



**HAL**  
open science

## Application of Lambda Distributions and Bootstrap analysis to the prediction of fatigue lifetime and confidence intervals

Maxence Bigerelle, Denis Najjar, Benjamin Fournier, Nicolas Rupin, Alain Iost

► **To cite this version:**

Maxence Bigerelle, Denis Najjar, Benjamin Fournier, Nicolas Rupin, Alain Iost. Application of Lambda Distributions and Bootstrap analysis to the prediction of fatigue lifetime and confidence intervals. *International Journal of Fatigue*, 2006, 28 (3), pp.223-236. 10.1016/j.ijfatigue.2005.06.033 . hal-01318580

**HAL Id: hal-01318580**

**<https://hal.science/hal-01318580>**

Submitted on 18 Dec 2017

**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.

# Application of Lambda Distributions and Bootstrap analysis to the prediction of fatigue lifetime and confidence intervals

M. Bigerelle<sup>a,b</sup>, D. Najjar<sup>b</sup>, B. Fournier<sup>b</sup>, N. Rupin<sup>b</sup>, A. Iost<sup>b,\*</sup>

<sup>a</sup>Roberval, FRE 2833, UTC/CNRS, Centre de Recherches de Royallieu, BP20529, 60205 Compiègne, France  
<sup>b</sup>Laboratoire de Métallurgie Physique et Génie des Matériaux, CNRS UMR 8517, Equipe Surfaces et Interfaces, ENSAM 8 Boulevard Louis XIV, 59046 Lille cedex, France

## Abstract

Dealing with fatigue lifetime prediction, this paper aims to report on a new statistical method combining the Lambda Distributions and the Bootstrap technique. This method is first applied for determining the Probability Density Function (PDF) of the  $C$  and  $n$  coefficients in the Paris relationship of a fatigue crack propagation curve. Then, introducing the initial crack's length distribution, the fatigue lifetime prediction is obtained and discussed considering various standard deviations of the initial crack's length. It is shown that the scattering of the initial crack's length needs to be taken into account in predicting lifetime, and that the stochastic nature of the crack's propagation is not self-sufficient to explain completely the experimental asymmetry of the PDF lifetime. This paper shows that the Lambda Distributions are a powerful tool for modelling the PDF lifetime, compared with traditional Gaussian or lognormal PDF.

*Keywords:* Lifetime prediction; Statistical analysis; Numerical simulation; Fatigue crack growth; Paris relationship; Bootstrap technique; Lambda Distribution

## 1. Introduction

The present work was motivated by a review of articles dealing with fatigue lifetime modelling. Indeed, on the one hand, the lifetime data under study are modelled by a two-parameter Weibull distribution in [1–8]. The authors justify this choice either by the weakest link assumption or because this is the most commonly used distribution in fatigue lifetime modelling. Caron and Ehlacher [9] use a three-parameter Weibull Distribution and Nagode and Fajdiga [10,11] choose multi-modal Weibull Distributions. On the other hand, the lifetime data under study are modelled by a Lognormal Distribution in [12–15], and even by a Gaussian Distribution in Liao et al. [15].

None of the commonly used distributions (Gaussian, Lognormal or Weibull) is able to model correctly all the shapes encountered in fatigue lifetime experimental data meaning that there is no universally admitted distribution

family for the modelling of fatigue lifetime. Moreover, the choice of the distribution used is often either due to the modelling background of the authors or to the ability of the considered distribution to fit correctly with the shape of the data under study. Furthermore, the distributions used are scarcely compared to others and are not always validated by statistical tests of adequacy. Therefore, the consequent variety in fatigue lifetime modelling may lead to discrepancies and comparison difficulties between the results thus obtained.

The present work aims to present a new alternative statistical method that avoids the previously cited drawbacks related to commonly used distributions. Based on a four parameter family of distributions called the Lambda Distributions, this study derives from the preliminary results obtained by the authors [16]. This first article [16] was based on the results of the fatigue crack growth experiment on an aluminium alloy obtained by Krausz et al. [17] and shown in Fig. 1.

This figure illustrates the fact that the behaviour of the fatigue crack propagation of engineering alloys can be generally divided into three stages. In stage II, stable fatigue crack growth conditions prevail and the fatigue crack growth rate (FCGR) is given by the well known

\* Corresponding author. Tel.: +33 3 2062 2233; fax: +33 3 2062 2957.  
E-mail address: iost@lille.ensam.fr (A. Iost).

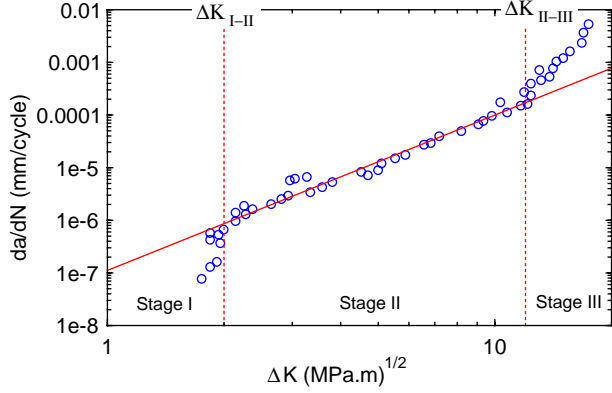


Fig. 1. Result of the fatigue crack growth on an aluminium alloy obtained by [17]. Dispersion on some experimental points can be observed. The Paris equation is given by  $\ln da/dN = -16.03 + 2,954 \ln(k)$ .

Paris–Erdogan relationship (Paris and Erdogan [18])

$$da/dN = C\Delta K^n \quad (1)$$

where  $C$  and  $n$  are two parameters depending on the material and on the experimental conditions.

In Bigerelle and Iost [16], the Paris coefficients and the interval limits  $[\Delta K_{I-II}, \Delta K_{II-III}]$  of stage II were determined combining the Bootstrap analysis and the linear regression method in the case of the data presented in Fig. 1. The values of  $\Delta K_{I-II}$  and  $\Delta K_{II-III}$  were statistically determined by finding the  $\{i, j, C_I, n_I, C_{II}, n_{II}, C_{III}, n_{III}\}$  values which minimize the residuals  $R_{i,j}$  defined by

$$\begin{aligned} R_{i,j} = & \sum_{k=1}^{i-1} \left( C_I \Delta K_k^{n_I} - \left( \frac{da}{dN} \right)_k \right)^2 \\ & + \sum_{k=i}^j \left( C_{II} \Delta K_k^{n_{II}} - \left( \frac{da}{dN} \right)_k \right)^2 \\ & + \sum_{k=j+1}^n \left( C_{III} \Delta K_k^{n_{III}} - \left( \frac{da}{dN} \right)_k \right)^2 \end{aligned} \quad (2)$$

where  $(\Delta K_k, (da/dN)_k)$  are  $n$  experimental data with an ascending  $\Delta K_k$  and  $(i,j)$  are the indices of the points bounding the stage II of the fatigue curve. Applying this method to the experimental results presented in Fig. 1, the authors found  $\Delta K_{I-II} = 2.3 \text{ MPa}\sqrt{m}$  and  $\Delta K_{II-III} = 10.9 \text{ MPa}\sqrt{m}$ .

The Lambda Distributions are described in the present work to propose a new, robust and powerful method to obtain an accurate and flexible analytical expression of the fatigue lifetime prediction. This method enables the calculation of any confidence interval needed and allows a better understanding of the influence of the considered parameters. Thus, instead of considering only a single-point value for each parameter (e.g. the mean initial crack's length), the presented method considers the whole

distribution of the studied parameters, through their associated Lambda Distribution.

Starting from the preliminary results obtained in Bigerelle and Iost [16], the Lambda Distributions are combined with the Bootstrap technique to determine firstly the Probability Density Function (PDF) of the Paris coefficients  $C$  and  $n$ . Then the Lambda Distributions are used in this article to model various shapes of fatigue lifetime data and turn out to be, in most cases, more accurate than the usual distributions used (i.e. Gaussian, Lognormal and Weibull) considering global goodness-of-fit criteria. Moreover, the adequacy of the Lambda Distribution modelling was checked and compared to those of other usual PDF around the distributions' tails (which are of particular interest in fatigue lifetime prediction) through the calculation of some extreme percentiles. Finally, to investigate the consistency of the distribution with the physics of the fatigue process, the corresponding hazard functions were plotted and compared to those of other usual distributions.

## 2. Bootstrap technique and Paris coefficients

Roughly speaking, the computer-based Bootstrap method (CBBM) [16,19,20] allows the replacement of statistical inference assumptions (therefore limiting the risk of asserting wrong conclusions) by intensive calculations while making the most of the power of modern computers.

The main principle of the Bootstrap method consists of generating a high number  $B$  of simulated Bootstrap samples from the original data points. The original dataset consists in either experimental or simulated points. A Bootstrap dataset of size  $N$ , noted  $(t_1^*, t_2^*, \dots, t_N^*)$  is a collection of  $N$  values simply obtained by randomly sampling with replacement from the original data points  $(t_1, t_2, \dots, t_N)$ , each of them with a probability of  $1/N$ . The Bootstrap dataset is, consequently, composed of elements from the original data points; some appearing zero time, some appearing one, some appearing twice, etc. Compared to the usual Monte Carlo and Latin Hypercube sampling methods, which both sample from an assumed PDF, the Bootstrap method presents the main advantage to avoid choosing a priori a PDF for the set of data points. There is no need to first identify an analytical PDF before sampling, which avoids the drawbacks related to the variability of the estimation of the parameters.

In this study, the Bootstrap method is applied to the linear regression model

$$\forall i \in [1; p] \quad y_i = x_i \beta + \varepsilon_i \quad (3)$$

where  $p$  is the number of experimental measurements,  $y_i$  is the  $i$ th element of  $Y = \ln da/dN$ ,  $x_i$  is  $i$ th element of  $X = (\ln \Delta K, 1)$ ,  $\beta = (n, \ln C)$  and  $\varepsilon_i$  is a random noise.

The application of the Bootstrap analysis to the experimental results presented in Fig. 1 enables us to obtain

a large set of possible pair coefficients ( $n$ ,  $\ln C$ ). The histograms related to these parameters are presented in Fig. 2.

### 3. The Lambda Distributions

#### 3.1. Definition

In this paragraph, the definition and some properties of this PDF family are briefly presented without giving any justification. For more details, the reader should refer to Karian [21]. The Generalized Lambda Distribution family with parameters  $(\lambda_1, \lambda_2, \lambda_3, \lambda_4)$ , noted  $GLD(\lambda_1, \lambda_2, \lambda_3, \lambda_4)$  is defined firstly by its percentile function

$$Q(y) = Q(y; \lambda_1, \lambda_2, \lambda_3, \lambda_4) = \lambda_1 + \frac{y^{\lambda_3} - (1-y)^{\lambda_4}}{\lambda_2} \quad (4)$$

with  $y \in [0, 1]$ ,  $(\lambda_1, \lambda_2)$  are, respectively, the position and scale parameters and  $(\lambda_3, \lambda_4)$  are, respectively, related to the skewness (asymmetry) and the kurtosis (flatness) of the  $GLD(\lambda_1, \lambda_2, \lambda_3, \lambda_4)$ . From the percentile function, the following PDF can be defined:

$$f(x) = f(Q(y)) = \frac{\lambda_2}{\lambda_3 y^{\lambda_3-1} + \lambda_4 (1-y)^{\lambda_4-1}} \quad (5)$$

The fact that this family of distributions is defined by its percentile function is fundamental and very useful as far as Monte Carlo simulations are concerned.

The Lambda Distributions present two main advantages. On the one hand, as they are defined by four parameters, they can approach accurately any kind of unimodal distribution; i.e. they can have various shapes. On the other hand, their percentile based definition enables easier and faster Monte Carlo simulations. Nevertheless, it must be outlined that the definition given above is restricted to domains where Eq. (5) corresponds to a valid PDF [21], imposing a careful determination of parameters  $(\lambda_1, \lambda_2, \lambda_3, \lambda_4)$ . Moreover, while the four parameters definition of the GLD enables a large flexibility for modelling a wide

variety of data sets, it must be stressed that estimating four parameters usually leads to bigger variability than estimating fewer parameters. This is particularly true for small sample sizes, meaning that for some practical cases where the amount of data is rather reduced, the choice of the method of determination of the GLD parameters must be carefully done. As it is a very usual method in distribution fitting, even for readers less familiar with statistics, only the method of moments will be presented in this paper.

#### 3.2. The method of moments

This part aims to describe one of the methods available to determine the parameters  $(\lambda_1, \lambda_2, \lambda_3, \lambda_4)$ , related to a given histogram of experimental or simulated values that characterize the population we wish to model. The method is based on the identification of the data moments with those of the  $GLD(\lambda_1, \lambda_2, \lambda_3, \lambda_4)$ . The first four moments of our experimental distribution are usually, as in [21], defined by:

$$\hat{\alpha}_1 = \bar{x} = \sum_{i=1}^n \frac{x_i}{n} \quad (6)$$

$$\hat{\alpha}_2 = \hat{\sigma}^2 = \sum_{i=1}^n \frac{(x_i - \bar{x})^2}{n} \quad (7)$$

$$\hat{\alpha}_3 = \sum_{i=1}^n \frac{(x_i - \bar{x})^3}{n \hat{\sigma}^3} \quad (8)$$

$$\hat{\alpha}_4 = \sum_{i=1}^n \frac{(x_i - \bar{x})^4}{n \hat{\sigma}^4} \quad (9)$$

In addition to the domains of validity of Eq. (5), the use of the method of moments imposes that the first four moments of the  $GLD(\lambda_1, \lambda_2, \lambda_3, \lambda_4)$  are defined, which leads to the condition  $\lambda_3 > -1/4$  and  $\lambda_4 > -1/4$ . Besides the first four moments of the  $GLD(\lambda_1, \lambda_2, \lambda_3, \lambda_4)$  are defined by

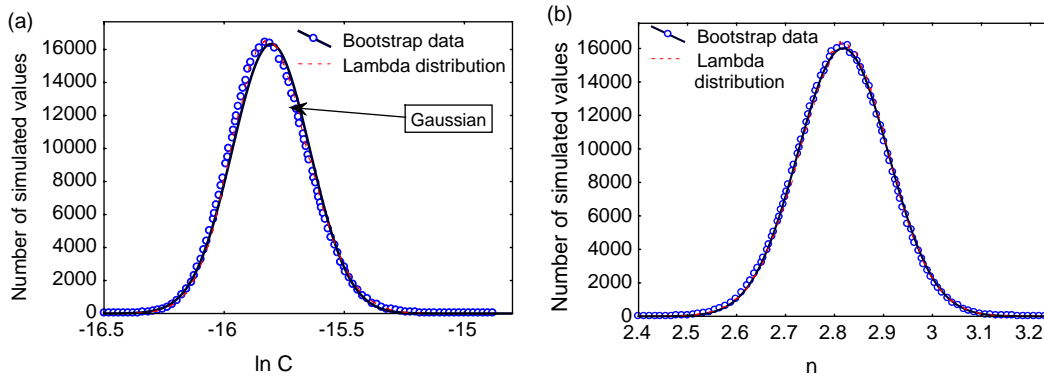


Fig. 2. Probability density function obtained by the Bootstrap technique on the values of  $\ln C$  (a) and  $n$  (b). These PDF are calculated from 100,000 Bootstrap iterations and fitted either by a Gaussian PDF or by a Lambda Distribution.

the following equations

$$\alpha_1 = \mu = E(X) = \lambda_1 + \frac{A}{\lambda_2} \quad (10)$$

$$\alpha_2 = \sigma^2 = E[(x-\mu)^2] = \frac{B-A^2}{\lambda_2^2} \quad (11)$$

$$\alpha_3 = \frac{E[(x-\mu)^3]}{\sigma^3} = \frac{C-3AB+2A^2}{\lambda_2^3\sigma^3} \quad (12)$$

$$\alpha_4 = \frac{E[(x-\mu)^4]}{\sigma^4} = \frac{D-4AC+6A^2B-3A^4}{\lambda_2^4\sigma^4} \quad (13)$$

with

$$A = \frac{1}{1+\lambda_3} - \frac{1}{1+\lambda_4} \quad (14)$$

$$B = \frac{1}{1+2\lambda_3} + \frac{1}{1+2\lambda_4} - 2\beta(1+\lambda_3, 1+\lambda_4) \quad (15)$$

$$C = \frac{1}{1+3\lambda_3} - \frac{1}{1+3\lambda_4} - 3\beta(1+2\lambda_3, 1+\lambda_4) + 3\beta(1+\lambda_3, 1+2\lambda_4) \quad (16)$$

$$D = \frac{1}{1+4\lambda_3} + \frac{1}{1+4\lambda_4} - 4\beta(1+3\lambda_3, 1+\lambda_4) + 6\beta(1+2\lambda_3, 1+2\lambda_4) - 3\beta(1+\lambda_3, 1+3\lambda_4) \quad (17)$$

where  $\beta$  is the Riemann Beta function given by:

$$\beta(a, b) = \int_0^1 x^{a-1}(1-x)^{b-1} dx \quad (18)$$

Knowing that an algorithm dedicated to the identification of the four parameters would be complex and time consuming, only the two parameters  $(\lambda_3, \lambda_4)$  are first considered, even though the four moments depend on both  $(\lambda_3, \lambda_4)$  and  $(\lambda_1, \lambda_2)$ . This identification is sufficient because, once a set of the  $(\lambda_3, \lambda_4)$  parameters is obtained, the

GLD $(\lambda_1, \lambda_2, \lambda_3, \lambda_4)$  is defined for any  $(\alpha_1, \alpha_2)$  [21]. Therefore, the optimization process can be restricted to the last two parameters  $(\lambda_3, \lambda_4)$  and the problem finally consists in minimizing the following surface:

$$S(\lambda_3, \lambda_4) = (\alpha_3 - \hat{\alpha}_3)^2 + (\alpha_4 - \hat{\alpha}_4)^2 \quad (19)$$

When the values of  $(\lambda_3, \lambda_4)$  are found so that GLD $(0, 1, \lambda_3, \lambda_4)$  has its third and fourth moments  $(\alpha_3, \alpha_4)$  defined, then  $(\lambda_1, \lambda_2)$  can be found solving these two equations

$$\hat{\mu} = \lambda_1 + \frac{A}{\lambda_2}, \quad \hat{\sigma}^2 = \frac{B-A^2}{\lambda_2^2}, \quad (20)$$

considering that  $\alpha_1 = \hat{\mu}$  and  $\alpha_2 = \hat{\sigma}^2$ , where  $\hat{\mu}$  is the estimated mean and  $\hat{\sigma}$  the estimated standard deviation.

Even if Eq. (19) apparently seems easy to minimize, the definition of the four moments of the GLD $(\lambda_1, \lambda_2, \lambda_3, \lambda_4)$  leads to heavy calculi and various problems.

### 3.3. The LambdaFinder software

One of the major difficulties encountered in the minimization of the surface defined by (19) is due to its highly non-linear shape. Indeed, this surface has several local minima and a usual Gauss-like algorithm of gradient may lead to the determination of wrong values of  $(\lambda_1, \lambda_2, \lambda_3, \lambda_4)$  that will correspond to a non-relevant GLD. Thus, the problem consists in finding an accurate point to initiate the gradient algorithm. With that aim in view, we resorted to a computer program called LambdaFinder designed to obtain a three dimensional visualization of the surface, in order to detect the potential different minima. Indeed, such a visualization enables us to avoid initiating the steepest descent algorithm near a local minimum that would not lead to a relevant density probability function. Fig. 3 shows one of these surfaces and highlights its non-linear shape.

The LambdaFinder software is programmed in C++ language and uses the OpenGL library to perform the 3D real time visualization of the surface defined in Eq. (19). It was designed to run easily on a modern personal computer and to be user-friendly. The flow chart of the LambdaFinder

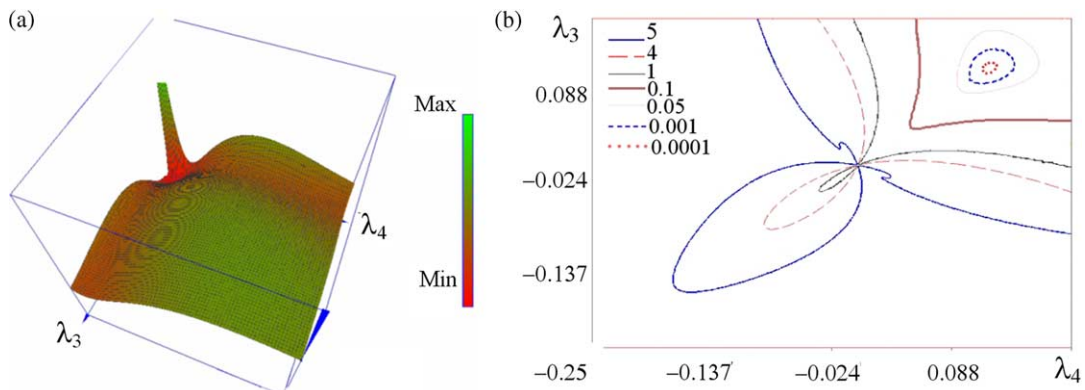


Fig. 3. 3D visualization of the surface  $s(\lambda_3, \lambda_4)$  calculated for the histogram of the  $n$  values (a) and its 2D projection (b).

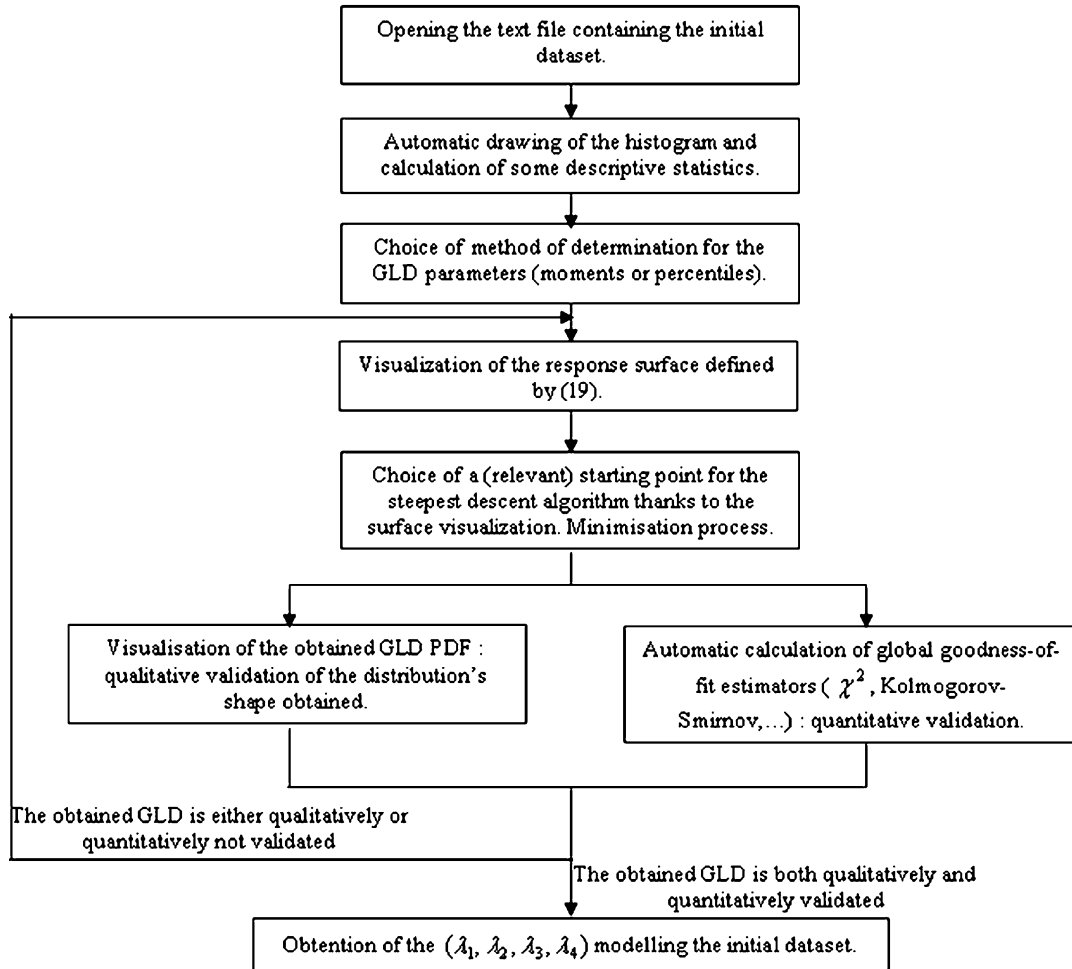


Fig. 4. Flow chart of the LambdaFinder software.

algorithm is represented on Fig. 4. After entering the initial data, their histogram and the corresponding surface (19) are plotted. The  $(\lambda_3, \lambda_4)$  space is discretized using a sizeable  $N \times N$  square grid and a lot of visualization features enable the user to zoom, rotate and move all over the surface, in order to obtain a correct idea of the global surface shape and to locate a relevant initial point.

After finding the initial point with much accuracy thanks to the visualization of the plotted surface, a steepest descent algorithm can be processed, taking into account the restrictions on the definition domains of the  $GLD(\lambda_1, \lambda_2, \lambda_3, \lambda_4)$ . The Gauss-like algorithm ends when either a given precision is reached (for example,  $S(\lambda_3, \lambda_4) \leq \varepsilon$ ) or a given number of iterations is reached, since the requested precision may be unreachable for the surface considered. It is necessary to notice that, when the moment method is used, different parameter values can give rise to the same moments and so, even if a set of parameters minimizes the surface defined in Eq. (19), it may in fact fail to adequately represent the actual distribution of data. In order to avoid selecting an unsuitable set of GLD parameters, the LambdaFinder software offers both qualitative and quantitative validation of the obtained GLD. Indeed, on the one hand, while plotting the probability

density function of the calculated GLD on the histogram of the initial data, the LambdaFinder software enables a qualitative visualization of the adequacy of the obtained parameters. On the other hand, a quantitative measurement of the agreement between the initial data and the determined GLD can be obtained by performing usual goodness-of-fit tests (which are the  $\chi^2$  and the Kolmogorov–Smirnov tests) as suggested in the literature [23].

Several additional features are included in the LambdaFinder software. One of which is the use of a method based on the identification of the percentiles instead of the moments of the GLD to determine the parameters of the GLD. This method proposed by Karian and Dudewicz [21,22] enables the identification of those members of the GLD family that do not possess the first four moments and yet which may provide superior fit to the data. Moreover, this method eases some of the computational difficulties associated with the calculation of the GLD moments. It must be noticed that the following results were obtained through the method of moments. Nevertheless, in every cases, the results were compared with those obtained with the method of percentiles and the levels of adequacy given by the  $\chi^2$  goodness-of-fit criterium were very close (the  $\chi^2$ -value was



a little smaller with the method of moments for some distributions, and inversely for some others. But these values remained very similar for both methods).

Further information on the LambdaFinder software and a demo-version can be sent by mail by the authors upon request.

### 3.4. Analytical PDF of the Paris coefficients $n$ and $\ln C$

To determine the analytical PDF of the Paris coefficients  $n$  and  $\ln C$ , the moments given by Eqs. (6)–(9) were estimated for the data plotted in the histograms presented in Fig. 2a and b.

Their values are given by Eq. (21):

$$\begin{aligned}\hat{\alpha}_1(n) &= 2.8174, & \hat{\alpha}_2(n) &= 0.0927, \\ \hat{\alpha}_3(n) &= -0.025, & \hat{\alpha}_4(n) &= 3.043\end{aligned}\quad (21)$$

Then, using the expression of the GLD( $\lambda_1, \lambda_2, \lambda_3, \lambda_4$ ) moments given in (6)–(9), the LambdaFinder software is used to perform the minimization procedure on the related surface defined by Eq. (19). Fig. 3.a presents the 3D surface obtained from the data of the  $n$  values, and Fig. 3.b shows its projected representation in 2D. The minimization leads to the couple  $(\lambda_3, \lambda_4) = (0.1302, 0.1239)$ , and using Eq. (20), the values  $(\lambda_1, \lambda_2) = (2.08199, 2.0292)$  are obtained. Using the same procedure, the following parameters  $(\lambda_1, \lambda_2, \lambda_3, \lambda_4) = (-15.83, 1.1611, 0.1080, 0.1060)$  are found for the histogram of the  $\ln C$  values. A very good agreement, in both qualitative (the distribution shape is correctly approximated) and quantitative (the  $\chi^2$  goodness-of-fit test is satisfied for the usual values of the  $\alpha$  parameter, and the values of the criterium are reported in Table 1, as detailed below) ways, between the experimental data and the plotted GLD( $\lambda_1, \lambda_2, \lambda_3, \lambda_4$ ) related to both  $n$  and  $\ln C$  histograms can be noticed in Fig. 2a and b. In Bigerelle [16], these histograms were modelled by a Gaussian distribution and a very satisfactory accordance between the analytical PDF and the experimental histograms was found. Nevertheless, the results reported in Table 1 reveal that, using the classical  $\chi^2$  goodness-of-fit criterion, the Lambda Distributions give an even better modelling of our histograms than under the assumed Gaussian distribution. Indeed, the  $\chi^2$ -values obtained for a Gaussian modelling are 15 to 55% higher than those obtained for the Lambda Distribution modelling, considering the same classes (in number, width, and limits).

The calculated  $\chi^2$ -values can be valuably compared in order to determine the most relevant data modelling. It must

Table 1  
Comparison of the goodness-of-fit for GLD ( $\lambda_1, \lambda_2, \lambda_3, \lambda_4$ ) and a Gaussian distribution with the  $n$  and  $\ln C$  histograms using  $\chi^2$ -values

	GLD( $\lambda_1, \lambda_2, \lambda_3, \lambda_4$ )	Gaussian distribution
$\ln C$	82	128
$N$	38	45

be mentioned that usually, the  $\chi^2$ -values cannot be compared directly for statistical laws having different number of parameters. However, the very high numbers of classes used for the experimental histograms led to high and almost equal degrees of freedom. Thus the direct comparison of the  $\chi^2$ -values provides a relevant estimation of the relative quality of fitting.

## 4. Application to fatigue lifetime prediction

Once all the various materials and structural inputs are obtained, the fatigue lifetime prediction (FLP) can be assessed by integrating the Paris relationship. As  $C$  and  $n$  do not depend on the initial crack length ( $a_i$ ), a simple closed-form equation can be obtained for the FLP. By applying the Irwin criterion, the expression of the stress intensity factor  $K = \alpha\tau\sqrt{\pi a}$  can be obtained, where  $a$  is the crack's length,  $\tau$  is the loading stress,  $\alpha$  is a geometry-load factor related to the crack's shape, the structure and the loading type. Then the integration of the Paris equation gives

$$N_f = \int_0^{N_f} dN = \int_{a_i}^{a_c} \frac{da}{C(\Delta K)^n} \quad (22)$$

and leads to the FLP

$$\begin{aligned}N_f &= \frac{1}{C(\Delta\tau\alpha\sqrt{\pi})^n} \int_{a_i}^{a_c} \frac{da}{a^{n/2}} = \frac{1}{C(\Delta\tau\alpha\sqrt{\pi})^n \left(\frac{n}{2} - 1\right)} \\ &\times \left[ \left(\frac{1}{a_i}\right)^{n/2-1} - \left(\frac{1}{a_c}\right)^{n/2-1} \right] \quad \text{for } n \neq 2\end{aligned}\quad (23)$$

where  $\Delta\tau$  is the stress range during the fatigue cycle. It must be noticed that Eq. (23) assumes that  $\alpha$  remains constant as the crack grows.

Two cases have to be considered whether the initial crack's length is supposed to be known or not. In order to illustrate our approach, we now propose to search the  $N_f$  PDF on a simple fracture mechanics problem. In the first case, we will assume that the initial crack's length  $a_i$  is perfectly known, and in the second case,  $a_i$  will be considered as a random parameter whose mean is known.

### 4.1. Engineering problem

Let us consider for example an infinite 7 cm diameter ( $\phi$ ) and 1 cm thick ( $e$ ) hollow cylinder loaded with a cyclic differential pressure of 20 Mpa ( $\Delta\tau$ ). The maximal initial crack's length located on the inner surface of the cylinder ( $a_i$ ) is supposed to be 1.5 mm in length. The cylinder, made of the aluminium alloy [17] for which the distribution of the parameters of the Paris–Erdogan relationship (1) were obtained, is considered to be perforated as soon as the crack reaches the outer surface. Taking  $\alpha = 0.71$ , an elementary

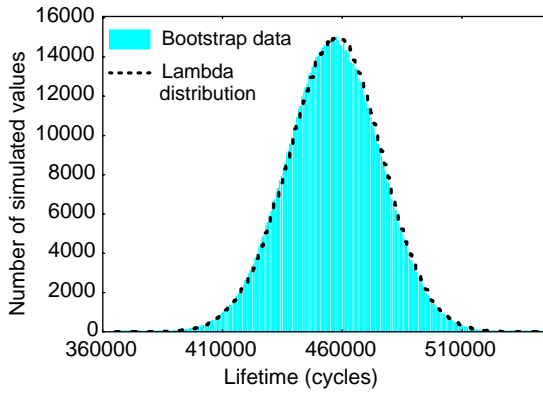


Fig. 5. Aluminium cylinder fatigue lifetime prediction histogram for  $a_i = 1.5$  mm. The corresponding Lambda Distribution is also shown.

mechanical analysis allows us to know that  $\Delta K$  varies between 3.4 and 8.8  $\text{MPa}\sqrt{\text{m}}$  so that the crack's propagation can be modelled by the Paris relationship.

#### 4.2. Lifetime prediction

To analyse the influence of the initial crack length variation on the cylinder's lifetime, 1,000,000 samples ( $n$ ,  $\ln C$ ) have been simulated, using the data shown in Fig. 1. After introducing these 1,000,000 couples in Eq. (23), the lifetime prediction histogram can be obtained and is plotted in Fig. 5.

The mean lifetime prediction is 456,400 cycles with the 98% confidence interval given by [410,677; 500,921] cycles

(Fig. 6). Using the procedure previously described, it is possible to determine the associated Lambda Distribution, so that an analytical expression of the FLP PDF can be obtained. The minimization procedure ends up in the following Lambda Distribution GLD [457,441; 0.0000105; 0.1468; 0.1337]. Fig. 5 highlights the fact that the Lambda Distribution accurately models the data obtained by the Bootstrap technique. As far as the  $\chi^2$  goodness-of-fit criterion is concerned, the GLD [457,441; 0.0000105; 0.1468; 0.1337] better estimates the data than a usual Gaussian law (Table 2). Nevertheless, both distributions have the same level of accuracy in this case.

Another way to quantify this goodness-of-fit is to consider the 1 and 99% centiles. Indeed, such an approach is a useful validation because lots of calculations are based on lifetime confidence intervals, as far as fatigue design is concerned. Therefore it is a major issue to know whether the used distribution gives correct confidence intervals or not. With that aim in view 100 series of 100,000 samples have also been simulated to calculate the 100 corresponding centiles. It is then possible to draw the PDF associated with these centiles (Fig. 6) and to calculate their means.

The values of the centiles obtained by the Bootstrap technique and those modelled by the related Lambda Distribution were then compared. The shapes of the centiles' histograms can be qualitatively said to be very close, and both are correctly modelled by a Gaussian law. Furthermore, the 98% symmetric confidence interval obtained using the Lambda Distribution expression is

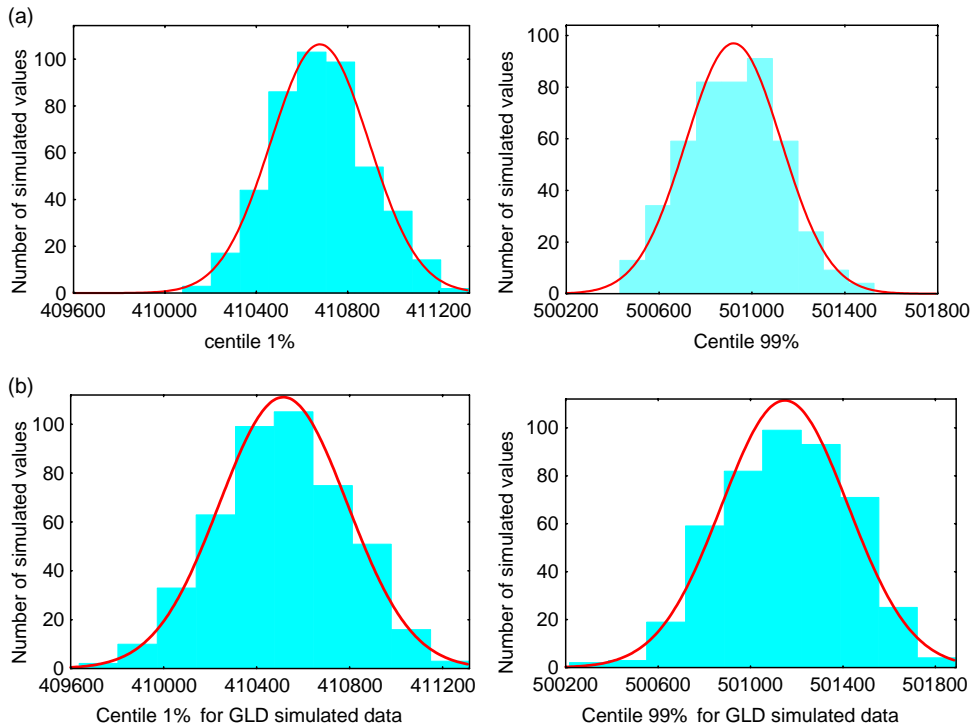


Fig. 6. 1 and 99% lifetime histograms obtained by the Bootstrap technique based on (a) the lifetime prediction histogram (Fig. 5) and on (b) the lifetime prediction modelled with a Lambda Distribution.



Table 2

Comparison of the goodness-of-fit between GLD ( $\lambda_1, \lambda_2, \lambda_3, \lambda_4$ ) and a Gaussian distribution for the FLP PDF using  $\chi^2$ -values

	GLD( $\lambda_1, \lambda_2, \lambda_3, \lambda_4$ )	Gaussian	Lognormal	Weibull
FLP PDF for $\sigma_i=0$	14	208	1570	39

[410,514, 501,148], giving an error of less than 0.43%, which means that the Lambda Distribution models the FLP PDF with a very high level of accuracy. Moreover, this agreement between the two methods indicates that the method used for the identification of the GLD parameters was quite robust. If the parameters calculated through the method of moments had suffered from a large variability, the lifetimes predicted by the GLD modelling would have been much more different from those obtained through the Bootstrap technique, especially in the tails of the distribution.

#### 4.3. Influence of the initial crack length PDF on lifetime prediction

In the previous case, the initial crack's length  $a_i$  was considered to be perfectly known. Let us now assume that  $a_i$  is a random variable the mean of which is perfectly known ( $\bar{a}_i = 1.5$  mm in this example). Such a kind of variability will add another degree of uncertainty to that resulting from the Paris relationship parameters. It can easily be conceived that if the initial crack's length is smaller than the mean crack's length considered previously on the same structure, the lifetime is expected to be longer and vice versa. The question which arises is to know if the mean lifetime corresponding to the mean crack's length is equal to the mean of the lifetimes calculated for the entire crack's population; in other words whether the lifetime calculation is biased or not.

To deal with the influence of the initial crack's length on the FLP PDF, it will be assumed that the initial crack length population is the consequence of a specific process (e.g. notch, threading, groove) for which this population obeys a Gaussian distribution with  $\bar{a}_i$  mean and  $\sigma_i$  standard deviation. To generate the population the following law is used

$$a_i = \bar{a}_i(1 + \Psi\sigma_i/100) \quad (24)$$

where  $\Psi$  is a standard Gaussian random number obtained through the Box and Muller transformation [24]. Now, by introducing the combination ( $a_i, n, \ln C$ ) in Eq. (23) and generating a high number of values, the FLP PDF can be obtained for different values of the standard deviation  $\sigma_i$  ( $\bar{a}_i = 1.5$  mm in this example).

Fig. 7 shows an example of seven modelled lifetime PDF using Lambda Distributions corresponding to seven values

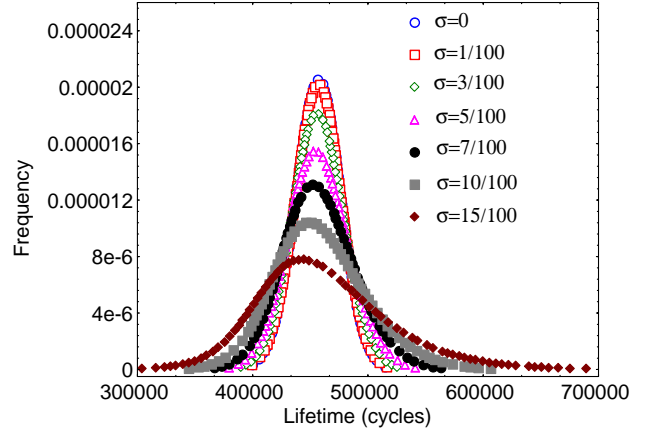


Fig. 7. Evolution of the lifetime fatigue prediction PDF with an increasing value of the initial crack length standard deviation  $\sigma_i$ .

of standard deviation and Fig. 8 presents the evolution of their first four moments.

The following observations can be drawn from these figures:

- The mean lifetime increases whereas the mean initial crack length remains constant. Moreover, the mode (maximum of the PDF shown in Fig. 7) decreases by approximately 20,000 cycles between  $\sigma_i=0$  and 0.15, meaning that for this latter value, the mode of the FLP PDF does no longer belong to the 68% confidence interval obtained for  $\sigma_i=0$ . It highlights the fact that if the initial crack length variation is not taken into account, the most probable lifetime value may not belong to the 68% confidence interval and that the mean is no longer a relevant indicator, which might lead to dramatic lifetime prediction mistakes. That is why it is fundamental to consider the stochastic variability of the initial crack, in the calculation of a lifetime prediction through fracture mechanics. Indeed, every deterministic fracture mechanics model that skips this kind of variability would be based on insufficient data and may lead to hazardous lifetime predictions.
- The lifetime scattering grows exponentially with the initial crack length standard deviation (Fig. 8b). Such an evolution can be described by the following equation obtained by using a non-linear regression

$$\sigma_N = \sigma_{0,N} + k\sigma_i^n \quad \text{with } r = 0.9997 \quad (25)$$

with  $\sigma_{0,N}=19,460$  cycles which is the lifetime's standard deviation considering no stochastic nature of  $a_i$ ,  $k=2.15 \times 10^{10}$  is a scaling coefficient and  $n=1.57 > 1$  is a power law exponent. This last value characterizes the magnitude of the exponential evolution, with the initial crack length standard deviation, of the scattering in fatigue lifetime. This result highlights the particular attention that should be paid to

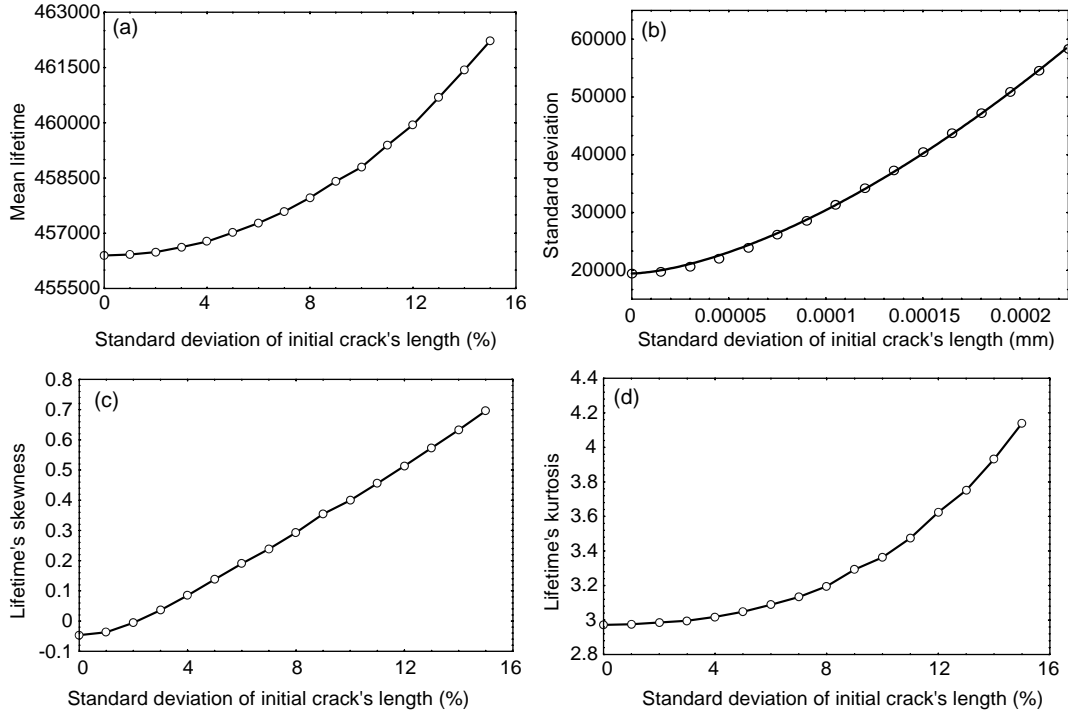


Fig. 8. Evolution of the FLP PDF first four moments with an increasing value of the initial crack length standard deviation  $\sigma_i$ .

the control of the initial crack length variation in order to minimize the scattering of the structure lifetime prediction.

- The higher the initial crack's length standard deviation, the more asymmetric the FLP PDF. Fig. 87c shows that the skewness, initially close to zero, increases much with  $\sigma_i$ . Moreover, the distribution's tail becomes longer as the initial crack length standard deviation increases. In this case, the experimental asymmetry thus noticed necessarily results from the stochastic distribution of initial cracks, and not from an eventual stochastic nature of the crack's propagation. We can even go further and say that taking into account the stochastic nature of crack's propagation ( $n$  and  $C$  coefficients PDF) is insufficient and cannot render the correct shape of the FLP PDF if the stochastic nature of the initial crack's distribution is not considered.

Fig. 8 shows that the evolutions of the first-four moments are highly non-linear and highlights the large variation of the FLP PDF shape due to a mere  $\sigma_i$  increase. The evolutions of the third and fourth moments of the FLP PDF cannot easily be deduced from those of the first two moments. This means that a simple model based on only these first two moments seems to be insufficient, which confirms results given in Bigerelle [16].

Fig. 9 shows the evolution of the four parameters of the Lambda Distribution that models the PDF lifetime. These four parameters also present non-linear variations with the initial crack's length standard deviation  $\sigma_i$ . These results emphasize the difficulty and inaccuracy in modelling fatigue

lifetime using a distribution defined by only two parameters, such as a Gaussian law.

In order to illustrate this last observation, Fig. 10.a and b show the modelling of the FLP PDF by three usual laws (Gaussian, Lognormal and Weibull laws) for  $\sigma_i=0$  and 0.15, respectively. If modelling by a Gaussian or a Lognormal law looks accurate for  $\sigma_i=0$ , none of the three usual laws is able to model accurately the FLP PDF for  $\sigma_i=0.15$ . On the contrary, Fig. 11b qualitatively shows that the Lambda Distribution is still able to model the shape of the FLP PDF in the latter case. This qualitative observation is quantitatively confirmed by the results from the  $\chi^2$  goodness-of-fit criterion reported in Table 3. Indeed, for  $\sigma_i=0$ , the Lambda Distribution gives a  $\chi^2$ -value that is 64% lower than the Weibull one which is the most accurate of the three usual fatigue lifetime distributions in this case. For  $\sigma_i=0.15$ , the Lambda Distribution gives a  $\chi^2$ -value that is 11% lower than the Lognormal one which is the most accurate of the three usual fatigue lifetime distributions in this case.

The above results enable us to draw a general remark. As can be seen on Fig. 10 the Lognormal and Gaussian laws look qualitatively more able to model accurately the obtained lifetime data than the Weibull one. Nevertheless, the  $\chi^2$  goodness-of-fit criterion gives the Weibull law to be the most adequate. This particular example shows that comparisons between the adequacies obtained for usual laws are often ambiguous or contradictory, meaning that they do not always enable us to conclude. The final choice among laws often depends on what is the most interesting to model (either the central tendency or a particular region of

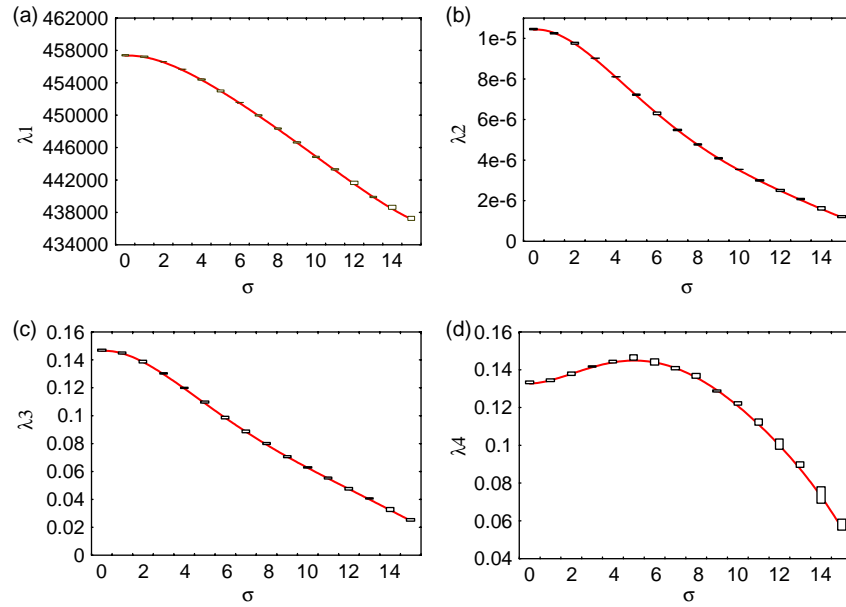


Fig. 9. Evolution of the ( $\lambda_1, \lambda_2, \lambda_3, \lambda_4$ ) parameters of the Lambda Distribution fitting with the FLP PDF with an increasing value of the initial crack length standard deviation  $\sigma_i$ .

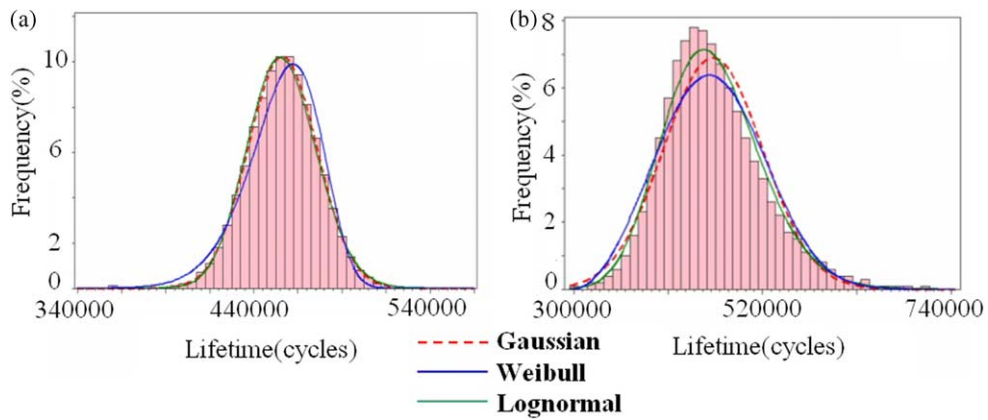


Fig. 10. Modelling of the FLP PDF by the Gaussian, Lognormal (three parameters) and Weibull (three parameters) distributions for (a)  $\sigma_1=0$  (b)  $\sigma_1=0.15$ .

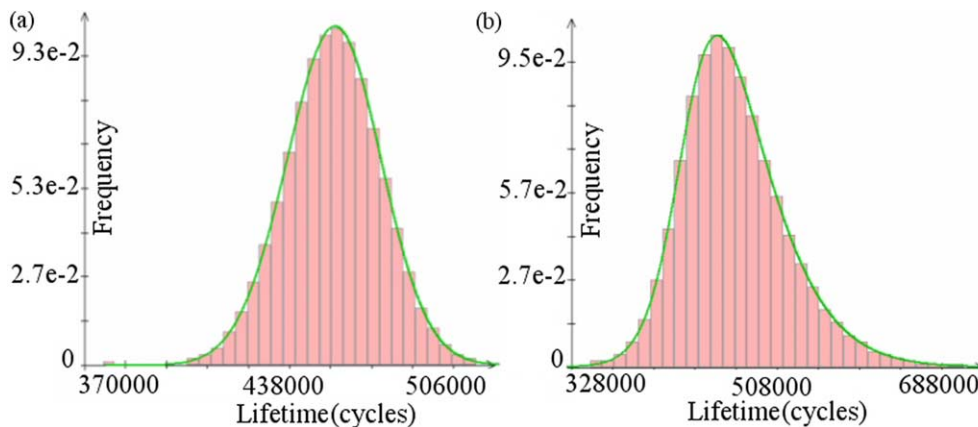


Fig. 11. Modelling of the FLP PDF by the Lambda Distributions for (a)  $\sigma_1=0$  (b)  $\sigma_1=0.15$ .

Table 3

Comparison of the goodness-of-fit between  $GLD(\lambda_1, \lambda_2, \lambda_3, \lambda_4)$  and a Gaussian distribution for the FLP PDF using  $\chi^2$ -values

	$GLD(\lambda_1, \lambda_2, \lambda_3, \lambda_4)$	Gaussian	Lognormal	Weibull
FLP PDF for $\sigma_i=0.15$	8	105	9	68

the data). Contrary to the usual laws, the Lambda Distributions gave the best qualitative and quantitative adequacy in every case of this study, giving a good modelling of both central and extreme data. This latter statement will be detailed by checking the quality of fit of the tails of the Lambda Distributions. Indeed, a good fit around the mean (guided by the bulk of the data) does not necessarily implies a good fit at the tails. As far as the authors know, there does not exist a statistical test universally admitted to account for the ‘goodness-of-fit’ of the tails. Thus, as a complete study would be needed to assess the quality of the tails’ modelling by the Lambda Distributions in a general case, only some results will be presented in the particular case under study in this article.

To evaluate the adequacy between the Bootstrap lifetimes data and their modelling by the Lambda Distributions and to compare it with other usual distributions, several values of ‘extreme’ percentiles were computed. These results are shown in Fig. 12 on which the absolute standard errors  $ASE = |(Bootstrap\ value - estimated\ value)/Bootstrap$

value|are plotted for several usual distributions modelling and the Lambda Distribution modelling at different values of  $\sigma_i$ . These standard errors are calculated for different percentiles characteristic of the distribution tails (i.e. 0.01, 0.1, 1.5 and 10 for the left tail and 90, 95, 99, 99.9 and 99.99 for the right tail). As it can be seen, for every values of  $\sigma_i$  the Lambda Distribution offers a much better fit of the right tail than any other usual distribution. Concerning the left tail, it can be noticed that, even if the Lambda Distribution modelling does not always offer the best adequacy, it is never in error of more than 20%, and this, only for the extreme tails. Nevertheless, it must be outlined that the quality of fit in the left tail is quite similar and even the most accurate with the Lambda Distribution for the values 0 and 5 of  $\sigma_i$ .

To complete the comparison between the use of the Lambda Distributions and other usual distributions to model fatigue lifetime data, their consistency with the physics of fatigue process was checked. A usual way used in reliability analysis to take this into account is to draw the hazard function of the modelled distribution and to compare it with the experimental one. In this context, the hazard is defined as the conditional probability that a component has survived until time  $t$  and fails within  $t + \Delta t$ . Practically speaking, the hazard function  $h$  is defined by

$$h(N) = \frac{PDF(N)}{1 - CDF(N)} \tag{26}$$

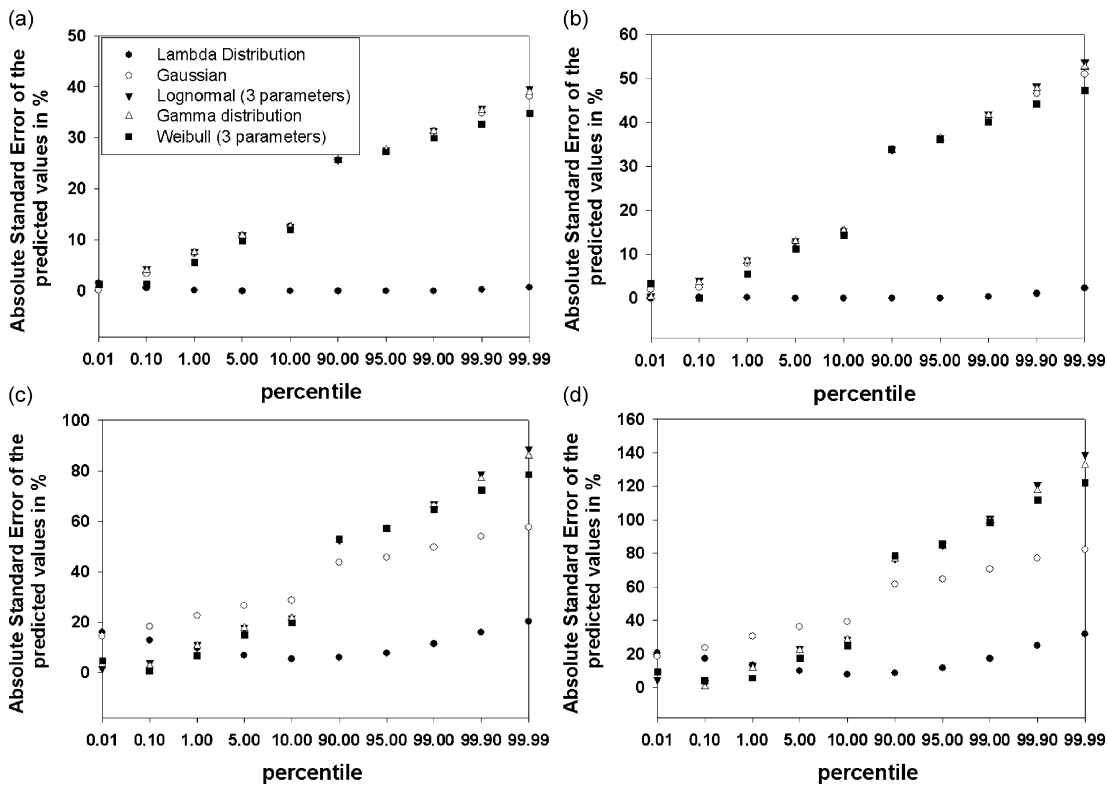


Fig. 12. Values of the Absolute Standard Error  $ASE = |(Bootstrap\ value - estimated\ value)/Bootstrap\ value|$  obtained with various usual modelling of the FLP PDF for (a)  $\sigma_i=0$ , (b)  $\sigma_i=0.05$ , (c)  $\sigma_i=0.10$  and (d)  $\sigma_i=0.15$ .

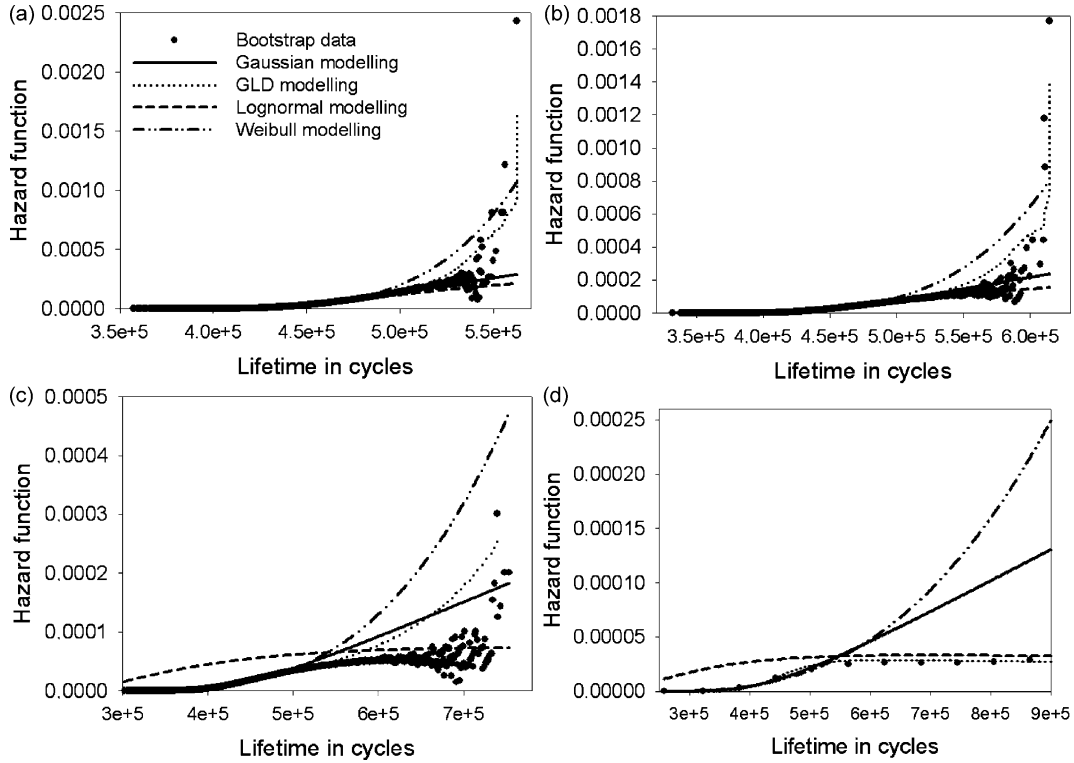


Fig. 13. Hazard functions obtained with various usual modelling of the FLP PDF and with the Bootstrap data for (a)  $\sigma_i=0$ , (b)  $\sigma_i=0.05$ , (c)  $\sigma_i=0.10$  and (d)  $\sigma_i=0.15$ .

where  $N$  is the lifetime (in cycles), and CDF is the Cumulative Density Function of the fatigue lifetime. Analytical expressions are known for the usual analytical probability laws, and Fig. 13 shows the curves obtained in our particular case. Once more if the ability of the Lambda Distributions in general reliability analysis were to be analysed, a complete study would be needed, and this is not the point of this article. Therefore, only a qualitative validation and comparison is presented here to show that there is no major incompatibility related to the use of the Lambda Distributions.

In fact, in the case where  $\sigma_i=0.15$ , if the Weibull and the Gaussian modelling offer a correct fitting with the hazard function for the left part of the curve, they completely overestimate the hazard values after  $6 \times 10^6$  cycles. On the contrary, the lognormal modelling, which gave the second best  $\chi^2$ -value, is unable to correctly estimate the left part of the hazard function, but offers a good approximation of the right part. Nevertheless, the GLD modelling is the only one able to render correctly the whole shape of the hazard function as shown in Fig. 13d. In Fig. 13a and b, i.e. for  $\sigma_i=0$  and 0.05, the GLD modelling seems to be able to account for the sharp increase of the hazard function for the high numbers of cycles, even though the Weibull modelling also offers a globally correct hazard function. Finally, for  $\sigma_i=0.1$ , there is clearly no analytical PDF able to render the whole hazard function obtained for the Bootstrap data. Both choices of the Weibull and the GLD modelling could be

valuably argued. These examples do not try to quantitatively account for the best modelling considering a criterion based on the ability to fit the experimental hazard function, nevertheless they show that the GLD modelling is definitely not incompatible with the physics of fatigue. It can even offer a very good modelling when the other usual distributions are completely in error (Fig. 13d).

As a final illustration, Fig. 14 represents the evolution of the 68% confidence interval and its boundaries as functions of  $\sigma_i$  for the usual Gaussian definition ( $\mu \pm \sigma_i$ ) and for the percentile definition (associated with the Lambda Distribution). This Figure emphasizes how hazardous it might be to consider that the FLP PDF can be modelled by a Gaussian law for the large values of  $\sigma_i$ . Indeed, as  $\sigma_i$  increases, it can be seen that the Gaussian confidence interval tends to overestimate the actual confidence interval, which means that a Gaussian model may be too optimistic in lifetime prediction. As already said, as far as fatigue design is concerned, confidence intervals are of major interest. If the selected model tends to overestimate these confidence intervals, it may lead to hazardous conclusions.

Finally, it must be mentioned that only a mere increase of  $\sigma_i$  was considered in this paper whereas many more parameters (such as shape, position and orientation factors) should be introduced to model more accurately the initial crack's population. It was shown that a model with only two parameters is unable to account accurately for the evolution

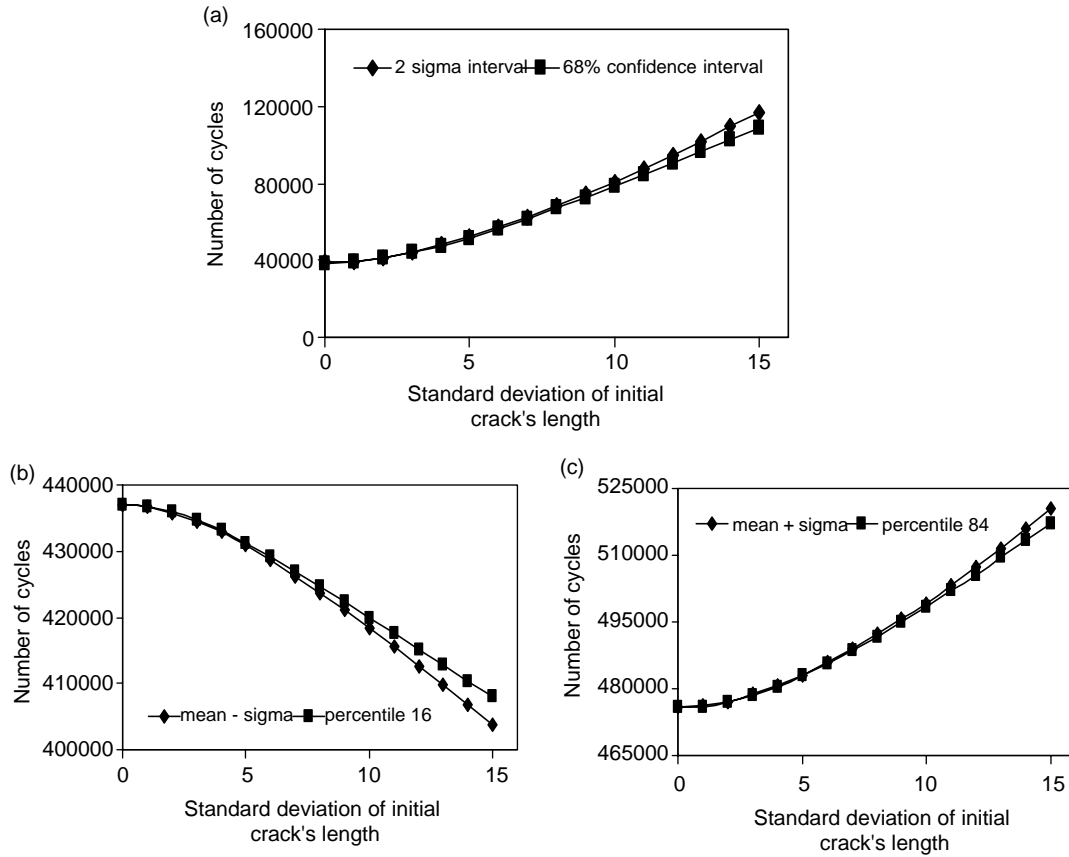


Fig. 14. Evolution of (a) the 68% confidence interval, (b) its lowest bound and (c) highest bound as functions of the value of the initial crack length standard deviation in the case of the FLP PDF for the Gaussian and the Lambda Distribution modelling.

of the FLP PDF induced by the variation of a single parameter in the initial crack's population. It is reasonable to conclude that a more complex description of this initial population would make a two-parameter model of the FLP PDF even more inaccurate.

## 5. Conclusion

The Lambda Distributions associated with the Bootstrap technique were first used to model the Paris coefficients PDF and turned out to be able to estimate accurately the experimental values. Then, Lambda Distributions were used to model the PDF lifetime of a basic structure under fatigue loading. This family of distributions is shown to be more suited than the usual Gaussian, Lognormal or Weibull laws to model the PDF lifetime and to follow accurately the evolution of its (skewed) shape due to the modification of initial crack length population. This comparison was drawn by the means of usual goodness-of fit tests. Moreover, the quality of fitting at the tails were checked by means of the calculation of several 'extreme' percentiles. The study of the modelled and experimental hazard functions account for the consistency of the distributions with the physics of the

fatigue process and the qualitative comparison with usual distributions still shows that the GLD modelling is definitely suited to model fatigue lifetime PDF. The experimental asymmetry of the lifetime PDF is found to result from the stochastic distribution of initial cracks and not from the stochastic nature of the crack's propagation. Despite the mathematical complexity of the Lambda Distribution determination, this method is made easy to apply thanks to the LambdaFinder software we especially designed to perform such a task. A main advantage of this technique is to allow the modelling of a large panel of various shapes and to facilitate Monte Carlo simulations thanks to the Lambda Distribution definition based on their percentile function.

## Acknowledgements

The authors would like to thank L. Minguet for its invaluable contribution to the design and programming of the LambdaFinder software, and V. Hague for the language corrections of this paper. The referees are particularly acknowledged for enabling us to complete our study and to present a more rigorous article.



## References

- [1] Soboyejo WO, Shen W, Lou J, Mercer C, Sinha V, Soboyejo ABO. A probabilistic framework for the modelling of fatigue in a lamellar XD gamma titanium aluminide alloy. *Int J Fatigue* 2002;24:69–81.
- [2] Hall PL, Strutt E. Probabilistic physics-of-failure models, for component reliabilities using Monte Carlo simulation and Weibull analysis: a parametric study. *Reliab Eng Syst Saf* 2003;80(3):233–42.
- [3] Ferreira JLA, Balthazar JC, Araujo APN. An investigation of rail bearing reliability under real conditions of use. *Eng Fail Anal* 2003; 10(6):745–58.
- [4] Lewis G, Janna S, Carroll M. Effect of test frequency on the in-vitro fatigue life of acrylic bone cement. *Biomaterials* 2003;24(6):1111–7.
- [5] Petryna YS, Pfanner D, Stangenberg F, Krätzig WB. Reliability of reinforced concrete structures under fatigue. *Reliab Eng Syst Saf* 2003;77(3):253–61.
- [6] Pauchard V, Grosjean F, Campion-Boulharts H, Chateauinois A. Application of a stress-corrosion-cracking model to an analysis of the durability of glass/epoxy composites in wet environments. *Compos Sci Technol* 2002;62(4):493–8.
- [7] Pauchard V, Chateauinois A, Grosjean F, Odru P. In situ analysis of delayed fiber failure within water-aged GFRP under static fatigue conditions. *Int J Fatigue* 2002;24(2–4):447–54.
- [8] Diao X. A statistical equation of damage evolution. *Eng Fract Mech* 1995;52(1):33–42.
- [9] Caron JF, Ehrlacher A. Modelling of fatigue microcracking kinetics in crossply composites and experimental validation. *Compos Sci Technol* 1999;59(9):1349–59.
- [10] Nagode M, Fajdiga M. On a new method for prediction of the scatter of loading spectra. *Int J Fatigue* 1998;20(4):271–7.
- [11] Nagode M, Fajdiga M. A general multi-modal probability density function suitable for the rainfall ranges of stationary random processes. *Int J Fatigue* 1998;20(3):211–23.
- [12] Gope PC. Determination of sample size for estimation of fatigue life by using Weibull or Lognormal distribution. *Int J Fatigue* 1999;21(8): 745–52.
- [13] Ferry L, Perreux D, Varchon D, Le Bras J. Tensile failure of filament-wound pipes under long-term creep loading: a probabilistic analysis. *Compos Sci Technol* 1997;57(9–10):1281–8.
- [14] Nikbin KM, Yatomi M, Wasmer K, Webster A. Probabilistic analysis of creep crack initiation and growth in pipe components. *Int J Pressure Vessels Piping* 2003;80(7–8):585–95.
- [15] Liao M, Shi G, Xiong Y. Analytical methodology for predicting fatigue life distribution of fuselage splices. *Int J Fatigue* 2001;23(Spl. 1):177–85.
- [16] Bigerelle M, Iost A. Bootstrap analysis of fatigue crack growth rate: application on the Paris' relationship and to lifetime prediction. *Int J Fatigue* 1999;21(4):299–307.
- [17] Krausz A, Zhiven XW, Krausz K. On the constitutive law of environment assisted fatigue: the physical meaning of the Paris type equations. *Zeitschrift der Metallkunde* 1992;83(5):356–63.
- [18] Paris PC, Erdogan F. A critical analysis of crack propagation laws. *J Basic Eng, ASME, Series D* 1963;55:528–34.
- [19] Efron B, Tibshirani R. Statistical data analysis in the computer age. *Science* 1991;253:390–5.
- [20] Najjar D, Bigerelle M, Iost A. The computer based Bootstrap method as a tool to select a relevant surface roughness parameter. *Wear* 2003; 254:450–60.
- [21] Karian ZA, Dudewicz EJ. Fitting statistical distributions: the generalized Lambda Distribution and generalized Bootstrap method.: CRC Press; 2000.
- [22] Karian ZA, Dudewicz EJ. Fitting the generalized Lambda Distribution to data: a method based on percentiles. *Commun Stat: Simul Comput* 1999;28(3):793–819.
- [23] Lakhany, Mausser. Estimating the parameters of Generalized Lambda distribution. *ALGO Res Q* 2000.
- [24] Box, Muller. *Annals of Statistics*; 1958.