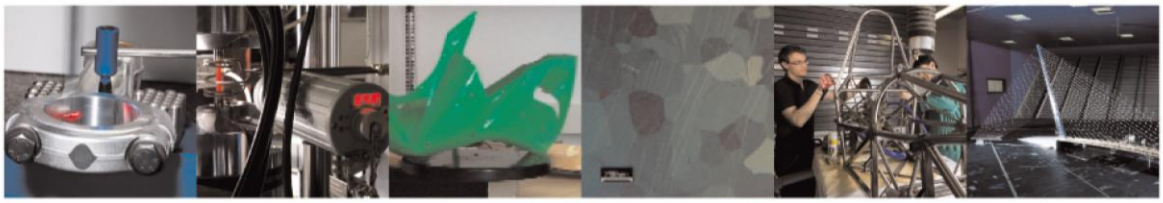
POLITECNICO
MILANO 1863

DIPARTIMENTO DI MECCANICA

mecc



Advanced Ultrasonic “Probability Of Detection” Curves For Designing In-Service Inspection Intervals

Michele Carboni, Stefano Cantini

This is a post-peer-review, pre-copyedit version of an article published in International Journal of Fatigue. The final authenticated version is available online at:
http://dx.doi.org/10.1016/j.ijfatigue.2015.07.018

DIPARTIMENTO DI MECCANICA ■ POLITECNICO DI MILANO
via G. La Masa, 1 ■ 20156 Milano ■ EMAIL (PEC): pecmecc@cert.polimi.it
http://www.mecc.polimi.it
Rev. 0

ADVANCED ULTRASONIC "PROBABILITY OF DETECTION" CURVES FOR DESIGNING IN-SERVICE INSPECTION INTERVALS

Michele CARBONI[1*], Stefano CANTINI[2]

[1] *Department of Mechanical Engineering, Politecnico di Milano, Via La Masa 1, 20156 Milano, Italy*

[2] *Lucchini RS SpA, Via G. Paglia 45, 24065 Lovere (BG), Italy*

Abstract

In order to ensure the safe service of modern railway vehicles, safety critical components are subjected to a dedicated maintenance and inspection plan. Railways axles, in particular, are periodically inspected by the ultrasonic testing method during maintenance interruptions in first level workshops, while the magnetic particles test is carried out in second level workshops. The reliability of such inspections is quantified in terms of "Probability of Detection" curves, which are traditionally related to a specific linear dimension of the defect to be detected. Actually, a fatigue crack is known to change its shape under cyclic loading, so affecting its detectability.

In the present paper, a novel approach to "Probability of Detection" takes into account of this effect, introducing a "Master" Probability of Detection curve, which is a function of the reflecting area of the crack. A comparative applicative example between standard and such advanced Probabilities of Detection curves is then presented considering high-speed railway applications, whose hollow axles are made of the medium strength EA4T steel grade.

Keywords: Ultrasonic testing, Probability of detection, Inspection intervals, Railway axles, Boreprobe equipment

[*] Corresponding author: Tel.: +39-02-23998253, Fax: +39-02-23998202, e-mail: michele.carboni@polimi.it.

1. INTRODUCTION

Considering mechanical components subjected to fatigue, and assuming during service some surface damages might occur and not be promptly repaired, it is licit to expect crack initiation and consequent propagation. To face this problem, some critical safety fields (such as railways, aerospace, automotive, …) employ the "Damage Tolerant" design approach as state-of-the-art. Its philosophy consists ([1]-[2]) in determining the most opportune in-service inspection interval given the "Probability of Detection" (POD) curve ([3]-[5]) of the adopted "Non-Destructive Testing" (NDT) method or, alternatively, in defining the needed NDT specifications given a programmed inspection interval. Structural integrity of safety components during service is then strictly related to the following factors ([1]-[2],[6]): i) the capability of the adopted NDT procedure, i.e. its POD curve; ii) the crack propagation behavior of the adopted material; iii) the influence of the geometry of the cracked body on crack driving force; iv) the reliable knowledge of service loads. An effective damage tolerant approach requires, then, well-defined crack growth lifetime predictions and reliable POD curves.

Focusing here on the capability of NDT, it may be qualitatively defined as "the probability of detecting a crack in a given size group under the specified inspection conditions and procedures" [7]. Even if many similar definitions can be found in the literature, it is well known that it is a statistical matter [7] and that the quantitative and the universally accepted underlying tool is the aforementioned POD curve of the adopted NDT method (a scheme is shown in Figure 1a). This characteristic statistical aspect of NDT derives from the experimental evidence that repeated inspections of the same flaw size or type do not necessarily result in consistent indications: this is the reason for the "realistic" curve shape shown in Figure 1a against the "theoretical" expected one. Such a realistic curve is then usually derived by experimental tests on components containing numerous artificial or natural defects. Moreover, traditionally ([3]-[5]), such probabilities are explicitly expressed and plotted in terms of a characteristic linear dimension of defects (depth, length, diameter, …). This because the depth is known to be the most relevant parameter for fracture mechanical assessment of superficially cracked bodies [2], but, actually, POD curves are also a function of many other physical and operative factors like the adopted NDT method, material, geometry, defect type and shape, equipment, human and environmental factors. This means that very rarely the POD

curve defined for a given inspection procedure can be used for other ones, even if similar. For all these reasons, their definition process is very demanding in terms of time and costs and, over the last years, experimental responses have begun to be partially substituted and integrated by numerical simulations ([8]-[12]).

Two statistical methods are available for analyzing NDT capability data and produce POD curves as functions of the linear flaw size “a”. The first one ([3]-[5]), also chronologically, is based on “hit/miss” data, where NDT results are only recorded in terms of whether the crack is detected or not. The second one ([3]-[5]) is based on the presence of more information within the NDT response, typically in terms of peak voltage in eddy currents NDT, the signal amplitude in ultrasonic NDT or the light intensity in fluorescent penetrant NDT. Since, in this case, the NDT signal response is someway correlated to flaw size, this method is named “signal response” or “â vs. a”, where “a” is the characteristic linear dimension of the defect and “â” its response to the inspection stimulus. The POD curve of the NDT procedure is, in this case, strictly related to the adopted calibration via the decision threshold $\hat{a}_{th}$ (Fig. 1b). Both methods rely on different probabilistic models to produce POD curves, for more details see [4] and [5]. Only the signal response approach is considered and described in the present research.

A very important aspect of POD curves is the need, for reliability and design of components, of a statistical characterization of the largest defect that can be missed and not of the smallest one that can be detected. For this reason, POD curves should always be provided along with a suitable lower confidence limit (typically 95%). Moreover, when dealing with real-life POD curves, it is important to distinguish [13] the intrinsic performance of the equipment from its application to different inspecting procedures and from all human factors affecting calibration and inspection operations (Fig. 1c).

Today, these concepts are also applied to railway axles, safety components designed to have an infinite lifetime [14], but showing occasional failures during service [15]. Such failures always occur, at the most stressed regions, as fatigue crack propagations whose initiation can be due to different causes [1]: for example (Fig. 2), wrong handling or maintenance practice, the presence of widespread corrosion ([16]-[17]) or the possible damage due to ballast impacts. For this reason, to guarantee adequate reliability and safety during service, NDT is performed during both production, in order to detect internal and surface

manufacturing defects, and maintenance, in order to detect in-service surface damages. Specifically, during service, railway axles are periodically inspected by means of ultrasonic testing (UT) at ordinary maintenance service interruptions and by UT and magnetic particles at overhauls. In particular, solid axles are manually inspected by traditional UT probes [18] or, in Italy, by means of a rotating UT probe [19] applied to both ends of the axle and scanning the critical regions (press-fit seats and geometrical transitions). In this case, POD curves are available in the literature ([8],[20]), but more work is still needed to define a generally accepted inspection procedure. Hollow axles are, instead, inspected using a highly automated UT boreprobe roto-translating along the longitudinal bore and scanning the whole external surface, while the availability of POD curves is more meager, an example is given in [21].

The present paper analyses a potential inconsistency regarding the traditional definition and representation of UT POD curves. Even if generally applicable to any UT inspection procedure, such an analysis is carried out considering the specific applicative case of hollow railway axles inspected by the boreprobe. The study is based on an approach recently proposed by the authors ([9],[22]), which is first briefly summarized. Then, a dedicated experimental full-scale fatigue campaign on axles is described: the availability of a large number of natural cracks, propagated under controlled fatigue loads, is the peculiarity of the here-presented research, especially considering this statistical population of natural defects in full-scale axles is currently unique. The derived set of natural defects and the novel approach to POD curves allowed defining those of the boreprobe, in a more flexible way, for its application to the inspection procedure adopted by Lucchini RS during production and maintenance [23]. Finally, some considerations are provided about the effect of the adoption of either traditional or advanced POD curves on the probability of failure of the inspection procedure of hollow axles by the boreprobe.

2. THE “REFLECTING AREA” APPROACH TO ULTRASONIC TESTING DATA

2.1. The traditional derivation of Probability of Detection curves

Considering the signal response approach, a POD(a) function is derived from the correlation of â vs. a data: between the four possible relations required to fulfil the conditions of the POD model [3], an approximate

linear relationship will be demonstrated to exist, for the present applicative case, between $\log_{10}(\hat{a})$ and $\log_{10}(a)$:

$$\log_{10}(\hat{a}) = \alpha + \beta \cdot \log_{10}(a) + \gamma \tag{1}$$

where γ is an error term distributed with zero mean and constant standard deviation σ_γ. Actually, Eq. (1) expresses the fact that $\log_{10}(\hat{a})$ is normally distributed as $N(\mu(a), \sigma_\gamma^2)$, i.e. with mean $\mu(a)=\alpha+\beta\cdot\log_{10}(a)$ and constant standard deviation σ_γ.

Generally, in signal response data, a flaw is regarded as "detected", if â exceeds some pre-defined "decision threshold" $\hat{a}_{th}$ (Fig. 1b) corresponding to the response of the flaw to be detected with 50% probability. The POD(a) function for signal response data can then be expressed as:

$$POD(a) = Pr[\log_{10}(\hat{a}) > \log_{10}(\hat{a}_{th})] \tag{2}$$

which represents the shaded areas shown in Figure 1b. Eq. (2) can be written as:

$$POD(a) = 1 - F\left\{\frac{\log_{10}(\hat{a}_{th}) - [\alpha + \beta \cdot \log_{10}(a)]}{\sigma_\gamma}\right\} = F\left\{\frac{\log_{10}(a) - \left[\frac{\log_{10}(\hat{a}_{th}) - \alpha}{\beta}\right]}{\frac{\sigma_\gamma}{\beta}}\right\} \tag{3}$$

where F is the cumulative log-normal distribution with:

$$\mu(a) = \frac{\log_{10}(\hat{a}_{th}) - \alpha}{\beta} \tag{4}$$

$$\sigma = \frac{\sigma_\gamma}{\beta} \tag{4'}$$

The estimates for α, β and σ_γ can be determined from â vs. a data by means of the maximum likelihood method ([24]-[26]).

2.2. An advanced approach to the analysis of "signal response" data

As mentioned in the Introduction, the traditional approach to POD curves suggests their definition in terms of a characteristic linear dimension “a” of defects (depth, diameter, …) in order to allow fracture mechanics assessments. Considering UT inspection capability, this is potentially inconsistent when dealing with different kind of defects (different causes, different morphologies, …) likely occurring on the same component during service, such as, for example, railway axles (Fig. 2). Previous studies by the authors ([9],[22]) show that a kind of homogenization of UT responses can be achieved analyzing data in terms of the effective reflecting area “A” of defects. This is supported by the intuitive idea the sound beam is physically reflected by an area and not by a linear dimension. The method is also effective because it can keep into account for crack areas and shapes varying during fatigue propagation. For the sake of clarity, a short summary of this approach is provided in the following, while full details can be found in [22].

The considered material is the medium strength EA4T steel grade (quenched and tempered 25CrMo4) typically used for the production of railway axles [14]. Twenty artificial defects were realized on the external surface of six chunks cut from hollow axles. Particularly, such defects (some examples are shown in Figure 3 and compared to real counterparts) were characterized by different geometries (saw-cut, convex, concave), different dimensions (depths ranging from 0.5 mm to 8 mm) and different manufacturing processes (traditional machining and EDM). It is worth remarking saw-cuts are the traditional notches adopted for realizing sample blocks of railway axles, convex defects represent classical fatigue cracks observed at the body or at the transitions and concave defects represent fretting-fatigue cracks usually observed at press-fit seats or scratches due to handling. For simplicity, each artificial defect was then inspected adopting a 2^{nd} leg inspection configuration, i.e. with an intermediate reflection on the bore (Fig. 4a). Actually, this approach was not representative of the common practice, because real UT inspection procedures consider, for hollow axles, only the 1^{st} leg configuration, i.e. a direct interaction between the defect and the sound beam (Fig. 4b), by means of the boreprobe.

UT inspections were also carried out on natural fatigue cracks induced, by means of proper artificial micro-defects, in the body of three full-scale axles (Fig. 4c) made of EA4T grade and fatigue tested by the dedicated bench available at the Dept. of Mechanical Engineering – Politecnico di Milano. Since the loading condition of axles mounted on the full-scale bench can be assimilated to a three point rotating bending, the

shape of natural fatigue cracks is expected to be similar to convex artificial defects, i.e. semi-elliptical or semi-circular. UT inspections were carried out during crack propagation tests so having the possibility to measure evolving cracks changing their dimensions and shape with the increasing number of cycles.
The obtained results are shown in Figures 4d and 4e in terms of depth (â vs. a data) and reflecting area (Â vs. A data), respectively. In the case of depth, the behaviour of fatigue cracks is very different from the one of artificial defects, as often reported in the literature [27]. Considering, instead, the reflecting area, the correlation looks much better, suggesting the reflecting properties of a natural fatigue crack, initiated by classical fatigue loading, are similar to those of artificial defects. It was eventually concluded reflecting area data are more robust than traditional ones.

2.3. A set of naturally fatigue-cracked hollow axles

Since UT inspection of hollow axles is carried out, in a 1st leg configuration, by means of a boreprobe roto-translating along their longitudinal bores, the UT responses, obtained through the experimental campaign shown in Figure 4d and 4e, do not represent the common practice. Moreover, the Master POD curve for the here adopted BAT ("Bore Axle Testing") device, fully described in [28], was not available at that time. For these reasons, a dedicated full-scale fatigue experimental campaign was first required.

Within the present research, in order to make available a significant sample of natural fatigue cracks in railway hollow axles, ten of them were subjected to constant amplitude rotating bending, again through the dedicated full-scale test bench available at the Department of Mechanical Engineering - Politecnico di Milano (Fig. 5a). In this case, axles were made of the high strength 30NiCrMoV12 steel grade, usually adopted for high speed applications and characterised by a longitudinal wave speed V_L=5900 m/s and a shear wave speed V_T=3200 m/s. It is worth remarking the ultrasonic attenuation of this particular grade is very similar to that of EA4T and quantifiable in 9 dB/m. Cracks were induced at the most dangerous geometrical T-transitions [14] between the body and the wheel press-fit seat, where the stress concentration factor K_t is approximately 1.2 and the in-service bending moments are at their maximum value. In order to obtain two different fatigue cracks in each axle, both T-transitions of each of them were prepared by introducing appropriate artificial notches. In particular, the first three axles were prepared by a combination of artificial micro-holes (five adjacent 0.3 mm diameter holes for a total notch length equal to

1.5 mm and depth equal to 0.5 mm) representing possible corrosion pits, while the remaining ones by semi-circular EDM notches (radius equal to 0.3 mm) representing possible impacts due to flying ballast or scratches due to handling. Moreover, all of the artificial notches were inclined, with respect to the longitudinal axis of the axles, of 5.28° so to be perpendicular to the local external profile of axles.

At the end of fatigue tests, which required a total amount of time of about two years, seventeen natural semi-circular fatigue cracks, characterized by a depth ranging from 0.4 to 12 mm, were made available (an example is shown in Fig. 5b).

3. ADVANCED "MASTER" PROBABILITY OF DETECTION CURVES

Advanced "Master" POD curves are defined applying the statistical procedure described in Section 2.1 to Â vs. A data instead of â vs. a ones. The most important feature of this kind of POD curves is, via the adoption of the reflecting area parameter, a prospective independence on the shape of considered cracks or defects. This might be particularly useful because it is well known from the literature [29] that, generally, any initial surface crack shape, subjected to fatigue, evolves towards a semi-elliptical shape. Figures 6a and 6b show [30] a comparative example between crack growth simulations and final crack shapes obtained on small-scale steel specimens subjected to plane bending fatigue and initiated by completely different artificial notch morphologies: two micro-holes representing a semi-circular notch and a shallow slot obtained by EDM. As can be seen, the final crack shapes are indeed morphologically very similar. This means a fatigue crack can change significantly its shape during the life of the component and supports the introduction of more robust approaches, such as the reflecting area, with respect to the traditional one.

Master POD curves in terms of reflecting area allow back calculating those in terms of depth assuming the proper instantaneous shape and value for the reflecting area of a propagating crack under fatigue loading. This is an indispensable feature for interfacing Master POD curves to crack growth algorithms for fracture mechanics assessment, which usually assume crack depth and/or semi surface length as parts of the driving parameters. Figure 7a shows schematically the back calculation procedure using a hypothetical Master POD curve, without its confidence band for the sake of clarity, and assuming the need to derive the traditional ones for semi-circular and shallow defects. Figure 7b, then, shows a validation of the back

calculation procedure based on the real experimental data described in Section 2.3: deriving the traditional POD curve for a semi-circular crack either from experimental acquisitions in terms of depth or by back calculation from the Master POD curve provides almost coincident curves and confidence bands. Similar results are reported in [22]. Figure 7b also shows the Master POD curve does not improve the probability to detect cracks or defects with respect to those in terms of depth, but, as explained above, it has characteristics of generality and versatility not available adopting the traditional ones. By the described process, it is then possible to estimate POD curves also for defect shapes no experimental data are available for: that is why it is possible to consider the POD curve in terms of reflecting area a “Master” (or “Parent”) POD curve.

Another important application of the Master POD curve consists in the possibility to determine the right decision thresholds for different times of flight. Actually, this aspect is today solved already using proper sample blocks where the same defect (typically corresponding to the one whose response has to be adopted as the decision threshold) is located at different times of flight. Figure 8a shows the example of a typical sample block for hollow railway axles: the $\hat{A}$ responses experimentally obtained for each time of flight correspond, then, to the decision thresholds $\hat{A}_{th}$. A more sophisticated and robust approach, again based onto the just described sample block, is implemented in modern boreprobe UT equipment [28] for inspecting railway axles. In particular, the dedicated software of the equipment builds, during the calibration procedure and knowing the drawing of the sample block, “Distance Amplitude Correction” (DAC) curves [31] already adjusted to keep into account for the effect of the time of flight. This procedure is equivalent (Fig. 8b), at least for modestly attenuating materials like steels for axles, to consider only one POD curve, i.e. to consider always the same time of flight, and suitably derive the area of an equivalent defect representing the effect of time of flight. In particular, shorter paths give equivalent defects larger than the calibration one, while, contrarily, longer paths give equivalent defects smaller than the calibration one. The experimental effort for deriving the Master POD curve is then equivalent to that needed for determining just one POD curve in terms of depth.

4. DETERMINATION OF INSPECTION INTERVALS AND NDT PROBABILITIES OF FAILURE USING A MASTER PROBABILITY OF DETECTION CURVE

The damage tolerant algorithm for the definition of maintenance inspection intervals of safety components is widely known in the literature ([1]-[2]) and has been presented by the authors, for the case of railway axles, in different papers such as [6]. What is discussed in the present study has a direct influence on the NDT part of the damage tolerant procedure, which:

1. is necessary to determine the dimension of the defect having a trustful probability to be detected and from which propagation will be considered to start ("Detection limit" in Figure 9);
2. is to be applied at each inspection interval in order to derive the probability of failure of the inspection procedure.

Considering point 2, Figures 9a and 9b show the comparison between the traditional and the proposed calculation procedures, respectively, for POD values at a given inspection N_i during the life (conveniently expressed in cycles, kilometres, hours, …) of the component. For the latter case, the reflecting area of the crack, to be introduced into the Master POD curve, is derived by crack growth simulations, which provide the depth and semi-surface length values for each possible inspection during the in-service life of the component. It is also worth remembering the compensation for the effect of the time of flight can be introduced by suitable DAC curves defined calibrating the equipment, as well.

During service, a well-designed inspection procedure allows inspecting safety components more than once before their final failure. Accordingly, given a length "L" of the inspection interval, the cumulative (total) probability of detection PC_{DET} of a propagating crack can be calculated [6] based on its fatigue crack growth life prediction, a given number of possible inspections "i" and the Master POD curve of the adopted UT procedure:

$$PC_{DET} = 1 - \left[\prod_i POND(A)_i\right] = 1 - \left\{\prod_i [1 - POD(A)]_i\right\} \quad (5)$$

where $POND(A)_i$ ("Probability of Non Detection") represents the probability to fail the i-th detection, equal to $1\text{-}POD(A)_i$. Figure 9c, in which three possible inspections are assumed for simplicity, shows schematically

this approach. It worth noticing the single values $POD(A)_i$ to be introduced in Eq. (5) will be different at each inspection due to crack growth and its consequent variation in shape/area.

The probability of failure P_f of the inspection procedure is then related to PC_{DET} by the following expression (Fig. 9c):

$$P_f = 1 - PC_{DET} = \prod_i POND(A)_i = \prod_i [1 - POD(A)]_i \quad (6)$$

5. AN APPLICATIVE EXAMPLE

An applicative example is presented, considering high-speed applications whose hollow axles (D_{ext}=176 mm and D_{int}=65 mm) are made of the medium strength EA4T steel grade already described in Section 2.

5.1. Derivation of the Master POD curve of the boreprobe

The adopted BAT is [28] a multi-channel system allowing the complete inspection of the external surface of the axle and of the internal one of the bore by means of up to eight probes mounted on a special probe holder (Fig. 10a). In the present research, two probes, all having frequency equal to 4 MHz and dimension of the piezo-electric transducer equal to 10x10 mm^2, have been considered: 45° "forward" and "backward", which represent the best configurations to detect defects, in a 1st leg approach, along the external surface of the body and of geometrical transitions. "Forward" means the sound beam is directed towards the probe holder tip, while "backward" along the opposite direction. The equipment was calibrated using the sample block described in [23] and characterized by 16x1 mm^2 concave defects.

The amplitude responses from the naturally fatigue-cracked axles described in Section 2.3 were then acquired adopting a "signal response" approach. This provided a total sample of 17 experimental data. Furthermore, the boreprobe was introduced, into each axle, from both ends so to inspect any crack from either side. In particular, since the defects are inclined with respect to the surface, they can be considered as "favourable" or "unfavourable" with respect to the inspection (Fig. 10b). Such definitions come from the observation that the relative inclination between probes and cracks, in some cases, increases the reflecting area, in other cases decreases it. Nevertheless, being the BAT a complex system formed by different

transducers, the data needed for the definition of its Master POD curve were collected considering only the best (highest) response obtained by the two probes for each inspected crack (Fig. 10c).

Figure 10d shows the Master POD curves, and their confidence intervals at 95%, obtained applying the statistical approach described in Section 2.1. It is important to note that two different decision thresholds were adopted (Fig. 10c). The first, related to the 16x1 mm^2 calibration defect, represents the application parameter of the boreprobe (Fig. 1c) according to the inspection procedure by Lucchini RS [23], while the second, related to a semi-circular defect having a radius equal to 1 mm, represents its intrinsic capability (Fig. 1c). As described in the Introduction, even if the intrinsic Master POD curve is much more performant, its drawback consists in detecting also a lot of factors and parameters (surface roughness, coatings, …) which could make the interpretation of responses more difficult and the probability of false calls higher. For these reason, only the application parameter Master POD curve will be considered in the following.

Finally, no human factors were considered in the present research, because the boreprobe is a highly automated system.

5.2. Fatigue crack growth lifetime estimations

The initial crack, considered in crack growth simulations, was located along the T-transition at the same position where natural cracks were developed in the fatigue-cracked axles. Three different scenarios were considered: i) a semi-circular shape with R=1 mm and reflecting area equal to 1.57 mm^2; ii) a shallow shape with depth a=0.44 mm, aspect ratio a/c=0.2 (defined as the ratio between the depth "a" and the semi surface length "c") and reflecting area equal to 1.57 mm^2; iii) a shallow shape with depth a=1 mm, aspect ratio a/c=0.2 and reflecting area equal to 7.85 mm^2. In particular, cases i) and ii) represent different shapes with same reflecting area, while cases i) and iii) different shapes with same depth.

Life predictions were carried out by the dedicated software package AFGrow v. 4.0012.15 [32], based on an implementation of the widespread and well-known NASGRO equations [33]. The first one is able to describe crack growth rate ("da/dN") as a function of the stress intensity factor ("ΔK") for all the propagation regimes (threshold, linear and critical), while the second one describes the threshold ("ΔK_{th}") variation as a function of the stress ratio ("R"). Calibration of the empirical parameters in these equations

was carried out by dedicated fracture mechanics experiments: in the case of EA4T grade, the authors, within previous researches, already achieved such a calibration [34].

A typical in-service load spectrum, available in the literature [35] and representative of about 57000 km of high-speed service, was considered.

Figure 11 shows the predicted results. First, a rigorous evidence on the evolution of initial crack shapes during propagation is shown in Figure 11a in terms of aspect ratio evolution: as expected, all initial cracks tend to the same shape. On the other hand, looking at the predicted lives (Fig. 11b to 11d, normalized with respect to the total life of the semi-circular crack), it seems the considered cracks spend most of their propagation lifetime to reach such a coincident shape, supporting again the reflecting area approach. Then, shallow initial cracks (Fig. 11c and 11d) provide a lifetime shorter than semi-circular ones with the same area or depth (Fig. 11b). In particular, the life of the 1.57 mm^2 shallow crack is shorter than, but comparable to, that of the 1.57 mm^2 semi-circular one, while the life of the 7.85 mm^2 (1 mm deep) shallow crack is significantly shorter than that of the 1.57 mm^2 (1 mm deep) semi-circular one. Nevertheless, while the former observation could be reasonably expected due to similar initial crack areas, the latter seems to suggest that, reasoning in terms of depth, simulations describe completely different prediction scenarios and this can have a strong influence on the determination of the probability of failure of the adopted inspection procedure, as discussed in the following section.

5.3. Inspection intervals and probability of failure of the adopted inspection procedure

The probabilities of failure, of the adopted inspection procedure for the present applicative case, were calculated based on different lengths of inspection interval. Moreover, the crack propagation curves shown in Figures 11b to 11d and the application parameter POD curves shown in Figures 7b (depth for a semi-circular crack) and 10d (reflecting area) were adopted. In all cases, the single POD values needed to define P_f were derived from 95% confidence limit curves.

Figure 12a shows the comparison of results, for the three considered initial cracks, obtained applying the experimental POD curve in terms of depth (Fig. 7b). This means a POD curve derived from the responses of experimental semi-circular cracks was applied to shallow ones, as well. First, as expected, the

wider the length of the inspection interval, the higher the probability of failure. Then, defects with the same reflecting area show very similar trends, while defects with the same depth, but different reflecting area, seem to provide significantly different probabilities of failure. This is expected and implicitly included into the reflecting area concept: even if a crack is represented by its depth, its experimental ultrasonic response, from which POD curves are derived, seems to be, in any case, physically based on the reflecting area. Actually, the effect of crack growth life predictions should be here considered, as well: 1.57 mm^2 initial cracks allow for ten to fifteen 1.000.000 km long inspections during their propagation life, while the 7.85 mm^2 initial crack just two. This means that, for the case of the 7.85 mm^2 initial crack, the length of the inspection interval should be significantly lowered to about 100.000-200.000 km in order to achieve the same number of inspections of 1.57 mm^2 initial cracks. The analysis of these synergic effects between crack growth predictions and NDT inspection intervals is beyond the aims of the present research and remains an open point for future developments.

Considering then the application of the Master POD curve (Fig. 10d) to just the 1.57 mm^2 initial cracks, Figure 12b compares the probabilities of failure of the adopted inspection procedure. The trends in terms of probabilities of failure coincides allowing concluding, again, the Master POD method is more consistent than the traditional one. The same Figure shows, in grey for further comparison, the curves reported in Figure 12a for 1.57 mm^2 cracks. As can be seen, the Master POD approach seems to provide significantly lower probabilities of failure of the inspection procedure, suggesting that a more precise description of the crack area/shape during a damage tolerant calculation enables an optimization (extension) of inspection intervals. As an example, considering a 10^6 km long inspection interval, a probability of failure equal to $6 \cdot 10^{-6}$ is achieved by the traditional POD curve and equal to $3 \cdot 10^{-7}$ by the Master POD one. The traditional POD curve could, then, be chosen to assume a more conservative approach, but the Master POD one provides a more accurate calculation.

6. CONCLUDING REMARKS

In order to ensure the safe service of safety critical mechanical components, they are usually subjected to a dedicated inspection plan. The capability of such inspections is quantified in terms of “Probability of

Detection" curves, which are traditionally related to a specific linear dimension of the defect to be detected.

The present research analyses some inconsistencies of such a traditional definition of ultrasonic POD curves and is focused, as an applicative example, on the capabilities of ultrasonic inspection, by means of a highly automated boreprobe equipment, of hollow axles for high speed application. The obtained results can be summarized:

- a set of naturally fatigue-cracked full-scale hollow axles was obtained at the bench making available a total of seventeen semi-elliptical fatigue cracks characterized by a depth ranging from 0.4 to 12 mm;
- the meaningful geometrical feature of a crack, inspected by ultrasonic testing, is its reflecting area and not a characteristic linear dimension of its, as traditionally assumed. This is even more significant for fatigue propagating inspected cracks, whose areas/shapes are continuously varying during service;
- Master POD curves, built in terms of a crack reflecting area, are independent from its shape. Consequently, they have characteristics of generality and versatility not available adopting traditional POD curves, especially considering evolving crack shapes;
- Master POD curves easily allow back calculating the traditional ones. This is an indispensable feature for interfacing the Master POD curve to the typical crack growth algorithms for fracture mechanics assessment;
- Master POD curves allow determining the right decision thresholds for different times of flight with less experimental effort than the traditional ones;
- fatigue crack growth lifetime estimations showed shallow initial cracks provide a lifetime shorter than semi-circular ones. Nevertheless, all initial cracks tend to the same shape increasing the number of fatigue cycles;
- considering an operative decision threshold and traditional POD curves in terms of depth, defects with the same reflecting area showed very similar trends of probability of failure, while defects with the same depth, but different reflecting area, seemed to provide significantly different ones;

- Master POD curves seem to provide significantly lower probabilities of failure of the inspection procedure with respect to traditional POD ones, suggesting that a more precise description of the crack area/shape during a damage tolerant calculation enables an optimization (extension) of inspection intervals.

AKNOWLEDGEMENTS

The authors sincerely thank Prof. S. Beretta (Dept. Mechanical Engineering – Politecnico di Milano) for the useful help and discussion.

REFERENCES

[1] Zerbst U, Vormwald M, Andersch C, Mädler K and Pfuff M. The development of a damage tolerance concept for railway components and its demonstration for a railway axle. Eng. Fract. Mech. 2003;72:209-239.

[2] Grandt AF Jr. Fundamentals of structural integrity. Hoboken: John Wiley & Sons Inc.; 2003.

[3] Georgiou GA. Probability of Detection (POD) curves: derivation, applications and limitations. UK: Health and Safety Executive Books, Research Report 454; 2006.

[4] ASM. ASM Handbook – Vol. 17: Non-destructive evaluation and quality control. 1997.

[5] MIL-HDBK-1823A. Nondestructive evaluation system reliability assessment. Department of Defense of the US; 2009.

[6] Carboni M, Beretta S. Effect of probability of detection upon the definition of inspection intervals for railway axles. J. Rail and Rapid Transit 2007;221:409-417.

[7] Rummel WD. Recommended practice for a demonstration of non-destructive evaluation (NDE) reliability on aircraft production parts. Mater. Eval. 1982;40:923-932.

[8] Carboni M, Cantini S, Gilardoni C. Validation of the Rotating UT Probe for In-Service Inspections of Freight Solid Axles by Means of the MAPOD Approach. Proc. 5th European-American Workshop on Reliability of NDE, Berlin, Germany, 2013.

[9] Cantini S, Beretta S (Editors). Structural reliability assessment of railway axles. Lovere: LRS-Techno Series 4; 2011.

[10] Thompson RB. A unified approach to the model-assisted determination of probability of detection. Proc. AIP Conf., Golden (Colorado), USA, 2008.

[11] Knopp JS, Aldrin JC, Lindgren E, Annis C. Investigation of a Model-Assisted Approach to Probability of Detection Evaluation. Proc. AIP Conf., Portland (Oregon), USA, 2006.

[12] Harding C, Hugo G, Bowles S. Model-assisted probability of detection validation of automated ultrasonic scanning for crack detection at fastener holes. Proc. 10th Joint Conference on Aging Aircraft, Palm Springs (California), USA, 2007.

[13] Mueller C, Bertovic M, Kanzler D, Heckel T, Holstein R, Ronneteg U, Pitkänen J, Rosenthal M. Assessment of the Reliability of NDE: A Novel Insight on Influencing factors on POD and Human Factors in an organizational Context. Proc. 11th European Conference on NDT (ECNDT2014), Prague, Czech Republic, 2014.

[14] EN 13261. Railway applications – Wheelsets and bogies – Axles – Product requirements. CEN; 2011.

[15] Lonsdale CP, Stone DH. North American Axle Failure Experience. J. Rail and Rapid Transit 2004;218:293-298.

[16] Beretta S, Carboni M, Fiore G, Lo Conte A. Corrosion-fatigue of A1N railway axle steel exposed to rainwater. Int. J Fatigue 2010;32:952-961.

[17] Beretta S, Carboni M, Lo Conte A, Palermo E. An investigation of the effects of corrosion on the fatigue strength of A1N axle steel. J. Rail and Rapid Transit 2008;222:129-143.

[18] VPI, VPI 04 – Maintenance of freight trains – Wheel-sets, 2nd edition, 1st modification, valid from 01/08/2008.

[19] Pettinato G. Il controllo ad ultrasuoni semiautomatico degli assi delle sale montate dei veicoli ferroviari in opera o fuori opera. La Metallurgia Italiana 1971;8:367-374. [In Italian]

[20] Benyon JA, Watson AS. The use of Monte-Carlo analysis to increase axle inspection interval. Proc. 13th International Wheelset Congress (IWC13), Roma, Italy, 2001.

[21] Carboni M, Beretta S, Cantini S, Gilardoni C. Probability of Detection of Ultrasonic In-Service Inspection of Hollow Axles. Proc. 5th European-American Workshop on Reliability of NDE, Berlin, Germany, 2013.

[22] Carboni M. A critical analysis of ultrasonic echoes coming from natural and artificial flaws and its implications in the derivation of probability of detection curves. Insight 2012;54:208-216.

[23] Patelli G. Technical Instruction QUA IT 065 Rev.7. Lovere (BG): Lucchini RS SpA; 2009.

[24] Lawless JF. Statistical Models and Methods for Lifetime Data. Second Edition, Wiley-Interscience; 2002.

[25] Taylor JR. An Introduction to Error Analysis: The Study of Uncertainties in Physical Measurements. Second edition, Herndon: University Science Books; 1997.

[26] Nelson W. Applied life data analysis. New York: John Wiley & Sons Inc.; 1982.

[27] Birchack JR, Gardner CG. Comparative ultrasonic response of machined slots and fatigue cracks in 7075 aluminum. Materials Evaluation 1976;34(12):275-280.

[28] Rocchi C, Patelli G, Gilardoni C. A new generation UT inspection device for high performance hollow axles. Proc. 15th Int. Wheelset Congress, Prague, Czech Republic, 2007.

[29] Sanford RJ. Principles of fracture mechanics. Upper Saddle River: Pearson Education Inc.; 2003.

[30] Beretta S, Carboni M, Madia M. Fatigue strength in presence of inhomogeneities: influence of constraint. J ASTM Int 2006;3(4):1-11.

[31] Krautkrämer J, Krautkrämer H. Ultrasonic Testing of Materials. 4th Ed. Berlin: Springer Berlin Heidelberg; 1990.

[32] AFGrow v. 4.0012.15, User's Manual, 2008.

[33] NASGRO Consortium. NASGRO v4.23 User's Manual. Website: www.nasgro.swri.org (accessed on 01/03/2015).

[34] Regazzi D, Beretta S, Carboni M. An investigation about the influence of deep rolling on fatigue crack growth in railway axles made of a medium strength steel. Eng Fract Mech 2014;131:587-601.

[35] Beretta S, Carboni M, Cervello S. Design review of a freight railway axle: fatigue damage versus damage tolerance. Mat.-wiss. U. Werkstofftech. 2011;42(12):1099-1104.

LIST OF FIGURES

Figure 11 - Fatigue crack growth predictions: a) evolution of the aspect ratio; b) 1.57 mm^2 semicircular crack; c) 1.57 mm^2 shallow crack; d) 7.85 mm^2 shallow crack.

Figure 12 - Probability of failure and inspection intervals: a) traditional POD curve; b) Master POD curve.

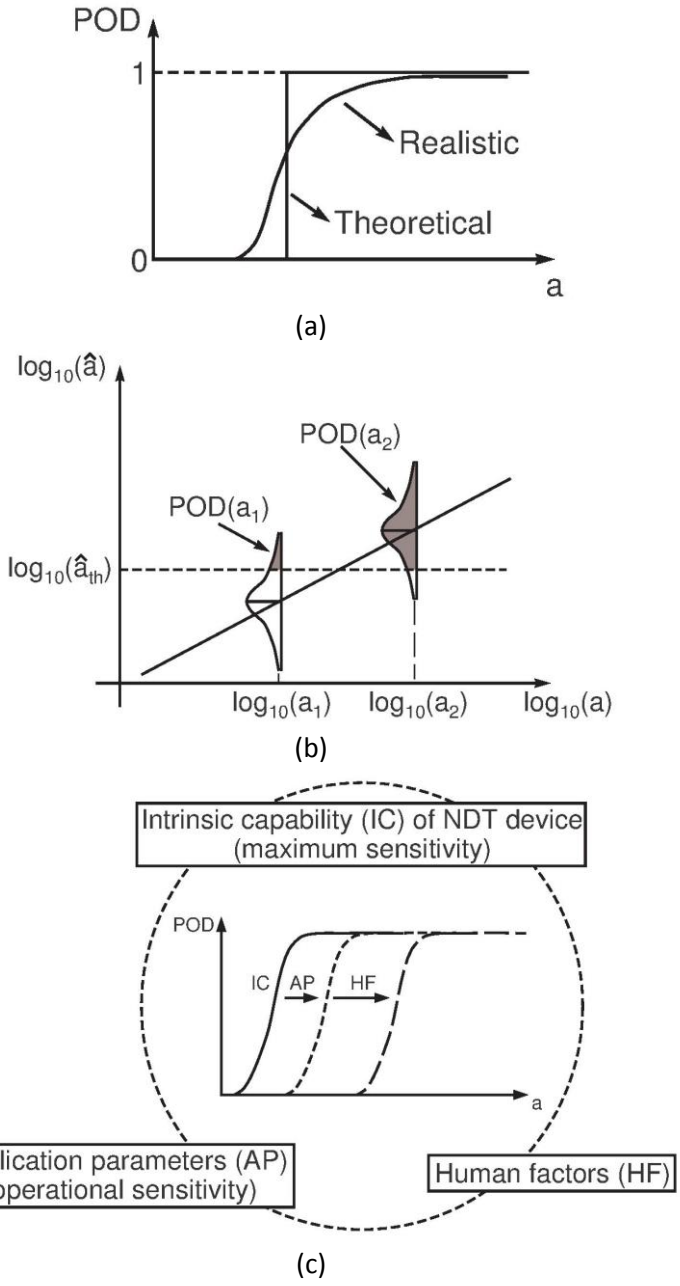
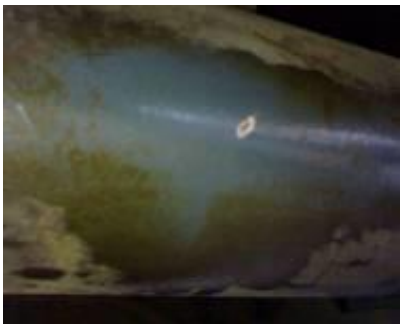

POD
1
0
a
Realistic
Theoretical
(a)
log10(â)
POD(a2)
POD(a1)
log10(âth)
log10(a1)
log10(a2)
log10(a)
(b)
Intrinsic capability (IC) of NDT device
(maximum sensitivity)
POD
IC
AP
HF
a
Application parameters (AP)
(operational sensitivity)
Human factors (HF)
(c)

Fig. 1

(a)

(b)



(c)

Fig. 2

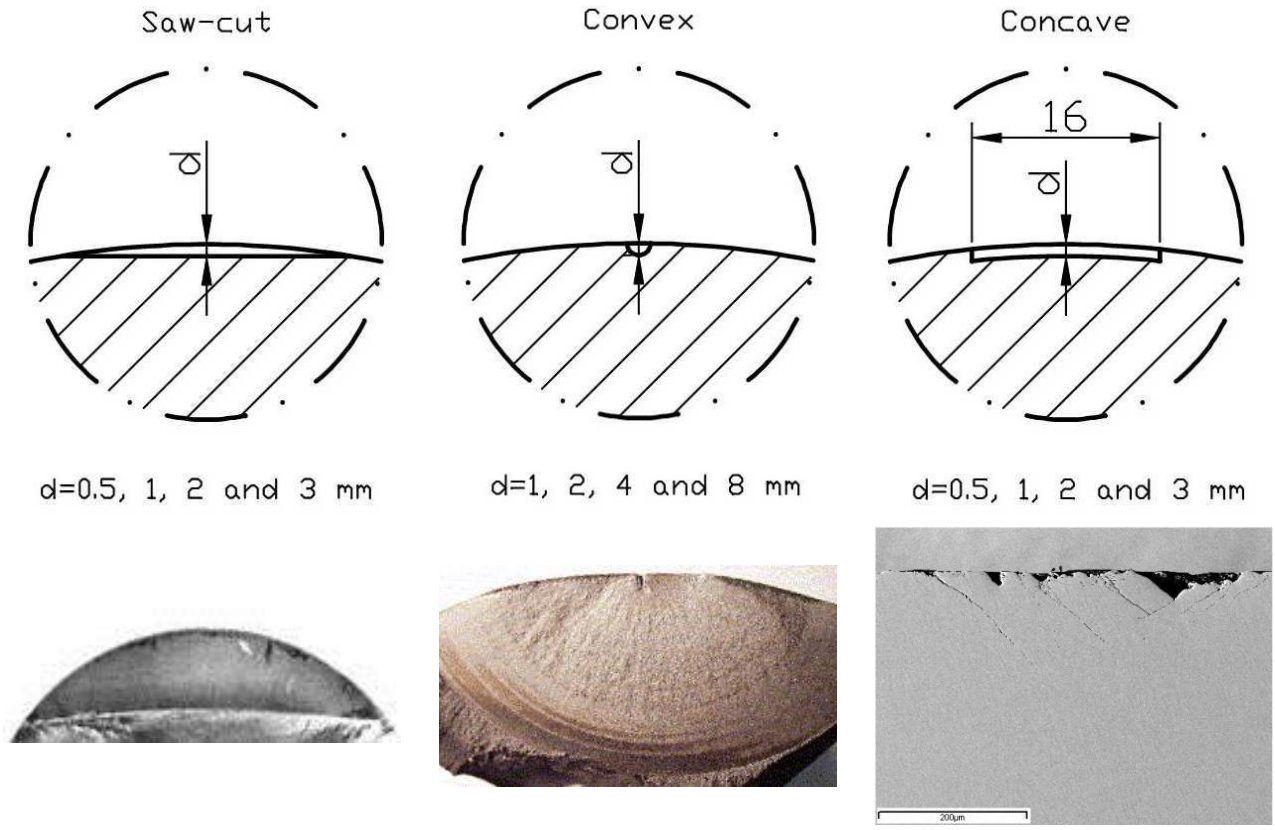

Saw-cut
Convex
Concave
d
d
16
d
d=0.5, 1, 2 and 3 mm
d=1, 2, 4 and 8 mm
d=0.5, 1, 2 and 3 mm
200µm

Fig. 3

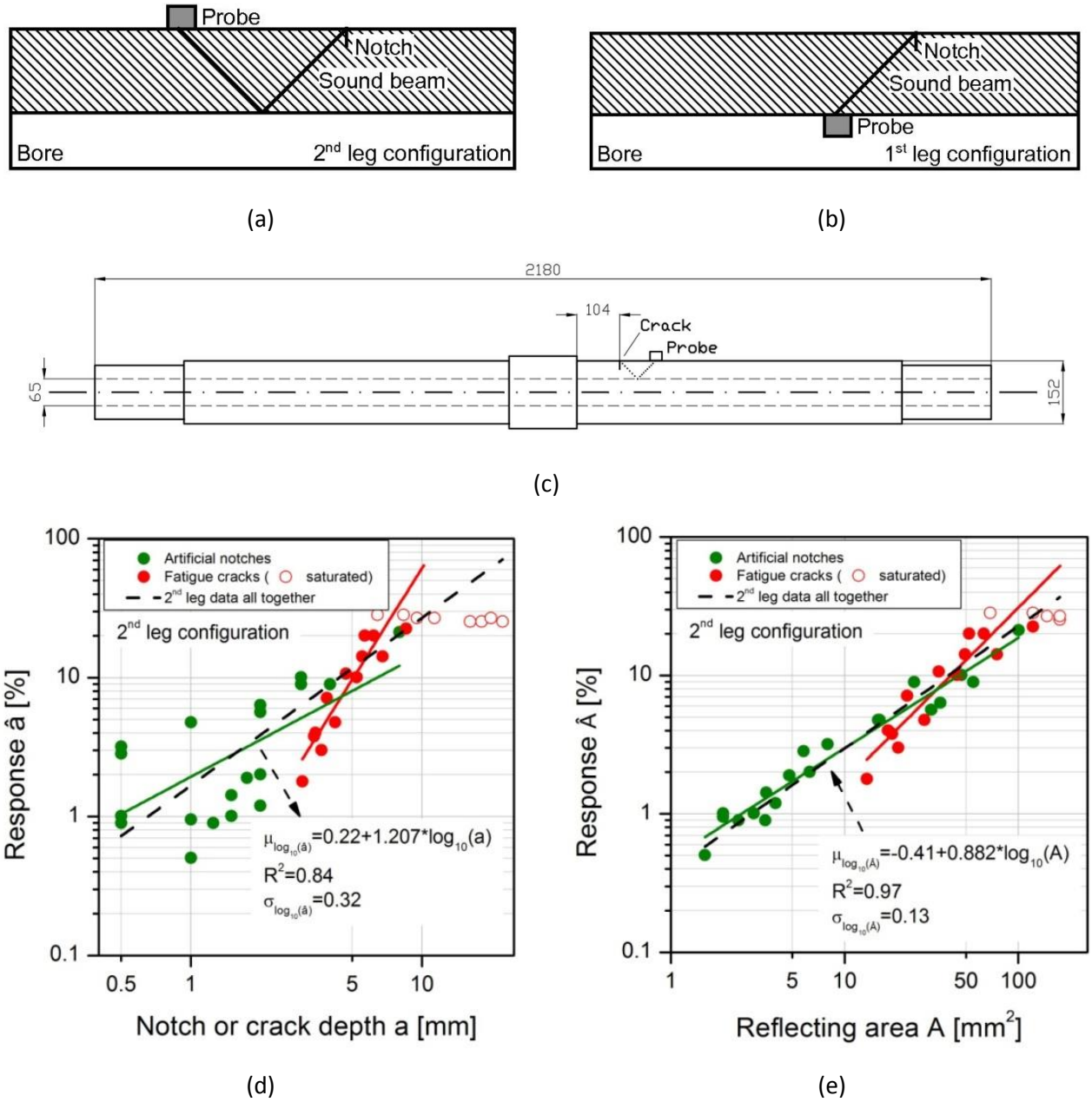
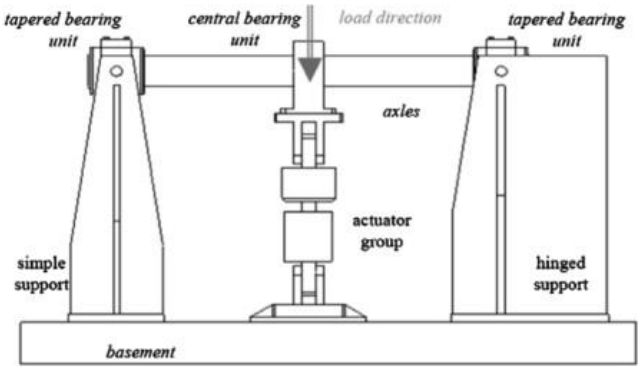
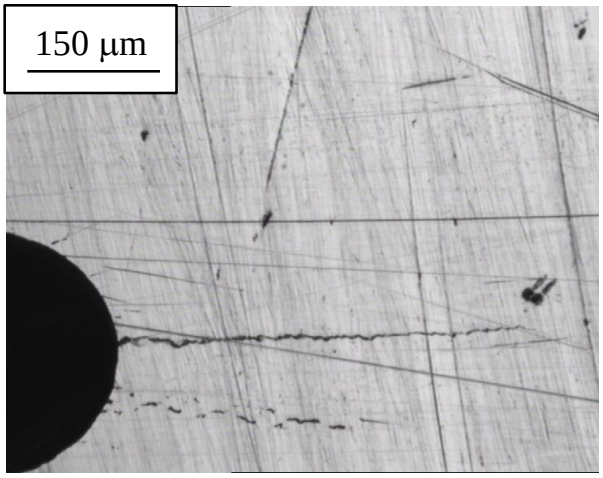

Probe
Notch
Sound beam
Bore
2nd leg configuration
(a)
Notch
Sound beam
Probe
Bore
1st leg configuration
(b)
2180
104
Crack
Probe
65
152
(c)
Artificial notches
Fatigue cracks (saturated)
2nd leg data all together
2nd leg configuration
μlog10(â)=0.22+1.207*log10(a)
R2=0.84
σlog10(â)=0.32
Response â [%]
Notch or crack depth a [mm]
(d)
Artificial notches
Fatigue cracks (saturated)
2nd leg data all together
2nd leg configuration
μlog10(Â)=-0.41+0.882*log10(A)
R2=0.97
σlog10(Â)=0.13
Response Â [%]
Reflecting area A [mm2]
(e)

Fig. 4

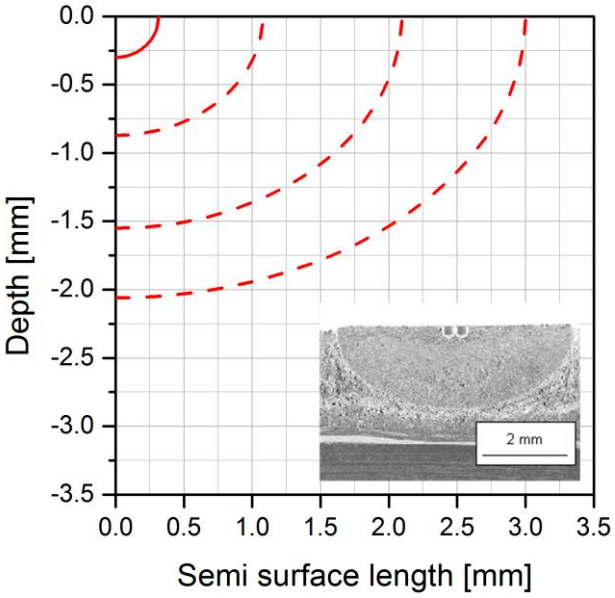

tapered bearing unit
central bearing unit
load direction
tapered bearing unit
axles
actuator group
simple support
hinged support
basement

(a)

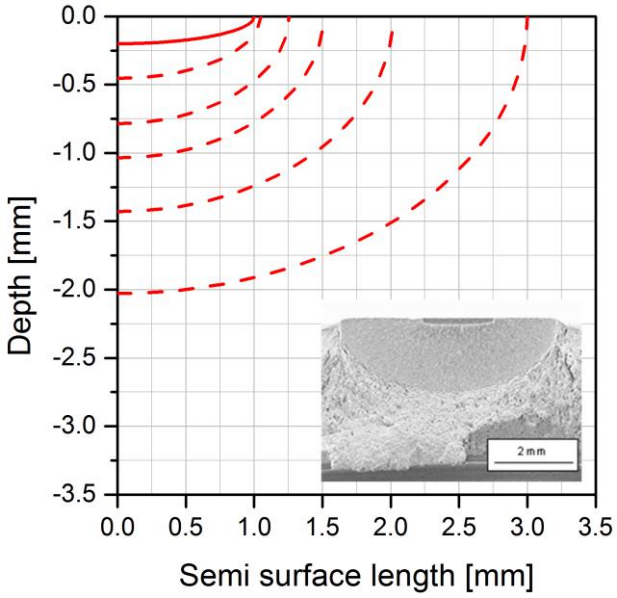

150 μm

(b)

Fig. 5

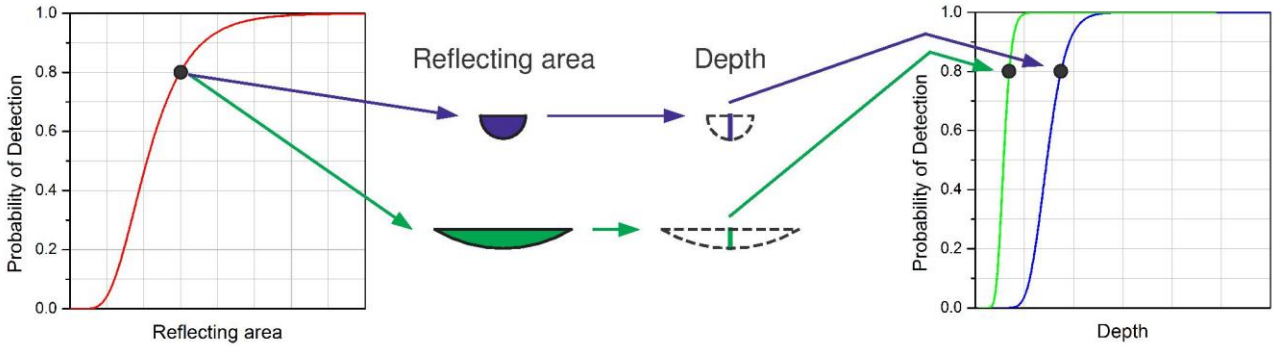

0.0
-0.5
-1.0
-1.5
-2.0
-2.5
-3.0
-3.5
Depth [mm]
0.0 0.5 1.0 1.5 2.0 2.5 3.0 3.5
Semi surface length [mm]
2 mm

(a)

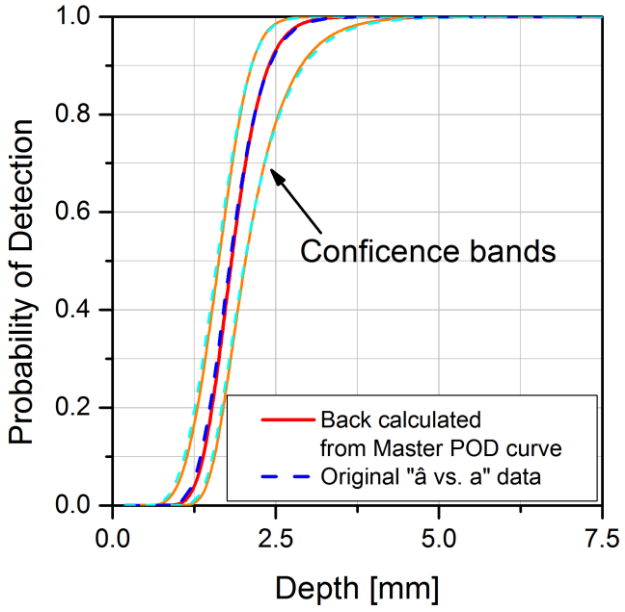

0.0
-0.5
-1.0
-1.5
-2.0
-2.5
-3.0
-3.5
Depth [mm]
0.0 0.5 1.0 1.5 2.0 2.5 3.0 3.5
Semi surface length [mm]
2 mm

(b)

Fig. 6

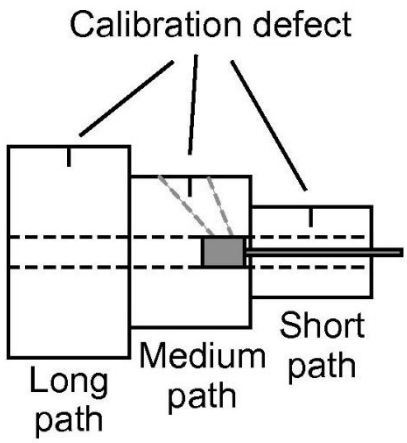

Probability of Detection
Reflecting area
Reflecting area
Depth
Probability of Detection
Depth

(a)

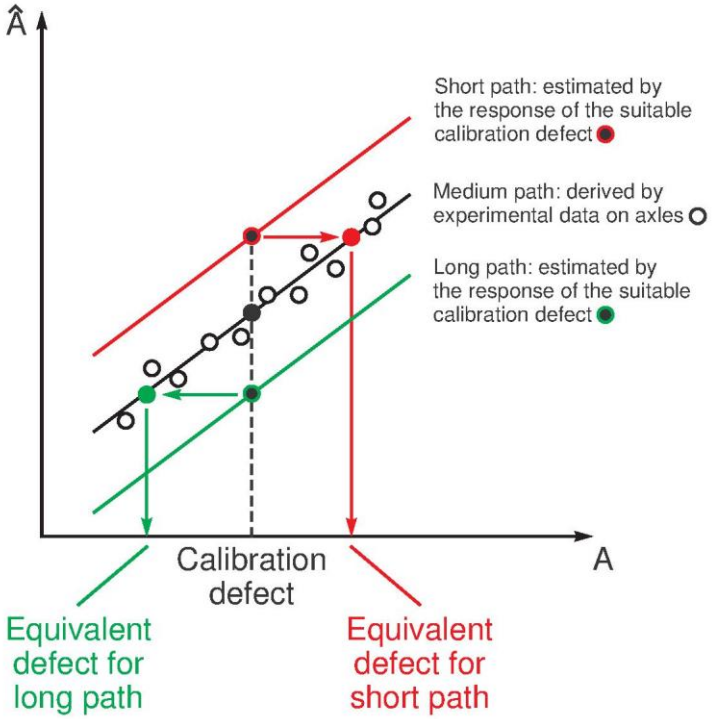

Probability of Detection
Conficence bands
Back calculated
from Master POD curve
Original "â vs. a" data
Depth [mm]

(b)

Fig. 7

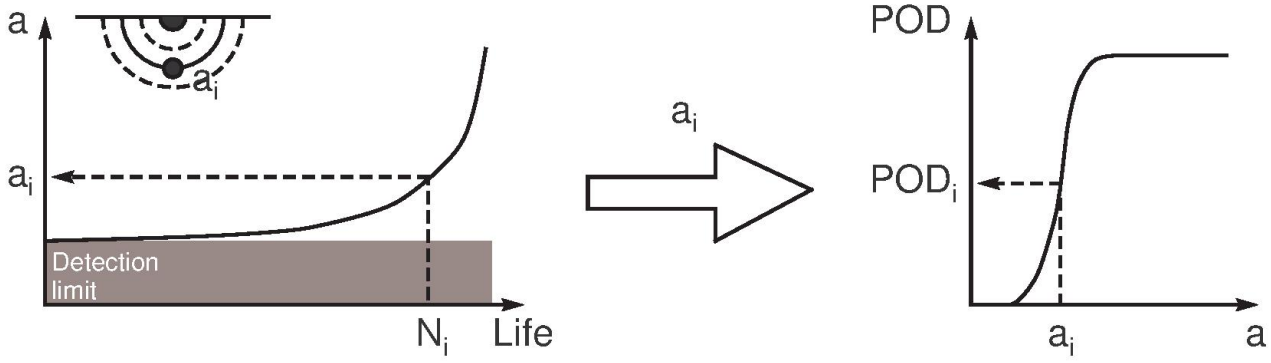

Calibration defect
Long path
Medium path
Short path

(a)

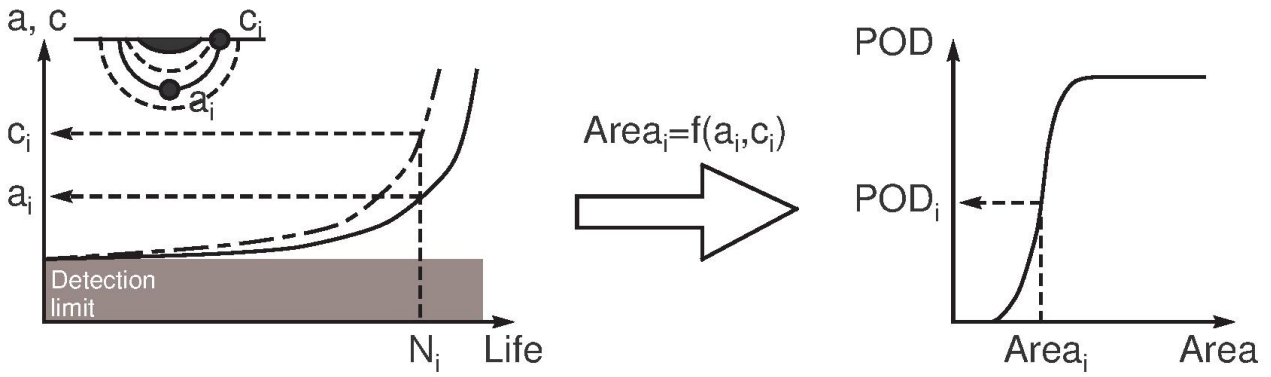

Â
Short path: estimated by the response of the suitable calibration defect
Medium path: derived by experimental data on axles
Long path: estimated by the response of the suitable calibration defect
A
Calibration defect
Equivalent defect for long path
Equivalent defect for short path

(b)

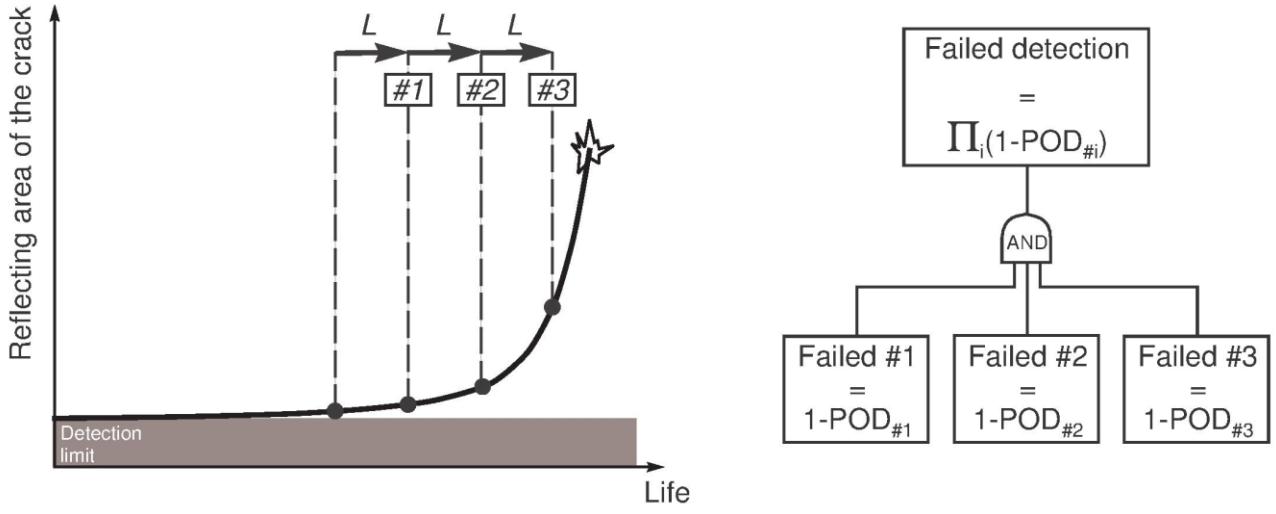
Fig. 8


a
a_i
a_i
Detection limit
N_i
Life
a_i
POD
POD_i
a_i
a

(a)

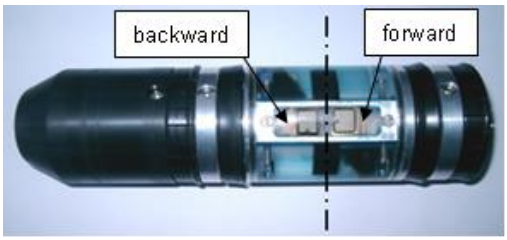

a, c
c_i
a_i
c_i
a_i
Detection limit
N_i
Life
$Area_i=f(a_i,c_i)$
POD
POD_i
$Area_i$
Area

(b)

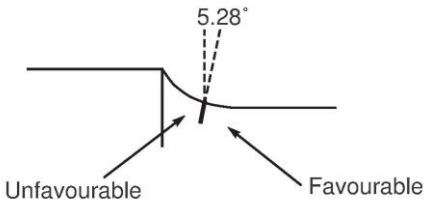

Reflecting area of the crack
L
L
L
#1
#2
#3
Detection limit
Life
Failed detection
=
$\Pi_i(1\text{-}POD_{\#i})$
AND
Failed #1
=
$1\text{-}POD_{\#1}$
Failed #2
=
$1\text{-}POD_{\#2}$
Failed #3
=
$1\text{-}POD_{\#3}$

(c)

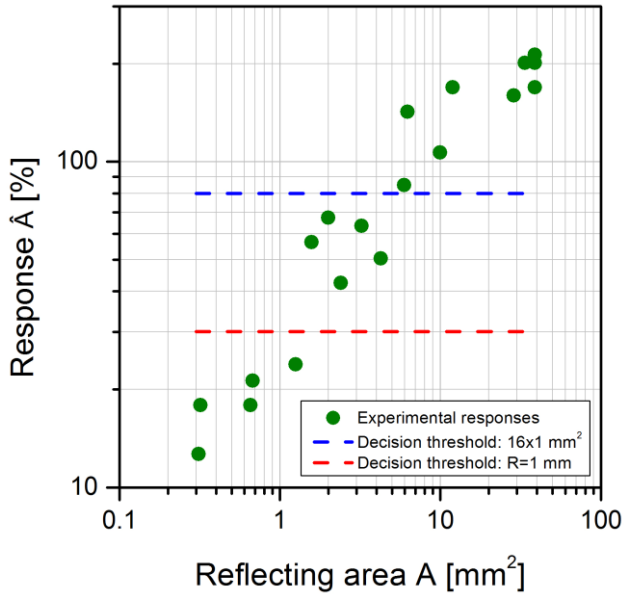
Fig. 9


backward
forward

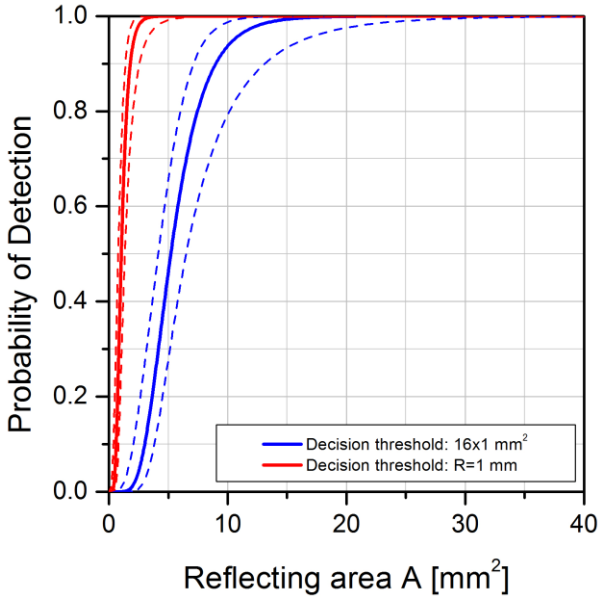
(a)


5.28°
Unfavourable
Favourable

(b)

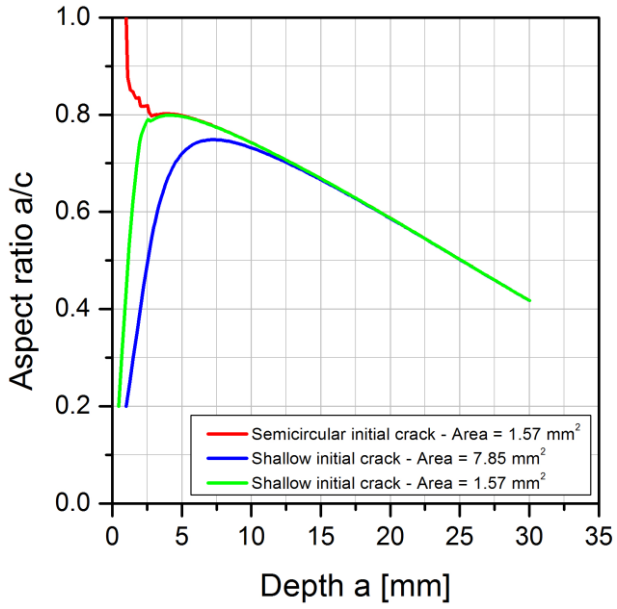

Response Â [%]
100
10
0.1
1
10
100
Reflecting area A [mm²]
Experimental responses
Decision threshold: 16x1 mm²
Decision threshold: R=1 mm

(c)

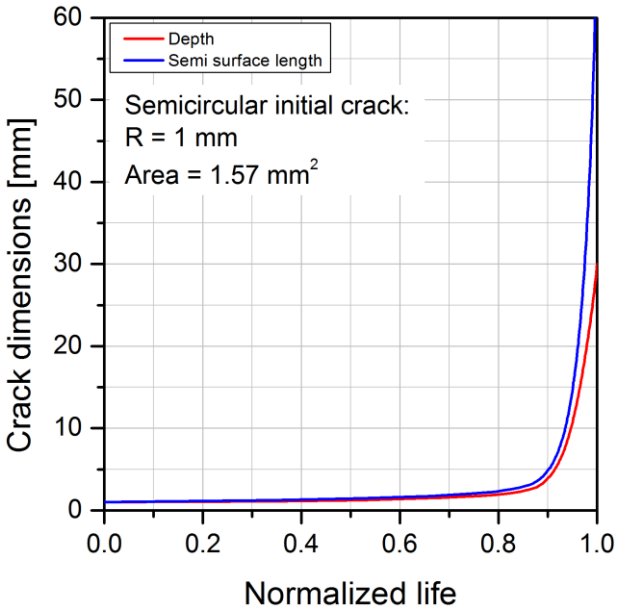

Probability of Detection
1.0
0.8
0.6
0.4
0.2
0.0
0
10
20
30
40
Reflecting area A [mm²]
Decision threshold: 16x1 mm²
Decision threshold: R=1 mm

(d)

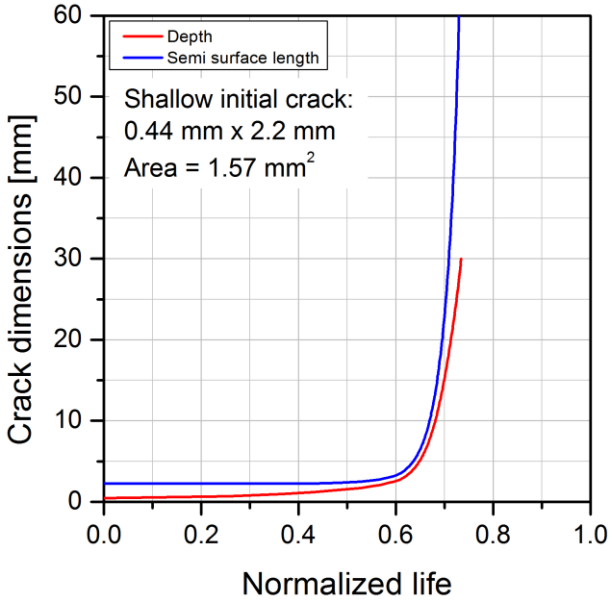
Fig. 10


Aspect ratio a/c
Depth a [mm]
Semicircular initial crack - Area = 1.57 mm²
Shallow initial crack - Area = 7.85 mm²
Shallow initial crack - Area = 1.57 mm²

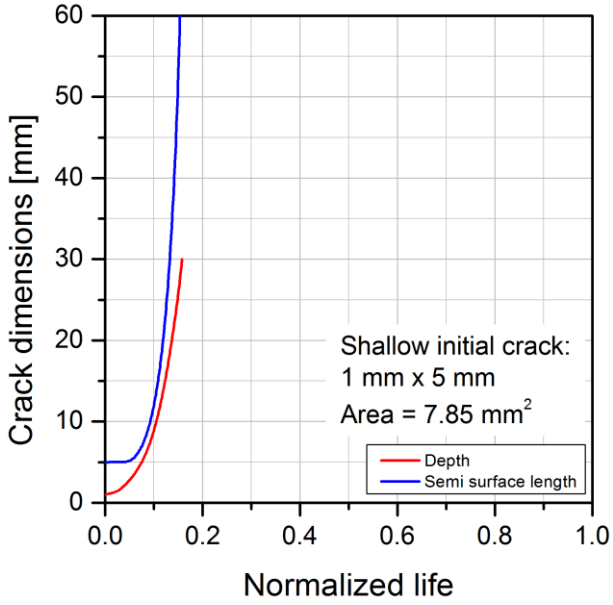
(a)


Crack dimensions [mm]
Normalized life
Depth
Semi surface length
Semicircular initial crack:
R = 1 mm
Area = 1.57 mm²

(b)

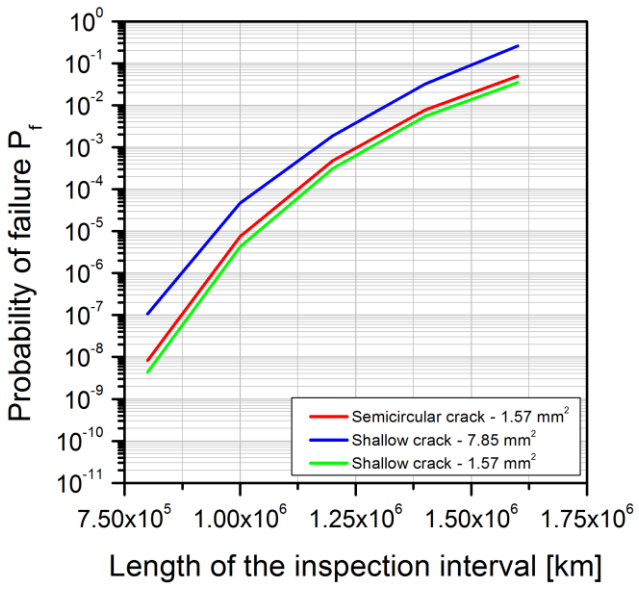

Crack dimensions [mm]
Normalized life
Depth
Semi surface length
Shallow initial crack:
0.44 mm x 2.2 mm
Area = 1.57 mm²

(c)

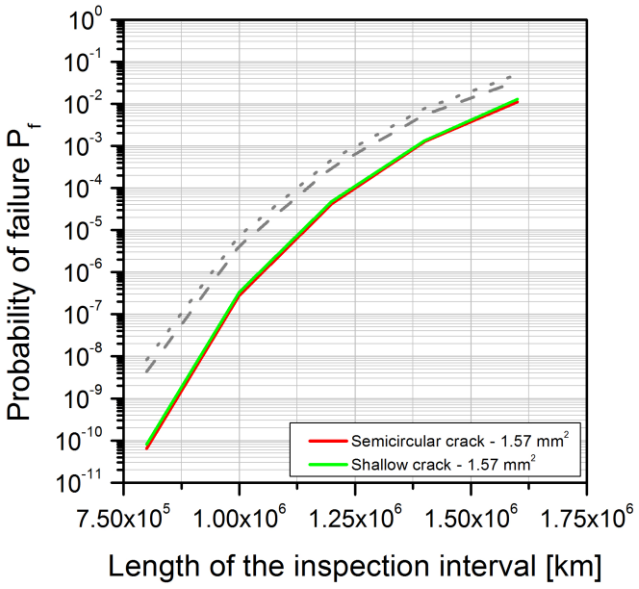

Crack dimensions [mm]
Normalized life
Shallow initial crack:
1 mm x 5 mm
Area = 7.85 mm²
Depth
Semi surface length

(d)

Fig. 11


Probability of failure P_f
10^{0}
10^{-1}
10^{-2}
10^{-3}
10^{-4}
10^{-5}
10^{-6}
10^{-7}
10^{-8}
10^{-9}
10^{-10}
10^{-11}
7.50x10^{5}
1.00x10^{6}
1.25x10^{6}
1.50x10^{6}
1.75x10^{6}
Length of the inspection interval [km]
Semicircular crack - 1.57 mm^2
Shallow crack - 7.85 mm^2
Shallow crack - 1.57 mm^2

(a)


Probability of failure P_f
10^{0}
10^{-1}
10^{-2}
10^{-3}
10^{-4}
10^{-5}
10^{-6}
10^{-7}
10^{-8}
10^{-9}
10^{-10}
10^{-11}
7.50x10^{5}
1.00x10^{6}
1.25x10^{6}
1.50x10^{6}
1.75x10^{6}
Length of the inspection interval [km]
Semicircular crack - 1.57 mm^2
Shallow crack - 1.57 mm^2

(b)

Fig. 12