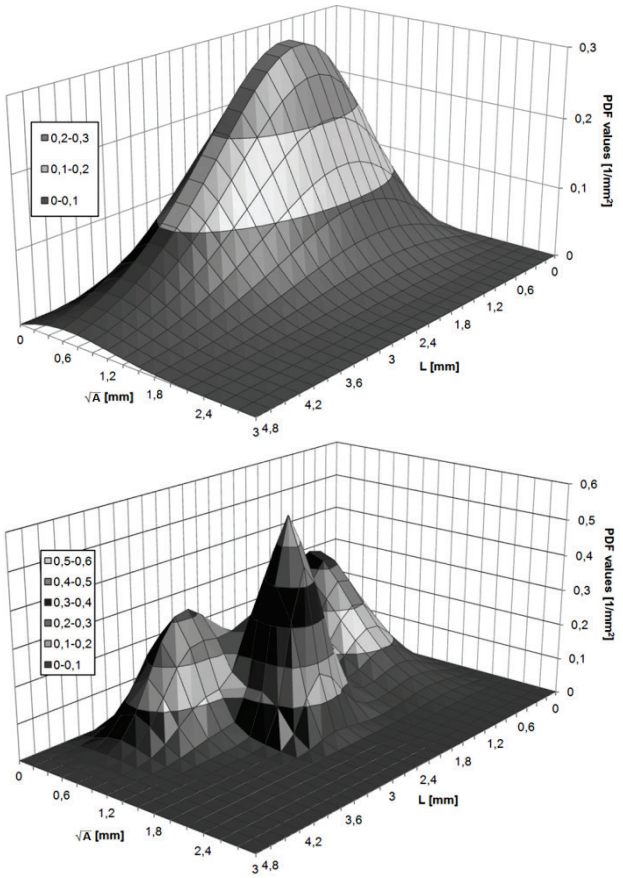


**Figure 5** Mixture of two-dimensional Weibull functions, 1 component (above) and 3 components (below), REBMIX algorithm

To identify the optimal PDF mixture also from the standpoint of the number of components in the mixture, the modified EM algorithm according to Fugueiredo and Jain [14] was applied. The initial number of the components in the mixtures was three, two and one component. The estimated mixtures of the two-dimensional Gaussian and *t*-functions are presented in Figs. 6 ÷ 11 for different number of components in the mixtures. The estimated parameters of the mixtures are presented in Tab. 5 for the mixtures of two-dimensional Gaussian functions and in Table 6 for the mixtures of two-dimensional *t*-functions.

Mixtures with one two-dimensional Gaussian function (Fig. 6, Tab. 4) and one two-dimensional *t*-function (Fig. 9, Tab. 5) are in good agreement with the results of the REBMIX algorithm. However, the corresponding covariance matrix (Tab. 5, Tab. 6) indicates a weak dependence of the parameters $\sqrt{A}$ and $L$.

Weights of components in mixture with two two-dimensional Gaussian functions are 0,6 and 0,4 (see Tab. 5). Both components are clearly visible in Fig. 7. We can see from Table 5 that the heavier component 1 shows strong dependence of variables $\sqrt{A}$ and $L$, while the dependence of the two parameters is less pronounced in component 2.

In the mixture with three two-dimensional Gaussian the weights are 0,55, 0,43 and 0,02 (see Tab. 5). The weight of the third component is very small. The value of the weight (0,02) implies that the component corresponds to a single sample point, making the component literally singular, if no restraints were applied in the modified EM to prevent singularity of covariance matrices.

**Table 5** Mixture parameters for normal PDF, modified EM algorithm

| $m$ | $j$ | $w$ | Covariance matrix $\Sigma$ / mm | Var. / mm | $\mu$ / mm | $\sigma$ / mm |
|---|---|---|---|---|---|---|
| 1 | 1 | 1 | $\begin{bmatrix} 0{,}23100 & 0{,}00613 \\ 0{,}00613 & 1{,}15649 \end{bmatrix}$ | $\sqrt{A}$ | 0,903 | 0,480 |
| | | | | $L$ | 2,092 | 1,075 |
| 2 | 1 | 0,6 | $\begin{bmatrix} 0{,}21127 & -0{,}15496 \\ -0{,}15496 & 0{,}34077 \end{bmatrix}$ | $\sqrt{A}$ | 1,001 | 0,459 |
| | | | | $L$ | 2,881 | 0,583 |
| | 2 | 0,4 | $\begin{bmatrix} 0{,}25861 & -0{,}03217 \\ -0{,}03217 & 0{,}15788 \end{bmatrix}$ | $\sqrt{A}$ | 0,763 | 0,475 |
| | | | | $L$ | 0,962 | 0,397 |
| 3 | 1 | 0,55 | $\begin{bmatrix} 0{,}12088 & -0{,}16196 \\ -0{,}16196 & 0{,}40211 \end{bmatrix}$ | $\sqrt{A}$ | 0,959 | 0,347 |
| | | | | $L$ | 2,910 | 0,634 |
| | 2 | 0,43 | $\begin{bmatrix} 0{,}22330 & -0{,}03237 \\ -0{,}03237 & 0{,}15829 \end{bmatrix}$ | $\sqrt{A}$ | 0,758 | 0,482 |
| | | | | $L$ | 0,962 | 0,397 |
| | 3 | 0,02 | $\begin{bmatrix} 0{,}66273 & -0{,}07682 \\ -0{,}07682 & 0{,}00089 \end{bmatrix}$ | $\sqrt{A}$ | 1,261 | 0,792 |
| | | | | $L$ | 2,667 | 0,094 |

Remarks: $m$ – number of components in mixture; $j$ – number of component; $w$ – weight of component; $\mu$ – mean value; $\sigma$ – standard deviation

**Table 6** Mixture parameters for t-function PDF, modified EM algorithm

| $m$ | $j$ | $w$ | Covariance matrix $\Sigma$ / mm | Var. / mm | $\mu$ / mm | $\sigma$ / mm |
|---|---|---|---|---|---|---|
| 1 | 1 | 1 | $\begin{bmatrix} 0{,}22945 & 0{,}00627 \\ 0{,}00627 & 1{,}15044 \end{bmatrix}$ | $\sqrt{A}$ | 0,904 | 0,479 |
| | | | | $L$ | 2,093 | 1,072 |
| 2 | 1 | 0,6 | $\begin{bmatrix} 0{,}20934 & -0{,}15431 \\ -0{,}15431 & 0{,}33929 \end{bmatrix}$ | $\sqrt{A}$ | 1,002 | 0,457 |
| | | | | $L$ | 2,878 | 0,582 |
| | 2 | 0,4 | $\begin{bmatrix} 0{,}22403 & -0{,}03251 \\ -0{,}03251 & 0{,}15671 \end{bmatrix}$ | $\sqrt{A}$ | 0,763 | 0,473 |
| | | | | $L$ | 0,963 | 0,395 |
| 3 | 1 | 0,08 | $\begin{bmatrix} 0{,}06654 & 0{,}00278 \\ 0{,}00278 & 0{,}00171 \end{bmatrix}$ | $\sqrt{A}$ | 1,318 | 0,257 |
| | | | | $L$ | 0,936 | 0,041 |
| | 2 | 0,68 | $\begin{bmatrix} 0{,}20837 & -0{,}15288 \\ -0{,}15288 & 0{,}33347 \end{bmatrix}$ | $\sqrt{A}$ | 1,002 | 0,456 |
| | | | | $L$ | 2,881 | 0,577 |
| | 3 | 0,24 | $\begin{bmatrix} 0{,}05429 & -04246 \\ -0{,}04246 & 0{,}24398 \end{bmatrix}$ | $\sqrt{A}$ | 0,462 | 0,233 |
| | | | | $L$ | 0,980 | 0,493 |

Remarks: $m$ – number of components in mixture; $j$ – number of component; $w$ – weight of component; $\mu$ – mean value; $\sigma$ – standard deviation

The two component mixture of the two-dimensional *t*-functions has the same component weights as the mixture with two normally distributed components: 0,6

and 0,4. Similar as well, one component implies statistical dependence of variables, and the other one does not (see Tab. 6).

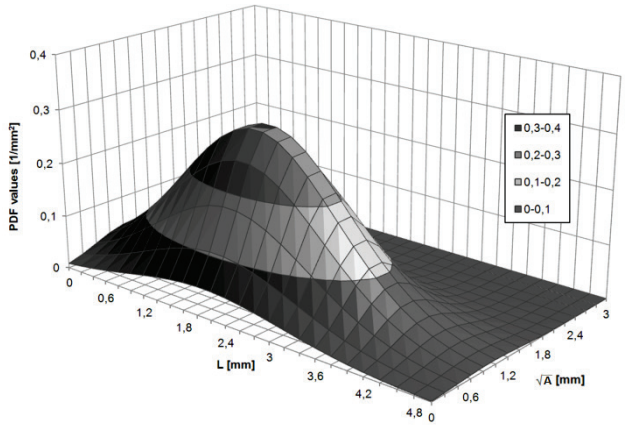


**Figure 6** Mixture of two-dimensional Gaussian functions, 1 component, modified EM algorithm

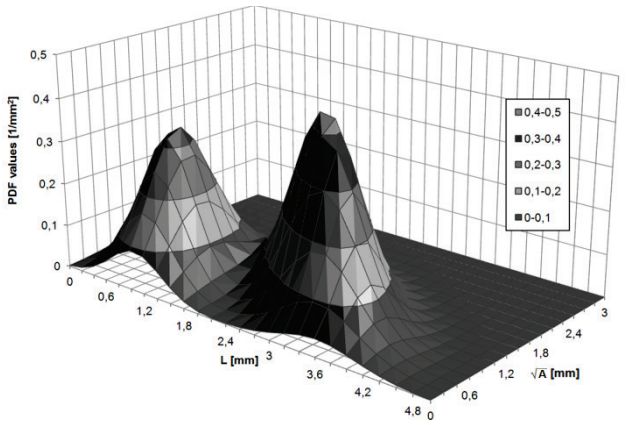


**Figure 7** Mixture of two-dimensional Gaussian functions, 2 components, modified EM algorithm

Discrepancy between mixtures of the two-dimensional Gaussian and *t*-function occurs if the mixture is composed of three components (see Tab. 5 and Tab. 6). This can be observed in Figs. 8 and 11. It can be concluded that the two PDFs are not alike. In the mixture of three *t*-functions the component with the smallest weight (component 1) is not singular and represents the part of the population that would otherwise belong to the component 3, if the number of components in the mixture would be reduced to two.

The results are comparable with those given by the REBMIX algorithm. However in contrast to REBMIX, the EM algorithm accounts for statistical dependence of variables within each component. Therefore, one component is sufficient to describe the second set of sample points ($w \approx 0{,}6$) instead of four (Gaussian PDF) or two (Weibull PDF) needed by the REBMIX algorithm.

Whether two-dimensional Gaussian or *t*-functions are used in the mixture, the third component always represents a relatively small part of the original sample. Therefore it was presumed that it is only a part of one of the two larger sets of sample points. On the basis of this finding we suggest that the actual PDF of the geometrical parameters of inclusions is modelled with a two component mixture. As mentioned earlier in the text the results for the two component mixtures of two-dimensional Gaussian and *t*-functions coincide, thus applying the mixture of Gaussian functions is a more justified choice, since it has less parameters than the mixture of *t*-functions (no degrees of freedom for each component). This is an advantage considering the small number of sample points available for estimating the parameters of the mixed PDF.

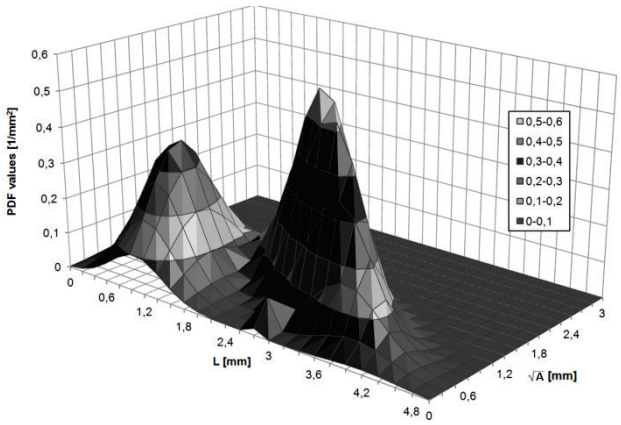


**Figure 8** Mixture of two-dimensional Gaussian functions, 3 components, modified EM algorithm

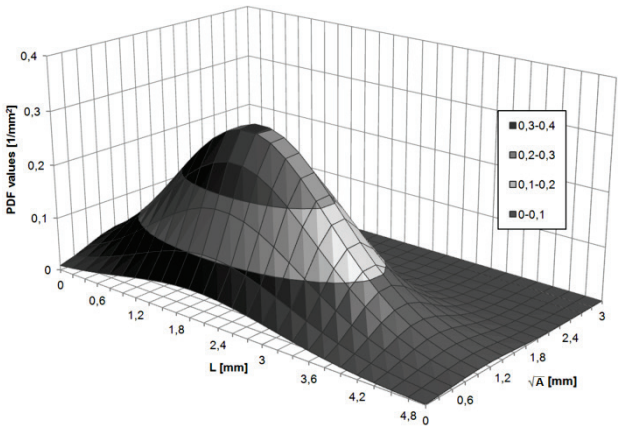


**Figure 9** Mixture of two-dimensional *t*-functions, 1 component, modified EM algorithm

The REBMIX inability to account for statistical dependence of variables within each component, when only a small number of samples are available, makes the modified EM algorithm more suitable to use in the case presented in this paper.

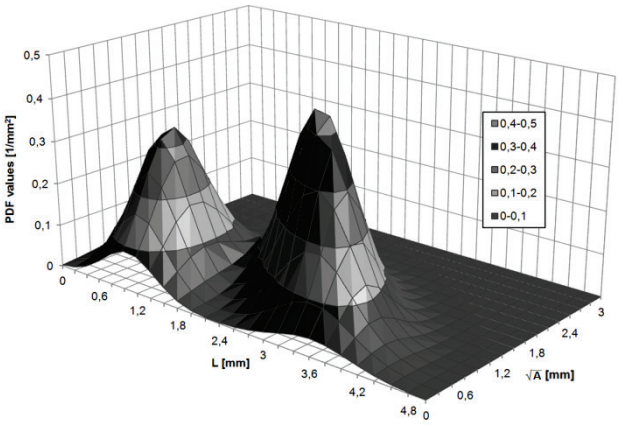


**Figure 10** Mixture of two-dimensional *t*-functions, 2 components, modified EM algorithm

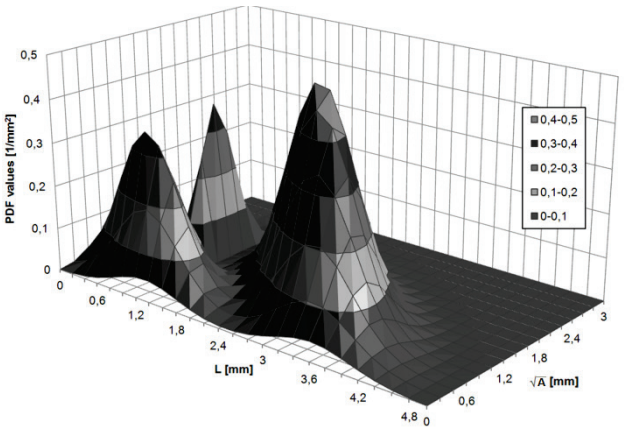


**Figure 11** Mixture of two-dimensional *t*-functions, 3 components, modified EM algorithm

### 4.3 Assessing the variability of the statistical estimates

For mixture parameter estimation to be reliable sample size is desired to be 10-times the number of estimated parameters [17]. In the presented case it turns out that a mixture of two multivariate Gaussian functions is the optimal choice for estimating the two-dimensional PDF of the vector variable ($\sqrt{A}$, $L$) based on 34 available samples. Each two-dimensional Gaussian function in the mixture contains 5 different parameters: means of $\sqrt{A}$ and $L$, two variances and one covariance of variables $\sqrt{A}$ and $L$. Therefore, the mixture with two such components is defined with 12 (5 + 5 + 2 weights) parameters, which must be estimated. Therefore the sample size is 3-times the number of estimated parameters. Due to a unique nature of investigated phenomenon (random presence of inclusion on critical cross section) inclusions could not be intentionally introduced into the specimens and sample size could not be increased. Therefore the variability of the 12 mixture parameter estimations was assessed using bootstrap analysis [27].

From original sample with 34 data points, 10 bootstrap sets were created using random sampling with replacement. Each bootstrap set contained 34 data points and the sets differed from one another due to replacement nature of the sampling process. For each bootstrap set the parameters of the mixture of two-dimensional Gaussian functions were estimated using the modified EM algorithm. Variability of mixture parameters was assessed by calculating their mean, standard deviation and variation coefficient for all 12 parameters. Results for component 1 and 2 are given in Tabs. 7 and 8, respectively.

They show that mean values vary the least, as opposed to variability of weights, variances and covariance. Greater variability of estimated parameters indicates lesser accuracy of parameter estimation. The variation coefficient value being quite high is expected in cases where the number of parameters being estimated is only 3-times smaller than the number of sample points available for their estimation. Should the number of sample points be increased, the values of variation coefficients would decrease in all instances. Variation coefficient of covariance is negative due to the value of covariance being negative and has a large value due to the small value of covariance itself.

**Table 7** Bootstrap analysis results, component 1

| | $w_{BS}$ | $\mu_{BS}$ $(L)$ | $\mu_{BS}$ $(\sqrt{A})$ | $\sigma^2_{BS}$ $(L)$ | $\sigma^2_{BS}$ $(\sqrt{A})$ | $COV_{BS}$ |
|---|---|---|---|---|---|---|
| st. dev. | 25,965 | 0,4082 | 0,2445 | 0,3910 | 0,0602 | 0,071 |
| mean | 51,610 | 1,0648 | 0,9194 | 0,3357 | 0,1197 | −0,002 |
| *VC* | 0,531 | 0,3833 | 0,2659 | 1,1645 | 0,5036 | −37,64 |

Remarks: $\mu_{BS}$ – mean value; $\sigma^2_{BS}$ – variance; $w_{BS}$ – weight of component; $COV_{BS}$ – covariance of variables; *VC* – variation coefficient

**Table 8** Bootstrap analysis results, component 2

| | $w_{BS}$ | $\mu_{BS}$ $(L)$ | $\mu_{BS}$ $(\sqrt{A})$ | $\sigma^2_{BS}$ $(L)$ | $\sigma^2_{BS}$ $(\sqrt{A})$ | $COV_{BS}$ |
|---|---|---|---|---|---|---|
| st. dev. | 25,965 | 0,4903 | 0,3225 | 0,3449 | 0,1023 | 0,115 |
| mean | 48,390 | 2,7056 | 0,9722 | 0,3513 | 0,1796 | −0,028 |
| *VC* | 0,536 | 0,1812 | 0,3317 | 0,9816 | 0,5698 | −4,081 |

Remarks: $\mu_{BS}$ – mean value; $\sigma^2_{BS}$ – variance; $w_{BS}$ – weight of component; $COV_{BS}$ – covariance of variables; *VC* – variation coefficient

## 5 Conclusion

Inclusions cannot be avoided in cast parts and components. Different authors suggest different models to take the most influential geometry parameters of inclusions into consideration with relation to fatigue life. We decided to follow the approach of Murakami and Endo [16] for describing the geometrical characteristics of inclusions. From the available experimental data on fatigue test of aluminium alloy Alsi9Cu3 we were able to isolate 34 specimens with obvious oxide inclusions in the fractured cross section. These samples were then used for statistical analysis of the inclusions' geometric parameters.

For modelling the statistical characteristics of the inclusions various mixtures of multivariate PDFs were used. The unknown parameters of the mixed PDFs were estimated with two different algorithms: the REBMIX algorithm and the modified EM algorithm according to Figueiredo and Jain [17]. On the basis of the estimated PDFs of the inclusions' geometrical parameters $\sqrt{A}$ and $L$ the modified EM algorithm was selected as the more suitable, since the REBMIX algorithm was unable to differentiate between two rather obvious clusters of sample points. The optimal number of components in the mixture was found to be two and the most appropriate basic function for the mixture was multivariate Gaussian function. It described the statistical distribution of the inclusions' geometrical parameters as good as the mixture of two multivariate *t*-functions with less parameters to be estimated, which is an advantage if the number of sample points is relatively small. A small number of sample points along with a high number of unknown parameters is also the very reason why values of variation coefficient calculated during bootstrap analysis are relatively high. Should the number of sample points be increased, the values of variation coefficients would decrease in all instances.

The research presents a fundamental work into linking the stochastic nature of inclusions appearing in the critical cross section, their geometrical features and fatigue life of the specimens.

**Acknowledgements**

The work presented in this paper is financially supported by the Ministry of Higher Education, Science and Technology of Slovenia.

## 6 References

[1] Mayer, H.; Papakyriacou, M.; Zettl, B.; Stanz- Tschegg, S. E. Influence of porosity on the fatigue limit of die cast magnesium and aluminium alloys. //International Journal of Fatigue. 25, (2003), pp. 245-56.

[2] Knott, J. F.; Bowen, P.; Luo, J.; Jiang, H.; Sun, H. L. The structural integrity of cast aluminium automotive components subjected to fatigue loads. //Material Science Forum. 331-337, (2000), pp. 1401-12.

[3] Evans, W. J.; Lu, Z. J.; Spittle, J. A.; Devlukia, J. Fatigue crack development from defects in cast aluminium alloy. // High cycle fatigue of structural materials/ editors Soboyejo, W. O. and Srivastan, T. The Minerals Metals and Materials Society, 1997. pp. 445-57.

[4] Debayeh, A. A.; Xu, R. X.; Du, B. P.; Topper, T. H. Fatigue of cast aluminium alloys under constant and variable- amplitude loading. // International Journal of Fatigue. 19, 2(1996), pp. 95-104.

[5] Linder, J.; Axelsson, M.; Nilsson. H. The influence of porosity on the fatigue life for sand and permanent mould cast aluminium. // International Journal of Fatigue. 28, (2006), pp. 1752-58.

[6] Wang, Q. G. Fatigue behavior of A356/357 aluminum cast alloys. Part II – Effect of microstructural constituents. // Journal of light metals. 1, 1(2001), pp. 85-97.

[7] Murakami, Y.; Beretta, S. Small defects and inhomogeneities in fatigue strength: experiments, models and statistical implications. // Extremes. 2, (1999), pp. 123-147.

[8] Juvonen, P. Effects of non-metallic inclusions on fatigue properties of calcium treated steels. PhD Thesis; 2004.

[9] Uhrs, L. O. Through-hardening steels for ball bearings-effect of inclusions on endurance. // Iron and Steel Institute, special report. 77, (1963), pp. 104-109.

[10] Duckworth, W. E.; Ineson, E. The effects of externally introduced alumina particles on the fatigue life of En24 steel, clean steel. // Iron and Steel Institute, special report. 77, (1963), pp. 87-103.

[11] Ramsey, P. W.; Kedzie, D. P. Prot fatigue study of aircraft steel in the ultra-high strength rang. // AIME Journal of metallurgy. 9, (1957), pp. 401-406.

[12] De Kazinczy, F. Effects of small defects on the fatigue Properties of medium- strength cast steel. // The journal of iron and steel institute. 208, (1970), pp. 851-855.

[13] Murakami, Y.; Endo, M. Effects of defects, inclusions and inhomogenities on fatigue strength. //International journal of fatigue, 16, (1994), pp. 163-182.

[14] Figueiredo, M.; Jain, A. Unsupervised learning of finite mixture models. // IEEE transactions on pattern analysis and machine intelligence. 24, 3(2002), pp. 381-396.

[15] Nagode, M.; Fajdiga, M. The REBMIX algorithm for the multivariate finite mixture estimation. // Communications in Statistics - Theory and Methods. 40, 11(2011), pp. 2022-2034.

[16] Richardson, S.; Green, P. J. On Bayesian analysis of mixtures with an unknown number of components. // Journal of Royal Statistical Society. 59, 4(1997), pp. 731-792.

[17] Titterington, D. M.; Smith, A. F. M.; Makov, U. E. Statistical analysis of finite mixture distribution. Wiley, University of California, 1985.

[18] Bishop, C. M. Neural Networks and their applications. // Review of Scientific Instruments. 65, 6(1994), pp. 1803-1832.

[19] Nagode, M.; Fajdiga, M. A general multi-nodal probability density function suitable for the rainflow ranges of stationary random processes. // International Journal of Fatigue. 20, 3(1998), pp. 211-223.

[20] Nagode, M.; Fajdiga, M. An alternative perspective on the mixture estimation problem. // Reliability Engineering and System Safety. 84, 3(2004), pp. 241-251.

[21] Nagode, M.; Fajdiga, M. The REBMIX algorithm for the multivariate finite mixture estimation. // Communications in Statistics - Theory and Methods. 40, 11(2011), pp. 2022-2034.

[22] Nagode, M.; Fajdiga, M. The REBMIX algorithm for the univariate finite mixture estimation. // Communications in Statistics - Theory and Methods. 40, 5(2011), pp. 876-892.

[23] Grabec, I.; Sachse, W. Synergetics of Measurement, Prediction and Control. Springer-Verlag, Berlin Heidelberg, 1997.

[24] Klemenc, J.; Fajdiga, M. Improved modeling of the loading spectra using a mixture model approach. // International Journal of Fatigue. 30, 7(2008), pp. 1298-1313.

[25] ASTM. E 606-92: standard practice for strain-controlled fatigue testing. 1996

[26] Brent Dove, S. ImageTool Version 3.0. UTHSCSA (1996 – 2002) URL: http://compdent.uthscsa.edu/dig/itdesc.html (22.02.2002)

[27] Efron, B. The Jackknife, the Bootstrap and Other Resampling Plans. Society for Industrial and Applied Mathematics, Pennsylavnia, 1982.

**Authors' addresses**

***Ana Bižal, BSc ME***
University of Ljubljana, Faculty of Mechanical Engineering
Askerceva 6
SI-1000 Ljubljana, Slovenia
E-mail: ana.bizal@fs.uni-lj.si

***Jernej Klemenc, associate professor***
University of Ljubljana, Faculty of Mechanical Engineering
Askerceva 6
SI-1000 Ljubljana, Slovenia
E-mail: jernej.klemenc@fs.uni-lj.si

***Matija Fajdiga, full time professor***
University of Ljubljana, Faculty of Mechanical Engineering
Askerceva 6
SI-1000 Ljubljana, Slovenia
E-mail: matija.fajdiga@fs.uni-lj.si