



Pacific Northwest
NATIONAL LABORATORY

Proudly Operated by Battelle Since 1965

Review of Literature for Model Assisted Probability of Detection

September 2014

RM Meyer
SL Crawford

JP Lareau
MT Anderson



Prepared for the U.S. Nuclear Regulatory Commission
under a Related Services Agreement with the U.S. Department of Energy
CONTRACT DE-AC05-76RL01830

U.S. DEPARTMENT OF
ENERGY

DISCLAIMER

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor Battelle Memorial Institute, nor any of their employees, makes **any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights.** Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof, or Battelle Memorial Institute. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

PACIFIC NORTHWEST NATIONAL LABORATORY

operated by

BATTELLE

for the

UNITED STATES DEPARTMENT OF ENERGY

under Contract DE-AC05-76RL01830

Printed in the United States of America

Available to DOE and DOE contractors from the
Office of Scientific and Technical Information,
P.O. Box 62, Oak Ridge, TN 37831-0062;
ph: (865) 576-8401
fax: (865) 576-5728
email: reports@adonis.osti.gov

Available to the public from the National Technical Information Service,
U.S. Department of Commerce, 5285 Port Royal Rd., Springfield, VA 22161
ph: (800) 553-6847
fax: (703) 605-6900
email: orders@ntis.fedworld.gov
online ordering: <http://www.ntis.gov/ordering.htm>



This document was printed on recycled paper.

(9/2003)

Review of Literature for Model Assisted Probability of Detection

RM Meyer
SL Crawford

JP Lareau
MT Anderson

September 2014

Prepared for
the U.S. Department of Energy
under Contract DE-AC05-76RL01830

Pacific Northwest National Laboratory
Richland, Washington 99352

Abstract

This report documents a literature survey of so-called “Model Assisted Probability of Detection” (MAPOD) approaches that may be useful in determining the probability of detection (POD), a metric for quantifying the performance of nondestructive evaluation (NDE) methods. The objective of this report is to summarize MAPOD concepts that have been proposed to date in order to assess specific approaches that may be appropriate for application to improve estimates of POD for field NDE of nuclear power plant components. The limitations of laboratory-based studies to replicate actual field conditions are well-recognized and not limited to the nuclear power industry. Probability of detection estimates based on laboratory studies generally provide ideal environments for performing examinations as compared field settings, which typically will result in non-conservative estimates of POD for field applications. As a consequence of this effort, the authors conclude that use of MAPOD concepts to improve estimates of field NDE performance may require access to certain field data, or alternatively, may require significant laboratory studies to assess the influence of human and environmental shaping factors. If the necessary field data cannot be made available, it is proposed that laboratory efforts will have greater chance of success by focusing on a specific examination technique and component application. The authors also find that MAPOD concepts may have a more immediate contribution to the nuclear power industry through their use in extending personnel and procedure qualifications beyond established limits. Finally, another basis was identified that may be used to adjust POD curves generated in previous reliability studies and performance demonstrations. Many of the previous studies calculated POD as the average of performance data which may be inappropriate for many applications, as the calculation may be non-conservative. A more conservative approach to estimating POD would include basing the calculation on a lower statistical quantile of the data set.

Acronyms and Abbreviations

AFRL	Air Force Research Laboratory
AMSE	American Society of Mechanical Engineers
AP	application parameters
CDF	cumulative distribution function
FAA	Federal Aviation Administration
FMA	full model assisted
HF	human factors
IC	intrinsic capability
ISI	inservice inspection
MAPOD	Model Assisted Probability of Detection
MLE	maximum likelihood estimation
MRR	Mini Round Robin
NASA	National Aeronautics and Space Administration
NDE	nondestructive evaluation
NDT	nondestructive testing
POD	probability of detection
UT	ultrasonic testing
XFM	transfer function

Contents

Abstract	iii
Acronyms and Abbreviations	v
1.0 Introduction	1
2.0 Parametric Probability of Detection Models	2
2.1 POD Model for Binary (Hit/Miss) Representation of NDE Response.....	3
2.2 POD Model for Continuous NDE Response (\hat{a} vs. a Models)	3
3.0 Model Assisted Probability of Detection.....	5
3.1 Non-Bayesian Demonstrations.....	6
3.2 Bayesian Approaches	7
3.3 Use of Field Inspection Data for POD Estimation	9
3.4 POD Based on Lower Quantiles	10
3.5 Data Pooling.....	10
4.0 Summary Human Factors Literature	11
5.0 Discussion and Analysis.....	13
5.1 MAPOD Discussion and Analysis	13
5.2 Discussion and Analysis of Human Factors and Integration with MAPOD	14
5.3 Other Considerations.....	15
6.0 Conclusions and Looking Forward.....	15
7.0 References	16

Figures

1	Illustration of a POD Curve	3
2	Linear Relationship between Natural Log of Response, \hat{a} and Defect Size, a	4

Tables

1	Summary of MAPOD Demonstrations Using Non-Bayesian Techniques for POD Parameter Estimation	6
2	Summary of MAPOD Demonstrations Using Bayesian Techniques for POD Estimation.....	8

1.0 Introduction

This report documents a literature survey of so-called “Model Assisted Probability of Detection” (MAPOD) approaches to assist in determination of probability of detection (POD), which is the de-facto metric for quantifying the performance of nondestructive evaluation (NDE). This literature survey is an effort to summarize the MAPOD concepts that have been proposed to date so that an understanding may be gained as to what approaches may or may not be appropriate for the application of improving the estimates of POD for the field inspection of nuclear power plant components. MAPOD is vaguely defined in the literature and has been described as an (Thompson 2008),

“...approach in which either experiment or physics-based theoretical models (previously validated by experiment) are used to extend the information gained in empirical measurements to new situations.”

In the Department of Defense Handbook for reliability assessment of nondestructive evaluation systems (DoD 2009, MIL-HDBK-1823A), MAPOD is defined as...

“methods for improving the effectiveness of POD models that need little or no further specimen testing”

Elsewhere, it is defined through its motivation for (Aldrin et al. 2012),

“...reducing POD sample number by better managing experimental variation using more accurate statistical models and appropriately designing experiments based on prior experience and data.”

Parts of all of these definitions reflect the goal of this effort, which is to maximize the efficiency of POD determination for the field inspection of nuclear power plant components by leveraging the most appropriate statistical models and existing knowledge and information.

The limitations of laboratory-based studies to replicate actual field conditions is well recognized and not limited to the nuclear power industry. Examples of field conditions that may be difficult or impossible to replicate in a laboratory study include stresses due to organizational pressures (e.g., influence of deadlines, revenue goals, etc.) or the environment (e.g., uncomfortable temperatures, anxiety over radiation exposure). In this context, laboratory studies provide a rather ideal setting for performing an inspection compared to the field setting, which could result in a non-conservative estimation of POD for field applications.

POD is usually presented as a function of a single variable representing flaw size, a , and depicted as an “S-shaped” curve that is monotonically increasing with a . This variable is denoted as $POD(a)$. Typically, the $POD(a)$ for a particular application is determined through the collection of a large amount of empirical data. Guidelines for performing an NDE reliability assessment study based on POD are described in MIL-HDBK-1823K (DoD 2009). Beginning around 1980, several studies were conducted to determine the POD for NDE of nuclear power plant components in an effort to quantify its reliability and effectiveness and understand how it could be improved. Such studies are typically performed by fabricating representative test blocks and having them inspected by several teams. The number of blocks and inspections required is dictated by the amount needed to achieve statistically relevant results. Many

publications allude to the “ $a_{90/95}$ ” criteria for reliability, which is the flaw size at which POD is 90% with a confidence of 95% or greater [although the $a_{90/95}$ has been criticized for its arbitrariness (ENIQ 2010)]. In order to achieve a small 95% confidence interval, a significant number of inspections can be required.

Thus, there has been motivation in recent years to leverage more relevant and appropriate statistical techniques to optimize the efficiency and/or accuracy of POD(a) determination, particularly in the aerospace and defense industries. The MAPOD Working Group was established in 2004 by the Air Force Research Laboratory (AFRL) in collaboration with the Federal Aviation Administration (FAA) and the National Aeronautics and Space Administration (NASA) (Thompson et al. 2009). Since then, similarly themed efforts have emerged, including efforts in France (SISTAE) and Europe (PICASSO) (Dominguez et al. 2012). Although there have been efforts to standardize the MAPOD concept, numerous demonstrations of MAPOD in the literature indicate that applications often have unique considerations or constraints.

The rest of this paper is organized such that an introduction to parametric POD models follows in Section 2.0 and an overview of literature presenting demonstrations of MAPOD is provided in Section 3.0, divided into non-Bayesian demonstrations and Bayesian demonstrations. Section 4.0 is a summary of human factors literature considered relevant to NDE of nuclear power plant components, and Section 5.0 provides a discussion and analysis of MAPOD approaches, and the potential integration of human factor effects with MAPOD. Section 6.0 includes conclusions based on the literature analyzed and possible next steps. References are contained in Section 7.0.

2.0 Parametric Probability of Detection Models

Probability of detection is a widely established method for measuring the performance of NDE systems. Because POD is dependent of flaw size, a , it is represented as a monotonic curve versus a that can range from 0% for flaws below the detection threshold up to 100% for larger flaws. The POD curve generally has an S-shaped appearance as illustrated in Figure 1. The POD curve can be created from curve fitting to NDE data, which can either be reported as binary responses (hit/miss) or as continuous responses in which the response intensity, \hat{a} , has an assumed relationship to the flaw size, a (Berens 1989; Schneider and Rudlin 2004; Guo et al. 2011). The statistical bases for POD curve fits is dependent on the nature of the NDE data responses, resulting in different formulations for POD curves for binary NDE responses in comparison to \hat{a} vs. a models (Guo et al. 2011).

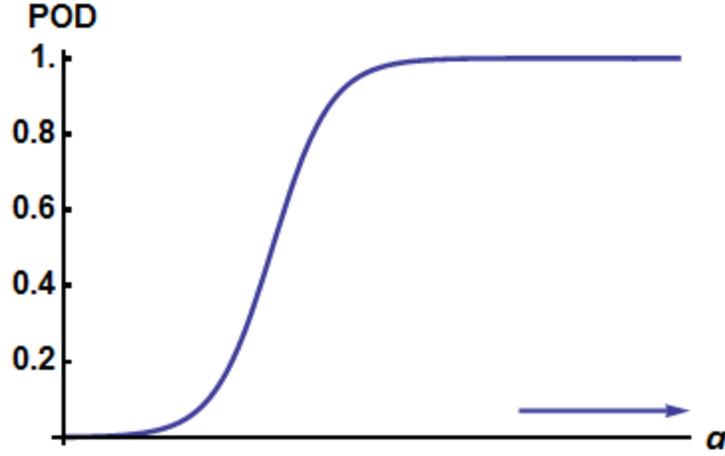


Figure 1. Illustration of a POD Curve

2.1 POD Model for Binary (Hit/Miss) Representation of NDE Response

POD curves resulting from binary NDE responses can be represented mathematically by a logistic function (Berens 1989),

$$\text{POD}(a) = \frac{1}{1 + \exp(-\beta_1 - \beta_2 a)} = \frac{\exp(\beta_1 + \beta_2 a)}{1 + \exp(\beta_1 + \beta_2 a)}. \quad (1)$$

This model also includes two parameters, β_1 , and β_2 , to be determined from curve fitting with empirical data. This model has been used to curve fit binary NDE data to estimate POD for NDE performance studies involving nuclear power plant components (Heasler and Doctor 1996). In this equation, the parameters β_1 and β_2 are determined using maximum likelihood estimation (MLE) (Forsyth and Fahr 1998). The basis for this model is discussed in ENIQ (2010) and it is noted that the parameters β_1 , and β_2 do not have a physical interpretation. This is opposed to POD models from the continuous signal responses (\hat{a} vs. a) in which the model parameters relate the NDE response, \hat{a} , to flaw size, a . In addition to the logistic function, the two parameter probit model has also been used to fit POD to binary NDE response data (Spencer 2001; Bode et al. 2012),

$$\text{POD}(a) = \Phi(\beta_1 + \beta_2 \log(a)). \quad (2)$$

2.2 POD Model for Continuous NDE Response (\hat{a} vs. a Models)

In the case of continuous NDE responses (\hat{a} vs. a models), proportional relationships between \hat{a} and a models are often assumed (see Figure 2), to be of the form (Berens 1989),

$$\hat{a} = \beta_0 + \beta_1 a + \varepsilon, \text{ or,} \quad (3)$$

$$\ln \hat{a} = \beta_0 + \beta_1 \ln a + \varepsilon. \quad (4)$$

where ε represents measurement variability with normal distribution, $\mathcal{N}(0, \delta^2)$ where δ^2 is the variance. The type of relationship depends on the specific type of measurement. If the scenario in Eq. (4) is assumed, the POD curve can be constructed from

$$\text{POD}(a) = P(\hat{a} > \hat{a}_{th} | a) = P(\ln \hat{a} > \ln \hat{a}_{th} | a) = \int_0^{\infty} f(\ln \hat{a} | a) d\hat{a}, \quad (5)$$

where $P(y|x)$ is the probability of y given x , \hat{a}_{th} is a threshold signal response, and $f(\ln \hat{a} | a)$ is a log-normal probability distribution of $\ln \hat{a}$ given a . Thus, the POD is given by the cumulative normal distribution function (Berens 1989),

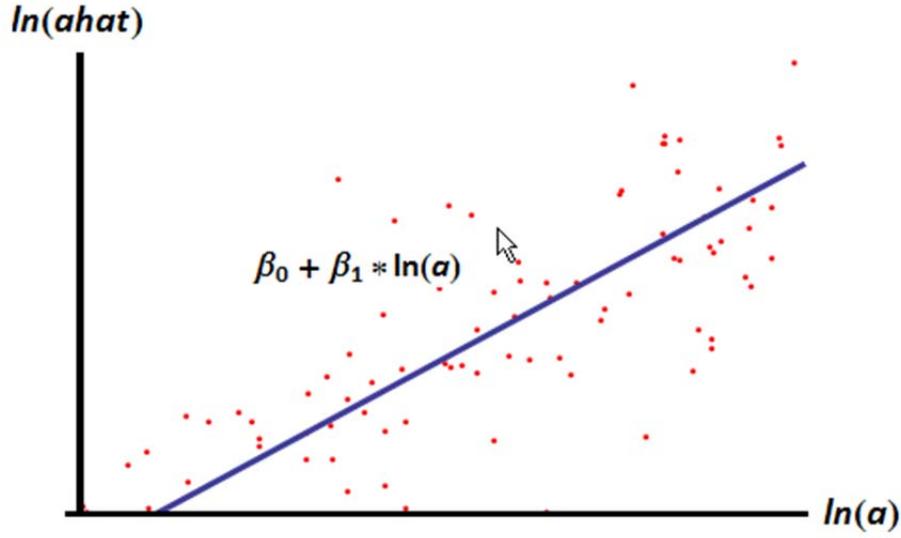


Figure 2. Linear Relationship between Natural Log of Response, \hat{a} and Defect Size, a

$$\text{POD}(a) = \Phi\left(\frac{\ln a - \mu_{th}}{\sigma}\right) = 1 - \Phi\left(\frac{\mu_{th} - \ln a}{\sigma}\right), \quad (6)$$

where $\mu_{th} = \frac{\ln \hat{a}_{th} - \beta_0}{\beta_1}$ and $\sigma = \frac{\delta}{\beta_1}$. Thus, this POD model is defined by four parameters, β_0 , β_1 , a_{th} , and

δ . There can be variations of the \hat{a} vs. a models to represent different relationships between the response and flaw size. For instance, Demeyer et al. (2012) consider an extension of \hat{a} vs. a models to accommodate piecewise linear relationships between $\ln \hat{a}$ and $\ln a$,

$$\ln \hat{a} = \begin{cases} \beta_0^A + \beta_1^A \ln a + \varepsilon_1, & \text{if } a \leq \tau \\ \beta_0^B + \beta_1^B \ln a + \beta_2 (\ln a - \tau) + \varepsilon_2, & \text{if } a > \tau \end{cases} \quad (7)$$

Explicit representation of POD as a function of multiple influencing parameters has been explored in works by Pavlović et al. (2012) and Aldrin et al. (2012). Instead of simply representing POD as a function of a single defect size parameter, it is suggested that POD can also be represented as a function of other relevant parameters such as flaw orientation, surface roughness, etc. In this case, the relationship between the NDE response and multiple influencing parameters, $a_1, a_2, a_3, \dots, a_i$ can be represented as,

$$\hat{a} = \beta_0 + \beta_1 f(a_1, a_2, a_3, \dots, a_i) + \varepsilon \quad (8)$$

The function, $f(a_1, a_2, a_3, \dots, a_i)$, in Eq. (8) can represent a complex physics-based relationship between the NDE response and relevant inspection parameters, in contrast to the linear relationship assumed previously.

3.0 Model Assisted Probability of Detection

As noted in the Introduction, the MAPOD Working Group formed in 2004 and began to formalize MAPOD approaches. MAPOD approaches were initially categorized as “Transfer Function (XFM)” approaches and “Full Model Assisted (FMA)” approaches (Thompson et al. 2009). In the XFM approach, an empirically derived baseline POD curve is leveraged in the estimation of POD for a similar inspection scenario, which differs by one significant controlling factor from the scenario for which the baseline curve was derived. In this case, the baseline curve is leveraged by augmenting the curve with information about how the significant controlling factor affects the POD. This information could be obtained through careful laboratory experiments or through physics-based computer modeling. This approach is most feasible when the controlling factor represents a well-understood physical phenomenon and is particularly applicable to extending qualifications beyond established limits.

In the FMA approach, factors that control variability of the inspection are systematically identified. Physics-based models can be used to estimate the signal response and variability due to factors that are represented by well-understood physical phenomena while the variability of other factors are determined empirically. These approaches have been merged into what has been called the “unified approach” to MAPOD in that the factors that control the variability of a scenario are divided into two groups: (1) those that must be assessed empirically and (2) those that are governed by well-understood physical phenomena.

The following two subsections (Subsections 3.1 and 3.2) are organized according to how information is combined for estimation of POD using MAPOD concepts. In this case, the classification is “Non-Bayesian” and “Bayesian” because many of the references cited in the former category do not explicitly identify how the information is combined, although it is assumed that the information is not combined using Bayesian techniques. However, some of the cited references in this category do describe in detail the combination of POD model parameters through the use of linear regression techniques. Subsection 3.3 discusses POD estimation for data collected by field inspections and some of the challenges associated with this approach. Subsections 3.4 and 3.5 reference documents that describe constructing POD curves from the lower quantiles of data and that describe some of the requirements for pooling different data sets, respectively.

3.1 Non-Bayesian Demonstrations

The MAPOD approach, as described above (either XFM, FMA, or unified), has been illustrated through numerous demonstrations, particularly on aircraft components, using either eddy current or ultrasonic techniques. Nearly all of these demonstrations reviewed in this effort were performed with \hat{a} vs. a data with the exception of Bode et al. (2012). A summary of the non-Bayesian demonstrations reviewed in this effort is provided in Table 1. Thompson et al. (2009) provide a MAPOD demonstration applied to eddy current for detection of fatigue cracks in a complex engine component. In this demonstration, due to the difficulty of artificially growing fatigue cracks in this complex geometry, the POD for electro-discharge machined notches in the complex engine components is determined and used as the “baseline” POD curve. Physics-based modeling is then used to determine the influence of fatigue cracks versus notches on this baseline curve. Thompson et al. (2009) also describe the application of MAPOD to assess the effect of microstructural variability on POD in different alloys for engine disks. In this case, the effects of material grain size on NDE noise level were evaluated using computer simulations while the effects of system variability were evaluated empirically. Thompson et al. (2009) and Smith et al. (2007) also discuss the application of MAPOD to estimate POD for fatigue cracks growing from fastener holes for aircraft wings using ET. In this case, computer modeling was used to estimate the influence of fatigue cracks growing outward from holes on eddy current response while the influence of hole geometry variability on eddy current response was determined empirically.

Table 1. Summary of MAPOD Demonstrations Using Non-Bayesian Techniques for POD Parameter Estimation

MAPOD Approach	NDE Response	NDE Method	Applied for...	Ref.
XFM	\hat{a} vs. a	ET	Fatigue cracks in complex engine components	Thompson et al. (2009)
FMA	\hat{a} vs. a	ET	Fatigue cracks in wing lap joints	Thompson et al. (2009)
FMA	\hat{a} vs. a	UT	Defects in engine disk alloys with microstructural variability	Thompson et al. (2009) Smith et al. (2007)
XFM	\hat{a} vs. a	UT	Fatigue cracks around fastener holes for aircraft	Harding et al. (2009)
XFM	\hat{a} vs. a	UT	Fatigue cracks in aluminum components	Demeyer et al. (2012)
XFM; FMA	\hat{a} vs. a	ET	Fatigue cracks in titanium plates	Rosell and Persson (2013)
FMA	\hat{a} vs. a	ET	Cracks at fastener sites	Aldrin et al. (2009)
XFM; FMA	\hat{a} vs. a	UT	Defects in railway axles	Carboni and Cantini (2012)
XFM	hit/miss	UT	Airplane lap joint specimen sets with multiple site fatigue damage	Bode et al. (2012)

Rosell and Persson (2013) demonstrate estimation of POD for fatigue cracks in titanium alloy plates using eddy current testing by both XFM and FMA approaches. The model is applied to an automated EC procedure in which the probe is raster scanned over the defect. This work noted three kinds of influencing parameters as 1) those associated with the handling and characteristics of the probe and equipment, 2) the sample and component features, and 3) the defect characteristics. However, variables associated with human factors and interpretations of signals were not considered. This work illustrated degradation of POD as the scan index of a raster scan is increased, and it is also noted that further work is

needed for modeling of natural cracks. Aldrin et al. (2009) also demonstrate an FMA approach for estimating POD of an eddy current inspection of cracks at fastener sites.

Carboni and Cantini (2012) apply MAPOD to ultrasonic inspection of defects in railway axles. CIVA, an ultrasonic inspection program developed by the French national laboratory Centre d'Etudes Atomique, is used to simulate ultrasonic inspections performed by the first leg and second leg methods, referring to the direct interaction of ultrasonic energy with a defect and the reflection of ultrasonic energy from the inside bore of the axle prior to defect interaction. Both XFM and FMA approaches were employed with the FMA approach employed to simulate experimental variability such as the impact of the probe location relative to the defect, etc. The XFM approach was applied to relate experimental data from a second leg inspection to a first leg inspection with the aid of computer modeling.

Harding et al. (2009) applied an XFM approach to estimate the POD for fatigue cracks around fastener holes in aircraft wings for an ultrasonic testing (UT) inspection. The model used data obtained from field trials and laboratory experiments taking into account effects of structural geometry, natural variability in fatigue cracks, and human factors in the inspection process. The XFM approach was applied such that the POD for real flaws in a real structure is determined from 1) a set of data for fabricated flaws in a real structure, 2) a set of real flaws in a simplified structure, and 3) fabricated flaws in a simplified structure. The estimation of POD for the target scenario from the other scenarios is based on a linear regression model of POD parameters from the other scenarios. This approach to POD estimation has been referred to as the "quadrant" approach. To incorporate the influence of human factors, the threshold for observation of a defect signal is also a random variable with a distribution that can be characterized. It's noted that this applies to procedures in which the influence of human factors is primarily on the interpretation of acquired data.

Demeyer et al. (2012) apply the XFM approach to estimate the POD for fatigue cracks in aluminum plates based on an empirically derived POD for fatigue cracks in titanium plates. CIVA is used to generate NDE response data for notches in both aluminum and titanium plates. The information from these computer simulations and the empirically derived data for fatigue cracks in titanium plates is extrapolated to estimate the POD for fatigue cracks in aluminum plates using a similar quadrant approach as applied by Harding et al. (2009).

Bode et al. (2012) apply the XFM approach to explore estimation of POD for airplane lap joint specimen sets with multiple site fatigue damage. In this case, a quadrant XFM approach is also applied for inspections using an ultrasonic linear array. In this case, POD parameter estimations for actual in-service cracks in real airplane structures are extrapolated from POD parameters generated for more ideal conditions. Unlike all of the MAPOD demonstrations described previously, this demonstration is based on using binary data ("hit/miss") fit to two parameter "probit" models. POD parameter estimates for the target scenario are constructed from linear regression models of POD parameters determined for the more ideal scenarios.

3.2 Bayesian Approaches

Bayes theorem has also been applied in the literature for the purposes of POD estimation. As simply stated in Leemans and Forsyth (2004), Bayes theorem states that,

“posterior information equals prior information plus new evidence.”

The mathematical formulation of Bayes theorem for POD estimation is presented in numerous papers, including Leemans and Forsyth (2004) and other works cited in this section. A summary of Bayesian demonstrations reviewed in this effort is provided in Table 2.

Leemans and Forsyth (2004) present a Bayesian framework for utilizing field data to estimate POD. The framework is demonstrated on a collection of 30 “hit/miss” visual inspection outcomes and an assumed distribution of POD model parameters (for binary data) with varying levels of uncertainty assumed for the model parameter distributions ranging from narrow to wide and “ignorant.” In this case, a bivariate normal distribution is assumed for prior information of model parameters and the outcomes of the 30 visual inspections are utilized in the likelihood function. The demonstration shows that when prior information is highly uncertain (“ignorant”), the prior information has little influence and the posterior estimation is mostly based on the new evidence. Conversely, when there is high confidence in the prior information, the influence of the prior information dominates the influence of new evidence.

Jenson et al. (2013) present a Bayesian approach for determination of POD of high frequency eddy current inspections on fatigue cracks. In this case, Bayes theorem is used to supplement experimental data with computer model generated data to reduce expenses and time associated with test block fabrication and inspections. Computer modeling is used to estimate the variability of several controlling factors that can be physically modeled. This modeling was used to generate prior information that was combined with some experimental data to obtain posterior estimates of model parameters. In this example, the computer-generated data and experimental data are both presented as binary responses and the logistics function is used for fitting the POD curve.

Table 2. Summary of MAPOD Demonstrations Using Bayesian Techniques for POD Estimation

NDE Response	NDE Method	Description	Reference
Hit/miss	Visual testing	Demonstrate Bayesian approach to determination of field POD based on limited field data	Leemans and Forsyth (2004)
Hit/miss	Eddy current testing	Demonstrate Bayesian approach for POD determination using computer models to generate additional information to supplement limited data from experiment.	Jenson et al. (2013)
\hat{a} vs. a	Radiographic testing	Demonstrate Bayesian approach to POD determination using information from artificial flaws to supplement information from a limited set of real flaws.	Kanzler et al. (2013)
\hat{a} vs. a	Eddy current testing	Demonstrate Bayesian approach for POD determination using computer model generated information.	Aldrin et al. (2012)
\hat{a} vs. a	Not specified	Demonstrate method for determination of both POD and crack size distribution from inservice inspection data.	Hovey (2009)
Hit/miss	Magnetic leakage	Apply a hierarchical Bayes approach to incorporate influence of spatial distributed uncertainties on in-line inspections of pipelines	Dann and Maes (2011)

Kanzler et al. (2013) apply a Bayesian approach for estimating the reliability of NDE for long-term storage of high-level radioactive waste. In this case, measurement data obtained from artificial flaws is combined with measurements obtained on a small amount of real flaws to estimate POD. The results from measurements on artificial flaws are considered the prior information that is updated based on limited real flaw data, through the likelihood function.

Aldrin et al. (2012) introduce a Bayesian method for estimating the POD model parameters for a high-frequency eddy current application based on \hat{a} vs. a responses. A Bayesian method is used to estimate the model parameters for different types of \hat{a} vs. a models including linear single parameter, linear multi-parameter, and physics-based models using data generated by computer simulation. In this case, the linear single parameter model relates the NDE response to a single controlling factor, such as the crack length. In the linear multi-parameter model, the NDE response is related to crack length and crack width. A separate calibration parameter is added to the linear models with each controlling factor that is included. The authors conclude that there is a natural advantage to using Bayes method (as opposed to MLE) for estimating model parameters for increasingly complex models (i.e., physics-based versus linear). The authors also introduced a hierarchical Bayesian method to estimate both the calibration parameters for a physics-based model and the distributions of a few model input parameters (in this case, probe lift-off and crack aspect ratios).

In addition, several documents explore the application of Bayesian techniques beyond the estimation of parameters for simple POD models. Hovey (2009) explored a Bayesian approach for simultaneously estimating the parameters of both the POD model and the crack size distribution from field inspection data for fatigue cracks in aging aircraft. A simulation sensitivity analysis is performed to observe the sensitivity of the posterior distributions to different prior information scenarios including informed prior distributions and incorrect prior distributions. Dann and Maes (2011) introduce a spatial hierarchical Bayes model for magnetic leakage in-line inspection data. The goal of this work is to develop POD estimates with spatial reference as typical POD models often ignore spatially distributed uncertainties.

The latter documents (Hovey 2009; Dann and Maes 2011; Aldrin et al. 2012) make reference to “hierarchical” Bayes methods. These are not formally defined but it may be inferred that hierarchical Bayes methods provide a framework for applying Bayesian approaches to POD models of almost arbitrary complexity. In the references cited, hierarchical Bayes methods are used to increase the descriptiveness of POD models and explicitly represent the variability of certain controlling factors. The development of hierarchical Bayes methods requires a systematic identification of parameters of interest along with models of their relationships to each other and to POD.

3.3 Use of Field Inspection Data for POD Estimation

POD can, in theory, be determined from field test data, based on an assumed flaw growth model (Leemans and Forsyth 2004). In this case, for each flaw detected in an inspection, one can estimate the size of the flaw in previous inspections for which the flaw was “missed.” This data can be used as “missed” data points for constructing a POD curve. The difficulty arises in that this technique has a non-conservative bias in that tested components could contain flaws that have yet to reach a detectable size and these flaws are never included in the data set. In addition, there are several practical challenges such as 1) a component may not have historical test data for estimating misses, 2) the flaw may not have been sized, 3) the testing procedure may change between tests, 4) sizing may be inaccurate, 5) crack growth

data may be unavailable, and 6) there may be too few large flaws in the detected population because large flaws would be detected in previous inspections and not allowed to grow further creating a bias in field inspection data. A report by the North Atlantic Treaty Organization (NATO 2005) echoes these challenges also noting that unknown misses of detectable cracks result in a non-conservative bias in estimated POD and that inspection records can vary significantly in fidelity and quality.

The bias introduced in field data by unknown misses can be corrected for using truncated regression models to fit the field data. The application of truncated regression models to correct for unknown misses in inspection data has been demonstrated for both \hat{a} vs. a NDE responses (Meeker and Thompson 2007) and binary NDE responses (Guo et al. 2011). However, the use of truncated regression models cannot help overcome the other practical challenges noted above.

The NATO report (2005) also describes the estimation of POD from field data from the cumulative distribution function (CDF) of the sizes of detected cracks. The report emphasis is that the crack size CDF is not the same as the POD but can be used as a first estimate of POD for situations in which information is not available for POD estimation. For the example cited in the report, the crack size CDF provided a conservative estimate of POD.

Khan and Liao (2013) apply an iterative procedure for estimating POD based on inspection data. In this method, the CDF of detected cracks is expressed as an integral over the probability density function of total crack size distribution and POD. The parameters of the CDF and POD are iteratively updated until the observed CDF matches the analytical CDF.

3.4 POD Based on Lower Quantiles

Li et al. (2012) illustrate the difference between the influence of uncertainty that results from limited sampling and the influence of the actual variability of data used to determine POD. A significant point of this work is that POD curves derived by averaging the performance results of many inspection teams may not be appropriate because such averaging is actually non-conservative. Depending on the variability of the inspection teams, several teams may perform worse than the average. Thus, Li et al. (2012) propose that it is more appropriate to compute POD based on the 0.05 quantile, instead of average or median (the 0.5 quantile). A curve based on the 0.05 quantile indicates the performance that would be achieved by 95% of the teams, and, thus, is more conservative than POD curves based on the median or average. In this report, a POD model is presented that includes an “operator random effect,” which is represented as a random variable.

3.5 Data Pooling

The NATO report (2005) also discusses “data pooling” which, in this context, refers to the combining of multiple separate NDE data sets to obtain a larger pool of data sufficient to make statistical inferences. The report emphasizes that the pooling of data is valid only if sufficient information is reported in sufficient fidelity and that the datasets to be combined should be based on the same or similar inspection procedures with similar fidelity in the data sets.

4.0 Summary Human Factors Literature

The importance of quantifying the role of human factors in NDE performance has been recognized for several decades. This section summarizes literature discussing human factors and their effects on NDE reliability. Many of the papers cited have been presented at the World Conference on Nondestructive Testing and the American-European Workshop on Non-Destructive Inspection Reliability. An analysis of the literature in this section is provided in Section 5.0.

Spanner et al. (1986) discuss a feasibility study to evaluate the human reliability of an inservice inspection (ISI). They defined five classes of variables that affect human performance: 1) task, 2) procedural, 3) training, 4) environmental, and 5) individual variables and concluded that no single variable is responsible for the wide performance variations noted.

Taylor et al. (1989) presents the findings of a human factors study performed during the mini-round robin (MRR) pipe inspection study. The study found that previous training and experience influenced performance in a positive manner and that coworkers may influence ISI decisions. Overall, it was found that organizational factors, the quality of supervision, the support of organization, and trust in teams all influenced the quality of UT/ISI. The general perception among technicians was that supervisors and utilities were concerned with false calls more than missed calls. In additions, other findings of the study included:

- Better written procedures would improve UT/ISI performance
- Practice on specimens with actual cracks and comparing techniques with other technicians was most valuable for training, whereas classroom lectures were least valuable
- Different technicians often reported different UT/ISI results.

Conclusions were that no single performance-shaping factor is responsible for the wide performance variations observed.

Pond et al. (1998) reviewed human factors research for consideration of upgrading the American Society of Mechanical Engineers (ASME) Boiler and Pressure Vessel Code Section XI requirements or the techniques in ASME Section V. The review included studies of the effects of heat, audible noise, task duration, fatigue, and social environment (co-workers, management, and organizational policies). The review concluded that audible noise had minimal impact on human performance and that the social environment had significant influence on inspector performance. Although it was believed that fatigue has a significant impact on inspection performance, the review was unable to draw any firm conclusions.

Harris and McCloskey (1990) performed a study that provided conclusive evidence that cognitive processes are important to the success of UT. Four cognitive elements were found to be “strongly associated with successful inspections.” These four cognitive elements include: 1) to develop and test explicit hypotheses as a part of the inspection; 2) avoid forming conclusions early in the inspection, before all information has been considered; 3) use if-then logic; and 4) avoid arbitrarily eliminating information while forming conclusions.

Stephens Jr. (2000) presents human factors in the context of an NDE reliability model with three elements referred to as intrinsic capability (IC), application parameters (AP), and human factors (HF). IC

is considered to represent the upper bound of an NDE system performance, which is then degraded by AP and HF. AP include those non-human factors related parameters that can influence inspection performance such as material conditions, and equipment. Wall et al. (2010) provided an overview of human factors analysis in NDE and noted that former NDE trials have shown that human factors lower nondestructive testing (NDT) reliability. One study showed a drop in reliability from 80% for an automated system to 50% for a manual procedure for weld inspection. They also noted that human factors can be accounted for using general models, such as that proposed by Stephens Jr. (2000), and also through multi-parameter POD analysis as discussed in Section 2.2. Wall et al. (2010) suggest that POD can be adjusted for human factors effects through multiplication by a constant multiplier less than 1. Mueller et al. (2012) note that the model elements introduced by Stephens Jr. (2000) are influenced by a higher level organizational context.

Bertovic et al. (2010) present a reliability model built according to principles in the Socio-Technical Systems Approach, which assumes that a system is composed of technical as well as social or human elements. This model interprets organizational context as having overarching influence on all other essential parameters. This paper also presents a study of the effects of time pressure (stress resistance) on the performance of ten manual ultrasonic operators. The operators were evaluated under three conditions: 1) no time pressure, 2) middle time pressure, and 3) highest time pressure. The testing showed that time pressure (and in particular a perceived time pressure) decreased the quality of the performance as measured by more scatter in the data. Higher demands of the task created mental workload, also decreasing the performance, but experienced operators better compensated for this load. Careful planning, good preparation with understandable instructions, all part of organizational working conditions, showed a positive influence on the performance of the operators.

Bertovic et al. (2012) presents a review of earlier human factors studies. The European Programme for the Inspection of Steel Components (PISC III) looked at variability between inspectors but was criticized for its poor design and statistical background. The Finnish Centre for Radiation and Nuclear Safety in the late 1990s showed no single human or organizational factor was responsible for the NDT performance variations and that attitudes of the inspectors significantly influenced their inspections. The Swedish Nuclear Power Inspectorate studies showed that trust in their own performance, motivation, optimal work conditions, and feedback influence performance. Finally, the United Kingdom Health and Safety Executive under the Programme for the Assessment of NDT in Industry in 2008 related ability and personality with performance in a manual ultrasonic evaluation. They also showed improvements in performance with debriefing, understanding of the geometry, systematic application of a procedure, and good preparation. In addition, in Bertovic et al. (2012), issues related to human redundancy were investigated. It is reported that social loafing occurs when an individual exerts less effort when working on a task, knowing their task will be duplicated or reviewed by another operator as opposed to working alone and being solely responsible for the outcome. Conversely, social compensation occurs when an individual works harder to compensate for less skilled oversight or lack of redundancy. The study found that participants made more sizing errors when working with redundancy in the system. Automation bias was also discussed in this paper.

Finally, in Harding et al. (2009), human effects are incorporated into a POD model by representing the threshold for observation of a defect signal as a distributed random variable. It's noted that this applies to procedures in which the influence of human factors is primarily on the interpretation of acquired data.

5.0 Discussion and Analysis

This section provides discussion and analysis of the literature review presented in the previous sections, in an attempt to identify themes and to identify possible paths forward. To facilitate, the section is organized such that discussion and analysis of MAPOD approaches is provided in Subsection 5.1, discussion and analysis of human factors and their integration with MAPOD is provided in Subsection 5.2.

5.1 MAPOD Discussion and Analysis

A review of the literature related to MAPOD to date identified three basic approaches for potentially estimating the field POD for nuclear power plant components:

- (1) An XFM approach that leverages empirically derived data from round-robin studies or performance demonstrations as a “baseline” curve, which would be augmented to obtain an improved estimation of field POD. In this case, the “baseline” curve could potentially be augmented with a) a small amount of field inspection data if it is available or b) information obtained empirically through careful laboratory experiments regarding the influence of relevant factors (e.g., human factors).
- (2) A Bayesian approach that leverages empirically derived data from round-robin studies or performance demonstrations as the “prior” condition. The field POD estimation represents the “posterior” condition. In this case, the “prior” condition is updated to the “posterior” condition using likelihood information that may consist of a) a small amount of field inspection data or b) information obtained empirically through careful laboratory experiments regarding the influence of relevant factors (e.g., human factors).
- (3) Build \hat{a} vs. a models for field inspection implementing the FMA approach. In this case, computer simulation could be leveraged to determine the effects of those factors whose influences on POD are physics-based and careful laboratory experimentation would be used to estimate the influence of factors that must be determined empirically (e.g., human factors).

The XFM approach would potentially allow empirical data collected from round-robin studies and/or from performance demonstrations to be leveraged toward determination of field POD. However, the demonstrations of the XFM approach are based on \hat{a} vs. a data. In the transfer function approach, POD model parameters are augmented based on physical interpretation of these parameters (i.e., slope and intercept of signal response). A demonstration of the transfer function approach applied to binary NDE response data is provided in Bode et al. (2012).

A Bayesian approach that relies on sparse field inspection data to update some prior information is similar to the approach demonstrated in Leemans and Forsyth (2004). The availability of any field data could pose a challenge to this approach. In addition, as demonstrated by Leemans and Forsyth (2004), the utility of the Bayesian approach depends significantly on the prior information. If the confidence bounds on the prior information are large, then the estimated posterior distribution will predominantly be influenced by the field data. Alternatively, if the confidence is high, then the prior information will dominate the posterior estimation and the additional small sample of field data may not have significant influence.

Alternatively, a FMA approach could be implemented by creating \hat{a} vs. a models. Computer simulation could be used to estimate a “baseline” POD curve based on those factors whose influence is physics-based. The influence of other factors (such as human factors) could be determined empirically and used to augment the “baseline” curve. This represents the “start-from-scratch” scenario as it does not directly leverage data collected through round-robin studies or performance demonstrations. However, this data could be used to benchmark the FMA approach providing some means of validation.

An important question is how to validate estimations of field POD using MAPOD approaches. In this scenario, field inspection data for comparison to MAPOD estimations may not be available. A possible option for validation in this case may be to compare estimations generated by two different MAPOD approaches. Alternatively, an FMA approach could be applied to estimate POD for round-robin studies and/or performance demonstration-generated data. Comparison of POD estimations from the FMA approach to the empirical results obtained from round-robin studies and performance demonstrations would provide some validation of the MAPOD estimation. Validation by both these methods is incomplete and, ultimately, availability of some field inspection data for validation would be most desirable. Further, caution must be exercised when attempting to pool data for reliability analysis, as emphasized in NATO (2005). The report emphasizes that the pooling of data is valid only if sufficient information is reported in sufficient fidelity and that the datasets to be combined should be based on the same or similar inspection procedures with similar fidelity in the data sets.

5.2 Discussion and Analysis of Human Factors and Integration with MAPOD

The review of human factors literature related to NDE, performed in Section 4.0, indicates that studies tend to focus on various human factors effects that can be categorized as related to 1) individual inspector competence (e.g., motivation, preparation, experience, etc.); 2) task complexity (e.g., time stresses, environmental stresses, component accessibility, etc.); and 3) social complexity (e.g., organizational influence, social loafing, social compensation). Several studies have been able to confirm hypotheses that certain human effects can have an influence on NDE performance, such that perceived time pressure can degrade performance, or that greater preparation or experience correlates positively with performance. However, the development of mature relationships between NDE performance and specific human effects remains elusive. In studies in which multiple human effects are present, it has not been possible to identify a dominant effect (Spanner et al. 1986; Taylor et al. 1989; Bertovic et al. 2012).

From the literature, there appear to be several ways in which human effects may be incorporated into POD models for MAPOD applications: 1) represent human effects as influencing parameters in the NDE response model, for instance, using a multi-parameter POD model (Aldrin et al. 2012; Pavlović et al. 2012), and/or 2) represent human effects explicitly as a random effects term in the POD model (Li et al. 2012), and/or 3) account for human effects through the definition of the threshold condition for observation of defects. In these scenarios, the challenge is to define the relationship between NDE response and the variable representing human effects and/or to quantify the variability contributed to the response by human effects. The model introduced by Bertovic et al. (2012) illustrates the complexity of general NDE systems and the numerous channels by which human effects are input into performance. As noted in the previous paragraph, prior studies have failed to identify dominant factors that account for the majority of human effects. In order to proceed with incorporating human effects into POD models, it will be necessary to capture human effects in just one or a few terms so that the model is not too cumbersome.

To facilitate this, it is recommended to focus narrowly on a specific NDE procedure and system and to develop a POD model for that specific system, rather than attempting to apply a general model. The first steps of this effort would include identifying the specific NDE procedure and system and performing a detailed analysis to identify and organize all of the human inputs.

5.3 Other Considerations

In the course of reviewing the literature summarized here, the authors identified some shortcomings with the MAPOD demonstrations in the context of application to nuclear power plant components. The majority of existing literature assumes a single response parameter is used to determine POD, typically signal amplitude. However, many additional features are available for data interpretation for a skilled operator. With eddy current testing, this is most commonly implemented with a multiple frequency approach and the evaluation of signal changes as a function of frequency. In ultrasonic testing, there are many signal characteristics besides amplitude, or approaches, which can provide additional information (e.g., rise time, persistence, multiple attack angles, or frequencies). Thus, POD models should be developed that not only capture multiple influencing parameters on signal response, but that also treat the signal response as a multi-parameter variable.

In addition, the existing literature does not address the assessment of inspection capability when there are prior results for comparison. This can apply when there is a digital record available for direct comparison. Field experience shows that there is a clear indication of performance improvement when such data are available and many of the more difficult inspections have prior data, including baseline inspections on replaced components.

6.0 Conclusions and Looking Forward

In several applications, the MAPOD approach has shown value for both assessing the flaw POD and for extending qualifications beyond established limits (e.g., wall thickness); however, none of the prior examples are directly applicable to nuclear power plant component inspections. The robust aerospace examples cited in the literature have generally included much larger experimental data sets for POD assessment for the inspection technique. There is scant literature on quantifying the human factors aspect on POD, which is considered to be a major contributor to POD degradation in field application. Extending qualification ranges beyond established limits does hold promise to simplify additional qualification and mockup fabrication. This approach of using modeling has been identified in ASME V, Article 14, as a viable approach and is worth pursuing.

Using MAPOD approaches to provide better estimates of field POD of nuclear power plant components will either require access to field inspection results and/or will require laboratory studies to empirically determine the influence of certain parameters for which the influence cannot be represented by physical models. For field inspections, human factors have a significant impact on performance and their influence will need to be determined empirically. The human factors literature reinforces the complexity of human effects with several studies unable to isolate human effects to a few dominant factors, and general models presented to capture human inputs exhibit complexity, as well. Performing laboratory studies to account for the influence of all of the human factors represented in general models is not practical; therefore, a better approach is to narrow focus on a specific NDE procedure and system with

the expectation that specificity will lead to simplification. Performing a detailed analysis of a specific NDE procedure and system to identify all sources of human influence and attempting to consolidate them into a manageable number of parameters for a POD model would represent a logical next step for this effort.

During the course of this review, the authors identified another basis upon which POD curves generated in previous reliability studies and in performance demonstrations should be adjusted. The paper by Li et al. (2012) explains that calculating POD as the average of performance data is inappropriate for many applications, as it may be non-conservative. It is suggested that a better way for determining POD is to calculate POD based on a lower quantile of the data. The consequence of applying this to POD curves generated for NDE of nuclear power components could be analyzed by recalculating the POD based on a lower quantile for an existing database from previous reliability studies or performance demonstrations.

In addition, in the course of this review, the authors noted some possible shortcomings in the POD models that have been used in many of the published demonstrations. In the literature reviewed, the authors have observed that POD models treat the NDE response as a single parameter variable. In a field application, a skilled operator may rely on multiple features of the signal response (i.e., not just the amplitude) for interpretation. Also, for the literature reviewed, POD is estimated based on absolute signal responses. In many applications, baseline data may be available for comparison and the differential response with respect to baseline can be used to estimate performance.

7.0 References

- Aldrin JC, JS Knopp, EA Lindgren and KV Jata. 2009. "Model-Assisted Probability of Detection Evaluation for Eddy Current Inspection of Fastener Sites." In *35th Annual Review of Quantitative Nondestructive Evaluation, Volume 28*, pp. 1784-1791. July 22-25, 2008, Chicago, Illinois. DOI 10.1063/1.3114175. American Institute of Physics, Melville, New York. AIP Conf. Proc. 1096.
- Aldrin JC, JS Knopp and HA Sabbagh. 2012. "Bayesian Methods in Probability of Detection Estimation and Model-Assisted Probability of Detection Evaluation." In *Review of Progress in Quantitative Nondestructive Evaluation*, pp. 1733-1740. July 15-20, 2012, Denver, Colorado. American Institute of Physics, Melville, New York.
- Berens AP. 1989. "NDE Reliability Data Analysis." In *ASM Handbook, Volume 17: Nondestructive and Quality Control*, pp. 689-701. ASM International, Materials Park, Ohio.
- Bertovic M, B Fahlbruch, C Mueller, J Pitkanen, U Ronneteg, M Gaal, D Kanzler, U Ewert and D Schombach. 2012. "Human Factors Approach to the Acquisition and Evaluation of NDT Data." In *Proceedings 18th World Conference on Non-Destructive Testing*, pp. 2003-2012. April 16-20, 2012, Durban, South Africa. British Institute of Non-Destructive Testing.
- Bertovic M, M Gaal, C Mueller and B Fahlbruch. 2010. "Investigating Human Factors in Manual Ultrasonic Testing: Testing the Human Factor Model." In *4th European-American Workshop on Reliability of NDE*. June 2009, Berlin, Germany. NDT.net, Rheintalstr, Germany.
<http://www.ndt.net/article/reliability2009/Inhalt/th4a3.pdf>

Bode MD, J Newcomer and S Fitchett. 2012. "Transfer Function Model-Assisted Probability of Detection for Lap Joint Multi Site Damage Detection." In *Proceedings of 38th Annual Review of Progress in Quantitative Nondestructive Evaluation*, pp. 1749-1756. July 17-22, 2011, Burlington, Vermont. DOI 10.1063/1.4716423. American Institute of Physics, Melville, New York. AIP Vol. 1430.

Carboni M and S Cantini. 2012. "A 'Model Assisted Probability of Detection' Approach for Ultrasonic Inspection of Railway Axles." In *Proceedings 18th World Conference on Non-Destructive Testing*, pp. 2457-2466. April 16-20, 2012, Durban, South Africa. British Institute of Non-Destructive Testing.

Dann MR and MA Maes. 2011. "Spatial Hierarchical POD Model for In-line Inspection Data." In *Applications of Statistics and Probability in Civil Engineering*, pp. 2274-2282, eds: M Faber, J Koehler and K Nishijima. Taylor & Francis Group, London.

Demeyer S, F Jenson, N Dominguez and E Iakovleva. 2012. "Transfer Function Approach Based on Simulation Results for the Determination of POD Curves." In *Proceedings of 38th Annual Review of Progress in Quantitative Nondestructive Evaluation*, pp. 1757-1764. July 17-22, 2011, Burlington, Vermont. DOI 10.1063/1.4716424. American Institute of Physics, Melville, New York. AIP Vol. 1430.

DoD. 2009. *Department of Defense Handbook - Nondestructive Evaluation System Reliability Assessment*. MIL-HDBK-1823A, U.S. Department of Defense (DoD), Washington, D.C.

Dominguez N, F Jenson, P Dubois and F Foucher. 2012. "NDT Performance Demonstration Using Simulation." http://www.extende.com/wp-content/uploads/publications/presentations/2012_ASNT_Dominguez_CivaForNDTPerformanceDemonstration_EN.pdf.

ENIQ. 2010. *Probability of Detection Curves: Statistical Best-Practices*. ENIQ Report No. 41, EUR 24429 EN, European Network for Inspection and Qualification (ENIQ), Luxembourg, Germany.

Forsyth DS and A Fahr. 1998. "An Evaluation of Probability of Detection Statistics." In *Research and Technology Organization Applied Vehicle Technology Workshop on Airframe Inspection Reliability under Field/Depot Conditions (RTO MP-10)*, pp. 10-1 to 10-5. May 13-14, 1998, Brussels, Belgium. RTO/NATO, Cedex, France.

Guo Y, K Yang and A Alaeddini. 2011. "A Truncated Logistic Regression Model in Probability of Detection Evaluation." *Quality Engineering* 23(4):365-377.

Harding CA, GR Hugo and SJ Bowles. 2009. "Application of Model-Assisted POD Using a Transfer Function Approach." In *35th Annual Review of Quantitative Nondestructive Evaluation, Volume 28*, pp. 1792-1799. July 22-25, 2008, Chicago, Illinois. American Institute of Physics, Melville, New York. AIP Conf. Proc. 1096.

Harris D and B McCloskey. 1990. *Cognitive Correlates of Ultrasonic Inspection Performance*. EPRI NP-6675, Electric Power Research Institute, Palo Alto, California.

Heasler PG and SR Doctor. 1996. *Piping Inspection Round Robin*. NUREG/CR-5068, PNL-10475, U.S. Nuclear Regulatory Commission, Washington, D.C.

Hovey PW. 2009. "Using the Information in Field Service Inspections to Assess the Damage of the Aircraft and the Detection Capability of the Inspection System." In *35th Annual Review of Quantitative*

Nondestructive Evaluation, Volume 28, pp. 1855-1861. July 22-25, 2008, Chicago, Illinois. American Institute of Physics, Melville, New York. AIP Conf. Proc. 1096.

Jenson F, N Dominguez, P Willaume and T Yalamas. 2013. "A Bayesian Approach for the Determination of POD Curves from Empirical Data Merged with Simulation Results." In *Review of Progress in Quantitative Nondestructive Evaluation, Vol. 32*, pp. 1741-1748. July 15-20, 2012, Denver, Colorado. DOI 10.1063/1.4789229. American Institute of Physics, Mellville, New York.

Kanzler D, C Muller, M Rosenthal, U Ewert and J Pitkanen. 2013. "An Approach to the POD Based on Real Defects Using Destructive Testing and Bayesian Statistics." In *9th International Conference on NDE in Relation to Structural Integrity for Nuclear and Pressurized Components*, pp. 191-199. May 22-24, 2012, Seattle, Washington. NDT.net, Rheintalstr, Germany. <http://www.ndt.net/article/jrc-nde2012/papers/32.pdf>.

Khan M and M Liao. 2013. "Estimating Probability of Detection from In-service Inspection Data." *Canadian Aeronautics and Space Journal* 59(02):59-69.

Leemans DV and D Forsyth. 2004. "Bayesian Approaches to Using Field Test Data in Determining the Probability of Detection." *Materials Evaluation* 62:855-859.

Li M, FW Spencer and WQ Meeker. 2012. "Distinguishing Between Uncertainty and Variability in Nondestructive Evaluation." In *Proceedings of 38th Annual Review of Progress in Quantitative Nondestructive Evaluation*, pp. 1725-1732. July 17-22, 2011, Burlington, Vermont. DOI 10.1063/1.4716420. American Institute of Physics, Melville, New York. AIP Vol. 1430.

Meeker WO and RB Thompson. 2007. "Sensitivity Analysis to Assess the Effects of Misses in the Estimation of POD from Field Inspection Data." In *Review of Progress in Quantitative Nondestructive Evaluation*, pp. 1799-1806. July 30-August 4, 2006, Portland, Oregon. American Institute of Physics, Melville, New York.

Mueller C, M Bertovic, M Pavlovic, D Kanzler, U Ewert, J Pitkanen and U Ronneteg. 2012. "Holistically Evaluating the Reliability of NDE Systems – Paradigm Shift." In *Proceedings 18th World Conference on Non-Destructive Testing*, pp. 653-666. April 16-20, 2012, Durban, South Africa. British Institute of Non-Destructive Testing.

NATO. 2005. *The Use of In-Service Inspection Data in the Performance Measurement of Non-Destructive Inspections*. RTO-TR-AVT-051, North Atlantic Treaty Organisation (NATO), Cedex, France.

Pavlović M, K Takahashi and C Müller. 2012. "Probability of Detection as a Function of Multiple Influencing Parameters." *Insight* 54(11):606-611.

Pond D, D Donohoo and RV Harris Jr. 1998. *An Evaluation of Human Factors Research for Ultrasonic Inservice Inspection*. NUREG/CR-6605, PNNL-11797, U.S. Nuclear Regulatory Commission, Washington, D.C.

Rosell A and G Persson. 2013. "Model Based Capability Assessment of an Automated Eddy Current Inspection Procedure on Flat Surfaces." *Research in Nondestructive Evaluation* 24(3):154-176.

Schneider CRA and JR Rudlin. 2004. "Review of Statistical Methods Used in Quantifying NDT Reliability." *Insight - Non-Destructive Testing and Condition Monitoring* 46(2):77-79(3).

Smith K, B Thompson, B Meeker, T Gray and L Brasche. 2007. "Model-Assisted Probability of Detection Validation for Immersion Ultrasonic Application." In *Review of Progress in Quantitative Nondestructive Evaluation, Vol. 26A/26B*, pp. 1816-1822. July 30-August 4, 2006, Portland, Oregon. DOI 10.1063/1.2718184. American Institute of Physics, Melville, New York.

Spanner JC, RW Badalamente, WL Rankin and TJ Triggs. 1986. *Human Reliability Impact on Inservice Inspection, Phase 1 Summary Report*. NUREG/CR-4436, PNL-5641, BHARC-400/85/016, Vol. 1, U.S. Nuclear Regulatory Commission, Washington, D.C.

Spencer FW. 2001. "Estimating Probability of Detection Curves from Regression Data." *Materials Evaluation* 59(7):866-870.

Stephens Jr. HM. 2000. "NDE Reliability - Human Factors - Basic Considerations." In *15th World Conference on Nondestructive Testing*. October 15-21, 2000, Rome, Italy. The Italian Society for Nondestructive Testing and Monitoring Diagnostics.
<http://www.ndt.net/article/wcndt00/papers/idn736/idn736.htm>.

Taylor TT, JC Spanner, PG Heasler, SR Doctor and JD Deffenbaugh. 1989. "An Evaluation of Human Reliability in Ultrasonic Inservice Inspection for Intergranular Stress Corrosion Cracking Through Round Robin Testing." *Materials Evaluation* 47(3):338-344.

Thompson RB. 2008. "A Unified Approach to the Model-Assisted Determination of Probability of Detection." In *34th Annual Review of Progress in Quantitative Nondestructive Evaluation*, pp. 1685-1692. July 22-27, 2007, Golden, Colorado. American Institute of Physics, Melville, New York.

Thompson RB, LJ Brasche, D Forsyth, EA Lindgren, P Swindell and W Winfree. 2009. "Recent Advances in Model-Assisted Probability of Detection." Presented at *4th European-American Workshop of Reliability of NDE, June 2009*.

Wall M, S Burch and J Lilley. 2010. "Human Factors in POD Modelling and Use of Trial Data." In *4th European-American Workshop on Reliability of NDE*. June 2009, Berlin, Germany. NDT.net, Rheintalstr, Germany. <http://www.ndt.net/article/reliability2009/Inhalt/th5a4.pdf>.



Pacific Northwest
NATIONAL LABORATORY

*Proudly Operated by **Battelle** Since 1965*

902 Battelle Boulevard
P.O. Box 999
Richland, WA 99352
1-888-375-PNNL (7665)

www.pnnl.gov



Prepared for the U.S. Nuclear Regulatory Commission
under a Related Services Agreement with the U.S. Department of Energy
CONTRACT DE-AC05-76RL01830

U.S. DEPARTMENT OF
ENERGY