3-1-2019

Quantile POD for nondestructive evaluation with hit–miss data

Yew-Meng Koh
Hope College

William Q. Meeker
Iowa State University, wqmeeker@iastate.edu

Quantile POD for nondestructive evaluation with hit–miss data

Abstract
Probability of detection (POD) is commonly used to measure a nondestructive evaluation (NDE) inspection procedure's performance. Due to inherent variability in the inspection procedure caused by variability in factors such as crack morphology and operators, it is important, for some purposes, to model POD as a random function. Traditionally, inspection variabilities are pooled and an estimate of the mean POD (averaged over all sources of variability) is reported. In some applications it is important to know how poor typical inspections might be, and this question is answered by estimating a quantile of the POD distribution. This article shows how to fit and compare different models to repeated-measures hit--miss data with multiple inspections with different operators for each crack and shows how to estimate the mean POD as well as quantiles of the POD distribution for binary (hit--miss) NDE data. We also show how to compute credible intervals (quantifying uncertainty due to limited data) for these quantities using a Bayesian estimation approach. We use NDE for the detection of fatigue cracks as the motivating example, but the concepts apply more generally to other NDE applications areas.

Keywords
Bayesian estimation, binary regression, have cracks, will travel, quantile POD, random effects

Disciplines
Materials Science and Engineering | Statistics and Probability

Comments
This is an Accepted Manuscript of an article published by Taylor & Francis in *Research in Nondestructive Evaluation* on March 1, 2019, available online: http://www.tandfonline.com/10.1080/09349847.2017.1374493. Posted with permission.

Quantile POD for Nondestructive Evaluation with Hit-Miss Data

Yew-Meng Koh
Department of Mathematics
Hope College
Holland, Michigan 49423

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

Abstract

Probability of detection (POD) is commonly used to measure a nondestructive evaluation (NDE) inspection procedure's performance. Due to inherent variability in the inspection procedure caused by variability in factors such as crack morphology and operators, it is important, for some purposes, to model POD as a random function. Traditionally, inspection variabilities are pooled and an estimate of the mean POD (averaged over all sources of variability) is reported. In some applications it is important to know how poor typical inspections might be and this question is answered by estimating a quantile of the POD distribution. This paper shows how to fit and compare different models to repeated-measures hit-miss data with multiple inspections with different operators for each crack and shows how to estimate the mean POD as well as quantiles of the POD distribution for binary (hit-miss) NDE data. We also show how to compute credible intervals (quantifying uncertainty due to limited data) for these quantities using a Bayesian estimation approach. We use NDE for the detection of fatigue cracks as the motivating example, but the concepts apply more generally to other NDE applications areas.

Keywords: Bayesian estimation; Binary regression; Have Cracks, Will Travel; NDE; Random effects

1 Introduction

1.1 Background and Motivation

Probability of Detection (POD) is the most commonly used metric to characterize the performance of a nondestructive evaluation (NDE) inspection procedure. POD curves (i.e., POD as a function of crack size) provide important inputs for making accept/reject decisions, scheduling inspections, doing lifing calculations, and performing risk analyses. There are many sources of variability that can affect the outcome of an NDE inspection, such as operator-to-operator differences, probe-to-probe differences, calibration, and crack morphology. It has been widely recognized that the dominant source of variability in NDE is crack morphology (i.e., if there are many cracks that have the same length, there is typically a large amount of variability in the results of an NDE inspection). If, however, there are important differences in the performance of different operators, different operators will have different POD curves. Traditionally, the POD that has been reported (e.g., POD methods described in MIL-HDBK-1823A (2009)) has been what we call the "mean POD." That is, what has been reported is an estimate of the mean (i.e., the expected value of) POD over all sources of variability. Usually, along with this estimate, a set of pointwise lower 95% confidence bounds for POD, reflecting the uncertainty from limited data, are also reported. We have seen confusion between the lower confidence bound for mean POD and the concept of a quantile POD (i.e., a quantile of the POD distribution indicating how poor a particular inspection might be due to factors like operator assignment.) They are not the same and can be very different.

1.2 Related Work

As described in Li, Meeker, and Spencer (2012), Li, Meeker, and Thompson (2014), and Li, Spencer, and Meeker (2015), for some applications, particularly when safety is of concern, it would be preferable to use an alternative form of POD that recognizes that it is possible, on a given inspection, to have a POD curve that is considerably less than the POD averaged over random effects like operator. Li, Meeker, and Spencer (2012) and Li, Meeker, and Thompson (2014) introduced the concept of a quantile POD for quantitative signal-response NDE data (sometimes known as $\hat{a}$ vs a data). In this paper we show how to properly analyze binary (hit-miss) NDE data for which individual cracks are inspected multiple times by different operators. Hit-Miss data arise naturally in many NDE applications, such as fluorescent penetrant inspection (FPI) and in some applications where there is no natural scalar quantitative response and an expert evaluates an image (e.g., in X-ray inspection). In some applications, precise quantitative NDE data are obtained but then reduced to binary data (even though this practice generally sacrifices information and results in a loss of estimation precision). We use a random-effects model and show how to compute an estimate of a given quantile of the POD distribution.

1.3 Contributions

In this paper, we show how to

- Do an exploratory analysis (i.e., an analysis with minimal assumptions to better understand the nature of hit-miss data)
- Fit a proper statistical model to hit-miss data from a POD study employing more than one operator
- Estimate a quantile of the POD distribution from hit-miss inspection data.

1.4 Description of the Hit-Miss Data

We use hit-miss data that were obtained from the examination of 52 cracks of various sizes in an aircraft component by 98 different operators (called "technicians" in Lewis et al. 1978). This dataset is now commonly known as the "Have Cracks, Will Travel" dataset. Specifically, the subset of the data used in this paper are from Sample B, examined using eddy current surface scans (see Figures 5-1, 5-4 and 5-5 of Lewis et al. 1978). The cracks in Sample B were from lower-surface segments of a C-130 center wing box. More details about the structure of the inspected parts can be found on Page 3-9 of Lewis et al. (1978). An important issue to be noted about this dataset is that most, but not all, cracks were inspected by all inspectors. Thus, there is some unbalance in the data (see Section 3 for more details on this unbalance). However, with the capabilities of modern statistical modelling methods, this unbalance causes no serious problems in the data analysis and we are able to estimate the different components of variability and other unknown model parameters without difficulty.

Figure 1 shows the number of inspections in the dataset versus crack length on a logarithmic axis. Figure 2 shows the distribution of hits and misses recorded by the operators at the various crack lengths. In Figure 2, the width of the bars at each crack length is proportional to the number of inspections made at that crack length.

The report in Lewis et al. (1978) also describes separate binary regression models fit to the "Have Cracks, Will Travel" data that had been grouped according to operator age and training. Page 22 of Berens and Hovey (1981) shows various proposed functional forms for modelling POD. A logistic regression model with identically and independently distributed error terms was found to fit the "Have Cracks, Will Travel" data well. See equation (5) on page 24 and equation (24) on page 37 of Berens and Hovey (1981). More details about this dataset, as well as an analysis (different from ours), can be found in Singh (2000).

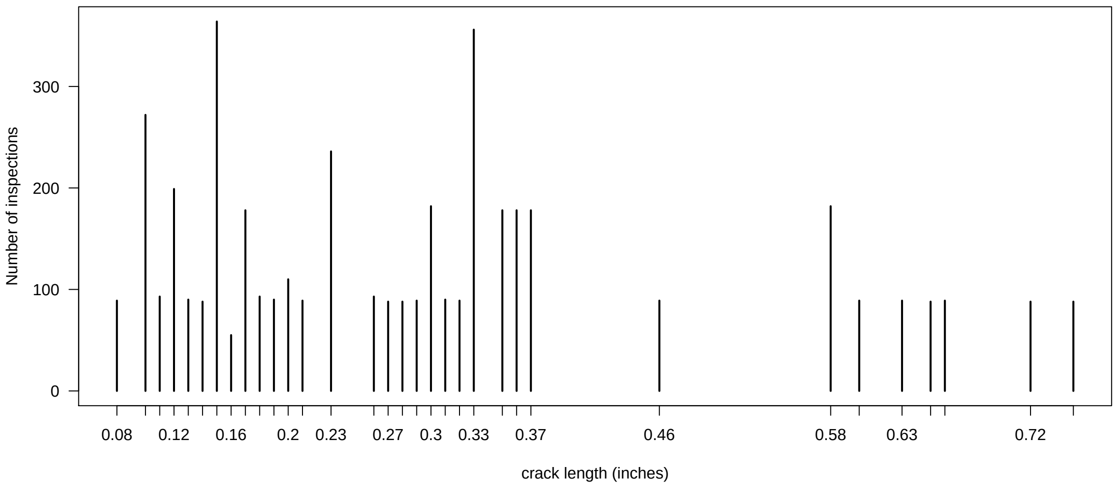

Figure 1: Number of inspections versus crack length in the dataset (on a log axis)

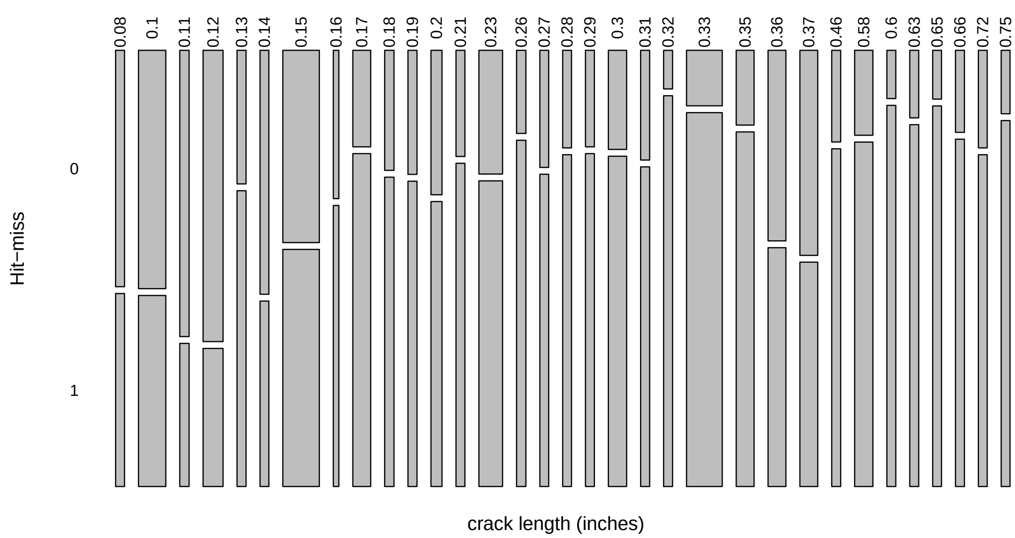

Figure 2: Proportion of hits (1) and misses (0) for various crack lengths in the dataset. The widths of the bars is proportional to the number of inspections made at that crack length

1.5 Overview

The remainder of this paper is organized as follows. Section 2 discusses the statistical models that have been used to describe hit-miss data. Section 3 is an exploratory analysis of the "Have Cracks, Will Travel" dataset that is used to provide modeling guidance. Section 4 introduces the model that we will use in this paper. Section 5 describes the Bayesian framework which we use to provide estimates for and do inference on the parameters in our statistical model. Section 6 shows how to estimate the POD for specific operators. The first part of Section 6 considers estimation and inference for mean POD, while the second part looks at estimation and inference for quantile POD. Finally, Section 7 presents a comparison between mean POD and quantile POD. Some technical details are given in the appendix.

2 Basic Binary Regression Model

Let Y denote a binary random variable, with

$$Y = \begin{cases} 1 & \text{for a hit} \\ 0 & \text{for a miss} \end{cases}.$$

We let $x = \log(\text{crack length})$. Then the basic binary regression model for POD is

$$\text{POD}(x) = \Pr(Y = 1; x) = \Phi(\beta_0 + \beta_1 x)$$

where $\Phi(\zeta)$ is generally taken to be the standard cumulative distribution function (cdf) of a location-scale distribution. For example, the commonly used logit binary regression model is

$$\Phi(\zeta) = \Phi_{\text{logis}}(\zeta) = \frac{\exp(\zeta)}{1 + \exp(\zeta)}$$

and the probit binary regression model is

$$\Phi(\zeta) = \Phi_{\text{nor}}(\zeta) = \int_{-\infty}^{\zeta} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{t^2}{2}\right) dt. \tag{1}$$

The binary linear regression model can also be expressed as a linear function of the explanatory variable x by

$$h[\text{POD}(x)] = \beta_0 + \beta_1 x$$

where the function $h(p) = \Phi^{-1}(p)$ is an inverse cdf (also known as the quantile function) of a standard location-scale distribution and, when used in modeling, is known in the statistical literature as a *link function*. In this paper, we will use the logit link function because it provides an adequate description of the data and is more commonly used in NDE applications. More details and a description of maximum likelihood methods of estimation of the model parameters for single-inspection data can be found in Appendix G, Section G.4.1, of MIL-HDBK-1823A (2009).

We note that the method and software for analyzing hit-miss data illustrated in Appendix G.3.6.1 and Appendix G.3.6.2 of MIL-HDBK-1823A (2009) can handle (in an ad-hoc manner) simple balanced repeated measures data where several operators all inspect each crack one time. The methods and software cannot, however, be used for unbalanced data, or to estimate quantile POD. Also, the coverage probabilities of resulting confidence intervals and procedures may not be close to the specified nominal confidence levels.

3 Exploratory Analysis

It is good statistical practice to explore initially the structure of available data with minimal assumptions and then use what is learned to build a model to describe the data. In this section we describe preliminary exploratory analyses of the subset of the "Have Cracks, Will Travel" dataset that we are using in this paper. There were 98 operators who performed an eddy current inspection on sample set B, which consisted of 52 crack-containing parts. There were 48 operators who inspected all of the 52 parts; one operator inspected 49 and another inspected 50 parts; 31 inspected 46 of the parts; the remaining 17 operators inspected between 6 and 45 of the parts. As part of our exploratory analysis, we analyzed separately the 91 inspections that had inspection results on 15 or more parts. In the rest of this section we explain the analyses done and present detailed results for three of the operators that roughly cover the range of results and illustrate the operator-to-operator variability.

3.1 Relationship between hit-miss binary regression data and mixed left and right censored data

Borrowing from ideas used in reliability data analysis (also used in survival analysis) we can use an alternative representation of the traditional binary regression data, providing alternative (equivalent) and new extended methods of analysis, based on left and right-censored data. In particular, suppose that for a given flaw morphology (i.e., aspect ratio and other characteristics other than length) that there is an *unobservable* critical crack length a_c and that there will be a Hit if the actual crack length $a \geq a_c$ and a Miss otherwise. That is, a hit can be viewed as a left-censored observation at a and a Miss can be viewed as a right-censored observation at a. We call such data "mixed left and right censored data." For a given parameterization of the POD model, this formulation has exactly the same likelihood function and thus provides the same POD estimates (for both parametric and nonparametric models). An advantage of this alternative formulation is that one can utilize widely available statistical software for reliability data analysis (e.g., JMP and MINITAB) to do hit-miss POD analysis, as will be illustrated in this section.

3.2 Nonparametric estimation of POD based on hit-miss data

A nonparametric estimate (i.e., an estimate of a cumulative distribution function or POD function in our application) is an estimate that does not use any assumptions about the shape of the distribution being estimated. Such an estimate can be obtained by using special statistical methods that use mixed left and right censored data (sometimes known by the more general name of "arbitrarily censored data"). Such methods were originally given in Peto (1973) and Turnbull (1976) and are also described in Chapter 3 of Meeker and Escobar (1998). With mixed left and right censored data the nonparametric estimate of POD is a non-decreasing function having steps, points, and often, gaps between these, determined by the mixture and location of the particular left and right censored observations. The plots in the first row of Figure 3 show nonparametric estimates of POD (step function and points) and 95% simultaneous confidence bands (dotted curves) on logistic probability paper on the left and lognormal probability paper on the right. The scales on these probability plots are drawn such that any line drawn on the plot would correspond to the particular probability distribution, as explained in detail in Chapter 6 of Meeker and Escobar (1998). It is useful to assess the departure that the nonparametric estimate has from a straight line. Also, if one can draw a straight line within the simultaneous confidence bands, then the data are consistent with the corresponding distribution. The confidence bands in this case are wide, due to the limited amount of data for the inspection (52 parts), and indicate that the data are consistent with both the loglogistic and lognormal distributions and that there is little information in the data to discriminate between these two distributions.

3.3 POD estimation based on fitting a log-location-scale distribution to the mixed left and right censored data

The plots in the second row of Figure 3, are similar to those in the first row showing the nonparametric estimates (points) and the *parametric* ML estimates of POD (lines) and 95% pointwise confidence intervals (dotted curves) for the logistic

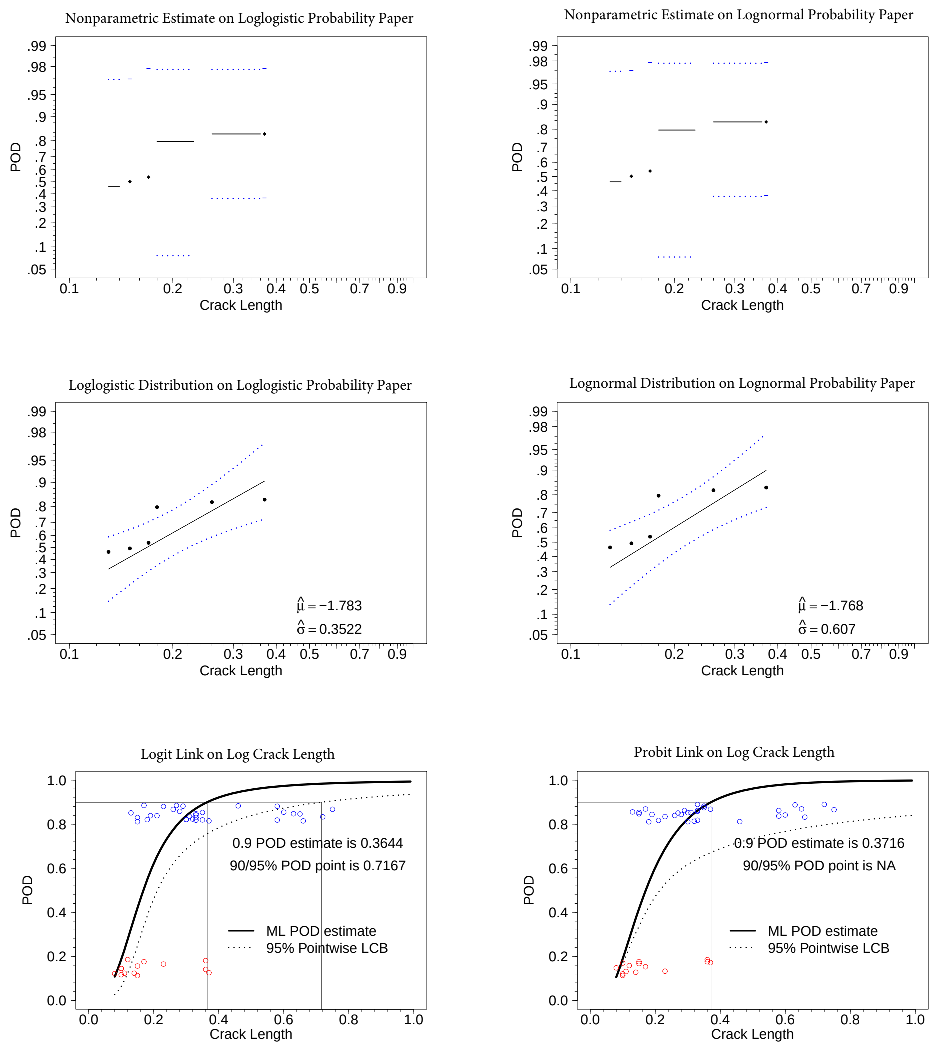

Figure 3: For operator 1301, plots of the nonparametric estimates of POD (step function and points) and 95% simultaneous confidence bands (dotted curves) on probability paper (top row); plot of the of the nonparametric estimates (points) and the parametric ML estimates of POD (lines) and 95% pointwise confidence intervals (dotted curves) on probability paper (second row); plot of the parametric ML estimates of POD (solid curves) with pointwise 95% lower confidence bounds (dotted curves) and jittered data (bottom row) for the logistic distribution (logit-link binary regression) on the left and lognormal distribution (probit-link binary regression) on the right.

distribution on logistic probability paper on the left and the lognormal distribution on lognormal probability paper on the right. The estimates of POD in both cases were computed from

$$\widehat{\mathrm{POD}}(a) = \Phi\left[\frac{\log(a) - \widehat{\mu}}{\widehat{\sigma}}\right]$$

where $\widehat{\mu}$ is the ML estimate of the location parameter and $\widehat{\sigma}$ is the ML estimate of the scale parameters of the underlying log-location-scale distribution (loglogistic and lognormal in our examples) and Φ is a cdf for the assumed location-scale distribution, as explained in Section 2. These ML estimates are shown on the plots in second row of Figure 3. These plots show that there is little difference between the POD estimates from these two distributions, within the range of the data.

3.4 POD estimation based on fitting a binary regression model to the Hit-Miss data

The plots in the third row of Figure 3 are based on fitting a binary regression model, using a generalized linear model approach with a log transformation on crack size and a logit (probit) link function, corresponding to fitting a loglogistic (lognormal) distribution to the mixed left and right censored data. The ML estimates of POD in these plots are exactly the same as the estimates shown in the plots in the second row of Figure 3. These plots also show the jittered hit-miss data (hits above and misses below) and the a_{90} estimates and the $a_{90/95}$ lower 95% confidence bounds for a_{90}, when they exist. In some cases neither exist.

For many (but not all) operators, the a_{90} value exists, but the $a_{90/95}$ values do not, due to a combination of poor POD and limited data, causing the lower bound on POD to be considerably less than the estimate of POD. For the operator 1301 data, the point estimates of POD for the two link functions are similar, but there is a large difference in the lower confidence bounds for POD, so much so that the $a_{90/95}$ does not exist for the probit-link model. Operator 1301 had inspection results typical of many of the 98 operators. There was, however, a considerable amount of variability across all operators. In the next section we present results for two other operators that illustrate estimates that were not so well behaved and that suggest the kind of operator-to-operator variability that existed in the Have Cracks, Will Travel experiment.

3.5 Other examples

Figure 4 is similar to Figure 3, giving results only for the loglogistic distribution (logit link function and log crack length) for operators 1806 on the left and 1118 on the right. Relative to operator 1301 (results in Figure 3), the results in the left-hand column of Figure 4 show that operator 1806 has larger POD for smaller cracks and smaller POD for larger cracks. This indicates, more specifically than operator-to-operator variability, an interaction between operator and crack length. That is, the *relative* ability of particular operators to detect cracks depends on the length of the crack.

The results for operator 1118, in the right-hand column of Figure 4 have results in the same direction, but even more extreme—so much so that the logistic regression estimate (binary regression based on the logit link and a log transformation on crack length) of POD in the bottom row is actually decreasing! The results using the reliability data analysis methods in the top two rows constrain the POD estimate to be non-decreasing, resulting in a flat estimate of POD as a function of crack length. This type of result occurred for 9 of the 96 operators who inspected at least 15 parts. Across all of the 96 sets of inspection data that we analyzed, we could see a few other POD curve shapes. For example, similar to Operator 1806, there were many POD curves that where relatively flat. Some of these were at high levels of POD and others were at lower levels. Most were increasing, but as mentioned above, 9 were decreasing. There were a few POD curves where POD rose very rapidly toward 1 because there were no misses for larger crack lengths.

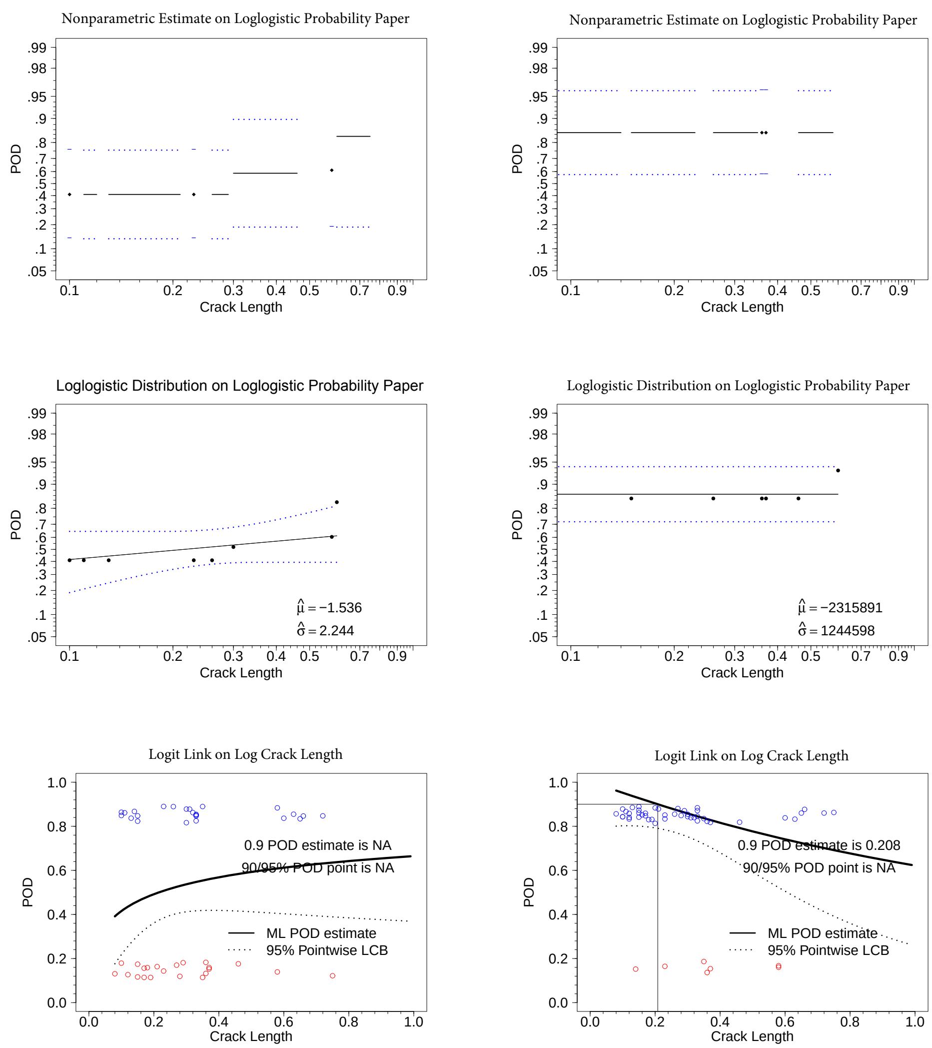

Figure 4: Plots of the nonparametric estimates of POD (step function and points) and 95% simultaneous confidence bands (dotted curves) on loglogistic probability paper (top row); plot of the nonparametric estimates (points) and loglogistic distribution ML estimates of POD (lines) and 95% pointwise confidence intervals (dotted curves) on loglogistic probability paper (second row); plots of the logit-link binary regression ML estimates of POD (solid curves) with 95% lower confidence bounds (dotted curves) and jittered data (bottom row) for operator 1806 on the left and operator 1118 on the right.

4 Binary Regression Model with Random Effects

We note here that in this application, x is taken as the standardized logarithm of the crack length, which is defined by

$$\text{Standardized crack length} = \frac{\text{log(crack length) - sample mean log(crack lengths)}}{\text{sample standard deviation of log(crack lengths)}}.$$

This standardized log crack length was used to improve the numerical stability in the Bayesian analysis which was used. In this paper, we follow the usual convention used in most of the data analysis literature and use natural (base e) logarithms. Neither of these conventions has an effect on the final estimates of POD, but will affect the values of some of the parameter estimates.

For standardized log(crack length) x, $Y(x)$ is a binary random variable. We note that inspections were carried out:

- Across different cracks (which, even if they are of the same size, will have different response distributions due to crack-to-crack variability),
- By different operators (e.g., with different skills and amounts of experience).

As we saw in Section 3, there was a substantial amount of operator-to-operator variability. POD curves estimates varied both in location and slope. As explained there, this implies that a model to describe operator effects should contain both a main effect and an interaction term between operator and crack size. We index the crack sizes by i and the operators by j. Thus the probability of a hit for operator j (with effect γ_j), and an operator-cracksize *interaction* effect ρ_{ij} can be expressed as

$$\text{POD}(x_i) = \Pr(Y(x_i) = 1 | \gamma_j, \rho_{ij}) = \Phi_{\text{logis}}\left(\frac{x_i - \mu}{\sigma} + \gamma_j + \rho_{ij}\right), \tag{2}$$

from which we have

$$\Phi^{-1}_{\text{logis}}\left[\text{POD}(x_i)\right] = \frac{x_i - \mu}{\sigma} + \gamma_j + \rho_{ij}, \tag{3}$$

where

- $\Phi^{-1}_{\text{logis}}(p) = \log\left(p/(1-p)\right)$ is the inverse logit function,
- $i = 1, \ldots, n_i$ (n_i is the total number of crack sizes)
- $j = 1, \ldots, n_j$ (n_j is the total number of operators).

In this model, we note that γ_j and ρ_{ij} are random and we will assume that $\gamma_j \sim N(0, \sigma^2_\gamma)$ and $\rho_{ij} \sim N(0, \sigma^2_\rho)$ are mutually independent. Here, μ, σ, σ^2_γ and σ^2_ρ are the parameters that need to be estimated from the experimental data. The model in (3) is similar to that described in Li, Meeker, and Thompson (2014), except that the observed data are hit-miss instead of signal-response ($\hat{a}$ versus a) data.

5 Bayesian Motivation, Estimation, and Results

5.1 Motivation for Using Bayesian Estimation

Following the approach in Li, Meeker, and Thompson (2014), we use the Bayesian method for estimation for the following reasons:

- The Bayesian framework provides a straightforward procedure to estimate important functions of the parameters, such as mean POD and quantile POD and to obtain corresponding credible bounds (similar to confidence bounds used in non-Bayesian inference), even with random effects.

- When diffuse prior distributions are used, the Bayesian analysis yields results that are close to those that would be produced by a maximum likelihood based analysis. In particular, it has been shown that Bayesian credible intervals, when used with diffuse prior information and reasonable sample sizes, usually have good frequentist properties (i.e., 95% credible intervals will contain the true POD with approximate probability 0.95).
- Bayesian estimation can still be carried out with models that are more complicated than (3) (e.g. for models that are non-linear in the parameters and have random effects).
- When valid informative prior information is available, it can be integrated with the data to improve estimation precision.

Bayesian methods for statistical inference are described in numerous places, including Carlin and Louis (2008), Gelman, Carlin, Stern, and Rubin (2004) and Gilks, Richardson, and Spiegelhalter (1996). Modern Bayesian analysis is conducted by using a Markov-Chain Monte Carlo (MCMC) procedure. See Pages 353-356 of Bernardo and Smith (2000) for details. Software packages are available to do the needed computations. We used the popular OpenBUGS software (see OpenBUGS (2015)).

5.2 Bayes' Theorem

Let $\boldsymbol{\theta} = (\mu, \sigma, \sigma_\gamma^2, \sigma_\rho^2)'$ be the vector of the parameters in model (3). A likelihood-based analysis estimates these parameters by using the observed data and identifying the value of $\boldsymbol{\theta}$ that maximises the likelihood of the data. The likelihood is denoted by $f(\mathbf{y}|\boldsymbol{\theta})$. A Bayesian analysis of the data, however, *integrates* prior knowledge about parameters with the available data. This prior knowledge is specified by using prior probability distributions [described by a joint density function $\pi(\boldsymbol{\theta})$]. When there is little or no knowledge about the parameters, it is possible to use *diffuse* prior distributions that are relatively flat over that part of the parameter space that has non-negligible likelihood, providing results that are close to those obtained by using likelihood-based methods. In this paper, we use diffuse prior distributions.

The conditional distribution of the parameters, given the data, $\boldsymbol{\theta}|\mathbf{Y} = \mathbf{y}$, is called the *posterior distribution* of $\boldsymbol{\theta}$. According to Bayes' theorem, the posterior density is

$$g(\boldsymbol{\theta}|\mathbf{y}) \propto f(\mathbf{y}|\boldsymbol{\theta})\pi(\boldsymbol{\theta}), \tag{4}$$

where $f(\mathbf{y}|\boldsymbol{\theta})$ is the likelihood of the data $\mathbf{y}$ when the true parameter vector is $\boldsymbol{\theta}$. Bayes' theorem then allows us to update the information that we have about the parameter vector *after* observing the data. The combined information is reflected in the joint posterior distribution of the parameters. More details about this model and method of estimation are given in the Appendix Section 9.4.

5.3 Prior Distribution Specification

When there is little or no available prior information about the parameters in $\boldsymbol{\theta}$, we can use a diffuse joint prior distribution. By doing this, the joint posterior distribution will be approximately proportional to the likelihood and the influence of the prior specification will be small as long as there is a sufficient amount of data. As mentioned in Section 5.2,' we specify diffuse marginal prior distributions for each of the model parameters, listed in Table 1. For μ, we used a normal distribution with a large variance. For σ, we used a gamma prior distribution with a large variance. For σ_γ and σ_ρ (non-negative parameters), following the recommendation of Gelman (2006), we used a Uniform$(0, 100)$ prior distribution.

Parameter	Prior Distribution
μ	normal with mean 0 and variance 1000
σ	gamma with mean 1 and variance 1000
σ_γ	Uniform$(0, 100)$
σ_ρ	Uniform$(0, 100)$

Table 1: Prior distributions on the parameters.

5.4 The Joint Posterior Distribution and Bayesian Parameter Estimates

In the modern approach to Bayesian inference, one uses draws from the joint posterior distribution of the parameters and the random effects, usually obtained by using a Markov Chain Monte Carlo (MCMC) procedure. A sufficiently large number of these draws allows computation of parameter estimates, estimates of functions of parameters, and credible intervals (similar to confidence intervals obtained in non-Bayesian methods) for these quantities. To enhance the efficiency of obtaining draws from the posterior distribution, we used the *hierarchical centering* method suggested in Gelfand et al. (1995) and Browne et al. (2009).

Often, one wishes to estimate a *scalar function* of the parameters $g(\boldsymbol{\theta})$. In our application, $g(\boldsymbol{\theta})$ will be either a mean POD or a quantile of the POD distribution. For a large integer M (typically on the order of tens of thousands to keep Monte Carlo error to an acceptable level), we denote the $k = 1, ..., M$ draws from the joint posterior distribution of $\boldsymbol{\theta}$ by $\boldsymbol{\theta}_1^*, ..., \boldsymbol{\theta}_M^*$. For guidance on how to choose M, see Liu, Nordman, and Meeker (2016). Then, $\{g(\boldsymbol{\theta}_1^*),\ g(\boldsymbol{\theta}_2^*), \ldots, g(\boldsymbol{\theta}_M^*)\}$ would be a set of M draws from the posterior distribution of $g(\boldsymbol{\theta})$. To simplify notation here, we use $\{g_1^*, ..., g_M^*\}$ to represent these draws. An estimate of $g(\boldsymbol{\theta})$ can be taken as the median of the marginal posterior distribution for $g(\boldsymbol{\theta})$ and we estimate the median of the marginal posterior distribution for $g(\boldsymbol{\theta})$ by computing the median of the set of simulated values $\{g_1^*, ..., g_M^*\}$.

Similarly, the lower and upper limits respectively of a 95% *credible interval* for the scalar function $g(\boldsymbol{\theta})$ are the 0.025 and 0.975 quantiles of the ordered set of simulated values $\{g_1^*, ..., g_M^*\}$. As mentioned in Section 5.1, because this analysis uses diffuse prior distributions and has a reasonably large sample size, the 95% credible interval found here provides a good approximation to a non-Bayesian *frequentist* 95% confidence interval for the scalar function $g(\boldsymbol{\theta})$.

5.5 Model Selection and Comparison

First, we fit a Bayesian hierarchical model for the POD, where

$$\text{POD}(x) = \Phi_{\text{logis}}(\mu_i + \sigma_i x),$$

and the (μ_i, σ_i) values have a joint bivariate normal distribution for $i = 1, \ldots, n$ (n is the total number of operators who examined the cracks in Sample B). We will compare this Bayesian hierarchical model (which we will call Model 1) to

- The random effects model for the POD in (2) (we will call this Model 2)
- A *reduced* random effects model for the POD which has operator random effects, γ_j, but does *not* contain operator-flaw size random effects, ρ_{ij}, given by $\text{POD}(x) = \Phi_{\text{logis}}\left(\frac{x-\mu}{\sigma} + \gamma_j\right)$. We will call this Model 3.

For any model, we define the marginal likelihood of the data $\boldsymbol{y}$ by

$$f_k(\boldsymbol{y}) = \int f_k(\boldsymbol{y}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})d\boldsymbol{\theta}, \tag{5}$$

where k indexes a particular model. Here, $f_k(\boldsymbol{y}|\boldsymbol{\theta})$ is the likelihood function for the data $\boldsymbol{y}$ under model k.

In a Bayesian framework, model comparison can be performed by computing ratios known as Bayes Factors. For two competing models i and j, the Bayes Factor for comparing model i to model j is

$$BF_{ij} = \frac{f_i(\boldsymbol{y})}{f_j(\boldsymbol{y})}, \tag{6}$$

where $f_k(\boldsymbol{y})$ is the marginal likelihood defined by (5). Large values of the Bayes Factor would provide evidence *against* model j, relative to model i. As mentioned in Section 5.4, the Bayesian framework allows us to obtain draws from the joint posterior distributions of the parameters $\boldsymbol{\theta^{(k)}}$ in model k. We denote these draws by $\boldsymbol{\theta_1^{(k)}}, \boldsymbol{\theta_2^{(k)}}, \ldots, \boldsymbol{\theta_M^{(k)}}$. The Weak Law of Large Numbers ensures that for large M, $\Sigma_{l=1}^{M} f_k(\boldsymbol{y}|\boldsymbol{\theta_l^{(k)}})/M$ converges in probability to $f_k(\boldsymbol{y})$. Thus (6) can be approximated by

$$BF_{ij} = \frac{\Sigma_{l=1}^{M} f_i(\boldsymbol{y}|\boldsymbol{\theta_l^{(i)}})}{\Sigma_{l=1}^{M} f_j(\boldsymbol{y}|\boldsymbol{\theta_l^{(j)}})}. \tag{7}$$

Rules of thumb for choosing one model over another based on Bayes Factors are provided in Kass and Raftery (1995) and are summarized in Table 2. From the values of the Bayes Factors from comparing these three models (Table 3), we conclude that Model 2, the random effects model from (2) would be the best model for our hit-miss data. We will use this model in subsequent analyses.

Log Bayes Factor, $\log_{10}(BF_{ij})$	Strength of evidence against model j
< 0	None
0 to 0.5	Weak
0.5 to 1	Substantial
1 to 2	Strong
> 2	Decisive

Table 2: Rules of thumb for using Bayes Factors in model selection

	Log Bayes Factor, $\log_{10}(BF_{ij})$ values
Model 2 vs Model 1	119.9
Model 2 vs Model 3	153.7
Model 1 vs Model 3	33.8

Table 3: Values of Bayes Factors in model comparison

6 Inference for Mean POD Distributions

We note that for any randomly selected operator from the population of operators and standardized log(crack length) x,

$$\text{POD} = \Phi_{\text{logis}}\left(\frac{x-\mu}{\sigma} + \gamma + \rho\right),$$

is a random variable because γ and ρ are random. (The i and j subscripts for the ith crack size and jth operator respectively have been suppressed for brevity.) For most applications of POD, one would not be interested in specific crack-operator combinations (although we have seen applications where there was a desire for separate POD curves for individual operators). Instead, one would typically be interested in situations in which cracks are random draws from a population of cracks. Similarly, in some applications, we would not be interested in a particular operator and therefore regard the operators as random draws from a population of operators. We would like to make inferences about the mean POD (averaging over many inspections by different operators) or perhaps some quantile of the POD distribution. Usually, this would be a lower-tail quantile to indicate the chance of having an operator-crack combination that might result in a poor POD for a crack of a given length.

6.1 Estimating the Conditional Probability of Detection for a Given Operator, Averaged Over Cracks

Consider the random variable defined by averaging over the distribution of interaction effects, but leaving operator to be random. That is,

$$\begin{aligned} h(\gamma; x) &= \int_{-\infty}^{\infty} \text{POD}(x) f_\rho(\rho) d\rho \\ &= \int_{-\infty}^{\infty} \left[\Phi_{\text{logis}} \left(\frac{x-\mu}{\sigma} + \gamma + \rho \right) \right] f_\rho(\rho) d\rho, \end{aligned} \tag{8}$$

where $f_\rho(\rho)$ is the $N(0, \sigma_\rho^2)$ pdf. We call this random variable the *operator-random mean POD.* Thus, $h(\gamma; x)$ is the POD(x) *averaged over* the population of interaction effects and is a function of the operator random effect γ. For a particular operator with effect γ_o, with standardized log(crack length) x, we would like to estimate $h(\gamma_o)$, the conditional probability of detection for operator with effect γ_o, averaged over the population of interaction effects, i.e.

$$h(\gamma_o) = \int_{-\infty}^{\infty} \left[\Phi_{\text{logis}} \left(\frac{x-\mu}{\sigma} + \gamma + \rho \right) \right] f_\rho(\rho) d\rho. \tag{9}$$

We obtain a Bayesian estimate of $h(\gamma_o)$ at a particular size x as well as credible bounds for the POD at a specified size x by first computing the values of

$$h_k(\gamma_o)^* = \int_{-\infty}^{\infty} \left[\Phi_{\text{logis}} \left(\frac{x-\mu_k^*}{\sigma_k^*} + \gamma_{ok}^* + \rho \right) \right] f_\rho(\rho) d\rho$$

for $k = 1, ..., M$, where the μ_k^*, σ_k^*, and γ_{ok}^* values are sample draws from the marginal posterior distributions of μ, σ, δ and γ_o (for operator o) respectively. These M $h_k(\gamma_o)^*$ values are then sorted from smallest to largest and the median, 0.025, and 0.975 quantiles of the empirical distribution are obtained. These would be, respectively, the point estimate of the conditional POD at log(crack length) x, and the lower 97.5% and the upper 97.5% credible bounds for the conditional POD at log(crack length) x for this particular operator o. The 97.5% lower and 97.5% upper credible bounds together give a 95% two-sided credible interval.

We repeat these computations over a range of values of x to obtain the conditional POD curve for an operator with effect γ_o. Figure 5 displays both point estimates and credible intervals for particular operators 1806 and 16E2. These particular operators were chosen to illustrate the wide range of operator ability.

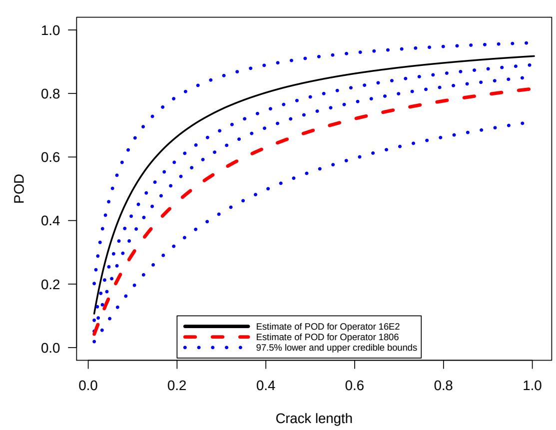

Figure 5: Point estimates and credible intervals for the conditional operator POD for operators 1806 and 16E2

6.2 Quantiles of the POD Distribution

Consider the random variable defined by averaging over the distribution of crack effects in (8). The p-quantile of the operator-random mean POD distribution for a given standardized log(crack length) x, denoted by $g_p(x)$, can be found by evaluating

$$g_p(x) = \int_{-\infty}^{\infty} \left[\Phi_{\text{logis}} \left(\frac{x-\mu}{\sigma} + \sigma_\gamma \Phi_{\text{nor}}^{-1}(p) + \rho \right) \right] f_\rho(\rho) d\rho, \tag{10}$$

where $\Phi_{\text{nor}}^{-1}(p)$ is the p-quantile of the $N(0,1)$ distribution. The derivation of this result is given in Appendix Section 9.3. Obtaining the Bayesian estimate and credible bounds for $g_p(x)$ is achieved by the same process described in Appendix Section 9.2 using the function $g_p(x)$ instead of $g(\mu, \sigma, \sigma_\gamma, \sigma_\rho, x)$. Again, repeating the computations over a range of values of x would give us estimates of the quantiles of the POD distribution (and credible bounds for these quantiles) as a function of standardized log(crack length) x. Figure 6 shows the logistic model Bayes' estimates of the 0.05, 0.10 and 0.50 quantiles of the POD distribution as a function of crack length for Sample A. For example, we estimate that in 5% of the inspections conducted in an inspection process similar to the "Have Cracks, Will Travel" study, the actual POD function would be smaller than the .05 quantile.

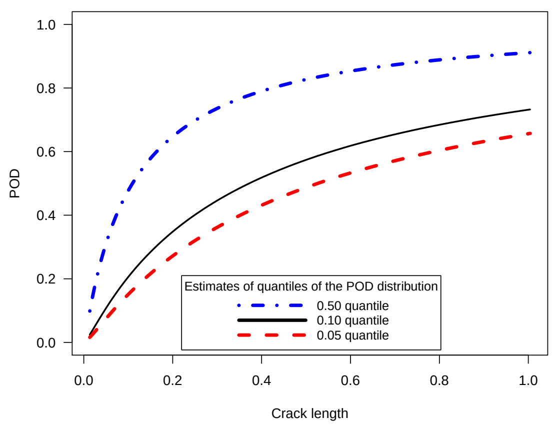

Figure 6: Logistic model point estimates of the 0.05, 0.10 and 0.50 quantiles of the POD distribution.

Figure 7 shows Bayes' estimates and lower and upper 97.5% credible bounds for the 0.10 quantile of the POD distributions as a function of crack length for Sample B. For example, we are approximately 95% confident that the 0.10 quantile of the POD distribution is within the interval (0.65, 0.75) when the crack length is 1 inch.

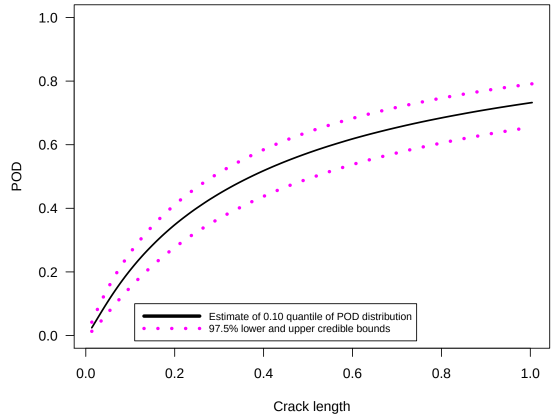

Figure 7: Bayes' estimates and 97.5% lower and 97.5% upper credible bounds for the 0.10 quantile of the POD distribution.

6.3 Mean of the POD Distribution

As shown in Appendix 9.1 and Appendix 9.2, the mean of the POD distribution (averaging over *all* of the random effects) for a given $x =$ standardized log(crack length) can be expressed as

$$g(\alpha, \beta, \sigma_\gamma, x) = 1 - \int_0^1 \Phi_{\text{nor}} \left(\frac{\Phi_{\text{logis}}^{-1}(s) - \frac{x-\mu}{\sigma}}{\sqrt{\sigma_\gamma^2 + \sigma_\rho^2}} \right) ds \tag{11}$$

where $\Phi_{\text{nor}}(\cdot)$ is the standard normal (Gaussian) cumulative distribution function defined in (1), used here due to the normal distribution assumption for the random effects. For given values of the parameters and standardized log(crack length) x, (11) can be evaluated using numerical integration.

To obtain a Bayesian estimate of the mean of the POD distribution at specified size x as well as credible bounds for the mean POD at size x, the values of

$$g_k^* = g\left[\mu_k^*, \sigma_k^*, \left(\sigma_\gamma^*\right)_k, \left(\sigma_\rho^*\right)_k, x\right]$$

are computed for $k = 1, ..., M$, where the μ_k^*, σ_k^*, $\left(\sigma_\gamma^*\right)_k$ and $\left(\sigma_\rho^*\right)_k$ values are draws from their respective posterior distributions. These g_k^* values are again ordered from smallest to largest and the median, 0.025, and 0.975 quantiles of the empirical distribution are obtained. These would be, respectively, the point estimate of the mean POD at standardized log(crack length) x, the lower 97.5% and the upper 97.5% credible bounds for the mean POD at standardized size x. Again, the 97.5% lower and 97.5% upper credible bounds together give a 95% two-sided credible interval. Figure 8 shows estimates and credible bounds for the mean POD for Sample B.

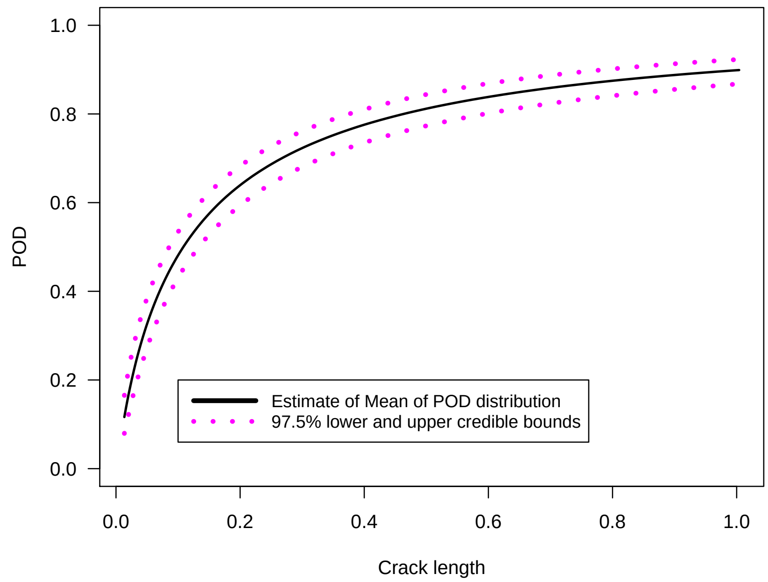

Figure 8: Mean of the POD distribution with credible bounds.

7 Conclusion

Some NDE applications report inspection data in binary (hit-miss) form. A common metric for an inspection procedure's performance is the POD. Different sources of variability in the inspection process suggest that POD(x) should be modeled as a random variable. That is, due to differences in factors such as operator and crack morphology, the POD for a crack of a given size may vary from inspection to inspection. For some applications, especially when safety is involved, treating POD as random from operator to operator might be more appropriate. We model the random POD using a generalised linear mixed model whose parameters can be estimated from the binary data. Standard methods to estimate POD summarise the resulting POD distribution by reporting an estimate of the POD mean (averaged over all sources of variability such as operator and crack morphology). In this paper, we show how to compute estimates of *quantiles* of the POD distribution. Both the mean and quantiles of the POD distribution are relatively complicated functions of the model parameters. For this reason, estimates and credible intervals are computed within a Bayesian framework with diffuse prior information. Figure 9 shows Bayes' estimates and credible limits for the mean of the POD distribution and the 0.10 quantile of the POD distribution as a function of crack length. We see considerable differences in these two summaries of the POD distribution. Each of these summaries answers different questions regarding the POD.

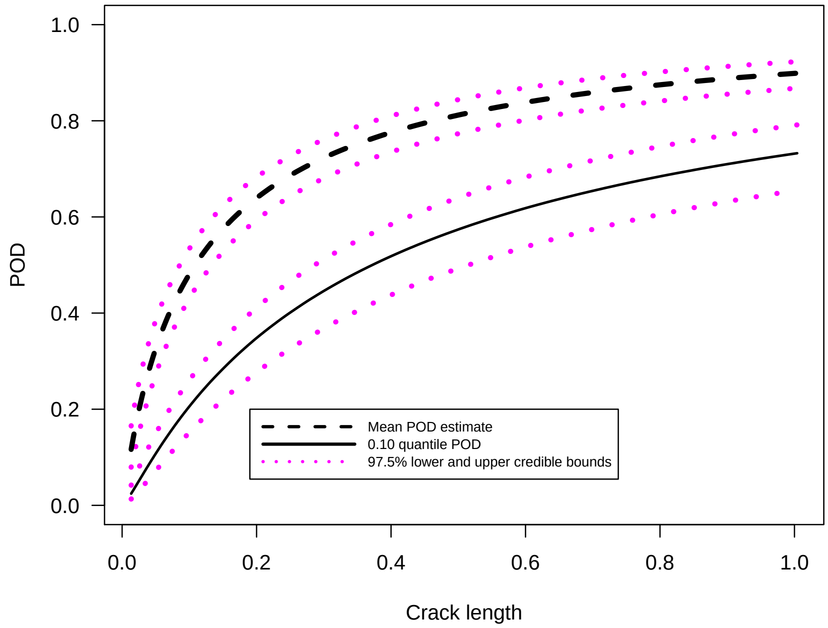

Figure 9: Bayes' estimates and credible limits for the mean of the POD distribution and the 0.10 quantile of the POD distribution

8 Acknowledgements

We would like to thank a referee who made numerous helpful comments on an earlier version of this article and who suggested the potential importance of using an interaction term in the model. These comments helped us to importantly improve the article. This material is based upon work that was partially supported by the Air Force Research Laboratory under Contract # FA8650-04-C-5228 at Iowa State University's Center for Nondestructive Evaluation.

9 Appendix

9.1 The POD distribution

As described in Section 4, the POD function is random due to the random effects γ and ρ. Thus, the POD at a given $x = \text{standardized}\log(\text{crack length})$ can be expressed as $S \equiv \text{POD}(x, \gamma, \rho)$. Because $0 \leq \text{POD} \leq 1$, we have

$$\Pr(S < s) = \begin{cases} 0 & s < 0 \\ 1 & s \geq 1 \end{cases}.$$

For $0 \leq s \leq 1$, we have

$$
\begin{aligned}
\Pr(\text{POD} < s) &= \Pr(\Phi_{\text{logis}}\left(\frac{x-\mu}{\sigma} + \gamma + \rho\right) < s) \\
&= \Pr\left(\left(\frac{x-\mu}{\sigma} + \gamma + \rho\right) < \Phi_{\text{logis}}^{-1}(s)\right) \\
&= \Pr\left(\frac{\gamma+\rho}{\sqrt{\sigma_\gamma^2+\sigma_\rho^2}} < \frac{\Phi_{\text{logis}}^{-1}(s) - \frac{x-\mu}{\sigma}}{\sqrt{\sigma_\gamma^2+\sigma_\rho^2}}\right) \\
&= \Phi_{\text{nor}}\left(\frac{\Phi_{\text{logis}}^{-1}(s) - \frac{x-\mu}{\sigma}}{\sqrt{\sigma_\gamma^2+\sigma_\rho^2}}\right).
\end{aligned}
\tag{12}
$$

Note that Φ_{logis} and Φ_{logis}^{-1} are the generic (e.g., for either the logit or probit models) standard cdf and standard quantile functions, respectively, corresponding to the binary regression model. The Φ_{nor} in (12) arises because of the linear combination of the normally distributed random effects γ and ρ. Under the assumption of mutual independence among the random effects γ and ρ,

$$
\frac{\gamma+\rho}{\sqrt{\sigma_\gamma^2+\sigma_\rho^2}} \sim N(0,1). \tag{13}
$$

Thus, from (12),

$$
\Pr(\text{POD} < s) = F_S(s) = \Phi_{\text{nor}}\left(\frac{\Phi_{\text{logis}}^{-1}(s) - \frac{x-\mu}{\sigma}}{\sqrt{\sigma_\gamma^2+\sigma_\rho^2}}\right) \tag{14}
$$

for $0 \leq s < 1$.

9.2 Derivation of the Mean of the POD Distribution

From the expression for the cdf of the POD in (14), the mean POD can be expressed as

$$
E(S) = \int_0^1 s\frac{d}{ds}F_S(s)ds = 1 - \int_0^1 F_S(s)ds
$$

$$
= 1 - \int_0^1 \Phi_{\text{nor}}\left(\frac{\Phi_{\text{logis}}^{-1}(s) - \frac{x-\mu}{\sigma}}{\sqrt{\sigma_\gamma^2+\sigma_\rho^2}}\right) ds = g(\mu, \sigma, \sigma_\gamma, \sigma_\rho, x).
$$

9.3 Derivation of p-quantile of Operator-Random Mean POD

From (8), we have the operator-random mean POD given by

$$
h(\gamma) = \int_{-\infty}^{\infty} \left[\Phi_{\text{logis}}\left(\frac{x-\mu}{\sigma} + \gamma + \rho\right)\right] f_\rho(\rho)d\rho.
$$

To obtain an expression for the p-quantile of the distribution of the random variable $h(\gamma)$, we proceed as follows. Using the fact that $\gamma \sim N(0, \sigma_\gamma^2)$, we know that

$$
\Pr\left(\frac{\gamma}{\sigma_\gamma} \leq \Phi_{\text{nor}}^{-1}(p)\right) = p,
$$

where $\Phi_{\text{nor}}^{-1}(p)$ is the p-quantile of the $N(0,1)$ distribution. Thus,

$$
\Pr(h(\gamma) \leq h(z_p\sigma_\gamma)) = p,
$$

because $h(\cdot)$ is a non-decreasing function. Thus, the $p-$quantile of the operator-random mean POD distribution is

$$g_p = h(z_p\sigma_\gamma) = \int_{-\infty}^{\infty} \left[\Phi_{\text{logis}}\left(\frac{x-\mu}{\sigma} + \sigma_\gamma \Phi_{\text{nor}}^{-1}(p) + \rho\right)\right] f_\rho(\rho) d\rho,$$

which is a function of the parameters μ, σ, σ_γ, and σ_ρ.

9.4 Likelihood and Posterior Distribution Definition

From (12), we have $\text{POD}(x) = \Phi_{\text{logis}}\left(\frac{x-\mu}{\sigma} + \gamma + \rho\right)$. We note that the hit-miss observation Y_i (conditional on γ_i and ρ_i) is Bernoulli($\text{POD}(x_i)$). Hence, the likelihood for observation i is

$$\Pr(Y_i = y_i | \gamma_i, \rho_i) = [\text{POD}(x_i)]^{y_i} [1 - \text{POD}(x_i)]^{1-y_i},$$

where $\text{POD}(x_i)$ is defined in (12). We also have that $\gamma_i \sim N(0, \sigma_\gamma^2)$ and $\rho_i \sim N(0, \sigma_\rho^2)$ are iid, with all these random variables being mutually independent. We write $\mathbf{y} = (y_1, ..., y_n)$ with similar definitions for the random vectors $\boldsymbol{\gamma}$ and $\boldsymbol{\rho}$. We then have

$$f(\mathbf{y}|\boldsymbol{\gamma}, \boldsymbol{\rho}) = \Pi_{i=1}^n \Pr(Y_i = y_i | \gamma_i, \rho_i).$$

The Bayesian method puts a joint prior distribution on the parameter vector $\boldsymbol{\theta} = (\mu, \sigma, \sigma_\gamma^2, \sigma_\rho^2)'$. The joint pror distribution is denoted by $\pi(\boldsymbol{\theta})$. It is important to note that $\boldsymbol{\theta}$ is not really random in nature but that the joint prior probability distribution expresses our knowledge about $\boldsymbol{\theta}$ *before* the data have been collected and observed.

From Bayes' theorem, the conditional distribution of $\mu, \sigma, \sigma_\gamma^2, \boldsymbol{\sigma_\rho^2}, \boldsymbol{\gamma}$, and $\boldsymbol{\rho}$ given the data $\mathbf{y}$ has a posterior distribution

$$g(\mu, \sigma, \sigma_\gamma^2, \boldsymbol{\gamma}, \boldsymbol{\rho} | \mathbf{y}) \propto f(\mathbf{y}|\boldsymbol{\gamma}, \boldsymbol{\rho}, \mu, \sigma) f(\boldsymbol{\gamma}|\sigma_\gamma^2) f(\boldsymbol{\rho}|\sigma_\rho^2) \pi(\boldsymbol{\theta}). \tag{15}$$

$g(\mu, \sigma, \sigma_\gamma^2, \sigma_\rho^2, \boldsymbol{\gamma}, \boldsymbol{\rho}|\mathbf{y})$ expresses our knowledge of $\boldsymbol{\theta}$, $\boldsymbol{\gamma}$, and $\boldsymbol{\rho}$ *after* the information in $\pi(\boldsymbol{\theta})$ and the observed data have been combined.

References

Berens, A. P. and P. W. Hovey (1981). Evaluation of NDE reliability characterization. *USAF Report No. AFWAL-TR-81-4160*. Available online at `http://www.dtic.mil/cgi-bin/GetTRDoc?AD=ADA114467`, accessed 5 Feb 2017.

Bernardo, J. M. and A. F. M. Smith (2000). *Bayesian Theory.* Chichester: Wiley.

Browne, W. J., F. Steele, M. Golalizadeh, and M. J. Green (2009). The use of simple reparameterizations to improve the efficiency of markov chain monte carlo estimation for multilevel models with applications to discrete time survival models. *Journal of the Royal Statistical Society 172*, 579–598.

Carlin, B. P. and T. A. Louis (2008). *Bayesian Methods for Data Analysis (Third Edition).* New York: Chapman & Hall.

Gelfand, A. E., S. K. Sahu, and B. P. Carlin (1995). Efficient parametrisations for normal linear mixed models. *Biometrika 82*, 479–488.

Gelman, A. (2006). Prior distributions for variance parameters in hierarchical models. *Bayesian Analysis 1*, 515–533.

Gelman, A., J. B. Carlin, H. S. Stern, and D. B. Rubin (2004). *Bayesian Data Analysis (Second Edition).* New York: Chapman & Hall.

Gilks, W. R., S. Richardson, and D. Spiegelhalter (1996). *Markov Chain Monte Carlo in Practice*. New York: Chapman and Hall.

Kass, R. E. and A. E. Raftery (1995). Bayes factors. *Journal of the American Statistical Association 90*, 773–795.

Lewis, W. H., B. D. Dodd, W. H. Sproat, and J. M. Hamilton (1978). Reliability of non-destructive inspections - final report. *USAF SA-ALC/MEE 76-6-38-1*. Available online at `http://www.dtic.mil/cgi-bin/GetTRDoc?Location=U2&doc=GetTRDoc.pdf&AD=ADA072097`, accessed 5 Feb 2017.

Li, M., W. Q. Meeker, and F. W. Spencer (2012). Distinguishing between uncertainty and variability in nondestructive evaluation. *Review of Progress in Quantitative Nondestructive Evaluation 31*, 1725–1732. D. O. Thompson and D. E. Chimenti (Eds) (American Institute of Physics, New York).

Li, M., W. Q. Meeker, and R. B. Thompson (2014). Physical model-assisted probability of detection of flaws in titanium forgings using ultrasonic nondestructive evaluation. *Technometrics 56*, 78–91.

Li, M., F. W. Spencer, and W. Q. Meeker (2015). Quantile POD: Distinguishing between uncertainty and variability in nondestructive evaluation. *Materials Evaluation 73*, 89–95.

Liu, J., D. J. Nordman, and W. Q. Meeker (2016). The number of MCMC iterations needed to compute bayesian credible intervals. *The American Statistician 70*, 275–284.

Meeker, W. Q. and L. A. Escobar (1998). *Statistical Methods for Reliability Data*. John Wiley & Sons.

MIL-HDBK-1823A (2009). Nondestructive evaluation system reliability assessment. Available online at `http://www.statisticalengineering.com/mh1823/MIL-HDBK-1823A(2009).pdf`, accessed 5 Feb 2017.

OpenBUGS (2015). Available online at `http://www.openbugs.net/w/Downloads`, accessed 5 Feb 2017.

Peto, R. (1973). Experimental survival curves for interval-censored data. *Applied Statistics 22*, 86–91.

Singh, R. (2000). Three decades of NDI reliability assessment. *Report No. Karta-3510-99-01*. Available online at `https://www.cnde.iastate.edu/mapod/Reference%20Documents/Karta%20pod%201970-1999.pdf`, accessed 5 Feb 2017.

Turnbull, B. W. (1976). The empirical distribution function with arbitrary grouped, censored, and truncated data. *Journal of the Royal Statistical Society 38*, 290–295.