

UNCERTAINTY QUANTIFICATION FRAMEWORK WITH INTERDEPENDENT
DYNAMICS OF DATA, MODELING, AND LEARNING IN NONDESTRUCTIVE
EVALUATION

By

Zi Li

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

Electrical and Computer Engineering—Doctor of Philosophy

2023

ABSTRACT

Even after extensive efforts to enhance our understanding of materials, modeling, and system processes, uncertainty continues to be an inevitable factor that impacts system behavior, especially at the operational limits. The evaluation of uncertainty is now a common practice in engineering and scientific fields, encompassing the analysis of experimental data, as well as numerous computational models and process simulations. Non-destructive evaluation (NDE) techniques are widely utilized across a range of industries and applications to guarantee the safety, quality, and dependability of components, systems, and structures. However, NDE processes are often challenged by uncertainties stemming from factors such as material variations, environmental conditions, and measurement limitations, which can introduce complexities into the assessment process. Therefore, there is a need to quantify uncertainties in NDE, which can enhance our comprehension of the constraints and potential inaccuracies linked to NDE inspections and aid in making NDE assessments more robust and reliable. In this thesis, a comprehensive uncertainty quantification (UQ) framework: the Three-Legged Stool (TLS) is proposed to provide systematic guidance in uncertainty analysis for NDE applications.

A Magnetic Flux Leakage (MFL) based defect characterization algorithm is proposed to classify the defect and handling uncertainties for pipeline inspection. The research compares Convolutional Neural Network (CNN) and Deep Ensemble (DE) methods for handling input uncertainties from MFL response data, while also employing Autoencoder for data augmentation to address limited experimental data. The study evaluates prediction accuracy and explores uncertainty analysis, emphasizing the importance of reliability assessment in MFL-based NDE decision-making.

To estimate the fatigue life of martensitic-grade stainless-steel turbine blades, a magnetic Barkhausen noise (MBN) technique is applied. This work involves the extraction of time and frequency domain features, followed by the application of techniques such as Principal Component Analysis (PCA) and probabilistic neural network (PNN) for classifying and estimating the remaining fatigue life.

An IMU-assisted robotic Structured light (SL) sensing system was developed for pipeline

detection. This system improves registration and defect estimation through a RANSAC-assisted cylindrical fitting algorithm, integrates inertial and odometry measurements for precise 3D profiling, and employs customized defect sizing techniques to offer a reliable 3D defect reconstruction solution for various defect shapes and depths.

The proposed TLS-based UQ framework highlights the interdependent dynamics among data, models, and learning when addressing uncertainties in NDE processes. Some advanced and commonly used techniques have been introduced to illustrate how uncertainties in the inputs or parameters of an NDE system, model, or measurement are propagated to the outputs or predictions. The uncertainty propagation is considered in terms of the forward modeling and inverse learning process separately. In order to demonstrate the efficiency and applicability of the proposed framework for NDE applications, the uncertainties in the previously mentioned NDE cases are investigated and quantified using the techniques outlined in the TLS model.

In summary, the proposed UQ framework is able to provide guidance in dealing with uncertainties in NDE inspection with efficient and reliable solutions. It holds great promise and opens up avenues for further research and advancement within the industry.

Copyright by
ZI LI
2023

I would like to express my dedication to this work to my parents, my aunt's family, my friend, and my entire family for their unconditional support throughout my life's journey, as well as their encouragement to confront and overcome all challenges, which have contributed to shaping the person I am today.

ACKNOWLEDGEMENTS

I want to express my sincere appreciation to all those who have been by my side and provided assistance throughout my doctoral journey. I extend my thanks to my advisor and mentor, Dr. Yiming Deng, for his exceptional support, guidance, and motivation during my doctoral studies. He consistently believed in me and guided not only my research but also my career and personal life. It is a great fortune for me to work with such a knowledgeable and caring advisor. I am grateful to my doctoral committee: Dr. Lalita Udpa, Dr. Xiaobo Tan, Dr. Ming Han, and Dr. Chih-Li Sung, for their valuable time and suggestions. I'd also like to express my gratitude to my fellow colleagues at NDEL for their help, support, and the insightful discussions we've had. Lastly, I'd like to express my deep gratitude to my family and friends for their support and encouragement.

TABLE OF CONTENTS

CHAPTER 1	INTRODUCTION	1
1.1	Uncertainty Quantification	1
1.2	Uncertainty Analysis in NDE	3
1.3	Objective and Scope	5
CHAPTER 2	UNCERTAINTY IN THREE-LEGGED STOOL FRAMEWORK	7
2.1	Uncertainty Sources in TLS UQ Framework	7
2.2	Uncertainty Propagation in Three-Legged Stool (TLS) UQ Framework	13
CHAPTER 3	FORWARD MODELING BASED UNCERTAINTY PROPAGATION	15
3.1	Methods of Uncertainty Propagation in Forward Modeling	15
3.2	TLS-based Forward Uncertainty Modeling Application	19
3.3	Conclusion	27
CHAPTER 4	INVERSE LEARNING BASED UNCERTAINTY PROPAGATION	29
4.1	Probability Theory for Inverse Uncertainty Quantification	29
4.2	Methods of Uncertainty Propagation in Physics-informed Learning	31
4.3	Methods of Uncertainty Propagation in Data-driven Learning	33
4.4	Methods of Uncertainty Propagation in Hybrid Learning	34
4.5	TLS-Based Inverse Uncertainty Learning Application	40
4.6	Conclusion	62
CHAPTER 5	RELIABILITY EVALUATION TO NDE PROCESS WITH UQ	65
5.1	Probability of Detection	65
5.2	GUM-based Measurement Uncertainty Evaluation	68
5.3	Magnetic Barkhausen Noise-based Material Fatigue Detection	72
5.4	Structure Light Sensing based Defect Reconstruction	84
5.5	Conclusion	103
CHAPTER 6	CONCLUSION AND FUTURE WORK	105
6.1	Conclusion	105
6.2	Future Work	107
BIBLIOGRAPHY		109

CHAPTER 1

INTRODUCTION

1.1 Uncertainty Quantification

Despite years of efforts to enhance knowledge of the material, modeling, and systems processes, uncertainty remains an unavoidable element affecting the behavior of systems and more so with respect to their limits of operation. Uncertainty assessment is becoming increasingly common in fields of engineering and science and includes almost all experimental data processing as well as computational modeling and simulation processes. Therefore, uncertainty analysis is a key component of model-based risk analysis and decision-making, as it provides risk assessors and decision-makers with information about the reliability of model outputs [1]. Uncertainty quantification (UQ) involves the quantitative assessment of uncertainty from various sources (input variables, parameters, or equations) throughout the process and derives the uncertainty distribution for each output variable [2]. The purpose of this analysis is to determine how uncertainty in certain components (eg: inputs, parameters, equations) translates into uncertainty in the final output. For example, it can be applied to compute the probability of an output variable of interest exceeding a certain threshold. The methodology and extent of uncertainty estimation, and result interpretation, vary widely depending on the nature and context of each assessment and the degree of uncertainty that exists. In the domain of prognostics and health management, uncertainties in mathematical and computational models of complex processes and data are analyzed following the steps of representation, quantification, and management [3–6]. Specifically, inspired by [7], the process for modeling and quantifying uncertainty can be divided into three main activities: Uncertainty Categorization, Uncertainty Handling, and Uncertainty Characterization and Estimation.

Uncertainty quantification (UQ) involves identifying and categorizing various sources of uncertainty that may affect prognostics and sensitivity analysis. It is an important step to incorporate these sources of uncertainty into models and simulations as accurately as possible. The goal of this step is to resolve each of these uncertainties separately and quantify them using probabilistic/statistical methods. The most commonly used forms of uncertainty sources are classified as

aleatory (statistical) uncertainty and epistemic (systematic) uncertainty [8–10]. The former refers to the notion of variability and inherent uncertainty, from the natural variability of the physical system. As opposed to aleatoric uncertainty, in theory, epistemic uncertainty is reduced on the basis of additional information, such as quality control [11], structural testing [12], non-destructive inspection [13], and etc.

There are several theories that guide uncertainty handling based on the interpretation of the input uncertainties and the detailed application. The most widely and classical uncertainty theory is probability theory [14]. Under the probability principles, the input parameters are random variables represented using probability distributions. Specifically, with the context of the probability framework, the Bayesian approach provides a more suitable uncertainty interpretation based on prior knowledge and provides a way to update the probability of that event (posterior probability) given new or additional evidence [15]. Besides the empirical probability theory, there are works that shed new light on various interpretations of fuzzy sets and clarify their links with probability theory; conversely, Zadeh's logical point of view on fuzzy sets suggests a set-theoretic perspective on uncertainty measures, that brings together numerical quantification and logic [16]. Evidence theory, also called Dempster–Shafer theory, is proposed as an alternative to the classical probability theory to handle limited and imprecise data situations as an alternative to the classical probability theory [17, 18]. Besides, some researchers extend the notion from risk to uncertainty by invoking the principle of bounded subadditivity: an event has greater impact when it turns impossibility into possibility, or possibility into certainty than when it merely makes a possibility more or less likely. A series of studies provide support for this possibility theory in decisions under both risk and uncertainty and shows that people are less sensitive to uncertainty than to risk [19, 20]. Further, various theories can be combined to take advantage of the ability of each theory. Elishakoff et al. illustrate that the probabilistic analysis and the non-probabilistic, interval analyses are compatible with each other in an applied mechanics problem [21]. Chutia et al. combined imprecise probability and fuzzy knowledge and applied it to atmospheric dispersion with a case study [22]

Uncertainty characterization is then considered to provide a statistical description of the iden-

tified uncertainty sources based on the categorized uncertainty theory. Under the probability framework, if the probability distribution can be determined based on the prior knowledge, it is considered a parametric approach in probabilistic handling; otherwise, for nonparametric methods, the distribution comes only from the data but does not follow a specific distribution [23, 24]. In the parametric approach, identifying the distribution fitting functions is the first step which could be performed based on expert knowledge or a model selection method such as the Bayesian method [25, 26], Bayesian information criterion [27], and Akaike information criterion [28]. Next, some state of art tests, such as chi-squared [29] and Anderson–Darling [30], are applied to fit the parameters of the estimated distribution function. Note that in most scenarios, it is hard to have accurate knowledge of the input probability distribution. Also, if the random variable follows a nonparametric distribution or the given amount of data is insufficient, the nonparametric method is preferred to provide a more reliable prior probability distribution [31]. However, one distinct limitation of the nonparametric study is that the quality is highly correlated with the quality of the given data, therefore, one usually applies some hybrid technique, such as interval analysis, to overcome this problem [32]. Note that uncertainty characterization is the first stage of UQ, further the input uncertainty will be propagated through the analytical or mathematical model to obtain random responses or output. The corresponding uncertainty analysis to input, model, and output construct a full uncertainty quantification procedure [33]. In all state-of-the-art UQ-based techniques, the selection of uncertainty modeling and analysis will vary depending on the application and the problem that needs to be solved.

1.2 Uncertainty Analysis in NDE

Non-destructive evaluation (NDE) is an important aspect of any facility maintenance program while ensuring a product’s integrity, reliability, safety, and long-term productivity by implementing a qualified testing and inspection program. NDE with UQ represents an advanced and holistic approach to the inspection and assessment of materials, components, and structures [34]. It goes beyond traditional NDE by incorporating the rigorous analysis and management of uncertainties that can influence the outcomes of inspections and measurements. By quantifying and addressing

uncertainties, the limitations and potential errors in NDE results can be understood, therefore, more reliable and informed decision-making can be made. Besides, it helps in the early detection of defects or potential failures, allowing for timely maintenance or replacement. Identifying uncertainties in UQ also contributes to cost savings and efficient resource utilization, and unnecessary maintenance or replacement could be avoided by distinguishing between false alarms and genuine issues.

A successful uncertainty analysis of an inspection system requires a deep comprehension of the system itself. Normally, a quantitative NDE system can be described through the forward and inverse modeling approach. A full NDE system is usually analyzed through forward and inverse approaches, which are the fundamental components of the assessment and analysis of materials, components, and structures [35]. In the forward learning process, some established knowledge, models, and methods are used to predict or simulate the expected outcomes of NDE inspections or measurements and make predictions based on known factors such as material properties, inspection parameters, and equipment characteristics. The inverse learning process, on the other hand, involves the analysis of actual NDE data or measurement results to extract meaningful information to infer the material properties, defects, or characteristics of the component being inspected. Some optimization techniques or statistical methods are usually involved to match the observed data with a model, allowing for the identification of defects or property variations [36].

The approach and scope of uncertainty analysis, as well as the interpretation of results, exhibit considerable variability. This variation is contingent on the specific characteristics and context of each NDE and diagnostic procedure, as well as the extent of inherent uncertainty. In the realm of prognostics and health management, the analysis of uncertainties in intricate mathematical and computational models and data follows a structured process encompassing representation, quantification, and management, as highlighted in previous works [3–6]. Additionally, the choice of the NDE method employed is a critical consideration for researchers in proposing effective UQ solutions based on the specific application's requirements. For example, a comprehensive examination of the factors that affect measurement uncertainty in the context of Ultrasonic Testing

(UT) has been proposed in [37]. They considered the influence of test material parameters and the choice of operating parameters, while also referencing relevant standards. In [38], the performance of Guide of Measurements (GUM) and Monte Carlo (MC) methods are compared for the estimation of measurement uncertainty in Electromagnetic (EM) compatibility testing. A hybrid random/fuzzy approach for uncertainty quantification was proposed in electromagnetic modeling, which combines probability and possibility theory to properly account for both aleatory and epistemic uncertainty, respectively [39]. Janousek et al. reduced the uncertainty in depth estimation of partially conductive cracks with eddy current testing [40]. Beyond its traditional applications in material characterization and pipeline inspection, UQ-based NDE has found extensive usage in diverse fields such as medical imaging, geophysics, computer vision, and various other industries [41–43].

1.3 Objective and Scope

As demonstrated in the last section, the context of uncertainty interpretation is varied for different applications and there are multiple ways to describe the uncertainty propagation in different NDE techniques. In this dissertation, I proposed a new framework to redefine the UQ in NDE scope by dividing the sources of uncertainty into data, model, and learning, which is called UQ Three-Legged Stool (TLS) Framework. Related NDE applications will be discussed within the scope to validate the feasibility of the proposed structure, which provides great potential for further NDE-related uncertainty analysis.

Chapter 2 demonstrates the proposed TLS UQ framework by explicitly explaining the uncertainty sources, which is called the "Leg" in this thesis work. Besides, based on the characteristics of the NDE process, a brief introduction to the propagation of uncertainty is presented.

Chapter 3 describes the forward "Modeling" based propagation of uncertainty within the proposed uncertainty sources under the TLS framework. State-of-the-art uncertainty propagation techniques are applied to provide a comprehensive introduction to this process. Besides, a Capacitive sensing-based experimental work is illustrated to validate the feasibility of the proposed TLS UQ framework. The relationship between the "Data" and the output signal is described through a "Learning" process based on a surrogate model, which provides a basis for studying the uncertainty

propagation from the "data" to the "modeling".

Chapter 4 addresses the inverse "Learning" based propagation of uncertainty within the proposed uncertainty sources under the TLS framework. The Bayesian theory, as the most popular foundation for uncertainty analysis, is introduced, which provides a good basis for further uncertainty propagation discussion. Effective methods are introduced to understand and analyze involved uncertainty depending on the characters of different scenarios. Further, a Magnetic flux leakage-based defect classification work is used as a practical example to quantify the uncertainty between "Data" and "Learning". The impact of uncertainties in learning and data is investigated through the Convolutional Neural Network (BCNN) and the Deep Ensemble-based(DE) technique with data-augmented techniques.

Chapter 5 introduced two typical approaches that have been extensively applied in NDE inspections. Probability of Detection (POD) is introduced to evaluate the effectiveness and reliability of defect detection under uncertainties. Additionally, the discussion extends to the topic of measurement uncertainty analysis, which aims to assess and quantify the overall uncertainties associated with NDE measurements. Two NDE applications regarding measurement uncertainty analysis are presented as supporting examples. The first application is for material fatigue detection and prediction using the Magnetic Barkhausen Noise (MBN) technique. In this work, The feasibility of the MBN technique is investigated in detecting early-stage fatigue, which is associated with plastic deformation in ferromagnetic metallic structures. Another one is to quantify the measurement uncertainty in structured light sensing-based defect characterization. The uncertainties in this sensing system are investigated and then the total reconstruction uncertainty and the measurement uncertainty are studied to illustrate the reliability of the proposed system.

CHAPTER 2

UNCERTAINTY IN THREE-LEGGED STOOL FRAMEWORK

In this chapter, possible sources and forms of uncertainty in the proposed Three-Legged Stool (TLS) framework are discussed to provide a thorough understanding of the influential parameters in the system. Uncertainties coming from Data, Modeling, and Learning lay a strong foundation for understanding how the uncertainties are propagated within the inspection process, which is presented in Fig. 2.1. Nonetheless, definitions and supporting discussions of uncertainty-related examples will be presented in this chapter.

2.1 Uncertainty Sources in TLS UQ Framework

2.1.1 Data Uncertainty

Under the NDE context, "Data" is divided into two parts: input and inspection output. Inspection output usually referred to as inspection data, is generated from simulation or experiment induced by the material under test or defects (if present). The uncertainties are considered mainly from three aspects: the inspected material, defect geometry, and sensing/experimental measurement.

The research on material uncertainties is an important part of material integrity applications. In the NDE scope, the structural design needs to consider overstress, fracture, and fatigue failure mechanisms in the context of material degradation, which is a random process possible for a crack to initiate. Uncertainties are highly related to incomplete knowledge of material property changes and

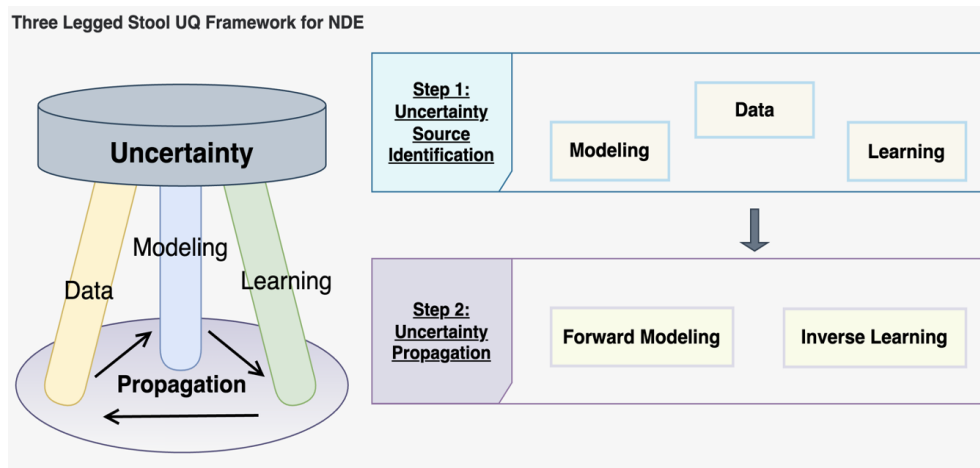


Figure 2.1 Illustration of UQ for NDE in TLS framework.

should be investigated to ensure the reliability of inspected material. Based on the literature review, the quantified data on material properties can be represented in the form of statistical ranges or probabilistic form based on stochasticity or possibilistic form with fuzzy variables to schematically distinguish one material from another while significant variability was observed in the respective material properties [44]. For example, Taddei et al. investigated the effects of material uncertainty for low-to-high frequency vibration analyses of thin plates utilizing a statistical moment-based probabilistic approach [45]. Silva et al. quantified the variability in material properties of sisal fibers using a Weibull distribution, which is a form of probabilistic distribution, to correlate sisal microstructures with tensile strength [46]. In [47], finite element-based probabilistic and possibilistic methods are discussed and compared in the modulus of elasticity analysis. Generally, the material properties are divided into microstructural properties, known as phase content and grain size, and mechanical properties, such as hardness, strength, residual stress, etc. For both microstructural and mechanical properties, Ultrasonic testing (UT) is a powerful NDE technique to provide parameters, which are correlated with grain size, hardness, yield strength, residual stress, etc [48]; electromagnetic techniques is a typical NDE method to characterize material using EM induction process. The interaction between the electromagnetic and microstructural properties, such as grain boundaries, precipitations, as well as mechanical properties, is a good determination of material state [49]. Besides, some hybrid NDE methods, the combination of several NDE testing methods, perform well in characterizing material residual stresses, hardness, yield strength, etc. Those hybrid techniques are applied with respect to specific problems while considering the material's physical information and the characters of measurement methods [50].

In the process of NDE-based defect identification, the forward problem is a direct method to obtain the response from the defect through experimental measurement or numerical simulation, while the solution of the inverse problem involves identifying the studied defect from the response. Therefore uncertainties of the defect geometry size, shape, or orientation result in uncertainties in the associated inspected response, which further affect the overall inspection uncertainty. Various mathematical and analytical procedures are used to reduce uncertainty in estimating parameters

of defect geometry. Houssem et al. addressed the effect of defect depth on response signals and then optimized 3D defects with eddy current measurements [51]. In [52], the variety of the defect volume, presented in different shapes, is directly related to the probability of failure to affect the system reliability for eddy currents control. In an X-ray Computed Tomography (XCT) based flaw detection case, the effects of flaw locations and orientations in the phantoms are incorporated and investigated with a multi-level Bayesian model [53]. In some cases, defects manifest as changes in the physical properties of conductive materials. Therefore, conductivity is considered as a random variable expanded in a series of Hermitian polynomials as one of the sources of uncertainty [54, 55].

During experiments, the inspection reliability depends not only on defect variability and material heterogeneity but also on the measurement quality, which is related to the noise resistance related to the sensor and experimental setup, as well as the superimposed measurement noise from the operation of inspectors. The noise resistance during the simulation or experiment is a very important criterion in NDE inspection estimation [56]. Lift-off, which is the distance between the sensor and the inspected material, is one of the most common sources to induce uncertainties, which are brought by cases like uneven surfaces, non-conducting coatings, or irregular paintings on the material [57]. Besides, there are other artifacts to affect the measured NDE data, such as the probe resolution and orientation, fluctuations in sampling rates, ambient electromagnetic disturbances etc [58–60]. The noise resistance from the aforementioned uncertainties will affect the sensitivity to capture changes in the material under the tests, which will further give rise to detection errors [61]. In another hand, the operator-influenced measurement uncertainty can be divided as systematic and random. Systematic measurement error is a consistent uncertainty related to incomplete knowledge of the inspection or setup, such as changes in frictional force between the probe and material surfaces as well as misalignment of the probe and coordinates of the material under the tests. The random error is only some incorrect measurements which rarely happen. The Guide to the Expression of Uncertainty in Measurement (GUM) [62], is the most commonly used method to analyze measurement uncertainty and evaluate the measurement quality, which will be discussed in the next chapter.

2.1.2 Modeling Uncertainty

In TLS framework, "Modeling" denotes describing the forward procedure from the input (Inspected material) to the obtained NDE inspection data. NDE inspection methods primarily rely on scientific principles of physics to model the sensing problems, which range from capillary action (dye/fluorescent penetrant methods) to wave propagation (ultrasound, microwave, and terahertz methods) to high energy interactions of elementary particles (radiography) [63].

The process of physics-based mathematical modeling is to decompose and refine the complex system and then attempt to reveal the basic principle behind a phenomenon. There are two types of uncertainties affecting the confidence of the given modeling procedure: parametric uncertainty and structural uncertainty. Parametric uncertainty arises because of insufficient knowledge of the predefined model parameters such as empirical quantities, initial conditions, boundary conditions, etc [64]. The effective definition of the initial condition and the selection of empirical quantity, which is considered as an initial system property, is an important uncertainty factor for "Modeling" design [65]. Some researchers have proved that boundary conditions are a significant source of uncertainty in structural dynamics modeling, where even small changes in boundary conditions can cause significant changes in the model predictions, such as estimated model parameters, dynamic time domain response, and frequency response functions [66–68]. Theoretically, parametric uncertainty can be identified and reduced by parameter refinement during the modeling, which is usually realized through the parametric study, such as the perturbation method [69], hierarchy method [70], and Neumann expansion method [71]. However, in most cases, without sufficient theoretical support coupled with complex physical systems or engineering problems, it's hard to access and restructure mathematical models directly to describe the NDE modeling procedure. In this case, some coarse or approximate equations are applied to simply the true underlying physics. Since the physics behind is hard to be fully understood, the lack of complete knowledge will make the outputs of the simplified modeling have a large disparity to the true results, therefore, other than the parametric uncertainty, structural uncertainty will exist in the system as well. Based on the definition, the structural uncertainty is aleatoric which can be reduced only by improving param-

eterized schemes, refining model dynamics, or implementing state-of-the-art numerical methods [64].

In the physics-based mathematical modeling procedure, the input and the material parameters are assumed known, whereas forward models can be developed to predict or estimate the output response function. In contrast, there are cases where the measured data are obtained from undefined input functions or the material properties, where the corresponding procedure can be described as empirical modeling. The uncertainty of the empirical modeling can not be identified and resolved directly, therefore, learning will be applied as the inverse procedure for uncertainty investigation, which will be described in the next section.

2.1.3 Learning Uncertainty

Inversion techniques are used to obtain quantitative estimates of the size, shape, and natural properties of defects in materials based on either measured or computed experimental data. Usually, the inversion is hard to be solved directly. Therefore, the approximate functional mapping between the input and the desired output can be described through a simplified model by fitting an optimal statistical distribution to data, which can be named as a surrogate model. In the TLS framework, the process of constructing an optimized surrogate model is described as "Learning". Two kinds of learning techniques are defined as the surrogate model, depending on the availability of physical prior information to the measured or inspected data during the "Modeling" stage. Measured data, from physics-based mathematical modeling, provide good physical theoretical support in the inverse problem, whose process is identified as physics-informed learning. Otherwise, the inversion process of empirical modeling's output is called data-driven learning. Artificial intelligence(AI) based approaches (such as Neural Network clusters [72, 73], fuzzy logic [74, 75], etc.) and the statistical-based approaches (such as Regression related methods [76, 77], hidden Markov model [78], etc.) are commonly applied as the data-driven learning. While Bayesian inference plays an important role in integrating the available physics model with the selection of estimation algorithms, such as the Bayesian method [79] and filtering related methods [80, 81], in the scope of physics informed learning [82].

The benefit of physics physics-informed learning model arises from its use of physics parameters to assist in describing the damage behavior. In this case, the integration of the physical model and measured data is able to assist in identifying model parameters and predicting future behavior. Therefore, even with limited data, the detection and characterization quality can be ensured. However, when constrained by the knowledge to describe the problem, information from previously collected data is needed for training data-driven learning process to identify the characteristics of the currently measured damage state and to predict the future trend. It is a fitting process for constructing the surrogate model by optimizing the learning without complex analytical theory cooperating. Compared with the physics-based learning model, this is computationally cheaper to evaluate and compute, but this requires large efforts in training and optimizing to find the optimal solution [83]. Besides, some works combine the concepts from both approaches as hybrid approaches to improve the performance [84, 85]. The aforementioned methods all have different properties that contribute to the preference of researchers' choice.

As "Learning" is identified as the optimization process, the optimal model and parameters are usually determined through the iterative process to reach a reliable solution. Therefore, similar to uncertainties in "Modeling", structural and parametric uncertainty exist in "Learning". Structural uncertainty is related to the selection of a learning model, and in physics-informed cases, an appropriate estimation model can ensure the optimal posterior distribution is learned and updated from the prior knowledge or information of the unknown parameters. When comes to data-driven learning, the structural uncertainty can be highly reduced by applying an efficient model to ensure the relationship between the training input and output can be learned and accurately described. Regarding the parametric uncertainty, updating and refining the learning model's parameters are essential for uncertainty reduction during optimization. For example, the determination of the number of nodes, weight parameters, and initial values in NN-related data-driven learning [86] and identification and correlation to parameters of the likelihood function and mechanism [87] in physics-informed case, are all important in ensuring the completeness and reliability in learning. The less uncertainty in learning will give us a better insight to uncertainties from the data itself.

2.2 Uncertainty Propagation in Three-Legged Stool (TLS) UQ Framework

Generally, an effective uncertainty propagation analysis of an NDE problem requires a thorough understanding of the system and detailed knowledge of all the influential parameters and their effects. The uncertainties during the NDE inspection process can be described through the forward "Modeling" and "inverse Learning":

1. Forward Modeling: The output signal from NDE sensors conveys the uncertainties from the geometric and material parameters of the problem. The aleatoric uncertainty during this stage referred to the variations in material properties, and geometry [88]. Uncertainties in material property, such as hardness [89], strength [90], and micro-structure [91], are usually analyzed in material characterization applications. Besides, in the application of damage detection, aleatoric uncertainties are not only from the material itself, but the geometric variance in defect design [92]. On the other hand, examples of epistemic uncertainty are related to the processing parameters for simulation as well as for experiment testing. For example, the mesh parameters in FEM-based simulation solutions [93]; the variability associated with setup procedures [94] and environmental noise [95] in the experimental inspection. The addressed variability will be incorporated in further modeling and analysis, which results in uncertainty propagation throughout the whole inspection system.
2. Inverse Learning: In this procedure, a mathematical or analytical framework is applied to obtain the predicted parameters to describe the system from the observed measurement or simulated output from the forward procedure [96]. The output often provides information about physical parameters that we cannot directly observe or infer. The variation of input uncertainty is not only propagated from the previous forward stage, there will be additional variability associated with measurement procedures [97]. Besides, the application of the inversion model will bring epistemic uncertainty to this process which is related to learning model parameters and model.

All sources of uncertainty have an impact on system response, and the way these errors propagate

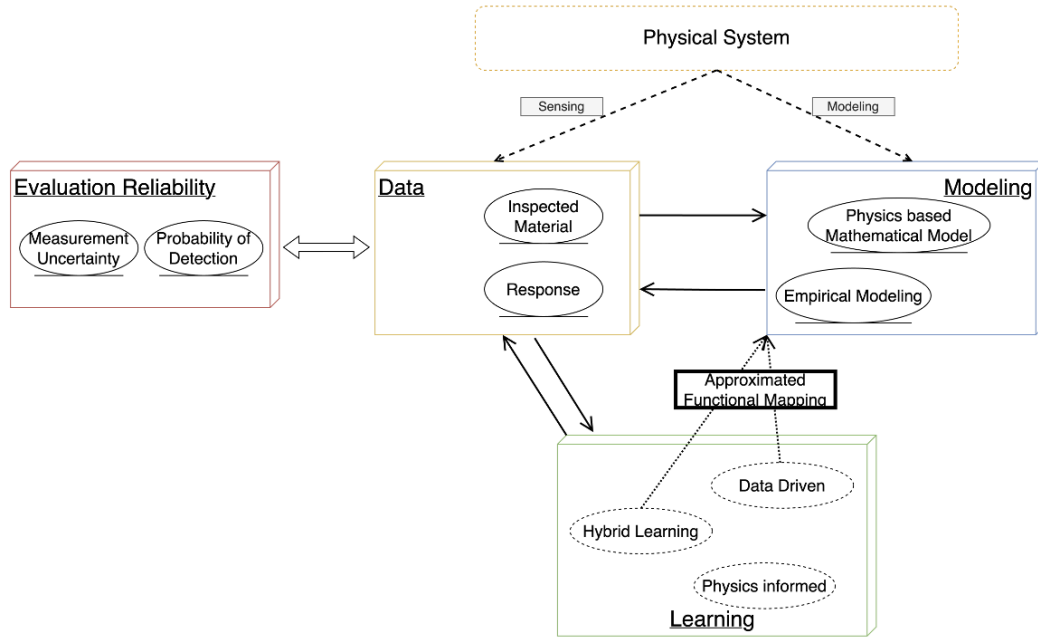


Figure 2.2 Illustration of the proposed TLS framework.

and interact can affect the overall accuracy of NDE detection. Therefore, the optimal design of NDE projects should quantify the uncertainty in system output performance propagated from uncertain inputs, which is described as Uncertainty Propagation (UP). A major problem when performing uncertainty analysis in many models is the large dimensionality of the uncertain parameters, which manifest themselves as multiple parameters in complex models, or as stochastic input fields. It is of great importance to decompose the uncertainty into unique, shared, and synergistic contributions originating from the different "Legs". The proposed TLS UQ framework is presented in Fig. 2.2. Specifically, forward UP modeling can be categorized into two primary problems, which depend on the presence or absence of a mathematical model. On the other hand, the assessment of the inverse learning process can be approached from the perspectives of physics-informed, data-driven, and hybrid scenarios. Detailed UP discussions for "Forward Modeling" and "Inverse Learning" will be presented in the following chapters.

CHAPTER 3

FORWARD MODELING BASED UNCERTAINTY PROPAGATION

3.1 Methods of Uncertainty Propagation in Forward Modeling

Considering the uncertainty in the input data, this section will describe how the uncertainty from the inputs is propagated through the model to affect the overall system response in the forward NDE process. For general UP problems, forward modeling can be classified into two main problems depending on the availability of a mathematical model: intrusive and non-intrusive. Theoretically, intrusive methods involve incorporating input uncertainties in the explicit model equations or algorithms; while non-intrusive methods treat the models as 'black boxes', which could introduce additional approximations during the uncertainty propagation process [98]. Moreover, if the input uncertainty could be described in terms of probability distributions or random variables within the model, the uncertainty propagation analysis is categorized as probabilistic. In contrast, there are situations where uncertainty inherently exhibits a fuzzy nature, and these are categorized as non-probabilistic approaches [99]. Popular UP methods have been categorized into different groups theoretically, such as simulation-based methods, local expansion-based methods, most probable point-based methods, functional expansion-based methods, numerical integration-based methods, and evidence theory [99]. Those methods have been widely investigated and compared based on their advantages and applications [100–106]. Based on the characteristics of different forward UP approaches, several popular UP methods are discussed and compared for different NDE applications.

The Monte Carlo simulation (MCS) method, also known as random event simulation technology, is a popular simulation-based mathematical method in UP [107–109]. MCS applies random sampling from a specific probability distribution of the random variable to predict the output of a model. With a known simulation or computational model derived from physics, the combined output is then collectively analyzed to understand the statistical variability of the random system [110]. As the focus of MCS and its variant techniques in UP is to assess how variations in input parameters impact the model's output in terms of relative relation, they do not have a strict requirement for given physical models. Even with the unknown true probability distribution of

the stochastic input parameter, given repeated sampling from the input quantities following an assumed probability distribution, the numerical output distribution in outputs is still able to provide useful insights into uncertainty [111]. For NDE inspections, based on the characteristics of MCS, MCS and its variants have been widely for investigating forward parametric-based uncertainty propagation. For example, MCS was applied in investigating uncertainties from material properties in Resonant Ultrasound Spectroscopy (RUS) inspection. Uncertainty bounds from MC analysis provide effective insights into how the uncertainty from modulus, density, Poisson's ratio, length, and diameter are affecting the measured results [112]. Matthias et al. repeated thousands of times of integrated MC simulation into a ray tracing simulation system for Flywheel Rotors to understand how the uncertainties from each influence parameter affect the final decision-making with given distribution [113]. Theoretically, it is able to deal with high-dimensional uncertainty inputs or complex mathematical models with high accuracy, but it will be computationally intensive, which makes it impractical for real-time applications or for large-scale problems.

There is an error in the statistics due to the error in each sample and due to the sampling error. Alternatively, we can use a small number of evaluations to approximate the response surface to realize UP, such as Polynomial Chaos Expansion (PCE), and Stochastic collocation (SC). PCE is a deterministic method that uses polynomials to represent the uncertain input variables and then employs these polynomial representations as inputs to solve the model or system [54]. The polynomial basis functions are chosen to be orthogonal with respect to the joint probability distribution of these random variables. Common choices of polynomials include Hermite, Legendre, or Laguerre polynomials, depending on the distribution of the inputs, such as normal, uniform, or exponential distributions, respectively. The model's response can be expanded as a series of expansion coefficients, which insights into the global sensitivity of the response concerning the expansion variables, resulting in a set of deterministic equations. When there's probabilistic uncertainty in the system parameters, the Polynomial Chaos Expansion is able to propagate uncertainty through the model, while the uncertainty can be quantified by calculating the moments (mean, variance, etc.) of the expansions. Generally, methods for calculating PCE coefficients of model outputs are classified

into intrusive and nonintrusive [114, 115]. In the intrusive approach, PCE needs to modify the specific mathematical model or simulation code by projecting the model on the polynomial basis, while for nonintrusive PCE, similar to MCS, the input uncertainty is approximated by the sampling strategy. Intrusive PCE offers more control and efficiency in UP but is difficult in solving real-time problems under many parametric uncertainties. Comparatively, non-intrusive PCE is modeled with reduced-order polynomial, which provides more flexibility in real applications. Noted that non-intrusive PCE are computationally expensive when many model evaluations are required, therefore many variant approaches have been developed to reduce the number of collocation points in the PCE process to reduce the computational costs [116, 117]. Non-intrusive PCE and its variants are efficient UP methods, which have been applied in material and defect characterization within NDE-related inspection, such as UT, EC, and EM [118–121]. Moreover, PCE has been widely applied for meta modeling (surrogate or approximate models) using experimental or simulation data when the mathematical behind is hard to compute or unavailable [122]. Meta-modeling replaces a complex time-consuming computational model with a statistically equivalent computationally inexpensive model by learning from a relatively small set of inputs[123]. Other common meta-models in UP include Kriging [124], Polynomial chaos Kriging [33], and Canonical low-rank approximations (LRA) [125], which have been widely applied and compared in NDE applications, such as NDE simulation modeling, sensitivity analysis, defect detection [111, 126, 127]. Within the UP discussion, Stochastic collocation (SC) is another stochastic expansion technique comparable to PCE. It is also a non-intrusive method to propagate uncertainties, where the collocation is evaluated at a fixed set of realizations, which are further used to approximate quantities. Different from PCE the polynomial coefficients are estimated as known orthogonal polynomial basis functions, SC relies on Lagrange interpolation functions to derive the expansion polynomials. The polynomial approximation from SC allows for interpolation and extrapolation of model responses between and beyond the collocation points, which is particularly useful for estimating responses at untested parameter values and dealing with high dimensional NDE problems [128].

If there is a lack of physics information due to ignorance or at the early stage of product develop-

ment, the uncertainty in data is inherently fuzzy. Therefore, the probabilistic nature of uncertainty is not well understood or justified, there are several useful non-probability-based methods to deal with non-probabilistic uncertainty. Common methods such as Fuzzy logic, allow for the representation of uncertainty using fuzzy sets and linguistic variables. Fuzzy sets can be used to describe the imprecise or qualitative nature of uncertainty. Fuzzy logic is a nonlinear mapping of an input feature vector into a scalar output, which can be expressed as a linear combination of fuzzy basis functions [129]. By integrating descriptive knowledge and numerical data into a fuzzy model, approximate models can be applied to describe the relationships between the fuzzy sets of input parameters and the fuzzy sets of output parameters for UP analysis. In NDE-based structural health monitoring applications, fuzzy logic-based UP has been widely applied in NDE-based structural health monitoring applications, usually realized with common model approximation techniques (e.g. generalized polynomial chaos, Monte Carlo, etc.) simulation [130–132]. Besides, interval analysis and evidence theory are also commonly used nonprobabilistic UP methods in NDE inspection [96, 133, 134].

Overall, only some of the popular NDE-based forward uncertainty propagation techniques are discussed and there's no correct way when deal with specific problems. Depending on the nature and amount of available data, the selection of appropriate UP approaches can be determined by considering the following aspects:

- Input data characteristic: probabilistic or non-probabilistic; low-dimension or high dimension;
- Uncertainty characteristic: parametric or non-parametric;
- Model availability: intrusive or non-intrusive; physics-informed or evidence-related.

Usually, to provide a more reliable UP analysis, more than one method is integrated when dealing with different applications. Overall, forward modeling uncertainty propagation can assist in identifying the most influential sources and parameters that affect the uncertainty, evaluate the level of confidence and risk, and prioritize them for further investigation or system design improvement. In

the next section, an MCMC-based 'Forward Modeling' uncertainty propagation is illustrated on a Capacitive sensing system.

3.2 TLS-based Forward Uncertainty Modeling Application

3.2.1 Introduction

Microwave imaging systems, such as capacitive imaging systems, have been and still are routinely used to characterize and image defects in composite materials[135–137]. In order to understand how uncertainty in data affects modeling performance, the process of error propagation is formulated and evaluated through a microwave-sensing-based experiment. The shortwave capacitive probe's low frequency enables it to penetrate deeply into the layers of carbon fiber, revealing defects at greater depths. One example is the parallel plate capacitor probe, which excels at detecting variations in dielectric permittivity. The sensing system under investigation employs capacitive sensors that can identify changes in the dielectric coefficient. This capability enables the measurement of high-resolution contrast images for detecting and identifying defects in carbon fiber-reinforced polymer composites (CFRP).

A widely used dielectric probe incorporates an inductance-capacitance (LC) circuit, forming a resonant tank. In this design, the inductance and capacitance are considered constant, while the resonant frequency shift relies entirely on an unfixed variable. The LC tank probe used in this study is detailed in [138]. It includes an I/Q demodulator and a local oscillator (LO) to capture the reflected power, from which the phase can be calculated using the total impedance Z . The LC tank probe architecture is known for its significantly higher Q factor, which results in increased sensitivity for detecting changes in dielectric permittivity (ϵ) and inferring the defect's location and size. In this Uncertainty Quantification (UQ) study, our focus is solely on the process of inferring impedance. The probe's impedance is defined as follows:

$$Z = \sqrt{R^2 + (X_L - X_C)^2} \quad (3.1)$$

where

$$X_L = 2\pi f_0 * L \quad (3.2)$$

and

$$X_C = \frac{1}{2\pi f_0 * C} \quad (3.3)$$

whereas the Q factor for an LC tank is simplified as [138]:

$$Q = \frac{2\pi f_0 * L}{R} \quad (3.4)$$

where f_0 is the resonant frequency of the capacitive probe, at which with the highest Q factor, the reflection coefficient of the probe is close to a maximum, allowing total reflection. Resistance R is a negligible constant, which is related to the sensor itself. The introduction to inspected material changes the resonant frequency according to the Q value and has a response to the dielectric value of common defects.

The designed ring separation d of the applied LC tank resonance probe is 2mm. The probe is designed to have a fixed inductance value L of $75\mu H$ and the resonant frequency is determined at $5MHz$ to realize the total reflection. Therefore, according to Eq.3.1, any change in total impedance is related to capacitance C . Based on the work from [], the effective permittivity ϵ_r of the sensing field of view, is proven to be the function of the sensor's spatial location. Since the design of the parallel plates contains a fixed area, the variance in spatial location is referred to as the variety in sensor liftoff (lo). Specifically, with the different sensor liftoff, the permittivity component under the inspected region will change accordingly, which can be described in Fig. 3.1.

Theoretically, the sensor capacitance is positively related to the relative permittivity ϵ_r , which can be described as:

$$\epsilon_r(\vec{r}) \propto \frac{1}{lo} \quad (3.5)$$

$$C(\vec{r}) \propto \frac{A\epsilon_0\epsilon_r(\vec{r})}{d} \quad (3.6)$$

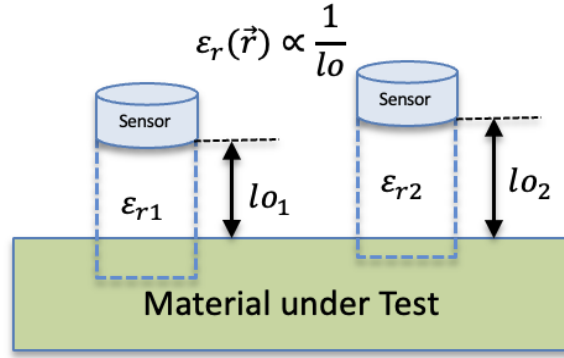


Figure 3.1 The relation between the liftoff and effective permittivity.

where ϵ_0 is the permittivity of vacuum; A is the electrode surface area; and d is the separation distance between two electrode rings, which is determined by sensor design and configuration. Therefore, combining Eq.3.7 and Eq.3.5, there exists a negative correlation between the sensor capacitance $C(\vec{r})$ and the sensor liftoff l_o , which is defined as:

$$C(\vec{r}) = f(l_o) \quad (3.7)$$

The function $f(.)$ is unknown in this case, so in the next section, several meta-modeling methods are applied to construct a reliable mathematical model to describe this process based on the experimental data in a simplified manner, for further forward uncertainty propagation investigation.

3.2.2 Meta-Modeling based "Empirical Modeling" in Capacitive Sensing System

3.2.2.1 Experimental Setup

To build up a true model describing capacitance versus liftoff, 39 groups of measurements are recorded with different sensor liftoff. The experimental setup is shown in Fig. 3.2:

The sensor is attached to an AGS1000 Direct-Drive Gantry for accurately controlling the liftoff changes; while an impedance analyzer is connected to obtain the real-time capacitive reactance reading. The liftoff change ranges from $0.1mm$ to $1.3mm$ at $0.1mm$ intervals and is repeated three times. The relation between the capacitance and sensor liftoff is presented in Fig. 3.3 with example readings, where a 2nd order polynomial fitting curve with 95% confidence curve is applied

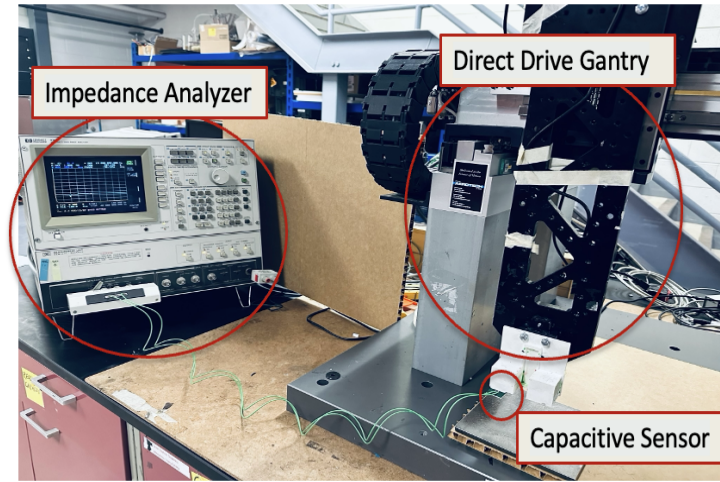


Figure 3.2 Experimental setup.

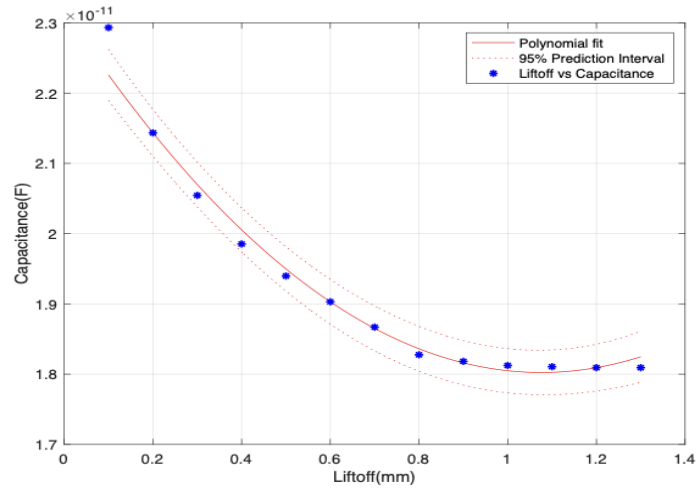


Figure 3.3 Example of experimental data for capacitance VS liftoff.

for a preliminary illustration. For building a more robust computational model to describe this correlation, several functional approximations are discussed in the next section.

3.2.2.2 Surrogate Modeling Methods

Meta modeling (surrogate modeling) is a process for reducing the associated computational costs by substituting an expensive computational model with inexpensive-to-compute surrogate models. It is an efficient modeling process by learning from a relatively small set of inputs and corresponding model responses to generate a high-confidence mathematical-based approximation, which has been

applied in a wide variety of engineering contexts[138–140]. Meta-model $M^{Meta}(x)$ can be considered as a statistical approximation process to the original finite variance computational model $Y = M(x)$. In this study, three popular surrogate modeling methods are applied to describe the experimental data, which are Kriging, Polynomial Chaos Expansion(PCE), and Canonical low-rank approximations (LRA).

Kriging is a stochastic interpolation algorithm that considers the computational model as a realization of a Gaussian process, indexed by the parameters in the input space x . It can be described as[141, 142]:

$$y = M(x) \approx M^K(x) = \beta^T f(x) + \sigma^2 Z(x, w), x \in R^M \quad (3.8)$$

where $\beta^T f(x)$ is the mean value of the Gaussian process, which is usually named as a trend; σ^2 is the variance of the Gaussian process, and $Z(x, w)$ is a zero mean, unit variance, stationary Gaussian process which is characterized by a correlation function R and its hyperparameters θ .

The selection of the functional basis of the Kriging trend is an important part of building up the Kriging model. In this work, linear trend, as one of the most commonly used polynomial basis trends is applied, where the mean value in Eq.3.8 can be expressed as:

$$\beta^T f(x) = \beta_0 + \sum_{i=1}^M \beta_i x_i \quad (3.9)$$

Besides, the process for estimating the unknown hyper-parameter from the available data is considered as the optimization problem, which is important for calculating other unknown Kriging parameters (e.g., β)[142]. In this work, the Covariance matrix adaptation–evolution strategy (CMA-ES)[143], is applied for solving the optimization problem. It is a derandomized stochastic search algorithm, where the covariance matrix of a normal distribution is adapted for improving the objective function in the recent past iterations are more likely to be sampled again.

Polynomial Chaos Expansion (PCE) has proven to be a powerful tool for developing meta-models in a wide range of applications, such as structural systems[144], computational dosimetry[145], sensitivity analysis[146], and so on. PCE is achieved by expanding the model response to a basis

consisting of multivariate polynomials as the tensor product[147]. Consider a random input vector X with independent components expressed by the joint PDF f_X and finite variance computational model can be expressed with the polynomial chaos expansion:

$$y = M(x) \approx M^{PCE}(x) = \sum_{\alpha \in A} y_{\alpha} \Psi_{\alpha}(x) \quad (3.10)$$

where $\Psi_{\alpha}(x)$ are multivariate polynomials orthonormal to f_X , $A \subset N^M$ is the set of selected multi-indices of multivariate polynomials, and y_{α} are the corresponding coefficients.

Canonical low-rank approximation (LRA), also known as separated representations, has recently been applied as a promising tool for effectively dealing with high-dimensional model input. The key idea is to approximate the response quantity of interest by the sum of a small number of appropriate rank-one tensors that are products of univariate functions. Under the scope of meta-modeling, LRA is built with polynomial functions. Similar to PCE, the model response is expanded over an orthogonal multivariate polynomial obtained as the tensor product of the univariate polynomials in the input parameters. By employing the definition of canonical rank and expanding the univariate function onto a polynomial basis, the rank- R approximation of $y = M(x)$ can be written as:

$$y = M(x) \approx M^{LRA}(x) = \sum_{l=1}^R b_l \left(\prod_{i=1}^M \left(\sum_{k=0}^{p_i} z_{k,l}^{(i)} \phi_k^{(i)}(x_i) \right) \right) \quad (3.11)$$

where $\phi_k^{(i)}(x_i)$ is the k th degree univariate polynomial in the i -th input variable, p_i is the corresponding maximum degree. $z_{k,l}^{(i)}$ is the coefficient of univariate polynomial in the l th rank-one component. $b_l, l = 1, \dots, R$ are scalars that can be viewed as weighing factors. Compared with Eq.3.10 and Eq.3.11, they use the same univariate polynomial family as the basis, while the main difference is that LRA retains the tensor-product form and PCE uses the spectral expanded form.

For building up the above meta-models, 70% of the available experimental data is split as training sets while the left is for validation to evaluate the model. After creating meta models $M^{Meta} = [M^K, M^{PCE}, M^{LRA}]$, it is of interest to predict model $M(x)$ for a new point x , given the (observed) experimental design $x = lo(1), \dots, lo(N)$ and the corresponding model responses

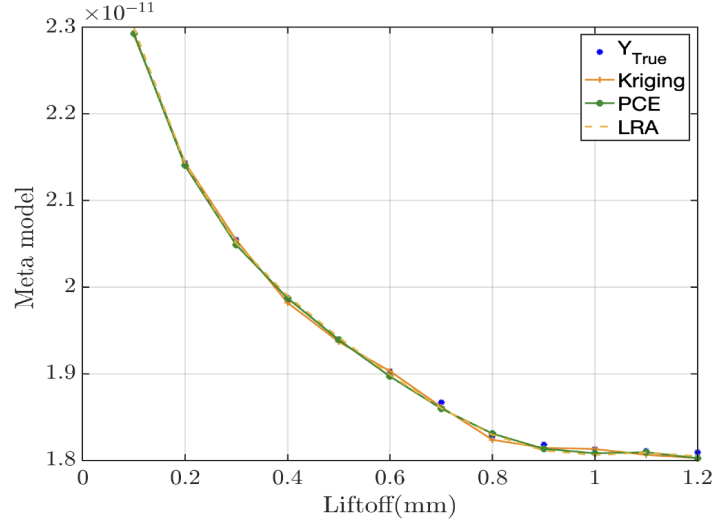


Figure 3.4 Comparison meta-modeling results for one example validation data set.

$y = X_C^{(1)} = M(lo^{(1)}), \dots, X_C^{(N)} = M(lo^{(N)})$. After all three meta-models are set up, the comparison modeling results for a new set of input are shown in Fig. 3.4

Results show that all three models can reveal the correlation between lo and C . The variation between the Y_{True} and the modeling results is very small. To further evaluate the performance of all meta-models in a statistical way, the relative generalization error on a new set of independent data is considered through the term validation error for the same validation data set. The input pair can be defined as $(x_{val}^{(i)}, y_{val}^{(i)}), i = 1 : N_{val}$, therefore the validation error can be referred as:

$$err_{val} = \frac{N_{val} - 1}{N_{val}} \left[\frac{\sum_{i=1}^{N_{val}} (M(x_{val}^{(i)}) - M^{meta}(x_{val}^{(i)}))^2}{\sum_{i=1}^{N_{val}} ((M(x_{val}^{(i)}) - \hat{\mu}Y_{val})^2)} \right] \quad (3.12)$$

where $\hat{\mu}Y_{val} = \frac{1}{N_{val}} \sum_{i=1}^{N_{val}} M(x_{val}^{(i)})$ is the sample mean of the validation set responses. In this work, the prediction is repeated five times with different liftoff-capacitance combinations for validating each meta-model, while the final validation error is obtained by taking the average among all five repetition results. The mean validation error is presented in Table 3.1

Table 3.1 Mean Validation for Three Meta Models

Validation Error	Kriging	PCE	LRA
Mean Error	0.0005	0.0009	0.0015

According to the calculated mean error, all three methods have shown high fidelity in learning from the input pair and making the prediction. Meta models, especially the Kriging model, have shown great advance in providing a substitution model to describe the relationship between the liftoff and capacitance, which provides a good basis for uncertainty propagation in the next section.

3.2.3 Uncertainty Propagation from Data to "Modeling"

In real experiments, because of the roughness of the material under the test, the sensor liftoff is not a strict constant, which is considered as the uncertainty imposed on data. As mentioned before, the liftoff variance is affecting permittivity, which further changes the resultant capacitance and total impedance. Therefore, in this section, we will investigate how the liftoff-induced data uncertainty propagates the modeling results. Based on the discussion in the previous discussion, the applied meta-model M^{Meta} is applied as the approximation mathematical model to make up the complex unknown true model $M(x)$ for realizing the uncertainty propagation.

Monte Carlo (MC) method, is a sampling-based approach that has been widely used for quantification and propagation of uncertainties [107–109]. As a classical UQ analysis method, MC is non-intrusive, robust, and simple for implementation [111]. The process of "Modeling" is considered as the black box which is evaluated at random samples of the input probability distribution. Outputs at these realizations are then used to approximate quantities such as expectation or variance.

When examining real CFRP pipes, we can assume that the standard liftoff is 0.5 mm. However, it's important to note that irregular bumps on the pipe surface introduce variations to the liftoff. We can characterize this variance as following a normal distribution with a known mean of 0.5 mm and a standard deviation of 0.15 mm. Consequently, the input probability distribution can be modeled as a Gaussian distribution. To analyze the system, we sample the inputs 10,000 times, which enables us to derive the distributions and statistics for the final output, represented by the total impedance Z . Utilizing equations from Eq. 3.1 to Eq. 3.3 and disregarding the resistance R of the sensor itself, the output can be described as:

$$Z = X_L - X_C = \omega L - \frac{1}{\omega M^{Meta}(lo)} \quad (3.13)$$

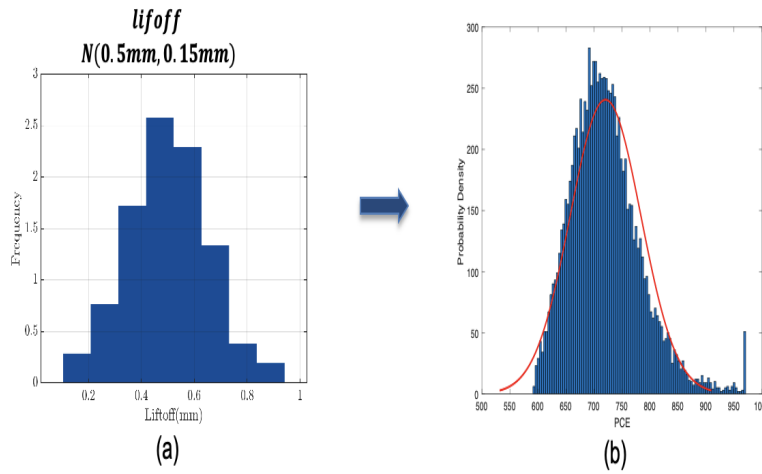


Figure 3.5 Illustration for uncertainty propagation process: a) Probability distribution of input: liftoff; b) Probability distribution of output from PCE meta modeling: total impedance.

The process of uncertainty propagation is depicted in Figure 3.5. This process reveals how uncertainty stemming from the liftoff stage is carried over to the "Modeling" output. The resultant output distributions from the three meta-models we proposed are displayed in Table 3.2. All of the meta-models yield similar mean and variance values. By averaging their results, it is determined that the probability distribution of the total impedance Z follows a normal distribution with parameters $N(721.187, 63.3013)$.

Table 3.2 Output probability distribution of Three Meta Models

Probability Distribution	Mean	Variance
Kriging	722.599	63.7365
PCE	720.706	63.0021
LRA	721.187	63.1652

3.3 Conclusion

This chapter delves into the critical aspect of uncertainty propagation in the context of forward 'Modeling' within the proposed TLS UQ framework. Recognizing the diversity of forward UP approaches, several widely employed UP methods, offer a detailed comparison across various NDE applications. These applications encompass Monte Carlo Simulation (MCS), Polynomial Chaos

Expansion (PCE), and Fuzzy Logic, each tailored to suit different scenarios and requirements. To exemplify the potency of the forward 'Modeling' scheme in action, we focus on a practical case study centered around Capacitive Sensing, an NDE method with significant real-world applications. The application outlines the application of MCS, emphasizing its role in unraveling the ramifications of liftoff uncertainty on the critical parameter, the total impedance Z . This in-depth analysis serves as a concrete illustration of the uncertainties introduced and their impact within this specific NDE context.

CHAPTER 4

INVERSE LEARNING BASED UNCERTAINTY PROPAGATION

In this chapter, the UP will be addressed in the inverse Learning process, which is an essential part of the proposed TLS UQ framework. Fundamental probability theory will be introduced to provide a basis to establish the groundwork for subsequent UP discussions and application examples.

4.1 Probability Theory for Inverse Uncertainty Quantification

The dominant view dealing with uncertainty assumes that expectations are based on statistical analyses of past data and market signals that provide information on objective probabilities. In another aspect, probability is a mathematical concept that allows predictions to be made in the face of uncertainty.

The goal of statistical inference, especially for UQ problem, is to draw conclusions about the distribution of a random variable (X) based on a particular statistical index (θ). This process can be described by a certain statistical model $f(X; \theta)$, $\theta \in \Theta$, where $f(x; \theta)$ denotes a probability mass function (pmf) or a probability density function (pdf) and Θ is the parameter space. Full knowledge of the true value of θ is equivalent to a complete understanding of the distribution of interest. However, due to incomplete knowledge or numerical approximation errors in the underlying physics, θ often deviates from the true value. Consequently, there is a need to infer θ as part of the inference problem. There are two most popular paths, commonly known as the frequentist, and the Bayesian approach.

Frequentist models are objective, where probabilities are defined as the frequency with which an event occurs if the experiment is repeated a large number of times. In the frequentist approach, the parameter θ is unknown but is fixed. Quantity and only the information coming from the sampling data are relevant for inference. Thus, we do not take into account the prior belief on the parameter [148]. Specifically, random data samples are collected from a consistent and repeatable process to find the θ that would maximize the likelihood $f(x|\theta)$ to find the optimal solutions. To statistically build confidence interval in the estimation process, one constructs estimators such as

OLS estimators or maximum likelihood estimators [149].

While frequentist approaches rely on ensembles of models empirically to approximate the posterior distribution, Bayesian methods can directly estimate the posterior distribution over the model's parameters. In contrast to frequentist theory, Bayesian theory treats probabilities as a distribution of subjective values. Priors in Bayesian modeling are critical for providing information about past experience in a sample space, where the prior distribution $\pi(\theta)$ is updated on the basis of the likelihood function through Bayes' theorem. Parameters θ are viewed as random variables with associated densities, which are quantified based on the observation $x = [x_1, x_2, \dots, x_n]$ and the solution to the parameter estimation problem is the posterior probability density, which summarizes the information in both the prior distribution and in the data.

Let $f(x|\theta)$ denote the conditional distribution of the sample given θ , which is equivalent to the likelihood function. We can apply the classical Bayes' theorem to derive the posterior distribution, denoted as $p(\theta|x)$. This posterior distribution represents the updated knowledge about θ and is determined based on the observed sample, denoted as x :

$$p(\theta|x) = \frac{f(x|\theta)\pi(\theta)}{m(x)} \quad (4.1)$$

where $m(x)$ is the marginal Probability density function(PDF) of the data as a normalization factor, which can be written as:

$$m(x) = \int_{-\infty}^{\infty} f(x|\theta)\pi(\theta) d\theta \quad (4.2)$$

The mean, which is often used to summarize the posterior distribution of the parameter, is a Bayesian point estimator of θ . Sometimes, the mode or the median of the posterior distribution is used instead. Even applied with different prior distributions, continued observations eventually force their conclusions to converge to different parts of the prior parameter space Θ , which can prove how prior beliefs should be changed by observing data [150]. Bayesian models have considerable flexibility in incorporating and modeling multiple sources of uncertainty. Standard inclusions are parameter uncertainty, residual process error, and observation uncertainty [149]. As it provides a

density that can be propagated through the model, thus, the Bayesian approach can provide a more significant value for model uncertainty quantification. which will be discussed in the next Chapter.

Both Bayesian and Frequentist methods offer substantial flexibility when it comes to incorporating and modeling multiple sources of uncertainty in the inverse NDE process. Typically, these sources include parameter uncertainty, residual process error, and observation uncertainty [149][151], often represented as $\delta(x)$. These methods are advantageous as they provide a density that can be propagated throughout the model. Consequently, statistical inference-based approaches prove to be highly valuable for conducting comprehensive uncertainty propagation analyses from 'Learning' to 'Data'. The physics basis is important for improving the reliability of the inverse analysis, especially for ill-posed NDE problems. As mentioned in the previous section, the 'Learning' based UP process can be categorized based on the availability of the physics information. Organizing all kinds of UP methods into a uniformly structured taxonomy is challenging since there are numerous potential approaches to consider. The following sections will introduce several classical and popular uncertainty propagation methods from measurements or observation data to determine the variations in posterior distribution during 'Learning' in NDE applications. They are classified with regard to the type of distribution hypothesis for NDE problems, which are presented in Fig. 4.1

4.2 Methods of Uncertainty Propagation in Physics-informed Learning

If the physics-based governing equations and essential conditions for describing the inverse NDE process could be fully understood, it is considered as a direct inverse NDE solution. Typically, Kubo.S et al. concluded indispensable information for constructing a solid direct inverse process, which is domain boundary, governing equation of the physical quantity, force or source term, and material properties [152]. With a full understanding of that information, the desired output could be obtained with traditional analytical or numerical techniques like the finite element method, the boundary element method, and the finite difference method. In this case, the structural uncertainty can be omitted, while the parametric uncertainty from measurement data can be investigated. W. Lord investigated three partial differential equation types to illustrate the use of numerical methods in

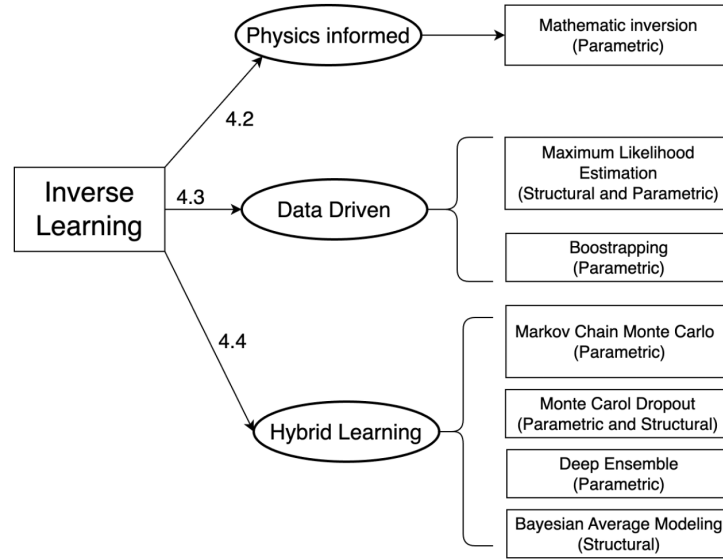


Figure 4.1 Overview of Learning based uncertainty propagation methods in NDE.

developing efficient inverse NDE applications, including Magnetic flux leakage based on the elliptic Poisson equation, Eddy current with parabolic diffusion equation, and the Ultrasonic inspection with a hyperbolic wave equation. The paper highlighted that a comprehensive understanding of the forward problem is essential for the development of effective inverse algorithms [153]. For example, Ultrasonic Testing needs to have known properties and geometry to detect defects [154]; X-ray Radiography should have knowledge of X-ray source parameters and material properties to obtain an expected image or radiograph of an object [155]; Eddy current testing with known coil properties and defect geometry is important to determine the electromagnetic field distribution in a conductive material with known coil properties and defect geometry [156]. A typical way for investigating uncertainty propagation in direct NDE inverse problems is to compare the differences between the measured and calculated signals by considering the variations in the given physics information. For example, K. Grondzak applied the optimization idea for investigating how the convergence criteria affect the uncertainty interval of the ECT-based defect property determination [157]. Generally, NDE methods are applied to find the location, shape, and structure of inclusions, defects, and sources (of heat, oscillations, stress, and pollution). Within such a wide range of engineering applications of NDE technique, it is hard to have full knowledge of the conditions of existence, or stability of the solution under small variations of the problem data. Therefore, most

NDE problems are ill-posed, and thus Data-driven 'Learning' is gaining more population in UP analysis.

4.3 Methods of Uncertainty Propagation in Data-driven Learning

Data-driven based 'Learning' is useful when there's no robust mathematical support from sufficient physics prior models. The inverse UP is realized based on the long-run collected data without knowledge of the prior distribution $\pi(\theta)$. Therefore, frequentist statistics is applied to make inferences about the unknown but fixed model parameters θ relying on the optimization theory from the available sample of data/observations [158]. For example, NDE-based fatigue life assessment approaches, the long-period monitoring data provide a good basis for frequentist statistical analysis for uncertainty quantification [159]. The concept of likelihood is a fundamental aspect of statistical inference, and maximum likelihood estimation (MLE) is a classical frequentist method applied in uncertainty propagation. MLE aims to find the parameter values θ within the selected model that maximize the likelihood of observing the given data X . It is basically an optimization algorithm (e.g., gradient descent) for adjusting the model parameters iteratively until the maximum likelihood is achieved [160]. The maximizer of the likelihood function can be described as follows:

$$U_{MLE} = CI[\theta, \hat{\theta}_{MLE}], \hat{\theta}_{MLE} = \operatorname{argmax}(f(X|\theta))$$

Therefore variability and uncertainty of parameter θ can be represented by the random variables of a statistical distribution from the given measured data $f(X; \theta)$, which are then evaluated using the Confidence Intervals(CI). As addressed in [160], MLE-based inverse UP could infer the 'Learning' model's uncertainty (structural uncertainty) through the resultant likelihood function shape: when the likelihood function is peaked, the inferred model parameter θ is with less uncertainty; while high uncertainty can be inferred when a flat likelihood function is obtained. Other than MLE, other common estimators such as Ordinary Least Squares(OLS) are able to statistically build confidence intervals in the estimation process [149]. The key in this uncertainty propagation is to calculate the variance of the estimator based on the probability distribution from independent observations.

Generally, it is hard to obtain a completely independent sample set, which requires a large

amount of repeated measurements. Therefore, more popular UP analysis can be realized through Bootstrap, which is an ensembling-based frequentist method [161]. Bootstrap empirically assesses the uncertainty associated with input (measurement) quantities in situations where modeling techniques and analytical solutions are not readily applicable [162, 163]. It stimulates the frequentist concept of obtaining the probability distribution of observations from repeated similar experiments based on resampling to approximate the posterior empirically. The selected bootstrap samples are considered 'almost-independent' and are able to approximate the variance of the estimator. This statistical technique consists of generating samples of size B (called bootstrap samples) from an initial dataset of size N by randomly drawing with replacement B observations. The selection of N needs to consider each subsample could approximate drawing samples from the real distribution; while the dataset N should be large enough to minimize correlations between samples. As the Bootstrap method follows standard probability estimation, it is mostly concerned with parametric uncertainty [160]. Kass et al proposed and derived the initial usage and formula for bootstrap-based uncertainty propagation [164]. For NDE applications, bootstrap has been integrated with other techniques to provide reliable confidence intervals for uncertainty evaluation. For example, Felix H. Kim et al. constructed the response probability of detection analysis in an X-ray Computed Tomography (XCT) system with Additive Manufacturing defects, where the bootstrap technique is applied to quantify the uncertainty for the POD curve [165]. Regarding a remaining useful life estimation of fatigue cracks, a bootstrap method was developed for calculating the lower confidence limit for parameter estimation from expectation maximization (EM) and stochastic EM algorithms [159].

4.4 Methods of Uncertainty Propagation in Hybrid Learning

While Data-driven approaches approximate the probability distribution of observation empirically, Bayesian approaches are able to incorporate the prior information from the available physical model, which is considered 'Hybrid Learning' in this framework. Prior physics information provides built-in regularization for the 'Learning' process with objective physical constraints. Even with limited measurements and ill-posed NDE problems, the stabilization of the inversion process

can be improved. In the Bayesian approach, in addition to the likelihood function, information regarding the prior physics model parameter is updated to find the posterior distributions based on a comparison of the inverse model and observation data [158]. In the context of deep learning, the Bayesian neural network (BNN) is a basic method for investigating uncertainty propagation from Learning to data. The uncertainty captured by BNN is realized through posterior weight distribution $P(\theta|x)$ by specifying a prior distribution over the weight parameters [166]. As mentioned before, the marginal data distribution $\pi(x)$ is hard to obtain analytically, the posterior must be approximated using the proportionality. The posterior distribution and model can be used to obtain the probability distribution of the prediction. Under the Bayesian inference, the uncertainties are obtained from the variance of the predictive posterior probability distribution $p(\theta|x)$, when approximations are made to fit the true posterior distribution.

In Bayesian inference, a typical sampling-based approach to derive the posterior distribution of parameters is called the Markov chain Monte Carlo (MCMC). MCMC reduces the calculation effort by sampling from complex probability distributions, particularly when analytical solutions are difficult to obtain [158]. Specifically, MCMC generates a Markov chain of samples, and as the chain progresses, it converges toward the target posterior distribution of interest. These samples can be used to estimate the statistical variation of the model parameters and predictions in terms of statistical terms, such as means, variances, and credible intervals. Moreover, there exists lots of MCMC variants in uncertainty analysis depending on the applicability, convergence time, or model formulation: such as Hamiltonian Monte Carlo which incorporates Hamiltonian dynamics to accelerate the convergence of the sample distribution; transdimensional MCMC which uses reversible jumps for model selection. A more comprehensive Monte Carlo-based UP method is addressed with examples in [111]. MCMC-based uncertainty propagation methods have been applied for NDE-based defect detection[167], characterization [168], and image reconstruction [169].

Although MCMC could provide high-fidelity results from drawing samples from the posterior, its computational cost is pretty high as thousands of samples are needed, which highly restricts its

UP-related application, especially for dealing with highly multimodal scenarios. Variational Inference (VI) is a widely applied posterior approximation method under the Bayesian framework, which could approximate the posterior over the model weights with a simpler variational distribution. In this process, a variational distribution $q_w(\theta)$ with variational parameter w , is used to approximate the true posterior $p(\theta|x)$ [170]. The distance between them can be realized by minimizing the Kullback-Leibler divergence with regard to w , which is equivalent to maximizing the log evidence lower bound [171], which can be expressed as:

$$L = \int q_w(\theta) \log(P(\theta|x, \theta)) d\theta - KL(q_w(\theta) | \pi(\theta)) \quad (4.3)$$

As this VI-based approximation is constructed over the 'Learning' model's parameter, it is able to capture the structural uncertainty during the process. There are various ways to realize VI approximation in the Bayesian framework, such as Gaussian distributions along with diagonal covariance matrices [172], stochastic gradient VI [173], empirical Bayes (EB)[174], and etc. A popular and basic example of VI in deep learning applications is Monte Carol Dropout-based variational inference. It establishes a relation between (variational) Bayesian inference and using Dropout as a learning technique for neural networks [175]. MC Dropout can be applied before the weighted layer in the neural network and is used as an approximating variational reasoning scheme in the deep Gaussian process as they are marginalized over its covariance function parameters [176]. Dropout layers are usually used as a regularization technique during training, where neurons are randomly dropped out at a certain probability during a forward pass through the network to create variation in the model's outputs. Specifically, in the context of a neural network, the L_2 regularization term is employed to enhance the dropout procedure with weighted decay λ . The minimization of this objective function is proven as a good approximation to the variational inference [177]. The output predictive distribution can be further approximated by sampling model weights from the estimated posterior distributions, the variance from this distribution can be quantified as a measure of total uncertainty [160]. The total uncertainty can be further decomposed to obtain a meaningful interpretation of the uncertainty in terms of the randomness of the model

(Epistemic) and the variability of the given data (Aleatoric)[178], which corresponds to parametric and structural uncertainty in this work. The total prediction uncertainty can be expressed as follows:

$$\begin{aligned} Var(q_w(\theta)) &= Aleatoric + Epistemic \\ &= Parametric + Structural \end{aligned} \tag{4.4}$$

With increasing prediction repetitions, the uncertainty from both aspects are converging in probability to system variance. Lots of studies have used the dropout-based method for UP analysis in the NDE area. For vision-based applications, including crack detection, local damage identification, and bridge component detection, the uncertainty for each application is quantified with MC dropout in terms of variations in softmax probability and entropy [179]. Li et al. applied a Dropout-assisted Convolutional Neural Network model in a magnetic flux leakage-based crack classification case, which is able to quantify the addressed aleatoric and epistemic uncertainties during the inspection [180].

BNN-based methods are effective but are computationally expensive when dealing with large-scale networks. In recent years, deep ensembles (DE), have been widely applied as a good alternative to traditional Bayesian NNs with readily parallelization and less hyperparameter tuning in deep learning area [181]. DE is a more straightforward procedure that combines multiple base models for delivering dependable predictive uncertainty estimates by randomly initializing their weights and training with the same dataset. Model ensembling can also be considered as a sampling process in Bayesian inference, and the variation of this process is considered an approximation to the uncertainty from prior model parameters. The idea of model aggregation is to reduce the bias and/or variance of weak learners to create a strong ensemble model that achieves better performances in uncertainty analysis. Generally, there are two typical ensemble learning techniques: Bagging and Boosting, which encourage diversity through different importance sampling schemes [182]. Bagging (Bootstrap Aggregating) involves training multiple instances of the same model in parallel on different subsets of the training data, typically using resampling with replacement (Bootstrap). The final prediction is often a combination of the predictions from each model, such as taking

a majority vote (for classification) or averaging (for regression). The random forest approach is a typical bagging method that combines bagging with an additional layer of randomness by considering only a random subset of features at each split in each tree. Sampling over features could ensure each tree doesn't rely on identical information to make decisions, thus reducing the correlation between their respective outputs. Thus, the concepts of bagging and random feature subspace selection of random forest are beneficial in reducing overfitting and enhancing generalization [182]. Different from Bagging, Boosting is an iterative ensemble scheme that focuses on training multiple models sequentially. Each consecutive model gives more weight to the data that the previous models predicted erroneously, therefore, the ultimate ensemble is comprised of all the weak learners, each assigned an appropriate weight based on their performance. Examples of boosting include Adaboost and Gradient boosting. Adaboosting, also known as Adaptive Boosting, uses weak learners as base models such as decision stumps (shallow decision trees with a single split). Instead of trying to find all the coefficients and weak learners that give the best overall additive model, it focuses on reweighting data points to correct misclassifications from each new weak learner. During the process, more weights are added to misclassified points to focus on the samples that are difficult to classify correctly. Different from Ada boosting, Gradient Boosting employs decision trees as base learners, while focusing on minimizing the residuals or errors made by the previous models. It adjusts the target values for each iteration to correct the errors made by the previous model. Gradient boosting typically includes a learning rate, often referred to as the shrinkage parameter, which scales the contribution of each base learner [183, 184]. Conclusively, adaptative boosting tries to address a specific 'local' optimization problem during each iteration, whereas gradient boosting employs a gradient descent approach and is more adaptable to a wide range of loss functions. Therefore, gradient boosting can be seen as an extension of Adaboost and has great potential for handling various differentiable loss functions. There are more techniques applied within ensemble learning for improved accuracy, and robustness, such as Stacking [185], Stochastic Gradient boosting [186], XGboosting [187], The choice of ensemble method depends on the specific problem and dataset. Specifically, Deep ensembles have been used for estimating

uncertainty for deep learning-based NDE applications, such as Ultrasonic Tomography [188], Magnetic Flux Leakage Testing [94], Guided wave array imaging [189] and etc. Li et al. investigate the impact of the liftoff uncertainty in measurement data in a magnetic flux leakage-based defect depth classification. Prediction accuracy and uncertainties are estimated and compared between the Bayesian Neural Network and Deep Ensemble methods, which demonstrated the efficacy and feasibility of DE in uncertainty propagation [94]. In an ultrasonic tomography-based speed of sound reconstruction application, the ensemble approach is able to provide robust uncertainty in terms of estimation variance and model variance [188].

Generally, there exists a set of candidate models to describe the inverse process, it is hard to obtain the optimal posterior model probability between them. Therefore, the Bayesian solution is able to simplify the model selection by taking the weighted average of all candidate models and weighing each by its marginal posterior probability [190, 191]. This strategy is called Bayesian model averaging (BMA), which is considered a transition method between Bayesian NN and Deep Ensemble method. Generally, for constructing BMA, the prior distributions over all parameters in all models and prior probabilities of all models must be specified. As the priors on model parameters need to be specified for determining the posterior model weights, any insufficient or uncertain knowledge of those prior model information will lead to uncertainty in final predictive probability distribution. Therefore, BMA is a useful method for investigating the structural uncertainty from the 'Learning' model. Theo S et al. investigated how different model prior distributions affect the economic growth determination results [192]. In an NDE application of strength estimation for aging materials, BMA is applied to understand the model uncertainty which also could provide a reliable approximation of the actual underlying model for predicting the bulk mechanical properties [193].

There are other popular uncertainty propagation methods for inverse learning, including Bayesian Active Learning (BAL), Deep Gaussian process, Bayes by Backprop(BBB), etc. Besides, lots of researchers often combine more than one method to enhance the estimation performance by combining the advantages of multiple methods, such as Cyclical stochastic gradient MCMC

(CsgMCMC) [194] or BNN with MC dropout [195], which are applied for handling different scenarios in investigating uncertainty propagation from learning to data. Only a selection of well-known NDE-based inverse learning uncertainty propagation techniques are covered and there's no one-size-fits-all approach for addressing specific problems. Conclusively, the selection of inverse UP approaches can take the following criteria into account:

- Measurement data: data amount; prior distribution;
- Learning Model: direct or indirect; model parameter distribution; bias and variance; Computation efficiency;
- Estimated Uncertainty type: structural, parametric, or total uncertainty.

4.5 TLS-Based Inverse Uncertainty Learning Application

4.5.1 Introduction

Pipeline infrastructure is a crucial part of society, with over 2.5 million miles of natural gas, petroleum, and hazardous liquid pipelines operating across all states, as reported by the Pipeline and Hazardous Material Safety Administration (PHMSA) of the United States Department of Transportation (DOT). Information about incidents on gas and liquid pipelines is accessible to the public [196]. Pipelines are susceptible to damage from internal and external corrosion, cracks, manufacturing defects, etc., which cause severe leakage to induce safety concerns, therefore, a reliable assessment scheme for monitoring and maintaining that critical infrastructure is imperatively critical. Cracks, which are caused by different mechanisms, can appear in the pipeline at any stage during manufacturing, installation, or throughout its service life [197].

Crack sizing and profiling are of critical importance in monitoring crack growth to define inspection intervals for safety-critical components, therefore, nondestructive evaluation (NDE) methods have been applied in detecting and characterizing cracks' size, shape, and orientation [198]. Ultrasonic testing (UT) methods are considered suitable for structural integrity damage monitoring systems and characterizing surface cracks by applying Rayleigh wave and acoustic

emission[199, 200]. Also, a variety of electromagnetic (EM) methods, such as eddy current testing (ECT), microwave, MFL, etc., are advanced in detecting and identifying surface and sub-surface cracks in metal, based on the electromagnetic principles [201–205]. In MFL-based pipe inspection, lots of effort has been put into characterizing metal loss defect inversion problem, which is associated with length, width, and wall loss (%WL), obtained from the measured three-axis MFL signals in terms of magnetic flux density [206–208].

For the widely used magnetic flux leakage (MFL) inspection method, the collected signal quality is greatly affected by various uncertainty factors, such as material property, inspection variations, shape irregularity, noises, etc. Therefore, a quantified uncertainty estimation in NDE is indispensable. Deep learning proves to be an effective approach for handling extensive NDE data at scale without requiring prior physics expertise [209–211]. Nevertheless, conventional deep learning methods lack the capability to assess the reliability of predictions. In contrast, Bayesian-based techniques excel in estimating uncertainty by incorporating Bayesian posterior inference across the neural network parameters [212].

This study delves into the impact of uncertainties arising from the dynamic magnetization process, specifically due to the relative motion between a magnetic flux leakage (MFL) sensor and the material being tested in both axial and circumferential directions. In the context of MFL inline inspection, the surface roughness of the material plays a pivotal role in influencing sensor liftoff and stands out as a significant source of uncertainty affecting inspection outcomes. Consequently, this research delves into the uncertainties stemming from sensor liftoff, examining their propagation throughout the sensing system and their influence on output data. Given the intricacies involved in describing the forward uncertainty propagation process, the study employs Deep Ensemble, a learning-based non-Bayesian uncertainty estimation method. This approach addresses the input uncertainty originating from the response MFL data. To assess performance, a three-dimensional finite element method (FEM) based model generates simulation data for MFL-based defect depth classification. Experimental data are then used to validate MFL-based defect size classification. The study conducts prediction accuracy and uncertainty calibration, proving its value in enhancing

prediction performance assessment and quantifying uncertainties. Additionally, an Autoencoder method is applied to compensate for the shortage of experimental data available for training the uncertainty estimation model. This approach extends to address the challenge of insufficient experimental data in generalized non-destructive evaluation (NDE) problems.

4.5.2 Uncertainty in MFL-based Defect Classification

Without the formulation of models based on complex physics knowledge, Deep Learning methods relied on available comprehensive NDE data, which have shown great advance in aiding nondestructive evaluation methods. NDE data is often complex, massive, discordant, and noisy, it is very necessary to jointly develop UQ with an appropriate deep learning model to efficiently deal with existing uncertainty to improve the safety of the inspection system. The predictive probability obtained from deep learning assists in deciding the probability interpretation and quantifying the uncertainties of the predictions to accomplish statistical inference [213]. Because of the uncertainty among the whole system, the network predictions are usually misleading, therefore a good predictive uncertainty score can quantify how reliable the model's prediction is, which is considered a good basis for assessing the performance.

To develop a reliable NDE-based inspection system, a thorough understanding of the system and its influential parameters is essential. The NDE system is characterized through the forward and inverse modeling process, as detailed in [36]. In the forward stage, variations in geometric parameters (e.g., defect size and shape) and material properties (e.g., hardness and strength) are considered as aleatoric uncertainty in applications related to material characterization and damage detection [88, 89, 91, 92]. Processing parameters related to simulation (e.g., mesh parameters, boundary conditions) and experimental testing (e.g., setup process, experimental noise) are regarded as epistemic uncertainty [93–95]. This variability constitutes input uncertainty, which is integrated into the subsequent inversion stage. In this stage, modeling and analysis are employed to derive predicted parameters describing the system based on observed measurements or simulated output from the forward procedure [96]. During the inversion process, epistemic uncertainty is introduced, associated with the learning model parameters and the model itself.

During MFL in-line inspections (ILI), surface irregularities like changes in coating thickness, welds, or hardness deposits introduce variations in the liftoff distance, complicating the inspection process. These fluctuations impact the amplitude of MFL signals, influencing detection sensitivity [214]. Therefore, exploring liftoff distance is a critical uncertainty factor in MFL inspection. Other considered uncertainty factors include sensor velocity [215], defect size and shape [180], microstructural changes, and mechanical properties [216], among others. Additionally, during the inversion process, NDE field inspection results are often sensitive to environmental conditions and signal processing methods [217]. In this study, uncertainties arising from liftoff are examined in the inverse process for defect classification, where uncertainty from MFL data and the machine learning model is quantified using the approximated Bayesian Inference modeling process.

4.5.3 Autoencoder for Data Augmentation

The application of machine learning algorithms in analyzing NDE experimental data is often hindered by the challenging and costly nature of data collection procedures. To tackle this issue, a potential solution is to employ data augmentation methods, which can extend the existing dataset by generating more diverse and comprehensive training data. Autoencoder, a typical unsupervised multilayer neural network, is utilized to compress and decompress input data for data augmentation purposes. The core concept of Autoencoders is not to perfectly recreate the input data; instead, a controlled amount of error or noise is introduced intentionally. The goal of the Autoencoder is to train the network to minimize the discrepancies between the input data and the reconstructed data with proper loss function while retaining some certain similarity between the original input and the recreated output for enriching the original dataset. Autoencoder neural networks' capability has been demonstrated in areas such as image reconstruction [218], feature extraction [219], augmenting data for anomaly detection [220], noise reduction in medical images [221]. It consists of a pair of an encoder and a decoder, where the encoder is able to generate the compact representation for a whole set of data, which is then passed to the decoder to reconstruct the original data from this simplified representation with high fidelity [222].

To boost the learning efficiency of the Autoencoder on MFL experimental data, an initial pre-

trained phase is conducted on simulation data. This utilizes transfer learning, a technique where knowledge acquired from one task is applied to a related task. The context of this study involves using a pre-trained model as a starting point and fine-tuning it for MFL classification on experimental data. Transfer learning proves advantageous when there is limited data for the new task, as it leverages knowledge from the original task [223, 224]. Widely applied in various ML studies, transfer learning has demonstrated effectiveness in tasks like translation, image recognition, and image classification [225, 226]. The similarity in format and sensing methods between experimental and simulated MFL data, along with the larger size of the simulated data, enhances the effectiveness of transfer learning in this study. Through transfer learning, we enhanced the performance of pre-trained Autoencoder models using our experimental dataset, leading to improved results and a reduced need for a high number of experiments.

The applied Autoencoder model architecture and transfer learning process is illustrated in Figure 4.2. Specifically, the encoder stage of the proposed model employs two pairs of convolutional layers with max pooling operations, activated by the ReLU non-linear activation function, to capture essential representations. To prevent overfitting, a dropout layer is added as regularization. The number of kernels ensures a consistent number of activations across layers. These layers serve as feature extractors, creating a compressed feature representation space (Z). The encoder parameters are initially pre-trained on a large simulation MFL dataset, establishing a foundation for learning general MFL signal features. The pre-trained encoder facilitates the model's adaptation to specific experimental data features due to the intrinsic connection between simulation and experimental data.

During the processing of MFL experimental data, the pre-trained encoder layers' weights remain frozen, preventing the loss of valuable information. The subsequent trainable decoding layers mirror the encoder, learning to reconstruct the original images. The decoder includes upsampling layers to restore Z to the original image size. This Autoencoder model transforms old MFL signal features, providing predictions trained with the experimental dataset. The transfer learning process fine-tunes the model, allowing it to adapt to the unique characteristics of experimental data. This approach,

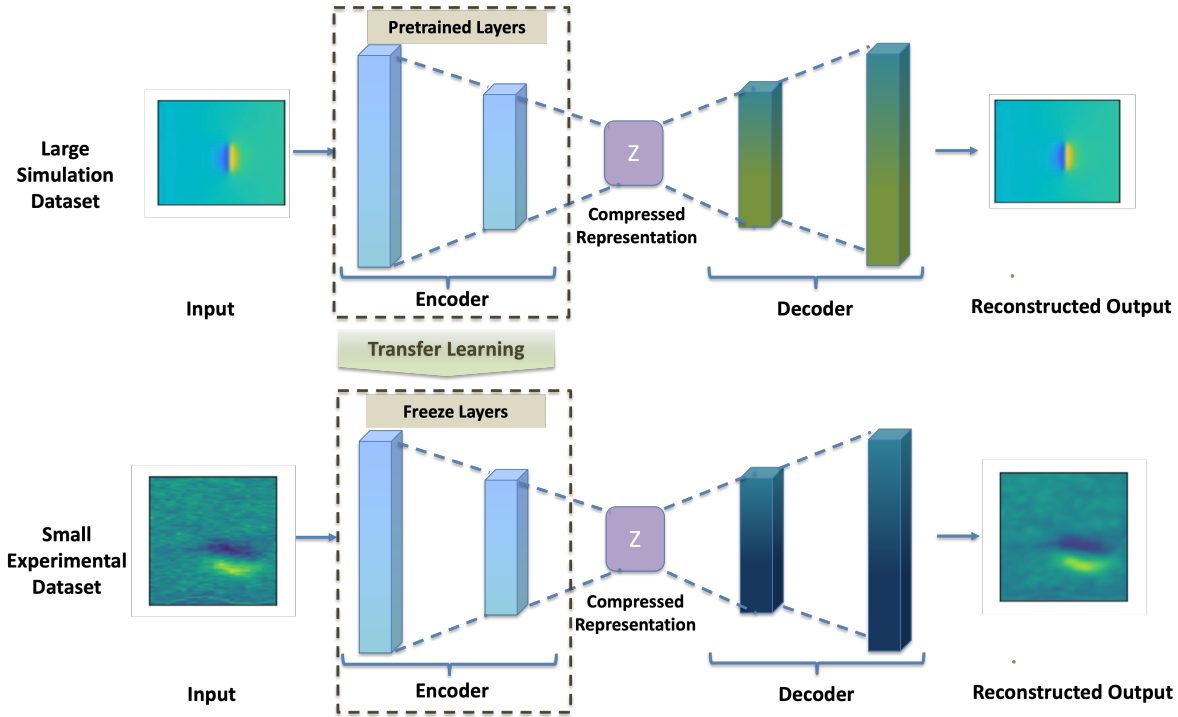


Figure 4.2 Schematic representation of the Autoencoder architecture and transfer learning process.

compared to training from scratch, typically results in improved performance.

Once the fine-tuned Autoencoder model is optimized, it serves the purpose of augmenting the experimental MFL data. This involves feeding MFL images into the Autoencoder for reconstruction, and then adding the resulting reconstructed images to the original experimental dataset. The original dataset is labeled "OR," while the combined dataset of the original and newly generated MFL data is labeled "GE." The primary goal is to enhance the training set for learning-based networks, leading to improved classification and prediction performance.

4.5.4 Applied "Learning" Models for Uncertainty Estimation

As mentioned in Chapter 4.1, Bayesian theory is considered as the primary approach to address uncertainties through the "Learning" model, which aims to comprehend and describe uncertainty in the inverse solution based on observations data, and other sources of information (e.g., prior distributions). As a result, probabilistic predictions can be made under the addressed uncertainties to assist in optimizing experimental design.

Bayes theorem is applied to the inference of a parameter given observed training input, which

is generated from a probability distribution depending on an unknown parameter ω . In this application, the obtained probability distribution $P(\omega | X_{mfl})$ is used to describe the relationship between the input MFL image data X_{mfl} and their associated defect classes D , $D \in 1, 2, 3$. The uncertainty in this process can be obtained from the variance of the predictive posterior probability distribution $P(\omega | X_{mfl})$, which can be expressed as:

$$p(\omega | X_{mfl}, D) = \frac{p(D | X_{mfl}, \omega)p(\omega)}{p(D | X_{mfl})}, \quad (4.5)$$

in which $p(D | X_{mfl}, \omega)$ is the likelihood of the model denoting the probability distribution of observed data X_{mfl} given the parameter ω ; $p(\omega)$ serves as the prior information of ω to describe the learning model, which is independent of any observation. During the modeling process of $p(\omega | X_{mfl}, D)$. Both prior and likelihood are considered known parts of the assumed model, while the probability distribution of the predictive distribution of the output for a new MFL input image x^* and its output class d^* , $p(D | X_{mfl})$ can be expanded as:

$$p(d^* | x^*, X_{mfl}, D) = \int p(d^* | x^*, \omega)p(\omega | X_{mfl}, D) d\omega \quad (4.6)$$

As the learning process of this posterior distribution is intractable in higher dimensions, some approximation techniques are used to fit the true posterior distribution to find an analytical way to evaluate the process with tractable approximating variational distribution $q_\theta(\omega)$ parametrized by variational parameters θ . This process is usually described as the Bayesian Inference [227].

As we discussed before, the sensor lift-off is the main aleatoric uncertainty source in this experimental MFL inspection, in order to investigate how these factors affect the inspection performance, two typical deep learning methods: Convolutional Neural Network (CNN) with Dropout and Deep Ensemble (DE) are proposed and compared to estimate predictive uncertainty with the realization of approximate Bayesian inference.

4.5.4.1 Convolutional Neural Network with Dropout

In the realm of ML-based networks, the Dropout technique, which randomly discards some of the model units during training, is not only effective in avoiding overfitting but can also serve as

an approximation of the Bayesian process [177]. In this CNN-based application, the conventional Bernoulli Dropout is applied for sampling each unit output with a certainty probability. Besides, considering this is a classification problem, the predictive probability $P(d | x, \omega)$ is a categorical distribution that corresponds to the softmax likelihood [228], which can be written as:

$$P_{\omega}(d_i | x_i) = \hat{d}_i = \text{Softmax}(f^{\omega}(d_i)) \quad (4.7)$$

where $f^{\omega}(x)$ represents the neural network. As proven in [178, 180] that the sampling process from $q_{\theta}(\omega)$ is the same as the dropout operation, the Bayesian inference process L as addressed in the previous section, that can be approximated as $L_{dropout}$. Therefore, a CNN model with dropout is applied as the realization of the Bayesian approximate learning process for addressing the uncertainty in this defect classification application.

The detailed learning process of the proposed CNN model is shown in Figure 4.3, where the Convolutional layers with Maxpooling are used to extract features from the MFL input image. The following fully connected layers with the Dropout layer are employed to combine extracted high-level features for classification purposes. The outputs from these layers are then passed through the softmax activation function, which assigns probabilities to each class label. For obtaining the posterior probability distribution for further uncertainty quantification, T times repetitive predictions are made for each MFL sample data. In the model learning process, all parameters are optimized by minimizing misclassification errors, resulting in a reliable output probability distribution for subsequent uncertainty estimation. This makes the approximation process a promising and effective universal approach for addressing uncertainties in classification-based NDE inverse problems within machine learning frameworks. It can lead to improved decision-making and risk mitigation across various NDE applications.

For obtaining the aleatoric and epistemic uncertainty of this work, as presented in [178, 180] the uncertainty is equivalent to the variance of the prediction probability of the network. Decomposing the prediction variance leads to a meaningful interpretation of the uncertainty, with the aleatoric uncertainty representing the randomness of the prediction defect class D and the epistemic uncertainty representing the variability coming from the proposed CNN model. Given the new MFL

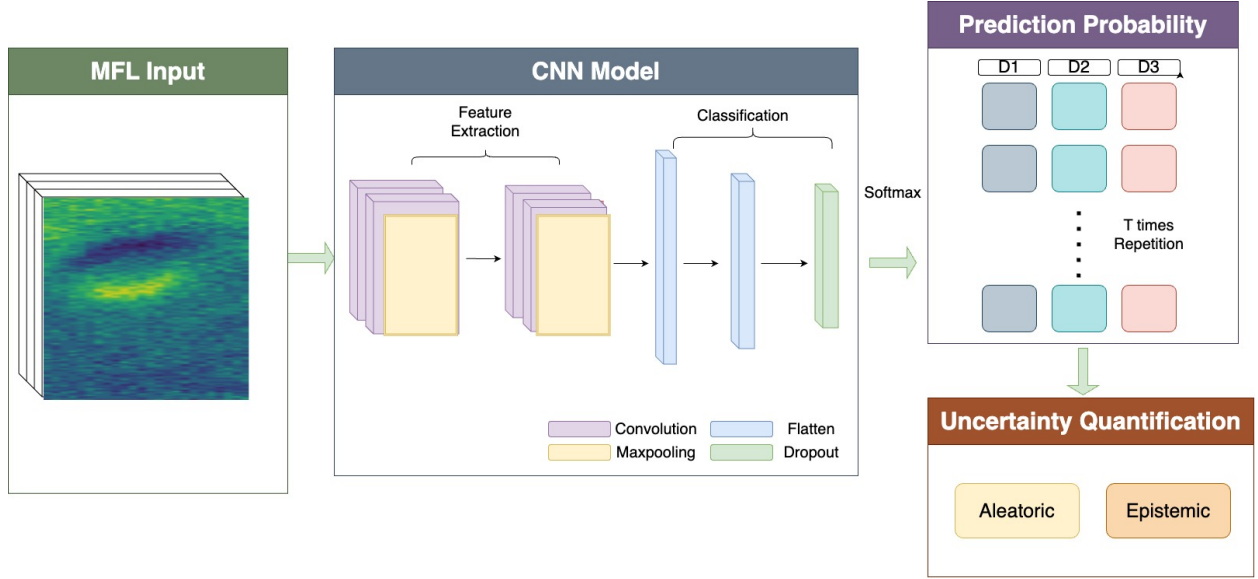


Figure 4.3 Schematic representation of the CNN process.

input image x^* , T times of prediction will be made for generating the corresponding new predictive probability d_t^* , ($t = 1, \dots, T$). Equation 4.8 introduces the correlation between the variance of the prediction probability and the uncertainty, which represents the total prediction uncertainty, comprising aleatoric and epistemic. the uncertainty can be expressed as:

$$\begin{aligned}
 Var_{q_{\hat{\theta}}(d^*)} &= Aleatoric + Epistemic \\
 &= \int \left[diag(E_p[d^*]) - E_p[d^*] E_p[d^*]^T \right] q_{\hat{\theta}}(\omega | X, D) d\omega \\
 &+ \int (E_p[d^*] - E_q[d^*])(E_p[d^*] - E_q[d^*])^T q_{\hat{\theta}}(\omega | X, D) d\omega.
 \end{aligned} \tag{4.8}$$

where $diag(\cdot)$ denotes the diagonal matrix. Since the softmax output is one-hot coded, in the variance, $E_p[d^*]$ square can be simplified as $diag(E_p[d^*])$. As addressed in [178], the first term is the expectation is over $q_{\hat{\theta}}$, which captures the inherent randomness of output defect classes, while the second term is only related to the network weight parameter ω , therefore, the Equation 4.8 can be split as Aleatoric A_i and Epistemic E_i of d_i^* :

$$A_t = \frac{1}{T} \sum_{n=1}^T diag \left[P_{\omega}(d_t^* | x^*) \right] - \left[P_{\omega}(d_t^* | x^*) \right]^{\otimes 2} \tag{4.9}$$

$$E_t = \frac{1}{T} \sum_{t=1}^T [P_\omega(d_t^* | x^*) - \bar{P}_\omega]^{\otimes 2} \quad (4.10)$$

where $\bar{P}_\omega = \frac{1}{T} \sum_{t=1}^T P_\omega(d_t^* | x^*)$. As the number of prediction repetitions (T) increases, the sum of A_t and E_t tends to converge in probability to the system variance. Aleatoric uncertainty is associated with liftoff variance, while epistemic uncertainty is linked to the model parameters of the proposed model. To evaluate aleatoric and epistemic uncertainty in this application, a dropout layer with the Softmax activation function is employed to generate predictions. During the prediction stage, each set of testing MFL data is predicted 10 times ($T = 10$) to obtain the variability distribution of the output. Consequently, for each testing MFL sample, there are ten aleatoric uncertainty results and ten epistemic uncertainty results, providing a manageable distribution to describe both types of uncertainties.

4.5.4.2 Deep Ensemble

In this MFL application, the random forest technique is utilized to classify crack depth and offer additional predictive uncertainty estimation. This helps examine the impacts of velocity variance and liftoff variance. Since the aleatoric uncertainty brought by the MFL signal is the focus of this work, the model uncertainty that comes from the hyperparameter is reduced by fixing the number of subtrees to ten. Besides, in order to achieve random initialization, k-fold Cross-Validation with a certain number of repeats is applied to split training and testing data into k-folds with a uniform probability distribution and randomized subsample to have an unbiased performance estimation. Unlike random train-test splits where a given example may be used to evaluate a model many times, this method is less biased because each example in the dataset is used only once in the test dataset to estimate model performance. Besides, addressing the limitation of k-fold cross-validation where the models tend to be highly similar in subsequent ensemble learning, random forest employs a bootstrapping technique to create different sub-datasets for each tree, bootstrap involves selecting examples randomly with replacement. Replacement refers to the practice of metaphorically returning the same example to the pool of candidate rows. This means that a specific example can be selected again, possibly multiple times, within a single sample from the

training dataset. Specifically, a set of decision trees is trained from a randomly selected subset of the new training data, which helps to reduce the correlation among the prediction results of the subtrees. It then grows a decision tree that is allowed to use only a random subset of features at each split. This diversity enhances the model's performance. Finally, the random forest averages the output of each decision tree to determine the final results. Calibrated outputs the predictive probability of each ensemble, which is set as a uniformly weighted mixed model. After repeating a certain number of evaluations to test data, the predictive probability is combined and averaged to make the final uncertainty prediction based on the scoring rules. Specifically, $k = 10$ in the DE model, while the first nine folds are used to train a model, the left holdout fold is used as the test set and each of the folds is given an opportunity to be used as the holdout test set. Totally ten models are fit and evaluated with three repetitions ($r = 3$), and the final performance of the model is calculated as the mean of these runs.

4.5.4.3 Predictive Uncertainty Estimation Scoring Index

Scoring rules are essential to evaluate the quality of predictive uncertainty, which is realized through a proper loss function in ML models. The training criterion of both CNN and DE is to minimize the cross-entropy loss for optimizing the predictive model and find the optimal model parameter ω , which shares the same formula as Log Loss, which can be presented as:

$$LL(\omega) = -S(\hat{d}, d) = -E(d) \log(\hat{y}) \quad (4.11)$$

where $\hat{d}$ and d are the predictive probability and true distribution respectively of input x . Each predicted probability is compared to the actual class output value (0 or 1) and a score is calculated that penalizes the probability. Besides, Brier score is a popular scoring rule in calculating the mean square error between the predictive probability and true classes, which can be expressed as follows:

$$Brier = E[(\hat{d} - d)^2] \quad (4.12)$$

Both the log loss and Brier score act as the evaluation of the predictive uncertainty, specifically, the higher the log loss and Brier score, the higher the uncertainty is buried in the system.

For ML modeling, the probability scores are overconfident or under-confident in some cases, which will bring bias to predictions that should be near zero or one and further affect the subsequent averaged prediction result. Therefore, calibration of predictions is an essential step to improve the reliability of probability predictions to make accurate probability estimates. Calibrated outputs the predictive probability of each ensemble, which is set as a uniformly weighted mixed model. It is a scaling operation to adjust the obtained probability distribution to match the expected distribution observed in the data [229]. Especially in the Random Forest method, because of the feature subset, the basic level trees are trained with a relatively high variance, which will bring errors to predictions that should be near zero or one and further affect the subsequent averaged prediction result. Therefore, the calibration of the log loss and Brier score is indispensable.

Platt scaling (Platt calibration) is a typical calibration method, which transforms predictions to posterior probabilities by passing them through sigmoid. Each calibrated predictive distribution can be presented as:

$$\hat{d}_{ci} = \frac{1}{1 + \exp(A\hat{d}_i + B)} \quad (4.13)$$

where $\hat{y}_i$ is the uncalibrated predictive output of true label of sample i . A and B are real numbers to be determined when fitting the regressor via maximum likelihood. The calibrated prediction is further applied to obtaining the calibrated scoring index for estimating the total predictive uncertainty in terms of Prediction Accuracy, Log Loss Score, and Brier score for evaluating and comparing the performance of the proposed CNN and DE model.

4.5.5 Performance Evaluation with Uncertainty Analysis

4.5.5.1 3D FEM Simulation Modeling of MFL

Maxwell's equations are applicable to the analysis of the electric as well as the magnetic field within MFL systems. In this work, permanent magnets are used for the generation of the magnetic field. For simulation study, the magnetostatic phenomena are governed by simplified Maxwell's equation illustrated below:

$$\nabla \times \left(\frac{1}{\mu} \nabla \times A \right) = J, \quad (4.14)$$

$$B = \nabla \times A, \quad (4.15)$$

where μ, A, J, B represent magnetic permeability constant, magnetic vector potential, the equivalent current density of permanent magnet, and magnetic flux density vector, respectively. The field equations are supplemented by the constitutive relation that describes the behavior of electromagnetic materials. In the permanent magnet region,

$$B = \mu H + \mu_0 M_0, \quad (4.16)$$

where M_0 denotes permanent intrinsic magnetization vector. The other region is governed by

$$B = \mu H, \quad (4.17)$$

On the other hand, the magnetic fields under the effect of mechanical motion are governed by Lenz's Law and Lorentz's Law. Lorentz's Law can be used for the analysis of the moving probe effect in a dynamic MFL inspection system. If the probe moves at a certain speed, the Lorentz force induces currents in the conductive specimen. Such currents in the specimen can be regarded as eddy currents dependent on the velocity at which the probe travels and the current density is expressed in

$$J_V = \sigma V \times B. \quad (4.18)$$

where J_V denotes eddy current density in the specimen; V denotes the speed of the applied magnetic field with magnetic flux density, σ represents conductivity of the sample. With respect to the dynamic EM system, the governing equation deduced from Maxwell's equations is added with eddy currents due to the movement of applied magnetic field. The modified equation for time-harmonic electromagnetic field is expressed as

$$\nabla \times \left(\frac{1}{\mu} \nabla \times A \right) = J_V - \sigma \frac{\partial A}{\partial t} + J_s. \quad (4.19)$$

The first, second, and third terms on the right represent velocity-induced eddy current, frequency-induced eddy current, and the current density due to the applied source respectively. Compared with the governing equation that precludes J_V , the modified equation implies that the eddy currents

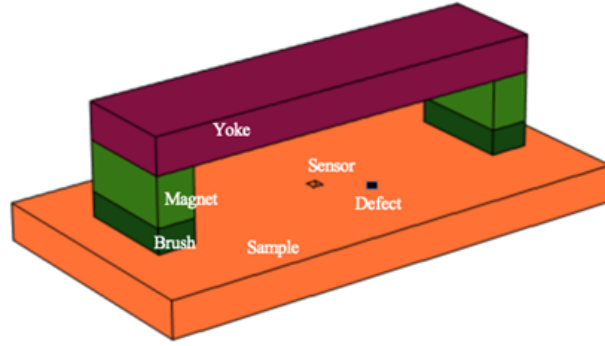


Figure 4.4 Geometry of MFL Simulation model.

generated by moving magnetic field influence not only the currents distribution in the conductive specimen but also the magnetic field profile, which results in distortion of the measured signal. Using the boundary condition, the magnetic potential vector can be solved. Then the distribution of the magnetic field can be obtained. 3-D FEM is applied to analyze flat samples by COMSOL.

Figure.4.4 shows the geometry of the problem in a 3D model, and the influence of the defect depth and lift-off on magnetic flux leakage density from the defect is being studied. The magnetic circuit is constituted by a yoke, magnets, brushes, and specimen, and a rectangular defect located at the center specimen. In the model, two permanent magnets, made of NdFeB material, are used as the magnetic flux induction, the yoke and brushes use the same material of mild steel, the relative permeability of which is 186,000, while the sample is made of Stainless Steel 416. In the process of calculating for the finite element model, magnetization clearance (clearance between brush and specimen) is equal to sensor liftoff. The most fundamental element of 3D is a tetrahedron. To have a precise result, we refine the elements near the flaw. The influence of eddy current under different speeds of 3,5,7m/s is taken as uncertainty. In each velocity case, the effect of liftoff variation is also considered. Consequently, it is practicable to arrange measurement of the magnetic field for defect detection at the liftoff ranging from 1mm to 9mm.

4.5.5.2 MFL experimental Setup

MFL experiment was conducted on a stainless steel sample containing three kinds of defect size which is presented in Table 4.1. Different sensor liftoff during the data collection ranges from 1mm, 2mm, and 3mm respectively. Specifically, each defect is subjected to 60 times testing under

each liftoff scenario. This results in the collection of 180 sets of MFL images for each type of defect with each image having dimensions of 217x217 pixels in RGB format. Consequently, a total of 540 sets of experimental MFL data are gathered and labeled as "OR" for subsequent analysis.

Table 4.1 Experimental MFL Defect Dimension

	Diameter x Depth (inch)
Class 1	0.367" x 0.15"
Class 2	0.505" x 0.12"
Class 3	0.633" x 0.12"

4.5.5.3 Performance Evaluation for Autoencoder-based Transfer Learning

To initiate the process of fine-tuning the network weights through pre-training, a set of 1500 MFL simulation images from the simulation model is employed for depth classification. Typically, three defect depths are considered, which are equally divided from varying depths from 2mm to 10mm. In the training of the pre-trained model, 70% of the total simulation MFL data is used for training, while the remaining 30% is set aside for validation. After updating the network layers and obtaining the optimally compressed representations, the Autoencoder model is further fine-tuned with the experimental data in which 70% are allocated for training purposes. Before evaluating the performance of the applied Autoencoder in data augmentation, the effectiveness of the proposed Autoencoder-based transfer learning approach in this application is addressed.

The Mean Squared Error (MSE) loss is a standard metric for assessing the accuracy of trained neural networks (NN). It calculates the average of the squared differences between predicted and actual target values. In this section, we compute the MSE loss for our proposed Autoencoder network on MFL data and compare two cases. In the first scenario, the Autoencoder is directly trained and tested on the experimental dataset. In the second case, the Autoencoder model is pre-trained on a larger simulated dataset and then fine-tuned with the experimental data. The results, shown in Figure 4.5, reveal a similar progression for both models, starting with relatively high loss values and gradually improving, with significant reductions, especially after 25 epochs. Notably, lower MSE in the testing data during validation indicates the model's effective generalization to

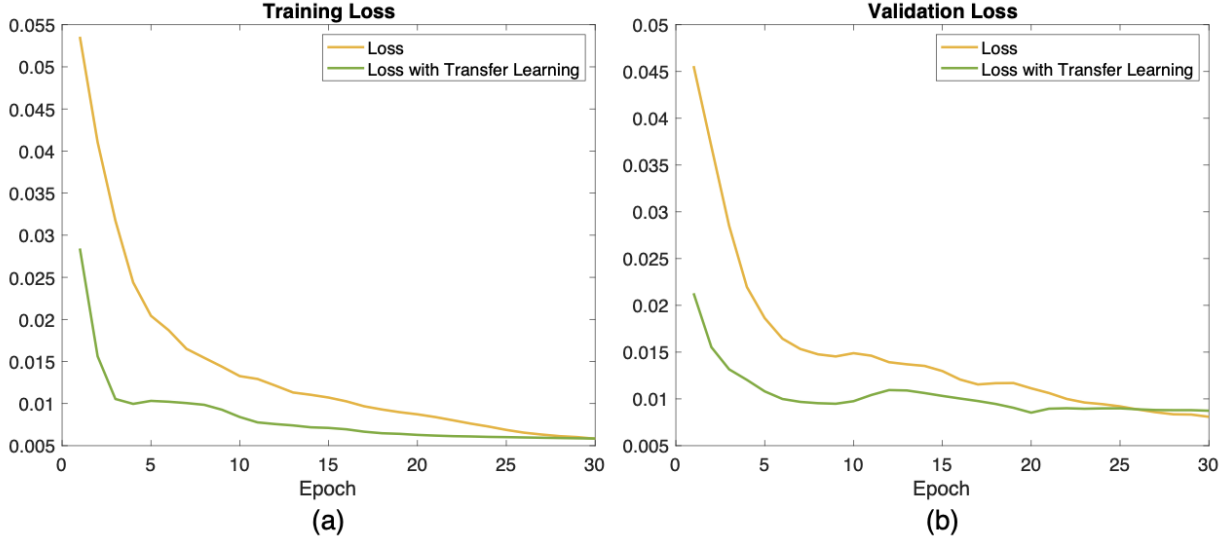


Figure 4.5 Comparative loss analysis with and without transfer learning: a) Train Loss; b) Validation Loss.

new, unseen data. The visual representations demonstrate that the Autoencoder excels at learning valuable features from MFL data. Additionally, transfer learning accelerates convergence to an acceptable loss level within a smaller number of training epochs, highlighting the benefits of leveraging existing knowledge for faster learning and improved generalization.

The optimal pre-trained Autoencoder network, obtained from training on experimental data, is used to validate the remaining 30% of the data. During this process, the newly generated data is employed to augment the original experimental MFL dataset. Consequently, 154 additional sets of newly generated data are combined with the original dataset labeled as "OR" and denoted as "GE." To assess the impact of data augmentation, the relationship between the "GE" and "OR" datasets is examined by analyzing the outcomes of the proposed CNN and DE models in terms of direction and strength.

The directional relation can be realized through the covariance matrix, which is expressed as:

$$COV(S_{GE}, S_{OR}) = \frac{1}{N-1} \sum_{i=1}^N (S_{GE}(i) - \bar{S}_{GE})(S_{OR}(i) - \bar{S}_{OR}),$$

$$S = \{Accuracy, LogLoss, Brier\} \quad (4.20)$$

where i denotes the number of liftoff variance. $\bar{S}(\cdot)$ is the averaged scoring index. Moreover, the correlation indicator is applied to determine how strongly two variables are related, which can be

written as follows:

$$CORR(S_{GE}, S_{OR}) = \frac{cov(S_{GE}, S_{OR})}{\sigma_{S_{GE}} \sigma_{S_{OR}}} \quad (4.21)$$

where $\sigma_{S(.)}$ denotes the standard deviation of the scoring index. Based on the "OR" and "GE" MFL data, the corresponding relation is evaluated in terms of classification accuracy and uncertainty scoring index: calibrated Log Loss and Brier score, which are presented in Table 4.2. The results from both CNN and DE models show that all the scoring indices' covariance indicators are positive, and their correlation indicators are close to 1. This confirms that there is a strong positive correlation between the "OR" and "GE" datasets, which further supports the feasibility and efficiency of using a pre-trained Autoencoder network to address data deficiency in MFL experimental scenarios. Therefore, the final augmented MFL dataset "GE" is used for further learning-based defect detection and uncertainty estimation.

Table 4.2 Performance Evaluation on Augmented MFL experimental data

	<i>S</i>	Direction	Strength
CNN	Accuracy	0.0025	0.9824
	Log Loss	0.0721	0.9806
	Brier	0.1963	0.9874
DE	Accuracy	0.0060	0.9965
	Log Loss	0.0725	0.9987
	Brier	0.0301	0.9955

4.5.5.4 Performance Evaluation for Uncertainty Analysis

As discussed before, network calibration is beneficial for improving the prediction reliability of modeling. For multiclass scenarios, Static Calibration Error (SCE) is usually applied to evaluate calibration performance by measuring the difference between the confidence and accuracy of a model [230]. Specifically, the model predictions are divided into N equally spaced bins separately for each class j and compute the calibration error within each bin. The final result is obtained by averaging the calibration error across all the bins. In the case of each bin B_{ij} , the accuracy $acc(B_{ij})$ represents the fraction of correct predictions, while the confidence $conf(B_{ij})$ corresponds to the

mean of the maximum probability for each data point. The SCE can be described as follows:

$$SCE = \sum_{i=1}^N \sum_{j=1}^M \frac{B_{ij}}{K} | acc(B_{ij}) - conf(B_{ij}) | \quad (4.22)$$

where N and M denote the number of bins and the total number of classes, respectively. K is the total number of data points. The corresponding calibration comparison results on MFL experimental classification problem for CNN and DE are illustrated in Table 4.3.

Table 4.3 The effects of calibration on the mean of SCE

	CNN	DE
Uncal	9.16%	8.72%
Cal	4.46%	3.64%

It can be seen that Platt Scaling leads to a noticeable decrease in prediction errors for both models. This reduction in prediction errors can help decrease uncertainty within the system and ultimately enhance the model's reliability, which provides a good basis for investigating the uncertainties brought by the liftoff variance.

To compare the uncertainty estimation of CNN and DE, we assess their performance on augmented MFL experimental data using the calibrated scoring index S with varying liftoff variance, as shown in Figure 4.6. In Figure 4.6 (a), the consistent decrease in prediction accuracy with increasing liftoff variation suggests sensitivity to liftoff changes in both models. However, CNN exhibits greater robustness, maintaining higher classification accuracy regardless of liftoff variation compared to DE. Regarding total predictive uncertainty in Figure 4.6 (b) and (c), an increase in liftoff variation raises total predictive uncertainty for both CNN and DE, showcasing their ability to assess uncertainties in this application. The results also highlight a more significant difference in accuracy and predictive uncertainty between liftoff 2 and liftoff 3 compared to liftoff 2 and liftoff 1, indicating an exponential decline in classification capability with liftoff changes. Although the Brier score implies less uncertainty in CNN than in DE, the overall results demonstrate that CNN achieves higher prediction accuracy and lower Log Loss than DE.

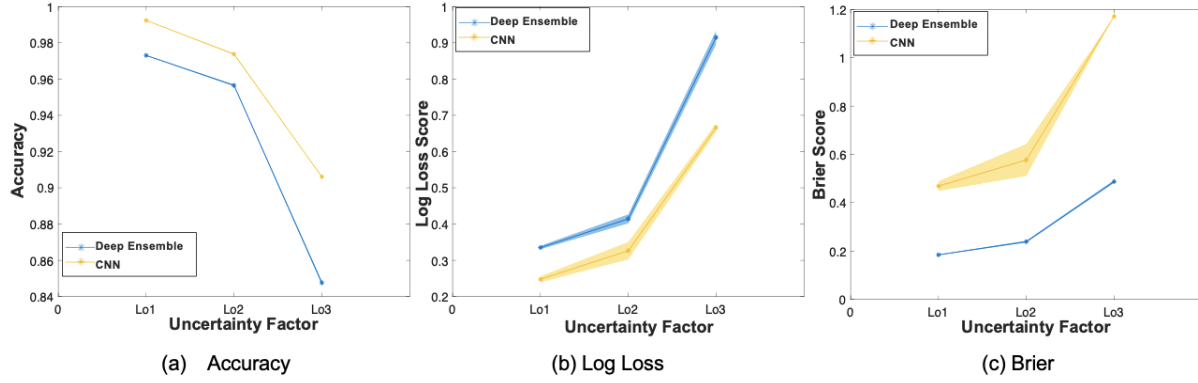


Figure 4.6 Comparison performance of mean (asterisk) and variance (shadowed bounds) for CNN and DE under different uncertainty: a) Prediction Accuracy; b) Log Loss; c) Brier.

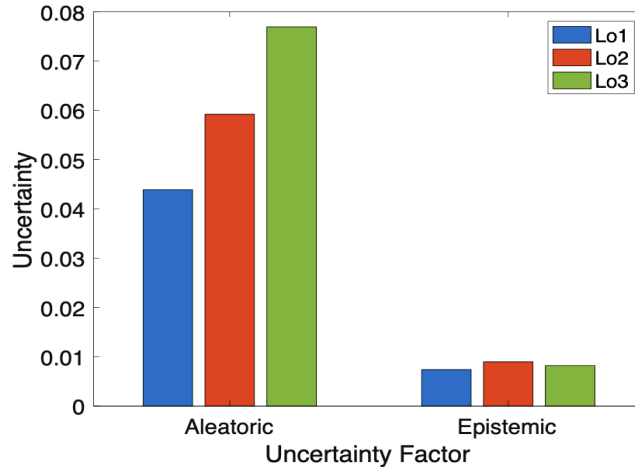


Figure 4.7 Uncertainty estimation of CNN.

Furthermore, as discussed earlier, the CNN model with Dropout can evaluate uncertainty components arising from the learning model and the data by identifying aleatoric and epistemic uncertainty during the prediction stage. The results are presented in Figure ???. The findings demonstrate a significant increase in aleatoric uncertainty from 0.04 to 0.08 under liftoff variation. Although there are fluctuations in epistemic uncertainty, aleatoric uncertainty remains around four times larger than epistemic uncertainty. Therefore, the uncertainty attributed to the model is negligible compared to the data uncertainty, and aleatoric uncertainty is mainly influenced by variations in the data.

The effectiveness of the proposed CNN model in uncertainty estimation has been discussed in previous sections. Now, we'll examine its capability to guide decision-making for new, unseen

MFL data.

Due to the time-intensive nature of collecting MFL data, four groups of new MFL data were obtained, comprising 36, 72, 108, and 144 sets of data, respectively. Liftoff variance was averaged across all samples in each group, and these four sample groups were input into the CNN model to evaluate its classification performance on defect size.

Figure 4.8 illustrates the percentage of wrongly classified instances among each sample concerning different liftoff variances. Larger liftoff values lead to higher prediction bias. On average, the percentage of incorrect classifications under varying liftoff uncertainties (from liftoff 1 to 3) is 25%, 30%, and 42%, respectively. It's important to note that this analysis used only four sample sets, and a more stable trend can be expected with a larger number of samples. However, a relatively consistent relation between liftoff variance and wrong classification percentage is observed, making results in Sample 4 a reliable representation of the true underlying trends in this application.

For better insights into the predictions under different liftoffs of Sample 4, the confusion matrix is utilized to provide a detailed breakdown of the model's predictions, as shown in Figure 4.9. With increasing liftoff uncertainty, class 1 defects maintain a high classification rate of 87.5%, while the other two defects, especially class 3, are more prone to misclassification. Additionally, using the confusion matrix, the F1 score is computed—a metric that considers false positives and false negatives, representing the harmonic mean of precision and recall. Among these three results, LO1 demonstrates the highest F1 score at 74.72%, indicating that increased uncertainty can reduce the model's ability to accurately identify positive instances and most of the actual positive instances. Overall, both the bar plot results and the confusion matrix support the previous observation in Figure 4.8 that the classification capability deteriorates exponentially with higher liftoff changes.

Further in order to propose a quantitative way to evaluate and determine the reliability of the classification to new input, two feature indexes are considered in this case:

- Confidence Index *CI*: indicate the degree of confidence in the classification performance, with higher CI values indicating lower uncertainty and vice versa [91]. The formula can be

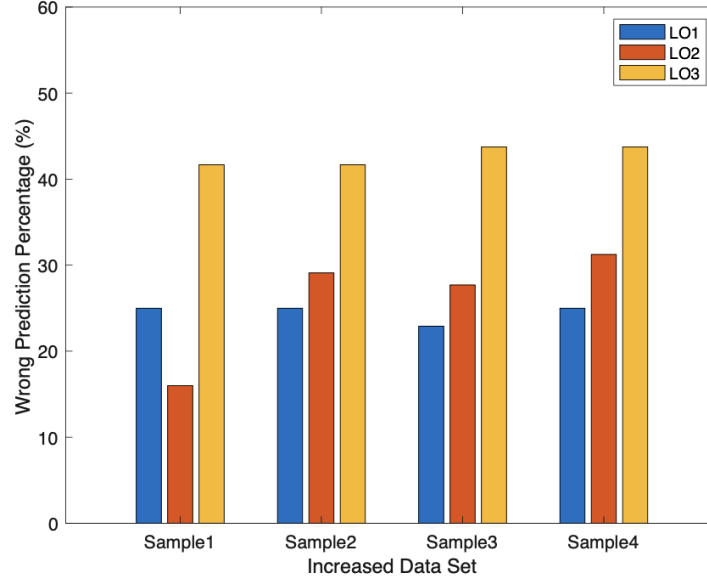


Figure 4.8 Prediction performance for new MFL data.

expressed as:

$$CI = abs(L1 - L2)$$

As each predictive probability is generated from the Softmax function, where a certain probability will be assigned to each class for one prediction. $L1$ is the negative log-likelihood of probability for the correctly classified class; while $L2$ is the negative log-likelihood of the maximum probability, calculated among the other wrong classes.

- **Weighted Predictive Uncertainty U :** The Log Loss, Brier, and Aleatoric uncertainty are all capable of revealing the uncertainty in classification to varying degrees. To combine these uncertainty indexes, the Minimum Redundancy Maximum Relevance (MRMR) method is used to rank them by finding the optimal feature set that can effectively represent the response variable while minimizing redundancy between features [231]. The resultant ranking determines the importance weights for each uncertainty index and thus generates the weighted total predictive uncertainty, which can be expressed as:

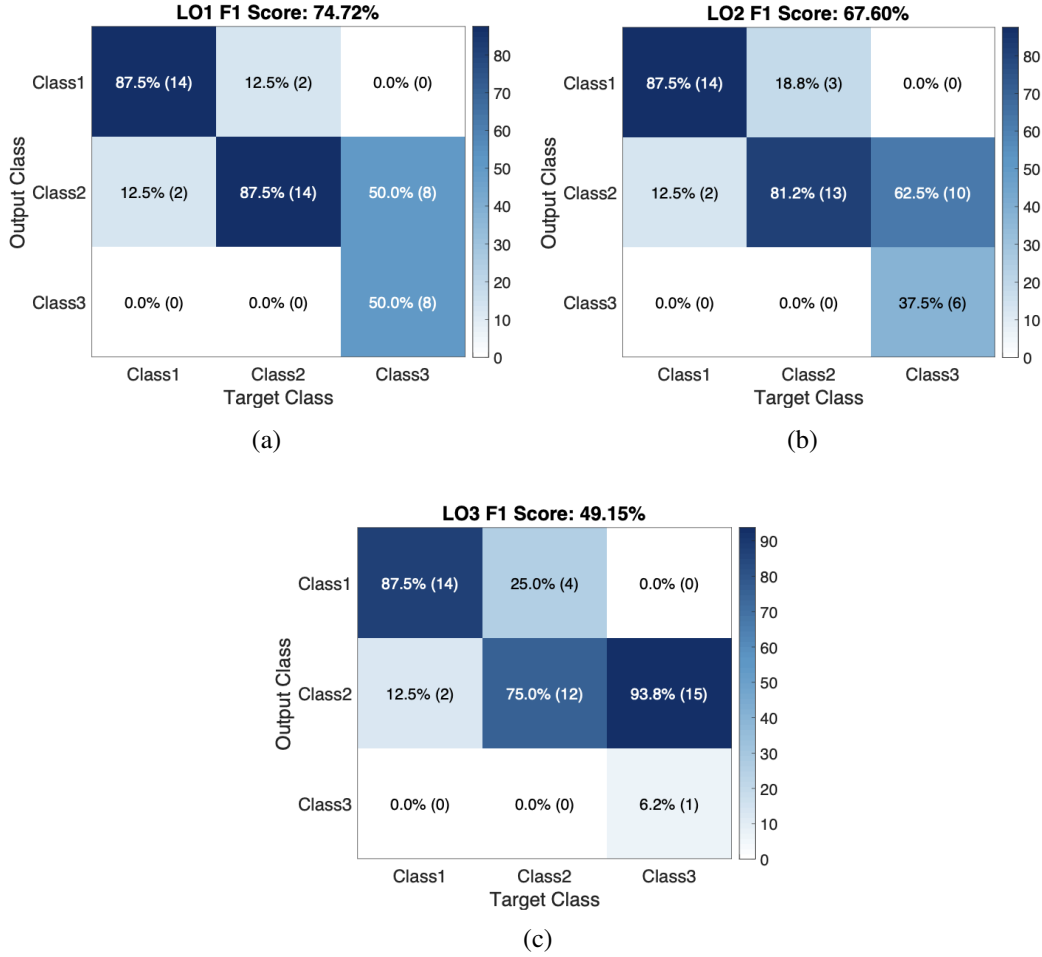


Figure 4.9 Confusion Matrix of prediction under different liftoff: a) LO1; b) LO2; c) LO3.

$$U = a * U_{loss} + b * U_{brier} + c * U_{aleatoric},$$

$$\{a, b, c\} = \{0.624, 0.318, 0.058\} \quad (4.23)$$

To assess the performance of defect classification on new MFL data, the corresponding feature indexes from Sample 4 are extracted and presented in Figure 4.10. In the figure, correctly classified data points are marked as red dots, while incorrectly classified ones are marked as green dots. Two boundaries are established to enhance the decision-making process guided by uncertainty: the Confidence Interval (CI) Threshold and the Weighted Total Uncertainty. The CI Threshold, marked with a brown line, serves as the initial boundary in uncertainty-guided decision-making.

It is determined based on the CI of the highest wrong prediction sample. Any samples with a CI higher than this threshold are considered to have a trustworthy classification, regardless of their uncertainty index. Otherwise, samples with a CI lower than this threshold will be evaluated using the uncertainty decision boundary, depicted with a pink curve. This boundary is generated using Quadratic Discriminant Analysis (QDA), a statistical algorithm used for classifying data into groups by modeling the distributions of the independent variables (predictors) for each group using a quadratic function [232]. When a new classification is made, the evaluation steps mentioned before should go through to determine the reliability of the classification. Based on these steps, examples of new MFL images with correct predictions and wrong predictions are presented in Figure 4.11 and 4.12. The corresponding feature indexes CI and U are listed, along with the True class and Prediction class. Based on the aforementioned steps, Figure 4.11 and 4.12 show examples of new MFL images with correct and wrong predictions, respectively. The corresponding CI and U, as well as the true class and predicted class, are listed. These examples emphasize the significance of considering both factors in the prediction process, demonstrating the effectiveness and practicality of the proposed uncertainty guidance process.

4.6 Conclusion

In this chapter, we delve into several classical and widely recognized methods for propagating uncertainty in NDE applications. These methods aim to assess how the posterior distribution varies during the 'Learning' phase, considering different distribution hypotheses relevant to NDE problems.

The inverse learning uncertainty propagation (UP) process is generally categorized into three scenarios: physics-informed, data-driven, and hybrid. Under these scenarios, different methods are employed to address two principal uncertainty types: parametric and structural uncertainties. Additionally, we explore an MFL-based defect classification problem which is an example of a hybrid learning-based UP analysis aimed at incorporating uncertainties into the final prediction. This uncertainty-guided decision-making process can provide more insights into the prediction results, therefore increasing the reliability of the classification results.

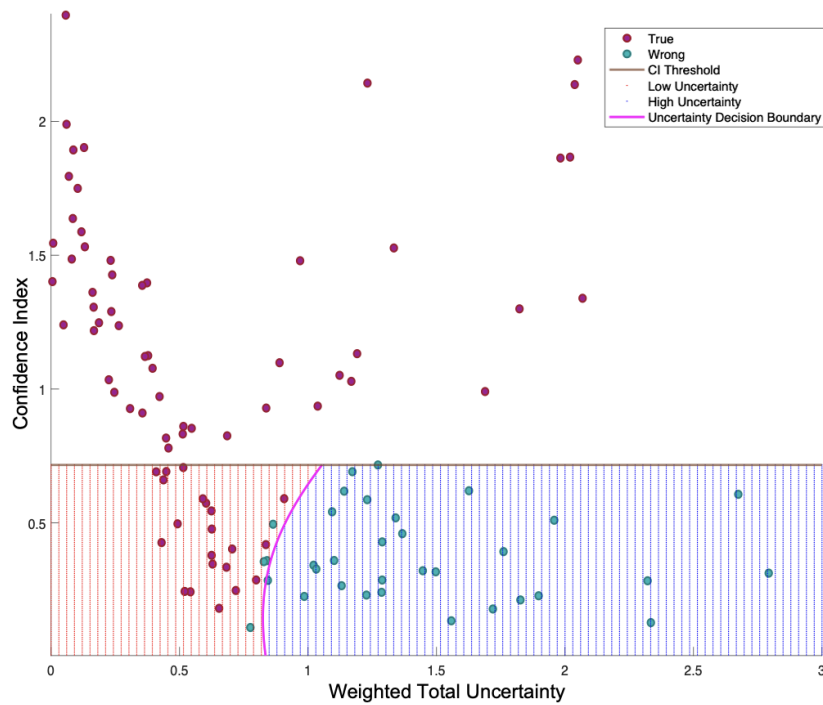


Figure 4.10 New MFL sample distribution based on the Confidence Index and Weighted Total Uncertainty with noted CI threshold (Brown) and Uncertainty Decision Boundary (Pink).

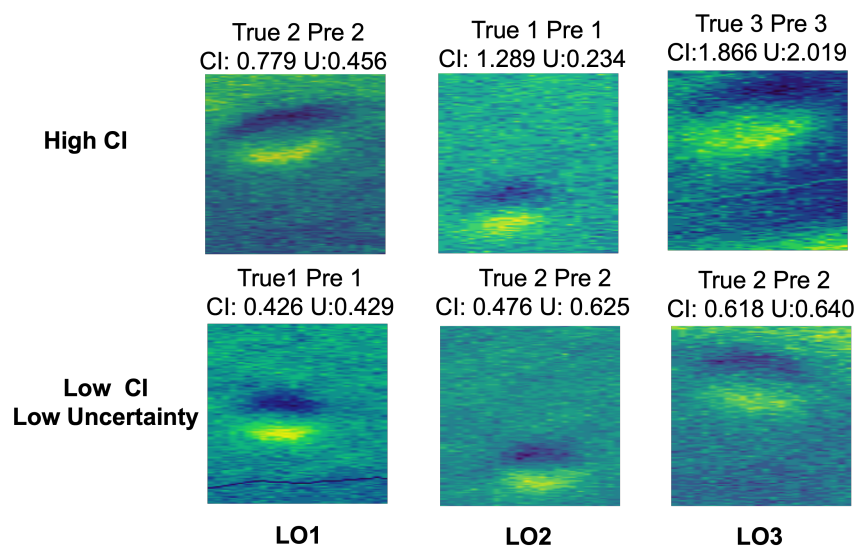


Figure 4.11 Examples of correctly predicted MFL images: First row images are of high CI, while the second row is of low uncertainty, even with low CI.

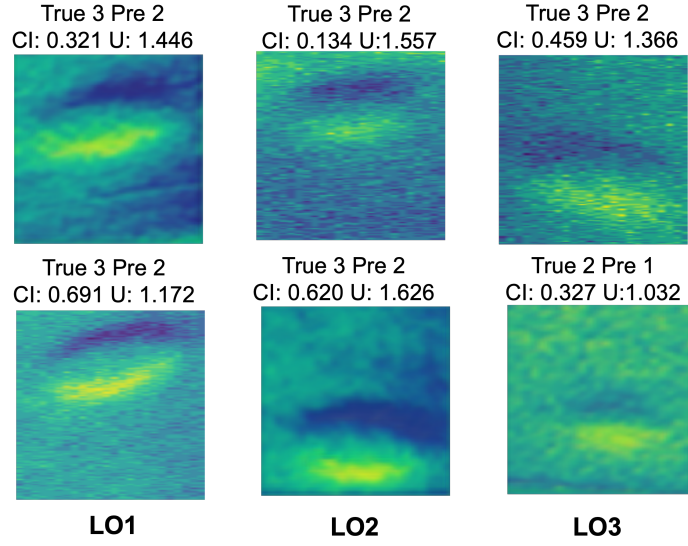


Figure 4.12 Example of wrongly predicted MFL images, which are of low CI and high uncertainty.

Further, a Bayesian approximation-based learning model is applied as a supportive case that provides a comprehensive and practical solution for uncertainty estimation in experimental MFL defect classification. This work has introduced a valuable research framework aimed at classifying and enhancing prediction reliability by incorporating transfer learning-assisted Autoencoder-based data augmentation, learning-based defect classification, and uncertainty analysis. Further, we proposed guidance to determine the reliability of the classification on new unseen MFL data with two features indexed confidence index CI and weighted total uncertainty. Those key elements are collectively contributing to the robustness and effectiveness of this work, and further enhancing the reliability of the classification outcomes.

CHAPTER 5

RELIABILITY EVALUATION TO NDE PROCESS WITH UQ

Reliability in NDE encompasses the ability of the NDE method to provide consistent and accurate information about the inspected material or component. Other than the aforementioned uncertainty among stools, there are numerous controlled and uncontrolled factors that are related to the reliability of the NDE inspection process. As mentioned in [233], relevant uncertainty may come from four aspects: specimen condition, sensor types and numbers, inspection setup and calibration, environmental conditions, and operator expertise. To have a better understanding of the system reliability, depending on the availability of specified uncertainty sources, popular uncertainty estimation technology: the measurement uncertainty and probability of detection (POD), are investigated in this section.

5.1 Probability of Detection

Probability of Detection (POD) is a widely adopted approach for quantifying the capability of an NDE inspection in terms of providing statistical description in identifying cracks or more generally defects in structures [234–236]. As mentioned in the previous section, given that parameters of defect characters or material properties are uncertain, and that all inspection equipment has noise, which will directly affect the sensing systems' sensitivity in defect characterization. In this case, POD is able to evaluate the likelihood that a given NDE method will correctly detect a flaw or defect of interest based on a specified uncertain characteristic developed through experimental testing, thus providing critical information for quality control, safety, and compliance in various industries [237].

Generally, two common variants of POD analysis are used to describe the detection results, referred to as $\hat{a}$ vs a and Hit/Miss. Hit/Miss analysis is a typical approach to measure how likely it is that the flaw will be detected. In cases where establishing a clear relation between flaw size and flaw response is challenging or quantifying the response is difficult, the target is to obtain the POD curve based on binary results based on a clearly defined hit/ miss criterion: hits (correctly found cracks) and misses (cracks not found in the inspection). In this binary detection scenario,

the testing results can be scaled in terms of four terms: True positive (TP), False negative (FN), False positive (FP), and True negative (TN). The corresponding POD can be modeled in terms of the probability of TP [236]:

$$P(TP) = \frac{TP}{TP + FN} \quad (5.1)$$

Different from the Hit/Miss method, the $\hat{a}$ vs a method, also named as signal-response data, deals with the response signal values directly, thus containing more information related to the detected flaw, such as flaw size and location. Specifically, where $\hat{a}$ is the measured magnitude response or signal amplitude (such as crack size), which is the response from crack size a through estimated POD(a) function [233]. Considering the uncertainty affecting the measurements, the objective is to determine a decision threshold ($\hat{a}_{th}$) that reduces false alarms caused by noise while maximizing the detection of cracks.

The fundamental assumptions for both POD analysis models are similar: the POD curve of a specific parameter of interest (such as flaw size) versus the probability of detection exhibits a rising trend, which is shown in Fig.5.1. Ideally, the probability of detection can reach up to 100% with a sufficiently large defect size [238]. Considering the uncertainties in the NDE inspection, the realistic POD curve can not provide such dependence on crack size. A 90% detection probability with a 95% confidence level is set as the inspection capability evaluation index $a_{90/95}$. The detection reliability can only be guaranteed when the detected parameter is above the $a_{90/95}$ threshold. Therefore, the determination of this threshold is of great importance in POD analysis, which varies in different NDT applications [233]. As concluded in [239], the log-logistic distribution and an approximately linear relationship between $\ln(\hat{a})$ and $\ln(a)$ provide the most applicable statistical basis to derive the POD curve respectively. Further, the proposed POD curve basis can be extended and customized for different NDE applications. For example, a groundbreaking recommended practice (RP) was introduced for fatigue cracks in offshore structures in May 2015, which recommended several common POD distribution functions for different NDE methods like EC, UT, and MPI [240].

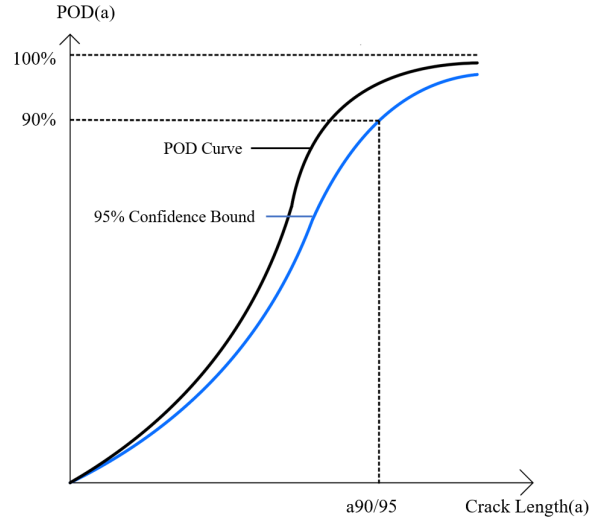


Figure 5.1 Typical POC Curve.

Based on the characteristics of each POD analysis model, the selection of the applied method varies for different types of NDE inspections, the data collected, and the specific goals of the analysis. Hit/Miss analysis has been widely applied in quantitative visual assessment of the flaw response for system reliability evaluation, such as visual inspection [241], penetration inspection [242], magnetic particle inspection [243], or ultrasonic testing [244]. The $\hat{a}$ vs a approach is applicable when a quantitative signal response can be correlated to flaw size as typically attainable with techniques, such as ultrasonic [244] or eddy-current inspection [245]. Further, to understand and compare those two POD modeling performances, a manual aerospace eddy-current inspection and a nuclear industry phased array ultrasonic weld inspection are analyzed of both models by Iikka et al [238]. Results have shown that uncertainties in inspector judgment or data format will provide significantly different POD curves.

Both the $\hat{a}$ vs a and Hit/Miss models have limitations in constructing a dependable full POD curve, restricted by the data availability, computation requirements, and costs. Therefore, a newer model-building technique, Model Assisted POD (MAPOD) is proposed, which requires less physical data and effort [246]. In MAPOD modeling, statistical models are employed to assist with true inspection data in improving the understanding of the nature of detected defects. Based on the information integration scenario, Meyer et al. have classified MAPOD into Bayesian and Nonbayesian

methods [247]. Non-Bayesian-based MAPOD approaches are chosen when there is a preference for using statistical or data-driven methods that may be better suited to the specific characteristics of the data or the nature of the defects being detected. For example, the physics-based modeling was performed with $\hat{a}$ vs a POD curve to determine the influence of fatigue cracks with ET [248]. Harding et al. used the transfer function method that relies on a data-driven linear regression model of POD parameters to estimate the POD for fatigue cracks in aircraft wings for UT inspection [249]. In contrast, Bayesian-based MAPOD relies on the availability of prior knowledge or expert opinion, which allows for a principled integration of existing knowledge with new data to make more accurate and reliable defect classification decisions. For example, in a high-frequency eddy current-based fatigue crack detection, the Bayes theorem is combined with computer modeling to provide 'prior information' to obtain posterior estimates of model parameters for generating more data. The resultant data are fitted with the logistics function generating a reliable Hit/Miss POD curve [250]. Besides, there are other model-based approaches to predict the POD curves in NDE area, such as signal/noise models, Image classification models, Human reliability index, and so on [251]. Overall, each POD model has its advantages and limitations within different NDE applications, it is hard to find uniform guidelines. However, Virkkunen et al. have divided the uncertainty sources in the POD estimation into five aspects [238]: statistical sampling error, measurement variation, configurational variation, variation in crack characteristics, and inspector judgment, which are also considered as the 'Data Uncertainty' in the proposed TLS model. They also concluded each POD model's capability in facing those uncertainties, which is beneficial in developing an optimal POD-based reliability analysis in NDE analysis.

5.2 GUM-based Measurement Uncertainty Evaluation

During the NDE measurement process, there may be uncertainties arising from factors like measurement instruments, environmental conditions, or operator influences. These combined uncertainties contribute to the variations in final measured outputs. Therefore, it is difficult to quantify the relationship between inputs and measured output directly. Instead of constructing a complex POD curve to describe each specified uncertainty factor, it is essential to provide an

overall estimate of uncertainty from the measurement process, which assists in evaluating how well the measurement accurately represents the quantity being measured. The Guide to the Expression of Uncertainty in Measurement (GUM) [62], is the most commonly used method to analyze measurement uncertainty, published by the Joint Committee for Guides in Metrology (JCGM). GUM is proposed to provide a standardized and systematic approach for accurately reflecting all propagated uncertainty relevant to the measurement and thus evaluating the quality of the numerical measurements.

In the content of the GUM model, the precision of the measurement system is decided from three aspects: repeatability, reproducibility, and maximum indication error of the instrument [252]. Repeatability focuses on quantifying the consistency of measurements of the same quantity under the same operations and equipment. Reproducibility assesses the consistency of measurements across varying conditions or different operators or using different equipment and methods. The indication error of the instrument pertains to the inherent accuracy and precision of the measurement instrument due to its limitations, which can be obtained from the producer's datasheet [253].

In the context of measurement uncertainty, Type A and Type B uncertainties are two distinct categories used to characterize and quantify different sources of uncertainty [254]. Theoretically, Type A uncertainty is associated with random or stochastic variations in measurement, which is quantified through statistical analysis of repeated and reproducible measurements or data. Type B uncertainty encompasses sources of uncertainty that are not characterized by random variations but are instead associated with systematic errors, approximations, or uncertainties in external parameters. These uncertainties are typically estimated through non-statistical methods, such as expert judgment, calibration data, or scientific knowledge. The detailed comparison between Type A and B uncertainty in terms of data source, probability interpretations, and statistical formula are summarized below:

- **Uncertainty Data Collection:** Type A relies on a series of observations in terms of Repeatability and reproducibility; while Type B relies on available relevant information such as manufacturer's specification, calibration certificate, Handbook, etc.

- Probability Interpretations: Type A is a frequentist method; while the Type B method is a Bayesian method.
- Statistically expression in terms of Standard Uncertainty:

$$\text{Type A: } U = \frac{S}{\sqrt{n}} \quad (5.2)$$

$$\text{Type B: } U = \frac{a}{\sqrt{3}} \quad (5.3)$$

Specifically, Type A doesn't require prior information about the system, and makes predictions only using data from the current experiment to obtain the probability density function; while the Type B method encodes prior knowledge of similar experiments, and combines it with current experiment data to make the probability density function. However, in some cases, prior information and mathematical modeling are hard to obtain. Also, the various uncertainty sources, in practical application, are correlated, which increases the difficulty in generating an accurate probability density function in Type A uncertainty analysis. Generally, Type A and B approaches may complement each other, and the choice of a particular analysis method should consider the specific needs of each application. To provide a statistical evaluation of the overall uncertainty, each type of uncertainty should be standardized to the same confidence level by converting them into standard uncertainty U_s . A standard uncertainty represents a range that can be conceptualized as 'plus or minus one standard deviation', which provides information about the uncertainty associated with an average value [254]. For type A, the standard uncertainty is calculated from the estimated standard uncertainty S , which is shown in Equ.5.2. For repeatability uncertainty analysis, n is the number of measurements; while for reproducibility uncertainty, n denotes the different measurement groups to make multiple independent tests. For type B, the standard uncertainty is represented with maximum allowable error a and normally, the upper and lower limits of uncertainty are assumed to fall into rectangular or uniform distribution. After each type of uncertainty is standardized, the easiest way to obtain the combined total measurement uncertainty U_c is to use 'summation in quadrature,

which can be expressed as follows:

$$U_c = \sqrt{U_{repeat}^2 + U_{reproduce}^2 + U_{system}^2} \quad (5.4)$$

For more complicated cases, there will be more variants for obtaining U_c relying on addition, multiplication, etc., which are discussed in [255]. Afterward, the expanded uncertainty U_e is needed to establish a confidence interval to describe total measurement uncertainty with a coverage factor k , which is expressed as:

$$U_e = k * U_c, \quad (5.5)$$

To provide coverage with 95% confidence when $k = 1.96$.

Following the GUM-guided measurement uncertainty analysis, the reliability of NDE inspections has been evaluated and tailored to the specific needs of each application. For example, Morana et al. evaluate measurement uncertainty associated with ultrasonic thickness measurements. The impact of response and contact surface roughness are investigated through the reproducibility tests [256]. In a structured light sensing-based defect detection, the relationship between the number of repetitions and the measurement uncertainty was analyzed by providing the uncertain range in defect size estimation. The calculated uncertainty is considered the best estimate of the correction with the measurement error, and the low uncertainty value illustrates the high reliability of the measurement system [257]. Besides, to track the relationship between the number of repetitions and the uncertainty, the statistical testing scheme, Analysis of Variance (ANOVA), was applied. In this process, the within-group mean square serves as an estimate of the variance of the measured quantity when measured under identical conditions, which provides a numerical estimate of how much each factor contributes to the overall uncertainty [258]. ANOVA is a powerful tool when dealing with multiple factors or sources of variation, which has been widely applied in NDE-based measurement uncertainty analysis [259–261]. The computational procedure of ANOVA can be illustrated in Fig.5.2. Noted that S is determined from the standard deviation group candidate with possible combinations of measurements $\binom{n}{m}$, which is decided based on different cases for obtaining the best selection for uncertainty estimation.

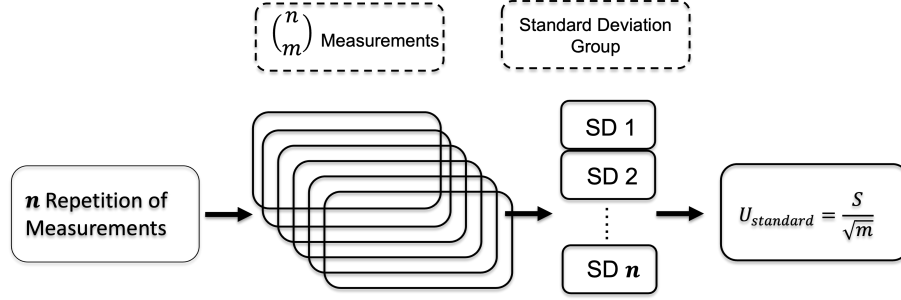


Figure 5.2 Block diagram of ANOVA procedure.

Generally, GUM-based measurement uncertainty analysis provides a good basis for ensuring the inspection quality. It is important to consider the uncertainty factors to choose the appropriate measurement uncertainty analysis model to ensure that the model accurately reflects the associated uncertainties and aligns with the goals and standards of specific NDE applications. Given the above measurement uncertainty scheme, it is then applied to two practical NDE-based cases for estimating the corresponding measurement uncertainty, which will be addressed in the next sections.

5.3 Magnetic Barkhausen Noise-based Material Fatigue Detection

5.3.1 Background

Martensitic-grade stainless steel is usually used to manufacture steam turbine blades in power plants. Due to cyclic loading applied on these blades, the stainless-steel material will get fatigued initiating cracks and ultimately leading to complete fracture. The failure of these turbine blades can contribute to expensive plant failures and safety concerns [262]. The fatigue crack formation process is a complex and dynamic sequence of events that involves the initiation, growth, and interaction of micro-cracks within the material[263]. An understanding of the fatigue crack formation process at the microstructure level is crucial for materials scientists, engineers, and researchers. It allows them to develop strategies to mitigate and manage fatigue-related issues, such as improving material properties, enhancing structural design, and implementing maintenance and inspection protocols to prevent failure due to fatigue cracking.

Magnetic Barkhausen Noise (MBN) is sensitive to microstructure changes in ferromagnetic materials. It has great potential in measuring surface residual stress and other microstructural parameters [264]. Specifically, for example, an MBN-based model is applied in characterizing

carbon content influence in plain steels [265]. The results reveal that the separation (gap) between the two peaks of the MBN envelop decreases with the increase in carbon content. Also, magnetic properties of ferrite-martensite dual-phase steels are evaluated using MBN signal, revealing that Barkhausen noise profiles are correlated with ferrite grain size and different percentages of martensite [266]. Furthermore, MBN is able to detect fatigue cracks from mild steel samples, where the influence of fatigue on fractal characteristics of Barkhausen noise is potential in the detection of fatigue crack initiation [267]. Therefore, MBN emerges as a promising technique for early-stage fatigue detection in steel samples.

Generally, MBN is a complex magnetoelectric phenomenon, and it is associated with a high degree of uncertainty. One criticism directed at MBN is its limited repeatability and stability. The primary reasons behind these issues include measurement instabilities arising from varying experimental conditions and the absence of a standardized normalization principle [258, 268]. Sources of uncertainty in MBN analysis can be categorized into two main categories: data-related and calculation model-related, as discussed in [180]. In the context of data-related uncertainties, three primary aspects are considered: the first is material physics properties, which denotes the variability in the material's physical properties, such as its composition, microstructure, and mechanical properties. The second is the data-generating method including the instrumentation and measurement setup. Besides, the presence of experimental noise and interference during data acquisition introduces uncertainty, like electromagnetic interference, sensor noise, and other environmental factors. To address these data-related uncertainties, it is crucial to employ a stable sensor that ensures consistent magnetization conditions and to implement accurate processing techniques for the obtained MBN signal. Additionally, the precision and reliability of the measurement system play a significant role in assessing the uncertainties associated with the collected data. A well-calibrated and high-precision measurement system contributes to a more accurate evaluation of the uncertainties, ultimately enhancing the quality and reliability of MBN analysis.

The primary objective of this work is to explore the applicability of MBN in detecting fatigue damage while diminishing and rectifying uncertainties that may arise during the measurement

process. The features derived from the processing of raw MBN signals are carefully selected from both the time and frequency domains, considering their relationship with microstructural material properties. Subsequently, Principal Component Analysis (PCA) is applied to extract advanced features. In addition, a Probabilistic Neural Network (PNN) is utilized for sample classification based on the percentage of remaining fatigue life, a factor expected to be closely linked to the specimen's fatigue life and with the potential for early fatigue onset prediction. Moreover, comprehensive uncertainty analysis is conducted to address and mitigate any noise present in the acquired MBN data. To further enhance measurement precision, a statistical technique, Analysis of Variance (ANOVA), is employed to assess repeatability.

5.3.2 Experimental Setup

The experimental setup can be described as follows: First, a wave generator produces a low-frequency sinusoidal excitation signal, which is then amplified using a power amplifier. This amplified signal is supplied to an excitation coil, generating changes in magnetization within the martensitic steel sample. A pick-up coil captures the resulting MBN voltage signal from the sample and transmits it to an NI DAQ card. LabView processes this data to create the MBN data file, which is further analyzed in MATLAB to derive features related to the steel sample's fatigue life. Given the sensitive and noise-like nature of the MBN signal, an efficient sensor is essential for accurate measurements. The MBN sensor assembly includes a U-shaped magnetic core, pick-up coil, and excitation coils. The pick-up coil is positioned on the sample and attached to the magnetic core with a customized holder. Optimization of the pick-up coil's turns and analysis of MBN signals in both the time and frequency domains are critical for obtaining the best pick-up signal results, as described in [269]. .

The test inventory consists of 36 samples, each representing various stages of fatigue, characterized by different loading cycles falling within the range of 0 to 2,000,000 cycles. Subsequently, EPRI conducts fatigue testing on these samples, ultimately leading to their failure, during which the lifespan of each sample is determined. It's worth noting that, out of the 36 samples, only 26 have information available regarding their total loading cycles at the point of failure. Based on the

loading cycles observed during the NDE test and the total loading cycles at the point of failure, a metric known as "percentage fatigue life" is introduced. This metric is further elucidated below:

$$\text{Fatigue Life} = \frac{\text{Total Loading Cycles} - \text{Cycles at NDE test}}{\text{Total Loading Cycles}} \quad (5.6)$$

The sample is re-categorized according to the Percentage Fatigue Life. The categorization is compared between the Cycles at NDE Test (without ground truth) and percentage Fatigue Life (ground truth), shown in Table 5.1. The samples are categorized into No-Fatigue, Mid-Fatigue, High-Fatigue and Cracked.

Table 5.1 Category for Ground truth and No Ground truth

Loading Cycles at NDE test (No Ground truth)	
Number of Sample	Sample Category
3	No-Fatigue (Untested)
19	Mid-Fatigue (150,000-750,000)
4	High-Fatigue (900,000-2000,000)
4	Cracked
Percentage Fatigue Life (Ground truth)	
Number of Sample	Sample Category
10	Low Fatigue (75% to 100%)
12	Mid-Fatigue (40% to 75%)
4	High-Fatigue (0% to 40%)
4	Cracked (0%)

MBN data are collected at 19 different points, where 18 of them on the 3×6 grid in the region of interest, representing the fatigued area. The last data point is in the un-fatigued area, which serves as the reference point. For each selected feature of a given sample, the normalization process of each feature value is through averaging values of 18 scanning points and divided by the corresponding sample's reference point feature value, which is described through Eqn.5.7:

$$N_{kl} = \frac{\sum_{i=1}^{18} S_{ikl}}{18 \cdot R_{kl}} \quad (5.7)$$

where i is each scanning point at fatigue area, ranges from 0 to 18; k denotes the sample number, ranges from 0 to 30 (including the cracked sample); l is the corresponding feature. S_{ikl} represents each sample point's feature, while R_{kl} is the corresponding reference point's feature.

5.3.3 Defect Classification in MBN

5.3.3.1 Defect Classification with Probabilistic Neural Network

Based on the obtained MBN signals from each sample, various features are evaluated in the time domain and frequency domains. In the time domain, the shape of MBN profile shows systematic and distinct variation in the magnetization process with respect to different microstructures[270]. Therefore, features are selected based on the information provided by the MBN signal profile. Since the signal profile exhibits symmetry around the x-axis, the upper part of each periodic MBN signal at every scanning point is represented using a combination of two Gaussian distributions. This modeling approach has been demonstrated to be effective in extracting features related to material microstructures, including metrics like the volume fractions of ferrite and pearlite, as well as the average grain size [270] [269]. The combination of two Gaussian fitted curves (depicted in blue and green) produces the resultant curve (shown in red), which characterizes the upper portion of the Magnetic Barkhausen Noise (MBN) signal profile.

As shown in Fig.5.3. (a), the selected time domain features are signal peak, the full width at half maximum (FWHM) of the upper MBN's profile, and the peak difference among the gaussian fitting curves. Also, fast Fourier transform (FFT) is applied to time domain MBN signals to obtain the corresponding frequency spectrum, where the frequency spectrum, the maximum spectrum amplitude noted as AMP, and the frequency for maximum spectrum amplitude noted as POS, are selected as features, which are shown in Fig.5.3. (b). Besides, the energy of MBN is associated with the misorientation angle of grain boundaries, affecting the alignment of magnetic domains along these boundaries. As per Parseval's theorem, the energy of a signal remains constant in both the time and frequency domains. Consequently, the energy of an MBN signal is calculated by summing across all the frequency components of the signal's spectral energy density.

Principal Component Analysis (PCA) is a method employed to accentuate variations and reveal

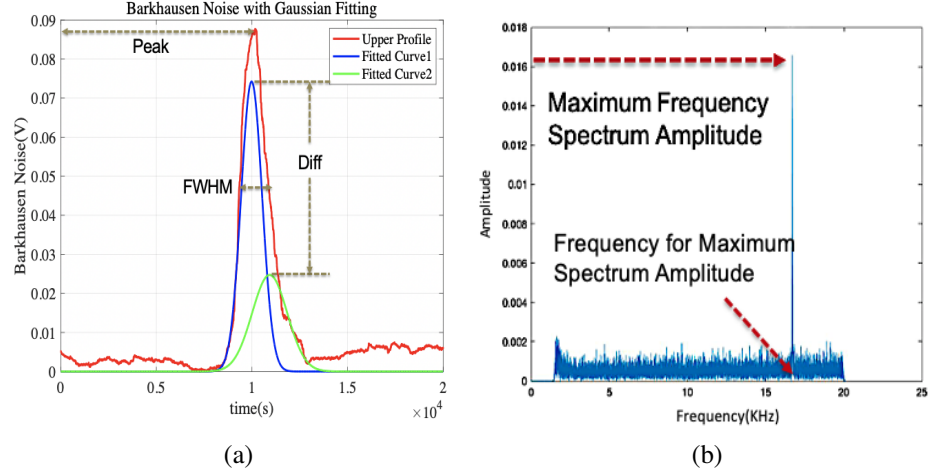


Figure 5.3 Illustration of extracted feature a) Time domain features: Peak, FWHM, and Diff; b) Frequency domain features: AMP and POS.

significant patterns within a dataset. In this case, where six distinct features have been introduced, PCA serves as a valuable tool for gaining a more insightful understanding of these features. PCA identifies higher-order features, known as principal components, which provide a more refined representation of the earlier findings. Additionally, PCA excels at reducing dimensionality, enhancing interpretability, and minimizing the loss of information [271]. The fundamental steps in the calculation procedure of Principal Component Analysis (PCA) are as follows: Begin by computing the eigenvalues of the covariance matrix of the data, which represents the correlations among both the original and new features. The next step involves multiplying the covariance matrix by the original feature space, resulting in the creation of a lower-dimensional feature space. This newly calculated feature space comprises the principal components, which serve as representations of the new features.

A probabilistic neural network (PNN) is a feedforward network formula for probability density estimation. Unlike other traditional back propagation neural networks, PNN shows excellent and efficient performance in dealing with limited datasets and has been widely used in NDE problems. In an Eddy Current-based defect detection case, the proposed PNN is able to identify the untrained eddy current patterns with 100% classification accuracy [272]. In [273], the proposed Kernel PCA followed with PNN algorithm is applied in acoustic data of rolling bearing type rotating machine

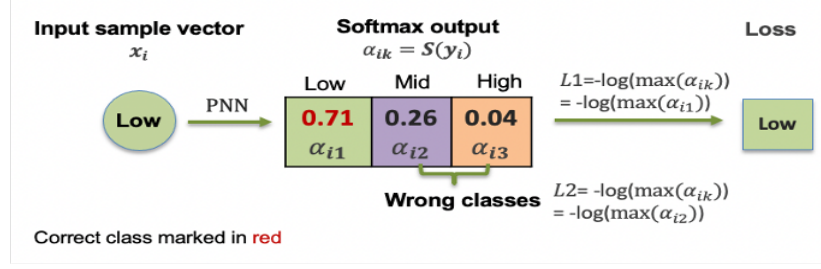


Figure 5.4 Process to obtain CI.

to classify the machine state.

In this case, the obtained MBN features are modeled by the PNN to classify the low-fatigue, mid-fatigue, and high-fatigue samples. During the network training process, the multi-Gaussian function is centered based on the associated input feature vector in each class, and total $k = 3$ classes are defined; for the output layer, the summed Gaussian output is defined. Then, for a test feature vector x_i , where i is the sample ID, ranges from 1 to 26, all Gaussian functional values at the hidden nodes will be computed and then be passed to the single output node for each group of hidden nodes. All of the inputs will be summed and multiplied by a weighted function. With the applied Softmax function, the corresponding output vector α_{ik} is obtained, which provides the probability belonging to each category. The maximum value within the output vector decides the classified result. Among all the test samples, the total classification accuracy is defined as $\text{Accuracy} = \frac{Y}{P}$, where Y is the corrected classified sample number and P is the total sample number. In order to describe the classification results with confidence evaluation, the Negative log-likelihood function is applied to the largest probabilities, defining the final output class $L = -\log(\max(\alpha_{ik}))$. For the accurately classified sample, Classification Index (CI), is involved to evaluate the classification confidence:

$$\text{CI} = \text{abs}(L1 - L2) \quad (5.8)$$

Specifically, $L1 = -\log((\max(\alpha_i)))$, is final output; $L2 = -\log((\max(\alpha_i)))$, is calculated among the other two wrong classes; which are illustrated in the Fig.5.4. A large CI value means classification performance is of higher confidence at the correct class and vice-versa.

Furthermore, the accuracy term can be misleading in unbalanced datasets, since it only focuses on the correctly identified cases. Considering in this early fatigue stage detection, not all errors are semantically equal, the evaluation of the misclassification risks is important. The confusion matrix is a common performance measurement technique for classification, which could provide a better understanding of mistakes made by the classifier and also mistake types. Other than accuracy, usually in binary classification, a confusion matrix could measure Recall (proportion of actual positives is correctly classified), Precision (proportion of predicted positives is truly positive), etc. Also, the F1-score takes a harmonic mean of Recall and Precision to punish extreme values, which is a better indicator for evaluating unbalanced class distribution, shown in Eq.(??). The higher the F1-score is, the classification procedure is of higher reliability.

In the context of unbalanced datasets, the term "accuracy" can be misleading since it primarily focuses on correctly identified cases. In this early fatigue stage detection, not all errors have the same significance. Therefore, it is important to assess the risks associated with misclassifications. To achieve this, the confusion matrix serves as a widely used performance measurement technique in classification. It offers a more comprehensive understanding of the errors made by the classifier, including the types of mistakes. In binary classification scenarios, a confusion matrix could measure Recall (proportion of actual positives is correctly classified), Precision (proportion of predicted positives is truly positive), etc. Additionally, the F1-score is a valuable metric that takes the harmonic mean of Recall and Precision. This helps in penalizing extreme values and serves as a more reliable indicator, particularly for assessing unbalanced class distributions. A higher F1 score signifies a classification procedure of higher reliability and effectiveness.

5.3.3.2 Comparison Results

Based on the percentage fatigue life, the normalized feature value for each sample N_{kl} is shown in Fig. 5.5. For each normalized feature, we use a box-and-whisker plot to describe sample distribution based on the ground truth category among 26 samples. Also, the average normalized feature of each category is presented. Results show that for all six MBN features, there exist monotonical trends in mean with percentage fatigue life. Except for POS plot, the other five

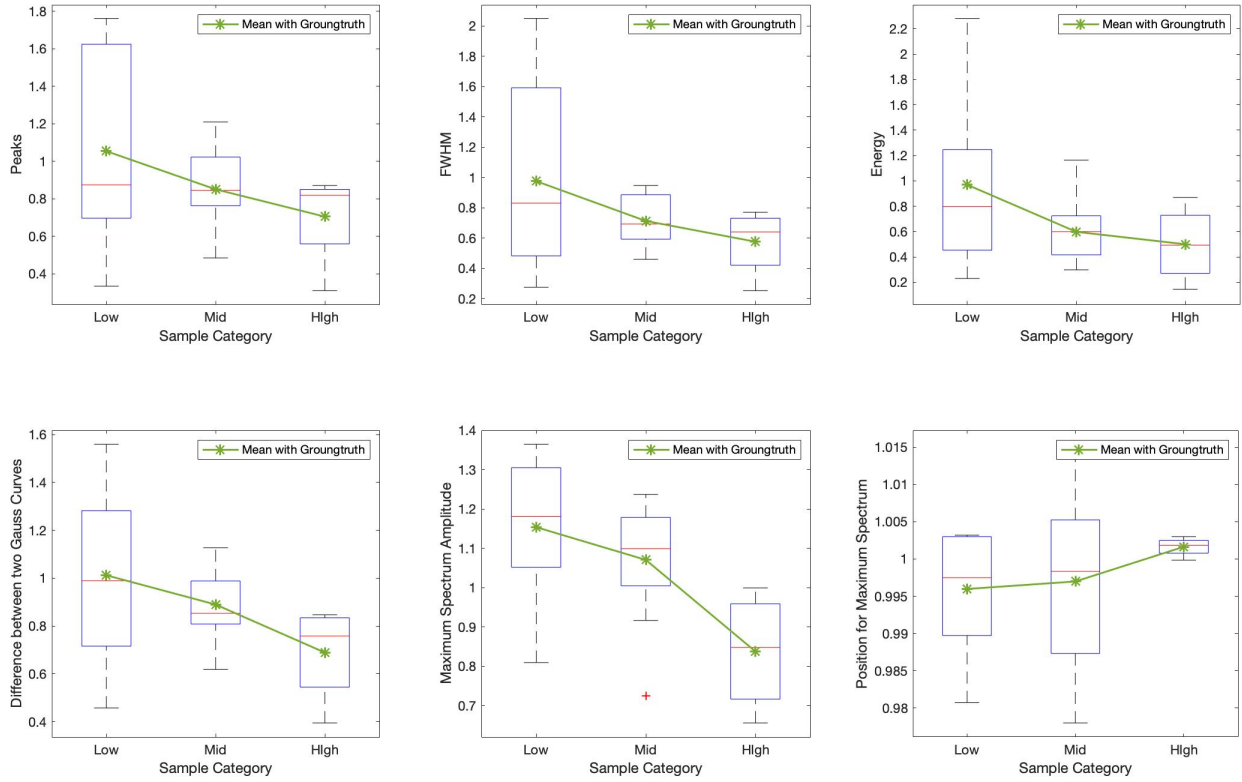


Figure 5.5 Comparison Results of Six features.

features are positively correlated with percentage fatigue life.

In the context of PCA, it's essential to address the potential impact of useless features on the results. To account for this, we employ two distinct original feature sets and conduct PCA separately for each, allowing for a comparison of the results. In the first scenario, the original features include signal peak, Full Width at Half Maximum (FWHM), differences between two peak curves, and signal energy. These features have previously been established to exhibit a strong relationship with microstructures. Through PCA, two new features, PC1 and PC2, are generated to represent the original feature space. In the second scenario, all six features are retained as the original feature set, which are then reduced to four principal components (PC1, PC2, PC3, and PC4). In both cases, the new principal components collectively account for 98% of the total variance. To construct new feature sets, we consider the first two components (PC1 and PC2) in Case 1 and the first three components (PC1, PC2, and PC3) in Case 2, as these components account for a higher proportion

of the variance. These new feature sets are defined and subsequently assessed using a PNN.

To enhance the reliability of the results, the PNN is trained and tested across 30 iterations, and the final accuracy is determined as the average of the values obtained from each iteration. In each train-test scheme, a random sample from each category is selected to ensure that every category is represented in the test sample. The dataset is divided into three categories, and balanced cross-validation is applied for the train-test split. For samples that are correctly classified, the CI is calculated. The final CI is derived as the average of all the obtained CIs. Given the necessity for three classifiers in our case, the classification performance is evaluated by computing the arithmetic mean of the per-class F1-scores, resulting in what is known as the macro-averaged F1-score. Based on the preceding discussions, three distinct input feature spaces for the Probabilistic Neural Network (PNN) are defined as follows:

1. Six normalized MBN features: Signal Peak, FWHM, Diff, Energy, AMP, POS;
2. PCA results(PC1 and PC2) for Case 1: Peak, FWHM, Diff, and Energy;
3. PCA results(PC1, PC2, and PC3) for Case 2: six MBN features;

The corresponding PNN classification results are presented in Table 5.2. Among all three cases, after PCA for all six features, higher order MBN features have 79% classification accuracy performance along with a high classification index, which proves that the extracted principal components (PC1, PC2, and PC3) in this scenario, are the most representative features to indicate remaining fatigue life of martensitic samples. Also, the macro-F1 score in the third input space is relatively higher than others, which shows that the selected classifier is of higher precision and robustness.

For these extracted principal components, Fig. 5.6 shows the averaged feature value's scatter plots based on the category of loading cycles at NDE test (No Ground truth) and that of percentage fatigue life (Ground truth). The results show that, compared with simply using the loading cycles as the classification criterion, the extracted principal components show a monotonic trend with ground truth information. Considering all the previous cases, though in Case 2, where the variance

Table 5.2 PNN Comparison Results

	Six features	PCA results for Case 1	PCA results for Case 2
PNN Accuracy	66.67%	66.67%	77.89%
CI	1.5896	1.0769	8.9922
Macro-F1 Score	0.4032	0.4859	0.5951

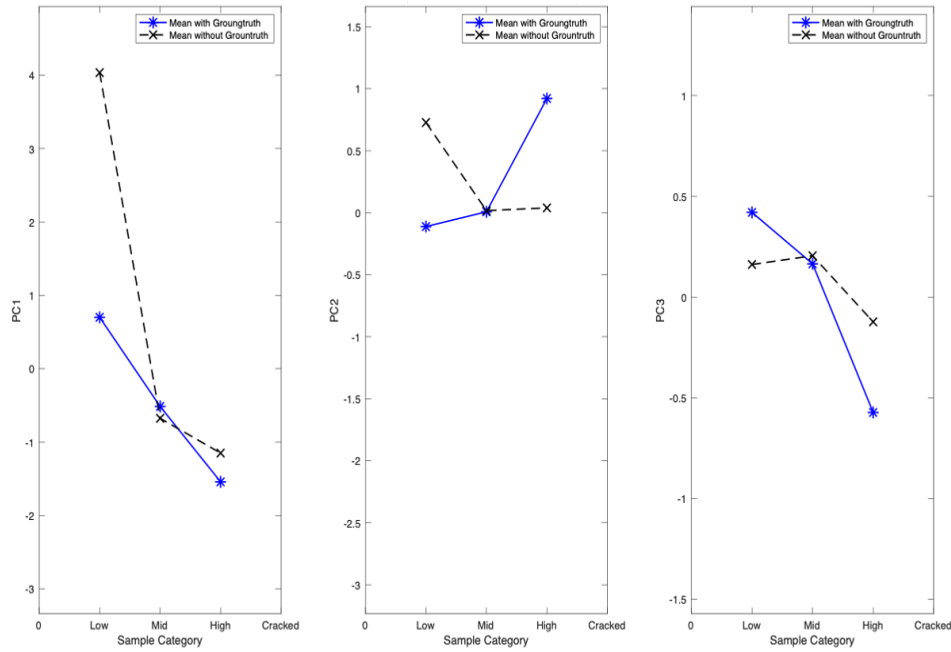


Figure 5.6 Averaged PCA results comparison among ground truth and no Ground truth.

distribution spreads out for more components with some information loss, the extracted new MBN feature space is better in categorizing samples based on the percentage fatigue life.

5.3.4 Uncertainty Analysis on MBN

The presence of uncertainties within the system can significantly impact prediction capabilities, underscoring the importance of uncertainty assessment in the MBN system to gauge the reliability of measurement results. Currently, the unavailability of complete material fatigue information and insufficient data on failures make it challenging to precisely identify the sources of uncertainty and quantitatively estimate them. To address this, efforts are concentrated on building a stabilized data collection system and obtaining reliable output to reduce the noises in the experiment and signal. These efforts encompass various aspects, such as stabilization and standardization of sensor

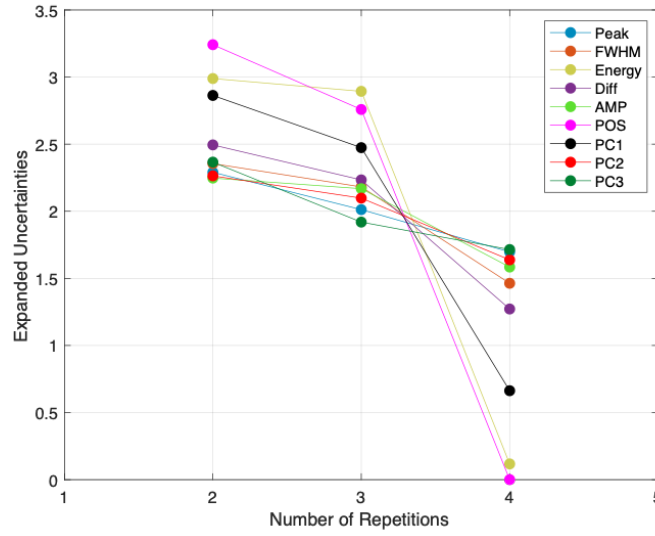


Figure 5.7 Comparison results of expanded uncertainty.

lift-off, circuit connections, and scanning areas; Multiple measurements are conducted, comprising four repetitions. Over ten half-periods of the MBN signal are averaged at each measurement point, followed by the application of Root Mean Square (RMS) to present signal intensity with normalized features.

In this study, by ignoring the Type B uncertainty, the measurement standard uncertainty is addressed in terms of repeatability only, through the Equ.5.2. Specifically, four repetitions ($n = 4$) are made for uncertainty analysis. The uncertainty associated with each feature is presented through the corresponding expanded uncertainty. The standard deviation S for every combination was determined by taking the maximum of those four chosen m standard deviations. The standard uncertainties were calculated for all selected features.

5.3.4.1 Analysis of Variance-based Uncertainty Quantification Results

Based on the previous discussion, PC1, PC2, and PC3 in PCA Case 2 as well as the original six features have a monotonic relationship to fatigue information. Therefore, all nine features are evaluated in this statistical uncertainty analysis. The comparison results of each feature's expanded uncertainty are illustrated in Fig. 5.7.

The generated plot illustrates a correlation between uncertainties and the number of repetitions

for each selected feature. As anticipated, the uncertainties exhibited a decreasing trend as the number of repetitions increased. In general, the uncertainties were observed to be quite low, suggesting that repeated measurements yield consistent results. Notably, features such as signal energy, POS, and PC1 initially displayed higher uncertainties. However, with the increasing number of repetitions, these uncertainties decreased rapidly. This implies that conducting additional repeated experiments could significantly reduce uncertainties among these features. It's worth noting that in this particular case, only four repetitions were performed. The robustness and comprehensiveness of this uncertainty analysis could be further enhanced with additional experiments.

5.4 Structure Light Sensing based Defect Reconstruction

5.4.1 Background

Plastic pipes have become the prevalent choice for the distribution of natural gas since the early 1970s, and this trend continued to be the primary material used as of 2017, as documented in [274]. However, it's important to note that the rigidity and strength of plastic pipes do not match those of steel pipes. This disparity renders plastic pipes vulnerable to damage caused by various factors, including improper excavation or installation, as well as excessive stresses within the pipe's operational environment [275]. Such damages can result in leaks and, in more severe cases, gas pipe explosions, which pose substantial risks. Consequently, the detection and identification of material degradation within the pipe walls hold significant importance. Various nondestructive evaluation-based methods have been developed and verified for inspecting plastic pipes. These methods encompass techniques like ultrasonic testing as highlighted in [276, 277], microwave testing methods as discussed in [278, 279], infrared thermography-based approaches detailed in [280, 281], and camera-based visual inspection techniques presented in [282, 283].

Optical inspection represents one of the earliest NDE methods, initially involving visual inspection with the naked eye to identify potential defects in the examined object [284]. Over time, with the evolution of digital photography and advancements in camera manufacturing, the preference shifted toward digital cameras equipped with automated detection algorithms. Optical inspection techniques exhibit several advantages. They are less influenced by the type of material being

inspected, in contrast to conventional NDE methods. Additionally, they do not necessitate the use of a coupling medium and can be scaled down in size through careful engineering to achieve a compact form factor, making them suitable for insertion into confined spaces.

The traditional visual inspection method, which employs various types of cameras, has a long history and remains popular. However, it relies heavily on the operator's expertise and lacks the capability to quantitatively measure the depth of damage in plastic pipes. To address these limitations, the authors have developed a structured light (SL) sensor for inline inspection of gas pipes[285]. SL technology is gaining widespread acceptance due to its numerous advantages, which include robustness, high precision, and the ability to shrink sensors to very compact sizes [286]. The structured light pattern serves to endow the inspected surface with the necessary features for triangulation, enabling SL systems to inspect surfaces even when unique surface characteristics are absent. This capability is especially valuable, as it facilitates the inspection of smooth and featureless surfaces, such as the walls of plastic gas pipelines.

The original prototype of the SL sensor is suitable for performing the inspection only when the sensor is moving linearly with no change in orientation. However, in a real-world industrial application, such stringent requirements are often not feasible because the sensor is typically fitted on a moving platform (e.g., a robot), which will not be able to maintain such a strict pose while traversing the length of the pipeline. The proposed registration algorithm addresses this design gap and allows the SL sensing system to dynamically correct for changes in pose, resulting in a stabilized and accurate 3D reconstruction of wall profiles. The designed SL sensor is attached to a scanning platform that moves along a pipeline during the internal 3D inspection. Every single frame from the sensor produces data for a sparse reconstruction of the pipe surface with a density that is dependent on the number of projected rings. In an ideal situation, the sensor's axis is aligned with the main axis of the pipe and always points in the direction of platform movement, which is defined as the z -axis. Therefore, the reconstructed 3D frames can be stacked sequentially by only adding a displacement in the z -direction that is dependent on the scanner speed at the time of acquisition. Experimentally, this assumption is not practical because it is difficult to maintain the sensor moving

exactly at the center of the pipe along the forward direction. Also, the platform's moving speed is hard to keep constant due to multiple uncertainties such as imperfections in mounting the sensor on the robot, vibration from the movement of the robotic platform, and slippage of the robot wheels. Therefore, a holistic registration algorithm is required to estimate both the orientation of the sensor and its real-time position inside the pipe, to realize accurate 3D reconstruction.

Simultaneous localization and mapping (SLAM) algorithms, a popular global positioning method, allow the incremental creation of maps using data from sensors while estimating real-time positions [287, 288]. While various methods have been applied to reduce mapping errors in SLAM, camera-based mapping with inertial navigation systems (INS) often had issues with the accuracy and drift of these systems [289]. For accurate global positioning in pipeline detection, the cylindrical nature of pipes is utilized as the basis for SL sensor-based sensor localization [290]. Also, encoder data from the robot can provide accurate estimations on how deep inside the pipeline the robot is [291]. After the location in the pipeline is refined, the performance of the 3D reconstruction is related to the local positioning of the sensor as well. In this work, information from wheel odometry and IMU are incorporated to estimate the speed and orientation of the sensor in real-time, which could realize a more reliable local positioning. The data are then fed to a registration algorithm to provide an initial guess about the sensor orientation and position inside the pipe; following that a RANSAC-assisted [292] cylindrical fitting-based registration approach is followed to provide high efficacy 3D point cloud registration to stabilize sensor. Furthermore, an intensity-based threshold search method is employed to determine the reconstructed defect size. Finally, the uncertainties associated with structured light sensing are examined to quantify both the total reconstruction uncertainty and the estimated measurement uncertainty, demonstrating the measurement precision. The effectiveness of the proposed algorithms is validated through experimental results in pipeline inspection.

5.4.2 Design of Structured Light Sensing System

To demonstrate the capability of wheel odometry in enhancing the performance of the proposed algorithm's performance, the robot-integrated sensing system with three main components: SL

sensor, IMU, and employed robot, are illustrated in Fig. 5.8. The robotic system has the capability to be deployed in the 4-inch to 6-inch PVC pipe for real-time data collection.

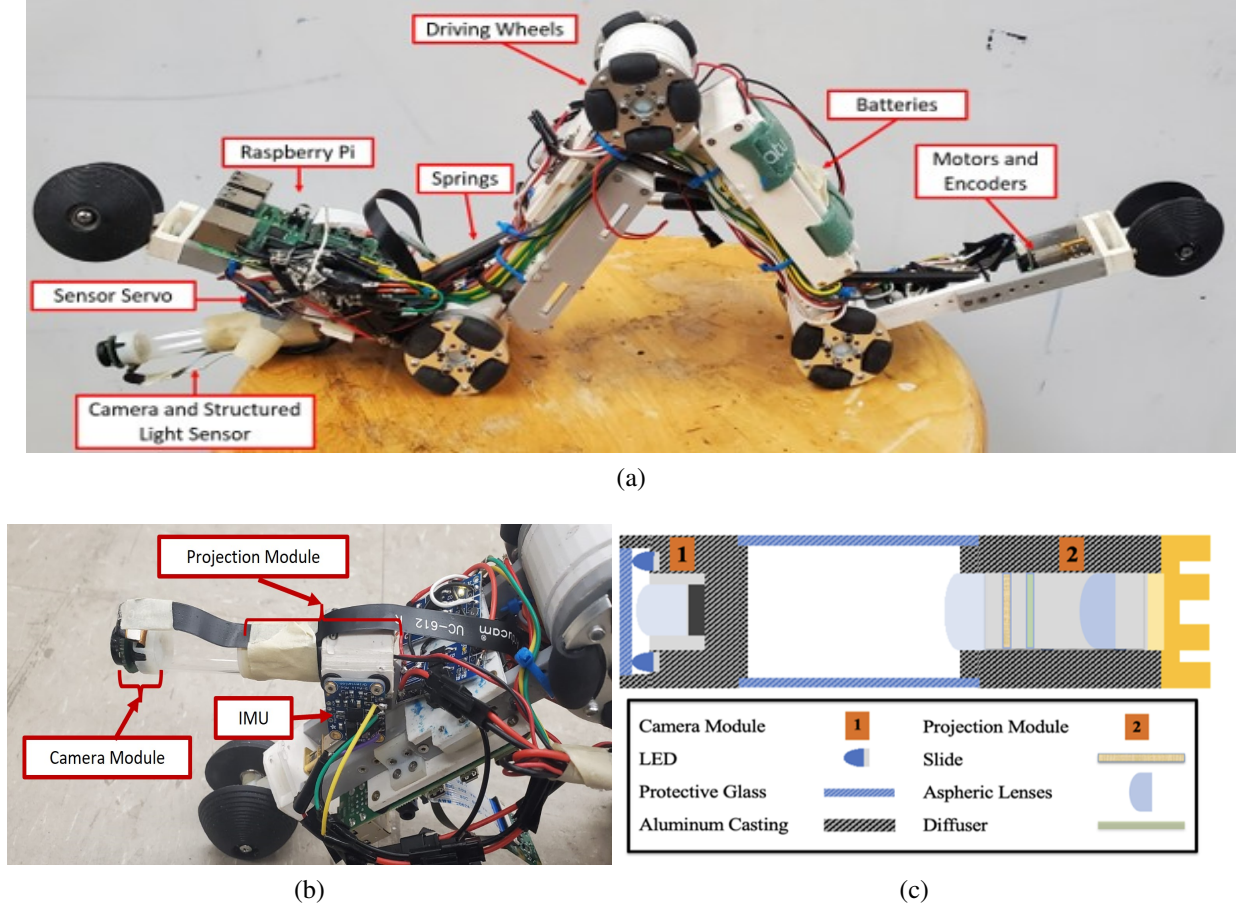


Figure 5.8 a) The robotic system with integrated SL sensor; b) Camera and SL sensor system; c) Schematic of the endoscopic SL sensor.

A structured light sensor consists of a projection module that projects a highly textured pattern and a camera that captures the deformations in the projected pattern [293]. The detailed description of the SL sensor design and fabrication is addressed in our previous work [285]. It consists of a camera module, a projector module, and a connected transparent glass tube for enabling the projection of the colored rings to the pipe walls. The projector module consists of a high-intensity light-emitting diode (LED), a collimation lens, a transparency slide, and a projection lens, which has been shown in Fig. 5.8. (c). The Complementary metal-oxide-semiconductor (CMOS) camera is used to monitor the pipe surface and capture deformations in the projected rings. The 3D

imaging reconstruction of the scanned object surface, as mentioned in [293], is the process of detecting, localizing, and matching the projected edges. In this process, the acquired image is converted to the polar domain to perform edge detection based on the predefined color coding of the slide pattern. With essential cleaning and filtering, the extracted edges of each acquired image will be reconstructed to a cylindrical shape in 3D domain, which provides a basis for point cloud registration between each data frame.

To find the geometric transformation, the registration algorithm depends on both inertial measurements and the matching of the common features in the fixed and moving frames. The main framework of the proposed algorithm for stabilization can be summarized in Figure 5.9 with two main interconnected tasks: local positioning and global positioning. In this scheme, the use of a synchronized acquisition framework is realized with real-time 3D data which are assisted by the IMU and wheel odometry data. Global positioning provides a vague pose (position and orientation) of the sensor inside the pipe by using wheel odometry and inertial measurements. Local positioning is then used to further improve the global position calculations especially when surface features exist. The data from the global positioning are fed to the registration algorithm to provide an initial guess about the sensor pose inside the pipe, and then 3D information is used to provide a more precise tuning. If defect features are found, the global position is updated and the data are registered; otherwise, the initial global position is used in addition to the constraints from the cylindrical 3D environment.

In this work, sensor characteristics are integrated into the 3D registration problem to improve the robustness of the fitting performance. The environment inside the pipe is described in Fig 5.10, where a structured light sensor is enclosed by a cylinder with a radius $\mathbf{R}_{Cyl}$ with an arbitrary orientation axis described by the unit vector A_{Cyl} for the SL sensor. In this environment, the camera is located at the origin ($C = (0, 0, 0)$) of the coordinate system and the camera is pointing along the z -axis. The projected ring is imaged by the camera to create a set of image points ($\mathbf{D}_C$) that can be represented by the camera ray ($\vec{w}$). The camera rays intersect with both the projected cone from the projector module and the surface of the bounding cylinder. Therefore, the intersection

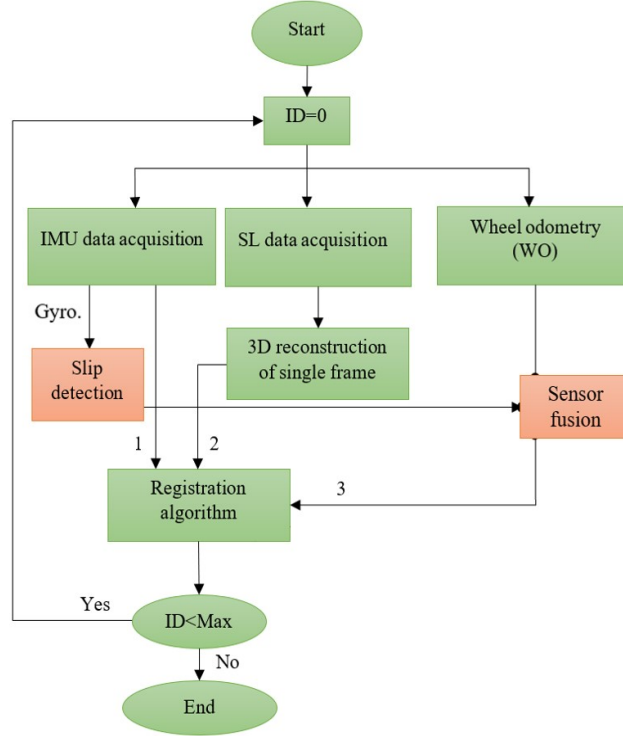


Figure 5.9 Proposed registration approach for sensor stabilization with data acquisition procedure.

points belong to both the cylinder and the cone surfaces. With known cylinder parameters, the intersection between the camera ray and the cylindrical surface can be easily calculated with the substitution of the ray equation in the cylinder equation. Therefore, the cylinder orientation can be calculated by minimizing the difference between $\mathbf{D}_C$ and $\mathbf{D}_{arb}$, which can be described by

$$(\phi_x, \phi_y, T_x, T_y) = \operatorname{argmin}(|(\mathbf{D}_{arb} - \mathbf{D}_C)|_{2^2}). \quad (5.9)$$

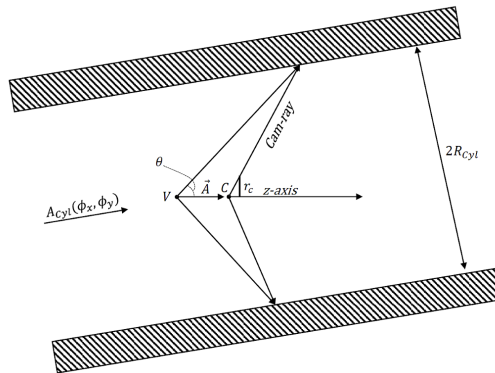


Figure 5.10 Triangulation of structured light sensor inside a pipe environment.

One of the error sources that affect the accuracy of the cylindrical fitting is the existence of artifacts on the pipe walls since the fitting process assumes an ideal cylindrical surface. The defect causes the fitting problem to be biased and results in an inaccurate estimation of the cylinder parameters; therefore, the problem is more prominent when having deep defects in the pipe wall [285]. To reduce the effect of wall defects, the defects are assumed to be outliers that need to be identified and removed from the fitting problem. In this case, random sample consensus (RANSAC) was applied. RANSAC is an iterative method that estimates the model parameters in the existence of outliers by separating them from inliers with repeated random sub-sampling [292]. Therefore, all the defects will be separated because they do not fit the cylindrical model that is assumed during the optimization process. The fitting process of simulation is presented in Figure 5.11, where the input frame to the fitting process with a cylinder diameter of 6 inches (76.2 mm) and a wall defect with a depth of 10.16 mm. Figure 5.11.(a) shows the isolated defect region with RANSAC, while the actual rotation angle is around z axis. The algorithm can successfully isolate the defect region from the rest of the cylindrical surface. After isolating the defect data, the cylindrical surface data are fitted and the rigid transformation parameters can be calculated.

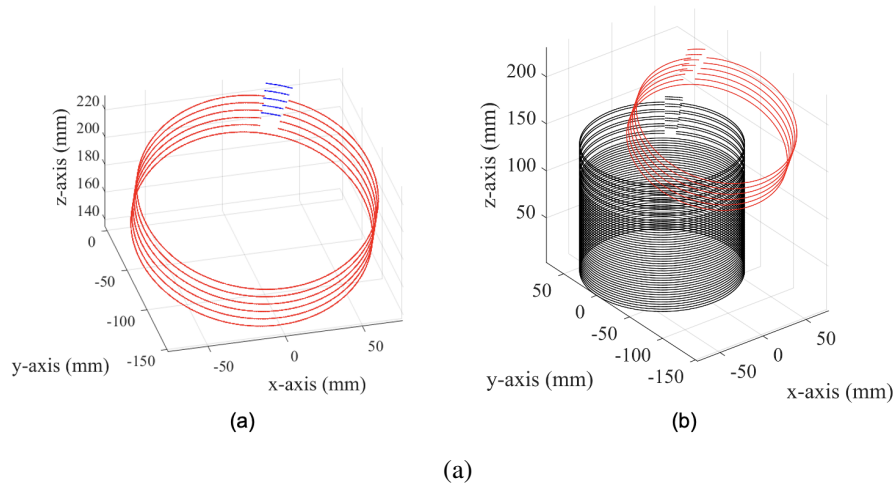


Figure 5.11 Alignment correction with cylindrical fitting, a) Moving frame with isolated defect by RANSAC in blue; b) Point clouds after alignment correction, Red: Moving frame, Black: Fixed frame.

The IMU cannot provide an absolute 3D position of the sensor but it can provide linear

acceleration, angular velocity, and orientation information, therefore these readings are integrated twice to estimate the instantaneous position of the sensor $r_{Acc}(t)$. The IMU combines readings from the magnetometer and gyroscope to estimate the orientation of the IMU in 3D space, which is used to estimate the rotation angle of the sensor inside the pipe. Through our testing, we found this type of data to be more reliable than the accelerator data but is still prone to deviation due to error accumulation. Therefore, it is only used as an initial point for the registration algorithm. It is worth noting that we are mainly interested in checking if our sensor has rotated around the main pipe axis (z -axis). Once the calibration parameters are known, we can start using the IMU data to monitor the sensor orientation, which is further integrated to minimize ellipsoid orientation and the gyroscope orientation plan for orientation correction.

Another set of sensors utilized for localization is the wheel odometers as an additional input to estimate the speed and position of the platform. The sensor position is estimated according to the number of wheel rotations at each frame and wheel diameter to obtain the estimated instantaneous position. The robot uses three pairs of wheels, with each wheel connected to a dedicated encoder. 3D printed material and the necessary electronics are equipped in the robot for operation. It is powered by two sets of 14.4v LiPo batteries attached to the robot. These batteries provide the necessary power to the motors, structured light projectors, and other electronics, enabling them to operate untethered to an external power source. In this work, the median distance between the frames of the three encoders is used as the reference distance estimate for the entire robot. Utilizing the median can reduce effects from wheel slippage producing an artificial distance increase or from motor stalling decreasing the distance.

5.4.3 Experimental Performance Evaluation

To demonstrate the IMU-assisted robotic sensing system, experiments were performed in a 6-inch PVC pipe with two defects, The first scanned segment is from point A to B, and the second is from point B to C as shown in Fig. 5.12. Both defects have the same dimensions of 70mm length, 35mm width, and 6mm height. The sensor is attached to a gantry to traverse the pipe. The inspection process starts at point A when reaching point B, the sensor is rotated, and the inspection

is continued till point C. This inspection scenario simulates the sensor rotation during the inspection and with off-center sensor misalignment. Fig. 5.13 shows example structured light image frames of the sensor rotation between points A to B and B to C.

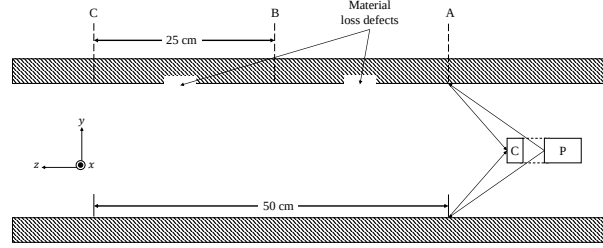


Figure 5.12 Schematic of the test pipe.

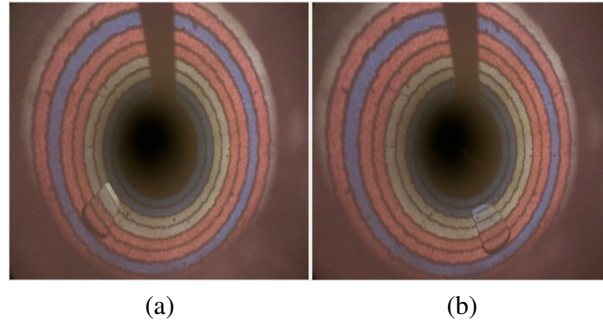


Figure 5.13 Example image frame illustration a) Point A to Point B; b) Point B to Point C.

5.4.3.1 Comparison of 3D reconstruction methods

To evaluate the performance of the proposed feature-based registration algorithm, we compare it to two other previous works. The first one is an ellipsoid fitting-based point cloud registration algorithm, which was developed previously by the authors [285]. In this method, an ellipsoid is used to fit the cylindrical surface to handle pipes with oval shapes and errors from the sensor calibration and then estimate both the orientation of the sensor and its position inside the pipe for each acquired frame with the following alignment correction. To register multiple frames, the corrected data are stacked by adding a constant displacement in the z -direction for each acquired frame, while keeping the z -direction displacement constant by having a fixed scanning speed.

Another method for comparison is the Iterative Closest Point (ICP) algorithm, a well-established registration technique often used to align 2D or 3D surfaces obtained from different scans and for

localizing robots in order to optimize path planning [294]. The ICP algorithm is widely employed in various applications, including the development of 3D models and the construction of 3D world maps for SLAM systems. Its primary function is to determine the transformation between a point cloud and a reference point cloud by minimizing the squared error between corresponding data points [295–297]. In this approach, the initial frame serves as the reference for establishing the initial transformation estimation, often involving the fitting of a plane. Subsequent frames utilize a point-to-plane distance metric minimization technique to align each source point cloud with the combined estimated rotation and translation. This aligned frame is then stacked in the z-direction to reconstruct the pipeline structure. The ICP algorithm’s versatility and capability to align data from various sources make it a valuable tool in pipeline inspection and other fields.

The comparison results are presented in Fig. 5.14. From the top view, it can be seen that the proposed registration algorithm retrieves a better pipe shape with a clearer and smoother boundary compared to the other two methods. Also, the marked defect area in the cylindrical-based method is more distinct and solid which is beneficial for defect isolation. Besides, the reconstructed 3D profile of the ICP-based method has a large misalignment after the sensor rotation, while the ellipsoid-based and the proposed cylindrical fitting-based method could fully reconstruct the 3D profile of the inspected pipe section. The main criterion of the registration algorithm is the ability to reconstruct the complete pipeline structure with less noise. Specifically, each data frame should be aligned vertically to build a straight and clear pipe surface. Therefore, we applied a plane to fit each 3D cloud point and extracted the normal vector $\vec{n}_i$ to obtain the directional information of each frame i . Since the pipe is a standard cylinder, projecting the 3D data frame onto the XY plane should theoretically be a circle. The projected points are then fitted to a circle to obtain the estimated center location $\mathbf{O}_{ci}$ and radius $\mathbf{R}_{ai}$. The center is able to reflect the location of each frame, which is highly related to the registration theory of each method. For a well-registered model, the differences in estimated centers and in directional vectors should be small to ensure height alignment between frames. Also, in the horizontal direction, the estimated circle should approximate the actual pipe size, so the estimated radius is a good criterion for evaluating registration performance, which is

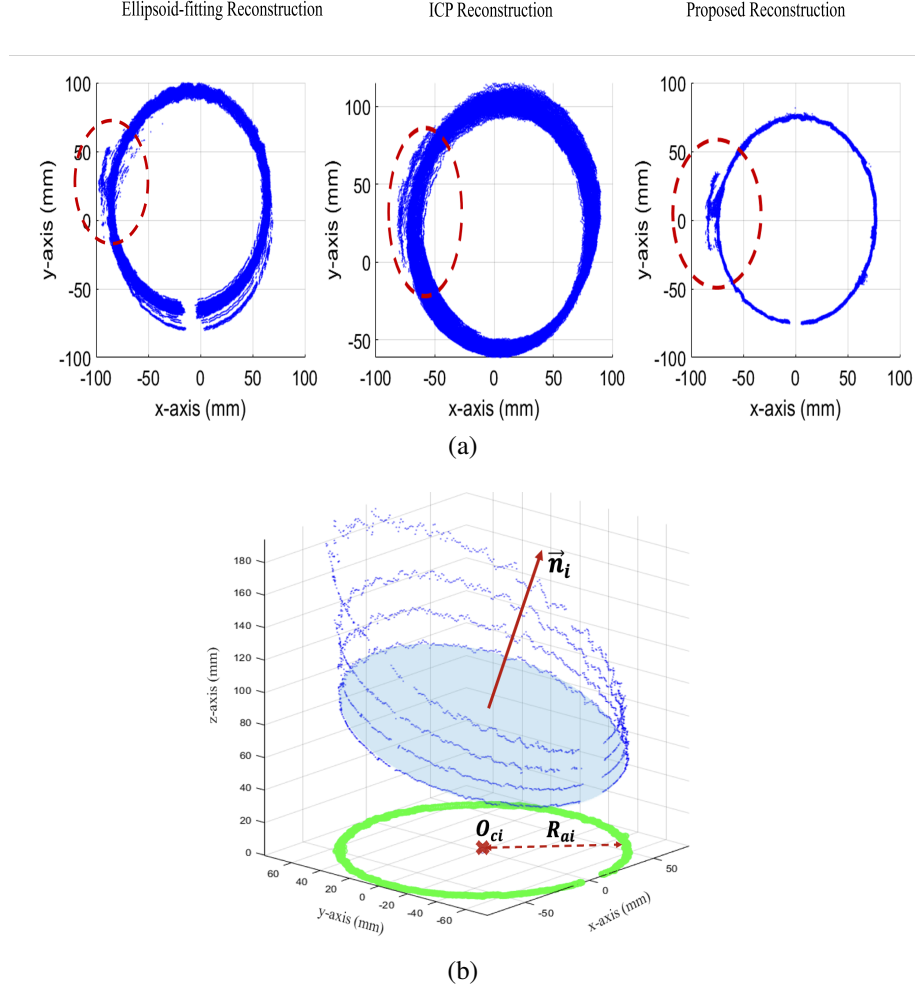


Figure 5.14 Reconstruction performance comparison among Ellipsoid based (left), ICP-based (middle) and proposed Cylindrical (right) registration algorithm: a) Top view: reconstructed defect area are marked by the red dotted circle; b) Performance evaluation parameters from one single frame.

interpreted as Closeness, and defined as:

$$Closeness = \frac{|\bar{R}_a - R_{GT}|}{R_{GT}} * 100\%. \quad (5.10)$$

where $\bar{R}_a$ is the average of the total estimated R_{ai} in each method. Therefore, we extracted the above shape-based parameters from the reconstructed pipe to quantitatively evaluate the reconstruction of those three registration techniques, which are illustrated in Fig.5.14.b. In this comparison work, total variances of normal vectors center location, and radius closeness are obtained from each registration technique, which is shown in Table 5.3.

The results indicate that all three methods effectively align the frame consistently with minimal

Table 5.3 Reconstruction Quality Evaluation for Registration Techniques

	Normal Variance(mm)	Center Variance(mm)	Radius Closeness
Ellipsoid	0.001	1.507	1.71%
ICP	0.002	9.112	1.55%
Cylindrical	0.001	0.025	0.95%

variation in the estimated orientation. Notably, the cylindrical-based algorithm demonstrates much greater reliability in aligning the frame vertically, as evidenced by the low variance at the center. Regarding the estimated pipe diameter, all methods managed to determine it with only minor differences, with the largest error being 1.71%. Notably, the proposed method exhibited the highest level of accuracy in this regard. In summary, the proposed method has demonstrated its robustness and dependability as an alignment correction technique. While ICP registration offers theoretically high accuracy, its performance and efficiency are constrained by the need for a precise initial value, minimal transformation between the two point clouds, and limited occlusion. This constraint is particularly pronounced in scenarios involving significant misalignment between point clouds, especially during rotations, which result in substantial misalignments and shifts in the frame baseline. Additionally, traditional ICP registration relies solely on geometry, color, or meshes and may struggle to perfectly reconstruct the defect area, particularly when the initial point cloud lacks comprehensive defect information. In contrast, the proposed algorithm treats the defect as a distinctive feature, enabling the reconstruction of both the pipe and the defect with higher reliability and resistance to baseline shifts. Considering the quality of the reconstructed pipe shapes and defects, the proposed cylindrical-based 3D registration algorithm outperforms current state-of-the-art methods.

All of the registration methods listed failed to correct the rotation of the robotic platform around the pipe's main axis. Therefore, through the use of inertial measurements from the IMU data, the pose was corrected. IMU data are acquired in real-time with the camera data and then used to correct the data alignments according to the procedure described in the background section. To illustrate the sensitivity and efficacy of the IMU, experiments were performed at different rotation angles (8deg) from point B to point C. The 3D reconstructed profile after incorporating the IMU

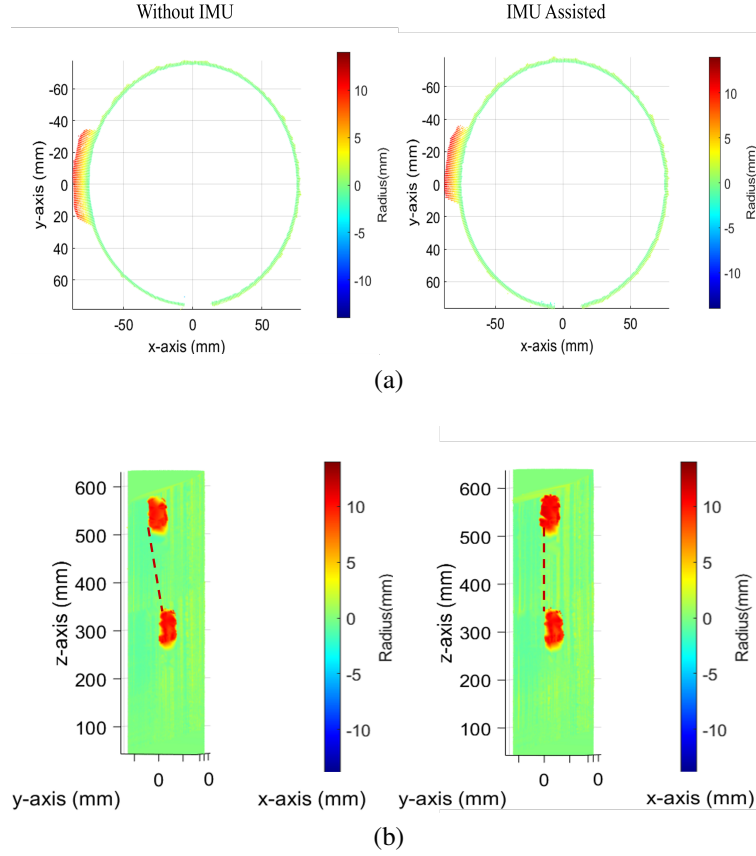


Figure 5.15 Reconstruction performance without IMU (left) and with IMU (right). Rotation angle: 8 degree: a) Top view; b) Side view.

data with the proposed cylindrical-based registration is shown in Fig.5.15. From the top and side views, we notice that after integrating the IMU information, the position of the second defect is corrected to a similar vertical orientation to the first defect in both cases. Specifically, in Fig. 5.15.b, the estimated angle between two defects in z -axis with IMU assistance is corrected from 8 degrees to 2.3 degrees. Even with such a relatively small rotation angle (8 degrees), the IMU is sensitive enough to capture changes in rotation and turned out to be very precise about its orientation. Experimental results show that the proposed registration algorithm with the IMU data incorporated was sufficient to reconstruct the defect adequately, which provides a good basis for applying cylindrical-based 3D registration to facilitate more reliable data reconstruction.

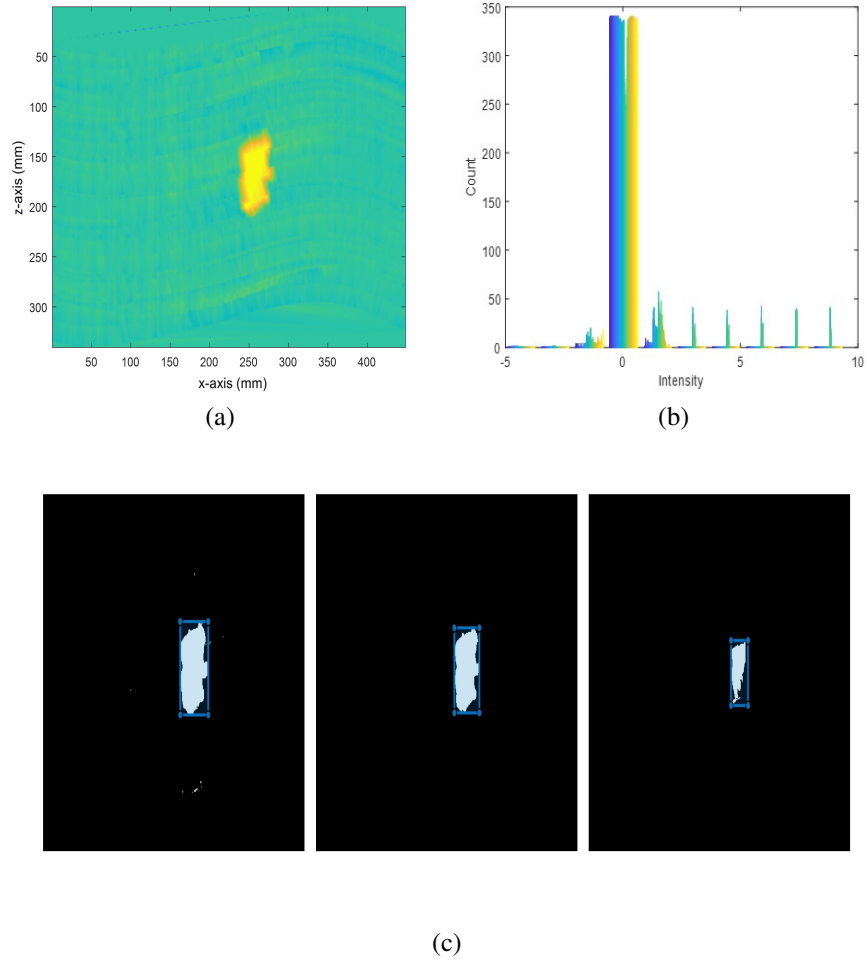


Figure 5.16 Intensity-based Threshold Searching Procedure: a) Cylindrical defect map; b) Intensity histogram; c) Binarized candidate examples with the segmented defect.

5.4.3.2 Reconstructed Defect Size Evaluation

To better illustrate the effectiveness and robustness of the robotic integrated platform, we estimate the reconstructed defect size for comparison with the ground truth defect. The idea of size estimation is converted to a segmentation problem, starting with flattening the reconstructed 3-D Point cloud to the cylindrical domain. The following defect estimation procedure is realized through a proposed Intensity-based Threshold Searching algorithm, which is presented in Fig. 5.16.

In detail, first, the intensity histogram is applied to Obtain N intensity clusters and then extract the mean of each cluster: $\mathbf{M}_i : \mathbf{M}_1, \mathbf{M}_2, \dots, \mathbf{M}_N$; Next, based on each $\mathbf{M}_i$, the binarized cylindrical

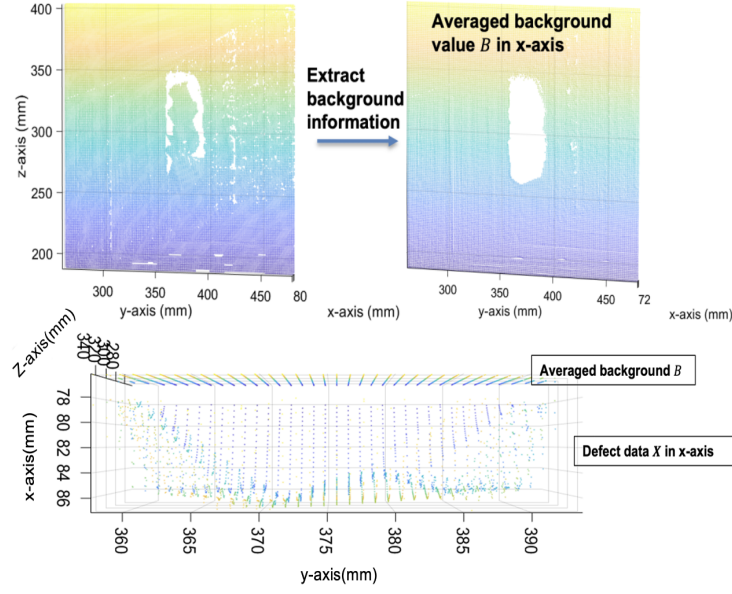
defect map $\mathbf{B}_i$ is generated for the following defect segmentation to obtain the estimated length $\mathbf{L}_i$ and Width $\mathbf{W}_i$; Then an error estimation is deployed for selecting the best candidate by evaluating the distance between the estimated size and the true defect size L_{GT} and W_{GT} , which is described as follows:

$$Err_i = 0.5 * \sqrt{\left(\frac{L_i - L_{GT}}{L_{GT}}\right)^2 + \left(\frac{W_i - W_{GT}}{W_{GT}}\right)^2}. \quad (5.11)$$

By selecting the minimal Err , the optimal threshold will be chosen and thus obtain the optimal estimated defect size length $\mathbf{L}^*$ and Width $\mathbf{W}^*$. Considering the reconstruction results are affected by various uncertainties, the calculated Err is a good estimation of the overall uncertainty in this 3D reconstruction-based sensing system.

The determination of defect depth holds significant importance as it plays a crucial role in evaluating the impact of a defect on the structural integrity of a component or material. In this specific application, the process of gauging defect depth commences with the extraction of a 3D point cloud that represents the defect area, as depicted in Figure 5.17. Initially, this 3D defect map is transformed onto the $Y - Z$ axis. Subsequently, the defect information is isolated to obtain the average background information B . This background information serves the purpose of filling in the defect region, resulting in the creation of a comprehensive 3D plot that accurately portrays the defect in relation to its reconstructed depth. Given that the actual defect surface possesses some inherent roughness, the variation in defect depth can be observed from the entire map. This variation provides a solid foundation for comprehending the inspected defect. Since the industry primarily focuses on determining the maximum wall loss, only the largest estimated depth measurement is considered in this scenario.

To investigate the accuracy of the reconstructed measurements, the estimated length L_i , width W_i , and depth D_i are compared to the ground truth defect size L_{GT} , W_{GT} , and D_{GT} . This evaluation involves an error estimation equation that considers the differences in length, width, and depth simultaneously. The equation for evaluating the overall size estimation performance can be obtained by extending the Equ.5.11 to:



(a)

Figure 5.17 Defect depth estimation procedure.

$$Error_i = \frac{\sqrt{\left(\frac{L_i - L_{GT}}{L_{GT}}\right)^2 + \left(\frac{W_i - W_{GT}}{W_{GT}}\right)^2 + \left(\frac{D_i - D_{GT}}{D_{GT}}\right)^2}}{3} \quad (5.12)$$

5.4.4 Uncertainty Analysis on SL Sensing System

5.4.4.1 Uncertainty Source

In the context of pipeline inspection and field testing, it is imperative to conduct a thorough investigation into the impact of various uncertainties. Uncertainty quantification (UQ) plays a pivotal role in quantitatively characterizing the quality and accuracy of non-destructive evaluation (NDE) and, consequently, the reliability of intricate systems. In this sensing system, any uncertainties or errors in data collection, feature extraction, or subsequent analysis can significantly influence the quality and reliability of the final 3D data reconstruction. Hence, uncertainty quantification serves as a means to impartially evaluate performance, offering a comprehensive analysis of the connection between uncertainty and the ultimate output. This, in turn, leads to the development of a highly reliable sensing system for pipeline inspection. The errors and uncertainties that impact the accuracy of this structured light 3D measurement system primarily stem from three sources: the instruments, the processing methods, and the environmental conditions.

1. From **Instrument Design**: A reliable and rigid mechanical design of the sensing system is important. The shadow effect of the sensor is one major uncertainty source, which deteriorates the reconstruction accuracy when dealing with abrupt height changes in the pipe surface. As mentioned in [285], this problem is caused by the current single camera and projector setup, which restricts the view angle. The low intensity of the light source and low resolution will cause poor imaging quality of the slide pattern and thus may affect the measurement accuracy. Also, the accurate distance and directional from IMU and encoder are key factors to determine the alignment between each frame and the accumulated error will lead to performance degradation of localization. For odometry input, possible error accumulation will be from either the measurements of the wheel diameter or speed estimation.

- a) Wheel diameter: non-circular wheel shape because of the usage of omnidirectional wheels.
- b) Wheel slippage: wheel slippage while passing over obstacles or during turning, will cause perturbations in velocity measurements. The relation between the acquired wheel slip velocity V_s and the measured velocity (rotational speed) V_r can be described through a linear function with an unbiased Gaussian uncertainty U [298], which can be written as:

$$\mathbf{V}_s = (f(V_r) + U) \quad (5.13)$$

while the uncertainty $U \sim \mathcal{N}(0, \sigma^2)$ is the unbiased Gaussian noise with zero mean and variance σ^2 ; $f(\cdot)$ is first-order polynomial fitting procedure with least square estimate. Therefore, if slip is considered in future work, the wheel slip distance T_{ods} , can be obtained by velocity and duration time Δt : $T_{ods} = V_s * \Delta t$.

2. From **Method**: The calibration of SL sensor's main components, such as the camera, projector, and IMU, is an essential part of obtaining the relative parameters for further data analysis. Errors introduced by the imprecise parameter estimation of the sensing system will

deteriorate a system's overall performance. Therefore, a reliable and efficient calibration method for each component is essential. For the following defect reconstruction and measurement, uncertainties from those processing models, such as the calculation error of the registration algorithm and threshold searching algorithm, also contribute to the uncertainty in the measurement result.

3. From **Environment**: The environment where tests and calibrations are performed can have an influence on uncertainty in measurement results. Vibration caused by the uneven pipe surface will introduce random errors to the measurements. Also, inadequate light conditions for measurements have a crucial impact on the imaging quality, which makes slide patterns difficult to distinguish.

The measurement uncertainties will result in the precision and accuracy of the reconstructed defect shape decreasing, which will be further investigated in the next section.

5.4.4.2 Measurement Uncertainty Evaluation

Similar to the previous MBN-based application, in this work, the GUM-based uncertainty analysis of repeatability is addressed to have a better evaluation to the reconstruction performance. Specifically, six repeated scans with the robotic platform were conducted. The estimated defect size and reconstructed *Err* of six tests are presented in Table 5.4. Results show that there exists variation in each reconstructed defect size, however, based on the estimated *Err* from Equ. 5.12, differences between the estimated defect size and the ground truth defect are relatively small, all within 13%. Therefore, the proposed registration and following defect estimation algorithm of this IMU-assisted SL sensing system is proven to reveal real defect information with high accuracy.

For estimating the uncertainty from measurement, the standard deviation S for every combination was determined by taking the mean of those six choose m standard deviations. The standard uncertainty to the estimated defect length, width and depth are computed separately, shown in Fig. 5.18. The results reveal a correlation between measurement uncertainties and the number of repetitions, indicating that uncertainties decrease as the number of repetitions increases. This

Table 5.4 Estimated Defect Size and *Error* in Repetition Tests

mm	Test 1	Test 2	Test 3	Test 4	Test 5	Test 6
Length	67.2	72.3	67.7	71.3	76.6	66.1
Width	34.2	35.7	33.1	31.3	35.2	32.6
Depth	5.7	8.6	9.4	4.7	5.6	4.8
Error	8.2%	5.1%	8.7%	12.9%	9.0%	12.4%

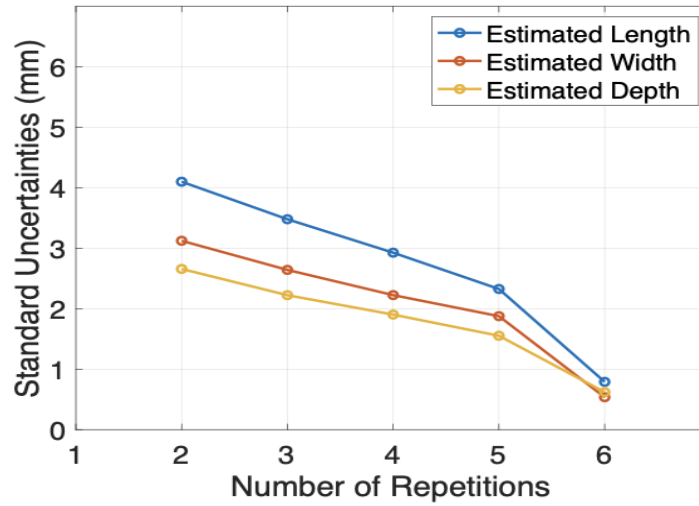


Figure 5.18 Comparison results of expanded uncertainty with GUM.

underscores the significance of conducting repeat measurements to reduce uncertainties effectively. Moreover, following the GUM, an expanded uncertainty is employed to ensure a 95% confidence level when the coverage factor k is set at 1.96, which is considered as the best estimate of the correction with the measurement error. The measurement uncertainty is expressed as::

$$\begin{aligned}
 U_{\text{extend}} &= k * U_{\text{standard}} = k * \frac{S}{\sqrt{N}} \\
 &= k * \sqrt{\frac{1}{N(n-1)} \sum_{i=1}^n (y_i - \bar{y})^2}
 \end{aligned} \tag{5.14}$$

where the mean value of N repeat measurements is considered as the optimal representation; y_i is the i th measurement value; and $\bar{y}$ is the average value of repeated measurements [259]. According to equation 5.14, the measurement uncertainty interval can be determined by considering the

estimated length and width of all six measurements. Specifically, the calculated confidence interval of length is $70.5 \pm 1.55\text{mm}$, the width is $33.7 \pm 1.05\text{mm}$, and the depth is $6.1 \pm 1.22\text{mm}$. Overall, the uncertainty of the length estimation is much smaller than that of the width, and since the frame distance is one of the main factors in determining the length, the accuracy and robustness of the odometer and IMU sensors in position correction are demonstrated.

While the number of repetitions is limited, the initial error estimates and the analysis of measurement uncertainty lay a solid groundwork for illustrating the potential of uncertainty estimates under the GUM. Additionally, the integrated robotics platform's ability to facilitate multiple Simultaneous Localization and Mapping (SLAM)-based data collections opens the possibility of constructing a more robust and comprehensive uncertainty estimation model through additional experiments.

5.5 Conclusion

In this chapter, we begin by introducing the concept of Probability of Detection (POD), which is a specific method aimed at assessing the effectiveness and reliability of defect detection under conditions of uncertainty within NDE. Additionally, we delve into the subject of measurement uncertainty analysis, which is the most popular means to evaluate and quantify the overall uncertainty inherent in NDE measurements. The information derived from both of these methods plays a critical role in shaping the interpretation of NDE results. Furthermore, the evaluated uncertainties resulting from both the POD and measurement uncertainty analysis methods are integral components of ensuring the reliability and robustness of NDE inspections.

Besides, two practical NDE applications are introduced to address the applicability of ANOVA-based measurement uncertainty:

In the first application, Magnetic Barkhausen Noise (MBN) is utilized, and relevant features are extracted to establish correlations with the microstructure information of martensitic-grade stainless-steel samples. A Probabilistic Neural Network (PNN) is subsequently employed to classify samples based on their remaining fatigue life percentage, effectively distinguishing between unfatigued and cracked samples.

The second application presents a comprehensive and practical Inertial Measurement Unit (IMU)-assisted robotic Structured Light (SL) sensing system with improved registration and defect estimation capabilities for pipeline detection. The results demonstrate that the proposed framework is capable of providing a robust and reliable 3D defect reconstruction solution, with valuable assistance from IMU and robot odometry.

Both of these sensing systems, whether it's the specially designed MBN sensor or the robot-enabled SL sensor, are capable of delivering consistent and reliable repeated measurements. This, in turn, forms a solid foundation for a more accurate estimation of measurement uncertainties.

CHAPTER 6

CONCLUSION AND FUTURE WORK

6.1 Conclusion

The main contributions of this thesis are concluded as follows:

1. A comprehensive TLS-based UQ framework has been proposed, which signifies an important step forward in the domain of Non-Destructive Evaluation (NDE) inspections. This framework expands the horizons of conventional NDE practices, offering a more holistic and comprehensive approach to system maintenance. With a clear focus on practicality, this proposed framework stands as a guiding light, facilitating the understanding and management of uncertainty propagation through a selection of advanced and widely adopted techniques tailored for NDE inspections. Within the framework, we've identified key criteria to guide the selection of appropriate methods for propagating uncertainty in NDE scenarios. These criteria are designed to ensure that uncertainty is comprehensively addressed, managed, and minimized, thereby enhancing the overall reliability and effectiveness of NDE inspections. The proposed uncertainty analysis scheme is a good basis for NDE system design optimization and quantification.

2. The comprehensive identification of uncertainty sources within each NDE application not only plays a pivotal role in ensuring the integrity of the inspection process but also serves as a foundational step for conducting more in-depth reliability and sensitivity analyses in this specialized domain. By diligently identifying and understanding these sources of uncertainty, NDE practitioners can enhance the accuracy and effectiveness of their inspections, ultimately contributing to safer and more dependable systems across various industries.

3. A capacitive sensing-based inspection is applied to illustrate the forward uncertainty propagation process from "Data " to "Modeling". The process of uncertainty propagation demonstrates how the uncertainty originating from the liftoff is systematically transferred through the various stages of the 'Modeling' process. The output distributions resulting from this propagation are derived from the implementation of the three proposed meta-models. Further, to comprehensively evaluate the influence of liftoff uncertainty on the final output, a widely used uncertainty prop-

agation method MCS is employed to generate a probability distribution for the total impedance, denoted as Z . This approach enables a thorough assessment of the potential variations in the total impedance due to the uncertainty in the measurements.

4. A Magnetic Flux Leakage-based defect characterization algorithm is used for addressing uncertainties during the inverse NDE process. It places a particular emphasis on the uncertainties stemming from sensing liftoff, which can influence the output signal of the sensing system. Given the intricate nature of the forward uncertainty propagation process, this research conducts a comparative analysis of two commonly used learning-based approximate Bayesian inference methods, Convolutional Neural Network (CNN) and Deep Ensemble (DE), to handle the input uncertainty derived from MFL response data. Moreover, the study employs an Autoencoder method to address the challenge of limited experimental data by augmenting the dataset. This approach involves pre-training on MFL simulation data and subsequently applying it to real-world data. In the context of defect size classification within experimental MFL-based applications, the study not only assesses prediction accuracy but also delves into uncertainty analysis. These efforts are crucial for evaluating the reliability of predictions. The proposed methodology for uncertainty quantification offers valuable insights into the assessment of reliability in decision-making and inverse problems related to MFL-based NDE.

5. While Probability of Detection (POD) and measurement uncertainty analysis serve distinct purposes in assessing uncertainty in NDE, it's important to note that the information derived from both methodologies plays a significant role in shaping the interpretation of NDE results within the proposed TLS UQ framework. The combined evaluation of uncertainty through these two methods is a critical component that enhances the overall reliability and robustness of NDE inspections. To assess the effectiveness of the introduced GUM-based measurement uncertainty analysis, two distinct sensing systems are employed: one based on structured light technology and the other on Magnetic Barkhausen Noise, which serves as practical test cases for measurement uncertainty analysis. In the MBN application, PCA can extract advanced features from both the time and frequency domain. A PNN is utilized for sample classification based on the percentage of

remaining fatigue life. The specimen's fatigue life can be successfully indicated with the potential for early fatigue onset prediction. In the second application, a comprehensive and practical IMU-assisted robotic SL sensing system with enhanced registration and defect estimation solutions is proposed for pipeline detection. The proposed framework utilizes a RANSAC-assisted cylindrical fitting registration algorithm to enhance the alignment of the SL system, ensuring accurate 3D profiling of the pipeline. Additionally, the system integrates inertial and odometry measurements to facilitate global and local positioning, further enhancing the precision and reliability of the 3D profiling process. Assisted with customized defect sizing techniques, The robot-enabled SL sensing system is able to provide a robust and reliable 3D defect reconstruction solution in terms of varying defect shape and depth. For both applications, the GUM-based uncertainty method is able to illustrate the reliability of the corresponding measurement process. Furthermore, the uncertainty sources in both sensing systems are described in detail, providing a guide for future uncertainty analysis.

6.2 Future Work

This section provides valuable insights into potential directions and ideas for the continuation of this research:

- TLS UQ framework:

It would be beneficial to incorporate additional UP methods within the proposed framework into NDE applications which can provide researchers with a deeper understanding of the versatility and applicability of the framework in various scenarios. Besides providing practical examples to demonstrate the potential and efficiency of this framework, which could further clarify the benefits and feasibility of utilizing these methods, ultimately strengthening the framework's value in the NDE field.

- Learning-based MFL defect characterization:

Exploring additional types of data augmentation techniques and incorporating more learning-based uncertainty propagation methods into the applied dataset can enhance the research by offering a broader perspective on the optimization of uncertainty propagation in NDE applications. Di-

versifying data augmentation methods allows for the evaluation of a wider range of approaches, potentially uncovering the most effective techniques for specific scenarios. Furthermore, integrating learning-based uncertainty propagation into the dataset can provide valuable insights into the nature of uncertainties and how they can be managed and minimized through advanced computational approaches. These efforts can lead to more robust, efficient, and data-driven solutions in the field of NDE.

- Velocity-induced MIEC effect on MFL signal:

Leveraging simulations provides a valuable means to gain deeper insights into the MIEC signal. By focusing on selected informative features, it becomes possible to explore the impact of velocity on signal characteristics and, concurrently, on defect sizing. Considering velocity as an uncertainty factor holds promise for incorporating it into a TLS-based uncertainty analysis, thus contributing to a more comprehensive understanding of the entire inspection process.

BIBLIOGRAPHY

- [1] Daniel Wallach, David Makowski, James W Jones, and Francois Brun. *Working with dynamic crop models: evaluation, analysis, parameterization, and applications*. Elsevier, 2006.
- [2] David Vose. *Quantitative risk analysis: a guide to Monte Carlo simulation modelling*. Wiley, 1996.
- [3] William L Oberkampf, Sharon M DeLand, Brian M Rutherford, Kathleen V Diegert, and Kenneth F Alvin. Error and uncertainty in modeling and simulation. *Reliability Engineering & System Safety*, 75(3):333–357, 2002.
- [4] Dan G Cacuci and Mihaela Ionescu-Bujor. A comparative review of sensitivity and uncertainty analysis of large-scale systems—ii: statistical methods. *Nuclear science and engineering*, 147(3):204–217, 2004.
- [5] Richard De Neufville, Olivier De Weck, Daniel Frey, Daniel Hastings, Richard Larson, David Simchi-Levi, Kenneth Oye, Annalisa Weigel, Roy Welsch, et al. Uncertainty management for engineering systems planning and design. In *Engineering Systems Symposium, MIT, Cambridge, MA*, 2004.
- [6] Shankar Sankararaman. Significance, interpretation, and quantification of uncertainty in prognostics and remaining useful life prediction. *Mechanical Systems and Signal Processing*, 52:228–247, 2015.
- [7] Erdem Acar, Gamze Bayrak, Yongsu Jung, Ikjin Lee, Palaniappan Ramu, and Suja Shree Ravichandran. Modeling, analysis, and optimization under uncertainties: a review. *Structural and Multidisciplinary Optimization*, 64(5):2909–2945, 2021.
- [8] F Owen Hoffman and Jana S Hammonds. Propagation of uncertainty in risk assessments: the need to distinguish between uncertainty due to lack of knowledge and uncertainty due to variability. *Risk analysis*, 14(5):707–712, 1994.
- [9] Shankar Sankararaman and Sankaran Mahadevan. Likelihood-based representation of epistemic uncertainty due to sparse point data and/or interval data. *Reliability Engineering & System Safety*, 96(7):814–824, 2011.
- [10] Yiping Li, Jianwen Chen, and Ling Feng. Dealing with uncertainty: A survey of theories and practices. *IEEE Transactions on Knowledge and Data Engineering*, 25(11):2463–2482, 2012.
- [11] Erdem Acar, Raphael T Haftka, and Theodore F Johnson. Tradeoff of uncertainty reduction mechanisms for reducing weight of composite laminates. 2007.
- [12] Erdem Acar, Raphael T Haftka, and Nam H Kim. Effects of structural tests on aircraft safety.

- AIAA journal*, 48(10):2235–2248, 2010.
- [13] Amit A Kale and Raphael T Haftka. Tradeoff of weight and inspection cost in reliability-based structural optimization. *Journal of Aircraft*, 45(1):77–85, 2008.
 - [14] Nikolaevich Kolmogorov, Andre and Albert T Bharucha-Reid. *Foundations of the theory of probability: Second English Edition*. Courier Dover Publications, 2018.
 - [15] Thomas Bayes. An essay towards solving a problem in the doctrine of chances. 1763. *MD computing: computers in medical practice*, 8(3):157–171, 1991.
 - [16] Didier Dubois and Henri Prade. Fuzzy sets, probability and measurement. *European Journal of Operational Research*, 40(2):135–154, 1989.
 - [17] Ha-Rok Bae, Ramana V Grandhi, and Robert A Canfield. An approximation approach for uncertainty quantification using evidence theory. *Reliability Engineering & System Safety*, 86(3):215–225, 2004.
 - [18] Ha-Rok Bae, Ramana V Grandhi, and Robert A Canfield. Epistemic uncertainty quantification techniques including evidence theory for large-scale structures. *Computers & Structures*, 82(13-14):1101–1112, 2004.
 - [19] Amos Tversky and Craig R Fox. Weighing risk and uncertainty. *Psychological review*, 102(2):269, 1995.
 - [20] Didier Dubois and Henri Prade. *Possibility theory: an approach to computerized processing of uncertainty*. Springer Science & Business Media, 2012.
 - [21] I Elishakoff and M Zingales. Contrasting probabilistic and anti-optimization approaches in an applied mechanics problem. *International journal of solids and structures*, 40(16):4281–4297, 2003.
 - [22] Rituparna Chutia. Uncertainty quantification under hybrid structure of probability-fuzzy parameters in gaussian plume model. *Life Cycle Reliability and Safety Engineering*, 6(4):277–284, 2017.
 - [23] Robert L Kosut, Ming Lau, and Stephen Boyd. Identification of systems with parametric and nonparametric uncertainty. In *1990 American Control Conference*, pages 2412–2417. IEEE, 1990.
 - [24] John McFarland and Sankaran Mahadevan. Error and variability characterization in structural dynamics modeling. *Computer Methods in Applied Mechanics and Engineering*, 197(29-32):2621–2631, 2008.
 - [25] Gideon Schwarz. Estimating the dimension of a model. *The annals of statistics*, pages

461–464, 1978.

- [26] Yoojeong Noh, KK Choi, and Ikjin Lee. Identification of marginal and joint cdfs using bayesian method for rbdo. *Structural and Multidisciplinary Optimization*, 40(1):35–51, 2010.
- [27] Andrew A Neath and Joseph E Cavanaugh. The bayesian information criterion: background, derivation, and applications. *Wiley Interdisciplinary Reviews: Computational Statistics*, 4(2):199–203, 2012.
- [28] Yosiyuki Sakamoto, Makio Ishiguro, and Genshiro Kitagawa. Akaike information criterion statistics. *Dordrecht, The Netherlands: D. Reidel*, 81(10.5555):26853, 1986.
- [29] Priscilla E Greenwood and Michael S Nikulin. *A guide to chi-squared testing*, volume 280. John Wiley & Sons, 1996.
- [30] Fritz W Scholz and Michael A Stephens. K-sample anderson–darling tests. *Journal of the American Statistical Association*, 82(399):918–924, 1987.
- [31] Young-Jin Kang, Jimin Hong, O Lim, Yoojeong Noh, et al. Reliability analysis using parametric and nonparametric input modeling methods. *Journal of the Computational Structural Engineering Institute of Korea*, 30(1):87–94, 2017.
- [32] Young-Jin Kang, Yoojeong Noh, O Lim, et al. Kernel density estimation with bounded data. *Structural and Multidisciplinary Optimization*, 57(1):95–113, 2018.
- [33] Roger Ghanem, David Higdon, Houman Owhadi, et al. *Handbook of uncertainty quantification*, volume 6. Springer, 2017.
- [34] Vistasp M Karbhari. *Non-destructive evaluation (NDE) of polymer matrix composites*. Elsevier, 2013.
- [35] D McA McKirdy, A Cochran, GB Donaldson, and A McNab. Forward and inverse processing in electromagnetic nde using squids. *Review of Progress in Quantitative Nondestructive Evaluation: Volume 15A*, pages 347–354, 1996.
- [36] Krishnan Balasubramaniam. Inverse models and implications for nde. In *Key Engineering Materials*, volume 321, pages 6–11. Trans Tech Publ, 2006.
- [37] Angelika Wronkowicz, Krzysztof Dragan, and Krzysztof Lis. Assessment of uncertainty in damage evaluation by ultrasonic testing of composite structures. *Composite structures*, 203: 71–84, 2018.
- [38] Marco A Azpurua, Ciro Tremola, and Eduardo Paez. Comparison of the gum and monte carlo methods for the uncertainty estimation in electromagnetic compatibility testing. *Progress in*

electromagnetics research B, 34:125–144, 2011.

- [39] Nicola Toscani, Flavia Grassi, Giordano Spadacini, and Sergio A Pignari. A possibilistic approach for the prediction of the risk of interference between power and signal lines onboard satellites. *Mathematical Problems in Engineering*, 2018, 2018.
- [40] Ladislav Janousek, Milan Smetana, and Marcel Alman. Decreasing uncertainty in size estimation of stress corrosion cracking from eddy-current signals. In *Electromagnetic Non-destructive Evaluation (XIV)*, pages 53–60. IOS Press, 2011.
- [41] Marc Horner, Stephen M Luke, Kerim O Genc, Todd M Pietila, Ross T Cotton, Benjamin A Ache, Zachary H Levine, and Kevin C Townsend. Toward estimating the uncertainty associated with three-dimensional geometry reconstructed from medical image data. *Journal of Verification, Validation and Uncertainty Quantification*, 4(4):041002, 2019.
- [42] Rambod Hadidi, Nenad Gucunski, and Ali Maher. Application of probabilistic approach to the solution of inverse problems in nondestructive testing and engineering geophysics. In *20th EEGS Symposium on the Application of Geophysics to Engineering and Environmental Problems*, pages cp–179. European Association of Geoscientists & Engineers, 2007.
- [43] Hajin Choi, Youngjib Ham, and John S Popovics. Integrated visualization for reinforced concrete using ultrasonic tomography and image-based 3-d reconstruction. *Construction and building materials*, 123:384–393, 2016.
- [44] SWEETY SHAHINUR. Quantification of uncertainty of material properties and its application. 2018.
- [45] F. Taddei, S. Martelli, B. Reggiani, L. Cristofolini, and M. Viceconti. Finite-element modeling of bones from ct data: Sensitivity to geometry and material uncertainties. *IEEE Transactions on Biomedical Engineering*, 53(11):2194–2200, 2006. doi: 10.1109/TBME.2006.879473.
- [46] Flavio de Andrade Silva, Nikhilesh Chawla, and Romildo Dias de Toledo Filho. Tensile behavior of high performance natural (sisal) fibers. *Composites Science and Technology*, 68 (15-16):3438–3443, 2008.
- [47] HAMZA Sekkak and MEHDI Modares. A comparison between probabilistic and possibilistic numerical schemes for structural uncertainty analysis. *Vol. 6 of Proc., Int. Structural Engineering and Construction*. Fargo, ND: International Structural Engineering and Construction Press. <https://doi.org/10.14455/ISEC.res>, 2019.
- [48] G Hübschen. Ultrasonic techniques for materials characterization. In *Materials Characterization Using Nondestructive Evaluation (NDE) Methods*, pages 177–224. Elsevier, 2016.
- [49] I Altpeter, R Tschuncky, and K Szielasko. Electromagnetic techniques for materials charac-

- terization. In *Materials characterization using nondestructive evaluation (NDE) methods*, pages 225–262. Elsevier, 2016.
- [50] R Tschuncky, K Szielasko, and I Altpeter. Hybrid methods for materials characterization. In *Materials Characterization Using Nondestructive Evaluation (NDE) Methods*, pages 263–291. Elsevier, 2016.
 - [51] Housseem Haddar, Zixian Jiang, and Mohamed Kamel Riahi. A robust inversion method for quantitative 3d shape reconstruction from coaxial eddy current measurements. *Journal of Scientific Computing*, 70(1):29–59, 2017.
 - [52] Zehor Oudni, Azouaou Berkache, Hamid Mehaddene, Hassane Mohellebi, and Jinyi Lee. Comparative study to assess reliability in the presence of two geometric defect shapes for non-destructive testing. *Przegląd Elektrotechniczny*, 95(12):48–52, 2019.
 - [53] Felix H Kim, Adam Pintar, Anne-Françoise Obaton, Jason Fox, Jared Tarr, and Alkan Donmez. Merging experiments and computer simulations in x-ray computed tomography probability of detection analysis of additive manufacturing flaws. *NDT & E International*, 119:102416, 2021.
 - [54] Roger G Ghanem and Pol D Spanos. *Stochastic finite elements: a spectral approach*. Courier Corporation, 2003.
 - [55] Zehor Oudni, Mouloud Féliachi, and Hassane Mohellebi. Assessment of the probability of failure for ec nondestructive testing based on intrusive spectral stochastic finite element method. *The European Physical Journal-Applied Physics*, 66(3), 2014.
 - [56] Milan Smetana, Lukas Behun, Daniela Gombarska, and Ladislav Janousek. New proposal for inverse algorithm enhancing noise robust eddy-current non-destructive evaluation. *Sensors*, 20(19):5548, 2020.
 - [57] Catalin Mandache, Mike Brothers, and Vivier Lefebvre. Time domain lift-off compensation method for eddy current testing. *NDT. net*, 10(6):1–7, 2005.
 - [58] Lukáš Behúň and Milan Smetana. Decreasing uncertainty in width estimation of edm cracking from eddy-current signals. In *2016 ELEKTRO*, pages 474–477. IEEE, 2016.
 - [59] Subrata Mukherjee, Xuhui Huang, Lalita Udpa, and Yiming Deng. Nde based cost-effective detection of obtrusive and coincident defects in pipelines under uncertainties. In *2019 Prognostics and System Health Management Conference (PHM-Paris)*, pages 297–302. IEEE, 2019.
 - [60] Julian Q Kosciessa, Ulman Lindenberger, and Douglas D Garrett. Thalamocortical excitability modulation guides human perception under uncertainty. *Nature communications*, 12(1): 1–15, 2021.

- [61] Mingyang Lu, Wenqian Zhu, Liyuan Yin, Anthony J Peyton, Wuliang Yin, and Zhigang Qu. Reducing the lift-off effect on permeability measurement for magnetic plates from multifrequency induction data. *IEEE Transactions on Instrumentation and Measurement*, 67(1):167–174, 2017.
- [62] I Iso and BIPM OIML. Guide to the expression of uncertainty in measurement. *Geneva, Switzerland*, 122:16–17, 1995.
- [63] B Venkatraman and Baldev Raj. Nondestructive testing: An overview of techniques and application for quality evaluation. *Non-Destructive Evaluation of Corrosion and Corrosion-assisted Cracking*, pages 1–55, 2019.
- [64] Menner A Tatang, Wenwei Pan, Ronald G Prinn, and Gregory J McRae. An efficient method for parametric uncertainty analysis of numerical geophysical models. *Journal of Geophysical Research: Atmospheres*, 102(D18):21925–21932, 1997.
- [65] Hélder S Sousa, John D Sørensen, Poul H Kirkegaard, Jorge M Branco, and Paulo B Lourenço. On the use of ndt data for reliability-based assessment of existing timber structures. *Engineering Structures*, 56:298–311, 2013.
- [66] Gabriel LS Silva, Daniel A Castello, Lavinia Borges, and Jari P Kaipio. Damage identification in plates under uncertain boundary conditions. *Mechanical Systems and Signal Processing*, 144:106884, 2020.
- [67] Gözde Sarı and Mehmet Pakdemirli. Effects of non-ideal boundary conditions on the vibrations of a slightly curved micro beam. In *AIP Conference Proceedings*, volume 1493, pages 883–890. American Institute of Physics, 2012.
- [68] James-A Goulet and Ian FC Smith. Structural identification with systematic errors and unknown uncertainty dependencies. *Computers & structures*, 128:251–258, 2013.
- [69] Canhai Lai and Xin Sun. Predicting flaw-induced resonance spectrum shift with theoretical perturbation analysis. *Journal of Sound and Vibration*, 332(22):5953–5964, 2013.
- [70] John C Aldrin, Enrique A Medina, Eric A Lindgren, Charles Buynak, Gary Steffes, and Mark Derriso. Model-assisted probabilistic reliability assessment for structural health monitoring systems. In *AIP Conference Proceedings*, volume 1211, pages 1965–1972. American Institute of Physics, 2010.
- [71] Ming Yang. *Efficient methods for solving boundary integral equation in diffusive scalar problem and eddy current nondestructive evaluation*. PhD thesis, Iowa State University Ames, IA, 2010.
- [72] Dezhi Li, Wilson Wang, and Fathy Ismail. Enhanced fuzzy-filtered neural networks for material fatigue prognosis. *Applied Soft Computing*, 13(1):283–291, 2013.

- [73] Jie Chen and Yongming Liu. Fatigue modeling using neural networks: A comprehensive review. *Fatigue & Fracture of Engineering Materials & Structures*, 45(4):945–979, 2022.
- [74] Pratesh Jayaswal, Sunita N Verma, and Arun K Wadhwani. Application of ann, fuzzy logic and wavelet transform in machine fault diagnosis using vibration signal analysis. *Journal of Quality in Maintenance Engineering*, 2010.
- [75] RE Silva, Rafael Gouriveau, Samir Jemei, Daniel Hissel, Loïc Boulon, Kodjo Agbossou, and N Yousfi Steiner. Proton exchange membrane fuel cell degradation prediction based on adaptive neuro-fuzzy inference systems. *International Journal of Hydrogen Energy*, 39(21):11128–11144, 2014.
- [76] Matthias Seeger. Gaussian processes for machine learning. *International journal of neural systems*, 14(02):69–106, 2004.
- [77] Alexandra Coppe, Raphael T Haftka, and Nam H Kim. Uncertainty identification of damage growth parameters using nonlinear regression. *AIAA journal*, 49(12):2818–2821, 2011.
- [78] Qinming Liu, Ming Dong, and Ying Peng. A novel method for online health prognosis of equipment based on hidden semi-markov model using sequential monte carlo methods. *Mechanical Systems and Signal Processing*, 32:331–348, 2012.
- [79] Jooho Choi, Dawn An, Jinhyuk Gang, Jinwon Joo, and Nam Ho Kim. Bayesian approach for parameter estimation in the structural analysis and prognosis. In *Annual Conference of the PHM Society*, volume 2, 2010.
- [80] Arnaud Doucet, Nando De Freitas, Neil James Gordon, et al. *Sequential Monte Carlo methods in practice*, volume 1. Springer, 2001.
- [81] Wei He, Nick Williard, MPAL Osterman, and Michael Pecht. Prognostics of lithium-ion batteries using extended kalman filtering. In *Proceedings of the IMAPS Advanced Technology Workshop on High Reliability Microelectronics for Military Applications, Linthicum Heights, MD, USA*, volume 1719, 2011.
- [82] Dawn An, Nam H Kim, and Joo-Ho Choi. Practical options for selecting data-driven or physics-based prognostics algorithms with reviews. *Reliability Engineering & System Safety*, 133:223–236, 2015.
- [83] Sun Hye Kim and Fani Boukouvala. Machine learning-based surrogate modeling for data-driven optimization: a comparison of subset selection for regression techniques. *Optimization Letters*, 14(4):989–1010, 2020.
- [84] Arinan Dourado and Felipe AC Viana. Physics-informed neural networks for corrosion-fatigue prognosis. In *Proceedings of the Annual Conference of the PHM Society*, volume 11, 2019.

- [85] Arinan Dourado and Felipe AC Viana. Physics-informed neural networks for missing physics estimation in cumulative damage models: a case study in corrosion fatigue. *Journal of Computing and Information Science in Engineering*, 20(6), 2020.
- [86] Nazri Mohd Nawi, RS Ransing, and MR Ransing. An improved conjugate gradient based learning algorithm for back propagation neural networks. *International Journal of Computer and Information Engineering*, 2(6):2062–2071, 2008.
- [87] Xuefei Guan, Yongming Liu, Abhinav Saxena, Jose Celaya, and Kai Goebel. Entropy-based probabilistic fatigue damage prognosis and algorithmic performance comparison. In *Annual Conference of the PHM Society*, volume 1, 2009.
- [88] Matthew Cherry, Harold Sabbagh, John Aldrin, Jeremy Knopp, and Adam Pilchak. Forward propagation of parametric uncertainties through models of nde inspection scenarios. In *AIP Conference Proceedings*, volume 1650, pages 1884–1892. American Institute of Physics, 2015.
- [89] Daniel Gauder, Michael Biehler, Johannes Gözl, Benedict Stampfer, David Böttger, Benjamin Häfner, Bernd Wolter, Volker Schulze, and Gisela Lanza. Development of a methodical approach for uncertainty quantification and meta-modeling of surface hardness in white layers of longitudinal turned aisi4140 surfaces. *tm-Technisches Messen*, 88(11):661–673, 2021.
- [90] Jeffrey A Kornuta, Nicoli M Ames, Mary W Louie, Peter Veloo, and Troy Rovella. Uncertainty quantification of nondestructive techniques to verify pipeline material strength. In *International Pipeline Conference*, volume 51869, page V001T03A046. American Society of Mechanical Engineers, 2018.
- [91] Zi Li, Bharath Basti Shenoy, Lalita Udpa, Satish Udpa, and Yiming Deng. Magnetic barkhausen noise technique for early-stage fatigue prediction in martensitic stainless-steel samples. *Journal of Nondestructive Evaluation, Diagnostics and Prognostics of Engineering Systems*, 4(4), 2021.
- [92] Xiaojing Wu, Weiwei Zhang, and Shufang Song. Uncertainty quantification and sensitivity analysis of transonic aerodynamics with geometric uncertainty. *International Journal of Aerospace Engineering*, 2017, 2017.
- [93] Sez Atamturktur, Saurabh Prabhu, and Gregory Roche. Predictive modeling of large scale historic masonry monuments: uncertainty quantification and model validation. In *9th International Conference on Structural Dynamics, EUROLYN 2014*, pages 2721–2727. European Association for Structural Dynamics, 2014.
- [94] Zi Li, Xuhui Huang, Obaid Elshafiey, Subrata Mukherjee, and Yiming Deng. Fem of magnetic flux leakage signal for uncertainty estimation in crack depth classification using bayesian convolutional neural network and deep ensemble. In *2021 International Applied*

Computational Electromagnetics Society Symposium (ACES), pages 1–4. IEEE, 2021.

- [95] Ming Hong, Zhu Mao, Michael D Todd, and Zhongqing Su. Uncertainty quantification for acoustic nonlinearity parameter in lamb wave-based prediction of barely visible impact damage in composites. *Mechanical Systems and Signal Processing*, 82:448–460, 2017.
- [96] John C Aldrin, Jeremy S Knopp, Mark P Blodgett, and Harold A Sabbagh. Uncertainty propagation in eddy current nde inverse problems. In *AIP Conference Proceedings*, volume 1335, pages 631–638. American Institute of Physics, 2011.
- [97] Hyung-Seop Shim. Performance evaluation of nde methods. *KSCE Journal of Civil Engineering*, 7(2):185–192, 2003.
- [98] Samira Mohammadi and Selen Cremaschi. Efficiency of uncertainty propagation methods for moment estimation of uncertain model outputs. *Computers & Chemical Engineering*, page 107954, 2022.
- [99] Sang Hoon Lee and Wei Chen. A comparative study of uncertainty propagation methods for black-box-type problems. *Structural and multidisciplinary optimization*, 37:239–253, 2009.
- [100] Cosmin Safta, Richard L-Y Chen, Habib N Najm, Ali Pinar, and Jean-Paul Watson. Efficient uncertainty quantification in stochastic economic dispatch. *IEEE Transactions on Power Systems*, 32(4):2535–2546, 2016.
- [101] XY Jia, C Jiang, CM Fu, BY Ni, CS Wang, and MH Ping. Uncertainty propagation analysis by an extended sparse grid technique. *Frontiers of Mechanical Engineering*, 14:33–46, 2019.
- [102] Martin Hunt, Benjamin Haley, Michael McLennan, Marisol Koslowski, Jayathi Murthy, and Alejandro Strachan. Puq: A code for non-intrusive uncertainty propagation in computer simulations. *Computer Physics Communications*, 194:97–107, 2015.
- [103] Mohammad Mahdi Rajabi. Review and comparison of two meta-model-based uncertainty propagation analysis methods in groundwater applications: polynomial chaos expansion and gaussian process emulation. *Stochastic environmental research and risk assessment*, 33: 607–631, 2019.
- [104] Chiara Tardioli, Martin Kubicek, Massimiliano Vasile, Edmondo Minisci, and Annalisa Riccardi. Comparison of non-intrusive approaches to uncertainty propagation in orbital mechanics. 2015.
- [105] D Ye, L Veen, A Nikishova, J Lakhilili, W Edeling, OO Luk, VV Krzhizhanovskaya, and AG Hoekstra. Uncertainty quantification patterns for multiscale models. *Philosophical Transactions of the Royal Society A*, 379(2197):20200072, 2021.
- [106] Jeff Duffy, Songquan Liu, Herbert Moskowitz, Robert Plante, and Paul V Preckel. Assessing

- multivariate process/product yield via discrete point approximation. *IIE transactions*, 30: 535–543, 1998.
- [107] Paul Glasserman. *Monte Carlo methods in financial engineering*, volume 53. Springer, 2004.
 - [108] Alex A Gorodetsky, Gianluca Geraci, Michael S Eldred, and John D Jakeman. A generalized approximate control variate framework for multifidelity uncertainty quantification. *Journal of Computational Physics*, 408:109257, 2020.
 - [109] Ajay Jasra, Kody JH Law, and Yan Zhou. Forward and inverse uncertainty quantification using multilevel monte carlo algorithms for an elliptic nonlocal equation. *International Journal for Uncertainty Quantification*, 6(6), 2016.
 - [110] John Hammersley. *Monte carlo methods*. Springer Science & Business Media, 2013.
 - [111] Jiaxin Zhang. Modern monte carlo methods for efficient uncertainty quantification and propagation: A survey. *Wiley Interdisciplinary Reviews: Computational Statistics*, 13(5): e1539, 2021.
 - [112] Eric Biedermann, Leanne Jauriqui, John C Aldrin, Alexander Mayes, Tom Williams, and Siamack Mazdiyasni. Uncertainty quantification in modeling and measuring components with resonant ultrasound spectroscopy. In *AIP Conference Proceedings*, volume 1706. AIP Publishing, 2016.
 - [113] Matthias Franz Rath, Bernhard Schweighofer, and Hannes Wegleiter. Uncertainty analysis of an optoelectronic strain measurement system for flywheel rotors. *Sensors*, 21(24):8393, 2021.
 - [114] Olivier Le Maître and Omar M Knio. *Spectral methods for uncertainty quantification: with applications to computational fluid dynamics*. Springer Science & Business Media, 2010.
 - [115] Dongbin Xiu. *Numerical methods for stochastic computations: a spectral method approach*. Princeton university press, 2010.
 - [116] Jun Xu and Fan Kong. A cubature collocation based sparse polynomial chaos expansion for efficient structural reliability analysis. *Structural Safety*, 74:24–31, 2018.
 - [117] Jeongeun Son and Yuncheng Du. An efficient polynomial chaos expansion method for uncertainty quantification in dynamic systems. *Applied Mechanics*, 2(3):460–481, 2021.
 - [118] Jethro Nagawkar and Leifur Leifsson. Applications of polynomial chaos-based cokriging to simulation-based analysis and design under uncertainty. In *International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*, volume 84010, page V11BT11A046. American Society of Mechanical Engineers, 2020.

- [119] TT Zygididis, AE Kyrgiazoglou, and TP Theodoulidis. Polynomial-chaos uncertainty modeling in eddy-current inspection of cracks.
- [120] Firooz Bakhtiari-Nejad, Naserodin Sepehry, and Mahnaz Shamshirsaz. Polynomial chaos expansion sensitivity analysis for electromechanical impedance of plate. In *International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*, volume 50206, page V008T10A037. American Society of Mechanical Engineers, 2016.
- [121] Arnold Binger and Sándor Bilicz. Sensitivity analysis using a sparse grid surrogate model in electromagnetic nde. In *Electromagnetic non-destructive evaluation (XXI)*, pages 152–159. IOS Press, 2018.
- [122] Nahid Sanzida and Zoltan K Nagy. Polynomial chaos expansion (pce) based surrogate modeling and optimization for batch crystallization processes. In *Computer Aided Chemical Engineering*, volume 33, pages 565–570. Elsevier, 2014.
- [123] K Konakli, C Mylonas, S Marelli, and B Sudret. Uqlab user manual—canonical low-rank approximations. *Report UQLab-V1*, pages 1–108, 2019.
- [124] Georges Matheron. Principles of geostatistics. *Economic geology*, 58(8):1246–1266, 1963.
- [125] Frank L Hitchcock. The expression of a tensor or a polyadic as a sum of products. *Journal of Mathematics and Physics*, 6(1-4):164–189, 1927.
- [126] Subrata Mukherjee, Xuhui Huang, Lalita Udpa, and Yiming Deng. A kriging-based magnetic flux leakage method for fast defect detection in massive pipelines. *Journal of Nondestructive Evaluation, Diagnostics and Prognostics of Engineering Systems*, 5(1):011002, 2022.
- [127] Leifur Leifsson, Xiaosong Du, and Slawomir Koziel. Multifidelity modeling of ultrasonic testing simulations with cokriging. In *2018 IEEE MTT-S International Conference on Numerical Electromagnetic and Multiphysics Modeling and Optimization (NEMO)*, pages 1–4. IEEE, 2018.
- [128] Michael Eldred. Recent advances in non-intrusive polynomial chaos and stochastic collocation methods for uncertainty analysis and design. In *50th AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference 17th AIAA/ASME/AHS Adaptive Structures Conference 11th AIAA No*, page 2274, 2009.
- [129] George Klir and Bo Yuan. *Fuzzy sets and fuzzy logic*, volume 4. Prentice hall New Jersey, 1995.
- [130] Ranjan Ganguli. A fuzzy logic system for ground based structural health monitoring of a helicopter rotor using modal data. *Journal of Intelligent Material Systems and Structures*, 12(6):397–407, 2001.

- [131] Giovanni Cascante, Homayoun Najjaran, and Paola Ronca. Relative ndt evaluation of the side walls of a brick channel. In *Advances in Engineering Structures, Mechanics & Construction*, pages 485–492. Springer, 2006.
- [132] Vahid Jahangiri, Hadi Mirab, Reza Fathi, and Mir Mohammad Ettefagh. Tlp structural health monitoring based on vibration signal of energy harvesting system. *Latin American Journal of Solids and Structures*, 13:897–915, 2016.
- [133] Bertrand Iooss and Loïc Le Gratiet. Uncertainty and sensitivity analysis of functional risk curves based on gaussian processes. *Reliability Engineering & System Safety*, 187:58–66, 2019.
- [134] Valérie Kaftandjian, Yue Min Zhu, Olivier Dupuis, and Daniel Babot. The combined use of the evidence theory and fuzzy logic for improving multimodal nondestructive testing systems. *IEEE Transactions on Instrumentation and Measurement*, 54(5):1968–1977, 2005.
- [135] Ram M Narayanan and Robin James. International journal of microwaves applications. *International Journal*, 7(1), 2018.
- [136] Zhen Li and Zhaozong Meng. A review of the radio frequency non-destructive testing for carbon-fibre composites. *Measurement Science Review*, 16(2):1, 2016.
- [137] Xiaodong Shi, Vivek T Rathod, Saptarshi Mukherjee, Lalita Udpa, and Yiming Deng. Multi-modality strain estimation using a rapid near-field microwave imaging system for dielectric materials. *Measurement*, 151:107243, 2020.
- [138] Paul Probst. *A Miniaturized Multi-Modality Imaging System for Dielectric Materials Evaluation*. Michigan State University, 2021.
- [139] Deng Huang, Theodore T Allen, William I Notz, and Ning Zeng. Global optimization of stochastic black-box systems via sequential kriging meta-models. *Journal of global optimization*, 34(3):441–466, 2006.
- [140] Roland Schöbi and Bruno Sudret. Pc-kriging: a new metamodeling method combining polynomial chaos expansions and kriging. In *Proc. 2nd Int. Symp. Uncertain. Quantif. Stoch. Model., Rouen, France*, 2014.
- [141] Thomas J Santner, Brian J Williams, William I Notz, and Brian J Williams. *The design and analysis of computer experiments*, volume 1. Springer, 2003.
- [142] R Schöbi, S Marelli, and B Sudret. Uqlab user manual–pc-kriging. *Report UQLab-V1*, pages 1–109, 2017.
- [143] Nikolaus Hansen and Andreas Ostermeier. Completely derandomized self-adaptation in evolution strategies. *Evolutionary computation*, 9(2):159–195, 2001.

- [144] Minas D Spiridonakos and Eleni N Chatzi. Metamodeling of dynamic nonlinear structural systems through polynomial chaos narx models. *Computers & Structures*, 157:99–113, 2015.
- [145] Pierric Kersaudy, Bruno Sudret, Nadège Varsier, Odile Picon, and Joe Wiart. A new surrogate modeling technique combining kriging and polynomial chaos expansions–application to uncertainty analysis in computational dosimetry. *Journal of Computational Physics*, 286: 103–117, 2015.
- [146] Thierry Crestaux, Olivier Le Maître, and Jean-Marc Martinez. Polynomial chaos expansion for sensitivity analysis. *Reliability Engineering & System Safety*, 94(7):1161–1172, 2009.
- [147] Katerina Konakli and Bruno Sudret. Polynomial meta-models with canonical low-rank approximations: Numerical insights and comparison to sparse polynomial chaos expansions. *Journal of Computational Physics*, 321:1144–1169, 2016.
- [148] F Bartolucci and L Scrucca. Point estimation methods with applications to item response theory models. 2010.
- [149] Ralph C Smith. *Uncertainty quantification: theory, implementation, and applications*, volume 12. Siam, 2013.
- [150] David Blackwell and Lester Dubins. Merging of opinions with increasing information. *The Annals of Mathematical Statistics*, 33(3):882–886, 1962.
- [151] Jadie Adams and Shireen Y Elhabian. Benchmarking scalable epistemic uncertainty quantification in organ segmentation. *arXiv preprint arXiv:2308.07506*, 2023.
- [152] Shiro Kubo. Inverse analyses and their applications to nondestructive evaluations. *Proc. 12th A-PCNDT*, 2006.
- [153] W Lord. Nondestructive evaluation inverse problems. In *Elsevier Studies in Applied Electromagnetics in Materials*, volume 6, pages 101–104. Elsevier, 1995.
- [154] F Honarvar and A Varvani-Farahani. A review of ultrasonic testing applications in additive manufacturing: Defect evaluation, material characterization, and process control. *Ultrasonics*, 108:106227, 2020.
- [155] Richard A Ketcham and William D Carlson. Acquisition, optimization and interpretation of x-ray computed tomographic imagery: applications to the geosciences. *Computers & Geosciences*, 27(4):381–400, 2001.
- [156] BA Auld and JC Moulder. Review of advances in quantitative eddy current nondestructive evaluation. *Journal of Nondestructive evaluation*, 18:3–36, 1999.

- [157] Karol Grondzak. Inverse problems of eddy current testing and uncertainties evaluation. *Advances in Electrical and Electronic Engineering*, 5(1):245–248, 2011.
- [158] Xu Wu, Ziyu Xie, Farah Alsafadi, and Tomasz Kozlowski. A comprehensive survey of inverse uncertainty quantification of physical model parameters in nuclear system thermal–hydraulics codes. *Nuclear Engineering and Design*, 384:111460, 2021.
- [159] Xiao Wang et al. Frequentist and bayesian approaches for probabilistic fatigue life assessment of high-speed train using in-service monitoring data. 2018.
- [160] Eyke Hüllermeier and Willem Waegeman. Aleatoric and epistemic uncertainty in machine learning: An introduction to concepts and methods. *Machine Learning*, 110:457–506, 2021.
- [161] Leo Breiman. Bagging predictors. *Machine learning*, 24:123–140, 1996.
- [162] John A Rice. *Mathematical statistics and data analysis*. Cengage Learning, 2006.
- [163] Anthony Kulesa, Martin Krzywinski, Paul Blainey, and Naomi Altman. Sampling distributions and the bootstrap. 2015.
- [164] Robert E Kass, Uri T Eden, Emery N Brown, Robert E Kass, Uri T Eden, and Emery N Brown. Propagation of uncertainty and the bootstrap. *Analysis of Neural Data*, pages 221–246, 2014.
- [165] Felix Kim, Adam Pinter, Jason Fox, Jared Tarr, Alkan Donmez, et al. Probability of detection of x-ray computed tomography of additive manufacturing defects. *Review of Progress in Quantitative Nondestructive Evaluation*, 2019.
- [166] Radford M Neal. *Bayesian learning for neural networks*, volume 118. Springer Science & Business Media, 2012.
- [167] Saptarshi Mukherjee, Hillary Fairbanks, Jordan Lum, David Stobbe, Gabe Guss, Andrew Townsend, Seemeen Karimi, and Joseph Tringe. A bayesian inference technique for ultrasound uncertainty quantification in metal additive manufacturing. *Available at SSRN 4250943*.
- [168] William C Schneck III, Heather Reed, Elizabeth D Gregory, and Cara AC Leckey. Sequential monte carlo based parameter estimation for structural health monitoring with an intel xeon phi optimized ultrasound kernel. In *AIP Conference Proceedings*, volume 2102, page 020035. AIP Publishing LLC, 2019.
- [169] Johnathan M Bardsley. Mcmc-based image reconstruction with uncertainty quantification. *SIAM Journal on Scientific Computing*, 34(3):A1316–A1332, 2012.
- [170] Alex Graves. Practical variational inference for neural networks. *Advances in neural infor-*

- mation processing systems*, 24, 2011.
- [171] David M Blei, Alp Kucukelbir, and Jon D McAuliffe. Variational inference: A review for statisticians. *Journal of the American statistical Association*, 112(518):859–877, 2017.
 - [172] Konstantin Posch, Jan Steinbrener, and Jürgen Pilz. Variational inference to measure model uncertainty in deep neural networks. *arXiv preprint arXiv:1902.10189*, 2019.
 - [173] Herbert E Robbins. An empirical bayes approach to statistics. In *Breakthroughs in Statistics: Foundations and Basic Theory*, pages 388–394. Springer, 1992.
 - [174] Christos Louizos and Max Welling. Multiplicative normalizing flows for variational bayesian neural networks. In *International Conference on Machine Learning*, pages 2218–2227. PMLR, 2017.
 - [175] Yarin Gal and Zoubin Ghahramani. Bayesian convolutional neural networks with bernoulli approximate variational inference. *arXiv preprint arXiv:1506.02158*, 2015.
 - [176] Durk P Kingma, Tim Salimans, and Max Welling. Variational dropout and the local reparameterization trick. *Advances in neural information processing systems*, 28, 2015.
 - [177] Yarin Gal and Zoubin Ghahramani. Dropout as a bayesian approximation: Representing model uncertainty in deep learning. In *international conference on machine learning*, pages 1050–1059. PMLR, 2016.
 - [178] Yongchan Kwon, Joong-Ho Won, Beom Joon Kim, and Myunghee Cho Paik. Uncertainty quantification using bayesian neural networks in classification: Application to biomedical image segmentation. *Computational Statistics & Data Analysis*, 142:106816, 2020.
 - [179] Seyed Omid Sajedi and Xiao Liang. Uncertainty-assisted deep vision structural health monitoring. *Computer-Aided Civil and Infrastructure Engineering*, 36(2):126–142, 2021.
 - [180] Zi Li. *Deep Learning Techniques for Magnetic Flux Leakage Inspection with Uncertainty Quantification*. Michigan State University, 2019.
 - [181] Balaji Lakshminarayanan, Alexander Pritzel, and Charles Blundell. Simple and scalable predictive uncertainty estimation using deep ensembles. *Advances in neural information processing systems*, 30, 2017.
 - [182] Malte Nalenz. *Characterizing model uncertainty in ensemble learning*. PhD thesis, lmu, 2022.
 - [183] Cheng Li. A gentle introduction to gradient boosting. URL: http://www.ccs.neu.edu/home/~vip/teach/MLcourse/4_boosting/slides/gradient_boosting.pdf, page 30, 2016.

- [184] Miguel Á Carreira-Perpiñán and Arman Zharmagambetov. Ensembles of bagged tao trees consistently improve over random forests, adaboost and gradient boosting. In *Proceedings of the 2020 ACM-IMS on foundations of data science conference*, pages 35–46, 2020.
- [185] Merrill B Rudd, James T Thorson, and Skyler R Sagarese. Ensemble models for data-poor assessment: accounting for uncertainty in life-history information. *ICES Journal of Marine Science*, 76(4):870–883, 2019.
- [186] Jerome H Friedman. Stochastic gradient boosting. *Computational statistics & data analysis*, 38(4):367–378, 2002.
- [187] Tianqi Chen and Carlos Guestrin. Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, pages 785–794, 2016.
- [188] Ossi Kajasalo. Deep ensemble-based speed of sound field estimation with uncertainty quantification in ultrasonic tomography. Master’s thesis, Itä-Suomen yliopisto, 2023.
- [189] Homin Song and Yongchao Yang. Uncertainty quantification in super-resolution guided wave array imaging using a variational bayesian deep learning approach. *NDT & E International*, 133:102753, 2023.
- [190] David Draper. Assessment and propagation of model uncertainty. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 57(1):45–70, 1995.
- [191] Max Hinne, Quentin F Gronau, Don van den Bergh, and Eric-Jan Wagenmakers. A conceptual introduction to bayesian model averaging. *Advances in Methods and Practices in Psychological Science*, 3(2):200–215, 2020.
- [192] Theo S Eicher, Chris Papageorgiou, and Adrian E Raftery. Default priors and predictive performance in bayesian model averaging, with application to growth determinants. *Journal of Applied Econometrics*, 26(1):30–55, 2011.
- [193] Jie Chen and Yongming Liu. Multimodality data fusion for probabilistic strength estimation of aging materials using bayesian networks. In *AIAA Scitech 2020 Forum*, page 1653, 2020.
- [194] Ruqi Zhang, Chunyuan Li, Jianyi Zhang, Changyou Chen, and Andrew Gordon Wilson. Cyclical stochastic gradient mcmc for bayesian deep learning. *arXiv preprint arXiv:1902.03932*, 2019.
- [195] Yuki Mae, Wataru Kumagai, and Takafumi Kanamori. Uncertainty propagation for dropout-based bayesian neural networks. *Neural Networks*, 144:394–406, 2021.
- [196] Chio Lam and Wenxing Zhou. Statistical analyses of incidents on onshore gas transmission pipelines based on phmsa database. *International Journal of Pressure Vessels and Piping*,

145:29–40, 2016.

- [197] Yao Yao, Shue-Ting Ellen Tung, and Branko Glisic. Crack detection and characterization techniques—an overview. *Structural Control and Health Monitoring*, 21(12):1387–1413, 2014.
- [198] Shiuh-Chuan Her and Sheng-Tung Lin. Non-destructive evaluation of depth of surface cracks using ultrasonic frequency analysis. *Sensors*, 14(9):17146–17158, 2014.
- [199] K Kimoto, S Ueno, and S Hirose. Image-based sizing of surface-breaking cracks by sh-wave array ultrasonic testing. *Ultrasonics*, 45(1-4):152–164, 2006.
- [200] Sony Baby, T Balasubramanian, R J Pardikar, M Palaniappan, and R Subbaratnam. Time-of-flight diffraction (TOFD) technique for accurate sizing of surface-breaking cracks. *Insight - Non-Destr. Test. Cond. Monit.*, 45(6):426–430, June 2003.
- [201] Weiying Cheng and Kenzo Miya. Reconstruction of parallel cracks by ECT. *Int. J. Appl. Electromagn. Mech.*, 14(1-4):495–502, December 2002.
- [202] Yiming Deng and Xin Liu. Electromagnetic imaging methods for nondestructive evaluation applications. *Sensors*, 11(12):11774–11808, 2011.
- [203] Maryam Ravan, Reza K Amineh, Slawomir Koziel, Natalia K Nikolova, and James P Reilly. Sizing of multiple cracks using magnetic flux leakage measurements. *IET science, measurement & technology*, 4(1):1–11, 2010.
- [204] Shamim Ahmed, Roberto Miorelli, Pierre Calmon, Nicola Anselmi, and Marco Salucci. Real time flaw detection and characterization in tube through partial least squares and svr: Application to eddy current testing. In *AIP conference proceedings*, volume 1949. AIP Publishing, 2018.
- [205] Kharudin Bin Ali, Ahmed N Abdalla, Damhuji Rifai, and Moneer A Faraj. Review on system development in eddy current testing and technique for defect classification and characterization. *IET Circuits, Devices & Systems*, 11(4):338–351, 2017.
- [206] PA Ivanov, V Zhang, CH Yeoh, H Udpa, Y Sun, SS Udpa, and W Lord. Magnetic flux leakage modeling for mechanical damage in transmission pipelines. *IEEE Transactions on magnetics*, 34(5):3020–3023, 1998.
- [207] Ameet Joshi, Lalita Udpa, Satish Udpa, and Antonello Tamburrino. Adaptive wavelets for characterizing magnetic flux leakage signals from pipeline inspection. *IEEE transactions on magnetics*, 42(10):3168–3170, 2006.
- [208] S Mukhopadhyay and GP Srivastava. Characterisation of metal loss defects from magnetic flux leakage signals with discrete wavelet transform. *Ndt & E International*, 33(1):57–65,

2000.

- [209] Gianni D’Angelo and Salvatore Rampone. Shape-based defect classification for non destructive testing. In *2015 IEEE Metrology for Aerospace (MetroAeroSpace)*, pages 406–410. IEEE, 2015.
- [210] Wei Lu, Yuning Wei, Jinxia Yuan, Yiming Deng, and Aiguo Song. Tractor assistant driving control method based on eeg combined with rnn-tl deep learning algorithm. *IEEE Access*, 8:163269–163279, 2020.
- [211] Peipei Zhu, Yuhua Cheng, Portia Banerjee, Antonello Tamburrino, and Yiming Deng. A novel machine learning model for eddy current testing with uncertainty. *ndt & e International*, 101:104–112, 2019.
- [212] David JC MacKay. Bayesian neural networks and density networks. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 354(1):73–80, 1995.
- [213] Andrey Malinin. *Uncertainty estimation in deep learning with application to spoken language assessment*. PhD thesis, University of Cambridge, 2019.
- [214] T Azizzadeh and MS Safizadeh. Investigation of the lift-off effect on the corrosion detection sensitivity of three-axis mfl technique. *Journal of Magnetism*, 23(2):152–159, 2018.
- [215] J Bruce Nestleroth and Richard J Davis. The effects of magnetizer velocity on magnetic flux leakage signals. *Review of progress in quantitative nondestructive evaluation: volumes 12A and 12B*, pages 1891–1898, 1993.
- [216] Ali Mirzaee, Sina Zahedifard, Iman Ahadi Akhlaghi, and Saeed Kahrobaee. Application of magnetic flux leakage (mfl) method to non-destructively characterize the microstructure and corrosion behaviour of api x65 grade steel. *Journal of Magnetism and Magnetic Materials*, 566:170311, 2023.
- [217] S Jäggi, H Böhni, and B Elsener. Macrocell corrosion of steel in concrete-experiments and numerical modelling. *European Federation of Corrosion*, 2001.
- [218] Muhammad Faheem, Syed Bilal Hussain Shah, Rizwan Aslam Butt, Basit Raza, Muhammad Anwar, Muhammad Waqar Ashraf, Md A Ngadi, and Vehbi C Gungor. Smart grid communication and information technologies in the perspective of industry 4.0: Opportunities and challenges. *Computer Science Review*, 30:1–30, 2018.
- [219] Yasi Wang, Hongxun Yao, and Sicheng Zhao. Auto-encoder based dimensionality reduction. *Neurocomputing*, 184:232–242, 2016.
- [220] Changro Lee. Data augmentation using a variational autoencoder for estimating property

- prices. *Property Management*, 39(3):408–418, 2021.
- [221] Lovedeep Gondara. Medical image denoising using convolutional denoising autoencoders. In *2016 IEEE 16th international conference on data mining workshops (ICDMW)*, pages 241–246. IEEE, 2016.
 - [222] Kasra Babaei, ZhiYuan Chen, and Tomas Maul. Data augmentation by autoencoders for unsupervised anomaly detection. *arXiv preprint arXiv:1912.13384*, 2019.
 - [223] Herim Han and Sunghwan Choi. Transfer learning from simulation to experimental data: Nmr chemical shift predictions. *The Journal of Physical Chemistry Letters*, 12(14):3662–3668, 2021.
 - [224] Matthew Welborn, Lixue Cheng, and Thomas F Miller III. Transferability in machine learning for electronic structure via the molecular orbital basis. *Journal of chemical theory and computation*, 14(9):4772–4779, 2018.
 - [225] Yifan Peng, Shankai Yan, and Zhiyong Lu. Transfer learning in biomedical natural language processing: an evaluation of bert and elmo on ten benchmarking datasets. *arXiv preprint arXiv:1906.05474*, 2019.
 - [226] Keunwoo Choi, György Fazekas, Mark Sandler, and Kyunghyun Cho. Transfer learning for music classification and regression tasks. *arXiv preprint arXiv:1703.09179*, 2017.
 - [227] Moloud Abdar, Farhad Pourpanah, Sadiq Hussain, Dana Rezazadegan, Li Liu, Mohammad Ghavamzadeh, Paul Fieguth, Xiaochun Cao, Abbas Khosravi, U Rajendra Acharya, et al. A review of uncertainty quantification in deep learning: Techniques, applications and challenges. *Information Fusion*, 76:243–297, 2021.
 - [228] Weiyang Liu, Yandong Wen, Zhiding Yu, and Meng Yang. Large-margin softmax loss for convolutional neural networks. *arXiv preprint arXiv:1612.02295*, 2016.
 - [229] Alexandru Niculescu-Mizil and Rich Caruana. Predicting good probabilities with supervised learning. In *Proceedings of the 22nd international conference on Machine learning*, pages 625–632, 2005.
 - [230] Jeremy Nixon, Michael W Dusenberry, Linchuan Zhang, Ghassen Jerfel, and Dustin Tran. Measuring calibration in deep learning. In *CVPR workshops*, volume 2, 2019.
 - [231] Chris Ding and Hanchuan Peng. Minimum redundancy feature selection from microarray gene expression data. *Journal of bioinformatics and computational biology*, 3(02):185–205, 2005.
 - [232] Ronald A Fisher. The use of multiple measurements in taxonomic problems. *Annals of eugenics*, 7(2):179–188, 1936.

- [233] John Brausch, Lawrence Butkus, David Campbell, Tommy Mullis, and Michael Paulk. Recommended processes and best practices for nondestructive inspection (ndi) of safety of flight structures'. Technical report, Report No. AFRL-RX-WP-TR-2008-4373, Air Force Research Laboratory, USA, 2008.
- [234] Alan P Berens. Evaluation of nde reliability characterization. *Report No. AFWAL-TR-81-4160*, 1981.
- [235] Peter W Hovey and Alan P Berens. Statistical evaluation of nde reliability in the aerospace industry. In *Review of Progress in Quantitative Nondestructive Evaluation: Volume 7B*, pages 1761–1768. Springer, 1988.
- [236] F Fücsök, C Müller, and M Scharmach. Measuring of the reliability of nde. In *8th International Conference of the Slovenian Society for Non-Destructive Testing „Application of Contemporary Non-Destructive Testing in Engineering, Portorož, Slovenia, September*, pages 1–3, 2005.
- [237] Eric Lindgren, David Forsyth, John Aldrin, and Floyd Spencer. Probability of detection. 2018.
- [238] Iikka Virkkunen, Tuomas Koskinen, Suvi Papula, Teemu Sarikka, and Hannu Hänninen. Comparison of $\hat{a}$ versus a and hit/miss pod-estimation methods: A european viewpoint. *Journal of Nondestructive Evaluation*, 38:1–13, 2019.
- [239] Arvind Keprate and RM Chandima Ratnayake. Probability of detection as a metric for quantifying nde capability: the state of the art. *J. Pipeline Eng*, 14(3):199–209, 2015.
- [240] GL DNV. Rp-0001: Probabilistic methods for planning of inspection for fatigue cracks in offshore structures. *Det Norske Veritas, Høvik, Norway*, 2015.
- [241] Fengyang Jiang, GUAN Zhidong, LI Zengshan, and WANG Xiaodong. A method of predicting visual detectability of low-velocity impact damage in composite structures based on logistic regression model. *Chinese Journal of Aeronautics*, 34(1):296–308, 2021.
- [242] CA Harding and GR Hugo. Review of literature on probability of detection for liquid penetrant nondestructive testing. 2011.
- [243] SK Burke and RJ Ditchburn. Review of literature on probability of detection for magnetic particle nondestructive testing. *Department of Defence, Australia*, 2013.
- [244] Jochen H Kurz, Anne Jüngert, Sandra Dugan, Gerd Dobmann, and Christian Boller. Reliability considerations of ndt by probability of detection (pod) determination using ultrasound phased array. *Engineering failure analysis*, 35:609–617, 2013.
- [245] Junzhen Zhu, Qingxu Min, Jianbo Wu, and Gui Yun Tian. Probability of detection for

- eddy current pulsed thermography of angular defect quantification. *IEEE Transactions on Industrial Informatics*, 14(12):5658–5666, 2018.
- [246] Michael Wright. How to implement a pod into a highly effective inspection strategy. *NDT Canada*, pages 15–17, 2016.
 - [247] Ryan M Meyer, Susan L Crawford, John P Lareau, and Michael T Anderson. Review of literature for model assisted probability of detection. 2014.
 - [248] R Bruce Thompson, Lisa J Brasche, Eric Lindgren, Paul Swindell, and William P Winfree. Recent advances in model-assisted probability of detection. In *4th European-American workshop on reliability of NDE*, number LF99-9094, 2009.
 - [249] CA Harding, GR Hugo, and SJ Bowles. Application of model-assisted pod using a transfer function approach. In *AIP conference proceedings*, volume 1096, pages 1792–1799. American Institute of Physics, 2009.
 - [250] F Jenson, N Dominguez, P Willaume, and T Yalamas. A bayesian approach for the determination of pod curves from empirical data merged with simulation results. In *AIP Conference Proceedings*, volume 1511, pages 1741–1748. American Institute of Physics, 2013.
 - [251] M Wall, FA Wedgwood, and S Burch. Modelling of ndt reliability (pod) and applying corrections for human factors. In *Proceedings of the 7th European Conference on NDT, Copenhagen, Denmark*, 1998.
 - [252] Sarah Muscat, Stuart Parks, Ewan Kemp, and David Keating. Repeatability and reproducibility of macular thickness measurements with the humphrey oct system. *Investigative ophthalmology & visual science*, 43(2):490–495, 2002.
 - [253] Barry N Taylor. *Guidelines for Evaluating and Expressing the Uncertainty of NIST Measurement Results* (rev. Diane Publishing, 2009).
 - [254] Stephanie A Bell. A beginner’s guide to uncertainty of measurement. 2001.
 - [255] United Kingdom Accreditation Service. *The expression of uncertainty and confidence in measurement*. United Kingdom Accreditation Service, 1997.
 - [256] Morana Mihaljević, Damir Markučić, Biserka Runje, and Zdenka Keran. Measurement uncertainty evaluation of ultrasonic wall thickness measurement. *Measurement*, 137:179–188, 2019.
 - [257] Mohand Alzuhiri, Zi Li, Adithya Rao, Jiaoyang Li, Preston Fairchild, Xiaobo Tan, and Yiming Deng. Imu-assisted robotic structured light sensing with featureless registration under uncertainties for pipeline inspection. *NDT & E International*, page 102936, 2023.

- [258] Robert Tomkowski, Aki Sorsa, Suvi Santa-aho, Per Lundin, and Minnamari Vippola. Statistical evaluation of barkhausen noise testing (bnt) for ground samples. *Sensors*, 19(21):4716, 2019.
- [259] Shuang Li, Zhongyu Wang, Jingyu Guan, and Jihu Wang. Uncertainty evaluation in surface structured light measurement. In *2021 IEEE 15th International Conference on Electronic Measurement & Instruments (ICEMI)*, pages 395–400. IEEE, 2021.
- [260] Fernando J Alamos, Jiahui C Gu, and Hyunok Kim. Evaluating the reliability of a non-destructive evaluation (nde) tool to measure the incoming sheet mechanical properties. In *Forming the Future: Proceedings of the 13th International Conference on the Technology of Plasticity*, pages 2573–2584. Springer, 2021.
- [261] Matthew R Cherry, Harold S Sabbagh, Adam L Pilchak, Jeremy S Knopp, and DAYTON UNIV RESEARCH INST OH. Characterization of a random anisotropic conductivity field with karhunen-loeve methods (postprint). 2013.
- [262] Yung-Li Lee, Mark E Barkey, and Hong-Tae Kang. *Metal fatigue analysis handbook: practical problem-solving techniques for computer-aided engineering*. Elsevier, 2011.
- [263] Kwai S Chan. Roles of microstructure in fatigue crack initiation. *International Journal of Fatigue*, 32(9):1428–1447, 2010.
- [264] Moorthy Vaidhianathasamy, Brian Andrew Shaw, Will Bennett, and Peter Hopkins. Evaluation of contact fatigue damage on gears using the magnetic barkhausen noise technique. In *Proc. 12th Int. Workshop Electromag. Nondestruct. Eval.*, pages 98–105, 2008.
- [265] Jose Alberto Perez-Benitez, J Capó-Sánchez, J Anglada-Rivera, and LR Padovese. A model for the influence of microstructural defects on magnetic barkhausen noise in plain steels. *Journal of magnetism and magnetic materials*, 288:433–442, 2005.
- [266] Sadegh Ghanei, M Kashefi, and Mohammad Mazinani. Comparative study of eddy current and barkhausen noise nondestructive testing methods in microstructural examination of ferrite–martensite dual-phase steel. *Journal of magnetism and magnetic materials*, 356: 103–110, 2014.
- [267] Krzysztof Miesowicz, Wiesław J Staszewski, and Tomasz Korbiel. Analysis of barkhausen noise using wavelet-based fractal signal processing for fatigue crack detection. *International Journal of Fatigue*, 83:109–116, 2016.
- [268] O Stupakov, J Pal’a, Toshiyuki Takagi, and Tetsuya Uchimoto. Governing conditions of repeatable barkhausen noise response. *Journal of Magnetism and Magnetic Materials*, 321 (18):2956–2962, 2009.
- [269] Shuo Zhang, Xiaodong Shi, Lalita Udpa, and Yiming Deng. Micromagnetic measurement

- for characterization of ferromagnetic materials' microstructural properties. *AIP Advances*, 8(5):056614, 2018.
- [270] M Vashista and V Moorthy. On the shape of the magnetic barkhausen noise profile for better revelation of the effect of microstructures on the magnetisation process in ferritic steels. *Journal of Magnetism and Magnetic Materials*, 393:584–592, 2015.
- [271] Svante Wold, Kim Esbensen, and Paul Geladi. Principal component analysis. *Chemometrics and intelligent laboratory systems*, 2(1-3):37–52, 1987.
- [272] P Burrascano, E Cardelli, A Faba, S Fiori, and A Massinelli. Application of probabilistic neural networks to eddy current non destructive test problems. In *EANN 2001 Conference*, pages 16–18, 2001.
- [273] S Kanemoto. Acoustic monitoring using kernel pca and probabilistic neural network. In *7th International Conference on NDE*, 2009.
- [274] Christian P Vetter, Laura A Kuebel, Divya Natarajan, and Ray A Mentzer. Review of failure trends in the us natural gas pipeline industry: An in-depth analysis of transmission and distribution system incidents. *Journal of Loss Prevention in the Process Industries*, 60: 317–333, 2019.
- [275] Julie Maupin, Michael Mamoun, et al. Plastic pipe failure, risk, and threat analysis. Technical report, Gas Technology Institute, 2009.
- [276] Juanjuan Zhu, Richard P Collins, Joby B Boxall, Robin S Mills, and Rob Dwyer-Joyce. Non-destructive in-situ condition assessment of plastic pipe using ultrasound. *Procedia engineering*, 119:148–157, 2015.
- [277] Qiang Wang, Haiting Zhou, Junwei Xie, and Xiaomeng Xu. Nonlinear ultrasonic evaluation of high-density polyethylene natural gas pipe thermal butt fusion joint aging behavior. *International Journal of Pressure Vessels and Piping*, 189:104272, 2021.
- [278] Tobias D Carrigan, Benjamin E Forrest, Hector N Andem, Kaiyu Gui, Lewis Johnson, James E Hibbert, Barry Lennox, and Robin Sloan. Nondestructive testing of nonmetallic pipelines using microwave reflectometry on an in-line inspection robot. *IEEE Transactions on Instrumentation and Measurement*, 68(2):586–594, 2018.
- [279] Andri Haryono, Mohamed A Abou-Khousa, et al. Microwave non-destructive evaluation of glass reinforced epoxy and high density polyethylene pipes. *Journal of Nondestructive Evaluation*, 39(1):1–9, 2020.
- [280] R Kafieh, T Lotfi, and Rassoul Amirfattahi. Automatic detection of defects on polyethylene pipe welding using thermal infrared imaging. *Infrared Physics & Technology*, 54(4):317–325, 2011.

- [281] Marjan Doaei and M Sadegh Tavallali. Intelligent screening of electrofusion-polyethylene joints based on a thermal ndt method. *Infrared Physics & Technology*, 90:1–7, 2018.
- [282] Cong Li, Hui-Qing Lan, Ya-Nan Sun, and Jun-Qiang Wang. Detection algorithm of defects on polyethylene gas pipe using image recognition. *International Journal of Pressure Vessels and Piping*, 191:104381, 2021.
- [283] Mansour Karkoub, Othmane Bouhali, and Ali Sheharyar. Gas pipeline inspection using autonomous robots with omni-directional cameras. *IEEE Sensors Journal*, 21(14):15544–15553, 2020.
- [284] Portia Banerjee, Rajendra Prasath Palanisamy, Mahmood Haq, Lalita Udpa, and Yiming Deng. Data-driven prognosis of fatigue-induced delamination in composites using optical and acoustic nde methods. In *2019 IEEE International Conference on Prognostics and Health Management (ICPHM)*, pages 1–10. IEEE, 2019.
- [285] Mohand Alzuhiri, Khalid Farrag, Ernest Lever, and Yiming Deng. An electronically stabilized multi-color multi-ring structured light sensor for gas pipelines internal surface inspection. *IEEE Sensors Journal*, 2021.
- [286] Christoph Schmalz, Frank Forster, Anton Schick, Elli Angelopoulou, Frank Forster, Anton Schick, and Elli Angelopoulou. An endoscopic 3D scanner based on structured light. *Medical Image Analysis*, 2012.
- [287] Tzu-Yi Chuang and Cheng-Che Sung. Learning and slam based decision support platform for sewer inspection. *Remote Sensing*, 12(6):968, 2020.
- [288] Shrijan Kumar. Development of slam algorithm for a pipe inspection serpentine robot. Master’s thesis, University of Twente, 2019.
- [289] Dennis Kryz and Homayoun Najjaran. Development of visual simultaneous localization and mapping (vslam) for a pipe inspection robot. In *2007 International Symposium on Computational Intelligence in Robotics and Automation*, pages 344–349. IEEE, 2007.
- [290] R Zhang, MH Evans, R Worley, SR Anderson, and L Mihaylova. Improving slam in pipe networks by leveraging cylindrical regularity. In *Annual Conference Towards Autonomous Robotic Systems*, pages 56–65. Springer, 2021.
- [291] Andreu Corominas Murtra and Josep M Mirats Tur. Imu and cable encoder data fusion for in-pipe mobile robot localization. In *2013 IEEE Conference on Technologies for Practical Robot Applications (TePRA)*, pages 1–6. IEEE, 2013.
- [292] Chi-Keung Tang. *Tensor voting in computer vision, visualization, and higher dimensional inferences*. University of Southern California, 2000.

- [293] Jason Geng. Structured-light 3d surface imaging: a tutorial. *Advances in Optics and Photonics*, 3(2):128–160, 2011.
- [294] Paul J Besl and Neil D McKay. Method for registration of 3-d shapes. In *Sensor fusion IV: control paradigms and data structures*, volume 1611, pages 586–606. Spie, 1992.
- [295] Rachid Tiar, Mustapha Lakrouf, and Ouahiba Azouaoui. Fast icp-slam for a bi-steerable mobile robot in large environments. In *2015 IEEE International Workshop of Electronics, Control, Measurement, Signals and their Application to Mechatronics (ECMSM)*, pages 1–6. IEEE, 2015.
- [296] Wilian França Costa, Jackson P Matsuura, Fabiana Soares Santana, and Antonio Mauro Saraiva. Evaluation of an icp based algorithm for simultaneous localization and mapping using a 3d simulated p3dx robot. In *2010 Latin American Robotics Symposium and Intelligent Robotics Meeting*, pages 103–108. IEEE, 2010.
- [297] Yue Wang, Rong Xiong, and Qianshan Li. Em-based point to plane icp for 3d simultaneous localization and mapping. *Int J Rob Autom*, 28:234–244, 2013.
- [298] Rakesh Kumar Sidharthan, Ramkumar Kannan, Seshadhri Srinivasan, and Valentina Emilia Balas. Stochastic wheel-slip compensation based robot localization and mapping. *Advances in Electrical and Computer Engineering*, 16(2):25–32, 2016.