18th World Conference on Nondestructive Testing, 16-20 April 2012, Durban, South Africa

A "Model Assisted Probability of Detection" approach for ultrasonic inspection of railway axles

Michele CARBONI [1], Stefano CANTINI [2]

[1] Department of Mechanical Engineering, Politecnico di Milano; Milano, Italy
Phone: +39 02 23998253, Fax: +39 02 23998202; michele.carboni@polimi.it
[2] Lucchini RS SpA; Lovere (BG), Italy
Phone: +39 035 963255, Fax: +39 035 963463; s.cantini@lucchinirs.it

Abstract
The reliability of non-destructive testing is usually quantified in terms of the "Probability of Detection" curves relating a characteristic linear dimension of the considered defects to the probability to observe them. Actually, POD curves are also function of other implicit factors, so they can be considered stochastic in nature and, consequently, the most rational and reliable approach to their derivation should also consider the determination of a suitable confidence level. The problem arises because such confidence level is usually requested to be equal to 95% for POD curves, so involving high costs and long times for the needed experimental tests.
The present paper is focused on the reliability of ultrasonic inspection applied to hollow railway axles. In particular, a novel method, recently proposed by the authors, for the interpretation of POD data is firstly described and experimentally validated. Eventually, with the aim to lower the overall experimental costs, the possibility to substitute part of the needed experimental tests with proper numerical simulations, so applying the so-called "Model Assisted Probability of Detection" approach, is successfully investigated.

Keywords: railway axles, Probability of Detection (POD) curves, POD numerical modelling (MAPOD)

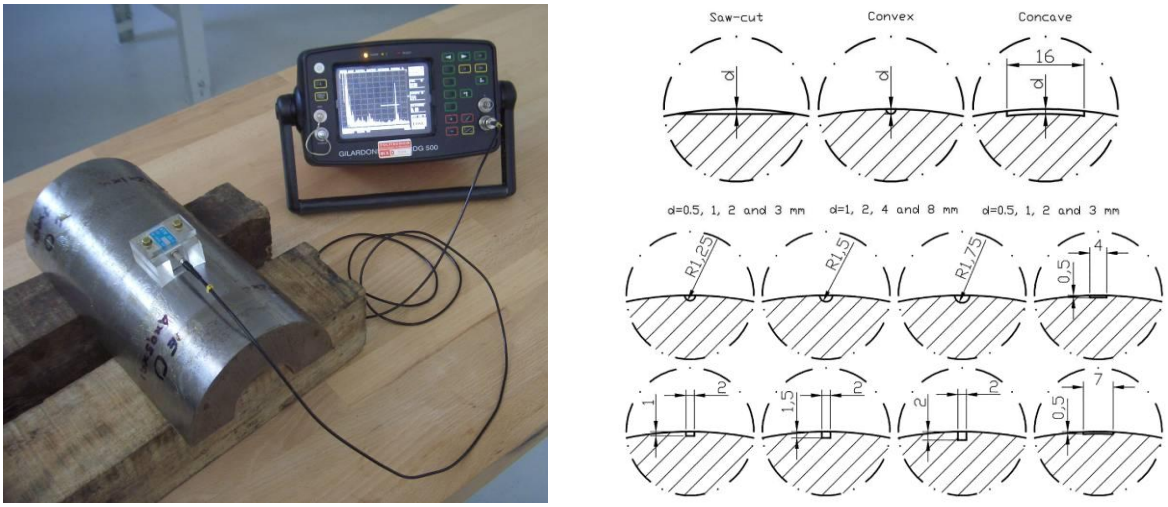
1. Introduction

Considering components subjected to fatigue, it is licit to expect crack initiation and consequent propagation during service. To face this problem, it is possible to employ the "Damage Tolerant" design approach whose 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]): i) the performance and the reliability of the adopted NDT procedure; ii) the crack propagation behavior of the adopted material; iii) the influence of the geometry of the cracked body on the crack driving force; iv) the reliable knowledge of service loads.
The present research is focused, between the just described points, on NDT effectiveness which is directly related to the reliable knowledge of the POD curve of the adopted inspection method. Traditionally, such probabilities are explicitly expressed and plotted in terms of a characteristic linear dimension of defects (depth, length, diameter, …). However, they are also a function of many other physical and operative factors such as: the adopted NDT method, the material, geometry, defect type, instrumentation, human and environmental effects. This means that very rarely the POD curve derived for a given configuration can be used for other configurations, even if similar. Another critical aspect of POD curves is the need, for reliability and design, of a statistical characterization of the largest defect that can be missed and not the smallest defect that can be observed. For this reason, POD curves should always be provided together with a suitable lower confidence limit (typically 95%) which needs a lot of experimental results to be achieved.
Considering, here, the specific case of ultrasonic (UT) NDT applied to railway axles (even if the approach is generally applicable to every UT inspection of mechanical components), a

novel approach, recently proposed by the authors [6-8], to the interpretation of UT responses is firstly built up and presented by means of a very high number of experimental responses coming from artificial and natural defects. Since experiments are expensive, it is then investigated the possibility to apply the "model-assisted probability of detection" (MAPOD) approach [9-11] which is very recent (the first publications come from 2003) and it is based on the idea to substitute part of the experimental responses with numerical ones obtained from suitable physical models. However, it is worth remembering that MAPOD does not allow to eliminate all the experimental activity because not all the relevant parameters are controlled by well-known physical models.

2. The "Reflecting Area" approach

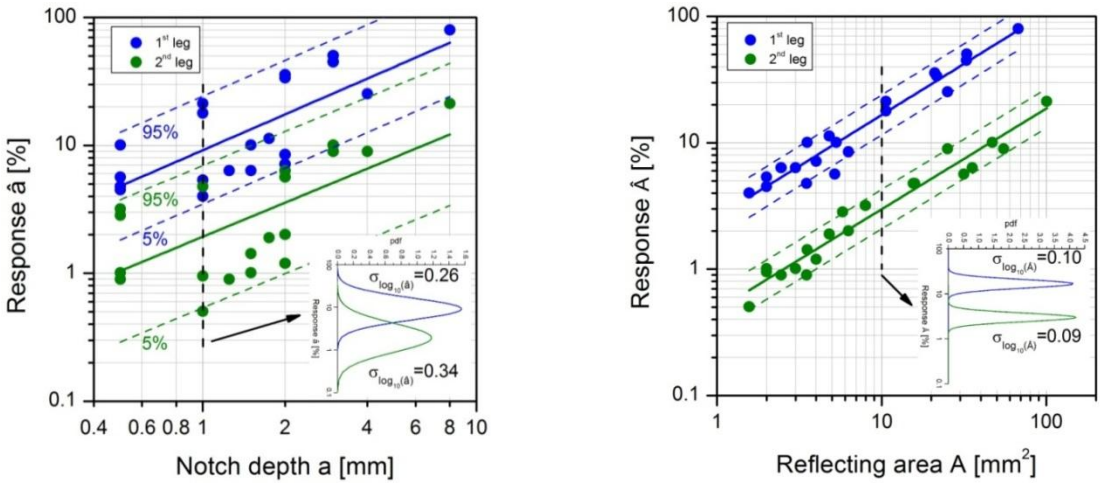
The considered material is an A4T steel grade typically used for the production of railway axles and characterized by a longitudinal wave speed V_L=5920 m/s and a shear wave speed V_T=3230 m/s. All the experimental tests were carried out by the same operator using a Gilardoni RDG500 flaw detector equipped with an ATM45/4 single 8x9 mm crystal probe. The UT response of the inspected defects was always recorded at a gain level equal to 48 dB, corresponding to the response echo of a big reflector (observed in a first leg configuration) set to the 80% of the display. The piece-probe coupling was guaranteed by means of grease and suitable Plexiglas wedges.
Twenty artificial defects were realized on the external surface of six chunks cut from axles made of A4T (Fig. 1a). Particularly, such defects are characterized by (Fig. 1b): different geometries (saw-cut, convex, concave), different dimensions (depths from 0.5 mm to 8 mm) and different manufacturing processes (traditional machining, EDM). All the defects could be inspected adopting both a 1st leg inspection configuration (direct reflection on the defect) and a 2nd leg one (with an intermediate reflection on the bore). It is important to remark that the industrial UT inspection, at fixed inspection intervals, of hollow axles during in-service maintenance considers only a 1st leg configuration by means of dedicated bore-probes [12]. The 2nd leg configuration is, actually, adopted for the inspection of solid axles and it should be added that, here, the intermediate reflection on the bore widens the sound beam lowering the sensibility.



(a) (b)

Figure 1. UT inspection of different artificial defects.

UT responses of artificial defects are reported in Figure 2a in terms of the traditional linear dimension of the defect (in this case, the depth). As it can be seen, the standard deviation is so high that the response coming from two different times of flight (one double of the other) intersect. This could be attributed [6-8] to the influence that the defect shape has on the UT response of defect having the same depth, but different morphology.
In order to eliminate the influence of the morphology of the defect, it was then proposed [6-8] to plot UT responses in terms of the reflecting area of defects, i.e. the intersection between the area of the sound beam at the considered time of flight and the area of the defect itself. More details on the procedure can be found in [6-8]. Results are reported in Figure 2b, where the influence of the morphology on UT responses (and, consequently, the standard deviation) is evidently decreased.



(a) (b)

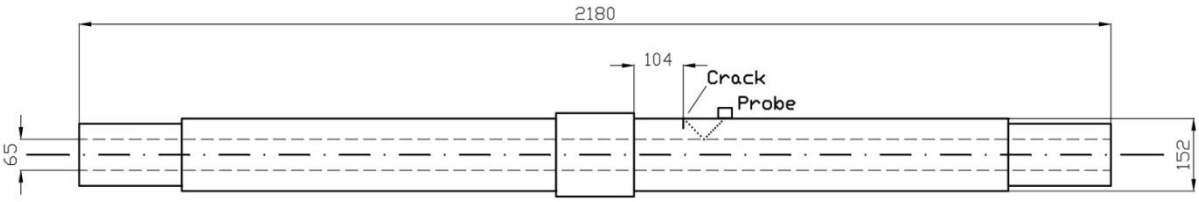
Figure 2. Comparison between UT responses in terms of (a) depth and (b) reflecting area.

UT inspections were then carried out on natural fatigue cracks induced, by means of proper artificial micro-defects, in the body of three full-scale axles (Fig. 3a) made of A4T and fatigue tested by the dedicated facility available at the Dept. of Mechanical Engineering of Politecnico di Milano. Since the loading condition of axles on the bench can be assimilated to a three point rotating bending, the shape of natural fatigue cracks is expected to be similar to the convex artificial defects shown in Figure 2b, i.e. semi-elliptical or semi-circular. Inspections were carried out during crack propagation tests so having the possibility to measure evolving cracks changing their dimensions with the increasing number of cycles. In this case, inspections could only be carried out considering the 2nd leg configuration.
The obtained results are shown in Figures 3b and 3c in terms of depth and reflecting area, respectively. As expected, it is possible a saturation level above which cracks and defects can be assimilated to big reflectors. In the case of depth, the behaviour of fatigue cracks is very different from that of artificial defects, as often reported in the literature [13]. Considering, instead, the reflecting area, the correlation seems to get much better, suggesting that the reflecting properties of a natural fatigue crack are similar to that of artificial defects (this conclusion cannot be generalised to other kinds of natural cracks).

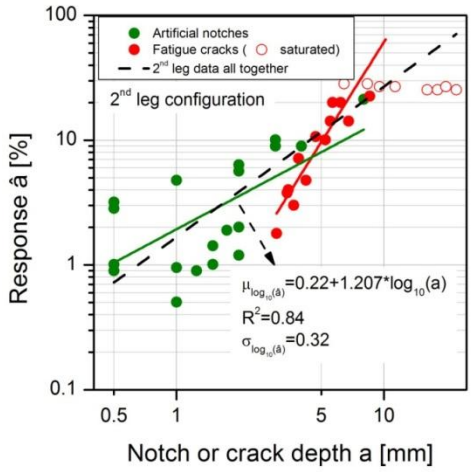
3. Numerical simulations for POD curves

The obtainment of the results described in the previous section required a big experimental effort in terms of time and costs. This problem of deriving POD curves is well known in the NDT community, so, recently, cheaper methods, alternative to the pure empirical one, are

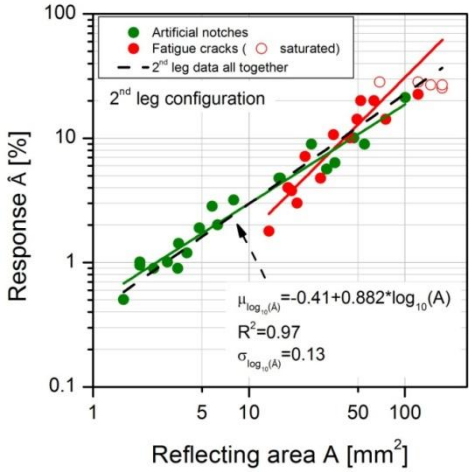
being researched and MAPOD is one of them. Particularly, POD curves are based on the statistical distribution of defect responses which, on the other hand, are controlled by a number of factors related to the details of the adopted NDT procedure. Today, the effects of many of these factors can be simulated by proper numerical models and MAPOD takes the maximum advantage of this possibility. Unfortunately, as already stated in the Introduction, MAPOD does not allow to eliminate all the experiments because today not all of the factors can be described by known physical models. It is also important to remark that MAPOD can be applied to every NDT technique, not only to UT considered in the present research.



(a)



(b)



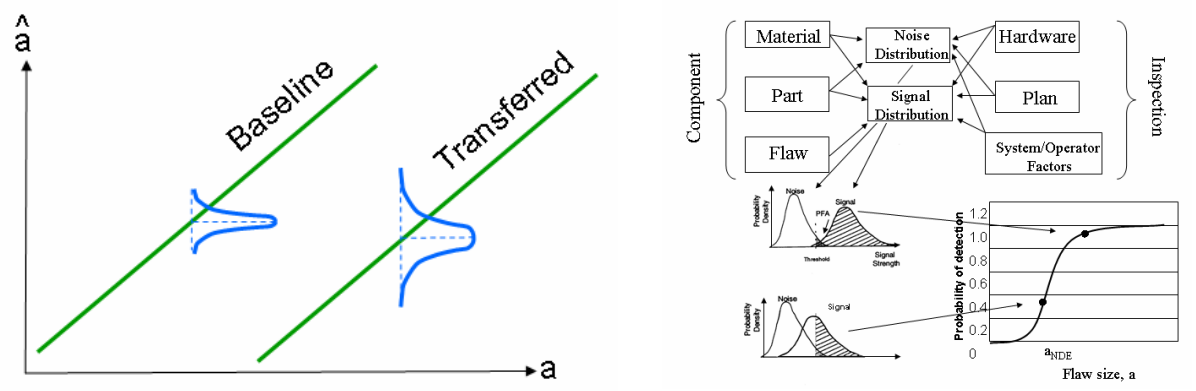
(c)

Figure 3. UT responses of natural fatigue cracks: a) full-scale axle; b) results in terms of depth; c) results in terms of reflecting area.

At the moment, two different MAPOD variations (Fig. 4) are reported in the literature [10], but, in the future, it is possible that they will become two interpretations of the same process. The first variation is called "transfer function" (Fig. 4a) because it suggests to use numerical models in order to ease the transferability of POD curves obtained in a given condition to other conditions where some parameters are changed. The second variation, called "complete MAPOD" (Fig. 4b), foresees a minimum set of experiments in order to characterise those inspection factors not described by known physical models together with numerical simulations for all the other factors.

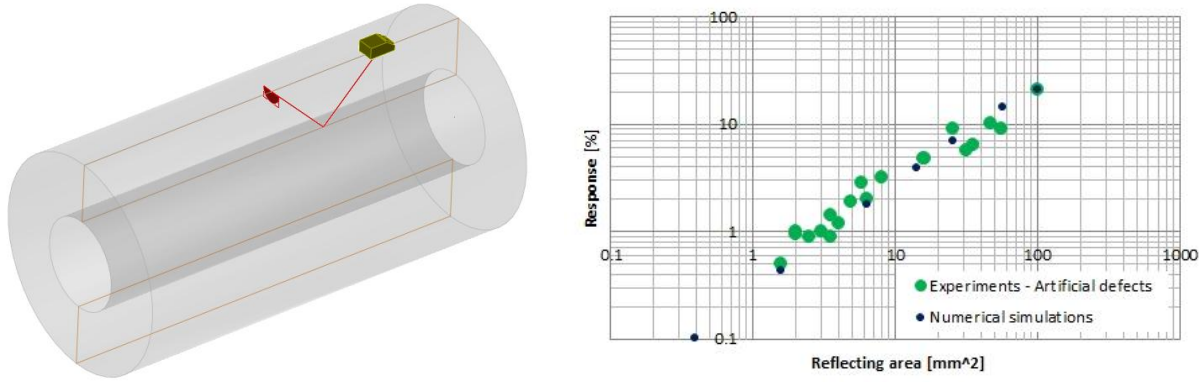
Both MAPOD variations can be successfully applied to the UT inspection of railway axles. In the present research, numerical simulations are carried out by means of the CIVA^{nde} 10.1 dedicated software package [14], able to simulate UT, ET and RT inspections. First of all, the calibration of the numerical model (Fig. 5a) was carried out simulating the 2nd leg inspection of the convex artificial defect having radius equal to 8 mm and consisted in looking for and setting the numerical gain able to yield an A-Scan height equal to the experimental one. Keeping constant the numerical gain, a series of convex semi-circular defects with increasing reflecting area (radius from 0.5 mm to 8 mm) was then simulated and the results are shown in Figure 5b. As it can be seen, numerical responses are very well aligned with the experiments,

suggesting the exactness of the calibration and the correct set-up of the numerical model for the following analyses. It is also worth remarking that the choice of considering convex numerical defects is not relevant in the reflecting area approach, the defect shape could have been anyone, but the convex one is the simplest to implement.



(a) (b)

Figure 4. Two MAPOD approaches [10]: a) transfer function; b) complete approach.



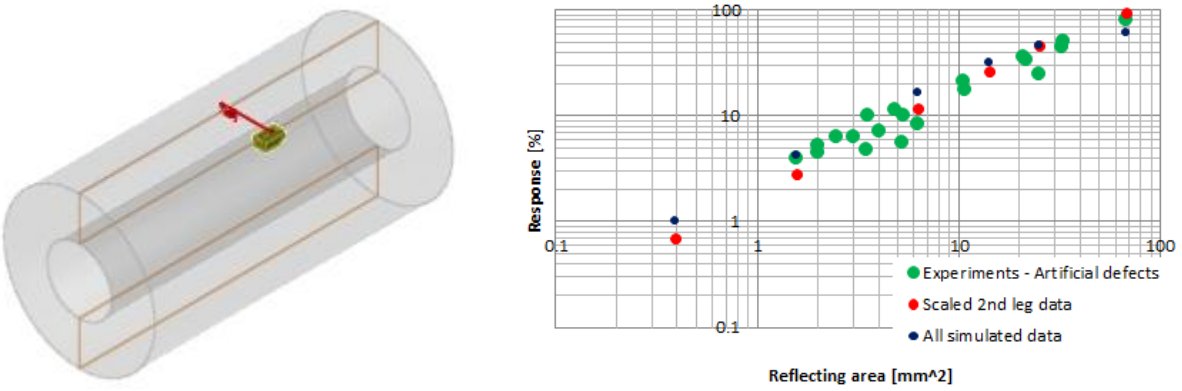
(a) (b)

Figure 5. Calibration of the numerical model: a) numerical set-up; b) comparison between 2nd leg experiments and numerical results.

3.1. "Transfer function" approach

As an applicative example of the "transfer function" approach, the calibrated numerical model was used in order to simulate the UT response of the same convex defects used for the calibration, but, in this case, inspected in the 1st leg configuration. In this way: i) a condition similar to the calibration is considered but with two significantly different factors (the time of flight and the absence of a sound reflection on the axle bore); ii) the experimental responses in 1st leg are available to validate the numerical results. The numerical set-up is shown in Figure 6a where the differences with the calibration set-up are the location of the probe (inside the bore) and the shape of the Plexiglas wedge (from concave to convex). All the other factors were unchanged, included the reference numerical gain determined during the calibration.
In Figure 6b, two different sets of numerical results are compared to experimental data: a) the first one, named "All simulated data", where each numerical datum was achieved by a dedicated CIVA simulation; b) the second one, named "Scaled 2nd leg data", where only the different amplitude response of the biggest reflector whose reflecting area is not saturated on the sound beam (i.e. the convex defect with R=4 mm and reflecting area=25.1 mm^2) is estimated by CIVA and then all the other amplitudes are achieved by vertically scaling the 2nd

leg numerical data of the calibration by this difference. As it can be seen, numerical simulations are very well aligned with experiments for both considered sets, with the best performance provided by the scaled data: this result supports the idea that, in absence of experimental data, a proper application of the numerical model could allow to determine the POD curve of 1st leg inspections starting from the experimental calibration of defect inspected in 2nd leg configuration.



(a) (b)

Figure 6. Application of the "transfer function" approach: a) numerical set-up; b) comparison between 1st leg experiments and numerical results.

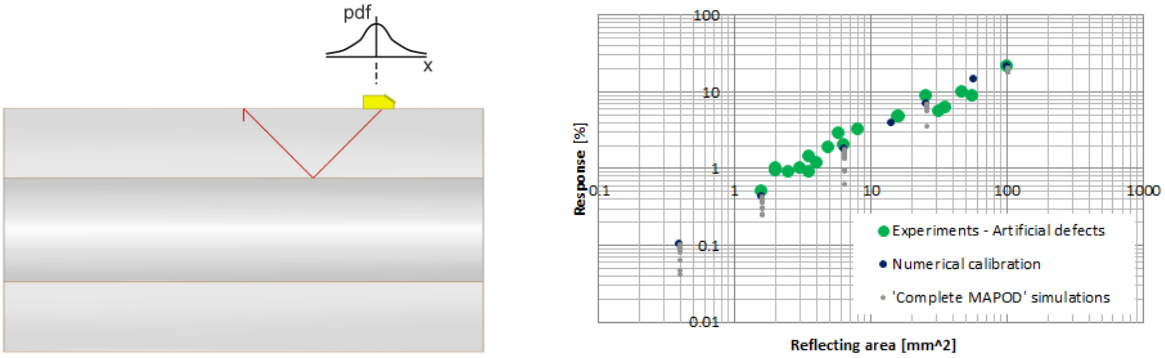
3.2. "Complete MAPOD" approach

The numerical results shown in Figures 5b and 6b are surely useful, interesting and represent a first effective MAPOD tool to evaluate the performance of the considered inspection methodology. Unfortunately, they do not provide information about the intrinsic variability of experimental results (as it is evident from the significantly different standard deviations between experiments and simulations) because they are obtained considering "ideal" conditions where, for example, the coupling is perfect, the probe position is exactly the one maximizing the UT response, etc. The consequence is that the confidence band of numerical results is fictitiously narrow and absolutely not representative of the experimental one. In the necessity to characterize also the confidence of POD curves, the application of the "complete MAPOD" approach then suggests to model, during simulations, every source of variability of the NDT response by means of a suitable statistical distribution. It is worth remarking again that not all the sources of variability are today describable by an effective and well known statistical model.

The numerical procedure adopted here is essentially based on the Monte Carlo method: before each numerical run, a value of the considered source (or sources) of variability is randomly extracted from a suitable statistical distribution and adopted to define the numerical set-up. At the end of all the scheduled runs, the obtained scattered results should be representative of the effects of the chosen source of variability on the results themselves which could be used to define the POD curves with the corresponding confidence band.

For simplicity and with a demonstration aim, a single source of variability is here considered: the position of the probe along the longitudinal axis of the axle with respect to the condition of maximum UT response from the defect. The first scenario consisted in simulating the manual inspection of axles, in a 2nd leg configuration, ideally carried out by a single crystal probe and a flaw detector, similar to those used for the experimental activity (see Section 2), applied from the external surface (Fig. 7a). In this case, the statistical distribution describing the longitudinal position of the probe was assumed to be a Gaussian one having mean equal to

the position of the maximum UT response and coefficient of variation CV=0.05. It is important to add that such distribution was assumed but not experimentally verified and the chosen low standard deviation represents the fact that the inspection should be carried out by qualified personnel able to avoid gross mistakes in individuating the maximum UT response. Five convex defects having R=0.5, 1, 2, 4 and 8 mm were considered and, for each of them, thirty numerical runs were carried out, so totalizing 150 runs. This number of analyses for each defect was so chosen because the minimum number of data, for a given defect, needed for the definition of the 95% confidence band is 29. Figure 7b shows the obtained results. The best performance, i.e. the highest numerical responses, corresponds to the calibration points since they were obtained positioning the probe in the maximising location. For all the other numerical results, the response is lower because the sound beam does not impact the defect in the optimum way. Generally, the obtained numerical standard deviation seems to be significant, comparable to that of experimental results and spread on about half a vertical axis decade.



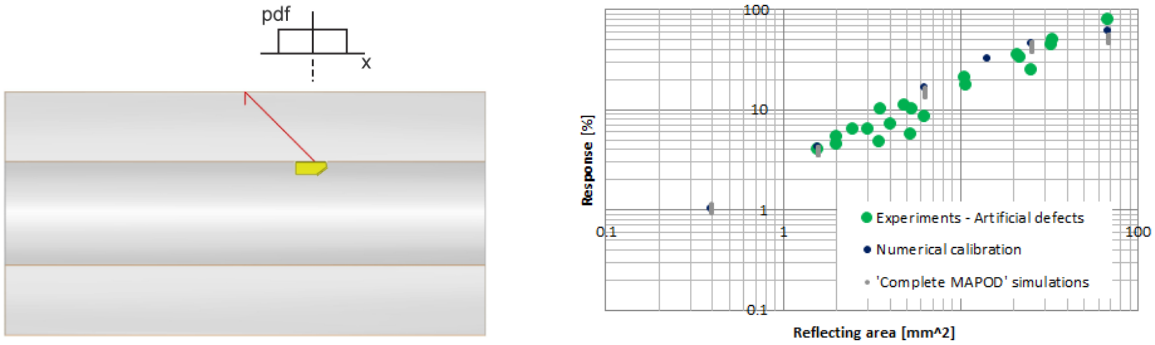
(a) (b)

Figure 7. "Complete MAPOD" approach applied to the manual inspection of axles from the external surface.

The second analyzed scenario considered the automatic inspection of the axle, by a boreprobe (a 45° probe with frequency equal to 4 MHz), in a 1st leg configuration (Fig. 8a). In this second case, the statistical distribution describing the longitudinal position of the probe was assumed to be a Uniform one characterized by a range of possible positions, around the maximizing one, equal to ±2.5 mm. This choice is due to the fact that, during automatic inspections, the probe advances in discrete steps (equal to 5 mm in the present research, a value corresponding to the advancement applied during in-service inspections) and, consequently, the defect can be located everywhere, inside this range, with the same probability. It is then intuitive that, in this scenario, the obtained standard deviation of the results is strictly related to the width of the chosen advancement step, more than to the operator's skill. All the other numerical parameters and the analysis procedure were kept equal to those of the first scenario. Figure 8b shows the obtained results. As it can be seen, the obtained numerical standard deviation seems to be significantly lower than that of the first scenario, suggesting that, as expected, the automatic inspection introduces less variability than the manual one. This conclusion is also supported by the observation that the 1st leg experimental data shown in Figure 8b were also obtained by manual inspection and their standard deviation is significantly higher than the one obtained by numerical results of automatic inspection.

It has to be added that another way to analyze the results obtained from the boreprobe, and to simulate them, is in a "hit/miss" fashion ([3-5]): the defect is only recorded in terms of "detected" or "not detected". This approach requires completely different mathematical and

statistical models and is not faced here, even if some work is being carried out in order to consider it, too.
It is, finally, worth remarking that more complex numerical results based on more sources of variability can be achieved generalising the described procedure: before a numerical run, a value must be extracted from each of the statistical distributions representing each of the considered sources of variability.



(a) (b)

Figure 8. "Complete MAPOD" approach applied to the automatic inspection of axles from the bore surface.

4. Derivation of POD curves

Signal response ("Â vs. A", where Â is the response obtained from the defect having reflecting area equal to A [4]) data were mathematically characterized applying a normal statistical model [3] based, on a bi-logarithmic plot, on a linear trend of the mean $\mu_{\log10(\hat{A})}$ with flaw dimension and a constant standard deviation $\sigma_{\log10(\hat{A})}$:

$$\begin{cases} \mu_{\log_{10}(\hat{A})} = \beta_0 + \beta_1 \cdot \log_{10}(A) \\ \sigma_{\log_{10}(\hat{A})} = \beta_2 \end{cases} \quad (1)$$

where the parameters β_0, β_1 and β_2 were determined interpolating data by means of the maximum likelihood method.
Generally, considering signal response data, a defect is regarded as "detected", if Â exceeds some pre-defined "decision threshold" $\hat{A}_{th}$ corresponding to the response of the defect to be detected with 50% probability ([3-5]). The POD curve can then be built, for each flaw dimension, calculating the probability of a given defect responding with an energy higher than the chosen threshold [3]:

$$POD(A) = Pr\left[\log_{10}(\hat{A}) > \log_{10}(\hat{A}_{th})\right] \quad (2)$$

which can be written as:

$$POD(A) = 1 - F\left\{\frac{\log_{10}(\hat{A}_{th}) - \left[\beta_0 + \beta_1 \cdot \log_{10}(A)\right]}{\beta_2}\right\} = F\left\{\frac{\log_{10}(A) - \left[\frac{\log_{10}(\hat{A}_{th}) - \beta_0}{\beta_1}\right]}{\frac{\beta_2}{\beta_1}}\right\} \quad (3)$$

Eq. (3) represents the cumulative log-normal distribution characterized by the following mean μ and standard deviation σ:

$$\mu = \frac{\log_{10}\left(\hat{a}_{th}\right) - \beta_0}{\beta_1} \qquad \sigma = \frac{\beta_2}{\beta_1} \tag{4}$$

where the β parameters are those previously derived. It is also evident that the choice of the decision threshold can be particularly critical on the achieved POD curves. The analysis of the effects of this choice is beyond the aims of the present research, so the following analyses are based on the response of the saw-cut having depth equal to 1 mm, one of the most traditional calibration defects for railway axles.

Figure 9a shows the comparison of the POD curves obtained from the artificial defects inspected in 2nd leg configuration and the numerical calibration results. No confidence bounds are reported for the numerical results due to the too few considered data. As it can be seen, the comparison is very satisfying because the numerical POD curve falls inside the 95% confidence band of the experimental one so making the numerical results statistically indiscernible. The steeper slope of the numerical POD curve can be ascribed to the significantly lower standard deviation due to the reasons already explained in Section 3.2.

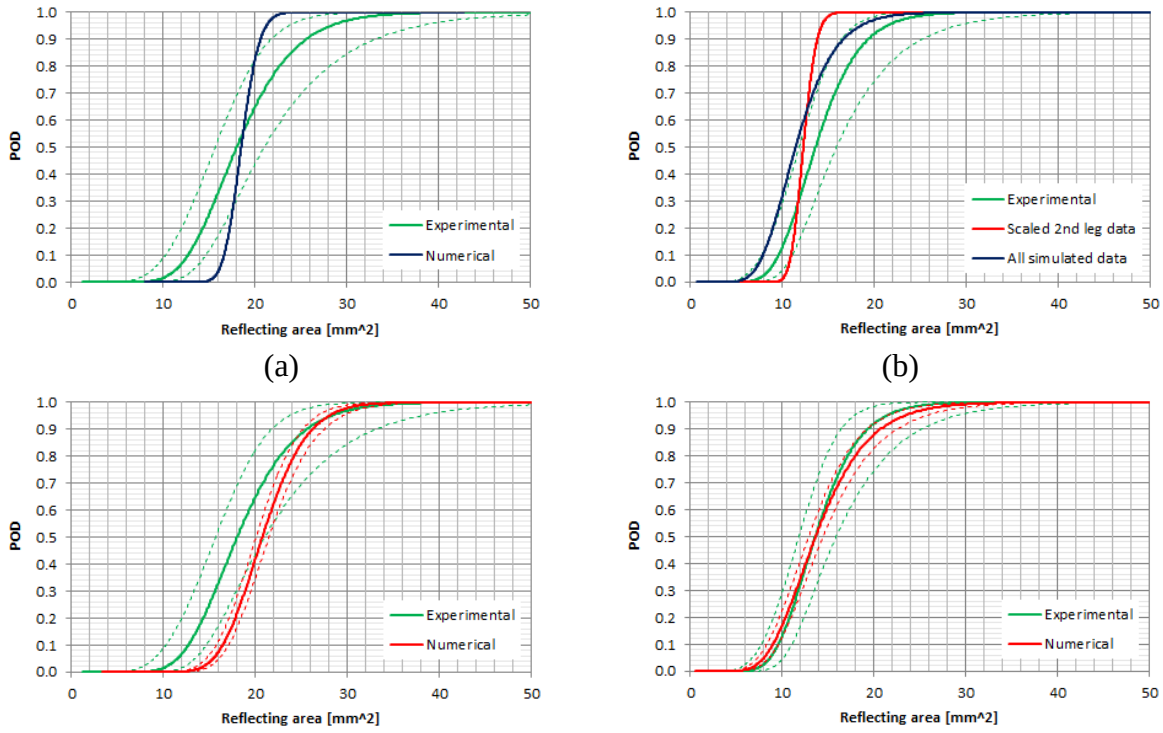


Figure 9. POD curves: calibration; b) transfer function; c) and d) complete MAPOD.

The same comments can be done observing Figure 9b, which shows the comparison between the experimental POD for the artificial defects inspected in 1st leg configuration and the two different sets of numerical results obtained using the "transfer function" approach: the numerical curves seem to intersect with the lower confidence limit of the experimental data, suggesting again the statistical identity of the POD curves. Moreover, in the case of the "All simulated data", the standard deviation of numerical results is higher than in both the calibration and the "scaled 2nd leg data" analyses, so that the slope of the curve is more or less the same of the experimental POD curve. Moreover, comparing the performance of the POD curve from 1st leg inspections and that of the 2nd leg ones (i.e. with the intermediate reflection on the bore), the former is significantly better.

Finally, Figures 9c and d show the comparison between the experimental and the numerical data considered in the "complete MAPOD" approach. In both cases, the numerical POD curves is very well aligned with the experimental one, even if the numerical standard deviations resulted to be somehow lower.

5. Concluding remarks

After a short summary of the "reflecting area" approach recently proposed by the authors for a more effective interpretation of UT responses, the application of the MAPOD approach, based on the numerical modelling of the phenomenon, was presented together with its usefulness, especially considering the possibility to lower the number of experiments needed to define POD curves. The obtained results seem to be very encouraging considering both the "transfer function" approach and the "complete MAPOD" one. In both cases, it was possible to obtain a good correspondence between experimental evidences and numerical results of UT inspection of railway axles, but more research is needed since the novelty of this modelling methodology.

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. John Wiley & Sons Inc., Hoboken, 2003.
3. Georgiou GA. Probability of Detection (POD) curves: derivation, applications and limitations. Research Report 454, HSE Books, Health and Safety, Executive, UK, 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. S. Cantini, M. Carboni, S. Beretta, 'NDT reliability and inspection intervals assessment for a safe service of railway axles', Proc. 16th Int. Wheelset Congress, 2010.
7. M. Carboni. A critical analysis of ultrasonic echoes coming from natural and artificial flaws and its implications in the derivation of probability of detection curves. Accepted for publication on Insight, 2012.
8. S. Cantini, S. Beretta (Editors), Structural reliability assessment of railway axles, LRS-Techno Series 4, 2011.
9. http://www.cnde.iastate.edu/MAPOD/index.htm.
10. R.B. Thompson, L. Brasche, J. Knopp, J. Malas, 'Use of physics-based models of inspection processes to assist in determining probability of detection', Proc. Aging Aircraft Conference, 2006.
11. J.S. Knopp, J.C. Aldrin, E. Lindgren, C. Annis, 'Investigation of a model-assisted approach to probability of detection evaluation', AFRL-ML-WP-TP-2006-494 Report, AIR FORCE RESEARCH LABORATORY, 2006.
12. S. Cantini, S. Beretta, M. Carboni, 'POD and inspection intervals of high speed railway axles', Proc. 15th Int. Wheelset Congress, 2007.
13. J.R. Birchack, C.G. Gardner, 'Comparative ultrasonic response of machined slots and fatigue cracks in 7075 aluminum', Materials Evaluation 1976;34(12):275-280.
14. CEA/CEDRAT, CIVA nde 10.1 User's Manual, 2011.