A generic statistical methodology to predict the maximum pit depth of a localized corrosion process

A. Jarrah [a,b,*], M. Bigerelle [c], G. Guillemot [b,d], D. Najjar [b,d], A. Iost [b,d], J.-M. Nianga [a]

[a] Equipe Mécanique des Structures, Hautes Etudes d'Ingénieur, 13 Rue de Toul, 59046 Lille Cedex, France
[b] Arts et Métiers ParisTech – Centre de Lille, 8, Boulevard Louis XIV, 59000 Lille Cedex, France
[c] Laboratoire Roberval, UMR CNRS 6253, UTC Centre de Royallieu, BP 20529, 60205 Compiègne, France
[d] Laboratoire de Mécanique de Lille (LML), UMR CNRS 8107, PRES Université Lille Nord de France, F-59650 Villeneuve d'Ascq, France

Keywords:
A. Aluminium
B. Modelling studies
C. Pitting corrosion

This paper outlines a new methodology to predict accurately the maximum pit depth related to a localized corrosion process. It combines two statistical methods: the Generalized Lambda Distribution (GLD), to determine a model of distribution fitting with the experimental frequency distribution of depths, and the Computer Based Bootstrap Method (CBBM), to generate simulated distributions equivalent to the experimental one. In comparison with conventionally established statistical methods that are restricted to the use of inferred distributions constrained by specific mathematical assumptions, the major advantage of the methodology presented in this paper is that both the GLD and the CBBM enable a statistical treatment of the experimental data without making any preconceived choice neither on the unknown theoretical parent underlying distribution of pit depth which characterizes the global corrosion phenomenon nor on the unknown associated theoretical extreme value distribution which characterizes the deepest pits.

Considering an experimental distribution of depths of pits produced on an aluminium sample, estimations of maximum pit depth using a GLD model are compared to similar estimations based on usual Gumbel and Generalized Extreme Value (GEV) methods proposed in the corrosion engineering literature. The GLD approach is shown having smaller bias and dispersion in the estimation of the maximum pit depth than the Gumbel approach both for its realization and mean. This leads to comparing the GLD approach to the GEV one. The former is shown to be relevant and its advantages are discussed compared to previous methods.

1. Introduction

Pitting corrosion is an extremely dangerous form of localized corrosion since a perforation resulting from a single pit can cause complete in-service failure of installations like water pipes, heat exchanger tubes or oil tank used for example in chemical plants or nuclear power stations [1–16]. The pits depth distribution is an important characteristic of the extent of such damage; the deeper the pits, the more dramatic the damage. In order to ensure safety and reliability of industrial equipments, statistical procedures have to be proposed to assess the maximum pit depth from data estimated from limited inspection.

In literature, the most common method for safety or reliability was found in the application of the statistical extreme value analysis using the Gumbel methodology to predict the maximum pit depth that will be found in a large scale installation by using a small number of samples with a small area [1–6,11–14,17,18]. This methodology is based on the estimation of the two parameters of the Gumbel distribution. It is worth noting that it has been extended to the three-parameter Generalized Extreme Value distribution (GEV) [8–11,19–26]. The GEV distribution is expressed such that:

$$G(x) = \exp\left(-\left[1+\xi\left(\frac{x-\mu}{\sigma}\right)\right]^{-1/\xi}\right), \quad 1+\xi\left(\frac{x-\mu}{\sigma}\right) > 0, \quad \xi \neq 0 \tag{1}$$

where μ is the location parameter, σ is the scale parameter and ξ is the shape parameter. Type II (Fréchet) and Type III (Weibull) correspond respectively to $\xi > 0$ and $\xi < 0$. It should be mentioned that Type II has a finite lower bound and Type III has a finite upper

* Corresponding author at: Equipe Mécanique des Structures, Hautes Etudes d'Ingénieur, 13 Rue de Toul, 59046 Lille Cedex, France. Tel.: +33 320622233; fax: +33 320622957.

E-mail addresses: adil.jarrah@yahoo.fr (A. Jarrah), maxence.bigerelle@utc.fr (M. Bigerelle), gildas.guillemot@ensam.eu (G. Guillemot), denis.najjar@ensam.eu (D. Najjar), alain.iost@ensam.eu (A. Iost), jean-marie.nianga@hei.fr (J.-M. Nianga).

bound. The subset of the GEV family with $\xi = 0$ corresponds to the Gumbel distribution and is expressed such that:

$$G(x) = \exp\left(-\exp\left(\frac{x-\mu}{\sigma}\right)\right), \quad -\infty < x < \infty \tag{2}$$

The use of the three-parameter GEV distribution in corrosion literature certainly reflects the difficulty in fitting the experimental distributions of maximum pit depths corresponding to corrosion phenomena of many types and many environments. In other words, there is neither a universally admitted distribution family nor a unique attraction domain for the modelling of the maximum pit depths corresponding to the overall corrosion experiments. It is worth noting that methodologies based on extreme value theory suppose that pit process has the property of homogeneity in law and that it requires occurring with the same frequency in time and space. However, it is acknowledged that pitting corrosion is a stochastic process mainly related to its initiation stage [17]. Furthermore, preliminary hypotheses related to homogeneity are difficult to verify before applying the extreme value methodologies. Moreover the distribution used is often chosen according 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. Finally, it should be noticed that some maximum value distributions do not belong to any of the three attraction domains (Type I, II or III). Therefore it appears that the application of this extreme value theory could lead to major limitations.

Because of these limitations, and others that will be mentioned later in the particular case of the most common Gumbel and GEV approaches, this paper proposes an alternative methodology. This methodology is based on the combination of two statistical methods: the Generalized Lambda Distribution (GLD) and the Computer Based Bootstrap Method (CBBM). Contrary to the Gumbel and GEV approaches that take into consideration inferred parent distributions constrained by specific mathematical assumptions both the GLD and the CBBM present the main advantage of avoiding to make any preconceived choice neither of the unknown theoretical underlying distribution which governs the corrosion phenomenon nor on the unknown associated theoretical extreme value distribution which characterizes the deepest pits [27,28]. Moreover, as far as GLD are concerned, it is worth noting that these distributions are highly flexible thanks to their ability to take a large variety of shapes within one distributional form. These four-parameter distributions can match with any mean, variance, skewness, kurtosis, tails that are truncated or extend to infinity on either or both sides. In this way, GLD can model many distributions often observed with engineering data [29,30] such as Weibull, Cauchy, Normal, Log-Normal, Gumbel, Pareto, bell-shaped distribution as well as inverted bell-shaped ones to name a few. Furthermore, the GLD is able to represent distributional characteristics such as moments (or combination of moments) or percentiles (or combinations of percentiles). Also due to its flexibility in modelling a range of different distributions, it is possible to directly model the underlying process, rather than relying on the central limit theorem and the mean of the process as in the case of traditional statistical analysis. This eliminates difficulties in choosing the appropriate distribution for the data set. It must be stressed that the goodness of fit by the GLD is particularly noticeable in the right tail region of this kind of distributions *i.e.* the region of interest for this study since it corresponds to the extreme values of any distribution. Indeed, it should be noticed that the estimations of GLD parameters are less influenced by the central values than the other distribution estimations. Because of these various advantages, the family of four-parameter GLD have been used in many fields where accurate data modelling is required such as insurance and inventory management [30], finance [31,32], meteorology [30,33], pipeline leakages [30], statistical process control [34,35], independent component analysis [36,37], simulation of queue systems [38] or for generating random number [39]. For several years, the authors of the present paper have also developed the use of this versatile family of distributions in materials science [28,30,40] and statistical control process [27,41].

In this study, a first algorithm was computed to determine the GLD that fits with an experimental distribution of pit depths that propagated from the surface of a A5 aluminium sheet during an accelerated corrosion test performed at free potential in an aqueous acid solution at room temperature. After showing statistically that the fitting of our experimental data is more relevant than any fitting by the most common laws (*i.e.* Normal [9], Lognormal [42], Weibull [43]) used in corrosion engineering literature to model pit depth distributions, the GLD model is considered to estimate the maximum pit depth. Then a second algorithm was computed to generate a high number of simulated datasets using the CBBM from the obtained GLD. Each extreme value of the simulated datasets is used to construct an empirical Probability Density Function (PDF) from which the mean of the maximum pit depth and a 90% confidence interval can be determined.

The GLD and the CBBM are successfully combined to assess the effect of the exposed surface size on the evolution of these statistic estimates. The relevance of this new approach is shown by comparison of the results with those obtained applying the usual Gumbel approach. Differences are explained by the analysis of the attraction domain of pit depth distribution. The results of the new approach presented in this investigation are finally compared with those obtained with the Generalized Extreme Value (GEV) approach and the advantages of combining GLD and CBBM methods are demonstrated.

2. Experimental procedure

2.1. Presentation of the material and the corrosion test

The material used in this experiment is A5 aluminium (*i.e.* for which purity in mass is higher than 99.5%). This purity level was checked by means of a scintillation spectrometer. The corrosion experiment consisted in immersing an aluminium sheet of 75 cm^2 by 0.8 mm in thickness at free potential in an aqueous acid solution at room temperature without any agitation. The as-received sheet was annealed for one hour at 300 °C, and cleaned with acetone before 6-h immersion period in the corrosion solution whose chemical composition was: 0.5 g of NaCl (38 mM), 4 g of $FeSO_4$ (117 mM), 25 cm^3 of H_2SO_4 (2.1 M) and 200 cm^3 of H_2O.

2.2. Characterization and distribution of pits formed during the corrosion test

Firstly, optical macroscopic observation was used to observe at a large scale the damage due to the accelerated corrosion test. Fig. 1a and b reveal respectively the spatial distribution of the corrosion pits and the binary image of the corroded sheet. The total number of pits observed on the surface of the corroded sample was counted visually without any cleaning process. 603 pits were counted and the corresponding depths were measured using an optical microscope with a 50 × magnification. Focus is firstly done on the top of the pit (*i.e.* surface of the sample) and secondary at its bottom. The displacement of the lens corresponds to the pit depth estimation. Fig. 1c shows the 3D topographical pit measurements corresponding to a small pit (71 μm), two medium pits (110, 130 μm) and the pit of maximum depth (163 μm). The 603 values of pit depth were used to determine the experimental frequency

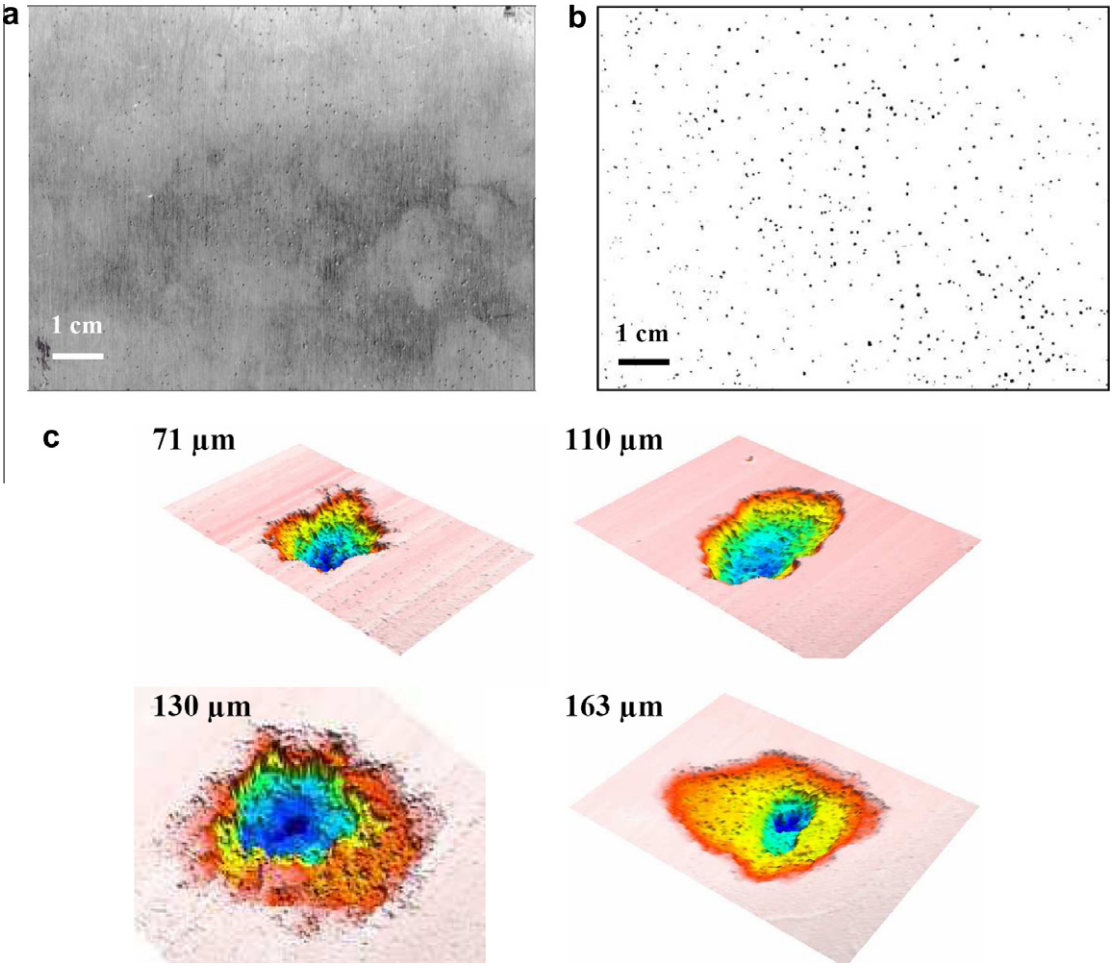

Fig. 1. (a) Optical macrograph of the pits formed during the accelerated corrosion test at the surface of the aluminium sample under study (603 pits). (b) Binarized image showing the indentified pits. (c) 3D topographical pit measurements obtained by means of an optical profilometer and corresponding to a small pit (71 µm), two medium pits (110 µm, 130 µm) and the pit of maximum depth (163 µm).

distribution of depths (Fig. 2). The maximum pit depth of this distribution is equal to 163 µm as show by the descriptive statistics reported in Table 1.

3. Determination of the pit depths distribution using the GLD method

3.1. Presentation of the GLD method [27,29,30]

Three ways can usually be considered to specify the probability for the various values x to occur for a random variable noted X: the distribution function $F_X(x) = Pr(X \leqslant x)$ which indicates the probability of X to be smaller than x, the PDF $f_X(x)$ which is obtained by differentiating the distribution function, the inverse distribution function $Q_X(y)$ which indicates the value of x such that $F_X(x) = y$ (for each y between 0 and 1).

The Generalized Lambda Distribution family is specified in terms of its inverse distribution function with four parameters (λ_1, λ_2, λ_3 and λ_4):

$$Q_X(y, \lambda_1, \lambda_2, \lambda_3, \lambda_4) = \lambda_1 + \frac{y^{\lambda_3} - (1-y)^{\lambda_4}}{\lambda_2} \tag{3}$$

The parameters λ_1 and λ_2 are, respectively, the location and scale parameters, while λ_3 and λ_4 are related respectively to the skewness and the kurtosis of the GLD. The definition range of Q_X is the [0, 1] interval (*i.e.* $0 \leqslant y \leqslant 1$). Q_X evolution as a function of the λ_i parameters values has been widely detailed in [27,29,30]. The evolution domain of Q_X, I_Q, is a function of the sign of λ_3 and λ_4 (Appendix A). PDF $f_X(x)$ can then easily be expressed from the inverse distribution function of the GLD:

$$f_X(x) = \frac{\lambda_2}{\lambda_3 y^{\lambda_3 - 1} + \lambda_4 (1-y)^{\lambda_4 - 1}} \tag{4}$$

where y is the solution of equation $Q_X(y) = x$ which can be solved numerically.

Eq. (3) defines a valid PDF if and only if Q_X meets the following conditions:

$$\begin{cases} f_X(x) \geq \forall x \in D \\ \int_{-\infty}^{+\infty} f_X(x) dx = 1 \end{cases}$$

where D is the definition domain of Q_X.

The main problem is to estimate the parameters λ_1, λ_2, λ_3 and λ_4 in order to have the best fitting of the GLD with the experimental frequency distribution of corrosion pit depths considered in this study. To solve this problem, the method of moments described in Appendix B is performed by a computer treatment of our experimental dataset $x_{i,i} \in \{1, \ldots, n\}$, of size $n = 603$.

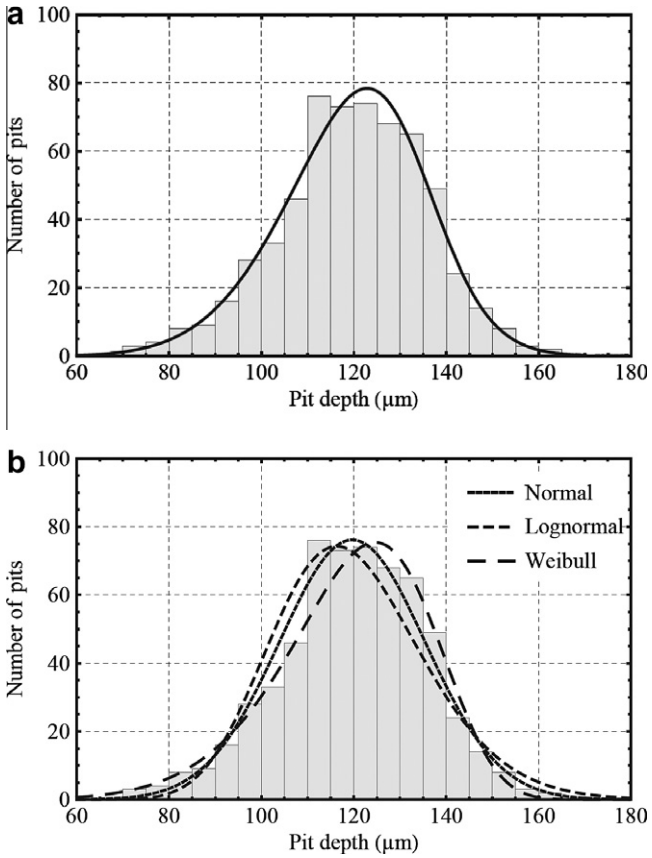

Fig. 2. Experimental frequency distribution of the 603 values of corrosion pit depths and estimated PDF: (a) of the corresponding Generalized Lambda Distribution (GLD), (b) of the three hypothesized parent laws: Normal, Lognormal and Weibull distributions.

Table 1
Descriptive statistics of the 603 pits depths measured at the surface of the corroded aluminium sheet under study.

Mean	Median	Mode	Minimum	Maximum	Std. Dev.	Skewness	Kurtosis
119.6	120.6	123.5	70.3	163.4	15.8	−0.285	0.141

3.2. Procedure and results

An algorithm was written and computed using the Statistical Analyses System (SAS™) software to determine the GLD and its related PDF from the experimental dataset by using the method of moments. This algorithm was used to perform a minimization process of function $\psi(\lambda_3, \lambda_4)$ (Appendix B) by a gradient decreasing method processed for $-0.25 < \lambda_3 < 0.2$ and $-0.25 < \lambda_4 < 0.2$. The value of this function is found to be minimum for the pair ($\lambda_3 = 0.150$, $\lambda_4 = 0.086$) which subsequently yields to the pair ($\lambda_1 = 124.3$, $\lambda_2 = 0.011$). Table 2 summarizes the values of moments as well as the confidence intervals.

Fig. 2a presents the PDF of the GLD corresponding to the experimental frequency distribution of corrosion pit depths. It has been obtained by replacing the four previous values determined by using the method of moments. In corrosion literature, it is reported that the pit depth distribution can be modelled by Normal [9], Lognormal [42] or Weibull law [43]; distributions which belong to the exponential type. Fig. 2b presents the fitting of our experimental pit depth values with those different most common laws and Table 3 summarizes the related results of goodness-of-fit Kolmogorov–Smirnov tests. These statistical test results show that these three distributions of exponential type are unfortunately rejected at a critical level of 0.05. On the contrary, it can be emphasized that a Kolmogorov–Smirnov test gives a value of 0.43 for the GLD. This means that this last distribution fits very well with the experimental data in comparison with Normal, Lognormal and Weibull distributions which are proposed in corrosion engineering literature to model pit depth distributions.

Table 2
Values of the fourth moments and the GLD parameters with their 90% confidence intervals (*i.e.* the difference between Q95 and Q5 that are the 95th and the 5th quantiles respectively) resulting from Bootstrap simulations computed from the original data set of 603 pit depths.

	Moments				GLD parameters			
	$\hat{\alpha}_1$	$\hat{\alpha}_2$	$\hat{\alpha}_3$	$\hat{\alpha}_4$	λ_1	λ_2	λ_3	λ_4
Value	119.7	15.7	-0.28	3.13	124.2	0.011	0.150	0.086
Q5	118.6	15.0	-0.43	2.85	121.7	0.008	0.102	0.061
Q95	120.7	16.5	-0.13	3.40	127.2	0.014	0.219	0.121

Table 3
Kolmogorov–Smirnov test results for goodness-of-fit tests for Normal, Weibull, Lognormal and GLD.

Normal	Weibull	Lognormal	GLD
0.012 (rejected)	0.021 (rejected)	0.001 (rejected)	0.43 (not rejected)

4. Estimation of the extreme value combining the GLD and the CBBM

4.1. Limitations about the application of the Gumbel approach

The Gumbel distribution [44] has been presented in Eq. (2). One must remember that this distribution is in fact an asymptotic limit form of the largest values extracted from a parent distribution of exponential type (*i.e.* Gumbel attraction domain). Hence, there are some mathematical requirements to use this analytical expression in a valid way. The first one relates to the nature of the parent distribution and the second one to the number of largest values of this distribution.

As far as the nature of the parent distribution is concerned, it has been shown just above using a Kolmogorov–Smirnov test that the most common laws (*i.e.* Normal, Lognormal, Weibull) used in corrosion literature to model the distributions of pit depths are unfortunately rejected at a critical level of 0.05. This means that, for these most common distributions of the Gumbel attraction domain, the first mathematical requirement on the nature of the parent distribution is not satisfied in the case of our experimental data. Moreover, even if one of these distributions was accepted, there are only few large values in our experimental data set. As a consequence, the difference between this expected model and the unknown true distribution would have been significant in the right tail region. In other words, the prediction in this right tail region, which corresponds to the extreme values, would have been far from being accurate considering the experimental data set of the present study.

Apart from the particular limitations aforementioned, two general limitations have been noted by several authors working in the field of corrosion [1,12,13,45]. The first limitation is the important property that the double exponential expression is unbounded, which implies a small but finite probability of observing a pit of exceptional depth. Up to now, this conflict between the unbounded

nature of the distribution and the physical limits on pit growth kinetics is subject of debate. The second limitation is that a concrete criterion for the number and size of samples to obtain a reasonable accurate extreme value prediction is not available except by making questionable assumptions. Therefore some criteria deduced from simulations can be proposed to estimate and reduce these uncertainties [46]. Nevertheless, they are based on hypotheses made on the unknown pit location and depth distributions.

Another limitation in the application of the Gumbel distribution comes from the limit theorem. Indeed, the Fisher theorem asserts that if $(X_1,\ldots, X_n)$ is a sequence of independent random variables representing the pit depths with a common distribution function, P, of the Gumbel attraction domain, and if a sequence of pairs of real numbers (a_n, b_n) exists, such that:

$$\lim_{n\to+\infty} P\left(\frac{M_n - b_n}{a_n} \leqslant x\right) = F_X \tag{5}$$

where $M_n = \max(X_1,\ldots, X_n)$, then F_X is the Gumbel distribution function presented in Eq. (2) with $\mu = 0$ and $\sigma = 1$. Therefore, this result supposes that the number of data (*i.e.* number of pits) must be large enough to accept this asymptotic law. For a distribution function, F, of PDF f, in the Gumbel attraction domain, a_n and b_n are given such that:

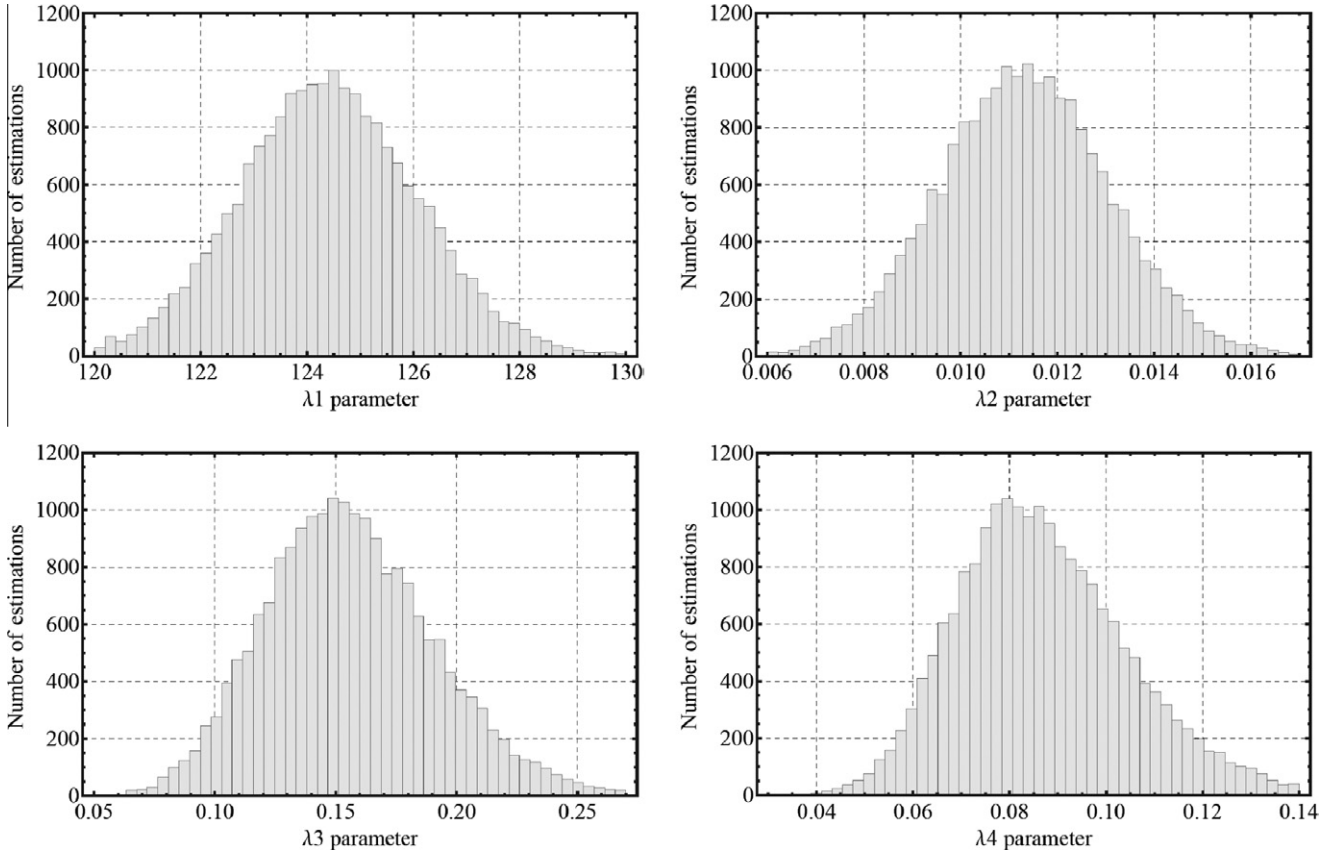

Fig. 3. Histograms of the four parameters of the Generalized Lambda Distribution PDF corresponding to the experimental frequency distribution of corrosion pit depths and obtained with 20,000 Bootstrap simulations.

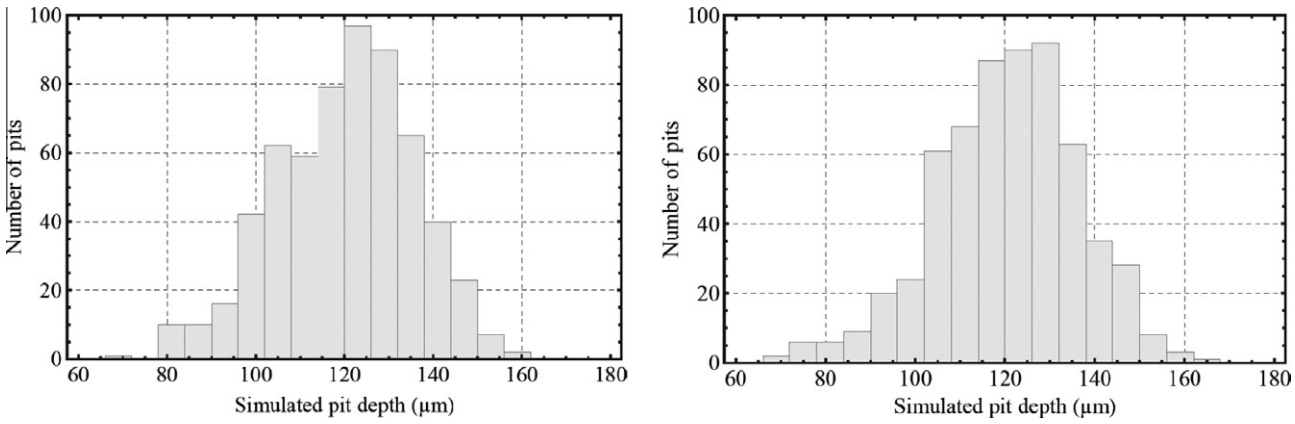

Fig. 4. Two examples of Bootstrap simulated distributions equivalent to the experimental one.

$$b_n = F^{-1}(1 - 1/n)$$
$$a_n = h(b_n) \quad (6)$$

where h is the function defined by:

$$h = (1 - F(x))/f(x) \quad (7)$$

In order to evaluate the convergence speed of the maximal number of samples coming from a Gaussian law, 10^6 simulations of the variable $\frac{M_n - b_n}{a_n}$ are generated with $n = 100$ for which the mean and the variance are calculated. The expected values of these parameters are equal to 0.5772 for the mean, and 1.64 for the variance, whereas the simulated values found are respectively 0.4854 and 1.3126. This simple example shows the problem of convergence toward the expected values of the Gumbel distribution and the dependence on the parent distribution, especially for the Normal distribution. Besides, when the theory supporting the application of the Gumbel distribution was developed as the number of samples becomes infinite, Shibata [1] mentioned that it might not be so plentiful in practice especially in engineering data because measurements are time consuming and expensive.

Because of these numerous limitations, an alternative approach combining the GLD and the CBBM is proposed hereafter to predict the maximum pit depth with a confidence interval for an exposed surface of a given size.

4.2. Presentation of the Computer-Based Bootstrap Method [47,48]

Efron introduced first the Computer-Based Bootstrap Method (CBBM) to avoid the risk of asserting wrong conclusions when analyzing experimental data that transgress the inference assumptions of the traditional statistical theory. One of the main reasons for making parametric assumptions is to ease out the derivation from textbook formulae for standard errors. Unfortunately, the traditional statistical theory does not provide formulae to assess the accuracy of most statistic estimates other than the mean. Since no formulae are necessary using the CBBM in non-parametric mode, restrictive and sometimes-dangerous assumptions about the form of underlying populations can be avoided. Moreover, a standard error (thus an assessment of the accuracy) can be calculated for any computable statistic estimate using the constructed empirical PDF. The CBBM does not work alone and its efficiency is indeed emphasized when applied to other statistical procedure, as will be shown hereafter for the GLD method.

Based on the mathematical resampling technique, the main principle of the CBBM consists in generating a high number B of simulated bootstrap samples from the original data points. The original dataset consists of either experimental or simulated points. A bootstrap dataset of size n, noted $(x_1^*, x_2^*, \ldots, x_n^*)$, is a collection of n values simply obtained by randomly sampling with replacement from the original data points $(x_1, x_2, \ldots, x_n)$, each of them with a probability of $1/n$. the bootstrap dataset thus consists of elements of the original data points; some appearing zero times, some appearing once, some appearing twice, etc.

4.3. Procedure and results

The CBBM has been applied onto the experimental dataset of size $n = 603$ to estimate the variability of the four coefficients λ_1, λ_2, λ_3 and λ_4 considering 20,000 bootstrap simulations. Then histograms of the four coefficients related to the 20,000 new GLD can be plotted (Fig. 3) from which the confidence intervals of these parameters are estimated (Table 2).

A second algorithm was computed using Mathematica™ software to generate simulated datasets of size $n = 603$ using a Monte-Carlo procedure considering the GLD equation corresponding to our experimental dataset:

$$Q_X(v) = \lambda_1 + \frac{v^{\lambda_3} - (1 - v)^{\lambda_4}}{\lambda_2} \quad (8)$$

where v is a uniform random number between 0 and 1 and $P(v)$ is the associated simulated corrosion pit depth. To illustrate the result of this procedure, two examples of simulated distributions are presented in Fig. 4. These distributions of corrosion pit depths obtained by the Monte-Carlo method simulate in fact two possible distributions equivalent to the experimental one.

The density function, $f_{\max}$, of the deepest corrosion pit distribution, corresponding to the GLD, presented in Eq. 8 can be calculated. Indeed, the inverse distribution function of this distribution, $f_{\max}$, is given by:

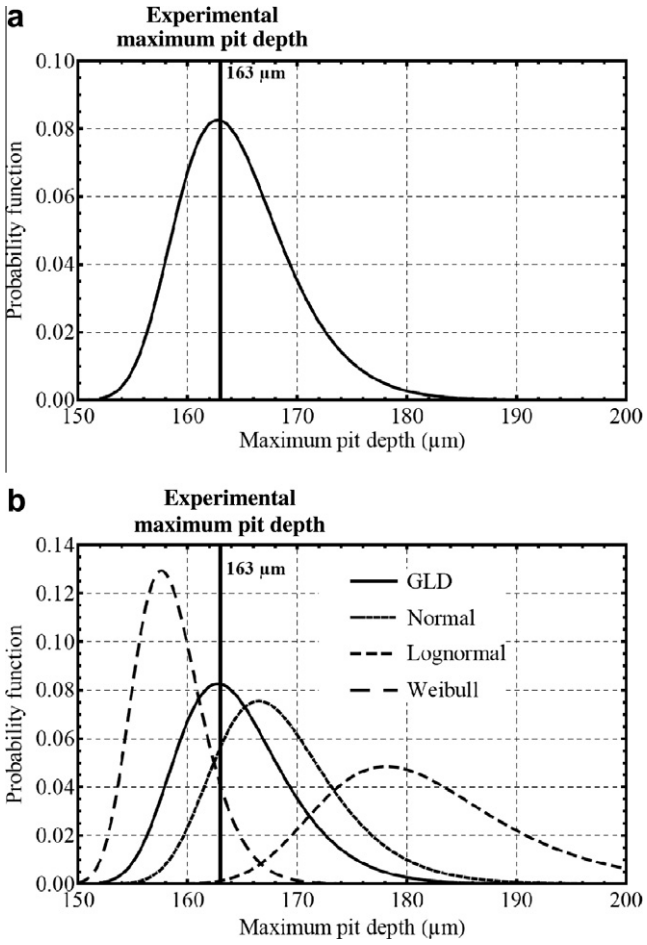

Fig. 5. (a) Representation of the analytical PDF of the maximum pit depths distribution corresponding to the GLD distribution (603 pits). The vertical solid line indicates the experimental maximum pit depth value (163 µm); (b) maximum pit depths probability functions for the four studied parent laws: Normal, Lognormal, Weibull and finally the GLD.

Table 4
Descriptive statistics of extremes values computed from the four distributions under study: Normal, Lognormal, Weibull and GLD.

	Mean	Std. Dev.	Skewness	Kurtosis	Median
Normal	168.5	5.7	0.764	1.060	167.8
Lognormal	176.2	7.7	0.874	1.393	172.2
Weibull	158.5	3.2	0.577	0.557	158.2
GLD	164.5	5.1	0.698	0.761	163.9

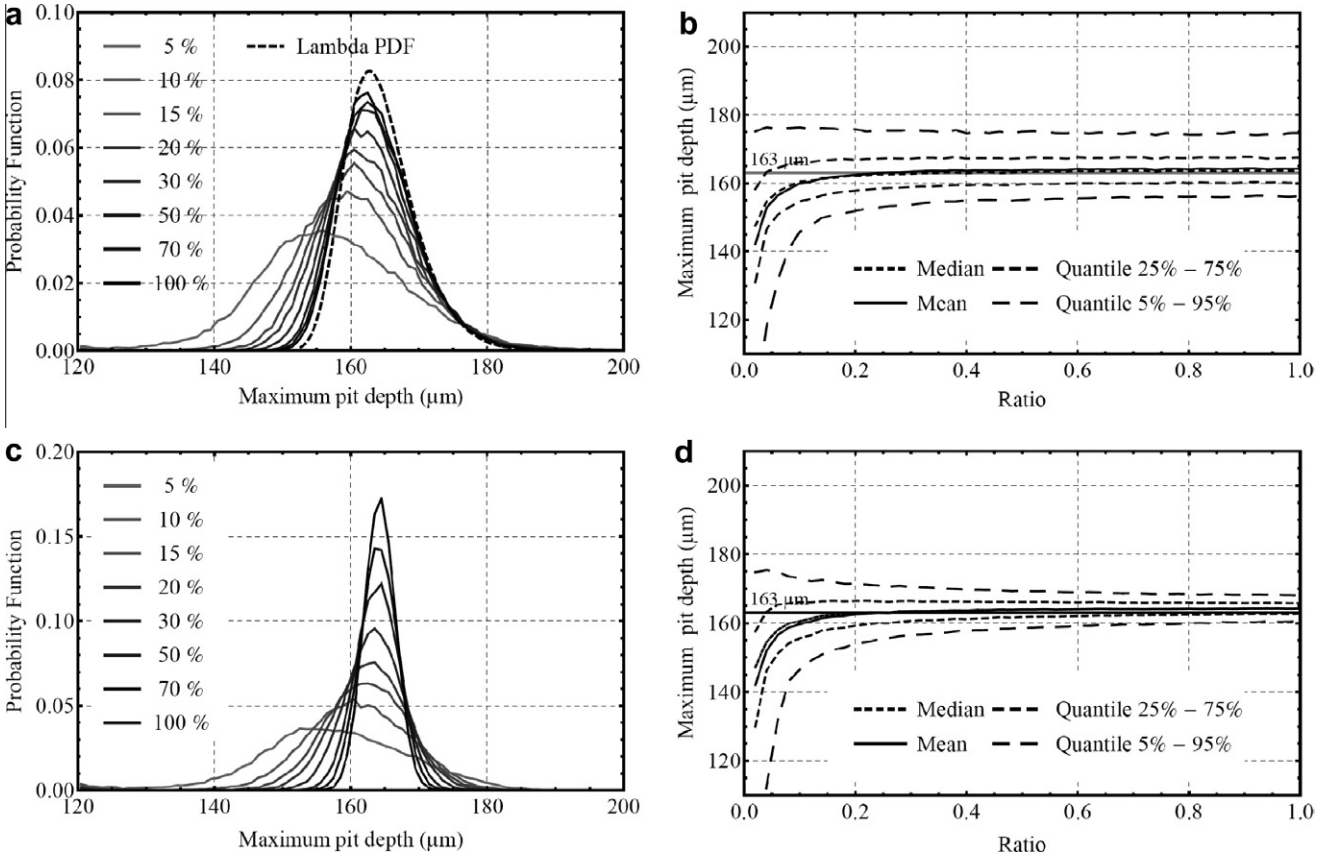

Fig. 6. Estimated PDFs from GLD combined with Monte-Carlo simulations (603 values) corresponding to the maximum pit depth realizations (a) and its mean (c) on area A_0 for various analysed area ratios. Evolutions of the maximum pit depth realizations (b) and its mean (d) as a function of the area ratio for the mean, the median, the 5th, 25th, 75th and 95th quantiles.

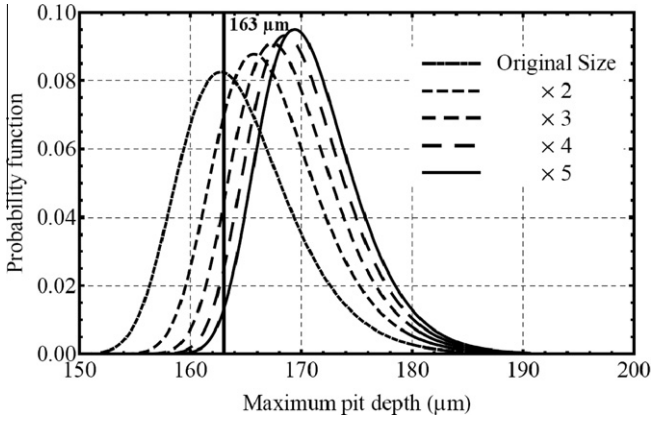

Fig. 7. PDFs of the maximum pit depth for larger exposed sample surface sizes. A surface coefficient of five corresponds to a surface size (375 cm^2) equal to five times that of the experimental aluminium sample (A_0 = 75 cm^2).

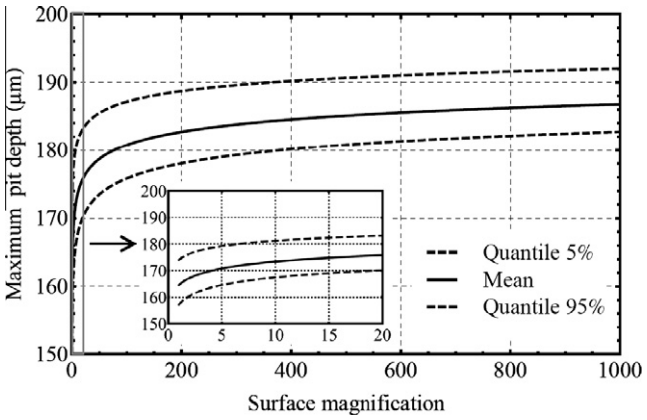

Fig. 8. Influence of the exposed sample surface size on the mean, 5th quantile and 95th quantile of the maximum pit depth. A surface coefficient of 1000 corresponds to a surface size (7.5 m^2) equal to 1000 times that of the experimental aluminium sample (A_0 = 75 cm^2).

$$Q_{\max}(v) = \lambda_1 + \frac{(v^{1/n})^{\lambda_3} - (1 - v^{1/n})^{\lambda_4}}{\lambda_2} \tag{9}$$

and the density function, f, corresponding to any GLD, Q, is:

$$f(x) = \frac{1}{Q'(F(x))} \tag{10}$$

where Q' is the derivative function of the Q function and F the cumulative distribution function (CDF) of the Q function (*i.e.* $F = Q^{-1}$). Therefore the $F_{\max}$ function can then be calculated for any depth value x by numerically inversing the $Q_{\max}$ function as well as the density function $f_{\max}$. Mathematica™ software has been used for this purpose. Fig. 5a presents the evolution of this density function. Both the mean and the 90% confidence interval (*i.e.* the difference between the 95th and the 5th quantiles) can then easily be determined from this density function to assess respectively the central tendency and the dispersion of the maximum pit depth. The calculated mean of the pit depth extreme values is 165 µm and 90% of these values lie between 157 µm and 174 µm.

Besides, it could be interesting to repeat this methodology considering the most comment parent laws (Gaussian, Lognormal and Weibull) used in literature to model pitting corrosion. This step is obviously of major interest to test the relevance of a GLD model to predict maximum pit depth in comparison to those most comment statistical laws. Fig. 5b shows the maximum pit depth distributions for the four previous laws and Table 4 summarizes the associated main estimates. As can be observed, the mean extracted from the GLD estimates very well the observed experimental maximum pit depth. On the contrary, the use of Weibull law leads to an underestimation of this experimental maximum pit depth whereas the use of Normal and Lognormal leads to an overestimation. Besides, the GLD allows a better estimation of the skewness and the kurtosis of the distribution that control the extreme value region. These results clearly emphasize the flexibility and the relevance of a GLD model to predict maximum pit depth in comparison to statistical laws classically used in literature to model pitting corrosion.

5. Comparison of the GLD method to the usual extreme values approaches

5.1. Validation of the relevance of a GLD modelling method

To validate our methodology, the following simulation has been carried out. An area, A, is randomly chosen onto the whole surface of the sample. This enables to define an area ratio, $r = A/A_0$, regarding the whole sample area, A_0 equals to 75 cm^2. The area A contains a subset of size p of the whole 603 pits with $p \leqslant 603$. The hypothesis is made that pits randomly nucleate at the surface of the aluminium sheet. Consequently p is the realization of a Poisson distribution of parameter ρA, where ρ is the mean number of pits per unit area (*i.e.* 8.04 pits/cm^2 considering our experimental data). Thus we can retain a subset of pits whose size follows a Poisson distribution in order to consider the fraction r of the whole surface. For each area ratio, r, and corresponding subset, p, a GLD is computed to represent pit depth evolution. The maximum pit depth of the surface is then estimated using a Monte-Carlo procedure. This corresponds to estimate the highest realization of the GLD (*i.e.* the highest values out a set of 603 realizations). We also computed the mean $\mu_{\max}$ of this highest value for the each GLD set of parameters (λ_1, λ_2, λ_3, λ_4) such that:

$$\mu_{max} = \int_0^1 nt^{n-1} \quad \lambda_1 + \frac{t^{\lambda_3} - (1-t)^{\lambda_4}}{\lambda_2} \Bigg) dt \tag{11}$$

or

$$\mu_{\max} = \lambda_1 + \frac{n}{\lambda_2(n+\lambda_3)} - \frac{n\Gamma(n)\Gamma(1+\lambda_4)}{\lambda_2\Gamma(1+n+\lambda_4)} \tag{12}$$

where n is the total number of pits equal to 603 and Γ (.) is the gamma Euler function.

For each of the selected area ratios corresponding to a partial analysis of the whole area, the CBBM has been performed to generate 50,000 bootstrap samples resulting from random sampling, and to compute the estimated PDFs of the maximum pit depth for area A_0 (Fig. 6a) and its mean (Fig. 6c). This latter value (Eq. (12)) can be considered as the best estimation of the mean of the maximum pit depth which is usually unknown. The curves in Fig. 6b and d represent respectively the continuous evolutions of the main statistics of the maximum pit depth realizations and mean of this maximum for each estimated GLD set of parameters. Each of the selected ratios for Fig. 6b and d corresponds to 5000 bootstrap simulations performed in order to estimate the mean, the median and quantiles. If the GLD model is relevant, it is expected that, on average, the predicted maximum pit depth value is closer to the experimental one (163 μm) for a sufficient ratio.

As can be observed in Fig. 6a, these PDFs converge towards the PDF determined from the total dataset $p = 603$ (called Lambda PDF). It should be observed in Fig. 6b that the maximum pit depth realizations are still scattered around the mean values even for an area ratio of 100%, for which the 90% confidence interval is [156 μm–175 μm]. It should be also noticed that this confidence interval is quite stable as soon as the area ratio is higher than 20%. However it should be emphasized that, in practical cases, statistical extreme value analysis is used to predict the maximum pit depth that will be found in a large scale installation by using a small number of samples with a small area. Thus Fig. 6b and d, as well as the following ones, should mainly be analyzed on the left

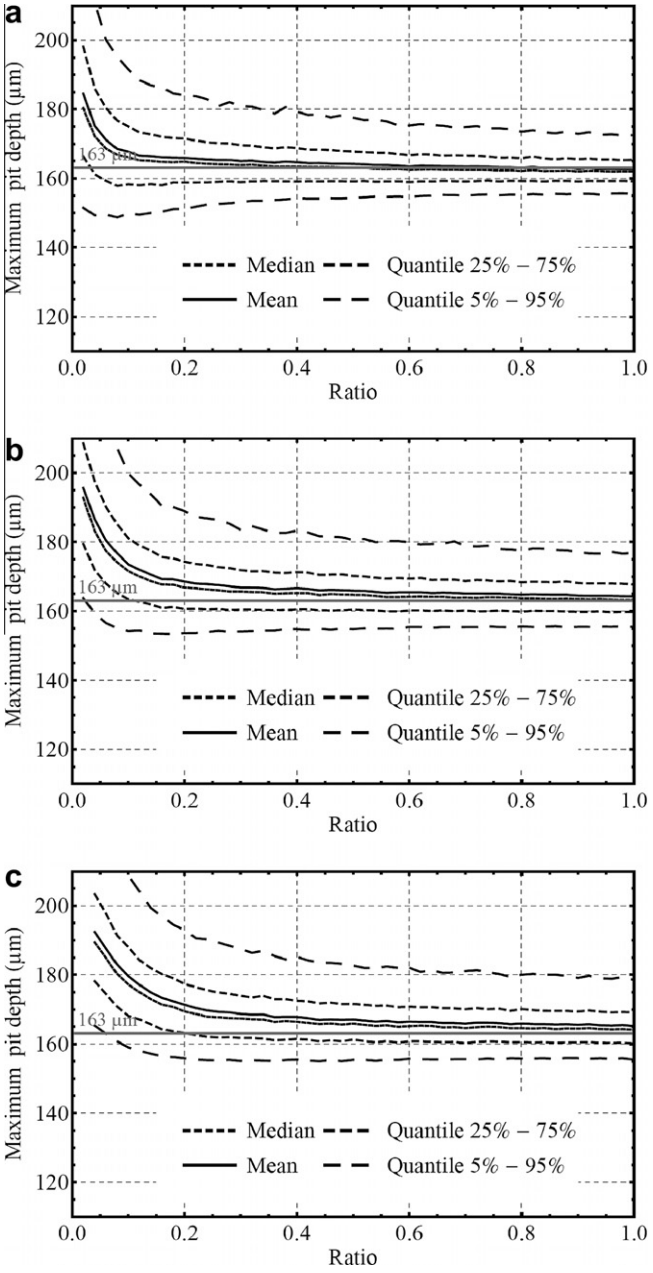

Fig. 9. Evolution of the maximum pit depth realizations on the whole sample as a function of the area ratio for the mean, the median, the 5th, 25th, 75th and 95th quantiles considering the Gumbel approach. The selected area has been divided into 5 (a), 10 (b) and 15 (c) samples in order to estimates parameters μ and σ.

side. Particularly, results corresponding to area ratios higher than 50% should not be considered as practical cases [12].

As shown in Fig. 6d, the visual convergence can also be quantitatively confirmed by the mean, the median, the 5th, 25th, 75th and 95th quantiles curves associated to the overall PDFs. This figure shows that the mean of the GLD modelling process provides an accurate estimation of the observed experimental maximum pit depth of 163 μm considering only the data collected on small fractions of the inspected surface for the statistical treatment. This emphasizes the relevance of GLD modelling for pitting corrosion and validates our methodology. In Fig. 6d, the 90% confidence interval is [154 μm–171 μm] for an area ratio of 20%. The width of this confidence interval decreases for higher area ratios. Nevertheless, it should be emphasized that, even if a good prediction of the mean of the maximum pit depth is done, a large confidence interval on their related realizations can still be observed.

The extreme value is often used to predict the maximum pit depth on an area larger than the inspected surface named A_0 in this investigation. Considering Eqs. (9) and (10), the density function of this maximum pit depth can be calculated considering that each one follows the previously defined GLD model. Fig. 7 represents the PDFs of the maximum pit depth on areas that are 2, 3, 4 and 5 times larger than the original one. Maximum pit depth continuously increases above the initial maximum pit depth of 163 μm with the exposed surface area. Fig. 8 shows the mean of the maximum pit depths on larger areas with magnification varying from 1 to 1000. For example, considering a surface area which is 1000 times larger than that of the experimental sample under study, the mean of the maximum pit depth is 187 μm and 90% of the extreme values lie between 183 μm and 192 μm. It is worth noting that the slope in the increase of the mean of the maximum pit depths is lower for the highest magnifications than for the lowest ones. Consequently the maximum pit depths do not increase much even for the highest magnification. For example, the pit depth increase is only of 20 μm for the 1000 magnification.

5.2. Comparison of the GLD approach to the usual Gumbel approach

As previously mentioned, the most common method for extreme pit depth prediction is found in literature in the application of the Gumbel distribution [44]. In this approach, n small samples with the same area, a, are selected from the whole sheet of area A_0. All the pit depths are measured and the maximum value of each sample is extracted. Considering that the number of pits per

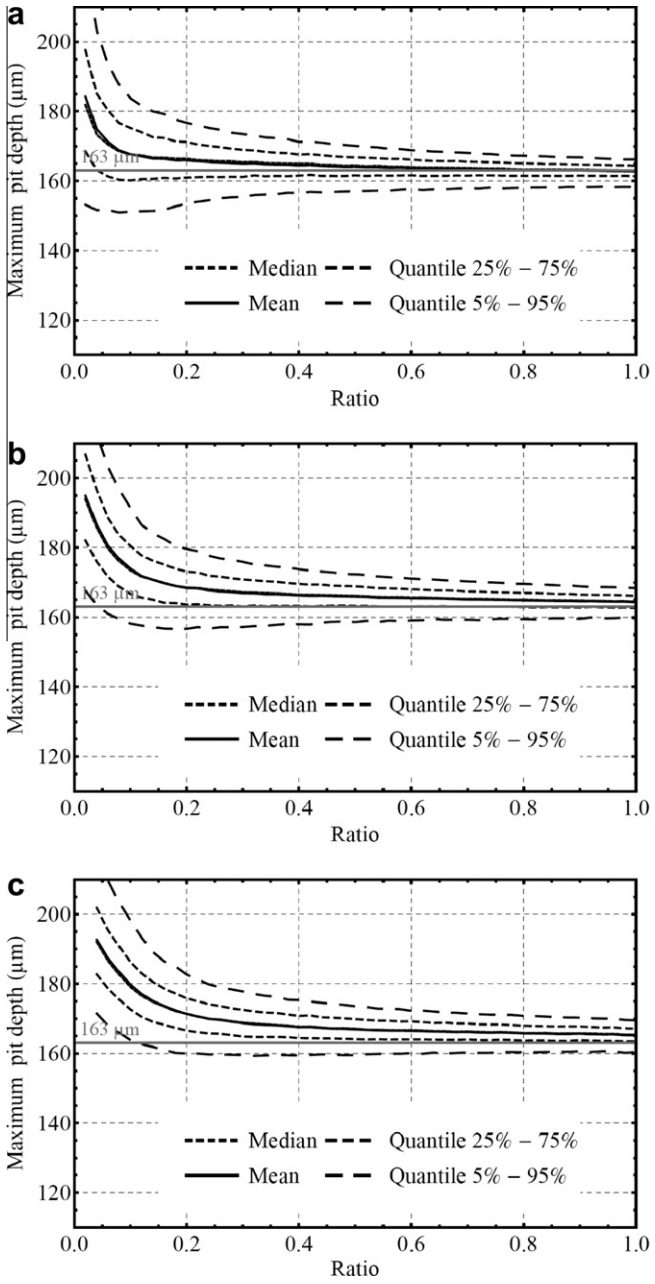

Fig. 10. Evolution of the mean of maximum pit depth on the whole sample as a function of the area ratio for the mean, the median, the 5th, 25th, 75th and 95th quantiles considering the Gumbel approach. The selected area has been divided into 5 (a), 10 (b) and 15 (c) samples in order to estimates parameters.

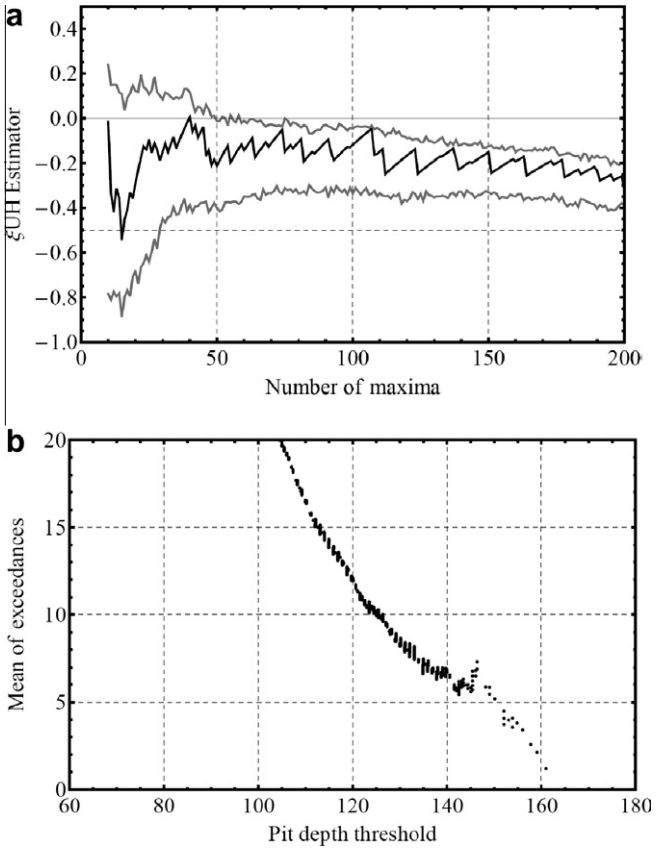

Fig. 11. (a) Evolution of the UH estimator, $\xi_{UH,}$ as a function of the number of maxima. The top and bottom grey lines represent the 90% confidence band. Only the 200 first maxima have been considered. (b) Evolution of the mean of exceedances as a function of the pit depth threshold.

sample is high, the $\{p_i\}_{1\leqslant i\leqslant n}$ maximum pit depths are the realizations of a Gumbel distribution (Eq. (2)) where location, μ, and scale, σ, parameters have to be determined. The method of moments is widely used in order to estimate these parameters. First and second order moments of the Gumbel distribution are compared to the estimated ones from the set of measurements, $\{p_i\}$. The two-equation system is solved leading to:

$$\begin{cases} \sigma = \frac{\sqrt{6}}{\pi} s \\ \mu = m - \sigma\gamma \end{cases} \tag{13}$$

where m and s are respectively the mean and the standard deviation of the $\{p_i\}$ values. γ is the Euler's constant ($\gamma \approx 0.577216$). These estimations enable to predict the maximum pit depth. Indeed, considering the CDF (Eq. (2)) of maximum pit depth distribution on each small sample, the maximum pit depth CDF on the whole sheet is:

$$F_{maxX}(x) = \exp\left(-T\exp\left(-\frac{x-\mu}{\sigma}\right)\right) \tag{14}$$

where T is the return period defined by:

$$T = \frac{A_0}{a} \tag{15}$$

This CDF and the estimated parameters, μ and σ, enable to generate a realization of the maximum pit depth process on area A_0. The mean of the distribution corresponds to the expected value of the mean of the maximum pit depth, $p_{\max}$, such that:

$$p_{max} = \mu + (\gamma + \log(T))\sigma \tag{16}$$

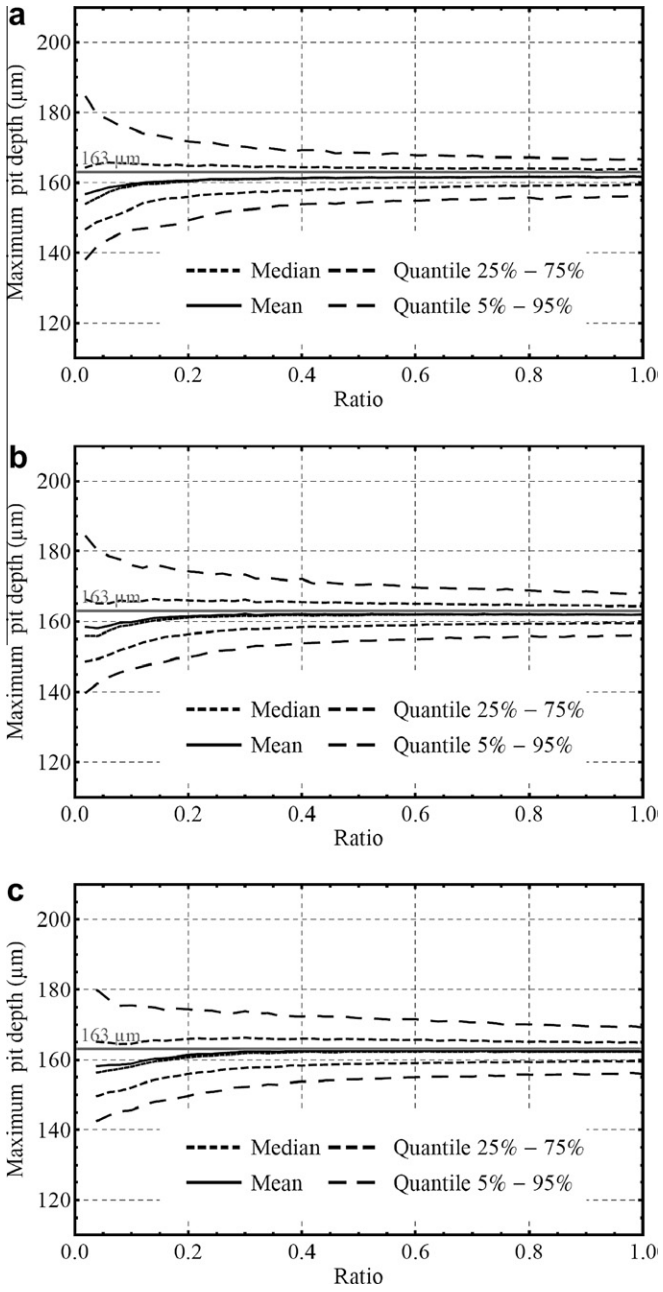

Fig. 12. Evolution of the maximum pit depth realizations on the whole sample as a function of the area ratio for the mean, the median, the 5th, 25th, 75th and 95th quantiles considering the GEV approach. The selected area ratio has been divided into 5 (a), 10 (b) and 15 (c) samples in order to estimates parameters μ, σ and ξ.

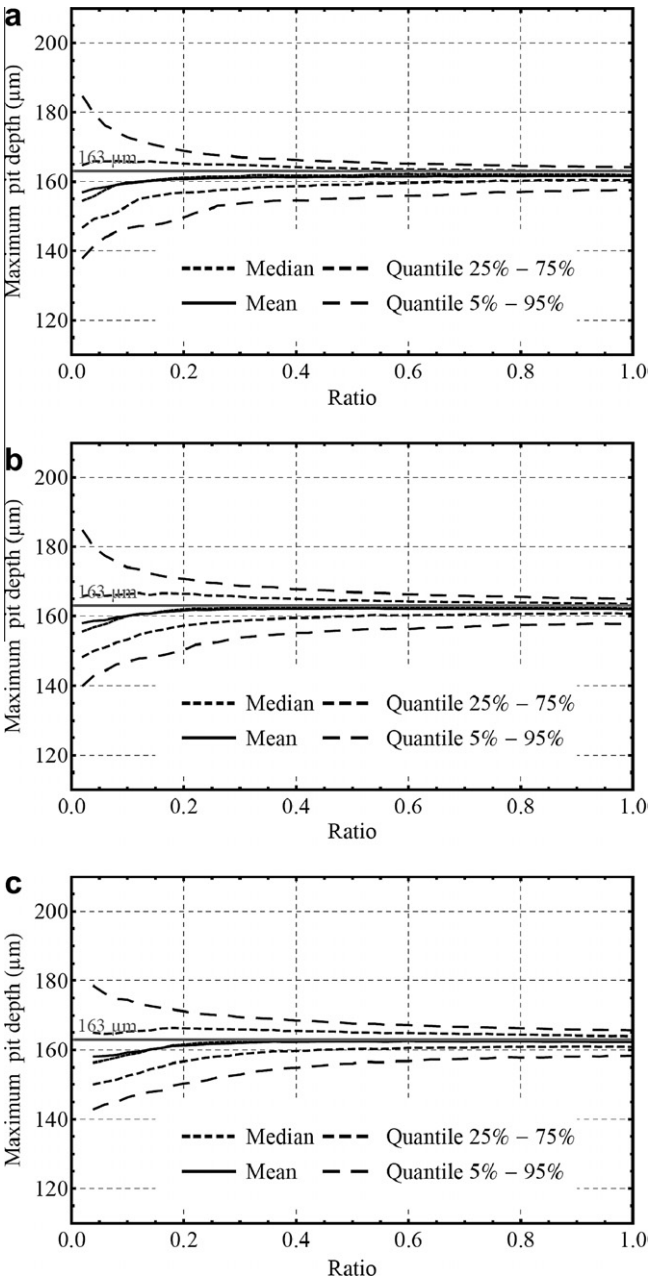

Fig. 13. Evolution of the mean of maximum pit depth on the whole sample as a function of the area ratio for the mean, the median, the 5th, 25th, 75th and 95th quantiles considering the GEV approach. The selected area ratio has been divided into 5 (a), 10 (b) and 15 (c) samples in order to estimates parameters μ, σ and ξ.

This approach has been developed in order to compare the maximum pit depth realizations (Fig. 9) and its mean estimation (Fig. 10) to the experimental ones and to GLD predictions (Fig. 6b and d) using the same number of bootstrap simulations. In order to avoid discussion on the optimal number of samples n that remains a subject of debate, three values have been chosen: 5, 10 and 15. It should be emphasized that the division with 15 samples corresponds to a small analysed surfaces. The small number of pits per area, in particular for the small ratio, involves a small number of measured pit depths. Consequently this situation does not correspond to valid application conditions of the Gumbel theory.

Firstly, Figs. 6b and 9 show that the pit depth realizations are less scattered with the GLD approach than with the Gumbel one, especially for the small area ratios (*i.e.* small surface analysis). Considering an area ratio of 20%, the 50% confidence interval is around 9 μm for the GLD approach (Fig. 6b) and 12 μm, 13 μm and 15 μm for 5, 10 and 15 analysed surfaces respectively in the case of the Gumbel approach (Fig. 9). Secondly, Figs. 6d and 10 show that the maximum pit depth means are also less scattered for GLD approach than the Gumbel one, especially for the small area ratios. For an area ratio of 20%, the 50% confidence interval is around 7 μm for the GLD approach (Fig. 6d) and 10 μm, 9 μm and 9 μm for 5, 10 and 15 analysed surfaces in the case of the Gumbel approach (Fig. 10). Thirdly, more generally, estimations presented in Fig. 6b and d are centred on the maximum pit depth measurement contrary to estimations in Figs. 9 and 10 where the maximum pit depth is estimated by excess in comparison to the experimental value. Contrary to the GLD approach, a bias is noticeable both in Figs. 9 and 10 whatever the number of analysed surfaces. Besides, it should be noticed that the difference between the confidence intervals for the GLD and the Gumbel approaches will increase while decreasing the area ratio.

These results may firstly be explained by the fact that pit depth PDF is not in the Gumbel attraction domain. Indeed, the parameter λ_4 of the GLD was found to be equal to 0.086 and implies that the PDF is bounded on the right hand side (Appendix A) and the pit depth should therefore not exceed an upper bound of 215 μm (*i.e.* $\lambda_1 + 1/\lambda_2$). Moreover, Fig. 3 shows that the confidence band of the estimation of this parameter is strictly positive. Thus the pit depth distribution could not be in the Gumbel attraction domain which is unbounded and may only be in the Weibull attraction domain. This attraction domain of the pit depth distribution is confirmed when looking at Fig. 11a presenting the evolution of the UH estimator of Beirlant et al. [14,49], ξ_{UH}, as a function of the number of maxima, k, which corresponds to the selected number of greatest pit depths. The confidence interval presented on this figure has been calculated by means of a bootstrap method. The mean of the UH estimator evolves in the negative domain of the graph whatever the number of extrema. Moreover the confidence interval is essentially in the same negative domain, especially for a number of extrema higher than 50, when ξ_{UH} slowly evolves. In the same way, the negative sign of the ξ parameter is confirmed considering the method based on the analysis of mean of exceedances as proposed by Laycock and Scarf [8]. The mean of exceedances has been plotted as a function of the selected pit depth threshold in Fig. 11b. This plot should follow a straight line with a slope such that:

$$a = \frac{\xi}{1-\xi} \tag{17}$$

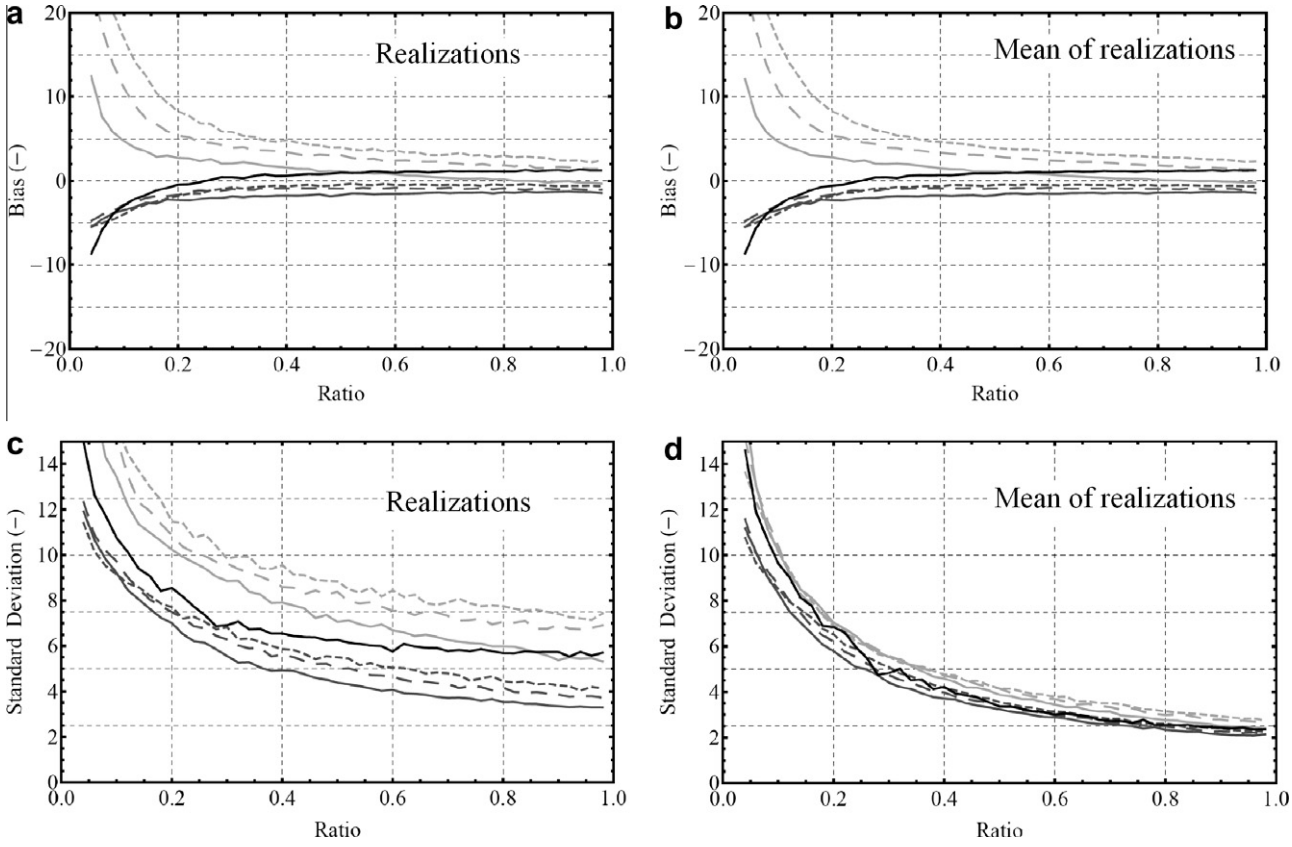

Fig. 14. Bias (a, b) and standard deviation (c, d) in the estimation of realization (a, c) and mean (b, d) of maximum pit depth for the various approaches: Gumbel (light gray lines), GEV (dark gray lines) and GLD (black lines). For the two first approaches, full lines, large dotted lines and small dotted lines represent respectively the various numbers of samples (5, 10 and 15) contained in the analyzed surface corresponding to the ratio of the whole surface.

The shape parameter ξ is equal to -0.56 for a selected threshold of 147 μm pit depth corresponding to a quite linear shape of the plot. This value is still negative and confirms that the pit depth distribution is in the Weibull attraction domain.

5.3. Comparison of the GLD and GEV approaches

Considering the previous result, it is proposed to estimate also the maximum pit depth using the GEV approach. The same procedure of division is used as for the Gumbel approach and the three GEV parameters are calculated using also the method of moments. First, second and third order moments of the GEV distribution are compared to the estimated ones from the measurements set, $\{p_i\}$, and a three-equation system is solved. The resolution gives the location parameter, μ, the scale parameter, σ and the shape parameter, ξ. These estimations enable to predict a realization of the maximum pit depth using Eq. (1) and the mean of this maximum pit depth considering the following expression:

$$p_{max} = \mu + \sigma \left(\frac{T^{\xi}\Gamma(1-\xi) - 1}{\xi} \right) \tag{18}$$

This expression can obviously be compared to Eq. (16) when ξ tends to 0.

Figs. 12 and 13 show respectively the evolution of the maximum pit depth realizations and their means obtained using 5000 bootstrap simulations as for the GLD (Fig. 6) and Gumbel (Figs. 9 and 10) approaches. Fig. 14 shows the evolutions of bias and standard deviation for the various approaches: GLD, Gumbel (5, 10 and 15 samples) and GEV (5, 10 and 15 samples). This figure summarizes the previous results and enables to compare the relevance of the different approaches to estimate the mean of maximum pit depths. In the region of interest corresponding to practical cases (area ratios lower than 50%), the Gumbel approach is shown to predict the maximum pit depth with a higher bias than the GLD and GEV approaches both for the realizations (Fig. 14a) and their means (Fig. 14b). This result can be explained considering the attraction domain of the pit depth distribution. Indeed, it has been shown that this distribution is in the Weibull attraction domain and that the maximum pit depth distribution is bounded. Moreover the same conclusion can stand for the standard deviation of the realizations of the maximum pit depth in the same region of interest (Fig. 14c) whereas approaches give similar results for its mean (Fig. 14d). Furthermore, Fig. 14 shows that the GEV and GLD approaches give similar estimations of the maximum pit depth in terms of bias and standard deviation both for the realization and its mean in this region providing that the analysed ratios are higher than 10%. The value of the bias is lower than 3 μm both for the realizations and their means. For an area ratio equal to 10%, the value of standard deviations of the realizations of the maximum pit depth and of their means is around 8 μm. For an area ratio equal to 50%, this value is equal to 6 μm in the case of the GLD approach and 5 μm in the case of the GEV approach for the realizations of the maximum pit depth. For their means, the standard deviation is equal to 3 μm both for the GLD and GEV approaches.

In summary, looking at these results related to the statistical treatment of the experimental data set considered in the present investigation, it can be claimed that there is no significant difference between GLD and GEV approaches in the region of practical interest. However, whatever the approach, an upper bound can be defined and this bound corresponds to the limit value which cannot be exceeded even for an infinite sized surface (*i.e.* considering an infinite set of pits). It is worth noting that this upper bound cannot be defined in the cases where the λ_4 parameter is found negative in the case of a GLD modelling or the shape parameter ξ is found positive in the case of a GEV modelling. Except these situations, the upper bound is defined by:

$$\begin{cases} \lambda_1 + \frac{1}{\lambda_2} & \text{for the GLD} \\ \mu - \frac{\sigma}{\xi} & \text{for the GEV} \end{cases} \tag{19}$$

Previous statistical treatments of our experimental data shown that the pit depth distribution was bounded whatever the approach. Fig. 15 shows the evolution of the upper bound value as a function of the area ratio for these two approaches considering 20,000 bootstrap simulations. It should be mentioned that the aforementioned situations where the calculated λ_4 parameter was negative (GLD modelling) or the shape parameter ξ was positive (GEV modelling) were excluded to represent this figure. Moreover, upper bound estimations corresponding to pit deeper than 10 mm which can be considered as an unrealistic pit depth value with regard to the experimental conditions were also excluded. These two conditions lead to define a failure proportion corresponding to situations where a realistic pit depth upper bound value cannot be estimated. Fig. 15a shows that, for area ratios higher than 10%, this failure proportion is significantly lower in the case of a GLD modelling than in the case of GEV one whatever the set of samples. Indeed, while this failure proportion is around few percents for the later method, it is equal to 0 for any area ratio higher than 20% for the former one. In other words, a GLD modelling enables a better prediction of the attraction domain of pit depth distribution and of realistic upper bound. Moreover Fig. 15b shows that the mean values of the upper bound predicted with the GLD

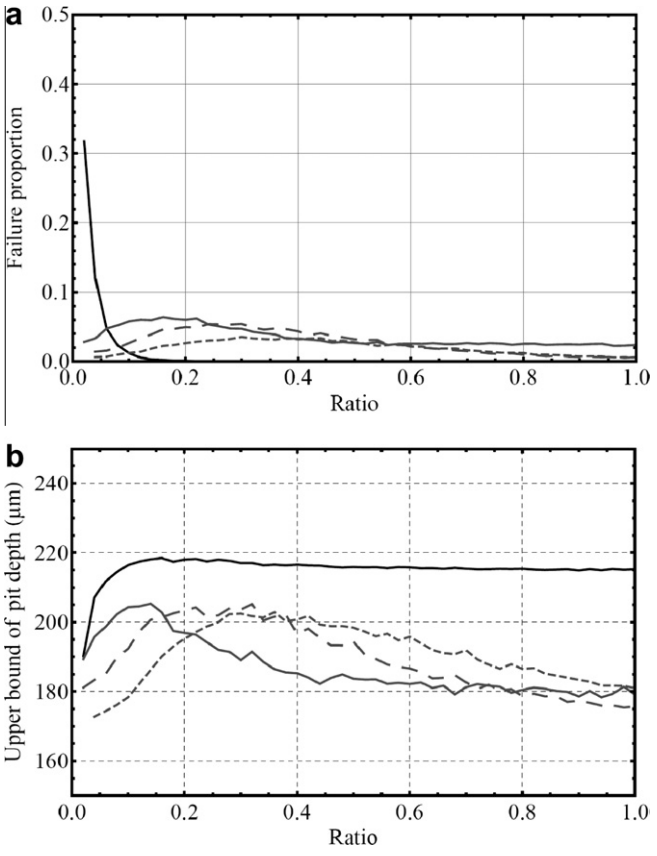

Fig. 15. (a) Proportion of failures to calculate the mean upper bound as a function of the area ratio. (b) Mean upper bound evolution estimated by mean of GLD (black lines) and GEV (dark gray lines) approaches as a function of the area ratio. For the GEV approach, full lines, large dotted lines and small dotted lines represent respectively the various numbers of samples (5, 10 and 15) contained in the analyzed surface corresponding to the ratio of the whole surface.

approach are higher than those predicted with the GEV one whatever the set of samples (5, 10 or 15). While the GEV approach predicts an upper bound closer to the experimental maximum pit depth estimated on the surface specimen, the GLD approach provides a safety margin corresponding to a higher upper bound prediction. Moreover this predicted upper bound shows a steady value around 217 μm for area ratios higher than 10%. In comparison, the GEV approach shows a decreasing value of this upper bound for the higher area ratios.

6. Conclusion

Because of the limitations of the Gumbel approach outlined in this paper, a new and alternative methodology was presented in order to estimate accurately the maximum depth of pits that propagated on an aluminium sheet during an accelerated corrosion test. It was shown statistically, both for the overall pit depth distribution and its maximum value, that a fitting of our experimental data by a Generalized Lambda Distribution (GLD) is more relevant than one carried out by the most comment laws (which belong to the Gumbel attraction domain) used in corrosion engineering literature.

Consequently, the GLD fitting was combined to the CBBM to simulate distributions of corrosion pit depths equivalent to the experimental one to predict the mean of the maximum pit depth as well as its 90% confidence interval that can be found in an installation having a larger surface than the inspected one. For analyzed areas larger than 20% of the whole surface, it is shown that estimations of variability in maximum pit depth realizations and means are quite stable. Therefore such a ratio is proposed in order to estimate maximum pit depth in the case of our experimental data set. Comparisons of maximum pit depth estimations show the superiority of the GLD approach with regards to the Gumbel one. The higher biases and wider confidence intervals estimated in the later case have been explained by the attraction domain of the pit depth distributions which is not in the Gumbel one. Indeed, it was shown that the experimental pit depth distribution considered in the present investigation belongs to the Weibull attraction domain. On the contrary, no significant difference was shown between GLD and GEV approaches. Indeed, whatever the approach, estimated biases and standard deviations are similar for area ratios smaller than 50% which correspond to practical cases with regards to the use of the extreme value statistics. However, as far as the estimation of the upper bound is concerned, the maximum pit depth value is more frequently estimated correctly using the GLD approach for area ratios higher than 10%. It should be emphasized that for are a ratio higher than 20%, proportion cases of failure in the upper bound estimation are extremely scarce in the case of the GLD approach. Moreover the upper bounds are higher than those estimated by the GEV approach which corresponds therefore to more safety situations in practice.

Contrary to the Gumbel and GEV approaches based on the use of inferred parent distributions constrained by specific mathematical assumptions that need to be verified to be safety applied, both the GLD modelling and the CBBM present the main advantage of avoiding to make any preconceived choice neither on the unknown theoretical parent underlying distribution of pit depth which characterizes the global corrosion phenomenon nor on the unknown associated theoretical extreme value distribution which characterizes the deepest pits. Moreover the GLD approach does not require making an a priori choice in the division of the surface and the number of selected samples which is still a subject of debate in the literature. Furthermore, it should be emphasized that the GLD are highly flexible because of their ability to fit, with a high degree of accuracy and within one distributional form, a large variety of shapes; not only in the central region of the considered distributions but also in their left and right tails. As far as the corrosion context is concerned, a high degree of accuracy is required in the right tail region and a GLD model is therefore well suited to provide an analytical expression of maximum pit depth distribution and mean of this distribution.

Considering industrial applications, this alternative approach based on the use of GLD can be applied to determine a minimum thickness of a piece of material in order to prevent its perforation by localized corrosion with a safety margin. Moreover, the alternative methodology presented in this paper could be also applied for pits generated with any type of material and any type of environment leading to localized corrosion since it only necessitates the knowledge of the distribution of a morphological feature (*i.e.* pit depth). Such a methodology, which is a contribution to perform proper maintenance from limited inspection data in the field of corrosion, can also be extended to the study of any kind of defects like inclusions, pores or cracks that may exist in engineering materials and conduct to fatigue failure.

Acknowledgment

The authors would like to thank V. Hague for her help in English.

Appendix A

The GLD expression and conditions lead to the specification of six regions of the (λ_3, λ_4) space [27,29,30]:

Region 1 = $\{(\lambda_3, \lambda_4) \mid \lambda_3 \leqslant -1, \lambda_4 \geqslant 1\}$,
Region 2 = $\{(\lambda_3, \lambda_4) \mid \lambda_3 \geqslant 1, \lambda_4 \leqslant -1\}$,
Region 3 = $\{(\lambda_3, \lambda_4) \mid \lambda_3 \geqslant 0, \lambda_4 \geqslant 0\}$,
Region 4 = $\{(\lambda_3, \lambda_4) \mid \lambda_3 \leqslant 0, \lambda_4 \leqslant 0\}$,
Region 5 = $\{(\lambda_3, \lambda_4) \mid -1 < \lambda_3 < 0, \quad \lambda_4 > 1, \quad \frac{(1-\lambda_3)^{1-\lambda_3}}{(\lambda_4-\lambda_3)^{\lambda_4-\lambda_3}} (\lambda_4 - 1)^{\lambda_4 - 1} < -\frac{\lambda_3}{\lambda_4}\}$,
Region 6 = $\left\{(\lambda_3, \lambda_4) \mid \lambda_3 > 1, \ -1 < \lambda_4 < 0, \ \frac{(1-\lambda_4)^{1-\lambda_4}}{(\lambda_3-\lambda_4)^{\lambda_3-\lambda_4}} (\lambda_3 - 1)^{\lambda_3 - 1} < -\frac{\lambda_4}{\lambda_3}\right\}$.

The GLD support can either be bounded or unbounded as a function of the signs of λ_3 and λ_4, as presented in the following table [27,29,30]:

λ_3	λ_4	Support
$\lambda_3 < -1$	$\lambda_4 > 1$	$]-\infty, \lambda_1 + 1/\lambda_2]$
$\lambda_3 > 1$	$\lambda_4 < -1$	$[\lambda_1 - 1/\lambda_2, \infty[$
$1- < \lambda_3 < 0$	$\lambda_4 > 1$	$]-\infty, \lambda_1 + 1/\lambda_2]$
$\lambda_3 > 1$	$-1 < \lambda_4 < 0$	$[\lambda_1 - 1/\lambda_2, \infty[$
$\lambda_3 > 0$	$\lambda_4 > 0$	$[\lambda_1 - 1/\lambda_2, \lambda_1 + 1/\lambda_2]$
$\lambda_3 > 0$	$\lambda_4 = 0$	$[\lambda_1 - 1/\lambda_2, \lambda_1]$
$\lambda_3 = 0$	$\lambda_4 > 0$	$[\lambda_1, \lambda_1 + 1/\lambda_2]$
$\lambda_3 < 0$	$\lambda_4 < 0$	$]-\infty, \infty[$
$\lambda_3 < 0$	$\lambda_4 = 0$	$]-\infty, \lambda_1]$
$\lambda_3 = 0$	$\lambda_4 < 0$	$[\lambda_1, \infty[$

Appendix B

The method of moments is performed in order to estimate the λ_i parameters of the GLD. This method consists in solving the four equations system of the four moments and their estimates. The analytical expressions of the estimated moments are:

$$\hat{\alpha}_1 = \sum_{i=1}^{n} x_i / n \tag{B.1}$$

$$\hat{\alpha}_2 = \sum_{i=1}^{n}(x_i - \hat{\alpha}_1)^2/n \tag{B.2}$$

$$\hat{\alpha}_3 = \sum_{i=1}^{n}(x_i - \hat{\alpha}_1)^3/n\hat{\alpha}_2^{3/2} \tag{B.3}$$

$$\hat{\alpha}_4 = \sum_{i=1}^{n}(x_i - \hat{\alpha}_1)^4/n\hat{\alpha}_2^2 \tag{B.4}$$

It is shown that if $\lambda_3 > -1/4$ and $\lambda_4 > -1/4$ [27,29,30] then:

$$\alpha_1 = \lambda_1 + \frac{A}{\lambda_2} \tag{B.5}$$

$$\alpha_2 = \sigma^2 = \frac{B - A^2}{\lambda_2^2} \tag{B.6}$$

$$\alpha_3 = \frac{C - 3AB + 2A^3}{\lambda_2^3\alpha_2^{3/2}} \tag{B.7}$$

$$\alpha_4 = \frac{D - 4AC + 6A^2B - 3A^4}{\lambda_2^4\alpha_2^2} \tag{B.8}$$

with

$$A = \frac{1}{1+\lambda_3} - \frac{1}{1+\lambda_4} \tag{B.9}$$

$$B = \frac{1}{1+2\lambda_3} + \frac{1}{1+2\lambda_4} - 2\beta(1+\lambda_3, 1+\lambda_4) \tag{B.10}$$

$$C = \frac{1}{1+3\lambda_3} - \frac{1}{1+3\lambda_4} - 3\beta(1+2\lambda_3, 1+\lambda_4) + 3\beta(1+\lambda_3, 1 + 2\lambda_4) \tag{B.11}$$

$$D = \frac{1}{1+4\lambda_3} + \frac{1}{1+4\lambda_4} - 4\beta(1+3\lambda_3, 1+\lambda_4) + 6\beta(1 + 2\lambda_3, 1+2\lambda_4) - 4\beta(1+\lambda_3, 1+3\lambda_4) \tag{B.12}$$

where

$$\beta(a,b) = \int_0^1 x^{a-1}(1-x)^{b-1}dx \tag{B.13}$$

The α_i moments are estimated by $\hat{\alpha}_i$ (Eqs. (B1–4)). It is necessary to solve a system of equations highly non linear (Eqs. (B5–8)) in order to calculate the four parameters $\lambda_{i,1\leqslant i\leqslant 4}$. As Eqs. (7) and (8) depend only on λ_3 and λ_4 (after simplification), the four equations system become a two equations system with more stable numerical convergence. The values of λ_3 and λ_4, solutions of the two equations systems, are found by a minimization process of the following function using a steepest gradient method:

$$\Psi'(\lambda_3, \lambda_4) = \sum_{i=3}^{4}(\hat{\alpha}_i - \alpha_i)^2 \tag{B.14}$$

After finding λ_3 and λ_4, λ_2 is calculated from Eq. (B6) and finally λ_1 from Eq. (B5).

References

[1] T. Shibata, Evaluation of corrosion failure by extreme value statistics, ISIJ Int. 31 (1991) 115–121.

[2] K.M. Deen, M.A. Virk, R. Ahmad, I.H. Khan, Failure investigation of heat exchanger plates due to pitting, Eng. Fail. Anal. 17 (2010) 886–893.

[3] E. Sato, H. Ito, T. Murata, Statistical approach to life time estimation of steel structures deteriorated by corrosion, Nippon Steel Tech. Rep. 32 (1987) 54–63.

[4] M.G. Stewart, A. Al-Harthy, Pitting corrosion and structural reliability of corroding RC structures: experimental data and probabilistic analysis, Reliab. Eng. Sys. Saf. 93 (2008) 373–382.

[5] S. Yamamoto, T. Sakauchi, An extreme-value statistical analysis of perforation corrosion in the lap joints of automotive body panels, Int. J. Mater. Prod. Tech. 6 (1991) 37–46.

[6] A. Valor, F. Caleyo, D. Rivas, J.M. Hallen, Stochastic approach to pitting-corrosion – extreme modeling in low-carbon steel, Corros. Sci. 52 (2010) 910–915.

[7] R.E. Melchers, Extreme value statistics and long-term marine pitting corrosion of steel, Probab. Eng. Mech. 23 (2008) 482–488.

[8] P.J. Laycock, P.A. Scarf, Exceedances, extremes, extrapolation and order statistics for pits, pitting and other localized corrosion phenomena, Corros. Sci. 35 (1993) 135–145.

[9] D. Rivas, F. Caleyo, A. Valor, J.M. Hallen, Extreme value analysis applied to pitting corrosion experiments in low carbon steel: comparison of block maxima and peak over threshold approaches, Corros. Sci. 50 (2008) 3193–3204.

[10] F. Caleyo, J.C. Velázquez, A. Valor, J.M. Hallen, Probability distribution of pitting corrosion depth and rate in underground pipelines: a Monte Carlo study, Corros. Sci. 51 (2009) 1925–1934.

[11] A. Turnbull, Review of modeling of pit propagation kinetic, Br. Corros. J. 28 (1993) 297–308.

[12] S. Komukai, K. Kasahara, On the requirements for a reasonable extreme value prediction of maximum pits on hot-water-supply copper tubing, J. Res. Nat. Inst. Stand. Technol. 99 (1994) 321–326.

[13] Y. Ishikawa, T. Ozaki, N. Hosaka, O. Nishida, Prediction of localized corrosion damage of some machine components by means of extreme value statistical analysis, ISIJ Int. 22 (1982) 977–983.

[14] A. Jarrah, J.M. Nianga, A. Iost, G. Guillemot, D. Najjar, On the detection of corrosion pit interactions using two-dimensional spectral analysis, Corros. Sci. 52 (2010) 303–313.

[15] A.K. Sheikh, J.K. Boah, D.A. Hansen, Statistical modeling of pitting corrosion and pipeline reliability, Corrosion 46 (1990) 190–196.

[16] A.K. Sheikh, J.K. Boah, M. Younas, Truncated extreme value model pipeline reliability, Reliab. Eng. Sys. Saf. 25 (1989) 1–14.

[17] A. Valor, F. Caleyo, L. Alfonso, D. Rivas, J.M. Hallen, Stochastic modeling of pitting corrosion: a new model for initiation and growth of multiple corrosion pits, Corros. Sci. 49 (2007) 559–579.

[18] A. Valor, D. Rivas, F. Caleyo, J.M. Hallen, Discussion: statistical characterization of pitting corrosion – Part 1: data analysis and Part 2: probabilistic modeling for maximum pit depth, Corrosion 63 (2007) 107–113.

[19] N.J. Laycock, P.J. Laycock, P. Scarf, D.P. Krouse, Applications of statistical analysis techniques in corrosion experimentation, testing, inspection and monitoring, in: R.A. Cottis, M.J. Graham, R. Lindsay, S.B. Lyon, J.A. Richardson, J.D. Scantlebury, F.H. Stott (Eds.), Shreir's Corrosion, fourth ed., Elsevier, 2009, pp. 1547–1580.

[20] P.A. Scarf, R.A. Cottis, P.J. Laycock, Extrapolation of extreme pit depths in space and time using the *r* deepest pit depths, J. Electrochem. Soc. 139 (1992) 2621–2627.

[21] A.L. Fougères, S. Holm, H. Rootzèn, Pitting corrosion: comparison of treatments with extreme-value-distributed responses, Technometrics 48 (2006) 262–272.

[22] M. Kowaka, Introduction to Life Prediction of Industrial Plant Materials: Application to the Extreme Value Statistical Method for Corrosion Analysis, Allerton Press, Inc., New York, 1994.

[23] R.E. Melchers, Representation of uncertainty in maximum depth of marine corrosion pits, Structural Safety 27 (2005) 322–344.

[24] F. Caleyo, J.C. Velázquez, A. Valor, J.M. Hallen, Markov chain modeling of pitting corrosion in underground pipelines, Corros. Sci. 51 (2009) 2197–2207.

[25] G. Engelhardt, D.D. Macdonald, Unification of the deterministic and statistical approaches for predicting localized corrosion damage: I. Theoretical foundation, Corros. Sci. 46 (2004) 2755–2780.

[26] S. Datla, M.I. Jyrkama, M.D. Pandey, Probabilistic modeling of steam generator tube pitting corrosion, Nucl. Eng. Des. 238 (2008) 1771–1778.

[27] B. Fournier, N. Rupin, M. Bigerelle, D. Najjar, A. Iost, R. Wilcox, Estimating the parameters of a generalized lambda distribution, Comput. Stat. Data Anal. 51 (2007) 2813–2835.

[28] D. Najjar, M. Bigerelle, C. Lefebvre, A. Iost, A new approach to predict the pit depth extreme value of a localized corrosion process, ISIJ Int. 43 (2003) 720–725.

[29] Z.A. Karian, E.J. Dudewicz, Fitting Statistical Distributions, The Generalized Lambda Distribution and Generalized Bootstrap Methods, Chapman and Hall, CRC Press, Boca Raton, Florida, 2000.

[30] Z.A. Karian, E.J. Dudewicz, Handbook of Fitting Statistical Distributions with R, Chapman and Hall, CRC Press, Boca Raton, Florida, 2010.

[31] A. Tarsitano, 2004. Fitting the generalized lambda distribution to income data, in: Antoch Jaromir, Proceedings of the 16th Symposium in Computational Statistics (COMPSTAT 2004), Prague, Czech Republic, 2004, pp. 1861–1867.

[32] C.J. Corrado, Option pricing based on the generalized lambda distribution, J. Fut. Mark. 21 (2001) 213–236.

[33] A. Öztürk, R.F. Dale, A study of fitting the generalized lambda distribution to solar radiation data, J. Appl. Meteorol. 21 (1982) 995–1004.

[34] S. Pal, Evaluation of non-normal process capability indices using generalized lambda distributions, Quality Eng. 17 (2005) 77–85.

[35] A. Negiz, A. Cinar, Statistical monitoring of multivariable dynamic processes with state space models, AIChE J. 43 (1997) 2002–2010.

[36] J. Eriksson, J. Karvanen, V. Koivunen, Source distribution adaptive maximum likelihood estimation of ICA model, in: Proceedings of Second International Workshop on Independent Component Analysis and Blind Signal Separation (ICA2000), Helsinki, Finland, 2000, pp. 227–232.
[37] J. Karvanen, J. Eriksson, V. Koivunen, Adaptive score functions for maximum likelihood ICA, J. VLSI Signal Process. 32 (2002) 83–92.
[38] B. Dengiz, The generalized lambda distribution in simulation of M/M/1 queue systems, J. Fac. Eng. Arch. Gazi Univ. 3 (1988) 161–171.
[39] R.R. Wilcox, Comparing the variances of two independent groups, Br. J. Math. Stat. Psychol. 55 (2002) 169–175.
[40] M. Bigerelle, D. Najjar, B. Fournier, N. Rupin, A. Iost, Application of lambda distributions and bootstrap analysis to the prediction of fatigue lifetime and confidence intervals, Int. J. Fatigue 28 (2006) 223–236.
[41] B. Fournier, N. Rupin, M. Bigerelle, D. Najjar, A. Iost, Application of the generalized lambda distributions in a statistical process control methodology, J. Process Control 16 (2006) 1087–1098.
[42] Y. Katano, K. Miyata, K. Shimizu, T. Isogai, Examination of statistical models for pitting on underground pipes and data analysis, in: Proceedings of International Symposium on Plant Aging and Life Predictions of Corrodible Structures, NACE International, Sapporo, 1995, pp. 161–167.
[43] J. Wallace, R. Reddy, D. Pugh, J. Pacheco, Sour service pit growth predictions of carbon steel using extreme value statistics, in: Proceedings of International Symposium on Plant Aging and Life Predictions of Corrodible Structures, NACE International (March 11–15), Nashville, Tennessee, US, 2007, Paper No. 07657.
[44] E.J. Gumbel, Statistical Theory of Extreme Values and Some Practical Applications, Applied Mathematics Series 33, National Bureau of Standards, Washington, DC, USA, 1954.
[45] A. Jarrah, Développement de méthodes statistiques et probabilistes en corrosion par piqûres pour l'estimation de la profondeur maximale: application à l'aluminium A5, Ph.D. Thesis, Arts et Métiers ParisTech, Lille, France, 2009.
[46] T. Shibata, Statistical and stochastic approaches to localized corrosion, Corrosion 52 (1996) 813–830.
[47] B. Efron, Bootstrap methods: another look at the Jackknife, Ann. Stat. 7 (1979) 1–26.
[48] B. Efron, R. Tibshirani, An Introduction to the Bootstrap, Chapman and Hall, New York, USA, 1993.
[49] J. Beirlant, P. Vynckier, J.L. Teugels, Tail index estimation, Pareto quantile plots, and regression diagnostics, J. Am. Stat. Assoc. 91 (1996) 1659–1667.