Enhancing Limited-Sample Probability of Detection Estimation Using Models and Advanced Regression Techniques

Qizheng Xia[1], John C. Aldrin[2,3], Qing Li[1,2*]

[1]Department of Industrial Engineering, Iowa State University, 3031 Black Engineering, Ames, 50014, IA, USA.
[2]Center for Nondestructive Evaluation Department of Industrial Engineering, Iowa State University, Applied Science Complex 2, 1915 Scholl Rd, Ames, 50014, IA, USA.
[3]Computational Tools, 4275 Chatham Ave, Gurnee, 60031, IL, USA.

*Corresponding author(s). E-mail(s): qlijane@iastate.edu;
Contributing authors: qxia25@iastate.edu;
aldrin@computationaltools.com;

Abstract

Probability of detection (POD) is a critical metric for assessing the effectiveness of nondestructive evaluation techniques. However, obtaining sufficient measurements for traditional empirical POD estimation is usually costly in terms of time, money, and resources. When data are limited, evaluations frequently fail to capture the true relationship between signal responses and crack sizes or to adequately reflect the variability inherent in testing procedures due to various influencing factors and uncertainties. Additionally, the assumptions underlying simple linear regression models are often violated in the context of insufficient sample sizes, resulting in inaccurate or imprecise POD estimates.

To address these challenges, this study explores advanced regression methods and their integration with physics models for limited-sample POD (LS-POD) estimation. Specifically, we investigate information-augmentation approaches, including physics-informed regression and Bayesian methods, which incorporate prior knowledge to better characterize the relationship and variability in testing procedures. We also adapt regression techniques such as Box-Cox transformation, robust regression, weighted linear regression, and bootstrapping to mitigate the effects of potential assumption violations. Furthermore, these techniques are integrated to simultaneously leverage prior information and address assumption

violations. To evaluate the performance of these methods under various LS-POD scenarios, we conducted extensive simulation studies using both synthetic and empirical data. The results provide a comparative analysis of these approaches against a baseline model.
Our findings demonstrate that incorporating prior knowledge or employing advanced regression techniques to address assumption violations significantly enhances LS-POD estimation. Additionally, combining information-augmentation methods with assumption-correction techniques further improves estimation accuracy and precision. These insights offer valuable guidance to practitioners, enabling the selection of appropriate LS-POD methods tailored to specific applications and data conditions.

Keywords: Bayesian statistics, Bootstrapping, Physics-informed methods, Robust regression, Weighted regression.

1 Introduction

Nondestructive evaluation (NDE) is a critical technology in modern engineering, aimed at assessing internal defects, physical properties, and the structural integrity of materials without causing damage or compromising performance. NDE plays a pivotal role in industries such as aerospace, petrochemical, and nuclear energy, ensuring the safety and durability of engineering structures.

In NDE, the probability of detection (POD) is a key statistical metric used to quantify the effectiveness of inspection procedures in detecting defects [1, 2]. POD evaluation can be classified as either hit/miss or continuous signal response [1, 2]. This paper focuses on continuous signal response, where the traditional approach involves modeling the relationship between the signal response ($\hat{a}$) and crack size (a) using linear regression [1]. The resulting POD curve represents the probability of detection as a function of crack size a, increasing monotonically. Additionally, key quantities such as a_{90} (the smallest crack size with 90% probability of detection) and $a_{90/95}$ (the upper 95% confidence bound of a_{90}.confidence) can be estimated. POD reflects the inherent variability in inspection outcomes, which can be influenced by factors such as crack morphology, operator skills, and equipment performance. To capture this variability, data collection for POD analysis requires sufficient sample sizes. According to the MIL-HDBK-1823A handbook [1], at least 30 flawed sites are needed for continuous POD analysis and 40 for hit/miss POD to obtain reliable estimates.

However, POD studies can be costly in time, money, and resources. For many new inspection applications, meeting the sample size requirements of MIL-HDBK-1823A is infeasible. The problem is that small sample sets in POD analysis often result in biased estimates and low precision, which can lead to unreliable decision-making, increased safety risks, and higher operational costs in practice. For instance, underestimating a_{90} due to insufficient samples could lead practitioners to overestimate detection capabilities, increasing the risk of missing large cracks for critical safety scenarios. In contrast, an overly large estimate of $a_{90/95}$ indicates low estimation precision, which could mandate more frequent inspections than necessary, thereby increasing costs. Limited samples produce these problems because they fail to capture the true relationship and variability of the NDE process, and can also violate the statistical assumptions underlying traditional POD regression methods. Accordingly, recent efforts have focused on developing advanced techniques to improve the efficiency and adequacy of POD assessments while minimizing the need for extensive empirical studies.

One solution to reduce the amount of test specimens and empirical data is to use physics models to augment the evaluation. Physics models can describe the true relationship between the signal response $\hat{a}$ and the crack size a, and can also capture the variability of the NDE process caused by different factors [3]. Hence, incorporating physics-based information can significantly improve the adequacy of POD estimation. Accordingly, the model-assisted probability of detection (MAPOD) evaluation approaches have been established to integrate physics information with empirical data [4, 5]. MAPOD approaches include the full model-assisted approach (FMA) [4] and the transfer function approach (TF).

FMA can use physics models to inform the functional form of the regression or generate large sets of synthetic data to investigate the effects of numerous factors on the variability of signals, hence reducing the number of samples required by empirical studies [6, 7]. FMA can also be combined with Bayesian methods to improve estimation further by leveraging prior experiments or knowledge. For instance, Aldrin et al. [8] utilized a physics-based model to determine the functional form of the regression, capturing the relationship between the response signal and variables such as crack length, crack depth, and liftoff variation in eddy current testing. This approach replaced the simple linear regression with the nonlinear regression, enhancing the accuracy of the estimations. Then, a hierarchical Bayesian method was employed to combine prior knowledge of regression coefficients from existing studies with experimental data. Other research on FMA include Aldrin et al. [9], Li et al. [10], Carboni et al. [11], Jenson et al. [12], Buethe et al. [13], Le Gratiet et al. [14], Du et al. [15], Sun et al. [16], Savli et al. [17], Lei et al. [18]. However, the effectiveness of these methods depends heavily on well-established knowledge of physics, a clear understanding of the factors influencing the signal response, or prior studies. For example, Bayesian approaches require suitable prior distributions for regression parameters to achieve adequate estimates. Without reliable physics knowledge to inform these priors, the posterior estimates may become biased and less precise, especially when sample sizes are limited.

The TF approach leverages a baseline POD estimation derived from prior empirical studies and transfers these results to a new test case with only a limited number of samples [3]. For example, Harding et al. [19] demonstrated the TF approach for POD estimation of cracks in wing structures by evaluating experimental data for three more economical test conditions and evaluating the transfer relationship. These experimental data include cracks and synthetic notches in specimens with simpler geometries, and synthetic notches in wing structures. Additional studies on the TF approach can be found in Demeyer et al. [20], Bode et al. [21], Rosell and Persson [22], Koshti [23, 24]. However, TF can only transfer the parameters of the regression model, such as coefficients and variance, or POD estimation like a_{90}. Still, TF cannot transfer the confidence interval estimation. Moreover, TF can only transfer data within a certain range but not extrapolate beyond that range.

Another solution for insufficient empirical studies is to augment data by leveraging unflawed data. These methods aim to generate a more conservative estimation of a_{90} to eliminate the risk of overestimation imposed by limited samples. For example, Koshti et al. [25] calculated the probability of false positives using unflawed data and the confidence interval of the probability of detection to determine a new threshold to make sure the probability of mistaking unflawed data as cracks is low. Koshti [26] validated this method using merit ratios, such as the net decision threshold-to-noise ratio which reflects the ability of a system to detect target cracks. Additional studies can be found in Koshti [27, 28]. However, this type of method requires knowledge of the distribution of unflawed data to generate realistic unflawed data. Moreover, although the approach provides more conservative results, it may underestimate the true capacity of NDE testing.

The MAPOD methods, the Bayesian approach, and the methods of leveraging unflawed data augment information to capture the true relationship and variability of the NDE process better. However, little research has been done to investigate the problem of limited samples violating statistical assumptions underlying traditional POD methods. The key assumptions of simple linear regression used in traditional POD analysis include: 1) the relationship between the response and the regressors should be linear, 2) the variance of the error term is a constant, and 3) the error is normally distributed. When these assumptions are violated, the regression analysis results in biased and unprecise estimations. Multiple statistical techniques have been developed to mitigate violations of assumptions for linear regression [29, 30]. For example, robust regression can reduce the influence of outliers on the estimation. However, these techniques have not been leveraged for LS-POD analysis. Accordingly, we propose to adapt these advanced regression methods for POD estimation under small samples.

To summarize, limited data cannot adequately capture the relationship and variability in the NDE procedure due to insufficient information. In addition, without enough empirical data, the statistical assumptions underlying traditional POD analysis are often violated, leading to low accuracy and precision in estimation. However, existing studies in LS-POD either fail to incorporate physics knowledge and prior information, or neglect the violation of assumptions. Moreover, there is no single technique that can address all the problems of a limited data set.

To address the existing gap, this study explores advanced regression methods and their integration with physics models. Specifically, we investigate information-augmentation approaches, including physics-informed regression and Bayesian methods, which incorporate prior knowledge to better characterize the relationship and variability in testing procedures. We also adapt regression techniques such as Box-Cox transformation, robust regression, weighted linear regression, and bootstrapping to mitigate the effects of potential assumption violations. Furthermore, these techniques are integrated to simultaneously leverage prior information and address assumption violations. To quantitatively evaluate the efficacy of our methodologies, the POD estimation is not only compared with the methods outlined in MIL-HDBK-1823A as prior studies did [11, 19, 20, 22, 31], but also with a presumed baseline derived from a sufficiently large data set. We use the standard POD estimation from a sufficiently large dataset as the baseline. Given that such a baseline might not be available in practice, we conducted extensive simulation studies using both synthetic and empirical data to evaluate and compare the performance of these methods against a presumed standard under various scenarios.

We provide some technical background here. For POD analysis of $\hat{a}$ vs. a, the simple linear regression model is described in MIL-HDBK-1823A [1]

$$\hat{a} = \beta_0 + \beta_1 a + \epsilon, \tag{1}$$

where error term follows $\epsilon \sim N(0, \sigma^2)$, and the coefficients are β_0 and β_1. Therefore, for a certain threshold a_{th}, the $POD(a_i)$ is the probability of a response signal larger than the threshold when the crack size is a_i. The function is:

$$POD(a_i) = P(\hat{a_i} > a_{th}) = P(\epsilon > a_{th} - (\beta_0 + \beta_1 a_i)) = \Phi(\frac{a_i - (\beta_0 - a_{th})/\beta_1}{\sigma/\beta_1}), \tag{2}$$

where $\Phi(\cdot)$ is the standard normal cumulative distribution function. The regression coefficients can be estimated by the least square approach. To calculate the 95% confidence interval, we can use the delta method [32].

The remainder of the paper is organized as follows. Our LS-POD methodologies including methodology description, model selection, POD calculations, and comparison metrics are described in Section 2. Synthetic-based and empirical-based studies are shown in Sections 3 to demonstrate the effectiveness of our work in LS-POD analysis. The paper provides the discussion and conclusion in Section 4.

2 Limited sample POD methodologies

In this section, we investigate multiple advanced regression methods and also explore combining them with physics-informed models to enhance the POD estimation given limited sample sizes. Our framework comprises five components. First, establish the baseline of POD estimation. As discussed in Section 1, a sufficiently large sample size is required to produce an accurate and precise POD estimation. Hence, the POD estimation for sufficiently large synthetic and empirical data sets can be used as the benchmark for comparison and validation. Second, generate limited samples and check assumptions. We select limited samples whose sample size does not meet the minimal requirement from the large data set, and diagnosis assumption violations for simple linear regression. Third, identify a proper regression technique tailored for a specific problem in the data, e.g., limited information, nonlinearity, outliers, and heteroscedasticity (the error term has non-constant variance). We explored a range of techniques. Fourth, conduct POD analysis for limited samples using these regression techniques to tackle assumption violations, and also combine regression techniques with information-augmentation methods. Lastly, compare and validate the results of these methods under various scenarios. We use several metrics to compare the performance of LS-POD estimation with the baseline. Figure 1 presents a diagram of the five components of our LS-POD framework. We will explain these components in detail in the following subsections.

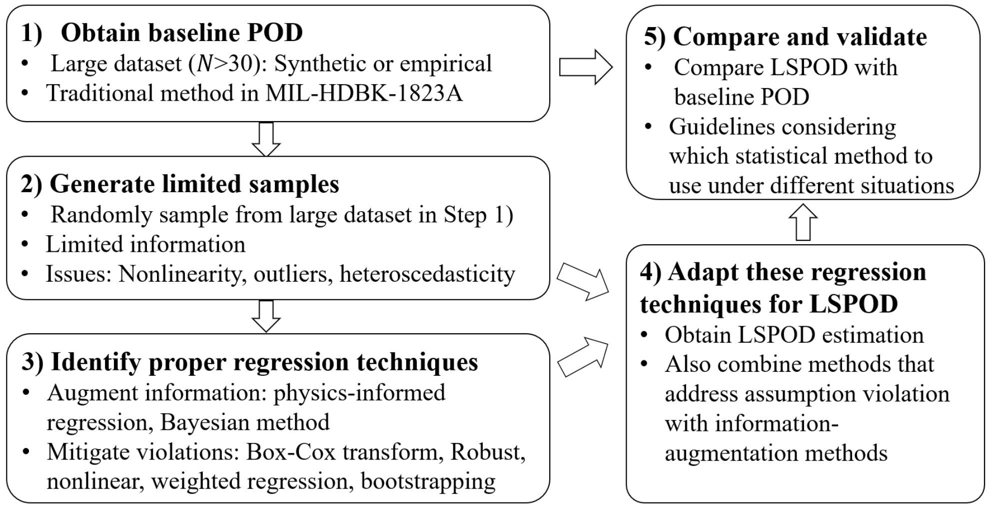


Fig. 1 Overview of the LS-POD framework in this study.

2.1 Generate limited samples and check assumptions

Simple linear regression requires the assumptions of linearity, homoscedasticity, and normality to ensure unbiased and precise estimates [33]. To represent limited samples, $n < 30$ observations will be sampled randomly with replacement from the baseline data. As discussed in Section 1, when the sample size is not large enough, the data often violates the assumptions for simple linear regression, leading to biased and imprecise estimates of parameters [34]. We explain the violation of these assumptions as follows.

- Nonlinearity. Nonlinearity refers to the fact that the relationship between an independent variable x and the dependent variable y in a regression model is not linear. Nonlinearity may also impact the normality and homoscedasticity of the regression model. To test the nonlinearity, the Ramsey regression equation specification error test (RESET) is one of the commonly used methods [35]. Ramsey's RESET adds nonlinear transformations of fitted values of y (e.g. $\hat{y}^2$, $\hat{y}^3$) as regressors to the original regression model and tests their significance. If the added regressors are statistically significant, it indicates the presence of nonlinearity in the regression model. A $p-value$ of less than 0.05 for the added regressors indicates significant nonlinearity, violating the linearity assumption.
- Outliers. Outliers refer to points in a dataset that deviate significantly from the majority of observed values, which may affect the accuracy and precision of the results. Outliers may influence the normality of the data and significantly change the estimation of the regression coefficients. To test for outliers, Cook's distance

is used to assess the overall influence of a single observation on the estimation of regression parameters [36, 37]. A relatively large Cook's distance value suggests that the observation has a substantial impact on the estimated coefficients and may be considered as an outlier. After identifying outliers, we need to investigate their causes. If an outlier is erroneous, such as due to instrument failure, improper operation, or data recording errors, we should remove the outlier. However, if an outlier results from normal fluctuations in the data and is completely valid, it should be appropriately accounted for in the analysis [38].

- Heteroscedasticity. Heteroscedasticity means the variance of the error term is not constant. In the presence of heteroscedasticity, ordinary linear regression estimation is still unbiased, but the estimator does not achieve the smallest variance, leading to less effective predictions [39]. To test heteroscedasticity, the White test is one of the most common methods for evaluating the correlation between the squared residuals and independent variables [40]. If the correlation is significant, the variance of the error term is not constant. Specifically, a $p - value$ less than 0.05 indicates that heteroscedasticity is present.

2.2 Identify proper regression techniques and conduct POD analysis

If some physics knowledge or prior studies are available, physics-informed regression or Bayesian methods can be used to integrate this information for LS-POD. When the assumptions are violated for a simple linear regression, several regression-type techniques can be adapted accordingly to enhance the adequacy of POD estimation.

2.2.1 Physics-informed regression

In NDE, the relationship between the signal response and the crack size can be nonlinear and complex. An opportunity exists when physics knowledge of the form $\hat{a} = f(a, \mathbf{X}'; \beta)$ is available. This knowledge describes the relationship between the signal response $\hat{a}$ and factors including the crack size a and other possible variables $\mathbf{X}'$. Hence, we can leverage this physics knowledge to inform the functional form of the regression:

$$\hat{a}_{Physics} = f(a, \mathbf{X}'; \beta) + \epsilon, \epsilon \sim N(0, \sigma_\epsilon^2). \tag{3}$$

The physics-informed regression captures the actual relationship between the signal response and different factors by incorporating existing information. This also addresses the non-linearity of the data. However, the quality of the physics information is crucial to formulate a proper regression model.

2.2.2 Box-Cox transformation

Another approach to address nonlinearity is the Box-Cox transformation, which transforms $\hat{a}$ by the power of λ as follows:

$$\hat{a}^{(\lambda)} = \beta a + \epsilon, \epsilon \sim N(0, \sigma_\epsilon^2), \tag{4}$$

where

$$\hat{a}^{(\lambda)} = \begin{cases} \dfrac{\hat{a}^{\lambda} - 1}{\lambda}, & \lambda \neq 0 \\ \ln(\hat{a}), & \lambda = 0. \end{cases} \tag{5}$$

The $POD(a_i)$ function becomes to

$$POD_{Box}(a_i) = P(\hat{a}^{(\lambda)} > a_{th}^{(\lambda)}) = \Phi(\frac{a_i - (\beta_0 - a_{th}^{(\lambda)})/\beta_1}{\sigma/\beta_1}) \tag{6}$$

This approach does not need extra information compared with the physics-informed regression and obtains the value of λ by minimizing the sum of squared error. However, if the true relationship is more complex than the form in Equation 4, this approach will lose efficacy.

2.2.3 Robust regression

Robust regression is a statistical method that reduces the impact of outliers by modifying the loss function, making it less sensitive to extreme values than ordinary linear regression. Instead of using a sum of squares, Robust regression uses a different loss function $\rho(\cdot)$. The Tukey function [41] is one example for Robust regression:

$$\rho_{Tukey}(e) = \begin{cases} e^2, & |e| \leq k \\ 2k|e| - k^2, & |e| > k, \end{cases} \tag{7}$$

where $e = y_i - (\beta_0 + \beta_1 x_i)$ and k is a pre-set parameter. Accordingly, Robust regression estimates the parameters by the following:

$$\min_{\beta_0, \beta_1} \sum_{i=1}^{n} \rho(y_i - (\beta_0 + \beta_1 x_i)). \tag{8}$$

For robust regression, the regression model is the same as in Equation 1, but the distribution of ϵ may be broadened to some heavy-tailed distributions $F(\cdot)$ instead of a normal distribution. For example, ϵ may follow a t location-scale distribution, namely, $\epsilon \sim t(n, \sigma)$, where n is the degree of freedom, and σ is a scale parameter decided by the standard deviation of the error term. Thus, the $POD(a_i)$ function becomes

$$POD_{Robust}(a_i) = P(\hat{a}_i > a_{th}) = T_n(\frac{a_i - (\beta_0 - a_{th})/\beta_1}{\sigma/\beta_1}). \tag{9}$$

where T_n is the cumulative density function of the t distribution with the degree of freedom n. To determine the distribution, we can use the Shapiro-Wilk test to check the normality of data. If the data does not follow a normal distribution, we need to consider the t distribution.

2.2.4 Weighted linear regression

Weighted linear regression (WLS) is a statistical method used to address heteroscedasticity (i.e., the variance of the error term is not constant) in linear regression models [42]. The WLS can be described in Equation 1 but the error term follows $\epsilon_i \sim N(0, \sigma_i^2)$, $\sigma_i \neq \sigma_j$ if $i \neq j$ instead of being a constant. To estimate the parameters, we introduce a weight w_i so that $\tilde{\epsilon}_i = w_i \epsilon_i \sim N(0, \sigma^2)$. Unlike the loss function in ordinary least squares, WLS estimates the parameters by minimizing the loss function

$$loss_{WLS}(e_i) = w_i e_i^2, \tag{10}$$

where $e_i = (y_i - \beta_0 - \beta_1 x_i)$. This estimator minimizes the weighted sum of squared differences between the observed values and the values predicted by the linear model, accounting for the varying variances in the observations as indicated by the weights w_i.

In the POD context, the variance of the error term can be related to the crack size, namely,

$$\epsilon_i \sim N(0, \sigma^2 g(a_i)), \tag{11}$$

where $g(a_i)$ is a function of crack size. Thus, we can set the weight $w_i = \frac{1}{\sqrt{g(a_i)}}$. The $POD(a_i)$ function is given by

$$POD_{WLS}(a_i) = P(\hat{a}_i > a_{th}) = \Phi(\frac{a_i - (\beta_0 - a_{th})/\beta_1}{\sigma\sqrt{g(a_i)}/\beta_1}). \tag{12}$$

2.2.5 Bayesian approach for POD

The Bayesian approach is used to incorporate prior knowledge about the parameters to improve the estimation [43]. The posterior distribution is derived by combining the prior distribution, informed by previous experiments or expert knowledge, with the likelihood based on observed data. In the context of NDE, the regression model has been described in Equation 1, so the likelihood function is:

$$L(\hat{a}_i|a_i, \beta_1, \beta_0, \sigma) = \prod_{i=1}^{n} (2\pi\sigma^2)^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}(\hat{a}_i - \beta_0 - \beta_1 a_i)^2/\sigma^2\right\}. \tag{13}$$

According to Bayes' theorem, the posterior distribution given the prior $p(\beta_0)$ and $p(\beta_1)$ is

$$p(\beta_1, \beta_0, \sigma|\hat{a}_i, a_i) \propto L(\hat{a}_i|a_i, \beta_1, \beta_0, \sigma)p(\beta_0)p(\beta_1). \tag{14}$$

By maximum likelihood estimation, we can get the posterior distribution $\beta_{post} = (\beta_{0,post}, \beta_{1,post})$ and σ_{post}. Generally, we assume the coefficients are independent and the prior distributions are normally distributed, namely

$$\begin{aligned} p(\beta_0) &\sim N(\mu_{\beta_0}, \sigma_{\beta_0}^2) \\ p(\beta_1) &\sim N(\mu_{\beta_1}, \sigma_{\beta_1}^2). \end{aligned} \tag{15}$$

Thus, the posterior distributions of β_0 and β_1 are also normal, and the $POD(a_i)$ becomes

$$POD_{Bayes}(a_i) = \Phi(\frac{a_i - (\beta_{0,post} - a_{th})/\beta_{1,post}}{\sigma_{post}/\beta_{1,post}}). \tag{16}$$

2.2.6 Bootstrapping for POD

Bootstrapping is a statistical resampling method used to estimate the distribution of certain statistical measures in a dataset, in order to obtain more robust statistical inferences [44]. Bootstrapping does not require data to follow a normal distribution or any other specific distribution, making it suitable for complex or skewed data [45]. Especially for limited samples, bootstrapping doesn't rely on a single limited set and, hence might provide more robust estimates.

Given an original sample $(a_1, \hat{a}_1), \ldots, (a_n, \hat{a}_n)$, we can generate a group of i.i.d bootstrapping observations with replacement $(a_1^{(b)}, \hat{a}_1^{(b)}), \ldots, (a_n^{(b)}, \hat{a}_n^{(b)})$, where $b = 1, \ldots, B$. B is the number of bootstrapping and can be a large number, e.g., 500, 1,000. We used 500 in our simulation studies. For each bootstrap sample, the new estimate is $\hat{\theta}^{*(b)} = (\hat{\beta}_0^{*(b)}, \hat{\beta}_1^{*(b)}, \hat{\sigma}^{*(b)})$. Thus, the $POD(a_i)$ is

$$POD_{Boot}(a_i) = \frac{\sum_{b=1}^{B} \mathbf{1}\{\hat{\beta}_0^{*(b)} + \hat{\beta}_1^{*(b)} a_i > a_{th}\}}{B}. \tag{17}$$

The final estimation will be the average of the results from these B limited sets.

2.2.7 Combine advanced regression techniques with the physics-informed regression

As discussed in Section 1, we propose to combine a regression technique with the physics model to tackle the assumption violation and enhance the evaluation through information augmentation. The first step involves replacing the original linear regression model with a form that aligns with the physics-informed model, i.e., transitioning from the regression model in Equation 1 to the model described in Equation 3. The next step is to integrate advanced regression methods with physics-informed regression to tackle assumption violations. For example, if outliers exist in the data, we should combine the physics-informed model with robust regression. Thus, the loss function in Equation 8 becomes

$$\min_{\beta_0, \beta_1} \sum_{i=1}^{n} \rho(y_i - f(a, \mathbf{X}'; \beta)), \tag{18}$$

and the resulting $POD(a_i)$ function is

$$POD_{Phy+Robust}(a_i) = P(\hat{a}_i > a_{th}) = T_n(\frac{a_i - f^{-1}(a_{th}, \mathbf{X}'; \beta)}{\sigma}). \tag{19}$$

The integration of other regression techniques with the physics-informed model will follow a similar procedure. The $a_{90/95}$ can be obtained by the Delta method, similarly as in Section 1.

2.3 Comparison and validation metrics

To compare and validate the performance of LS-POD methods, we choose six metrics to measure the accuracy and precision of the estimation. Specifically, a_{50}, a_{90}, and percentage bias (PB) are metrics for accuracy; the confidence interval length CI and confidence interval ratio R are metrics for precision. The definitions of these metrics are described as follows.

From the curve $POD(a)$, a_{50} or a_{90} can be calculated as the smallest flaw size that can be detected with a 50% or 90% probability level, respectively, given by

$$a_{50} = \hat{\mu}_{pod}, \tag{20}$$

and

$$a_{90} = \hat{\mu}_{pod} + F^{-1}(0.9)\hat{\sigma}_{pod}. \tag{21}$$

Note, if F is the normal distribution, $F^{-1}(90) = \Phi^{-1}(90) = 1.286$. a_{50} represents the midpoint of the POD curve, indicating that at this flaw size, the chance of detection is equal to the chance of non-detection. a_{90} shows the ability of the NDE technique to detect defects. Smaller a_{50} and a_{90} values indicate better ability to detect smaller cracks, while larger a_{50} and a_{90} indicate degraded performance.

To validate the accuracy of models used in limited samples, the percentage bias (PB) of a_{90} is calculated as

$$PB = \frac{a_{90,m} - a_{90,baseline}}{a_{90,baseline}} = \frac{a_{90,m}}{a_{90,baseline}} - 1, \tag{22}$$

where $a_{90,m}$ is estimated by a specific LS-POD method, and $a_{90,baseline}$ is for the baseline. As we explained at the beginning of Section 2.1, POD estimation from a sufficiently large dataset is used as the baseline. If the value of PB is closer to 0, the estimated a_{90} is closer to the baseline, indicating a better performance. A value of PB less than 0 indicates that the estimated a_{90} is smaller than the baseline. It may also indicate that the model is overfitting, and the estimation of a_{90} is not conservative. A value of PB larger than 0 indicates that the estimated a_{90} is larger than the baseline. It also may indicate that the model is underfitting.

A 95% confidence interval of a_{90} is calculated to measure the precision of the NDE technique:

$$a_{90/95} = a_{90} + 1.64\sigma_{a_{a90}}. \tag{23}$$

From Equation 23, the 95% confidence interval length (CI) and ratio (R) are calculated as

$$CI = a_{90/95} - a_{90} = 1.64\sigma_{a_{a90}}, \tag{24}$$

and

$$R = \frac{a_{90/95}}{a_{90}}. \tag{25}$$

A narrower confidence interval indicates a more accurate estimate, while a wider confidence interval reflects greater uncertainty. The ratio metric also shows the uncertainty of the estimated a_{90}. However, unlike the length of the confidence interval, the ratio is not affected by the estimated value a_{90}, so it can be used to compare the confidence intervals of different methods.

3 Synthetic-based simulation studies

In this section, we used synthetic data generated from a representative physics-informed regression model as the baseline to illustrate the use of physics models in POD analysis. Notice that we chose a simple physics model since exploring different types of physics models is beyond the scope of this paper.

We first generated $N = 1,000$ samples without censoring from the physics model as the benchmark data. We then obtained limited sets ($n = 10$) with replacements from the benchmark data and adapted multiple regression techniques to conduct POD analysis. We also analyzed limited samples with $n = 20$ to demonstrate how the sample size of the limited sets affects the estimation, shown in Appendix A. Appendix B showed the results using a linear model to generate the benchmark data to demonstrate the performance of these LS-POD methods on linear data. Appendix C shows the results of empirical data.

3.1 Nonlinear benchmark data and the baseline POD estimation

We generated $N = 1,000$ samples without censoring from the physics model as the benchmark data. In this $\hat{a}$ vs. a model, the signal responses are assumed as a power function of crack size,

$$\hat{a} = \beta_1 f(a) + \epsilon, \epsilon \sim N(0, 0.3^2), \tag{26}$$

where $f(a) = a^b$. The nonlinear function $f(a)$ is used as the functional form of the regression in our study to compare the difference of POD with and without incorporating this model information. This form of physics-informed regression is one of the common relationships between crack size and signal amplitude for eddy current testing [46]. Specifically, the physics-model representation used in this paper is associated with the nonlinear trend observed with simulated results for the low-frequency eddy current inspection of fastener sites for second-layer cracks of varying size, presented in [9]. While the relationship is usually more complicated with additional parameters associated with various probe properties, material under test, and discontinuity state, the goal of this study is to evaluate the feasibility and highlight the benefits of leveraging the underlying physics model as a simple test case. We chose $\beta = 0.7$ and $b = 0.5$ as an illustration since these values result in clear non-linearity in the generated data. We also chose the decision threshold of crack vs. non-crack to be 0.38 so that different regression models produce the same value of a_{50}.

Three regression models are adapted to calculate the POD results from the benchmark data, as shown in Figure 2. The first is the linear model described in Equation 1. The second is the physics-informed regression leveraging physics knowledge about the relationship between the signal response and the crack size $\hat{a} = f(a)$ with known power b:

$$\hat{a}_{\text{known power}} = \beta_1 a^{0.5} + \epsilon, \epsilon \sim N(0, \sigma_\epsilon^2). \tag{27}$$

The third model is the physics-informed regression with unknown power b since in practice, the power is usually unknown:

$$\hat{a}_{\text{unknown power}} = \beta_1 a^b + \epsilon, \epsilon \sim N(0, \sigma_\epsilon^2). \tag{28}$$

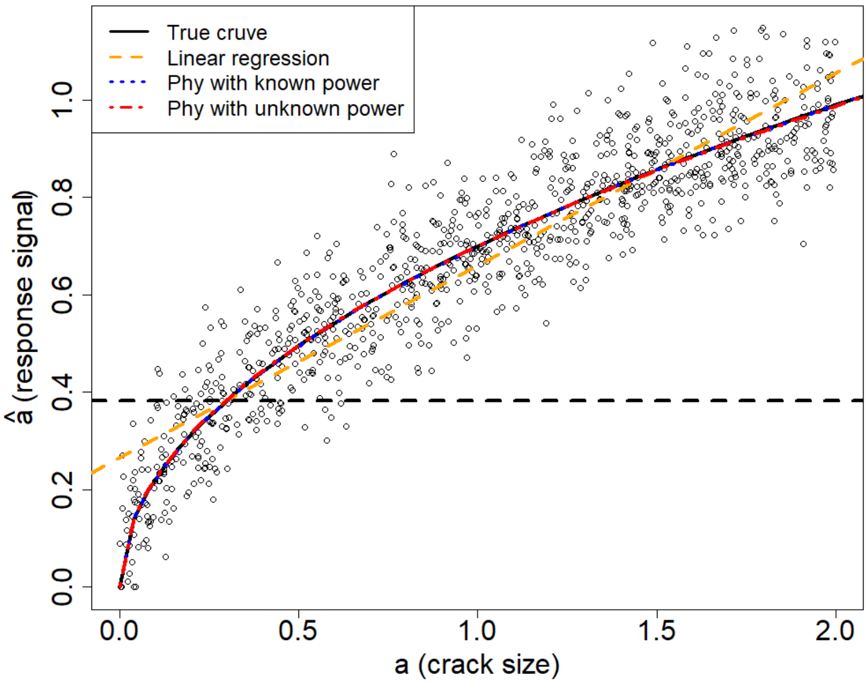


Fig. 2 The simulated benchmark data with the true curve (black solid), and the fitted lines from three LS-POD methods. The sample size is $N = 1,000$. The horizontal dashed line is the pre-chosen detection threshold of 0.38.

The simulated baseline data and fitted curves of these three regression models are illustrated in Figure 2. The plot shows that when the sample size is large, no matter whether we know the power parameter b or not, the two physics-informed regression methods can approximate the true curve. However, the linear regression line is a bit off from the true curve. This phenomenon confirms the importance of incorporating physics knowledge to inform the functional form of the regression. Even with unknown power, the physics-informed regression provides adequate estimation. We used the POD estimation from the physics-informed regression with known power as the baseline in the remaining studies.

In addition, when $a > 1.5$, the fitted lines of physics-informed regressions are under the linear model fit. It indicates that the POD results, like a_{90}, of physics-informed regressions, are smaller than the linear model. We can see that using linear regression when the actual relationship is nonlinear is likely to underestimate the capability of the NDE procedure.

The estimation results are shown in Table 1. R^2 is the coefficient of determination, and larger values of R^2 indicate better fitting. We can see that R^2 of physics-informed regression is 0.95, larger than that of linear regression, 0.91, indicating physics-informed regression fits data better than linear regression. The $a_{90/95}$ values for the physics-informed regression are also much smaller than for the linear model. We can see that using linear regression when the actual relationship is nonlinear is likely to incorrectly state the capability of the NDE procedure.

Table 1 Regression and POD estimation of the benchmark data ($N = 1,000$). We also included the average estimation from 1,000 data sets with $n = 30$ generated from the baseline model. We use phy as an abbreviation of physics in the following tables, figures, and text.

	Regression estimation			POD estimation		
Regression models	β_1	b	R^2	a_{50}	a_{90}	$a_{90/95}$
Linear regression	0.40	1.000	0.91	0.30	0.62	0.65
Phy with known power ($N = 1,000$)	0.70	0.500	**0.95**	0.30	0.50	**0.51**
Phy with unknown power ($N = 1,000$)	0.70	0.496	**0.95**	0.30	0.50	**0.51**
Phy with known power ($n = 30$)	0.70	0.500	0.97	0.30	0.50	0.60
Phy with unknown power ($n = 30$)	0.70	0.49	0.96	0.30	0.500	0.61

Since 30 is the required sample size for empirical POD studies, we also generated 1,000 data sets with sample size $n = 30$ and calculated their average POD estimation to compare with the large benchmark data. From Table 1, we can find that when the sample size meets the requirement in the MIL-HDBK-1823A handbook, the regression and POD estimations are as accurate as the results from the benchmark data with sample size $N = 1,000$. The only difference is when the sample size is smaller, the $a_{90/95}$ is much larger. This is because more data leads to less uncertainty in the estimation. We chose the results from the benchmark data ($N = 1,000$) as the baseline in the following sections.

3.2 Case studies of limited sets

Analysis of limited data requires advanced statistical techniques since the assumptions of simple linear regression are often violated. In this subsection, we considered three cases to show how our methodologies work under different assumption violations. We sampled small data sets ($n = 10$) from the benchmark data with replacement to represent limited samples in the real world. Statistical testing methods were used to identify assumption violations. We ensured that each limited set has only one of the following issues including nonlinearity, outliers, or heteroscedasticity to examine the performance of different methods specific to each type of assumption violation. A collection of 1,000 instances of limited samples were obtained for each case.

Case 1: Limited sets are nonlinear

As we discussed in Sections 1 and 2, we conducted linear regression, physics-informed regression, and Box-Cox transformation for nonlinear limited data sets.

Figure 3 presents the nonlinear dataset POD results (solid line) along with the lower confidence bounds (dashed line). The physics-informed regression with known power provides the best fit, closely aligning with the baseline curve. If we treat the power as an unknown parameter, the estimated value of b is 0.4, significantly smaller than the true value of 0.5. This is because small samples cannot afford to estimate more

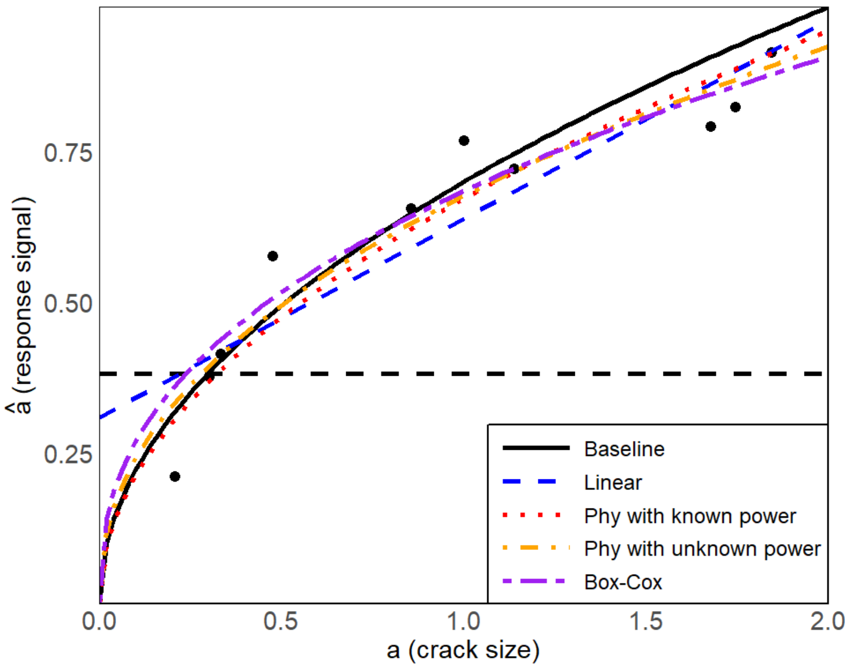


Fig. 3 Case 1: An example of the regression fitting for a nonlinear limited set. The black solid line is the baseline. The horizontal dashed line is the pre-chosen detection threshold.

parameters. The Box-Cox transformation yields results similar to the physics-informed regression with unknown power, but its performance is inferior to the physics-informed regression with known power. The linear fit is a bit off from the baseline curve and hence doesn't fit the data well.

The POD curves of this limited set are shown in Figure 4. We can see that the POD curve by the physics-informed regression with known power is the closest to the baseline, and the band is narrower than other regression models. So this method provides the best POD estimation, producing a more accurate and conservative POD fit. In contrast, the linear model provides the poorest fit to the data and has a smoother curve than the baseline, demonstrating that linear models are unsuitable for nonlinear datasets. Even though the Box-Cox transformation is not as good as physics-informed regression with known power, it is still better than the linear regression. The result indicates that using linear regression methods is not ideal for a nonlinear limited set. Table 2 shows the POD estimation for this limited set. The results from this example indicate that the physics-informed regression with known power provides adequate POD estimation for the nonlinear limited set.

We also evaluated the average estimation from 1,000 nonlinear limited data sets shown in Table 2 to study the performance patterns of different LS-POD methods. The same conclusion from the average estimation can be obtained from the example nonlinear data set, as can be seen from the patterns of the metrics. For example,

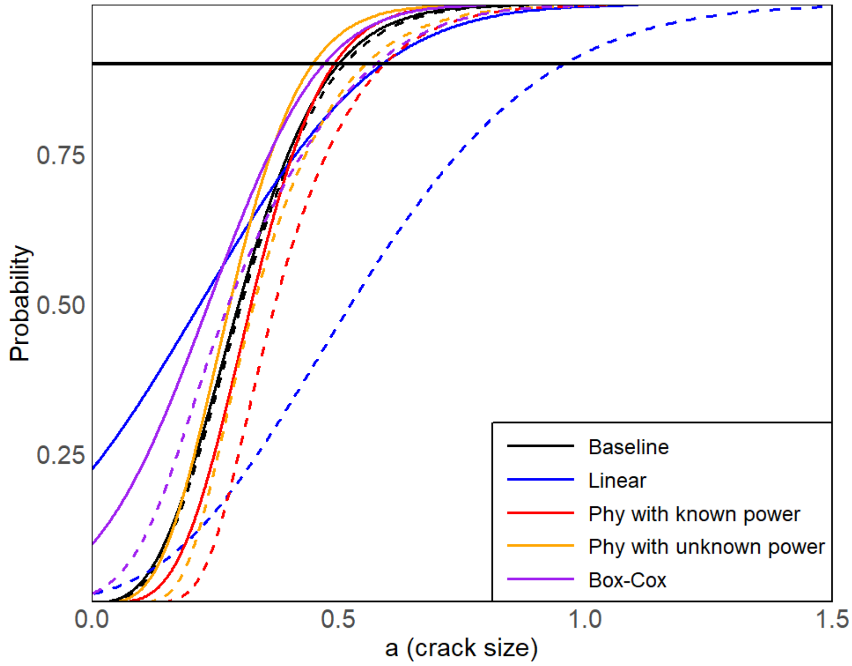


Fig. 4 Case 1: The POD curves for fitted models of the nonlinear limited set in Figure 3. The solid lines are the POD curves, the dashed lines are the lower 95% confidence bound, and the black horizontal line is the probability equal to 0.9.

Table 2 Case 1: Metrics of POD estimation for different regression methods for the limited set in Figure 3 and the average results of 1,000 nonlinear limited data sets.

	Regression models	a_{50}	a_{90}	PB	$a_{90/95}$	CI	R
Baseline	Phy with known power	0.30	0.50	0.00	0.51	0.01	1.02
One set results	Linear	0.22	0.59	0.18	0.96	0.37	1.63
	Phy with known power	**0.32**	**0.49**	−0.02	0.59	**0.10**	**1.20**
	Phy with unknown power	0.28	0.45	−0.10	0.55	**0.10**	1.22
	Box-Cox	0.24	0.47	−0.06	0.58	0.11	1.23
1,000 sets average results	Linear	0.14	0.48	−0.04	0.85	0.37	1.77
	Phy with known power	**0.30**	0.47	−0.06	0.58	0.11	1.23
	Phy with unknown power	0.26	0.42	−0.16	0.53	0.11	1.26
	Box-Cox	0.21	0.52	0.04	0.66	0.14	1.27
	Phy with known power +Bayes	0.32	**0.51**	**0.02**	0.61	0.10	**1.19**
	Phy with unknown power +Bayes	0.28	0.47	−0.06	0.57	0.10	1.21
	Phy with known power +Boot	**0.30**	0.45	−0.10	0.54	0.09	1.20
	Phy with unknown power +Boot	0.26	0.39	−0.22	0.48	0.09	1.23

the physics-informed regression with known power demonstrates superior performance with the most accurate a_{50} and a_{90}, as well as the smallest CI.

Next, we combined the Bayesian approach or Bootstrap with the physics-informed regression to examine whether combining multiple methods can improve the POD estimation further. We utilized the estimations from 100 random samples drawn from the large dataset to define the prior as follows:

$$\beta_1 \sim N(0.68, 0.04^2)$$
$$b \sim N(0.45, 0.02^2), \text{ when the power is unknown.} \tag{29}$$

Table 2 presents results when different methods were combined. Notably, the CI obtained by combining Bayesian methods with physics is smaller than those of physics-informed regression alone. For instance, the R value of physics-informed regression with unknown b decreases from 1.26 to 1.20 when Bayesian methods are incorporated. Furthermore, the estimated a_{90} is also improved. Specifically, the a_{90} of physics-informed regression with known b increases from 0.47 to 0.51, while that of physics-informed regression with unknown b increases from 0.42 to 0.45 after combining with the Bayesian method. In contrast, combining the Bootstrap method with physics does not yield good results. The a_{90} of physics-informed regression with known b decreases from 0.47 to 0.45, and with unknown b, it decreases from 0.42 to 0.39 after incorporating physics. These findings suggest that the Bootstrap method may not be suitable for LS-POD.

Case 2: Limited sets have outliers

As mentioned in Section 2, valid outliers should be kept in the analysis. We conducted physics-informed regression and also combined it with regression techniques from robust regression, Bayesian approach, or bootstrap for limited data sets with outliers.

Figure 5 presents an example dataset with an outlier and fit results. The slope of the physics-informed regression is overestimated due to the influence of the outlier. By combining physics information with robust regression (Phy+Robust), we significantly reduced the influence of the outlier and fit a regression model closer to the baseline. By combining physics information with the Bayesian approach (Phy+Bayes), the influence of the outlier was reduced also by having a closer fit to the baseline, but the slope is underestimated.

The POD curves for this limited set are shown in Figure 6. Table 3 shows the POD estimation for this limited set. We can see that the POD curve by combining the Bayesian approach or robust regression with the physics model is closer to the baseline compared with physics-informed regression. The confidence band by the Phy+Bayes model is narrower than that of combining other regression models. Although the 95% confidence interval of physics-informed robust regression is wider than that of combining the Bayesian approach, the interval contains the baseline curve. So the Phy+Bayes model provides the best POD estimations. Though the Phy+Robust model is not

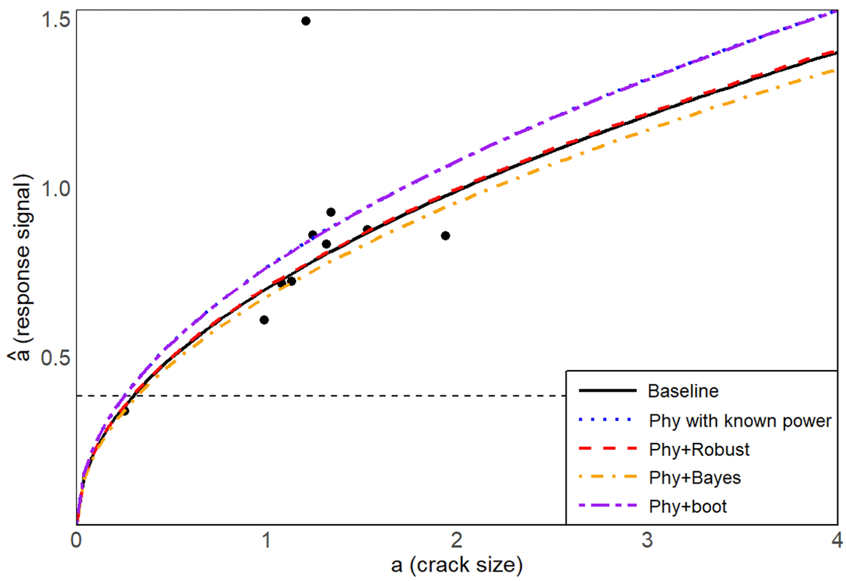


Fig. 5 Case 2: An example of the regression fitting for a nonlinear limited set with an outlier. The black solid line is the baseline. The horizontal dashed line is the pre-chosen detection threshold.

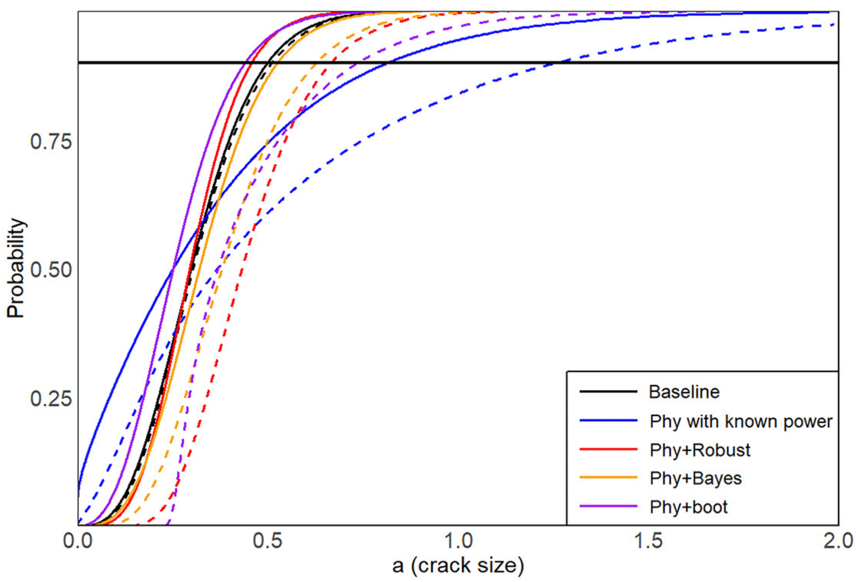


Fig. 6 Case 2: The POD curves for fitted models of a limited set with the outlier. The solid lines are the POD curves, the dashed lines are the lower 95% confidence bound, and the black horizontal line is the probability equal to 0.9

as good as the Phy+Bayes model, it is still better than only implementing physics-informed regression. The Phy+Boot model also improves POD estimation, but it requires significantly longer computational time.

Table 3 Case 2: Metrics of POD estimation from different regression methods for the limited set in Figure 5. We also showed the average results of 1,000 limited data sets with outliers.

	Regression models	a_{50}	a_{90}	PB	$a_{90/95}$	CI	R
Baseline	Phy with known power	0.30	0.50	0.00	0.51	0.01	1.02
One set results	Phy with known power	0.25	0.82	0.64	1.26	0.44	1.54
	Phy+Robust	**0.29**	0.46	−0.08	0.67	0.21	1.46
	Phy+Bayes	0.32	**0.52**	**0.04**	0.60	0.08	1.15
	Phy+Boot	0.25	0.44	0.74	−0.12	0.3	1.68
1,000 sets average results	Phy with known power	0.27	0.56	0.12	0.75	0.19	1.34
	Phy+Robust	**0.30**	0.49	−0.02	0.65	0.16	1.33
	Phy+Bayes	0.33	0.54	0.08	0.67	0.13	1.24
	Phy+Boot	0.27	0.55	0.10	0.65	0.13	1.25
	Phy+Bayes+Robust	0.29	**0.50**	0.00	0.59	**0.09**	**1.18**
	Phy+Boot+Robust	0.28	0.53	0.07	0.58	0.10	1.20

We also evaluated the average estimation results from 1,000 limited datasets with outliers in Table 3 to verify the performance patterns of different LS-POD methods. On average, the Phy+Robust model outperforms the Phy+Bayes model in terms of PB, with a value of −0.02 compared to 0.08 for the Phy+Bayes model. However, the Phy+Bayes model demonstrates better performance in terms of CI and R, achieving smaller values, 0.16 and 1.33 respectively, compared to robust regression. This difference arises because the robust method reduces the influence of outliers on the estimation but does not help the estimation of the variance, whereas the Bayesian approach incorporates prior information to reduce both bias and variance. Combining robust regression or the Bayesian approach with physics information improves the performance of physics-informed regression alone for limited datasets with outliers. The conclusions drawn from the average estimation of 1,000 limited sets are similar to those from the example dataset in Figure 6. We also generated a limited set where an outlier is below the threshold. The results are similar.

Table 3 presents the results of combining more than one regression technique with physics information. Notably, the estimated a_{50} and a_{90} for the Phy+Bayes+Robust model are 0.29 and 0.50, respectively, outperforming the Phy+Bayes model. Additionally, the CI of the Phy+Bayes+Robust model is 0.09 which is smaller than that of the Phy+Robust model. Similarly, the results of the Phy+Boot+Robust model are better than those of the Phy+Boot model. For instance, the CI and R of the Phy+Boot+Robust model decrease from 0.13 to 0.10 and from 1.25 to 1.20, respectively. These results demonstrate that combining multiple advanced regression

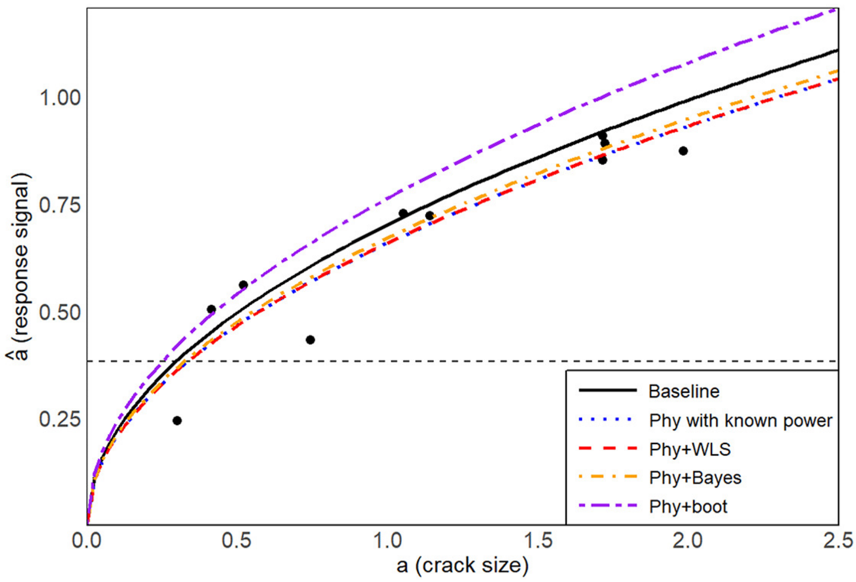


Fig. 7 Case 3: An example of the regression fitting for a limited set with heteroscedasticity. The black solid line is the baseline fitting. The horizontal dashed line is the pre-chosen detection threshold.

techniques with the physics information yields superior performance compared to combining a single regression technique. The results highlight the benefits of integrating different regression approaches, as each regression technique has its own advantages.

Case 3: Limited sets have heteroscedasticity

The third case study is where a limited set exhibits heteroscedasticity. Again, we conducted physics-informed regression and also combined it with regression techniques from weighted regression, the Bayesian approach, and bootstrap for limited data sets with heteroscedasticity.

Figure 7 illustrates a dataset exhibiting heteroscedasticity alongside its fitted lines. The slope estimated by the physics-informed regression is slightly underestimated due to the impact of heteroscedasticity. The Phy+WLS model does not improve fitted lines since it has minimal effect on the coefficient estimation. However, the Phy+Bayes model reduces the influence of heteroscedasticity, resulting in a fitted line closer to the baseline. In contrast, combining the Bootstrap method with physics-informed regression (Phy+Boot) does not enhance the fitted line, resulting in an overestimated outcome.

The POD curves of this limited set are shown in Figure 8. Table 4 shows the POD estimation for this limited set. We can see that the POD curve by the Phy+WLS model is closer to the baseline than other regression models. The band of the Phy+WLS model is narrower than only using physics-informed regression. So Phy+WLS provides the best POD estimations. The results from this example indicate that the

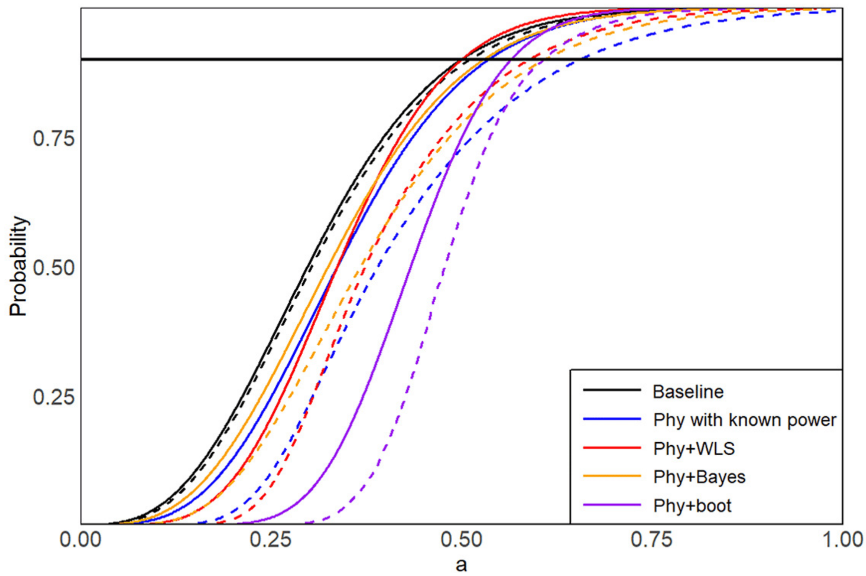


Fig. 8 Case 3: The POD curves for fitted models of a limited set with heteroscedasticity. The solid lines are the POD curves, the dashed lines are the lower 95% confidence bound, and the black horizontal line is the probability equal to 0.9

Phy+WLS model delivers adequate POD estimation for limited datasets exhibiting heteroscedasticity.

Table 4 Case 3: Metrics of POD estimation from different LS-POD methods in Figure 7. The average results of 1,000 limited data sets with heteroscedasticity are also included.

	Regression models	a_{50}	a_{90}	PB	$a_{90/95}$	CI	R
baseline	Phy with known power	0.30	0.50	0.00	0.51	0.01	1.02
One set results	Phy with known power	0.34	0.54	0.08	0.66	0.12	1.22
	Phy+WLS	0.34	**0.50**	0.00	0.58	0.08	1.16
	Phy+Bayes	**0.32**	0.53	0.06	0.61	0.08	1.15
	Phy+Boot	0.38	0.56	0.12	0.61	0.05	1.09
1,000 sets average results	Phy with known power	0.31	0.47	−0.06	0.58	0.11	1.23
	Phy+WLS	0.31	0.47	−0.06	0.54	0.07	1.15
	Phy+Bayes	0.32	0.53	0.06	0.61	0.08	1.15
	Phy+Boot	0.31	0.46	−0.08	0.54	0.08	1.17
	Phy+Bayes+WLS	0.32	**0.51**	**0.02**	0.56	**0.05**	**1.10**
	Phy+Boot+WLS	**0.30**	0.46	−0.08	0.53	0.07	1.15

We also evaluated the average estimation results from 1,000 limited datasets with heteroscedasticity in Table 4. On average, the Phy+WLS model outperforms the

Phy+Bayes model in terms of PB, with a value of 0.00 compared to 0.08 for the Phy+Bayes model. However, the Phy+Bayes model demonstrates better performance on estimating a_{50}, achieving a value of 0.32, respectively, compared to the Phy+WLS model. Combining WLS regression or the Bayesian approach with physics information improves the performance of physics-informed regression alone for limited datasets with heteroscedasticity. These conclusions drawn from the average estimation are similar to those from the example dataset in Figure 8.

Table 4 also presents the results of combining different methods. Notably, the estimated a_{90} for the Phy+Bayes+WLS model is 0.51, respectively, outperforming the Phy+WLS model. Additionally, the R of the Phy+Bayes+WLS model is 1.10, smaller than that of the Phy+WLS or Phy+Bayes model. However, the performance of the Phy+Boot+WLS model does not improve a lot compared with the Phy+Boot model. These results demonstrate that the Phy+Bayes+WLS model yields superior performance compared to the other regression techniques for limited sets exhibiting heteroscedasticity, while the Bootstrap method is not suitable for the limited sets.

3.3 Average performance of 1,000 randomly generated limited sets

To test the generality and robustness of these regression methods, we randomly selected 1,000 limited sets from the full benchmark data without grouping them by violated assumptions. We applied all possible regression methods to get POD estimation and calculate the average values.

Table 5 The average POD estimation of 1,000 random limited sets regardless of assumption violations.

Regression models	a_{50}	a_{90}	PB	$a_{90/95}$	CI	R
Baseline	0.30	0.50	0.00	0.51	0.01	1.02
Linear	0.23	0.53	0.06	0.88	0.35	1.66
Phy with known power	0.31	**0.50**	0.00	0.66	0.16	1.32
Phy with unknown power	0.27	0.45	−0.10	0.61	0.16	1.36
Box-Cox	0.29	0.60	0.20	0.69	0.09	1.15
Phy with known power + Bayes	0.32	0.47	−0.06	0.60	0.13	1.30
Phy with unknown power + Bayes	0.28	0.43	−0.14	0.56	0.13	1.30
Phy + Robust	0.31	0.48	−0.04	0.66	0.18	1.38
Phy + WLS	0.31	0.49	−0.02	0.60	0.11	1.22
Phy + Boot	0.31	0.48	−0.04	0.58	0.10	1.21
Phy + Bayes + Robust	0.31	0.47	−0.06	0.58	0.11	1.23
Phy + Boot + Robust	0.29	0.51	0.02	0.62	0.11	1.22
Phy + Bayes + WLS	0.31	0.51	0.02	0.59	**0.08**	1.16
Phy + Boot + WLS	0.31	0.48	−0.04	0.57	0.09	1.19

As seen in Table 5, both the average estimation of a_{50} and a_{90} of the physics-informed regression with known power are close to the baseline, but the physics-informed regression with unknown power has relatively large bias. This indicates that if we do not have enough data, it is hard to get accurate estimations by using models with more unknown parameters. The Box-Cox can improve the estimation of a_{50}, but a_{90} is much larger than the baseline.

The Phy+Robust with physics gets similar results as the physics-informed regression with known power. This is because if there are no outliers, the loss function of robust regression is the same as that of simple regression. The Phy+WLS model can reduce $a_{90/95}$, since if the data have heteroscedasticity, the variance can be reduced. If the variance is a constant, the model reduces to a simple regression. The estimated a_{50} and a_{90} of combining the Bayesian approach with physics are 0.32 and 0.47 because the prior distribution has more impact on the posterior distribution when the sample size is limited. However, when we combine robust regression or weighted regression with the Bayesian approach, the contribution of data increases. Thus, they perform better than if we only use the Bayesian approach with the physics-informed regression. Using the Phy+Boot model can also get similar results to those obtained by only using a physics-informed regression. Therefore, combining advanced regression approaches with physics information has great potential to improve the LS-POD significantly.

We also checked the proportions of data sets with different assumption violations, shown in Table 6. The proportion of limited sets with outliers or nonlinearity is much larger than that of limited sets with heteroscedasticity. Only a few limited sets violate more than one assumption. We can see that in our experiment settings, outliers and nonlinearity are more common, and usually one type of assumption violation dominates. Therefore, it should be sufficient to tackle the dominant violation instead of concerning multiple violations.

Table 6 The distribution of different assumption violations for 1,000 limited sets sampled from the baseline data.

Violations	Nonlinear	Outliers	Heteroscedastic
No. (%)	130 (13%)	148 (14.8%)	53 (5.3%)
Violations	Nonlinear+Outliers	Nonlinear+Heteroscedastic	Outliers+Heteroscedastic
No. (%)	15 (1.5%)	8 (0.8%)	3 (0.3%)

3.4 Summary of results

Overall, the results presented in the tables and figures of Section 3 demonstrate how various regression models affected the performance of LS-POD estimation across different cases. In cases when assumptions are violated, advanced regression models significantly enhance the accuracy and precision of POD estimation. This was evident since the estimates of a_{50} and a_{90} are closely aligned with the baseline and the 95% confidence intervals are narrower. Furthermore, combining multiple advanced regression techniques with information-augmentation models yielded superior performance compared to using a single regression method. However, the results obtained through the bootstrap method showed minimal improvement over simple linear regression, suggesting that bootstrap may not be suitable for LS-POD. Moreover, different LS-POD methods mainly affect the accuracy of the estimation. The precision is mainly determined by the sample size. Even though methods such as Bayesian analysis and physics-informed regression can improve the precision, the improvement is limited.

Appendix B provided results based on a linear baseline. In this scenario, advanced regression techniques still performed better than the simple linear regression model. We also incorporated some misleading physics information that describes a nonlinear relationship. Although this physics-informed regression yielded smaller POD estimations, the estimated a_{50} and a_{90} deviated significantly from the baseline, highlighting that inaccurate physics information can lead to problematic estimation. We also analyzed limited samples with $n = 20$ in Appendix A. The results indicated that advanced regression techniques significantly enhance the LS-POD estimation in limited samples, consistent with the results in Section 3. As the sample size increases, the performance of simple regression is improved. However, the advanced regression approach consistently delivers superior results.

4 Conclusion and discussion

This paper establishes a comprehensive framework to enhance POD estimation when the sample size is limited. Specifically, we adapt multiple advanced regression methods when the statistical assumptions are violated. We also combine these regression techniques with physics-based knowledge to address assumption violations and incorporate physics simultaneously. To evaluate the effectiveness of our approaches under different scenarios, we conducted extensive simulation studies using both synthetic and empirical data, enabling us to compare the performance of these methods against a presumed baseline.

The simulation studies demonstrate that either adapting appropriate regression methods to address assumption violations or incorporating physics-based information can significantly enhance POD estimation. Furthermore, combining advanced regression techniques with physics models offers even greater improvements in POD estimation accuracy and precision.

To the best of our knowledge, this study is the first to apply robust regression and weighted regression in LS-POD to mitigate the effects of assumption violations on POD estimation. It is also among the pioneering efforts to integrate advanced regression methods with physics-based knowledge, simultaneously addressing assumption violations and leveraging physics insights. Through systematic evaluation, our findings provide valuable insights into how different LS-POD approaches impact POD estimation across various scenarios, offering practitioners practical guidance for selecting the proper LS-POD methods for specific conditions. For example, practitioners can first determine the relationship between the signal response and crack size. If the relationship is not strictly linear, they can initially apply some transformation or conduct an appropriate physics-informed regression. If suitable prior knowledge is available, the Bayesian approach can be used to enhance the estimation. Next, they should assess whether key assumptions are violated in limited sample scenarios. If violations exist, selecting appropriate advanced regression techniques can help address these issues. These methods can be combined to further improve the estimation.

Note that it is problematic to conclude that smaller values of a_{90} indicate better estimation for LS-POD. As we can see from our simulation studies, the baseline a_{90} can be larger than the a_{90} obtained by a regression method. In this case, the smaller the value of a_{90} compared with the baseline, the worse the estimation. Nevertheless, a sensible baseline is often unavailable in practice, making it difficult to evaluate the adequacy of LS-POD estimation. Determining POD performance with limited sample sizes remains a critical challenge. Simulation studies, like those conducted in this research, provide valuable insights into the adequacy of methods under different scenarios. Furthermore, we used a relatively simple physics-informed model in our simulation studies. In the real world, more complex physics models with varying parameters and uncertainty will arise. Future research can focus on investigating more intricate physics models or developing new methods to incorporate parameter variability with model uncertainty. Another reminder is that although Bayesian methods help to incorporate prior knowledge, the priors affect, and may even dominate the estimation when the sample sizes are limited. Caution needs to be taken when obtaining priors for Bayesian analysis in LS-POD. Additionally, we only considered the continuous response problem. In the future, this LS-POD framework can also be adapted for hit/miss data analysis.

5 Acknowledgment

The authors would like to express their sincere gratitude to the Model-Assisted Probability of Detection Initiative Team for their invaluable assistance and constructive suggestions throughout this research. Their expertise and support significantly contributed to the successful completion of this work. Any errors or omissions are the sole responsibility of the authors.

6 Declarations

This work was supported by the Center for Nondestructive Evaluation (CNDE) at Iowa State University.

Appendix A Synthetic-based study results: limited sets with 20 samples

As described in Section 3, we have shown the results when sample sizes are 10. To determine the influence of sample size on LS-POD estimation, we showed the average results of 1,000 limited sets when $n = 20$ in this appendix.

Table A1 shows the average results of 1,000 nonlinear data sets whose sample size is 20. The table shows that the estimated a_{50} and a_{90} of the physics-informed regression with known power are closer to the baseline than the physics-informed regression with unknown power and Box-Cox method. These results are similar to what we got in Section 3.2. Like the conclusion in Section 3.2, when we combined the physics-informed regression with the Bayesian approach, the estimated a_{90} and its confidence interval $a_{90/95}$ are improved.

Table A1 Metrics on different combination model POD comparisons of average results of 1,000 nonlinear data sets ($n = 20$).

Models	a_{50}	a_{90}	PB	$a_{90/95}$	CI	R
Baseline	0.30	0.50	0.00	0.51	0.01	1.02
Linear	0.20	0.48	−0.04	0.75	0.27	1.56
Phy with known power	0.30	0.47	−0.06	0.55	0.08	1.17
Phy with unknown power	0.26	0.45	−0.10	0.53	0.08	1.18
Box-Cox	0.21	0.52	0.04	0.60	0.08	1.15
Phy with known power	0.32	0.49	−0.02	0.57	0.08	1.16
Phy with unknown power	0.28	0.45	−0.10	0.55	0.10	1.22
Phy with known power +Bayes	0.32	0.51	0.02	0.56	0.05	1.10
Phy with unknown power +Bayes	0.28	0.47	−0.06	0.53	0.06	1.13
Phy with known power +Boot	0.30	0.45	−0.10	0.54	0.09	1.20
Phy with unknown power +Boot	0.26	0.39	−0.22	0.48	0.09	1.23

Table A2 shows the average results of 1,000 data sets with outliers whose sample size is 20. When the outliers exist, if we only use a physics-informed regression, the bias of estimated a_{50} and a_{90} will increase. Robust regression and the Bayesian approach can reduce the influence of outliers. Compared with the Bayesian approach, robust regression does not need more information. Moreover, when we combine robust regression and the Bayesian approach, this combination method can get better results than robust regression or the Bayesian approach. These conclusions are consistent with those described in Section 3.2.

Table A2 Metrics on different combination model POD comparisons of average results of 1,000 data sets with outliers ($n = 20$).

Models	a_{50}	a_{90}	PB	$a_{90/95}$	CI	R
Baseline	0.30	0.50	0.00	0.51	0.01	1.02
Phy with known power	0.27	0.54	0.08	0.70	0.16	1.30
Phy+Robust	0.30	0.49	−0.02	0.60	0.11	1.22
Phy+Bayes	0.33	0.52	0.04	0.65	0.13	1.25
Phy+Boot	0.27	0.52	0.04	0.63	0.11	1.21
Phy+Bayes+Robust	0.29	0.51	0.02	0.57	0.06	1.12
Phy+Boot+Robust	0.28	0.48	−0.04	0.65	0.17	1.35

Table A3 shows the results of 1,000 heteroscedasticity data sets whose sample size is 20. When the variance is not constant, the 95% confidence interval, $a_{90/95}$ of the physics-informed regression is large. WLS method can improve the estimated $a_{90/95}$ to get a smaller 95% confidence interval than the physics-informed regression. When we combine WLS and the Bayesian approach, the PB of the combination method is smaller than that of the other methods. These conclusions are consistent with those described in Section 3.2. These conclusions are consistent with those described in Section 3.2.

Table A3 Metrics on different combination model POD comparisons of average results of 1,000 heteroscedasticity data sets ($n = 20$).

Models	a_{50}	a_{90}	PB	$a_{90/95}$	CI	R
Baseline	0.30	0.50	0.00	0.51	0.01	1.02
Phy with known power	0.31	0.48	−0.04	0.56	0.08	1.17
Phy+WLS	0.31	0.48	−0.04	0.53	0.05	1.10
Phy+Bayes	0.32	0.53	0.06	0.59	0.06	1.11
Phy+Boot	0.31	0.46	−0.08	0.54	0.08	1.17
Phy+Bayes+WLS	0.32	0.50	0.00	0.54	0.04	1.08
Phy+Boot+WLS	0.30	0.46	−0.08	0.53	0.07	1.15

In conclusion, compared to the results of $n = 10$, when we increase the sample size to 20, the 95% confidence interval $a_{90/95}$ is smaller, and the average estimated a_{50} and a_{90} are not changed significantly.

Appendix B Synthetic-based Study with a linear benchmark data

In Section 3, we showed the results of the LS-POD methodology using a nonlinear baseline. Here we also showed the results based on a linear synthetic data set using the linear model in Equation 1. The parameters of baseline are $\beta_1 = 0.55$, $\beta_0 = 0$, and $\epsilon \sim N(0, 0.2^2)$.

Table B4 Average POD results of 1,000 nonlinear limited data sets ($n = 10$) sampled from the linear benchmark data ($N = 1,000$).

Models	a_{50}	a_{90}	PB	$a_{90/95}$	CI	R
Baseline	0.73	1.12	0.00	1.14	0.02	1.02
Linear	0.77	1.28	0.14	1.62	0.34	1.27
Phy	0.63	1.11	−0.01	1.46	0.35	1.32
Box-Cox	0.65	1.08	−0.04	1.42	0.34	1.31
Bayes	0.65	1.08	−0.04	1.30	0.22	1.20
Boot	0.76	1.27	0.13	1.51	0.24	1.19

Table B4 shows the average POD results of 1,000 nonlinear limited data sets ($n = 10$) sampled from the linear benchmark data ($N = 1,000$). Compared with the baseline, if the data set is nonlinear, the estimated a_{50} and a_{90} of linear regression are underestimated. When we apply the data transformation methods, like the physics model or Box-Cox method, estimated a_{90} decreases and is closer to the baseline. However, the estimated a_{50} is overestimated. The Bayesian approach yields similar a_{50} and a_{90} values but a smaller $a_{90/95}$. The bootstrap method cannot improve the POD estimation, similar to the nonlinear case study in Section 3.

Table B5 shows the average POD estimation of 1,000 limited sets with outliers in the linear baseline. Compared with the baseline, if the data set has outliers, the estimated a_{50} and a_{90} of linear regression are overestimated. Robust regression and the Bayesian approach can improve estimated a_{50} and a_{90} so that the estimations are closer to the baseline. Although the estimated a_{50} of the Bayesian approach is better than robust regression, robust regression does not need more information. The combination methods of robust regression and the Bayesian approach can not only improve estimated a_{50} and a_{90}, but also get a smaller $a_{90/95}$. The bootstrap method cannot improve the POD estimation significantly.

Table B5 Average POD results of 1,000 limited data sets ($n = 10$) with outliers sampled from the linear benchmark data ($N = 1,000$)

Models	a_{50}	a_{90}	PB	$a_{90/95}$	CI	R
Baseline	0.73	1.12	0.00	1.14	0.02	1.02
Linear	0.57	0.94	−0.16	1.25	0.31	1.33
Robust	0.68	1.00	−0.11	1.35	0.35	1.35
Bayes	0.71	0.99	−0.12	1.29	0.30	1.30
Boot	0.58	0.90	−0.20	1.22	0.32	1.36
Bayes+robust	0.70	1.00	−0.11	1.20	0.20	1.20
Boot+robust	0.62	0.97	−0.13	1.25	0.28	1.29

Table B6 shows the results of 1,000 heteroscedasticity data sets in linear baseline. Compared with the baseline, if the variance is not constant, the estimated a_{50} and a_{90} of linear regression are underestimated and the 95% confidence interval is large. WLS can reduce $a_{90/95}$, but cannot improve a_{50}. Although the Bayesian approach can get a smaller a_{50} and a_{90}, its a_{90} is overestimated. The combination methods of WLS and the Bayesian approach can improve estimated a_{90} and get a smaller $a_{90/95}$. The bootstrap method cannot improve the POD estimation significantly.

Table B6 Average POD results of 1,000 heteroscedasticity limited data sets ($n = 10$) sampled from the linear benchmark data ($N = 1,000$)

Models	a_{50}	a_{90}	PB	$a_{90/95}$	CI	R
Baseline	0.73	1.12	0.00	1.14	0.02	1.02
Linear	0.83	1.31	0.17	1.64	0.33	1.25
WLS	0.84	1.16	0.04	1.42	0.26	1.22
Bayes	0.78	1.06	−0.05	1.26	0.20	1.19
Boot	0.83	1.30	0.16	1.57	0.27	1.21
Bayes+WLS	0.82	1.10	−0.02	1.32	0.22	1.20
Boot+WLS	0.82	1.29	0.15	1.52	0.23	1.18

In conclusion, the advanced regression method can improve the POD estimation compared with simple linear regression in this case study. Similar conclusions have been obtained in the nonlinear baseline case study in Section 3.

Appendix C Empirical-based study

This appendix used empirical data instead of synthetic data to demonstrate the feasibility of our methodology in the real world. Other procedures are similar to Section 3. This dataset came from Koshti et al. [25] with 92 measurements of varying flaw sizes and no censored observations as empirical data. The threshold is set as 265. According to the recommendations, we can model the signal response as a function of the transformed flaw size as $f(a) = \log(a)$, so we treat it as our physics information. Log transformation of the data may automatically apply in some research, but in our study, we focus on how to use advanced regression techniques and combine them with physics-informed regression in LS-POD. Thus, it is proper to use log transformations as the physics information.

C.1 Benchmark analysis

The simulated data and fitted curves of two regression models are illustrated in Figure C1. The fitted curve of the physics-informed regression lies above the fitted linear line beyond the threshold until $a = 37$, resulting in a smaller estimated a_{90} for the physics-informed regression compared to the linear regression. This indicates that the probability of detecting a crack is higher with the physics-informed regression than with the linear model.

Table C7 shows the details of estimation. Although the coefficient of determination R^2 of the linear model is better than the physics-informed regression, the estimated a_{90} of physics-informed regression is 13.9, smaller than that of the linear regression. The POD estimation is more important than model fitting in the NDE field. Thus, in the limited sample POD analysis, we will use physics-informed regression as the baseline to measure the performance of different models.

Table C7 Regression and POD estimation of the empirical benchmark data.

	Regression estimation			POD estimation		
Regression models	β_0	β_1	R^2	a_{50}	a_{90}	$a_{90/95}$
Physics-informed regression	−960.1	535.5	0.88	9.6	13.9	15.2
Linear regression	0	26.9	0.95	9.5	17.0	18.2

C.2 Limited sample POD

Since the threshold is 265, our crack size of interest is between 12 to 32. We randomly sampled a limited set ($n = 10$) in this range. For the Bayesian approach, the prior

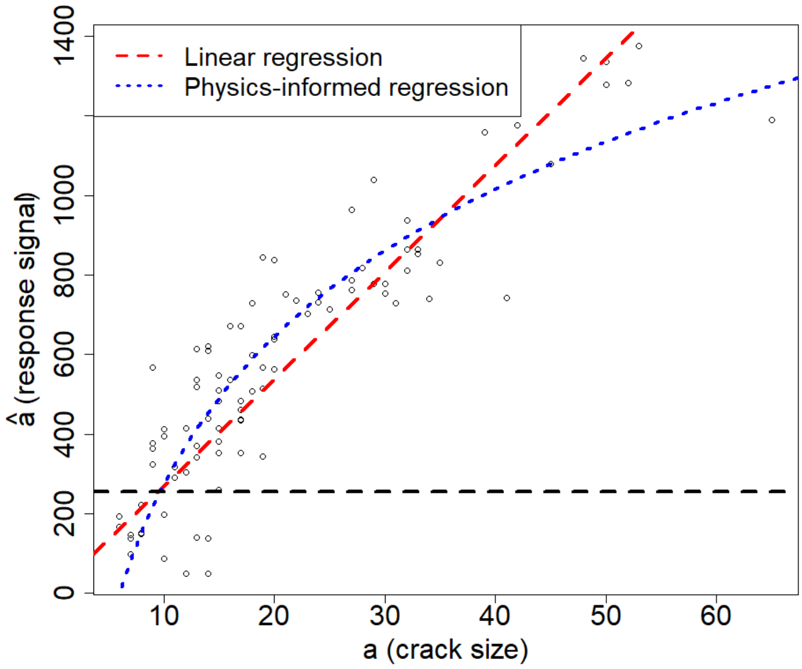


Fig. C1 An empirical data with linear regression (red) and physics-informed regression (blue) fitting results. The horizontal dashed line is the pre-set detection threshold.

distribution was estimated from 50 data points randomly sampled from the benchmark data. To be specific, the prior distributions are

$$\begin{aligned} \beta_0 &\sim N(-1134, 180.5^2), \\ \beta_1 &\sim N(588, 60.5^2). \end{aligned} \tag{C1}$$

By the White test, the $p-value$ of $0.662 > 0.05$ implies the variance of the errors is constant. One outlier was detected whose Cook's distance was larger than 3. Therefore, we used robust regression in this limited set but did not use WLS regression. And by Ramsey's RESET, the $p-value$ is 0.02 less than 0.05 which indicates that the nonlinearity exists in the limited set. Figure C2 presents the fitted lines of different LS-POD methods on the limited set ($n = 10$). Compared with the Box-Cox transformation, the physics-informed regression fits the data better. However, the physics-informed regression fit with the data is deviated by the outlier from the underlying relationship. Through using the Phy+Robust model, the fitted curve is improved, resulting in the fitted line closer to the baseline. The result is consistent with the conclusions in Section 3. Moreover, by integrating robust regression into the Bayesian approach with physics information (Phy+Bayes+Robust), the fitted line more closely approximates the baseline. In contrast, using the Phy+Boot+Robust model does not enhance the results.

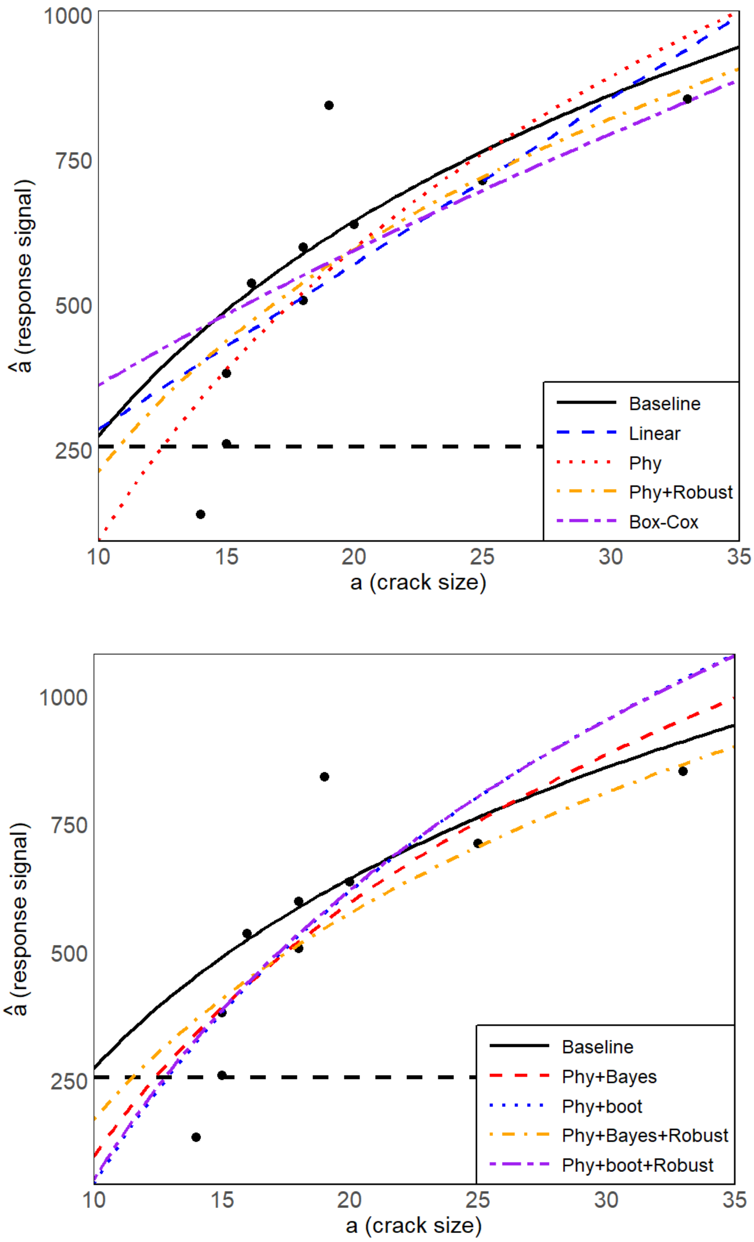
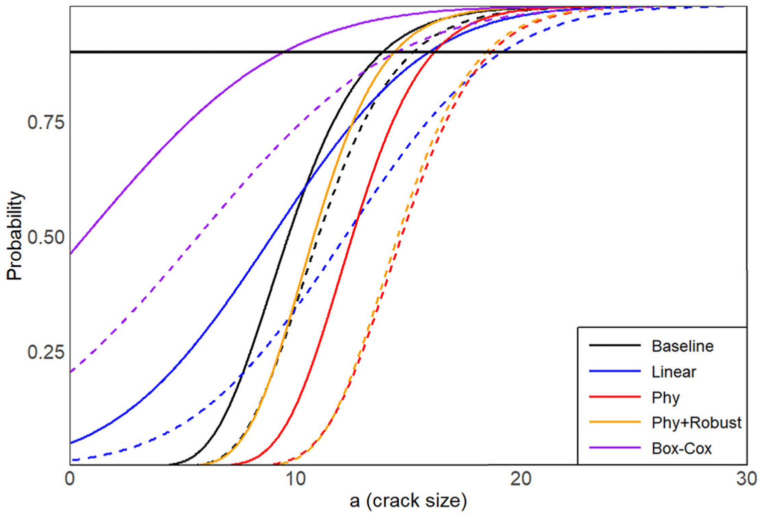


Fig. C2 Results for fitted models of limited samples in empirical data. The black solid line is the baseline. The horizontal dashed line is the pre-chosen detection threshold.

Figure C3 shows the POD curves and Table C8 gives a detailed estimation. The POD curves by the linear model and Box-Cox model are smoother compared with



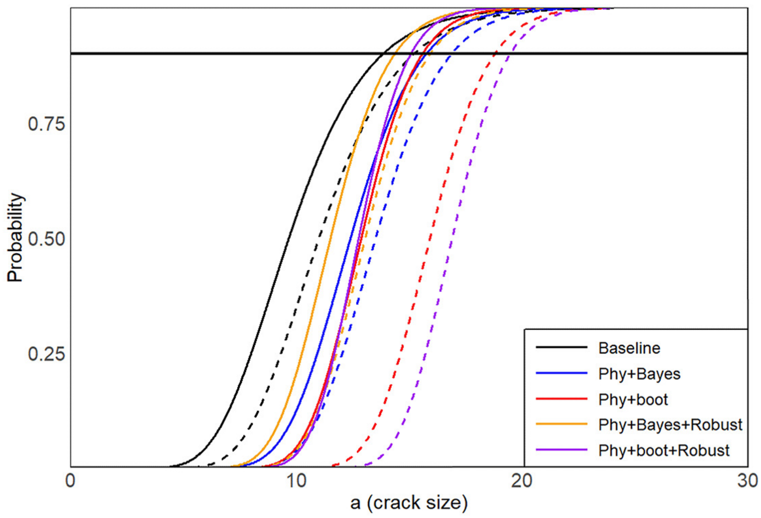


Fig. C3 POD curves of the limited set ($n = 10$) generated from the empirical data ($N = 92$) by multiple LS-POD methods. The solid lines are the POD curves, the dashed lines are the lower 95% confidence bound, and the corresponding probability of the black horizontal line is 0.9.

the baseline, while if we use physics information, curves are steeper. Moreover, the estimated a_{50} of the linear model and Box-Cox model are too small but their estimated a_{90} are large. The results indicate that the estimated σ is too large, so it is not proper to use a linear model or Box-Cox model in this experiment. By using a physics-informed regression, although the estimated a_{50} is larger than the baseline, it is conservative. The Phy+Robust model can improve the a_{50} significantly, but $a_{90/95}$ is too large. This is because although the robust method can reduce the influence of outliers, it cannot reduce the variance of estimated parameters. Bootstrap methods do not seem suitable to be used in this experiment, since the estimated a_{50} and a_{90} are not improved and the 95% confidence interval of a_{90} increases. By incorporating prior knowledge, the 95% confidence interval of a_{90} is significantly reduced; however, compared with the baseline, it is overestimated and not conservative. When we use the Phy+Bayes+Robust model, the a_{50} is better than just the Phy+Bayes model; a_{90} and its 95% confidence interval are both improved. The Phy+Bayes+Robust model performs best among the models.

Table C8 Metrics of different LS-POD methods from the limited set ($n = 10$).

Regression models	a_{50}	a_{90}	PB	$a_{90/95}$	CI	R
Baseline	9.60	13.90	0.00	15.20	1.30	1.09
Linear	8.98	15.93	0.15	19.18	3.25	1.20
Phy	12.50	16.16	0.16	18.67	2.51	1.16
Box-Cox	6.19	14.98	0.08	19.60	4.62	1.31
Phy+Robust	**10.81**	14.39	0.04	18.46	4.07	1.28
Phy+Bayes	12.42	15.89	0.14	16.90	**1.01**	**1.06**
Phy+Boot	12.89	15.61	0.12	18.81	3.20	1.20
Phy+Bayes+Robust	11.53	**14.38**	**0.03**	15.94	1.56	1.11
Phy+Boot+Robust	12.78	15.09	0.09	19.49	4.40	1.29

From the above results, we can conclude that when an outlier exists in a limited dataset, the POD estimations from simple linear or physics-based regression deviate significantly from the baseline. However, combining advanced regression techniques with physics models improves the accuracy of the results. Furthermore, employing multiple regression techniques enhances the estimations even further. These findings are consistent with those presented in Section 3. Through this empirical-based study, we demonstrate the application of our methodology to a real dataset. The results highlight the feasibility of the proposed approach, highlighting its effectiveness in improving POD estimation under limited samples. While results are promising, future work is needed to validate the LS-POD approach for cases using numerical physics-based measurement models.

References

[1] Nondestructive evaluation system reliability assessment. United States Department of Defense, Wright-Patterson AFB, Handbook MIL-HDBK-1823A, 9 (2009)

[2] Burkhardt, G.L., Beissner, R.: Asm handbook, volume 17: Nondestructive evaluation and quality control asm handbook committee, p 136-142. Nondestructive Evaluation and Quality Control **17**, 136 (1989)

[3] Meyer, R.M., Crawford, S.L., Lareau, J.P., Anderson, M.T.: Review of literature for model assisted probability of detection (2014)

[4] Thompson, R.B., Brasche, L.J., Lindgren, E., Swindell, P., Winfree, W.P.: Recent advances in model-assisted probability of detection. In: 4th European-American Workshop on Reliability of NDE (2009)

[5] Meeker, W.Q.: Rb thompson's contributions to model assisted probability of detection. In: AIP Conference Proceedings, vol. 1430, pp. 83–94 (2012). American Institute of Physics

[6] Jenson, F., Iakovleva, E., Dominguez, N.: Simulation supported pod: methodology and hfet validation case. In: AIP Conference Proceedings, vol. 1335, pp. 1573–1580 (2011). American Institute of Physics

[7] Ribay, G., Artusi, X., Jenson, F., Reece, C., Lhuillier, P.-E.: Model-based pod study of manual ultrasound inspection and sensitivity analysis using metamodel. In: AIP Conference Proceedings, vol. 1706 (2016). AIP Publishing

[8] Aldrin, J.C., Knopp, J.S., Sabbagh, H.A.: Bayesian methods in probability of detection estimation and model-assisted probability of detection evaluation. In: AIP Conference Proceedings, vol. 1511, pp. 1733–1740 (2013). American Institute of Physics

[9] Aldrin, J.C., Knopp, J.S., Lindgren, E.A., Jata, K.V.: Model-assisted probability of detection evaluation for eddy current inspection of fastener sites. In: AIP Conference Proceedings, vol. 1096, pp. 1784–1791 (2009). American Institute of Physics

[10] Li, M., Meeker, W.Q., Hovey, P.: Using a bayesian model to jointly estimate the flaw size distribution and the pod function. In: AIP Conference Proceedings, vol. 1211, pp. 687–694 (2010). American Institute of Physics

[11] Carboni, M., Cantini, S., *et al.*: A model assisted probability of detection approach for ultrasonic inspection of railway axles. In: 18th World Conference on Nondestructive Testing, pp. 16–20 (2012)

[12] Jenson, F., Dominguez, N., Willaume, P., Yalamas, T.: A bayesian approach

for the determination of pod curves from empirical data merged with simulation results. In: AIP Conference Proceedings, vol. 1511, pp. 1741–1748 (2013). American Institute of Physics

[13] Buethe, I., Dominguez, N., Jung, H., Fritzen, C.-P., Ségur, D., Reverdy, F.: Path-based mapod using numerical simulations. In: Smart Intelligent Aircraft Structures (SARISTU) Proceedings of the Final Project Conference, pp. 631–642 (2016). Springer

[14] Le Gratiet, L., Iooss, B., Blatman, G., Browne, T., Cordeiro, S., Goursaud, B.: Model assisted probability of detection curves: New statistical tools and progressive methodology. Journal of Nondestructive Evaluation **36**(1), 8 (2017)

[15] Du, X., Yan, J., Laflamme, S., Leifsson, L., Tesfahunegn, Y., Koziel, S.: Model-assisted probability of detection for structural health monitoring of flat plates. In: Computational Science–ICCS 2018: 18th International Conference, Wuxi, China, June 11-13, 2018, Proceedings, Part II 18, pp. 618–628 (2018). Springer

[16] Sun, K., Tao, Y., Tao, X., He, Y., Ye, C.: Model-based pod evaluation with parameters from experiment: A comparative study of tmr and coil array probes. Journal of Nondestructive Evaluation **41**(1), 27 (2022)

[17] Savli, E., Tschöke, K., Neubeck, R., Schubert, L.: A model-assisted approach to sensor network design in guided wave based shm systems. In: European Workshop on Structural Health Monitoring, pp. 945–952 (2022). Springer

[18] Lei, X., Wirdelius, H., Rosell, A.: Simulation-based investigation of a probability of detection (pod) model using phased array ultrasonic testing (paut) technique. Journal of Nondestructive Evaluation **41**(2), 40 (2022)

[19] Harding, C., Hugo, G., Bowles, S.: Application of model-assisted pod using a transfer function approach. In: AIP Conference Proceedings, vol. 1096, pp. 1792–1799 (2009). AIP Publishing

[20] Demeyer, S., Jenson, F., Dominguez, N., Iakovleva, E.: Transfer function approach based on simulation results for the determination of pod curves. In: AIP Conference Proceedings, vol. 1430, pp. 1757–1764 (2012). American Institute of Physics

[21] Bode, M.D., Newcomer, J., Fitchett, S.: Transfer function model-assisted probability of detection for lap joint multi site damage detection. In: AIP Conference Proceedings, vol. 1430, pp. 1749–1756 (2012). American Institute of Physics

[22] Rosell, A., Persson, G.: Model based capability assessment of an automated eddy current inspection procedure on flat surfaces. Research in nondestructive evaluation **24**(3), 154–176 (2013)

[23] Koshti, A.M.: Assessment of flaw detectability using transfer function. In: Nondestructive Characterization and Monitoring of Advanced Materials, Aerospace, Civil Infrastructure, and Transportation XV, vol. 11592, pp. 153–166 (2021). SPIE

[24] Koshti, A.M.: Transfer function models for using empirical and physics-based simulation signal response data. In: Health Monitoring of Structural and Biological Systems XVIII, vol. 12951, pp. 496–510 (2024). SPIE

[25] Koshti, A., Parker, P.A., Forsyth, D.S., Suits, M.W., Walker, J.L., Prosser, W.H.: Guidebook for limited sample probability of detection (ls-pod) demonstration for signal-response nondestructive evaluation (nde) methods. Technical report (2021)

[26] Koshti, A.M.: Nde flaw detectability validation using smaller number of signal response data-points. In: Nondestructive Characterization and Monitoring of Advanced Materials, Aerospace, Civil Infrastructure, and Transportation XIII, vol. 10971, pp. 148–160 (2019). SPIE

[27] Koshti, A.M.: Using requirements on merit ratios for assessing reliability of nde flaw detection. In: Health Monitoring of Structural and Biological Systems XV, vol. 11593, pp. 370–384 (2021). SPIE

[28] Koshti, A.M.: Assessing risk due to small sample size in probability of detection analysis using tolerance intervals. In: Health Monitoring of Structural and Biological Systems XVI, vol. 12048, pp. 373–384 (2022). SPIE

[29] Speed, R.: Regression type techniques and small samples: A guide to good practice. Journal of Marketing Management **10**(1-3), 89–104 (1994)

[30] Morgan, C.J.: Use of proper statistical techniques for research studies with small samples. American Journal of Physiology-Lung Cellular and Molecular Physiology **313**(5), 873–877 (2017)

[31] Falcetelli, F., Yue, N., Di Sante, R., Zarouchas, D.: Probability of detection, localization, and sizing: The evolution of reliability metrics in structural health monitoring. Structural Health Monitoring **21**(6), 2990–3017 (2022)

[32] Knott, C.E.: Advanced statistical methodology for the modern probability of detection (2023)

[33] Stapleton, J.H.: Linear Statistical Models vol. 719. John Wiley & Sons (2009)

[34] Montgomery, D.C., Peck, E.A., Vining, G.G.: Introduction to Linear Regression Analysis, John Wiley & Sons (2021)

[35] Long, J.S., Trivedi, P.K.: Some specification tests for the linear regression model. Sociological Methods & Research **21**(2), 161–204 (1992)

[36] Kim, M.G.: Sources of high leverage in linear regression model. arXiv preprint arXiv:2006.04024 (2020)

[37] Cook, R.D.: Detection of influential observation in linear regression. Technometrics **19**(1), 15–18 (1977)

[38] Rocke, D.M., Woodruff, D.L.: Identification of outliers in multivariate data. Journal of the American Statistical Association **91**(435), 1047–1061 (1996)

[39] Cunia, T.: Weighted least squares method and construction of volume tables. Forest Science **10**(2), 180–191 (1964)

[40] Weeks, M.: Introductory econometrics: a modern approach. JSTOR (2002)

[41] Ruppert, D.: Robust statistics: The approach based on influence functions. Taylor & Francis (1987)

[42] Knaub Jr, J.R.: Alternative to the iterated reweighted least squares method: Apparent heteroscedasticity and linear regression model sampling. In: Proceedings of the International Conference on Establishment Surveys, pp. 520–525 (1993)

[43] Carlin, B.P., Louis, T.A.: Bayesian Methods for Data Analysis, CRC press (2008)

[44] Efron, B., Tibshirani, R.J.: An introduction to the bootstrap, chapman & hall/crc. New York, NY (1993)

[45] Davison, A.C., Hinkley, D.V.: Bootstrap Methods and Their Application vol. 1, Cambridge university press (1997)

[46] Mohseni, E., Habibzadeh Boukani, H., Ramos França, D., Viens, M.: A study of the automated eddy current detection of cracks in steel plates. Journal of nondestructive evaluation **39**(1), 6 (2020)