Corrosion Science 53 (2011) 2453–2467

Contents lists available at ScienceDirect

Corrosion Science

journal homepage: www.elsevier.com/locate/corsci

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

A R T I C L E   I N F O

Article history:
Received 9 June 2010
Accepted 28 March 2011
Available online 1 April 2011

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

A B S T R A C T

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).

0010-938X/$ - see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.corsci.2011.03.026

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