R. B. Thompson's contributions to model assisted probability of detection

W. Q. Meeker

Citation: AIP Conf. Proc. **1430**, 83 (2012); doi: 10.1063/1.4716215
View online: http://dx.doi.org/10.1063/1.4716215
View Table of Contents: http://proceedings.aip.org/dbt/dbt.jsp?KEY=APCPCS&Volume=1430&Issue=1
Published by the American Institute of Physics.

R. B. THOMPSON'S CONTRIBUTIONS TO MODEL ASSISTED PROBABILITY OF DETECTION

W. Q. Meeker[1]

[1]Center for Nondestructive Evaluation and Department of Statistics, Iowa State University, Ames, IA 50011

ABSTRACT. Traditional empirical studies to estimate probability of detection (or POD) are expensive and time consuming. Over the past thirty years, much progress has been made in the use of physics-based models to predict POD. A deterministic model for flaw response can be combined with a probability distribution for inspection variabilities to provide a model-based POD. Actual inspections, however, involve complicated variabilities from a variety of sources and modeling all of the important ones, and especially human factors variabilities, would be difficult or impossible. Bruce Thompson's knowledge of physics, probability, statistics and industry needs gave him the insights to pioneer and subsequently serve as the leader in the important area that is now called "Model Assisted POD" or MAPOD. The basic idea of MAPOD is to find an appropriate combination of a physics-based model, combined with limited (usually by time and cost constraints) experimental data and statistical modeling to establish POD. This talk will outline Bruce Thompson's important contributions to this area.

Keywords: MAPOD, Physics Based Model, POD, Hard Alpha Inclusion, Censored Data
PACS: 43.60.UV, 43.60.Cg, 81.70.Cv, 02.50.Sk

INTRODUCTION

Background and Overview

From 1993 to 2010, I had the opportunity and great pleasure of working directly with Bruce Thompson, providing statistical support on a large number of projects. I learned a tremendous amount from Bruce, particularly about nondestructive evaluation (NDE) and probability of detection, and a little bit about physics. In this paper I will provide a brief description of the basic ideas behind probability of detection (POD) and the various ways by which one can obtain POD. This will be followed by a description of model-assisted POD (MAPOD) and an outline of Bruce's important contributions to the area. I will also present several abbreviated examples of applications of MAPOD that Bruce and I (and others) worked on.

Review of Progress in Quantitative Nondestructive Evaluation
AIP Conf. Proc. 1430, 83-94 (2012); doi: 10.1063/1.4716215

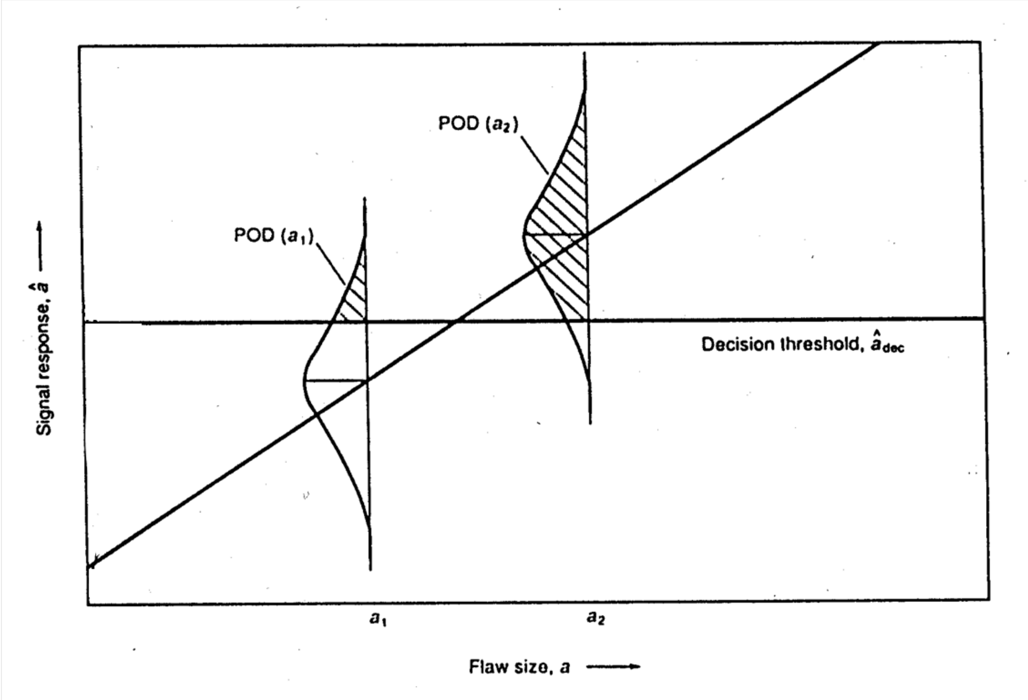

FIGURE 1. Graphic showing the relationship between inspection signal response and probability of detection.

What is POD

POD is the most commonly used metric for inspection capability in NDE. For example, in aerospace applications POD is an input to the processes for making accept/reject criteria decisions, scheduling inspections, doing lifing calculations, and performing risk analyses. It is important to obtain good estimates of POD and also to quantify uncertainty in POD.

I am sure that Figure 1 was originally drawn by Bruce by hand. This plot illustrates the basic ideas behind POD by showing the relationship between signal response and flaw size and how the decision threshold will affect POD. For certain kinds of inspection technologies, as the flaw size increases, so does the signal response. If the signal from the inspection is above the decision threshold, we would say that there has been a detection. If we were to add a distribution of noise to the plot, we could also visualize the probability of a false alarm (PFA). It is important to recognize that in any kind of detection problem there is a tradeoff between the probability of detection and probability of a false alarm. Decreasing the detection threshold will increase the probability of detection, but also increase the probability of a false alarm. Usually the detection threshold is chosen in order to control the PFA to an acceptable level.

Proper estimation of POD involves thinking about variability, as described by the probability distributions in Figure 1. Potentially, there are multiple sources of variability in NDE applications. These include the setup of the inspection involving such operations as positioning a transducer and the part to be inspected in the inspection system. Different transducers have different characteristics and will give different strength signals for a given target, even after calibration. And of course there is variability among calibration blocks and in the calibration process itself. There is also variability in material grain structure and in flaw morphology. Many different kinds of human factors variabilities also come into play. For example, in some applications, certain operators get consistently higher signals than others. The variability in all of these factors helps to determine whether

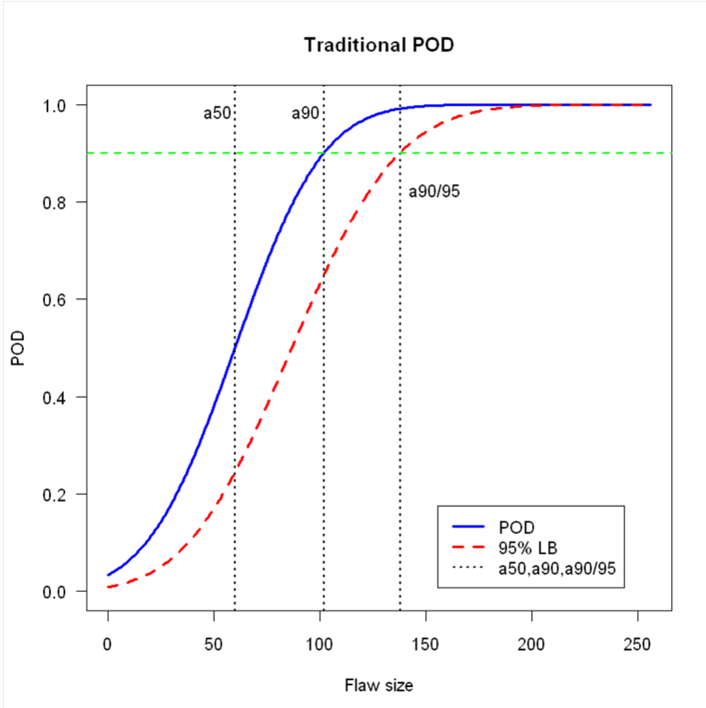

FIGURE 2. Example of a POD curve with a lower 95% confidence bound.

one will detect a crack or not, and can be characterized by probability distributions like those shown in Figure 1.

If the true flaw is at size a_1 in Figure 1, the probability of detection, represented by the relative area above the threshold is rather small, perhaps 0.15. For the larger flaw of size a_2, the probability is detection is much larger, perhaps 0.90. Using this approach to computing POD for all flaw sizes would lead to a plot like Figure 2 where POD is represented by the solid curve.

The vertical line on the left intersects the POD curve at probability 0.50 and the corresponding value on the Flaw size axis gives what is commonly called a_{50} (approximately 57). The middle line intersects the POD curve at probability 0.90 and the corresponding value on the Flaw size axis gives what is commonly called a_{90} (approximately 102).

When we have limited data (actually, we always have limited data), there is limited information about probability of detection. From the limited data, we can estimate the probability of detection, but there is statistical uncertainty in the estimate. The dashed curve in Figure 2 gives a lower 95% confidence bound on POD, expressing the "statistical uncertainty" due to limited data. For any given flaw size, we are 95% confident that the true (unknown) POD is larger that the dashed line. The vertical line on the right intersects the POD lower confidence bound curve at probability 0.90 and the corresponding value on the Flaw size axis gives what is commonly called $a_{90/95}$ (approximately 140). The $a_{90/95}$ value is a 95% upper confidence bound on the flaw size that can be detected with probability 0.90.

HOW DOES ONE FIND POD?

As mentioned in the introduction, there are three methods that might be used to compute POD. In this section I will briefly describe each one.

The Traditional Empirical Method

When a new inspection technology is to be employed, it is generally necessary to quantify the POD of the new technology. Typically this has been done with a purely empirical POD study. The empirical method involves the inspection of a substantially large number of flaws (typically 60 to 100) in a specimen set or in a block containing multiple flaws (e.g., flat-bottom holes or inclusions), each of a known size (e.g., length of a crack or area of an inclusion). Then the specimens or block is inspected, perhaps multiple times.

Usually a POD study is an expensive experimental program in which a set of specimens or an inspection block needs to be produced. Furthermore one must design the study and it is very important to capture all the important sources of variability that are present in the actual inspection process. If all sources of variability are not represented in the study, the POD estimate will be biased. In some cases, in order to obtain information on more sources of variability the specimens or block may be inspected multiple times by different operators and perhaps even at different locations.

It is also possible to assess POD from field finds, but doing so requires special considerations and special statistical methods. Technical details are given in Burkel, Sturges, Tucker, and Gilmore (1996) and Wang and Meeker (2006).

Figure 3 illustrates empirical POD analysis. These eddy current data come from the first edition of Mil-HDBK 1823 (1989). The two smaller probability distributions represent the probability distribution of noise signals that one would see when inspecting material with no cracks. As in Figure 1, the other probability distributions represent signals from cracks of different sizes. The line going through these probability distributions represents the 50% (or median) of the probability distributions. These distributions and the line are estimated from the available POD data, indicated by the X's and triangles.

The X's represent actual measurements of signal level for different flaw sizes. The downward-pointing triangles represent misses indicating that the signal was below threshold and "in the noise." The misses provide important information and should not be thrown away. Such observations are known as "left censored" and there are special statistical methods for handling such observations.

Similarly, the upward-pointing triangles are "saturated observations," meaning all we know is the signal strength is higher than the saturation level. Technically, we refer to these observations as "right censored" and again, they contain useful information but require special statistical methods.

Again, the probability of detection at any given crack size is computed by the relative area above the detection-threshold line. For a crack of size 0.005 inches, this probability is approximately 0.50. For a crack of size 0.01 inches, the probability of detection is very close to 1.0. For the distribution on the right, the probability detection is, for all practical purposes, equal to 1. The entire POD curve is shown in Figure 4.

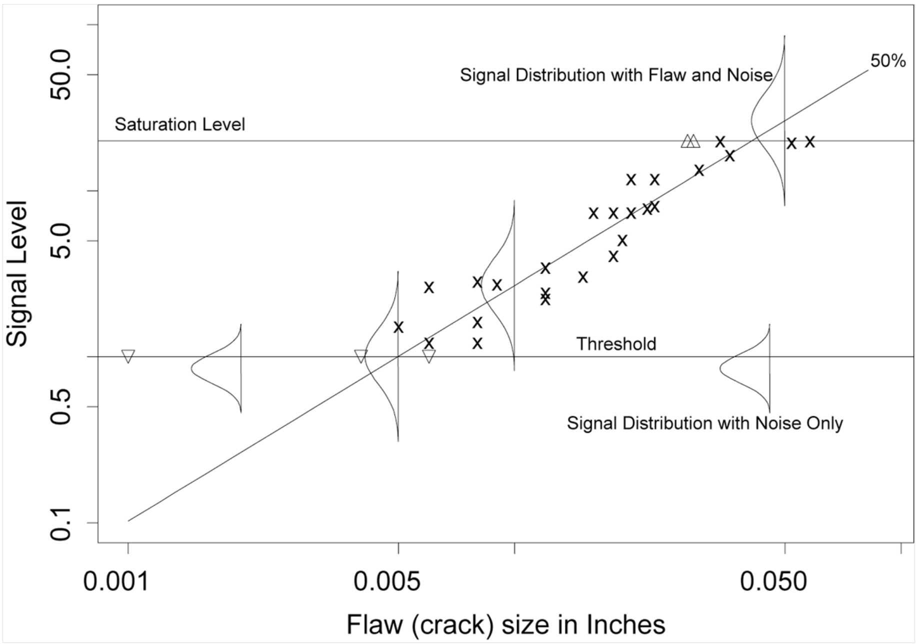

FIGURE 3. Bolt hole eddy current signal response and noise data versus crack length.

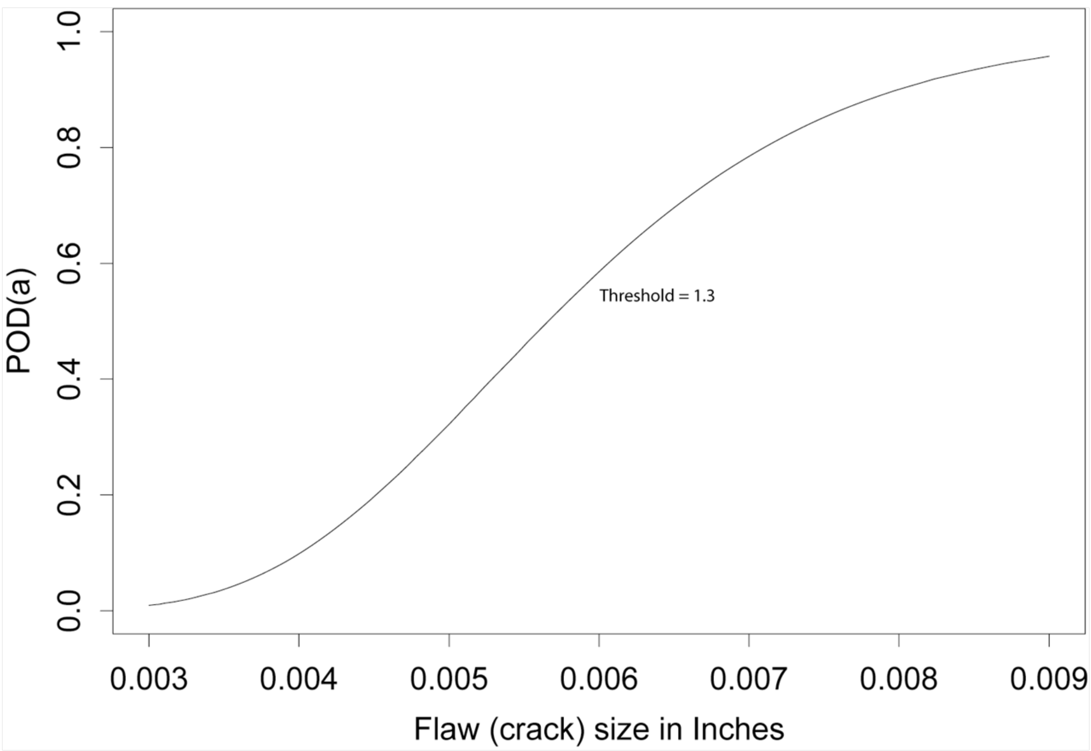

FIGURE 4. Bolt hole eddy current inspection probability of detection versus crack length.

Overall, conducting a POD study is expensive and time consuming. If you are using the empirical method and you do not have the right kind of data (e.g., if your specimen set or experimental block does not contain flaws of sizes over a range of interest) there is danger of incorrect results if your estimates involve extrapolation.

Physics Based Models For POD

Another approach for obtaining a POD is to use knowledge of the physics behind the inspection method. Given a particular type of inspection, characteristics of the target (e.g., materials, size, shape, orientation), the material in which the target resides, and details about the inspection system itself (e.g., in ultrasonic testing further inputs would be characteristics of the ultrasonic transducer and the electrical system) it is possible to build a physics-based model that will predict the signal strength. Much of Bruce Thompson's research was focused on how to do this. After these deterministic models had been developed, the next question was, could we apply these models to get probabilities of detection? In principle, one can use a deterministic model to obtain POD if there is knowledge (described by probability distributions) about the variability in all of the inputs to the model.

I remember seeing Bruce draw a simple diagram on the white board and explaining how to do this. If you take a transducer and change the angle, physics can predict the change in the signal strength. So if one understands the variability in adjustment of these transducers, it would be possible to deduce POD from that source of variability just from physics, without having to do a long complicated experiment. The needed information about variability of the transducer positioning could be obtained from simple experiment. The same kind of thing could be done for flaw morphology and flaw angle. As you rotate that flaw around, the signal strength will change, and this change could be translated into a probability distribution for the signal.

There are several reasons why the physics-based approach is not practical. First the physics-based models are not exact. The famous statistician George Box is often quoted as having said "All models are wrong, but some are useful." I worked with Bruce on the problem of quantifying model error. Although it is reasonably easy to quantify "statistical error" arising from limited data, it is more difficult to quantify model errors. A more difficult problem one would face in using physics alone to quantify POD is that it is difficult to get information on and quantify all of the different sources of variability. Human factors variabilities are particularly difficult to quantify.

Model Assisted POD

Bruce Thompson recognized the limitations of both the empirical method and the physics based method of computing POD. Bruce's idea was to combine the knowledge of physics of inspection with data in a way that would lessen the cost of or extend the power of the empirical method. This approach has led to what we now call "model assisted POD" or MAPOD. As I look back at the 15 joint publications that Bruce and I have had over the years, and each one had elements of MAPOD in them. We did not use the term MAPOD in the early days, but that is what we were doing. Over the years, Bruce demonstrated the value of these ideas, leading teams working on a number of real application. One of the first applications in which I was involved was described in Sarkar et al. (1996).

BRUCE THOMPSON'S CONTRIBUTIONS TO MAPOD

In addition to his leadership in numerous applications of MAPOD, Bruce Thompson, for many years lead the MAPOD Working Group. This group of interested NDE professionals would meet once or twice a year to share experiences and discuss methodology. The MAPOD Working Group web pages (maintained by the ISU CNDE) contain a large amount of information about MAPOD, including agendas and most of the materials that were presented at the meetings. The MAPOD Working Group, with Bruce's guidance, identified two general approaches for thinking about and implementing MAPOD. These methods have been codified in documents in the MAPOD Working Group web pages.

The Transfer Function approach to MAPOD uses experimentation to understand the relationship between the signal response of relatively easy to produce synthetic flaws and naturally-occurring flaws that would generally be impossible to much more expensive to manufacture. For example, several studies have been done comparing the signal responses cracks versus notches in certain materials. Such studies provide an understanding of the differences in signal responses (and POD) between how cracks and notches. Then the idea is you can run less expensive experiments on notches in some other material and use that same kind of transfer function model to predict POD for the new material without having to create a new specimen set.. And the same kind of idea holds, for example, when comparing synthetic hard alphas inclusions versus naturally occurring hard alpha defects or the flat bottom holes versus other kinds of real flaws. Because such inferences require a kind of extrapolation, knowledge of the physics of inspection would be important in the implementation of such a method.

A potentially more powerful approach to MAPOD is the Full Model Assisted, or FMA, method. I remember working with Bruce and other scientists in a series of meetings in the large conference room at the CNDE, brainstorming about how to pull together things that we knew into a workable framework for full model assisted POD. Bruce was always intellectual leader for this work, but he liked exchange of thoughts and experiences with others to help refine the ideas. The results of this effort can be found in the MAPOD Working Group web pages.

In order to implement the FMA MAPOD method, it is necessary to have a rather complete model for the particular inspection type that is to be used. Bruce (and his students and collaborators) had done fundamental research on models for ultrasonic testing. We've heard about that work in earlier talks in this session. At CNDE, Bruce also encouraged and supported the same kind of research in other kinds of NDE technologies (e.g. X-ray and eddy current inspection). Thus FMA MAPOD is a generic framework that can be applied to any inspection technology, once model are available to predict signal strength as a function of inspection parameters and flaw characteristics.

In addition to his knowledge about the physics of inspection, Bruce Thompson had a deep understanding of industrial needs, how POD would be used in practice, and he had a solid understanding of the importance of physical experiments and statistical methods. With this combination of knowledge, Bruce successfully led the MAPOD working group from 2003 to 2010.

EXAMPLES OF MAPOD

In this section I will briefly outline two MAPOD applications that Bruce led and where I was involved in the statistical work. Other examples can be found in the MAPOD Working Group web pages and among Bruce’s many publications.

Figure 5 shows Multizone ultrasonic test data on naturally occurring hard alpha flaws. These data were taken in an experiment during the Contaminated Billet Study. The purpose of the study was to obtain “default POD curves” for ultrasonic inspection of titanium billets and is described in full in Thompson et al. (2008). The data in Figure 5 have the usual relationship with amplitude increasing in increasing flaw area and a line that has been fit through the data. The line on the left has a dramatically steeper slope and came from a Physics-based model. The corresponding POD curve is shown in Figure 6.

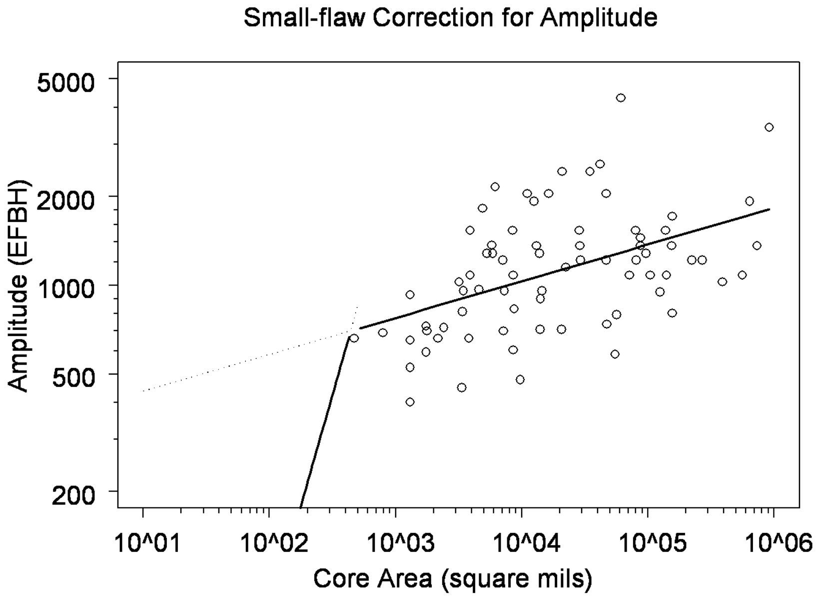

FIGURE 5. Signal response of hard alpha inclusions in a titanium billet.

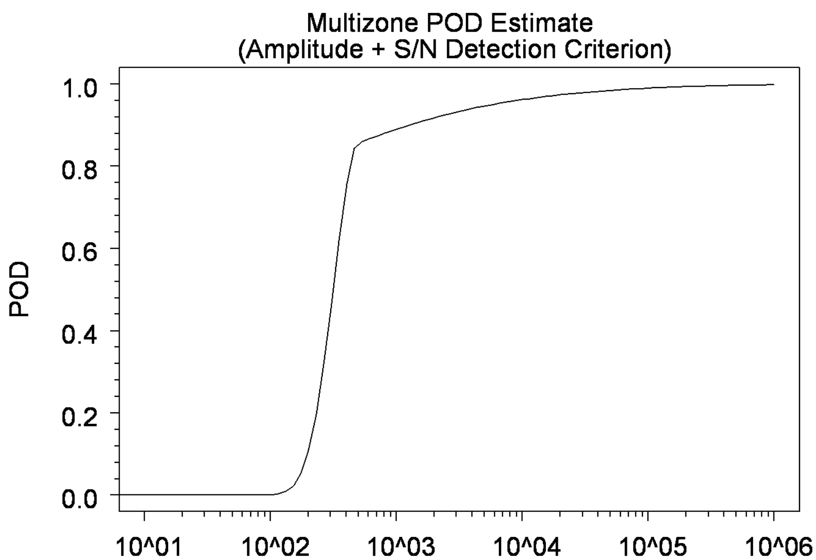

FIGURE 6. Estimate of POD for hard alpha inclusions in titanium billets.

When we were working on this project I, being a statistician, was responsible for the data analysis. I fit a line through the data and my line came down to the left to a point where a flaw with 10 square mils area has a predicted response of 500 EFBH (a common scale for reporting ultrasonic testing amplitude), an extrapolation quite far outside the range of the data. The corresponding POD curve that I presented had a value of approximately 0.40 at 10 square mils. I recall presenting those results to the team and Dick Berkel from GE commented "That POD curve does not make sense because we have never found any flaws anywhere near that size."

Bruce though about this for about 20 seconds and then concluded "You can't extrapolate that line because the reflections need to be described by a different 'small-flaw regime.'" Tim Gray and Bruce developed a physics-based relationship that lead to the more appropriate relationship shown in Figure 5 and the corresponding POD curve in

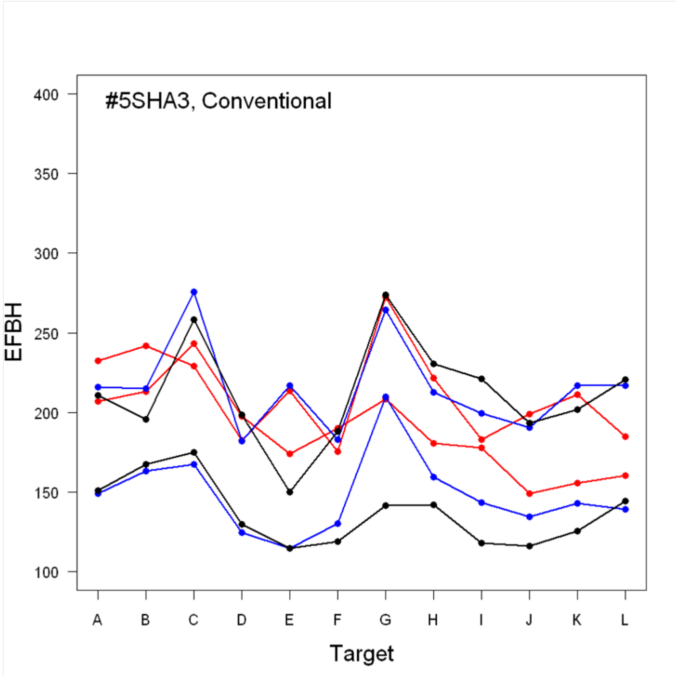

FIGURE 7. Target plot showing variability in nominally similar #5 SHAs and different operators.

Figure 6. This is an excellent example of an application where if you tried to do a purely empirical approach, the answers would be substantially incorrect.

The second example describes a similar study that was conducted to obtain default POD curves for synthetic hard alpha (SHA) inclusions in forgings (few, if any naturally occurring hard alpha inclusions in forgings would have been available for such an experiment) and is described completely in Thompson et al. (2011). The experiment was to be conducted on a sample consisting of a forged titanium-alloy disk (known as the Synthetic Inclusion Disk or SID), produced in a previous research program. This disk contains the artificial flaws (or targets), both SHAs and flat bottom holes. One problem with the SID is that it had only two sizes of SHAs and only two levels of weight percent nitrogen among the targets. It was necessary, however, to estimate POD for SHA sizes outside the range of target sizes. As illustrated in the previous example, such extrapolation can be dangerous. Moreover, Bruce knew that the relationship in this application would also be nonlinear for larger flaws, due to the "beam-limiting" effect, but with only two SHA sizes estimating the nonlinear curve would be difficult or impossible (depending on un-testable assumptions we might be able to make).

The other complicating factor is there was the need to model several sources of variability, illustrated in Figure 7 where we can see these are the differences from target to target and from operator to operator. Such variability can have an important effect on POD.

To capture the nonlinearity due to the beam-limiting effect, we used the Kirchhoff approximation with a reflectance factor, which can be summarized by

$$\mathrm{EFBH} = |R|\frac{\pi w^2}{2}\left(1 - e^{-2(b/w)^2}\right)$$

where b is the diameter of the target, w is the diameter of the ultrasonic beam as it meets the target, and the reflectance factor

$$R = \frac{\rho_i v_i - \rho_m v_m}{\rho_i v_i + \rho_m v_m} \qquad (1)$$

is used to describe the differences in reflectance for flat bottom holes (where R=1) and SHAs with different amounts of weight percent nitrogen. The coefficients in (2) were obtained from experimental work described by Gigliotti, Gilmore, and Perocchi (1994).

Combining (2) with a probability model for the target-to-target random effects described by τ and the operator-to-operator random effects described by γ gives the statistical model

$$\log_{10}\left(\mathrm{EFBH}(x)\right)=\log_{10}(\alpha)+\log_{10}\left(R\bullet\frac{\pi}{2}w^2\left(1-e^{-2(x/w)^2}\right)\right)+\tau+\gamma+\varepsilon$$

$$\text{with}\quad \tau\sim N\left(0,\sigma_\tau^2\right),\ \gamma\sim N\left(0,\sigma_\gamma^2\right),\ \varepsilon\sim N\left(0,\sigma_\varepsilon^2\right)$$

where $N\left(0,\sigma^2\right)$ indicates a Gaussian (or normal) distribution with mean 0 and variance σ^2. This model was fit to all of the available data (including the measurements taken on the flat bottom holes). The fitted values along with the data for the 3% nitrogen SHAs are shown in Figure 8.

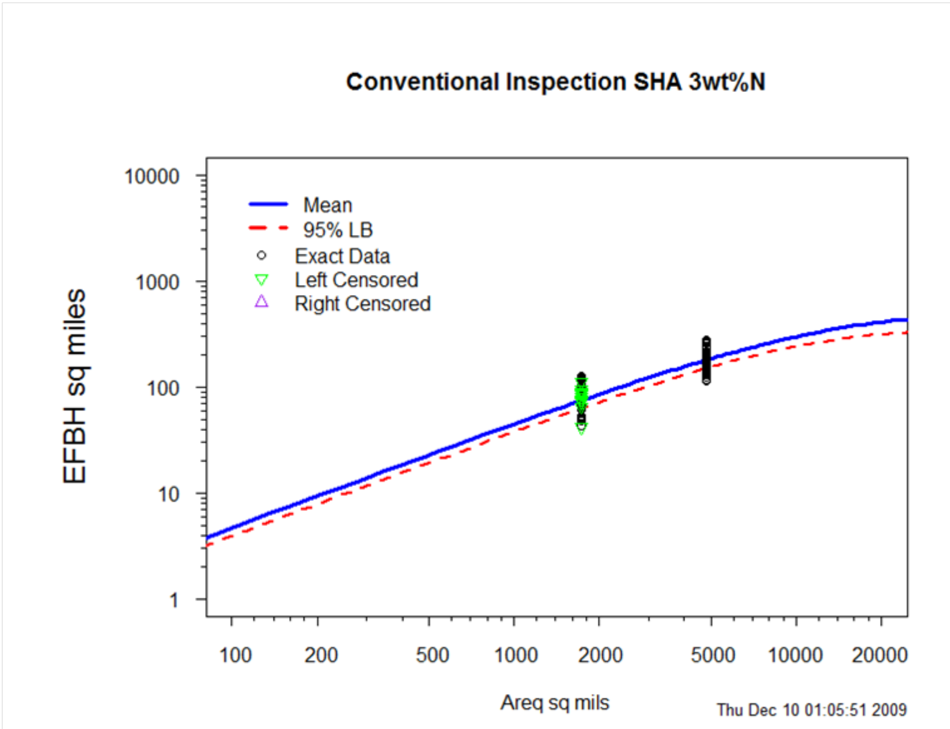

FIGURE 8. Signal response for conventional ultrasonic inspection of 3% synthetic hard alpha inclusions in a titanium forging.

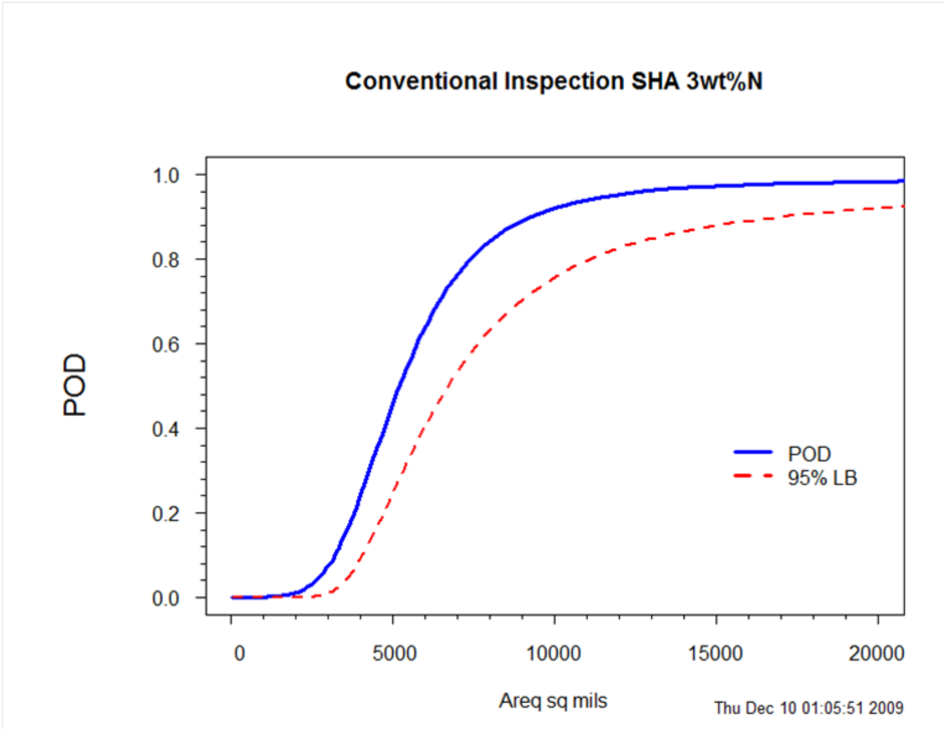

FIGURE 9. POD for Conventional Inspection of Synthetic Hard Alpha Inclusions with 3% Weight Percent Nitrogen.

Figure 9 shows the corresponding plot of the POD function for SHAs with 3% nitrogen, along with a lower confidence bound, again quantifying the uncertainty due to a limited number of targets and a limited number of measurements on those targets. A complete set of figures for all of the target types is given in Thompson et al. (2011).

SOURCES OF FURTHER INFORMATION ON MAPOD

One of the best sources of information about MAPOD is the MAPOD Working Group web site: http://www.cnde.iastate.edu/MAPOD where the MAPOD working group meetings have been fully documented. At this web page there is a list of all of the MAPOD Working Group meetings. If you click on a particular meeting the agenda is displayed. Then clicking on agenda items brings up a copy of the corresponding presentation.

There is also a brief description of the MAPOD procedures in Appendix H of the new version of MIL-HDBK-1823A (2009). Additionally, there are various reports and papers in the QNDE proceedings, beyond those referenced here that have documented particular examples. The early documents will not, however, have the keyword MAPOD, because that is a relatively new term.

CONCLUDING REMARKS

Bruce Thompson made fundamental contributions to the physical theory of ultrasonic testing and probability of detection. He also had the vision for MAPOD and how it would have to be structured to work well. Bruce was primarily responsible for this very generalized version of MAPOD which we call the full model assisted (FMA) approach, which is very carefully described in some of the documents in the MAPOD working group web pages. MAPOD will continue to be important in the modern application of NDE because of the added value and economy that it brings to the discipline.

REFERENCES

1. R. H. Burkel, D. J. Sturges, W.T. Tucker, and R.S. Gilmore (1996) "Probability of Detection for Applied Ultrasonic Inspection," *Review of Progress in Quantitative NDE*, Vol. 15, edited by D. O. Thompson and D. E. Chimenti, Plenum Press, New York, NY, 1996, pp.1991-1998.
2. M.F.X Gigliotti,., R.S. Gilmore, and L.C. Perocchi (1994), "Microstructure and Sound Velocity of Ti-N-O Synthetic Inclusions in Ti-6Al-4V," Metallurgical and Materials Transactions A, 25, 2321-2329.
3. MIL-HDBK-1823 (1989), Non-Destructive Evaluation System Reliability Assessment, Standardization Order Desk, Building 4D, 700 Roberts Avenue, Philadelphia, PA 1911l-5094.
4. MIL-HDBK-1823A (2009), Nondestructive Evaluation System Reliability Assessment, Standardization Order Desk, 700 Roberts Avenue, Philadelphia, PA 19111-5094.
5. R. B. Thompson, T. A. Gray, and W. Q. Meeker (2006), Use of Physics-Based Models to Guide the Extrapolation of Aircraft Engine Ultrasonic POD Data to Small Flaw Sizes, *Review of Progress in Quantitative Nondestructive Evaluation*, Vol. 25, D. O. Thompson and D. E. Chimenti, Eds. (American Institute of Physics, NY), 1878-1885.
6. R. B. Thompson, W. Q. Meeker, L. J. H. Brasche, Ming Li, R. Klaassen, J. Umbach, H. Wasan, W. Hassan, S. Singh, , K. Smith, T. Patton, (2011). Update of Default *Probability of Detection Curves for the Ultrasonic Detection of Synthetic Hard-Alpha Inclusions in Titanium Alloy Forgings.* U.S. Department of Transportation, Federal Aviation Administration (awaiting final review and release).
7. R. B. Thompson, B. Meeker, M. Keller, J. Umbach, T. Chiou, Y. Wang, D. Burkel, W. Hassan, K. Smith, T. Patton, and L. Brasche. (2008), "*Update of Default Probability of Detection Curves for the Ultrasonic Detection of Hard Alpha Inclusions in Titanium Alloy Billets*." DOT/FAA/AR-07/63, U.S. Department of Transportation, Federal Aviation Administration.
8. P. Sarkar, W. Q. Meeker, R. B. Thompson, and W. Junker (1998), Probability of Detection Modeling for Ultrasonic Testing. In Thompson, D.O. and Chimenti, D.E., editors, *Review of Progress in Quantitative Nondestructive Evaluation* 17B, 2045-2052, Plenum Press.