New methods for statistical modeling and analysis of nondestructive evaluation data

Ming Li
Iowa State University

Follow this and additional works at: https://lib.dr.iastate.edu/etd

Part of the Statistics and Probability Commons

Recommended Citation
Li, Ming, "New methods for statistical modeling and analysis of nondestructive evaluation data" (2010). Graduate Theses and Dissertations. 11513.
https://lib.dr.iastate.edu/etd/11513

This Dissertation is brought to you for free and open access by the Iowa State University Capstones, Theses and Dissertations at Iowa State University Digital Repository. It has been accepted for inclusion in Graduate Theses and Dissertations by an authorized administrator of Iowa State University Digital Repository. For more information, please contact digirep@iastate.edu.
New methods for statistical modeling and analysis of nondestructive evaluation data

by

Ming Li

A dissertation submitted to the graduate faculty
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Major: Statistics

Program of Study Committee:
William Q. Meeker, Major Professor
Stephen D. Holland
Ranjan Maitra
Daniel John Nordman
Huaiqing Wu

Iowa State University
Ames, Iowa
2010

Copyright © Ming Li, 2010. All rights reserved.
To Ruixue and my parents
# TABLE OF CONTENTS

ACKNOWLEDGEMENTS vii  
ABSTRACT viii  

CHAPTER 1. GENERAL INTRODUCTION 1  
1.1 Background 1  
1.2 Motivation 2  
1.2.1 Noise Interference Model 2  
1.2.2 Noise Interference Model for Vibrothermography Data 2  
1.2.3 Matched Filter for Vibrothermography Data 3  
1.2.4 Physical Model Assisted NDE Analysis 3  
1.2.5 Bivariate Normal Joint Estimation Method 4  
1.3 Dissertation Organization 4  

CHAPTER 2. A NOISE INTERFERENCE MODEL FOR ESTIMATING PROBABILITY OF DETECTION FOR NONDESTRUCTIVE EVALUATIONS 6  
Abstract 6  
2.1 Introduction 7  
2.1.1 Background 7  
2.1.2 Motivation and Overview 7  
2.1.3 Related Literature 8  
2.1.4 Eddy Current Experiment Description 8  
2.2 Standard POD Methods 9  
2.2.1 Hit/Miss POD 9  
2.2.2 Traditional α-Versus-a POD 10  
2.2.3 The Floor Threshold POD 13  
2.3 The Noise Interference Model 13  
2.3.1 The Complementary Risk Model 13  
2.3.2 Model Setup 14  
2.3.3 Likelihood Functions 15  
2.3.4 POD and Confidence Lower Bound 16  
2.4 Eddy Current Example 17  
2.5 Conclusion 18  
Acknowledgements 19  
References 19  

CHAPTER 3. QUANTITATIVE MULTI-INSPECTION-SITE COMPARISON OF PROBABILITY OF DETECTION FOR VIBROThERMography NONDESTRUCTIVE EVALUATION DATA 21  
Abstract 21  
3.1 Introduction 22  
3.2 Experimental Setup 24
CHAPTER 4. STATISTICAL METHODS FOR AUTOMATIC CRACK DETECTION BASED ON VIBROTHERMOGRAPHY SEQUENCE-OF-IMAGES DATA

Abstract

4.1 Introduction
  4.1.1 Background
  4.1.2 Related Literature
  4.1.3 Overview

4.2 Standard Statistical Methods in NDE
  4.2.1 Signal Response
  4.2.2 Detection Threshold
  4.2.3 POD

4.3 Vibrothermography Inspection System

4.4 Concept of Matched Filter
  4.4.1 A One Dimensional Matched Filter
  4.4.2 A Three Dimensional Matched Filter

4.5 Matched Filter for Vibrothermographic Crack Detection

4.6 Dimension Reduction
  4.6.1 Regions of Signal and Noise
  4.6.2 Feature Extraction

4.7 Detection Criteria
  4.7.1 Temperature Increase
  4.7.2 A Detection Criterion Based on Signal and Noise
    4.7.2.1 The matched filter signal response
    4.7.2.2 The noise response
    4.7.2.3 The detection criterion

4.8 POD Comparison

4.9 Summary and Conclusion

Acknowledgements
References

CHAPTER 5. PHYSICAL MODEL ASSISTED PROBABILITY OF DETECTION IN NONDESTRUCTIVE EVALUATION FOR DETECTING OF FLAWS IN TITANNIUM FORGINGS

Abstract
ACKNOWLEDGEMENTS

Eight years have been passed since my first day at this lovely mid-west small town, Ames. Shortly after accepted a TA offer from Department of Statistics, I got email from Dr. Meeker asking whether I was interested in his NDE projects as RA. After several long talks with Dr. Meeker, I decided to pursue a Statistics Ph.D. focusing on engineering applications. I came to realize the great opportunity of interdisciplinary research fields of Statistics, Physical Sciences and Engineering during these in-depth discussions. Step by step, Dr. Meeker lead me into the fantastic world of Statistics. Little by little, I learned how to apply my statistics and physics knowledge to real world problems. Dr. Meeker gave me enormous help and valuable advice not only in academic research but also in personal life. I have learned a lot from his deep understanding and sharp vision in statistics research fields. Here I would like to sincerely thank my advisor Dr. Meeker.

Also during my Statistics Ph.D. study, I learnt a lot from my study committee: Dr. Holland explained the details of the NDE experimental setup and helped me to develop statistical models; Dr. Maitra suggested numerical and computational strategies for my model simulation; Dr. Nordman pointed potential applications of my research results; and Dr. Wu discussed the future development of statistics applications in engineering, economy and finance fields. I would like to thank my study committee for their insightful suggestion, numerous help and continuous encouragement. I would also like to thank Dr. Bruce Thompson and Dr. Peter Hovey for valuable discussions and help at the MAPOD project and BVN project. I would also like to thank all my friends during my eight years stay in Ames. Last I thank my wife and parents who always encourage and support my study.
ABSTRACT

Statistical methods have a long history of applications in physical sciences and engineering for design of experiments and data analyses. In nondestructive evaluation (NDE) studies, standard statistical methods are described in Military Handbook 1823A as guidelines to analyze the experimental NDE data both in carefully controlled laboratory setup and field studies. However complicated data structures often demand non-traditional statistical approaches. In this dissertation, with the inspiration and needs from actual NDE data applications, we introduced several statistical methods for better description of the problem and more appropriate modeling of the data. We also discussed the potential applications of those statistical methods to other research areas.

The dissertation is organized as following. First a brief background introduction and overview are presented at Chapter 1. Then the complementary risk noise-interference model is discussed in Chapter 2 to better describe the noise and signal relation. In Chapter 3, a direct application of the noise interference model to vibrothermography NDE experiment scalar data is presented. In Chapter 4, the matched filter technique is used to increase signal-to-noise ratio for sequence of image analysis. In Chapter 5, the physical model assisted probability of detection analyses are introduced where the underlying physical mechanism plays an important role in the data interpretation. In Chapter 6, a bivariate normal Bayesian approach is studied to efficiently handle missing information. Finally we summarize these recent NDE developments at Chapter 7.
CHAPTER 1. GENERAL INTRODUCTION

1.1 Background

Statistical methods have a long history of applications in physical sciences and engineering for design of experiments and data analyses. A set of standard statistical methods have been developed by statisticians, scientists and engineers, such as regression, analysis of variance, hypothesis testing, reliability assessment and robust analysis. Least squares and likelihood based estimation techniques are widely used in these applications. With the fast development of computer hardware and algorithm, Bayesian methods through Markov Chain Monte Carlo simulation are becoming more and more popular.

Nondestructive evaluation (NDE) is a research area in many industries, such as aerospace applications, to detect defects or cracks enclosed in structures by non-intrusive physical measurements. There exists random measurement noise for most NDE applications and statistical methods are needed for NDE data analysis. Standard statistical methods such as regression can be applied to well formed NDE data sets. In actual field studies, however, there exist complicated data structures such as missing information, longitudinal data, repeated measure, random effects, and high-dimensional data. Advanced statistical methods are needed for such problems. Sometimes a physics principle based statistical analysis is needed for better interpretation of the NDE data set. In this dissertation, we discuss several newly developed statistical methods for NDE data analysis. These recently developed methods provide better description of the data and the results help to improve the product reliability.
1.2 Motivation

1.2.1 Noise Interference Model

A traditional way to estimate probability of detection (POD) from quantitative inspection data involves estimating the relationship between signal response and flaw size using a linear regression model of the (possibly) transformed data. Noise response data, when available, are used to estimate the detection threshold to control the probability of a false alarm. One of the direct results from the traditional POD method is that the POD will be close to zero where there is no flaw. However, the POD for the limiting case of a very small flaw (or no flaw) is actually the probability of a false alarm. To better describe the whole data set, we introduced the complementary risk noise interference model to use both the noise data and the measured signal data. The resulting POD estimate from the noise interference model provides a correct estimate of the positive probability of detection for small and zero flaw sizes.

1.2.2 Noise Interference Model for Vibrothermography Data

Vibrothermography is a technique widely used in nondestructive evaluation for finding cracks through frictional heat generated from crack surface vibrations under external excitations. The vibrothermography inspection method provides a sequence of infrared images as output, and a scalar reduction in units of heat increase from the sequence-of-images data is obtained through a physical model of heat dissipation. The scalar reduction of heat increase for specimens with small cracks is close to the experimental noise level. Thus the noise interference model is needed to describe the heat increase data set. Our result shows
that despite the substantially different vibrothermography configurations and experimental measurement responses, the estimated PODs as function of crack length and dynamic stress were similar for all different configurations, which makes quantitative POD comparisons possible across different vibrothermography setups.

1.2.3 Matched Filter for Vibrothermography Data

Beside the scalar reduction of the sequence-of-images data obtained through the vibrothermography inspection, a direct analysis the sequence of images is possible with the help of a matched filter technique. A matched filter is the optimal linear filter in terms of signal-to-noise ratio (SNR) under a stationary white noise process. We have developed a 3D matched filter to greatly increase the SNR of the vibrothermography sequence-of-images data. With the increased SNR, a noise threshold detection criterion using the largest contrast frame of image is used to define a detection criterion. An automatic detection algorithm can be developed based on the SNR detection criterion to increase the detection sensitivity.

1.2.4 Physical Model Assisted NDE Analysis

For some NDE applications, a simple empirical approach such as linear regression is inadequate to describe the data. An important alternative approach is to use knowledge of the physics of the inspection process to provide information about the underlying relationship between the response and explanatory variables. Use of such knowledge can greatly increase the power and accuracy of the statistical analysis and enable, when needed, proper extrapolation outside the range of the observed explanatory variables. A set of physical model-assisted statistical analyses are developed to study the capability of two different
ultrasonic testing inspection methods to detect synthetic hard alpha inclusion defects in titanium forging disks.

1.2.5 Bivariate Normal Joint Estimation Method

Life prediction and inspection interval decisions in aerospace applications require knowledge of the size distribution of unknown existing cracks and the probability of detecting a crack, as a function of crack characteristics (e.g., crack size). The POD for a particular inspection method is usually estimated on the basis of laboratory experiments on a given specimen set. These experiments, however, cannot duplicate the conditions of in-service inspections. Quantifying the size distribution of unknown existing cracks is more difficult. If NDE signal strength is recorded at all inspections and if crack size information is obtained after a “crack find” inspection, it is possible to estimate the joint distribution of crack size and signal response. This joint distribution can then be used to estimate both the in-service POD and the crack size distribution at a given period of service time.

1.3 Dissertation Organization

This dissertation consists of 5 main chapters, preceded by the general introduction (Chapter 1) and followed by a general conclusion (Chapter 7). Each chapter corresponds to a submitted or to-be-submitted paper. Chapter 2 describes the concept of the complementary risk noise interference model and Chapter 3 applies the noise interference model to a vibrothermography NDE scalar reduction data. Chapter 4 describes the matched filter technique to increase the signal-to-noise ratio for sequence-of-images inspection data. Chapter 5 presents the physical model assisted probability of detection analyses where the
underlying physical mechanism plays an important role in the data interpretation. Chapter 6 presents a bivariate normal Bayesian approach to efficiently handle missing information in NDE applications. The Appendix A and B include WinBUGs codes for Chapter 5 and 6.
CHAPTER 2. A NOISE INTERFERECE MODEL FOR
ESTIMATING PROBABILITY OF DETECTION FOR
NONDESTRUCTIVE EVALUATIONS


Ming Li and William Q. Meeker
Center for Nondestructive Evaluation and Department of Statistics
Iowa State University, Ames, IA 50011

Abstract

A traditional way to estimate probability of detection (POD) from quantitative inspection data involves estimating the relationship between signal response and flaw size using a linear regression model of the (possibly) transformed data. Noise response data, when available, are used to estimate the detection threshold to control the probability of a false alarm. One of the direct results from the traditional POD method is that the POD will be close to zero where there is no flaw. However, the POD for the limiting case of a very small flaw (or no flaw) is actually the probability of a false alarm. In this paper, we will use both the noise data and the measured signal data to estimate the parameters of a combined model for signal and noise. The resulting POD estimate from the noise interference model provides a correct estimate of the positive probability of detection for small and zero flaw sizes.
2.1 Introduction

2.1.1 Background

Nondestructive evaluation (NDE) methods are widely used in many industries to detect flaws such as cracks or inclusions or otherwise to determine the integrity of materials, components or structures by measuring some physical responses (such as impedance, heat or mechanical vibration). Often such data are obtained from carefully designed experiments. The most significant feature of NDE methods is that the measurement should not cause permanent physical damage to the specimen which makes repeating measurement of the same specimen through different experimental approaches possible.

2.1.2 Motivation and Overview

In most applications when an NDE measurement is taken in a place where we know there are no target flaws, the reading can still be of some value to quantify measurement and background noise. Such noise data can be used to estimate the probability of a false alarm (PFA). When there are very small flaws, small signals close to the noise level will be obtained from the measurements. Based only on the measurements, we cannot be sure that such measurements were from the flaw or some artifact of the specimen or the test setup that would cause noise.

However, the traditional â-versus-a analysis usually returns very small (around zero) POD when the flaw size approaches zero. In this paper, we extend the â-versus-a POD analysis to our noise interference model (NIM) which returns a more accurate POD estimate.
for small flaw sizes. The NIM has a better statistical interpretation provided by the simultaneous analysis of noise and signal data.

### 2.1.3 Related Literature

A good review article about the development of NDE especially for statistical inference of NDE data can be found at Olin and Meeker (1996). The Department of Defense Handbook (MIL-HDBK-1823A, 2009) provides some guidelines for NDE industry standard along with a basic introduction of NDE techniques and statistical methods for NDE data analysis. This paper follows the MIL-HDBK-1823A (2009) nomenclature and notation.

### 2.1.4 Eddy Current Experiment Description

We will apply the NIM to an experimental eddy current dataset which was obtained at Center for Nondestructive Evaluation at Iowa State University. The experimental units are several alloy panels that were fabricated at Sandia National Laboratory to be used in POD studies. Each panel has approximately 20 rivet holes. Some of the rivet holes have cracks and others do not. The rivet holes are divided into two groups based on their position: at a tear strap and not at a tear strap. These two groups show different measurement characters. The eddy current method was applied to the rivet holes providing an impedance trace as a response. Characteristics of the impedance plot are used to make decisions of crack existence.

We also have access to the actual crack size information. The rivet holes without cracks were also measured to provide information about the noise distribution. In this paper, only the rivet holes at the tear straps are studied.
2.2 Standard POD Methods

2.2.1 Hit/Miss POD

The name “hit/miss POD” comes from some NDE procedures that return binary decision results of flaw detection or not. The inspection operator uses knowledge and experience to make a decision of flaw existence or not for specimens with different flaw properties such as size and geometry. The decision that there is a flaw is called a “hit” represented by numerical value 1; otherwise there is a “miss” represented by numerical value zero.

With the information of flaw size, the hit/miss POD is calculated by binary regression using the hit/miss data, modeling the probability of a hit as a function of flaw size. It is possible to fit such a binary regression as a special case of the generalized linear model (GLM) regression procedure. A typical binary regression POD with logit link function can be written as: $\text{POD}(x) = \frac{\exp(\beta_0 + \beta_1 x)}{1 + \exp(\beta_0 + \beta_1 x)}$ where POD(x) is the POD of the flaw size or some transformation of flaw size (x), $\beta_0$ and $\beta_1$ are the binary-model regression coefficients. For the eddy current data example, the typical “S” shape binary regression hit/miss POD with logit link is plotted at Figure 2-1 (left). The POD estimate is a smooth increasing function with flaw size as explanatory variable. The hit (open circles) and miss (solid circles) binary decisions used in the binary regression are also plotted at Figure 2-1 (left).

In general the binary response from the operator is subjective and may contain more human-factors variability and as such may lack repeatability. Thus, such data may be less informative than continuous response data. That is, reducing good quantitative information to
hit/miss data will sacrifice information. On the other hand, hit/miss reflects the operator’s comprehensive understanding of the whole inspection system, and may allow use of visual information that is difficult to quantify. Two examples where hit/miss data are commonly used include fluorescent penetrant inspection and inspections resulting in complicated images.

![Binary regression model hit/miss POD (Logit Link)](image1)

![Histogram of response for all specimens without flaw](image2)

**Figure 2-1.** Binary regression hit/miss POD, and the noise distribution with decision threshold

### 2.2.2 Traditional â-Versus-a POD

The traditional â-versus-a approach is usually used instead of hit/miss POD when continuous response (“â”) is measured in the experiment for each specimen where “a” is the flaw size or certain transformations of flaw size and â is the response or some transformation of the response (e.g., log transformations are often used on a and â). There may also be measurement responses for inspection targets without any flaws (i.e., providing noise data). Figure 2-1 (right) is the noise distribution of the measurements from the eddy current example rivet holes without flaws. When flaws are present in the specimen, the
measurement response will have some probability to indicate the existence of flaws. The expected measurement response is usually a simple increasing function of flaw size or certain transformations of flaw size. Often the relationship between the expected log signal versus log flaw size is assumed to be linear. Figure 2-2 (left) shows the increasing measurement response (open circles) as function of flaw size for the eddy current example. The solid line in Figure 2-2 (left) is the maximum likelihood (ML) estimate of the mode 
\[
\hat{a} = \beta_0 + \beta_1 a + \epsilon, \quad \epsilon \sim N\left(0, \sigma^2\right).
\]
That is, we fit a simple linear regression with normally distributed residuals where \( \hat{a} \) is the signal response and \( a \) is the actual crack size. The dashed lines are the 95% point-wise confidence intervals for the true regression line. The noise data from the rivet holes without a crack are also plotted in Figure 2-2 (left) as cross symbols.

The decision threshold is the measurement response above which a flaw-detection (hit) decision will be made. Usually the decision threshold is chosen so that the PFA is small (typically less than 0.01) based on the noise data. We use PFA 0.1 in the eddy current example because the noise level is very high, and if PFA = 0.01 is used, most of the data points will be below the decision threshold. The decision threshold that results in a 10% PFA for the eddy current dataset is represented by the vertical dashed line in Figure 2-1 (right). The decision threshold is also plotted as a horizontal dotted line in Figure 2-2 (left), and if the noise distribution is symmetric, the intersection of the decision threshold with the ML regression line is the location of flaw size where the POD is 50%.

After the decision threshold is set and the ML estimates \((\hat{\beta}_0, \hat{\beta}_1, \hat{\sigma})\) are obtained, the POD estimate as a function of flaw size \((a)\) can be determined by
POD(a) = Pr(\(\hat{a} > a_\text{th} \mid a\)) = 1 - \Phi\left(\frac{(a_\text{th} - \hat{\beta}_0 - \hat{\beta}_1 a)}{\hat{\sigma}}\right)

where \(a_\text{th}\) is the decision threshold and \(\Phi\) is the standard normal cumulative distribution function. The 95% POD lower bound can also be found with the variance-covariance matrix of the ML estimates \(\hat{\beta}_0, \hat{\beta}_1, \hat{\sigma}\), using the delta method.

Using the 10% PFA decision threshold, the traditional POD (solid curve) and the corresponding 95% lower bound (dashed curve) for the eddy current example are plotted at Figure 2-2 (right). Lines also plotted at Figure 2-2 (right) are the 90% POD level (horizontal dashed line) and the flaw sizes where 50% POD (a50), 90% POD (a90) and 95% lower bound of the 90% POD (a90/95) are achieved (vertical dotted lines).

![Figure 2-2](image.png)

**Figure 2-2.** Maximum likelihood fitting of response vs. flaw size, and the traditional \(\hat{a}\)-versus-\(a\) POD estimate and a 95% lower bound

The traditional POD approach returns the probability of a “hit” that is almost zero when the flaw size approaches zero. But from the eddy current example data, we know that the noise level is very high and we set the decision threshold to return PFA 10%, which
means even for zero flaw size the probability of a “hit” should be around 10%, and the POD for a small non-zero flaw size should be even larger. This is an indication of lack of fit for the simple linear regression model.

2.2.3 The Floor Threshold POD

A four-parameter POD model with “floor threshold ($\alpha$)” is proposed in the appendix I.4 of MIL-HDBK-1823(A) as: 
\[
\text{POD}(a) = 1 - (1 - \alpha)\Phi\left(\frac{a_{\text{th}} - \beta_0 - \beta_1a}{\sigma}\right).
\]
In this model the POD is allowed to be larger than the asymptotic limit of traditional â-versus-a POD when flaw size approaching zero. The ML method can be used to obtain the ML estimates of those parameters including the floor threshold $\alpha$. Although the floor threshold can remedy the asymptotic feature of POD for very small flaws, this approach still uses separate analyses of the noise and signal data.

2.3 The Noise Interference Model

2.3.1 The Complementary Risk Model

Before we give details of the NIM, we first briefly introduce the idea of a complementary risk model. For $p$ given random variables $X_1, X_2, X_3, \ldots, X_p$ with corresponding cumulative distribution functions $F_1(x), F_2(x), F_3(x), \ldots, F_p(x)$, the competing risk is defined as $U = \min(X_1, X_2, \ldots, X_p)$ and the complementary risk is defined as $V = \max(X_1, X_2, \ldots, X_p)$. Basu and Ghosh (1980) used ML method to show that one could uniquely determine and estimate the marginal distributions when $p$ equals two given data on $U$ or $V$, respectively. In reliability applications, the competing risk model is used to model
the failure-time distribution of a series system and the complementary risks model is used to model a parallel system. Chan and Meeker (1999) applied competing failure model to reliability analysis with more than one failure mode.

Because in some NDE applications, such as the rivet-hole data, the response is the maximum of a signal from the crack or other flaw and the noise, we will apply the complementary risk model of noise and signal to POD analysis and will call this the noise interference model (NIM). The NIM gives us a method to untangle the mixture of noise and signal data (which is what we usually have from a POD experiment) and to obtain good estimation of parameters of both noise and signal distributions simultaneously. From the estimation of noise and signal, the POD will have a more reasonable limiting value when the flaw size is small and the estimated NIM POD will have smaller standard errors compared with the traditional model.

### 2.3.2 Model Setup

Suppose again that the signal response or some transformation of the signal response ($y_{\text{signal}}$) follows a simple relationship: $y_{\text{signal}} = \beta_0 + \beta_1 x + \varepsilon_s$, where $x$ is flaw size or a specified transformation of flaw size (often a log transformation) and $\varepsilon_s$ follows a normal distribution with mean zero and variance $\sigma_{\text{signal}}^2$ (i.e., $\varepsilon_s \sim N(0, \sigma_{\text{signal}}^2)$). Also suppose that the noise response or some transformation of the noise response ($y_{\text{noise}}$) follows a normal distribution with mean $\mu_{\text{noise}}$ and variance $\sigma_{\text{noise}}^2$ (i.e., $y_{\text{noise}} \sim N(\mu_{\text{noise}}, \sigma_{\text{noise}}^2)$). Then the raw data ($y$) from the experiment measurement can be described as $y \sim \max(y_{\text{signal}}, y_{\text{noise}})$. 
2.3.3 Likelihood Functions

The likelihood is a function of data that is proportional to the probability of the data. For the complementary risk NIM, there are three possible types of data: signal, noise, or uncertain signal or noise. For observations known to be either signal or noise data, it is easy to write the likelihood contributions as following:

\[
L_{\text{signal}} = \prod_{i \in \text{signal}} \frac{1}{\sigma_{\text{signal}}} \phi \left( \frac{y_i - \beta_0 - \beta_1 x}{\sigma_{\text{signal}}} \right)
\]

\[
L_{\text{noise}} = \prod_{j \in \text{noise}} \frac{1}{\sigma_{\text{noise}}} \phi \left( \frac{y_j - \mu_{\text{noise}}}{\sigma_{\text{noise}}} \right)
\]

where \( \phi \) is the standard normal distribution density function. For the uncertain data the model \( y \sim \max(y_{\text{signal}}, y_{\text{noise}}) \) is used and the corresponding likelihood contribution can be calculated from the following equation:

\[
L_{\text{uncertain}} = \prod_{k \in \text{uncertain}} \left( \frac{1}{\sigma_{\text{signal}}} \phi \left( \frac{y_k - \beta_0 - \beta_1 x}{\sigma_{\text{signal}}} \right) \Phi \left( \frac{y_k - \mu_{\text{noise}}}{\sigma_{\text{noise}}} \right) + \frac{1}{\sigma_{\text{noise}}} \Phi \left( \frac{y_k - \beta_0 - \beta_1 x}{\sigma_{\text{signal}}} \right) \phi \left( \frac{y_k - \mu_{\text{noise}}}{\sigma_{\text{noise}}} \right) \right)
\]

where \( \Phi \) is the standard normal cumulative distribution function. Finally the total likelihood function for the whole dataset is \( L = L_{\text{signal}} \times L_{\text{noise}} \times L_{\text{uncertain}} \).

In real applications, when the flaw size is large enough, the corresponding signal data will always be larger than the noise data, and the “actual” responses (i.e. the experiment measurements) are solely signal. We call those data points identifiable signal and the likelihood contribution \( L_{\text{signal}} \) will be used. In some POD experiments, measurements are
made on specimens without any flaws or on parts of the specimen that do not have a flaw, providing known noise data for which the likelihood contribution $L_{\text{noise}}$ will be used. For a relatively small flaw sizes, the signal and noise data are mixed together. The “actual” response will be the maximum of those two, and likelihood contribution $L_{\text{uncertain}}$ should be used. It is not always possible to identify whether an observation came from a flaw or some artifact of the specimen that would be noise. When we can not determine the data type the likelihood function $L_{\text{uncertain}}$ should be used. By maximizing the log-likelihood function, the five ML estimates $\hat{\theta} = (\hat{\mu}_{\text{noise}}, \hat{\sigma}_{\text{noise}}, \hat{\beta}_0, \hat{\beta}_1, \hat{\sigma}_{\text{signal}})$ can be obtained. Being able to identify signal and noise observations for some of the data will numerically help the ML procedure to be more efficient and accurate. The variance-covariance matrix $(\Sigma_{\hat{\theta}})$ of the ML estimates can also be estimated by evaluating the Hessian matrix of the log-likelihood function at the ML estimates, as described, for example, in Meeker and Escobar (1998).

2.3.4 POD and Confidence Lower Bound

We can find the NIM POD estimate through the following equation:

$$\text{POD}(x) = P(x \left ( \hat{\beta}_0, \hat{\beta}_1, \hat{\sigma}_{\text{signal}}, \hat{\mu}_{\text{noise}}, \hat{\sigma}_{\text{noise}} \right )$$

$$= \Pr(y > y_{\text{th}}) = \Pr(\max(y_{\text{signal}}, y_{\text{noise}}) > y_{\text{th}})$$

$$= 1 - \Pr(\max(y_{\text{signal}}, y_{\text{noise}}) < y_{\text{th}})$$

$$= 1 - \Pr(y_{\text{signal}} < y_{\text{th}}) \Pr(y_{\text{noise}} < y_{\text{th}})$$

$$= 1 - \Phi \left( \frac{y_{\text{th}} - \hat{\beta}_0 - \hat{\beta}_1 x}{\hat{\sigma}_{\text{signal}}} \right) \Phi \left( \frac{y_{\text{th}} - \hat{\mu}_{\text{noise}}}{\hat{\sigma}_{\text{noise}}} \right)$$

(2-3)

where $\Phi$ is again the standard normal cumulative distribution function. The limiting NIM POD estimate when flaw size approaches zero will now be a more reasonable positive value.
In particular, the limit of the POD for small flaws is
\[ \text{POD}(0) \approx 1 - \Phi \left( \frac{y_{\text{th}} - \mu_{\text{noise}}}{\sigma_{\text{noise}}} \right) \]
which is also the probability of a false alarm (PFA).

As the POD expression \( P_x \) is a function of the five ML estimates, the variance of the POD estimate can be found through delta method:
\[ \text{var}(P_x) = \left( \frac{\partial P_x}{\partial \theta} \right)^T \hat{\Sigma}_\theta \left( \frac{\partial P_x}{\partial \theta} \right), \]
where
\[ \hat{\Sigma}_\theta = \left[ -\frac{\partial^2 \log L}{\partial \theta^2} \right]^{-1}. \]
To ensure that the POD lower bound is always positive we compute the confidence interval in the scale of the monotone increasing logit transformation of \( P_x \):
\[ g_x = \text{logit}[P_x] = \log \left\{ \frac{P_x}{1 - P_x} \right\}, \]
based on the asymptotic normal property of \( g_x \) (see for example Meeker and Escobar 1998). The 100(1 - \( \alpha \))% POD lower confidence bound \( P_{\text{LCL}} \) can then be calculated from
\[ P_{\text{LCL}} = P_x \left( \frac{1}{P_x + (1 - P_x) \cdot w} \right) \]
where
\[ \hat{\text{se}}_{P_x} = \sqrt{\text{var}(P_x)} \text{ and } z_{1-\alpha} \text{ is the 100(1 - \( \alpha \))% quartile of standard normal distribution.} \]
In recent work, Hong, Meeker and Escobar (2008) describe the \( \hat{\xi} \) approach and compare it with other normal-approximation methods. An adaptation of this \( \hat{\xi} \) method might be a better approach to get the POD lower bound. But we have not yet attempted to adapt the \( \hat{\xi} \) approach to the NIM. Likelihood-based methods for computing confidence intervals could also be used.

### 2.4 Eddy Current Example

For this example, we have identified data from measurements taken at rivet holes that had no cracks. Thus we can now use the noise data (shown as crosses of Figure 2-3, left) and 10% PFA to get the decision threshold (horizontal dotted line). The ML estimate of the NIM (solid line) and the traditional model (dashed line) are compared in Figure 2-3 (left) along
with the estimated noise mean (horizontal solid line) and the 99\% upper bound of noise
distribution (horizontal dashed line).

![Comparison of maximum likelihood fitting](image1)

**Figure 2-3. Eddy current example results: ML regression estimates and ML POD estimates**

The corresponding POD estimates of the NIM (dark black solid curve) and the
traditional model (light gray solid curve) are plotted at Figure 2-3 (right). The NIM has a
limiting POD approaching that is close to 0.1, while the traditional model has limiting POD
approaching zero. By comparing the standard errors of the estimated POD estimates (detail
not given here), we confirm the NIM has smaller standard errors over most of the crack size
range. This is because the NIM model provides a better fit to the data.

### 2.5 Conclusion

In this paper, we have reviewed the basic ideas of nondestructive evaluation data
analysis as they relate to POD estimation. We find that the NIM can be applied to the POD
analysis to solve the small flaw size asymptotic problem presented at the traditional POD
analysis. The NIM model estimates the noise and signal distributions simultaneously using the method of maximum likelihood to provide a POD estimate and corresponding confidence intervals. The standard error of the estimated NIM POD is smaller than the traditional model, which indicates better statistical inferences from the NIM. The complementary risk NIM approach was illustrated with data from an eddy current experiment. Because many NDE experiments involve signals that are potentially mixed with noise, the complementary risk noise interference model will provide a useful option for POD assessment. Scientists and engineers in other research areas where noise is involved and the underlying physical principle of noise plus signal is to be modeled may also find this approach to be helpful.

Acknowledgements

This material is based upon work supported by the Federal Aviation Administration under Contract #DTFA03-98-D-00008, Delivery Order #0058 and performed at Iowa State University's Center for NDE as part of the Center for Aviation Systems Reliability program.

References


CHAPTER 3. QUANTITATIVE MULTI-INSPECTION-SITE COMPARISON OF PROBABILITY OF DETECTION FOR VIBROTHERMOGRAPHY NONDESTRUCTIVE EVALUATION DATA


Ming Li\textsuperscript{1}, Stephen D. Holland\textsuperscript{2}, William Q. Meeker\textsuperscript{1}

\textsuperscript{1}Center for Nondestructive Evaluation and Department of Statistics, Iowa State University, Ames, Iowa, USA

\textsuperscript{2}Center for Nondestructive Evaluation and Department of Aerospace Engineering, Iowa State University, Ames, Iowa, USA

Abstract

This paper describes the estimation of probability of detection (POD) for a vibrothermography inspection procedure. The results are based on a large scale experiment on specimens with two different kinds of metal containing fatigue cracks. The specimens were tested independently at three locations: Iowa State University (ISU), Pratt and Whitney (PW) and General Electric (GE). Despite the substantially different vibrothermography configurations and experimental measurement responses, the estimated PODs as function of
crack length and dynamic stress were similar for all three locations, which makes quantitative POD comparisons possible across different locations.

3.1 Introduction

Vibrothermographic inspection, also known as sonic infrared or thermosonics, is a nondestructive evaluation (NDE) method that can be used to detect delaminations in composite materials or cracks in metals (see for example Maldague 2001, Rantale, Wu and Busse 1996, and Favro et al. 2000). There has been, however, little quantitative research to study the transduction from vibration into heat that underlies vibrothermography and the effect that these mechanisms will have on probability of detection (POD). There is also concern about the large amount of experimental setup variability across and within vibrothermography systems (see for example Morbidini et al. 2006 and Ibarra-Castanedo et al. 2009).

To address these concerns, a large experimental study involving vibrothermography inspection was conducted on two specimen sets at three different locations. This study is described in Holland et al. (2009). A collection of 63 Titanium Ti-6Al-4V specimens and 63 Inconel-718 specimens containing fatigue cracks were independently tested at ISU, PW and GE. Each location has a different vibrothermography system. In this paper we describe the statistical models and methods used to estimate vibrothermography POD from the experimental data.

Holland et al. (2009) and Holland and Renshaw (2009) have developed an algorithm, based on a physical model, to reduce the vibrothermography sequence-of-image data in each experimental measurement into a scalar measure of temperature increase. The algorithm
performs a surface-fit of the heat from the crack to an elliptical Gaussian envelope. The heating temperature is estimated by integrating the observed heat over the peak of the Gaussian envelop and dividing by an enlarged area. The temperature calculated in this way is approximately 95% of the peak value of the surface fit.

As illustrated later in this paper, for small cracks, the amount of heat generated is close to the noise level of the inspection system and there are many such observations for both the Titanium and the Inconel specimens. In applications like this where there is a large amount of noise in the data, the traditional statistical methods for estimating POD can lead to unconservative bias in POD estimates as the noise can lift the regression line which in turn lifts POD (see for example MIL-HDBK-1823A 2009 and Li and Meeker 2009). In this paper, we apply a noise interference model (e.g. Li and Meeker 2009) to the vibrothermography data, providing POD estimates as function of crack length and dynamic stress. Our results show that estimates of POD for the different vibrothermography experimental configurations are similar. These results support the viability of using vibrothermography to detect and evaluate cracks inside metals.

This paper is organized as follows: In Section 3.2 we briefly describe the experimental configuration of the vibrothermography systems and the reduction of the complete experimental sequence-of-image data to a scalar temperature increase value. In Section 3.3 we present graphical displays of the scalar temperature increase as function of crack length and dynamic stress for the whole data set. Section 3.4 describes the noise interference model (NIM). Section 3.5 presents the detailed statistical analysis of the vibrothermography data sets for both materials. In Section 3.6, we present some conclusions.
3.2 Experimental Setup

The particular vibrothermography inspection system used in our experiments is illustrated conceptually in Figure 3-1. This system involves an excitation source to excite the sample, an infrared camera to record heating of the specimen, and a laser vibrometer to monitor vibration in the specimen. The excitation source (a piezo stack at ISU, an ultrasonic welder at PW and GE) is pneumatically pressed to the sample, and the sample itself is gripped with a rigid or compliant clamp. A coupling medium, such as paper, plastic, or cardstock is usually used to separate the tip of the vibration source from the sample. The specimen is typically excited from 1 to 2 seconds duration. The goal is to cause the crack surfaces to rub and generate heat.

![Figure 3-1. Common configuration for a vibrothermography inspection system](image)

The sample surface temperature profile is captured by a sequence of images recorded by an infrared camera and the sample surface velocity is measured by a laser vibrometer. Both the temperature profile and the surface velocity are typically recorded at short time intervals for each measurement. The vibrometer sampling rates in the experiment were 1
MHz for ISU and 500 kHz for GE and PW, and the infrared camera sampling rates were 90 Hz for ISU and 189 Hz for GE and PW.

The ultrasonic welder and piezo stack that are used as excitation sources typically generate 1 to 2 kW of vibrational power at a fixed frequency such as 20 kHz. For this study the specimens were tuned to a natural resonance near 20 kHz. During an inspection, the vibrational excitation power is coupled into the specimen near the natural resonance and frictional rubbing between crack surfaces generates heat. The known mode shape of the natural resonance allows calculation of the dynamic vibrational stress on the crack from the transverse velocities measured with the vibrometer. For each vibrothermographic inspection, the scalar heat-increase response and dynamic stress were obtained from the sequence of infrared images and the vibrometer records, as described in Holland et al. (2009).

### 3.3 Heat-increase Response Data

The scalar heat-increase response was modeled as a function of the dynamic stress and crack length. The inspection data sets for all three locations are shown in Figure 3-2 where different symbols represent various dynamic stress ranges as indicated by the legend. The inspection system noise level is around 0.03K which is indicated by the horizontal dashed lines. We also chose to use a temperature increase of 0.03K as the detection threshold to be used in POD analysis. It is clear that a large portion of the data is below the noise level and the traditional statistical method for POD analysis is no longer valid. To better retrieve the signal response from the noisy data, the noise interference model can be implemented for more efficient and reliable statistical analysis.
3.4 Noise Interference Model

The traditional statistical method for estimating POD from a NDE study with a quantitative response is the $\hat{a}$-versus-$\alpha$ method described in MIL-HDBK-1823A (2009). The traditional $\hat{a}$-versus-$\alpha$ method has, for small targets, an asymptotic limit for POD that approaches zero. This characteristic is in contradiction to the fact that for zero crack size (i.e. specimens without cracks) the POD should be approximately the probability of false alarm. When NDE measurements are taken in locations where there are no target flaws, the reading can still be of some value to quantify measurement and background noise. Such noise data are usually used to estimate the probability of false alarm. In locations where there are very small flaws, the observed response could be the result of a noise-causing artifact rather than
the small flaw. Based only on the measurements, we cannot be sure whether the measurement came from a flaw or a noise-causing artifact.

To account for possible mixture of flaw and noise responses, we extend the $a$-versus-$a$ POD analysis to our noise interference model (NIM). A detailed derivation of NIM can be found in Li and Meeker (2009). Before fitting a model, transformation of the original physical quantities are often needed. For example, one might use a logarithm or square root transformation, depending on the data itself and its variance structure. We define the observed measurement response or its transformation as $y$, the signal response or its transformation as $y_{signal}$, and the noise response or its transformation as $y_{noise}$. The NIM components are as follows:

- The signal response is modeled as
  \[ y_{signal} = f(\beta, x) + \epsilon_s \]
  where $\beta$ is a vector of regression parameters, $x$ is a vector of explanatory variables such as crack length and dynamic stress or their transformations, and $\epsilon_s$ is the residual term, assumed to be normally distributed with mean zero and variance $\sigma_s^2$, i.e.
  \[ \epsilon_s \sim \mathcal{N}(0, \sigma_s^2). \]

- The noise response $y_{noise}$ is assumed to be normally distributed with mean $\mu_n$ and variance $\sigma_n^2$.

- The observed measurement response (i.e. the experimental measurement) is the maximum of the signal and noise:
  \[ y = \max(y_{noise}, y_{signal}). \]

With the measurement data and specified $f(\beta, x)$, estimates of the parameter vector $(\hat{\beta}, \hat{\mu}_n, \hat{\sigma}_n^2, \hat{\sigma}_s^2)$ and the estimated variance covariance matrix of these estimates can be
obtained through standard maximum likelihood methods described, for example, in Pawitan (2001). The POD estimate, as a function of the explanatory variable $x$ can be calculated through

$$
POD(x) = \Pr(y > y_{th}) = 1 - \Phi \left( \frac{y_{th} - f(\hat{\beta},x)}{\hat{\sigma}_y} \right) \Phi \left( \frac{y_{th} - \hat{\mu}_n}{\hat{\sigma}_n} \right)
$$

where $y_{th}$ is the detection threshold and $\Phi(x)$ is the standard normal cumulative distribution function.

We have shown theoretically and by simulation that the NIM asymptotic POD for zero crack size is very close the probability of false alarm. The standard error of the estimated NIM POD is smaller than the traditional model, which indicates that the NIM model fits better and will provide better statistical inferences (see for example Li and Meeker 2009).

### 3.5 NIM Applied to Vibrothermography Data

First, we apply the NIM to the vibrothermography Titanium and Inconel data sets separately for each inspection locations. Then we compare the POD results across locations. Different choices of $f(\beta,x)$ are used for Titanium specimens and Inconel specimens because material differences affect the underlying generation of heating. Although for the same material the estimated parameters are different across locations, due to the significant variation of inspection system setup, the final estimated POD functions for the three locations are similar.
3.5.1 Inspections on the Titanium Specimens

As mentioned previously, the inspection system designs at the three locations are different. Each time a unit is energized, we obtain both the signal response (amount of temperature increase) and the amount of dynamic stress in the specimen. The differences among the locations are partly reflected in the differences in the distributions of the dynamic stress values at each location. Figure 3-3 shows the distribution of dynamic stress for Titanium specimens at the three locations: the ISU data have dynamic stress range from 0 to around 120MPa; the PW data have dynamic stress up to 200MPa, while the GE data has many inspections with dynamic stress much larger than 200MPa with a bi-modal distribution.

![Histogram of Dynamic Stress for Titanium](image)

**Figure 3-3. Distributions of dynamic stress for Titanium specimens at ISU, PW and GE.**

Both the physical model relating heat-increase to vibration (see for example Holland et al. 2009 and Holland and Renshaw 2009) and our statistical analysis suggest that the
logarithm of heat-increase is a linear function of logarithm of crack length for a fixed level of dynamic stress. Based on knowledge of the heat generation mechanism and comparisons among a number of different models we found that the interaction model

\[ y_{signal} = \beta_0 + \beta_1 \left( x_1 - x_1^0 \right) + \beta_2 \left( x_1 - x_1^0 \right) \times \left( x_2 - \bar{x}_2 \right) + \epsilon \]

provides a good description of the data. Here

\[ y_{signal} = \log_{10}(heat); \quad x_1 = \log_{10}(crack \ length); \quad x_1^0 = \log_{10}(25) \]

a fixed intersection position at crack length 25 mils; \( x_2 = \log_{10}(stress) \); and \( \bar{x}_2 \) is the average of all measured \( x_2 \) values.

Figure 3-4. Separate NIM analyses for the Titanium data at ISU, PW and GE.

This model implies a linear relationship between logarithm transformation of heat and crack length for each level of dynamic stress. For a given dynamic stress the slope of the regression line is \( \beta_1 + \beta_2 \left( x_2 - \bar{x}_2 \right) \). All of these lines are constrained to intersect at a common point of 25 mils. We found it necessary to use this fixed point of intersection to avoid having
the lines intersect within the range of the data. The signal response parameters $\beta_0, \beta_1, \beta_2,$ and $\sigma_z$, as well as the noise parameters $\mu_\eta$ and $\sigma_n$, can be estimated from the inspection data by using the maximum likelihood method.

Figure 3-5. Separate POD curves for Titanium specimens at three levels of dynamic stress at ISU, PW and GE.

We fit this same model to the data from each location. Figure 3-4 shows the data using different symbols to represent different dynamic stress ranges and fitted lines for dynamic stress levels 35MPa (solid lines), 60MPa (dashed lines) and 100MPa (dotted lines). At location GE, the regression line with dynamic stress 100MPa (dotted line) is much lower than the data points of dynamic stress larger than 80MPa. This is because at GE many of the inspections in this range had dynamic stress values that were much larger than 80MPa, as indicated at the dynamic stress histogram (Figure 3-3). To keep comparisons consistent, we used the same dynamic stress levels for the regression lines for each location. The parameter
estimates from statistical analysis such as regression line intercept and slopes differ across locations. These differences are a reflection of inspection system variations.

With the parameter estimates as well as their variance covariance matrix, the POD for fixed dynamic stress and its corresponding 90% lower bound can be calculated easily as described parting Section 3.4. Figure 3-5 shows the POD estimates for inspections at 35MPa, 60MPa and 100MPa for all three locations (solid lines) along with the corresponding 90% lower confidence bounds (dashed lines). The POD estimates and their lower bounds have similar patterns across all locations except for the GE POD estimate at dynamic stress 35MPa. This estimate is, however, in doubt, because there were very few data points around 40MPa for the GE inspections.

### 3.5.2 Inconel Data Sets

The Inconel specimens present different dynamic stress distributions when compared with Titanium specimens, as shown at Figure 3-6. There are some inspections with dynamic stress larger than 200MPa for PW data and many more inspections at very high dynamic stress for GE data. For Inconel specimens, the statistical model $y_{\text{signal}} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \epsilon$, provides a better description of the data. Here $y_{\text{signal}} = \log_{10}(\text{heat})$, $x_1 = \log_{10}(\text{crack length})$, and $x_2 = \log_{10}(\text{stress})$. The signal response parameters $\beta_0, \beta_1, \beta_2$, and $\sigma$, as well as the noise parameters $\mu_n$ and $\sigma_n$, can be estimated from the inspection data by using the maximum likelihood method. In this model, the regression lines for different fixed dynamic stress will be parallel as illustrated at Figure 3-7. The estimated POD curves and their
corresponding 90\% lower confidence bounds for fixed dynamic stress levels at each location are shown at Figure 3-8.

Figure 3-6. Distribution of dynamic stress levels for Inconel specimen across ISU, PW and GE.

Figure 3-7. Separate NIM analyses for the Inconel data at ISU, PW and GE.
As with the Titanium data, the heat-increases in the Inconel specimens as function of crack length, behave differently across three locations as shown at Figure 3-7. The ISU and PW data have many observations with low heat-increase (below 0.01K), while GE data have many observations with large heat-increase (above 1.0K). The inspection system variations lead to the different distributions of dynamic stress which in turn causes the discrepancy of heat-increases. Again, to keep comparisons consistent, we continue to use the same dynamic stress levels (35 MPa, 60MPa, and 100 MPa) for the regression line and POD estimates at each location. For the GE location, the estimated regression line and POD for the 35 MPa dynamic stress level are again in doubt, because there were very few data points below 40 MPa.

Figure 3-8. Separate POD curves for the Inconel specimens at three levels of dynamic stress at ISU, PW and GE.
3.5.3 POD Comparisons

After analyzing both the Titanium and the Inconel vibrothermographic experimental data for each inspection location, we now compare the POD curves for a fixed dynamic stress level of 100MPa across the three inspection locations for both materials at Figure 3-9. The ISU and PW POD results are almost the same while the GE results are close but with some degree of offset which may due to one or both of the following reasons: (1) GE inspections have fewer data points around the dynamic stress level 100MPa, (2) GE inspections have a bi-modal distribution of dynamic stress and there are many data points at very high dynamic stress range as shown in Figure 3-3 and Figure 3-6. We believe that if the GE system could eliminate the very high dynamic stress inspections (for example, use a lower power excitation source), then their POD results would likely be closer to those of ISU and PW.

Figure 3-9. POD comparisons across all three locations with fixed dynamic stress.
3.6 Conclusions

In this paper, we have applied the noise interference model to a large set of vibrothermography inspection data of metal specimens with two different materials at three different inspection locations. Despite the large difference in the experimental configurations at three locations, similar estimates of POD as a function of crack length for fixed values of dynamic stress were obtained for all locations. This is the first quantitative, multi-inspection-site demonstration of the reliability for vibrothermography method for fatigue crack detection. The estimated POD obtained at this paper only applies to these particular cracks on these 126 specimens. Further investigation of cracks and materials variability is required to extend the estimated POD to field applications.

Acknowledgements

This material is based upon work supported by the Federal Aviation Administration under Contract #DTFA03-98-D-00008, Delivery Order #0058 and performed at Iowa State University's Center for NDE as part of the Center for Aviation Systems Reliability program. We would like to acknowledge Thadd Patton (GE), Tom Bantel (formerly of GE), Zhong Ouyang (Pratt and Whitney), and Christopher Uhl (GE, formerly of ISU) for their participation in the multi-inspection-site comparative experiment.

References


CHAPTER 4. STATISTICAL METHODS FOR AUTOMATIC CRACK DETECTION BASED ON VIBROTHERMOGRAPHY SEQUENCE-OF-IMAGES DATA

A paper to be submitted to *Applied Stochastic Models in Business and Industry.*

Ming Li\(^1\), Stephen D. Holland\(^2\), William Q. Meeker\(^1\)

\(^1\)Center for Nondestructive Evaluation and Department of Statistics, Iowa State University, Ames, Iowa, USA

\(^2\)Center for Nondestructive Evaluation and Department of Aerospace Engineering, Iowa State University, Ames, Iowa, USA

Abstract

Vibrothermography is a relatively new nondestructive evaluation technique for finding cracks through frictional heat generated from crack surface vibrations under external excitations. The vibrothermography inspection method provides a sequence of infrared images as output. We use a matched filter technique to increase the signal-to-noise ratio of the sequence-of-images data. An automatic crack detection criterion based on the features extracted from the matched filter output greatly increases the sensitivity of the vibrothermography inspection method. In this paper, we develop a three dimensional matched filter for the sequence-of-images data, present the statistical analysis for the
matched filter output, and evaluate the probability of detection. Our results show the crack detection criterion based on the matched filter output provides improved detection capability.

## 4.1 Introduction

### 4.1.1 Background

Nondestructive evaluation (NDE) methods are widely used in many industries, such as aerospace applications, to detect defects or cracks enclosed in structures by non-intrusive physical measurements. There exists random measurement noise for most NDE applications and statistical methods are needed for NDE data analysis. MIL-HDBK-1823A (2009) describes the standard statistical methods used in NDE applications. Vibrothermography is an NDE inspection method based on the heat generation and temperature change around the defects or cracks under external sonic or ultrasonic wave excitations. The measurement response of a vibrothermography inspection is a sequence of images taken by an infrared camera. The sequence of images record the temporal trend and spatial pattern of temperature changes for the region inspected. Although scalar reduction of the sequence-of-images data is possible (see for example Holland et al., 2010), direct analysis of a set of features presented in the sequence-of-images data has the potential to importantly increase crack detection power.

In this paper we use data obtained by the vibrothermography method, taken on a collection of 63 titanium Ti-6Al-4V specimens containing fatigue cracks. Those titanium specimens were specially fabricated with cracks of known sizes. The background noise for vibrothermography measurements is usually high and a direct view of the sequence-of-images data after standard background removal procedures has, for small cracks, only limited
power of discriminating inspection regions that do or do not have cracks. A matched filter can be used to increase the signal-to-noise ratio (SNR) if one has knowledge of the expected signal profile. For the vibrothermography measurement data used in this paper, we use an empirically-derived spatial-temporal profile of the temperature changes to construct a matched filter and the output from the matched filter provides important improvements in the SNR when compared with the sequence-of-image input data with background removal.

4.1.2 Related Literature


4.1.3 Overview

The rest of this paper is organized as follows. Section 4-2 presents the standard statistical methods used in NDE and the concept of probability of detection (POD). Section 4-3 describes the experimental setup for the vibrothermography inspection system. Section 4-
4 summaries the matched filter technique. Section 4-5 describes the construction of a matched filter for the vibrothermography sequence-of-images data. Section 4-6 presents the matched filter output dimension reduction. Section 4-7 describes the two detection criteria. Section 4-8 presents the POD comparison results. Section 4-9 contains some concluding remarks and extensions for future research work.

### 4.2 Standard Statistical Methods in NDE

In this section, we outline the standard statistical methods and procedures for a continuous scalar response in NDE applications, as described at MIL-HDBK-1823A (2009). We will use these methods in our comparison with the methods developed here.

#### 4.2.1 Signal Response

We use \( Y \) to denote the NDE measurement response and \( a \) to denote the crack size. Then the statistical model is \( h_Y(Y) = \beta_0 + \beta_1 h_o(a) + \varepsilon \) where \( h_Y(Y) \) and \( h_o(a) \) are specified transformations of the response and flaw size, respectively, \( \beta_0 \) and \( \beta_1 \) are regression parameters, and \( \varepsilon \) is the measurement error following a normal distribution \( N(0, \sigma^2_\varepsilon) \). With the measurement data (possibly censored or truncated), estimates of the parameter vector \( (\hat{\beta}_0, \hat{\beta}_1, \hat{\sigma}^2_\varepsilon) \) and the estimated variance covariance matrix of these estimates can be obtained through standard maximum likelihood (ML) statistical methods. Annis (2009) provided an R package based on the ML method for this and more general linear regression models with censored observations.
It is common to use a normal distribution to describe the variability in $\varepsilon$, although it is possible to use an alternative appropriate distributions when needed. It is possible to use more sophisticated models that are, in some cases, suggested by the physics of the inspection method. In many such cases, however, a simple linear regression model will provide an adequate approximation.

### 4.2.2 Detection Threshold

For specimens without any defects there are still measurement responses due to background noise and other measurement variations. We use $Y_n$ to denote the resulting noise response (or its transformation). Often the noise, which we denote by $Y_n$, can be modeled adequately with a normal distribution. That is, $Y_n \sim N(\mu_n, \sigma_n^2)$. NDE data taken on units without cracks or from those regions of a unit not containing a crack provide noise data from which estimates $(\hat{\mu}_n, \hat{\sigma}_n^2)$ of the noise parameter can be obtained. The detection threshold $(y_{th})$ is typically set to provide an acceptably small probability ($p_f$) of false alarm (e.g., $p_f = 0.01$ or $0.05$).

In particular, the detection threshold can be chosen such that $\Pr(Y_n > y_{th}) = p_f$. Thus the detection threshold is then chosen as $y_{th} = \hat{\mu}_n + \hat{\sigma}_n \Phi^{-1}(1 - p_f)$ where $\Phi^{-1}(x)$ is the standard normal quantile function.

### 4.2.3 POD

For a specified detection threshold, the probability of detection as function of crack size can be obtained as follows:
\[
\text{POD}(a) = \Pr(Y > y_{th}) = \Phi\left(\frac{\beta_0 + \beta_1 h_{vt}(a) - h_Y(y_{th})}{\sigma_z}\right)
\]

(4-1)

where \( \Phi(z) \) is the standard normal cumulative distribution function. With knowledge of the estimated variance-covariance matrix of \((\hat{\beta}_0, \hat{\beta}_1, \hat{\sigma}_y^2)\), confidence bounds for the POD can be obtained by using the delta method (see for example Meeker and Escobar 1998, Appendix B).

## 4.3 Vibrothermography Inspection System

The particular vibrothermography inspection system that was used in our experiments is illustrated conceptually in Figure 4-1 (left). This system involves an excitation source, an infrared camera and a laser vibrometer. The excitation source (a piezo stack) is pneumatically pressed to the sample, and the sample itself is gripped with a rigid or compliant clamp. A coupling medium, such as plastic, is used to separate the tip of the vibration source from the sample. The energy provided by the excitation source causes vibration which in turn causes the crack surfaces to rub and generate heat. The sequence of infrared images, reflecting the sample-surface temperature, is recorded by the infrared camera and the sample surface velocity is measured by the laser vibrometer. The piezo stack that is used as the excitation source typically generates 1 to 2 kW of vibrational power at a fixed frequency such as 20 kHz. The excitation amplitude is tunable in our system and we used three excitation amplitudes \((1.5, 2.2, 3.0)\) in the experiments. Higher excitation amplitude generates more vibrational power.
During an inspection, the vibrational excitation power is coupled into the specimen and the frictional rubbing between crack surfaces generates heat. Both the temperature and the surface velocity are typically recorded at short time intervals for each measurement. In our experiments, the infrared camera sampling rate was 90 Hz and the vibrometer sampling rate was 1 MHz.

The infrared camera takes 150 frames of image for each measurement with the excitation source turning on at frame 20 and turning off at frame 110. The temperature background was obtained by averaging the first 10 frames of the image. The background removal procedure was performed by subtracting the temperature background from each of the 150 frames. Frame 109 was the last frame acquired before the excitation source is turned off and this frame has the highest image contrast.

Figure 4-1. The vibrothermography inspection system setup (left) and a typical spatial pattern image at the frame with highest contrast for a relatively large crack (right).
Figure 4-1 (right) shows frame 109 (i.e. the frame with the highest image contrast) of the sequence-of-images data after background removal for a relatively large crack. We can clearly see the higher temperature at the center of the picture compared with the surrounding areas, and there would be no problem detecting the existence of a crack from the sequence of images from that particular inspection. In general, however, it is important to identify relatively small cracks where the signal within the sequence of images is usually at or close to the noise level and we cannot easily identify the existence of such cracks. Thus the use of statistical methods to boost SNR is needed to setup crack detection criteria with improved sensitivity needed to develop an automatic crack detection algorithm.

4.4 Concept of Matched Filter

The matched filter technique is widely used in signal processing to increase the SNR. An introduction to the matched filter concept can be found in Turin (1960). A matched filter is the optimal linear filter in terms of improving SNR under a stationary white noise process (see for example Engelbery 2007). One requirement for using a matched filter is the need to construct the filter based on the knowledge of the profile of the signal to be detected. In this section we first show conceptually how a matched filter works in a simple one dimensional (1D) example. Then we extend the method to a more complex 3D situation, corresponding to our application.

4.4.1 A One Dimensional Matched Filter

Suppose that in 1D, the signal we are expecting to receive is represented by a set of discrete data points $f[k], k = 1, ..., N$ as shown by filled circles at Figure 4-2 (left) with
$N = 50$. The 1D white noise is represented by filled triangles and the actual measurement (i.e. signal plus noise) is represented by open squares. We denote the discrete input data (e.g. filled triangles or open squares) by $x[k]$, $k = 1,...,N$. By looking at Figure 4-2 (left) directly, it is difficult to distinguish between the signal plus noise data (open squares) and the noise only data (filled triangles), especially when the noise level is high. The matched filter technique utilizes the information of the signal signature to increase the SNR by computing the convolution

$$y[j] = \sum_{k=-\infty}^{\infty} f[j-k] \cdot x[k]$$

as output, where $x[k]$ is the input data (which could be either signal with noise or noise only) and $f[j-k]$ is the reversed known signal (i.e., the matched filter) to be detected. The input data and the matched filter are all discrete-time finite-length arrays. Thus the infinite summation is truncated to be finite and the number of nonzero elements of the output is twice the number of input elements (i.e. $y[j]$, $j = 1,...,2N$).

The matched filter output results for signal plus noise (open squares) and pure white noise (filled triangles) in Figure 4-2 (left) are shown in Figure 4-2 (right) with the same symbolic representation. The matched filter output results for the pure signal are shown in Figure 4-2 (right) also by filled circles. There is a significant difference between the matched filter output of signal plus noise (open squares) and the pure noise (filled triangles). By applying the matched filter, the SNR for the output of the actual measurement is increased importantly. The best discrimination occurs at the output sequence index $N = 50$, just after all of the information has entered the filter convolution (see for example Engelberg 2007).
With the matched filter output results, a reliable automatic classification algorithm can be developed to separate measurement with expected signal and measurement of pure noise.

![Figure 4-2. The one dimensional signal, noise and actual measurement input (left) and the output of signal, noise and actual measurement after applying matched filter (right). The discrete symbols in the plot on the right show only the odd number output elements.](image)

**4.4.2 A Three Dimensional Matched Filter**

One can extend the 1D matched filter to higher dimensions such as 2D for image analysis and 3D for our sequence-of-images analysis. For our vibrothermography data the known signal profile \( f[k_1,k_2,k_3] \) and the input data \( x[k_1,k_2,k_3] \) are now three dimensional arrays with \( k_1 = 1,...,N_1, k_2 = 1,...,N_2 \) and \( k_3 = 1,...,N_3 \). The matched filter is represented by the reversed 3D signal \( f[j_1-k_1,j_2-k_2,j_3-k_3] \) and the convolution is now a three-fold summation:

\[
y[j_1,j_2,j_3] = \sum_{k_1=-\infty}^{\infty} \sum_{k_2=-\infty}^{\infty} \sum_{k_3=-\infty}^{\infty} f[j_1-k_1,j_2-k_2,j_3-k_3] \cdot x[k_1,k_2,k_3]. \quad (4-2)
\]
For large arrays, the three-fold summation in the convolution is computationally intensive. Fortunately, the Fast Fourier Transformation (FFT) algorithm can be used to reduce the computation time dramatically (see, for example, Brigham 1988). For our sequence-of-images data, the whole computation time to finish one sequence-of-images convolution is less than 10 seconds on a standard PC when using the FFT.

4.5 Matched Filter for Vibrothermographic Crack Detection

To construct a 3D matched filter for the sequence-of-images data with background removal, we need to describe the temperature change profile for both the temporal trend and spatial pattern with the presence of a crack. Two approaches can be adopted to get the temperature change profile: (1) use empirical measurements of temperature changes over time and space or (2) use underlying heat-dispersion theory to find the analytical temperature change function.

Figure 4-3. The expected temperature-change profile without noise for matched filter construction: the empirical temporal trend (left) and the Gaussian kernel spatial pattern (right).
For the pixels at the center of a medium-size crack, the temperature change trend at each time frame can be obtained with the following procedures. First we take the average of the 2x2 group of pixels at the center of the crack region at each frame to get a sequence of 150 temporal responses. Then we normalize the sequence of temporal responses by dividing the response of frame 109 (i.e. the frame just before the excitation source is turned off). The empirical temporal trend for the center of a typical medium-size crack is shown at Figure 4-3 (left).

The spatial pattern to represent the temperature changes around the crack is obtained from the normalized 2D Gaussian peak shown at Figure 4-3 (right). The combination of temporal and spatial characteristics taken together provides the 3D matched filter for the sequence-of-images data.

4.6 Dimension Reduction

4.6.1 Regions of Signal and Noise

In this paper, we use an SNR-based detection criterion for crack detection. For the specimens used in this laboratory study, we know the location and size of the crack. We will compare the use of the matched filter technique for regions with and without crack. Also, because we have no data from specimens without cracks, we will use responses from regions of the specimens without a crack to obtain the noise data. The raw 2D image of frame 109 from the vibrothermography measurement is shown in Figure 4-4 with the dashed box indicating the location of the entire specimen.

The crack is located in the region of the left solid box and the pixels inside the left solid box provide an example of matched filter input data when there is a signal. The pixels
inside the right solid box provide an example of matched filter input when there is no signal (i.e., noise data). The same matched filter and detection procedures are used for both regions to assess the performance of our crack detection criterion.

Figure 4-4. The raw data of the 2D image at frame 109 with the specimen location indicated in the dashed box region. The known crack is located in the region of the left solid box. The pixels in the left solid box are used for the signal matched filter input and the pixels in the right solid box are used for the noise matched filter input.

4.6.2 Feature Extraction

After applying the matched filter to a background-removed sequence-of-images input, the output is a 3D array with 300 frames of 2D images with the highest contrast image at or near frame 150 (i.e., the frame corresponding to the time at which all the information has entered the filter, similar to sequence index 50 in the 1D example shown in Figure 4-2 right).
Figure 4-5 (top left) shows the input image of frame 109 (i.e., the last frame with the excitation source turned on) for an inspection region with a small crack (i.e., the left solid box in Figure 4-4), and Figure 4-5 (top right) shows the matched filter output image of frame 150 with a clear “hot spot” indicating the location of the crack.

Figure 4-5. Highest contrast 2D images before (left) and after (right) use of the matched filter for an inspection region with a crack (top) and without a crack (bottom).
To compare the highest contrast images for noise data, the image of frame 109 of the input data and the image of frame 150 of the output data for an inspection region without a crack (i.e., the right solid box in Figure 4-4) for the same measurement are shown at Figure 4-5 (bottom).

Based on a comparison of matched filter outputs for inspection regions with and without a crack for all of the specimens used in the experiment, we developed a noise-threshold detection criterion based on two types of features: (1) the maximum value (MV) in the image of frame 150 after the matched filter as indicated by a cross in Figure 4-5 (top and bottom right) and (2) an empirical characterization of the noise in a rectangle in the general vicinity of the MV (regions between the inner and outer boxes in Figure 4-5 top and bottom right), represented by the noise peak value and the average noise value of averaging all the pixels in the vicinity region between the two boxes.

In vibrothermography inspection applications, the features obtained from an inspection region (i.e. the MV and the peak and average noise values) can be used to make crack existence decisions by comparing the MV signal to a threshold that depends on the local noise level. The inspection region is often a rectangle (e.g., the solid boxes illustrated in Figure 4-4) that is smaller than the specimen. The inspection region is moved around the specimen to cover the entire surface. Adjacent inspection regions are overlapped to avoid edge effects. The detail of the noise threshold detection criterion based on the MV signal, the noise peak value, and average noise value is described at Section 4.7.2.
4.7 Detection Criteria

4.7.1 Temperature Increase

Holland et al. (2010) developed an algorithm to reduce the vibrothermography sequence-of-images data in each measurement into a scalar measure of temperature increase and Li, Holland and Meeker (2010) compared the PODs using the scalar temperature increase for different vibrothermography inspection systems. The algorithm was based on a physical model to perform a surface-fit of the heat from the crack to an elliptical Gaussian envelope. Here we review the method they used to compute POD so that we can compare it with the POD from the matched filter method presented here.

Figure 4-6. The plot on the left shows scalar temperature increase as function of crack size and excitation amplitude (different symbols). The detection threshold is shown as the horizontal dot-dashed line on the left. The crosses shown on the left are the corresponding noise temperature increase taken in regions of the images where there is no crack. These noise data are also shown in the lognormal probability plot on the right.
The scalar temperature increase as function of crack size and excitation amplitude is shown at Figure 4-6 (left). For small cracks, the amount of heat generated is close to the noise level of the inspection system. The same algorithm was applied to regions without a crack to find the noise distribution of the temperature increase. These noise data are independent of the excitation amplitude and are indicated by crosses in Figure 4-6 (left). The lognormal probability plot for the noise data is shown at Figure 4-6 (right) indicates the log transformed temperature increase noise data can be described well by a normal distribution.

The detection threshold for temperature increase was determined such that the probability of a false alarm was 0.02 (i.e., no more than 2% of the temperature increase from regions without a crack exceeded the detection threshold). The detection threshold is shown as a horizontal dot-dashed line in Figure 4-6 (left). The linear regression between temperature increase $T$, crack size $a$ and excitation amplitude $b = 1.5, 2.2, 3.0$ in the log-log scale is:

$$
\log_{10}(T) = \beta_0 + \beta_1 \log_{10}(a) + \beta_2 \log_{10}(b) + \varepsilon_T
$$

(4-3)

where $\beta_0$, $\beta_1$, $\beta_2$ are regression parameters to be estimated from the data and the random variation $\varepsilon_T$ is assumed to have a normal distribution $N\left(0, \sigma_T^2\right)$.

### 4.7.2 A Detection Criterion Based on Signal and Noise

The concept of a variable-noise threshold was used by Nieters et al. (1995) to increase the detection power for C-scan images from ultrasonic inspections of titanium billets. The basic idea is to compare the signal amplitude to a noise threshold that is a function of the estimated noise level in the vicinity of the detection location. By using such a criterion weak signals caused by small flaws in low-noise areas can still be detected. Here we
adapt their idea of a noise threshold detection criterion to make crack-detection decisions based on the output of the matched filter.

Following Nieters et al. (1995), the SNR is defined as $\text{SNR} = \frac{(S_p - N_a)}{(N_a - N_p)}$

where $S_p$ is the signal MV, $N_a$ is the average noise in the surrounding area, and $N_p$ is the peak noise in the surrounding area. The SNR detection criterion is $\text{SNR} > \alpha$ where $\alpha$ may differ depending on the application.

### 4.7.2.1 The matched filter signal response

The signal $S_p$ from our matched filter output corresponds to the MV in the image of frame 150 (e.g., the crosses in the middle of the squares in the right-hand plots in Figure 4-5). The signal $S_p$ from regions with a crack has a strong dependency on crack size and excitation amplitude.

### 4.7.2.2 The noise response

We estimate the distribution of noise in the output of our matched filter by using the vicinity region surrounding the MV (i.e. the region between the two boxes in Figure 4-5 top and bottom right). The maximum value from such region is used to determine the noise peak $N_p$. The average-noise value from such region is defined as $N_a = \sum \eta_i / M$ where $\eta_i$ is the response of each pixel in the region and $M$ is the total number of pixels in the region.

Following Nieters et al. (1995), we define the noise threshold as $N_{th} = \alpha \times N_p + (1 - \alpha) \times N_a$ where $\alpha = 2.5$ has been used as the SNR detection criterion in the Multizone ultrasonic inspection of billets and forgings (e.g., Margetan 2007).
From the model for our matched filter results, the choice of $\alpha = 3.0$ returns 0.02 probability of a false alarm and $\alpha = 3.0$ is thus used in the SNR detection criterion to compare the performance of noise threshold detection criterion with the scalar temperature increase detection criterion.

4.7.2.3 The detection criterion

The SNR based detection criterion that declares a find when $\text{SNR} > \alpha$ is equivalent to $S_p > N_{th}$ or $\log_{10}(S_p) > \log_{10}(N_{th})$. We define $D = \log_{10}(S_p) - \log_{10}(N_{th})$ and the observed values of $D$ are shown in Figure 4-7 (left).

![Normal Probability Scale](image)

**Figure 4-7.** The observations of $D$ as function of crack size and excitation amplitude (left), and the normal probability plot for the observed values of $D = \gamma_0 - \gamma_1 \log_{10}(a) - \gamma_2 \log_{10}(b)$ (right).

The relationship between $D$ and crack size $a$ and the excitation amplitude $b$ is

$$D = \gamma_0 + \gamma_1 \log_{10}(a) + \gamma_2 \log_{10}(b) + \epsilon_D$$
with regression parameters $\gamma_0$, $\gamma_1$, $\gamma_2$ and the random variation $\varepsilon_D$ following a normal distribution $N\left(0, \sigma_D^2\right)$. The detection criterion becomes $D > 0$ and it follows from our model that

$$D \sim N\left(\gamma_0 + \gamma_1 \log_{10}(a) + \gamma_2 \log_{10}(b), \sigma_D^2\right).$$

Figure 4-8. The SNR detection criterion for each excitation amplitude. A detection corresponds to having a MV signal $S_p$ larger than the noise threshold $N_{th}$ (i.e. a point above the diagonal dashed line). The dots correspond to inspection regions with a crack and the crosses correspond to inspection region without a crack.
The normal probability plot for the observed values of $D - \hat{\gamma}_0 - \hat{\gamma}_1 \log_{10}(a) - \hat{\gamma}_2 \log_{10}(b)$ is shown in Figure 4-7 (right) indicating that the normal distribution assumption in (4-4) provides a good description of the data. The relationship between the observed MV signal and the noise threshold is shown in Figure 4-8 for each excitation amplitude with a crack detection criterion being a point above the diagonal dashed line. The dots correspond to inspection regions with a crack and the crosses correspond to inspection regions without a crack. The crosses above the diagonal line are false alarms and the probability of a false alarm.

### 4.8 POD Comparison

With the log-log linear relationship between the scalar temperature increase, crack size and excitation amplitude in (4-3) and the detection threshold set such that the probability of a false alarm is 0.02, the POD for the scalar temperature increase detection criterion can be found by using the approach described in Section 4.2. That is,

$$\text{POD}(a) = \Phi \left( \frac{\beta_0 + \beta_1 \log_{10}(a) + \beta_2 \log_{10}(b) - \log_{10}(y_{T,jh})}{\sigma_T} \right)$$

where $a$ is the crack size, $y_{T,jh}$ is the detection threshold for temperature increase, and $\Phi$ is the standard normal cumulative distribution function. For the detection procedures described in Section 4.7.2.3, the POD for the SNR based noise threshold detection criterion is obtained as

$$\text{POD}(a) = \Pr(D > 0) = \Phi \left( \frac{\gamma_0 + \gamma_1 \log_{10}(a) + \gamma_2 \log_{10}(b)}{\sigma_D} \right).$$
Figure 4-9. The POD and its 95% LB for temperature increase detection and the matched-filter noise threshold detection with a90/95 value for each excitation amplitude (top left, top right and bottom left), and POD mean comparison for all three excitation amplitudes (bottom right).

The estimated POD and its 95% lower confidence bound (LB) based on both the scalar temperature increase and the matched-filter SNR-based detection criteria are shown at Figure 4-9 (top left, top right and bottom left) with a solid line for POD and a dashed line for the POD LB for each excitation amplitude. A POD comparison for all three excitation amplitudes.
amplitudes is shown at Figure 4-9 (bottom right). The comparison shows that the matched-filter SNR-based noise threshold detection criterion provides an overall better POD.

Based on the matched-filter SNR-based detection criterion POD, an automatic detection algorithm with SNR > 3.0 will have, with 95% confidence, a probability of at least 0.90 to detect a crack with size 1.21 mm (known as the a90/95 value in the NDE community) for high excitation amplitude. The a90/95 value for the temperature increase detection criterion is 1.65 mm for the high excitation amplitude. Both a90/95 values are indicated by vertical dotted lines in Figure 4-9.

Compared with the temperature increase criterion, for the same detection confidence (e.g., 95% confidence, a probability of at least 0.90 to detect), the noise threshold criterion can detection cracks 0.44 mm smaller if using high excitation amplitude, and 1.22 mm smaller if using low excitation amplitude.

4.9 Summary and Conclusion

In this paper we have developed a 3D matched filter to greatly increase the SNR of the vibrothermography sequence-of-images inspection data. We suggested a matched-filter SNR-based detection. With detection thresholds set to have the same probability of a false alarm, the SNR detection criterion based on the output of the matched filter has better overall detection performance when compared with the scalar temperature increase results from our previous study.

There are a number of possible extensions for the methodology presented in this paper that suggest future research directions. These include the following:
• Our procedure has been applied to vibrothermography sequence-of-images data from just one system configuration. The performance of our procedures with other vibrothermography detection systems on different kinds of specimens needs to be evaluated.

• We now use a normalized Gaussian peak to represent the temporal-spatial profile. Other types of spatial profiles can be developed to detect particular types of defects such as elongated or triangle shaped cracks.

• It might be possible to find an alternative SNR-based detection criterion that uses the matched filter output in a different manner and that would improve POD without increasing the probability of a false alarm.

Acknowledgements

This material used in the paper is based upon work supported by the Air Force Research Laboratory under Contract # FA8650-04-C-5228 at Iowa State University's Center for NDE.

References

Annis, C. (2009), R package: mh1823, Version 2.5.4.2,


CHAPTER 5. PHYSICAL MODEL ASSISTED PROBABILITY OF DETECTION IN NONDESTRUCTIVE EVALUATION FOR DETECTING OF FLAWS IN TITANIUM FORGINGS

A paper to be submitted to Technometrics.

Ming Li\textsuperscript{1}, William Q. Meeker\textsuperscript{1} and R. Bruce Thompson\textsuperscript{2}

\textsuperscript{1}Center for Nondestructive Evaluation and Department of Statistics, Iowa State University, Ames, Iowa, USA

\textsuperscript{2}Center for Nondestructive Evaluation and Department of Aerospace Engineering, Iowa State University, Ames, Iowa, USA

Abstract

Nondestructive evaluation is used widely in many engineering and industrial areas to detect defects or flaws such as cracks inside parts or structures during manufacturing or for products that need to be inspected while in service. The commonly-used standard statistical model for such data is a simple empirical linear regression between the (possibly transformed) signal response variables and the (possibly transformed) explanatory variable(s). For some applications, such a simple empirical approach is inadequate. An important alternative approach is to use knowledge of the physics of the inspection process to provide information about the underlying relationship between the response and the
explanatory variable or variables. Use of such knowledge can greatly increase the power and accuracy of the statistical analysis and enable, when needed, proper extrapolation outside the range of the observed explanatory variables. This paper describes a set of physical model-assisted analyses to study the capability of two different ultrasonic testing inspection methods to detect synthetic hard alpha inclusion defects in titanium forging disks.

5.1 Introduction

5.1.1 Background

Nondestructive evaluation (NDE) is used to characterize the status or properties of components or structures without causing any permanent physical damage. The aerospace industry is one important NDE application area where failing to detect defects inside airplane components can lead to disasters [see for example NTSB/AAR-89/03 (1989) and NTSB/AAR-90/06 (1990)]. In virtually all NDE applications, there are random effects and errors involved in the measurements and statistical models are needed to analyze the NDE data sets. MIL-HDBK-1823A (2009) describes the standard statistical approaches used in NDE studies. Given a sufficient amount of data over an appropriate region of interest for the explanatory variables (e.g. flaw size and depth), simple empirical statistical models are often adequate to describe the relationship between the response and the explanatory variables. In many applications, however, including the one that motivated this research, the available data are not sufficient to address the questions that need to be answered. Under such circumstances, a physics-based statistical model can sometimes be used to extract the needed information from the limited data. In addition, the physics-based model enables us to extrapolate outside the range of the available data.
As exemplified in NTSB/AAR-90/06 (1990), hard alpha inclusions in titanium alloy aircraft engine disks can lead to serious accidents. A hard alpha inclusion is a brittle nitrogen-based contamination that could cause fatigue cracks to grow more rapidly than what would be otherwise expected in the usually ductile titanium alloy. To develop better NDE tools for detection of hard alpha inclusions, a synthetic inclusion forging disk (known as the SID) was fabricated (details are given in Margetan et al. 2007). The SID contains numerous types of synthetic hard alpha (SHA) inclusions and flat bottom holes (FBHs) of different known sizes. For each inclusion type, there are multiple copies which we refer to as “targets.” These targets are under different surfaces and at different depths.

This paper describes a round-robin experiment in which the SID was inspected by two different ultrasonic testing (UT) methods, with different operators at different locations. We describe the modeling and statistical analyses that were used to estimate the probability of detection (POD) for the synthetic hard alpha inclusions and provide the needed extensions to standard methods that have been used traditionally in the analysis of NDE data. Our modeling and analysis include the use of a physics-based model to describe the relationship between NDE signals and flaw characteristics and the use of a mixed effect model to describe random effects in the inspection process. We also introduce the important concept of making inferences on a quantile of the POD distribution.

5.1.2 Related Literature

procedures for NDE data analyses and Annis (2009) provided an R package to implement these procedures through maximum likelihood (ML) method.

5.1.3 Overview

The rest of this paper is organized as follows. Section 5.2 presents the standard statistical methods used in NDE and the concept of POD. Section 5.3 gives a summary description of the experimental data. Section 5.4 describes the details of the physical models used in the analyses. Section 5.5 presents the physics-based statistical model. Section 5.6 describes the estimation procedures of the statistical model. Section 5.7 presents detailed POD results for different types of defects. Section 5.8 contains some concluding remarks and extensions for future research work.

5.2 Standard Statistical Methods in Nondestructive Evaluation

In this section, we outline the standard statistical methods and procedures that are commonly used in NDE applications, as described at MIL-HDBK-1823A (2009). There are two types of responses in NDE applications: hit and miss binary responses and continuous responses such as voltage. Given the fact that the UT measurements from the titanium forging SID are continuous, we focus on the statistical model for a continuous response.

5.2.1 Statistical Models for NDE

We use \( Y \) to denote the NDE measurement response (or its transformation) and \( x \) to denote the defect size (or its transformation). Other explanatory variables (or their transformation), some of which might be random effects, are denoted by a vector \( z \). Then the statistical model is \( Y = f(x,z,\beta) + \varepsilon \) where \( \beta \) is a vector of regression parameters and \( \varepsilon \)
is the measurement error following a normal distribution $N(0, \sigma^2)$. With the measurement data (possibly censored or truncated) and specified $f(x, z, \beta)$, estimates of the parameter vector $(\hat{\beta}, \hat{\sigma}^2)$ and the estimated variance covariance matrix of these estimates can be obtained through standard ML methods described, for example, in Pawitan (2001). MIL-HDBK-1823A (2009) discussed the commonly used simplest case with $Y = \beta_0 + \beta_1 x + \varepsilon$ and Annis (2009) provided an R package based on the ML method with censored observations for this and more general linear regression models. It is common to use a normal distribution to describe the variability in $\varepsilon$, although it is possible to use alternative appropriate distributions when needed.

### 5.2.2 Detection Threshold

For specimens without any defects there are still measurement responses due to background noise and other measurement variations. We use $Y_n$ to denote the resulting noise response (or its transformation) Often the noise (generally using the same transformation as the response) can be modeled adequately with a normal distribution of $Y_n \sim N(\mu_n, \sigma_n^2)$. NDE noise data can be obtained by taking measurements on units without flaws or from those parts of a unit not containing flaws. These data can then be used to compute ML estimates $(\hat{\mu}_n, \hat{\sigma}_n^2)$ of the noise parameter. The detection threshold $(y_{th})$ is typically set to provide an acceptably small probability ($p_f$) of a false alarm (e.g., $p_f = 0.01$ or $0.05$). In particular, the detection threshold can be chosen such that $\Pr(Y_n > y_{th}) = p_f$. Specifically, the detection
threshold is then chosen as \( y_{th} = \hat{\mu}_n + \hat{\sigma}_n \Phi^{-1}(1 - p_f) \) where \( \Phi^{-1}(x) \) is the standard normal distribution quantile function.

### 5.2.3 Probability of Detection

For a specified model \( f(x, z, \beta) \) and detection threshold, the probability of detection as function of defect size can be obtained as follows:

\[
POD(x) = \Pr(Y > y_{th}) = 1 - \Phi\left( \frac{y_{th} - f(x, z_0, \beta)}{\sigma_y} \right) \tag{5-1}
\]

where \( z_0 \) is a set of fixed explanatory variables and \( \Phi(x) \) is the standard normal cumulative distribution function. Confidence bounds for the POD can be obtained by using delta method (see for example Appendix B in Meeker and Escobar 1998), requiring as inputs, the estimated variance and covariance matrix of the parameter estimates \( (\hat{\beta}, \hat{\sigma}_y^2) \).

### 5.3 Data Description

#### 5.3.1 Data Overview

The titanium SID that was used in the experiments described in this paper contained a large number of cylindrical FBH and SHA targets. A cross section diagram of the SID is shown in Figure 5-1 with longer rods indicating FBH inclusions and shorter rods indicating cylindrical SHA inclusions. For the FBH targets, there were three sizes: #1, #3 and #5 (corresponding to 1/64, 3/64 and 5/64 inches in diameter, respectively). For the SHA targets, there were only two different sizes: #3 and #5. The SHA targets had two different weight
percent nitrogen concentrations \((N_w)\) for each size: 3% and 17%. Thus there were seven different target types. We denote these by #1FBH, #3FBH, #5FBH, #3SHA3, #3SHA17, #5SHA3, and #5SHA17. Detailed information about the SID can be found in Margetan et al. (2007).

The SID was inspected with two different UT inspection methods which are commonly known as the Conventional method (Figure 5-2 top) and the Multizone method (Figure 5-2 bottom). The Conventional method set the focal point near the surface of the SID, and the Multizone method uses several transducers simultaneously each of which has a focal point at certain depth of the SID. Both methods have software depth compensation such that the measurement response has little or no dependency on the depth of a target. The UT response from each measurement within an inspection was a voltage that was, for purposes of statistical analysis, converted, through a scale change, to an Effective Flat Bottom Hole (EFBH) response. The EFBH response is defined as the flat bottom hole area that would give
a signal response equal to the observed response, assuming a common calibration to certain size FBH. In the case of the SID experiment the comparison was to a #1 FBH, corresponding to the specified calibration level that was used for all runs of the experiment (i.e., gain was set such that a #1 FBH would have a response that is 80% of a signal that would cause saturation). This kind of standardized response is often used when it is necessary to combine data with differences in calibration level. For the Multizone method, which uses a signal-to-noise ratio detection criterion, there were additional noise measurements also converted to EFBH units. Noise data was also acquired in the Conventional inspections and used to define detection limits, so that missed targets could be treated as left-censored observations.

Figure 5-2. Conceptual illustration of the Conventional inspection system (top) and the Multizone inspection system (bottom) for billet inspection.

For most observations on individual targets within an inspection, we have exact readings that were translated to EFBH. In some of the inspections, however, the signal was below the noise floor and therefore determined to be a “miss.” These observations are left
censored in that we know only that the actual EFBH response is less than the noise floor EFBH. The noise floor varies from target to target. In some of the Multizone method inspections, the operator did not follow the protocol with respect to saturated observations. The protocol required that, in the case of saturated observations, the operator should reduce the gain in a sequence of steps to a known level where an actual reading could be made. Then this reading could be converted to the actual voltage and corresponding EFBH. When the operator did not follow the protocol, we know only that the EFBH response is larger than the EFBH corresponding to the smallest voltage level that would cause saturation. Figure 5-3 is a summary plot of the data sets used in the analyses for both the Conventional and the Multizone methods, showing the seven different target types. Because the systems used UT probes that were operating at the same nominal frequency (10 MHz) and were calibrated in the same manner, it is not surprising that the amplitude values are similar for the two methods.

Figure 5-3. A summary plot of the data from the Conventional (left) and Multizone (right) inspections.
5.3.2 Operator Plots and Targets Plots

The SID disk was inspected with both the Conventional UT method (two locations, six operators) and the Multizone UT method (three locations, seven operators). Figure 5-4 (operator plots) shows the EFBH response for target type #5SHA3 plotted versus operator, with one line for each target with Conventional results on the left and Multizone results on the right. These plots show that there is an important amount of operator-to-operator variability (i.e., random operator effect) in the EFBH responses for a given target type.

Figure 5-5 (target plots) shows the EFBH response for target type #5SHA3 plotted versus the individual targets, with one line for each operator, again with the Conventional results on the left and the Multizone results on the right. These plots show that there is an important amount of target-to-target variability (i.e., a random target effect) in the EFBH responses for a given target type. The operator plots and target plots for target types other than #5SHA3 are similar and thus not shown here.

![Figure 5-4. Operator plots for the #5SHA3 targets for Conventional (left) and Multizone (right) inspections, with one path for each target.](image-url)
5.4 Physical Model Detail

A typical UT system includes a pulser, a transducer, and a display screen. Driven by the electrical pulses generated by the pulser, the transducer generates an ultrasonic wave. The ultrasonic wave is coupled into and propagates through the SID being tested. When there is a discontinuity such as one of the SHA or FBH targets in the ultrasonic wave propagation path, part of the energy will be reflected. The reflected energy is then transformed into an electrical signal by the transducer and is shown in the display screen. By analyzing the results at the display screen, the existence of defects (SHA or FBH in the SID study considered here) can be determined, and the location and size of the defects can be further evaluated. In this section, several physical models are discussed to describe the principles behind the UT responses for defects with different composition and various sizes.

Figure 5-5. Target plots for the #5SHA3 targets for Conventional (left) and Multizone (right) inspections with one path for each operator.
5.4.1 Reflectance Factor

A key characteristic affecting ultrasonic (and other kinds of) reflection from a discontinuity and the resulting signal strength is a function of the material properties on both sides of the discontinuity. The reflectance factor \( R \)

\[
R = \frac{\rho_i v_i - \rho_m v_m}{\rho_i v_i + \rho_m v_m}
\]  

(5-2)

is used to describe this characteristic. Here \( \rho \) denotes density, \( v \) denotes the ultrasonic wave speed and subscripts \( i \) and \( m \) refer to inclusion and titanium alloy matrix (host material), respectively. Because the density of a flat bottom hole target is essentially zero and the density of the host titanium materials is much larger (i.e., \( \rho_i << \rho_m \)), it follows that \( |R| \) is unity for a FBH. Thus \( R \) can be expressed as a function of weight percent nitrogen concentration for a SHA (3% or 17% in this study) target through the coefficients \( \rho_i \) and \( v_i \).

The effects of SHA nitrogen concentration on the values of \( \rho_i \) and \( v_i \) in titanium alloys were studied experimentally by Gigliotti, Gilmore, and Perocchi, (1994). Based on their experiments and analysis, they reported that

\[
\rho_m = 4461 \text{ kg/m}^3
\]
\[
v_m = 6175 \text{ m/s}
\]
\[
\rho_i = \left(4490.9 + 5.03 \times N_{at} - 0.01 \times N_{at}^2\right) \text{ kg/m}^3
\]
\[
v_i = \left(6002.2 + 61.86 \times N_{at}\right) \text{ m/s}
\]  

(5-3)

where \( N_{at} \) is the atomic percent nitrogen concentration. The relationship between atomic percent nitrogen concentration and weight percent nitrogen concentration \( (N_w) \) is
\[ N_{aw} = \frac{342 \times N_w}{100 + 2.42 \times N_w}. \]  

(5-4)

Thus \( R \) can be used to link the signal responses of the data from the FBH, SHA3, and SHA17 targets and make predictions for intermediate values of weight percent nitrogen.

### 5.4.2 Kirchhoff Approximation

#### 5.4.2.1 General background

When the duration of the incident ultrasonic pulse is sufficiently small with respect to the delay of the back surface echo of the targets, the echoes from the front and back surfaces of the targets can be resolved in time. Under such cases the elastodynamic Kirchhoff approximation (Adler and Achenbach 1980) is appropriate to model the measurement response. With the 10MHz UT system that was used for the Conventional and Multizone inspections, the seven types of target studied in this paper fall in the Kirchhoff regime. Thompson and Lopez (1984) introduced the beam radiation pattern Gaussian approximation concept and concluded the electrical signal \( \text{Voltage}(\omega) \) observed in a pulse-echo experiment for a circular planar surface target can be described by using the following form:

\[
\text{Voltage}(\omega) = A(\omega, z) R \frac{\pi w^2}{2} \left(1 - e^{-2(b/w)^2}\right)
\]  

(5-5)

where \( \omega \) is ultrasonic frequency, \( z \) is the propagation distance, \( b \) is the circular target radius, and \( w \) is the ultrasonic beam radius.
5.4.2.2 Effective flat bottom hole

In production inspections, calibrations are performed to eliminate the effects of the factor $A(\omega, z)$ in (5-5) which account for variations in transducer performance and the effects of propagation distance. Especially when there is need to combine data from measurements that are taken under different calibration levels, it is common practice to scale UT data into what is known as an EFBH response. The EFBH response is intended to represent the FBH area that would produce a signal equal to that which was observed from the target. More precisely, the EFBH is defined as

$$\text{EFBH} = \frac{S}{S_c} \frac{\pi}{4} \left( \frac{D_c}{64} \right)^2 = \frac{S}{S_c} \pi b_c^2$$  \hspace{1cm} (5-6)

where $S$ is the peak defect signal strength [proportional to Voltage(\omega) but in units of percentage of full screen height], $S_c$ is the peak calibration signal strength (in units of percentage of full screen height), $D_c$ is the diameter of the calibration hole (in units of 1/64 inch diameter), and $b_c$ is the radius of the calibration hole in inches. It is easy to show that, for a FBH with size in the Kirchhoff approximation regime, in the absence of noise, the EFBH would be equal to the area of the FBH, consistent with the intent of the definition. Combining (5-5) and (5-6), and assuming calibration to a #1 FBH, and targets in the Kirchhoff approximation regime, the predicted response in units of EFBH would be

$$\text{EFBH} = R \frac{\pi w^2}{2} \left(1 - e^{-2(b/w)^2} \right).$$  \hspace{1cm} (5-7)
This is a powerful result in the context of this study because (5-7) can be used to predict the response of all types of targets in the SID. The size enters through the value of $b$ and the composition (i.e. weight percent nitrogen concentration) through the factor $R$ (which is taken to have a value of 1 for a FBH). This approach allows the data from all of the targets in the SID to be described by a single statistical model, thereby increasing the power of the regression analysis and tightening the confidence bounds. Based on the physical model, valid extrapolation of EFBH values for targets with a radius between #1 and #5 and beyond #5 can be obtained for a range of weight percent nitrogen concentrations in SHAs.

5.4.2.3 Beam limiting Kirchhoff approximation

From (5-7), we can see the beam limiting effect that arises when the defect size $b$ becomes large compared to the ultrasonic beam size $w$. When $b \gg w$, the EFBH response in (5-7), can be simplified as $\text{EFBH}_{b\gg w} = |R| \pi w^2 / 2$. That is, the response is no longer a function of target size $b$ but is only a function of reflectance factor $R$ and beam size $w$.

5.4.3 Rayleigh Scattering Regime

The Kirchhoff approximation is appropriate for the targets that are present in the SID under study in this paper. An additional consideration is the Rayleigh scattering regime where the defect size is small with respect to the ultrasonic wavelength. Although none of the targets in the SID fall in the Rayleigh scattering regime, it is necessary to consider the different response mechanism to avoid improper extrapolations of the Kirchhoff model to defect sizes smaller than the size of a #1 target. Huang, Schmerr, and Sedov (2006)
developed the modified Born approximation from which the model for EFBH is given by

$$\text{EFBH}_B = 2 \pi b^2 \left( \frac{\nu_i}{\nu_m} \right) \left| R \sin \left( \frac{2b\omega}{\nu_i} \right) \right|.$$  

When the defect radius $b$ is sufficiently small such that $\sin \left( \frac{2b\omega}{\nu_i} \right) \approx \frac{2b\omega}{\nu_i}$, the corresponding EFBH model is given by $\text{EFBH}_R = 4\pi b^3 \left| R \frac{\omega}{\nu_m} \right|$ and we say that the response from a target or defect in this size region is described by the Rayleigh limit regime.

### 5.4.4 Physical Model Summary

There are several regimes of scattering determined by the relative values of the target radius $b$, the ultrasonic wavelength $\lambda$, and the beam radius $w$. As the flaw size grows from very small to very large, one will respectively pass through the following regimes:

- **Rayleigh limit**: if $b << \lambda$, the signal is proportional to $b^3$.
- **Modified Born approximation**: if $b < \lambda$, transition from the Rayleigh limit to the Kirchhoff regime with a complex signal pattern is dependent on the spectrum of the ultrasonic pulse.
- **Kirchhoff regime without beam limiting**: if $\lambda < b < w$, the signal is proportional to $b^2$.
- **Kirchhoff regime with beam limiting**: if $\lambda < w = b$, the signal is independent of $b$.

In this work, experimental measurement and the sizes of the SHA and FBH targets in the SID fall within the Kirchhoff regime. Thus in the following statistical modeling, only the Kirchhoff approximation is used.
5.5 Statistical Model

5.5.1 Mean Response

Section 5.4 described the physical models for different size regimes with respect to wave length and beam size. The targets in the SIDs fall into the Kirchhoff regime and the physical response function is written in (5-7) in units of EFBH. By adding a fitting parameter and taking log transformation of (5-7) we have:

\[
\log_{10}\left[\text{EFBH}(x)\right] = \log_{10}(\alpha) + \log_{10}\left[\frac{R \pi w^2}{2} \left(1 - e^{-2(x/w)^2}\right)\right]
\]  

(5-8)

where \(\alpha\) is the scaling fitting parameter that accounts for the overall factor of the Kirchhoff approximation, \(R\) is the reflectance factor, \(w\) is the beam radius, and \(x\) is the target radius. The beam radius \((w)\) is to be estimated from the data and the target radius \((x)\) is in units of mils (a mil is .001 inch).

5.5.2 Weight Percent Nitrogen Concentration Correction

The reflectance factor \((R)\) is equal to 1 for FBH targets and is a function of weight percent nitrogen concentration for SHA targets, as described at Section 5.4.1. The original weight percent nitrogen concentration \((N_w)\) values for SHA targets were 3% and 17%. However the correction to the original weight percent nitrogen concentration is needed because when the SID was HIPped (Hot Isostatic Pressing) after inserting the SHA targets into the forging, the high temperature and pressure caused some unknown amount of the nitrogen to diffuse into the titanium alloy matrix. There was strong evidence of this effect in the SID data. Both experimentally and from physical theory it is known that the amount of
diffusion will depend on the original concentration of nitrogen. The amount of diffusion is related to the complicated HIPing process. Here we assume a typical quadratic correction term to the original weight percent nitrogen concentration as

$$N_{wc} = \left[1 - \beta \times \left(\frac{N_w}{100}\right)^2\right]N_w$$  \hspace{1cm} (5-9)

where $\beta$ is a parameter to be estimated from the data. Then instead of using the original weight percent nitrogen concentration, the corrected weight percent nitrogen concentration in (5-9) is now used in (5-4) as follows

$$N_{wr} = \frac{342 \times N_{wc}}{100 + 2.42 \times N_{wc}}.$$  \hspace{1cm} (5-10)

### 5.5.3 Random Effects

At each inspection location, there were several operators, each of whom inspected the entire disk. There were operator-to-operator variations in the measurement responses even for the same target. There were also target-to-target variations, probably due to variability in the SID fabrication processes and spatial variability in materials properties throughout the SID. To account for these variations, we assumed a random operator effect and a random target effect in addition to the measurement error. We also assume that any differences from site-to-site were due primarily to differences among the operators.

To account for these random effects, the physical model in (5-8) was extended as follows:

$$\log_{10} (\text{EFBH}(x)) = \log_{10} (\alpha) + \log_{10} \left(\frac{R \cdot \pi}{2} w^2 \left(1 - e^{-2(x/w)^2}\right)\right) + \tau + \gamma + \varepsilon$$  \hspace{1cm} (5-11)

with $\tau \sim N(0, \sigma_\tau^2)$, $\gamma \sim N(0, \sigma_\gamma^2)$, $\varepsilon \sim N(0, \sigma_\varepsilon^2)$
where \( \tau \), \( \gamma \) and \( \varepsilon \) are the corresponding operator random effect, target random effect and measurement error, respectively. We assume a normal distribution with mean zero for the operator random effect, the target random effect, and the measurement error. The variances for operator random effect, target random effect and measurement error are \( \sigma_\tau^2 \), \( \sigma_\gamma^2 \) and \( \sigma_\varepsilon^2 \), respectively. Thus, in addition to the three parameters \((\alpha, \beta, w)\) in the physical model in (5-8), we now have three more variance component parameters to be estimated.

To simplify the expression of the statistical model in (5-11), we define the
\[
\mu_{\log_{10}(EFBH)}(x) = \log_{10}(\alpha) + \log_{10}\left(\frac{R}{2\pi}w^2\left(1 - e^{-2(x/w)^2}\right)\right). \tag{5-12}
\]

Then the statistical model can be expressed as
\[
Y(x) = \log_{10}(EFBH(x)) = \mu_{\log_{10}(EFBH)}(x) + \tau + \gamma + \varepsilon.
\]
By defining the total variance as \( \sigma^2_{\text{total}} = \sigma_\tau^2 + \sigma_\gamma^2 + \sigma_\varepsilon^2 \), we can write the log response function in terms of a normal distribution as
\[
Y(x) \sim N\left(\mu_{\log_{10}(EFBH)}(x), \sigma^2_{\text{total}}\right). \tag{5-13}
\]

## 5.6 Estimation

### 5.6.1 Estimation of the Model Parameters

The features in our statistical model and data involve a non-linear response function from the physical model, left and right censored data, random effects, and a need to provide point estimate and bounds to reflect statistical uncertainty. Likelihood based methods (e.g.,
Pawitan 2001) could be used to handle all the above needs and data/model features. No commercial software, however, exists to do such an analysis, and developing such software was not feasible within the timing constraints of our funding sponsor. Bayesian methods (e.g., Gelman, Carlin, Stern, and Rubin. 2003) provide a useful alternative method of analysis.

It is well known that with flat prior distributions the joint posterior distribution is proportional to the likelihood function. Thus with a moderately large amount of data, and diffuse prior distributions, Bayesian methods will produce inferences on functions of the parameters that are similar to what would be obtained by using likelihood-based methods. Furthermore, the software package WinBUGs (2007) is flexible enough (with just a little programming being needed) to handle the data/model features needed for the analysis of the SID data.

In our Bayesian analysis, a Markov Chain Monte Carlo (MCMC) algorithm is used, through WinBUGs, to generate a large number of sampling draws from the joint posterior distribution of the model parameters. After the MCMC algorithm has converged, we have $M$ sampling draws for each model parameter. These $M$ sampling draws are samples from the joint posterior distribution of the parameters. These can in turn be used to compute statistics of interest such as mean, standard deviation, median, 2.5% and 97.5% quantiles of the posterior distribution for each model parameter.

Summary results for all model parameters are shown at Table 5-1 for both Conventional method and Multizone method.
Table 5-1. Posterior mean and standard deviation for all of the model parameters

<table>
<thead>
<tr>
<th>Model Parameter</th>
<th>Conventional Method</th>
<th>Multizone Method</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Posterior Mean</td>
<td>Standard Dev.</td>
</tr>
<tr>
<td>( \alpha )</td>
<td>1.168</td>
<td>0.1441</td>
</tr>
<tr>
<td>( \beta )</td>
<td>14.77</td>
<td>0.9054</td>
</tr>
<tr>
<td>( w )</td>
<td>81.46</td>
<td>9.407</td>
</tr>
<tr>
<td>( \sigma_r )</td>
<td>0.1016</td>
<td>0.04644</td>
</tr>
<tr>
<td>( \sigma_y )</td>
<td>0.05398</td>
<td>0.005752</td>
</tr>
<tr>
<td>( \sigma_e )</td>
<td>0.05343</td>
<td>0.002183</td>
</tr>
</tbody>
</table>

5.6.2 Estimation of Functions of Model Parameters

Besides the model parameters, we can also find the posterior distribution for functions of the model parameters. For example the corrected weight percent nitrogen concentration (\( N_{wc} \)) is a function of the model parameter \( \beta \) defined in (5-9). By substituting in the \( M \) sampling draws of \( \beta \) into (5-9) we can get the \( M \) sampling draws of \( N_{wc} \) for any fixed \( N_w \). We can further get the \( M \) sampling draws of reflectance factor (\( R \)) based on the sampling draws of \( N_{wc} \) through (5-2), (5-3) and (5-10). The posterior mean and standard deviation (in parenthesis) for corrected weight percent nitrogen concentration and reflectance factor are shown at Table 5-2 for both the Conventional method and the Multizone method.
Table 5-2. Posterior mean and standard deviation for the corrected weight percent nitrogen concentration and reflectance factor for 3% and 17% original weight percent nitrogen concentrations

<table>
<thead>
<tr>
<th>$N_w$</th>
<th>Conventional Method</th>
<th>Multizone Method</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$N_{wc}$</td>
<td>$R$</td>
</tr>
<tr>
<td>3%</td>
<td>2.96 (0.0024)</td>
<td>0.041 (0.00004)</td>
</tr>
<tr>
<td>17%</td>
<td>9.74 (0.445)</td>
<td>0.125 (0.0044)</td>
</tr>
</tbody>
</table>

5.6.3 Estimation of the Response Function

In an inspection process with random effects, the true response function and true POD are random (e.g., in our application there would be a different response function for each target/operator combination). The NDE community traditionally focuses on the average quantities in reporting the response function and POD, in effect, averaging over the random effects. We refer to these averages as the *mean response function* and the *mean POD function*, respectively. In some applications, however, there is interest in the worst case scenario among the population of operators and targets. Under such cases a small quantile of response function distribution and a small quantile of POD function distribution for operator and target random effects would be more appropriate metrics to report. In this section we describe the procedures to estimate the mean response function and a quantile of response function distribution. Section 5.7 describes procedures to estimate the mean POD and a quantile of the POD distribution.
5.6.3.1 Mean of the response function distribution

As described in Section 5.6.1, we used the internal MCMC simulation algorithm in WinBUGs (2007) to generate sampling draws for all of the parameters in the statistical model (i.e., $\alpha$, $\beta$, $w$, $\sigma^2_z$, $\sigma^2_\gamma$ and $\sigma^2_\varepsilon$). We then used these sampling draws to generate the sampling draws of $\mu_{\log_{10}(\text{EFBH})}(x)$ through (5-12). Figure 5-6 shows the mean response functions (solid lines) with 95% lower credible bounds (LCBs) (dashed lines) versus target areas for each target type with Conventional results on the top and Multizone results on the bottom. The amplitude detection criterion is the same for Conventional and Multizone inspections and is indicated as horizontal solid lines at these plots. The Multizone inspection uses, in addition, a signal-to-noise ratio criterion, as described in Section 5.7. Also shown in these plots are the exact, left censored and right censored data points denoted by circles, down triangles and up triangles, respectively.

5.6.3.2 Quantile of the response function distribution

In many applications it is important to obtain estimates of quantities in the tail of a distribution, as opposed to the mean or other measure of central tendency. For example, in the SID inspection experiment the data tell us that some targets and some operators tend to result in weaker signals than others. Consider a random draw of an operator $\tau \sim N\left(0,\sigma^2_\tau\right)$ and a target $\gamma \sim N\left(0,\sigma^2_\gamma\right)$. Important functions of these random effects such as $\log_{10}(\text{EFBH}(x))|_{\tau,\gamma}$ and $\text{POD}(x)|_{\tau,\gamma}$, where $x$ is the target area, will have their own distributions.
Figure 5-6. Estimates of the mean response functions, the 0.05 quantiles of response function distribution and their corresponding 95% LCBs for the Conventional (top) and Multizone (bottom) inspection methods.
The mean response for a particular operator and target (averaging over measurement error) can be described by the random variable

\[ Y(x)_{\tau, \gamma} = \mu_{\log(EFBH)}(x) + \tau + \gamma. \]  

with mean \( \mu_Y(x) = \mu_{\log(EFBH)}(x) \) and variance \( \sigma_Y^2 = \sigma_{\tau}^2 + \sigma_{\gamma}^2 \). The \( p \) quantile of \( Y(x)_{\tau, \gamma} \) is \( y_p(x) = \mu_Y(x) + z_p \sigma_Y \) where \( z_p \) is the standard normal \( p \) quantile. Here \( \varepsilon \sim N(0, \sigma_\varepsilon^2) \) is the consolidation of all other variations in the measurement after a particular operator and target are selected.

In our examples we focus on the 0.05 quantile of the response function. This quantile can be interpreted as the mean response value that will be exceeded by 95% of the target and operator combinations from the population of targets and operators. Because \( y_{0.05}(x) \) is a function of the model parameters we can estimate its mean and compute a corresponding 95% LCB by using the sampling draws of \( y_{0.05}(x) = \mu_Y(x) + z_{0.05} \sigma_Y \). Figure 5-6 shows the mean of the 0.05 quantiles of the response function distribution (dashed-dotted lines) and their LCBs (dotted lines) versus target areas for the 3% SHA, 17% SHA, and FBH targets respectively with Conventional results on the top and Multizone results on the bottom. Compared to the tight 95% LCBs on the mean response function, the LCBs for the 0.05 quantiles of the response function distribution are further away from the estimate of the response quantile.

### 5.7 Diagnostics

It is important to assess how well the statistical model fits the experimental data. Figure 5-7 shows the residuals versus fitted values with Conventional results on the left and
Multizone results on the right. The residuals are evenly distributed except for those from the #5FBH targets. The reason for this deviation is that there are relatively few data points for #3FBH and #5FBH and thus these observations are not influential in fitting the model. For #5FBH targets the residuals are below zero which indicates an upward bias in estimation for a #5FBH. This does not raise serious practical concerns because there is little practical interest in predicting POD in the target space region anywhere near to the #5 FBHs. We also compared the results between including #5FBH targets and excluding #5FBH targets when doing the analysis. The results showed little change in the mean response function, and the PODs were more conservative for analysis that includes the #5 FBH targets and thus all the results in this paper are based on analyses that include the #5FBH targets.

Figure 5-7. Residual plots as function of fitted value for Conventional (left) and Multizone (right) inspection methods.
5.8 Probability of Detection

Given the response function and the detection threshold, the mean POD, the quantile of POD distribution and the corresponding LCBs can be obtained. In this section we first describe the procedures to estimate POD for Conventional and Multizone respectively. Then we present the POD plots of both inspection methods for all types of targets.

5.8.1 The Conventional Inspection Method

5.8.1.1 Mean POD

For the Conventional method, the detection threshold is set as
\[ y_{th} = \log_{10}(191.75) = 2.2827, \]
where 191.75 is the area of a #1 FBH in units of square mils. Sensitivity to a #1 FBH was the inspection sensitivity agreed upon by jet engine manufacturers and the Federal Aviation Administration. POD can be found by computing
\[ \text{POD}(x) = \Pr(Y(x) > y_{th}) \]
where \( x \) is the target area and the random variable \( Y(x) \) is defined in (5-13). Specifically, the \( \text{POD}(x) \) is evaluated as follows:
\[ \text{POD}(x) = \Pr(Y(x) > y_{th}) \Phi \left( \frac{\mu_{\log_{10}(\text{FBH})}(x) - y_{th}}{\sigma_{\text{total}}} \right) \]
(5-15)
where \( \Phi(x) \) is the standard normal cumulative distribution function and \( \mu_{\log_{10}(\text{FBH})}(x) \) is defined in (5-12). With the sampling draws of \( \mu_{\log_{10}(\text{FBH})}(x) \) and \( \sigma_{\text{total}} \), we can compute the corresponding sampling draws of \( \text{POD}(x) \). Estimates of the mean POD and a
corresponding 95% LCB can be found by computing the sampling draws of $POD(x)$ over a range of $x$ values.

### 5.8.1.2 Quantile of the POD distribution

Again, consider a random draw of an operator and a target. Some combinations will result in higher POD than others. As with the derivation of the quantile of response function distribution in Section 5.6.3.2, we can take account of this variability by computing a quantile of the POD distribution. An expression for the $p$ quantile of the POD distribution for the Conventional method is obtained by replacing $\mu_{\log_{\text{EFBH}}}(x)$ with $y_p(x) = \mu_p(x) + z_p \sigma_y$ and replacing $\sigma^2_{\text{total}}$ with $\sigma^2_\varepsilon$ in (5-15). In particular, the $p$ quantile of the POD distribution for target size $x$ is

$$\left[ POD(x) \right]_p = \Phi \left( \frac{y_p(x) - y_{\text{th}}}{\sigma_\varepsilon} \right).$$

(5-16)

Again, estimates of the 0.05 quantile of the POD distribution and corresponding 95% LCB were obtained by computing the sampling draws of the 0.05 quantile of the $POD(x)$ distribution for different values of $x$.

### 5.8.2 The Multizone Inspection Method

#### 5.8.2.1 Mean POD

The Multizone inspection method uses a signal-to-noise ratio (SNR) detection rule in addition to the amplitude detection criterion used in the Conventional method. Nieters et al. (1995) used the following definition for SNR and a corresponding detection limit.
\[ \text{SNR} = \frac{Y - N_a}{N_p - N_a} \]

Here \( Y \) is the UT signal measurement. \( N_a \) is the noise average and \( N_p \) is the noise peak in a defined rectangular region with the rectangle containing the target signal cut out. The industry standard detection criterion for SNR detection in Multizone inspection is \( \text{SNR} > 2.5 \). Then the SNR criterion is equivalent to \( Y > N_{th} = 2.5N_p - 1.5N_a \) where \( N_{th} \) is defined as the noise threshold. Instead of modeling the SNR, it is easier to estimate the signal distribution and the noise-threshold distribution directly. The noise threshold varies from target to target and from disk to disk and can be computed from the results of the Multizone experimental results. The variability in the noise threshold data can be described by a normal distribution:

\[
N_{th} \sim N(\mu_{\text{noise}}, \sigma_{\text{noise}}^2).
\] (5-17)

Figure 5-8 illustrates this two-dimensional Multizone detection criterion. There is an amplitude detection if the amplitude is above the horizontal line. There is also a SNR detection if the amplitude is above the noise threshold (i.e., if the amplitude/noise threshold point lies above the diagonal line in Figure 5-8).

The POD for SNR noise threshold detection criterion is:

\[
\text{POD}_i(x) = \Pr(Y(x) > N_{th}) = \Phi \left( \frac{\mu_{\text{log(ar)[EFBH]}}(x) - \mu_{\text{noise}}}{\sqrt{\sigma_{\text{total}}^2 + \sigma_{\text{noise}}^2}} \right)
\]
with the random variable \( Y(x) \) defined in (5-13) and \( N_{th} \) is defined in (5-17). Given the independent relationship between \( Y(x) \) and \( N_{th} \), the joint density for the response function and noise threshold is:

\[
f(y, n_{th}) = \phi(y, \mu_{\log(EFBH)}, \sigma_{\text{total}}^2) \phi(n_{th}, \mu_{\text{noise}}, \sigma_{\text{noise}}^2)
\]

with \( \phi(x, \mu, \sigma^2) \) the normal density function with mean \( \mu \) and variance \( \sigma^2 \).

Figure 5-8. Illustration of the Multizone detection criteria.

The POD for regions with \( Y(x) \leq N_{th} \) but \( Y(x) > y_{th} \) (i.e. the triangle at right edge of Figure 5-8) is

\[
POD_2(x) = \Pr(Y(x) > y_{th} \text{ and } Y(x) \leq N_{th}) = \int_{y_{th}}^{\infty} \int_{y_{th}}^{n_{th}} f(y, n_{th}) dy
\]
which is calculated by numerical integration. Then the Multizone mean POD\((x)\) is determined as

\[
POD(x) = POD_1(x) + POD_2(x).
\]  

The estimate of the noise mean is 1.7990. The target-to-target noise variance estimate is 0.02860. We do not have data that would provide a disk-to-disk noise variance estimate, but, for purposes of illustration, we assume the disk-to-disk variance estimate is 0.01430. Thus the estimate of total noise variance is 0.04290. The sampling draws of \(\mu_{\text{noise}}\) and \(\sigma_{\text{noise}}^2\) were used to compute the mean POD estimate and the corresponding 95% LCB for Multizone inspection through (5-18).

### 5.8.2.2 Quantile of POD distribution

Similar to the quantile of POD distribution for the Conventional method, by replacing \(\mu_{\log,\text{EFBH}}(x)\) with \(y_p(x)\) and replacing \(\sigma_{\text{total}}^2\) with \(\sigma_e^2\) in (5-18) we can get the Multizone \(p\) quantile of POD distribution for any target size \(x\).

### 5.8.3 POD Plots

Figure 5-9 contains plots of the estimates of the mean of the POD distribution (solid lines), the corresponding 95% LCBs (dashed lines), the 0.05 quantile of POD distribution (dashed-dotted lines) and the corresponding 95% LCBs (dotted lines) for 3% SHA, 17% SHA and FBH targets for both inspection methods.
Figure 5-9. Estimates of the mean PODs, the 0.05 quantiles of POD distribution and their corresponding 95% LCBs for the Conventional (top) and Multizone (bottom) inspection methods.
Figure 5-9 shows that although there was little difference between the inspection methods when looking at the signal-response functions estimates in Figure 5-6, there are large differences between the estimates of the POD functions. This is due to the important increase in detection power provided by the more complicated SNR detection criterion used in the Multizone inspection method and to some degree because there is less operator-to-operator variability in the Multizone inspection method.

5.9 Summary and Conclusion

In this paper, we described the establishment and application of a statistical model for quantifying inspection capability and estimating POD, based on the physical mechanisms of an ultrasonic testing process. The physics-based statistical model enabled needed information extraction from data taken on the limited types and sizes of the synthetic inclusion targets in the synthetic inclusion titanium disk that was available for the experiment. The physics-based model further made possible the needed interpolation and extrapolation for a wider range of flaw sizes and nitrogen concentrations. The nonlinear response function, random effects, and the censored observations were accommodated in the statistical part of the physics-based model. The Markov Chain Monte Carlo based Bayesian software WinBUGs was utilized with a diffuse prior distribution for estimation of the model-tuning parameters. The mean and 0.05 quantile of the response functions and the POD curves for a representative set of target areas and target types were presented. The results from this study provide useful information about the ability to detect hard alpha inclusions in titanium forgings. The methodology provided here is, however, more general and could be used to study NDE inspection capability in other areas of application and for other kinds of inspection.
There are a number of extensions for the methodology presented in this article that suggest future research directions. These include the following:

1. The target sizes and flaw sizes of interest in this study were within the range where the Kirchhoff approximation provides a good description of ultrasonic testing signals. Although not adopted in this study, an explicit extrapolation procedure based on the Rayleigh scattering regime could be developed when needed, allowing extrapolation to smaller flaw sizes.

2. The quadratic term correction for the weight percent nitrogen concentration was used to account for nitrogen diffusion during HIPping process. A correction based on a physical principle could be implemented if we had more knowledge about the mechanism behind the diffusion arising in the HIPping process.

3. For applications where there is useful prior information about the model parameters (e.g., from previous experience with a particular kind of inspection), a Bayesian analysis with informative prior distributions could be implemented.

4. The current model showed some lack of fit for the #5 FBH condition. Further experimentation on a different inclusion sample with targets having reflectance in the gap between the 17% nitrogen and the flat bottom holes might make it possible to resolve the reasons for this deviation from the physics-based model. Presently, no such sample block is known to exist.

**Acknowledgments**

This work was performed at the Center for Nondestructive Evaluation at Iowa State University with support of the Federal Aviation Administration under contract number 08-C-
We would also like to thank other members of the project team for their valuable inputs and guidance and for providing details of the inspection methodology that was important when making decisions about data processing and model development. These include Lisa Brasche (Iowa State), Waled Hassan (Rolls-Royce), Richard Klaassen, (GE Aviation), Tim Mouzakis (FAA), Cu Nguyen (FAA), Surendra Singh (Honeywell), Floyd Spencer (FAA consultant), and Harpreet Wasan (Pratt and Whitney). We would like to thank Floyd Spencer for suggesting that, for some purposes, it is important to estimate the quantile of the POD distribution.

The experimental results presented in this paper are based on the efforts of the aforementioned team. However, as a final report has not been completed and approved by all, these results should be considered to represent the current views of the authors of this paper and not an endorsement by the total team or the funding agency.

References

Annis, C. (2009), R package: mh1823, Version 2.5.4.2,


CHAPTER 6. JOINT ESTIMATION OF NDE INSPECTION CAPABILITY AND FLAW-SIZE DISTRIBUTION FOR IN-SERVICE AIRCRAFT INSPECTIONS

A paper to be submitted.

Ming Li¹, William Q. Meeker¹, and Peter Hovey²

¹Center for Nondestructive Evaluation and Department of Statistics, Iowa State University, Ames, Iowa, USA
²Department of Mathematics, University of Dayton, Dayton, OH 45469

Abstract

Nondestructive evaluation (NDE) is widely used in the aerospace industry during scheduled maintenance inspections to detect cracks or other anomalies in structural and rotating components. Life prediction and inspection interval decisions in aerospace applications require knowledge of the size distribution of unknown existing cracks and the probability of detecting a crack (POD), as a function of crack characteristics (e.g., crack length). The POD for a particular inspection method is usually estimated on the basis of laboratory experiments on a given specimen set. These experiments, however, cannot duplicate the conditions of in-service inspections. Quantifying the size distribution of unknown existing cracks is more difficult. If NDE signal strength is recorded at all
inspections and if crack-length information is obtained after “crack find” inspections, it is possible to estimate the joint distribution of crack length, noise response and signal response. This joint distribution can then be used to estimate both the in service POD and the crack-length distribution at a given period of service time. In this paper, we present a statistical model and methodology to do this estimation.

6.1 Introduction

6.1.1 Background

Nondestructive evaluation (NDE) methods that use non-intrusive physical measurements are widely used in aerospace applications to detect flaws or cracks inside structures or parts. Depending on the situation (e.g., a designed laboratory study versus a field study that is based on actual inspection data) and the particular structure of the data that are collected, different statistical models and methods are needed to analyze the NDE data. MIL-HDBK-1823A (2009) describes the standard statistical methods that are used in laboratory studies.

Carefully designed laboratory experiments are expensive, but provide flexibility to study the effect of particular experimental factors. Laboratory experiments are usually based on artificial cracks or other flaws in test specimens (e.g., Li, Meeker and Thompson 2010). The measurement response is modeled as a function of crack length and this model is used to estimate the probability of detection (POD). The laboratory studies are usually for validation and quantification of inspection capability for new NDE methods. After a detection method is developed, tested in the laboratory, and put into use, there is often a desire or a need to do
a field study to assess actual field performance and to monitor the inspection process over time to assure that it is being done effectively.

Regularly-scheduled in-service nondestructive inspections look for cracks in aircraft components such as engine fan blades and lap-splice rivet holes. Such inspections are, for example, an integral part of the FAA Aging Aircraft program. The purpose of these inspections is to determine whether there is a crack in a part and if a crack is detected there is usually a need to determine the approximate size of the crack. For a particular inspection method, there is a detection threshold, often based on previous field inspections, laboratory experience, model-based theory, and operator experience. For parts with a signal response above the detection threshold, a crack detection decision is made and that part is either repaired or removed from future service. The crack-length information could be obtained during repairing or other post detection procedures.

In most applications when an NDE measurement is taken in a place where we know there are no cracks, the reading can still be some value to quantify background and measurement noise. When there are very small flaws, small signals close to the noise level will be obtained from the measurements. Based only on the measurements, we cannot be sure that such measurements were from the crack or some artifact of the part or the test setup that would cause noise. We use a noise-interference model to describe the relationship between signal and noise.

6.1.2 Motivation

This work was motivated by the need to use information from in-service inspections of lap-splice rivet holes used on aircraft bodies. If a crack signal exceeds a crack-detection
threshold, then crack-length information is obtained during the repair process. Although measurements are taken on all holes during each scheduled inspections before the “crack find” inspection, currently there are no data recorded for these measurements. Modern inspection and communications technology will, however, make it possible to record these measurements with little effort, allowing better estimation of POD and crack-length distributions. For small cracks, the measurement response could be signal from the crack or the noise artifact. We use a noise-interference model to describe the rivet-hole measurement data by assuming that the response is the maximum of the signal and the noise.

The estimation methods proposed here require keeping the repeated measurement response records for all holes with a growing crack. There is one crack-length reading for holes at the “crack find” inspection but no crack-length reading for holes that have not had an above-threshold measurement response. There are no standard statistical methods to analyze such inspection data. In this paper we develop a statistical method to jointly estimate the measurement (i.e., maximum of the signal response and the noise response) and crack length based on assumed knowledge of the crack growth model. The joint estimation increases the power of the statistical analysis and improves the overall reliability assessment. Although the research in this paper was motivated by the rivet-hole inspection applications, the methods presented here should have broad applicability into other areas of NDE inspections.

6.1.3 Related Literature

these procedures through maximum likelihood method. Li and Meeker (2009) introduced the
noise interference model to extract the signal response from the NDE measurement. Hovey,
Meeker and Li (2009) discussed a similar crack growth NDE problem with one fixed crack
growth rate. There are a number of books that have discussion about statistical methods for
repeated measurement data (e.g. Davidian and Giltinan 1995). Johnson and Wichern (2001)
summarize properties of the multivariate normal distribution that is used in our joint
estimation model.

6.1.4 Overview

The rest of this paper is organized as follows. Section 6-2 describes the inspection
procedures and the data structure for scheduled aircraft maintenance inspection. Section 6-3
presents the standard statistical methods used in NDE and the concept of POD. Section 6-4
describes the crack growth model and the measurement response model. Section 6-5
describes the statistical model for the simulated field data. Section 6-6 describes the Bayesian
estimation of the parameters and functions of the parameters of the statistical model. Section
6-7 contains some concluding remarks and extensions for future research. A summary of the
bivariate normal distribution properties used in this paper is presented at the Appendix.

6.2 In-service Inspection of Aircraft Lap-splice Rivet Holes

The current procedures for aircraft maintenance require measuring every hole with an
eddy current inspection method at each scheduled inspection, but only find or no-find
information is recorded. As a result, currently, there are no available field study data sets
based on our proposed data recording scheme. Therefore we use a simulated data set for rivet
holes used in lap-splices on aircraft bodies to illustrate our proposed inspection procedures and to present the joint estimation statistical methodology. The parameters used in the simulation are based on previous experience with eddy current NDE inspections for rivet holes, as described in Hovey, Meeker, and Li (2009) and Li, Nakagawa, Larson, and Meeker (2010).

The proposed inspection procedures are outlined as follows.

- First, one measurement is taken for each rivet hole at each scheduled inspection. For any rivet holes with measurement below the detection threshold, we will assume the crack is small enough that the rivet hole can, without risk, be continued in service without repair. Thus there is no direct crack-length information for those rivet holes with measurement below the detection threshold.

- Second, at any scheduled inspections, if a rivet hole has a measurement above the detection threshold, the rivet hole is repaired and the crack-length information in units of inches is obtained during the repair procedure.

This paper considers only the time to a first detectable crack at each hole.

The eddy current measurement from each hole could come from the signal response of a crack or the noise artifact response (e.g., an innocuous scratch). In eddy current inspection, the log signal response is usually described adequately with a linear relationship with the log crack length. The noise response can be described by a log normal distribution and is independent of crack length. For small cracks, the signal responses are usually smaller than noise response (i.e., below the noise floor of the eddy current inspection output). We therefore model eddy current measurement responses by using the noise interference model.
(i.e., the maximum of the signal response and noise response). The simulated results from the proposed inspection procedures are illustrated in Figure 6-1.

Figure 6-1. Simulated aircraft rivet hole field data: the full data set of signal and noise response as function of inspection time (top), the full data set of measurement results as function of crack length (bottom left) and the actual observed data structure for proposed field inspection procedures (bottom right).
Figure 6-1 (top) shows the full data set of signal response (open circles) and noise response (crosses) for each scheduled inspection with service time in thousand hours. Figure 6-1 (bottom left) shows the relationship between the measurement result (i.e. the maximum of signal response and noise response) and the crack length for the full data set. Figure 6-1 (bottom right) shows the structure of the actual data that would be observed in real applications based on the proposed inspection procedures. The preset detection threshold \( y_{th} = \log_{10}(1000) \) and noise mean \( \mu_{\text{noise}} = \log_{10}(316) \) are also indicated in Figure 6-1 by horizontal dashed lines and dotted lines respectively. An estimate of the probability of false alarm for this data set is 0.028 (i.e., the proportion of crosses that are above the detection threshold in the top of Figure 6-1 is 0.028).

### 6.3 Standard Statistical Methods in Nondestructive

In this section, we outline the standard statistical procedures in NDE for accessing inspection capability in applications, as described in MIL-HDBK-1823A (2009). There are two types of response in NDE applications: hit and miss binary responses and continuous responses such as voltage. Because the measurements from the rivet-hole field data are continuous, we focus on the statistical model for a continuous response. In subsequent sections, we will present extensions to these existing methods that will allow joint estimation of inspection capability and a flaw size distribution by using data coming from regularly scheduled in-service inspections.
6.3.1 Statistical Models for NDE

We use $Y$ to denote the NDE measurement response (or its transformation) and $x$ to denote the crack length (or its transformation). Then the statistical model is $Y = \beta_0 + \beta_1 x + \varepsilon$ where $\beta_0$ and $\beta_1$ are the regression parameters and $\varepsilon$ is the measurement error with a normal distribution $N(0,\sigma^2_\varepsilon)$. With the measurement data (possibly censored or truncated), estimates of the parameter vector $(\hat{\beta}_0, \hat{\beta}_1, \hat{\sigma}_\varepsilon^2)$ and the estimated variance covariance matrix of these estimates can be obtained by using standard maximum likelihood (ML) methods described, for example, in Pawitan (2001). MIL-HDBK-1823A (2009) discussed this model and Annis (2009) provided an R package based on the ML method with censored observations. It is common to use a normal distribution to describe the variability in $\varepsilon$, although it is possible to use alternative appropriate distributions when needed.

6.3.2 Detection Threshold

For rivet holes with cracks that are very small (even newly drilled holes can be considered to have micro-cracks of size on the order of grain boundary sizes), there are still measurement responses due to background noise and other measurement variations. We use $Y_{\text{noise}}$ to denote the resulting log noise response which we assume to have a normal distribution of $Y_{\text{noise}} \sim N(\mu_{\text{noise}},\sigma_{\text{noise}}^2)$. NDE data taken on new rivet holes provide noise data from which ML estimates $(\hat{\mu}_{\text{noise}}, \hat{\sigma}_{\text{noise}}^2)$ of the noise parameters can be obtained. The detection threshold $(y_{\text{th}})$ is typically set to provide an acceptably small probability $(p_f)$ of false alarm (e.g., $p_f = 0.01$ or $0.05$). In particular, the detection threshold can be chosen
such that $\Pr(Y_{\text{noise}} > y_{th}) = p_f$. Specifically, the detection threshold is chosen as

$$y_{th} = \mu_{\text{noise}} + \sigma_{\text{noise}} \Phi^{-1}(1 - p_f)$$

where $\Phi^{-1}(x)$ is the standard normal distribution quantile function.

### 6.3.3 Probability of Detection

With a specified detection threshold, POD as a function of crack length can be obtained as follows:

$$\text{POD}(x) = \Pr(Y > y_{th}) = \Phi \left( \frac{\beta_0 + \beta_1 x - y_{th}}{\sigma_y} \right)$$

(6-1)

where $\Phi(x)$ is the standard normal cumulative distribution function. Confidence bounds for the POD can be obtained by using the delta method or by using the likelihood directly (e.g., Meeker and Escobar 1998, Appendix B) where the estimated variance and covariance matrix of the model parameters is needed as an input.

### 6.4 Crack-length and Measurement-response Models

In this paper, we focus on the fatigue crack growth in aircraft lap-splices rivet holes. The crack growth models can be used to compute reliability properties (see, for example, Chapter 13 of Meeker and Escobar 1998). Crack growth models can be very complicated and are usually function of geometry, materials properties, and usage environmental variables. Many of the more sophisticated models are developed for particular applications and are proprietary. We use a simple fatigue crack growth model that assumes exponential growth
over time. This simple model could be extended to more complicated crack growth models when needed.

We denote the crack length by $a$. We assume that cracks have a random initial size and grow deterministically with rates that are random from aircraft to aircraft. This is a standard model, used in fatigue-fracture aerospace applications.

### 6.4.1 Initial Crack Length Distribution

We assume that there is a crack at each rivet hole location at time $t_0$. Those cracks are generally very small and the log crack length for rivet hole $i$ in aircraft $j$ is $x_{ij}(t_0) = \log_{10}(a_{ij}(t_0))$ with $i = 1,\ldots,I$ and $j = 1,\ldots,J$. In our model, the log initial crack length follows a normal distribution $N(\mu_x,\sigma_x^2)$.

### 6.4.2 Crack Growth Model

In our model, the size of the crack at rivet hole location $i$ in aircraft $j$ at inspection time $t$ is denoted by $a_{ij}(t) = a_{ij}(t_0)\exp(\lambda_{ij}(t-t_0))$. To take into account the different crack growth rates from aircraft-to-aircraft, a random crack growth model is needed. We assume the log crack growth rates $\log_{10}(\lambda_{ij})$ for $j = 1,\ldots,J$ follow a normal distribution $N(\mu_\lambda,\sigma_\lambda^2)$, although it is possible to use an alternative appropriate distribution when needed. The inspection time $t$ is the same for all rivet holes in the same aircraft at each scheduled inspection (with index $k = 1,\ldots,K$). For rivet holes in different aircraft at the same scheduled inspection, the actual service time may be different because of inspection-scheduling variability. Thus we identify the service time $t = t_{jk}$ by index $(j,k)$. The log crack length
\[ x_y(t_{jk}) = \log_{10}(a_y(t_{jk})) \] at time \( t_{jk} \) is \[ x_y(t_{jk}) = x_y(t_0) + \lambda_j(t_{jk} - t_0) \] where \( \lambda_j \) is the crack growth rate for aircraft \( j \).

### 6.4.3 Eddy Current Response Model

#### 6.4.3.1 Signal response

In our model, the log signal response (open circles in top of Figure 6-1) for the rivet hole at location \( i \) in aircraft \( j \) at scheduled inspection \( k \) is

\[ Y_y(t_{jk}) = \beta_0 + \beta_1 x_y(t_{jk}) + \varepsilon_{ijk} \]

with \( \varepsilon_{ijk} \sim N(0, \sigma_y^2) \). Here we assume that the signal response errors \( \varepsilon_{ijk} \) are independently and identically distributed. Finally, recalling that the log initial crack length follows a normal distribution \( N(\mu_x, \sigma_x^2) \) and using the bivariate normal distribution results (A-1) and (A-2) in the Appendix, the crack-size/NDE signal can be modeled through a random vector \((Y(t_{jk}), X(t_{jk}))^T\) with a bivariate normal distribution:

\[
\begin{bmatrix}
Y(t_{jk}) \\
X(t_{jk})
\end{bmatrix}
\sim \text{BVN}\left(\begin{bmatrix}
\beta_0 + \beta_1 (x_y + \lambda_j(t_{jk} - t_0)) \\
\mu_x + \lambda_j(t_{jk} - t_0)
\end{bmatrix}, \begin{bmatrix}
\sigma_y^2 + \beta_1^2 \sigma_x^2 & \beta_1 \sigma_x^2 \\
\beta_1 \sigma_x^2 & \sigma_x^2
\end{bmatrix}\right).
\]

(6-2)

#### 6.4.3.2 Noise response

The log noise response for rivet-hole inspections at any inspection time can be described by a normal distribution

\[ Y_{\text{noise}}(t_{jk}) \sim N(\mu_{\text{noise}}, \sigma_{\text{noise}}^2) \]

(6-3)
with mean $\mu_{\text{noise}}$ and variance $\sigma_{\text{noise}}^2$. The log noise response is independent of crack length and service time. The proportion of noise data above the detection threshold is 0.028 (i.e. the PFA for our simulated data is 0.028).

### 6.4.3.3 Noise interference model

The actual eddy current NDE response is the maximum of the signal response and the noise response. That is,

$$Y_{\text{actual}}(t_{jk}) = \max(Y(t_{jk}), Y_{\text{noise}}(t_{jk})). \quad (6-4)$$

Li and Meeker (2009) and Li, Nakagawa, Larson, and Meeker (2010) used the noise interference model to describe NDE measurement responses in other applications.

### 6.4.4 Simulation Parameters

Based on available expert knowledge and previous experience with the crack growth in lap-splice rivet holes in aircraft bodies (e.g., Li, Nakagawa, Larson, and Meeker 2010), our simulation parameters were chosen as follows:

- Initial log crack length distribution: $N\left(\mu_x, \sigma_x^2\right)$ with $\mu_x = -7.39$ and $\sigma_x^2 = 0.51$.
- Linear model for the log measurement response: $y_{ij}(t_{jk}) = \beta_0 + \beta_1 x_{ij}(t_{jk}) + \epsilon_{ijk}$ and $\epsilon_{ijk} \sim N\left(0, \sigma_y^2\right)$ with $\beta_0 = 4.50$, $\beta_1 = 0.50$ and $\sigma_y^2 = 0.065$.
- Log random crack growth rate distribution: $N\left(\mu_\lambda, \sigma_\lambda^2\right)$ with $\mu_\lambda = -0.35$ and $\sigma_\lambda^2 = 4.0 \times 10^{-4}$. 
Log noise response distribution: \( N(\mu_{\text{noise}}, \sigma_{\text{noise}}^2) \) with \( \mu_{\text{noise}} = 2.50 \) and \( \sigma_{\text{noise}}^2 = 0.064 \).

We assume there are \( I = 5 \) particular rivet holes under study in each aircraft (i.e., group) and \( J = 20 \) aircraft in the fleet (actual numbers could be expected to be much larger, but we use these smaller numbers for our illustrative example so that we can present informative plots using all of our data). All rivet holes from the same aircraft have the same crack growth rate which is sampled from the random crack growth distribution. We further assume there are \( K = 9 \) scheduled inspections for service at nominal times 1000, 2000, ..., and 9000 operating hours. The actual inspection time for each aircraft at each nominal inspection time is determined by the sum of the nominal scheduled inspection time and a random number generated from a uniform \((-50,50)\) distribution to account for inspection-scheduling variability. The simulated data (including parts of the data that are not observable) are shown in Figure 6-1.

### 6.5 Statistical Model

When measurements are available on the noise response, signal response, and crack length, a bivariate normal distribution can be used to model the data. In this section we illustrate the details of our joint bivariate normal statistical model.

In order to explain our estimation procedure, we separated the rivet holes into two categories, based on whether a crack is eventually found or not in the sequence of scheduled inspections. The first category includes rivet holes for which a crack existence decision was made at one of the scheduled inspections. For these rivet holes, both the NDE measurements
and crack length measurements are available at the “crack find” inspection. The second category includes rivet holes that have a measurement below the detection threshold at every scheduled inspection. No crack-length information is available for these rivet holes. For both categories, the actual measurement results are described by the noise interference model (6-4) with the signal response from (6-2) and noise response from (6-3). The relationship between the signal response and crack length follows from (6-2) and we assume that the noise response is independent of crack length.

6.5.1 Rivet Holes with a “Crack Find” Inspection

6.5.1.1 Statistical model for the “crack find” inspection data

Some locations within an aircraft will eventually have a crack-find event. Suppose the measurement for rivet hole $i$ in aircraft $j$ is above the detection threshold at the scheduled inspection $\kappa(i, j)$ ($1 \leq \kappa(i, j) \leq K$) with service time $t_{jk(i,j)}$. The log crack length

$$x_{ij}(t_{jk(i,j)}) = x_0(t_0) + \lambda_j(t_{jk(i,j)} - t_0)$$

is measured at the time of repair where $x_{ij}(t_0)$ is from the $N(\mu_x, \sigma_x^2)$ log initial crack length distribution and $\log_{10}(\lambda_j)$ is from the log crack growth rate distribution $N(\mu_\lambda, \sigma_\lambda^2)$. We assume that measurement error is negligible, although this would be easy to generalize if the measurement error has a known distribution.

Thus from the result of (6-2) and the bivariate normal distribution properties given in (A-1), and (A-2) from the appendix, the conditional distribution of the signal response for a
rivet hole of a given crack length \( x_{ij}(t_{jk(i,j)}) \) at the “crack find” inspection can be modeled with a random variable \( Y(t_{jk(i,j)}) \) through the normal distribution:

\[
Y(t_{jk(i,j)}) \sim N\left(\beta_0 + \beta_1 \left(\mu_x + \lambda_j \left(t_{jk(i,j)} - t_0\right)\right), \sigma_y^2\right).
\]  \tag{6-5}

The eddy current NDE signal response for rivet hole \( i \) in aircraft \( j \) at scheduled inspection \( \kappa(i,j) \) is modeled with the random variable \( Y_{\text{actual}}(t_{jk(i,j)}) \) through the noise interference model:

\[
Y_{\text{actual}}(t_{jk(i,j)}) = \max\left(Y(t_{jk(i,j)}), Y_{\text{noise}}(t_{jk(i,j)})\right)
\]  \tag{6-6}

where \( Y(t_{jk(i,j)}) \) is defined in (6-5) and \( Y_{\text{noise}}(t_{jk(i,j)}) \) is defined in (6-3).

### 6.5.1.2 Statistical model before the “crack find” inspection

The crack length at any scheduled inspections before a “crack find” inspection is

\[
x_{ij}(t_{jk}) = x_{ij}(t_{jk(i,j)}) + \lambda_j \left(t_{jk} - t_{jk(i,j)}\right) \quad \text{for} \quad k = 1,...,\kappa(i,j).
\]

The log signal responses for these scheduled inspections are \( Y_{ij}(t_{jk}) = \beta_0 + \beta_1 \left[x_{ij}(t_{jk(i,j)}) + \lambda_j \left(t_{jk} - t_{jk(i,j)}\right)\right] + \epsilon_{ijk} \) and are modeled by the normal distribution:

\[
Y(t_{jk}) \sim N\left(\beta_0 + \beta_1 \left[x_{ij}(t_{jk(i,j)}) + \lambda_j \left(t_{jk} - t_{jk(i,j)}\right)\right], \sigma_y^2\right).
\]  \tag{6-7}

The actual measurement result \( y_{\text{actual,ij}}(t_{jk}) \) for rivet hole \( i \) in aircraft \( j \) at scheduled inspection \( k \) with \( k = 1,...,\kappa(i,j) \) can be modeled through a random variable \( Y_{\text{actual,ij}}(t_{jk}) \) in the noise interference model.
\[ Y_{\text{actual}}(t_{jk}) = \max \left( Y(t_{jk}), Y_{\text{noise}}(t_{jk}) \right) \]  

(6-8)

where \( Y(t_{jk}) \) is defined in (6-7) and \( Y_{\text{noise}}(t_{jk(i,j)}) \) is defined in (6-3).

### 6.5.2 Rivet Holes without a “Crack Find” Inspection

The inspection results for rivet holes in the second category are below the detection threshold at all scheduled inspections. Thus no direct crack-length information is available for these rivet holes. What we know about these cracks is that they follow the crack growth model that says that the crack length at each scheduled inspection is

\[ x_0(t_{jk}) = x_0(t_0) + \lambda_j(t_{jk} - t_0) \quad \text{for} \quad k = 1, ..., K \]

where \( x_0(t_0) \) is the unknown initial log crack length having a normal distribution \( N(\mu_x, \sigma_x^2) \). The log signal response at each scheduled inspection is then modeled by a normal random variable

\[ Y(t_{jk}) \sim N(\beta_0 + \beta_1 \left[ x_0(t_0) + \lambda_j(t_{jk} - t_0) \right], \sigma_y^2). \]  

(6-9)

The actual eddy current log response is thus modeled with the noise interference model random variable

\[ Y_{\text{actual}}(t_{jk}) = \max \left( Y(t_{jk}), Y_{\text{noise}}(t_{jk}) \right) \]  

(6-10)

where \( Y(t_{jk}) \) is defined in (6-9) and \( Y_{\text{noise}}(t_{jk(i,j)}) \) is defined in (6-3).

### 6.6 Bayesian Estimation

Likelihood based methods could be developed to estimate the model parameters \( (\mu_x, \sigma_x^2, \beta_0, \beta_1, \sigma_y^2, \mu_\lambda, \sigma_\lambda^2) \) with the likelihood contributions corresponding to the two
different types of observations described in Section 6-5. Computation of the likelihood for this model is, however, difficult because of the multiple-fold integrals needed to represent the random effects. Bayesian methods (e.g., Gelman, Carlin, Stern and Rubin 2003), which are closely related to likelihood methods, provide an easy-to-use and versatile alternative approach to do the estimation for the field data from the proposed rivet-hole inspection procedures. Bayesian methods also provide a formal way to incorporate useful prior information such as physics-based theory, information from previous studies, or expert opinion into the statistical analysis. In our analysis, however, we use diffuse (approximately non-informative) prior distributions. We have used WinBUGs (2007) to do the Bayesian analysis.

6.6.1 Model Specification

We use the statistical model described at Section 6-5 and diffuse prior distributions. The WinBUGs Markov Chain Monte Carlo (MCMC) algorithm is used to generate a large number of sampling draws from the joint posterior distribution for the model parameters

\[
\left( \mu_{\text{noise}}, \sigma_{\text{noise}}^2, \mu_x, \sigma_x^2, \beta_0, \beta_1, \sigma_y^2, \mu_y, \sigma_y^2 \right).
\]

After the MCMC algorithm has converged, we have \(M\) sampling draws for each model parameter from the joint posterior distribution. Based on the \(M\) sampling draws, we can calculate statistics of interest such as the mean, standard deviation, and the 5% and 95% posterior quantiles for each model parameter or functions of the model parameters (e.g., POD and PFA). The 5% and 95% posterior quantiles also determine the 90% credible bounds for the model parameter. Summary results for all model parameters, comparing estimates with the true parameters used in the simulation are given in Table 6-1.
Table 6-1. Posterior mean and standard derivation for all model parameters

<table>
<thead>
<tr>
<th>Model Parameter</th>
<th>$\mu_x$</th>
<th>$\sigma_x^2$</th>
<th>$\beta_0$</th>
<th>$\beta_1$</th>
<th>$\sigma_y^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>True Value</td>
<td>-7.39</td>
<td>0.51</td>
<td>4.50</td>
<td>0.50</td>
<td>0.065</td>
</tr>
<tr>
<td>Posterior Mean</td>
<td>-7.19</td>
<td>0.46</td>
<td>4.68</td>
<td>0.55</td>
<td>0.057</td>
</tr>
<tr>
<td>95% Credible Bounds</td>
<td>(-7.65,-6.76)</td>
<td>(0.30,0.68)</td>
<td>(4.31,5.11)</td>
<td>(0.44,0.67)</td>
<td>(0.037,0.079)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Model Parameter</th>
<th>$\mu_\lambda$</th>
<th>$\sigma_\lambda^2$</th>
<th>$\mu_{\text{noise}}$</th>
<th>$\sigma_{\text{noise}}^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>True Value</td>
<td>-0.35</td>
<td>0.00040</td>
<td>2.50</td>
<td>0.064</td>
</tr>
<tr>
<td>Posterior Mean</td>
<td>-0.38</td>
<td>0.00145</td>
<td>2.51</td>
<td>0.069</td>
</tr>
<tr>
<td>95% Credible Bounds</td>
<td>(-0.44,-0.32)</td>
<td>(0.00035,0.00409)</td>
<td>(2.49,2.53)</td>
<td>(0.061,0.078)</td>
</tr>
</tbody>
</table>

6.6.2 Estimate of the Response Function and POD

The mean log signal response function $\mu(x)$ for rivet holes with log crack length $x$ can be expressed as:

$$\mu(x) = \beta_0 + \beta_1 x.$$  \hspace{1cm} (6-11)

The noise interference model POD, as described by Li and Meeker (2009), as a function of crack length, is

$$\text{POD}(x) = 1 - \Phi \left( \frac{y_{th} - \mu(x)}{\sigma_y} \right) \Phi \left( \frac{y_{th} - \mu_{\text{noise}}}{\sigma_{\text{noise}}} \right).$$  \hspace{1cm} (6-12)
By substituting the $M$ sampling draws of $\beta_0, \beta_1, \sigma_y^2, \mu_{\text{noise}}$ and $\sigma_{\text{noise}}^2$ into (6-11) and (6-12), we can get the $M$ sampling draws of $\mu(x)$ and $\text{POD}(x)$ respectively, for any specified log crack length $x$. The posterior mean, standard deviation and 90% credible bound for $\mu(x)$ and $\text{POD}(x)$ can be obtained through their respective $M$ sampling draws. The estimated relationship between the posterior mean signal response and crack length and corresponding pointwise two-sided 90% credible bounds are shown in Figure 6-2 (left) along with the detection threshold (horizontal dashed line) and the posterior noise mean (horizontal dotted line). The posterior mean POD estimate, as a function of crack length, and corresponding 95% lower credible bounds are shown in Figure 6-2 (right). The estimated asymptotic POD as crack size approaches zero is 0.031 and it is close to the actual (unobserved) PFA of 0.028.

Figure 6-2. Posterior mean signal response and two-sided pointwise 90% credible bounds as a function of crack length (left) with detection threshold (horizontal dashed line) and posterior noise mean (horizontal dotted line), and posterior mean POD and 95% lower credible bound as functions of crack length (right).
6.6.3 Estimate of the Length Distribution and the Growth Model

One of the main advantages of our proposed inspection procedures is that it provides the information needed to estimate the noise distribution, the crack growth rates and the crack-length distribution at any point in time. In our model, the log noise response distribution has the normal distribution $N\left(\mu_{\text{noise}}, \sigma_{\text{noise}}^2\right)$. The initial log crack length has a normal distribution of $N\left(\mu_x, \sigma_x^2\right)$ and the log crack growth rates have a normal distribution of $N\left(\mu_\lambda, \sigma_\lambda^2\right)$.

![Lognormal Probability Scale](image)

Figure 6-3. Crack length distribution at last scheduled inspection (9000 hours in service).

Given the measurement data from the proposed inspection procedures, we can accurately estimate $\mu_{\text{noise}}, \sigma_{\text{noise}}^2, \mu_x, \sigma_x^2, \mu_\lambda$ and $\sigma_\lambda^2$ as shown at Table 6-1. With the estimates
of the crack-length distribution and crack growth rates, we can also predict the expected number of rivet holes needed to be replaced for a future inspection. The estimated log crack-length distribution at the last scheduled inspection in our data set (i.e., 9000 hours in service) and its 90% credible bounds are shown in Figure 6-3. Such information provides not only guide lines for spare parts inventory but also a criterion to detect any unusual behaviors (such as extreme larger numbers of rivet hole replacement at certain period of services), and in turn improves the overall aircraft reliability.

6.7 Concluding Remarks and Areas for Future Research

In this paper, we have proposed modified scheduled maintenance inspection procedures for crack detection in aircraft components through nondestructive evaluation techniques and show how to properly analyze the resulting data. We developed a joint estimation statistical method to model the data obtained from the procedures. We used the Bayesian analysis software WinBUGs to model jointly, crack growth rates, a crack length distribution, and the probability of detection. The proposed inspection procedures and the joint statistical analysis would provide much better understanding for the cracks inside aircraft components and improve the overall reliability assessment.

There are a number of extensions for the methodology presented in this article that suggest future research directions. These include the following:

- Crack growth rates vary within an aircraft, perhaps with several types of locations within an aircraft type. The hierarchical model used in this paper could be extended in a straight-forward manner to allow for this.
• It would also be possible to extend the hierarchical model in this paper to pool data across different types of aircraft.

• In our presentation we have assumed that, at the time of a detection event, crack length is measured precisely. Actually, when a crack is detected, crack-size information is obtained by drilling the rivet hole with successively larger drill bits until the crack can no longer be detected. Thus the crack-size observation is actually interval censored. Such interval-censored data can be easily accommodated in either a likelihood or a Bayesian estimation framework.

• It would be possible to use our approach to model NDE-signal/crack-growth data with a more complicated crack-growth model.

Acknowledgments

This material is based upon work supported by the Federal Aviation Administration under Contract #DTFA03-98-D-00008, Delivery Order #0058 and performed at Iowa State University's Center for Nondestructive Evaluation as part of the Center for Aviation Systems Reliability program.

Appendix Bivariate Normal Distribution

A.1 Density Function

The multivariate normal distribution is widely used to model the joint distribution of more than two random variables. The multivariate normal distribution has nice mathematical properties, described, for example, in Johnson and Wichern (2001). The bivariate normal is a
special case of multivariate normal with dimension two. For a random vector \((Y, X)^T\) following a bivariate normal distribution \(\text{BVN}(\mu, \Sigma)\), if we denote \(\mu = (\mu_1, \mu_2)^T\) as the mean vector and denote

\[
\Sigma = \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{pmatrix}
\]

as the variance-covariance matrix, then the density function for the random vector \((Y, X)^T\) is:

\[
f(y, x; \mu_1, \mu_2, \sigma_{11}, \sigma_{22}, \rho) = \frac{1}{2\pi \sqrt{\sigma_{11} \sigma_{22}} \sqrt{1 - \rho^2}} \times \\
\exp \left\{ -\frac{1}{2(1 - \rho^2)} \left[ \frac{(y - \mu_1)^2}{\sigma_{11}} - 2\rho \frac{(y - \mu_1)(x - \mu_2)}{\sqrt{\sigma_{11} \sigma_{22}}} + \frac{(x - \mu_2)^2}{\sigma_{22}} \right] \right\}
\]

(A-1)

where \(\rho = \frac{\sigma_{12}}{\sqrt{\sigma_{11} \sigma_{22}}}\) is the correlation between \(Y\) and \(X\). Given data in form of \((y, x)^T\) pairs, the estimate of parameters \((\mu_1, \mu_2, \sigma_{11}, \sigma_{12}, \sigma_{22})\) can be obtained through likelihood or Bayesian methods.

### A.2 Relationship to Linear Regression

An important property of the bivariate normal distribution used in this paper is that the distribution of one of the random variables, conditional on a fixed value of the other random variable, is a univariate normal distribution. For example, conditional on a fixed value of \(X = x\), the distribution of \(Y\) is normal with mean
\[ \mu_{Y|x=x} = \left( \mu_i - \frac{\sigma_{12}}{\sigma_{22}} \mu_x \right) + \left( \frac{\sigma_{12}}{\sigma_{22}} \right) x \] and variance \[ \sigma^2_{Y|x=x} = \sigma_{11} - \frac{\sigma_{12}^2}{\sigma_{22}} \] (see for example, Chapter 4 of Johnson and Wichern 2001).

Suppose we have observations in the form of \((y, x)^T\) pairs. We can model the relationship between \(Y\) and \(X\) with linear regression as \(Y = \beta_0 + \beta_1 x + \varepsilon\) with \(x\) the observation from \(X \sim N\left( \mu_x, \sigma^2 x \right)\) and \(\varepsilon \sim N\left( 0, \sigma^2 y \right)\). Thus traditional ML and linear regression methods can be applied to estimate the regression model parameters \((\mu_x, \sigma^2 x)\) and \((\beta_0, \beta_1, \sigma^2 y)\). Equivalently we can treat \((Y, X)^T\) as following a bivariate normal distribution \(BVN(\mu, \Sigma)\) of \((A-1)\) with parameters \((\mu_1, \mu_2, \sigma_{11}, \sigma_{12}, \sigma_{22})\). The relationship between the two sets of parameters is summarized as follows:

\[
\begin{align*}
\beta_0 &= \mu_i - \mu_2 \sigma_{12} / \sigma_{22} \\
\beta_1 &= \sigma_{12} / \sigma_{22} \\
\mu_x &= \mu_2 \\
\sigma^2_x &= \sigma^2_{22} \\
\sigma^2_y &= \sigma^2_{11} - \sigma_{12}^2 / \sigma_{22}
\end{align*}
\]

\[
\begin{align*}
\mu_i &= \beta_0 + \beta_1 \mu_x \\
\mu_2 &= \mu_x \\
\sigma_{11} &= \sigma^2_y + \beta_1^2 \sigma^2_x \\
\sigma_{12} &= \beta_1 \sigma^2_x \\
\sigma_{22} &= \sigma^2_x
\end{align*}
\]  \quad (A-2)

**References**


CHAPTER 7. GENERAL CONCLUSIONS

The existence of random measurement noise for most nondestructive evaluation (NDE) applications requires statistical methods to analyze NDE data. The Military Handbook 1823A described standard statistical methods for scalar NDE data analysis. However, special statistical methodologies are needed for complex NDE measurement responses and complicated data structures in real world applications. One of the main metrics in NDE analyses is probability of detection (POD). In this dissertation, we developed several new statistical methods for NDE applications and POD estimation. The overall introduction of these statistical methods is presented at Chapter 1.

Chapter 2 extended the standard scalar NDE data analysis with noise interference model (NIM) to describe the relationship between the signal response and noise response. We find that the NIM can be applied to the POD analysis to solve the small flaw size asymptotic problem presented at the traditional POD analysis. Because many NDE experiments involve signals that are potentially mixed with noise, the complementary risk NIM will provide a useful option for POD assessment. Scientists and engineers in other research areas where noise is involved may also find this approach to be helpful.

In Chapter 3, we applied the NIM to a large set of heat-increase scalar reduction of vibrothermography inspection data from metal specimens with two different materials at three different inspection locations. Despite the large difference in the experimental configurations at three locations, similar estimates of POD as a function of crack length for fixed values of dynamic stress were obtained for all locations. This is the first quantitative, multi-inspection-site demonstration of the reliability for vibrothermography method for
fatigue crack detection. We expect that these quantitative results will be useful for the future development of model-assisted POD analysis.

In Chapter 4, we developed a 3D matched filter to greatly increase the signal-to-noise ratio (SNR) of the vibrothermography sequence-of-images NDE inspection data. With the increased SNR, a noise threshold detection criterion using the largest contrast frame of image is used to define a crack detection criterion. With detection thresholds set to have the same probability of a false alarm, the SNR detection criterion based on the output of the matched filter has better overall detection performance when compared with the scalar heat-increase results described at Chapter 3.

In Chapter 5, we described the establishment and application of a statistical model for quantifying inspection capability and estimating POD, based on the physical mechanisms of an ultrasonic testing process. The physics-based statistical model enabled needed information extraction from data taken on the limited types and sizes of the synthetic inclusion targets in the synthetic inclusion titanium disk that was available for the experiment. The physics-based model further made possible the needed interpolation and extrapolation for a wider range of flaw sizes and nitrogen concentrations. The results from this study provide useful information about the ability to detect hard alpha inclusions in titanium forgings. The methodology provided here is, however, more general and could be used to study NDE inspection capability in other areas of application and for other kinds of inspection.

In Chapter 6, we proposed scheduled maintenance inspection procedures for crack detection in aircraft components through nondestructive evaluation techniques. We developed a joint estimation statistical method to model the data obtained from the procedures. We used the Bayesian analysis software WinBUGs to model jointly, crack
growth rates, a crack size length distribution, and the probability of detection. The proposed inspection procedures and the joint statistical analysis would provide much better understanding for the cracks inside aircraft components and improve the overall reliability assessment.
### APPENDIX A. WINBUGS CODE FOR CHAPTER 5

```winbugs
### This is the WinBUGs code for Chapter 5.
### This is the model for Conventional inspection data set.
### The model for Multizone is similar and thus not included here.

model {
  for (j in 1:66) {
    ID[j] ~ dnorm(0,tau.ID)      ### define ID random effect
  }

  for (k in 1:6) {
    OP[k] ~ dnorm(0,tau.OP)  ### define OP random effect
  }

  NF2 ~ dnorm(0,0.001)                 ### N weight concentration correction coef.

  wt.N3 <- (1.0 - NF2*0.03*0.03)*3
  wt.N7 <- (1.0 - NF2*0.17*0.17)*17

  at.N3 <- 3.42*wt.N3/(1+0.0242*wt.N3)
  at.N7 <- 3.42*wt.N7/(1+0.0242*wt.N7)

  Ref[1] <- 1.0
  Ref[2] <- abs(((4490.9+5.03*at.N3-0.01*at.N3*at.N3)*(6002.2+61.86*at.N3) -
                  4461*6175)/((4490.9+5.03*at.N3-.01*at.N3*at.N3)*(6002.2+61.86*at.N3) +
                  4461*6175))
  Ref[3] <- abs(((4490.9+5.03*at.N7-0.01*at.N7*at.N7)*(6002.2+61.86*at.N7) -
                  4461*6175)/((4490.9+5.03*at.N7-.01*at.N7*at.N7)*(6002.2+61.86*at.N7) +
                  4461*6175))

  ### Total of 381 data points, mu[i] is the physical model
  ### y[i] is the actual EFBH measurement,
  ### I(low[i],upp[i]) defines the lower and upper bounds for censored obs.
  for (i in 1:381) {
    mu[i] <- beta + ID[id.locator[i]] + OP[op.locator[i]]
    + 0.434294*log(1.5708*Ref[Rlocator[i]]*ww*ww*
                    (1.0-exp(-2*size[i]*size[i]/(ww*ww))))
    y[i] ~ dnorm(mu[i],tau0)I(low[i],upp[i])
  }
}
```
### Define prior distributions for model parameters
### Here diffuse priors are used
### \( \text{tao0, tau.ID, tau.OP} \) are the precision parameter, i.e. 1/variance

\[
\begin{align*}
\text{beta} & \sim \text{dnorm}(0, 0.0001) \\
\text{ww} & \sim \text{dnorm}(0, 0.0001) \\
\text{tau0} & \sim \text{dgamma}(0.001, 0.001) \\
\text{tau.ID} & \sim \text{dgamma}(0.001, 0.001) \\
\text{tau.OP} & \sim \text{dgamma}(0.001, 0.001)
\end{align*}
\]

\[
\begin{align*}
\text{sigma0} & \leftarrow \text{sqrt}(1/\text{tau0}) \\
\text{sigma.ID} & \leftarrow \text{sqrt}(1/\text{tau.ID}) \\
\text{sigma.OP} & \leftarrow \text{sqrt}(1/\text{tau.OP})
\end{align*}
\]

\[
\begin{align*}
\text{R.factor} & \leftarrow \text{pow}(10, \text{beta}) \\
\text{sigma.all} & \leftarrow \text{sqrt}(1/\text{tau0}+1/\text{tau.ID}+1/\text{tau.OP})
\end{align*}
\]

### WinBUGs code for Chapter 5 end here.
### This is the WinBUGs code for Chapter 6.

```winbugs
model {
    ### Define diffuse priors distribution for all model parameters.
    a0 ~ dnorm(0, 0.001)                     ### mean of initial crack size distribution
    tau.x ~ dgamma(0.001, 0.001)
    sigma.x <- 1/tau.x                      ### variance of initial crack size distribution

    for (j in 1:20) {
        lamda[j] ~ dnorm(mu.lamda, tau.lamda)  ### mean of crack growth rate distribution
    }
    mu.lamda ~ dnorm(0, 0.001)                ### mean of crack growth rate distribution
    tau.lamda ~ dgamma(0.001, 0.001)
    sigma.lamda <- 1/tau.lamda                ### variance of crack growth rate distribution

    for (j in 1:20) {
        lamda.10[j] <- pow(10, lamda[j])  ### crack growth rate follow a log normal
    }

    b0 ~ dnorm(0, 0.001)                     ### intercept of signal response function
    b1 ~ dnorm(0, 0.001)                     ### slope of signal response function
    tau.y ~ dgamma(0.001, 0.001)            ### slope of signal response function
    sigma.y <- 1/tau.y                      ### variance of signal response
    noise.mean ~ dnorm(0, 0.001)             ### mean of noise response
    noise.tau ~ dgamma(0.001, 0.001)        ### variance of noise response
    noise.sigma <- 1/noise.tau              ### variance of noise response

    ### NIM for specimens at "crack find" inspections
    ### The "zero" tricks is used to find likelihood contribution of max(y1,y2)
    for (iii in 1:53) {
        mu[iii] <- a0 + lamda.10[lamda.bvn[iii]]*t.bvn[iii]
        yx.bvn[iii,2] ~ dnorm(mu[iii], tau.x)
        aaa[iii] <- b0 + b1*yx.bvn[iii,2]
        signal[iii] <- (yx.bvn[iii,1]-aaa[iii])/sqrt(sigma.y)
        noise[iii] <- (yx.bvn[iii,1]-noise.mean)/sqrt(noise.sigma)

        zeros[iii] <- 0
        zeros[iii] ~ dpois(tmp[iii])
        tmp[iii] <- -log(1/sqrt(sigma.y)*exp(-0.5*pow(signal[iii],2))*phi(noise[iii]))
                     + 1/sqrt(noise.sigma)*exp(-0.5*pow(noise[iii],2))*phi(signal[iii]) + 10000
    }
}
```
### NIM for the “crack find” rivet holes before the “crack find” inspection

```r
for (M in 1:290) {
  mu.cat1.y[M] <- b0 + b1*(yx.bvn[bvn.index[M],2] +
  lamda.10[lamda.cat1.slr[M]]*(t.cat1.slr[M]-t.bvn[bvn.index[M]]))
  signal.cat1[M] <- (y.cat1.slr[M]-mu.cat1.y[M])/sqrt(sigma.y)
  noise.cat1[M] <- (y.cat1.slr[M]-noise.mean)/sqrt(noise.sigma)

  zeros.cat1[M] <- 0
  zeros.cat1[M] ~ dpois(tmp.cat1[M])
  tmp.cat1[M] <- -log( 1/sqrt(sigma.y)*exp(-0.5*pow(signal.cat1[M],2))
    *phi(noise.cat1[M])+1/sqrt(noise.sigma)*exp(-
    0.5*pow(noise.cat1[M],2)))+phi(signal.cat1[M]) )+10000
}
```

### NIM for the inspections for rivet hole without “crack find” inspections

```r
for (jjj in 1:47) {
  x.last[jjj] ~ dnorm(mu.last[jjj],tau.x)
  mu.last[jjj] <- a0 + lamda.10[lamda.cat2.main[jjj]]*t.last[jjj]
}
```

```r
for (M in 1:423) {
  mu.cat2.y[M] <- b0 + b1*(x.last[last.idx[M]] + lamda.10[lamda.cat2.slr[M]]
    *(t.cat2.slr[M]-t.last[last.idx[M]]))
  signal.cat2[M] <- (y.cat2.slr[M]-mu.cat2.y[M])/sqrt(sigma.y)
  noise.cat2[M] <- (y.cat2.slr[M]-noise.mean)/sqrt(noise.sigma)

  zeros.cat2[M] <- 0
  zeros.cat2[M] ~ dpois(tmp.cat2[M])
  tmp.cat2[M] <- -log( 1/sqrt(sigma.y)*exp(-0.5*pow(signal.cat2[M],2))
    *phi(noise.cat2[M])+1/sqrt(noise.sigma)*
    exp(-0.5*pow(noise.cat2[M],2)))*phi(signal.cat2[M]) )+10000
}
```

### WinBUGs code for Chapter 6 end here.