Predictive probability of detection curves based on data from undamaged structures

Alexander Mendler, Michael Döhler, Christian U Grosse

Journal Title
XX(X):1–14

DOI: 10.1177/ToBeAssigned
www.sagepub.com/
SAGE

Predictive probability of detection curves based on data from undamaged structures

Alexander Mendler[1], Michael Döhler[2], Christian U. Grosse[1]

Abstract

This paper develops a model-assisted approach for determining predictive probability of detection (P-POD) curves. The approach is 'model-assisted,' as the damage-sensitive features are evaluated in combination with a numerical model of the examined structure. It is 'predictive' in the sense that probability of detection (POD) curves can be constructed based on measurement records from the undamaged structure, avoiding any destructive tests. The approach can be applied to a wide range of damage-sensitive features in structural health monitoring and non-destructive testing, provided the statistical distribution of the features can be approximated by a normal distribution. In particular, it is suitable for global vibration-based features, such as modal parameters, and evaluates changes in local structural components, e.g., changes in material properties, cross-sectional values, prestressing forces, and support conditions. The approach explicitly considers the statistical uncertainties of the features due to measurement noise, unknown excitation or other noise sources. Moreover, through confidence intervals, it considers model-based uncertainties due to uncertain structural parameters and a possible mismatch between the modelled and the real structure. Experimental studies based on a laboratory beam structure demonstrate that the approach can predict the probability of detection before damage occurs. Ultimately, several ways to utilize predictive POD curves are discussed, e.g., for the evaluation of the most suitable measurement equipment, for quality control, for feature selection, or sensor placement optimization.

Keywords

Structural health monitoring, damage detection, global features, probability of detection, confidence interval

Introduction

Probability of detection (POD) curves are commonly used to assess the performance of non-destructive testing (NDT) methods, with applications in safety-critical systems, such as aircraft and spacecraft, high-speed railway systems[1], military applications[2], or nuclear power plants[3]. The POD is defined as the probability of detecting damage, given that it is present. Since the measurement signals are typically scattered, the POD can be understood as the relative number of results beyond a prescribed threshold, cf. the black area in Fig. 1 (left). It is a function of the flaw size and a representative result is written in the following format: the 90/95 POD is 2 mm. That means that the minimum detectable crack is 2 mm, where the required POD is 90%. This statement is subject to uncertainty in the estimation process of the POD curve, and the reliability is given with a confidence of 95%, Fig. 1 (right). For the construction of POD curves, measurements from damaged specimens are usually required, and multiple destructive tests have to be performed, regardless of the employed POD method. However, this data is typically not available for large or expensive engineering systems, such as bridges, dams, and power plants, and this is why POD curves are barely used in structural health monitoring (SHM). The goal of the paper is to develop a method, appropriate for SHM or NDT, that can "predict" POD curves based on measurements from the undamaged specimen and a numerical model of the examined structure. Before doing that, the existing methods are critically reviewed.

NDT methods

The 29/29 method[4] was among the first attempts to quantify the POD. It counts the number of damages detected n and divides it by the total number of inspected damages n_{tot} to obtain an estimate for the POD = n/n_{tot}. Such binary approaches are appropriate for crack detection, fluorescent

[1]Technical University of Munich, TUM School of Engineering and Design, Department of Materials Engineering, Chair of Non-destructive Testing, Munich, Germany
[2] Univ. Gustave Eiffel, Inria, COSYS-SII, I4S, Campus de Beaulieu, 35042 Rennes, France

Corresponding author:
Alexander Mendler, Technical University of Munich, Franz-Langinger-Str. 10, 81245 Munich, Germany.
Email: alexander.mendler@tum.de

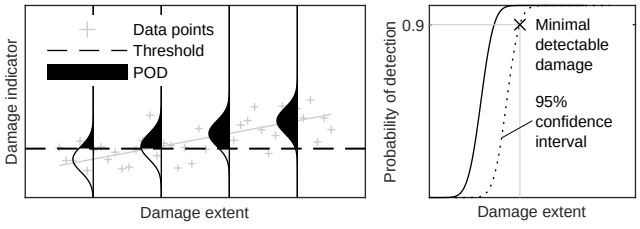

Figure 1. Creating a POD curve by counting the relative number of tests beyond a safety threshold for varying damage extents

penetrant testing, magnetic particle testing, and ultrasonic testing. They require a clear criterion to distinguish faults, but the actual magnitude of the response signal is not important once it exceeds the safety threshold. The approach can be explained at the opening example of the $90/95$ POD: the 90% POD is given if more than 9/10 flaws are found, but the confidence level of 95% is reached once 29/29 specimens are tested and all damages are found. Later, such binomial approaches were considered inadequate, as the POD changes with flaw size.

The hit/miss method[5] considers the flaw size using logarithmic regression. The development led to seminal works that remained the industry standard for many years[6]. Initially, the confidence intervals were determined using the Wald statistic, but shortly after, the likelihood ratio took its place, as it appeared to deliver more accurate, and less conservative results[2,7]. The data from hit/miss analyses contains little information regarding the correlation between crack size and damage indicator, or external factors such as the variance of the inspector's judgment. Hence, many data sets from damaged specimens are needed for the POD determination and the confidence interval is typically wider than for other methods.

The $\hat{a}$ vs. a method[8] is a suitable POD method for NDT techniques that quantify the damage extent through a more pronounced signal response. Typical examples include eddy current methods, infrared measurements, or radar-based approaches. The underlying assumption is that the apparent damage size $\hat{a}$ (the signal of the measurement device or the extracted damage indicators) and the physical damage size a are linearly related, or can be linearized by plotting $\hat{a}$ against a on semi-log or log-log paper. As for the hit/miss method, linear regression is used in combination with the maximum likelihood operator to evaluate the mean curve as well as the confidence intervals, which describe the uncertainty in the regression model.

SHM methods

Virkkunen et al.[9] outline that the $\hat{a}$ vs. a may be more suited for automated monitoring applications but further enhancements are necessary for SHM applications. The major differences of SHM, in comparison to NDT, are the fixed sensor layout, repeated measurements with correlated diagnosis results, and time-dependent effects[10]. This disqualifies all previously discussed methods for the generation of POD curves. Originally developed by Schubert Kabban et al.[11], the linear-mixed-effects method can deal with correlated data. They replaced the ordinary least square estimator and the maximum likelihood estimator with generalized square models, which can handle dependencies in measurement results. Kessler et al.[12] improved the model by assuming not only a random intercept but also a random slope and called it the random effects model.

The length-at-detection (LAD) method, developed by Roach et al.[13], is another suitable method to evaluate crack growth for permanently installed sensors and repeated inspections. This method only uses the crack length values when cracks are first detected. A crack is detected if a clear and stable signal response is produced, and the remaining task is to determine the probability density function of the crack length at detection, e.g., using the normal or log-normal distribution. The resulting cumulative density function is then identical to the POD, and confidence intervals can be computed based on standard statistical methods. Representative applications are documented by Falcetelli et al.[14] who applied the LAD method in combination with fiber-optical strain measurements.

Model-assisted methods

Another group of methods, suitable for NDT or SHM, are based on numerical models and they are typically referred to as model-assisted POD (MAPOD) methods. Their purpose is to synthetically generate measurement data, which are evaluated in combination with real laboratory data, and to obtain reliable confidence intervals with a minimum number of data sets. Thompson et al.[15] summarize the first group of methods as transfer function-based approaches. The main idea is to determine a baseline POD curve in a fully empirical laboratory experiment. Then, a physics-based model of the inspection process is employed so the POD curves can be transferred to a similar inspection process on a different component or material. Applications of this method are documented by Harding et al.[16] and Demeyer et al.[17], who applied the approach for ultrasonic fatigue crack detection in aeroplane wings and aluminum plates.

Thompson et al.[15] juxtapose those methods with so-called fully model-assisted methods. The models predict the signal response based on the examined component, material, and flaw type with well-understood damage mechanisms, and superimpose it with various noise sources, e.g., due to different measurement systems or operators. As shown by Thompson et al.[15], transfer function-based methods and fully model-assisted methods can be incorporated into a common framework. Both approaches were applied by Carboni and Cantini[18], who performed ultrasonic tests on railway axles, and Rosell and Persson[19], who examined fatigue cracks in titanium plates.

Metamodels, also known as surrogate models, are simplified models of the physics-based model and can also be used for data generation and POD estimation. Foucher et al.[20,21] implemented this in software packages and gave an overview of existing methods. Spencer[22] employed meta-models to derive non-parametric POD curves, and Dominguez et al.[23] developed an approach to determine POD curves and confidence intervals for generic structures such as beams.

Bayesian methods

Bayesian methods are convenient because they inherently include information on the confidence bounds, they can treat multiple measured signal responses and damage phenomena at the same time, and may further reduce the number of required data sets, as the prior knowledge on the distribution can be considered (engineering judgment or model-assisted information). In general, Bayesian models can be combined with purely data-driven methods or model-assisted ones. Knopp et al.[24], for example, used a data-driven approach and combined the hit/miss method with Bayesian models. They concluded that the confidence interval can be defined more accurately in comparison to non-Bayesian models and this was also the driving factor behind other

studies by Abdessalem et al.[25]. Aldrin et al.[26] presented a Bayesian model for estimating POD curves for eddy-current testing that mixed real measurement data with synthetically generated ones. Jenson et al.[27] supplement experimental data from eddy current-based fatigue crack inspections with synthetically generated data and applied Bayesian methods. Hovey[28] used Bayesian models to simultaneously identify the POD curve and the crack size distribution.

Problem statement

The goal of this paper is to develop a method, suitable for NDT and SHM, to predict the POD based on data from the undamaged state. The literature review shows that most of the existing methods are not suitable for this task, as they require a substantial number of data sets from damaged structures. The hit/miss method requires the largest number of data sets of 60[2] to 90[29], and conservative analysts[24] propose up to 300 data sets for an accurate estimation of confidence intervals. In contrast, the $\hat{a}$ vs. a method requires fewer data sets of 30 to 40, as the magnitude of the signal response is considered. If Bayesian methods are employed, the confidence interval can be determined with even fewer data sets, and model-informed methods postulate enhancing experimental data with synthetically generated data to reduce the required number of destructive tests to a minimum. In this sense, fully model-assisted methods are the most appropriate group of methods to develop POD curves without data from the damaged state. However, they require synthetically generated data sets, and correctly modelling the noise properties is challenging. In many cases, the synthetically generated measurements exhibit different properties than the real ones, and this introduces bias into the POD curves. Moreover, they do not offer a framework to evaluate several features at the same time and to link changes in local structural components to global damage-sensitive features, such as natural frequencies. This is why an appropriate POD method is developed in this paper that considers the real noise properties from non-destructive measurements, recorded on the undamaged specimen, and predicts the POD based on model-based sensitivity matrices that map changes onto local structural components.

The proposed method is based upon a statistical framework for change detection that considers the uncertainties in damage-sensitive features, and maps changes in the mean values onto local structural components using sensitivity vectors[30,31]. Recently, an approach was developed in this framework to analyse the detectability of damage for subspace-based features[32]. This paper also employs this framework, and extends the previous approaches to predict entire POD curves as continuous functions over hypothetical damages. Moreover, the approach is generalized from subspace-based features to arbitrary features in NDT and SHM—whose distribution can be approximated as Gaussian—and uncertainties in the physics-based model are considered through confidence intervals for the first time.

The paper is organized as follows: the first section recaps how damage can be detected based on statistical tests. The subsequent section develops a method to predict the test response to damage based on numerical models and measurement data from the undamaged structure, and to construct POD curves without any measurement data from the damaged structure. In the application section, the approach is applied to a laboratory beam structure for proof of concept. Ultimately, possible ways to utilize the predictive POD curves for the optimization of NDT and SHM systems are discussed, followed by the conclusions.

Statistical damage detection

This section recaps the statistical test the developed method is based upon. In general, the damage detection process is split into three basic steps: the acquisition of measurement data $\mathbf{Y}$ through sensors, the signal processing and extraction of damage-sensitive features $\mathbf{f}$, and their subsequent statistical evaluation[33] where the processed features are visually inspected or decision-making tools are applied to summarize the analysis in a scalar diagnostic test result t. To detect damage, the diagnostic can be compared against a safety threshold value t_{crit}, and an alarm is raised if the threshold is exceeded.

Damage assessment requires a comparison between the undamaged state and the tested state. The undamaged state is often divided into a *training* state, in which the statistical models are built, and a *validation* state to verify the model. Afterwards follows the *testing* state, as shown in Fig. 2. In this paper, it is assumed that the statistical distribution of the extracted feature can be approximated by a normal distribution in all states, which is the case for many features, and that the effects of environmental changes have already been removed. In the training state, the reference value of the feature vector $\mathbf{f}^0$ is determined based on some baseline model, i.e., a numerical model or as the mean value from measurements. During validation and testing, a damage-sensitive residual is formed by extracting the feature vector and subtracting the baseline vector from training $\mathbf{r} = \hat{\mathbf{f}} - \mathbf{f}^0$. Then, it is tested whether or not the residual approximates the distribution from the validation state and the significance of possible deviations is assessed. The individual steps are explained in detail in the following paragraphs.

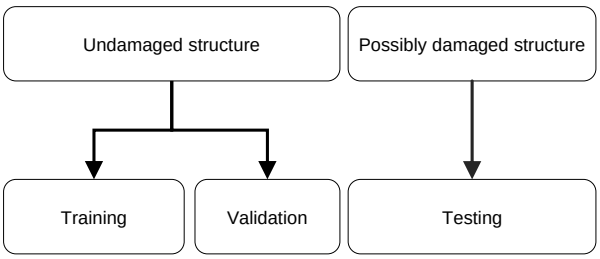

Figure 2. The three states of each monitoring campaign

Residual vector

It is assumed that the residual follows a normal distribution

$$\mathbf{r} = \hat{\mathbf{f}} - \mathbf{f}^0 \sim \mathcal{N}(\mathbf{m}, \mathbf{\Sigma}), \tag{1}$$

where $\mathbf{m} \in \mathbb{R}^{N_f}$ is the mean vector and $\mathbf{\Sigma} \in \mathbb{R}^{N_f \times N_f}$ is the covariance matrix. If no structural changes occurred, the mean vector is zero, i.e. $\mathbf{m} = \mathbf{0}$. Damage, on the other hand, manifests itself through changes in the estimated features $\hat{\mathbf{f}}$, leading to a shift in the mean value of the residual's probability density function. For problems with

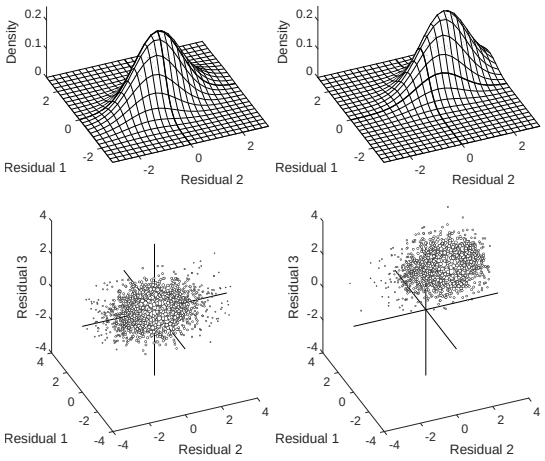

Figure 3. Distribution of the residual vector with two features (top plots) and three features (bottom plots), where damage manifests itself in a shift in the residual vector from the origin

two or three features, the statistical damage detection problem is visualized in Fig. 3. The covariance matrix describes the statistical uncertainties in the residual, e.g., due to measurement noise (noisy sensors, cables, and data acquisition systems), unknown excitation, or other epistemic uncertainties in the feature estimation process. It can, for example, be evaluated based on the sample covariance using n_b realizations from training data

$$\mathbf{\Sigma} = \frac{1}{n_b - 1} \sum_{k=1}^{n_b} \mathbf{r}_k \, \mathbf{r}_k^T. \tag{2}$$

Depending on the feature, other techniques such as the delta method[34] are available, which allows one to estimate the covariance based on a single measurement record.

Many features from SHM or NDT satisfy the Gaussian assumption (1). For example, features computed from vibration time histories are often asymptotically Gaussian distributed after proper normalization with the square root of the data length $\sqrt{N}$, due to the central limit theorem. Representative features include modal parameters from operational modal analysis[35,36] and experimental modal analysis[37], features based on covariance functions[38], or residuals that are formed in the subspace of covariance functions[39]. When environmental and operational variables disturb the Gaussian assumption, their effect on the feature vector $\mathbf{f}$ can be eliminated before the estimation of the covariance matrix, using well-known methods for data normalization such as multivariate linear regression, principal component analysis[40], or auto-associative neuronal networks[41]. In this sense, the removal of EOVs can be decoupled from the damage detection. Examples of normally distributed signals in NDT include eddy current methods, infrared measurements, or radar-based approaches. In POD curves based on the $\hat{a}$ vs. a method, the signal response from NDT devices is drawn on linear, semi-log or log-log paper, and it is assumed that they can be approximated by a normal distribution in one of these scales before applying regression methods.

Damage detection test

Statistical tests evaluate the significance of changes in the mean of the feature vector, and break all changes down into a scalar test value for change detection, which can be depicted in control charts, as in Fig. 4. A standard test statistic is

$$t = \mathbf{r}^T \mathbf{\Sigma}^{-1} \mathbf{r}, \tag{3}$$

which is known as the squared Mahalanobis distance or the Hotelling T^2-test. It can also be derived from the generalized likelihood ratio test. Its distribution can be approximated by a chi-squared distribution,

$$t \sim \chi^2(\nu, \lambda), \tag{4}$$

with the number of degrees of freedom

$$\nu = \text{rank}(\mathbf{\Sigma}), \tag{5}$$

and the non-centrality parameter λ. It should be noted that the inverse covariance matrix $\mathbf{\Sigma}^{-1}$ can be replaced with the pseudo-inverse if the matrix is rank-deficient. The non-centrality is zero in the undamaged state, and unequal to zero in the damaged state. It describes the shift in the mean value of the test, and therefore, it is called the "mean test response" from here onward.

The diagnostic test (4) is repeatedly evaluated for each data set during validation and testing, and plotted in control charts, with a representative example in Fig. 4. The figure shows the test for 50 observations before and 50 observations after system changes occurred, and draws the empirical test distribution in histograms. In the reference state, the mean value is identical to the number of degrees of freedom ν, and the more severe the structural changes, the more pronounced is the mean test response λ.

Safety threshold

A safety threshold t_{crit}, sometimes also called a control limit[42], helps to distinguish the "normal" system state from abnormal states, and an alarm is raised if the safety threshold is exceeded. Typically, the threshold is determined based on the test distribution in the validation state, based on one of the following metrics.

The false negative (FN) is an appropriate metric to define the safety threshold. It is defined as a significant test response beyond the safety threshold although no system changes have occurred (in the validation state). False negatives can cause an intact structure to be taken out of operation for no

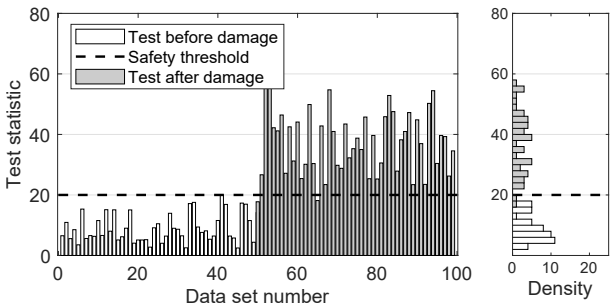

Figure 4. Empirical distribution of diagnostic test before and after a damage event

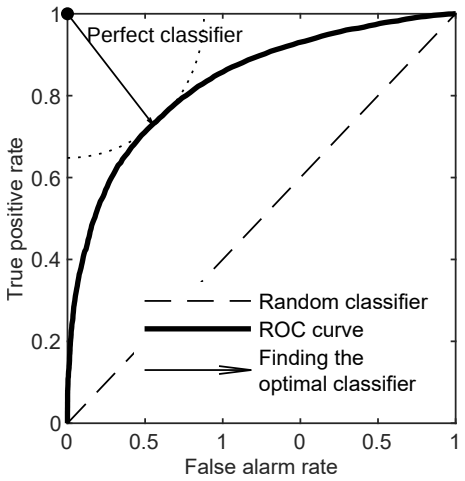

Figure 5. Receiver operating curves: finding the optimal safety threshold (classifier) as the one that maximizes the POD and minimizes false alarms

safety-relevant reasons, so it is desirable to minimize them by setting the safety threshold sufficiently high. An appropriate false negative rate depends on the time scale of the diagnosis results. If multiple diagnostics (3) are generated within a short period of time (e.g. in one second), a higher false alarm rate may be justifiable, as the majority of diagnostics indicate that no system changes occurred. A monitoring system that yields a diagnostic value every hour, day, month, or year should minimize false negatives, as the course of action (and possible negative consequences for the structure) depend on a single diagnostic. In this paper, a false alarm rate of 1% is chosen, meaning on average one out of 100 tests is a false alarm.

The true positive (TP) rate is another important quantity that depends on the safety threshold. It quantifies the relative number of test responses beyond the safety threshold after damage occurred (in the testing state). The true positive rate is equivalent to the probability of detecting damage.

Receiver operating curves (ROC) are an appropriate tool to determine a safety threshold if data is available from the damaged structure[43]. The optimal threshold is chosen as the one that minimizes the false alarm rate while maximizing the true positive rate. Graphically speaking, the optimal point can be found by plotting the true positive rate over the false positive rate for different thresholds, and by choosing the point with the smallest Euclidean distance to the perfect classifier, where the TP is 100% and the FN rate is 0% (i.e., the point in the top left of Fig. 5).

Predictive POD curves

This section introduces an approach for constructing predictive probability of detection curves based on non-destructive data from the examined system and a parametrized numerical model.

Parametric system model

Numerical models are suitable tools to physically interpret "damage," for example, based on changes in structural parameters. It is assumed that the feature vector $\hat{\mathbf{f}}$ can be modelled as

$$\hat{\mathbf{f}} = h(\boldsymbol{\theta}) + \boldsymbol{\varepsilon} \tag{6}$$

where h is the (non-linear) function that maps structural parameters $\boldsymbol{\theta}$ onto the feature vector, and $\boldsymbol{\varepsilon}$ is a noise term that can be characterized based on data (2). Damage is defined as a relative change in the structural parameters from their reference state $\boldsymbol{\theta}^0$ (material properties, cross-sectional values, etc.), with

$$\boldsymbol{\Delta\theta} = \boldsymbol{\theta} - \boldsymbol{\theta}^0. \tag{7}$$

Using Taylor series expansion, the mapping function $h(\boldsymbol{\theta})$ (6) can be linearized, and changes in features can be expressed as

$$h(\boldsymbol{\theta}) \approx \mathbf{f}^0 + \boldsymbol{\mathcal{J}}\boldsymbol{\Delta\theta}, \tag{8}$$

where $\boldsymbol{\mathcal{J}}$ is the first-order derivative of data-driven features with respect to structural parameters

$$\boldsymbol{\mathcal{J}} = \left.\frac{\partial h}{\partial \boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}^0} = \begin{bmatrix}\boldsymbol{\mathcal{J}}_1 & \boldsymbol{\mathcal{J}}_2 & \dots & \boldsymbol{\mathcal{J}}_{N_p}\end{bmatrix}. \tag{9}$$

Each column $\boldsymbol{\mathcal{J}}_h$ in the Jacobian describes the rate of change of the feature vector $\mathbf{f}$ for the single parameter change $\Delta\theta_h$. The linearization (8) allows one to "predict" the changes in features based on hypothetical damage scenarios, without having to generate measurements and estimate features in the damaged state. This simplification is justified for small structural changes, which are the primary focus when constructing POD curves and for applications, where the mapping function from features to parameters $h(\boldsymbol{\theta})$ exhibits moderate non-linearities.

Predicted test response

In contrast to the traditional damage detection approach, where the test distribution is evaluated empirically (as in Fig. 4), the presented approach is based on the theoretical distribution of the diagnostic test (3), rendering empirical evaluations unnecessary. As explained above, the test's distribution can be approximated by a chi-squared distribution $\chi^2(\nu, \lambda)$ with a mean value ν in the undamaged state and a mean test response λ. Let us assume that the analyzed damage scenario manifests itself through a change in a single parameter $\boldsymbol{\theta}_h$. Then, the mean test response can be calculated analytically[32] as

$$\lambda = (\boldsymbol{\theta}_h - \boldsymbol{\theta}_h^0)^2 F_{hh}, \tag{10}$$

where F_{hh} is the Fisher information

$$F_{hh} = \boldsymbol{\mathcal{J}}_h^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\mathcal{J}}_h, \tag{11}$$

and $\boldsymbol{\mathcal{J}}_h$ is the sensitivity vector that corresponds to the analyzed parameter $\boldsymbol{\theta}_h$. Note that all quantities in Eq. (10) are known from the training state, so empirical data from the validation or testing state is not required. In principle, multiple parameter changes can be analyzed simultaneously with the presented framework, but the simplification of a single parameter change is made here to reduce the complexity of the mathematical expressions.

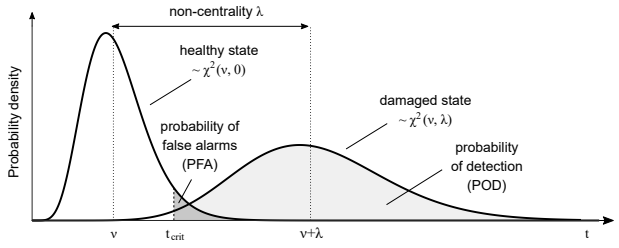

Figure 6. Theoretical distribution of the diagnostic test before and after a damage event

Probability of detection

The traditional approach quantifies the probability of detection (POD) as the relative number of diagnostic tests beyond the safety threshold when damage has occurred, see Fig. 4. In contrast, the POD is now quantified based on the theoretical distributions of the diagnostic test, as the area under the probability density function beyond the safety threshold

$$\text{POD} = \int_{t_{crit}}^{\infty} f_{\chi^2}(\nu, \lambda)(t)dt, \tag{12}$$

i.e. the grey area in Fig. 6, where $f_{\chi^2}(\nu, \lambda)$ is the probability density function of the chi-squared distribution. The POD is the complement of the cumulative distribution function, evaluated at the safety threshold $\text{POD} = 1 - F_{\chi^2}(\nu, \lambda)(t_{crit})$, which is uniquely defined through the number of degrees of freedom ν (5) and the mean test response λ. In other words, the POD depends on the mean test response λ, and the mean test response depends on the parameter change due to damage $\boldsymbol{\theta}_h - \boldsymbol{\theta}_h^0$ (10), so there is a mathematical relation between the POD and the parameter change.

The developed framework creates a mathematical relation between the POD and the parameter change, and the most intuitive way to visualize this relation is by plotting it in so-called predictive probability of detection (P-POD) curves, see Fig. 13. The P-POD curves are drawn over a change in single structural parameters θ_h based on Eq. (10) and (12), and the location of this parameter is known from the numerical model. That means that the POD can be evaluated for local structural components, even if global features (such as modal parameters) are evaluated.

Confidence intervals

Due to a possible mismatch between the real and the modelled structure, numerical models are an additional source of uncertainty for P-POD curves, which can be considered through confidence intervals. The predictive formula (10) contains the Fisher information (11), which in turn considers the statistical uncertainties in the estimation of the features through the covariance matrix $\boldsymbol{\Sigma}$. Model-based uncertainties would affect the Jacobian matrix $\boldsymbol{\mathcal{J}}$, but are not considered yet. A straightforward way to derive confidence intervals is to calculate the Jacobian matrix based on probabilistic finite element models (P-FEM), and to describe the P-POD curves through the mean curve and, for example, the 95% confidence interval from Fig. 1. Therefore, the confidence interval is not determined through the accuracy of the regression techniques (as it is the case for other POD methods) but the uncertainties in the model.

Probabilistic finite element models are a powerful tool to capture model-based uncertainties, and in theory, all structural parameters that affect the Jacobian matrix should be treated as random variables. The randomness in one or multiple structural parameters can be represented through random fields. Besides point discretization methods (e.g., the midpoint method, the integration point method, and the average discretization point method), series expansion methods can be employed, such as the Karhuen-Loève expansion, the spectral representation method, and others, with a concise introduction in the referenced literature[44]. In addition to the field representation, an appropriate evaluation method has to be selected, where the most widely employed methods are the perturbation method[45], the spectral finite element method[46], and the Monte Carlo method[47].

For simplicity, an approach is chosen in this paper that can be reproduced by readers that work with deterministic FEM software without stochastic modelling capabilities. Only the monitored structural parameters are assumed to be random. Each parameter is modelled as a Gaussian random variable and no correlation between different parameters is assumed

$$\boldsymbol{\theta} \sim \mathcal{N}(\mu_\theta, \boldsymbol{\Sigma}_\theta), \tag{13}$$

where μ_θ is the mean value and $\boldsymbol{\Sigma}_\theta$ is the diagonal covariance matrix. A Monte Carlo approach is pursued, meaning the sensitivity calculation is performed multiple times with varying structural input parameters. First, a sample is drawn from the distribution (13), and the Jacobian is computed with the corresponding model, e.g., based on the finite difference method. After completing the computation, the analysis is rerun for many other samples, resulting in a suite of sensitivity matrices, which can be translated into a suite of POD curves based on Eq. (10) to (12). Finally, the mean curve and the confidence interval can be computed.

Laboratory Application

For proof of concept, the P-POD method is applied to a laboratory steel beam, where added masses take the role of damage. After elaborating on the predictive POD curve, a validation study is presented to demonstrate that the P-POD is close to the empirical POD after damage is applied.

Laboratory experiment

The considered structure is a hollow structural steel beam (HSS 152x51x4.78 mm) with a length of 4.1 m. The beam is pin-supported on both ends and the material is characterized by a modulus of elasticity of E = 210,000 MPa and a total mass of m = 56.8 kg, Fig. 7. The beam is located at the University of British Columbia and has already been described in the referenced literature[32,48]. The instrumentation includes eight equispaced accelerometers that measure the vibration in the vertical direction at a sampling rate of $f_s = 330$ Hz, and one shaker in the center of Segment 3 that injects white noise vertically into the beam. Each sensor has a weight of 1.28 kg, and the shaker exhibits a static mass of 3.6 kg and a moving mass of 360 g. The expected damage scenario is a change in mass in any of the nine beam segments. To detect mass changes, natural frequencies f_n of the first three modes of vibration

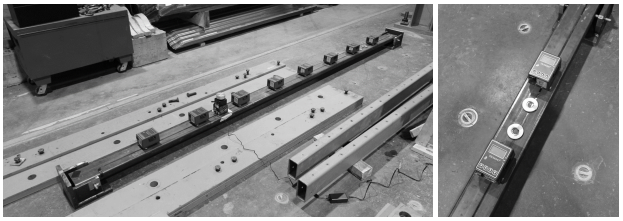

Figure 7. Laboratory HSS beam with eight sensors and one shaker (left) and extra masses applied (right)

are monitored, Fig. 9. They are estimated using covariance-driven operational modal analysis (SSI-Cov), and stored in a feature vector

$$\mathbf{f} = \begin{bmatrix} f_1 \\ f_2 \\ f_3 \end{bmatrix}. \tag{14}$$

Predictive POD curves

Before the POD curves can be constructed, a finite element model is built in ANSYS, so damage can be parametrized and sensitivities can be calculated. Note that the model is only used to extract numerical modal parameters for the sensitivity computation, but no vibration data is generated. To simulate different damage locations, a different mass parameter is assigned to each beam segment, and the parameter vector is defined as

$$\boldsymbol{\theta} = \begin{bmatrix} m_1 \\ \vdots \\ m_9 \end{bmatrix}. \tag{15}$$

Various methods exist to compute the derivative of modal parameters with respect to structural parameters[49,50], with the most straightforward one being the finite difference method. To compute the first-order derivatives, the baseline features $\mathbf{f}^0$ are created through numerical modal analysis of the undamaged structure. Then, as many finite element analysis runs are performed as there are structural parameters in $\boldsymbol{\theta}$. In each run, a single parameter θ_h is perturbed, and the resulting modal parameters $\mathbf{f}$ are extracted, so each Jacobian entry can be calculated as

$$\mathcal{J}_{ih} = \frac{f_i - f_i^0}{\theta_h - \theta_h^0}, \tag{16}$$

where $i = 1, \ldots, N_f$ is the number of features and $h = 1, \ldots, N_p$ is the number of parameters (15).

The covariance matrix is estimated based on real measurement data and describes the uncertainties related to the feature extraction process. As explained above, it can be estimated based on a single measurement record if the uncertainties are propagated through the feature estimation process (the operational modal analysis) using the delta method[34]. In this paper, a single measurement record of 60 s is recorded on the beam to evaluate the covariance matrix.

With the Jacobian matrix and the covariance matrix at hand, the Fisher information can be calculated using Eq. (11). Moreover, the theoretical distribution of the diagnostic test can be determined as a central chi-squared distribution $\chi^2(\nu)$ with $\nu = 9$ degrees of freedom (5). The safety threshold is set based on an allowable false-alarm rate of 1% to $t_{crit} = 18.2$, meaning one out of 100 tests is beyond the threshold even if the structure is undamaged. Ultimately, the sample size during testing is fixed to $N = 990$ (T = 3 s), and the POD curve is predicted using Eq. (10) to (12), where Fig. 10 shows the resulting P-POD curves for the two parameters θ_5 and θ_8. By looking at the figure, it can be appreciated that the POD curve is different for different structural parameters. The locations of the parameters are well-defined in the model, with θ_5 being at midspan and θ_8 closer to the right support, see Fig. 8. That means that the POD is evaluated for local structural components, even if it is applied to global vibration-based features.

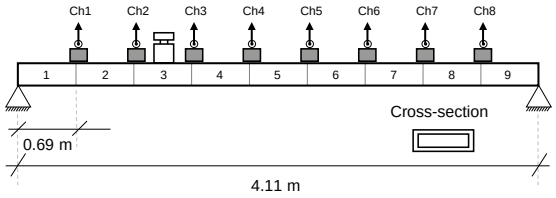

Figure 8. Laboratory beam with eight segments

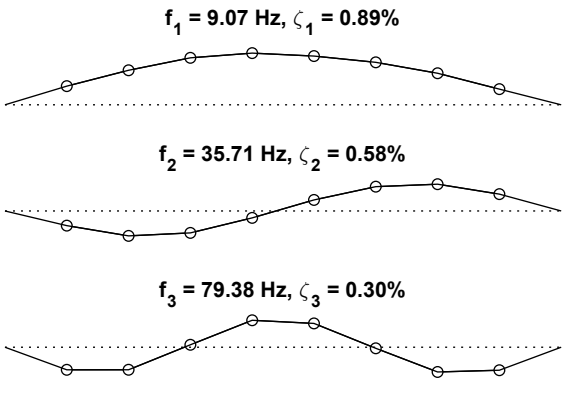

Figure 9. Modal parameters estimated based on OMA

Empirical validation

The results from the previous section are based on vibration data from the undamaged structure, with no extra masses applied to the beam. In this section, extra masses are applied to analyze whether the "predictive" POD are close to the empirical ones, and thus, to validate that the developed method is accurate. The main idea is to apply the mass, run the diagnostic test (3) one-hundred times, and draw the distribution of the test. If the prediction is correct, the empirical POD (i.e., the relative number of tests beyond the threshold) should be close to the analytical prediction using P-POD curves.

The first validation study sets out to verify that an increasing mass leads to a higher POD, see Fig. 11. The figure displays the empirical distributions of the test diagnostic for four different damage scenarios with extra masses on a beam segment close to the support, together with a χ^2-distribution that is fitted to the histograms. Increasing the mass from $\Delta\theta_8 = 2.5\%$ to 5.0%, leads to an increase in the POD from 28% to 57%, similar to how it was predicted based on the analytical P-POD curves, see Fig. 10. Further increasing the damage extent to 7.5% and 10% leads to

a POD of 96% and 100%, respectively. Consequently, an increasing damage extent leads to a higher POD, and the developed approach could accurately predict the POD for all considered mass changes.

The second study compares the POD for mass changes at different locations along the beam. Where an extra mass of $\Delta_8 = 2.5\%$ close to the support leads to a POD of 28%, the same mass leads to a POD of only 3% at midspan ($\Delta_5 = 2.5\%$), see Fig. 12. Regardless of the considered damage scenario (that is, an extra mass of 2.5%, 5.0%, 7.5% or 10%), a higher extra mass close to the support leads to a higher POD in comparison to the scenarios with masses at midspan, confirming the prediction of the P-POD curve. That means that the detectability of damage is lower at midspan, and it varies depending on the considered beam element. No empirical studies in the damaged state are necessary, because the information on the detectability of damage is already contained in the P-POD curve from Fig. 10, and they are only performed here for validation.

The prediction accuracy is indicated through the blue lines between the P-POD curve and the empirical POD in Fig. 10. In all cases, the empirical POD for masses close to the support (black squares) is higher than for masses at midspan (black dots), so the data points follow the same trends as the P-POD curves, where the prediction error ranges between 0.5 and 21%. The largest deviations occur for the smallest mass changes of 2.5%, where 155 g are added to a beam with a total beam mass of 65.8 kg, in which case the modelling errors may have the most significant effect. Other reasons for the deviation might be approximation errors due to limited data length, possibly leading to slightly non-Gaussian components in the feature distribution or a slightly changing covariance over different damage states. For real data these are acceptable errors, as the predictions are mostly more conservative than the empirical values. Related studies based on numerical simulation led to a prediction error below 0.5%, confirming the predictive qualities of the underlying statistical framework[32]. Therefore, the validation is concluded, and model-based uncertainties are discussed in the next section on confidence intervals.

Confidence intervals

To evaluate the confidence intervals, the sensitivity computation is repeated 500 times, while assuming a normal distribution for each mass parameter. The parameter distributions are assumed to be uncorrelated, with the mean vector μ_θ from Eq. (15), a coefficient of variation of 0.05 each, and a normal distribution of $\boldsymbol{\theta} \sim \mathcal{N}(\boldsymbol{\mu}_\theta, \mathrm{diag}(0.05 \cdot \boldsymbol{\mu}_\theta)^2)$. The 500 POD curves are drawn in Fig. 13 through grey solid lines. Assuming a symmetrical distribution about the mean POD curve, standard techniques can be applied to estimate the 95% confidence interval and to draw the lower bound (solid black line). By looking at the plot, it can be appreciated that the POD is unequal to zero for a parameter change of zero. This is because the safety threshold was set based on an allowable false negative rate of 1%, and the false negative rate is identical to the POD for a parameter change close to zero. Furthermore, the POD for a 2.5% parameter change varies between 7.3% and 17.7% for the individual runs, highlighting the importance of the confidence interval

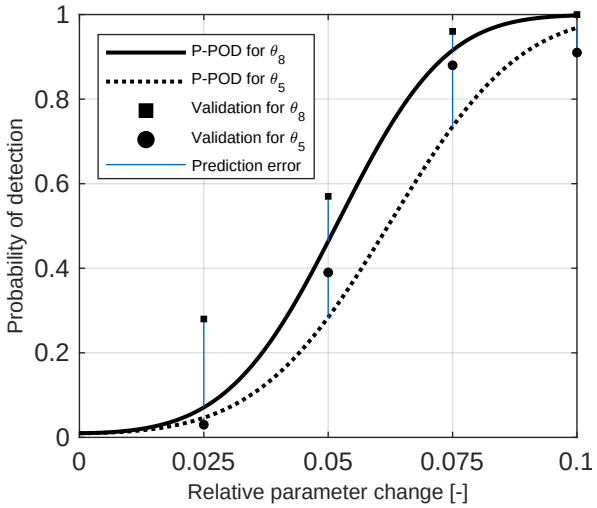

Figure 10. Predictive POD curve for changes in mass parameter θ_5 at midspan and θ_8 close to the support, together with empirical POD values for validation

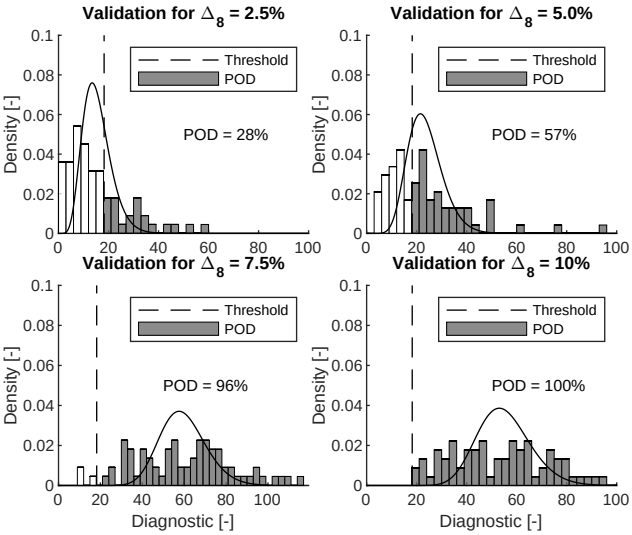

Figure 11. Validation of the empirical POD for mass changes close to the support (parameter θ_8)

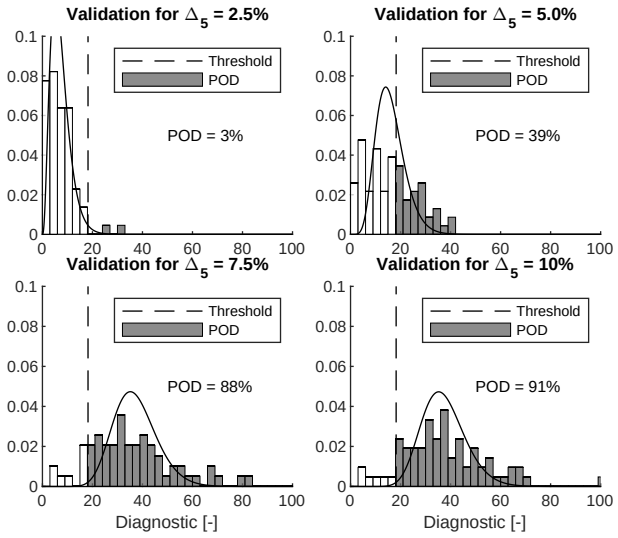

Figure 12. Validation of the empirical POD for mass changes at midspan (parameter θ_5)

and explaining the possible discrepancies between predicted and empirical POD in Fig. 11 and 12.

Discussion

The presented P-POD method exhibits distinct differences to traditional approaches, such as $\hat{a}$ vs. a. The method is based on a different statistical framework, underlying assumptions, and data requirements, which are critically discussed in this section. Moreover, the strengths and limitations of the developed P-POD method are elaborated on.

Method comparison

One of the fundamental differences between the P-POD and the $\hat{a}$ vs. a is definition of the POD and the number of damage-sensitive features that are evaluated. In the $\hat{a}$ vs. a, the POD is defined based on the Gaussian probability density function of each feature individually, which would be drawn on the y-axis of Fig. 1, and it is uncommon to evaluate multiple features at the same time. The developed P-POD method combines the information from all Gaussian features into a single diagnostic test value. It combines the features while considering their uncertainties, see Eq. (3), and the POD is defined based on the distribution of the diagnostic and not the features. Therefore, the P-POD method can evaluate changes in multiple features at the same time.

Due to these differences, a meaningful method comparison is not straightforward. Where the $\hat{a}$ vs. a assumes Gaussian distributions for each feature, the single diagnostic test value in the P-POD method exhibits a χ^2-distribution, as it squares up and scales multiple Gaussian features in Eq. (3). Consequently, the safety thresholds for each of the features in $\hat{a}$ vs. a cannot be directly compared to the threshold of the diagnostic in P-POD even for identical quantile values. However, both methods could be compared if the diagnostic (which evaluates and combines the features statistically) is considered as the feature itself for which the $\hat{a}$ vs. a curves are drawn. For this purpose, the diagnostic is drawn on logarithmic paper (y-axis in Fig 1), so the χ^2-distribution of the diagnostic approximates a normal

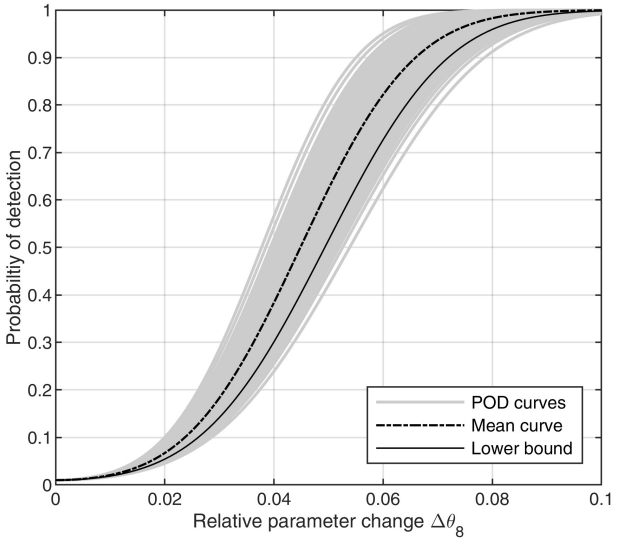

Figure 13. POD curve for mass changes close to the support (parameter θ_8) with confidence interval

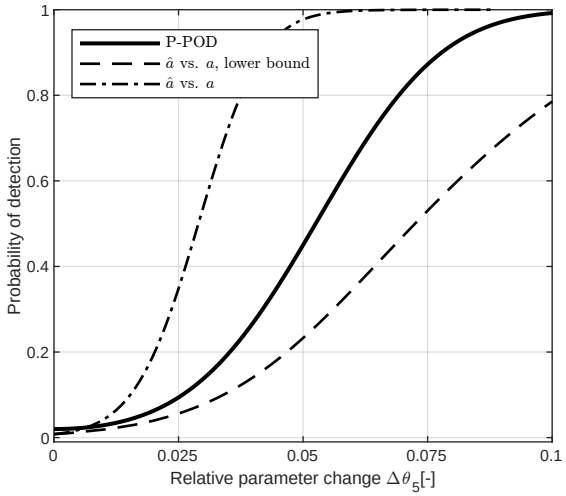

Figure 14. Comparing the P-POD to the $\hat{a}$ vs. a curve for changes in beam segment 5. Both curves are drawn for the diagnostic test from Eq. (3)

distribution. The results are shown in Fig. 14. The figure shows the $\hat{a}$ vs. a curve and its confidence bounds based on identical safety thresholds as for the P-POD method (the 99%-quantile value). The P-POD method leads to lower probabilities of detection than the $\hat{a}$ vs. a, but higher POD than the lower bound. This evokes another question, as to what curve the P-POD is to be compared to.

The confidence bound of the $\hat{a}$ vs. a describes both statistical uncertainty in the diagnostic as well as epistemic uncertainty in the regression model. The P-POD method, on the other hand, already explicitly considers the statistical uncertainty in the features in the POD computation based on the covariance matrix, see Eq. (3). An evaluation of confidence bounds that result from epistemic uncertainty is not necessary for the P-POD method, since the distribution properties can be described through theoretical investigations even for the damaged states, see Eq. (4), unlike in the $\hat{a}$ vs. a approach. The confidence bounds for P-POD curves, developed in this paper, relate to a different source of uncertainty, that is, modelling errors in the employed numerical model, whereas modelling uncertainty in MAPOD approaches is not reflected in $\hat{a}$ vs. a curves. Therein, a separation of the two sources of uncertainty is not possible either, and that is why the P-POD curve (without considering modelling errors) is comparable to the lower bound of the $\hat{a}$ vs. a. As a result, the P-POD leads to higher PODs than the $\hat{a}$ vs. a, see Fig. 14. This is because no (empirical) regression model is used but the theoretical distribution of the diagnostic is exploited. Epistemic uncertainty is only introduced if large parameter changes are analyzed, because then the first-order Taylor series expansion (9) loses accuracy. However, for such large parameter changes the POD is already close to 100%, so the computation of the POD is hardly affected. More information on this is given in the next section.

Model validation

The way the structural model is used in traditional MAPOD approaches is different than in the developed P-POD method. MAPOD approaches require the generation of many data

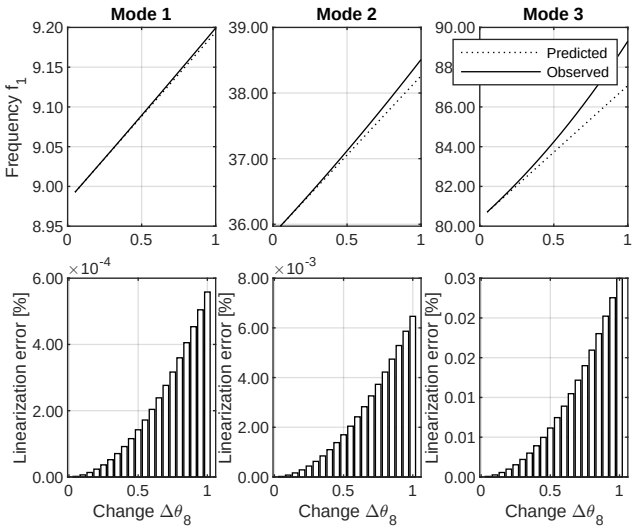

Figure 15. Linearization error from Eq. (17) for changes in θ_8

sets from different damage scenarios, whereas P-POD constructs POD curves based on data from the undamaged structure and sensitivity vectors, which are calculated based on the numerical model in the undamaged state alone. Hence, model validation with MAPOD approaches would be necessary for each damaged state, whereas in the P-POD approach, the validity of the sensitivity vectors (computed in the undamaged state) should be verified for the damaged states. Suppose that the numerical model is calibrated in the undamaged state, using standard techniques such as model updating based on sensitivity vectors[51], Bayesian approaches, interval, or Fuzzy approaches[52]. In the underlying statistical framework of sensitivity-based statistical tests[30], the sensitivity vectors are valid in the damaged state under the assumption of small parameter changes, since the relation between structural parameter changes and measured damage-sensitive features is linearized in Eq. (9). The small change assumption is relevant for the computation of the POD curve in P-POD methods; however, a structural change is in general small when the POD is smaller than 100%, so this is a reasonable assumption.

A simple way to assess the limits of the "small damage assumption" is the evaluation of the *linearization error*

$$\varepsilon_i(\Delta\theta_j) = \frac{f_i^{lin} - f_i}{f_i}, \tag{17}$$

which quantifies the relative deviation between the linearly predicted feature f_i^{lin} and the actual feature f_i as the output of a non-linear function. The subscripts i and j indicate that this error is evaluated for each feature and parameter individually, as shown in Fig. 15. For example, the maximum linearization error for a mass change at beam segment 8 is 3.0% for the third natural frequency, even for large parameter changes of Δm = 100%. That means that the linearization error is negligible in this specific case study.

Besides the linearization error, small changes in the covariance matrix between undamaged and damaged states could possibly contribute to deviations in the predicted POD. Hence, a more comprehensive way to assess the validity of the small damage assumption is the deviation between the predicted and the empirical distribution of the diagnostic over gradually increasing damage extents, which may lead to deviations in the predicted POD. For this purpose the *non-centrality error* (NCE) is introduced

$$\mathrm{NCE}(\Delta\theta_j) = \frac{\lambda - \lambda^{obs}}{\lambda^{obs}}, \tag{18}$$

where λ describes the predicted non-centrality (10) and λ^{obs} the observed non-centrality after evaluating the diagnostic test several times, in a Monte Carlo experiment. In contrast to the linearization error, the NCE allows one to assess the validity of the small damage assumption and possible covariance changes for parameter changes up to 100%, see Figure 16 (bottom plots). This plot clarifies that that the small damage assumption leads to significant errors for mass changes above 20%, with a NCE of 9.7%. However, as shown in Fig. 16 (top plot), the POD converges toward 100% for a mass change above 10%, so the errors are not relevant for the computation of the POD. In summary, the small damage assumption is reasonable for the critical domain of parameter changes, where the POD is smaller than 100%. Note that the presented NCE takes on the largest values for parameter θ_8 and smaller values have been observed for all other parameters.

Assumptions

For conciseness, the underlying assumption of the P-POD method are summarized in the following list:

- Gaussian features. It is assumed that the measured damage-sensitive features can be approximated through a normal distribution (1).
- Linearized models. The relation between damage-sensitive features and changes in structural parameters can be linearized using first-order Taylor series expansion (9). For strongly non-linear relations, this assumption limits the P-POD method to the analysis of small structural damages.
- Calibrated models. The numerical model used for the sensitivity computation should be sufficiently calibrated to ensure an accurate sensitivity computation.

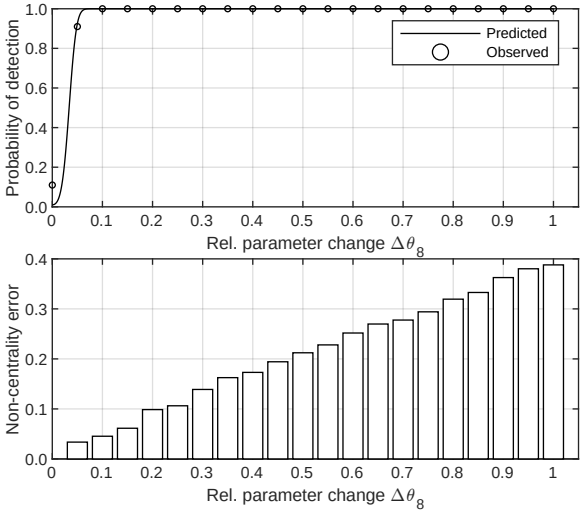

Figure 16. Non-centrality error from Eq. (18) for changes in θ_8

- Identical variance. Due to the small change assumption, the covariance of the damage-sensitive features is assumed to remain constant between undamaged and damaged states.

It should be noted that the Gaussianity and the linearity assumption do not hold for some features in SHM or NDT, but after their transformation on semi-log or log-log paper. For some features, e.g. in acoustic emission testing, the linearization may not be accurate enough even before the POD converges to 100%. Therefore, the assumptions should always be critically verified.

Fields of Application

This section discusses possible ways to utilize P-POD curves for the performance evaluation and optimization of NDT or SHM systems. For example, the P-POD can serve as a performance criterion for feature selection and sensor placement optimization, and can assist with choosing the most appropriate measurement equipment.

Feature selection

P-POD curves quantify the probability of detecting damages in model-based parameters, based on measured changes in damage-sensitive features. Therefore, they can be employed to compare different damage-sensitive features, and to select the one that is the most sensitive to small and critical structural changes. In the following, three different feature vectors are examined, including natural frequencies, mode shapes, and a combined vector,

$$\mathbf{f}_1 = \begin{bmatrix} f_1 \\ f_2 \\ f_3 \end{bmatrix}, \quad \mathbf{f}_2 = \begin{bmatrix} \mathbf{\Phi}_1 \\ \mathbf{\Phi}_2 \\ \mathbf{\Phi}_3 \end{bmatrix}, \quad \mathbf{f}_3 = \begin{bmatrix} f_1 \\ f_2 \\ f_3 \\ \mathbf{\Phi}_1 \\ \mathbf{\Phi}_2 \\ \mathbf{\Phi}_3 \end{bmatrix}. \tag{19}$$

Note that all available information should be used in practice ($\mathbf{f}_3$ in this case), and that this example is merely for demonstration. The corresponding mean curves for parameter θ_5 are shown in Fig. 17 (bottom). For small changes of $\Delta\theta_5 = 5\%$, the combined vector exhibits the highest POD of close to 100%, followed by the mode shapes and the frequencies in descending order. Looking at the lower bounds in Fig. 17 (top) reveals that features with higher damage sensitivity can also have wider confidence bounds, as the curves for mode shapes and frequencies are swapped. Consequently, the feature selection should always be based on the lower bound of the POD curve.

Another aspect that affects the POD is the number of monitored features. An increased number of features increases the available information on the structural state, provided that no redundant information is measured, so the sensitivity to small damages is increased. In this sense, the developed approach can be employed to decide on an appropriate number of features.

Sensor placement optimization

The POD is different for each sensor layout, so it may be an appropriate optimization criterion for sensor placement optimization. Optimization criteria are often categorized into three groups, i.e., methods for optimized data acquisition, optimized feature extraction, or optimized damage detectability[48]. Following this train of thought, POD curves can be categorized in the third group, as they optimize the detectability for the diagnostic test from Eq. (3). Typically, the sensor layout is determined before sensors are installed, so the optimization has to be done based on numerically generated data in the undamaged state. Obviously, the generated measurements should exhibit similar noise properties as experimental data to represent the true measurement conditions on the structure. This could either be ensured based on preliminary tests on the examined structure, using identical sensors as for long-term monitoring, or based on engineering judgment with similar structures. Then, the P-POD is a function of the sensor layout, and can be used as a performance criterion for the sensor placement optimization. Since the P-POD is drawn for individual structural parameters, a decision rule has to be implemented on how to merge the information from multiple curves, e.g. based on the least sensitive parameter.

Hardware noise specification

A powerful aspect of the developed POD method is that uncertainties due to measurement noise are explicitly considered. Lower measurement noise leads to a shift in

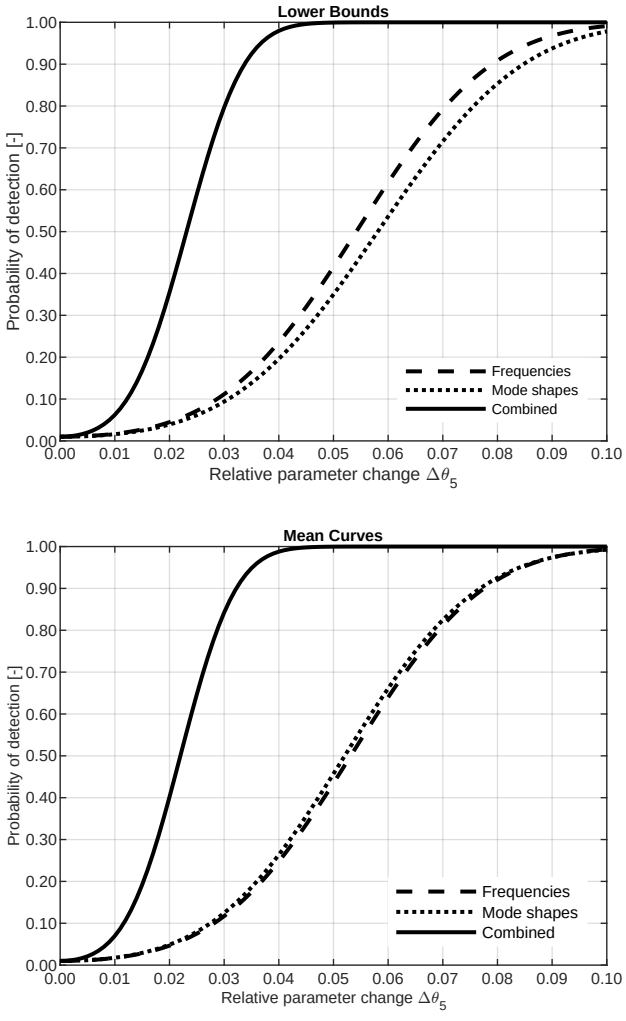

Figure 17. Feature selection based on POD curves

the POD curve to the left side, meaning the POD increases for identical parameter changes. Therefore, POD curves are an appropriate tool for the performance assessment of measurement equipment, as sensors and data-acquisition systems with different specifications lead to varying signal-to-noise ratios, and thus, different POD curves. This may assist in the selection of the most appropriate hardware as the one with the highest POD curves.

Quality control

POD curves offer a convenient way to verify the efficacy of the entire measurement chain, even after several years of operation. In this paper, the POD curves are validated by comparing the analytical to the empirical POD after extra masses have been applied. This verifies that the damage detection method is calibrated. Static load tests based on extra masses are a standard procedure to evaluate the load-bearing capacity of bridges, so it could be repeated in regular intervals to ensure that the entire measurement chain, including hardware and software components, is fully functional, even after several years of operation.

Conclusion

This paper introduces a model-assisted approach for determining predictive probability of detection (P-POD) curves. Traditional probability of detection methods require a minimum of 30 destructive tests with varying damage extents, which can lead to high experimental costs. In contrast, the developed approach requires no destructive tests but a single measurement record from the undamaged specimen. This can lead to significant economic benefits for applications where the costs for laboratory testing are greater than the costs for numerical modelling. Therefore, the approach is particularly suited for large, expensive, or unique engineering structures, where destructive tests are impossible, or structures where numerical models are readily available from the design phase.

In contrast to many existing POD methods, no regression models are employed to approximate the signal response over varying damage extents. Instead, the method employs statistical tests that combine the information from multiple feature changes into a single diagnostic, and this diagnostic explicitly considers the uncertainty in the data-driven features. Additional confidence intervals have been developed to account for modelling uncertainty, for example, due to a mismatch between the modelled and the real specimen, which are not considered in traditional approaches. Additional differences to existing methods, such as the $\hat{a}$ vs. a, as well as all underlying assumptions are elaborated on in the "Discussion" section.

Acknowledgements

The experiments to obtain the beam vibration data were conducted by the first author as part of his doctoral studies at the Earthquake Engineering Research Facility (EERF) of the University of British Columbia, Canada, under the direction of Dr. Carlos E. Ventura.

Funding

This research study is funded by dtec.bw - Digitalization and Technology Research Center of the Bundeswehr. Furthermore, the BayFrance seed fund of the Bavarian State Ministry of Science and Arts is gratefully acknowledged, as well as the French Ministery of Europe and Foreign Affairs.

References

1. Carboni M and Cantini S. Advanced ultrasonic probability of detection curves for designing in-service inspection intervals. *International Journal of Fatigue* 2016; 86: 77–87.
2. MIL-HDBK. Department of defense handbook, nondestructive evaluation system reliability assessment. Technical report, Department of Defense, United States, 2009.
3. Gandossi L and Annis C. Probability of detection curves: Statistical best-practices. Technical Report No. 41, European Network for Inspection and Qualification, 2010.
4. Rummel WD, Christner BK, Mullen SJ et al. Characterization of structural assessment testing. Technical report, SA-ALC/MMEI/1/86: January, 1986.
5. Berens AP and Hovey PW. Evaluation of NDE reliability characterization. Technical Report Volume I, University of Dayton Research Institute, Dayton, United States, 1981.
6. Berens AP. *NDE reliability data analysis–Metals handbook*. United States: ASM International, 1989.
7. ASTME-2862. Standard practice for probability of detection analysis for hit/miss data. Technical report, ASTM International, West Conshohocken, United States, 2018.
8. Berens A and Hovey P. Flaw detection reliability criteria. Volume 1. Methods and results. Technical report, University of Dayton University Research Institute, 1984.
9. Virkkunen I, Koskinen T, Papula S et al. Comparison of $\hat{a}$ versus a and hit/miss POD-estimation methods: A European viewpoint. *Journal of Nondestructive Evaluation* 2019; 38(4): 1–13.
10. Falcetelli F, Yue N, Di Sante R et al. Probability of detection, localization, and sizing: The evolution of reliability metrics in structural health monitoring. *Structural Health Monitoring* 2021; 21(6): 2990–3017.
11. Schubert Kabban CM, Greenwell BM, DeSimio MP et al. The probability of detection for structural health monitoring systems: Repeated measures data. *Structural Health Monitoring* 2015; 14(3): 252–264.
12. Kessler SS, Dunn CT, Swindell P et al. Detection sensitivity analysis for a potential drop (PD) structural health monitoring (SHM) system. In *Proceedings of the 12th International Workshop on Structural Health Monitoring*. Stanford, United States: DEStech Publications, Inc.
13. Roach DP, Delong WA, White S et al. Use of composite materials, health monitoring and self-healing concepts to refurbish our civil and military infrastructure. Technical Report SAND2007-5547, Sandia National Laboratories, Albuquerque and Livermore, United States, 2007.
14. Falcetelli F, Cristiani D, Yue N et al. Qualification of distributed optical fiber sensors using probability of detection curves for delamination in composite laminates. *Structural Health Monitoring* 2023; In press.
15. Thompson RB, Brasche LJ, Lindgren E et al. Recent advances in model-assisted probability of detection. In *4th European-American workshop on reliability of NDE*. LF99-9094, Berlin,

Germany.
16. Harding C, Hugo G and Bowles S. Application of model-assisted POD using a transfer function approach. In *AIP Conference Proceedings*, volume 1096. Chicago, United States: American Institute of Physics, pp. 1792–1799.
17. Demeyer S, Jenson F, Dominguez N et al. Transfer function approach based on simulation results for the determination of POD curves. In *AIP Conference Proceedings*, volume 1430. Burlington, United States, pp. 1757–1764.
18. Carboni M and Cantini S. A model assisted probability of detection approach for ultrasonic inspection of railway axles. In *18th World Conference on Nondestructive Testing*. Durban, South Africa, pp. 16–20.
19. Rosell A and Persson G. Model based capability assessment of an automated eddy current inspection procedure on flat surfaces. *Research in Nondestructive Evaluation* 2013; 24(3): 154–176.
20. Foucher F, Fernandez R, Leberre S et al. New tools in CIVA for model assisted probability of detection (MAPOD) to support NDE reliability studies. In *Nondestructive Evaluation of Aerospace Materials and Structures 2018*. St. Louis, United States, pp. 32–43.
21. Foucher F, Lonné S, Toullelan G et al. An overview of validation campaigns of the CIVA simulation software. In *12th European Conference on Non-destructive Testing*. Gothenburg, Sweden, pp. 11–15.
22. Spencer FW. Nonparametric POD estimation for hit/miss data: a goodness of fit comparison for parametric models. In *AIP Conference Proceedings*, volume 1335. San Diego, United States, pp. 1557–1564.
23. Dominguez N, Reboud C, Dubois A et al. A new approach of confidence in POD determination using simulation. In *AIP Conference Proceedings*, volume 1511. Denver, United States, pp. 1749–1756.
24. Knopp J, Grandhi R, Zeng L et al. Considerations for statistical analysis of nondestructive evaluation data: hit/miss analysis. *E-Journal of Advanced Maintenance* 2012; 4(3): 105–115.
25. Abdessalem AB, Jenson F and Calmon P. Quantifying uncertainty in parameter estimates of ultrasonic inspection system using Bayesian computational framework. *Mechanical Systems and Signal Processing* 2018; 109: 89–110.
26. Aldrin JC, Knopp JS and Sabbagh HA. Bayesian methods in probability of detection estimation and model-assisted probability of detection evaluation. In *AIP Conference Proceedings*, volume 1511. Denver, United States, pp. 1733–1740.
27. Jenson F, Dominguez N, Willaume P et al. A Bayesian approach for the determination of POD curves from empirical data merged with simulation results. In *AIP Conference Proceedings*, volume 1511. Denver, United States, pp. 1741–1748.
28. Hovey PW. Using the information in field service inspections to assess the damage state of the aircraft and the detection capability of the inspection system. In *AIP Conference Proceedings*, volume 1096. Chicago, United States, pp. 1855–1861.
29. Annis C, Gandossi L and Martin O. Optimal sample size for probability of detection curves. *Nuclear Engineering and Design* 2013; 262: 98–105.
30. Benveniste A, Basseville M and Moustakides G. The asymptotic local approach to change detection and model validation. *IEEE Transactions on Automatic Control* 1987; 32(7): 583–592.
31. Döhler M, Mevel L and Zhang Q. Fault detection, isolation and quantification from Gaussian residuals with application to structural damage diagnosis. *Annual Reviews in Control* 2016; 42: 244–256.
32. Mendler A, Döhler M and Ventura CE. A reliability-based approach to determine the minimum detectable damage for statistical damage detection. *Mechanical Systems and Signal Processing* 2021; 154: 107561.
33. Farrar C and Worden K. *Structural health monitoring: A machine learning perspective*. Oxford, United Kingdom: Wiley, 2012.
34. Casella G and Berger RL. *Statistical Inference*. Pacific Grove, California: Duxbury, 2002.
35. Reynders E, Pintelon R and De Roeck G. Uncertainty bounds on modal parameters obtained from stochastic subspace identification. *Mechanical Systems and Signal Processing* 2008; 22(4): 948–969.
36. Döhler M and Mevel L. Efficient multi-order uncertainty computation for stochastic subspace identification. *Mechanical Systems and Signal Processing* 2013; 38(2): 346–366.
37. Mellinger P, Döhler M and Mevel L. Variance estimation of modal parameters from output-only and input/output subspace-based system identification. *Journal of Sound and Vibration* 2016; 379: 1–27.
38. Greś S, Döhler M, Andersen P et al. Subspace-based Mahalanobis damage detection robust to changes in excitation covariance. *Structural Control and Health Monitoring* 2021; : e2760.
39. Balmès E, Basseville M, Bourquin F et al. Merging sensor data from multiple temperature scenarios for vibration-based monitoring of civil structures. *Structural Health Monitoring* 2008; 7(2): 129–142.
40. Magalhães F, Cunha Á and Caetano E. Vibration based structural health monitoring of an arch bridge: From automated OMA to damage detection. *Mechanical Systems and Signal Processing* 2012; 28: 212–228.
41. Sohn H, Worden K and Farrar CR. Statistical damage classification under changing environmental and operational conditions. *Journal of Intelligent Material Systems and Structures* 2002; 13(9): 561–574.
42. Montgomery DC. *Introduction to Statistical Quality Control*. 6th ed. Jefferson City, United States: John Wiley & Sons, 2007.
43. Schoefs F, Clément A and Nouy A. Assessment of ROC curves for inspection of random fields. *Structural Safety* 2009; 31(5): 409–419.
44. Papadopoulos V and Giovanis DG. Stochastic finite element methods. *An Introduction* 2018; .
45. Liu WK, Belytschko T and Mani A. Random field finite elements. *International Journal for Numerical Methods in Engineering* 1986; 23(10): 1831–1845.
46. Ghanem RG and Spanos PD. *Stochastic finite elements: a spectral approach*. Courier Corporation, 2003.
47. Astill C, Imosseir S and Shinozuka M. Impact loading on structures with random properties. *Journal of Structural Mechanics* 1972; 1(1): 63–77.
48. Mendler A, Döhler M and Ventura CE. Sensor placement with optimal damage detectability for statistical damage detection. *Mechanical Systems and Signal Processing* 2022; 170: 108767.

49. Nelson RB. Simplified calculation of eigenvector derivatives. *AIAA journal* 1976; 14(9): 1201–1205.
50. Heylen W and Sas P. *Modal analysis theory and testing*. Leuven, Belgium: Katholieke Universiteit Leuven, 1997.
51. Mottershead JE, Link M and Friswell MI. The sensitivity method in finite element model updating: A tutorial. *Mechanical Systems and Signal Processing* 2011; 25(7): 2275–2296.
52. Simoen E, de Roeck G and Lombaert G. Dealing with uncertainty in model updating for damage assessment: A review. *Mechanical Systems and Signal Processing* 2015; 56-57: 123–149.