A unifying review of NDE models towards optimal decision support

Elizabeth Bismut[a,*], Daniel Straub[a]

*[a]Engineering Risk Analysis Group, Technische Universität München.
Arcisstraße 21, 80290 München, Germany*

Abstract

Non-destructive evaluation (NDE) inspections are an integral part of asset integrity management. The relationship between the condition of interest and the quantity measured by NDE is described with probabilistic models such as PoD or ROC curves. These models are used to assess the quality of the information provided by NDE systems, which is affected by factors such as the experience of the inspector, environmental conditions, ease of access, and the precision of the measurement device. In this paper, we review existing probabilistic models of NDE and show how they are connected within a unifying framework. This framework provides insights into how these models should be learned, calibrated, and applied. We investigate and highlight how the choice of the model can affect the maintenance decisions taken on the basis of NDE results. In addition, we analyze the impact of experimental design on the performance of a given NDE system in a decision-making context.

Keywords: NDE, PoD, ROC, decision analysis, value of information

*Corresponding author.
Email addresses: `elizabeth.bismut@tum.de` (Elizabeth Bismut), `straub@tum.de` (Daniel Straub)

1. Introduction

Probabilistic models have been developed to measure the quality and performance of non-destructive evaluation (NDE) methods. These models include probability of detection (PoD) curves, receiver operating characteristic (ROC) curves, or simply a probability of false positives/false negatives. In specific application domains, one of these models typically prevails (e.g., Packman et al., 1968; Hovey and Berens, 1988; Somoza et al., 1990; Sarkar et al., 1998), since they historically emerged independently in different disciplines to address specific problems. PoD curve models were formulated in the 1950s as a statistical method to investigate the dose-response effect in biological tests (Finney, 1978; Rudemo et al., 1989; Ritz et al., 2015). ROC curve models grew out of the signal detectability theory developed in the 1940s (Shannon, 1948; Woodward and Davies, 1952; Peterson and Birdsall, 1953), aimed at measuring the capacity of a receiver to distinguish the presence of a signal from noise. The purpose of this paper is to review and investigate these models for NDE quality by means of a unifying framework, which shows formally how they are connected. This enables us to provide insights into how they should be used and learned.

Historically, a first major application of NDE was the identification of flaws with systematic inspections during the manufacturing process of parts. This was especially a concern in the nuclear industry, which strived to improve quality assurance in the fabrication of pressure vessel components (International Atomic Energy Agency, 1965). In the aeronautic industry, the high cost associated with discarding parts with small defects during manufacture motivated the use of NDE also during the service life. By the 1970s, the US Air Force had launched detailed investigations on quantifying the performance (also called reliability) of defect detection measures (Packman et al., 1968; National Materials Advisory Board, 1969). Its program "Have-Cracks-Will-Travel" laid the foundation for systematic inspection and maintenance of aircraft (Berens and Hovey, 1981; Hovey and Berens, 1988; Singh, 2000), where fatigue cracks are allowed to develop as long as they are monitored and repaired regularly. These studies provide the analytical framework to derive PoD curves from NDE data.

In parallel, the research on lifetime extension of aircraft structures (Graham and Tetelman, 1974; Yang and Trapp, 1974), nuclear reactor components (Harris and Lim, 1983), and large deteriorating civil infrastructures (Frangopol et al., 1997; Hong, 1997; Sheils et al., 2012) started to incorporate these probabilistic NDE models into Bayesian reliability analysis of the structure. This was also coupled with inspection and maintenance plans, notably in risk-based inspection (RBI) planning (Yang and Chen, 1985; Madsen et al., 1987; Straub, 2004). In the context of highly reliable structures, optimizing the time and location of an inspection and the eventual subsequent repair can result in significant savings (Nielsen and Sørensen, 2010; Goulet et al., 2015; Luque and Straub, 2019; Bismut and Straub, 2021). This explains why NDE studies now also focus on quantifying the added value of a specific NDE method on the expected total life-cycle cost.

The accuracy and reliability of NDE depends on environmental conditions (humidity, temperature, experimental setting) and inspector expertise, among other factors (e.g., Lentz et al., 2002). NDE quality models establish a probabilistic relationship between the condition of the structure and the testing outcome (or outcome assessment). They can be used to rate NDE methods and ensure their compliance with norms and standards in place (e.g., Deutsches Institut für Normung, 2014; Pavlovic et al., 2008).

In this contribution, we give an overview of the existing types of models for NDE quality, and show how they are connected within a unifying framework. While some existing literature does discuss relationships between two NDE quality models, we are not aware of any previous attempt to formalize the relationship between all these models. This framework enables a better understanding of the assumptions associated with each model type and, as we demonstrate, can help in obtaining better models for NDE quality for specific applications. We particularly draw attention to the dependence of the NDE quality model on the experimental design and the risks of misinterpreting and generalizing quality indicators of NDE quality models from experimental studies. This formalization allows us to identify systematically issues and potential pitfalls when establishing and applying these models. Decision analysis and the concept of value of information (VoI) (Raiffa and Schlaifer, 1961) shed further light on what the implications are of not choosing the optimal model. The joint optimization of the calibration of the NDE system and the repair actions is formally addressed.

After some definitions in Section 2.1, Section 2.2 introduces the framework and the models and their connections are reviewed and discussed in Sections 2.3 to 2.6. Considerations on model availability and uncertainty are made in Section 2.7. Section 3 focuses on the exploitation of the NDE data within decision analysis. The findings are illustrated with two examples: In Section 4, we apply the framework and solve a basic decision problem for a purely theoretical NDE system; In Section 5 we consider a real NDE technique, the half-cell potential, and we analyze the effect of model choice on the outcomes of optimal decision making for a one-step and a two-step decision problem.

2. A unifying framework for NDE quality

2.1. NDE systems

NDE quality models establish a relationship between the true *condition* and the measurement, which we call the *observed signal*. The latter includes noise. Fig. 1 depicts this relationship, inspired by Shannon's diagram of a communication system (Shannon, 1948).
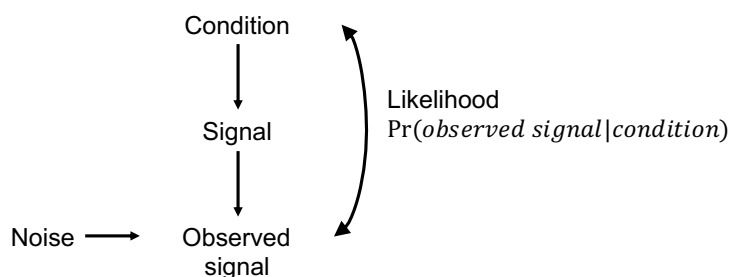

Figure 1: The observed signal measures the signal emitted by the condition and is also affected by noise. The model for NDE quality expresses the relationship between the observed signal and the condition, in the form of a likelihood.

The relationship between the true condition and the observed signal can be expressed probabilistically with *NDE quality models*. Here the term "NDE" covers all non-destructive information collection methods, ranging from visual inspection to automated data collection. An NDE quality

3

model is defined by the conditional probability Pr(*observed signal*|*condition*), which in statistics is known as the likelihood function. This relationship can be derived empirically by performing a number of tests (Packman et al., 1968; Berens and Hovey, 1981; Berens, 2000). While the primary purpose of these models was to establish a measure of reliability and performance of NDE techniques, they can be used for Bayesian analysis and decision analysis, see Section 3.

The *NDE system* encapsulates the process of collecting the data (type of data collected, inspection technique) and the interpretation of this data (Berens, 2000). The NDE system can encompass several measuring devices and data processors or interpreters (Sheils et al., 2010). The observed signal is the quantity on which the repair decision is taken. An example of an NDE system is an inspector going on-site, visually inspecting a wall, and appraising its state of damage (here, the condition) with a rating (here, the observed signal) (Quirk et al., 2018). The NDE quality model characterizes the NDE system.

2.2. Unifying NDE quality models

The NDE quality model probabilistically relates the observed signal with the condition. The condition can take continuous values (e.g., a crack size), or discrete values (e.g., "functioning" or "not functioning"). The observed signal can similarly take continuous (e.g., maximum vibration amplitude, measured crack length) or discrete values (e.g., "red" or "green", "suitable" or "not suitable"). Here, we limit the consideration of the discrete case to binary states. As the number of discrete states increases, the multinomial case approaches the continuous state.

We denote the continuous observed signal by S and the binary observed signal by I. Similarly, if the condition takes continuous values, it is denoted by X, and when it takes binary states, by Y. The condition can express a degree of damage or failure, although this remains an abstract concept and might not be related to anything failing as such. Here, the observed signal is considered as a scalar quantity, which can result from processing a multivariate signal, e.g., a time-series or an image (Kurz et al., 2012; Webb et al., 2015). Table 1 gives an overview of the four main NDE quality model categories for the possible combinations of continuous or binary condition X/Y and observed signal S/I.

Table 1: Monitoring models for binary or continuous signal and condition

		Condition	
		Continuous X	Binary Y
Signal	Continuous S	(1) $f_{S|X=x}(s)$	(3) ROC curve
	Binary I	(2) $PoD(x)$ curve	(4) PoD / PFA

The connection between the models comes from the fact that *the binary/discrete variables are the result of imposing one or more thresholds on the underlying continuous variables*. Specifically, a binary signal I is the result of a classification of the underlying continuous signal S, by means of a threshold s_{th}, that assigns $I = 1$ for $S > s_{th}$. A binary condition Y represents two domains of a continuous condition X, classified such that $Y = 1$ when $X > x_{th}$.

If one has access to the full continuous/continuous description of NDE quality, one can establish the link between all four models, as illustrated by Fig. 2. Model (1) is the *base model*. Models (2–4) are derived by imposing thresholds s_{th} and x_{th} on continuous signal S and condition

4

X, respectively. As we discuss in Section 2.7, almost all NDE systems and applications can be connected to the base model.
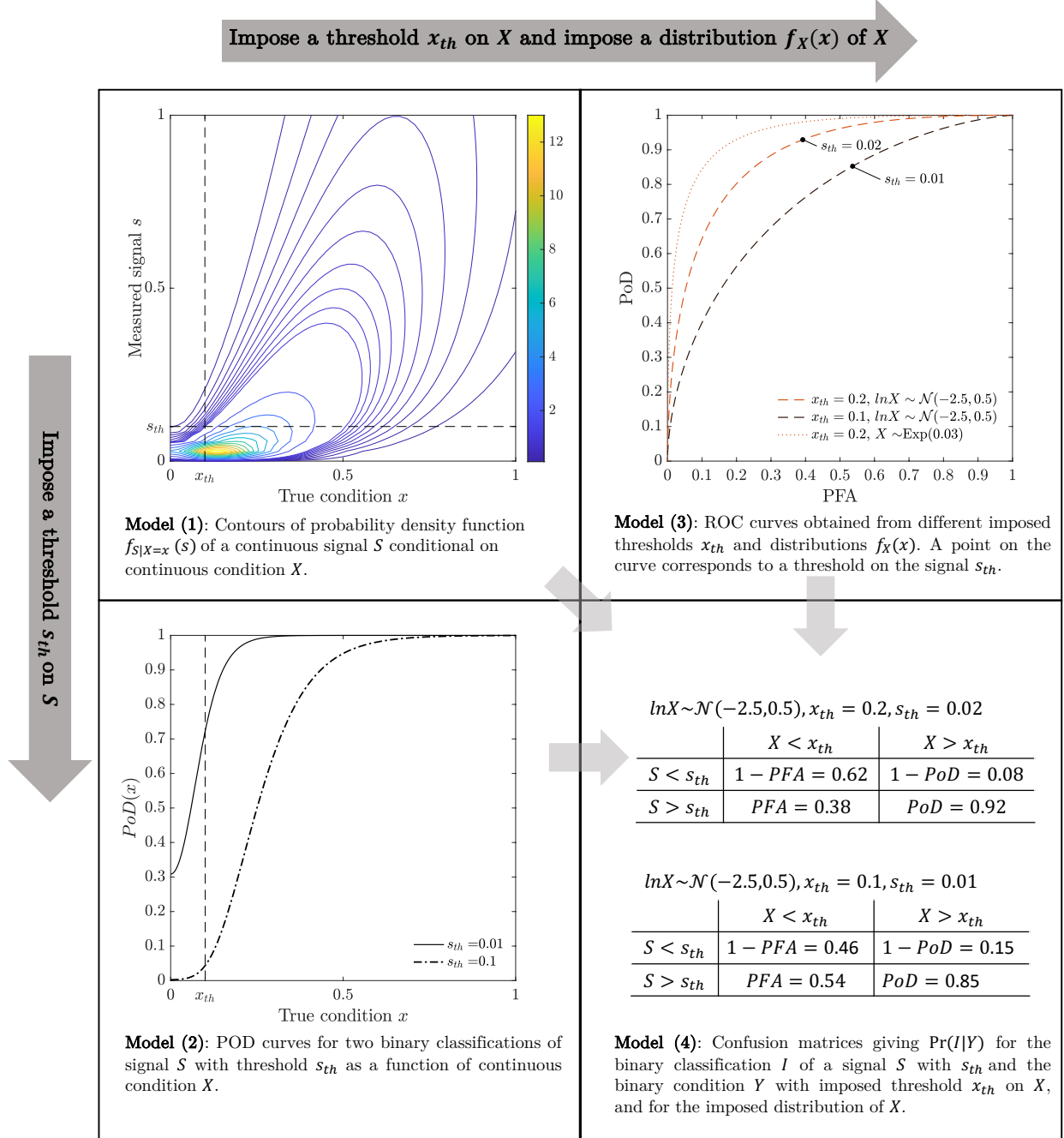

Figure 2: Unifying framework of models for NDE quality. Models (2–4) with a binary signal or condition can in principle be linked to Model (1). This link is established by fixing a threshold either on the signal, s_{th}, or on the condition, x_{th}, to classify the continuous signal or condition into binary states. The link between Models (1–2) to (3–4) requires additional information on the distribution of the continuous condition X. The specific NDE quality model depicted here is defined in Section 4.1.

The mathematical formulation of this unifying framework is derived in Sections 2.3 to 2.6. Importantly, we show that the links between models require an understanding of the population of defects in the experimental design. This affects the validity of the NDE models outside of the experimental setting in which they are learned. This has ramifications on the optimal interpretation of data and the decisions taken based on NDE.

2.3. Model (1): Base model – S continuous, X continuous

Many NDE methods relate a continuous condition to a continuous observed signal. For example, ultrasonic testing (UT) detects discontinuities inside a metal plate by emitting a high frequency ultrasonic pulse towards the plate and recording the echo. The amplitude of this echo (the observed signal) relates to the thickness (the condition) of defect-free material (Lavender, 1976). Another example described by this model is crack detection and measurement using magnetic particle inspection (MPI): MPI reveals the crack to the inspector who must then perform a visual inspection under UV lights and evaluate the crack size (Clark et al., 1987).

In the configuration of Model (1), the NDE system is fully characterized by the probability distribution of the observed signal S given the condition X, through the conditional probability density function (pdf), $f_{S|X}(s|x)$, or the associated cumulative distribution function (cdf), $F_{S|X}(s|x)$. Fig. 2 gives an example of such a conditional pdf.

This probabilistic model can be obtained from experimental data (experimental test blocks), if possible in different experimental settings. A traditional approach, which would allow to find such a model, is called "$\hat{a}$ vs. a," where $\hat{a}$ is the continuous observed signal and a the continuous condition (Berens, 2000). A relationship of the form $\hat{a} = f(a) + \varepsilon(a)$ has been proposed, where f is the mean response function with some fixed parameters and $\varepsilon(a)$ is a random variable representing the measurement noise. A linear log-logistic relationship is a common choice for the function f (Berens and Hovey, 1981; Annis, 2009). It originates from biological tests investigating the dose-response effect (Finney, 1978; Rudemo et al., 1989; Ritz et al., 2015). $\varepsilon(a)$ is commonly modeled as a Gaussian random variable (e.g., Kurz et al., 2012; Goulet et al., 2015). Annis (2009) provides some guidance as to how the noise should be considered.

Simulation and meta-models of NDE processes have also given rise to model-assisted PoD (MAPOD) (Aldrin et al., 2013; Calmon, 2012; Calmon et al., 2016). In this procedure, physics-based models are used to determine the relationship f and are validated with the experimental data. One of the advantage of MAPOD is that it does not require large datasets and can include numerous experimental settings and influential parameters.

Our review of the existing literature on NDE models shows that continuous/continuous probabilistic models are often learned in an ad-hoc manner. An example is the probability of (correct) sizing (POS), which describes the error in the measurement by an inspector of the continuous condition (e.g., a crack length) (Brennan, 2013; da Silva and de Padua, 2012; Granville and Charlton, 2016; Visser, 2002; Nath, 2021). Models for POS are continuous/continuous, but definitions vary and no application to reliability analysis is documented in the literature.

For many NDE techniques, this model remains abstract, as a continuous signal or a continuous condition might not be easily identifiable. In this case, the NDE quality model is chosen among the other three categories described below.

2.4. Model (2): PoD curve – I binary, X continuous

The probability of detection curve, or *PoD curve*, has been adopted for many NDE techniques, such as UT, MPI, Eddy current testing, or impulse radar, which aim at identifying cracks or more generally defects in structures (Berens and Hovey, 1981; Hovey and Berens, 1988; Sarkar et al., 1998; Feistkorn and Taffe, 2011). It is

$$PoD(x) = \Pr(I = 1 | X = x). \tag{1}$$

As previously noted, one can interpret I as a classification of a continuous signal S by fixing a threshold s_{th}, i.e., $\{I = 1\} = \{S > s_{th}\}$. In NDE literature pertaining to PoD curves, s_{th} is called the decision or detection threshold (Berens and Hovey, 1981; Sheils et al., 2012). The PoD function in Eq. (1) can thus be written as a function of the continuous/continuous Model (1).

$$PoD(x) = \Pr(S > s_{th} | X = x) = \int_{s_{th}}^{+\infty} f_{S|X}(s|x)\mathrm{d}s = 1 - F_{S|X}(s_{th}|x). \tag{2}$$

By changing the threshold s_{th}, the PoD curve changes (Sarkar et al., 1998); at the limit, it is $PoD(x, s_{th} = -\infty) = 1$ and $PoD(x, s_{th} = +\infty) = 0$ for any value x.

PoD curves have indeed been obtained from "$\hat{a}$ *vs.* a" models (Berens, 2000; Annis, 2009; Virkkunen et al., 2019). In aircraft integrity management, the threshold s_{th} is chosen so that a critical crack size fixed by expert judgment (Wood and Engle Jr, 1979) is detected with a 90% probability, with a 95% confidence level (Berens, 2000). According to Wood and Engle Jr (1979), the basis for these target probability and confidence level is arbitrary and relates to a required degree of conservatism and to the practical implementation of NDE testing programs. The characterization of a PoD curve with this 90/95 target value is considered best practice, until today (e.g., Annis, 2009; Tschöke et al., 2021). However, as we show below in Section 3, calibrating the PoD curve and fixing s_{th} in this way without a comprehensive decision analysis may not ensure that the test is exploited to its full potential. Furthermore, although the term decision threshold originally refers to the fact that a repair action systematically follows the detection of a defect with $S > s_{th}$, it may not be optimal to do so in all circumstances.

Eq. (2) does not preclude the PoD curve from taking a non-zero value when $x = 0$. A PoD curve for which $PoD(0) = 0.3$ is depicted in Fig. 2 for $s_{th} = 0.01$. Straub (2004) notes that PoD curves typically proposed in the literature, such as those obtained from the log-logistic model mentioned in Section 2.3, pass through the point $PoD(0) = 0$, and thus do not include the possibility of false detection. (Packman et al., 1968; Berens and Hovey, 1981) have motivated this choice with damage tolerant design philosophy by presuming that unnecessary repairs can only improve the reliability of the system. However, Heasler and Doctor (1996) argues that for risk-based decision analysis it is more appropriate to consider PoD models for which $PoD(0) > 0$ since unnecessary repairs lead to additional costs. Straub (2004) has proposed the term probability of indication (PoI) for PoD curves that include false alarms, i.e., those for which $PoD(0) > 0$, which include the effect of noise on the indication of the condition. Nevertheless, we use the term PoD curve throughout this paper.

7

2.5. Model (3): ROC curve – S continuous, Y binary

This model is commonly used to describe how the continuous observed signal S from NDE can be interpreted to discriminate between the absence ($Y = 0$) and presence ($Y = 1$) of a flaw. Most traditional NDE methods can be described by this model (Olin and Meeker, 1996), which can also be used to compare the ability of inspectors (human or machine) to interpret NDE results (Swets, 1983). Recently, image processing techniques for crack detection and monitoring systems of pipe corrosion have been evaluated with this model (Pakrashi et al., 2010; Jarvis et al., 2018).

In this framework, the discrete condition Y is defined by setting a threshold on the continuous condition X such that $\{Y = 1\} = \{X > x_{th}\}$ and $\{Y = 0\} = \{X \leq x_{th}\}$. x_{th} is called the critical threshold (Sheils et al., 2010; Schoefs et al., 2012). The likelihood function is formed by the two conditional pdfs of the observed signal S. They are represented in Fig. 3 and can be derived from the base Model (1):

$$f_{S|Y=1}(s) = \frac{1}{1 - F_X(x_{th})} \cdot \int_{x_{th}}^{+\infty} f_{S|X}(s|x) f_X(x) \mathrm{d}x, \tag{3}$$

$$f_{S|Y=0}(s) = \frac{1}{F_X(x_{th})} \cdot \int_{-\infty}^{x_{th}} f_{S|X}(s|x) f_X(x) \mathrm{d}x, \tag{4}$$

where $f_X(x)$ and $F_X(x)$ are the pdf and cdf of the condition X.
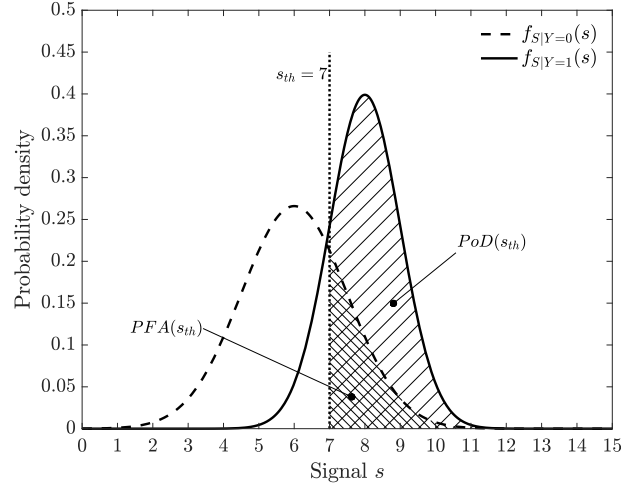

Figure 3: Conditional probability density functions. The indicated areas under the two curves defined by the threshold s_{th} corresponds to a point $PoD(s_{th} = 7)$, $PFA(s_{th} = 7)$ on the ROC curve shown in Fig. 4.

This model is commonly visualized by the corresponding receiver (or relative) operating characteristic (ROC) curve, which plots the *PoD* against the *PFA*. This curve is parametrized by a threshold on the signal s_{th}, also called the cut-off point (Fluss et al., 2005). Note that the *PoD* is not expressed as a function of the continuous condition X as in Section 2.4 above, but is here a function of s_{th}. The *PoD* and *PFA* as a function of s_{th} are

8

$$PoD(s_{th}) = \Pr(S > s_{th}|Y = 1) = \int_{s_{th}}^{+\infty} f_{S|Y=1}(s)\mathrm{d}s, \tag{5}$$

$$PFA(s_{th}) = \Pr(S > s_{th}|Y = 0) = \int_{s_{th}}^{+\infty} f_{S|Y=0}(s)\mathrm{d}s. \tag{6}$$

The ROC curve is illustrated in Fig. 4, wherein the *PoD* and *PFA* for threshold values s_{th} are obtained from the conditional pdfs of Fig. 3. Note that studies often derive the ROC curve for a single flaw size $X = x$, rather than for a domain defined by threshold x_{th} (e.g., Rajesh, 1993). This can be useful to qualitatively compare NDE methods on a standardized flaw size but does not provide a complete model linking the observed signal with the condition.
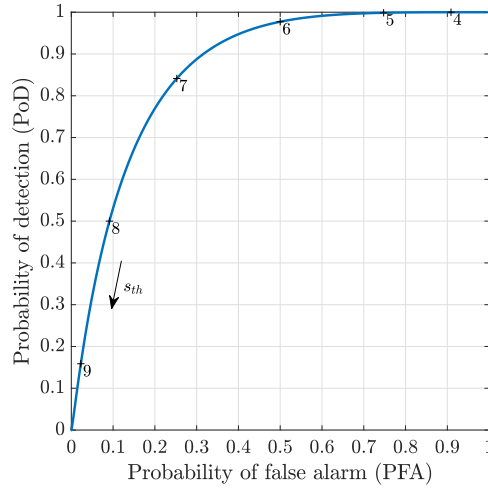

Figure 4: Each point on the ROC curve corresponds to a value of the signal threshold s_{th}.

One can alternatively express the *PoD* and *PFA* as a function of the conditional cdf $F_{S|X}$ from Model (1):

$$PoD(s_{th}) = \Pr(S > s_{th}|X > x_{th}) = \frac{1}{1 - F_X(x_{th})} \cdot \int_{x_{th}}^{+\infty} \left(1 - F_{S|X}(s_{th}|x)\right) f_X(x)\mathrm{d}x, \tag{7}$$

$$PFA(s_{th}) = \Pr(S > s_{th}|X < x_{th}) = \frac{1}{F_X(x_{th})} \cdot \int_{-\infty}^{x_{th}} \left(1 - F_{S|X}(s_{th}|x)\right) f_X(x)\mathrm{d}x. \tag{8}$$

Several indicators have been proposed to qualify the performance of an NDE system with a ROC curve (Taner and Antony, 2000; Greiner et al., 2000; Fluss et al., 2005). They include the *area under the curve* (AUC), the shortest distance between the curve and $(0,1)$, also expressed in polar coordinates (Schoefs et al., 2012), and the *Youden Index* (Youden, 1950). The latter is computed as the maximum vertical distance between the ROC curve and the 45 degree line starting at $(0,0)$.

Eqs. (3) to (8) show that the *PoD* and *PFA* on the ROC curve are a function of the distribution of the condition X. This signifies that even if the ROC curve is evaluated directly from experiments, it is only strictly valid for the distribution of the defects from which it is derived. Therefore, the ROC curve for the same NDE method can vary when applied to different situations. Boero et al. (2009) illustrates this effect for NDE of deteriorating structures, where the ROC curves change over time along with the distribution of the progressing damage condition. A problem appears when the given ROC curve is derived from a previous experiment because the distribution of X in the experiment, which we call *experimental design*, might not match the distribution of X in the specific application. To formally represent this, we distinguish between the pdf $f_X(x)$ and associated cdf $F_X(x)$ of the condition X for a specific application, and the pdf $f_{X,exp}(x)$ and associated cdf $F_{X,exp}(x)$ of X for the experiment from which the ROC curve is derived. Although mentioned in earlier studies (Berens and Hovey, 1981), the effect of the experimental design has largely been ignored in the recent literature. To show this effect, we consider three experimental designs and show that the resulting ROC curves change significantly in Fig. 5. An ROC curve from an NDE system provider should therefore be associated with information of the underlying experimental design.
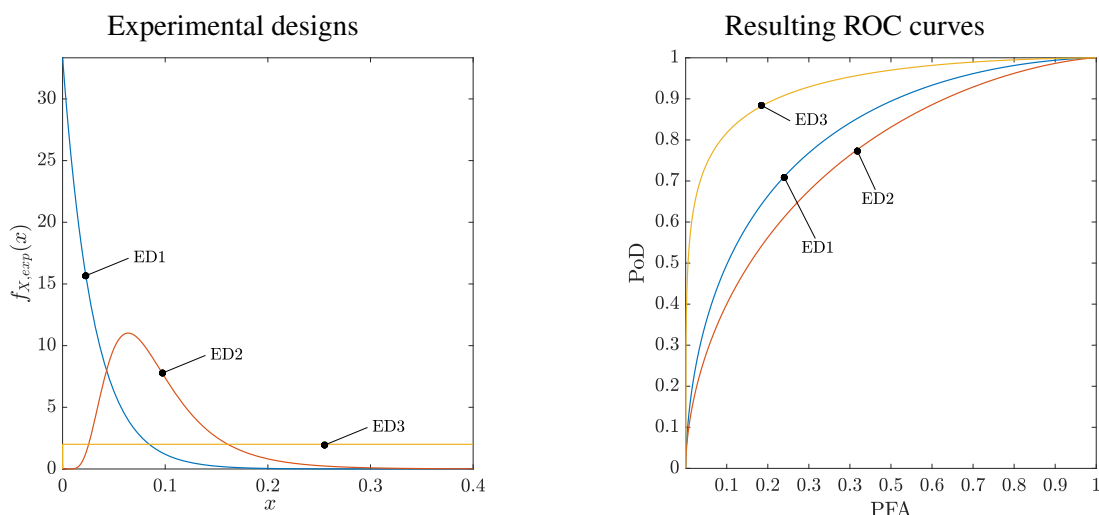

Figure 5: ROC curves for the same NDE derived for different experimental designs, ED1, ED2, and ED3. The model and distributions are presented in the example of Section 4 and their impact is investigated in Section 4.8. Here, $x_{th} = 0.1$.

2.6. Model (4): PoD/PFA – I binary, Y binary

This is the most elementary NDE quality model associated with an NDE, which identifies whether the system is in a certain state or not. It is obtained by fixing thresholds both on the condition X and the signal S. The associated likelihood is described by a confusion matrix, involving the operating *PFA* and *PoD*, as presented in Table 2.

An example of an NDE system that is described by such a model is ultrasonic flooded member detection, which detects the presence/absence of water in tubular steel members of underwater support structures as an indication of through-thickness cracks or other severe defects (Hayward et al., 1993; Visser, 2002). The performance of inspectors for visual inspection or using NDE

Table 2: Likelihood of the binary–binary case: The confusion matrix $\Pr(I|Y)$.

$$\Pr(I|Y)$$

		$Y = 0$	$Y = 1$
I	0	$1 - PFA$	$1 - PoD$
	1	PFA	PoD

devices is also typically represented through the PoD/PFA model (Swets, 1992; Sheils et al., 2010; da Silva and de Padua, 2012; Quirk et al., 2018).

The transition between Model (3) to Model (4) corresponds to calibrating the NDE system to an operating point on an ROC curve, by fixing the threshold s_{th}. The *PoD* and *PFA* are obtained from Eqs. (5) and (6).

The transition from Model (2) to (4) is obtained by combining Eqs. (7) and (8) with Eq. (2). The *PoD* and *PFA* for Model (4) are

$$PoD = \Pr(I = 1 | X > x_{th}) = \frac{1}{1 - F_X(x_{th})} \cdot \int_{x_{th}}^{+\infty} PoD(x) f_X(x) \mathrm{d}x, \tag{9}$$

$$PFA = \Pr(I = 1 | X < x_{th}) = \frac{1}{F_X(x_{th})} \cdot \int_{-\infty}^{x_{th}} PoD(x) f_X(x) \mathrm{d}x. \tag{10}$$

The transition from Model (1) to (4) is obtained by transitioning from Model (1) to (3), then from Model (3) to (4), by fixing thresholds s_{th} and x_{th}. Depending on the nature of the condition and the observed signal, it is possible that both thresholds s_{th} and x_{th} are assigned the same value (Sheils et al., 2012), however they play very distinct roles in the NDE system and should not be confused with one another.

2.7. *Some comments on learning the models*

In the ideal case, one would learn the continuous/continuous Model (1) directly, giving a probabilistic relationship between condition *X* and signal *S*. From this base model, the three other model categories could be derived for specific applications. Surprisingly, review of the existing literature reveals that, when continuous/continuous probabilistic models are obtained from experimental or simulated data (see Section 2.3), they are often not used for reliability analysis, or for explicitly deriving PoD curves or other models.

Fig. 6 summarizes how the four continuous or binary variables *X*, *Y*, *S*, and *I* interact through the four model types.

It is not always possible to reveal a continuous condition or a continuous signal, in which case the NDE system is described by one of Models (2) to (4). Still, the base model linking *X* to *S* can be considered at an abstract level to ensure correct interpretation of the signal and good experimental design.

The correct interpretation of the observed signal is also affected by the uncertainties associated with the model of NDE considered. Several studies point to the fact that experimental data and NDE performance obtained from this experimental data do not translate into similar performance once in-situ (e.g., Rouhan and Schoefs, 2003). The reasons given typically involve aleatoric and
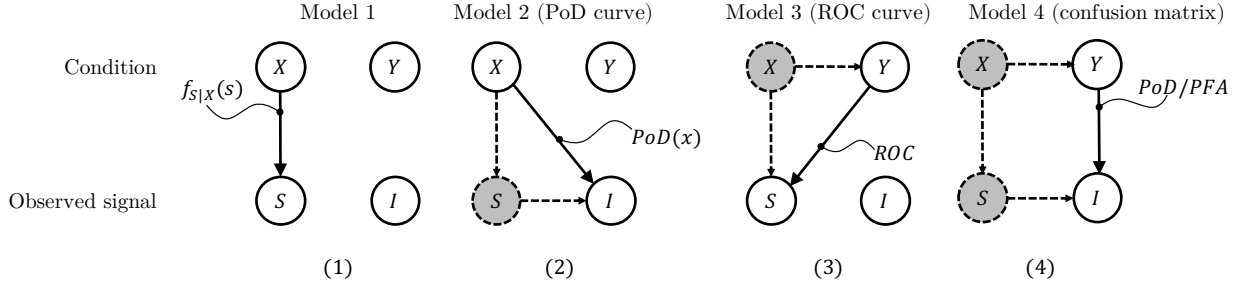
Figure 6: Interaction of the four continuous or binary variables X, Y, S, I through the NDE quality Models (1–4). The continuous edges illustrate the relationships between condition and observed signal described by the models. The grayed-out nodes designate the variables hidden when the models are directly evaluated from the experiments. For each model, the link between these hidden variable and the variables of the model can be expressed with Model (1) and an imposed threshold s_{th} or x_{th}.

epistemic uncertainties in influential parameters affecting the NDE quality, which cannot always be reproduced in an experimental context (Wall and Wedgwood, 1994; Straub, 2004). For instance, temperature, humidity or lighting can affect the performance of an NDE technique (Annis, 2009) and introduce a bias in the model. Boero et al. (2009) investigated the spatial variability of NDE models. These model uncertainties can be partly quantified by assigning a probability distribution to the model parameters, thus explicitly accounting for the model uncertainty in the NDE quality model performance (Straub, 2004). This hierarchical dependence in the NDE models can also be used to update the models and their parameters with the information collected through Bayesian inference (Aldrin et al., 2013; Hamida and Goulet, 2020).

However, even if model uncertainties can be mitigated, the derivations in Sections 2.5 and 2.6 show that incorrect assumptions on the distribution of the true condition X lead to erroneous ROC curves and misinterpretation of the performance of NDE systems. In Section 4.8 we investigate the effect of such errors on the optimality of decisions.

3. NDE systems and decision analysis

3.1. Possibility of improved decision-making

Section 2 shows that several choices are made when modeling the performance of an NDE system. There is no unique description of NDE quality and the quality models depend on threshold values on the measured signal and on the condition, s_{th} and x_{th}. These thresholds are set by either the designer or the user of the NDE system, often according to standard practice and predefined requirements (Wood and Engle Jr, 1979; Annis, 2009; Kurz et al., 2012). When the decision context is specifically considered for the choice of the thresholds, it is usually through the minimization of mis-classification rates or through minimum requirements on the detection performance of the NDE method (Rajesh, 1993). In some studies, the consequences of actions or failure are accounted for (e.g., Swets, 1992). In general, it is acknowledged that the thresholds affect the quality of the NDE system and that they should be calibrated (Kurz et al., 2012; Webb et al., 2015). Furthermore, the repair decision is often directly conditioned on the NDE outcome (recall the name "decision threshold" for s_{th}), and the quality of NDE systems is usually assessed

without optimizing this decision (Berens and Hovey, 1981; Sheils et al., 2010). In this paper, we distinguish between the interpretation of observed signal *S* (or calibration of the NDE system) and the decision and scrutinize these choices with a formal analysis.

This section employs formal decision analysis and the value of information (VoI) concept to investigate how NDE quality models affect the decisions taken based on NDE results. As is shown in the numerical results, employing Models (2) to (4) can lead to suboptimal decisions relative to using Model (1).

To evaluate the performance of NDE systems with Bayesian decision analysis, we introduce the following:

– *F* is the critical condition, e.g., a system failure.

– The decision maker can choose among a set of actions $\{a_0, a_1, ...\}$, typically to mitigate the consequences of failure. The "do nothing" action is denoted by a_0.

– All actions a_i have an associated cost; we assume here that action a_0 "do nothing" has an associated cost of 0. The consequence of *F* is c_F. To be directly comparable, all costs and consequences are expressed in monetary terms.

3.2. Bayesian analysis

As stated in Section 2.1, the probabilistic NDE quality models are likelihood functions. When a probabilistic model of the system state is available, Bayesian analysis can be performed, and the reliability of the structure can be evaluated in light of the NDE results. Such reliability updating with likelihoods from NDE quality models was first investigated by Tang (1973), and many studies integrating reliability analysis with inspection models have been published (e.g., Madsen, 1987; Sindel and Rackwitz, 1996; Onoufriou and Frangopol, 2002; Straub, 2011).

With Bayesian analysis, the posterior probability distribution of the condition Θ is

$$p(\boldsymbol{\theta}|z) \propto \mathcal{L}(\boldsymbol{\theta};z)p(\boldsymbol{\theta}), \tag{11}$$

where Θ is either *X* or *Y* depending on the setting. *z* is the measurement, which is either *s* or *i*. $\mathcal{L}(\boldsymbol{\theta};z)$ is the likelihood function, i.e., one of the NDE quality models of Section 2.2. The normalizing constant of Eq. (11) is the model evidence, $p(z)$:

$$p(z) = \mathbf{E}_{\Theta}\left[\mathcal{L}(\boldsymbol{\theta};z)\right], \tag{12}$$

where $\mathbf{E}_{\Theta}[\cdot]$ is the expectation with respect to $p(\boldsymbol{\theta})$.

One is typically interested in identifying a critical condition, or failure. Conditional on the NDE outcomes, one obtains the posterior probability of this critical condition $\Pr(F|Z=z)$ as

$$\Pr(F|Z=z) = \frac{1}{p(z)}\mathbf{E}_{\Theta}\left[\Pr(F|\boldsymbol{\theta})\mathcal{L}(\boldsymbol{\theta};z)\right]. \tag{13}$$

13

3.3. Optimal decision and value of information

While the methods developed in the NDE community focus on comparing inspection techniques, the actual value added is related to how the information from NDE leads to better decisions (Raiffa and Schlaifer, 1961; Straub, 2014). In decision analysis, it is assumed that the decision maker selects an action that maximizes the expected utility after obtaining information Z through NDE. Here we consider utility to be negatively proportional to costs. Hence, the optimization problem is written in as

$$a_{opt}(z) = \arg \min_{a \in \{a_0, a_1 \dots\}} \mathbf{E}_{\Theta|z}[C_T(a, \Theta)]. \tag{14}$$

The total cost C_T includes the cost of the actions and consequences of failure. $\mathbf{E}_{\Theta|z}[\cdot]$ is the expectation with respect to the conditional distribution $p(\theta|z)$ from Eq. (11).

The influence diagram of Fig. 7 represents the probabilistic relationships between the true condition, the observation and the failure event (round nodes), the actions and choice of NDE system (square nodes), as well as the costs incurred (diamond nodes). For example, e can be a
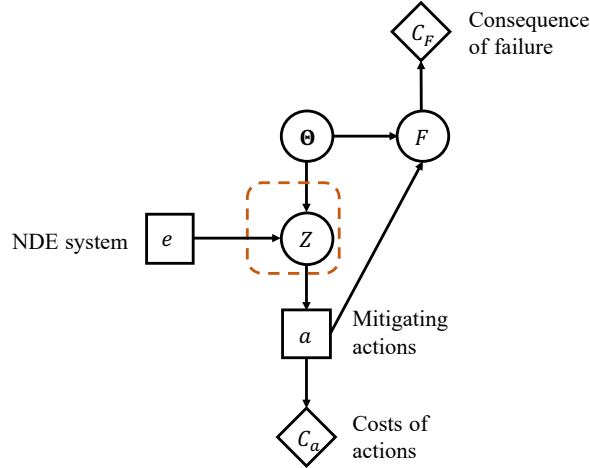

Figure 7: Influence diagram for the basic decision problem. The true condition, the observation, and the failure event are represented by round nodes, the actions and choice of NDE system by square nodes, and the costs incurred as diamond nodes.

visual inspection of a crack, the outcome Z can be the signal I "detection" or "no detection," and the possible actions a "repair" or "no repair."

The observation Z is unknown a priori and is described by the distribution of Eq. (12). Hence, the expected cost resulting from implementing the NDE system e, C_e, is obtained by deriving an expected value with respect to $p(z)$, i.e.,

$$C_e = \mathbf{E}_Z \left[\mathbf{E}_{\Theta|Z}[C_T(a_{opt}(Z), \theta)] \right]. \tag{15}$$

Raiffa and Schlaifer (1961) denote with e_0 the case where no information is collected and the decisions are based on the prior knowledge of the system. The optimal action and expected cost under e_0 are

14

$$a_{e_0} = \arg\min_a \mathbf{E}_{\Theta}[C_T(a, \Theta)], \tag{16}$$

$$C_0 = \mathbf{E}_{\Theta}[C_T(a_{e_0}, \Theta)]. \tag{17}$$

The value of information (VoI) (also called expected value of sample information) for an NDE system e is the difference between the total expected prior cost associated with e_0 and the total expected cost associated with e.

$$VoI(e) = C_0 - C_e. \tag{18}$$

Notably, if for any outcome z obtained through e, the optimal action $a_{opt}(z) = a_{e_0}$, then $VoI(e) = 0$. In other terms, an NDE system only brings value if it changes the decision on which action to take. The VoI does not include the cost of implementing the NDE system e (e.g., cost of mobilizing a team of inspectors). This cost can be added to C_e in Eq. (18) to obtain the net VoI (NVoI). In the Bayesian decision analysis framework, an NDE system with a positive NVoI is worth implementing, and the higher the NVoI the better. Somoza et al. (1990) employed this concept to calibrate an NDE system for optimal maintenance planning.

Although the concept of VoI appeared in the mid-20th century, it was not used to support inspection and maintenance decisions until recently, mainly because of the complexity associated with its calculation. Reliance on established processes also explains why it was left aside. Historically, the primary focus of NDE performance evaluation has been to support systematic inspection plans or condition-based maintenance, where monitoring outcomes systematically result in one action, of the "find nothing, do nothing" and "find a defect then repair" type (Berens and Hovey, 1981; Kerntechnischer Ausschuss, 2016). The performances of NDE techniques are codified by standards, based on traditional rule-of-thumb criteria, that are not specific to the problem. An example of NDE for systematic maintenance is the automated Eddy current inspection system for detecting and repairing cracks in engine components which has been used by the US air force since the 1980s (Berens, 2000). In the nuclear industry, standards prescribe fixed time intervals between inspections and the NDE techniques to adopt, assuming that any detected defect is systematically repaired (Kerntechnischer Ausschuss, 2016). The conservative conditions for inspection and repair are adopted without a risk-based decision analysis. In this context, false alarms and unnecessary repairs are judged as negligible when compared to preserving the safe operation of the system.

Currently, NDE is used for integrity management and service life extension of large infrastructure such as bridges. The cost of unnecessary repair is not negligible and can vary greatly with the accessibility of the structure. The VoI allows accounting for the economic aspects in the decision process. The computational and theoretical advances in the computation of the VoI in recent years provide the opportunity to evaluate the performance of an NDE system in the real decision-making context.

In Sections 4 and 5 we perform optimizations using Eqs. (14) to (18), where the decision maker can decide whether or not to repair the system based on observations provided by an NDE system. In the one-step decision problem, Eq. (14) reduces to a simple condition that links the likelihood $\mathcal{L}(\boldsymbol{\theta}; z)$ with the different costs and prior probabilities.

4. Example 1: Hypothetical NDE system

The purpose of this example is to demonstrate the unifying framework and investigate the impact of each of the four NDE models on the optimal decision outcome and cost in a basic decision problem. We highlight the effect of experimental design on the assessment of NDE performance.

4.1. Base model

We consider a hypothetical NDE system which measures a continuous condition and outputs a continuous signal, in analogy with crack detection systems such as UT.

The conditional pdf of the continuous measurement signal S given true condition X is

$$f_{S|X}(s|x) = \frac{1}{s\sqrt{2\pi}} \exp\left(-\frac{\left(\ln s - \ln\left(2x^3 + x^2 + 10^{-2}\exp(-1/2)\right)\right)^2}{2}\right). \tag{19}$$

This is the lognormal distribution with parameters $\left[\ln\left(2X^3 + X^2 + 10^{-2}\exp(-1/2)\right), 1\right]$. This conditional probability density is shown in Fig. 2 (Model (1)).

In the following we consider a basic decision problem, where one needs to decide on a repair action. We adopt the four different NDE quality models described in Section 2.2 in turn and assess their impact on the repair decision. We additionally investigate the effect of learning the NDE models with different experimental designs.

4.2. Solution of the one-step decision problem

The decision problem consists in selecting a repair action to mitigate the consequence of failure, after obtaining imperfect NDE outcome Z on state Θ. It is illustrated in Fig. 7. The condition Θ is either X or Y. The observation variable is Z, which is either the continuous signal S or the binary signal I. $\mathcal{L}(\Theta; z)$ indicates the likelihood function. In this setup, we consider two possible actions: either do nothing a_0, for a cost $c_A(a_0) = 0$, or repair a_R, for a cost $c_A(a_R) = c_R$. The probability of system failure F is defined conditional on the state Θ and the action A. The consequence of failure is c_F.

The a priori optimal action a_{e_0} is the one that minimizes the total expected cost as given by Eq. (17). Here,

$$\mathbf{E}_{\Theta}[C_T(a,\Theta)] = c_A(a) + c_F \Pr(F|a) \text{ with } \Pr(F|a) = \mathbf{E}_{\Theta}\left[\Pr(F|\Theta,a)\right]. \tag{20}$$

From Eq. (16), one finds that

$$a_{e_0} = a_R \iff c_F \Pr(F|a_0) > c_R + c_F \Pr(F|a_R). \tag{21}$$

As per Eq. (14), the a posteriori optimal action is the one that minimizes the conditional expected total cost,

16

$$\mathbf{E}_{\Theta|z}[C_T(a,\Theta)] = c_A(a) + c_F\Pr(F|Z=z,a), \text{ with } \Pr(F|Z=z,a) = \mathbf{E}_{\Theta|z}[\Pr(F|\Theta,a)]. \quad (22)$$

Making the likelihood $\mathcal{L}(\Theta;z)$ explicit in the conditional expectation, one obtains that

$$a_{opt}(z) = a_R \iff \mathbf{E}_{\Theta}[\mathcal{L}(\Theta;z)c_F\Pr(F|\Theta,a_0)] > \mathbf{E}_{\Theta}[\mathcal{L}(\Theta;z)\{c_R+c_F\Pr(F|\Theta,a_R)\}]. \quad (23)$$

When the condition is binary, i.e., $\Theta = Y$, $\mathcal{L}(Y;z)$ is the likelihood for Models (3) or (4). The condition of the above equation can be transformed into a condition on $\frac{\mathcal{L}(Y=1;z)}{\mathcal{L}(Y=0;z)}$, also called the "likelihood ratio test" (Peterson and Birdsall, 1953; Green and Swets, 1966). If $a_{e_0} = a_R$,

$$a_{opt}(z) = a_R \iff \frac{\mathcal{L}(Y=1;z)}{\mathcal{L}(Y=0;z)} > \frac{\{c_R+c_F\Pr(F|Y=0,a_R)-c_F\Pr(F|Y=0,a_0)\}\Pr(Y=0)}{\{c_F\Pr(F|Y=1,a_0)-[c_R+c_F\Pr(F|Y=1,a_R)]\}\Pr(Y=1)}. \quad (24)$$

The expected total cost associated with the optimal action following an NDE is given by Eq. (15). When the observation is binary, i.e., $Z = I$ and $\{I=1\} = \{S > s_{th}\}$, the expected cost depends on the choice of threshold s_{th}.

4.3. Condition model and a priori optimal action

The prior pdf of X is the exponential pdf with mean 0.03.

$$f_X(x) = \frac{1}{0.03}\exp\left(-\frac{x}{0.03}\right). \quad (25)$$

We fix the cost of repair $c_R = 1$. Under the do-nothing action a_0, the probability of failure of the system is defined conditional on X (Fig. 8),

$$\Pr(F|X=x) = 10^{-5} + (1-10^{-5}) \cdot \left(\frac{1}{2} + \frac{1}{2}\mathrm{erf}\left(\frac{\log x - 0.1}{\sqrt{2}}\right)\right), \quad (26)$$

where erf is the error function.

When the system is repaired, the probability of failure is reduced such that $\forall x$, $\Pr(F|X = x, a_R) = p_{F|R} = 10^{-4}$. The cost of failure is $c_F = 800$.

Eq. (21) gives the condition to identify the a priori best decision. The probability of failure of the system conditional on action a_0 is evaluated as $\Pr(F|a_0) = 1.2 \cdot 10^{-3}$, and the probability of failure conditional on action a_R is $\Pr(F|a_R) = p_{F|R} = 10^{-4}$. Since $c_F\Pr(F|a_0) = 0.94$ is smaller than $c_R + c_F p_{F|R} = 1.08$, the a priori best decision is therefore to do nothing, $a_{e_0} = a_0$, with associated cost 0.94.

4.4. A posteriori optimal actions with Model (1)

The likelihood $\mathcal{L}(x;s)$ is given by Eq. (19). By solving Eq. (23) for $z = s$, we find that the optimal action is $a_{opt}(s) = a_0$ for $s < s_{th,1} = 1.6 \cdot 10^{-2}$, and $a_{opt}(s) = a_R$ for $s > s_{th,1}$. Here, $s_{th,1}$ is effectively a repair threshold. The comparison of expected costs for both actions as a function of the measured signal S is depicted in Fig. 9. The expected total cost computed with Eq. (15) is 0.65. The VoI of this NDE system is $0.94 - 0.65 = 0.29$.
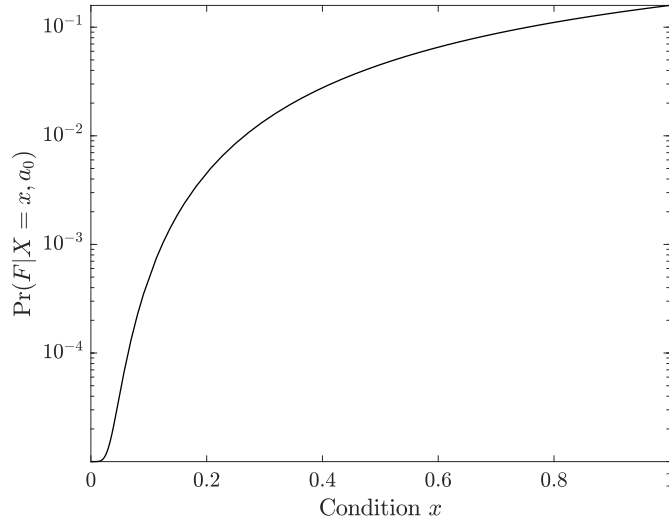
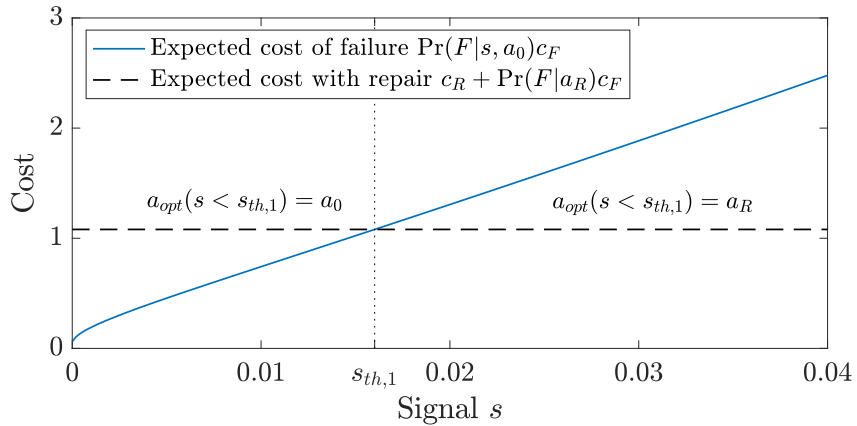Figure 8: Probability of system failure conditional on the condition X.

Figure 9: Expected cost for action a_0 and a_R conditional on the observed signal S.

4.5. A posteriori optimal actions for a given PoD curve

We now determine the optimal a posteriori action identified with the PoD curve (Model (2)) and the associated VoI. We remind that in Model (2), a binary signal I is considered, which is related to the continuous signal S by $\{I = 1\} = \{S > s_{th}\}$. The likelihood $\mathcal{L}(x; I)$ is given by the PoD curve of Eq. (2), which depends on the fixed threshold s_{th}. PoD curves for different thresholds are depicted in Fig. 10.

PoD curves are typically given and not specific to the application, and hence are likely suboptimal for a given decision context. For example, if the given PoD curve is calibrated with $s_{th} = 0.03$, the optimal actions are $a_{opt}(I = 0) = a_0$, $a_{opt}(I = 1) = a_R$ and the expected cost is 0.70. The resulting the VoI is $0.94 - 0.70 = 0.24$, which is below the potentially achievable VoI= 0.29.

For each imposed threshold s_{th} and associated PoD curve, Fig. 11 indicates the optimal actions
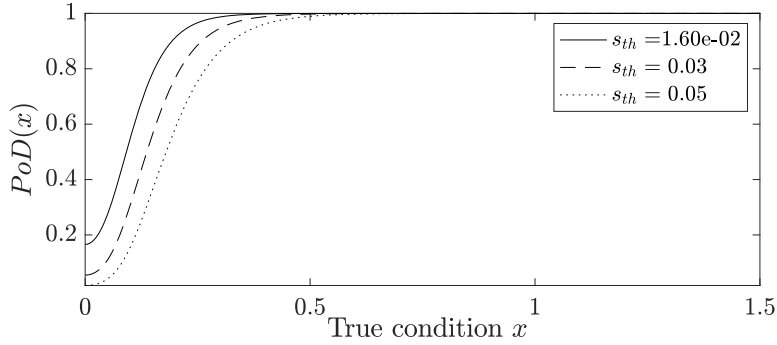
Figure 10: PoD curves for different thresholds s_{th}, including $s_{th,2}$.

$a_{opt}(I)$ and the expected cost calculated with Eq. (15). A threshold exists that maximizes the VoI and minimizes the expected cost for this case study. It is the decision threshold $1.6 \cdot 10^{-2}$, with the associated expected cost of 0.65, which corresponds to the result obtained with Model (1). However, when describing the NDE by the PoD curve, this optimal solution will only be obtained by coincidence. In general, the PoD curve leads to a suboptimal decision.
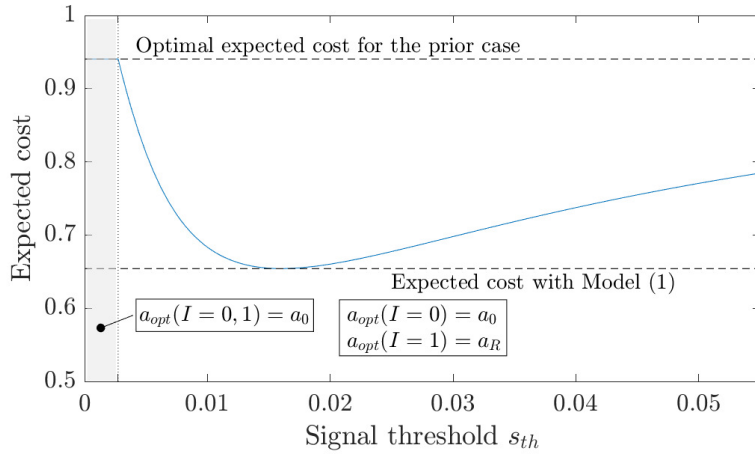

Figure 11: Expected cost using the PoD curve model as a function of the fixed signal threshold s_{th}.

4.6. A posteriori optimal actions for a given ROC curve

When ROC curves are utilized to represent the quality of NDE, the continuous condition X is replaced by the binary condition $\{Y = 1\} = \{X > x_{th}\}$. Here we fix $x_{th} = 0.1$. The corresponding ROC curve is depicted in Fig. 5, where the distribution of X corresponds to the experimental design ED1.

To compare the performance of Models (1) and (3), we must first ensure that the relationships between binary condition Y and failure F are compatible with the relationship between continuous condition X and failure F. This compatibility is ensured by the following equations,

19

$$\Pr(Y=1) = 1 - F_X(x_{th}) = 3.6 \cdot 10^{-2}, \tag{27}$$

$$\Pr(F|Y=1,a_0) = \frac{\int_{x_{th}}^{+\infty} \Pr(F|X=x) f_X(x) \mathrm{d}x}{1 - F_X(x_{th})} = 1.7 \cdot 10^{-2}, \tag{28}$$

$$\Pr(F|Y=0,a_0) = \frac{\int_0^{x_{th}} \Pr(F|X=x) f_X(x) \mathrm{d}x}{F_X(x_{th})} = 5.8 \cdot 10^{-4}. \tag{29}$$

Additionally, it is $\forall y \in \{0,1\}$, $\Pr(F|Y=y,a_R) = p_{F|R}$.

The solution of the decision problem is derived from Eq. (24):

$$a_{opt}(s) = a_R \iff \frac{\mathcal{L}(Y=1;s)}{\mathcal{L}(Y=0;s)} = \frac{f_{S|Y=1}(s)}{f_{S|Y=0}(s)} > 1.31. \tag{30}$$

The likelihoods $\mathcal{L}(Y;s) = f_{S|Y}(s)$ are computed with Eqs. (3) and (4). Both Models (1) and (3) provide the same model evidence $p(s)$.

From Eq. (30), we find the optimal action $a_{opt}(s)$ is a_0 when $s < 1.7 \cdot 10^{-2}$ and a_R otherwise. This threshold $s_{th,3} = 1.7 \cdot 10^{-2}$ is shown as the optimal operating point on the ROC curve in Fig 12. Furthermore, Fig. 12 indicates that ROC performance indices, such as the Youden Index, are associated with certain operating points. But using either of these points, as proposed in some references (e.g., Schoefs and Clement, 2004), is not optimal in view of the specific decision.

The total expected cost for the optimal operating point is 0.76. The VoI of this NDE system is 0.18, lower than for Model (1). Even though the actions are optimized, their efficiency are limited by the fixed threshold x_{th} and associated "failure" domain of the NDE device, through Eq. (30). Other values of x_{th} can yield a higher or lower VoI: E.g., with $x_{th} = 0.2$ the associated optimal expected cost is 0.91 and the VoI is only 0.03. However, the threshold x_{th} can typically not be influenced for a given NDE device.

4.7. A posteriori optimal actions for a given point on the ROC curve

Here, the NDE system is described by a point on the ROC curve defined above in Section 4.6, such that $\{I = 1\} = \{S > s_{th}\}$. The likelihood of Model (4) $\mathcal{L}(Y;I)$ is given by a confusion matrix as per Table 2, with $\mathcal{L}(Y=1;I=1) = PoD$ and $\mathcal{L}(Y=0;I=1) = PFA$. As in Eq. (30), the likelihood ratio verifies the conditions

$$a_{opt}(I=1) = a_R \iff \frac{\mathcal{L}(Y=1;I=1)}{\mathcal{L}(Y=0;I=1)} > 1.31 \iff \frac{PoD}{PFA} > 1.31, \tag{31}$$

$$a_{opt}(I=0) = a_R \iff \frac{\mathcal{L}(Y=1;I=0)}{\mathcal{L}(Y=0;I=0)} > 1.31 \iff \frac{1-PoD}{1-PFA} > 1.31. \tag{32}$$

For example, for $s_{th} = 1 \cdot 10^{-2}$ (and $x_{th} = 0.1$ as above), $PoD = 0.82$, $PFA = 0.37$, $a_{opt}(I=1) = a_R$, $a_{opt}(I=0) = a_0$ and the expected cost is 0.79.

Fig. 12 shows the expected cost for any fixed *PoD* and *PFA* as well as the optimal actions $a_{opt}(I)$. Two zones can be distinguished: (a) for any (PFA, PoD) in the yellow zone, the optimal course of action is to do nothing, whatever the observation outcome; this is what the prior optimal action also prescribes, therefore the VoI at these points is 0; (b) for any (PFA, PoD) to the left of this yellow zone, detection triggers repair a_R, and no detection entails a_0.
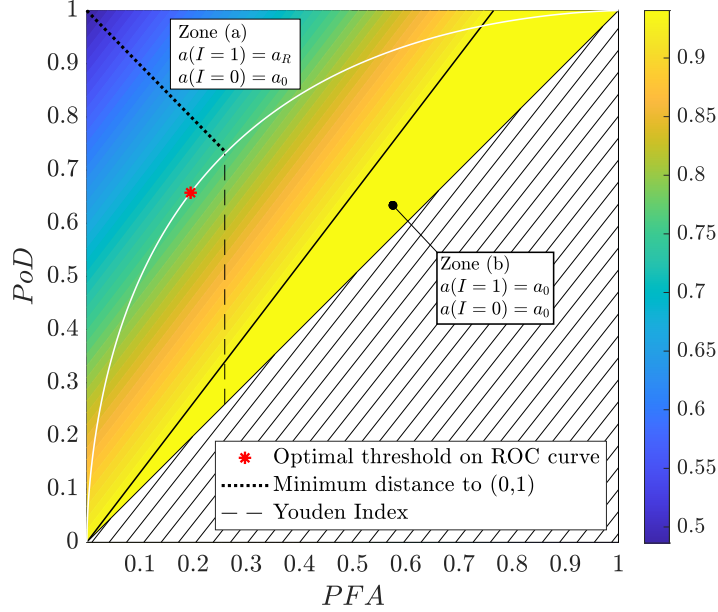

Figure 12: Optimal expected cost as a function of *PFA* and *PoD*. Zone (b) corresponds to *VoI* = 0: this means that for any decision threshold s_{th} on the ROC curve that falls within that zone, the best course of action is the prior optimal action, i.e., do nothing. The hatched area corresponds to the situation where the significance of $I = 0$ and $I = 1$ are inverted.

4.8. Influence of the experimental design

We investigate how the experimental design $f_{X,exp}$ used to learn the ROC curve can affect the decision taken and the resulting expected total cost and VoI. We consider three experimental designs ED1, ED2, and ED3, where the test conditions follow an exponential, lognormal, and uniform distribution, respectively (Eqs. (33) to (35)). The imposed threshold on the condition remains $x_{th} = 0.1$ as above, with $\{Y = 1\} = \{X > x_{th}\}$.

$$f_{X,exp,1}(x) = \frac{1}{0.03} \exp\left(-\frac{x}{0.03}\right), \tag{33}$$

$$f_{X,exp,2}(x) = \frac{2}{x\sqrt{2\pi}} \exp\left(-2(\ln x + 2.5)^2\right), \tag{34}$$

$$f_{X,exp,3}(x) = 2 \text{ if } 0 \leq x \leq 0.5, \text{ 0 otherwise.} \tag{35}$$

The corresponding likelihoods $f_{S|Y,exp,i}(s)$ are obtained with Eqs. (3) and (4). The resulting ROC curves for $x_{th} = 0.1$ are plotted in Fig. 5.

For each likelihood $f_{S|Y,exp,i}(s)$, the model evidence $p_{exp,i}(s)$ is evaluated with Eq. (12), where the prior probability of the condition Y is obtained with Eq. (27):

$$p_{exp,i}(s) = \sum_{y=0}^{1} f_{S|Y=y,exp,i}(s) \Pr(Y = y) = \frac{1 - F_X(x_{th})}{1 - F_{X,exp,i}(x_{th})} \cdot \int_{x_{th}}^{+\infty} f_{S|X}(s|x) f_{X,exp,i}(x) \mathrm{d}x \qquad (36)$$

$$+ \frac{F_X(x_{th})}{F_{X,exp,i}(x_{th})} \cdot \int_{-\infty}^{x_{th}} f_{S|X}(s|x) f_{X,exp,i}(x) \mathrm{d}x.$$

The posterior probabilities $\Pr(Y|s)$ and $\Pr(F|s)$ are similarly computed with Eqs. (11) to (13).

The first experimental design ED1$= f_{X,exp,1}(x)$ corresponds to the distribution of the condition in this application $f_X(x)$ defined in Eq. (25). Eq. (36) shows that in this case the model evidence $p_{exp,1}(s)$ and the posterior probabilities coincide with the model evidence and posterior probabilities obtained by applying Model (1) and the continuous/continuous likelihood $f_{S|X=x}(s)$. $f_{S|Y,exp,1}(s)$ is therefore the true likelihood and provides the correct evaluation of expected costs. This case is examined in Sections 4.6 and 4.7, and is called the reference case here.

When the experimental design differs from the distribution in the application, i.e., when $f_{X,exp}(x) \neq f_X(x)$, the model evidence and posterior probabilities deviate from the reference case and the conditional expected costs calculated with Eq. (22) are incorrect. This in turn affects the results of the decision problem and changes the optimal action attributed to each observation outcome by Eq. (23). It also impacts the expected cost calculated with Eq. (15).

One can evaluate this impact by overlaying the corresponding ROC curves on Fig. 12 and by plotting the expected cost and optimal actions of the decision problem with binary observation $\{I = 1\} = \{S > s_{th}\}$ as a function of the fixed signal threshold s_{th}, as in Fig. 13.

Fig. 13 shows that each experimental design gives a different evaluation of the expected cost for a given s_{th}. The constant part of each curve corresponds to the portion of the ROC curve in zone (a), and the other part falls in zone (b) (see Fig. 12). These actions zones differ for each ROC curve. For example, when the ROC curve is learned from experiments with ED2$= f_{X,exp,2}(x)$, one finds that imposing a threshold $3.5 \cdot 10^{-3} < s_{th} < 6 \cdot 10^{-3}$ leads to the action "do nothing," whatever the outcome s. In contrast, under the reference case the optimal action is to repair if the signal is higher than the fixed threshold, and do nothing otherwise. By misevaluating the conditional expected costs, the experimental design leads to suboptimal actions.

Another problem is that the optimal calibration, in the form of a threshold $s_{th,exp,i}$, minimizes the wrong expected cost function. The effective expected cost at this threshold, obtained with the reference case, can in fact be higher or lower. In Fig. 13, the optimal threshold for ED2 is $s_{th,exp,2} = 2.6 \cdot 10^{-2}$. At this point, the effective expected cost given by ED1 is lower than the perceived expected cost for ED2. Table 3 summarizes these results for all investigated likelihoods.

In summary, ignoring the underlying experimental design of a given ROC curve can lead to suboptimal decisions and to an NDE system that is not used to its full potential. Additionally, the expected costs can be over- or underestimated, hence they affect the perceived VoI. Ultimately, this
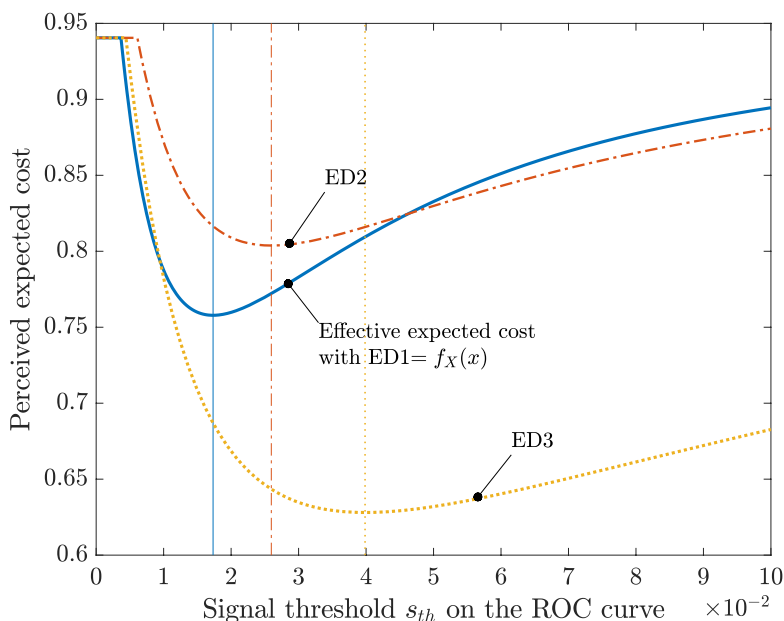
Figure 13: Expected cost calculated for experimental designs ED1, ED2, and ED3. ED1 corresponds to the distribution of the condition in this application $f_X(x)$. ED2 and ED3 differ from $f_X(x)$. For a fixed threshold s_{th}, only ED1 provides the correct values for *PoD* and *PFA* and the associated expected cost. ED2 and ED3 give wrong *PoD* and *PFA*, thus the perceived expected cost associated with s_{th} is also wrong. The vertical lines of corresponding line style locate the perceived optimal threshold $s_{th,exp,i}$ for each experimental design, where optimal action a_{opt} is a_0 if $s < s_{th,exp,i}$ and a_R otherwise. The effective expected cost for these thresholds are read on the reference curve (ED1): The perceived expected cost for the optimal threshold for ED2 overestimates the effective expected cost; the perceived expected cost for the optimal threshold for ED3 underestimates it (see Table 3).

Table 3: Effect of the experimental design on recommended actions and on the assessment of expected cost.

Experimental design	Optimal threshold	Perceived expected cost	Effective expected cost
$f_{X,exp,1}(x)$ – Reference	$1.7 \cdot 10^{-2}$	0.76	0.76
$f_{X,exp,2}(x)$	$2.6 \cdot 10^{-2}$	0.80	0.77
$f_{X,exp,3}(x)$	$4.0 \cdot 10^{-2}$	0.63	0.81

can lead the decision maker to favor one or another NDE system based on an erroneous evaluation of the NVoI of the NDE systems.

5. Example 2: Half-cell potential measurement for corrosion detection

The purpose of this example is the investigation of the effects of different model choices on a real NDE technique. We consider the half-cell potential measurement of reinforcement corrosion in concrete structures, and we analyze the effect of choosing Model (3) or Model (4), with different calibration choices, on the outcomes of optimal decision making for a one-step and a two-step decision problem.

5.1. Description and model of the inspection technique

Corrosion of reinforcement bars (rebars) in concrete structures is one of the leading deterioration mechanisms in civil infrastructure. Visible signs of corrosion on the concrete surface occur when corrosion of the rebars is already extensive and major repairs are needed. This demonstrates the need for early detection of rebar corrosion.

As the rebars are encased in the concrete, NDE methods are required to monitor their condition. Half-cell potential measurement is such a method and it is used to detect whether corrosion has initiated. This test measures the difference of electric potential between an electrode, directly connected to an exposed rebar, and a half-cell (reference electrode) placed on the concrete surface. The intensity of the chemical reaction responsible for corrosion at the half-cell is translated into a negative potential. The higher the amplitude of this electric potential, the higher the probability that corrosion has initiated in the rebar under the concrete cover (Elsener et al., 2003).

Empirical testing has revealed that the amplitude of the signal is clearly affected by environmental factors including humidity, chloride content, temperature, concrete cover, or concrete strength (e.g., Lentz et al., 2002; Kessler and Gehlen, 2017). This is why obtaining relevant and accurate probabilistic models a priori is a challenge. In practice, the electric potentials are measured on a regular grid on the concrete structure and are summarized in a frequency plot. The measurement data is then classified into two sets, "corroded" and "not corroded," by applying a threshold on the potential. This is typically done by expert judgment. The downside of this method is that it relies on relatively few sample measurements and misclassification can occur at a high rate.

Some studies have proposed to model the performance of this NDE method with a non-site-specific likelihood (Lentz et al., 2002; Faber and Sorensen, 2002). Typically, the chosen model involves a continuous signal (the potential) and a binary measured condition (corrosion initiated/no corrosion), thus follows Model (3) described in Section 2.2. However, these distributions are not used as such to perform Bayesian analysis. Instead, a threshold value on the potential is proposed, and the continuous signal transformed into a binary one (Lentz et al., 2002). Interestingly, the ROC representation is not common in this domain, even if it highlights best how the *PoD* and *PFA* vary with the chosen threshold.

This example uses the NDE quality model proposed by Faber and Sorensen (2002). The absence and presence of corrosion is represented by $Y = 0$ and $Y = 1$, respectively. We adopt normal pdfs for the likelihood after (Faber and Sorensen, 2002). The parameters of the normal distribution of the measured electric potential S given that no corrosion has initiated ($Y = 0$) are $\mu_0 = -0.207$[Volt] and $\sigma_0 = 0.0804$[Volt]. The parameters of the normal distribution given that corrosion has initiated ($Y = 1$) are $\mu_1 = -0.354$[Volt] and $\sigma_1 = 0.08$[Volt]. The distributions are shown in Fig. 14a and the corresponding ROC curve in Fig. 14b. These distributions are (where φ is the standard normal) pdf:

$$f_{S|Y=0}(s) = \mathcal{L}(Y=0;s) = \varphi\left(\frac{s-\mu_0}{\sigma_0}\right), \tag{37}$$

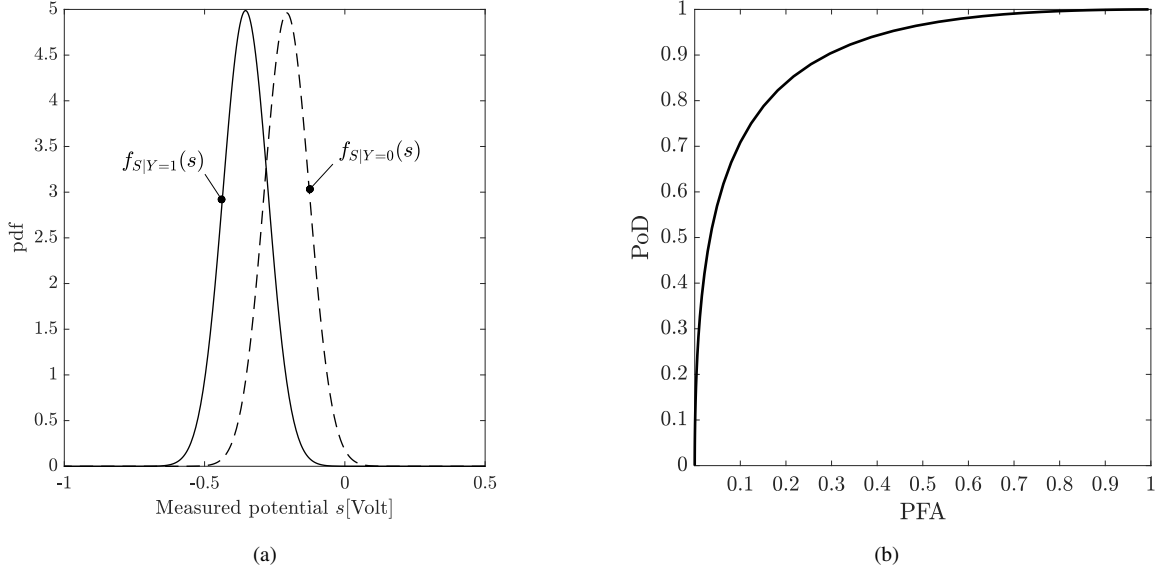$$f_{S|Y=1}(s) = \mathcal{L}(Y=1;s) = \varphi\left(\frac{s-\mu_1}{\sigma_1}\right). \tag{38}$$

Figure 14: (a):The two likelihoods are depicted. (b): Corresponding ROC curve.

We first investigate the one-step decision problem, where a repair action is taken based on the NDE result. In the second part, we consider a two-step decision problem, where information about the condition of the rebar is collected in two separate instances and repairs can be carried out each time. For both problems, we find the optimal course of action considering the continuous signal with Model (3). We show that fixing a threshold for Model (4), irrespective of the specific application is not optimal, for instance the one recommended by Faber and Sorensen (2002). We demonstrate how this fixed threshold can be optimized for both problems.

5.2. One-step decision problem

5.2.1. Decision model and a priori optimal action

Corrosion of the system ($Y = 1$) can eventually lead to failure. The decision maker has the opportunity to repair, a_R, at a cost $c_R = €5$ million, or to do nothing, a_0, at cost of $€0$. If the system is repaired, there are no further consequences. If the system is corroded ($Y = 1$) but not repaired, the consequences are $c_F = €50$ million. If the system is not corroded ($Y = 0$), there are no consequences. The prior probability of corrosion is $\Pr(Y = 1) = 0.05$.

The results are obtained following Section 4.2. Without inspections, the expected total cost conditional on action a_0 is $c_F \Pr(Y = 1) = €2.5$ million. Conditional on action a_R, it is $c_R = €5$ million. Hence, the a priori optimal action is $a_{e_0} = a_0$, do nothing.

5.2.2. Optimal a posteriori action considering the continuous signal

Eq. (24) applied with $Z = S$ is here adapted to

$$a_{opt}(s) = a_R \iff \frac{\mathcal{L}(Y = 1; s)}{\mathcal{L}(Y = 0; s)} > \frac{c_R \Pr(Y = 0)}{(c_{F,0} - c_{F,R} - c_R) \Pr(Y = 1)} = 2.11. \tag{39}$$

25

We derive

$$a_{opt}(s) = a_R \iff \left(\frac{s-\mu_0}{\sigma_0}\right)^2 - \left(\frac{s-\mu_1}{\sigma_1}\right)^2 > 2\log\left(\frac{\sigma_1}{\sigma_0}\frac{1}{2.11}\right). \tag{40}$$

Finally,

$$a_{opt}(s) = a_R \iff s \in [s_1^* = -29.72[\text{Volt}]; s_2^* = -0.31[\text{Volt}]]. \tag{41}$$

We see here that the signal domain is divided into three action zones. This is due to a mathematical artifact caused by the fact that the likelihoods are normal pdfs with unequal (although very similar) standard deviations. In practice, it does not make sense to not repair when the observed potential is very negative. The unsuitability of normal likelihoods for some NDE systems has been highlighted by Green and Swets (1966). The left bound s_1^* in Eq. (41) is, however, very far from the expected values of the signal S, and does not significantly affect the results.

The expected cost associated with the a posteriori optimal action is evaluated by numerical integration from Eq. (15) and is €1.4 million.

The VoI of this NDE system is $2.5 - 1.4 = €1.1$ million, equivalent to a gain of 45% relative to the a priori expected cost.

5.2.3. Optimal a posteriori action for a binary signal with decision threshold s_{th}

Here, the continuous signal S is processed to a binary outcome $\{I = 0\} = \{S > s_{th}\}$ (no detection) and $\{I = 1\} = \{S < s_{th}\}$ (detection), i.e., $Z = I$, and $\mathcal{L}(Y = 1; I = 1) = PoD(s_{th})$ and $\mathcal{L}(Y = 0; I = 1) = PFA(s_{th})$.

As in Eq. (39), the optimal actions are determined by a condition on PoD and PFA:

$$a_{opt}(I = 1) = a_R \iff \frac{PoD}{PFA} > 2.11, \tag{42}$$

$$a_{opt}(I = 0) = a_R \iff \frac{1-PoD}{1-PFA} > 2.11. \tag{43}$$

Fig. 15 shows the expected cost given by Eq. (15) as a function of PFA and PoD. It also represents the ROC curve obtained from the pdfs of Eqs. (37) and (38).

Fig. 15 also locates the cut-off point of the Youden Index, $s_{th} = -0.28[\text{Volt}]$, which coincides for this ROC curve with the point closest to the origin. These are suboptimal operating points with an associated expected cost of €1.5 million, which is higher than the optimal cost found with Model (1).

The optimal threshold on the ROC curve that minimizes the expected cost is $s_{th,fix} = -0.31[\text{Volt}]$, with associated cost €1.4 million. We note that this threshold is equal to bound s_2^* found in Section 5.2.2. For this setup, Model (3) and Model (4) achieve the same performance without loss of information if the threshold is optimized for the problem at hand.

We can also evaluate the NDE system with cut-off point $-0.2515[\text{Volt}]$ that Faber and Sorensen (2002) directly apply on the measured potential for the repair decision criteria. It is shown on Fig. 15. The expected total cost associated to that threshold is €1.9 million, which is 34% more than the cost found with Model (3). It still yields a non-zero VoI.
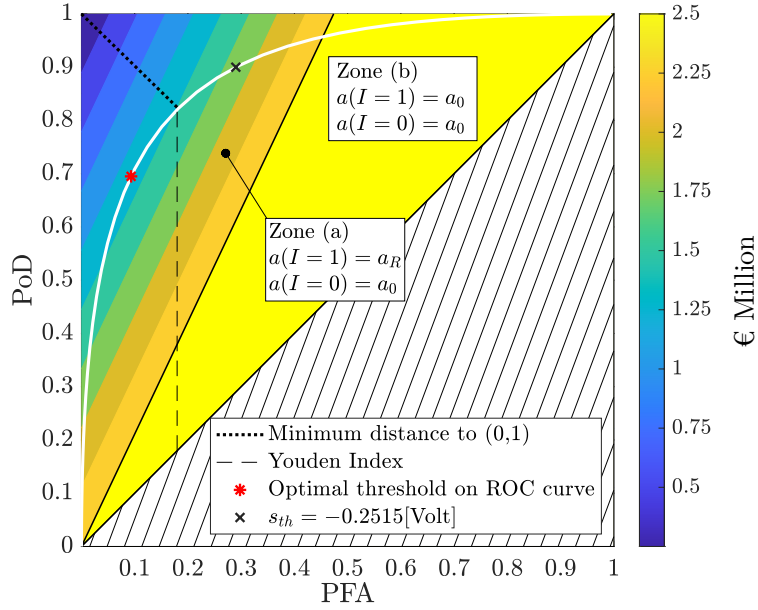
Figure 15: Expected cost as a function of *PFA* and *PoD*, corresponding optimal actions, and the ROC curve in white. In zone (b), the selected action is a_0, hence the VoI is 0. The hatched area corresponds to a monitoring system where the significance of $I = 0$ and $I = 1$ are inverted. The optimal point on the ROC curve obtained by Bayesian analysis differs from the Youden Index threshold and the point closest to (0,1). As a comparison, the threshold -0.2515[Volt] recommended by Faber and Sorensen (2002) is indicated.

5.3. Two-step decision problem

5.3.1. Decision model and a priori optimal actions

We now consider a two-step decision problem. The repair actions are taken at each step based on the NDE data obtained up to that step. Additionally, the actions affect the state of the system at the next step. The initial condition of the system is Y_1. The effect of the action $A_1 \in \{a_0, a_R\}$ at time step 1 on the state of the system is reflected in Y_1'. The condition of the system can evolve from Y_1' to Y_2 at time step 2. Both Y_1 and Y_2 are binary, with states "corroded" or "not corroded." Similarly, an action A_2 is taken that affects the state of the system, resulting in Y_2'. The immediate consequence of the concrete being corroded at time step i is $C_{F,i}(Y_i' = 1) = c_F$, where $c_F = €50$ million. If $Y_i' = 0$, there are no associated consequences, i.e., $C_{F,i}(Y_i' = 1) = €0$. The action cost is $C_{A,i}(a_R) = c_R = €5$ million and $C_{A,i}(a_0) = €0$.

The random variable Y_1 is characterized by $\Pr(Y_1 = 1) = 0.1$. The conditional probabilities reflecting the decision taken at time step i are given by $\Pr(Y_i'|A_i = a_0, Y_i \in \{0,1\}) = \mathbb{1}_{Y_i'=Y_i}$, where $\mathbb{1}$ is the indicator function, and $\Pr(Y_i' = 1|A_i = a_R, Y_i \in \{0,1\}) = 0$. The transition from Y_1' to Y_2 is expressed by the conditional probability in Table 4.
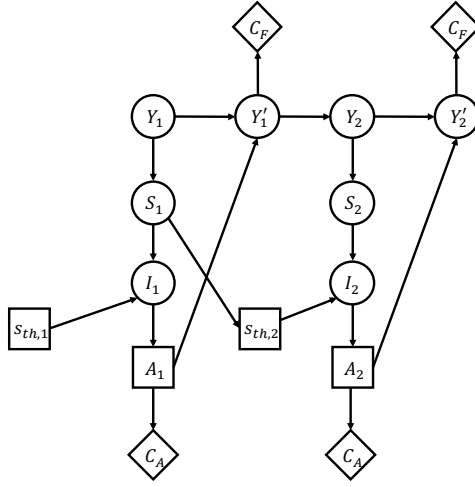
We solve the two-step decision problem assuming three models of NDE quality. First, the NDE quality is described by Model (3) (Fig. 16a) and the thresholds and actions are optimized sequentially. The second quality model is Model (4), where a fixed point on the ROC curve in Fig. 14b is given (Fig. 16b). Finally, we optimize the point on the ROC curve for this specific applica-

Table 4: Conditional probability $\Pr(Y_2|Y_1')$.

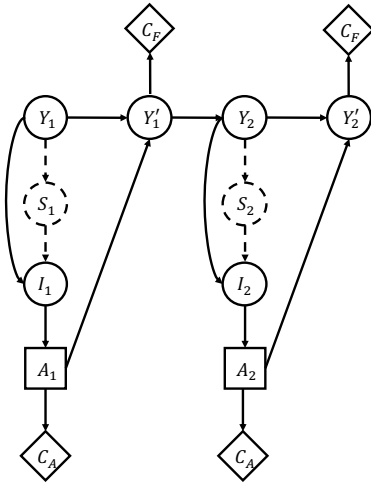|	$Y_1' = 0$	$Y_1' = 1$
$Y_2 = 0$	0.95	0
$Y_2 = 1$	0.05	1

tion (Fig. 16c). The details of the derivations are omitted here but are included as Supplementary Material.
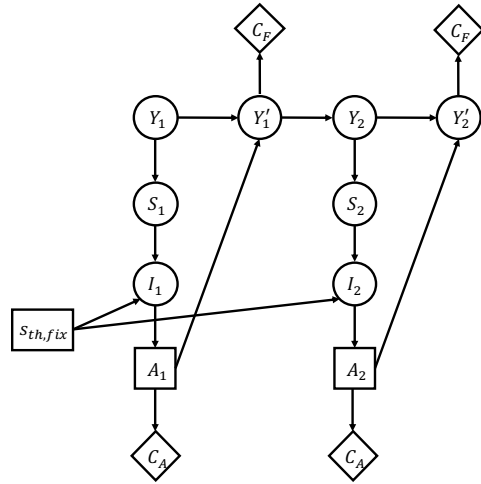
Optimal actions for Model (3)

(a)

Fixed threshold s_{th} for Model (4)

Optimized threshold $s_{th,fix}$ for Model (4)

(b)

(c)

Figure 16: Influence diagrams of the two-step decision problem where the quality of NDE is described: (a) by Model (3), where the thresholds $s_{th,i}$ such that $\{I_i = 1\} = \{S_i < s_{th,i}\}$ are optimized sequentially; (b) by Model (4) with a given threshold defining the likelihood $\Pr(I|Y)$; and (c) by Model (4), where the fixed threshold is optimized for the specific application.

The optimal actions based only on prior knowledge are to first repair at time step 1, and then do nothing at time step 2, i.e., $(a_1, a_2)_{e_0} = (a_R, a_0)$. The associated expected cost is €7.5 million.

5.3.2. Optimal a posteriori actions considering the continuous signal

The actions are first optimized sequentially, based on the continuous half-cell potential measurements S_1 and S_2 at time step 1 and 2, respectively. This means that the NDE method is represented by Model (3).

This multi-step sequential decision problem with NDE results is in general difficult to resolve. A heuristic approach to the optimization can provide an approximate solution (Bismut and Straub, 2021). For this two-step problem, an analytical solution is obtained by backward induction (Bellman, 1957).

The optimal actions are shown in Fig. 17, as a function of the measured signal S_1 and S_2 obtained sequentially. We can see that the optimal solution is equivalent to applying a threshold $s_{th,1} = -0.28$[Volt] on S_1, then a threshold $s_{th,2}$ on S_2 that depends on S_1, as illustrated by Fig. 16a. We see that if $S_1 > s_{th,1}$, the optimal threshold at time step 2 depends on the outcome at time step 1. When the system is repaired at time step 1, the optimal decision $a_{2,opt}$ is independent of outcome S_1.
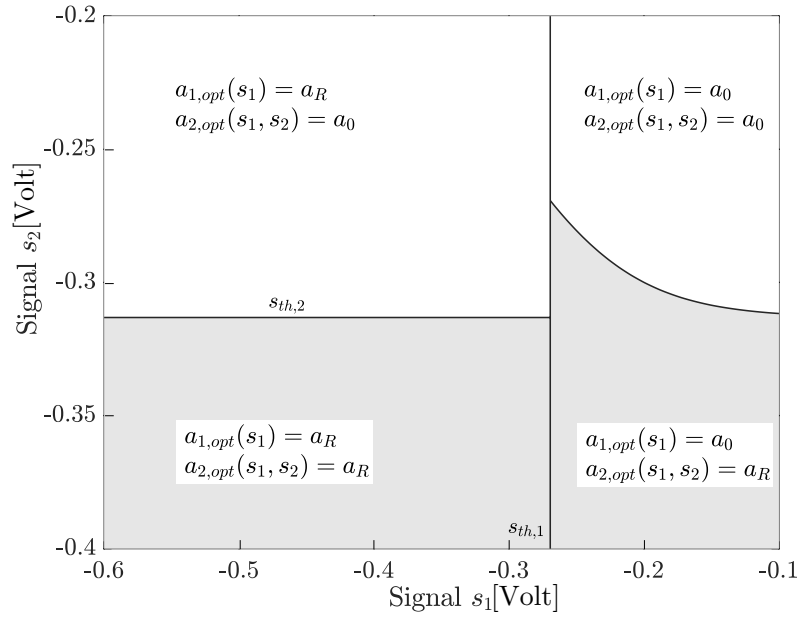

Figure 17: Optimal policy given monitoring history S_1 and S_2. The optimal threshold on signal S_1 is $s_{th,1} = -0.27$[Volt]. The optimal threshold $s_{th,2}$ on signal S_2 depends on the observation S_1.

The expected total cost of reacting optimally at each time step to the continuous signal is €3.7 million. The VoI for this NDE system is $7.5 - 3.7 = $ €3.8 million. This result shows that considering the continuous signal is more optimal compared to classifying the signal according to a fixed rule.

5.3.3. Optimal a posteriori actions for a binary signal with decision threshold

In this configuration, the NDE system is described by Model (4). The corresponding influence diagram is drawn in Fig. 16b. We investigate, for example, the fixed threshold -0.2515[Volt] from Faber and Sorensen (2002). The NDE system is described by the point $PoD = 0.90$ and $PFA = 0.29$ on the ROC curve (Fig. 14b). The expected total cost is €4.2 million and the VoI is $7.5 - 4.2 =$ €3.3 million. This fixed threshold -0.2515[Volt] is not optimal when compared to the VoI of Model (3) above.

Using Model (3) for the considered NDE method provides the optimal actions for the two-step decision problem. However, one can optimize the fixed threshold and point on the ROC curve for Model (4), as illustrated by the modified influence diagram of Fig. 16c. To do so, we first determine sequentially the optimal actions $a_{1,opt}(I_1)$ and $a_{2,opt}(I_1, I_2)$, for any PFA and PoD values. The resulting expected cost is depicted in Fig. 18. The optimal threshold is found on the ROC curve that maximizes the expected cost along the curve. One finds $s_{th,fix} = -0.28$[Volt] and the associated expected cost is €3.8 million. The VoI for this NDE system is $7.5 - 3.8 =$ €3.7 million. The ROC curve and optimal operating point are also shown in Fig. 18. As a comparison, the previously evaluated threshold -0.2515[Volt] is also plotted on the ROC curve.
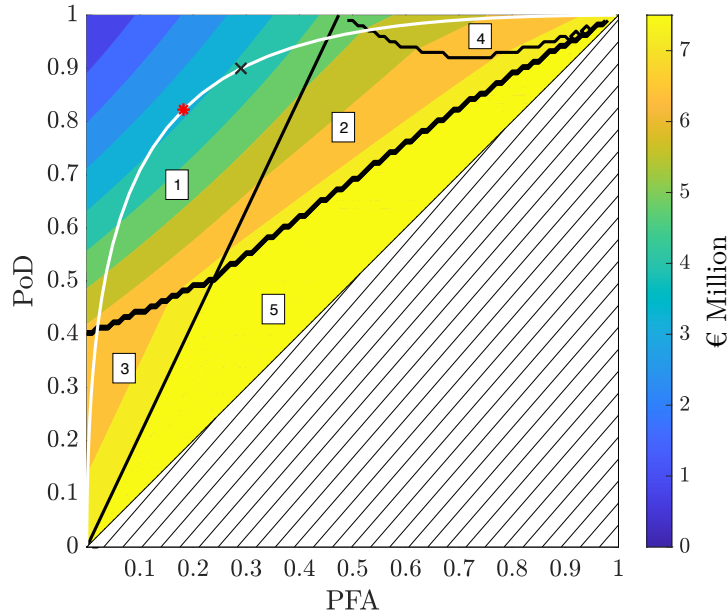

Figure 18: Expected total cost as a function of *PFA* and *PoD*. The ROC curve is in white. The areas numbered 1 to 5 correspond to action zones. In zone 1, $a_{i,opt}(I_i = 1) = a_R$, $a_{i,opt}(I_i = 0) = a_0$; in zone 2, $a_{1,opt}(I_1 = 1) = a_R$, $a_{1,opt}(I_1 = 0) = a_0$, $a_{2,opt}(I_2 = 1, I_1 = 0) = a_R$, $a_{2,opt}(I_2 = 0, I_1 = 0) = a_0$, $a_{2,opt}(I_1 = 1) = a_0$; in zone 3, $a_{1,opt} = a_R$, $a_{2,opt}(I_2 = 0) = a_0$, $a_{2,opt}(I_2 = 1) = a_R$; in zone 4, $a_{1,opt}(I_1 = 1) = a_R$, $a_{1,opt}(I_1 = 0) = a_0$, $a_{2,opt} = a_0$; in zone 5, $a_{1,opt} = a_R$, $a_{2,opt} = a_0$. Zone 5 corresponds to *VoI = 0*. The red asterisk locates the optimal fixed operating point with threshold $s_{th,fix} = -0.28$[Volt]. As a comparison, the cross locates the point for threshold -0.2515[Volt] from (Faber and Sorensen, 2002).

For given *PFA* and *PoD* values, the optimal actions of the two-step problem with Model (4) are defined by the zone, numbered 1 to 5, to which the *PFA* and *PoD* values belong in Fig. 18. In

zone 1, the optimal actions are condition-based and memoryless, which corresponds to $a_{opt,i}(I_i = 1) = a_R$ and $a_{opt,i}(I_i = 0) = a_0$. This is not true when the *PFA* and *PoD* values belong to zones 2 to 5. It is therefore necessary to solve the decision problem, as any point on the ROC curve does not result in optimal actions that are condition-based and memoryless. This conclusion echoes the results by Bertovic (2016), who warns that the outcomes of mechanized NDE may be blindly trusted and interpreted as a call to action, to the detriment of a formal information assessment.

The effect of ignoring the different action zones is illustrated in Fig. 19. This figure compares the optimal expected total costs along the ROC curve and the expected total cost of following a condition-based and memoryless maintenance strategy (as in zone 1). The curve corresponding to the optimal expected cost along the ROC curve stays within the lower and upper bounds, which are respectively the optimal expected cost with Model (3) at €3.7 million (Section 5.3.2) and the optimal a priori expected cost of €7.5 million (Section 5.3.1). In contrast, implementing a condition-based and memoryless maintenance strategy with a fixed NDE threshold, for example -0.4[Volt] in zone 3, results in an expected cost of €8 million. This cost is higher than the expected cost of €6.9 million associated with the optimal actions for Model (4) with this same threshold, and even higher than the a priori expected cost.
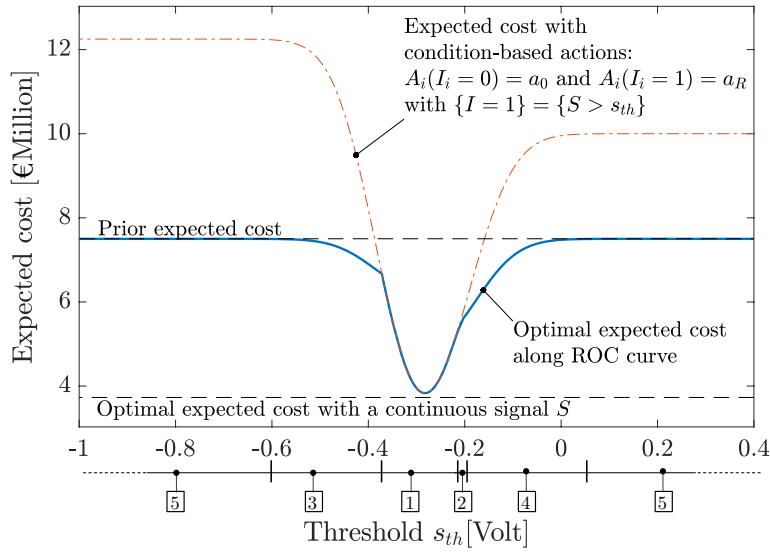

Figure 19: Expected cost associated with a condition-based and memoryless maintenance strategy and with the optimal actions for Model (4). The action zones, 1 to 5, of Fig. 18 are indicated as a function of the threshold value. For certain threshold values, the expected cost associated with the condition-based memoryless maintenance strategy is higher than the prior expected cost.

6. Concluding remarks

This paper reviews existing models NDE quality in the context of a unifying framework .

We have clarified the connection between the different models of NDE quality, such as PoD curves or ROC curves, through a base model in which both the condition and the NDE outcome are

modeled continuously. The models with a binary observed signal or binary condition are derived by imposing thresholds on the observed signal and the condition. For models that use a binary condition, one must also impose a distribution on the underlying continuous condition.

In practice, there currently is no systematic approach to choose and calibrate NDE models and handle NDE data for reliability assessment or maintenance planning (Kurz et al., 2012). The presented framework can support the development and appropriate application of NDE models to real-life settings. Ideally, the base model is learned from experimental data and any other model can then be derived from this base model. When the base model is not available, it can still be considered at an abstract level to appraise the quality of given NDE models. As we demonstrate, the use of given PoD curves or ROC curves are likely to yield suboptimal actions for a specific decision context, since the signal and condition thresholds are fixed without considering the decision problem. The framework also highlights the importance of explicitly accounting for the experimental design conditions that underlie ROC curve models, which should be assessed within the decision context.

Calibration of an NDE system in function of the decision settings (e.g., cost model) is beneficial. Such a calibration is often done implicitly. In particular, on-site inspectors tend to report a larger damage when the occurrence of failure is cost-critical, and vice-versa will be more cautious to diagnose damage when the costs of repair are very high (da Silva and de Padua, 2012; Bertovic, 2016). This paper derives the formal analysis associated with these decisions. Bayesian decision analysis provides the means to compute the VoI, which allows direct comparison between NDE systems. This analysis gives the opportunity to calibrate an NDE system to suit the decision parameters such as the cost of mitigating actions and the expected consequences of failure. In Sections 4 and 5 we have evaluated the VoI in simple one-step and two-step decision problems. We find that using the model with continuous variables leads to the best decisions. In contrast, systematically associating NDE system outcomes with maintenance actions can lead to sub-optimal decisions and a detrimental use of the inspection data.

Finding the VoI of NDE systems in sequential decision problems remains, however, a challenging task. The formal and complete decision analysis for an NDE system applied to continuous monitoring is not tractable (Papadimitriou and Tsitsiklis, 1987). As an alternative, a decision rule, or heuristic, can be applied for the treatment of information, and some heuristic parameters optimized such that the information from NDE system is collected, interpreted and exploited in a good manner (Sheils et al., 2010; Bismut and Straub, 2021).

Acknowledgments

Supplementary Material

The following supplementary material is available online at `https://arxiv.org/src/2103.12853v1/anc/Supplementary_Material.pdf`: **A**. Solution of the two-step decision problem of Example 2.

References

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

Annis, C., 2009. Nondestructive evaluation system reliability assessment. Report MIL-HDBK-1823A, Department of Defense.

Bellman, R., 1957. Dynamic Programming. Princeton University Press.

Berens, A. P., 2000. Probability of Detection (POD) Analysis for the Advanced Retirement for Cause (RFC)/Engine Structural Integrity Program (ENSIP) Nondestructive Evaluation (NDE) System-Volume 1: Pod Analysis. Report, University of Dayton Research Institute.

Berens, A. P., Hovey, P. W., 1981. Evaluation of nde reliability characterisation. Air Force Wright Aeronautical Laboratories, AFWAL-TR-81-4160, November 1981.

Bertovic, M., 2016. A human factors perspective on the use of automated aids in the evaluation of ndt data. In: AIP Conference Proceedings. Vol. 1706. AIP Publishing LLC, p. 020003.

Bismut, E., Straub, D., 2021. Optimal adaptive inspection and maintenance planning for deteriorating structural systems. Reliability Engineering & System Safety 215, 107891.

Boero, J., Schoefs, F., Capra, B., 2009. Effect on risk based inspection of spatio-temporal dependence of roc curves: Study of the corrosion of steel harbour structures. In: NDTCE 09.

Brennan, F., 2013. Risk based maintenance for offshore wind structures. Procedia CIRP 11, 296–300.

Calmon, P., 2012. Trends and stakes of NDT simulation. Journal of Nondestructive evaluation 31 (4), 339–341.

Calmon, P., Chapuis, B., Jenson, F., Sjerve, E., 2016. The use of simulation in POD curves estimation: An overview of the IIW best practices proposal. In: 19th World Conference on Non-Destructive Testing. Munich, Germany.

Clark, R., Dover, W. D., Bond, L. J., 1987. The effect of crack closure on the reliability of NDT predictions of crack size. NDT international 20 (5), 269–275.

da Silva, R. R., de Padua, G. X., 2012. Nondestructive inspection reliability: state of the art. Nondestructive Testing Methods and New Applications, 1.

Deutsches Institut für Normung, 2014. In-service inspections for primary coolant circuit components of light water reactors - Part 6: Eddy current testing of steam generator heating tubes. Report DIN 25435-6:2014-01.

Elsener, B., Andrade, C., Gulikers, J., Polder, R., Raupach, M., 2003. Half-cell potential measurementspotential mapping on reinforced concrete structures. Materials and Structures 36 (7), 461–471.

Faber, M. H., Sorensen, J. D., 2002. Indicators for inspection and maintenance planning of concrete structures. Structural Safety 24 (2-4), 377–396.

Feistkorn, S., Taffe, A., 2011. Die POD-eine Vorgehensweise zum qualitativen Gütenachweis zerstörungsfreier Prüfverfahren im Bauwesen am Beispiel des Impulsradars. In: DGZfP-Jahrestagung 2011, Bremen, Germany.

Finney, D. J., 1978. Statistical method in biological assay, 3rd Edition. Griffin, High Wycombe.

Fluss, R., Faraggi, D., Reiser, B., 2005. Estimation of the Youden Index and its associated cutoff point. Biometrical Journal: Journal of Mathematical Methods in Biosciences 47 (4), 458–472.

Frangopol, D. M., Lin, K.-Y., Estes, A. C., 1997. Life-cycle cost design of deteriorating structures. Journal of Structural Engineering 123 (10), 1390–1401.

Goulet, J.-A., Der Kiureghian, A., Li, B., 2015. Pre-posterior optimization of sequence of measurement and intervention actions under structural reliability constraint. Structural Safety 52, 1–9.

Graham, T., Tetelman, A., 1974. The use of crack size distribution and crack detection for determining the probability of fatigue failure. In: 15th Structural Dynamics and Materials Conference. p. 394.

Granville, R., Charlton, P., 2016. Detection and sizing of baffle plate erosion and fretting using eddy current array technology. ndt. net 21 (09).

Green, D. M., Swets, J. A., 1966. Signal detection theory and psychophysics. Vol. 1. John Wiley & Sons, Inc.

Greiner, M., Pfeiffer, D., Smith, R. D., 2000. Principles and practical application of the receiver-operating characteristic analysis for diagnostic tests. Preventive veterinary medicine 45 (1-2), 23–41.

Hamida, Z., Goulet, J., 2020. Modeling infrastructure degradation from visual inspections using networkscale statespace models. Structural Control and Health Monitoring 27 (9), e2582.

Harris, D. O., Lim, E. Y., 1983. Applications of a probabilistic fracture mechanics model to the influence of in-service inspection on structural reliability. ASTM International, pp. 19–41.

Hayward, G., Pearson, J., Stirling, G., 1993. An intelligent ultrasonic inspection system for flooded member detection in offshore structures. IEEE transactions on ultrasonics, ferroelectrics, and frequency control 40 (5), 512–521.

Heasler, P. G., Doctor, S. R., 1996. Piping inspection round robin. Report, U.S. Nuclear Regulatory Commission, Washington, DC (United States).

Hong, H. P., 1997. Reliability analysis with nondestructive inspection. Structural Safety 19 (4), 383–395.

Hovey, P. W., Berens, A. P., 1988. Statistical evaluation of NDE reliability in the aerospace industry. Springer, pp. 1761–1768.

International Atomic Energy Agency, 1965. Non-Destructive Testing in Nuclear Technology Vol I Proceedings of a Symposium on Non-Destructive Testing in Nuclear Technology. IAEA, International Atomic Energy Agency (IAEA), (Austria).

Jarvis, R., Cawley, P., Nagy, P. B., 2018. Permanently installed corrosion monitoring using magnetic measurement of current deflection. Structural Health Monitoring 17 (2), 227–239.

Kerntechnischer Ausschuss, 2016. Safety Standards of the Nuclear Safety Standards Commission (KTA) .

Kessler, S., Gehlen, C., 2017. Reliability evaluation of half-cell potential measurement using POD. Journal of Infrastructure Systems 23 (2), B4016007.

Kurz, J. H., Jüngert, A., Dugan, S., Dobmann, G., 2012. Probability of detection (pod) determination using ultrasound phased array for considering ndt in probabilistic damage assessments. In: South-African Insitute for Nondestructive Testing: World Conference on Nondestructive Testing (18), WCNDT. Citeseer.

Lavender, J. D., 1976. Ultrasonic testing of steel castings. Steel Founders' Society of America.

Lentz, A., Faber, M. H., Johnsen, T., 2002. Half-cell potential measurements for condition assessment. In: Proceeding of 1st International Conference on Bridge Maintenance.

Luque, J., Straub, D., 2019. Risk-based optimal inspection strategies for structural systems using dynamic Bayesian networks. Structural Safety 76, 68–80.

Madsen, H. O., 1987. Model updating in reliability theory. In: Proc. ICASP5. Vancouver, Canada, pp. 565–577.

Madsen, H. O., Skjong, R. K., Tallin, A. G., 1987. Probabilistic fatigue crack growth analysis of offshore structures, with reliability updating through inspection. In: Marine Structural Reliability Engineering Symposium. pp. 45–55.

Nath, S., 2021. Estimates of probability of detection and sizing of flaws in ultrasonic time of flight diffraction inspections for complex geometry components with grooved surfaces. Journal of Nondestructive Evaluation, Diagnostics and Prognostics of Engineering Systems 4 (2), 021003.

National Materials Advisory Board, 1969. Nondestructive evaluation. Report NMAB-252.

Nielsen, J. J., Sørensen, J. D., 2010. Risk-based operation and maintenance planning for offshore wind turbines. In: Proceedings of the Reliability and Optimization of Structural Systems. pp. 7–10.

Olin, B. D., Meeker, W. Q., 1996. Applications of statistical methods to nondestructive evaluation. Technometrics 38 (2), 95–112.

Onoufriou, T., Frangopol, D. M., 2002. Reliability-based inspection optimization of complex structures: a brief retrospective. Computers & Structures 80 (12), 1133–1144.

Packman, P. F., Pearson, H. S., Owens, J. S., Marchese, G. B., 1968. The applicability of a fracture mechanics-nondestructive testing design criterion. Report, Lockheed-Georgia Company Material Sciences Research Laboratory.

Pakrashi, V., Schoefs, F., Memet, J. B., O'Connor, A., 2010. ROC dependent event isolation method for image processing based assessment of corroded harbour structures. Structures & Infrastructure Engineering 6 (3), 365–378.

Papadimitriou, C. H., Tsitsiklis, J. N., 1987. The complexity of Markov decision processes. Mathematics of operations research 12 (3), 441–450.

Pavlovic, M., Takahashi, K., Müller, C., Boehm, R., Ronneteg, U., 2008. Reliability in non-destructive testing (NDT) of the canister components. NDT Reliability - Final Report. Report, SKB Technical report R-08-129, ISSN 1402-3091, Swedish Nuclear Fuel and Waste Management Co.

Peterson, W. W., Birdsall, T. G., 1953. The theory of signal detectability: Part I. The general theory. Report, Technical Report 13, June 1953. Electronic Defense Group, University of Michigan.

Quirk, L., Matos, J., Murphy, J., Pakrashi, V., 2018. Visual inspection and bridge management. Structure and Infras-

tructure Engineering 14 (3), 320–332.

Raiffa, H., Schlaifer, R., 1961. Applied Statistical Decision Theory. Harvard Business Review Press.

Rajesh, S. N., 1993. Probability of detection models for eddy current NDE methods. Report, Ames Lab., IA (United States).

Ritz, C., Baty, F., Streibig, J. C., Gerhard, D., 2015. Dose-response analysis using R. PloS ONE 10 (12).

Rouhan, A., Schoefs, F., 2003. Probabilistic modeling of inspection results for offshore structures. Structural Safety 25 (4), 379–399.

Rudemo, M., Ruppert, D., Streibig, J., 1989. Random effect models in nonlinear regression with applications to bioassay. Biometrics 45 (2), 349–362.

Sarkar, P., Meeker, W. Q., Thompson, R. B., Gray, T. A., Junker, W., 1998. Probability of detection modeling for ultrasonic testing. In: D.O. Thompson, D. C. (Ed.), Review of Progress in Quantitative Nondestructive Evaluation. Vol. 17. Springer, pp. 2045–2052.

Schoefs, F., Boéro, J., Clément, A., Capra, B., 2012. The $\alpha\delta$ method for modelling expert judgement and combination of non-destructive testing tools in risk-based inspection context: application to marine structures. Structure and Infrastructure Engineering 8 (6), 531–543.

Schoefs, F., Clement, A., 2004. Multiple inspection modeling for decision making and management of jacket off-shore platforms: effect of false alarms. In: International Forum on Engineering Decision Making, Switzerland.

Shannon, C. E., 1948. A mathematical theory of communication. Bell system technical journal 27 (3), 379–423.

Sheils, E., O'Connor, A., Schoefs, F., Breysse, D., 2012. Investigation of the effect of the quality of inspection techniques on the optimal inspection interval for structures. Structure and Infrastructure Engineering 8 (6), 557–568.

Sheils, E., OConnor, A., Breysse, D., Schoefs, F., Yotte, S., 2010. Development of a two-stage inspection process for the assessment of deteriorating infrastructure. Reliability Engineering & System Safety 95 (3), 182–194.

Sindel, R., Rackwitz, R., 1996. Calculation of the failure probability and inspection optimization for cracks in ships. Technische Universität München.

Singh, R., 2000. Three decades of NDI reliability assessment. Contract 41608 (99-C), 404.

Somoza, E., Mossman, D., McFeeters, L., 1990. The INFO-ROC technique: a method for comparing and optimizing inspection systems. In: Review of Progress in Quantitative Nondestructive Evaluation. Springer, pp. 601–608.

Straub, D., 2004. Generic approaches to risk based inspection planning for steel structures. Vol. 284. vdf Hochschulverlag AG.

Straub, D., 2011. Reliability updating with equality information. Probabilistic Engineering Mechanics 26 (2), 254–258.

Straub, D., 2014. Value of information analysis with structural reliability methods. Structural Safety 49, 75–85.

Swets, J. A., 1983. Assessment of NDT systems. I. The relationship of true and false detections. II. Indices of performance. Materials Evaluation 41 (11), 1294–1303.

Swets, J. A., 1992. The science of choosing the right decision threshold in high-stakes diagnostics. American Psychologist 47 (4), 522.

Taner, T., Antony, J., 2000. The assessment of quality in medical diagnostic tests: a comparison of ROC/Youden and Taguchi methods. International Journal of Health Care Quality Assurance.

Tang, W. H., 1973. Probabilistic updating of flaw information. Journal of Testing and Evaluation 1 (6), 459–467.

Tschöke, K., Mueller, I., Memmolo, V., Moix-Bonet, M., Moll, J., Lugovtsova, Y., Golub, M., Venkat, R. S., Schubert, L., 2021. Feasibility of model-assisted probability of detection principles for structural health monitoring systems based on guided waves for fibre-reinforced composites. IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control.

Virkkunen, I., Koskinen, T., Papula, S., Sarikka, T., Hänninen, H., 2019. Comparison of â Versus a and Hit/Miss POD-Estimation Methods: A European Viewpoint. Journal of Nondestructive Evaluation 38 (4), 89.

Visser, W. P., 2002. POD/POS Curves for Non-Destructive Examination. Report Offshore Technology Report 2000/018, HSE Books, Norwich, United Kingdom.

Wall, M., Wedgwood, F., 1994. NDT-value for money. Insight 36 (10), 782–90.

Webb, G. T., Vardanega, P. J., Middleton, C. R., 2015. Categories of shm deployments: technologies and capabilities. Journal of Bridge Engineering 20 (11), 04014118.

Wood, H. A., Engle Jr, R. M., 1979. USAF Damage Tolerant Design Handbook: Guidelines for the analysis and

Design of Damage Tolerant Aircraft Structures. Revision A. Report, Air Force Flight Dynamics Lab Wright-Patterson Air Force Base, Ohio.

Woodward, P. M., Davies, I. L., 1952. Information theory and inverse probability in telecommunication. Proceedings of the IEE-Part III: Radio and Communication Engineering 99 (58), 37–44.

Yang, J. N., Chen, S., 1985. Fatigue reliability of gas turbine engine components under scheduled inspection maintenance. Journal of Aircraft 22 (5), 415–422.

Yang, J.-N., Trapp, W. J., 1974. Reliability analysis of aircraft structures under random loading and periodic inspection. AIAA journal 12 (12), 1623–1630.

Youden, W. J., 1950. Index for rating diagnostic tests. Cancer 3 (1), 32–35.