

Theoretical Explanation for the Empirical Probability of Detection (POD) Curve: A Neural Network-Motivated Approach

Roberto Osegueda¹ and Yohans Mendoza^{2,1}

¹FAST Center for Structural Integrity of Aerospace Systems

²Department of Computer Science

University of Texas at El Paso, El Paso, TX 79968, USA, osegueda@utep.edu, ymendoza@cs.utep.edu

Abstract—For non-destructive testing of aerospace structures, it is extremely important to know how the probability of detecting a fault depends on its size. Recently, an empirical formula has been found which described this dependence. In this paper, we provide the theoretical justification for this formula by using methods motivated by the neural network approach.

I. FORMULATION OF THE PROBLEM

For non-destructive testing of aerospace structures (see, e.g., [3]–[7], [9], [11]), it is extremely important to know how the probability $p(a)$ of detecting a fault of linear size a depends on this size a . This dependence is called a *probability of detection* (POD) curve. Recently, an empirical formula has been found which described this dependence [2], [7], [8]:

$$p(a) = \frac{A \cdot a^\beta}{1 + A \cdot a^\beta}. \quad (1)$$

Since important decisions are based on this formula, it is desirable to find out how reliable it is, i.e., whether it is a crude empirical approximation or a precise formula which has deep theoretical justifications.

II. WHAT WE ARE PLANNING TO DO

In this paper, we show that this formula (1) can indeed be theoretically justified. Our justification for this formula will use methods motivated by the neural network approach (see, e.g., [10]).

III. WE MUST CHOOSE A FAMILY OF FUNCTIONS, NOT A SINGLE FUNCTION

A. POD can be, in principle, experimentally determined

For practical applications, we need the function $p(a)$ which would determine the probability that if a sample with a fault size a is presented to a certain NDE technique, then this fault will be detected. In order to determine this function empirically, we must have a statistics of samples which were presented to this techniques and for which, later on, the fault was discovered; from this statistics, we can determine the desired probability.

B. POD depends on the pre-selection procedure

This probability, however, depends on how we select the samples presented to the NDE techniques. For example, most structures are inspected visually before using a more complicated NDE technology. Some aerospace structures are easier to inspect visually, so we can detect more faults visually, and only harder-than-usual faults are presented to the NDE technique; as a result of this pre-selection, for such structures, the success probability $p(a)$ is lower than in other cases. Other structures are more difficult to inspect visually; for these structures, all the faults (including easy-to-detect ones) are presented to the NDE techniques, and the success probabilities $p(a)$ will be higher. In view of this pre-selection, for one and the same NDE technique we may have different POD functions depending on which structures we apply it to. So, instead of looking for a *single* function $p(a)$, we should look for a *family* of POD functions which correspond to different pre-selections.

C. Relation between POD curves corresponding to different pre-selection procedures

How are different functions from this family related to each other? Pre-selection means, in effect, that we are moving from the original *unconditional* detection probability to the *conditional* probability, under the condition that this particular sample has been pre-selected. In statistics, the transformation from an unconditional probability $P_0(H_i)$ of a certain hypothesis H_i to its conditional probability $P(H_i|S)$ (under the condition S that a sample was pre-selected) is described by the Bayes formula

$$P(H_i|S) = \frac{P(C|H_i) \cdot P_0(H_i)}{\sum_j P(S|H_j) \cdot P_0(H_j)}.$$

In mathematical terms, the transformation from $p(a) = P_0(H_i)$ to $\tilde{p}(a) = P(H_i|S)$ is *fractionally linear*, i.e., has the form $p(a) \rightarrow \tilde{p}(a) = \varphi(p(a))$, where

$$\varphi(z) = \frac{k \cdot z + l}{m \cdot z + n}$$

for some real numbers k , l , m , and n .

D. Resulting description of the desired family of POD functions

So, instead of looking for a single function $p(a)$, we should look for a family of functions $\{\varphi(p(a))\}$, where $p(a)$ is a fixed function and $\varphi(z)$ are different fractionally linear transformations. In the following text, when we say “a family of functions”, we will mean a family of this very type.

IV. WHICH FAMILY IS THE BEST?

A. We can have many different optimality criteria

Among all such families, we want to choose the best one. In formalizing what “the best” means we follow the general idea outlined in [10]. The criteria to choose may be:

- approximation accuracy (i.e., accuracy with which these functions approximate the empirical data about the dependence of the POD of the fault size),
- computational simplicity, or
- something else.

B. Non-numeric criteria are possible

In mathematical optimization problems, *numeric* criteria are most frequently used, when to every family we assign some value expressing its performance, and choose a family for which this value is maximal. However, it is not necessary to restrict ourselves to such numeric criteria only. For example, if we have several different families that have the same approximation accuracy A , we can choose between them the one that has the minimal computational complexity C . In this case, the actual criterion that we use to compare two families is not numeric, but more complicated:

A family F_1 is better than the family F_2 if and only if:

- either $A(F_1) > A(F_2)$,
- or $A(F_1) = A(F_2)$ and $C(F_1) < C(F_2)$.

C. A general description of optimality criteria

A criterion can be even more complicated than above. What a criterion *must* do is to allow, us for every pair of families (F_1, F_2) , to tell:

- whether the first family is better with respect to this criterion (we’ll denote it by $F_1 > F_2$),
- or the second is better ($F_1 < F_2$),

- or these families have the same quality in the sense of this criterion (we’ll denote it by $F_1 \sim F_2$).

Of course, it is necessary to demand that these choices be consistent, e.g., if $F_1 > F_2$ and $F_2 > F_3$ then $F_1 > F_3$.

D. A criterion must choose a unique optimal family

A natural demand is that this criterion must choose a *unique* optimal family (i.e., a family that is better with respect to this criterion than any other family).

The reason for this demand is simple:

- If a criterion does not choose any family at all, then it is of no use.
- If several different families are “the best” according to this criterion, then we still have a problem to choose among those “best”. Therefore, we need some additional criterion for that choice. For example, if several families turn out to have the same approximation accuracy, we can choose among them a family with minimal computational complexity. So what we actually do in this case is abandon that criterion for which there were several “best” families, and consider a new “composite” criterion instead: F_1 is better than F_2 according to this new criterion if either it was better according to the old criterion or according to the old criterion they had the same quality and F_1 is better than F_2 according to the additional criterion.

In other words, if a criterion does not allow us to choose a unique best family it means that this criterion is not ultimate; we have to modify it until we come to a final criterion that will have that property.

E. A criterion must be scale-invariant

The next natural condition that the criterion must satisfy is that the relative quality of the two families should not depend on the choice of the units in which we measure the size of the fault.

Suppose that instead of the original unit of length, we consider a new unit of length which is λ times larger than the original one. How will the POD curve change, i.e., what will be the new function $\tilde{p}(\tilde{a})$ describing the dependence of the probability of detection on the size \tilde{a} in the new units?

One new unit is equal to λ old units, therefore, the length \tilde{a} in the new units means the length $a = \lambda \cdot \tilde{a}$ in the old units. So, the probability $\tilde{p}(\tilde{a})$ is equal to $\tilde{p}(\tilde{a}) = p(\lambda \cdot a)$.

This argument can be used to motivate that the criterion is invariant with respect to rescaling transformations.

We arrive at the following definitions:

Definition 1.

- By a *probability function*, we mean a smooth monotonic function $p(a)$ defined for all $a \geq 0$ for which $p(0) = 0$ and $p(a) \rightarrow 1$ as $a \rightarrow \infty$.
- By a *family of functions* we mean the set of functions that is obtained from a probability function $p(a)$ by applying fractionally linear transformations.
- A pair of relations $(<, \sim)$ is called *consistent* if it satisfies the following conditions: (1) if $F < G$ and $G < H$ then $F < H$; (2) $F \sim F$; (3) if $F \sim G$ then $G \sim F$; (4) if $F \sim G$ and $G \sim H$ then $F \sim H$; (5) if $F < G$ and $G \sim H$ then $F < H$; (6) if $F \sim G$ and $G < H$ then $F < H$; (7) if $F < G$ then $G < F$ or $G \sim F$ are impossible.
- Assume a set \mathcal{F} is given. Its elements will be called *alternatives*. By an *optimality criterion* we mean a consistent pair $(<, \sim)$ of relations on the set \mathcal{F} of all alternatives. If $F > G$, we say that F is *better* than G ; if $F \sim G$, we say that the alternatives F and G are *equivalent* with respect to this criterion. We say that an alternative F is *optimal* (or *best*) with respect to a criterion $(<, \sim)$ if for every other alternative G , either $F > G$ or $F \sim G$.
- We say that a criterion is *final* if there exists an optimal alternative, and this optimal alternative is unique.

In the present section we consider optimality criteria on the set \mathcal{F} of all families.

Definition 2. Let $\lambda > 0$. By the λ -rescaling $S_\lambda(p)$ of a function $p(a)$, we mean a function $\tilde{p}(a) = p(\lambda \cdot a)$. By the λ -rescaling $S_\lambda(F)$ of the family F , we mean the family of the functions that are obtained from $p \in F$ by λ -rescaling.

Definition 3. We say that an optimality criterion on \mathcal{F} is *scale-invariant* if for every two families F and G and for every number $\lambda > 0$, the following two conditions are true:

- if F is better than G in the sense of this criterion (i.e., $F > G$), then $S_\lambda(F) > S_\lambda(G)$;
- if F is equivalent to G in the sense of this criterion (i.e., $F \sim G$), then $S_\lambda(F) \sim S_\lambda(G)$.

Theorem. If a family F is optimal in the sense of some optimality criterion that is final and scale-invariant, then every function p from F is equal to (1) for some A and $\beta > 0$.

1. Let us first show that the optimal family is scale-invariant.

Indeed, we assumed that the optimal family F_{opt} exists and is *scale-invariant* in the sense that $S_\lambda(F_{opt}) = F_{opt}$ for all real numbers $\lambda > 0$. Indeed, we assumed that the optimality criterion is final, therefore there exists a unique optimal family F_{opt} . Let's now prove that this optimal family is scale-invariant (this proof is practically the same as in [10]). The fact that F_{opt} is optimal means that for every other F , either $F_{opt} > F$ or $F_{opt} \sim F$. If $F_{opt} \sim F$ for some $F \neq F_{opt}$, then from the definition of the optimality criterion we can easily deduce that F is also optimal, which contradicts the fact that there is only one optimal family. So for every F either $F_{opt} > F$ or $F_{opt} = F$.

Take an arbitrary λ and let $F = S_\lambda(F_{opt})$. If $F_{opt} > F = S_\lambda(F_{opt})$, then from the invariance of the optimality criterion we conclude that $S_{\lambda^{-1}}(F_{opt}) > F_{opt}$, and that conclusion contradicts the choice of F_{opt} as the optimal family. So $F_{opt} > F = S_\lambda(F_{opt})$ is impossible, and therefore $F_{opt} = F = S_\lambda(F_{opt})$, i.e., the optimal family is really scale-invariant.

2. Let us now show that the functions from the optimal family satisfy a certain functional equation.

Due to Part 1 of this proof, if the function $p(a)$ belongs to the optimal family F_{opt} , then, for every $\lambda > 0$, the re-scaled function $p(\lambda \cdot a)$ also belongs to F_{opt} , i.e., due to definition of a family, there exist values $k(\lambda)$, etc., for which

$$p(\lambda \cdot a) = \frac{k(\lambda) \cdot p(a) + l(\lambda)}{m(\lambda) \cdot p(a) + n(\lambda)}. \quad (2)$$

3. Let us now deduce the actual form of the functions p from the optimal family.

The solution to the functional equation (2) is, in essence, described in [1]. For completeness, let us describe the proof in detail.

For $\lambda = 1$, we have $n = 1$, so, since p is smooth (hence continuous), for $\lambda \approx 1$, we have $n(\lambda) \neq 0$; hence, we can divide both the numerator and the denominator of (2) by $n(\lambda)$ and thus, get a similar formula with $n(\lambda) = 1$.

If we multiply both sides of the resulting equation by the denominator, we get the following formula:

$$m(\lambda) \cdot p(a) \cdot p(\lambda \cdot a) + p(a) = k(\lambda) \cdot p(a) + l(\lambda).$$

If we fix λ and take three different values of a , we get three linear equations for determining three unknowns $k(\lambda)$, $l(\lambda)$, and $m(\lambda)$, from which we can determine these unknowns using Cramer's rule. Cramer's rule expresses every unknown as a fraction of two determinants, and

these determinants polynomially depend on the coefficients. The coefficients either do not depend on λ at all (like $p(a)$) or depend smoothly ($p(\lambda \cdot a)$ smoothly depends on λ because $p(a)$ is a smooth function). Therefore, these polynomials are also smooth functions of λ , and so are their ratios $k(\lambda)$, $l(\lambda)$, and $m(\lambda)$.

Now that we know that all the functions in the equations (2) are differentiable, we can differentiate both sides with respect to λ and set $\lambda = 1$. As a result, we get the following differential equation:

$$a \cdot \frac{dp}{da} = C_0 + C_1 \cdot p + C_2 \cdot p^2$$

for some constants C_i . To solve this equation, we can separate the variables, i.e., move all the terms related to a to one side and all the terms related to p to the other side, and get the differential equation

$$\frac{dp}{C_0 + C_1 \cdot p + C_2 \cdot p^2} = \frac{da}{a}. \quad (3)$$

Let us first show that $C_2 \neq 0$. Indeed:

- If $C_2 = 0$ and $C_1 = 0$, then $p/C_0 = \ln(a) + \text{const}$, which contradicts to our assumption that $p(0) = 0$.
- If $C_2 = 0$ and $C_1 \neq 0$, then we get $C_1^{-1} \cdot \ln(C_1 \cdot p + C_0) = \ln(a) + \text{const}$ hence $C_1 \cdot p + C_0 = A \cdot a^\alpha$, which for $\alpha < 0$, contradicts to the assumption that $p(0) = 0$, and for $\alpha > 0$, contradicts to the assumption that $p(a) \rightarrow 1$ as $a \rightarrow \infty$.

Thus, the case $C_2 = 0$ is impossible, and $C_2 \neq 0$. For $C_2 \neq 0$, in general, the left-hand side of the equation (3) can be represented as a linear combination of elementary fractions $(p+z_1)^{-1}$ and $(p+z_2)^{-1}$ (where z_i are – possibly complex – roots of a quadratic polynomial $C_1 + C_1 \cdot p + C_2 \cdot p^2$):

$$\frac{1}{C_0 + C_1 \cdot p + C_2 \cdot p^2} = c \cdot \left(\frac{1}{p + z_1} - \frac{1}{p + z_2} \right)$$

(the case of a double root can be handled in a similar manner). Thus, integrating the equation (3), we conclude that

$$c \cdot \ln \left(\frac{p + z_1}{p + z_2} \right) = \ln(a) + \text{const},$$

and

$$\frac{p + z_1}{p + z_2} = P \cdot a^\beta$$

for some A and β . So, the expression $A \cdot a^\beta$ can be obtained from $p(a)$ by a fractional linear transformation; hence, by applying the inverse transformation (and it is known that the inverse to a fractionally linear transformation is also fractionally linear) we conclude that

$$p(a) = \frac{A \cdot a^\beta + B}{C \cdot a^\beta + D}$$

for some numbers A , B , C , and D . One can easily check that only for *real* values $A \div D$ and β , we get a monotonic everywhere defined function $p(a)$.

If $\beta < 0$, then we can multiply both numerator and denominator by $a^{-\beta}$ and get a similar expression with $\beta > 0$. Thus, without losing generality, we can assume that $\beta > 0$.

Now, the condition that $p(0) = 0$ leads to $B/D = 0$ and hence, to $B = 0$. The condition leads to $A = C$, i.e., to

$$p(a) = \frac{A \cdot a^\beta}{A \cdot a^\beta + D}.$$

Since $p(a)$ is not identically equal to 1, we have $D \neq 0$. Therefore, we can divide both the numerator and the denominator of this fraction by D , and get the desired expression (1). The theorem is proven.

ACKNOWLEDGMENT

This work was supported in part by the Future Aerospace Science and Technology Program (FAST) Center for Structural Integrity of Aerospace Systems, effort sponsored by the Air Force Office of Scientific Research, Air Force Materiel Command, USAF, under grant number F49620-95-1-0518. The authors are thankful to Vladik Kreinovich for valuable discussions.

REFERENCES

- [1] J. Aczel, *Lectures on functional equations and their applications*, New York, London: Academic Press, 1966.
- [2] P. Barbier and P. Blondet, *Using NDT techniques in the maintenance of aeronautical products*, Aerospatiale France Report No. 93-11587/1/GAL, 1993.
- [3] D. E. Chimenti, "Guided waves in plates and their use in materials characterization", *Appl. Mech. Rev.*, vol. 50, pp. 247–287, 1997.
- [4] R. W. Clough and J. Penzien, *Dynamics of Structures*, New York: McGraw Hill, 1986.
- [5] C. Ferregut, R. Osegueda, and A. Nuñez (eds.), *Proceedings of the International Workshop on Intelligent NDE Sciences for Aging and Futuristic Aircraft*, El Paso, TX, September 30–October 2, 1997.
- [6] I. Grabec and W. Sachse, *Synergetics of Measurement, Prediction, and Control*, Berlin-Heidelberg: Springer-Verlag, 1997.
- [7] X. E. Gros, *NDT Data Fusion*, London: J. Wiley, 1997.
- [8] P. W. Hovey and A. P. Berens, "Statistical evaluation of NDE reliability in the aerospace industry", *Review of Progress in QNDE*, New York: Plenum, vol. 7B, pp. 1761–1768, 1988.
- [9] A. K. Mal and S. J. Singh, *Deformation of Elastic Bodies*, Englewood Cliffs, NJ: Prentice Hall, 1991.
- [10] H. T. Nguyen and V. Kreinovich, *Applications of continuous Mathematics to Computer Science*, Dordrecht: Kluwer, 1997.
- [11] I. A. Viktorov, *Rayleigh and Lamb Waves: Physical Theory and Applications*, N.Y.: Plenum Press, 1967.