Bournemouth University

Adaptive Simulation Modelling Using The Digital Twin Paradigm

A Thesis

Submitted to

The Academic Faculty of

Faculty of Science & Technology

by

Madhu Sudan Sapkota

In partial fulfillment

of the requirements for the

degree of Doctor of Philosophy

February 2023

Copyright Statement

Abstract

Structural Health Monitoring (SHM) involves the application of qualified standards, by competent people, using appropriate processes and procedures throughout the structure's life cycle, from design to decommissioning. The main goal is to ensure that through an ongoing process of risk management, the structure's continued fitness-for-purpose (FFP) is maintained – allowing for optimal use of the structure with a minimal chance of downtime and catastrophic failure.

While undertaking the SHM task, engineers use model(s) to predict the risk to the structure from degradation mechanisms such as corrosion and cracking. These predictive models are either physics-based, data-driven or hybrid based. The process of building these predictive models tends to involve processing some input parameters related to the material properties (e.g.: mass density, modulus of elasticity, polarisation current curve, etc) or/and the environment, to calibrate the model and using them for the predictive simulation. So, the accuracy of the predictions is very much dependent upon the input data describing the properties of the materials and/or the environmental conditions the structure experiences.

For the structure(s) with non-uniform and complex degradation behaviour, this process is repeated over the life-time of the structure(s), i.e., when each new survey is performed (or new data is available) and then the survey data are used to infer changes in the material or environmental properties. This conventional parameter tuning and updating approach is computationally expensive and time-consuming, as multi-simulations are needed and manual intervention is expected to determine the optimal model parameters. There is therefore a need for a fundamental paradigm shift to address the shortcomings of conventional approaches. The Digital Twin (DT) offers such a paradigm shift in that it integrates ultra-high fidelity simulation model(s) with other related structural data, to mirror the structural behaviour of its corresponding physical twin. DT's inherent ability to handle large data allows for the inclusion of an evolving set of data relating to the structure with time as well as provides for the adaptation of the simulation model with very little need for human intervention.

This research project investigated DT as an alternative to the existing model cali-

bration and adaptation approach. It developed a design of experiment platform for online model validation and adaptation (i.e., parameter updating) solver(s) within the Digital Twin paradigm. The design of experimental platform provided a basis upon which an approach based on the creation of surrogates and reduced order model (ROM)-assisted parameter search were developed for improving the efficiency of model calibration and adaptation. Furthermore, the developed approach formed a basis for developing solvers which provide for the self-calibration and self-adaptation capability required for the prediction and analysis of an asset's structural behaviour over time.

The research successfully demonstrated that such solvers can be used to efficiently calibrate ultra-high-fidelity simulation model within a DT environment for the accurate prediction of the status of a real-world engineering structure.

Contents

List of Figures

List of Tables

Acknowledgements

This PhD research journey including composing this dissertation has been a challenging but ultimately fulfilling undertaking. Unforeseen obstacles resulting from the COVID pandemic added a unique dimension to the experience, and I could not have completed the journey without the support of others. The encouragement and faith of certain beautiful souls were invaluable to me, both in my professional and personal life, and I would like to express my gratitude to them for their valuable contributions throughout this endeavour.

First and foremost, I would like to express my heartfelt appreciation to my esteemed academic mentors, Dr. Mark Hadfield, Dr. Vasilis Katos, Dr. Roya Haratian and Dr. Edward Apeh for generously providing their time and offering regular feedback throughout my tenure. Their unwavering support and guidance were instrumental in enabling me to undertake this endeavour. I am particularly indebted to Dr. Edward, who went above and beyond the role of an academic advisor, providing me with invaluable advice, steadfast encouragement, and valuable mentorship throughout the challenging situations during the journey.

I would like to express my gratitude to CM Beasy Ltd, the collaborating company that provided technical support, experimentation, and feedback throughout the course of this research project. In particular, I extend my appreciation to Robert Adey and John Baynham for their inspiration and insightful advice, which have proven invaluable.

The collaborative efforts of both the University and Industrial teams have been crucial in achieving the highest quality of work, resulting in exceptional outcomes in this project.

Above all, I am deeply grateful to my mom, whose unwavering love, faith, and sacrifices have been the foundation of my achievements. I will remain forever indebted to her for everything she has provided for me.

Last but not least, I want to express my heartfelt gratitude to my caring life partner, my friends, and family members, including my sisters and in-laws, for always being there for me. Their consistent support has made my studies and life much easier, despite being away from my home country. I would also like to share this special moment with my nephews and nieces, whose presence has been a source of comfort and joy during difficult times. I send my love and sincere appreciation to all of them.

Original Work Declaration

This thesis is based on my own research activities, except where stated, in accordance with University regulations.

1 Introduction

1.1 Overview

The engineering infrastructures (assets), including bridges, buildings, industrial plants, offshore structures, and others, play a vital role in the functioning of societies including private and public business. While most of these structures are typically designed to withstand expected loading and environmental conditions during their lifespan, unforeseen uncertainties can still arise that do not follow the design rules. Therefore, it is crucial to have effective risk assessment procedures in place to mitigate potential risks associated with such uncertainties and ensure timely maintenance (Brownjohn 2007). Moreover, effective risk assessment will allow for optimal usage of the structure with minimal chance of downtime or catastrophic failure by maintaining their fitness-for-purpose (FFP).

Structural Health Monitoring (SHM) technologies aim to assess the risks associated with engineering infrastructures (assets) during their operational phase (Farrar and Worden 2007a). To achieve this, predictive models are frequently used in SHM, which can also be utilised further for system optimisation, design control, and other related purposes. These models can be physics-based (Gopalakrishnan et al. 2011), data-driven (Azimi et al. 2020), or hybrid (Chao et al. 2022). The physics-based model makes predictions by capturing the dynamic processes of the system through solving equations that describe the underlying dynamics, whereas the data-driven model relies on previous system data to assess the current and future state of damage. Hybrid models that combine both approaches can leverage the strengths of each. While data-driven models are popular in many domains due to their lower computational complexity and the abundance of data available, SHM-related domains still rely on physics-based models (Malveiro et al. 2018, An et al. 2019). This is because the SHM data available is often limited, and structures can undergo varying new scenarios with different patterns of depletion and deterioration, making it difficult to track these patterns with limited data. Moreover, physics-based models are capable of providing insight into the situations type that have not yet occurred, which is critical for predicting and preventing structural failures in SHM-related applications (Farrar and Lieven 2007b).

To rely on any type of model for predictive analysis, including risk assessment, the credibility of the model is essential. For a physics-based model, credibility depends on its accuracy in emulating real-world dynamics and the use of the most precise input variables to run the simulation. The physics-based model building process in various domains is often aided by the availability of commercial process simulators. These simulators can simulate the real-world dynamics for designated processes with acceptable accuracy (Brynjarsdóttir and O'Hagan 2014). Using such simulators, parametric simulation models for real systems are frequently constructed. However, before these models can be used, they must be calibrated by providing the best set of material and environment-related input parameters (e.g.: mass density, modulus of elasticity, polarisation current curve, etc) so that the model accurately represents the real structure or asset (Agami Reddy 2006).

The commonly used calibration approaches for structural models are often computationally expensive, time-consuming, and may not be effective for real-world structures due to their complexity and variability. The parameter estimation and updating task presents challenges related to obtaining real-time data, selecting appropriate algorithms, and understanding the complexity associated with parameters. Although technology has led to more precise and cost-efficient data acquisition in many domains, data related to SHM often remain unstructured and obtained from inspections and surveys (Gong et al. 2016, Gulgec et al. 2017, Barni et al. 2018). Furthermore, there is a lack of a standardised approach to determine the quality, quantity, and variability of data required from sensors (surveys) to ensure model performance during calibration (Fabrizio and Monetti 2015, Kang et al. 2021). Additionally, the traditional trial-and-error calibration approach relies on manual involvement, and multiple simulation runs are required before determining the best set of parameters (Kim et al. 2019, Cao et al. 2020, Silva et al. 2021). These issues can lead to delays in developing a credible predictive model that represents a real-time dynamics for the structure, and also increase the risk of the model being out of sync.

To address the challenges of parametric calibration in developing a credible model utilising simulator(s), there is a need for automated data extraction from unstructured data sources and fully automate the calibration task. Additionally, an appropriate framework for benchmarking the validation data during calibration, as well as an efficient parametric calibration strategy, are required to ensure an accurate and reliable model.

Maintaining the predictive ability of the calibrated model presents another significant challenge, particularly in the context of structure-related models. The model often diverges from the true behaviour of the asset over time due to complex changes in the material properties of the structure, which can vary even among seemingly similar components of the structures (Sohn 2007, Sehgal and Kumar 2016). As a result, the engineer must repeatedly re-calibrate (i.e., adapt) the model throughout the structure's operational lifespan when each new survey is performed (or new data is available) to infer changes in the material and/or environmental properties.

Despite significant progress in physics-based modelling, the challenges in calibration and adaptation hinder the development of practical and durable prognostic tools for real-world structures. The need for frequent adaptation of the model demands a self-adaptive model that is supported by analytic (e.g.: Machine-Learning tools) and able to account for the evolving material properties of the structure over time (Sohn 2007, Gabor et al. 2016). This paradigm shift would overcome the limitations of conventional approaches to model calibration and adaptation.

The concept of Digital Twin (DT) is an evolving concept in the modelling world that represents a virtual model replicating the real-time behaviour of an existing system, which is referred to as its physical twin (Glaessgen and Stargel 2012, Boschert and Rosen 2016, Ye et al. 2020). Moreover, the DT concept involves a comprehensive model that combines an ultra-high fidelity simulation tool with physical twin-related data to provide the online simulation of the physical twin (Barricelli et al. 2019, Rasheed et al. 2020).

Therefore, this research aimed to explore the potential of DT in addressing challenges related to model calibration and adaptation. The findings suggest that although the concept of DT has been widely adopted in manufacturing for over a decade, its applicability in SHM is gaining momentum in recent years only (Seshadri and Krishnamurthy 2017, Tao et al. 2018, Ye et al. 2020). One of the reasons for this is that the practical limitations of the DT concept, particularly regarding the lack of an appropriate framework for its application within the context of SHM (Barricelli et al. 2019, Aivaliotis et al. 2019b, Broo et al. 2022).

This research therefore leveraged existing DT related features and technologies to address the model calibration and adaptation related challenges, and aimed to improve its characteristics, for which suitable frameworks are still lacking.

1.2 Research Goals and Scopes

The primary objective of this research was to address issues related to model calibration and adaptation using the Digital Twin paradigm. This involved leveraging the inherent capabilities of DTs and identifying areas where they may fall short in terms of parameter-related calibration and adaptation.

The model calibration and adaptation related challenges demand various solution approaches and/or frameworks, while the DT concept can provide a platform that can incorporate these frameworks when available. The DT concept allows for a holistic approach to reach the solution, utilising the benefits among the frameworks to assist each other.

In line with the DT concept, research fields were categorised based on gaps in the calibration and adaptation characteristics of DT, as well as other related DT features. From this categorisation, five key topics were selected for further research activities. Then research objectives and milestones were set for each of the research areas that are in accordance with the problems and scopes corresponding to the research areas (Figure 1.1).

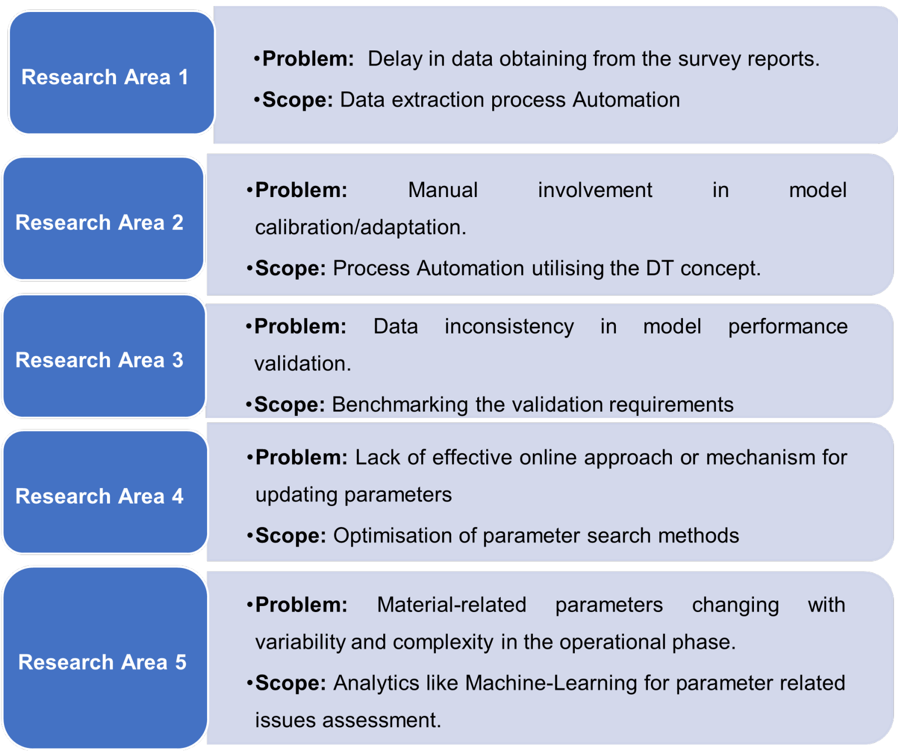


Figure 1.1: Research areas categorised to address the research challenges categorically.

Following the establishment of objectives and milestones, further research activities were undertaken. These include analysis, development/adoption/design, and implementation of various strategies to address research gaps in each area.

1.3 Contributions to the Field

This research has been able to make the following contributions to the field of SHM-related modelling and DT:

1. An Ontology and Natural Language Processing based data extracting framework to obtain possible near real-time data from the survey reports (discussed in Section 4.2).
2. An Integrated Platform for online simulation model validation and parameter updating, eventually also assisting in the automation of DT enabling (Sapkota et al. 2021a).
3. An approach to benchmark the data and the metric requirements before tailoring a parametric model to represent the real physical asset's behaviour (Sapkota et al. 2021b).
4. The surrogate-assisted model calibration approach utilising the DT concept that ultimately reduces the number of required simulation cases for optimisation-assisted parameter estimation. This approach also assists in reducing the issue of local minima during optimisation-based calibration (Sapkota et al. 2022a).
5. A ROM and modularisation-assisted approach to deal with the problem of parametric spatial variability arising with time (Sapkota et al. 2022b).
6. Machine-Learning (M/L) in trend capturing, offering a substitute for the physics-based model during the operational lifetime of DT (discussed in Section 8.4).

The developed codes to have the above-mentioned outcomes up to their current development are available in the Adaptive-Digital Twin repository at https://github.com/Adaptive-DigitalTwin.

The outcomes of this research has provided solutions for model-related challenges in parametric calibration and adaptation, and also benefit the practical implementation of DT within SHM. Furthermore, successful deployment of the research-adopted DT concept (including architecture) has established it as an enhanced adaptive modelling concept. The DT architecture together with the outcomes (approaches and frameworks) from this research could also be considered as a contribution to the standardisation of SHM-related DT.

1.4 List of Publications

The outcomes and contributions of the research so far have been published as follows (including papers in progress):

1. Sapkota, M. S., Apeh, E.., Hadfield, M., Adey, R., & Baynham, J. (2021a). “Design of experiments platform for online simulation model validation and parameter updating within digital twinning “. *WIT Transactions on Engineering Sciences, 130(12)*, 3-14.

2. Sapkota, M. S., Apeh, E., Hadfield, M., Haratian, R., Adey, R., & Baynham, J. (2021b). “An approach for adaptive model performance validation within digital twinning”. *International Journal of Computational Methods and Experimental Measurements, 9(3)*, 213-225

3. Sapkota, M. S., Apeh, E., Hadfield, M., Haratian, R., Adey, R., & Baynham, J. (2022a). “Surrogate-Assisted Parametric Calibration Using Design Of Experiment Platform Within Digital Twinning”. *International Journal of Computational Methods and Experimental Measurements, 10(2)*, pp.158-171.

4. Sapkota, M.S., Apeh, E., Adey, R., Baynham, J. & Peratta, C., (2022b) “Improving Integrity Management Decision Making Using Automated Calibration of Model Predictions from Cathodic-Protection (CP) Survey Data.” *In AMPP Annual Conference+ Expo. OnePetro,* 2022

5. Sapkota, M.S., & Apeh, E. (2023a). “An Ontology and Natural Language Processing based data extracting framework within Digital Twinning”. (*In progress*)

6. Sapkota, M. S., Apeh, E., Hadfield, M., Haratian, R., Adey, R., & Baynham, J. (2023b). “Machine Learning in Simulator-based Cathodic-Protection Digital Twin Adaptation and Online prediction”. (*In progress*)

7. Sapkota, M.S., Apeh, E., Adey, R., Baynham, J. & Peratta, C., (2023) “Cathodic-Protection (CP) Digital-Twin with automated calibration for Integrity Related Decision Making.” (*In Progress for Material’s Performance*)

1.5 Thesis Structure

The rest of this thesis is outlined as follows (Figure 1.2).

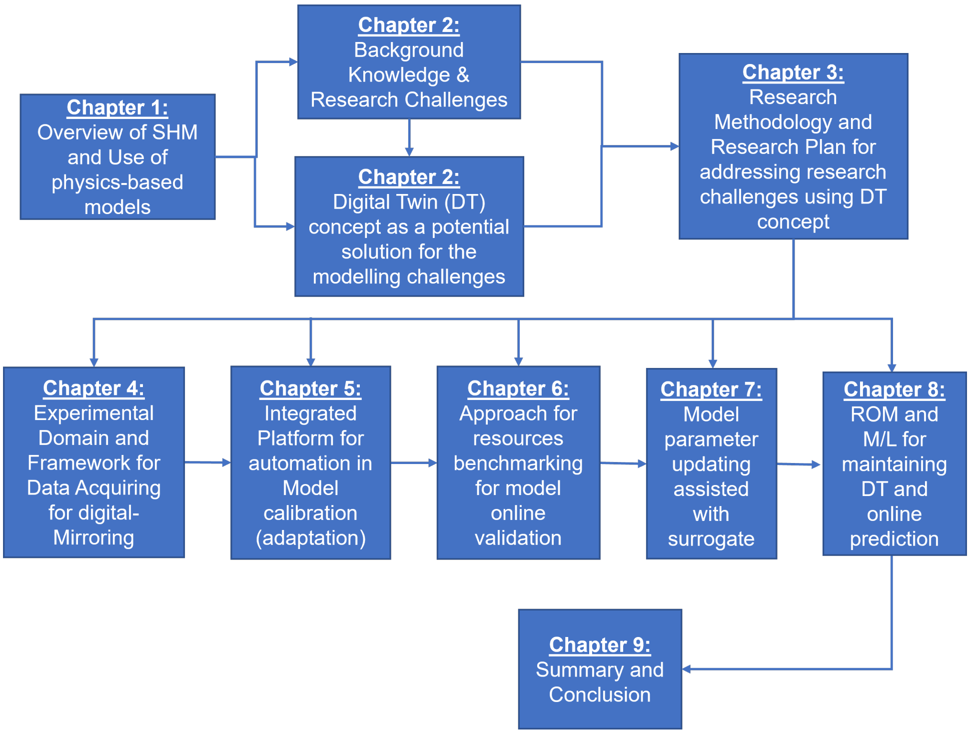


Figure 1.2: Representation of the thesis structure

Chapter 2 provides an overview of *Predictive Simulation in Structural Health Monitoring* (SHM) together with the modelling-related challenges, particularly during the calibration and/or adaptation phase. For this, the techniques for building models, with a focus on those related to structural health, are surveyed. Additionally, the chapter explores the sources of parametric uncertainty and examines current developments in dealing with this type of uncertainty. The chapter also investigates the potential of the DT concept as a means of addressing key research challenges and provides an outline of an appropriate DT architecture for this purpose. Research questions are then presented considering the benefits and limitations DT provides based on a discussion of the problems from the DT perspective.

Chapter 3 covers the industry-as-academy research methodology, which outlines the roles of both industry and academia during various phases of the research process. These phases include dealing with sources of motivation, design and development, experimentation, and outcome testing. The chapter also details the research plan, including research objectives and milestones set for each of the identified research areas.

Chapter 4 provides an overview of modelling in the Cathodic-Protection (Corrosion) domain, which will be utilised in the experimentation and results analysis during the research. The chapter also contains the results of investigated approaches to address the first research problem, which is *"delay in acquiring data when available in survey reports"*.

Chapters 5, 6, 7, and 8 present the findings from the investigation guided by the research objectives and milestones for model calibration and adaptation-related uncertainties and complexity handling. Chapter 5 focuses on the automation requirements of the model calibration/adaptation process, while Chapter 6 discusses the results of research related to resources benchmarking. Chapter 7 explores an efficient parameter search technique, and Chapter 8 investigates addressing the complexity of model adaptation during the operational phase. In addition to presenting their findings, each chapter details the procedure for designing and developing solutions proposed to address the identified problems. Finally, each chapter also provides a demonstration of the proposed solutions' applicability.

Chapter 9 concludes this thesis with a summary of the research activities and the outcomes. The chapter summarises the proposed solutions and their applicability, highlighting the benefits of each approach (framework). Additionally, future works are outlined for the real-world application, extension, and enhancement of the proposed solutions.

2 Background and Research Challenges

This chapter provides a comprehensive overview of the simulation modelling and the Digital Twin (DT) concept, with a particular focus on those related to Structural Health Monitoring (SHM). The chapter furthermore includes the findings from the related literature analysis on the existing challenges in model calibration and adaptation for two different simulation applications- i) applications before incorporating DT and ii) applications where DT is incorporated as the potential solution to the challenges.

It begins with an introduction to SHM and is followed by the role of the physics-based model(s) for predictive simulation within SHM. The process of model generation and the uncertainties involved during the process are then explored to provide insight into issues related to the predictive models used in SHM. Furthermore, the section presents the common challenges in model calibration and adaptation within SHM.

The next section presents the findings from an investigation into recent literature on the current developments in the key challenges related to model calibration and adaptation in SHM-related domains. The aim of this section is to outline the current state-of-art associated with challenges in implementing predictive models in real-world scenarios through calibration and adaptation.

In the next section, the chapter discusses the idea of Digital Twin as a suggested strategy for resolving the difficulties associated with model calibration and adaptation. Additionally, the challenges in integrating DT into SHM are examined, with a focus on the difficulties in online adaption because it has the greatest potential for application. This outlines the DT gaps that must be evaluated to address the model calibration and adaptation-related issues.

Finally, the chapter concludes by raising research questions to address the challenges associated with the use of DT in SHM.

2.1 Simulation Modelling in SHM and Research Challenges

2.1.1 Structural Health Monitoring (SHM): An Introduction

The use of civil, aerospace, and other mechanical engineering infrastructure often involves some level of risk, including the potential for loss of life and economic loss. Therefore, these structure must not only be designed and produced with proper supervision but their integrity must also be monitored and maintained throughout their operational life cycle.

The term Structural Health Monitoring (SHM) itself is typically applied during the operational phase of the structure. It is the process of implementing a damage identification strategy for the assets (structures) during their operational phase (Farrar and Worden 2007a). Early damage identification allows for the appropriate maintenance strategies to be implemented to address the identified issue in accordance with the design rules. A reliable SHM process not only maintains the structure's fitness-for-purpose but also minimises the chance of downtime and catastrophic failure, allowing for optimal use of the structure.

Overall, SHM-related tasks can be classified into two types: diagnostic (identifying after happening, a traditional method) or prognostic (predicting before time) (Farrar and Worden 2007a, Daigle and Goebel 2012). While diagnosis is straightforward relying upon sensors' equipment and frequency of inspection, the prognosis-related task depends upon the data availability and/or knowledge about the physics of the behaviour.

In practice, almost all industries (private and government) aim to detect damage at the earliest possible time in their products, equipment and infrastructures. If maintenance activities can be carried out proactively it will reduce the maintenance cost and also maintain the asset's life at the highest possible level. This suggests a transformation of maintenance strategy from the traditional fail-and-fix practices (diagnostics) to a predict-and-prevent (prognostics) one (Lee et al. 2014). With the assistance received from evolving related technology SHM concept is now achieving maturity and it is increasingly being approached from a prognostic perspective (Farrar and Lieven 2007b, Daigle and Goebel 2012, Abbas and Shafiee 2018, An et al. 2019). Additionally, prognosis (prediction) of asset's performance enables optimal operational time planning, which enhances revenue-generating potential. Therefore this research focus on the prognosis aspect of SHM.

2.1.2 Prognosis in SHM

Prognosis sometime in SHM is also defined as, *'the estimate of an engineered system's remaining useful life (RUL) utilising principles, expert knowledge and/or data from the past'* (Farrar and Lieven 2007b). This suggest prognosis task is not straightforward and can be complicated in cases where behaviour and loading conditions are constantly varying, typically encountered in complex structural applications.

For the prognosis, predictive models are frequently used in SHM, together with the anticipated future environmental and loading conditions, information from usage monitoring and system-level testing, as well as past maintenance-related data (Farrar and Worden 2007a). Though term "*model*" is used with many different meanings in the sciences and philosophy (Blum and Ferri 2009), this research refers to it from the mathematical model's perspective as a '*representation of the dynamics occurring to the real-world system*'. The types of models used for prognosis can be physics-based, data-driven, or hybrid. The characteristics of these model types are briefly discussed below:

Physics-based model: The physics-based model provides an imitation of the dynamic process by solving the corresponding equations of the underlying dynamics (Gopalakrishnan et al. 2011). To make predictions using such model, analysts estimate the loading conditions (including initial and boundary) together with material and environmental related variables, and feed them into the model for the simulation run (Farrar and Lieven 2007b).

Data-driven Model: Data-based techniques, on the other side, depends on the previous data from the system to assess the current and future damage state, using typical pattern recognition method(s). Comparatively, data-driven ones are less computationally intensive than physics-based numerical models. That is why the recent trend in most of the domains is data-driven analysis in the situation of wider data availability and accessibility (Azimi et al. 2020).

Significant challenges are, however, posed to the data-driven models as they usually generalise the pattern with a larger set of situations.

Hybrid model: Using both numerical simulation and data analysis supported with inspection (sensor) data from the structure has been proven to be more promising than relying upon one technique (Ling and Mahadevan 2012, Neerukatti et al. 2014, Chao et al. 2022).

2.1.3 Factors determining the choice of Model for Prognosis in SHM

The utilisation of the physics-based or historical data-driven or hybrid approaches for structural risk assessment within SHM tends to be dependent on a few factors such as the complexity of the structure, data to track the trend, availability of physics-based models, etc.(An et al. 2015, Chao et al. 2022). The ultimate objective is to reduce resource requirements, including the costs associated with prognostic tasks (Balageas 2006).

While data-driven methods are becoming more popular in multiple domains in recent years, physics-based models are still necessary to understand situations that have not yet occurred (Daigle and Goebel 2012, Malveiro et al. 2018, An et al. 2019). The conditions and rates of damage progression in structures can vary even among structures of the same type due to differences in manufacturing and exposure to varying loading and environmental conditions during operation. Therefore, when it comes to the structural health monitoring (SHM) of complex structures, a physics-based model is necessary to carry out SHM using a prognostic approach (Daigle and Goebel 2012).

The effectiveness of prognostic solutions in SHM heavily relies on the ability of the physics-based model to provide accurate predictions. Therefore, it is crucial to ensure that the numerical simulations of the model are credible and meet the accepted threshold. However, modelling structural composites requires a deep understanding of the underlying mathematics of phenomena within the discretised domains, making the numerical process itself a crucial task. Even if the model sufficiently replicates the phenomena, the accuracy of the predictions heavily relies on the input data that describes the properties of the materials and environmental conditions that the structure experiences.

Since the credibility of the model is essential before using it for the prognostic task within SHM, the process of developing the model and potential stage (areas) for errors and uncertainties is discussed in the next section. The ultimate goal is identification of the potential aspects for the enhancement of the model's robustness.

2.1.4 Simulation Modelling: Overview

The physics-based model from mathematical model's perspective is often understood as an '*interpretation of the theory's calculus*' (Hartmann 1996). The outputs from such models are the dynamic response data values anticipated in the real system and obtained via solving the equations (such as Partial Differential Equations) of the underlying dynamics. Simulation is the process of imitating the dynamic process for a real system under a given load, boundary, and environmental conditions. When the numerical solution is performed using a computer, then it is called computer simulation. The process of developing such a mathematical model is termed modelling.

The development of a valid simulation model i.e., modelling requires an iterative process (Figure 2.1).

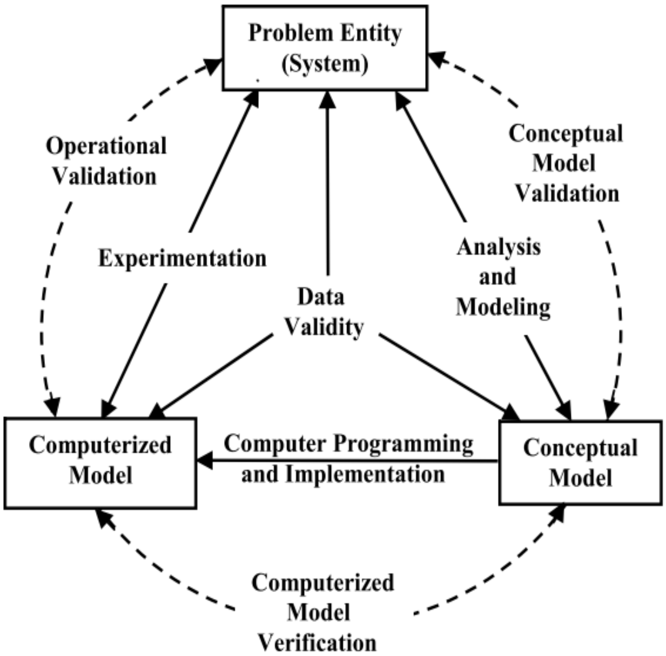


Figure 2.1: General model developing process (Sargent 1984)

Conceptual modelling

It is the process of abstracting a model from a real or proposed system (Robinson 2008). Moreover, it is a non-software with a specific description regarding the model (that will be or has been developed), describing the objectives, inputs, outputs, content, assumptions and simplifications of the model (Robinson et al. 2015). The followings are the common procedures of conceptual modelling (Oberkampf et al. 2002):

1. Mathematical modelling: It is the process of identifying governing equations for the dynamic physical, chemical and/or biological phenomena, and establishing a set of mathematical relations for the representation of the system (Hartmann 1996). The governing equations are mostly partial differential equations (PDEs) as most of the phenomena can be described in general by PDEs (Farlow 1993).

2. Discretisation and algorithm selection (Numerical approximation method): Numerical approximation is the process to solve the underlying Partial Differential Equations (PDEs) (Morton and Mayers 2005). Some commonly used methods for numerical approximation for models involved in SHM are the Finite Element Methods (FEM) (Rao 2017), Finite Difference Methods (FDM) (Thomas 2013), and Boundary Element Methods (BEM) (Kythe 2020).

Computerised modelling

The next step is the development of the computerised model from the conceptual model. Computerised modelling for structure-related dynamics usually involves three major tasks to obtain the numerical approximation using the computer programme (Aliabadi 2002, Zienkiewicz et al. 2005):

1. geometrical modelling of the structure(s),
2. meshing of the geometry, and
3. computational approximation of the Partial Differential Equations (PDEs) solution.

With recent developments, modelling steps 1 and 2 can be easily performed with CAD and meshing software tools respectively. Then, step 3 requires an additional numerical solver (algorithms) with appropriate solution (numerical approximation) methods.

Moving forward, the iterative process illustrated in Figure 2.1 are often undertaken to determine and assess the gaps in the model's performance during the conceptual and/or computerised model building process. A well-designed model should incorporate most of the salient features of the system and is a judicious trade-off between realism and simplicity (Maria 1997).

Dealing with Model's errors and Uncertainties

In modelling a structure or process, it is essential to account for errors in the measurement of representative values and the uncertainty of the model's performance compared to the real asset.

American Society of Mechanical Engineers (ASME), defines the error and uncertainty in the modelling process as follows (ASME 2009):

Error: *It is the result difference between the observed or calculated value from the true value. An error (δ) is thus a quantitative value that has a particular sign and magnitude.*

Uncertainty: *It is the effect of parametric and/or modelling quality seen in the model output. Uncertainty characterises the dispersion of the output values that could reasonably be attributed as observation values.*

Statistically, uncertainty is the estimated amount or percentage by which an observed or calculated value may differ from the true value. Intuitively it can be understood as the issues of non-determinism, particularly about the prediction of future events and the estimation of the reliability of systems.

Each phase of the modelling contains sources of error and uncertainty (Oberkampf and Trucano 2002). During re-assessment on each iteration after validation (Figure 2.1), the tasks to be performed are determining and addressing uncertainty that typically involves (Sargent 2010):

1. the inadequate theories (physics) of the model and/or
2. the approximation of the input parameters involved.

However, there are also situations where uncertainties can not be removed/reduced. Some of the uncertainties are reducible but appear (remain) in the model mostly due to the lack of knowledge (expertise), and are known as epistemic uncertainties. While other types are irreducible due to probabilistic uncertainties and are known as aleatory uncertainties (Helton et al. 2010).

The validation or verification of the output from each phase is essential to ensure the accurate representation of the real system to the requirement. The accuracy, as well as the complexity of the model (making it more comprehensive), can be increased by the iterative process.

Model Validation and/or Verification

Determining the correctness of the model or measuring the uncertainty the model possess is the goal of model validation. However, the different insights about conceptual modelling and computerised modelling lead to distinctions in the validation process. Conceptual model validation is dealing with the correctness and/or reasonability of the theories and assumptions that are followed. While computerised model verification is providing assurance of the correct implementation of the conceptual model into the computerised programme.

The validation task is defined by ASME (2009) as *"the process of determining the degree to which a model is an accurate representation of the real world from the perspective of the intended uses of the model"*

Substantial advances have been made in the field of benchmarking the definition of validation distinguished from the task of verification and accreditation has been provided (Oberkampf and Trucano 2008, ASME 2009, Sargent 2010). Among the different methods used to validate simulation models, operational validation is widely applied at the last stage of model building i.e., during the realisation of a virtual replica of a physical asset (Sargent 2010). It is done by determining the error and/or uncertainty of the model's prediction with a comparative analysis of predicted data to the corresponding physical system-related data. This measurement of accuracy between model outputs and real data is made using performance validation metrics also termed as performance criteria (Sarin et al. 2010).

The process of model-building has been undertaken in many domains for several years and has progressed significantly. Different methods (techniques and tools) for capitalising on knowledge, experience, and expertise are now in practice. These includes the commercial simulators (simulating software) for numerical simulation across multiple domains and the adoption of the parametric modelling concept.

2.1.5 Simulator-based Simulation Modelling

The use of a previously validated conceptual model (discretisation and numerical approximation) assists in the design of not just one simulation model but many within the problem domain (Robinson et al. 2015, Abdelmegid et al. 2020). The simulation model building in multiple domains is already facilitated by the availability of commercial process simulators (simulating software). Such simulators are generally based on current scientific understanding (physics of phenomena), often involving numerical approximation of differential equations, and are implemented in complex computer programs (Brynjarsdóttir and O'Hagan 2014). Commercially available simulators in structural analysis (e.g.: AKSELOS™, ANSYS ™, BEASY™) provide the functionality of Computer-aided-design (CAD) modelling, meshing, and also behavioural process simulation i.e., a numerical approximation of PDE's.

2.1.6 Parametric Modelling Concept

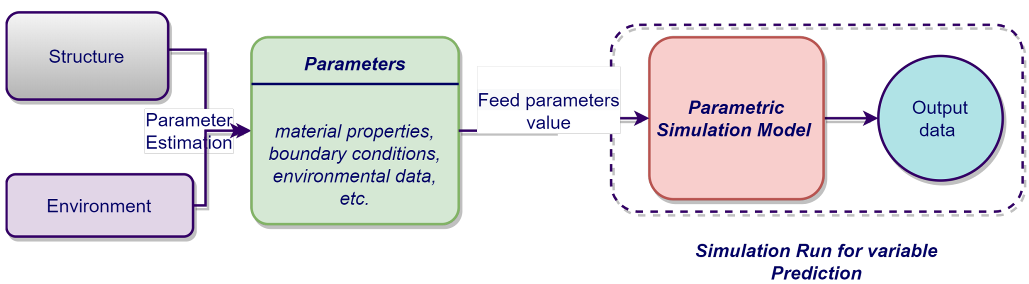


Figure 2.2: Simulation execution with parameter feeding into the parametric model to make the prediction.

In general engineering practice, the knowledge of a structure such as material properties, deterioration and so on is usually uncertain. This lack of knowledge of the true parameter values motivates the idea of building a parametric model sufficiently representative of the phenomenon of interest in the field. Such a model requires feeding into appropriate parameter-related values to execute predictive simulation (Figure 2.2). The parameters can be a variable representing the material properties of the system and/or the surroundings that have a significant effect on the simulation output (Hartmann 1996). The parameterisation of the model allows for the implementation of generic functions to solve similar problems where the model can be tuned.

As discussed earlier (Section 2.1.4), the process of enhancing a model involves two major re-iterative tasks: i) determining and addressing uncertainty by updating parameters, and ii) accounting for inadequacies in the model's physics. When predefined simulators that have been verified and validated are adopted, the re-iterative task is mostly limited to the first task (Higdon et al. 2008, Martinez et al. 2018), which is necessary for the model's performance to converge with the actual behavior of the system/structure. Therefore, it is recommended that the validation and updating mechanism focuses on ensuring the correctness of the parameter(s). For tuning the parametric model to the real-world scenario, the best set of parameters are required that best correlates the model output (i.e., prediction made by the model) to the available measurement data from the real-asset. This parameter estimation (tuning) is termed as calibration of the model.

2.1.7 Calibration of Parametric Model

Parametric models built using pre-existing simulation computer programmes required to be tuned and calibrated with various inputs to closely match the observed data with the predicted simulation run results (Agami Reddy 2006). Structure-related data from sensors or inspections, together with knowledge of deterioration mechanics and behavioral laws are combined and used to calibrate the model. The model calibration, which involves updating the parameters, is performed through multiple simulation runs by varying the input parameter values until a match is found (Figure 2.3). In most scenarios, engineers or analysts may not have a clear understanding of the model's working principle, but they use their engineering judgment, along with available data interpretation, to choose appropriate values. However, manual calibration using trial-and-error approaches is often inefficient (Taylor et al. 2010).

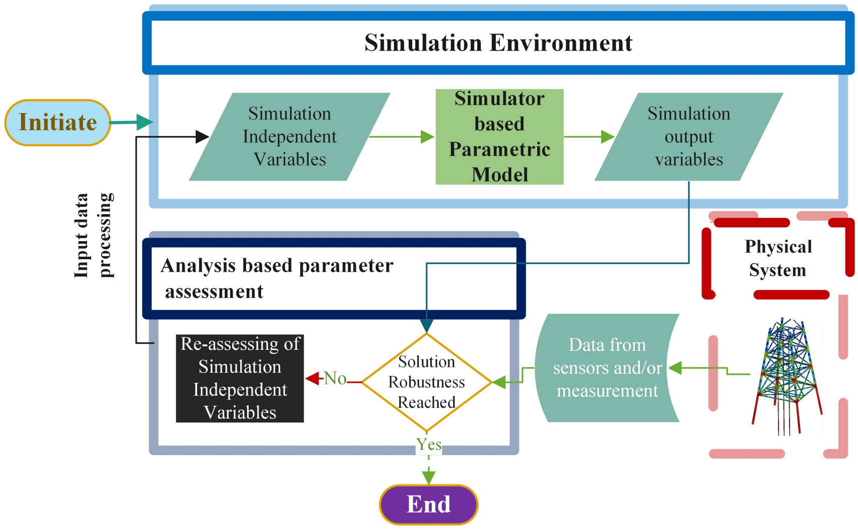


Figure 2.3: The model parameter estimation process during its calibration with comparative analysis between model's output and real-world data (trial-and-error method).

Common Issues associated with model calibration

Calibrating a parametric model (pre-validated to its conceptual level) through parameter tuning can still present challenges associated with either data-related and/or procedure-related issues. In Section 2.1.9, the former will be discussed, while the latter pertains to the selection of appropriate methods or algorithms. Some of the most common and relevant procedure-related challenges for simulator-based model's calibration in SHM are listed below.

1. Manual involvement mostly while implying the traditional trial-and-error methods, and/or implementation of calibration algorithm (Zambrano-Bigiarini and Rojas 2013, Silva et al. 2021).
2. Difficulty in selecting appropriate performance metrics to evaluate the goodness-of-fit of the model (Liu et al. 2011, Sarin et al. 2010).
3. The computational cost can be high if required multiple simulation runs, especially for large-scale or high-dimensional systems (Yang et al. 2013).
4. The model can be sensitive to the initial conditions and/or the choice of model parameters. This often leads to the risk of ending up at the local minima during implementation of calibration algorithms (such as mathematical optimisation) making it difficult to find the global minimum (Yang et al. 2013, Cao et al. 2020).

2.1.8 Model's adaptation requirements in SHM

The issue of non-uniform and complex changes in the material properties of a structure is a significant challenge for SHM (Sohn 2007, Sehgal and Kumar 2016). Even if structures share the same design parameters, variations in manufacturing precision, material properties, and operational loading conditions lead to differences in behaviour. As a result, even after initial calibration and provided design rules suggesting future trend of parameters, the future performance of a structure-related model cannot be guaranteed. This is why the model's predictions tends to drift-away from the real-world data over time. Analysts may need to perform re-calibration (adaptation) for each model corresponding to different physical assets, and the adaptation route would vary (Belostotsky et al. 2018). Since adaptation mainly involves re-calibration, calibration-related challenges are also relevant from a model adaptation perspective.

Other major issue that arise over time in the case of the structural model is the variability of the model's parameter where a material considered uniform before behaves varyingly among the position of the structure (Marques et al. 2012). For structure on operation, this variability in dynamic properties can be a result of varying environments and/or operational conditions (Sohn 2007). This demands further parameterisation of the model with an increased number of parameters and repetition of the parameter updating.

2.1.9 Real-world data for model calibration and adaptation, and common issues

Real-world data is a crucial requirement for validating model's performance during its calibration or adaptation. However, the complexity of physical systems can result in varying data requirements to ensure accurate validation of operational performance (Pace 2004, Oden et al. 2013). Obtaining structural data is often restricted to data from sensors or physical inspections, and it may not always be readily available. As a result, the shortage of necessary data or information during calibration or adaptation can lead to increased costs, time, and the risk of incorrect assumptions.

Common challenges associated with timely acquisition of quality data for model calibration and adaptation include:

1. Inconsistency in SHM-related data resources with a wide variety of formats, terminologies, and concepts (Gong et al. 2016, Gulgec et al. 2017, Bayraktarov et al. 2019), which results in a risk of knowledge-data mismatch (Ashino 2010).

2. Delays in obtaining measurement data when obtained from the surveys (Adey et al. 2020).
3. Incompleteness and/or error in the measurement data (Catbas et al. 2013).

Due to the challenges in obtaining data and the varying data requirements caused by the behavioural complexity of the physical system, it is crucial to ensure that the collected data are sufficient for model calibration (Fabrizio and Monetti 2015). Therefore, it is essential to benchmark the data based on their quantitative, qualitative, and diversity requirements to monitor the model's performance and avoid incorrect assumptions about its performance.

2.1.10 Summary on SHM related simulation modelling and challenges

The research background overview indicates that structure-related model (such as the FE model) development is made easier by software packages that include simulating tools (simulators) (Quintana et al. 2014, Nguyen et al. 2018, Ezzat et al. 2020). Furthermore, research has acknowledged that even with the most advanced software packages available, developing a comprehensive representative model via. calibration remains a challenge (Aktan and Brownjohn 2013, Nguyen et al. 2018). This tends to be due to uncertainties regarding the structure's material properties, as well as uncertain boundary conditions (Sabamehr et al. 2018, Nguyen et al. 2018).

While this section presented the overall potential SHM-related model calibration and adaptation issues, the next section will present the current developments related to these issues based on the findings of the research investigation into recent research work.

2.2 Current developments in model calibration and adaptation related key challenges in SHM-related domains

The previous section discussed the common issues that may arise in the process of creating a trustworthy physics-based model in SHM, with a specific emphasis on problems related to calibrating and adapting the model. It is important to note that researchers are not ignoring these issues. The complexity of the modelling process, including the need for calibration and adaptation, has been recognised by researchers and developers for many years (Sargent 1984, Papalambros and Wilde 2000, Agami Reddy 2006).

The following sub-sections present the findings from the analysis of recent relevant literature. The process details on the literature search and analysis findings of the representative literature are presented in tabular format in Appendix A.

2.2.1 Real-world data for model calibration

The advancement of technology, such as high-performance sensors and precision signal conditioning units has led to more precise and cost-efficient data acquisition in SHM (Cremona and Santos 2018). The optimisation of the sensors placement strategies are also being discussed to ensure adequate data requirement (Barthorpe and Worden 2020, Liu et al. 2021). This is more applicable in situations where only a small number of sensors can be placed at the limited positions of a structure in practice due to the restrictions of physical space and economic conditions. Additionally, the model-data mismatching risk caused due to heterogeneity of data is being addressed with the adoption of related developments such as the material and asset-facilitated Ontology concept (Ashino 2010, Stark and Pförtner 2015).

However, there is still a need for a standardised approach to determine the quality, quantity, and variability of data required from sensors (survey) to ensure model performance during calibration (Fabrizio and Monetti 2015, Kang et al. 2021). Another issue that still persists is the acquisition of past, unstructured SHM-related data that could still play an essential role in information mirroring and analytics (Gong et al. 2016, Gulgec et al. 2017). Furthermore, there is currently a lack of a comprehensive platform to organise the heterogeneity of data and utilise it for model validation during calibration or adaptation (Barni et al. 2018).

2.2.2 Tools and Techniques for model calibration

Calibration in SHM typically involves using response data to estimate parameter(s) related to the material properties, also referred to as "*system identification*" (Sabamehr et al. 2018, Barthorpe and Worden 2020). This is considered an "*inverse problem*" due to its inherent nature, i.e., model output data are used to estimate the input parameters (Sabamehr et al. 2018, Liu et al. 2021). Following the concept, calibration may prove helpful in identifying errors in data obtained from the physical system and monitoring sensor (survey equipment) malfunctions (Buethe et al. 2014).

With the development in tools and technologies, the calibration approach is being advanced from the traditional trial-and-error approach. To overcome the shortcomings of the traditional trial-and-error approach, systematic calibration procedures such as experimental design (Seltman 2012), sensitivity analysis (Christopher Frey and Patil 2002), and design (parameter) optimisation (Roy et al. 2008) are often used as alternatives for model calibration (Law et al. 2007). Design optimisation in this scenario can be understood as an approach that combines mathematical optimisation algorithms with a parametric simulation model to search the design (parameter) space for the optimal solution (Papalambros and Wilde 2000). When it comes to optimisation-based parametric calibration, residual minimisation is commonly employed which involves minimising the difference between the observed and predicted values of a system. This can be accomplished using either deterministic or probabilistic approaches for optimisation.

Sensitivity-based (Nguyen et al. 2018, Peč et al. 2019) and/or gradient-based approaches (Cao et al. 2020) are used under the deterministic approach where the information on the sensitivity of the model to the parameters is used to update them iteratively until the residual is minimised. However, such an approach has a common issue of reaching local minima during optimisation mostly in the situation with errors in measurement data (Cao et al. 2020). Therefore, probabilistic approaches like Bayesian methods are often preferred because they account for uncertainties in both the parameter and calibration data (Green et al. 2015, Behmanesh and Moaveni 2016, Ye et al. 2020). Nevertheless, the implementation of probabilistic techniques may result in significant computational expenses, making them impractical for complex structures.

To address the challenge of high computation costs in calibration, surrogate-assisted parameter calibration has emerged as a viable solution (Yondo et al. 2018, Vincenzi et al. 2019). However, the role of surrogates in addressing issues related to data and the risk of local minima during optimisation problems is not yet well-established.

Automation of calibration is being achieved in some cases through the use of optimisation algorithms (Huang et al. 2010, Peč et al. 2019), however, this approach has limitations when the model and the optimisation algorithms operate in different design environments. In such cases, the algorithms need to be heavily modified to establish links with simulator inputs and outputs (Zambrano-Bigiarini and Rojas 2013), requiring specialised expertise. Another approach to automation is the collaboration of scientific software, analytical tools, and simulators (Murphy and Yarnold 2018, Benaouali and Kachel 2019), but this approach still requires manual involvement in most situations (Cao et al. 2020, Silva et al. 2021). To fully automate model validation and updating, a framework is needed that establishes the role of software(s) in assisting simulators.

Given the importance of data and expertise-related costs in establishing a model with validated performance, it is essential to benchmark the calibration process. Before implementing an algorithm for parameter identification in the real world, it is recommended to validate it experimentally (McGetrick et al. 2015).

The analysis shows that the key gap in the literature is how to best use observed data to idealise the physics-based model that replicates the responses of the operational structure (Gregory et al. 2019). Moreover, the dynamic operational and environmental conditions of the structure make it challenging to develop efficient prognostic models that are resilient enough to tolerate uncertainty under diverse conditions (Cross 2012, Javed et al. 2017, Xia et al. 2018).

2.2.3 Development on model adaptation-related issues

The model and/or parameter adaptation requirement in SHM though acknowledged by the researchers/engineers, the progress in adaptive modelling is still at a slow pace (Gude et al. 2015, Behmanesh and Moaveni 2016, Rabiei et al. 2018, Zhang et al. 2020). Typically, model adaptation is approached similarly to calibration (Belostotsky et al. 2018), but evolving material properties and changing surroundings over time requires a comprehensive model that can gather information and reduce resource requirements over time (Cross 2012, Gabor et al. 2016, Javed et al. 2017, Kita et al. 2019). Some research has attempted to incorporate environmental effects on structure-related parameters to create a holistic model (Murphy and Yarnold 2018, Zheng et al. 2020), however, the approaches are limited to experimentation and cannot address the continuous adaptation requirement of a practical model.

In regular operation, the dynamics of the structural environment and loading become non-uniform, which induces "*spatial variability*" within a parameter (Spiridonakos et al. 2016, Ehret et al. 2020). The model adaptation in such a situation is suggested to be assessed by increasing the parameter count to the full-order model or by embracing the module-based concept (Sohn 2007, Jesus et al. 2017). However, increasing the parameter count and/or still relying upon the full-order model with manual involvement can be complex and time-consuming and pose risks to the structure if the model fails. Hence, more efficient approaches to adapting model parameters are needed.

On the other hand, testing model performance during the operational phase with parametric variability issues requires a robust tool for probabilistic assessment (Aldrin et al. 2011). This also necessitates a framework for data collection (sensor placement) to calibrate spatially varying parameters (Nath et al. 2017), which advocates for an online data benchmarking framework.

The adoption of Machine learning (M/L) and other data-driven approaches either to support physics-based model's updating (Montáns et al. 2019) or having a data-driven predictive tool in such a complex scenario is also gaining prominence in SHM (An et al. 2015, Azimi et al. 2020, Katam et al. 2022, Omar et al. 2022). This is due to recent advancements in sensor technology, as well as fast progress in internet-based cloud computation which promises more similar structure-related data (Katam et al. 2022). However, completely replacing physics-based models seems unlikely in the near future for SHM-related prognostics (An et al. 2015, Montáns et al. 2019, Omar et al. 2022).

2.2.4 Summary on current development

While researchers have made progress in addressing individual challenges related to model calibration and adaptation, these advancements have not been sufficient to enable the practical and continuous implementation of prognostic models for real structures. Despite progress in modelling and calibration-related literature with experimental data and models, real simulator-based parametric models remain scarce (Javed et al. 2017, Ozer and Feng 2019, Ezzat et al. 2020). The difficulty in creating accurate prognostic models is not due to a single factor, but rather a combination of previously discussed factors, such as inherent uncertainties related to the deterioration process, insufficient data quantities, sensor noise, unknown environmental and operating conditions, and engineering variations (Javed et al. 2017).

The frequent adaptation requirement of the structural model presents the most significant challenge in practical application of the model. To address this challenge, a self-adaptive model is needed, which is most likely to be a more comprehensive model that also encompasses the structure's evolving material properties as well as its evolving surroundings over time (Gabor et al. 2016, Kita et al. 2019).

To successfully develop such a robust but practical predictive tool, integrating large and heterogeneous real-world data resources in multiple SHM-related domains with physics-based and data-driven approaches would be crucial (Stark and Pförtner 2015, Cremona and Santos 2018). However, this requires a holistic approach to address the existing categorical challenges in real-world data collection, arrangement, and integration into the model for its calibration and/or adaptation.

The need to address challenges related to model calibration and adaptation holistically has motivated this research to investigate the potential features of a Digital Twin, which has emerged as a more comprehensive model in the field of modelling in recent years.

2.3 Digital Twin Concept in Modelling

The Digital Twin (DT) concept is a novel development in simulation and modelling field (Boschert and Rosen 2016, Barricelli et al. 2019). Over the last decade, the term DT has been used to describe a model that has a corresponding physical twin with real-time information (Glaessgen and Stargel 2012, Ye et al. 2020). In addition, the DT concept offers to provide a comprehensive asset management model that differs from a purely mathematical approach. It incorporates real-world data, analytics and simulation model to provide predictive insights (Macchi et al. 2018, Wright and Davidson 2020).

This section therefore aims to explore the potential of using DT as an approach to address the SHM-related modelling challenges discussed in Sections 2.1 and 2.2. To achieve this, the current state-of-the-art of DT will be discussed, and the following research questions will guide the findings:

1. *What is the current state of the art for undertaking simulation tasks within the DT paradigm?*
2. *What are the current approaches for simulation model calibration, predictive simulation, and online adaptation within DT*
3. *What are the existing challenges in model calibration, predictive simulation and online adaptation-related assessment, within SHM-related DT?*

To answer these research questions, this section presents the findings from a review of DT-related papers, primarily focused on SHM and dedicated to predictive simulation assessment.

2.3.1 Digital Twin's Definition

The definition of Digital twin (DT) also known as Cyber-Physical System tends to vary depending on the domain of application. For instance, DT has been defined by NASA (in 2012) as an " *integrated multi-physics, multi-scale, probabilistic simulation of a vehicle or system that uses the best available physical models, sensor updates, fleet history, etc., to mirror the life of its flying twin. It is ultra-realistic and may consider one or more important and interdependent vehicle systems*" (Glaessgen and Stargel 2012).

However, the DT concept was informally introduced in 2003 within product life-cycle management by Michael Grieves. At the time, digital representations of actual physical products were relatively new and immature. Grieves (2014) elaborated on the digital twin concept from the product life-cycle management (PLM) perspective and defined it with three major components: physical products, corresponding virtual products of the physical one and the data and information connections between them.

Rosen et al. (2015) proposed the adoption of the digital twin concept in manufacturing to realise Industry 4.0. as one of the substantial technical and technological solutions. Parrott and Warshaw (2017), in the context of Industry 4.0 has defined DT as, *"an evolving digital profile of the historical and current behaviour of a physical object or process that helps optimise business performance"*.

According to Negri et al. (2017), *"the digital twin is based on massive, cumulative, real-time, real-world data measurements across an array of dimensions, and can be used for forecasting and optimisation of production systems at each life cycle phase in real-time."*

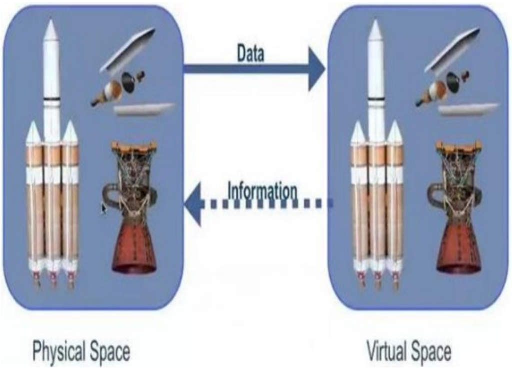
Similarly, Rasheed et al. (2020) defines Digital Twin as *"an adaptive model of a complex physical system"*

The literature shows, there has been a different understanding of the Digital Twin concept in academia. While the few available international standards for Digital Twin to date, (for example ISO/DIS-23247-1 (2020)) are limited to the manufacturing domain, an analyst in the structures (assets) related domain could understand the definition in their way. Still, the best part is the concept of DT has evolved significantly in the last decade (Figure 2.4). Ongoing research and analysis on the DT concept, as well as its practical suggestion, can be considered as an effort towards standardisation of the DT concept for real-world problems.

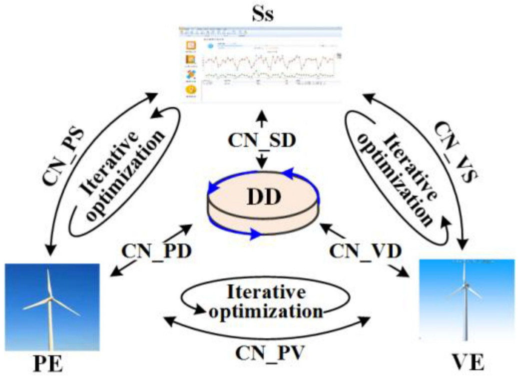
Also, there are other terms often appearing in the literature to represent the virtual replica and/or following the above-mentioned DT-related conceptual definitions.

Digital Thread: The digital thread refers to linking and integrating models from various aspects through common inputs and data flows to speed up the design time and enable trades across isolated disciplines (Siedlak et al. 2018). Often mentioned together with Digital Twin, digital thread especially refers to the communication framework that links the entity in support of synchronising the real and virtual world (West and Blackburn 2017).

Avatar: The concept of "*product avatar*" was introduced by Hribernik et al. (2006), and the concept was similar to digital twin. The concept is intended to support the establish-



(a)



(b)

Figure 2.4: Evolution of DT- 3-dimensional concept (Grieves 2014) to 5-dimensional concept (Tao et al. 2018)

ment of an information management architecture that furthermore enables a bidirectional information flow. The term is used for the digital counterparts as targeted digital representations of products (Hribernik et al. 2013). Also, some research recognises both the term '*product avatar*' and '*digital twin*' and differentiates them (Ríos et al. 2015). However, the implication of the term '**Digital Twin**' in most cases is slowly replacing the product avatar concept.

Internet Of Things (IoT): IoT is usually understood as the interconnection of objects to interact and cooperate to achieve a common objective. Also, the concept of IoT is still evolving, much like the concept of DT (Li et al. 2015). While DTs are typically understood as the virtual representations of the physical world that organise and manage information and are integrated with models and analytics, IoT is mostly about connecting resources and collecting data about the physical world (Jacoby and Usländer 2020).

2.3.2 Development of DT in Manufacturing

DTs are considered one of the key enablers for Industry 4.0. in the scenario where Industry 4.0 is envisioned as interlinked and autonomous manufacturing systems, self-organising the production of small batch sizes (Stark et al. 2017). Today, manufacturing covers more than half the number of DT research, though DT was introduced for asset health monitoring. The industrial application of the DT could range from real-time monitoring (Soares et al. 2019), production control (Zhao et al. 2019), process evaluation and optimisation (Sun, Bao, Li, Zhang, Liu and Zhou 2020). Moreover, from the asset's health perspective, the application ranges from asset management (Zhuang et al. 2018) to predictive maintenance (Aivaliotis et al. 2019a, Feng et al. 2023), fault detection (Wang et al. 2019), performance prediction (Seshadri and Krishnamurthy 2017) and more in the service phase.

In the industrial context, the initial simulation-based concept is being replaced with the data-driven concept, especially while having DT on the Industrial-floor (Schroeder et al. 2016, Lee et al. 2017).

2.3.3 DT Application in SHM

In the initial phase, DT roles were foreseen as the multi-physics models of the aircraft's digital counterpart (Tuegel et al. 2011, Glaessgen and Stargel 2012). The role of DT anticipated was to design and predict the structural life of an air-frame and to enhance the existing approaches for aircraft certification and sustaining. DT at the beginning was limited to the conceptual level, but still enough to have insight as a future powerful tool applicable for structural health prognosis. However, unlike the other field of manufacturing, the progress of the DT concept and its related technologies in the engineering infrastructures related field is not fast enough and is still nascent (Broo et al. 2022).

During the last few years, the research about the structural DT technology applied to civil or engineering structures has increased to some extent within the scope of SHM. Bazilevs et al. (2015) uses the term DT though lacking proper elaboration on the concept and used the term in the context of developing the framework for aircraft components' fatigue-damage prediction that uses an advanced computational model (virtual replica) informed by in situ SHM data. Next, Seshadri and Krishnamurthy (2017) developed the methodologies as part of the digital twin concept, where multi-physics models, sensor information and input data are integrated to mirror the life of its corresponding physical twin (aircraft).

The application of the DT concept seems to have expanded to other SHM-related domains in the last few years, which includes product/machinery-related predictive maintenance (Scaglioni and Ferretti 2018, Qiao et al. 2019, Aivaliotis et al. 2019a, Feng et al. 2023) to the offshore/onshore assets' structural integrity management task (Knezevic et al. 2019, Tygesen et al. 2018, Adey et al. 2020). The need to monitor and control manufactured assets like buildings, bridges, etc. throughout their life-cycle has moved several researchers into investigating Digital Twin's potential for the role of SHM (Ye et al. 2019, Angjeliu et al. 2020, Chiachío et al. 2022). However, the lack of a proper definition for infrastructure-related DT exists which creates confusion, for example in building infrastructure-related papers, authors use the term digital twin simply as a synonym for Building-Information-Modelling (Sacks et al. 2020).

In summary, DT is gaining increasing attention within SHM in recent years, possibly due to the risk associated with the assets and with the support available with recent technology. The approaches found in the literature contribute to one or some of the functionalities generally, such as digital representation and/or predictive simulation within or assisted by DT.

2.3.4 DT and Simulation features

The lack of a standardised concept of a digital twin leads to misunderstandings with simulation aspects within digital twins. Most researchers believe that a Digital Twin should be an ultra-high-fidelity simulation (Glaessgen and Stargel 2012, Ye et al. 2020, Gardner et al. 2020, Ganguli et al. 2023). Although the simulation seems to be a key aspect related to the DT concept, some authors do not mention it (Negri et al. 2017, Parrott and Warshaw 2017).

On summarising the important characteristics of DT after related literature analysis, Barricelli et al. (2019) has concluded: *”DT should provide modelling and simulation applications for representing, realistically and naturally, both the current status of the physical twin and different “what-if” scenarios.”* The trend shows DT concept is following two directions - one focuses on the physics-based simulation aspect (Glaessgen and Stargel 2012, Boschert and Rosen 2016, Wright and Davidson 2020) while the other is on the data-based modelling (Lee et al. 2017, Booyse et al. 2020, Fahim et al. 2022). DT based upon both types of the predictive model i.e., data-driven and physics-based is termed a hybrid Digital Twin (Chinesta et al. 2020, Azangoo et al. 2022). A hybrid DT uses a combination of both types of models to have a more accurate predictive tool. Again, in the structure-related DT, a physics-based model is suggested for the task of predictive simulation (Wright and Davidson 2020).

2.3.5 Digital Twin’s Architecture

The DT’s architecture in literature is also evolving along with the concept but is varied within the application domain. The general and less challenged architecture of a DT is the one proposed by Grieves (2014) which includes physical space, virtual space and the connection for the flow of data and information between them. Higher-dimensional DT architectures are also being introduced which are aligned to the initial 3D concepts with the categorisation of the virtual space into different sub-modules. For example, the five layers of DT architecture presented by Ponomarev et al. (2017) and also adopted by Bazaz et al. (2019) are: (1) cyber-physical layer, (2) primary processing layer/store data layer, (3) distributed computing and storage layer, (4) models and algorithms layer and (5) visualisation and user interfaces layer. The architectures are highlighting the data storage modules as well as the modules for multi-purpose algorithms that include analytics and optimisation within DT.

Likewise, Tao et al. (2018) proposed another 5-d concept (Figure 2.4) for Digital Twin which is composed of: (1) Physical entity (PE); (2) Virtual entity (VE); (3) Services (Ss) for PE and VE; (4) Data (DD) and (5) Connection (CN) among PE, VE, Ss and DD.

In summary, a DT architecture can be understood and elaborated differently, however, the components and technologies can be organised into three major spaces: the physical space, the communicating/networking space and the virtual space. The digital replica or the DT itself exists in the virtual space.

Physical Space:

The physical space usually consists of a product or a device, a physical system (structure), or even an entire organisation including the activities, process and/or phenomenon. DT is about having virtual models for the physical entities in a virtual space to simulate/replicate their behaviours. For this, the physical space at first is anticipated to have the sensing capability so that the structural features of the physical entity can be perceived which is required in many application scenarios.

During both directional information flowing, reliable sensors and actuators (when applicable) are vital for effective analysis (diagnosis or prognosis) and the control chain.

Virtual/Computing Space:

The virtual space incorporates the virtual entities required to represent the corresponding physical entities, which primarily include the 3D CAD model and data element to infer the properties of the observable physical asset.

In advanced form, it comprises other virtual models and tools such as physics-based models, data-driven models and analytics. Furthermore, data storage and services systems required in establishing and maintaining the DT system comes within this space, when DT-concept is applied with 3 dimensions.

Communicating/Networking Space

Both the physical and the digital twins should be essentially equipped with connecting devices to guarantee a seamless connection for the continuous exchange of data and signals. With the recent cutting-edge technologies in the development of sensor and communication tools, advanced quantities of data can be acquired. Though in some cases sensors may not be connected, such cases anticipate near-time real-world data, otherwise, the term digital twin may not be appropriate to represent the virtual replica of the physical twin (Jones et al. 2020).

2.3.6 DT concept on Adaptive Modelling

The concept of model adaptation came into existence together with the notion of predictive DT. The DT concept as a "*living model*" presented by Tuegel et al. (2011), is about the model that continually adapts to changes in the environment or operation using real-time sensors data. The information model updating is provided with some frameworks within Digital Twinning such as ontology-based support (Bao et al. 2022) to overcome the issue created due to the heterogeneity of data. However, this doesn't mean only updating the physical representation model (CAD model) of DT is enough. The predictive model also needs to be updated/adapted to emulate the real-time behaviour of the related physical twin. Moreover, the maintenance and re-assessment of the model are essential tasks for real-life applications requiring reliable reference data from physical space.

Following the adaptation concept, some research suggests DT over mathematical models for the case when an object/structure is changing over time (Rasheed et al. 2020, Wright and Davidson 2020). This undoubtedly demands the ability of DT to accurately simulate events on different scales of space and time. This everlasting prediction credibility of DT is possible by relying upon the ultra-high-fidelity advanced physics-based simulation, but also on the collection of data from all deployed systems and thus aggregating the experience gained in the field to update the model based on collected data (Liu et al. 2018). From the object/structure changing perspective, the online adaptation is mostly about updating the parameter in moving time windows.

This model adaptation concept within DT is also supported by the extended 5D DT proposed by Tao et al. (2018). In the 5-dimensional architecture for DT proposed by Tao et al. (2018), the role of the services module (Ss) is also to ensure the high fidelity of the Virtual Entity (VE) by calibrating the VE parameters during its running to maintain its performance with the Physical Entity (PE). Similar to Tao et al. (2018), Vrabič et al. (2018) has proposed the concept of the Learning Model as a partial module of DT, that assists in enabling the simulation of what-if scenarios. This concept can then be exploited for tasks such as the optimisation of control parameters or synchronisation of the digital twin models with the physical asset.

The ongoing research around the concept of Digital Twin (DT) demonstrates its increasing recognition and acceptance as an advanced modelling approach, supported by cutting-edge technologies (Barricelli et al. 2019, Zhang, Zhou and Horn 2021, Ganguli et al. 2023). Moreover, these cutting-edge technologies include advancements in Machine Learning and Artificial Intelligence (AI), which are often utilised for physics-based DT-related adaptations (Chakraborty et al. 2021a, VanDerHorn and Mahadevan 2021).

This chapter has already discussed the advantages of selecting a physics-based model over a data-driven approach (in Section 2.1.3). However, the complexity of implementing physics-based models can sometimes lead to an inability to precisely represent the actual system. Additionally, environmental noise present in measurement data can pose further challenges. In such scenarios, data-driven Digital Twins (DTs) may offer an alternative solution for mitigating these issues (Booyse et al. 2020). Nevertheless, it's important to note that data-driven DTs also come with their own set of limitations (as discussed in Section 2.1.3).

While data-driven DTs can factor in noise within the data, they may struggle to accurately predict behaviour in unfamiliar environments. This limitation emphasises the need for a more comprehensive approach. As a potential solution, the concept of a hybrid DT, which blends elements of both physics-based and data-driven DTs, is also introduced (Chinesta et al. 2020, Azangoo et al. 2022). This hybrid approach aims to balance between the strengths of each method. The hybrid DT concept which incorporate data-driven and physics-based model is about utilising machine learning techniques to learn and compensates for incomplete physics from the real-time streaming data (Chakraborty et al. 2021a, Tripura et al. 2023).

On the other hand, within the realm of DT being model with real-time representation, the concepts of self-learning and self-adaptation were introduced to the DT paradigm a few years ago (Barricelli et al. 2019). While research in the earlier and mid-2010s focused on establishing the DT concept, a trend for DTs' self-adaptive features establishment began to gain prominence in the late 2010s and early 2020s (Rasheed et al. 2020, Julien and Martin 2021, Alnowaiser and Ahmed 2023). Specifically, for this notion of self-adaptability of DT in automotive contexts, recent developments in Machine Learning and AI are being leveraged and expanded (Orlova 2022, Tripura et al. 2023). This is why such developments in M/L and AI is emerging as a benchmark for DT concept, getting established as major characteristics of Digital Twin (Rasheed et al. 2020, Julien and Martin 2021, Chakraborty and Adhikari 2021b).

2.4 DT Features adoption to address challenges in adaptive Modelling

From this developments within the areas the research activities should be directed towards addressing practical limitations associated with DT implementations (Orlova 2022, Alnowaiser and Ahmed 2023, Ganguli et al. 2023), which is happening to some extent. For instance, the adoption of machine learning (M/L) models to complement physics-based model (Chakraborty and Adhikari 2021b, Feng et al. 2023) or as core predictive model (Fahim et al. 2022) has been considered as a potential solution to overcome the practical limitations of DT in recent years.

Moreover, the current state-of-the-art in DT technologies shows certain features can be utilised in SHM to address issues related model calibration and adaptation. Thus, DT is considered a viable solution for the challenges discussed in Sections 2.1 and 2.2.

Before utilising the DT concept followed by its development, it is first necessary to benchmark the DT concept, especially for its predictive notion. This research from this stage will embrace the DT concept as, *a predictive tool that is integrated with physics-based models and adapt on its own*.

Additionally, it is important to adopt and adhere to a suitable DT architecture that can maximise the benefits of the current state-of-the-art.

2.4.1 DT Architecture to support adaptive simulation and predictive features

The research at this phase adopts the DT architecture for SHM-related predictive role with the given artefacts as shown in Figure 2.5. The DT architecture in the virtual space (One dimension of DT discussed above) comprises the following modules:

1. **Physical-Twin-related data for data-mirroring:** This artefact will be facilitated by the recent cutting-edge technologies related to the development of sensor and com-

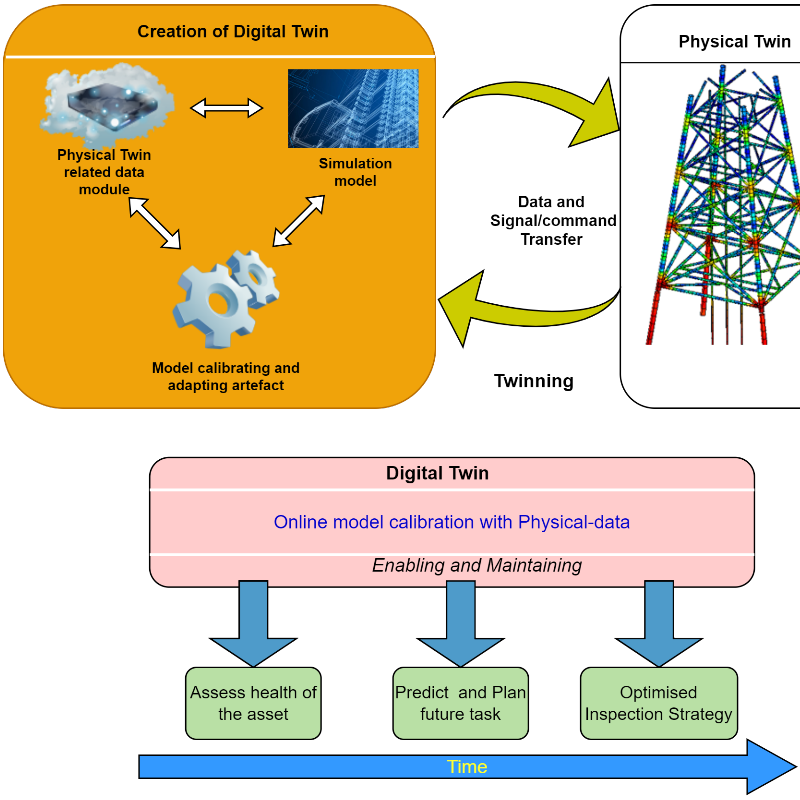


Figure 2.5: The adopted DT concept with the required data-mirroring, simulation, and calibration/adaptation artefact

munication tools so that data can be obtained in near-real time. For the geometrical representation of the physical twin, CAD model will be utilised.

2. **Parametric Simulation model for prognosis:** This artefact provides the role of behavioural simulation and can be achieved using simulation software (simulators) in most cases.

3. **Model calibration/adaptation artefact for parameter tuning**: This artefact will remain the most essential artefact required during tuning the model to represent the real system. This artefact will host the model calibration and adaptation related mechanisms (frameworks, tools).

After benchmarking the DT architecture, the analysis of the research-challenges discussed above (Section 2.2) is now again required but from the DT perspective. The analysis of the issues is presented below where the extra columns provide insight into the model calibration and adaptation-related challenges from the DT perspective.

The analysis (Table 2.1) shows how DT offers a solution. This can be achieved by either providing near-real-time physical asset-related data or facilitating the integration of

Table 2.1: Analysis of the adaptive modelling-related research challenges from the DT perspective.

Related Area	Adaptive Modelling related		Utilising DT concept	
	Issues	**Scope**	**Related Artefact**	**DT Support**
Data for model representation and validation	Delay in data obtaining from the survey reports	Automation in data acquisition.	Data-Mirroring	The ontology concept of DT (Bao 2022) can be utilised for data mapping.
	Data Inconsistency	Benchmarking of the validation resources requirements.	Performance Validation	Resources benchmarking approach is still required within DT.
Model Calibration and Adaptation	The traditional manual process of parameter estimation.	Process Automation	Calibration/ Adaptation	Achievable with artefacts integration as DT offers a comprehensive model (Barricelli et al. 2019).
	Time-consuming parameter search methods	Optimisation of parameter search methods.	Calibration/ Adaptation	Analytics under hybrid DT concept (Chinesta et al. 2020, Azangoo et al. 2022), can be utilised but requires the appropriate mechanism(s).
Online Adaptation related	Parameter uncertainty with spatial variability and complexity	Efficient adaptivity of model.	Adaptation	DT is accepted as an adaptive model (Rasheed et al. 2020), but still requires appropriate framework(s)/ mechanism (s).

different supporting artifacts (such as frameworks or tools) needed to adapt the model. However, some issues still demand additional framework(s) within the DT's artefacts.

2.4.2 Additional benefits offered by the Architecture

The adopted DT architecture also facilitates establishing simulator (or parametric model) based DT for which the frameworks are still missing in the literature. The DT establishing i.e., enabling and maintaining tasks are also mostly related to the calibration and adaptation of the parametric model.

Simulator-based DT enabling

In today's engineering fields that involve structures, the presence of simulators and tools for creating geometric models requires a unified system to support all users in enabling simulator-based Digital Twin (DT) technology. By utilising the DT architecture shown in Figure 2.5, various components such as data mirroring, performance validation, and calibration can aid the simulator or conceptual simulation model in creating a real-time virtual replica of the physical asset.

Maintaining DT

While calibration is about dealing with parameter estimation, maintaining a DT means online updating of the parameter whenever required. The efficient and effective model updating artefact(s) within the architecture is anticipated to address the issue of model performance getting drift-out with time.

2.4.3 Challenges in DT features Adoption

The analysis (Table 2.1) suggests DT offers a significant role in addressing model calibration and adaptation-related issues, however, posed some implementation challenges. The practical limitation of the DT concept are summarised below:

1. The absence of data acquisition and management frameworks for DT data (Barni et al. 2018).
2. A general platform and a common methodology for a physical model-based DT creation is still missing (Aivaliotis et al. 2019b).
3. An enhanced and online tuning mechanism with more efficient algorithms to tune the modelling parameters is anticipated (Macchi et al. 2018, Aivaliotis et al. 2019a, Chakraborty and Adhikari 2021b, Feng et al. 2023) in order to fully realise the model calibration/adaptation artefact (Figure 2.5) of DT.
4. The proper framework for utilising historical and real-time data for the continuous adaptation of the digital twin is very limited in the literature (Gabor et al. 2016, Ye et al. 2019, Alnowaiser and Ahmed 2023).
5. Additionally, the solvers (frameworks or approaches) should provide the self-detection and self-adaptation features for the DT (Alnowaiser and Ahmed 2023).

2.4.4 Uniqueness of the research

This research explores predictive simulation modelling, especially in the field of SHM. It brings two important aspects to the forefront. Firstly, it suggests using the idea of a DT to solve challenges in adapting models. Additionally, it aims to offers practical solutions to overcome the difficulties of applying simulator-based DT in real-world scenarios.

The second significant aspect revolves around the unique area of study, experimentation, and showcasing outcomes within the SHM framework. While existing SHM research mainly revolves around elasticity models (as seen in the Table in Appendix A) for handling structural integrity against fatigue and identifying cracks in assets, this research expands the scope. It focuses on addressing another critical area of concern – the impact of material depletion on structural integrity and material properties. This choice is not guided by the previous reason alone, with other following rationales considered for choosing the corrosion domain for research and overall experimentation:

1. The domain tends to rely upon physics-based models for prognosis (Adey 2005),
2. The development of a simulator that can accurately replicate the behaviour and phenomena of corrosion, and
3. There is very little research on corrosion-related DT techniques (Adey et al. 2020, Peratta et al. 2021).

By concentrating on the area of electro-chemical reactions, specifically in the context of corrosion affecting engineering structures, this research also advocates for monitoring corrosion severity and incorporating it into SHM considerations. A detailed explanation of how the corrosion process model works and an exploration of the state-of-the-art use of the Digital Twin concept in this domain will be presented in the next chapter (Chapter 4).

2.5 Research Questions

The DT related analysis shows that it can effectively facilitate the model's calibration and adaptation requirements in SHM, but there are still implementation challenges (Barricelli et al. 2019, Rasheed et al. 2020).

In the scenario, where the DT architecture (Figure 2.5) is considered as the solution for model- calibration and adaptation-related issues, next, the goal should be addressing the limitations which hinder the application of the DT architecture. For this, the research from here onward will investigate the following research questions (RQ) posed aligning to the previous-discussed challenges and considering the DT's current developments:

Data requirement and collection:

RQ1: What is the recent development in data-obtaining related technologies that can be utilised to obtain near-real-time data for Digital Mirroring, especially when data is not readily available from direct sensors (for example data in survey reports)?
RQ2: What are methods that can be utilised to set the standard of the data resources required for model's performance validation during simulator-based DT enabling?

Efficient Parameter updating during the model calibration/adaptation process:

RQ1: What are the available techniques/tools that can be implemented for realising the integrated DT with the features of full automation for the calibration/adaptation process?
RQ2: How the solver can be utilised within the DT perspective so that it also assists in simulator-based DT enabling?

Model parametric adaptation and maintaining predictive capability over time:

RQ1: What are the benefits and challenges involved in adopting the analytical features under the DT concept to solve the issue of performance drift of the model?
RQ2: How can the features of AI and/or M/L under the hybrid DT concept be utilised to address the issue of parameters-related uncertainty and complexity arising with time?

2.6 Conclusion

This chapter included a brief outline of the physics-based model's function in SHM as well as some of its limitations concerning the need for calibration and adaptation. The model building process and the sources of uncertainties involved in it were explored, to gain the isight associated with the challenges related to model calibration and adaptation.

Next, the chapter discussed current developments related to the key challenges of model calibration and adaptation in SHM-related domains. Additionally, the state-of-the-art of Digtial Twin was presented, with a focus on domains linked to SHM, to address concerns related to model calibration and/or adaptation.

The DT idea offers a significant role in addressing model calibration and/or adaptation-related issues, however, posed some implementation challenges. The limitations of DT integration into SHM are then discussed, given that DT's online adaption has the most scope for use. The research questions are then put forth to address the model calibration and adaptation-related issues from the DT's perspective.

The research methodology will be covered in the following chapter. This methodology will serve as a guide for carrying out the research activities to achieve its goal. Based on the research challenges described and the questions provided in this chapter, the research plan will then be developed. The objectives and milestones will be established to help the research stay on track with the main goal, which is to address problems with SHM-related model's calibration and adaptation.

3 Research Methodology and Research Plan

Engineering and Design science research usually looks for bringing the theoretical and practical aspects of the research together. The research challenges discussed, and the research questions posed in Chapter 2 were approached from both theoretical and practical perspectives. The next task is to direct research to reach the solution(s) guided by research challenges and the questions.

This chapter provides a short overview of the research methodology that will be followed hereafter. It also covers the steps followed to reach this state of the research. Then in the next section, the research plan is discussed, which includes the categorisation of the research area followed by setting up the research objectives and milestones for each research area.

3.1 Research Methodology

The research methodologies used in engineering and information systems have been known by different names for many years (Gregor and Hevner 2013, Peffers et al. 2018). Nevertheless, these methodologies usually aim to create or develop artifacts for engineering and/or information technology. These typically involve the creation and testing of innovative solutions to address specific problems or challenges in a particular domain as they arise or are identified.

Among different understandings, this research follows the industry-as-academia methodology (Potts 1993) which is supported by continuous interaction between the industrial world (real-world problem) and academia. This mixed approach allows for literature-based research as well as the creation and testing of practical solutions that can be implemented in real-world settings. The process typically involves multiple iterations and the use of feedback from stakeholders to refine the solution. This methodology is chosen over others because of its demand of academic novelty and emphasis on practical solutions at the same time.

3.1.1 Industry-as-Academia Method

The benefit offered by the Industry-as-academia method is emphasising the intervention during the research increasing the potential of the practical utility of the output, which is unlikely in typical research-then-transfer methods with detached analysis. However, it also posed a challenge to balance the demands of methodological rigour that they share with purely curiosity-driven scientists, with the demands of practical utility that they share with utility-driven engineers (Wieringa 2010).

The major goal of this collaborative research is learning with practice and enhancing knowledge from performing research to systematically building the artefacts. Furthermore, the substantial outcome of the research should be innovative and bring interesting knowledge.

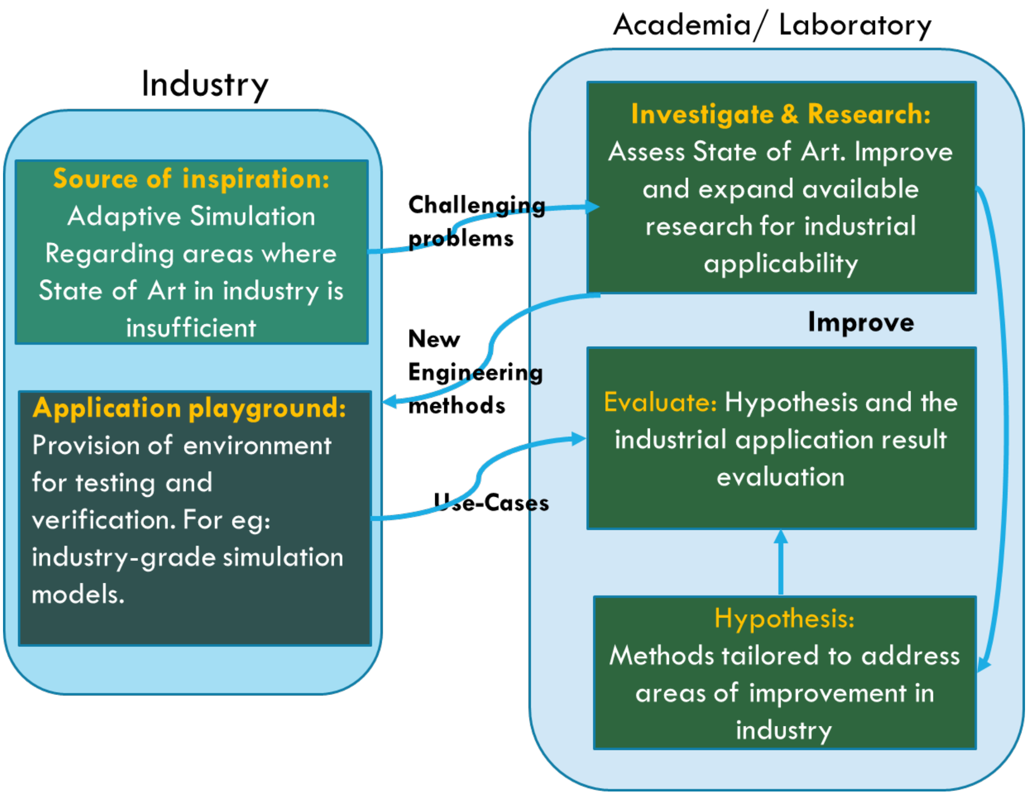


Figure 3.1: Project adopted Industry-as-academia methodology of research (Potts 1993).

Industrial Role

The source of inspiration for this research is the state of the art in adaptive simulation from an industrial applicability perspective existing in most of the domains. For example: In the field of corrosion and Cathodic Protection (CP) modelling, a commercial simulator like BEASY™ is available for CAD modelling, meshing and the simulation of corrosion phenomena. The current state of the art in BEASY-based Cathodic-Protection (CP) modelling will be discussed in the next chapter which will further justify this research requirement in a certain domain like CP modelling.

The next industrial role after passing the challenging problems to the academic researcher will be providing the environment for testing and verification of the proposed solvers (frameworks, approaches or methods). During this research, the modelling and the DT-related experimentation will receive such aid from the BEASY™ tool and CP modelling.

Academia-Research Task

The role of academia is to approach problem(s) from an academic perspective to have a novel solution. This research conducted an integrative literature review (Torraco 2005) to explore the state of the art from literature perspective in SHM-related adaptive simulation and associated research challenges. The integrative review was chosen over other types of literature reviews because it provides a more comprehensive approach, allowing for a broader understanding of the issues and the generation of new frameworks and perspectives. This can help design researchers to build on existing knowledge, identify research gaps, and develop more effective research strategies.

Following this integrative approach, novel Digital Twin (DT) concept in the field of modelling was also investigated and gaps were identified within it. Findings suggest DTs offer some benefits in addressing the model calibration and/or adaptation-related issues, however, there still are areas for improvement within DTs. Additionally, the state-of-art of a specific SHM-related domain, namely Cathodic-Protection modelling was investigated (presented in Chapter 4), to elaborate on the source of inspiration for this research following the integrative literature review approach.

Other academic-related tasks include setting research objectives, making hypothesis(es) of possible solutions, giving the shape of artefact(s), evaluating, etc. (Figure 3.1) in addressing the areas of improvement in both academia and Industry. The output of the academia-research could range from design theories (Gregor et al. 2007) to concepts, models, methods/frameworks, or instantiations (March and Smith 1995).

3.1.2 Experimentation and Evaluation

While conducting such research, it is crucial to have an evaluation process for artefacts that are created (March and Smith 1995, Venable et al. 2012). Moreover, a rigorous evaluation will assure that the output of the research fulfils the desired goal.

The cathodic-protection model(s) will be used in the evaluation of the research outcomes. The rationale behind selecting this particular domain is already discussed in Chapter 2 (under sub-section 2.4.4). The elaborated explanation regarding working principles of this CP model is presented in Chapter 4.

Simulator adoption for Modelling

From the current state of the art, this research takes into consideration the availability of a simulator such as BEASY (Danson et al. 1982), that can be utilised in CP modelling. BEASY, is a commercial parametric simulator specifically designed to simulate the behav-

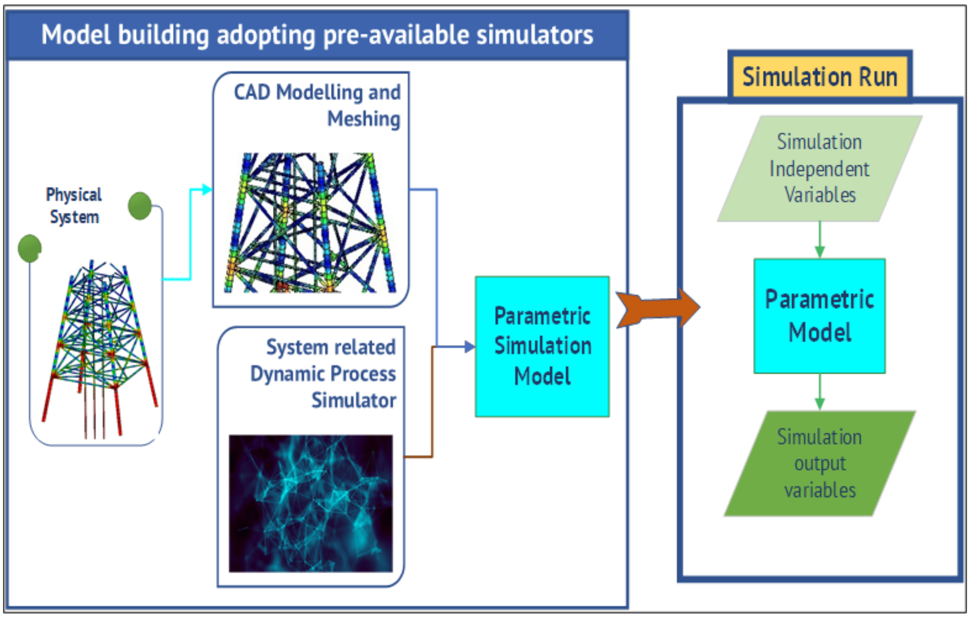


Figure 3.2: CP model building with the adoption of pre-available simulator and requirements for its simulation run.

ior of galvanic corrosion problems and cathodic protection designs. The software utilises the Boundary Element Method (BEM) to provide a numerical approximation of Laplace's equation (Brynjarsdóttir and O'Hagan 2014, Zienkiewicz et al. 2005). In this research, BEASY will be used to assist in the computerised modelling stages (Section 2.1.4). The software is selected based on its BEM-based numerical simulation abilities and its practical prognostic applicability in the field (Al-Otaibi 2010, Cui et al. 2015, Kim et al. 2017).

It will be assumed that the tool has been validated for its numerical approximation ability, meaning that its accuracy is accepted for mathematical modelling-related uncertainties, and only geometric and parametric uncertainties will need to be addressed.

CP model(s) to represent the real CP system(s) will be built using the simulator and the solution on addressing the SHM-related model calibration and adaptation challenges will be tested. In the entire research process, evaluation mechanisms can be iterative, however, the final evaluation will take place after the complete development of the artefact. Evaluating an artefact sometimes may be a complex task as the performance of the artefact is tightly coupled with the reason why the artefact was created.

Data for CP model calibration and adaptation related analysis

The calibration of the model for an existing physical system is usually performed against the data from the real system. In the case of the CP system, the data types that can be practically feasible from the structure and also predicted by the CP model are: a) Surface Potential (mV), b) Normal Current density (mA/m^2) and c) electric field (mV/m) (Adey 2005). However, the quantitative and qualitative data dependency for validation as well as the calibration of the model is always influenced by the complexity of the model.

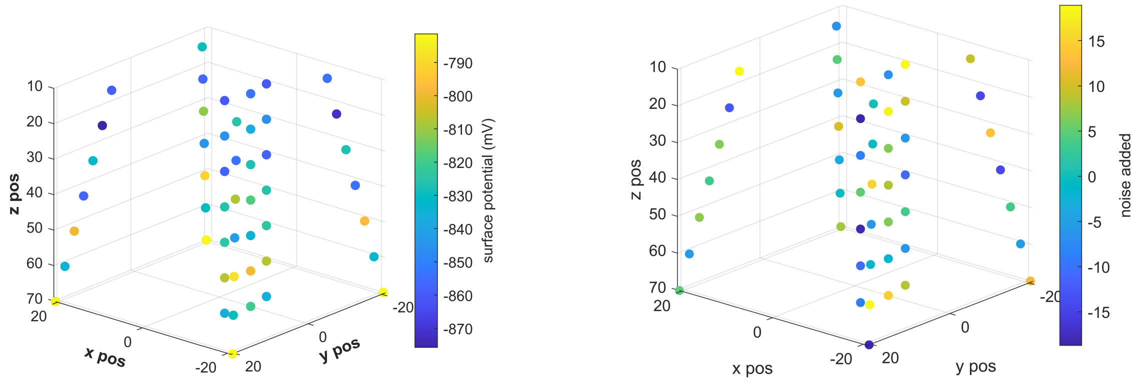
While the model will be built for the real-existing CP system for an offshore structure, obtaining real-world data might be limited (Soomro et al. 2022). Therefore, calibration data during experimentation and analysis will be generated from a virtual reference model simulation run with fixed parameters suggested by design rules. This approach which involves the use of artificially generated data is implied by researchers in the similar situations (Bi et al. 2017, Jensen et al. 2017, Zhou and Tang 2021). A benefit of such procedure is it avoids the influence of other factors than the required one into the experiment.

Given that a substantial portion of CP-related data is obtained from surveys, these data are susceptible to errors. These errors may arise due to various factors, such as the inability to position or orient the measurement probe accurately on the structure or anode (Adey et al. 2012), or errors due to the positional misplacement of the probe in terms of elevation (Stutzmann 2017). However, recent advancements in precise measuring tools and data handling techniques have led to the underrepresentation of survey errors during calibration related experiment within research (Marcassoli et al. 2013, Stutzmann 2017). This can be attributed to the ability to identify and removing of outliers (noise) during experimentation. Other significant attribute is the development of tools and techniques to nullify errors, ultimately limiting errors negligible level when equipment handled properly (Melios et al. 2023). Despite these developments, anticipating an inherent level of error within measurement data and considering it during the calibration would be a good practice. As synthetic CP-related simulation data will be used for calibration during this research, the incorporation of anticipated synthetic near real-world measurement errors to the data would be preferred. This emulation can be achieved through the introduction of noise into the data (Amaya et al. 2014). Various methods can be employed to induce noise in such situations, encompassing systematic errors with consideration of measurement uncertainties from historical data, random noise, and adjustments to uncertain parameter values (Inigo et al. 2021, Burés and Larrosa 2023).

During the presence of unrecognised pattern of errors, with ample data are available it is often assumed that the nature of the error (uncertainty) follows a Gaussian distribution (Soomro et al. 2022). When considered this one, the noise will be added to each synthetic data (x) at each data point to have data with noise given as: $x\_noisy = x + noise\_factor$, where, the $noise\_factor$ follows Gaussian distribution with given mean and standard deviation of the anticipated noise (error) value. However, in scenarios where data availability is limited, this assumption may not hold true, and would be preferable to follow subjective patterns specific to the given context. For example, Sapkota et al. (2022b) undertook random noise injections ranging from $+2\%$ to -2% to replicate real-world errors within the data. This approach aligns with the broader research discourse on refining data accuracy in CP-related studies (Amaya et al. 2014, Stutzmann 2017) and state-of-art in development with CP related data collection (Soomro et al. 2022, Melios et al. 2023). Following the similar approach, this research enriched noise into artificially generated calibration data by introducing minor perturbations/errors (approximately uniform within the range of ± 2). The noise is added to each synthetic data at each data point using the following mathematical formula and is slightly different to above.

$x\_noisy = x(1+noise\_factor)$ with, x representing the original data point and $noise\_factor$ as a random number in the range of [-0.02, 0.02], corresponding to a $\pm 2\%$ noise level. This approach ensures that each data point is perturbed by a small fraction of its value, mimicking the variability observed in actual measurements. The representative example is presented in Figure 3.3.

It is equally, if not more, important to consider the errors in measurement data during the calibration process, i.e., the effect of this error into the calibration accuracy. Techniques such as the Monte Carlo method and/or Bayesian methods are often employed



(a) (b)

Figure 3.3: a)The initial synthetic surface potential data obtained after a simulation run with a reference CP model b) synthetic noise to be incorporated to the initial data (in mV)

to address such situations during the calibration as an Inverse-Problem (Zapoměl et al. 2016, Ramancha et al. 2020, Soomro et al. 2022).

3.2 Setting research objectives

The research endorses the use of the Digital Twin (DT) concept and adopts a DT architecture (Figure 2.5) to address the model calibration and adaptation related challenges, but also acknowledges that limitations in DT application must also be addressed. Therefore, the objective of this research, in other words, is to contribute to overcoming the challenges in the practical implementation of the suggested artefacts for the DT architecture discussed in Chapter 2. The focus among them will be more on the calibration/adaptation artefact.

To reach its objective, the research challenges-related areas are required to be investigated separately but following the DT concept. It will enable this research project to reach the solutions within each area but in a holistic way i.e., under DT architecture. The research areas are therefore categorised considering the model calibration and adaptation-related problems in SHM.

Research Areas:

Following are the categorised research areas (RA):

The research activities within each of the areas will be carried out in overall 3 steps:

- Setting research area-specific objectives and goals (milestones)
- Design and Development,

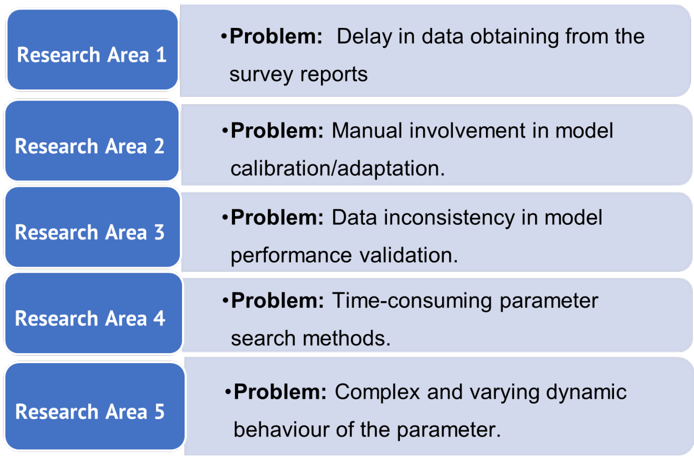


Figure 3.4: Research Area Categorisation according to the model calibration and adaptation related Research Problems

- Evaluation with Experimental demonstration

Aligning to the research questions, the research-area-specific objectives and milestones are set. While the objectives are towards undertaking the research in a theoretical way, the milestones set will also undertake the practical applicability of the research outcome.

RA1: Data-Acquiring for Digital-Mirroring

Objective:

1. To investigate the recent development in data-obtaining related technologies that can be utilised to obtain near-real-time data for Digital Mirroring when data is not readily available from direct sensors (for example data in survey reports).

Milestone:

1. Data-extraction and data-acquiring framework aligned with the DT concept that can be automated.

RA2: Automation in Model Calibration and Adaptation process:

Objectives:

1. To investigate the applicability of scientific software and/or tools in having a platform for automated calibration and/or adaptation following DT concept.
2. Experimentation analysis on the benefits offered by such an integrated platform i.e., platform with scientific software for analytics and the simulator for simulation task.

Milestones:

1. A platform with incorporated analytics performing the model calibration with automation.
2. The utilisation of the platform for automated DT enabling and maintaining.

RA3: Data Benchmarking for Model Calibration and Adaptation

Objective:

1. To investigate the DT concept and its associated recent development (such as integrated platform) on standardisation of the data resources required for model calibration and adaptation.

Milestone:

1. Approach for benchmarking the required quantity, quality, and variability of validation data and performance metric(s) for model online calibration and/or adaptation within DT.

RA4: Online Model calibration/adaptation: Addressing Input Parameters Uncertainties

Objectives:

1. To investigate the significant methods/procedures in dealing with parameter uncertainties during adaptive simulation.
2. Providing a standard framework for efficient and reliable parameter updating during model calibration and/or adaptation.

Milestone:

1. The online model updating/adapting artefact (a mechanism) aligned to the DT concept, that uses the best suitable optimisation algorithms/methods.

RA5: Maintaining the predictive capability of the Model over time

Objectives:

1. To investigate the significant analytical features that aid DT in maintaining its predictive capability despite changes and uncertainty arising with time.
2. To investigate the possibility to pave the way for data-driven prediction towards providing an alternative to the physics-based model in order to avoid its complexity arising with time.

Milestones:

1. Features facilitated by analytical and database that support in providing DT with the predictive capability despite changes and uncertainty arising with time.

2. The past pattern-based i.e., data-driven techniques and/or tools for online prediction, also capable of extending the range of predictive applicability utilising the Digital Twin aspect.

The suggestions/solutions will be proposed aligning with the research objectives in the subsequent chapters. Additionally, the industrial goals are set up to ensure the practical applicability of this Industry-as-Academia research which are presented in Appendix B

3.3 Conclusion

This chapter presented a description of the research methodology being followed and to be followed. Different aspects of the Industry-as-Academia methodology were discussed which will be used as a process guide by this research to reach its goal. Also, this chapter presented the research plan with categorical research objectives and milestones for each of the research area identified.

The next task for this research includes making hypothesis(es), investigating from different aspects, giving the shape of artefact(s) in addressing the areas of improvement and evaluating. Before moving to this, Cathodic-Protection domain that follows SHM and also uses the physics-based model for it will be discussed to have more insight into the research problem. This domain will not only provide to explain a source of inspiration in SHM but will also provide the application playground (Figure 3.1) to evaluate the outcomes of this research.

4 Problem Domain: Modelling and Digital Data Mirroring

Chapter 3 discussed the research methodology being undertaken and presented the research plan. The industrial source of inspiration and necessity of evaluation of the outcomes of the research were two of the discussion topics.

This chapter in the first section discusses the state-of-art of adaptive simulation and DT in the domain of Cathodic-Protection (CP) Modelling. This will elaborate on the source of inspiration for this research by explaining the specific SHM-related domain which also provides the experimental platform to analyse and validate the outcomes.

In the next section, this chapter presents the findings from the work under the first research area i.e., *Data-Acquiring for digital-mirroring* with the DT concept utilised. One of the issues in Digital Twinning presented in Chapter 2 is the hindrance in near-real-time data for the situation where physical-twin-related data lies in the survey reports. This hindrance to the establishment of a crucial feature of DT i.e., data mirroring in CP modelling is required to be assessed. So, before continuing the research on model calibration and adaptation issues, this chapter provides a framework for the acquisition of the data utilising the existing DT concept together with the other applicable tools (frameworks).

4.1 Corrosion and Cathodic-Protection Modelling: State-of-Art

Corrosion is the degradation/depletion of the material as a result of exposure and interaction with the surrounding environment. In metals, corrosion occurs due to chemical or electrochemical reactions occurring with their surroundings. Corrosion is a common problem for most of the existing infrastructure adversely affecting the structural load-withstanding ability (Chen et al. 2017). When corrosion-caused damage reduces the load-bearing capacity of a structure to extreme conditions, catastrophic failures like sudden collapse may occur. Therefore, during prognostic analysis i.e., predicting infrastructure's behaviour over the service life, it is important to consider the effects of corrosion on the structure.

During electrochemical corrosion, numerous anodic (oxidation) and cathodic (reduction) reactions occur. The principal anodic reaction during the corrosion process for structures with steel is given by (Ahmad 2003):

$$Fe \rightarrow Fe^{2+} + 2e^{-} \tag{4.1}$$

while the principal cathodic reaction is represented by:

$$O_2 + H_2O + 4e^{-} \rightarrow 4OH^{-} \tag{4.2}$$

The depletion or degradation of the metal components occurs through the anodic reaction (Equation 4.1), which is complemented by a cathodic reaction (Equation 4.2).

The structures especially metallic and buried/submerged such as offshore turbines, storage tanks, pipelines, etc at high risk of corrosion should be protected from it. The cathodic protection (CP) method and well-bonded coating/paintings are the most effective and most frequently used measures for protecting the surface from corrosion (DeGiorgi 1993, Adey 2005).

Among the two, Cathodic Protection (CP) method is very often used for the protection of underground or underwater (seawater) metallic infrastructures from corrosion (Angst 2019). The principle of CP is the conversion of the protection required areas on a metal surface into the cathodes, as electrochemical corrosion occurs only in the anode. When the structure is coupled with anodes, the corrosion i.e., the anodic reactions occur to the anodes instead of the main component of the structure. This means the flow of current is reversed with the cathodic reaction (Equation 4.2) occurring on the structure's surface. In the case of sacrificial anodes (for example zinc), the anode metal depletes due to the inherent potential difference between the anodes and the structure.

There are two general types of CP systems: Sacrificial anodes CP (SACP) and impressed current CP (ICCP) (Adey 2005).

Sacrificial anode Cathodic Protection (SACP): A sacrificial anode is a metal that is more reactive (anodic) than the metal of the structure it is protecting. This anode is connected to the structure, forming a galvanic couple, which allows the anode to corrode instead of the structure. The most commonly used sacrificial anodes materials are: Zinc and Aluminium (Adey 2005). The effectiveness of cathodic protection depends on the anode material having a large enough natural voltage difference to produce an electrical current flow. If applied effectively, this cathodic protection can provide complete protection to any section of the structure for its entire lifespan.

Impressed current Cathodic Protection (ICCP): Unlike sacrificial anode systems, impressed current cathodic protection (ICCP) uses an external DC current to protect a structure from corrosion. This involves applying a current through long-lasting anodes without any loss of material from the anodes. A corresponding cathodic reaction (Equation 4.2) which generates hydrogen gas occurs on the structure and prevents corrosion. When an AC power source is used to supply current, it is converted to DC by a rectifier before supplying to the ICCP system. The electrical current is adjusted throughout the lifespan of the ICCP system to achieve an efficient level of protection for the structure.

The appropriate design of any of the CP system is always anticipated for effective and cost-efficient protection of the system (Tezdogan and Demirel 2014) while the selection of one of the choices is decided by factors such as the source availability, feasibility, maintenance requirement, etc. The CP designs are mostly reliant on the shape of the structure, which means the shape could demand more protection in some areas than others. Furthermore, the design of the CP system depends upon the design rules about the performance of other protection measures such as coatings, in particular the rate at which coatings are to be assumed to degrade over the life of the structure.

A computerised mathematical model can be used to design a proper CP system (Adey et al. 1990 2012). Before this, the model requires to accurately emulate the current and potential distributions on the structure's (electrodes) surfaces in contact with the electrolyte (soil, water, concrete).

4.1.1 Numerical Modelling

With the assumption that the electrolyte is homogeneous, the distribution of electric potential obeys the following Laplace equation which is the governing partial differential equation (PDE) for electrochemical corrosion (Newman and Balsara 2021).

$$-\nabla\left(k\nabla\phi\right) = 0 \tag{4.3}$$

where,

k = electric conductivity, ϕ = electric potential and ∇ is Nabla operator

The PDE equation (Equation 4.3) represents the distribution of electric potential on the electrode-electrolyte surface. To obtain predictions for the response data, i.e., surface potential together with current density data by solving the PDE , a numerical method is often necessary. Computational numerical methods such as the Finite Element Method

(FEM) and Boundary Element Method (BEM) have been successful in electrochemical responses modelling (Liu and Kelly 2019). However, the BEM method is better suited for this purpose since it only requires modelling the interface of the surface and the electrolyte, unlike the FEM which considers the entire volume (Adey et al. 1990, DeGiorgi et al. 1992, Liu and Kelly 2019).

Computerised modelling utilising one or multiple numerical methods has been established as a powerful tool for the study of complex system responses and is therefore an appropriate means for simulating corrosion (DeGiorgi 1993, Adey et al. 2012, Liu and Kelly 2019). These predicted response data not only facilitate monitoring the system's protection state but can also be integrated to calculate the rate of depletion of the anode in Sacrificial Anode Cathodic Protection (SACP) systems. Moreover, the model can be used in the design process to optimise the protection provided to structures and has also been recognised as an effective tool to reduce unplanned downtime and failures in long-term integrity management (Stutzmann 2017, Liu and Kelly 2019, Adey et al. 2020).

However, creating a virtual replica of an existing CP system through model calibration is challenging, resulting in a lack of practical implementation of the model. Direct in-situ inspection data are limited (Hawari et al. 2020), and matching data from multiple inspections can be difficult due to the use of different tools and techniques with varying levels of resolution and accuracy (Stutzmann 2017, Kim et al. 2021). Continuous adaptation requirement of CP model and optimisation of the CP survey data collection procedure are other issues lacking solution till the date (Adey et al. 2020).

The steps involved in CP modelling including the calibration process can be divided into three stages:

1. **Pre-Processing Stage:** This stage includes defining the geometry of the problem, creating the mesh, and the categorisation of the elements to be able to assign properties and corresponding parameters to them.

2. **Model Setting and Simulation**: This phase includes specifying the boundary conditions, setting the value of the parameter and implementing the numerical method for process simulation. The polarisation behaviour of the materials serves as a boundary condition for running simulations, and other parameters such as the conductivity of the surrounding environment are also required (Adey 2005).

3. **Post-Processing Support:** This stage facilitates the visualisation and analysis of the simulation-predicted data.

4.1.2 Simulator-based CP modelling

In this research, BEASY tool is adopted for the CP modelling process, assisting in all three of the aforementioned stages required for CP modelling.

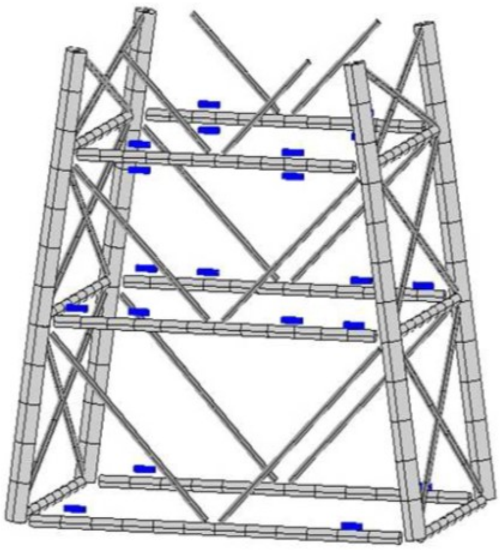


Figure 4.1: A geometrical model for a CP system provided with the sacrificial anodes (represented by blue cylinders) (Danson et al. 1982)

A geometrical representation of the representative SACP model that can be built with the BEASY software is presented in Figure 4.1. The data about geometry, data about meshing can be exported to text files together with the materials and surrounding related parameters data and fed to the solver (Figure 4.2) for the numerical approximation.

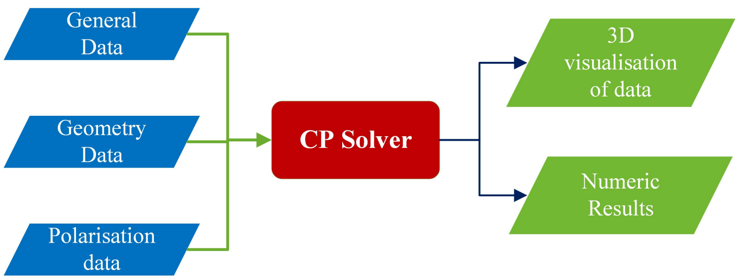


Figure 4.2: Inputs-Outputs for BEASY tool based Cathodic-Protection Simulation

The availability of simulator(s) in corrosion modelling motivates the idea of building a parametric simulation model of a CP system. The parametric model requires to feed into the parameter values before performing the simulation run.

4.1.3 Parameters and Calibration of the CP model

The common parameters required to run the simulator-based CP model are:

- **Polarisation Behaviour:** It gives the relationship between potential and current density and represents the electrode kinetics of the metal in the seawater. This polarisation related values (data) provides the boundary conditions to the model which are essential for solving the numerical problem.
- **Conductivity/Resistivity:** The conductivity of the surrounding medium/material are directly involved parameter in the numerical model.

Calibration of a CP model now means providing correct polarisation values (curve) related to the material and conductivity value related to the surroundings. While, providing the conductivity value is straightforward, changing the polarisation values for the materials would be tricky as it is often represented as a graph.

Polarisation curve-related parameterisation

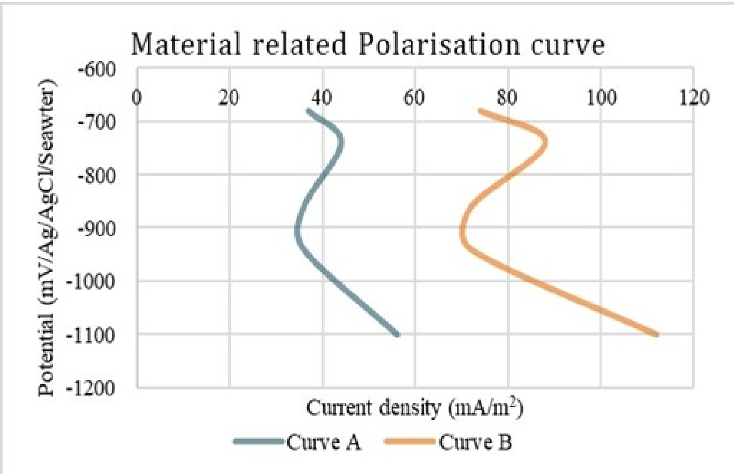


Figure 4.3: Two different polarisation curves, which can transformed to each other with the transformation p-value.

The requirement of a quantitative representation of the graphical polarisation curves is felt so that modification of the polarisation data can be achieved easily during the calibration. For this, a curve transformation factor is chosen as a variable (parameter) to allow the polarisation curve to be modified from the original design values. This parameterisation concept can be understood as a modification of the diffusion limiting current in the polarisation behaviour of the materials involved and also reflects the coating breakdown factor. The transformation vector or parameter is termed the *p-value*. *P-value*s are multiplicative factors that approximately account for new polarisation behaviour from a reference curve.

The task of calibrating the CP model for a physical CP system, is however, challenging as it is often done with the trial-and-error method and also data not being consistent as obtained from the survey (Jain et al. 2011, Stutzmann 2017, Adey et al. 2020).

4.1.4 Performance Prediction with the calibrated CP Model

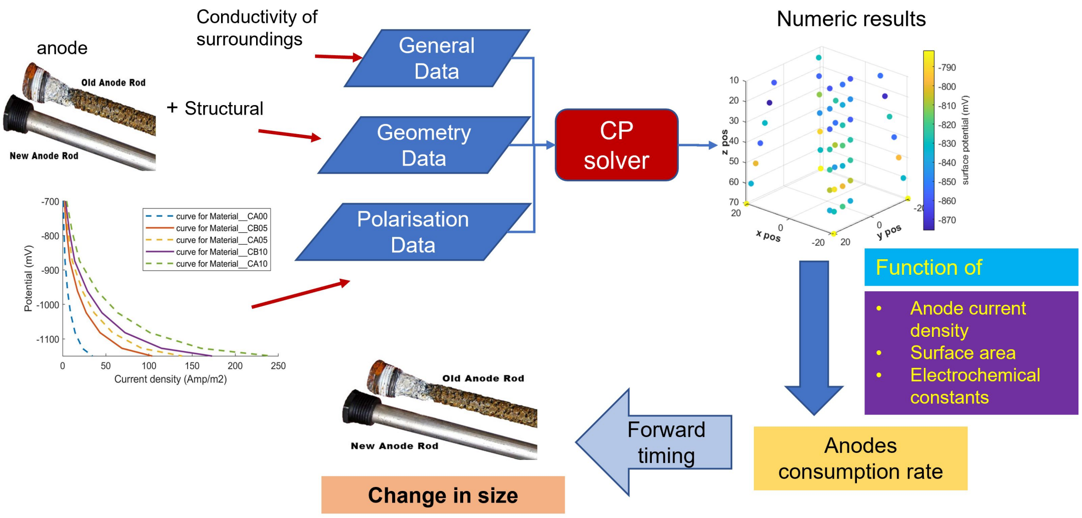


Figure 4.4: Performance prediction i.e., predicting anode status of the SACP system using the CP simulation model

The calibrated CP simulation model emulates the distribution of electrical potential and current density on the structural surface. From there, in the case of a sacrificial anode cathodic protection (SACP) system, the anode consumption rate is calculated by integrating the delivered current from the anodes over time (Figure 4.4). Making the forward prediction then helps to assess the immediate impact, plan to retrofit, improve design, find effective RUL of the CP system and ultimately have a cost-effective CP system.

While in the case of ICCP, on performance replication after calibration of the model, the required protection potential and current for the system are determined and thus the impressed current can be maintained efficiently.

4.1.5 Digital Twin concept in CP modelling

The Digital Twin concept has also appeared in the domain of CP modelling but in limited numbers to date (Adey et al. 2020, Peratta et al. 2021). The concept is understood as a 3D virtual twin of the current status of the asset supporting the engineer responsible for integrity management. Furthermore, it is also a computer model which simulates the physics of galvanic corrosion and the features of a cathodic protection system. This digital twin is anticipated to provide a clearer understanding of the protection provided to the asset and can be used to make predictions of the present and future protection of all parts of the structure (Adey et al. 2020). For example, the engineer can use the software to systematically monitor the differences between the model predictions and survey data to identify anomalies and to provide early identification of problems that will require mitigation.

The CP method though an effective way to prevent corrosion, it's monitoring still relies on diagnostic surveys. Hence, there is a need to focus on CP modelling and integrating the concept of a CP Digital Twin. Towards the realisation of DT for a physical CP system, the following model calibration and adaptation-related challenges need to be addressed (Adey et al. 2020, Stutzmann 2017, Kim et al. 2021, Peratta et al. 2021):

- Model Performance Validation relies upon non-standard data.
- Data from physical structures available from Inspection is often unstructured.
- Manual calibration during the realisation of the simulation model.
- Time-consuming simulation run requirement for Calibration.
- Model adaptation (re-calibration) demands a similar procedure to initial calibration.

4.2 Data Mirroring for CP System

From the DT-related literature analysis in Chapter 2, it is unquestionably accepted a DT should be a mirror of the physical twin. Though mirroring i.e., digital twinning is more than just data, data (state) representation is the core functionality anticipated for a DT. Literature suggests employing the proper data models such as ontologies to fully exploit the potential of Cyber-Physical Systems and IoT (Negri et al. 2017). Such data models should be explicit and semantic, and represent the formal concept in a particular domain. The electronic data measured by the sensors and flowing to the DT through the connection are the best form of data anticipated for the real-time data mirroring process.

4.2.1 Data acquiring challenge

The first benefit offered by the DT concept i.e., data mirroring in most cases is enabled by the sensors and the communicating space between physical and virtual space. However, not in every case data are obtained from sensors or remained in the best applicable form for making an assumption or performing the analysis.

Manufacturing and survey big data are of three major types, 1 – structured, 2 – unstructured, and 3 – semi-structured with unstructured data making up the major portion of the physical asset-related data (Gulgec et al. 2017). Data from physical structures available from the inspection and lying in the survey reports is often unstructured (Gulgec et al. 2017, Bayraktarov et al. 2019). For example, during a corrosion-related survey of offshore assets, data is normally contained in reports and EXCEL spreadsheets often with different measurement locations and inconsistent naming of the locations between reports (Adey et al. 2020). This leads to delay in obtaining the structural monitoring data from such a format that is essential for the information mirroring as well as the calibration of the model.

This hindrance on real-time or near real-time data acquisition not only affects the data mirroring via. the digital twinning, but also the predictive role of the DT. Moreover, it is not about obtaining only near-real-time data, the past data lying on such reports after extraction also plays role in enabling the DT at the present. As timely relevant data acquisition and asset components mapping together are the aspects desired within DT, there is a requirement for the automated data extraction module within DT to obtain near-time data from the reports.

4.2.2 Ontology Concept in Structural DT data Mirroring

Ontology defined as "*an explicit specification of a shared conceptualisation*" by Gruber (1993) is useful to define the common vocabulary and sharing for the reuse of formally represented knowledge. The heterogeneity in the structural and manufacturing data demanding an effective method of data organising motivates the utilisation of the Ontology concept in the system Digital-Twinning (Erkoyuncu et al. 2020). Ontology-based data management could help the integration of big amounts of sensed data which further can be accessed through smart analytics tools during decision-making.

Differences among individual parts should be highly valued and proper relations should be established to have a comprehensive model. Likewise, the Digital Twin with years

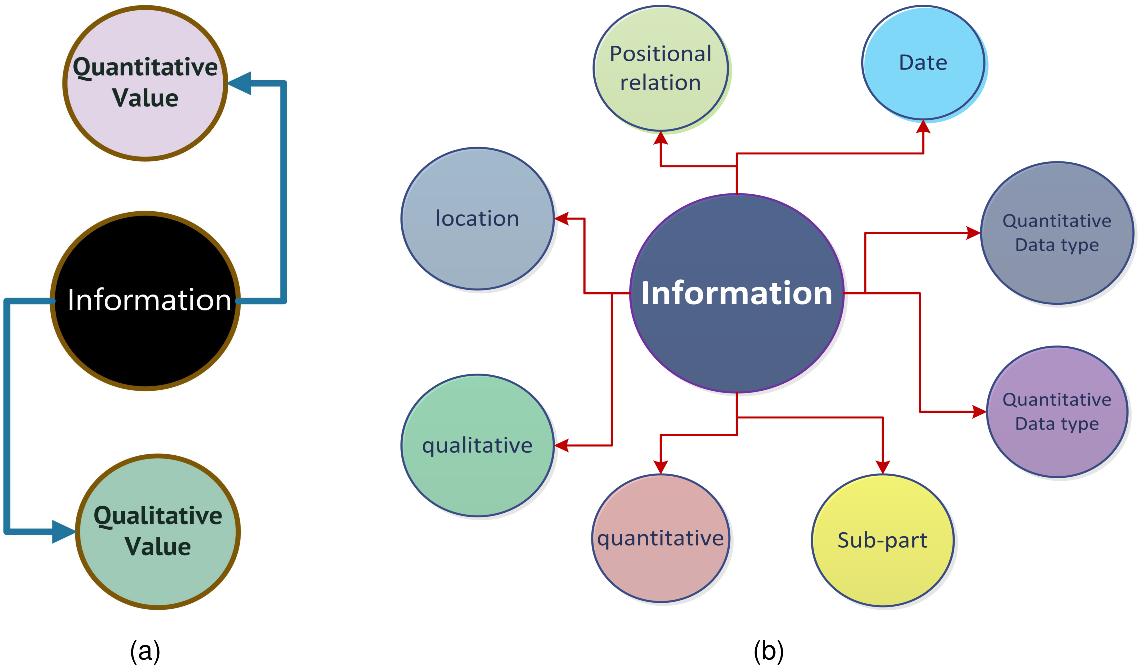


Figure 4.5: (a) Ontology initiated with only two attributes. (b) Ontology with multiple attributes provided.

contains a vast amount of data. This could be the data from sensors or inspection of the physical system or the self-generated data by the simulation at multiple stages. A data aggregation layer is required for the Digital Twin to organise the data hierarchically (Bazaz et al. 2019). This is where the ontology concept is attracting interest (Erkoyuncu et al. 2020, Bao et al. 2022). With the adopted ontology concept for providing the relationship of the data entries, the digital twins can offer effective historical data search and convenient real-time analysis for the required sub-model. Ontologies provide benefits of representing knowledge that can be shared between different entities establishing a common understanding of information. Furthermore, the ontology-based data organising method can provide expansibility to accommodate the new assembly or analytical measure that may arise in the digital twin/asset during its life cycle. This concept is also illustrated in Figure 4.5, which shows that as new data types emerge, ontologies with a higher number of attributes can be created.

Moreover, a semantic model based on ontology can be utilised to realise the fusion of multi-source heterogeneous data (Liu and Cai 2020). Additionally, an ontology can be developed to define the information architecture of the digital twin by accommodating the possible modelling framework (Figure 4.6).

Though, Ontology idea has begun to get adopted in DT, the benefits of its data acquisi-

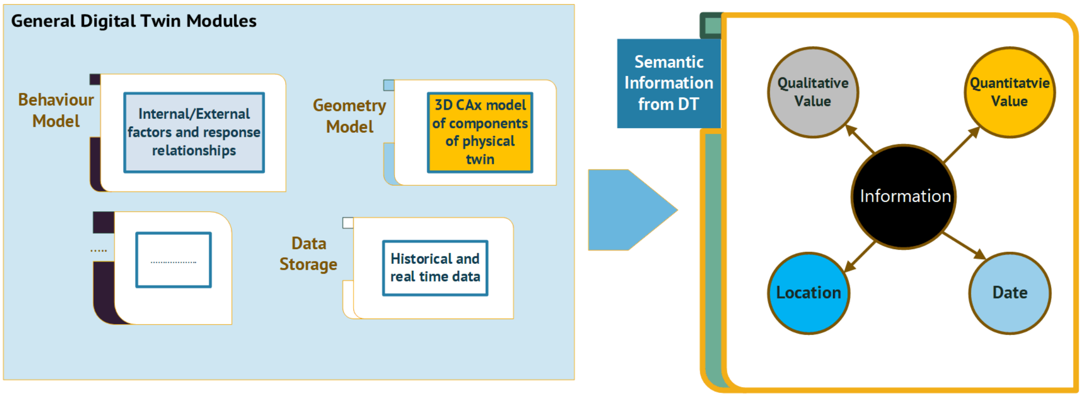


Figure 4.6: Semantic Ontology creation with pre-available information from Digital Twin

tion especially when data are in unstructured formats have not been sufficiently explored and utilised within Digital Twinning Concept.

Literature suggest, the use of Machine-Learning (M/L) techniques for text/image mining can significantly improve the discovery of information from unstructured data (Sun, Shang, Xia, Bhowmick and Nagarajaiah 2020). Similarly, the application of *Natural Language Processing* (NLP) is becoming popular for extracting quantitative and qualitative information from large heterogeneous datasets (Collobert et al. 2011).

Despite the potential benefits of Ontology and NLP, their combined use in acquiring and/or organising SHM related data remains limited to date (Gardner et al. 2021).

Considering the problem and the benefits that can be leveraged under the DT concept, a Natural Language Processing supported framework is proposed that utilises ontology from the DT to extract the data from unstructured resources.

4.2.3 Proposed approach: Ontology and N-L-P-based Data Extraction

The proposed framework utilises support from both Ontology and NLP concepts. The ontology developed from the DT is used to define the information architecture, while NLP is used to perform data-processing and refining tasks that align with the ontology to extract data from unstructured resources (Figure 4.7) and map it back to the DT.

In overall, the data-extracting process can be categorised into following four major steps:

1. **Pre-Processing :** Raw data available from the data resources mostly in physical report files, are first digitalised. Different available tools such as "*pdfminer*" (Shinyama 2015), and *python-pickle* (van Rossum and Drake 2009) within Python can be utilised for the task.

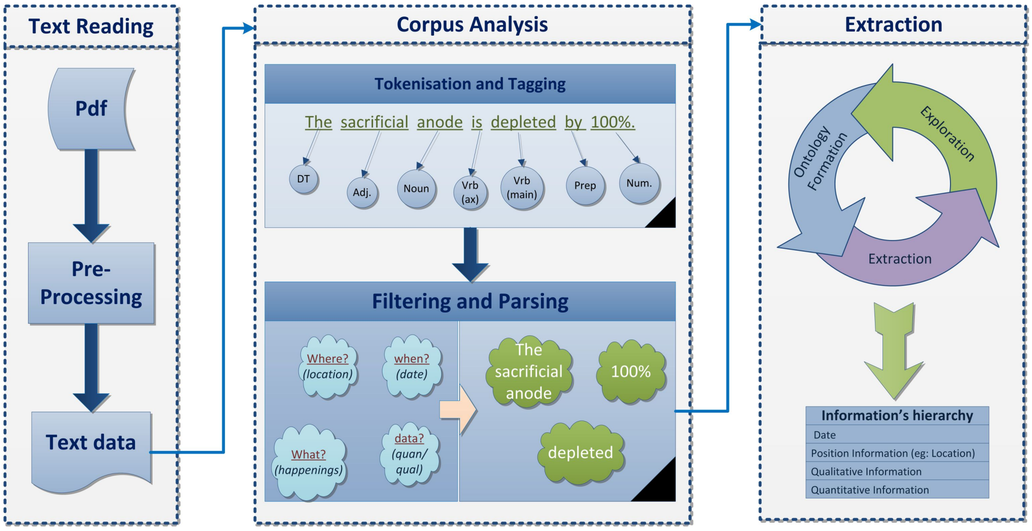


Figure 4.7: Overall process of Ontology and N-L-P supported data extraction.

2. **Corpus Analysis:** After having the digital format of data, NLP is initiated with corpus analysis which is performed with *"tokenisation"* and followed by *"tagging"* (Collobert et al. 2011) as shown in Figure 4.7. For the automation of the NLP, open-source tools such as python-based "*NLTK*" (Bird 2006), etc. can be utilised.
3. **Filtering and Parsing:** The digitalised and tagged data from the Step 2 are then further categorised into two types, whether the data consist of crucial information (i.e., qualitative or quantitative value) or action or relationship explaining secondary information (example: date, position, or action).
4. **Extraction**
   - **Ontology from DT:** The major challenges of Ontology creation could be leveraged with the concept of DT and vice-versa. When DT is provided with information from its design to commissioning to operation, the information serves on coping with the challenges of ontology creation (Figure 4.6). With the pre-defined components and attributes from DT, the ontologies from lesser to higher attributes can be initiated (Figure 4.5).
   - **Exploration and extraction :** The information that falls into the crucial category (for example quantitative or qualitative) is first mapped to the initialised information ontology and provided to their corresponding attribute(s). Then secondary data for the crucial information are explored, but within the limited range to avoid false information. Following the process when information is detected, they are stored within the attributes for the realisation of the complete Information Ontology.

These ontologies can later be exported and saved as structured data formats such as Excel, CSV, etc. using the tools like pandas in Python.

4.2.4 Data Extraction tool

A computer programme in “python” platform is built following the above-discussed (Section 4.2.3) algorithms and its demonstration is performed on extracting CP-related data.

Python packages (tool) used at different stages:

1. “pdfminer” (Shinyama 2015), “pickle” with python codes for reading the pdf file and for the pre-processing task respectively.
2. “Camelot” -python tool for extracting tables in pdf files.
3. “nltk” (Bird 2006), “re” for corpus analysis
4. Python code for filtering and ontology creation.
5. “numpy” and “pandas” on quantitative data handling and data exporting to standard formats like excel.

4.2.5 Performance evaluation of the tool:

Currently, the approach of the framework (tool) is semi-unsupervised, meaning that few of the latent thematic information are passed into the system which will be taken as reference response text during the clustering process.

Dataset Description:

Two different survey data report with the data after the inspection of two different offshore structure/components protected with Cathodic-Protection system are considered. Each of the report are provided in pdf format with the length of 75 and 200 pages each respectively. One out of two is the inspection report for a Mooring chain. A mooring devices for an offshore system refers to the arrangement of equipment and devices used to secure and anchor floating structures.

Anomaly		Status
AW	Anode Wastage	New
Date Reported	20/06/2018	
Criticality	High	
Details		

Chain 7 Connector Tube Anode

New

The anode located on the inboard side chain 7 connector tube is 100% depleted.

A contact CP reading of -659mV was obtained.

Figure 4.8: An example of unstructured data lying within the CP survey report

Ground Truth Data and Evaluation Metrics:

Ground truth dataset for each report is prepared manually which contains the correct data and anticipated extraction from the source datasets mentioned above. This dataset will be used to compare the data extracted by the NLP-assisted tool from the unstructured reports. To evaluate the performance by the comparative analysis, precision metric is used. Precision (also called positive predictive value) measures the accuracy of the positive predictions made by the model. It is the ratio of true positive (TP) predictions to the sum of true positive and false positive (FP) predictions. This metric is chosen over others for the dataset size limitation and manual preparation of ground data, which will again facilitate in precision analysis.

Results and Analysis:

The raw dataset was passed into the model for the extraction of useful information. Additionally, few of the information required by the tool are provided to track and explore the data, such as terms to recognise crucial ontology attributes. This information for ontology initialisation is believed to be obtained from DT later on (Section 4.2.3 Step 4a). With this initial information provided to the tool, the required information is then explored and extracted on matching from the parsed data (Section 4.2.3 Step 4b). The data extracted also get structured and then can be exported to formats like EXCEL, CSV, etc. (Figure 4.9).

date	location	sub_part	quantitative	Quan data type	qualitative	Qual data type
20/06/2018	Chain 7 Connector Tube	the anode			100%	depleted
20/06/2018	Chain 7 Connector Tube		-659	mv		obtained

Figure 4.9: Extracted data in a structured format from the survey report (Figure 4.8).

The model generates two output files: one containing data from the tables, which can be considered semi-structured data if present in the report, and the other for the remaining unstructured text data. A comparative analysis with the ground data revealed that over 90% of the tabular data was successfully extracted, while for other data, the extraction of over 50% of data was achieved with a precision of approximately 75% when limited thematic information was provided to the system. However, when a few additional spatial information elements obtained from the Ontology were passed into the system, the extraction accuracy improved to over 75%. Furthermore, with the inclusion of this semantic information, the precision of the extracted data reached over 95%.

Discussion:

The result shows that N-L-P with automation and supervision can be implemented to extract data from different report formats. The highlights of this results and analysis is how the model's performance improves with additional information from the Ontology. Additionally, it opens the possibility of supervised (and unsupervised) M/L algorithms' application on training the tool with time, i.e., on implementation to the additional data formats. This should gradually enhance the robustness of the tool and ultimately make it applicable to most cases and reducing the supervision requirement.

4.3 Conclusion

This chapter discussed the state-of-art for cathodic-protection modelling in the first section, including challenges in CP Digital Twin realisation. In the next section, an Ontology and N-L-P-supported data acquisition framework is discussed to provide the real-data mirroring features of the Digital Twin. The framework together with the Digital Twinning concept is for addressing the difficulties in obtaining SHM-related data (for example, corrosion and CP-related) where most of the data are obtained from surveys and lies in different paper formats.

From here, the research activities are directed towards achieving the other milestones set in Chapter 3, i.e., addressing the model calibration and adaptation issues with the DT concept utilised. This includes the steps: making hypothesis(es), investigating from different aspects, giving the shape of artefact(s) and evaluating. The steps will be repeated being guided by the research objective under each research area and will be presented in subsequent chapters.

5 A Design of Experiments Platform for Online Simulation Model Validation and Parameter Updating within Digital Twinning

This chapter presents the findings from the work under the second research area i.e., *Automation of Model Calibration/adaptation Process* with the Digital Twin concept utilised.

The research background analysis in Chapter 2 has demonstrated that simulation tools are available in several domains to reproduce the process phenomenon. Additionally, DT concept has evolved enough to offer data mirroring characteristics utilising the cutting-edge technology. Then, establishing the simulator-based DT requires the tailoring of the simulator-based parametric model within DT.

The aim is to provide a solution for addressing the manual calibration task but with the adoption of the DT concept. This demands the solver which also offers to address the practical limitation of DT particularly in implementing the automated model validation and calibration/adaptation artefacts within its architecture (Figure 2.5).

The objectives and milestones set in Chapter 3 for the corresponding research area are presented below which were set by considering the promise offered by the DT and the recent development in tools and technology:

Objectives

1. To investigate the applicability of scientific software and/or tools in having a platform for automated calibration and/or adaptation following DT concept.
2. Experimentation analysis on the benefits offered by such an integrated platform i.e., platform with scientific software for analytics and the simulator for simulation task.

Milestones

1. A platform with incorporated analytics performing the model calibration with automation.
2. The utilisation of the platform for automated DT enabling and maintaining.

The outcomes of the research activities within this research area are already published in Sapkota et al. (2021a), which includes the proposed solution approach and a case study for its demonstration.

5.1 DT concept as a Comprehensive Tool- Motivation

This research has accepted the DT concept as a comprehensive tool with the incorporated essential features like the potential to handle data, perform experimentation, and implementation of algorithms to calibrate and update the model. Also, Chapter 2 asserted some practical limitations for such comprehensive tool and propose an architecture to overcome the issue of DT practical implementation.

The recent advancements in analytical tools suggest that they can be utilised for the necessary analytical tasks involved in model calibration and adaptation (Coleman et al. 1999, López 2014). These pre-existing tools and algorithms offering analytical aids such as sensitivity analysis, data-sampling, and design optimisation, are promising for efficient model calibration and adaptation. This creates an incentive to integrate these tools into the model validating and calibrating/adapting components of the DT architecture (Figure 2.5). Additionally, a platform is expected to be developed to implement the appropriate experimental design within the DT concept (Aivaliotis et al. 2019b). Experimental design or Design of Experiment (DOE) is the term used for the techniques used to guide such experiments needed for the model calibration and/or adaptation process in an efficient manner (Cavazzuti and Cavazzuti 2013). Therefore, the potential of collaboration of the scientific software with the simulator and analytics is investigated to achieve the DT architecture anticipated in addressing the manual calibration issues.

Study shows, Scientific software like MATLAB and/or PYTHON have demonstrated their applicability to have an integrated platform where multiple modules can be utilised (Cruz 2016, Inzillo et al. 2017, Benaouali and Kachel 2019). This suggests the idea of software integrated platform to give the shape of the DT architecture which will ultimately assist in the online model validation and calibration task. This will also assist as an approach for establishing a simulator-based DT with automation in a situation where process simulators are already available in multiple domains.

5.2 The platform for Digital Twinning with Software Integration

A Design of Experiments (DOE) platform is proposed, to enable the practical application of the DT and also to address the above mentioned challenges. The platform will be created by integrating the process simulator and scientific software, providing essential features such as the ability to handle data, conduct experiments, and implement adaptive algorithms to update the model. This approach will facilitate the establishment of a simulator-based DT with automation, particularly in domains where process simulators are already available.

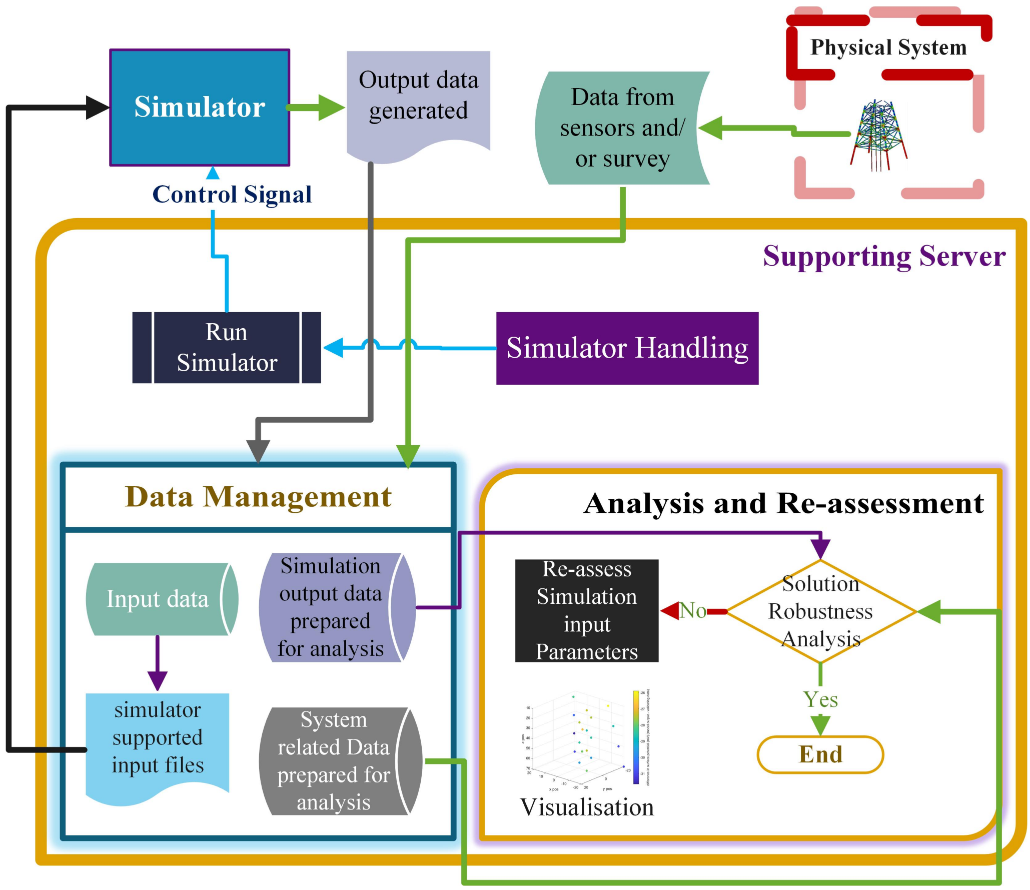


Figure 5.1: Integration of server and simulator for automation of the process of data management and analytical support required for the calibration/adaptation process.

The analytical tools that are typically used for model calibration/adaptation often employ design optimisation algorithms to recommend new parameters based on an analysis of the model's output for previous input parameters and real-world data. Such algorithms and tools are commonly available in scientific software packages such as MAT-

LAB, PYTHON, and others. This availability of optimisation algorithms and tools within scientific software motivates the use of such software for analytical tasks in addition to the automation of the process.

Likewise, when process simulation(s) can be achieved from commercial software (simulator(s)), the collaboration of the simulator(s) with the scientific software(s) offers to leverage the benefits provided by the analytical tools in predictive simulation. This leads to the idea of an integrated platform which combines both software and simulator and offers automation for model validation and calibration. The comprehensive integrated platform (Figure 5.1) is supposed to pave a path towards establishing a self-adaptive Digital Twin. For example, the applicable benefit of the proposed integration is a call for optimisation from analytics parallel to the simulation run.

5.2.1 Roles of Scientific Software as a Server

In a simulator-based DT, the simulator plays the primary role in simulation and prediction, while scientific software serves as a supporting tool. Given that DT may need to accommodate multiple process simulators and other tools, the scientific software can act as a server that facilitates communication between the tools. It provides a platform for experimentation and analysis, using internal or external interfaced tools and algorithms to assist in the analytical tasks required for parametric model establishment and performance enhancement.

Scientific software is also useful for pre-simulation tasks such as managing input-output data, filtering and mapping data, and preparing the dataset for the simulator. It can also connect to analytical tools to facilitate analysis during repetitive simulation. The server software can enable data visualisation to provide insight into the model's performance, either with or without comparison to data from the actual system.

The other crucial roles of the server in a simulator-based DT is to handle the switch of the simulation running process and provide the interface for relevant calibration/adaptation algorithms to the model. To accomplish this, the scientific software should either include a tool (Figure 5.1) or provide an interface to external tools for analytical support, including design optimisation. Scientific software such as MATLAB or PYTHON can be used to run user-customised optimisation algorithms or call external commercial or open-source optimisers, regardless of the optimisation type chosen.

Moreover, automated data flow between the platform and the data server allows for retro-perspective analysis if relevant time-series data are available. The platform should also provide a model performance validating criterion/algorithm, either separately or integrated into the platform, as the calibration (or adaptation) task cannot be performed without synchronised validation.

5.3 Case Study: Automated Calibration of a CP Model for an Offshore Jacket Structure

5.3.1 Experimental Model

A CP model (Section 4.1) for an offshore Jacket structure (Figure 5.2) protected by sacrificial anodes is built using the BEASY tool (Section 3.1.2). More details of the structure (or the CP system) can be found in Appendix C. The geometry and the discretisation task required before the simulation running is achieved with the aid of the tool as discussed in Chapter 4.

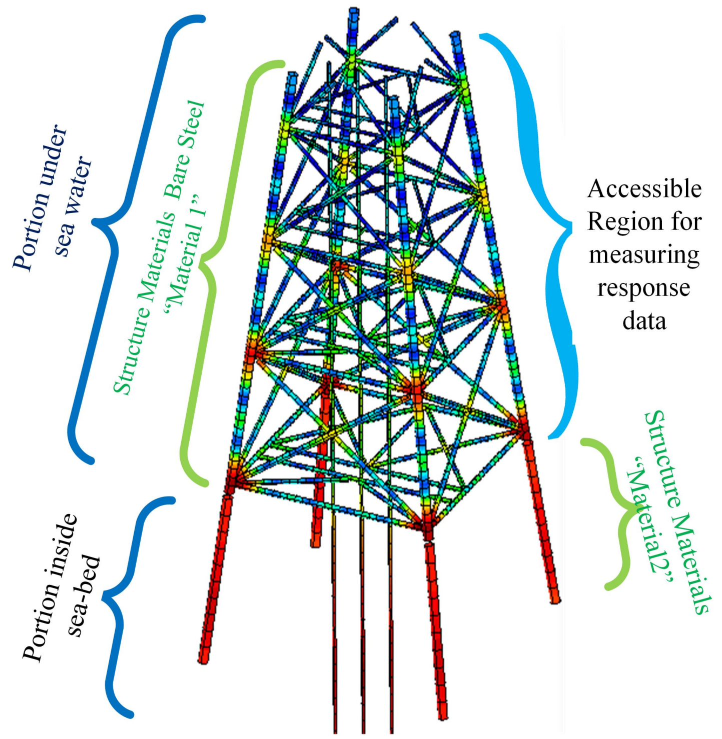


Figure 5.2: The geometry of the marine structure protected with sacrificial anodes adopted for the case study.

For the calibration experiment, the parameters related to geometry are not within the scope of the case study as such values can be obtained from design data and/or from the structure's geometrical measurements.

Then the parameters required to run the simulation for the CP model built for the structure are:

- Polarisation Behaviour
  1. P-value (Section 4.1.3) for *Material 1* related Polarisation curve of the CP system
  2. P-value (Section 4.1.3) for *Material 2* related Polarisation curve of the CP system
- Conductivity/Resistivity
  1. Sea water-related conductivity (Siemens/m)
  2. Sea-bed-related conductivity (Siemens/m)

5.3.2 MATLAB as a Server

For the approach demonstration towards the Cathodic-Protection DT enabling, integration of MATLAB-BEASY is chosen, where BEASY is the simulating tool and MATLAB will perform as a Server. The reason for selecting MATLAB for the role of server during the concept demonstration is its extensive data handling and analysis capability, plotting capability and the availability of different optimisation algorithms within it. Such features enable the assess-modify-check loops and can be completed in reduced computational time when programmed for automation.

Optimisation toolbox like *fmincon* and *fminunc* (Coleman et al. 1999) within MATLAB provides benefits in the continuous optimisation for the constrained and unconstrained cases respectively. These optimisation tools synchronised to the simulation environment fulfill the requirements of the model's parameter updating.

MATLAB also facilitates app building with automatic code-generation features. By utilising these features, along with the algorithms and tools required for calibration and adaptation, a user-friendly Graphical User Interface (GUI) (Figure 5.3) can be created. This GUI can be used by non-developers (engineers) for the calibration and adaptation task, providing a straightforward functional platform for automated model calibration and adaptation.

Figure 5.3: User-friendly GUI example built using the MATLAB which can be utilised for model calibration with automation.

5.3.3 CP Model's Operational Validation data

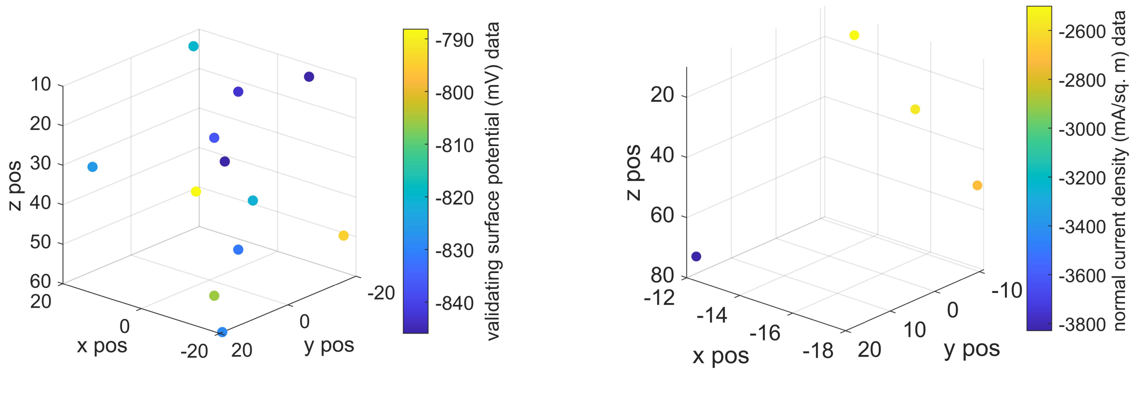


Figure 5.4: The calibration/validation data generated with simulation run of the reference model a) surface potential, and b) normal current density, corresponding to the selected data-points from the structure's surface (Figure 5.2)

Two types of validating data are considered- *surface potential* (mV) and *normal cur-*

rent density (mA/m^2). The validating data positions count from the structure's surface for two datatypes are 12 and 4 respectively (Figure 5.4). The benchmark of the validation/calibration data to ensure model's calibration is subjective and depends upon the different aspects. The justification for selecting the given data positional count will be presented in the next Chapter.

Due to limitations in applying (obtaining) real-world data, validating/calibration data for this experimentation are generated from a virtual reference model with fixed parameters suggested by design rules. Summary of the response data obtained from the simulation run of the reference model is presented in Section C.3 of Appendix C. Additionally, to make the data more realistic, $\pm 2\%$ error is introduced to the data considered for calibration from the reference model's simulation output.

The calibration data generated at this stage or even when obtained from the real world will be stored in a data-storage server so that can be accessed via the GUI (Figure 5.3) during the optimisation-based calibration process.

5.3.4 Tool selection for optimisation

Parametric optimisation is usually used to find the best set of design parameters where parameters $x = x_1, x_2, x_3, \ldots x_n$ can in some way be defined as optimal (Coleman et al. 1999). In a simple case, it is done with the minimisation or maximisation of some system characteristic termed as an objective function ($f(x)$) that is dependent on the variable x. In a more advanced formulation, constraints are added to the optimisation task, where x or $f(x)$, may be subject to constraints in the form of equality and/or inequality constraints.

To undertake the task of optimisation, different algorithms are available offering to provide effectiveness in the process. The Optimisation Toolbox within MATLAB provides several algorithms for solving a wide range of optimisation problems (López 2014). However, the selection of algorithms is based upon the model behaviour such as the parameters involved, linearity, non-linearity in the input-output, the search-space limits, simulation time, the constraints, etc. For example, many simulation models including the CP models built for a structure require a substantial amount of computation time. A single simulation run might require several minutes to several hours or even days. Therefore, the problem-suitable algorithm is required during the objective minima or maxima searching optimisation as the computational expense from the simulation restricts the total number of objective function evaluations.

At the initial stage of the research, the continuous optimisation method is adopted to establish the applicability of the integrated platform for model calibration with automation. The iterative optimisation-suggested simulation runs will occur for which the optimisation tool will be operated together with the simulator in a synchronous way. The algorithm chosen for this continuous optimisation is a gradient-based "*quasi-newton*" algorithm (Venter 2010). The reason for choosing a gradient-based approach as opposed to other optimisation techniques, such as genetic algorithms or neural networks, is that the CP model problem space is mostly monotonic. The monotonic behaviour of the CP model was found, after sensitivity analysis of the response data for the CP model vs. input parameters.

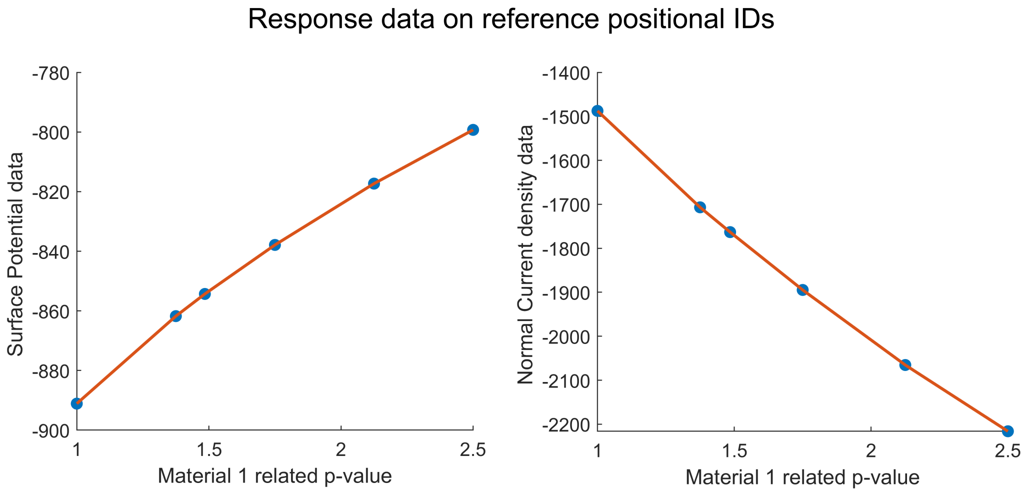


Figure 5.5: Sensitivity Analysis for Material 1 related P-value against two data types obtained from simulation at reference positional IDs

For this sensitivity test, multiple simulation outputs are obtained by varying the parameter *Material 1-related p-value*, while the other parameters including the sea-water conductivity are kept constant. Now for the analysis, two reference IDs selected from the above positional IDs (Figure 5.4) are considered and then the response data for the corresponding IDs obtained from the simulation are plotted (Figure 5.5). For both the data types considered, the monotonic behaviour is obtained against the material-related p-values. With such behaviour of the model, gradient methods decide the appropriate direction of the search during best parameter search by using information about the slope of the function (Venter 2010). This will save the simulation run time significantly as opposed to the random-step (stochastic) optimisation method.

Optimised values of parameters can, therefore, be quickly found with the quasi-newton, i.e., in a lower amount of completed laps. On this basis, the unconstrained optimisation tool “*fminunc*” within MATLAB provided with a “*quasi-newton*” algorithm is utilised. A comparison analysis is made between ‘*fminunc*’ against ‘*fmincon*’ before implementing the tool, both found within the MATLAB optimisation toolbox.

Table 5.1: Comparative analysis between two continuous optimisation tools from MATALB’s optimisation toolbox

Features/Tools	***fmincon***	***fminunc***
Minimisation of Non-linear functions	Yes	Yes
Toolbox designed for global minima	No	No
Search space reducing with bound & constraint	Yes	No
Gradient-based algorithm’s applicability to have a direction of search	Can be applied, however, could terminate before reaching solution due to constraints given	The newton-gradient step can be best utilised for medium-scale

In the experiments like this, where the constraints are not sufficient enough to reduce the search space, an unconstrained method with an efficient algorithm method (such as ‘*fminunc*’ with a *quasi-newton* algorithm) would perform better over the constrained one. However, both the tools implementing the gradient-based search could end up with the local minima. These tools/algorithms from MATLAB can be linked to the GUI (Figure 5.3) so that the user does not need to have more expertise and can use the tool appropriately.

5.3.5 Formulation of the Optimisation problem

Reducing the discrepancies between the validating/calibration data and the model’s output from the simulation is the goal of this minima-based optimisation problem. While the validation (calibration) data remains the same during the process, the input parameters of the model are changed which ultimately varies the discrepancy between the model output with the validation data. This discrepancy representing model performance measuring criteria, as a function of the relevant parameters, will serve as the objective function for the optimisation problem.

Objective function

The Normalised Root Mean Square Error (NRMSE) is taken as an Objective function during this experiment. The reason for selecting this criterion is its wider application in the engineering field for the minimisation problem. Assuming two different validating data types (Section 5.3.3), NRMSE here means Normalised Root Mean Square difference between validating data and model output data with the weightage constants (2:1) provided for the two data types. This weightage constant is decided based on the sensitivity of the data, reliability of the data type,etc. when obtained from the physical world.

$$f(x) = \sum_{i=1}^{n} k_i * \sqrt{\frac{\sum_{j=1}^{j=m} \frac{X_j^i - Y_j^i}{X_j^i}}{m_i}} \tag{5.1}$$

with,
”X” as model output data for the input parameter value ”x”
”Y” as validating data
”n” as total data types (n=2 in this case),
”k_i” as the weightage constant given to the data type i.
and ”m_i” as the response data total positional count for the data type i.

The continuous optimisation method or the minimisation problem (Equation 5.2) will seek the best parameter combination that gives the minimum objective function (5.1) value.

$$\min_{x} \quad f(x) \tag{5.2}$$

Initial State Primary Model and Parameter’s Value

At this stage, the research deals with two parameters case, as increasing the number of parameters could potentially lead to greater complexity. It can be easily obtained from sensitivity analysis that the parameters “*p-value of **Material 1** related polarisation curve*” and “*Sea-water conductivity*” are more sensitive to response data for the CP model. While focusing on these two highly sensitive parameters, other parameters’ values discussed above will be fixed i.e., kept constant.

As continuous optimisation is usually initiated with some best-known parameters value, the values are considered as in the Table 5.2. This is under the assumption that the involved engineer can always make some initial guess that can be based on the expertise and/or the design rules. These initial parameter values in this case and the bound of the parameter values during the implementation of constrained optimisation (for example

Table 5.2: Parameter's value provided to the initial primary model presumably from design data rules.

'Material 1" Polarisation curve's p-value	***Sea-water's Conductivity (Siemens/m)***
1.7500	3.0000

using *fmincon* tool) can be fed via the GUI.

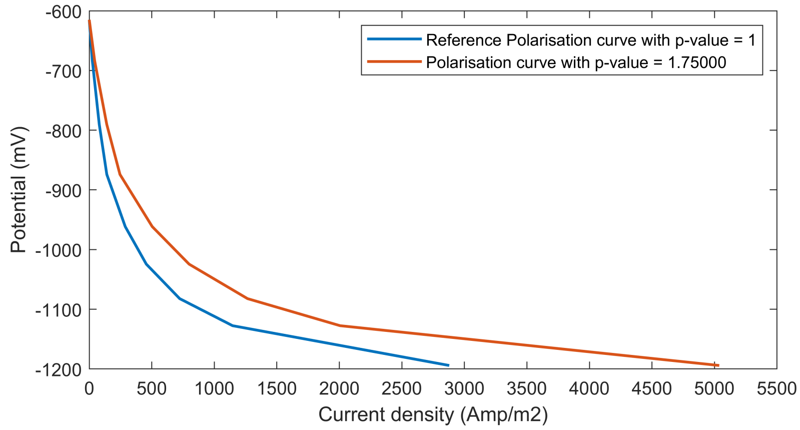
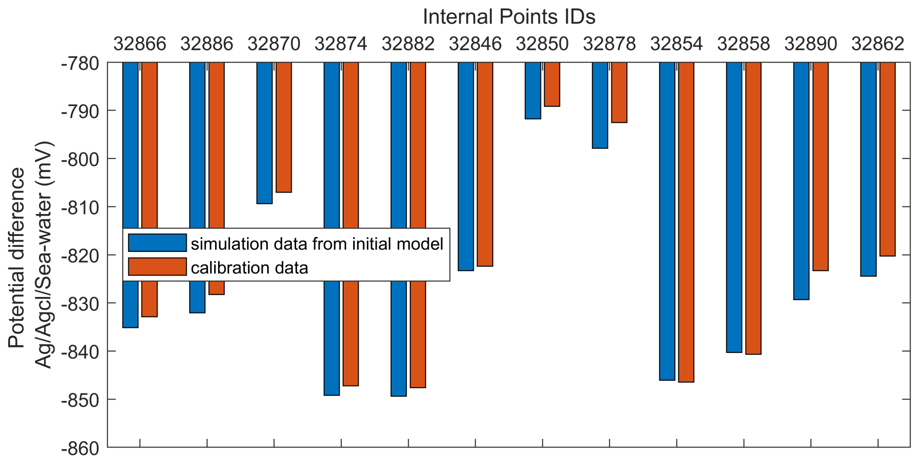


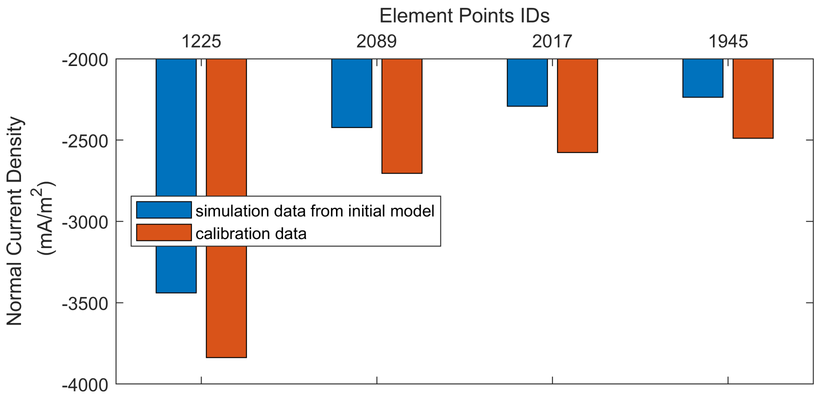
Figure 5.6: Polarisation curves with the associated *P-values*

With the initial guess of the parameter, the simulation data can be obtained by feeding the parameter for the simulation run. The polarisation curve represented by the *p-value* in Table 5.2 is shown in Figure 5.6. On obtaining the simulation data, comparative analysis (Objective calculation) can be done against the calibration/validation data to have insight into the performance of the model with the given parameters.

A graphical plot might offer some insights to the concerned engineer through the process of visualisation even though the automatic method simply takes the quantitative requirement (objective calculation) into account. The graphs represent the response data value in the y-axis against the IDs for the response data presented in Figure 5.7. The IDs are the denotation of the meshes and elements created during the discretisation stage of the CP modelling. For this case, the bar graph (Figure 5.7a) shows that there is a less significant discrepancy between the calibration *surface potential* reference data and the data from the model with the initial parameters' values, while the other type of data i.e., the *normal current density* (Figure 5.7b) has got significant discrepancy between the calibration and the data from the simulation. This demands an update of the parameter



(a)



(b)

Figure 5.7: The comparison plot between validating/calibrating response surface data against the simulation data from the initial primary model with parameters value given in Table 5.2, a) surface potential and b) Normal current density.

to match the simulation output with the reference data.

The parametric CP model with the initial parameter (Table 5.2) and with the corresponding output response data as displayed in Figure 5.7 is now considered and selected for optimisation-based best parameter search.

5.3.6 Optimisation with Minimisation for Best Parameter Finding

In the situation, where the objective function is dependent on the two independent parameters/variables "*p-value of Material 1 related polarisation curve* (p_value_1)" and "*Sea-water conductivity* (σ)", the minimisation problem (Equation 5.2) will be assessed as Equation 5.3.

Calibrating data, objective function and optimisation tool(s) discussed in Sections 5.3.3, 5.3.5 and 5.3.4 respectively are selected and applied for the requirement of minimum searching.

$$\min_{p\_value_1,\sigma} \quad f(p\_value_1, \sigma) \tag{5.3}$$

Utilising the given data resources, tools and algorithms, continuous optimisation is undertaken. The iterative state reached during the optimisation process is presented in Table 5.3. The graphical discrepancy between the state's model's output with the calibration data can also be visualised within the GUI. The optimisation process for parameter search ended as the algorithm cannot further decrease the objective function, in the search direction.

Table 5.3: Iteration stages and corresponding parameters' values during the optimisation problem before reaching a solution.

Iteration	F-count	Material 1 related p-value	Sea-water's Conductivity (Siemens/m)	Objective values (F(x))
0	3	1.7500	3.0000	5.5168e-03
1	9	1.8145	3.0443	3.5445e-03
2	12	2.0516	3.2340	3.5730e-05
3	15	2.0515	3.2422	2.8556e-05
4	18	2.0425	3.2655	1.5382e-05
5	21	2.0229	3.2959	4.5574e-06
6	24	2.0027	3.3186	1.5651e-06

Result and Analysis

In this case, it took a total of seven iterations to converge to the best set of parameter values. The total count for objective function calculations, represented as F-count in Table 5.3, was 24, which corresponds to the total number of data exchanges between the server and the simulation. This also means that the equivalent number of simulation runs took place during the entire process. The comparison between the model's output obtained with the best fit input parameters against the calibration data is presented in Figure 5.8.

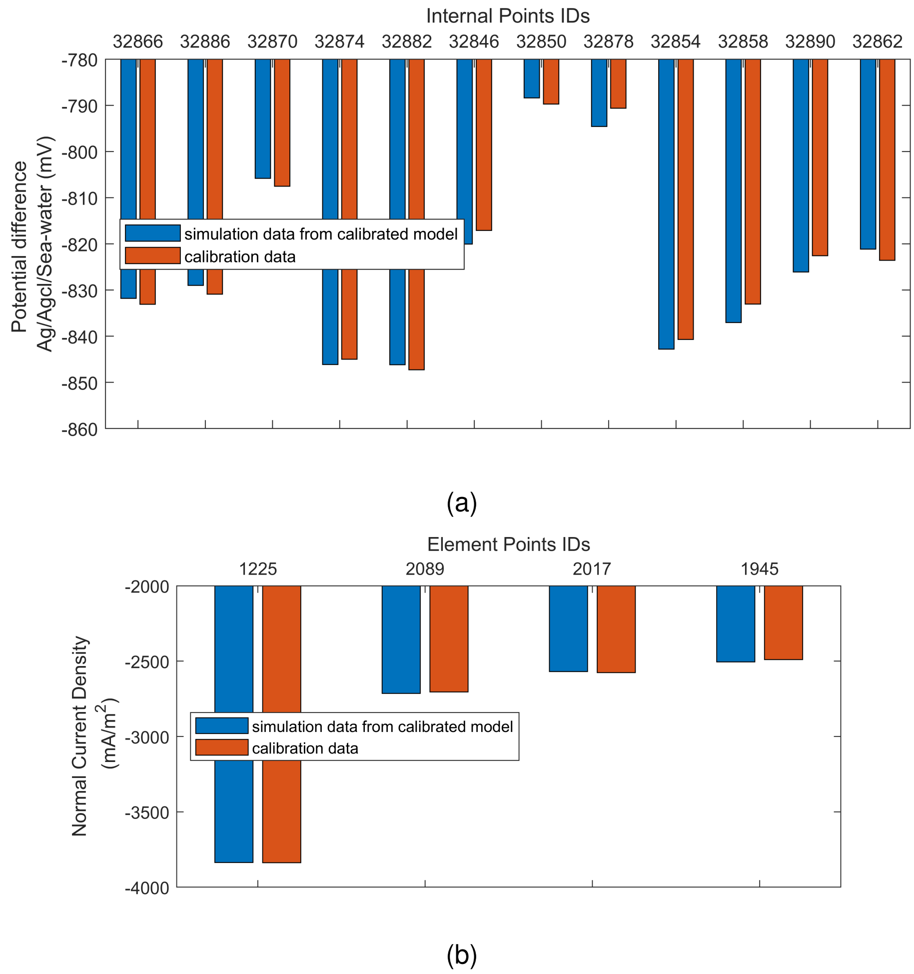


Figure 5.8: The comparison plot between validating/calibrating response surface data and simulation data from the solution model with solution parameters reached (Table 5.3): a) surface potential data and b) current density data.

Table 5.4: Calibration time analysis for the case study

Total count for simulation runs	24
Time for each Simulation Run	Approx. 8 min.
Time for data transfer and analysis at each iteration	Approx. 3 sec.
Total time	**Approx. 3 hrs. and 37 min**

Utilising the platform and the approach, this parameter updating task which could take many hours or days when performed manually is reduced to less than a few hours including all the simulation running time (Table 5.4). All the experimentation (including the simulation runs) were performed on a Lenovo-ThinkPad with a 2.30GHz CPU and 32 GB RAM. The importance of the platform is highlighted by the significant reduction of model calibrating time using the automated experimental platform compared to the manual approach.

5.4 Conclusion

In summary, the proposed approach aims to reduce the need for human involvement in model calibration and adaptation by introducing a Design of Experiments (DOE) platform that integrates a process simulator and scientific software. This platform enables the handling of data, experimentation, and implementation of adaptive algorithms for updating the model, while also providing a way to integrate multiple process models and simulators. By using this platform, a digital twin of a physical asset can be realised using a process simulator that is already in use, making it a practical approach for Structural Health Monitoring (SHM) that relies on physics-based prognosis.

The case study demonstrates the usability of the platform towards achieving a Cathodic-Protection DT. The continuous optimisation procedure was undertaken using the platform for the parameter estimation task essential during DT enabling and maintaining. Overall, the proposed software integration approach combines advantages offered by scientific and commercial software(s) to have a comprehensive Digital Twin in reduced time capable to predict the present and future health of a structure.

An issue felt during the above experimentation is the lack of benchmark of qualitative and quantitative validating/calibration data besides selecting the optimisation tool (algorithm). This type of benchmark is necessary to ensure the optimisation process is successful, assuming that all other aspects have been appropriately selected. A similar issue was felt in determining the weightage constant of the objective function during the CP model-related minimisation-based calibration. While the major of the further work will be focused on the utilisation of the integrated platform on addressing DT adaptation requirements, the issue of model validation/calibration resources benchmarking also needs to be assessed. This necessity of the resource benchmarking approach is highly felt when the data from the real structure is obtained from the costly survey and inspection as in the CP system's case for offshore structure. The next chapter will discuss the issue and propose an approach for resource benchmarking for optimal usage of the experimentation.

6 Benchmarking the Validation Resource Requirements for Adaptive Modelling Within Digital Twinning

This chapter presents the findings from the work under the third research area i.e., *Data Benchmarking for Model Calibration and Adaptation*.

The benefits of Digital Twins have been already demonstrated in Chapter 4 and 5 in obtaining the data from the real world and in automated model calibration respectively. Under situations where the model operational validation formed a basis for the iterative calibration, deterministic methods were utilised for both calibration and validation. Every model update utilises the latest available observation data to reduce epistemic modelling and/or parametric uncertainties. As SHM subsequently should also have the goal of reducing resource requirements for the prognosis as discussed in Chapter 5, the resources required for the performance validation of the model demand a proper trade-off between the accuracy of validation and the cost involved (Alaswad and Xiang 2017).

The followings are the objective and milestones set associated with resources (particularly data) benchmarking during model calibration and/or adaptation:

Objective

1. To investigate the DT concept and its associated recent development (such as integrated platform) on standardisation of the data resources.

Milestone

1. Approach(es) for benchmarking the required quantity, quality, and variability of validation data and performance metric(s) for model online calibration/adaptation within DT.

6.1 Quantitative and Qualitative Real-world Data Requirement for Model Operational Validation – Motivation

Tailoring any parametric simulation model (pre-accepted for behavioural simulation) to represent the physical system requires a suitable model adaptation route (Park and Schneeberger 2003, Tahmasebi et al. 2012). Data is one of the crucial factors for validating model performance, upon which the route is dependent. Advanced sensor and data measurement technologies have enabled online access to operational data from physical assets, which is required for operational validation. However, the information requirements for operational validation may vary with the behavioural complexity of the system (Oberkampf and Trucano 2008). Therefore, there is an anticipated trade-off between the validating data and the cost involved in data collection, particularly when data are obtained from inspections and surveys. For example, in the case of cathodic-protection model validation, the data dependency ranges from surface-potential, normal current density, and potential gradient-related data (Adey et al. 2020). These data in the case of the offshore structure are often obtained from the survey/inspection which gets costly if too many positional inspections of the structure are required.

Studies on standardising model verification and validation methods (Oberkampf and Trucano 2008, Sargent 2010, ASME 2009) lack guidance on selecting or defining resources (data) standards for validation (Pace 2004, Fabrizio and Monetti 2015). Operational validation becomes more challenging for complex models and/or immature domains, such as offshore systems and wind turbines. This is because in such a situation, with a shortage of quality-assured, and high-quality measurement data, the lack of the standard escalates the challenges (Doubrawa et al. 2019, Adey et al. 2020).

A similar situation of data standardisation was felt during the case study in Chapter 5 while undertaking the minima-based optimisation problem for model calibration. The problem was regarding the fixation of the positional data counts required to validate/calibrate the model ensuring avoidance of the wrong assumption about the model's performance. Though data for all small mesh points corresponding to the offshore structure can be obtained from CP simulation, the issue is with obtaining the data from the physical system and the cost involved in obtaining the data. Fixing the weightage constant of the objective function for two types of validation data was another similar issue but related to the formulation of the validating criterion encountered during the CP model-related minimisation-based calibration.

Furthermore, with the growth of the degree of freedom of parameters in any simulation model also puts a strain on the validation and calibration of the model due to increased complexity. Thus, establishing a proper benchmark of the data to their quantitative, qualitative, and diversity requirements to track the model's performance and provide an efficient adaptation route is anticipated.

While the behaviour of the structure is changing, so should be reflected in the model thus leading to the resource's requirement might also need updating. Thus, rather than fixing to benchmarked resources, an approach to benchmark the resources would be better preferred within the DT. The research, therefore, instead of finding a benchmark for each case proposes an approach to benchmark the resources. This chapter includes the details of this approach and its demonstrative analysis. The outcome of the research activities including a case study for the proposed approach demonstration has been already published in Sapkota et al. (2021b).

6.2 Approach for Benchmarking Online Operational Model Validation Requirements

A virtual and repetitive experimental-based approach is proposed to address the issue of resource standardisation during model calibration or adaptation. The approach in other words does resources benchmarking required to detect the parameter(s) uncertainty of a predictive model.

The approach primarily involves multiple reference virtual experiments (Figure 6.1) and can be undertaken with automation utilising the integrated platform presented in Chapter 5. The benefit offered by this experimentation-based approach is that it can be undertaken in the offline phase, even though it requires plenty of simulation-based experiments. It involves generating virtual models within the DT platform and continues with the performance evaluation and updating the primary model relying upon the selected validating data and/or the formulated metrics. The successful repetitive experiments ultimately form the basis of increasing trust to rely upon the undertaken routes i.e., the resources dependency.

The accepted model validating data can be used as the benchmark of the minimum data required from the real asset. This means procedures and resources utilised in reference experiments can be adopted in parameter optimisation of the virtual model when

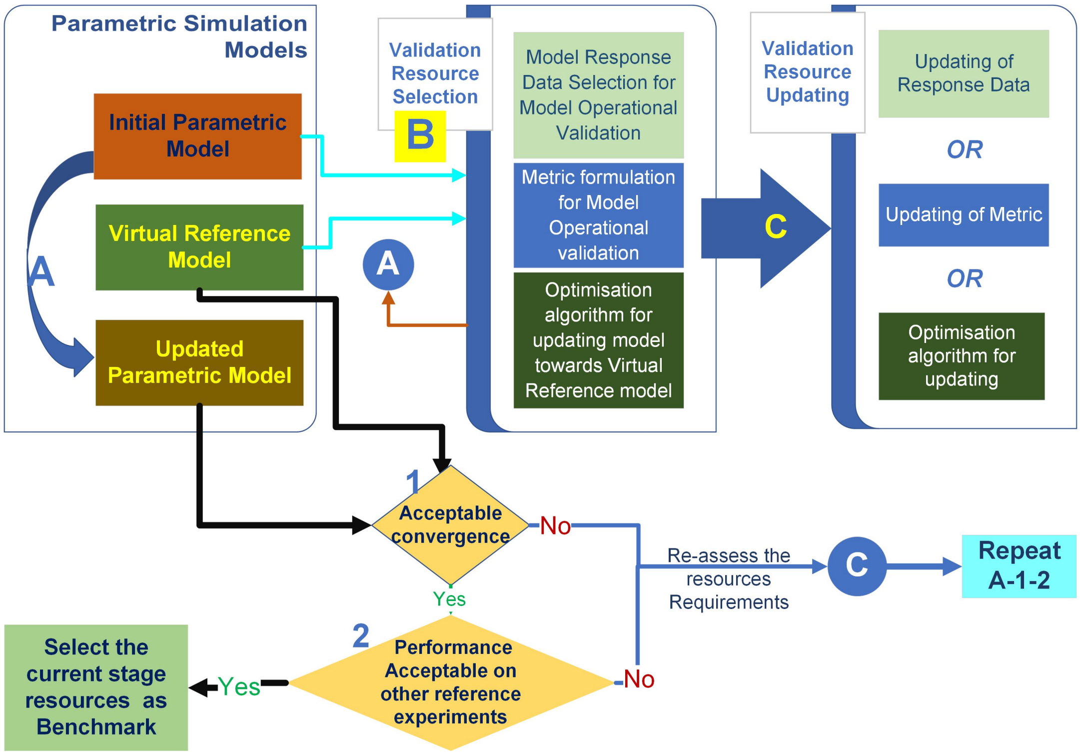


Figure 6.1: Illustration of reference experiment for Benchmarking Operational Validation Requirements.

data is available from the real physical system. A similar approach can be employed to formulate, the performance-validating metric.

The major steps performed during reference experimentation for benchmarking the model operational validation requirements are:

- Virtual Reference Model Generation
- Validation Data Selection to Standardisation
- Validation Metric Assessment
- Performance of Selected attributes on Model Convergence to the Reference Model.

6.2.1 Virtual Reference Model Generation

The effectiveness of the calibration route (resources and algorithms) can be analysed by comparing the adapted/calibrated model either to the higher precision model or the physical prototype (or real system) but with all details known including the parameter-related information. When conducting simulation experiments using a pre-accepted simulator, a reference model with established behavioural simulation performance and known parameter values will serve as a virtual prototype. This research also endorses utilising the virtual environment as virtual testing is the key to reduction of the cost and effort but also avoiding inconvenience for hardware tests. While at this stage, the research is focused on assessing the parametric uncertainties of the model, the higher precision model is generated with some selected parameter(s) values but within a reasonable range(s). As error (noise) in the data from the physical system cannot be neglected in the real-world scenario, some error (noise) can be introduced to the input or output data for the reference model's simulation to make the data more realistic to the real world.

On the other side, there exists a similar virtual model but with different parameter values than the reference model and termed as Primary Model. The Primary model is the one which will be updated (or calibrated), while the reference model is the one used as the representative of the real structure. Analytical support within DT discussed in Chapter 5 can generate such virtual models, thus such virtual model-generating and virtual experimentation features are proposed within the Digital Twin. However, limiting the parameters within reasonable ranges might require external support/expertise.

6.2.2 Validation Data Initial Selection to Final Benchmarking

Data sensitivity tracking for corresponding parameters i.e., response data mapping to the parameters is usually achieved by sensitivity analysis and usually requires no reference model. After fixation of other involved models validating and updating attributes (e.g.: metric, optimisation algorithm), the initial data-set required for the model validation/calibration is selected. The experience with the domain and data could guide tracking the ability of data on validating and adapting performance, but also the cost and time for data collection should be considered when applicable. Likewise, the response data from the reference model needs to be provided with the noise(error), to have the data as realistic (with some technical data measurement error) from the real asset.

If the convergence of the initial primary model to the reference model within the acceptance range is reached in the reference experiment i.e., passing *Test 1* (Figure 6.1) , also other similar convergence tests are required (Section 6.2.4). If fails either of the tests, the resources (for example data) are updated to ensure the data capability on converging the primary model to the reference one. The non-risk involved balance between the cost involved in data collection and model optimisation accuracy using the data is identified by the series of virtual experiments. Not in every case, the data counts are increased, also should be reduced to find the trade-off between the data for validation and the cost involved in data collection.

6.2.3 Validation Metric Selection/Formulation

The performance of the quantitative metric to assess the goodness of fit between the simulation response output and the real-world data is another important aspects as like the data in the process. While the nature of the simulation is known and the uncertainty types to assess are known, the metric selection/formulation will be easier. The nature of the model in most of the application domains for their designated purposes is pre-known.

When performing model updating of complex structures deterministic approaches are preferred (Goller et al. 2011) due to the associated computational benefits. Therefore, metrics performance evaluation and metric formulation are confined to the deterministic metrics at this phase of the research keeping the probabilistic assessment of the simulation model outside the scope.

Multiple types of deterministic criteria (metrics) are available in the literature for performance validation of a deterministic model. Among them, the magnitude-discrepancy based, and the correlation-based are two wide categories of metrics for deterministic validation. It is not possible to conclude that one criterion is better than another, as they assess different aspects of the dataset. The choice of criteria to employ in validation should start with the features that are intended to be assessed (Ni et al. 2004).

The role of validation metric formulation persists even after a metric selection when multi-response data types available from simulation can be used for the model's performance validation. For example, a situation encountered in the case study of minima-based optimisation in Chapter 5 can be considered. As multiple data types were involved, the weightage constant was an important aspect, that will decide the role of specific data types during the performance validation. Determining the dependency of multiple data-types during model performance validation is the further step of metric formulation in such cases. The selection of the weightage constant is not straightforward and might require some experimental analysis before reaching the decision.

The best model's performance measuring metric required during optimisation-based parameter updating can be formulated using similar pre-mentioned repetitive virtual experiments by fixing the other influencing attributes. Once the metric is formulated, it is further adopted as an objective function during optimisation-based model calibration when real-world data are available.

6.2.4 Models' Performance Convergence Analysis

For the performance validation of the selected resources supported with the reference model updating experiments (*Step A*), the convergence *Tests (1 and 2)* (Figure 6.1) are performed. The convergence of the response data as well as the value of the parameter between the solution model and the reference model should be confirmed.

If the initially selected validating data and/or metrics fail in *Test 1*, the steps are repeated after updating the data-set and/or metrics (*Step C*) until the data-set and/or metric(s) are promising to converge the Primary model to the reference one. Passing *Test 1* is not enough as the performance need to be validated for other parametric situations as well. *Test 2* is about repetitive testing with the different cases of the virtual models (Initial and/or reference) by varying the parameters of the virtual models.

6.3 Case-Study: Benchmarking CP model Validation and Calibration Data Requirement

6.3.1 Experimental setup

The experimental CP model (Section 5.3.1) built for offshore structure (Figure 5.2) is adopted for the approach demonstration. Similarly, the Software-Simulator platform (previously discussed) is utilised for the implementation of the proposed virtual and offline experimentation-based approach. Most of the resources except the response data count will be fixed to the experiment discussed in Chapter 5.

The considered parameters required to run the simulation are:

1. *P-value* (Section 4.1.3) for Material-1 (Figure 5.2) related Polarisation curve of the CP system.
2. Sea water-related *conductivity* (*Siemens/m*).

In this case study, the approach application for benchmarking response data requirements will be demonstrated. The benchmarked data should be sufficiently enough to validate the model's performance and calibrate the above-mentioned parameters.

6.3.2 Data types and other resources Selection

Considering the practical feasibility of the data, response data that would be selected as validation/calibration data are: a) *Surface Potential* (mV) and b) *Normal Current density* (mA/m^2) and c) *electric field* (mV/m) with the first type being more preferred and so on. While at this stage the approach is being demonstrated to find the benchmark of data requirements only, other attributes involved such as the metric and optimisation tool are fixed beforehand. Normalised Mean Square difference is fixed as the validating metric and as an objective function for optimisation with a fixed weightage constant (2:1) when two different data types are considered. Similarly, for optimisation-based parameter updating, the gradient-based "*Newton-quasi*" algorithm and MATLAB-provided tool "*fminunc*" are selected.

6.3.3 Virtual Models in the reference Experiments

Primary and *Reference* CP models are generated using the adopted structure related geometrical data and meshing. The parameter value(s) of the "*Reference model*" provided differ from those of the "*primary initial model*" in each reference experiment. To make the data more realistic to the real-world data $\pm 2\%$ noise is added to the data of the reference model's after simulation.

Table 6.1: A representative example of the parameters' values provided for the models in the reference experiment

	'Material 1' related p-value	Sea-water's Conductivity $(Siemens/m)$
Primary Initial Model	1.5000	2.7500
Reference Model	2.0000	3.3333

A representative example of the *Primary initial* and *Reference model* is presented in Table 6.1 together with the corresponding parameters' values. During the process of continuous optimisation utilising the selected resources, the virtual experiment will be aiming to converge the *Initial model* to the *Reference Model*.

6.3.4 Approach Implementation

An experiment with initial data dependency as Case-I (Table 6.2) is initiated. The lesser data variety would be always preferred in practical phase as more data varieties being obtained from the inspection requires more survey equipment, expertise and time. The Surface Potential data positions (Case I) for both the primary and reference model is presented in Figure 6.2b.

Utilising the validating data suggested by Case-I (Table 6.2) i.e., only one data type,

Table 6.2: Response data counts considered for model validation/calibration for 2 different cases

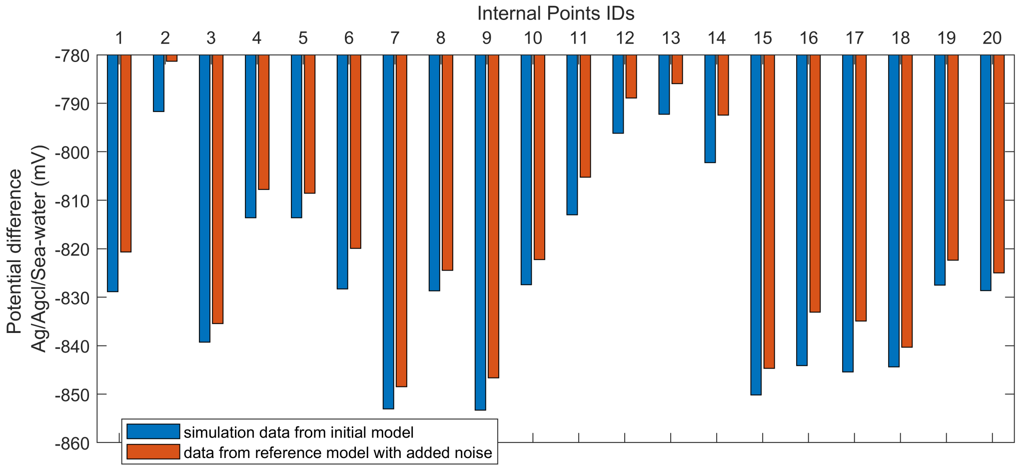
Data Counts		
	Surface Potential (mV)	*Normal Current density* (mA/m^2)
Case I	20	0
Case II	15	5

together with the other optimisation resources, the continuous optimisation-based parameter search similar to one in Chapter 5 is undertaken. The process is towards converging to the reference model (parameters) starting from the initial primary model. After the parameter update (Step A) and analysis 1-2 (Figure 6.1), the likelihood to reach the wrong solution i.e., local minima during optimisation-based parameter updating is found in this case. It is detected after making the parameter comparison of the solution model reached with resources under Case-I (Table 6.3) against the Reference model (Table 6.1).

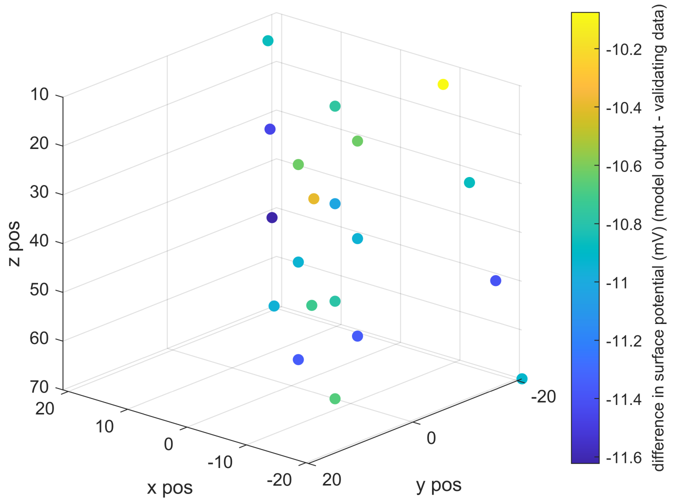
Table 6.3: Parameter's value in solution model reached for both cases.

	'Material 1' related p-value	Sea-water's Conductivity ($Siemens/m$)
Case-I	1.6055	2.6525
Case-II	2.0099	3.3313

Then, the next set of response data is selected by giving diversity in the data type i.e., normal current density data is further added keeping the initial quantity the same as Case-I. After the analysis like before, the result shows Case-II favours Case-I regarding data dependency. These two representative cases form the basis of the virtual-experimentation-based standardisation of model validation data requirements. If passed performance evaluation with re-testing (Section 6.2.4), the Case-II suggested response data will be set as a standard resource for validation and adaptation of the CP model when data are available in the online phase. The experimentation and the approach can be further exploited to reach the lowest possible response-data count ensuring the diversity and distribution over the structure. However, the cost of the experimentation though mostly undertaken in the offline phase and the risk of the wrong assumption needs consideration.



(a)

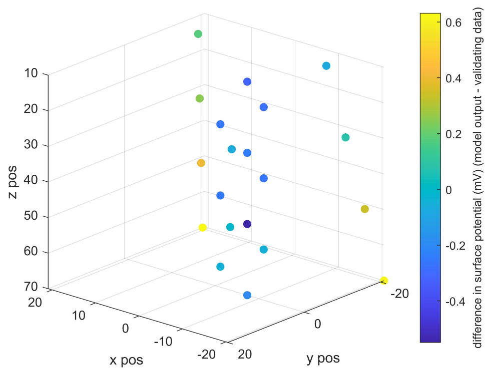


(b)

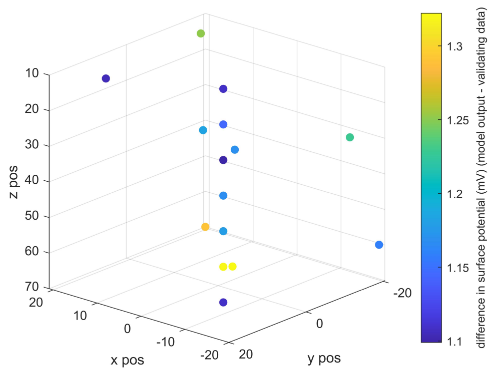
Figure 6.2: a) The graphical comparison, b) Difference shown at real 3D positions of the data, between surface potential data from the simulation run of the Initial primary model and the data for reference model.

Result Analysis

The result shows diversification in data selection could minimise the risk converging to local minima during the given CP model's validation and adaptation route from any primary state. When relying upon data from Case-I for model operational validation one can reach the wrong assumption about the performance of the model. This situation can be more understood by analysing the solution models' surface potential (i.e. first data type) discrepancies with the reference data for both cases (Figure 6.3). The response data (only one type) suggests the solution from Case-I is better than Case II, which is a wrong assumption.



(a)



(b)

Figure 6.3: The discrepancies for Surface Potential data, between solution model output and Reference model output (validating data) a) Case I solution model, b) Case II solution model.

A similar approach can be implemented in metric formulation the performance of the provided attributes differs.

The benchmarks provided might not only vary between models but also for the same model at different stages of its operational lifespan due to changes in the integrity of the system. This underlines the integration of such benchmarking approach within the DT artefact, to support adaptation required at any instance of its operational life span.

6.4 Conclusion

This chapter presented the solution for the issue of model validation resource standardisation being an important aspect which cannot be discarded during model calibration and/or adaptation. An approach is proposed to benchmark the data and the metric requirements before going to the online phase of tailoring the parametric model as the representation of the real physical asset. The approach utilises the virtual reference model and multiple virtual experiments in the offline phase to set the standard of the validating resources.

Moreover, the proposed approach can be achieved within the integrated platform proposed in Chapter 5 for the automation of model validation, calibration, and adaptation under the DT concept. This highlights that the experimentation-based approach is also essential to the DT architecture (Figure 2.5) and can be accomplished within the validation artefact. Additionally, the applicability of the approach is demonstrated to standardise the quantity, quality, and diversity of data requirements for validating the adopted CP model during calibration. However, a limitation of the approach is that proper justification for requiring the benchmarked resources may require high expertise in the domain and related data, as it is experimentation-based.

The above experiments also encountered one of the issues discussed in Chapter 2 i.e., time-consuming simulation runs highly affect this experimentation-based resource benchmarking. Each update of the data or metric repeats the entire process of continuous optimisation, therefore, increase in time for the approach implementation, though won't be a serious issue in the offline phase. However, with the complexity arising in the system's behaviour during operational time, the requirements of the approach are also in the online phase. The re-assessment of the experimentation-based approach thus requires the solver for addressing the time-related issue.

As, a future work within the context, the approach needs to be implemented in a more complex model. Also, the benchmark reached needs to be applied to the real-world problem i.e., to calibrate the model based on the response data from the physical CP system.

7 Surrogate-Assisted parametric calibration within Digital Twinning

This chapter presents the findings from the work under the fourth research area i.e., *Online Model Calibration/adaptation: Addressing Input Parameters Uncertainties*.

Chapter 5 and Chapter 6 have provided the basis for the automation of the parameter updating task and an approach to benchmark the resources required to calibrate a model before applying the real-world data from the real system. From here, the research activities are focused on the core research area and are guided by the following Objectives and milestones.

Objectives

1. To investigate the significant methods/procedures in dealing with parameter uncertainties during adaptive simulation.
2. Providing a standard framework for efficient and reliable parameter updating during model calibration/adaptation.

Milestone

1. The online model updating/adapting artefact (a mechanism) aligned to the DT concept, that uses the best suitable optimisation algorithms/methods..

The state-of-art discussed in Chapter 2 about modelling in SHM has already established the role of calibration to prepare the model for its predictive role. The optimisation-based calibration (parameter estimation) is also discussed to some extent in the previous chapters including the application of a type of them.

7.1 Issues with Simulation-Based Optimisation and Role of Surrogate Modelling: Motivation

Chapter 2 concluded with the anticipation of an enhanced and online tuning mechanism with more efficient algorithms to tune the modelling parameters during the establishment of the predictive DT. Also, it is acknowledged that the deterministic model calibration at an advanced level is achieved with gradient-based or gradient-free optimisation methods, requiring output from the simulation runs for multiple situations.

7.1.1 Limitation of Continuous Optimisation Method for Parametric Calibration of Model

The continuous optimisation methods requiring the objective calculation (with simulation run) for each iteration can be divided into 2 categories: 1) the gradient-based method and 2) the gradient-free method. The gradient-based methods such as Quasi-Newton Method (Gill and Murray 1972), Conjugate Gradient method (Hager and Zhang 2006), Interior Point Method (Nocedal and Wright 1999), etc. though efficient than trail-and-error based approach, has got the demerits like the results being too sensitive to the initial values and tending to get trapped in the local optimum. Such issues were also noticed in the CP model-related experimentation during the case study of Chapters 5 and 6.

On the other side, the gradient-free optimisation method such as Genetic Algorithm (Grefenstette 1993), Simulated Annealing (Kirkpatrick et al. 1983) and Particle Swarm Optimisation (Poli et al. 2007) can find the global optimal solution but require hundreds or even thousands to millions of objective evaluations in the whole design space. While relying upon such algorithms, real-time parameter updating becomes almost unachievable due to the computational expense associated with higher-order physics-based models.

7.1.2 Surrogate Concept in Modelling

Surrogate models (also known as surrogates) are the approximation models to the full-order simulation model with significantly less computation time than the full-order physics-based one (Barton 1992).

Surrogate models are broadly categorised into three different classes: hierarchical models (simplified models), projection-based reduced models, and data-fit models. A brief insight into the method of obtaining the approximation model is provided below (Eldred and Dunlavy 2006, Benner et al. 2015):

1. **Model Simplification method:**. It is built by simplifying the underlying physics and required model building and domain expertise. This includes, mesh simplifying, reducing the number of parameters and/or ignoring nonlinear terms in the solver. This approximation modelling is often referred to as multi-fidelity, hierarchical, component-based, sub-structuring, etc.

2. **Projection-based Methods:** These approximations are based on mathematical derivation, rather than professional knowledge. Generally, a subspace is created to project the governing equations into the subspace to achieve dimensionality reduction of the model space. Some of the methods that follow this approximation technique on having ROM are:
    - Proper Orthogonal Decomposition (POD) (Chatterjee 2000)
    - Balance Truncation method (Gugercin and Antoulas 2004)
    - Reduced Basis Method (Haasdonk and Ohlberger 2008)
3. **Data fitting method:** This approximation is known as the black-box modelling and is also termed response data or scalar surrogate.

This research from this stage will assess the role of surrogates in the calibration and/or adaptation of the model within the DT concept. This will be expanded to the simulator-based DT enabling and maintaining concept. Once the surrogate-assisted approach is established in the realisation of the DT for a physical system, it will be an attempt to standardise the roles of surrogates within the DT concept.

This surrogate-assisted calibration method using the DT concept including a case study for the approach demonstration has been already published in Sapkota et al. (2022a).

7.2 Surrogate Models in Digital Twin's Realisation

In overcoming the issue of time-consuming calibration tasks, replacing the computationally expensive model with an approximation model that takes less running time offers a solution. The applicability of such an approximation model or surrogate to circumvent intractability caused due to high computational costs in many-query applications has been discussed for years (Barton 1992, Robinson et al. 2008, Yin et al. 2019). The surrogate, as an alternative to a time-consuming numerical simulation model, could assist in the estimation/update of design parameters by obtaining the global optimal solution within a reasonable time.

As, the motive is utilising the surrogate on calibration rather than using it for understanding the dynamics of the system, the data-driven black-box models that are typically trained by using datasets obtained from a physics-based (white-box) model's simulation run will be considered at this phase. *Black-box* models are models which provide no information about the behaviour (physical process) of the system, unlike the *grey* (or *white*) box models with some (or complete) information available about the working principle of the process behaviour for the system. The response surface method (RSM) also known as polynomials regression fit (Myers et al. 2016), artificial neural network (ANN) (Shalev-Shwartz and Ben-David 2014, Zhang, Xie, Ji, Zhu and Zheng 2021) and Kriging (Van Beers 2005) are the commonly used methods in constructing a surrogate in engineering fields. Any of the methods require generating simulation datasets using the appropriate Design of Experiments (DOEs) (Giunta et al. 2003) method that decides the combination of input variables for the simulation run to generate data snapshots. Then the simulation snapshots are used to train the surrogates (data-driven) with the simulation input parameters as the input-variable and the simulation response data as the output. The surrogate is accepted as an alternative model once the performance of the model is evaluated and is within the required threshold

Although the cost of running a surrogate is insignificant compared to the full-order physics-based model, the building of a surrogate may get relatively expensive due to the multiple simulations requirement for data generation. Two terms *”offline”* and *”online”* are often used by the literature to highlight the benefits of surrogates. *Offline* cost refers to the up-front cost to create the surrogate which could be large enough, while *online* cost refers to the cost of running the surrogate during the parameter exploration with real-time calibration data. The inexpensive computational cost offered by the surrogates illustrates the benefits of surrogates in the online phase when real-time evaluation is essential.

Towards, the real-world realisation of the DT concept, experimentation-based adaptation is essential to tailor the physics-based simulator to represent the physical twin. The necessity of surrogates in structural DT realisation has been discussed since the same time when the concept of DT was introduced to the field of simulation and modelling (Tuegel et al. 2011). However, in the lack of standardisation of the DT concept, the discussion is mostly limited to the conceptual level.

Leveraging the concept of hybrid DT discussed in Chapter 2, the surrogates here are considered as the data-driven models within the DT but serve as a substitute for the physics-based model. For the practical DT endorsing surrogate, the integrated platform discussed in previous chapters with automated data and control signal flow between them also offers the feasibility. For example the MATLAB-based platform proposed in Chapter 5 is capable of performing the tasks essential for surrogate building such as DOE, model training, performance analysis, etc.

Though surrogates are suggested as an alternative when faster prediction is required

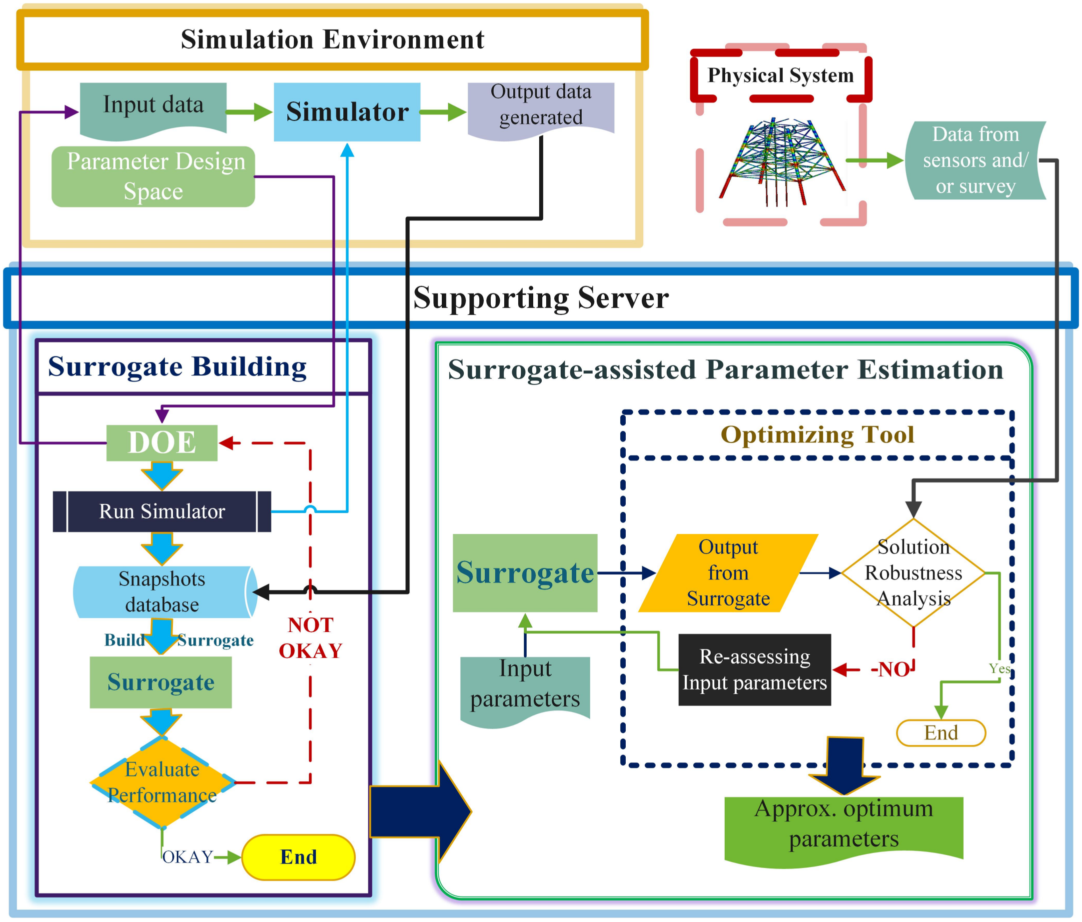


Figure 7.1: Illustration of role of surrogates in parameters approximation with the utilisation of the sever-simulator integrated platform.

or to ease the process of parameter estimation, physics-based full-order model(s) will remain the core and robust predictive tool of the DT.

7.2.1 Design of Experiments for Surrogate Model building

DOE is a procedure of choosing a set of samples in the design space, intending to gather the maximum amount of information from a limited number of samples (Giunta et al. 2003). The quality of the samples (training data) significantly impacts the accuracy of the data-driven model. Hence, an appropriate DOE method to select the optimum number and region of the sample training data points can enhance the surrogate modelling process (Alam et al. 2004, Davis et al. 2018). Latin Hypercube, Box–Behnken, and Central Composite Design (CCD) (Giunta et al. 2003) are a few of the common DOE methods implemented during the space-filling design to ensure adequate coverage of the space of the variables within the given range.

Based on the data generated from the initial experimental design, a surrogate model is trained and the accuracy of the model is evaluated using the error testing metrics accompanied by graphical analysis (if required). Then based on the performance accuracy of the surrogate, the requirement of additional sample points is decided (if the samples from the initial design were not enough) (Figure 7.1). The success of having the best surrogate depends upon proficiency in finding the trade-off between the computational cost associated with the data generation by simulation runs suggested by the DOE and the accuracy of the surrogate (Alizadeh et al. 2020). Furthermore, the ability to represent the complex behaviour of the simulation model is another considerable factor during sampling in the cases when higher accuracy of surrogate is desired.

7.2.2 Surrogate Model Building with Simulation Parameters as the Input Variables

Most of the surrogate modelling methods have a common procedure to follow, which includes DOE, model training, assessing the quality, and updating to enhance the quality (if required). Among the multiple types of surrogate modelling methods available, the appropriateness depends on the system's behaviour that is being modelled. A polynomial fit model also known as Response Surface (RS) methodology among one of the surrogate building methods is considered as an example and the data-fitting meaning for the method is explained.

Response surface (RS) methodology is a collection of mathematical and statistical techniques based on the fit of a polynomial equation to the input-output dataset. The regression coefficients (mostly second-order) are usually used to provide the relation between the output responses and the input variables. A representative second-order polynomial RS equation is presented as equation 7.1 (Bezerra et al. 2008, Myers et al. 2016).

$$y = \beta_0 + \sum_{i=1}^{k} \beta_i x_i + \sum_{i=1}^{k} \beta_{ii} x_i^2 + \sum_{i<j}^{k} \sum_{j=2}^{k} \beta_{ij} x_i x_j + \epsilon \tag{7.1}$$

where, y is the estimated output of the RS model, and x are the input parameters or variables with a total count equal to k. Then, β represents the regression coefficients also known as hyper-parameters of the data-driven model.

The training of the RS models in this case means calculating the regression coefficients, which are usually performed using the method of curve-fitting. Curve-fitting is the process of creating a curve or formulation of the mathematical function that best fits the provided group of the dataset (Arlinghaus 1994).

The server offering to enable and represent DT proposed in Chapter 5 can also facilitate utilising an "*offline*" phase to create the surrogate model of the system using the simulation input-output dataset from the physics-based model.

7.2.3 Surrogate Model's Performance Evaluation and Updating

The performance of the constructed surrogate model should always be assessed before using it as a representative of the full-order model. Usually, a set of testing points other than those used during the surrogate training are used for the performance validation of the surrogate. The qualitative assessment is made using criteria such as the Normalised Mean Square Error (NMSE). If the error is higher than the threshold, sample data points are added to the existing training data set and the surrogate is updated to enhance its prediction accuracy. Sequential adaptation (Golzari et al. 2015, Marque-Pucheu et al. 2019) is often preferred during surrogate updating in which new points are selected in the region where the prediction made by the surrogate model is not within an acceptable threshold.

7.2.4 Surrogate as an Alternative model in Calibration

The surrogate i.e., data-driven model on performance validation can now be used in the "*online*" phase. In the "*online*" phase, the adequately efficient surrogate(s) is used for the objective evaluation during the process of optimisation for optimum parameter approximation. When the computation time for each surrogate-based prediction can be within milliseconds to seconds, sufficient counts of searches to reach the global minima are attainable in significantly reduced time. This allows for an exploratory search in contrast to the exploitatory search limited by the gradient or non-gradient optimisation technique, thus avoiding the trap in local minima.

7.3 Case Study I- Surrogate Assisted Cathodic Protection (Sacrificial Anode) System's Digital Twin Realisation

7.3.1 Experimental Setup

The experimental CP model (Section 5.3.1) built for offshore structure (Figure 5.2) using the simulator BEASY is adopted. Similarly, the Software-Simulator integrated platform discussed in previous chapters is utilised for the expedition of experiments with automation. The analytical support required for surrogate building such as DOE, data generation, and performance analysis, are also provided by MATLAB's tools, within the integrated platform.

As the same CP model built for the structure is adopted, the core parameters considered to run the simulation as well as to build the surrogates are:

1. *P-value* (Section 4.1.3) for Material-1 related Polarisation curve of the CP system
2. Sea water-related conductivity (Siemens/m).

The goal of this case study is to demonstrate the role of the surrogate in the estimation of the above two mentioned parameters. The selection of the same parameters as of the previous experiment (Section 5.3.1) will allow for the comparative analysis of two different approaches implemented for parameter estimation.

7.3.2 Response data points for Surrogate building

The role of the surrogate in the calibration task is to provide an approximation of the simulation for the corresponding calibration/validation data points that could be available from the real system. Considering the feasibility of the data, two types of response data (as in Section 5.3.3) are taken for surrogate building: surface potential (mV) and normal current density (mA/m^2). The response data-position counts (IDs given as per count) are 12 and 4 respectively from the structure's surface (Figure 5.4). For these considered response data points (total n = 16), the data-fit models will be built.

7.3.3 Design of Experiments for Surrogate building

Regarding experimental design for Response-Surface-Modelling, the central composite design (CCD) has been frequently discussed (Montgomery 2017) and adopted in this case study. The reason CCD is the most implemented sampling method and prominent in DOE in most engineering problems, made it to be used as the first choice in the CP model. The inscribed *central composite design (CCD)* is selected which gives 8+1 sample points for 2 variables case. MATLAB based tool '$ccdesign$' is used to generate the sampling points for the two independent variables *"Material 1 related p-value"* and *"sea-water conductivity"*. The central point is used multiple times to provide high impact of the centre of the parameter space. However, deciding the range over which the samples will be generated requires some level of prior knowledge about the parameters. When the sample count is fixed, a narrower range will result in a higher accuracy of the model for predicting within the parameter range.

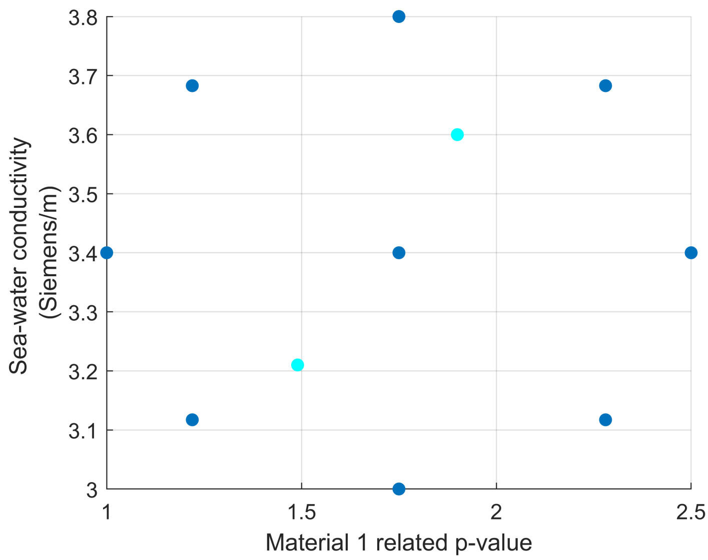


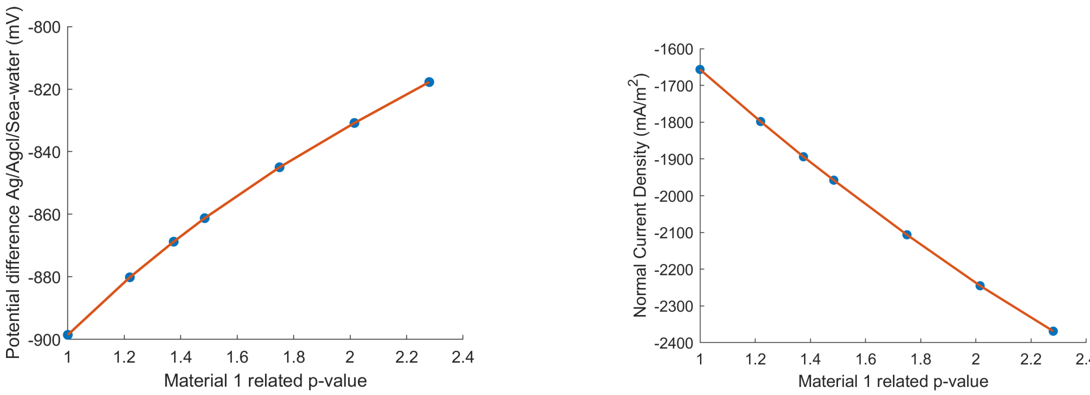
Figure 7.2: Plot showing the sample parameter points (in blue) for surrogate building from CCD with corresponding value for the two selected parameters and surrogate performance testing parameter points (in cyan)

The sample range is selected (Figure 7.2) and then samples are generated using the *CCD* method within the range. Assuming the interest is towards the centre i.e., the probability of the solution parameter falling towards the centre is maximum, the data-set corresponding to the central point is taken two times. Likewise, a few surrogate testing samples points are also generated (Figure 7.2).

7.3.4 Surrogate Model for CP system

The response surface method (RSM) with 2^{nd} order polynomials regression fit (equation 7.1) is implemented for surrogate building. When the effect of the independent variables and their interactions can be tracked, the polynomial response surface model (RSM) would be the first choice as a data-fitting model.

To investigate the applicability of the polynomial-fit model for the CP surrogate modelling, sensitivity analysis is performed. Few reference data positional IDs are considered for the output response data, and then response data at the positions obtained after simulation are taken into account by varying the target input parameter(s) while keeping the other parameters (if involved) constant. From the sensitivity analysis, the effect of the independent variables (*p-value* and *conductivity*) on the response data (*surface-potential* and *normal current density*) is obtained. The analysis (also represented by the Figure 7.3 for one of the related IDs in each analysis) shows the non-linear but uniform relation between *Material 1-related p-value* with both the *surface potential* and *normal current density*. The analysis (Figure 7.3) thus suggests that the 2^{nd}-order polynomial fit model could capture the relation between the input-output variables for such a CP modelling case.



(a)

(b)

Figure 7.3: Material 1-related p-value vs a) surface potential, b) normal current density, with other input variables fixed, for two different reference data positions for the two data-types.

Now the parameter (independent variables) involved for RSM in the case study are the polarisation behaviour represented by the *p-value* of the curve and the *sea-water's conductivity*. If *p_value* and *conductivity* ("σ") represents the two variables involved, each positional response data (y) for the corresponding ID (with index i) according to equation 7.1 can be represented as:

$$y_i = \beta_{0i} + \beta_{1i} * p\_value + \beta_{2i} * \sigma + \beta_{3i} * p\_value * \sigma + \beta_{4i} * p\_value^2 + \beta_{5i} * \sigma^2 \quad (7.2)$$

Using the simulation data snapshots generated from the aforementioned DOE, RS models are constructed for each node (i.e., i = 1, 2,... n). MATLAB based tool '*fit*' is used for the second-order polynomial fit, i.e., to determine the beta-coefficients for each equation for each response data position. In total, the '*n*' number of the second-order data-fit models are trained with the data corresponding to the data IDs. The trained surrogate can now provide a prediction of the response data (*Surface Potential* or *Normal current density*) with varying input control variables (*p_value* and σ).

For performance evaluation of the surrogate, the comparative analysis between surrogate output and the full order model simulation output for the testing parameters (one testing case Figure 7.4) is made. The developed surrogate model is capable of representing the CP model responses with acceptable accuracy ($NMSE < 0.002$). The acceptance threshold is taken after the performance analysis and considering the experts opinion from the related field. This often tends to be varying from the model and the type of data considered for the analysis.

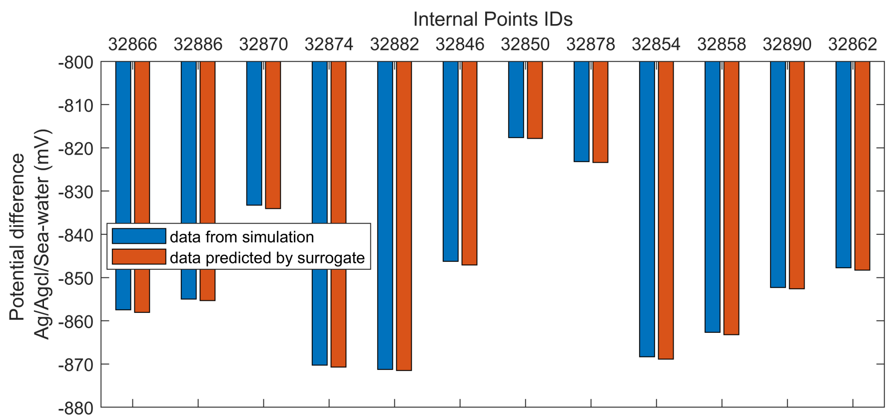


Figure 7.4: Surrogate output for one of the testing point compared with simulation output data (Surface Potential only)

The surrogate after performance evaluation and ensuring it's predictive efficiency within the margin of error provided can now be used as an alternative model for Optimisation based parameter estimation.

7.3.5 Parameter Estimation with surrogate

Having the polynomial-based surrogate now enables performing the exploratory search for the possible combinations of the parameters to obtain the objective function output with less computation time.

The calibration/validating data set (Section 5.3.3) is repeated from the case study in Chapter 5 where the data are used for continuous optimisation. With two different validating (calibrating) data types, the Normalised Mean Square difference (equation 5.1) between calibrating and model output data with weightage constant (2:1) is taken as validating criteria (optimisation objective).

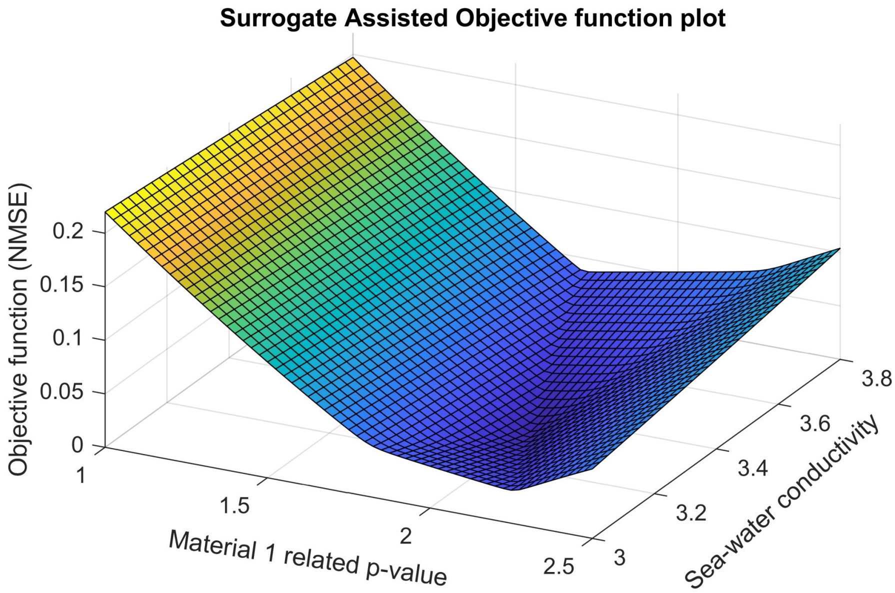


Figure 7.5: Objective plot over the parametric space with response data obtained from the surrogate for different parameter inputs, while validating response data remains fixed.

The plot in Figure 7.5 represents the objective function's output over a range of parameter combinations. The surrogate is used to assist in the generation of the plot by predicting the response data for the combination of input parameter values, while calibration (validation) data remains fixed. During the objective calculation, the minimum value is stored, and an exploratory search method is used to find the global minimum. The plot also demonstrates the effectiveness of the surrogate in avoiding local minima with surface smoothening.

However, in other situations when the exploratory search tends to induce a delay even while assisted with surrogate, gradient or non-gradient based optimisation algorithms/tools are suggested.

Result Analysis

The performance of the full-order simulation model with the parameters "*polarisation curve for Material 1*" and "*sea-water conductivity*" approximated using the above-mentioned calibrating resources and the surrogate is demonstrated in Figure 7.6.

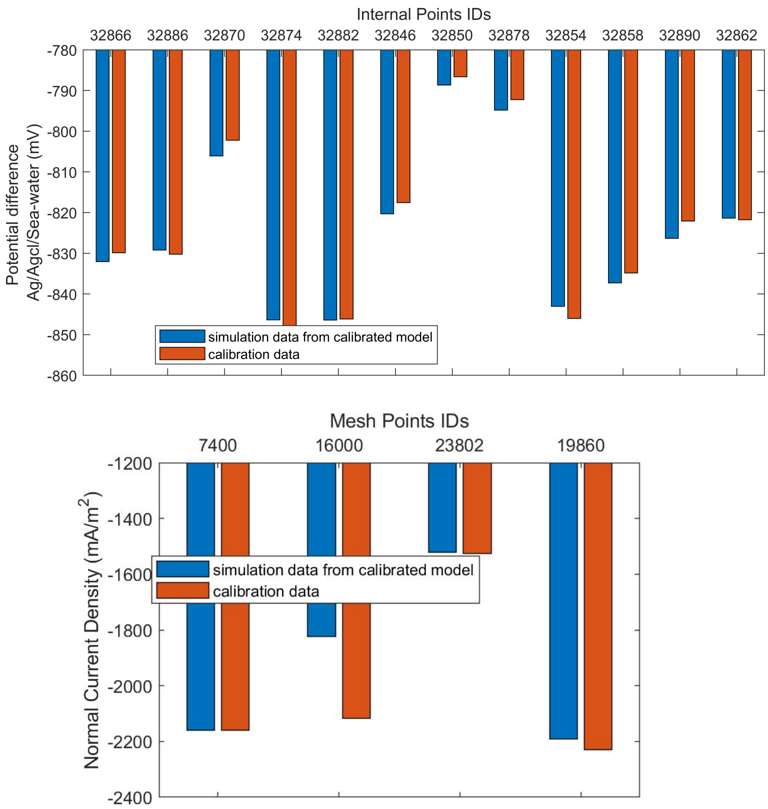


Figure 7.6: The comparison plot between calibration response surface data against the simulation results with solution parameter achieved a) surface potential and b) Normal current density.

This parameter estimation which could take hours (Table 5.4) while entirely relying upon the physics-based simulation is reduced to less than a few hours including the offline cost of surrogate building. The calibration time requirement for the case-study is shown in Table 7.1, with all the experimentation (including the simulation runs) performed on a Lenovo ThinkPad with a 2.30GHz CPU and 32 GB RAM.

Beyond parameter estimation, surrogate also assists in problem and data understanding. For example, from the surrogate-assisted objective plot for the case, only one type of calibration data is available (Case I from Section 6.3), the more clarity into the issue can be reached.

The objective function with only one data type gives the non-converging types of plot (Figure 7.7), that suggest two possible reasons for the problem: 1) the similar effect of the input parameters to the response data (the surface potential data) 2) the correlation between the involved parameters. While the objective plot with both types of data involved suggests a proper minimum. This analysis highlights the importance of the both

Table 7.1: Surrogate Performance Table for the Case-Study.

Offline Cost	Total Simulation Run for training sample generation	9
	Total Simulation Run for testing sample generation	2
	Time for each Simulation Run	Approx. 8 min.
	Time for data collection and surrogate building	Approx. 30 sec.
Online cost	Time for surrogate-assisted parameter search	Approx. 1 min.
Total time	-	Approx. 1 hr. and 30 min.

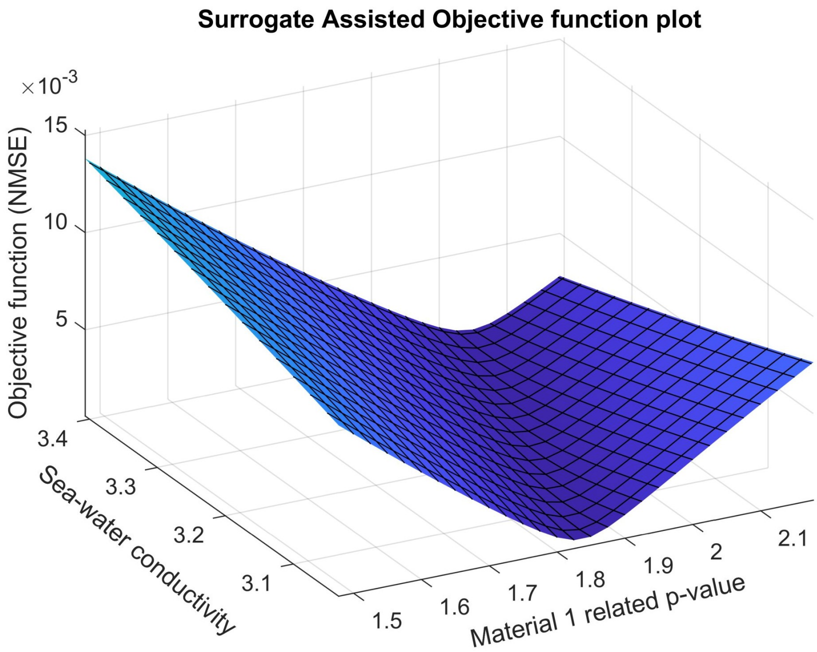


Figure 7.7: Objective plot over the parametric space for only one validating response data-type case (Case I from Section 6.3).

data benchmarking approach as well as the surrogate in efficient real-time virtual model building.

7.4 Case Study II- Surrogate-assisted calibration for Cathodic Protection System's (ICCP type) DT realisation

This case study is about the demonstration of the approach in another type of CP system i.e., ICCP, while case study I presents the experimentation with the SACP type CP system. For this, the tool and methods will remain the same as the first case study except for the model used for the experimentation.

7.4.1 Experimental Setup

The experimental ICCP model built for offshore tank structures (Figure 7.8) is considered. The model represents a buried gas tank of size length = $44m$, width = $5m$, height = $5m$, protected by an ICCP System which includes 4 linear ICCP anodes (represented as A1, A2, A3 and A4 in Figure 7.8) arranged as loops around the tank at different elevations. The normal current density and the current from these anodes are adjusted such that the protection potential is achieved. The current and normal current density values considered can be found in Appendix D.

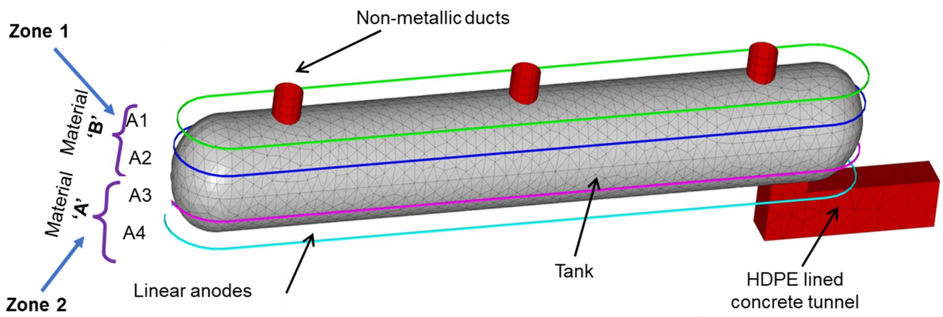


Figure 7.8: The geometry of the offshore Tank protected with ICCP system adopted for the case study.

The data measurement points also termed as internal-points are located at three positions along the tank axis and four positions around the circumference (Figure 7.9), located close to the tank surface (distance: 15 m, element size: 0.6 m). The ICCP model for this system is built with the BEASY tool following the procedure of geometrical modelling and discretisation discussed in Chapter 4. The geometry-related parameters

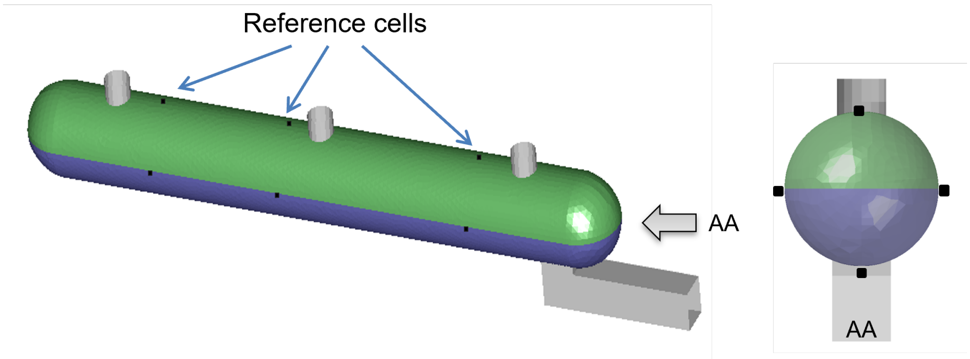


Figure 7.9: The location of the Internal Points for response data measurement with the reference cells.

are not within the scope of the case study as such values can be obtained from design data and/or from the structure geometrical measurements.

Then for the ICCP model built for the tank structure, the modelling parameters required to run the simulation are:

- Polarisation Behaviour
  1. *P-value* (Section 4.1.3) for *Material A* related Polarisation curve of the CP system (Figure 7.8)
  2. *P-value* (Section 4.1.3) for *Material B* related Polarisation curve of the CP system (Figure 7.8)

- Conductivity/Resistivity
  1. *Zone 1 surroundings's conductivity* (Siemens/m)
  2. *Zone 2 surrounding's conductivity* (Siemens/m)

In this case study, the parameter *p-value* will be assessed differently i.e., by relating it to the coating-breakdown factor as the coating breakdown is the main cause of fluctuation in the polarisation behaviour of the metal (Adey 2005). The coating breakdown for the bare metal i.e., without any paint will be considered equal to 1 $(i.e.100\%)$ while for the fully insulated painting case, it will be taken as 0. Now, the *p-value* which is a multiplicative factor approximately account for new polarisation behaviour by taking the reference of the curve for the bare Material (*p-value* = 1) fall in the range of 0 and 1.

Comparatively, estimating the *p-value* is more challenging during the calibration of a CP model, while conductivity can also be measured directly using the equipment available. Thus, in this case-study focus will be on estimating the *p-values for Material A and Material B*.

7.4.2 Response data points for Surrogate building

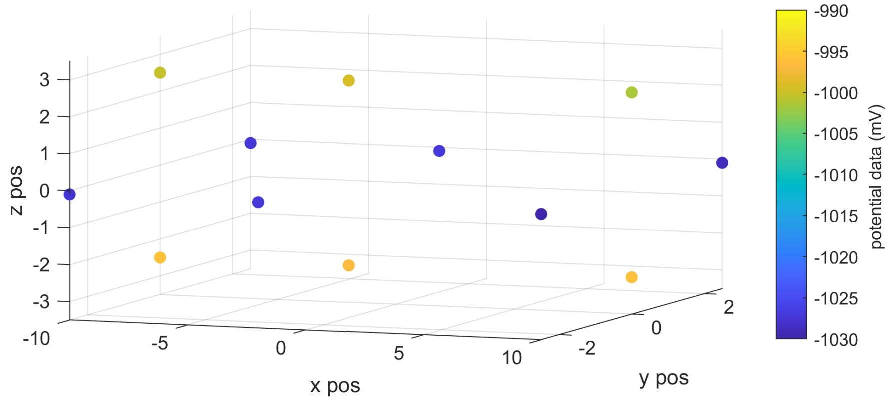


Figure 7.10: The calibration data (surface potential) for selected validating positions of the structure's surface (Figure 7.9).

In this case study, only one response data type i.e., surface potential (mV) data will be considered for surrogate building with total counts = 12 (Figure 7.10). The decision to use only surface potential data is based on the finding that this data type with appropriate data-count is sufficient for calibrating the model when it is about estimating polarisation behaviour only. This finding was reached through preliminary analysis about benchmarking the resources required for calibration and validation, as discussed in previous chapter. For all the considered response data points (total n = 12), the 2^{nd} order polynomial-fit model will be built.

7.4.3 Design of Experiments for Surrogate building

The lower and upper limit of the polarisation-related *p-value* for both curves is assumed to be taken according to the design rules for breakdown depletion. The polarisation behaviour for the corresponding *p-values* is represented in the Figure 7.11.

The inscribed central composite design (CCD) selected before for Case Study I is adopted. Likewise, the MATLAB-based tool '*ccdesign*' is used for generating the sampling points with CCD for the two independent variables "*Material A related p-value*" and "*Material B related p-value*" (Figure 7.12).

The required number of simulations for the 2-variable case to generate the surrogate training data-set is 9. Also, one additional testing sample is generated (Figure 7.12).

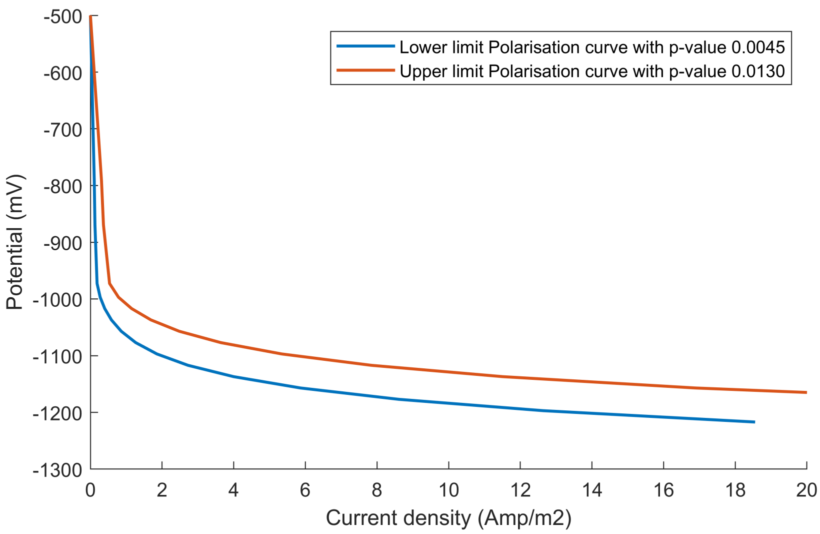


Figure 7.11: The upper and lower limit of the polarisation behaviour selected together with the breakdown factor

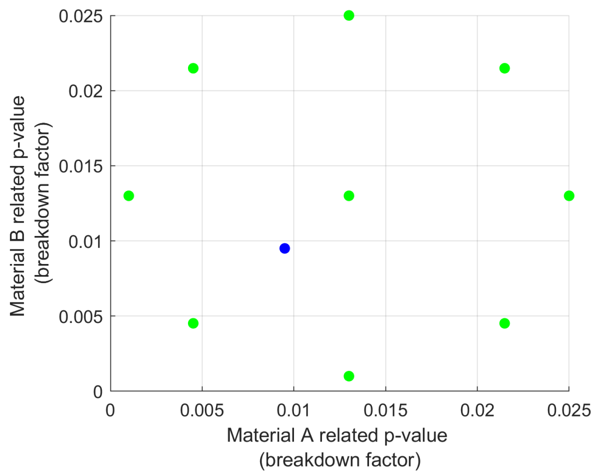


Figure 7.12: Sample points from CCD for two selected parameters for surrogate building (in green), and generated parameter values for surrogate model performance testing (in blue).

7.4.4 Surrogate Model for the ICCP system

The surrogate building follows the similar method as in Case study I. The Response Surface data-fit models are built for each of the nodes (n = 12) for which the data are considered to be accessible/available.

Let's denote the polarisation curve related *p-value* for Material A and Material B as p_value_A and p_value_B respectively. Now, if p_value_A and p_value_B are two variables involved, each positional response data (y) for the corresponding IDs(i) according to equation 7.1 can be represented as:

$$y_i = \beta_{0i} + \beta_{1i} * p\_value_A + \beta_{2i} * p\_value_B + \beta_{3i} * p\_value_A * p\_value_B + \beta_{4i} * p\_value_A^2 + \beta_{5i} * p\_value_B^2 \quad (7.3)$$

MATLAB based tool ‘fit’ is used for the second-order polynomial fit surrogate, i.e., to determine the beta-coefficients (β) for each data-fit model (Equation 7.3) at each response data point.

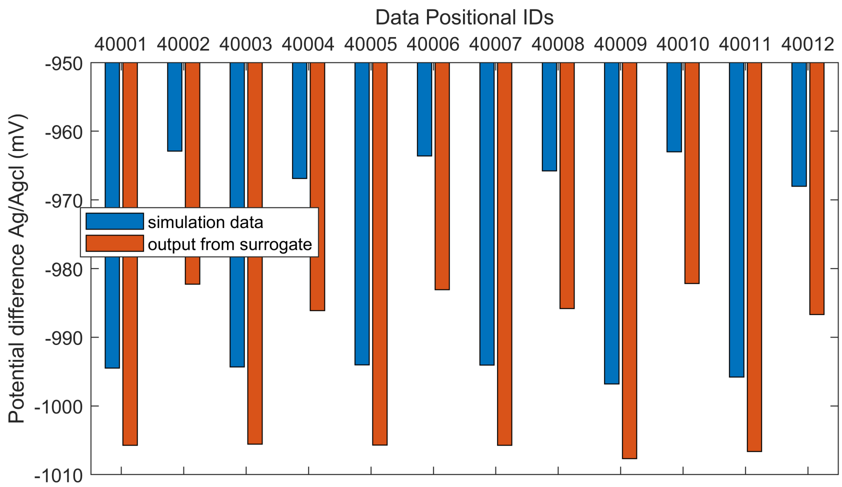


Figure 7.13: Comparison of data from simulation with surrogate output (The testing point is shown as a blue dot in Figure: 7.12.)

For the performance evaluation of the surrogate, the comparative analysis is made between the surrogate output and the full-order model simulation output for one testing parameter value (For example Figure 7.12).

Figure 7.13 presents an analysis indicating that the surrogate model does not adequately represent the CP model responses being $NMSE >> 0.002$, i.e., higher than the acceptable threshold. As a result, the surrogate model requires an update that involves additional sampling points, corresponding simulation runs, and retraining of the model. While the surrogate is not suitable for determining deterministic values, it can still be used to approximate the solution parameter neighbourhood. This approach will avoid the computational burden that could arise from having samples that are not relevant to the solution parameters.

7.4.5 Parameter Space Approximation with the surrogate

With only one type of validating (calibrating) data type, the optimisation objective i.e., the Normalised Mean Square difference between calibrating and model output data will take the weightage constant of 1.

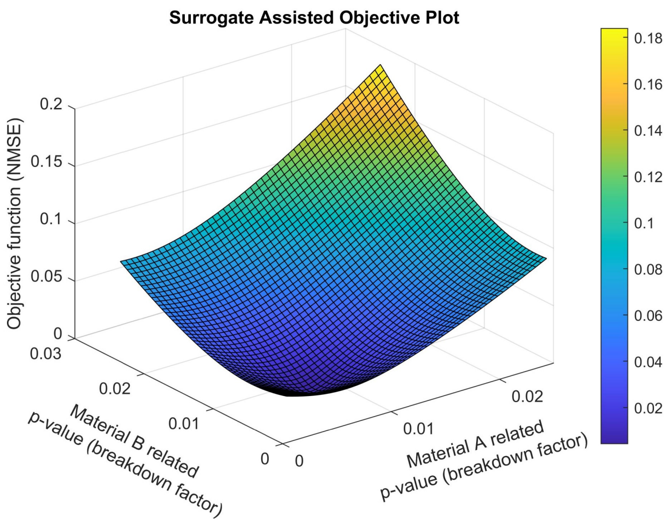


Figure 7.14: Objective plot over the parametric space with response data obtained from the surrogate for the parameter inputs and fixed validating response data.

The objective function's output is plotted (Figure 7.14) over the parameter range for the respective combination of the parameters. The process of finding global minima is repeated as in the Case study I.

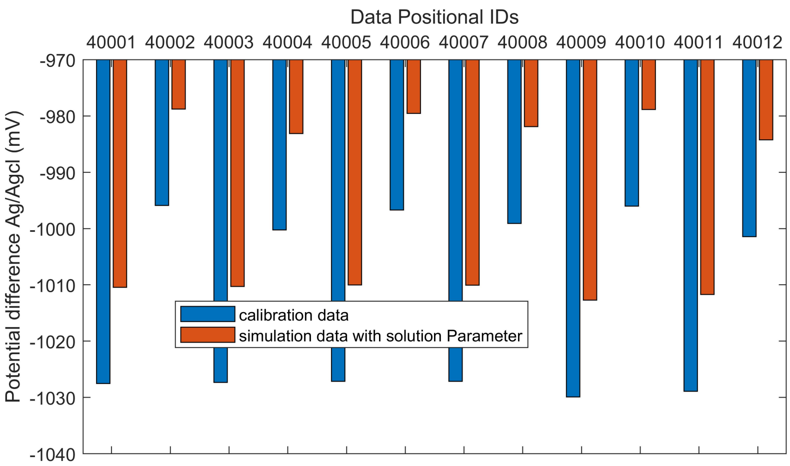


Figure 7.15: Comparison of data from simulation results with solution parameters achieved and the calibration data

The performance of the model with approximated parameters i.e., "*Material A related p-value*" and "*Material B related p-value*" obtained using above mentioned calibrating resources and the surrogate is demonstrated in Figure 7.15.

The analysis of the model's output still shows a significant discrepancy between the solution model and the calibration data. However, from the objective plot and analysis, it also can be obtained that the true solution exists in the neighbourhood of the current approximated one.

7.4.6 Surrogate updating with additional sample Points

As the necessity of the surrogate updating was felt beforehand though utilised to approximate the solution space, the next task is to update the surrogate efficiently. When solution space is narrowed down from the previous DOE space, the additional sample points are generated within the approximated space.

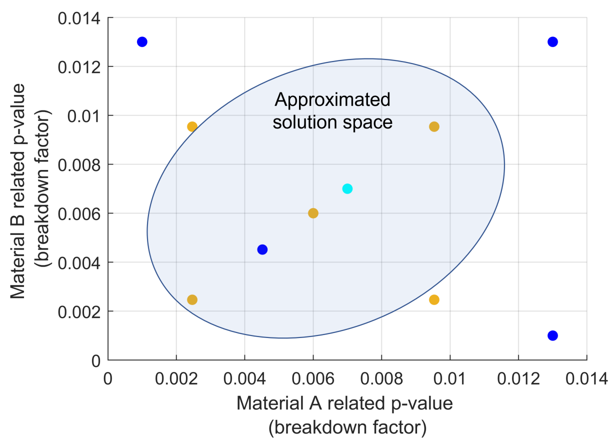


Figure 7.16: Sample points updated (added shown in yellow, previous in blue) in the neighbourhood of the solution reached (in cyan) with the previous surrogate.

Following the previous sample points counts (n = 9) considered for the surrogate building, 5 sample points are added within the space and then the sample points in the far neighbourhood of the space are removed (Figure 7.12 and 7.16). This is done to enhance the efficiency of the surrogate within the solution neighbourhood space, but keeping the data count same as before.

Again the performance evaluation of the updated surrogate is made, by making the comparative analysis (Figure 7.17) between the output from updated surrogate and the full-order model simulation output (simulation output same as in Figure 7.13) for the same testing parameter value. The surrogate model after performance evaluation on their pre-

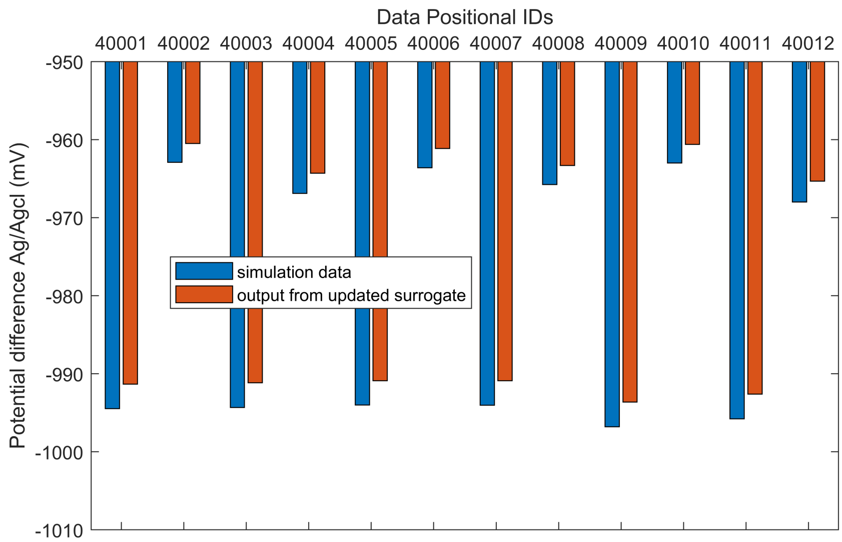


Figure 7.17: Comparison of data from simulation with updated surrogate output.

dictive efficiency though with some error ($NMSE \approx 0.002$) now will be used for minima-based parameter estimation and the solution is anticipated to be better than the previous one.

7.4.7 Parameter Estimation assisted with the updated surrogate

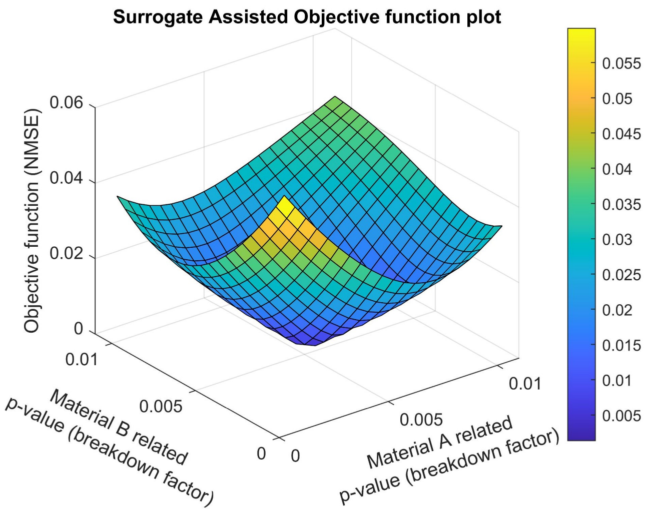


Figure 7.18: Objective plot for over the narrowed solution space with response data obtained from the updated surrogate for combination of parameter inputs.

The objective function's output is plotted over the reduced parameter space (Figure 7.18) for the respective combination of the parameters. The process of finding global minima is repeated as before.

The conical shape of the objective plot (Figure 7.18) within the solution region suggests that the solution minima is more accurate than previous one (Figure 7.14) with fewer closer value parameters located in the neighbourhood.

Results and Discussion

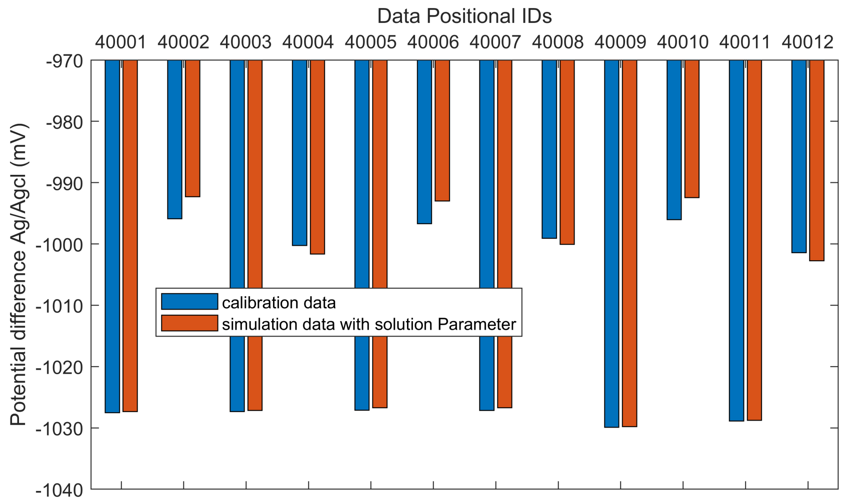


Figure 7.19: Comparison of data from simulation results with solution parameter achieved and the calibration data.

The performance of the full-order model with the solution parameters approximated using above mentioned calibrating resources and the updated surrogate is demonstrated in Figure 7.19. The comparative analysis of the model's output shows that the discrepancy between the solution model and the calibration data are within the acceptable threshold (Objective value ≈ 0.0019), thus the estimated solution is accepted as the input parameter value for the CP model.

This case study has highlighted, the surrogate updating requirement with sample additions for higher accuracy. Moreover, not every time, is the surrogate capable of calibrating the full-order model to the required accuracy. However, a less accurate surrogate still can be used to narrow down the possible solution space in the situation with no idea regarding the solution parameter values.

Additionally, if the surrogate is sufficiently accurate at least for the limited parameter space, it can be used for predictive analysis saving the simulation running time.

7.5 Conclusion

This chapter presented the findings from the research area "*Online Model Calibration/ adaptation: Addressing Input Parameters Uncertainties*" towards fulfilling the requirement of an enhanced and online parameter tuning mechanism. When the different optimisation methods for model calibration are already discussed in the literature including the previous chapters of this thesis, the research aims to ease the issue of the time-consuming simulation-based calibration.

Following the idea of surrogate modelling, this chapter then discussed the role of surrogates in model calibration and adaptation within the DT concept. Surrogate models are proposed as an approximation to full-order models, thus ultimately reducing the number of required simulation cases, unlike the continuous simulation-based parameter estimation (discussed in Chapters 5 and 6 case studies). This benefit is achieved with access to the analytical support required for the surrogate building within the integrated platform. A surrogate model, built offline using simulation input-output data, is used as an alternative model for online parameter estimation. Multiple surrogate building tools are available with easy access to the integrated platform, and the benefits of surrogates in model calibration make them an important feature of the DT architecture.

Not only in parameter estimation but surrogate models also could assist in other ways. Surrogate models can assist in problem and data understanding in addition to parameter estimation, with the computational benefits they offer. Furthermore, surrogate models with not enough approximation of the full-order model to the required accuracy can still aid in identifying solution neighbourhoods . This is particularly useful when there is no idea about the solution neighbourhood space.

The case study presented in this chapter highlights the benefits of using a surrogate model in the calibration of CP models, which is a crucial task for enabling a digital twin. Surrogate-assisted optimisation was utilised to estimate the polarisation data and sea-water conductivity of the CP models. The results showed a significant reduction in calibration time compared to relying solely on full-order model simulation runs. Additionally, the case studies illustrated the other benefits of utilising a surrogate model in various situations.

The findings and analysis of this study suggest that surrogate modelling is a crucial component of the structural Digital Twin ecosystem. This is relevant particularly in the situation where there is lack of a standardised definition of what constitutes a structural

Digital Twin. Surrogate models are not only useful in Digital Twin development but also could play an important role in adaptation, which is a key aspect of the Digital Twin concept. Further research is needed to assess the role of surrogate modelling in dealing with operational uncertainties that may arise during the life of the Digital Twin.

With respect to Digital Twin establishment and its calibration issues, the results of this study are promising. The next step of research is to investigate the real-time predictive capabilities of the Digital Twin. The subsequent chapter will focus on the online adaptation-related issues that may arise during the operational phase.

8 Maintaining Digital Twin for Online Prediction

This chapter presents the findings from the work under the fifth research area i.e., *Maintaining the predictive capability of the model over time*.

One of the major issues that arise over time in the case of the structural model is the spatial variability of the model's parameter (Sohn 2007, Marques et al. 2012). This material-related uncertainty issue during DT's operational phase is discussed in the first section of this chapter. The approach discussed in previous chapters including the surrogate-assisted model calibration will be exploited and expanded together with other analytics for the role of maintaining the predictive accuracy of the Digital Twin.

In the following section, the chapter discusses the potential role of data-driven modelling within DT during its operational phase. The role is on addressing the complexity and also provide a substitute for the time-consuming simulation issue with a physics-based model. This is moreover motivated by the previous approach of utilising the data-driven models (surrogate) as a solution to time-consuming calibration tasks. The promise of data-driven models providing a substitute for the simulation model during calibration also suggests expanding the concept by utilising different Machine Learning (M/L) tools to have an alternative model with the lesser computational time required for (Chakraborty et al. 2021a). The hybrid DT concept (Chinesta et al. 2020, Azangoo et al. 2022) also promotes having an alternative data-driven tool for the structural damage detection task even in the newly arrived situation.

Let us recall the research objectives and milestones set to guide the research activities for this phase of research:

Objectives:

1. To investigate the significant analytical features that aid DT in maintaining its predictive capability despite changes and uncertainty arising with time.
2. To investigate the possibility to pave the way for data-driven prediction towards providing an alternative to the physics-based model in order to avoid its complexity arising with time.

Milestone:

1. Features facilitated by analytical and database that support in providing DT with the predictive capability despite changes and uncertainty arising with time.
2. The past pattern-based i.e., data-driven techniques and/or tools for online prediction, also capable of extending the range of predictive applicability utilising the Digital Twin aspect.

8.1 Maintaining Digital Twin for Online Prediction: Challenges and Scopes

Categorising the different phases of the DT across the life-cycle, 3 major phases can be considered:

- **Phase 1:** Digital Twin enabling by calibrating physics-based model (simulator(s))
- **Phase 2:** Digital Twin operation (Prediction of RUL)
- **Phase 3:** Maintaining the Digital Twin across the lifetime.

Different aspects that support enabling the DT (Phase 1) with automation are already discussed in the previous chapters.

An enabled DT after calibration is now supposed to provide the prediction features but is limited for some time period. The maintenance and re-calibration of the model (Phase 3) are essential due to the uncertainty and complexity arising from time (Gabor et al. 2016, Belostotsky et al. 2018, Kita et al. 2019). It is also an equally challenging and cost involving task as the calibration with the requirement of some adaptive features to be incorporated within the DT architecture. With such features incorporated, a DT with real-time adaptive ability even under the situation of change is anticipated.

From the SHM-related state-of-art discussed in Chapter 2, it is established that predictive modelling in SHM has the goal of cost reduction (including inspection cost) beyond ensuring the safety related to the structure (Alaswad and Xiang 2017). This goal should be highly embraced during the adoption of the DT concept for the ultimate reduction of the physical inspection requirement (Figure 8.1). DT, with real-time predictive capabilities, should provide a long-term prognosis or tool for estimating remaining useful life (RUL) for SHM, but making it less dependent on inspections (i.e., diagnostic).

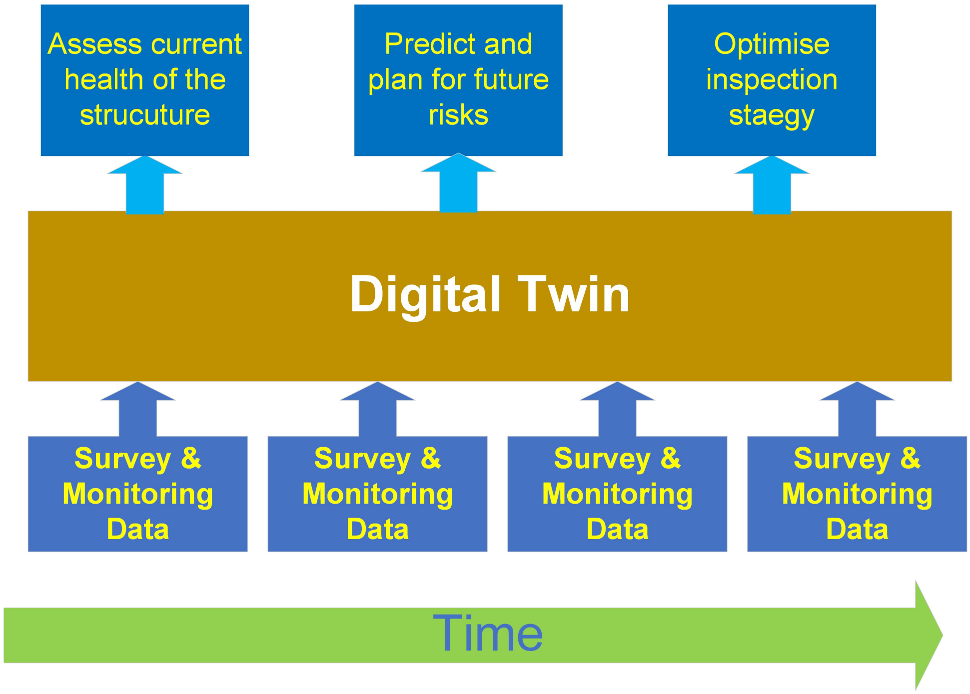


Figure 8.1: Digital Twin's roles anticipated in SHM (Adey et al. 2020)

8.1.1 Uncertainty with time

Though significant progress has been established in having high fidelity and interpretable physics-based models, complex systems still face model-related uncertainties over time. One particular form of uncertainty related to structural models is parametric spatial variability, which arises due to a range of factors, such as manufacturing tolerances, material differences, and variations in wear patterns. For instance, the polarisation behaviour of a material in response to its environment, which is often used as an input parameter in CP models, may exhibit spatial variability due to differences in coating performance. This uncertainty necessitates periodic reassessment of the model while it is in use, either through updating the model or replacing it with a more complex one. However, as models become more complex, it demands further expertise which raises costs, their accuracy may be compromised, and is particularly problematic for real-time applications.

8.1.2 Machine-Learning (M/L) model and DT

Machine Learning (M/L) and Artificial Intelligence (AI) are the other trends coming up together with the DT concept in recent years. Also, a collaboration of machine learning models (including Reduced Order models (ROM)) to the DT is being appeared in recent years (Lee et al. 2020, Chakraborty and Adhikari 2021b, Ritto and Rochinha 2021, Orlova

2022).

Under the scenario of frequent adaptation requirements of the physics-based model, and with the promise the M/L model and ROMs offer in modelling, the potential role of M/L within the DT concept is investigated and discussed.

The role of the Reduced-Order-Modelling is discussed around the parametric variability (a type of uncertainty) issue while M/L models are discussed as the solution to the complexity issue.

8.2 Modularisation and Reduced-Order-Modelling for Model Adaptation

Reduced Order Modelling together with domain modularisation is investigated to address the above-mentioned issue of the model re-assessment requirements due to spatial variability in parameters.

The Reduced Order Modelling (ROM) technique is one of the major accomplishments in theoretical and applied mathematics at the end of the last and beginning of the 21^{st} century (Antoulas 2005). These models rely on an adequate approximation of the solution without sacrificing the model solution accuracy but accommodating the real-time constraints. The term ROM sometime might get confused with the term surrogates as understood differently. To prevent confusion, the research will now refer to the projection-based techniques, which are also discussed briefly in Chapter 7, as the reduced-order modelling. ROMs are used for predicting field quantities providing an approximation solution for the governing partial differential equations (PDEs) or underlying mathematical structure of the system (Rajaram 2020). It is reasonable to assume that there exists no definite answer to the question regarding the choice of approximation modelling. ROMs are believed to be powerful only if the sufficient approximation of input to output map can be achieved by a low-rank subspace (Antoulas 2005). A common framework under this ROMs method consists of extracting the major influencing modes involved in the model solution offline and then projecting the solution to similar problems in that reduced space (Quarteroni et al. 2014).

This ROM and modularisation-based model calibration approach to deal with the parametric variability issue following the DT concept has been already published in Sapkota et al. (2022b).

8.2.1 Domain Modularisation with Data grouping

The diversity of the data location distributions must be taken into account to prevent drawing incorrect conclusions about the performance of the model. To ensure the best uniform distribution of the data, data grouping based on the sub-components offers a solution. On the DT application, the part-modelling is suggested to form a basis for the reduction of the data entries required for analytical needs (Bao et al. 2022). Additionally, the sensitivity of the measurement data to the parameters will differ with the position when the model incorporates varying materials-related characteristics. Thus, domain sub-grouping is essential to estimate/update the parameter with their best-influenced data.

The primary role of the modularisation of the structural model will be finding the appropriate parameters that influence the response data in each module. In other words, modularisation splits and maps the inspection response data with the input parameters based on the sensitivity of the parameters to the module response data. In the case of SHM-related models, when most of the parameters are related to the materials the structure is constructed, the task of mapping the parameter to response data will be straightforward. Modularisation not only benchmarks the data dependency for parameter estimation but also finds the best calibration subset of data for the selected parameter(s). Moreover, this will also provide the remedy for the calibration complexity arising due to the higher count of parameters and dealing with them together during optimisation-based calibration.

Beyond this, the utilisation of the modularisation concept is suggested in assessing the variability of parameters arising with time (Jesus et al. 2017). The factors reasonable for the parametric-variability can form the basis for modularisation. For instance, the region most likely to get affected by load or environmental factor can be separated from others, even if they had similar structural integrity at the time of commissioning of the asset. Modularity further can assist in generating sub-models to minimise the effect in one module on others.

8.2.2 Module-based Reduced-Order Modelling (ROM)

The benefits offered by the approximation (surrogate/reduced order) models such as providing fast predictions to enable computationally efficient design space exploration has been already discussed in Chapter 7. In this section, ROM together with domain modularisation is proposed so that the past simulation data from the full-order model can also be utilised in the ROM model building when modularised.

Parametric model reduction is one of the approaches within the more general area of the surrogates and will be achieved with the projection-based approaches. The goal of parametric model reduction is to generate accurate ROMs that characterise system responses for different values of the parameters (Benner et al. 2015). This explaination of the concept in broader way is provided in Appendix E.

The idea is to utilise the previously generated simulation snapshots matrix even after modularisation and even after variability in the parameter. Digital Twin at its operational phase is assumed to have the simulation snapshots created at a different phase of time for a different combination of the parameters. For example, the parameter and simulation response data during continuous or surrogate-assisted calibration, discussed in Chapters 5 and 7. Now, when ROM is implemented together with the modularisation concept (Figure 8.2) and if more than one module exists, the response data ($U = n * m$) will be categorised into the corresponding modules. For example for module-1, the snapshot matrix (U_1) will have the dimension $n_1 * m$ with $n_1 < n$. If 'k' is the total count of modules, then $n_1 + n_2 + \cdots + n_k = n$. Now, during the parameterisation stage of ROM building, the input variable(s) that was considered to be uniform at the beginning but are believed to have a different pattern of change in each module can be taken in a unique way for each module. However, the sample count (m) need not be varied unless any other less sensitive and parameter-varied snapshot data are removed. The dimension of snapshot matrix can be reduced for the module to reduce the ROM building time by ignoring the insignificant variable of that module.

8.2.3 ROMs' Performance Evaluation and Updating

The performance of the constructed ROM required a validation assessment before using it as a representative of the full-order model. Usually, a set of testing points and simulation snapshots other than those used during the ROM building are used for performance validation, similar to surrogate validation discussed in Chapter 7. The difference, in this case, would be the validation simulation data required to be generated with the variability of the parameter for the different modules. This demands additional parameters for the full-order simulation model from different modules.

If the inaccuracy is more than the threshold during the comparative performance analysis, sample data points are added to the current training data set, and the ROMs are updated to improve their prediction accuracy (Figure 8.2). For the sample data update,

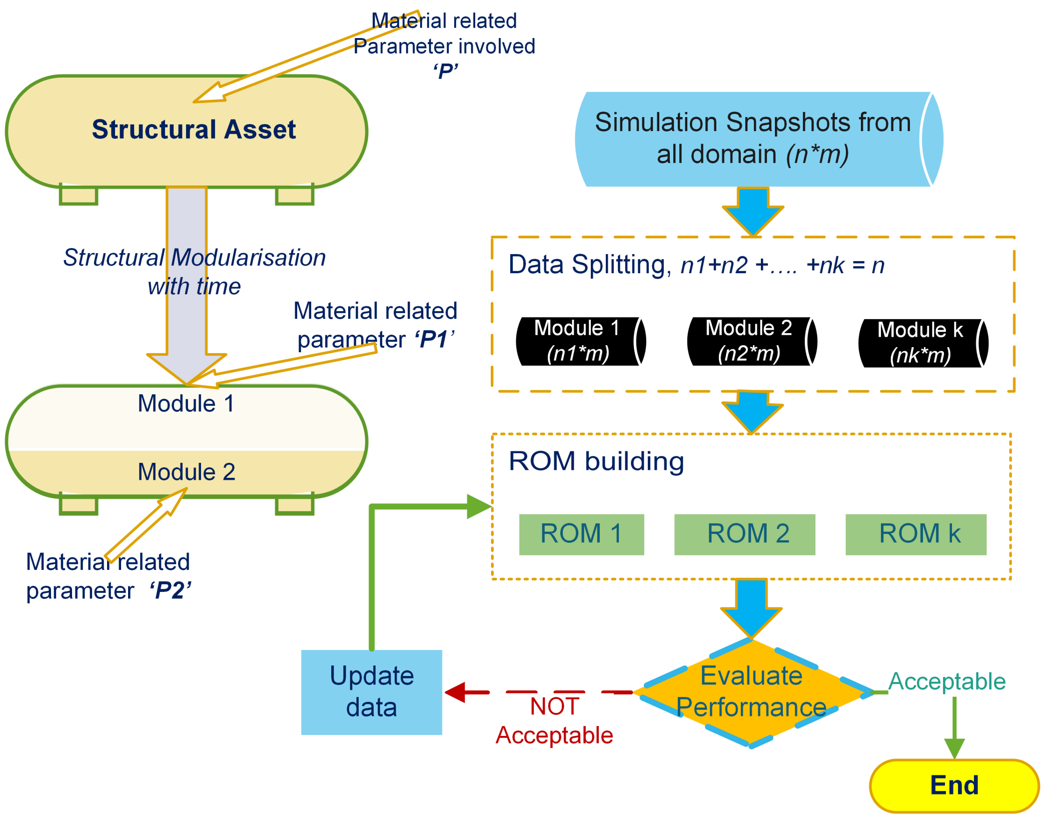


Figure 8.2: Modularisation of the structure and ROM building for each modules

the corresponding material parameter(s) value can be varied among the modules but snapshots will be obtained with the full-order model.

8.2.4 ROMs for Optimisation-Based Parameter Estimation/Adjustment

The ROMs after performance validation can now be used in the "*online*" phase. In the "*online*" phase, adequately efficient ROMs are used for the objective evaluation during the process of optimisation like the usage of polynomial surrogates in Chapter 7. Provided the ROMs are available for each module, parameter estimation would be also performed separately for each module by only relying upon the calibration data belonging to the modules. For each module, the objective function has the corresponding module-related parameter as an input variable plus other parameters that exist from the beginning. It takes less time to conduct enough counts of searches to reach the global minima when the forecast time for each ROM can be within milliseconds to seconds. This also allows for an exploratory search in contrast to the exploitatory search limited by the gradient or non-gradient optimisation technique, thus avoiding the trap in local minima.

The comprehensive approach for real-time and faster calibration of the model that also

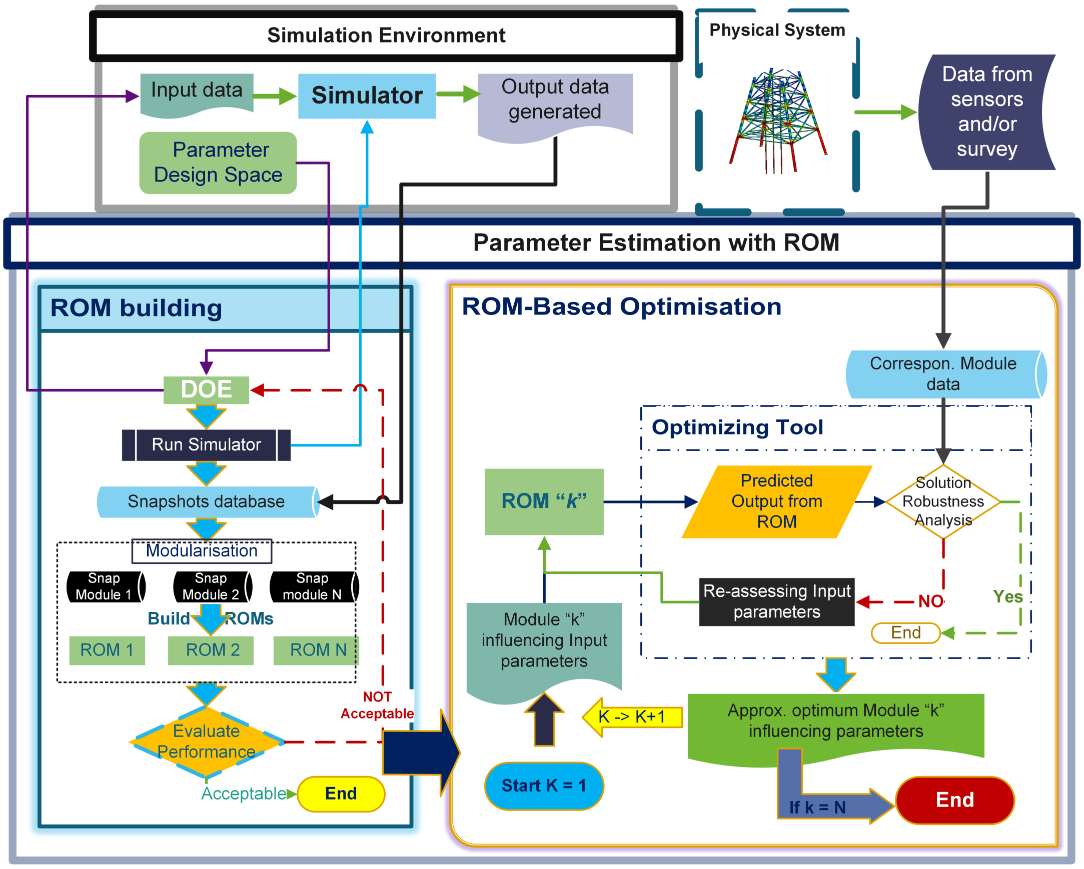


Figure 8.3: Module-wise Parameter estimation with Reduced-Order-Modelling for the physics-based model.

addresses the issue of variability issue of the parameter is illustrated in Figure 8.3. The above-discussed approach offers the core functionalities essential within the DT system required for real-time prediction, i.e., the potential of parameter-related adaptation.

8.3 Case Study- ROM for Re-calibration with Parameter variability assessment of CP model

8.3.1 Experimental Setup and Parameters

The experimental CP model (used in the previous chapter) built for offshore structure (Figure 5.2) using the BEASY simulator is adopted. Similarly, the Software-Simulator integrated platform is utilised for the expedition of experiments with automation. The analytical support required will be received, within the integrated platform.

Though the same CP model built for the structure is adopted, the core parameters considered to run the simulation are limited to the polarisation-related:

1. *P-value* (Section 4.1.3) for Material-1 related Polarisation curve of the CP system.
2. *P-value* (Section 4.1.3) for Material-2 related Polarisation curve of the CP system

The goal of this case study is to demonstrate the benefit of the ROMs in the variability assessment of the above material-related parameter (p-value for Material 1).

8.3.2 Modules and parameter mapping

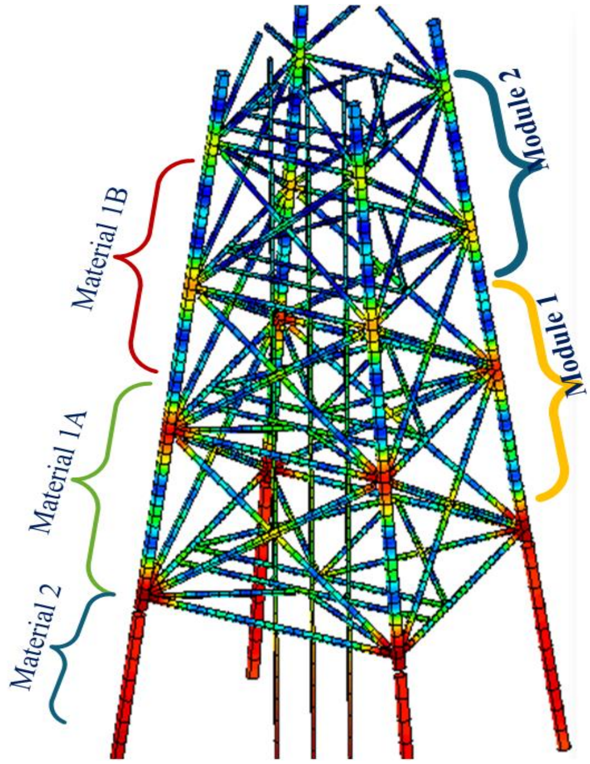


Figure 8.4: The adopted offshore structure (Figure 5.2) with modules categorised based upon the material 1 sub-categorisation for its polarisation behaviour assessing.

The primarily focused parameters are the parameters representing the polarisation data corresponding to the main structural steel in the structure. This polarisation behaviour might change over time due to localised coating breakdown, calcareous deposits, and mechanical damage. Assuming the pattern of degradation varies with depth i.e., considering possible variability of the polarisation and/or coating, *material 1* is sub-categorised into two sub-materials i.e., *material 1A* and *Material 1B*, and correspondingly two modules are separated as in Figure 8.4.

8.3.3 Model Validation/Calibration data.

Two types of calibration data are considered: surface potential (mV) and normal current density (mA/m^2). The count of data positions on the structure surface (Figure 8.4) is 49 and 10 respectively. A similar procedure of generating calibration data from a virtual reference model as in the case study in Chapters 5 and 7. However, in this reference full-order model, possible parameter variability along the depth is taken into account and simulation data is obtained.

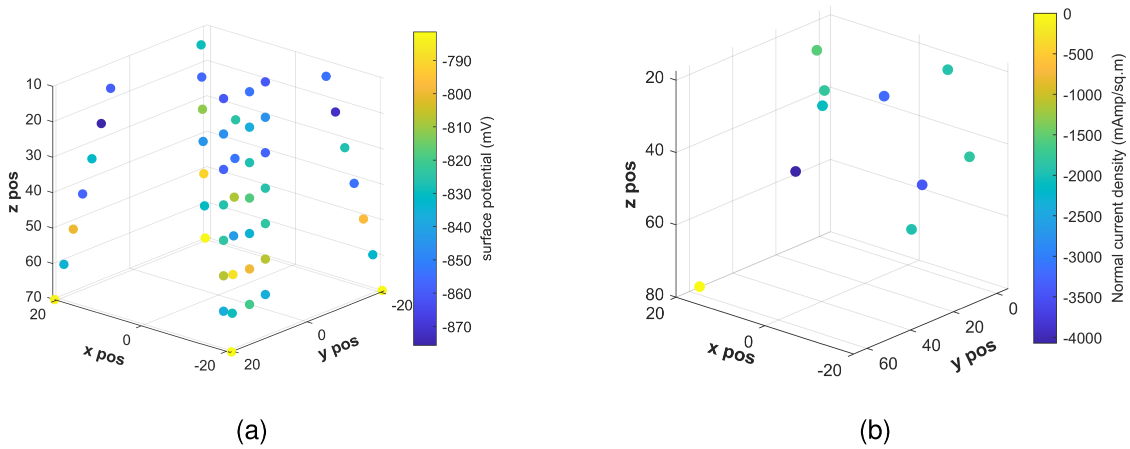


Figure 8.5: The data for calibration obtained after a simulation run from the reference model before introducing noise a) surface potential, b) normal current density, from a selected validating position of the structure surface (Figure: 8.4).

Also, to represent the type of errors expected in real-world inspection surveys noises are introduced to the calibration data (Figure 8.5) before using the data for validation and calibration. Then, the calibration data are separated for the modules categorised based on the region they belong to.

8.3.4 Module-wise Reduced-Order Models

In the context of implementing a DT approach and having access to simulation snapshots over time, the data obtained from the previous surrogate-assisted calibration experiment will be utilised. The sample data, generated with an inscribed central composite design (CCD) approach making 9 sample points (m) for the two-variable case is considered.

Then, the data snapshots for POD (or SVD) based Reduced-Order-Modelling are stored in matrix X of dimension, dim(X) = *total data positional counts* ($n = 49 + 10$) $*$ *sampling counts* ($m = 9$). This entire snapshot matrix will then be split into the matrices equal to the total number of modules. For example, in the two modules case, *X* is split into X_1 and X_2 with dim(X_1) $=$ *data points in first module* $(k = 28 + 6)$ $*$ *sampling counts* $(m = 9)$ and dim(X_2) $=$ *data points in second module* $(n - k = 21 + 4)$ $*$ *sampling counts* $(m = 9)$.

For the parametric ROM building, an open-source python-based tool EZyRB (Demo et al. 2018) is adopted that utilises the SVD method for ROM and RBF for parametric interpolation. The reason for relying upon the particular tool is due to its applicability in building ROM only from the input-output data i.e., without assessing the Operator Matrix of the full-order model. Also, this research is aiming towards utilisation of the previously available analytical tools in the adaptation requirement. The tool can be interfaced with the simulation software within the integrated platform simply and can be used to generate the parametric ROM using the simulation snapshots.

SVD-based ROMs are then built for both modules using the corresponding snapshot matrix, and the tool. During the parameterisation phase, the *p-value for Material 1* will be assessed as the *p-value for Material 1A* and the *p-value for Material 1B* in *Module 1* and *Module 2* respectively.

8.3.5 ROM assisted Parameter Estimation: Result and Discussion

Having the ROMs now enables performing the exploratory search for the solution parameters i.e., to obtain the objective function output for different parameter combinations with less computation time.

Assuming two different validating/calibrating data types, the Normalised Mean Square difference between validating/calibrating and model output data with weightage constant (2:1) for the two data types, is undertaken as validating criteria and ultimately as the objective function for the minimisation problem.

The categorised module-based calibration data for the corresponding module is fixed and an objective surface plot is obtained over the two-dimensional parameter space for each module (Figure 8.6). For the both modules, during the objective calculation over the parameter space, the minimum value is stored, which means an exploratory search method is made for all possible combinations of the parameters to reach the global minima in the presented case study. The plot also signifies the similar role of ROMs as of polynomial-surrogates discussed in Chapter 7 in surface smoothening and avoiding local minima.

The exploratory search method could find the global minima in this case study. How-

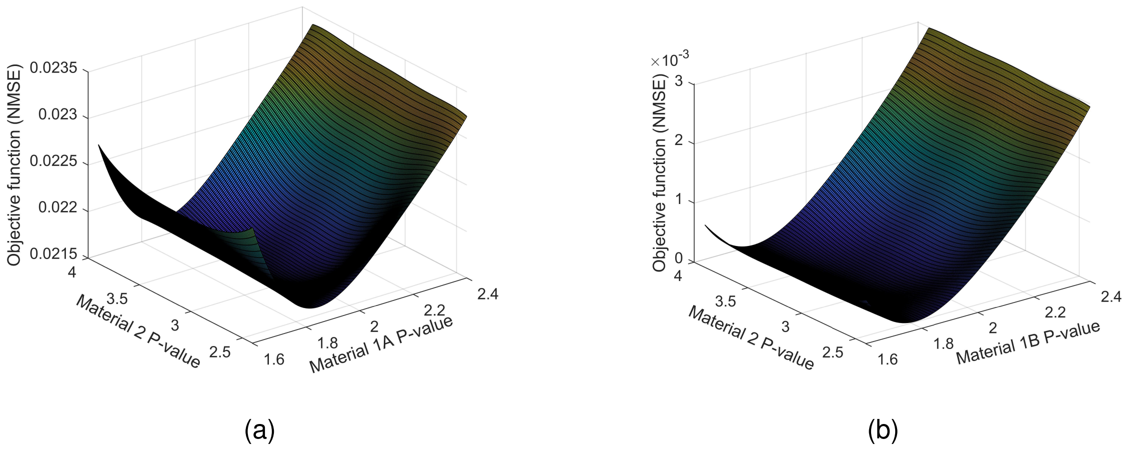


Figure 8.6: ROM-assisted objective plot over the parametric space with validating, response data and ROM of the corresponding module a) module 1 b) module 2

.

ever, in other situations when the exploratory search tends to consume more time and induces a delay in the process, different gradient and non-gradient-based optimisation algorithms are required.

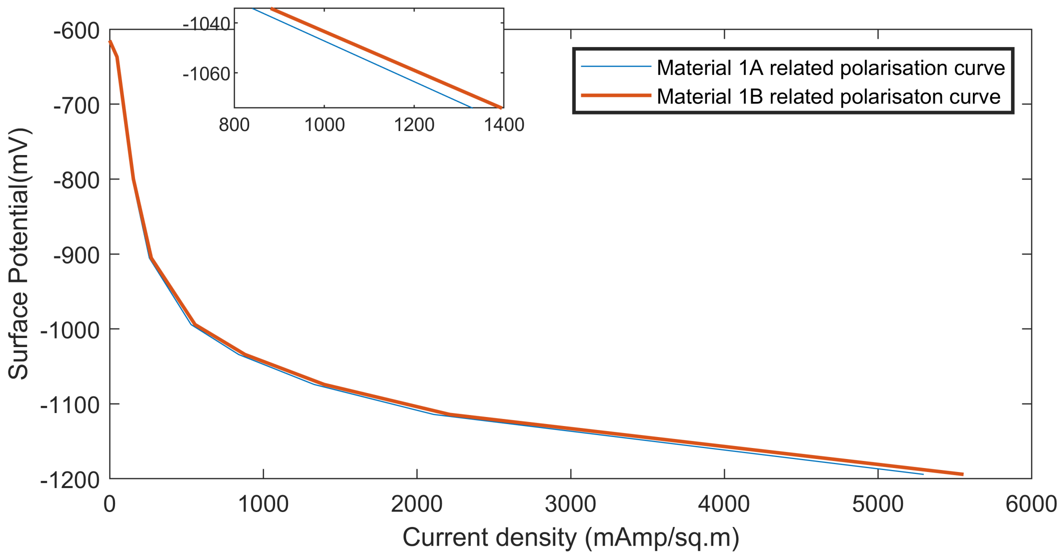


Figure 8.7: Polarisation behaviour for the respected materials suggested after ROMs assisted minima finding-based parameter estimation

The polarisation curves set for the two modules that are obtained with the above-implemented p-value estimation approach can be visualised in Figure 8.7. The performance of the model with estimated/updated polarisation curves is validated against the calibration data as in Chapters 5 and 7. As the solution response data are within the error threshold when compared to the calibration data, the above solutions are accepted.

The results show that the ROM-assisted model re-calibration (adaptation) approach is capable of replicating the inspection results in the DT model under the situation of parameter variability. The ROMs built by utilising the past simulation snapshots, on the other side reduce the computational burden significantly. Not only this, modularisation forms the basis of having ROMs that take in parameters that influence the response data of the module, which is why parameter variability among the modules can be traced.

The quality and amount of the calibration data that are available among the modules would be the approach's restriction. The quality and quantity of the calibration data at the beginning during the situation of non-variation among the parameter won't be enough in the new situation where another parameter-related uncertainty i.e., the variability of the parameter appears with time. This situation thus requires the benchmarking of the resources for the reached situation. This though induced complexity in the process can be achieved by using the resources benchmarking approach proposed in Chapter 6.

8.4 Machine Learning within DT as alternative to simulation model with time

The conceptual DT introduced was physics-based and so is the focus of this research as a physics-based model is essentially required for some complex domains. The flip side of the physics-based digital twin is the complexity of its implementation since it requires a more detailed structural model of the asset, calibration of the model, and also often consume more time during computation.

On the other side, DT towards endowing Artificial-Intelligence (AI) to it requires high integrity, awareness, self-learning, and adaptability. These intelligence-providing features cannot be achieved without the advanced analytics features incorporated within. These analytics with features like analysis and learning to bring self-configure, self-adapt, and self-learning functionalities have already been adopted in manufacturing to increase productivity, speed, flexibility, and efficiency of the process (Lee et al. 2017).

It is fair to say, if a DT driven by data analytics with slightly less predicting accuracy is available as an alternative to the physics-law-driven DT, they are largely complementary. Moreover, the data collected from real-system during the operational time-span together with simulation data can be used to predict an approaching unplanned situation after adequate curation of the data. This will further enable data-based decision-making about predictive maintenance, control, and efficient inspection possible in real-time. Also, it is already accepted that the future of "*intelligent*" DT applications relies on the combination of these two, uniting the strengths of each side. In this context, also a concept of a hybrid twin (Chinesta et al. 2020) is proposed that encompasses a digital twin including all physics of the structure to the purely data-based digital twin with the reduced order models/surrogates. However, realisation of such DT demands integrated advanced analytics for example Machine Learning to the DT architecture (Barricelli et al. 2019, Lee et al. 2020). When facilitated with advanced analytics like M/L, DT can also analyse the trend of the physical and the simulation data across the multiple stages of the product life cycle to enhance on-demand predictive services.

M/L could assist in constructing surrogates, ROMs and/or other supervised and unsupervised models within DT in several ways that ultimately assist in reducing the computational cost. To some extent, the previously discussed ROM has a common goal with this M/L in the adaptation of the predictive model within the DT concept. Though the M/L-based model offers computational benefits over the physics-based, the major drawback of such a data-driven model is the deterioration of its predictive capabilities outside the training set scope rapidly. Similarly, the models fail to perform when little complexity arises requiring new model building with much more updated training data to capture the new trend (Chan et al. 2020). Therefore, such a model also requires the mechanism to identify the drift in their performance as well as the appropriate model updating mechanisms. However, until there are enough data to represent the emerging trend, the computational advantages of such an M/L model, as well as updating them, still favour the data-driven model over the physics-based one.

8.4.1 Data collection and Storage over Time Series

A well-maintained physics-based DT that has been in operation for several years is a reliable resource for understanding the operational past of an asset. It provides a wealth of physical data from sensors and inspections, as well as virtual data from simulations, that can be used to analyse the root causes and trends of the asset's behaviour over time.

In order to develop a successful M/L model, it is necessary to collect data specifically related to the deterioration, fatigue, or damage patterns experienced by the structural components over a certain period of time. It is important to determine the optimal data collection interval to minimise unnecessary expenses and to gain a thorough understanding of the structure's complex patterns and performance. Although the fatigue index may exhibit non-uniform trends due to seasonal and loading variations, it should still be recorded. Furthermore, any outliers or noise in the data, such as sensor malfunctions or measurement-related technical issues, should be eliminated before relying on it for prediction.

Once large amounts of data related to assets with similar materials and performance characteristics are available globally, accessing this data, would be useful for identifying the trends.

8.4.2 M/L tools for trend capturing

Materials scientists are increasingly adopting the machine learning tools to discover hidden trends in data and make predictions (An et al. 2015, Wagner and Rondinelli 2016, Azimi et al. 2020). M/L algorithms can be trained on data from both physical systems and their virtual counterparts, including both historical and real-time data. The trained model is then tested on a portion of the data, and if the performance is not satisfactory, the process is repeated until the testing results meet benchmarked criteria. Once the model is validated, it can be used to make real-time and future assessments.

Different types of Machine-Learning techniques such as data mining, deep learning, regression, manifold learning, etc are available to build such data-driven predictive tools. M/L can be broadly categorised into three types: supervised, unsupervised and reinforcement learning (Li 2017).

Supervised learning: For the supervised learning model, labelled data-sets are provided i.e., the independent and dependent variables are fixed. It is mostly applicable in situations where the complete behaviour representative data (for example all failure modes and expected behaviours) are available so that the M/L model map the independent variable to the dependent variables by finding a generalised function that maps inputs to desired outputs. Some commonly used supervised algorithms are Polynomial Regression (Ostertagová 2012), Logistic Regression (Kleinbaum et al. 2002), Decision Trees (Charbuty and Abdulazeez 2021), Support Vector Machine (SVM) (Hearst et al. 1998), Random Forest (Biau and Scornet 2016) and Artificial Neural Networks (ANN) (Zupan 1994).

Unsupervised learning: In unsupervised algorithms, the data-set is not labelled, so the appropriate internal representation of the input is desired from M/L.

Reinforcement learning: It is provided with feedback evaluation algorithms that learn to act given an observation. The computer then learns how to achieve the goal through trial-and-error interactions with its environment (Harmon and Harmon 1997).

8.4.3 Supervised M/L models for time-step prediction

The role of the M/L model such as the polynomial-regression model and the ROMs for DT realisation is already established in this research. Furthermore, within the DT, supervised M/L model with time has the potential to extract the patterns linking different loads and environmental conditions and structural performance to them. For this, the simulation data obtained within the DT pave the way to have the parameterised and time-dependent M/L model (Figure 8.8).

The parameterisation process mostly followed while enabling physics model-based structural DT would favour the usage of Supervised Learning among different M/L techniques considering the parameters as the independent variables. Also, to make the time-step predictions/projections (time as one of the input independent variables) polynomial regression (for simpler) and/or Artificial Neural Networks (for higher complexity) can be exploited.

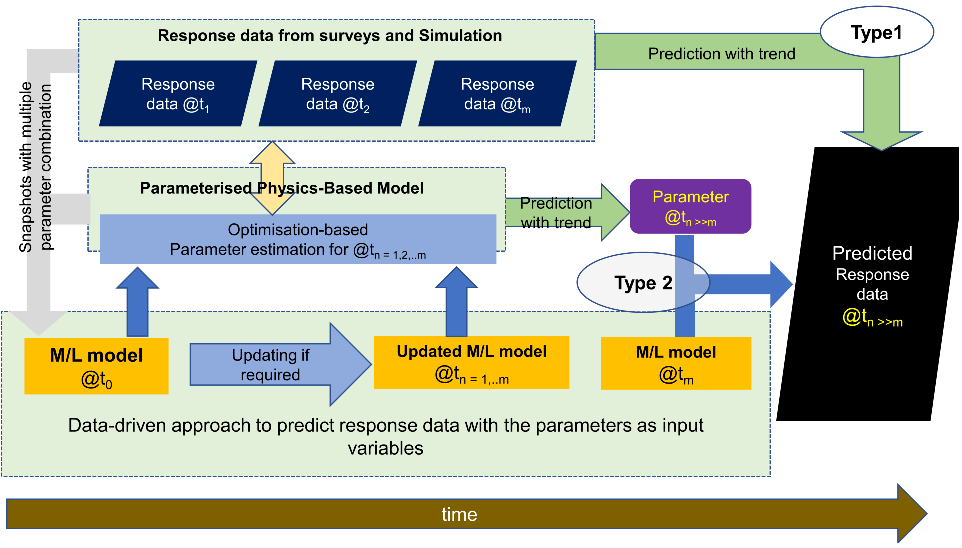


Figure 8.8: Data accumulation with time and usage of M/L model (Two types) for trend capturing utilising the past data

Utilising the data accumulated with time, the supervised M/L model that can predict future response data can be built in two ways (Figure 8.8):

Type 1- This type of M/L model considers time as an input variable, and only response data that is associated with the corresponding discrete-time will be utilised for model training. This type of trend-based predicting M/L model may become obsolete with even a slight disturbance in the trend, as it does not take into account other parameters and response data are less likely to be directly dependent to the time.

Type 2- This type of model is parametric, similar to a physics-based model, but with a different working principle. The M/L models are trained with the simulation snapshots as simulation parameter as the input and the response data from the real-structure or simulation as the output data. The surrogate model for CP system discussed in Chapter 7 is a typical example for such M/L model. This model also requires estimation of parameters, which must be tracked over time by relating them to other environmental factors. The prediction with the M/L model tends to remain reliable for a longer period compared to *Type 1*, as long as it is built with the snapshots data of possible future parametric changes and other relevant factors. However, over time, this type of data-driven model will still become obsolete and require updating.

Performance-drift of supervised M/L models over time is a known and common issue, as a new influential independent variable might appear or the response to the parameter might change. In such a situation, the identification of the independent variable using feature extraction methods should be reassessed to continue with supervised learning. Moreover, obtaining a large number of failure datasets for SHM related to critical assets can be problematic, as damages are usually only recorded for a fraction of the asset's total life. Additionally, supervised learning approaches are inapplicable for newly procured assets as there is no historical data to leverage.

Deep learning, or deep neural networks, is a another advance M/L scheme that is also receiving attention for its data-driven prognostic role in structural health monitoring (Khan and Yairi 2018). Deep learning can be integrated with supervised or unsupervised learning with reinforcement learning as a function approximator. It has the advantages of data classification and feature extraction even in unsupervised cases, making it useful for detecting unprecedented complex situations for assets arising with time. However, this research will be limited to the supervised M/L on establishing the role of the M/L within structural DT but also foresee the potential of reinforced or deep learning approach as a future scope.

8.5 Case Study- Machine Learning for CP System's Predictive Analysis

This research has previously discussed the simulator-based CP model for the CP system. Once such model is calibrated to the required accuracy, it can generate virtual data that closely resembles real-world data. Additionally, physical data can be obtained from sensors or surveys. The polarisation data, response data, and anode consumption rates obtained from simulation, survey, or calibration at different time spans can now be studied to identify trends using the M/L concept.

This section will present a case study to demonstrate the application of the M/L model for a CP system.

8.5.1 Experimental Setup

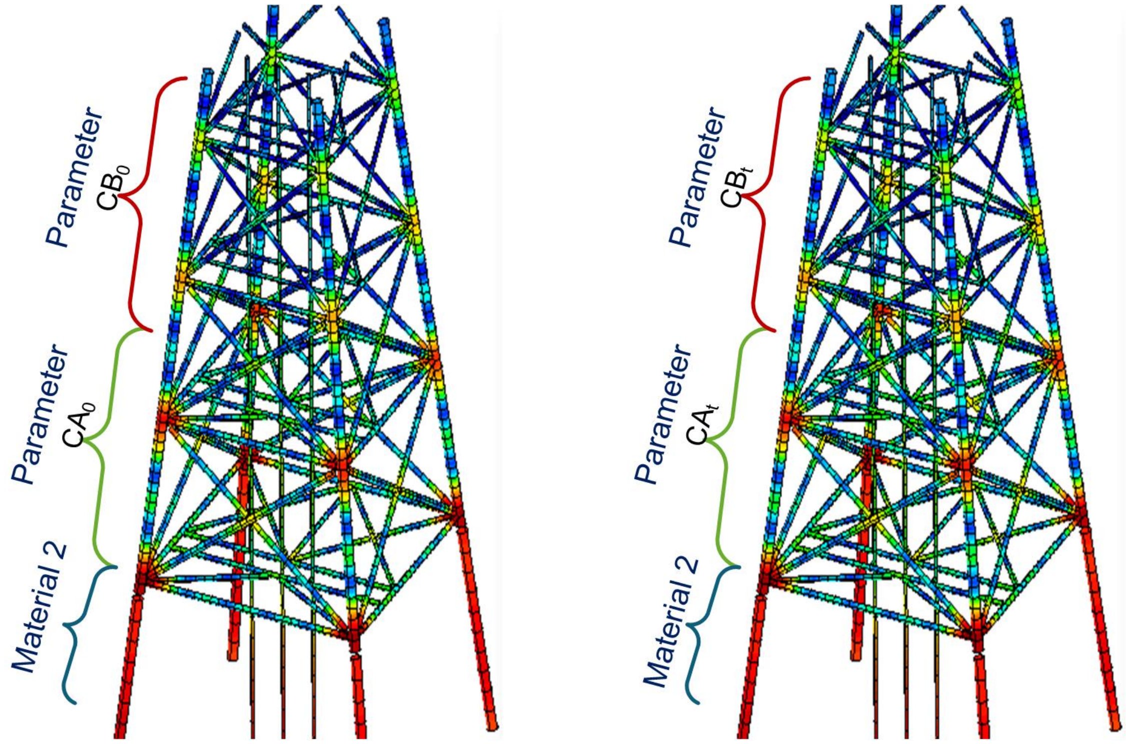
The experimental CP model built for offshore structure (Figure 5.2) using the simulator is adopted. As, in the previous case study of this chapter, the core parameters considered to run the simulation are limited to the materials-related polarisation behaviour. Furthermore, in this case-study the *Material 1 related p-value* will be split into two parameters, while *Material 2 related p-value* will be treated as a constant, limiting the parameters count to two. The categorisation of the material aligns with the modularisation concept discussed in the previous case-study, which considered possible variability over time.

1. Parameter CA_t (coating breakdown for Material-1A) of the CP system
2. Parameter CB_t (coating breakdown for Material-1B) of the CP system)

New notations are used for the parameters to reflect the corresponding relative years from the time of reference (Figure 8.9). CA_t and CB_t will represent the coating breakdown for the *Material-1A-related Polarisation curve* and *Material-1B-related polarisation curve* of the CP system (and model) with the suffix '*t*' representing the corresponding relative years. For year $t = 0, 1, 2, \ldots N,$ the parameter will be updated either when new calibration data from the physical structure are obtained or with other method (tool) to predict or interpolate the parameter value when necessary.

8.5.2 Calibration and Model updating with time

In practical applications, data obtained from physical inspections are collected at different intervals, and the model are adapted accordingly by updating parameter values in



(a) (b)

Figure 8.9: Materials related parameter denoted to reflect the change with time

the event of changes in material properties (such as polarisation behaviour or coating breakdown) and/or environmental factors (such as conductivity).

In this case study, the previously discussed polynomial regression approach is utilised to update the model over time. A similar calibration-data generation approach is implemented, where data are obtained from the reference model while considering possible changes in polarisation behaviour and introducing some error to make the data more realistic. Surrogates are built at different time intervals using the same procedure outlined in Chapter 7, incorporating the above-mentioned parameters. In the previous experiments, surrogates were built for a fixed time by neglecting other factors, i.e., only varying the parameters of interest. Moving forward with time, the usability of previously built surrogates should be investigated to identify and benchmark the threshold and/or criteria for surrogate updating requirements, which can be accomplished using the automated platform and resource benchmarking approaches discussed in earlier chapters.

For the SACP-related model mentioned above, anodes size also influence the response data (Figure 4.4), but were previously deemed less sensitive and therefore disregarded. Anode-related data were held constant during surrogate building, meaning that changes in anode size were not captured by the two-variable input surrogates. This raises the question of the surrogate's usability over time. For example, when surrogates are built for two breakdown-related parameters at particular anodes size, the predictive capacity of the surrogate may decrease and uncertainty may increase if the anodes size changes. The experiment suggests that the surrogate need not be discarded until anodes depletion reaches $15 - 20\%$ from the stage of surrogate building, as the predictive capacity of the surrogate depreciates with uncertainty increasing by less than 5% error. However, if the change in anode size is more significant (i.e., $> 20\%$), the surrogate (M/L model) should be updated (Figure 8.8). The analysis of this surrogate updating requirement corresponding to the anode status related status can be found in Appendix F.

8.5.3 Predicting important variables with calibration

It is well known that predicting the anode size itself is the goal of any CP-related model whether it is physics-based or data-driven. Also, when the anode sizes have depleted significantly, the reduced size should be considered as the input to make a further time-step simulation. Regarding, the usage of surrogates in such a case, the incorporation of anode size-related variables into the surrogate (M/L model) is required. However, this task is somehow limited by two reasons: 1) the count of anodes which ultimately increases the number of input variables if considered, 2) the anode sizes are already dependent variables to the other input parameters which will invite complexity to the model when they are considered as input parameters (usually independent variable are taken as input to any model). Though having limitations, the model can be rebuilt with further parameterisation utilising the supervised approach. However, the number of parameters should be limited in order to avoid the complexity which means the anode sizes will be considered as other input parameters but limiting them to one or a few. For example, taking an average of the anode sizes or adding an extra parameter with the average or initial size to reflect their distribution.

Data for surrogate modelling with a 3-variables case

The polarisation-related data (breakdown factors) and anode consumption factor data together with the corresponding response output are obtained at different time-span. The data collected from different time-step ($t_0, t_1, \dots t_n$ years) has three varying parameters when included anodes related parameter. For example, in this case, the data are obtained during surrogate-assisted calibration (two variables case) for the time instants where the anode consumption factor remains fixed for each time instant but get updated with years. Now, these data being converted to three variables case (Figure 8.10) can be used to train the data-fit model with three input parameters.

While the input-data pattern is changed as new factor is considered, the output response data type will remain same.

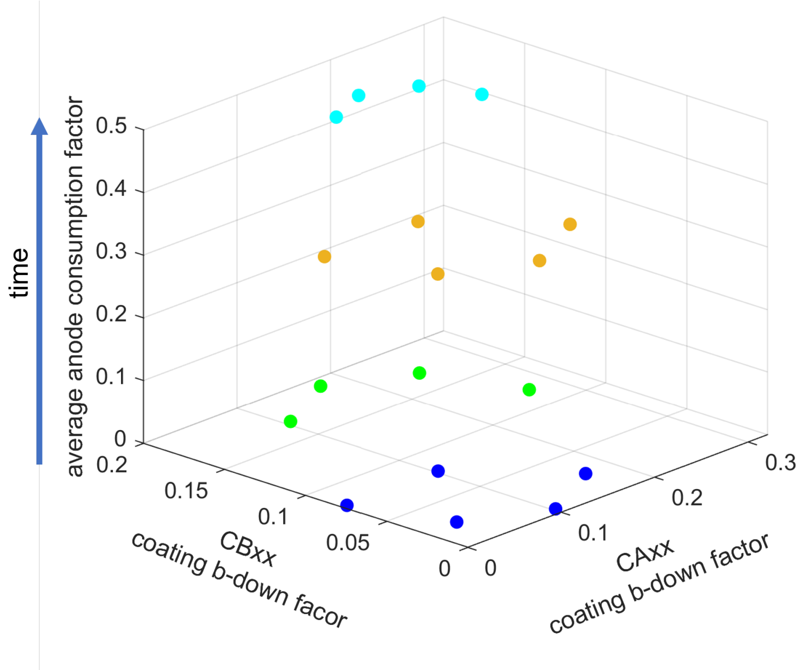


Figure 8.10: Data samples collected at different time-span, and being utilised in surrogate-building for the three variables case

Data-driven Model building

For the 3-variable case, the representative 2^{nd}-order polynomial data-fit model will be:

$$y_i = \beta_{0i} + \beta_{1i} * CA_{xx} + \beta_{2i} * CB_{xx} + \beta_{3i} * \sigma + \beta_{4i} * CA_{xx}^2 + \beta_{5i} * CB_{xx}^2 + \beta_{6i} * \sigma^2 + \beta_{7i} * CA_{xx} * CB_{xx} + \beta_{7i} * CA_{xx} * \sigma + \beta_{7i} * CB_{xx} * \sigma \quad (8.1)$$

where, β represents regression coefficients, σ represents the average consumption factor with CA_{xx} and CB_{xx} representing coating breakdown for Material-1A and Material-1B respectively. The surrogates are built for each of the nodes $(i = 1, 2, \dots n)$ considering the availability or accessibility of corresponding nodes related validation (calibration) data.

Calibration Step

When the calibration response data from the physical system be available, let's say at time "t_{n+1}", the calibration task can be performed. The calibration/validating data set, objective computing metrics and the procedure are repeated from the case study in Chapter 7. Calibration data is generated by making the forward time-step simulation where the two coating breakdown related parameters' values will be assumed to represent the change for the time (Table 8.1). Additionally, anode consumption data for the state (time "t_{n+1}") can be obtained from the time-step (5 years for this case) simulation for the analysis.

Table 8.1: Variable values for calibration and solution data

Parameter Cases	CA_{xx}	CB_{xx}	Average Anode Consumption (σ)
Values for Calibration data generation	0.26	0.18	0.6904
Values at Minima point (Solution reached)	0.2575	0.2000	0.6925

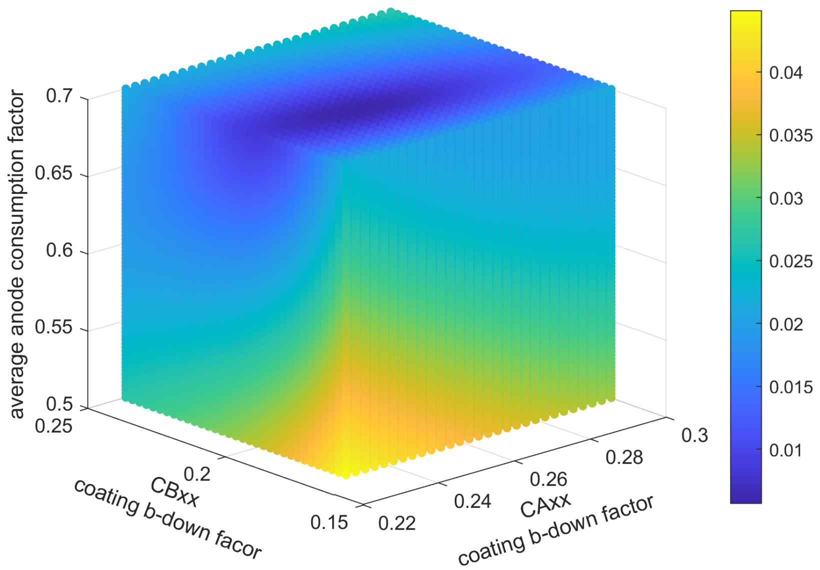


Figure 8.11: Objective plot over 3 dimensional space for 3 variables case

Now, the variables estimation will be undertaken in a similar way to the previous surrogate-assisted calibration approach. However, this time the search will take place in 3 dimensions (Figure 8.11). During the objective calculation, the minimum value is stored, for all possible combinations of the parameters to reach the global minimum. In this case, the anode status in the CP system is considered together with the breakdown factors of the materials during the surrogate-assisted calibration (inverse problem).

The solution parameters' values obtained from this procedure are also presented Table 8.1. Likewise, the performance of the M/L model (polynomial surrogate) with the approximated solution for σ, CA_{xx} and CB_{xx} using the above-mentioned calibrating resources and the surrogates is demonstrated in Figure 8.12.

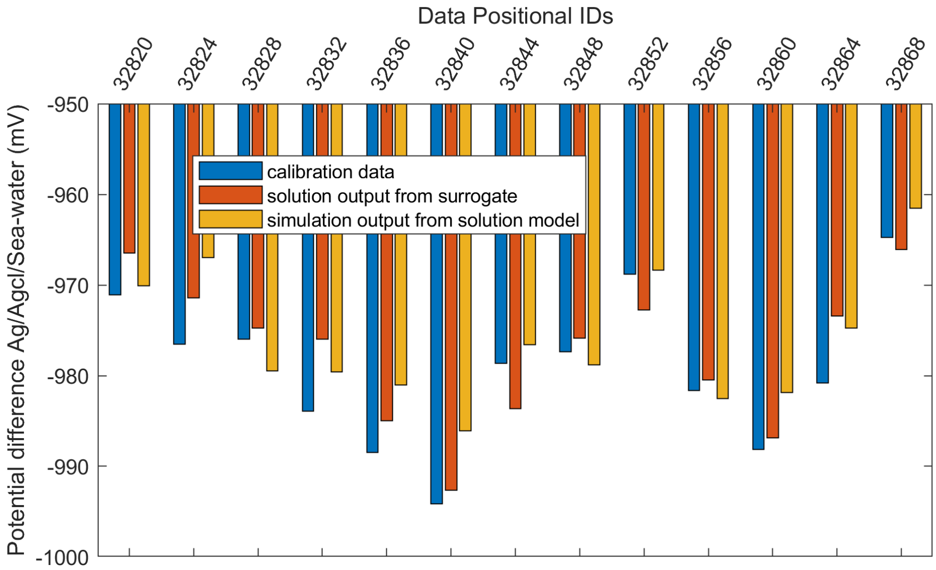


Figure 8.12: Comparison of data from simulation and surrogate with input solution parameters to the calibration data

The prediction of anode states for a CP system, which previously relied on physics-based simulation, now appears to be directly achievable through the use of historical data and the analytics. This approach is different to the previously discussed approach, which also used historical data to adapt the model even under parametric variability, but still relied on the full-order model for predicting anode-related status. In contrast, this case study demonstrates a situation where anode-related data can be predicted using past data (e.g., from time t_0 to t_n) and analytics. However, calibration data from the real system for time $t_{n+1} > t_n$ is still needed.

The results seen in CP modelling shows simulation and real-world data collected over time got the potential to reduce the dependence on the simulation model for predictive tasks. In the next subsection, the broader concept of machine learning will be explored for its use as a substitute for the physics-based model in predicting the state of CP systems.

8.5.4 Supervised Learning on Prediction

The calibrated simulation model is usually used to predict the anode depletion rate in the case of the SACP model. The prediction about the anode status is made by integrating the response data over time which are obtained from the simulation. This means the ultimate anode states are the function of the parameters involved. On the other hand, multiple data sets are generated over time either during Design of Experiments (DOE) for calibration tasks or while performing predictive simulations. Now the motive is to predict the anode depletion utilising the data-set.

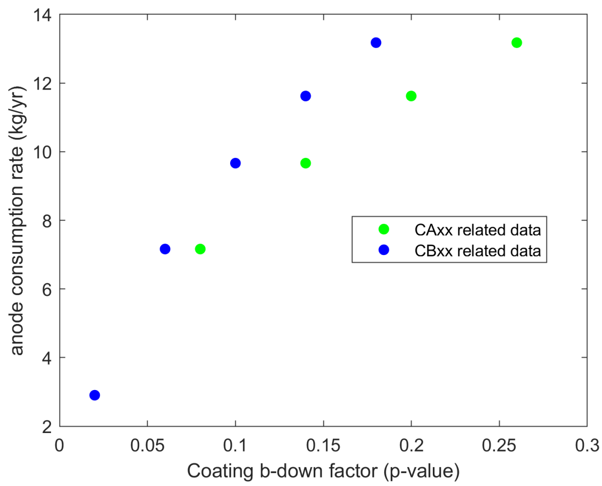


Figure 8.13: sensitivity analysis for pvalue vs anode consumption rate

For this, at the first step, the correlation between the polarisation behaviour (break-down factors) and anode consumption rate is studied (for one or more anodes) by fixing the other parameters. This analysis is done by assuming the uniformity of the change in the breakdown factor. Likewise, it is assumed the system is not encountering noise or some localised issue that could invite extreme abnormality in the anode consumption rate.

In the given condition, the analysis (Figure 8.13), indicates that the relation can can be captured using a 2^{nd}-order polynomial fit model. A M/L model accurately capturing the relation would be preferable to running simulations for prediction. To build the model, simulation snapshots were used from time t_0, for the two input parameters considered (CA_0 and CB_0), following the approach outlined in Chapter 7.

Upon analysis, it was found that the performance of the model was within an acceptable threshold ($NMSE \approx 0.005$) set for the anode consumption rate output. Figure 8.14 provides a representative example case, with input parameters of $CA_0 = 0.02$ and $CB_0 = 0.02$.

The findings suggest for the utilisation of the generated past data with the available M/L model building algorithms to make anode-data related prediction for the CP system, reducing the prediction time. However, the polynomial-fit and other neural network M/L

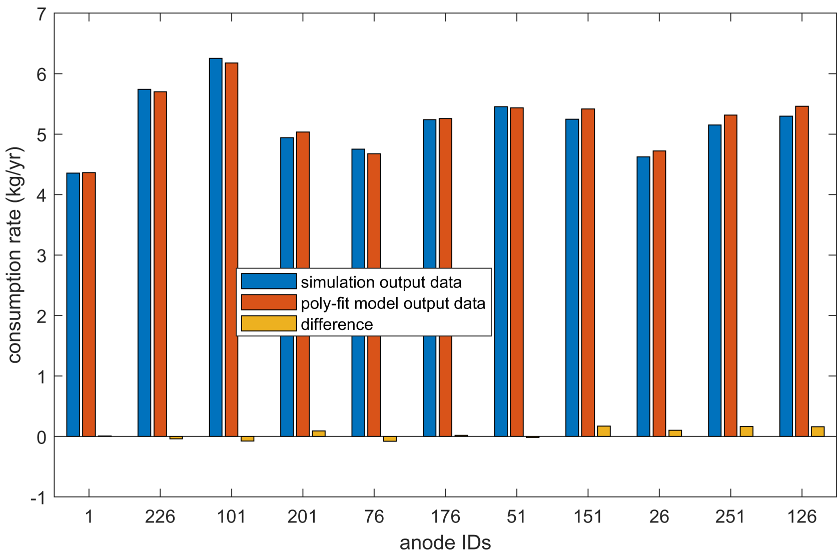


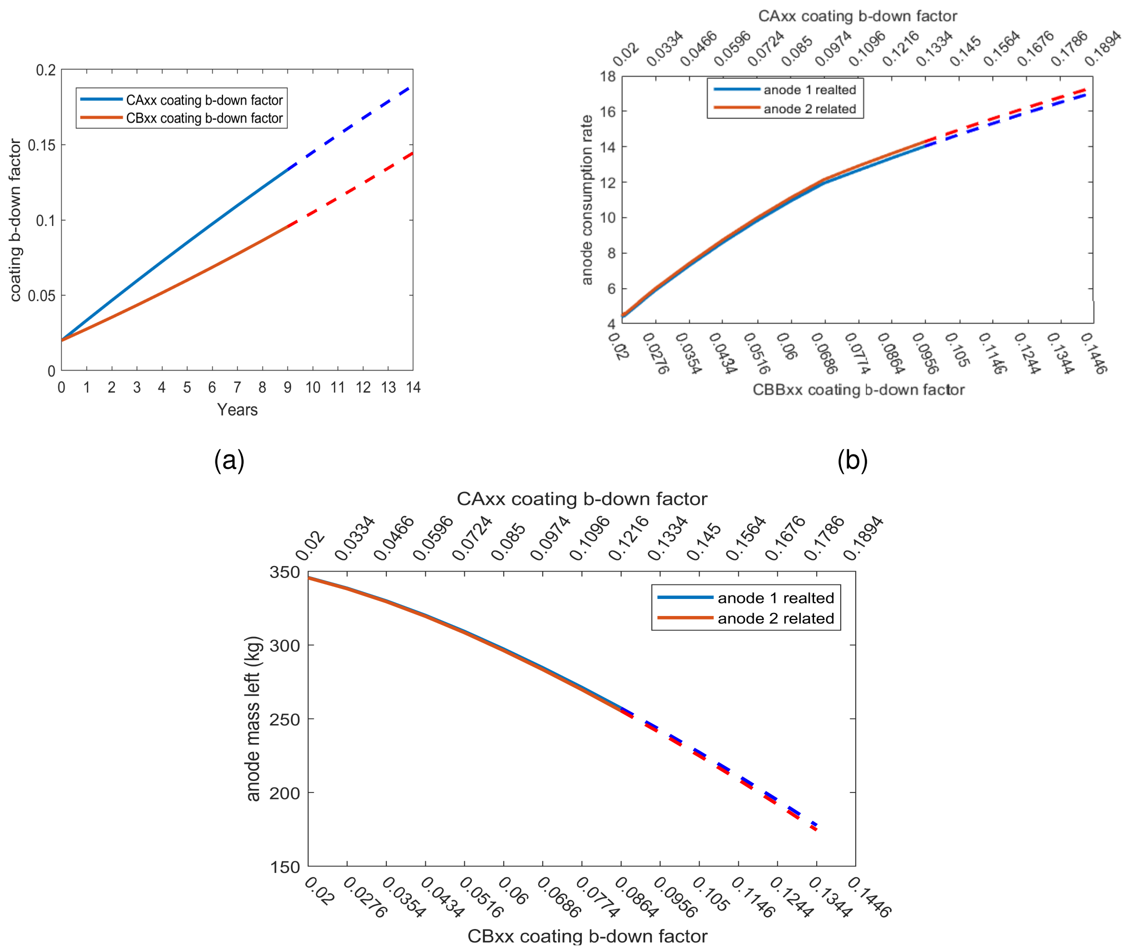
Figure 8.14: Anode consumption rate prediction comparison between simulation and poly-fit model output

models can only assist in the predictive role once the abundant related virtual and physical data are available to train the model.

When the trend of parameters can be captured for a certain time ahead, the applicability of the *Type 2* model (Figure 8.8) is facilitated. Once future parameter values are achieved (Figure 8.15a), the rate of anode consumption or anode mass left can be predicted in a few seconds with the machine learning model that was built. Figures 8.15b and 8.15c represent the anode-related performance prediction for the CP system with the *Type 2* M/L model (Figure 8.8). However, in order to utilise the *Type 2* M/L model, the condition discussed in Section 8.4.3 must be met, which involves ensuring the reliability of the M/L model for a certain time period before the influence of other neglected factors becomes significant.

Furthermore, the *Type 1* M/L model (Figure 8.8) can also be achieved using supervised and unsupervised learning M/L algorithms. However, to establish the credibility of this model, a wider range of data is required, including maintenance-related data and other complex scenarios.

This result has demonstrated the benefits of the supervised learning algorithm to implement the M/L model for CP system, so that the computation time for predictive model can be significantly reduced. This has opened the scope of M/L in reducing the dependency upon the complex physics-based simulators in CP Digital Twinning.



(a) (b)

(c)

Figure 8.15: Machine Learning in the role of making prediction for CP system performance (dotted line representing future prediction)

8.6 Conclusion

In this chapter, the focus was on the difficulties of maintaining physics-based DT in an online setting, i.e., during the operational phase. One major challenge is the emergence of parameter-related uncertainties that can exhibit spatial variability and complex patterns, therefore requiring more intricate adaptation methods beyond the scope of previous calibration techniques.

The modularisation and ROM-assisted re-calibration (adaptation) of the model is suggested to address the issue of variability in the input parameter values. The benefit provided by the ROM over other methods is its ability to replicate the physics to some extent, while modularisation ensures the usability of the historical data to assess the parameter(s) separately among the modules. Moreover, the module-based approach facilitates an expressive framework(s) required for rapid model adaptation and provides significant computational time reduction for larger complex systems. Thus, the approach will contribute to achieving the crucial feature of the digital twin i.e., less time-consuming model adaptation.

Likewise, the next section discussed replacing the physics-based model with the M/L model eventually with time. The facilitating features available within the DT paradigm make this possible even though physics-based models are essential in many situations at the beginning. The M/L model as a substitute for the physics-based not only provides an online prediction (with computation time in seconds) but also avoids the complexity in physics-based model's calibration arising with time. A case study to demonstrate the role of supervised M/L in trend capturing for the prognosis of the working status of the CP system is also presented. This forms the basis for M/L models offering to replace the physics-based model with time i.e., after capturing the possible trends and scenario of the assets.

While this experimentation is limited to the use of supervised machine learning, it is crucial to explore the potential of other algorithms offering promise in more complicated modelling circumstances. This will be significant for example in CP modelling when more data about other similar assets are available but with different patterns of anode depletion.

9 Thesis Summary, Conclusion and Future Work

This chapter provides a summary of the research activities undertaken and presents the conclusion of the study. Furthermore, it outlines some potential future avenues of this research project.

9.1 Research Summary

This research investigated the issues with the simulation model required for prognosis during Structural Health Monitoring (SHM).

The SHM process demands a high-performance model that accurately reflects the structure in real-time for the purpose of risk assessment, optimising system usage, design control, etc. Recent advancements in simulation modelling have made it possible to use 3D CAD design software and commercial simulators to simulate structural degradation caused fatigue, corrosion, and/or cracking. This facilitate for simulator-based parametric model to make real-world predictions after calibration. However, the model often deviates from the actual behaviour of the asset due to parameter-related uncertainty and other complexities that emerge over time. Consequently, repetitive calibration or adaptation of the model is necessary, which not only increases costs but also poses a risk to the structure if the model fails to perform. Despite significant progress in physics-based modelling, the current advancements have not been sufficient to enable the practical and continuous implementation of prognostic models for real structures.

The challenges involved in the calibration and/or real-time adaptation of the SHM-related model, are outlined in Chapter 2 which includes:

1. Unstructured data obtained from surveys is required for calibration/adaptation.
2. Calibration and adaptation requiring manual involvement.
3. Lack of benchmark of data for model calibration and adaptation.
4. Lack of effective online approach or mechanism for updating parameters..
5. Material-related parameters changing with variability and complexity in the operational phase.

The concept of Digital Twin, a novel, promising and holistic approach in the modelling field, was then explored to discover its potential benefits in addressing model calibration and adaptation-related challenges. The findings revealed that while DT holds promise in addressing these challenges, it still lacks a practical benchmark for implementation. Thus, the research aimed to leverage the advantages of DT to address the research challenges, and to propose suitable approaches (frameworks) within the DT domain. A research plan was developed accordingly which is presented in Chapter 3 Section 3.2. Also, the research areas were categorised based on the research challenges and the scopes they hold.

The Industry-as-Academia methodology was followed to address the research challenges, categorically for each of the research areas. The research challenges, the corresponding scopes they hold and the reached solutions to the research challenges are outlined in Figure 9.1.

The key findings from the research to which resulted in the outcomes highlighted in Figure 9.1 are summarised and concluded as follows:

1. Ontology and N-L-P assisted automated data-extraction

Recent advancements in sensors and data-transfer technologies, support the idea of DT's online mirroring concept. However, still, in some SHM-related domains, data are collected from the survey and lie in unstructured report formats. To deal with this issue and make the data available to the DT in near-real time, an ontology and Natural-Language-Processing assisted framework was proposed.

This framework leverages the ontology-based DT concept which provides to built information ontologies from the DT and utilises them for automated N-L-P-based data exploration and extraction. The solution for automated data extraction utilising this framework was presented with an example case in Section 4.2 of Chapter 4. The proposed framework for data acquisition and arrangement, is also equally important for the practical implementation of the first feature of DT, i.e., digital mirroring (Barni et al. 2018, Barricelli et al. 2019).

2. Scientific software-simulator integrated platform for automated calibration

When the real-world data are available via data-mirroring, the uncertainty of the model's parameters can be addressed through the use of multiple design optimisation algorithms and tools (Huang et al. 2010, Peč et al. 2019, Benaouali and Kachel 2019). However, to fully take advantage of these tools, it is necessary to automate the calibration process.

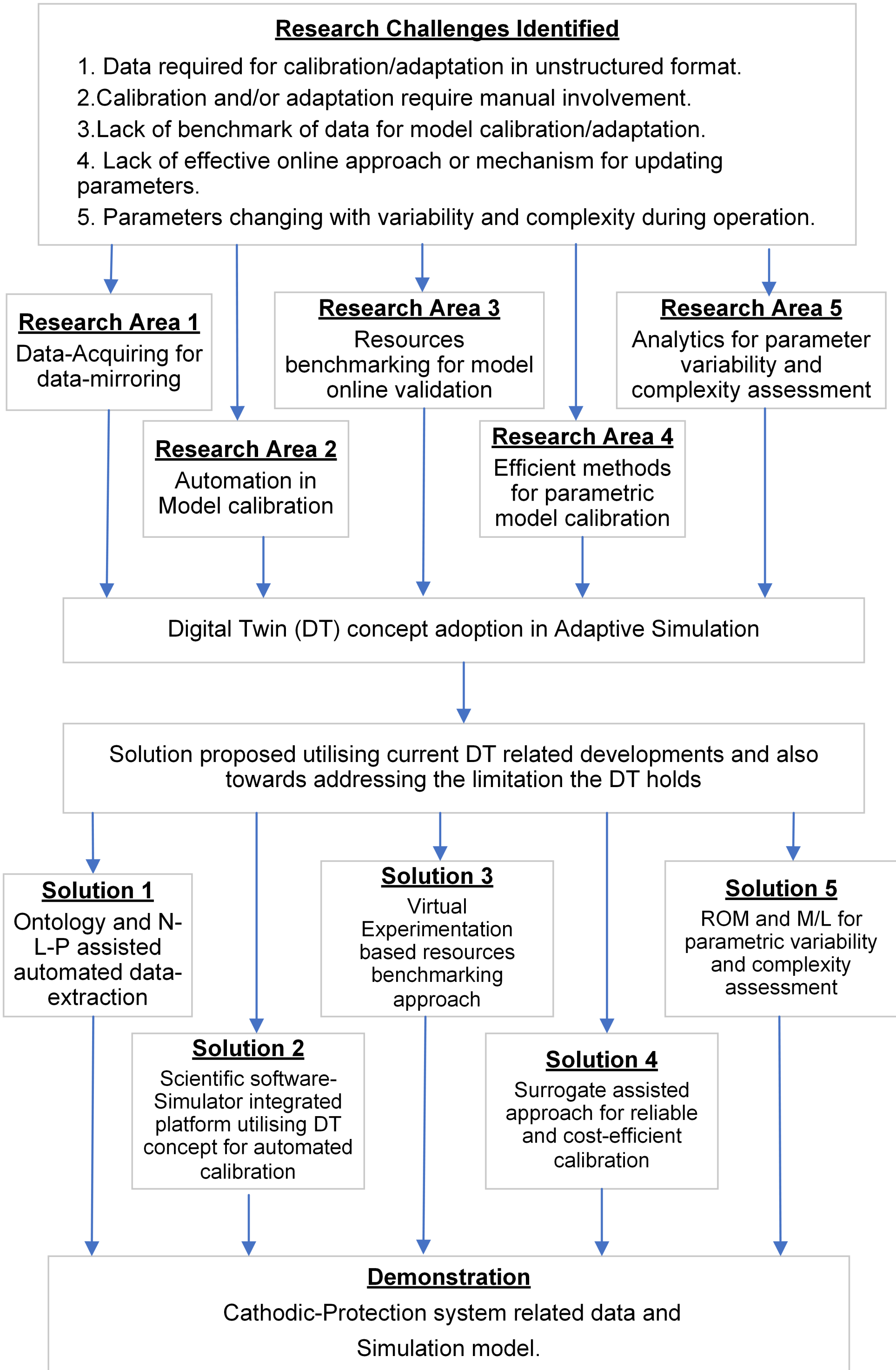


Figure 9.1: The outline of the research from research challenges, the scopes they hold to the proposed solutions.

Furthermore, there was also a requirement of a framework that provides a common platform for creating a predictive DT from physical models (Aivaliotis et al. 2019b, Barricelli et al. 2019).

A Design of Experiment platform was proposed to enable the practical use of real-world data, physical model and design optimisation tools for DTs. This platform, which includes analytical support, reduces the need for manual involvement during the calibration and adaptation process. The development and advantages of this platform were demonstrated in Chapter 5 using a MATLAB-BEASY integrated platform to represent a Cathodic-Protection system. Furthermore, this platform provides the groundwork for automating the creation of DT from existing simulator(s).

3. Virtual Experimentation based resources benchmarking approach

Despite the availability of real-world data and automation tools for calibration, challenges still exist due to a lack of benchmark data needed for accurate model's calibration. Inadequate data can lead to incorrect solutions, while an over-reliance on data may increase costs (Fabrizio and Monetti 2015, Kang et al. 2021).

To address this challenge, an approach was proposed for benchmarking the quantity, quality, and variability of validation data required for model calibration/adaptation. This approach, which is based on virtual experimentation and presented in Chapter 6, aims to optimise data collection and minimise the costs involved in the process, particularly when data is obtained from costly surveys (or sensors). The approach assists in standardising the necessary data before moving on to the online phase of parametric model tailoring. The benefit of the approach was demonstrated for the cathodic-protection model-related data standardisation.

4. Surrogate-assisted approach for reliable and cost-efficient calibration

Though calibration is achieved with automation, other challenges still remain, such as time-consuming simulation runs during calibration and the risk of getting stuck at local minima. To address this research challenge, Chapter 7 proposed a cost-effective but reliable parameter search technique.

The proposed surrogate-assisted calibration approach utilises the approximation model (i.e., surrogate) for prediction during the parameter tuning process, rather than relying upon a full-order model. The case study also demonstrated that using an offline-built surrogate model during parameter estimation in BEM based CP modelling reduces the online time required from hours when using a physics-based simulation to just a few minutes.

The surrogate model offers additional benefits such as avoiding local minima during calibration with optimisation, solution space approximation and providing problem understanding over the parameter and solution spaces. Considering the analytical support available with the hybrid DT concept (Chinesta et al. 2020, Azangoo et al. 2022) and the benefits surrogate offers, this research, therefore, advocates for the use of surrogate modelling as a crucial feature within the structural DT ecosystem.

5. ROM and M/L for parametric variability and complexity assessment

Following the beneficial results observed in terms of model calibration and DT enabling processes, the research then turned its attention to the challenges that arise when implementing physics-based DT during their operational phase. These challenges include spatially varying parameter-related uncertainty and the complexity posed by the dynamic pattern of those material-related parameters over time.

Modularisation and ROM-assisted model adaptation were suggested towards addressing the issue of variability in the input parameter values arising during the operational phase. ROM got the benefit over the polynomial surrogate that it retains some of the physics of the dynamic phenomenon. Furthermore, the proposed approach facilitates an expressive framework(s) required for rapid model adaptation and also provides significant computational time reduction for larger complex systems.

Additionally, the role of M/L algorithms were discussed to provide a substitute for the physics-based model eventually with time. This is because the adaptation of the simulation model often gets more complex with time, making it difficult to obtain accurate predictions. M/L models including above mentioned surrogate built-in offline phase with sufficient accuracy for the parameter values within the DOE can be used for online prediction requirements. This concept which was presented in Chapter 8, Section 8.4 is also about the utilisation of the DT concept as it provides a wealth of physical and virtual data, that can be used to analyse the root causes and trends of the asset's behaviour over time.

The implementation of the approaches discussed in this research for a model in a comprehensive manner seems to be challenging as dealt with multiple aspects of modelling. However, following the DT concept and the features it provides for such comprehensive modelling, makes it achievable. The outcomes have demonstrated the feasibility of such a DT applications (Sapkota et al. 2021a 2022a b) in situations where the frameworks for dealing with parameter related uncertainties and complexities under DT were missing (Barni et al. 2018, Aivaliotis et al. 2019a, Broo et al. 2022). In addition, this research presents a solution towards using predictions of cathodic protection performance to estimate the lifetime extension of the offshore structure.

The outcomes achieved in addressing the research challenges, still have opportunities for future work, which will be recommended in the next section. The recommendations will be based upon either the scope to test the usability of the proposed approaches on a variety of problems or the limitations (time, goals) of this research.

9.2 Future Work Recommendations

The potential avenues beyond the promise offered by this research outcomes are presented below as future scopes:

1. Data collection and processing in DT

The Digital Twin (DT) concept has gained significant momentum in recent years to provide real-world data mirroring in several fields that require Structural Health Monitoring (SHM). Despite this, past data lying in unstructured formats when acquired still holds valuable benefits for DT's prognostic tool. This research developed a tool which was presented in Chapter 4 that utilises ontologies from the DT for N-L-P-based data exploration and extraction.

It is recommended to conduct usability testing of the tool with supervision across a range of data as a future task. With further implementation, the tool's robustness should improve, making it applicable to a wide range of cases with less supervision. Additionally, implementing such a tool opens up the possibility of training it with supervised/unsupervised M/L algorithms over time, as additional data formats become available. This possibility should be explored in line with the trend of using M/L techniques for extracting information from structured and unstructured data, including text and image mining (Barni et al. 2018, Sun, Shang, Xia, Bhowmick and Nagarajaiah 2020).

2. Expansion of Scientific software-simulator integrated platform

The demonstration of the Design of Experiment platform proposed in Chapter 5 for online parameter validation and calibration, is currently limited to MATLAB and BEASY tools. As a future work, it is suggested to test the concept of the integrated platform with other software and simulator(s) on having an automated platform for self-calibration. Also, the platform has the potential to be developed and expanded to achieve a more comprehensive Digital Twin as anticipated by researchers (Barricelli et al. 2019, Aivaliotis et al. 2019b). To accomplish this, additional modelling-providing simulators (or algorithms) and analytical capabilities (tools or algorithms) required for co-simulations can be incorporated into the platform.

3. Extension of the surrogate-assisted calibration approach

In this research, only a 2^{nd} order polynomial-fit surrogate and the CCD method for sampling were used to demonstrate the surrogate-assisted calibration approach. However, there exist other data-driven surrogate modelling and sampling methods that are also being applied to SHM-related simulation alternatives (Yondo et al. 2018, Westermann and Evins 2019, García-Macías et al. 2021). This indicates a vast potential for further investigation into the applicability of these methods in the CP domain and beyond within the DT concept. Thus, the research recommends exploring the extension of surrogate-assisted calibration as another avenue for future work.

4. M/L algorithms for dealing with uncertainties and complexities during the operation

Another potential future scope is to expand the use of Machine-Learning (M/L) for dealing with parameter-related uncertainties and complexities during the model-operation period. The experimentation and results presented in Chapter 8 demonstrated the benefits of utilising a supervised learning algorithm of M/L for assessing the complexity of the Cathodic-Protection system. This has already opened up the possibility of reducing dependence on complex physics-based simulators in CP Digital Twinning.

While M/L-based SHM concept has been introduced, current techniques are not yet sufficient for tracking damage with respect to the type of structural components, materials, locations, and other environmental conditions (Azimi et al. 2020, Omar et al. 2022, Willard et al. 2022). To address this, it is worth exploring the potential of using reinforcement learning or deep learning approaches to track such complex scenarios over time in an asset. For example, in the context of CP modelling, where anode size itself plays the role of the input variable(s) and an output variable to be predicted, a deep learning approach can be explored to deal with this complexity.

5. Wider potential application

To implement practical solutions in existing CP system-related prognostic, the model must be tailored to the real specific environmental and boundary conditions. For this real-world data are anticipated. Gathering individualised information for each measurement date and location, along with detailed and continuous measurements, would result in significant benefits.

Apart from the real-world application, the proposed solutions can be further tested and evaluated in more complex models. While the research outcomes presented in Chapters 5, 6, and 7 have demonstrated their effectiveness using two parametric variables cases, the applicability of these approaches in higher parameter situations can be investigated and their efficacy and efficiency can be studied.

Furthermore, to achieve real-world practicality test, the solutions in more complex scenarios such as time-step models can be utilised. Although Chapter 8 demonstrated the approaches for a few time-step cases, further evaluation can be conducted using time-dependent parametric models, which are generally more complex than steady-state one. By applying the proposed solutions to time-dependent parametric models, their potential can be assessed in a broader context.

Additionally, exploring to adopt probabilistic parameter values instead of assuming determinism is another scope of this research. The incorporation of probabilistic values can aid in the propagation of uncertainty throughout the model, resulting in a range of possible outcomes rather than a single estimate (Honarmandi and Arróyave 2020, Chai et al. 2022). This is particularly useful for robust or reliability-based design, where accurately quantified uncertainties in the predicted outcomes of a design choice are crucial (Arróyave and McDowell 2019). However, it is important to note that this approach can be computationally expensive, particularly when using a physics-based model.

9.3 Concluding Remarks

To conclude, this research study proposed various approaches to enhance the calibration and adaptation procedures of a simulator-based model to create a digital twin of a physical asset. These solutions are particularly useful in situations where standard frameworks for digital twin implementation are lacking. Therefore, the research outcomes, including the suggested digital twin architecture (shown in Figure 2.5), can also be understood as a contribution on standardisation of SHM-related digital twins. Additionally, while existing SHM research mainly revolves around elasticity models this research focused on the impact of material depletion due to corrosion on structural integrity.

The demonstration of the outcomes, particularly in corrosion-related SHM that uses BEM models, also shows their applicability from an industrial perspective. This is particularly noteworthy as most SHM-related research focuses on FEM models. Additionally, the research outcomes offer the potential to equip SHM engineers with a predictive tool for real-time and future analysis in real-world applications. The outcomes being promising in the use of digital twin concepts for SHM-related prognosis, also still has opportunities for future work, which are outlined above.

Bibliography

Abbas, M. and Shafiee, M. (2018), ‘Structural health monitoring (shm) and determination of surface defects in large metallic structures using ultrasonic guided waves’, *Sensors* **18**(11), 3958.

Abdalla, M. and Ibrahim, H. (2021), ‘A physical approach to simulate the corrosion of ceramic-coated magnesium implants’, *Applied Sciences* **11**(15), 6724.

Abdelmegid, M. A., González, V. A., O’Sullivan, M., Walker, C. G., Poshdar, M. and Ying, F. (2020), ‘The roles of conceptual modelling in improving construction simulation studies: A comprehensive review’, *Advanced Engineering Informatics* **46**.

Adey, R. A. E. (2005), *Modelling of Cathodic Protection Systems (Vol. 12)*, WIT press.

Adey, R. A., Peratta, C. and Baynham, J. M. (2012), Using modeling to interpret and expand cp survey data, *in* ‘CORROSION 2012’, OnePetro.

Adey, R., Brebbia, C. and Niku, S. (1990), Applications of boundary elements in corrosion engineering, *in* ‘Electrical Engineering Applications’, Springer, pp. 34–64.

Adey, R., Peratta, C. and Baynham, J. (2020), ‘Corrosion data management using 3d visualisation and a digital twin’, *CORROSION 2020* .

Agami Reddy, T. (2006), ‘Literature review on calibration of building energy simulation programs: Uses, problems, procedure, uncertainty, and tools’, *ASHRAE Transactions* **112**(4844), 226–240.

Ahmad, S. (2003), ‘Reinforcement corrosion in concrete structures, its monitoring and service life prediction—-a review’, *Cement and concrete composites* **25**(4-5), 459–471.

Aivaliotis, P., Georgoulias, K., Arkouli, Z. and Makris, S. (2019a), ‘Methodology for enabling digital twin using advanced physics-based using physics-based modelling in predictive maintenance’, *Procedia CIRP* **81**, 417–422.
URL: *https://doi.org/10.1016/j.procir.2019.03.072*

Aivaliotis, P., Georgoulias, K. and Chryssolouris, G. (2019b), 'The use of digital twin for predictive maintenance in manufacturing', *International Journal of Computer Integrated Manufacturing* **32**(11), 1067–1080.

Aktan, A. and Brownjohn, J. (2013), 'Structural identification: opportunities and challenges', *Journal of Structural Engineering* **139**(10), 1639–1647.

Al-Otaibi, M. (2010), The application of BEASY software to simulate cathodic protection of pipelines and storage tanks, PhD thesis, University of British Columbia.

Alam, F. M., McNaught, K. R. and Ringrose, T. J. (2004), 'A comparison of experimental designs in the development of a neural network simulation metamodel', *Simulation Modelling Practice and Theory* **12**(7-8), 559–578.

Alaswad, S. and Xiang, Y. (2017), 'A review on condition-based maintenance optimization models for stochastically deteriorating system', *Reliability engineering & system safety* **157**, 54–63.

Aldrin, J. C., Medina, E. A., Lindgren, E. A., Buynak, C. F. and Knopp, J. S. (2011), Case studies for model-assisted probabilistic reliability assessment for structural health monitoring systems, *in* 'AIP Conference Proceedings', Vol. 1335, American Institute of Physics, pp. 1589–1596.

Aliabadi, M. H. (2002), *The boundary element method, volume 2: applications in solids and structures (Vol. 2)*, Wiley, John & Sons.

Alizadeh, R., Allen, J. K. and Mistree, F. (2020), 'Managing computational complexity using surrogate models: a critical review', *Research in Engineering Design* **31**(3), 275–298.

Alnowaiser, K. K. and Ahmed, M. A. (2023), 'Digital twin: Current research trends and future directions', *Arabian Journal for Science and Engineering* **48**(2), 1075–1095.

Amaya, K., Yoneya, N. and Onishi, Y. (2014), 'Obtaining corrosion rates by bayesian estimation: numerical simulation coupled with data', *The Electrochemical Society Interface* **23**(4), 53.

An, D., Choi, J.-H. and Kim, N. H. (2013), 'Prognostics 101: A tutorial for particle filter-based prognostics algorithm using matlab', *Reliability Engineering & System Safety* **115**, 161–169.

An, D., Kim, N. H. and Choi, J.-H. (2015), ‘Practical options for selecting data-driven or physics-based prognostics algorithms with reviews’, *Reliability Engineering & System Safety* **133**, 223–236.

An, Y., Chatzi, E., Sim, S.-H., Laflamme, S., Blachowski, B. and Ou, J. (2019), ‘Recent progress and future trends on damage identification methods for bridge structures’, *Structural Control and Health Monitoring* **26**(10), e2416.

Angjeliu, G., Coronelli, D. and Cardani, G. (2020), ‘Development of the simulation model for digital twin applications in historical masonry buildings: The integration between numerical and experimental reality’, *Computers & Structures* **238**, 106282.

Angst, U. M. (2019), ‘A critical review of the science and engineering of cathodic protection of steel in soil and concrete’, *Corrosion* **75**(12), 1420–1433.

Antoulas, A. C. (2005), *Approximation of large-scale dynamical systems*, SIAM.

Arlinghaus, S. (1994), *Practical handbook of curve fitting*, CRC press.

Arróyave, R. and McDowell, D. L. (2019), ‘Systems approaches to materials design: past, present, and future’, *Annual Review of Materials Research* **49**, 103–126.

Ashino, T. (2010), ‘Materials ontology: An infrastructure for exchanging materials information and knowledge’, *Data Science Journal* **9**, 54–61.

ASME (2009), *Standard for Verification and Validation in Computational Fluid Dynamics and Heat Transfer: An American National Standard*, American Society of Mechanical Engineers.

Azangoo, M., Sorsamäki, L., Sierla, S. A., Mätäsniemi, T., Rantala, M., Rainio, K. and Vyatkin, V. (2022), ‘A methodology for generating a digital twin for process industry: a case study of a fiber processing pilot plant’, *IEEE Access* **10**, 58787–58810.

Azimi, M., Eslamlou, A. D. and Pekcan, G. (2020), ‘Data-driven structural health monitoring and damage detection through deep learning: State-of-the-art review’, *Sensors* **20**(10), 2778.

Balageas, D. (2006), Introduction to Structural Health Monitoring.

Bao, Q., Zhao, G., Yu, Y., Dai, S. and Wang, W. (2022), ‘Ontology-based modeling of part digital twin oriented to assembly’, *Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture* **236**(1-2), 16–28.

Barni, A., Fontana, A., Menato, S., Sorlini, M. and Canetta, L. (2018), Exploiting the digital twin in the assessment and optimization of sustainability performances, *in* '2018 International conference on intelligent systems (IS)', IEEE, pp. 706–713.

Barricelli, B. R., Casiraghi, E. and Fogli, D. (2019), 'A survey on digital twin: Definitions, characteristics, applications, and design implications', *IEEE access* **7**, 167653–167671.

Barthorpe, R. J. and Worden, K. (2020), 'Emerging trends in optimal structural health monitoring system design: From sensor placement to system evaluation', *Journal of Sensor and Actuator Networks* **9**(3), 31.

Barton, R. R. (1992), Metamodels for simulation input-output relations, *in* 'Proceedings of the 24th conference on Winter simulation', pp. 289–299.

Barzegari, M., Mei, D., Lamaka, S. V. and Geris, L. (2021), 'Computational modeling of degradation process of biodegradable magnesium biomaterials', *Corrosion Science* **190**, 109674.

Bayraktarov, E., Ehmke, G., O'Connor, J., Burns, E. L., Nguyen, H. A., McRae, L., Possingham, H. P. and Lindenmayer, D. B. (2019), 'Do big unstructured biodiversity data mean more knowledge?', *Frontiers in Ecology and Evolution* **6**, 239.

Bazaz, S. M., Lohtander, M. and Varis, J. (2019), '5-dimensional definition for a manufacturing digital twin', *Procedia Manufacturing* **38**, 1705–1712.

Bazilevs, Y., Deng, X., Korobenko, A., Lanza di Scalea, F., Todd, M. and Taylor, S. (2015), 'Isogeometric fatigue damage prediction in large-scale composite structures driven by dynamic sensor data', *Journal of Applied Mechanics* **82**(9).

Behmanesh, I. and Moaveni, B. (2016), 'Accounting for environmental variability, modeling errors, and parameter estimation uncertainties in structural identification', *Journal of Sound and Vibration* **374**, 92–110.

Belostotsky, A., Akimov, P., Negrozov, O., Petryashev, N., Petryashev, S., Sherbina, S., Kalichava, D. and Kaytukov, T. (2018), 'Adaptive finite-element models in structural health monitoring systems', *Magazine of Civil Engineering* (2 (78)), 169–178.

Benaouali, A. and Kachel, S. (2019), 'Multidisciplinary design optimization of aircraft wing using commercial software integration', *Aerospace Science and Technology* **92**, 766–776.

Benner, P., Gugercin, S. and Willcox, K. (2015), 'A survey of projection-based model reduction methods for parametric dynamical systems', *SIAM review* **57**(4), 483–531.

Bezerra, M. A., Santelli, R. E., Oliveira, E. P., Villar, L. S. and Escaleira, L. A. (2008), 'Response surface methodology (rsm) as a tool for optimization in analytical chemistry', *Talanta* **76**(5), 965–977.

Bi, S., Prabhu, S., Cogan, S. and Atamturktur, S. (2017), 'Uncertainty quantification metrics with varying statistical information in model calibration and validation', *AIAA journal* **55**(10), 3570–3583.

Biau, G. and Scornet, E. (2016), 'A random forest guided tour', *Test* **25**, 197–227.

Bird, S. (2006), Nltk: the natural language toolkit, *in* 'Proceedings of the COLING/ACL 2006 Interactive Presentation Sessions', pp. 69–72.

Blum, W. and Ferri, R. B. (2009), 'Mathematical modelling: Can it be taught and learnt?', *Journal of mathematical modelling and application* **1**(1), 45–58.

Booyse, W., Wilke, D. N. and Heyns, S. (2020), 'Deep digital twins for detection, diagnostics and prognostics', *Mechanical Systems and Signal Processing* **140**, 106612.

Boschert, S. and Rosen, R. (2016), Digital twin—the simulation aspect, *in* 'Mechatronic futures', Springer, pp. 59–74.

Broo, D. G., Bravo-Haro, M. and Schooling, J. (2022), 'Design and implementation of a smart infrastructure digital twin', *Automation in Construction* **136**, 104171.

Brownjohn, J. M. (2007), 'Structural health monitoring of civil infrastructure', *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* **365**(1851), 589–622.

Brynjarsdóttir, J. and O'Hagan, A. (2014), 'Learning about physical parameters: The importance of model discrepancy', *Inverse problems* **30**(11), 114007.

Buethe, I., Eckstein, B. and Fritzen, C.-P. (2014), 'Model-based detection of sensor faults under changing temperature conditions', *Structural Health Monitoring* **13**(2), 109–119.

Buhmann, M. D. (2000), 'Radial basis functions', *Acta numerica* **9**, 1–38.

Burés, J. and Larrosa, I. (2023), 'Organic reaction mechanism classification using machine learning', *Nature* **613**(7945), 689–695.

Cao, Z., Fei, Q., Jiang, D., Zhang, D., Jin, H. and Zhu, R. (2020), 'Dynamic sensitivity-based finite element model updating for nonlinear structures using time-domain responses', *International Journal of Mechanical Sciences* **184**, 105788.

Catbas, N., Gokce, H. B. and Frangopol, D. M. (2013), 'Predictive analysis by incorporating uncertainty through a family of models calibrated with structural health-monitoring data', *Journal of Engineering Mechanics* **139**(6), 712–723.

Cavazzuti, M. and Cavazzuti, M. (2013), 'Design of experiments', *Optimization methods: from theory to design scientific and technological aspects in mechanics* pp. 13–42.

Chai, X., Rozsas, A., Slobbe, A. and Teixeira, A. (2022), 'Probabilistic parameter estimation and reliability assessment of a simulated sheet pile wall system', *Computers and Geotechnics* **142**, 104567.

Chakraborty, S. and Adhikari, S. (2021b), 'Machine learning based digital twin for dynamical systems with multiple time-scales', *Computers & Structures* **243**, 106410.

Chakraborty, S., Adhikari, S. and Ganguli, R. (2021a), 'The role of surrogate models in the development of digital twins of dynamic systems', *Applied Mathematical Modelling* **90**, 662–681.

Chan, S., Reddy, V., Myers, B., Thibodeaux, Q., Brownstone, N. and Liao, W. (2020), 'Machine learning in dermatology: current applications, opportunities, and limitations', *Dermatology and therapy* **10**, 365–386.

Chao, M. A., Kulkarni, C., Goebel, K. and Fink, O. (2022), 'Fusing physics-based and deep learning models for prognostics', *Reliability Engineering & System Safety* **217**, 107961.

Charbuty, B. and Abdulazeez, A. (2021), 'Classification based on decision tree algorithm for machine learning', *Journal of Applied Science and Technology Trends* **2**(01), 20–28.

Chatterjee, A. (2000), 'An introduction to the proper orthogonal decomposition', *Current science* pp. 808–817.

Chen, Z., Koleva, D. and van Breugel, K. (2017), ‘A review on stray current-induced steel corrosion in infrastructure’, *Corrosion Reviews* **35**(6), 397–423.

Chiachío, M., Megía, M., Chiachío, J., Fernandez, J. and Jalón, M. L. (2022), ‘Structural digital twin framework: Formulation and technology integration’, *Automation in Construction* **140**, 104333.

Chinesta, F., Cueto, E., Abisset-Chavanne, E., Duval, J. L. and Khaldi, F. E. (2020), ‘Virtual, digital and hybrid twins: a new paradigm in data-based engineering and engineered data’, *Archives of computational methods in engineering* **27**(1), 105–134.

Christopher Frey, H. and Patil, S. R. (2002), ‘Identification and review of sensitivity analysis methods’, *Risk analysis* **22**(3), 553–578.

Coleman, T., Branch, M. A. and Grace, A. (1999), ‘Optimization toolbox’, *For use with MATLAB. User’s guide for MATLAB* **5**.

Collobert, R., Weston, J., Bottou, L., Karlen, M., Kavukcuoglu, K. and Kuksa, P. (2011), ‘Natural language processing (almost) from scratch’, *Journal of machine learning research* **12**(ARTICLE), 2493–2537.

Cremona, C. and Santos, J. (2018), ‘Structural health monitoring as a big-data problem’, *Structural Engineering International* **28**(3), 243–254.

Cross, E. (2012), On structural health monitoring in changing environmental and operational conditions, PhD thesis, University of Sheffield.

Cruz, C. D. (2016), ‘Genes software-extended and integrated with the r, matlab and selegen’, *Acta Scientiarum. Agronomy* **38**, 547–552.

Cui, G., Li, Z., Yang, C. and Wei, X. (2015), ‘Study on the interference corrosion of cathodic protection system’, *Corrosion Reviews* **33**(5), 233–247.

Daigle, M. J. and Goebel, K. (2012), ‘Model-based prognostics with concurrent damage progression processes’, *IEEE Transactions on Systems, man, and cybernetics: systems* **43**(3), 535–546.

Danson, D., Brebbia, C. and Adey, R. (1982), The beasy system, *in* ‘CAD82’, Elsevier, pp. 482–493.

Davis, S. E., Cremaschi, S. and Eden, M. R. (2018), Efficient surrogate model development: impact of sample size and underlying model dimensions, *in* 'Computer Aided Chemical Engineering', Vol. 44, Elsevier, pp. 979–984.

DeGiorgi, V. (1993), Corrosion basics and computer modeling, *in* 'Industrial applications of the boundary element method', pp. 47–79.

DeGiorgi, V., Lucas, K., Thomas, E. and Shimko, M. (1992), Boundary element evaluation of iccp systems under simulated service conditions, *in* 'Boundary Element Technology VII', Springer, pp. 405–422.

Demo, N., Tezzele, M. and Rozza, G. (2018), 'Ezyrb: Easy reduced basis method', *Journal of Open Source Software* **3**(24), 661.

Doubrawa, P., Debnath, M., Moriarty, P. J., Branlard, E., Herges, T., Maniaci, D. and Naughton, B. (2019), Benchmarks for model validation based on lidar wake measurements, *in* 'Journal of Physics: Conference Series', Vol. 1256, IOP Publishing, p. 012024.

Ehret, U., van Pruijssen, R., Bortoli, M., Loritz, R., Azmi, E. and Zehe, E. (2020), 'Adaptive clustering: reducing the computational costs of distributed (hydrological) modelling by exploiting time-variable similarity among model elements', *Hydrology and Earth System Sciences* **24**(9), 4389–4411.

Eldred, M. and Dunlavy, D. (2006), Formulations for surrogate-based optimization with data fit, multifidelity, and reduced-order models, *in* '11th AIAA/ISSMO multidisciplinary analysis and optimization conference', p. 7117.

Erkoyuncu, J. A., del Amo, I. F., Ariansyah, D., Bulka, D., Roy, R. et al. (2020), 'A design framework for adaptive digital twins', *CIRP annals* **69**(1), 145–148.

Ezzat, A. A., Tang, J. and Ding, Y. (2020), 'A model-based calibration approach for structural fault diagnosis using piezoelectric impedance measurements and a finite element model', *Structural Health Monitoring* **19**(6), 1839–1855.

Fabrizio, E. and Monetti, V. (2015), 'Methodologies and advancements in the calibration of building energy models', *Energies* **8**(4), 2548–2574.

Fahim, M., Sharma, V., Cao, T.-V., Canberk, B. and Duong, T. Q. (2022), 'Machine learning-based digital twin for predictive modeling in wind turbines', *IEEE Access* **10**, 14184–14194.

Farlow, S. J. (1993), *Partial differential equations for scientists and engineers*, Courier Corporation.

Farrar, C. R. and Lieven, N. A. J. (2007b), 'Damage prognosis: The future of structural health monitoring', *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* **365**(1851), 623–632.
URL: *https://doi.org/10.1098/rsta.2006.1927*

Farrar, C. R. and Worden, K. (2007a), 'An introduction to structural health monitoring', *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* **365**(1851), 303–315.

Feng, K., Ji, J., Zhang, Y., Ni, Q., Liu, Z. and Beer, M. (2023), 'Digital twin-driven intelligent assessment of gear surface degradation', *Mechanical Systems and Signal Processing* **186**, 109896.

Gabor, T., Belzner, L., Kiermeier, M., Beck, M. T. and Neitz, A. (2016), A simulation-based architecture for smart cyber-physical systems, *in* 'Proceedings - 2016 IEEE International Conference on Autonomic Computing', ICAC 2016, pp. 374–379.

Ganguli, R., Adhikari, S., Chakraborty, S. and Ganguli, M. (2023), *Digital Twin: A Dynamic System and Computing Perspective*, CRC Press.

García-Macías, E., Ierimonti, L., Venanzi, I. and Ubertini, F. (2021), 'An innovative methodology for online surrogate-based model updating of historic buildings using monitoring data', *International Journal of Architectural Heritage* **15**(1), 92–112.

Gardner, P., Bull, L., Gosliga, J., Dervilis, N. and Worden, K. (2021), 'Foundations of population-based shm, part iii: Heterogeneous populations–mapping and transfer', *Mechanical systems and signal processing* **149**, 107142.

Gardner, P., Dal Borgo, M., Ruffini, V., Hughes, A. J., Zhu, Y. and Wagg, D. J. (2020), 'Towards the development of an operational digital twin', *Vibration* **3**(3), 235–265.

Gill, P. E. and Murray, W. (1972), 'Quasi-newton methods for unconstrained optimization', *IMA Journal of Applied Mathematics* **9**(1), 91–108.

Giunta, A., Wojtkiewicz, S. and Eldred, M. (2003), Overview of modern design of experiments methods for computational simulations, *in* '41st Aerospace Sciences Meeting and Exhibit', p. 649.

Glaessgen, E. and Stargel, D. (2012), The digital twin paradigm for future nasa and us air force vehicles, *in* '53rd AIAA/ASME/ASCE/AHS/ASC structures, structural dynamics and materials conference 20th AIAA/ASME/AHS adaptive structures conference 14th AIAA', p. 1818.

Goller, B., Broggi, M., Calvi, A. and Schueller, G. (2011), 'A stochastic model updating technique for complex aerospace structures', *Finite Elements in Analysis and Design* **47**(7), 739–752.

Golzari, A., Sefat, M. H. and Jamshidi, S. (2015), 'Development of an adaptive surrogate model for production optimization', *Journal of petroleum Science and Engineering* **133**, 677–688.

Gong, J., Heaslip, K., McNeil, S., Farzan, F., Brink, S. et al. (2016), 'Big data: opportunities and challenges in asset management'.

Gopalakrishnan, S., Ruzzene, M., Hanagud, S. et al. (2011), *Computational techniques for structural health monitoring*, Springer.

Green, P. L., Cross, E. J. and Worden, K. (2015), 'Bayesian system identification of dynamical systems using highly informative training data', *Mechanical systems and signal processing* **56**, 109–122.

Grefenstette, J. J. (1993), Genetic algorithms and machine learning, *in* 'Proceedings of the sixth annual conference on Computational learning theory', pp. 3–4.

Gregor, S. and Hevner, A. R. (2013), 'Positioning and presenting design science research for maximum impact', *MIS quarterly* pp. 337–355.

Gregor, S., Jones, D. et al. (2007), The anatomy of a design theory, Association for Information Systems.

Gregory, A., Lau, F. D.-H., Girolami, M., Butler, L. J. and Elshafie, M. Z. (2019), 'The synthesis of data from instrumented structures and physics-based models via gaussian processes', *Journal of Computational Physics* **392**, 248–265.

Grieves, M. (2014), 'Digital twin: manufacturing excellence through virtual factory replication', *White paper* **1**, 1–7.

Gruber, T. R. (1993), 'A translation approach to portable ontology specifications', *Knowledge acquisition* **5**(2), 199–220.

Gude, M., Filippatos, A., Langkamp, A., Hufenbach, W., Kuschmierz, R., Fischer, A. and Czarske, J. (2015), 'Model assessment of a composite mock-up bladed rotor based on its vibration response and radial expansion', *Composite Structures* **124**, 394–401.

Gugercin, S. and Antoulas, A. C. (2004), 'A survey of model reduction by balanced truncation and some new results', *International Journal of Control* **77**(8), 748–766.

Gulgec, N. S., Shahidi, G. S., Matarazzo, T. J. and Pakzad, S. N. (2017), Current challenges with bigdata analytics in structural health monitoring, *in* 'Structural Health Monitoring & Damage Detection, Volume 7: Proceedings of the 35th IMAC, A Conference and Exposition on Structural Dynamics 2017', Springer, pp. 79–84.

Haasdonk, B. and Ohlberger, M. (2008), 'Reduced basis method for finite volume approximations of parametrized linear evolution equations', *ESAIM: Mathematical Modelling and Numerical Analysis* **42**(2), 277–302.

Hager, W. W. and Zhang, H. (2006), 'A survey of nonlinear conjugate gradient methods', *Pacific journal of Optimization* **2**(1), 35–58.

Harmon, M. E. and Harmon, S. S. (1997), 'Reinforcement learning: A tutorial.'.

Hartmann, S. (1996), The world as a process, *in* 'Modelling and simulation in the social sciences from the philosophy of science point of view', Springer, pp. 77–100.

Hawari, A., Alkadour, F., Elmasry, M. and Zayed, T. (2020), 'A state of the art review on condition assessment models developed for sewer pipelines', *Engineering Applications of Artificial Intelligence* **93**, 103721.

Hearst, M. A., Dumais, S. T., Osuna, E., Platt, J. and Scholkopf, B. (1998), 'Support vector machines', *IEEE Intelligent Systems and their applications* **13**(4), 18–28.

Helton, J. C., Johnson, J. D., Oberkampf, W. L. and Sallaberry, C. J. (2010), 'Representation of analysis results involving aleatory and epistemic uncertainty', *International Journal of General Systems* **39**(6), 605–646.

Higdon, D., Gattiker, J., Williams, B. and Rightley, M. (2008), 'Computer model calibration using high-dimensional output', *Journal of the American Statistical Association* **103**(482), 570–583.

Honarmandi, P. and Arróyave, R. (2020), 'Uncertainty quantification and propagation in computational materials science and simulation-assisted materials design', *Integrating Materials and Manufacturing Innovation* **9**, 103–143.

Hong, M.-S., So, Y.-S. and Kim, J.-G. (2019), 'Optimization of cathodic protection design for pre-insulated pipeline in district heating system using computational simulation', *Materials* **12**(11), 1761.

Hribernik, K. A., Rabe, L., Thoben, K.-D. and Schumacher, J. (2006), 'The product avatar as a product-instance-centric information management concept', *International Journal of Product Lifecycle Management* **1**(4), 367–379.

Hribernik, K., Wuest, T. and Thoben, K.-D. (2013), Towards product avatars representing middle-of-life information for improving design, development and manufacturing processes, *in* 'IFIP International Conference on Digital Product and Process Development Systems', Springer, pp. 85–96.

Huang, Y., Seck, M. D. and Verbraeck, A. (2010), 'Towards automated model calibration and validation in rail transit simulation', *Procedia Computer Science* **1**(1), 1259–1265.

Inigo, B., Colinas-Armijo, N., de Lacalle, L. N. L. and Aguirre, G. (2021), 'Digital twin-based analysis of volumetric error mapping procedures', *Precision Engineering* **72**, 823–836.

Inzillo, V., Santamaria, A. F. and Quintana, A. A. (2017), Integration of omnet++ simulator with matlab for realizing an adaptive beamforming system, *in* '2017 IEEE/ACM 21st International Symposium on Distributed Simulation and Real Time Applications (DS-RT)', IEEE, pp. 1–2.

ISO/DIS-23247-1 (2020), 'Automation systems and integration–digital twin framework for manufacturing–part 1: Overview and general principles'.

Jacoby, M. and Usländer, T. (2020), 'Digital twin and internet of things—current standards landscape', *Applied Sciences* **10**(18), 6519.

Jain, A. K., Peratta, C., Baynham, J. M. and Adey, R. A. (2011), Optimization of retrofit cathodic protection (cp) systems using computational modeling by evaluating performance of remnant and retrofit cp systems, taking into account long-term polarization effects, *in* 'CORROSION 2011', OnePetro.

Javed, K., Gouriveau, R. and Zerhouni, N. (2017), 'State of the art and taxonomy of prognostics approaches, trends of prognostics applications and open issues towards maturity at different technology readiness levels', *Mechanical Systems and Signal Processing* **94**, 214–236.

Jensen, H. A., Esse, C., Araya, V. and Papadimitriou, C. (2017), 'Implementation of an adaptive meta-model for bayesian finite element model updating in time domain', *Reliability Engineering & System Safety* **160**, 174–190.

Jesus, A., Brommer, P., Zhu, Y. and Laory, I. (2017), 'Comprehensive bayesian structural identification using temperature variation', *Engineering Structures* **141**, 75–82.

Jones, D., Snider, C., Nassehi, A., Yon, J. and Hicks, B. (2020), 'Characterising the digital twin: A systematic literature review', *CIRP Journal of Manufacturing Science and Technology* **29**, 36–52.

Julien, N. and Martin, E. (2021), 'How to characterize a digital twin: a usage-driven classification', *IFAC-PapersOnLine* **54**(1), 894–899.

Kang, F., Wu, Y., Li, J. and Li, H. (2021), 'Dynamic parameter inverse analysis of concrete dams based on jaya algorithm with gaussian processes surrogate model', *Advanced Engineering Informatics* **49**, 101348.

Katam, R., Kalapatapu, P. and Pasupuleti, V. D. K. (2022), A review on technological advancements in the field of data driven structural health monitoring, *in* 'European Workshop on Structural Health Monitoring: EWSHM 2022-Volume 3', Springer, pp. 371–380.

Khan, S. and Yairi, T. (2018), 'A review on the application of deep learning in system health management', *Mechanical Systems and Signal Processing* **107**, 241–265.

Kim, C., Chen, L., Wang, H. and Castaneda, H. (2021), 'Global and local parameters for characterizing and modeling external corrosion in underground coated steel pipelines: A review of critical factors', *Journal of Pipeline Science and Engineering* **1**(1), 17–35.

Kim, Y.-S., Kim, J., Choi, D., Lim, J.-Y. and Kim, J.-G. (2017), ‘Optimizing the sacrificial anode cathodic protection of the rail canal structure in seawater using the boundary element method’, *Engineering Analysis with Boundary Elements* **77**, 36–48.

Kim, Y.-S., Park, I.-J. and Kim, J.-G. (2019), ‘Simulation approach for cathodic protection prediction of aluminum fin-tube heat exchanger using boundary element method’, *Metals* **9**(3), 376.

Kirkpatrick, S., Gelatt Jr, C. D. and Vecchi, M. P. (1983), ‘Optimization by simulated annealing’, *science* **220**(4598), 671–680.

Kita, A., Cavalagli, N. and Ubertini, F. (2019), ‘Temperature effects on static and dynamic behavior of consoli palace in gubbio, italy’, *Mechanical Systems and Signal Processing* **120**, 180–202.

Kleinbaum, D. G., Dietz, K., Gail, M., Klein, M. and Klein, M. (2002), *Logistic regression*, Springer.

Knezevic, D., Fakas, E. and Riber, H. J. (2019), Predictive digital twins for structural integrity management and asset life extension–jip concept and results, *in* ‘SPE Offshore Europe Conference and Exhibition’, OnePetro.

Kromanis, R., Kripakaran, P. and Harvey, B. (2016), ‘Long-term structural health monitoring of the cleddau bridge: evaluation of quasi-static temperature effects on bearing movements’, *Structure and Infrastructure Engineering* **12**(10), 1342–1355.

Kythe, P. K. (2020), *An introduction to boundary element methods*, CRC press.

Law, A. M., Kelton, W. D. and Kelton, W. D. (2007), *Simulation modeling and analysis*, Vol. 3, Mcgraw-hill New York.

Lee, J., Azamfar, M., Singh, J. and Siahpour, S. (2020), ‘Integration of digital twin and deep learning in cyber-physical systems: towards smart manufacturing’, *IET Collaborative Intelligent Manufacturing* **2**(1), 34–36.

Lee, J., Jin, C. and Liu, Z. (2017), Predictive big data analytics and cyber physical systems for tes systems, *in* ‘Advances in Through-life Engineering Services’, Springer, pp. 97–112.

Lee, J., Wu, F., Zhao, W., Ghaffari, M., Liao, L. and Siegel, D. (2014), 'Prognostics and health management design for rotary machinery systems—reviews, methodology and applications', *Mechanical systems and signal processing* **42**(1-2), 314–334.

Li, S., Xu, L. D. and Zhao, S. (2015), 'The internet of things: a survey', *Information systems frontiers* **17**(2), 243–259.

Li, Y. (2017), 'Deep reinforcement learning: An overview', *arXiv preprint arXiv:1701.07274* .

Liang, Y., Lee, H., Lim, S., Lin, W., Lee, K. and Wu, C. (2002), 'Proper orthogonal decomposition and its applications—part i: Theory', *Journal of Sound and vibration* **252**(3), 527–544.

Ling, Y. and Mahadevan, S. (2012), 'Integration of structural health monitoring and fatigue damage prognosis', *Mechanical Systems and Signal Processing* **28**, 89–104.

Liu, C. and Kelly, R. G. (2019), 'A review of the application of finite element method (fem) to localized corrosion modeling', *Corrosion* **75**(11), 1285–1299.

Liu, G. and Cai, M. (2020), 'Modeling time-dependent deformation behavior of brittle rock using grain-based stress corrosion method', *Computers and Geotechnics* **118**, 103323.

Liu, J., Ouyang, H., Han, X. and Liu, G. (2021), 'Optimal sensor placement for uncertain inverse problem of structural parameter estimation', *Mechanical Systems and Signal Processing* **160**, 107914.

Liu, Y., Chen, W., Arendt, P. and Huang, H.-Z. (2011), 'Toward a better understanding of model validation metrics', *Journal of Mechanical Design* **133**(7).

Liu, Z., Meyendorf, N. and Mrad, N. (2018), The role of data fusion in predictive maintenance using digital twin, *in* 'AIP conference proceedings', Vol. 1949, AIP Publishing LLC, p. 020023.

López, C. P. (2014), Optimization techniques via the optimization toolbox, *in* 'MATLAB optimization techniques', Apress, Berkeley, CA, pp. 85–108.

Macchi, M., Roda, I., Negri, E. and Fumagalli, L. (2018), 'Exploring the role of digital twin for asset lifecycle management', *IFAC-PapersOnLine* **51**(11), 790–795.

Malveiro, J., Ribeiro, D., Sousa, C. and Calçada, R. (2018), 'Model updating of a dynamic model of a composite steel-concrete railway viaduct based on experimental tests', *Engineering Structures* **164**, 40–52.

Marcassoli, P., Bonetti, A., Lazzari, L. and Ormellese, M. (2013), Modeling of potential distribution of subsea pipeline under cathodic protection by finite element method, *in* 'NACE CORROSION', NACE, pp. NACE–2013.

March, S. T. and Smith, G. F. (1995), 'Design and natural science research on information technology', *Decision support systems* **15**(4), 251–266.

Maria, A. (1997), Introduction to modeling and simulation, *in* 'Proceedings of the 29th conference on Winter simulation', pp. 7–13.

Marque-Pucheu, S., Perrin, G. and Garnier, J. (2019), 'Efficient sequential experimental design for surrogate modeling of nested codes', *ESAIM: Probability and Statistics* **23**, 245–270.

Marques, S., Badcock, K., Khodaparast, H. and Mottershead, J. (2012), 'How structural model variability influences transonic aeroelastic stability', *Journal of Aircraft* **49**(5), 1189–1199.

Martinez, G. S., Sierla, S., Karhela, T. and Vyatkin, V. (2018), Automatic generation of a simulation-based digital twin of an industrial process plant, *in* 'IECON 2018-44th Annual Conference of the IEEE Industrial Electronics Society', IEEE, pp. 3084–3089.

McGetrick, P. J., Kim, C.-W., González, A. and Brien, E. J. (2015), 'Experimental validation of a drive-by stiffness identification method for bridge monitoring', *Structural Health Monitoring* **14**(4), 317–331.

Melios, C., Dimitriou, A., Kokkinos, N., Kioupis, N., Manolis, T. and Charalambous, C. (2023), 'Instant eoff measurement error in cathodically protected pipelines: A parametric assessment study', *Electric Power Systems Research* **223**, 109576.

Mohammadrahimi, A. and Sayebani, M. (2019), 'Using the bayesian updating approach to develop time-dependent corrosion wastage model for deck panel of bulk carriers', *Marine Structures* **64**, 92–109.

Mojtahedi, A., Yaghin, M. L., Hassanzadeh, Y., Ettefagh, M., Aminfar, M. and Aghdam, A. (2011), 'Developing a robust shm method for offshore jacket platform using model updating and fuzzy logic system', *Applied Ocean Research* **33**(4), 398–411.

Montáns, F. J., Chinesta, F., Gómez-Bombarelli, R. and Kutz, J. N. (2019), 'Data-driven modeling and learning in science and engineering', *Comptes Rendus Mécanique* **347**(11), 845–855.

Montgomery, D. C. (2017), *Design and analysis of experiments*, John wiley & sons.

Morton, K. W. and Mayers, D. F. (2005), *Numerical solution of partial differential equations: an introduction*, Cambridge university press.

Murphy, B. and Yarnold, M. (2018), 'Temperature-driven structural identification of a steel girder bridge with an integral abutment', *Engineering Structures* **155**, 209–221.

Myers, R. H., Montgomery, D. C. and Anderson-Cook, C. M. (2016), *Response surface methodology: process and product optimization using designed experiments*, John Wiley & Sons.

Nath, P., Hu, Z. and Mahadevan, S. (2017), 'Sensor placement for calibration of spatially varying model parameters', *Journal of Computational Physics* **343**, 150–169.

Neerukatti, R. K., Liu, K. C., Kovvali, N. and Chattopadhyay, A. (2014), 'Fatigue life prediction using hybrid prognosis for structural health monitoring', *Journal of Aerospace Information Systems* **11**(4), 211–231.
URL: *https://doi.org/10.2514/1.I010094*

Negri, E., Fumagalli, L. and Macchi, M. (2017), 'A review of the roles of digital twin in cps-based production systems', *Procedia Manufacturing* **11**, 939–948.

Newman, J. and Balsara, N. P. (2021), *Electrochemical systems*, John Wiley & Sons.

Nguyen, A., Kodikara, K., Chan, T. and Thambiratnam, D. (2018), 'Toward effective structural identification of medium-rise building structures', *Journal of Civil Structural Health Monitoring* **8**, 63–75.

Ni, D., Leonard, J. D., Guin, A. and Williams, B. M. (2004), 'Systematic approach for validating traffic simulation models', *Transportation Research Record* **1876**(1), 20–31.

Nocedal, J. and Wright, S. J. (1999), *Numerical optimization*, Springer.

Oberkampf, W. L., DeLand, S. M., Rutherford, B. M., Diegert, K. V. and Alvin, K. F. (2002), 'Error and uncertainty in modeling and simulation', *Reliability Engineering & System Safety* **75**(3), 333–357.

Oberkampf, W. L. and Trucano, T. G. (2002), 'Verification and validation in computational fluid dynamics', *Progress in aerospace sciences* **38**(3), 209–272.

Oberkampf, W. L. and Trucano, T. G. (2008), 'Verification and validation benchmarks', *Nuclear engineering and Design* **238**(3), 716–743.

Oden, J. T., Prudencio, E. E. and Bauman, P. T. (2013), 'Virtual model validation of complex multiscale systems: applications to nonlinear elastostatics', *Computer Methods in Applied Mechanics and Engineering* **266**, 162–184.

Omar, I., Khan, M. and Starr, A. (2022), 'Compatibility and challenges in machine learning approach for structural crack assessment', *Structural Health Monitoring* **21**(5), 2481–2502.

Orlova, E. V. (2022), 'Design technology and ai-based decision making model for digital twin engineering', *Future Internet* **14**(9), 248.

Ostertagová, E. (2012), 'Modelling using polynomial regression', *Procedia Engineering* **48**, 500–506.

Ozer, E. and Feng, M. Q. (2019), 'Structural reliability estimation with participatory sensing and mobile cyber-physical structural health monitoring systems', *Applied Sciences* **9**(14), 2840.

Pace, D. K. (2004), 'Modeling and simulation verification and validation challenges', *Johns Hopkins APL technical digest* **25**(2), 163–172.

Papalambros, P. Y. and Wilde, D. J. (2000), *Principles of optimal design: modeling and computation*, Cambridge university press.

Park, B. and Schneeberger, J. (2003), 'Microscopic simulation model calibration and validation: case study of vissim simulation model for a coordinated actuated signal system', *Transportation research record* **1856**(1), 185–192.

Parrott, A. and Warshaw, L. (2017), 'Industry 4.0 and the digital twin', *Deloitte University Press* pp. 1–17.

URL: *https://dupress.deloitte.com/dup-us-en/focus/industry-4-0/digital-twin-technology-smart-factory.html*

Peč, M., Šebek, F., Zapletal, J., Petruška, J. and Hassan, T. (2019), 'Automated calibration of advanced cyclic plasticity model parameters with sensitivity analysis for aluminium alloy 2024-t351', *Advances in Mechanical Engineering* **11**(3), 1687814019829982.

Peffers, K., Tuunanen, T. and Niehaves, B. (2018), 'Design science research genres: introduction to the special issue on exemplars and criteria for applicable design science research'.

Peratta, C., Adey, R., Baynham, J. and Dede, A. (2021), Supporting integrity management with a cp digital twin, *in* 'CORROSION 2021', OnePetro.

Poli, R., Kennedy, J. and Blackwell, T. (2007), 'Particle swarm optimization', *Swarm intelligence* **1**(1), 33–57.

Ponomarev, K., Kudryashov, N., Popelnukha, N. and Potekhin, V. (2017), Main principals and issues of digital twin development for complex technological processes, *in* 'Proceedings of the 28th DAAAM International Symposium', pp. 0523–0528.

Potts, C. (1993), 'Software-engineering research revisited', *IEEE software* **10**(5), 19–28.

Qiao, Q., Wang, J., Ye, L. and Gao, R. X. (2019), 'Digital twin for machining tool condition prediction', *Procedia CIRP* **81**, 1388–1393.

Quarteroni, A., Rozza, G. et al. (2014), *Reduced order methods for modeling and computational reduction*, Vol. 9, Springer.

Quintana, J., Carrión, F. and Crespo, S. (2014), Damage detection on a cable stayed bridge using wave propagation analysis, *in* 'EWSHM-7th European Workshop on Structural Health Monitoring'.

Rabiei, E., Droguett, E. L. and Modarres, M. (2018), 'Fully adaptive particle filtering algorithm for damage diagnosis and prognosis', *Entropy* **20**(2), 100.

Rajaram, D. (2020), Methods for Construction of Surrogates For Computationally Expensive High-Dimensional Problems, PhD thesis, Georgia Institute of Technology.

Ramancha, M. K., Astroza, R., Conte, J. P., Restrepo, J. I. and Todd, M. D. (2020), Bayesian nonlinear finite element model updating of a full-scale bridge-column using sequential monte carlo, *in* 'Model Validation and Uncertainty Quantification, Volume 3: Proceedings of the 38th IMAC, A Conference and Exposition on Structural Dynamics 2020', Springer, pp. 389–397.

Rao, S. S. (2017), *The finite element method in engineering*, Butterworth-heinemann.

Rasheed, A., San, O. and Kvamsdal, T. (2020), 'Digital twin: Values, challenges and enablers from a modeling perspective', *Ieee Access* **8**, 21980–22012.

Ríos, J., Hernandez, J. C., Oliva, M. and Mas, F. (2015), 'Product avatar as digital counterpart of a physical individual product: Literature review and implications in an aircraft', *Transdisciplinary Lifecycle Analysis of Systems* pp. 657–666.

Ritto, T. and Rochinha, F. (2021), 'Digital twin, physics-based model, and machine learning applied to damage detection in structures', *Mechanical Systems and Signal Processing* **155**, 107614.

Robinson, S. (2008), 'Conceptual modelling for simulation part i: definition and requirements', *Journal of the operational research society* **59**(3), 278–290.

Robinson, S., Arbez, G., Birta, L. G., Tolk, A. and Wagner, G. (2015), 'Conceptual modeling: definition, purpose and benefits', *2015 Winter Simulation Conference (WSC)* pp. 2812–2826.

Robinson, T., Eldred, M. S., Willcox, K. E. and Haimes, R. (2008), 'Surrogate-based optimization using multifidelity models with variable parameterization and corrected space mapping', *AIAA journal* **46**(11), 2814–2822.

Rogers, C. A., Kassab, A. J., Divo, E. A., Ostrowski, Z. and Bialecki, R. A. (2012), 'An inverse pod-rbf network approach to parameter estimation in mechanics', *Inverse Problems in Science and Engineering* **20**(5), 749–767.

Rosen, R., Von Wichert, G., Lo, G. and Bettenhausen, K. D. (2015), 'About the importance of autonomy and digital twins for the future of manufacturing', *Ifac-papersonline* **48**(3), 567–572.

Roy, R., Hinduja, S. and Teti, R. (2008), 'Recent advances in engineering design optimisation: Challenges and future trends', *CIRP annals* **57**(2), 697–715.

Rozza, G., Malik, H., Demo, N., Tezzele, M., Girfoglio, M., Stabile, G. and Mola, A. (2018), 'Advances in reduced order methods for parametric industrial problems in computational fluid dynamics', *arXiv preprint arXiv:1811.08319* .

Sabamehr, A., Lim, C. and Bagchi, A. (2018), 'System identification and model updating of highway bridges using ambient vibration tests', *Journal of Civil Structural Health Monitoring* **8**, 755–771.

Sacks, R., Brilakis, I., Pikas, E., Xie, H. S. and Girolami, M. (2020), 'Construction with digital twin information systems', *Data-Centric Engineering* **1**.

Sapkota, M. S., Apeh, E., Adey, R., Baynham, J. and Peratta, C. (2022b), Improving integrity management decision making using automated calibration of model predictions from cathodic-protection (cp) survey data, *in* 'AMPP Annual Conference+ Expo', OnePetro.

Sapkota, M. S., Apeh, E., Hadfield, M., Adey, R. and Baynham, J. (2021a), 'Design of experiments platform for online simulation model validation and parameter updating within digital twinning', *WIT Transactions on Engineering Sciences* **130**(12), 3–14.

Sapkota, M. S., Apeh, E., Hadfield, M., Haratian, R., Adey, R. and Baynham, J. (2021b), 'An approach for adaptive model performance validation within digital twinning', *International Journal of Computational Methods and Experimental Measurements* **9**(3), 213–225.

Sapkota, M. S., Apeh, E., Hadfield, M., Haratian, R., Adey, R. and Baynham, J. (2022a), 'Surrogate-assisted parametric calibration using design of experiment platform within digital twinning', *International Journal of Computational Methods and Experimental Measurements* **10**(2), 158–171.

Sargent, R. G. (1984), 'A tutorial on verification and validation of simulation models'.

Sargent, R. G. (2010), Verification and validation of simulation models, *in* 'Proceedings of the 2010 winter simulation conference', IEEE, pp. 166–183.

Sarin, H., Kokkolaras, M., Hulbert, G., Papalambros, P., Barbat, S. and Yang, R.-J. (2010), 'Comparing time histories for validation of simulation models: error measures and metrics'.

Scaglioni, B. and Ferretti, G. (2018), 'Towards digital twins through object-oriented modelling: a machine tool case study', *IFAC-PapersOnLine* **51**(2), 613–618.

Schroeder, G., Steinmetz, C., Pereira, C. E., Muller, I., Garcia, N., Espindola, D. and Rodrigues, R. (2016), Visualising the digital twin using web services and augmented reality, *in* '2016 IEEE 14th international conference on industrial informatics (INDIN)', IEEE, pp. 522–527.

Sehgal, S. and Kumar, H. (2016), 'Structural dynamic model updating techniques: A state of the art review', *Archives of Computational Methods in Engineering* **23**(3), 515–533.
URL: *https://doi.org/10.1007/s11831-015-9150-3*

Seltman, H. J. (2012), 'Experimental design and analysis'.

Seshadri, B. R. and Krishnamurthy, T. (2017), Structural health management of damaged aircraft structures using digital twin concept, *in* '25th aiaa/ahs adaptive structures conference', p. 1675.

Shalev-Shwartz, S. and Ben-David, S. (2014), *Understanding machine learning: From theory to algorithms*, Cambridge university press.

Shinyama, Y. (2015), 'Pdfminer: Python pdf parser and analyzer', *Retrieved on* **11**.

Siedlak, D. J., Pinon, O. J., Schlais, P. R., Schmidt, T. M. and Mavris, D. N. (2018), 'A digital thread approach to support manufacturing-influenced conceptual aircraft design', *Research in Engineering Design* **29**(2), 285–308.

Silva, R., Ribeiro, D., Bragança, C., Costa, C., Arêde, A. and Calçada, R. (2021), 'Model updating of a freight wagon based on dynamic tests under different loading scenarios', *Applied Sciences* **11**(22), 10691.

Soares, R. M., Câmara, M. M., Feital, T. and Pinto, J. C. (2019), 'Digital twin for monitoring of industrial multi-effect evaporation', *Processes* **7**(8), 537.

Sohn, H. (2007), 'Effects of environmental and operational variability on structural health monitoring', *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* **365**(1851), 539–560.

Soomro, A. A., Mokhtar, A. A., Kurnia, J. C., Lashari, N., Sarwar, U., Jameel, S. M., Inayat, M. and Oladosu, T. L. (2022), 'A review on bayesian modeling approach to quantify failure risk assessment of oil and gas pipelines due to corrosion', *International Journal of Pressure Vessels and Piping* p. 104841.

Spiridonakos, M., Ou, Y., Chatzi, E. and Reiter, U. (2016), Wind turbines structural identification framework for the representation of both short-and long-term variability, *in* '7th International Conference on Structural Health Monitoring of Intelligent Infrastructure (SHMII 2015)', Vol. 3, International Society for Structural Health Monitoring of Intelligent . . . , pp. 2488–2497.

Stark, R., Kind, S. and Neumeyer, S. (2017), 'Innovations in digital modelling for next generation manufacturing system design', *CIRP annals* **66**(1), 169–172.

Stark, R. and Pförtner, A. (2015), 'Integrating ontology into plm-tools to improve sustainable product development', *CIRP Annals* **64**(1), 157–160.

Stutzmann, J. (2017), Cathodic Corrosion Protection in the Context of Lifetime Extension of Monopile-based Offshore Wind Turbines, PhD thesis.

Sun, L., Shang, Z., Xia, Y., Bhowmick, S. and Nagarajaiah, S. (2020), 'Review of bridge structural health monitoring aided by big data and artificial intelligence: From condition assessment to damage detection', *Journal of Structural Engineering* **146**(5), 04020073.

Sun, X., Bao, J., Li, J., Zhang, Y., Liu, S. and Zhou, B. (2020), 'A digital twin-driven approach for the assembly-commissioning of high precision products', *Robotics and Computer-Integrated Manufacturing* **61**, 101839.

Tahmasebi, F., Zach, R., Schuß, M. and Mahdavi, A. (2012), Simulation model calibration: An optimization-based approach, *in* 'Proceedings of Fourth German-Austrian IBPSA Conference, BauSIM', pp. 386–91.

Tao, F., Zhang, M., Liu, Y. and Nee, A. Y. (2018), 'Digital twin driven prognostics and health management for complex equipment', *Cirp Annals* **67**(1), 169–172.

Taylor, D. C., Pawar, V., Kruzikas, D., Gilmore, K. E., Pandya, A., Iskandar, R. and Weinstein, M. C. (2010), 'Methods of model calibration', *Pharmacoeconomics* **28**(11), 995–1000.

Tezdogan, T. and Demirel, Y. K. (2014), 'An overview of marine corrosion protection with a focus on cathodic protection and coatings', *Brodogradnja: Teorija i praksa brodogradnje i pomorske tehnike* **65**(2), 49–59.

Thomas, J. W. (2013), *Numerical partial differential equations: finite difference methods*, Vol. 22, Springer Science & Business Media.

Torraco, R. J. (2005), 'Writing integrative literature reviews: Guidelines and examples', *Human resource development review* **4**(3), 356–367.

Tripura, T., Desai, A. S., Adhikari, S. and Chakraborty, S. (2023), 'Probabilistic machine learning based predictive and interpretable digital twin for dynamical systems', *Computers & Structures* **281**, 107008.

Tuegel, E. J., Ingraffea, A. R., Eason, T. G. and Spottswood, S. M. (2011), 'Reengineering aircraft structural life prediction using a digital twin', *International Journal of Aerospace Engineering* **2011**.

Tygesen, U. T., Jepsen, M. S., Vestermark, J., Dollerup, N. and Pedersen, A. (2018), The true digital twin concept for fatigue re-assessment of marine structures, *in* 'International Conference on Offshore Mechanics and Arctic Engineering', Vol. 51203, American Society of Mechanical Engineers, p. V001T01A021.

Van Beers, W. C. (2005), Kriging metamodeling in discrete-event simulation: an overview, *in* 'Proceedings of the Winter Simulation Conference, 2005.', IEEE, pp. 7–pp.

Van Buren, K. L., Mollineaux, M. G., Hemez, F. M. and Atamturktur, S. (2013), 'Simulating the dynamics of wind turbine blades: part ii, model validation and uncertainty quantification', *Wind Energy* **16**(5), 741–758.

van Rossum, G. and Drake, F. L. (2009), 'pickle–python object serialization', *Available from World Wide Web: http://docs. python. org/lib/module-pickle. html* .

VanDerHorn, E. and Mahadevan, S. (2021), 'Digital twin: Generalization, characterization and implementation', *Decision support systems* **145**, 113524.

Venable, J., Pries-Heje, J. and Baskerville, R. (2012), A comprehensive framework for evaluation in design science research, *in* 'International conference on design science research in information systems', Springer, pp. 423–438.

Venter, G. (2010), 'Review of optimization techniques'.

Vincenzi, L., Bassoli, E., Ponsi, F., Castagnetti, C. and Mancini, F. (2019), 'Dynamic monitoring and evaluation of bell ringing effects for the structural assessment of a masonry bell tower', *Journal of Civil Structural Health Monitoring* **9**, 439–458.

Vrabič, R., Erkoyuncu, J. A., Butala, P. and Roy, R. (2018), 'Digital twins: Understanding the added value of integrated models for through-life engineering services', *Procedia Manufacturing* **16**, 139–146. Proceedings of the 7th International Conference on Through-life Engineering Services.
URL: *https://www.sciencedirect.com/science/article/pii/S2351978918312897*

Wagner, N. and Rondinelli, J. M. (2016), 'Theory-guided machine learning in materials science', *Frontiers in Materials* **3**, 28.

Wang, J., Ye, L., Gao, R. X., Li, C. and Zhang, L. (2019), 'Digital twin for rotating machinery fault diagnosis in smart manufacturing', *International Journal of Production Research* **57**(12), 3920–3934.

West, T. D. and Blackburn, M. (2017), 'Is digital thread/digital twin affordable? a systemic assessment of the cost of dod's latest manhattan project', *Procedia computer science* **114**, 47–56.

Westermann, P. and Evins, R. (2019), 'Surrogate modelling for sustainable building design–a review', *Energy and Buildings* **198**, 170–186.

Wieringa, R. (2010), Design science methodology: principles and practice, *in* 'Proceedings of the 32nd ACM/IEEE International Conference on Software Engineering-Volume 2', pp. 493–494.

Willard, J., Jia, X., Xu, S., Steinbach, M. and Kumar, V. (2022), 'Integrating scientific knowledge with machine learning for engineering and environmental systems', *ACM Computing Surveys* **55**(4), 1–37.

Wohlin, C. (2014), Guidelines for snowballing in systematic literature studies and a replication in software engineering, *in* 'Proceedings of the 18th international conference on evaluation and assessment in software engineering', pp. 1–10.

Wright, L. and Davidson, S. (2020), 'How to tell the difference between a model and a digital twin', *Advanced Modeling and Simulation in Engineering Sciences* **7**(1), 1–13.

Xia, T., Dong, Y., Xiao, L., Du, S., Pan, E. and Xi, L. (2018), 'Recent advances in prognostics and health management for advanced manufacturing paradigms', *Reliability Engineering & System Safety* **178**, 255–268.

Yang, X.-S., Koziel, S. and Leifsson, L. (2013), 'Computational optimization, modelling and simulation: Recent trends and challenges', *Procedia Computer Science* **18**, 855–860.

Ye, C., Butler, L., Calka, B., Iangurazov, M., Lu, Q., Gregory, A., Girolami, M. and Middleton, C. (2019), 'A digital twin of bridges for structural health monitoring'.

Ye, Y., Yang, Q., Yang, F., Huo, Y. and Meng, S. (2020), 'Digital twin for the structural health management of reusable spacecraft: A case study', *Engineering Fracture Mechanics* **234**, 107076.

Yin, H., Ma, J., Dong, K., Peng, Z., Cui, P. and Yang, C. (2019), 'Model updating method based on kriging model for structural dynamics', *Shock and Vibration* **2019**.

Yondo, R., Andrés, E. and Valero, E. (2018), 'A review on design of experiments and surrogate models in aircraft real-time and many-query aerodynamic analyses', *Progress in aerospace sciences* **96**, 23–61.

Yu, B., Yang, L., Wu, M. and Li, B. (2014), 'Practical model for predicting corrosion rate of steel reinforcement in concrete structures', *Construction and Building Materials* **54**, 385–401.

Zambrano-Bigiarini, M. and Rojas, R. (2013), 'A model-independent particle swarm optimisation software for model calibration', *Environmental Modelling & Software* **43**, 5–25.

Zapomĕl, J., Stachiv, I. and Ferfecki, P. (2016), 'A novel method combining monte carlo–fem simulations and experiments for simultaneous evaluation of the ultrathin film mass density and young s modulus', *Mechanical Systems and Signal Processing* **66**, 223–231.

Zhang, L., Zhou, L. and Horn, B. K. (2021), 'Building a right digital twin with model engineering', *Journal of Manufacturing Systems* **59**, 151–164.

Zhang, X. and Hao, Z. (2021), Computational models of magnesium medical implants degradation: A review, *in* 'Journal of Physics: Conference Series', Vol. 1838, IOP Publishing, p. 012012.

Zhang, X., Xie, F., Ji, T., Zhu, Z. and Zheng, Y. (2021), 'Multi-fidelity deep neural network surrogate model for aerodynamic shape optimization', *Computer Methods in Applied Mechanics and Engineering* **373**, 113485.

Zhang, Z., Zhao, Z., Li, X., Zhang, X., Wang, S., Yan, R. and Chen, X. (2020), ‘Faster multiscale dictionary learning method with adaptive parameter estimation for fault diagnosis of traction motor bearings’, *IEEE Transactions on Instrumentation and Measurement* **70**, 1–13.

Zhao, R., Yan, D., Liu, Q., Leng, J., Wan, J., Chen, X. and Zhang, X. (2019), ‘Digital twin-driven cyber-physical system for autonomously controlling of micro punching system’, *IEEE Access* **7**, 9459–9469.

Zheng, W., Qian, F., Shen, J. and Xiao, F. (2020), ‘Mitigating effects of temperature variations through probabilistic-based machine learning for vibration-based bridge scour detection’, *Journal of Civil Structural Health Monitoring* **10**, 957–972.

Zhou, K. and Tang, J. (2021), ‘Structural model updating using adaptive multi-response gaussian process meta-modeling’, *Mechanical Systems and Signal Processing* **147**, 107121.

Zhuang, C., Liu, J. and Xiong, H. (2018), ‘Digital twin-based smart production management and control framework for the complex product assembly shop-floor’, *The international journal of advanced manufacturing technology* **96**(1), 1149–1163.

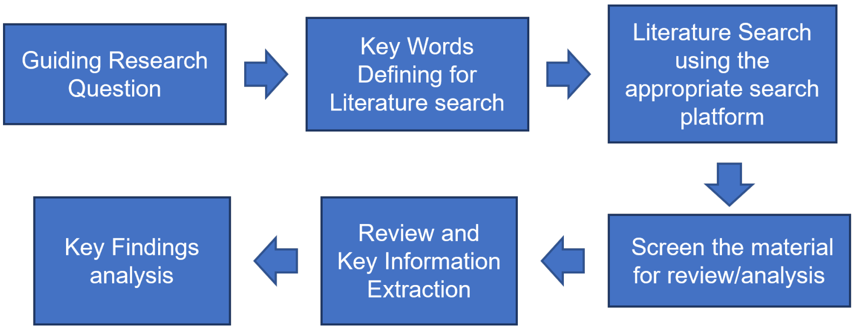
Zienkiewicz, O. C., Taylor, R. L. and Zhu, J. Z. (2005), *The finite element method: its basis and fundamentals*, Elsevier.

Zupan, J. (1994), ‘Introduction to artificial neural network (ann) methods: what they are and how to use them’, *Acta Chimica Slovenica* **41**, 327–327.

Appendix A - Analysis of related representative literature

A review of the SHM-related literature is done to have insights into current developments in model calibration and adaptation-related challenges in SHM.

The following procedure is followed for the related literature study. The search, review



and analysis are motivated by the following research question:

RQ: What is the current development in physics-based model parametric calibration and adaptation particularly those involved in SHM?

The keywords groups identified related to the research question are:

1. ”structural health” OR “SHM”
2. “physics-based model” OR ”simulation model”
3. ”model calibration” OR ”model adaptation” OR ”parameter updating” OR “parameter estimation”

Table A.1: Materials count yielded from Search Engines

Scopus	WebOfScience	Google Scholar
121	3	840

The next step involved determining the source of the publication databases. Since searching individual publication databases was found to be insufficient in yielding comprehensive results (Table B.1), a general source such as Google Scholar was taken into account. Additionally, Google Scholar, addresses the biases of specific scientific publishers, as suggested by Wohlin (2014). The search was limited to papers published up to 2021 and required at least one keyword from all three groups mentioned earlier. As Google Scholar restricts searches to either the title or the entire body of the material, the search was conducted across the entire body.

The search yielded 834 potential research items, which included research papers and related theses. These items were then screened for quality and relevance to the research topics. The initial screening process involved reviewing the abstracts, which resulted in 116 scholarly papers and relevant theses being selected. However, the number of relevant papers was reduced due to the keyword being present only in the referencing sections of most papers.

Out of the 116 selected papers, 78 works were chosen for the survey based on their quality (as determined by citation received) and appropriateness to the survey topic. Additionally, 10 other articles from other domains, such as “corrosion modelling” related to SHM, were added to the list as they were not found in the previous search. The tables below provide information on the representative works, including their categorisation by application domain, simulation, and calibration method, as well as key findings from the corresponding papers.

Table A.2: Summary of Representative SHM Research Discussing Physics-based Model and Parametric Calibration

Studies	Predictive Model type	model application Domain	Experimental Setup					Key Relevant Information /Suggestion
				Calibration related				
			Tool(s) for modelling	Parameters	method/technique		data	
Jain et al. (2011)	CP performance (BEM)	Catodic-protection offshore structure	BEASY	polarisation curve	Manual	trial and error	from survey	**Requirement of re-calibration** with time
Mojtahedi et al. (2011)	elastic (FEM)	offshore jacket	ANSYS	Young's modulus , Poisson's ratio, and density	Manual	Iterative	frequency response function measurements	Data required for calibration not standardised
An et al.(2013)	crack growth (MATLAB)	aero-structure	ANSYS	damage model parameters	Analytical (semi-automated) with MATLAB	Particle filter	sensor data	The trail-and-error based calibration can be somehow achieved with analytical way (with codes in software)
	degradation	battery	codes on MATLAB	mathematical parameters	Analytical (semi-automated) with MATLAB	Particle filter	measurement data	
Catbas et al. (2013)	FEM deflection model	Laboratory set bridge	-	spring constant, modulus of elasticity, and moment of inertia	Analytical with MATLAB code	ANN model with reverse experimental sample data	measured with gauges	**Abundant (10000) samples** FEM model run required to train ANN model. Continuous calibration is discussed and suggested for operational phase
Van Buren et al. (2013)	vibration dynamics using FEM	wind turbine blade	ANSYS	Trailing edge (density),spring stiffness, Young's Modulus	Statistical effect screening (ANOVA)	Uncertainty quantification using emulators of the model	Vibration measurements	**FE model integration** with other model such as **computation fluid dynamics model** is suggested as future task
Quintana et al.(2014)	FE wave propagation model	Cable Stayed Bridge	StaDyn software	mass, damping and stiffness matrices	Strdent software, complementary of StaDyn	Software to solve inverse problem	sensors placed at a different spatial location	**Inspections and test** are recommended as data from sensors may not be sufficient to track mechanical, geometric and physical conditions.
Yu et al. (2014)	mathematical corrosion model	concrete structure		W/C ratio, chloride content, concrete thickness ,relative humidity	Analytical	Curve-fitting	experimental data	The capitalisation of previous achievement in corrosion modelling **is lacking**.
Gude et al. (2015)	elastic and expansion model using FEM	composite mock-up bladed rotor	ANSYS	modulus, Poisson's ratio, Area density	Curve fitting	Response surface (surrogate) assisted	measured responses from experiment set up	**The model extension concept** is discussed to analyse the relation between speed and temperature
Behmanesh and Moaveni (2016)	FE model for response prediction and damage identification	footbridge	-	Young's modulus , ambient temperature	Analytical support	Hierarchical Bayesian framework	8721 sets of data measured from footbridge	**modelling environmental effects** anticipated on reducing variability of parameter

Studies	Predictive Model type	model application Domain	Experimental Setup					Key Relevant Information /Suggestion
			Tool(s) for modelling	Calibration related				
				Parameters	method/technique		data	
Kromanis et al. (2016)	bearing movement for thermal load with FEM	Bridge	FEDEASLab	Young's modulus, Poisson's ratio, density & thermal coefficient	Manual	trial and error	temperature and displacement data from measurement campaign	Data from operational phase and **knowledge on env. temp distribution** in anticipated to enhance performance of model.
Bi et al. (2017)	elasticity model using simplified FE (1d)	Frame structure	ANSYS	Young's modulus, Overall mass density, Offset of vertical elements	Analytical (semi-automated)	Deterministic and Stochastic	simulated experimental data	Highlights the issue with deterministic metric leading to **local minima**
Jesus et al. (2017)	Thermal model	aluminium bridge at laboratory	ANSYS	Young modulus, Poisson coefficient and thermal coefficient	Analytical with MATLAB code	comprehensive (module stepwise) Bayesian with Genetic algorithm	temperature and strain data from laboratory measurement	**Module based approach** adopted to address variations in parameters caused due to temperature difference.
Jensen et al. (2017)	elastic model	reinforced concrete structure	ANSYS	Young's modulus, Poisson ratio, mass density	Homemade code based on a MATLAB - C++ platform	Markov-chain Monte-Carlo assisted with adaptive surrogate	simulated response data	**adaptive surrogate** idea is suggested for operational time model calibration
Belostotsky et al. (2018)	FEM for load-bearing capacity analysis	Buildings	-	materials related parameters	-	sensitivity-analysis based	-	**same method** is considered for **calibration and adaptation**.
Malveiro et al. (2018)	Finite element Model to predict deck slab's dynamic response.	steel-concrete railway	ANSYS	Young's modulus , Poisson's ratio, density, etc	Automated with MATLAB code and OptiSlang	genetic algorithm for optimisation	With 26 high sensitivity piezoelectric accelerometers	Forward prediction is not enough discussed for prognostic application.
Murphy and Yarnold (2018)	FE model	Bridge	Strand7 finite element analysis software	Stiffness with temperature-driven signature	Developed MATLAB program and interfaced to FE software (Strand 7)	-	-	Automation is suggested to simulate many models without spending excessive amounts of time on data analysis. The environmental (Temperature) effect on the structure is analysed.
Nguyen et al. (2018)	FE vibration model	Building structure	software package SAP2000	Young's Modulus, Mass Density, Cross-Sectional Area, Torsional Stiffness, 2nd Moment of Area, shell thickness	Analytical using sensitivity matrix	Sensitivity-based and a pseudo-inverse parameter estimation	-	Point out the **accuracy requirement for initial analytical models**, in sensitivity-based calibration. Also, the exclusion of low-sensitivity parameters is suggested to avoid ill conditions of the model calibration process.

Studies	Predictive Model type	model application Domain	Experimental Setup					Key Relevant Information /Suggestion
			Tool(s) for modelling	Calibration related				
				Parameters	method/technique		data	
Sabamehr et al. (2018)	FE vibration model	Highway bridge	SAP2000 - MATLAB based FEM	-	MATLAB tool for ANN	Neural Network and Genetic Algorithms	-	A grouping scheme is used in this article to reduce the number of design variables, and the number could affect the quality of ANN.
Hong et al. (2019)	BEM model for Cathodic-Protection	Pre-insulated pipeline	BEASY	Cathodic polarisation curve	Manual analysis	Potentiodynamic polarisation test	-	Discuss the importance of computational-based analysis for real-case scenarios in CP system.
Kim et al. (2019)	BEM model for Cathodic-Protection	Aluminium-fin tube heat exchanger	BEASY	anodic and cathodic Tafel slopes	Manual analysis	Traditional potentiodynamic polarisation test	-	Suggest for practical application of numerical simulation of the multi-galvanic situation to improve the corrosion design of products.
Kita et al. (2019)	concrete damage plasticity (CDP) model using FEM	Buildings	ANSYS	Young's modulus and Poisson's ratio	Manual	trial-and-error based tuning	capture experimental vibration data	**environmental factor (temperature) affect on elastic model** is core topic of discussion and suggested modelling with/for it.
Knezevic et al. (2019)	potential fatigue and RUL	offshore jacket structures	Akselos (RB-FEA)	mass, damping, stiffness	-	-	strategically placed accelerometers on structure	**The model with real-time parameters termed as DT** and also discuss model updating should be case-by-case specific
Mohammadrahimi (2019)	A basic time step exponential Corrosion model	Ship structure	-	coefficient from equation	-	Bayesian method in MATLAB	thousand of measured points	The implemented model do lack the physics of the phenomenon and required 1000 of sample points to calibrate which is not feasible.
Ozer and Feng (2019)	FE vibration model	Bridge	OpenSees Script	-	MATLAB integration with OpensSees	-	experimental data collected with the aid of smartphone sensor	Discuss the importance of adequate data from the structure and propose mobile app technology for data collection. Automation of the platform is considered a future scope.
Pec et al. (2019)	plasticity model using FE codes	aluminium alloy	Code_Aster	elasticity and plasticity related	Automated	sensitivity analysis-based	experimental (user-defined) data	**Simulation tool also allowing calibration** with built in routines, but with limited application
Cao et al. (2020)	elasticity model FEM	oscillator, magnetometer boom, and a cantilever	NASTRAN	damping and system parameter	Analytical using MATLAB code	sensitivity analysis based optimisation	data from magnetometer	**Discussed limitation of sensitivity based optimisation (to local minima) common** on the situation with error in measurement data.
Liu and Cai(2020)	grain-based stress corrosion model	brittle rock	-	elastic modulus, Poisson's ratio, density	Manual	curve fitting (failure time vs time) with trail and error	data obtained from damage-controlled tests	anticipate modelling of time-dependent failure of blocky rocks

Studies	Predictive Model type	model application Domain	Experimental Setup					Key Relevant Information /Suggestion
			Tool(s) for modelling	Calibration related				
				Parameters	method/technique		data	
Ezzat et al. (2020)	FE model for fault prediction	aluminium cantilevered plate that	ANSYS	-	-	statistical calibration with Surrogate	experimental data	Prediction fault with a surrogate is discussed, and automation of the process is suggested. The environmental effect and its separate assessment are suggested as future work.
Liu and Cai(2020)	grain-based stress corrosion model	brittle rock	-	elastic modulus, Poisson's ratio, density	Manual	curve fitting (failure time vs time) with trail and error	data obtained from damage-controlled tests	Anticipate modelling of time-dependent failure of blocky rocks
Ye et al. (2020)	FEA for fatigue and crack-growth	airframe	ABAQUS/ Standard solver	parameters of Walker's law	Analytical with a python toolkit *"sklearn"*	dynamic Bayesian network	observed crack data	**DT concept integrating online monitoring data** to enhance predictive capability and real-time simulation techniques anticipated in future
Abdalla & Hambdy (2021)	2D FEM, solving Fick's law	-	ANSYS APDL	diffusion coefficient	-	trial and error	-	Model phasing out with geometrical and parametric change, making the process of re-calibration more complex with time.
Barzegari et al.(2021)	FEM together with corrosion degradation model	A cuboid of Magnesium	open-source PDE solver FreeFEM	corrosion related parameters (diffusion coefficient)		sensitivity based and assisted with Bayesian	Data obtained from the experimental tests	Time step simulation is undertaken.
Kang et al. (2021)	FEM to simulate seismic performance	Concrete dam	ANSYS	dynamic elastic modulus and	Analytical but simulation data fed to MATLAB	Jaya algorithm with Gaussian surrogate model	vibration measurement data selected randomly	**The optimal selection of measurement points** (placing sensors) is considered for future research, which is preeminent for higher parameter count.
Liu et al. (2021)	degradation model	FEM model for crack plate problem	Not mentioned	-	Analytical	Bayesian assisted with Monte Carlo method	sensor data	Discuss the importance of adequate sensor data and their spatial placement optimisation.
Silva et al. (2021)	Components flexibility model with FEM	Freight Wagon	ANSYS	Springs' stiffness, deformability modulus, equivalent density, etc.	Analytical with simulation data feeding to MATLAB	Genetic Algorithm with the interaction of software MATLAB & ANSYS	experiment data obtained from 14 high-sensitivity accelerometers	Data from the ANSYS are fed to MATLAB in text files which also demand automation for self-data transferring.
Zhang and Hao (2021)	magnesium corrosion degradation	Magnesium medical implant	-	related to corrosive environment properties	Manual	curve fitting (global damage vs time) with trail and error	-	**model accounting for the parameter change (material degradation)** is suggested to be hugely beneficial in future prediction.
Zhou and Tang (2021)	Finite element model for structural analysis	multi-plate Structure	MATLAB codes	stiffness, mass variation coefficient	Analytical using MATLAB code	assisted with Gaussian Process meta modelling	synthetic data from FE simulation	Same software is preferred for modelling to facilitate a the calibration (**difficulty with software heterogeneity)**

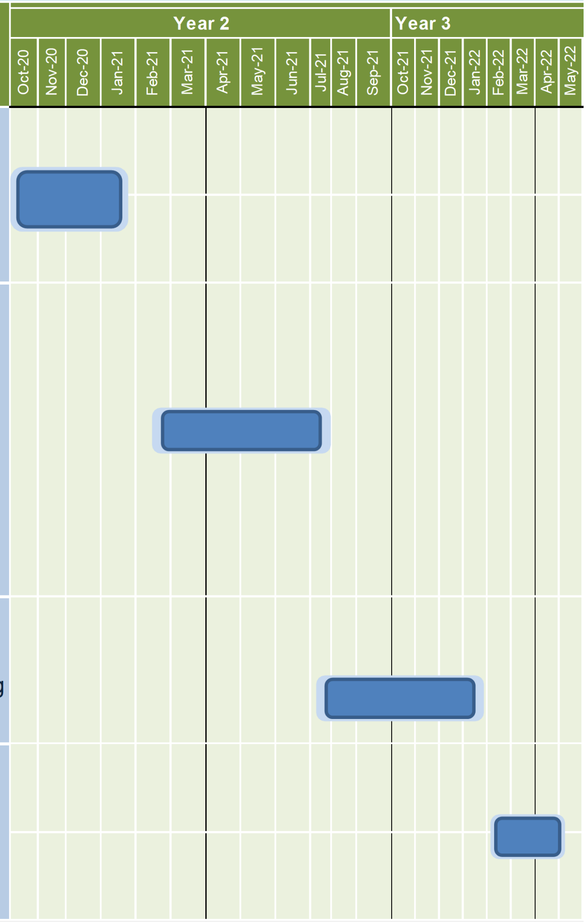
Appendix B - Gantt-Chart to addressing Industrial problem

Table B.1: Gantt-Chart with the Industrial Milestones set for this Research-project

Adaptive Simulation Modelling Using the Digital Twin Paradigm - Industrial Goals																
			Year 1											Year 2		
Scope	**Activities- Targets**	**Milestone**	Nov-19	Dec-19	Jan-20	Feb-20	Mar-20	Apr-20	May-20	Jun-20	Jul-20	Aug-20	Sep-20	Oct-20	Nov-20	Dec-20
Domain-data specialisation	Investigate on the current state of CP modelling including: i. Spatial data generated from the BEASY-based cathodic-protection (CP) model.	domain's data specialisation														
Data Obtaining	Benchmark: 1. Unstructured data (text, images, etc.) analysis tools (e.g. Python, Nvivo, etc.) 2. Algorithms for the following unstructured data extraction steps: i. Corpus analysis ii. Word sense disambiguation iii. Sentence Segmentation iv. Image processing for text extraction v. Word Tokenisation, etc. 3. Techniques for structuring and storing extracted data	Data extracting tool														
	Implementation Implement best-unstructured data extraction algorithms for extracting survey data for Corrosion Model.															
Software Benchmarking	Investigation and testing of Scientific software on: i.Correlating model output and measured data. ii.Running simulation tool (eg: BEASY) within the software iii.Optimisation algorithms (eg: gradient descent, genetic algorithm, etc) iv.Visualisation (3d , 4d data visualisation) v.Data extraction from unstructured/semi-structured data set files (eg: pdf, excel sheet) vi.Integration of optimisation algorithm to simulator-based model.	Experimental platform Standardisation														

Adaptive Simulation Modelling Using the Digital Twin Paradigm - Industrial Goals

Scope	Activities- Targets	Milestone
Model prediction Correlation and Validation	Investigation 1. Investigate techniques for correlating and validating design predictions with survey data. Benchmark: i. Techniques (including validation metrics) for measurement of goodness of fit for continuous Validation of the model prediction.	Techniques for measurement of goodness of fit for validation
Synchronisation of operational online validation to online calibration/ adaptation	Investigate and Benchmark: i. Method for identification of Model's functional parameters (i.e. sensitivity analysis) and feature selection, to select and map the sensitive parameters, data categorisation with modularity for sensitive parameters. ii. Requirement on online simulation model updating facilitating/accelerating techniques. (eg: Surrogate Modelling, Reduced Order Modelling etc) . Implementation: Implementation and validation of the methods/algorithms identified/ developed.	Approach for facilitating of model validating and calibrating artefact, with automation
Online Adaptation	Investigate and Benchmark: i. Algorithms on continuous Model adaptation/re-calibrating methods Addressing Input Parameters Uncertainties that arise with time. Implementation: Implementation with time-dependence model.	Automated Predictive Model real-time updating solver
Final Evaluation	Investigation: i. Deviations from expectations, both quantitative and qualitative will be carefully noted. Explanation: i. Causes of deviation identified. Ii. Consideration of future work.	Comprehensive Solution Provider



Appendix C - Offshore SACP System

C.1 Offshore Structure Description

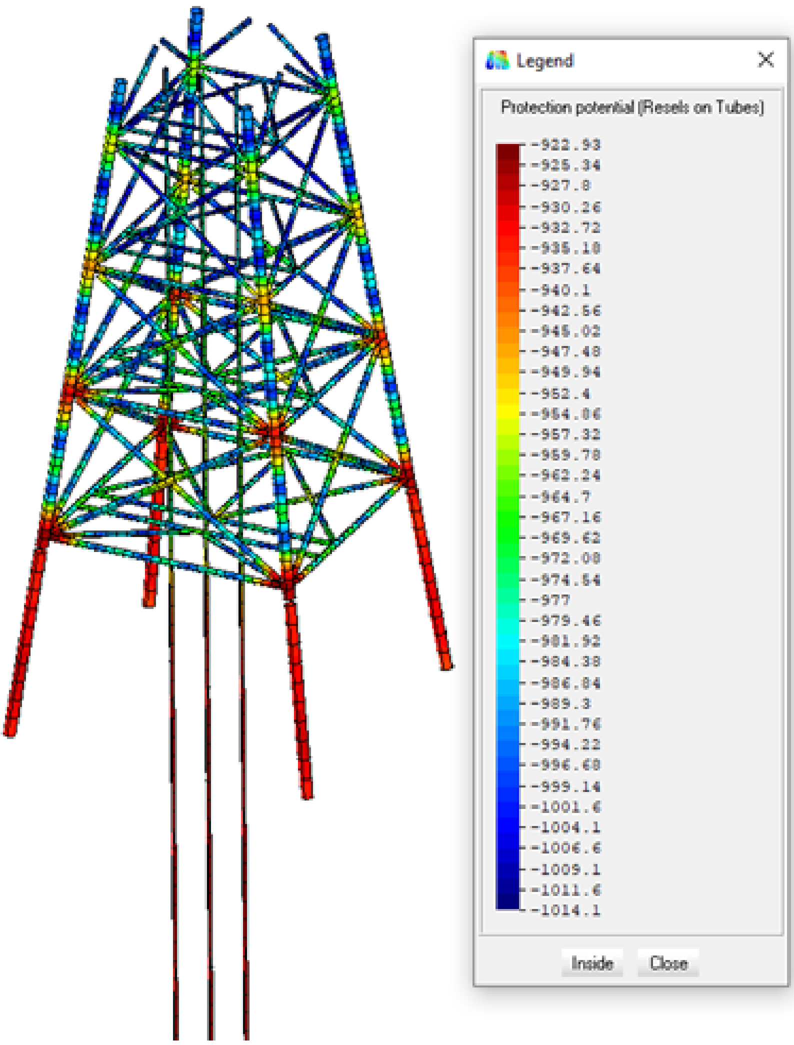


Figure C.1: Real-world Offshore Jacket structure, usually provided with the protection potential on the wet and buried members ranged from -923 to -1014 mV (Ag/AgCl/seawater)

Following are the conditions considered for the offshore structure:

- Water depth 75 m
- Four legs each with a pile underneath it.
- Four elevations: $-10m$, $-30m$, $-50m$, $-73m$,
- 3 conductors and well casings
- Survey points down the legs and down the conductors, at 10m intervals
- 266, stand off, sacrificial (Aluminium-Zinc) anodes

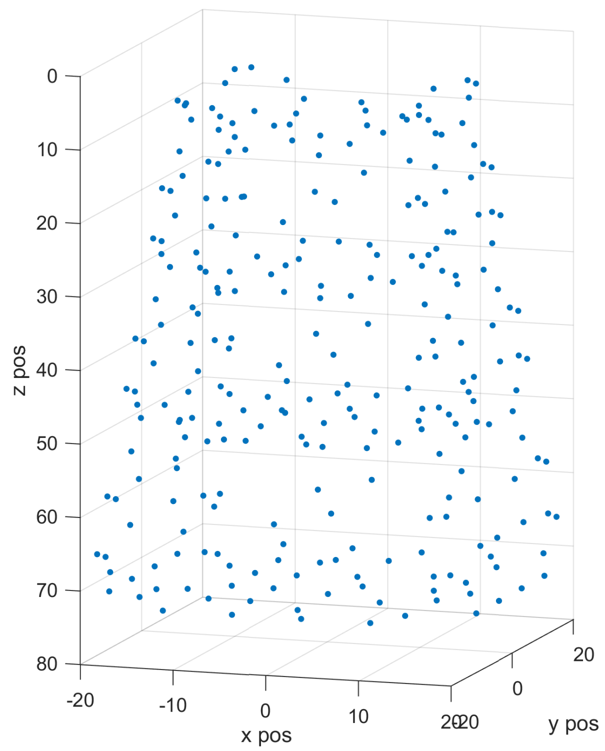


Figure C.2: Plot with the dot-points representing the centroid of the spatial position of the anodes (n =266)

The sacrificial anodes (266 in number) connected to the offshore structure uniformly (Figure C.2) got following properties:

- Length = 3.0m, height = 0.228m, width = 0.228m, mounting bar diameter = 0.1143 m, stand off distance = 0.325m
- Mass = 356 kg, utilisation factor = 0.9, consumption factor = 0.0, density = $2710kg/m^3$
- Electrochemical capacity = 0.003504 kg/mA*year (2500 A*hrs/kg).

The definition for the terms used above are provided below:

1. **Stand off distance:** It is the minimum distance needed to prevent electrical interference or arcing between them.
2. **Utilisation factor:** The ratio of the mass that can be utilised during the cathodic-protection.
3. **Consumption factor:** The ratio of the mass-consumed to the initial total mass.

C.2 CP Model related General Information

C.2.1 Mesh-related

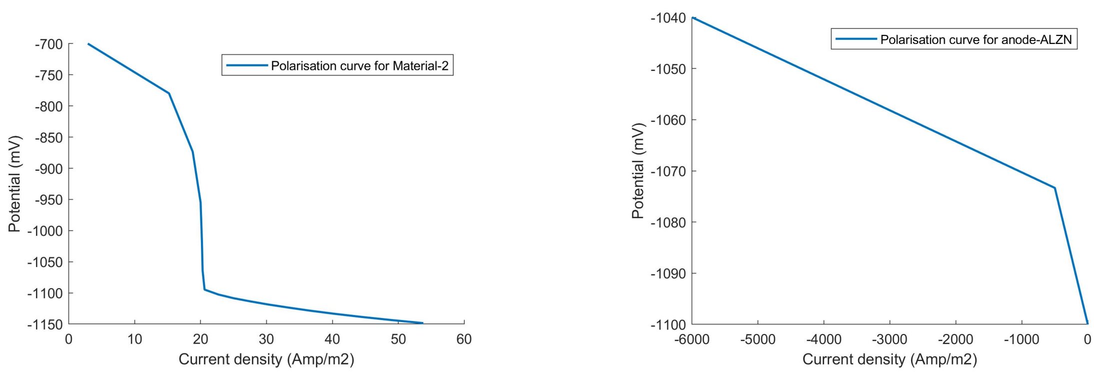
Following are the conditions considered for the CP model discretisation phase using the BEASY tool:

- Number of Elements (Meshes) : 8820
- Number of Mesh-points: 32894
- Number of Zones: 2.

C.2.2 Polarisation Curves

The list of polarisation curves used in this model is as follows:

- **Material 1** : This will be achieved by calibration as considered dynamic due to factors like coating breakdown.
- **Material 2** : Will be considered consistent (constant) as represented by Figure C.2(a).
- **Anode ALZN** : Will be considered consistent as represented by Figure C.2(b).



(a) (b)

Figure C.3: Polarisation curves considered for two materials a) Material 2 related, b) Anode related, during the CP simulation run obtained from design rule.

C.2.3 Conductivity data

- Zone 1 (portion in sea-water) : will be considered uniform but as unknown variable for calibration, otherwise with the value 3.3333 (Siemens/m)
- Zone 2 (portion in sea-bed) : will be considered uniform and known variable with value 0.667 (Siemens/m)

Important Points to be noted about the Model:

1. The geometrical as well as material related data are obtained from the domain expertise (BEASY Ltd.), and also the design rules are considered.
2. The illustration showing the structure (for example Figure C.1 only represents the metal-body components and not the sacrificial anodes, though will be connected to have the CP system.)

C.3 Reference Simulation summarised data

The reference simulation run is corresponding to the model considered during the calibration data generation phase, discussed in Chapter 5 and Case Study I in Chapter 6. The summary of the output from the reference simulation run data is presented below:

Model Current Balance

Summation over non-interface elements:

- Current flowing into model: 1.58543e+006 mA
- Current flowing out of model: 1.58544e+006 mA
- Net Current flowing into model: -9.21036 mA
- Percentage Current Error: 0.0000058

Table C.1: Zone wise Current (mA) results

Zone ID	Current Inwards (mA)	**Current Outwards (mA)**	**Net Current Inwards (mA)**	**Percentage Current Error (%)**
1	1.59E+06	1.59E+06	-1.5346	-9.70E-05
2	21175.1	21164.3	10.7449	0.051

Table C.2: Performance Result over each group categorised during meshing step

Group name	**Type**	**Boundary Condition**	**Area**	**Area x AF**	**Current**			**Protection Potential**		**Anodic Current Density**		**Cathodic Current Density**	
					Anodic	**Cathodic**	**Net**	**Most Positive**	**Most Negative**	**Max**	**Min**	**Max**	**Min**
Anodes_BOTTOM	ANODE	ALZN (PC)	180.918	180.918	401977	0	-401977	-1055.5	-1070.71	3869.73	1777.4	NA	NA
Anodes_TOP	ANODE	ALZN (PC)	570.176	570.176	1.18E+06	0	-1.18E+06	-1020.9	-1073.36	8621.37	1413.5	NA	NA
coated_conductors_diameter_0.600	METAL	CM00 (PC)	11.3097	11.3097	0	25.8479	25.8479	-859.42	-860.156	NA	NA	2.28863	2.2802
coated_legs_diameter_1.800	METAL	CM00 (PC)	45.8629	45.8629	0	108.763	108.763	-863.42	-871.315	NA	NA	2.41637	2.326
coated_members_0.800	METAL	CM00 (PC)	58.792	58.792	0	141.282	141.282	-860.37	-879.64	NA	NA	2.51167	2.2911
conductors_diameter_0.600	METAL	Material 1 (PC)	412.947	412.947	0	73784.5	73784.5	-786.9	-845.326	NA	NA	211.891	151.13
legs_diameter_1.800	METAL	Material 1 (PC)	1674	1674	0	323969	323969	-726.08	-905.423	NA	NA	280.829	109.93
members_diameter_0.400	METAL	Material 1 (PC)	24.1274	24.1274	0	4569.16	4569.16	-787.88	-847.185	NA	NA	214.019	151.79
members_diameter_0.600	METAL	Material 1 (PC)	342.875	342.875	0	80437.5	80437.5	-817.8	-922.595	NA	NA	337.857	180.38
members_diameter_0.800	METAL	Material 1 (PC)	2516.78	2516.78	0	540509	540509	-724.74	-918.175	NA	NA	323.18	109.03
members_diameter_1.000	METAL	Material 1 (PC)	2735.93	2735.93	0	544196	544196	-721.5	-896.647	NA	NA	270.645	106.83
piles_diameter_1.800	METAL	Material 2 (PC)	688.607	688.607	0	9474.38	9474.38	-765.2	-782.167	NA	NA	14.9286	12.417
wells_diameter_0.600	METAL	Material 2 (PC)	566.476	566.476	0	8215.13	8215.13	-772	-814.15	NA	NA	16.406	13.424
Total			9828.8	9828.8	1.59E+06	1.59E+06	-9.21036						

Notes:

Net current = Cathodic current - Anodic current

PC = Polarisation curve
AF = Area factor (=1)

Appendix D - ICCP model related Current Supply

Table D.1: The adjusted normal current density applied to the anodes such that the most positive potential on the tank are $-850mV$.

Linear Anodes	AREA (m^2)	Current (mA)	Normal Current density (mA/m^2)
A1	0.500934	-218	-435.187
A2	0.500934	-81	-161.698
A3	0.500934	-82	-163.803
A4	0.500934	-162.71	-324.813
	Total Current	-544	

Appendix E - Parametric Reduced Order modelling

This section attempts to provide some insights on Parametric Reduced Order Modelling.

E.0.1 Model reduction concept

For this, let's begin with the concept of numerical approximation of the PDEs. During the numerical approximation of the PDE(s), the response data is represented with the basis functions and the coefficient vectors in most situations. The response variables (U) for each snapshot (temporal or spatial) can be represented as :

$$u \approx \sum_{k=1}^{N} \phi^k \alpha_k \tag{E.1}$$

With coefficient vector α and basis functions ϕ of count $k = 1, 2, 3, \ldots N$.

For example, the numerical solution for the heat equation is usually achieved by solving the following PDE form (Morton and Mayers 2005):

$$\frac{\partial u}{\partial t} = \Delta u \tag{E.2}$$

The approximation for the response variables (eg: temperature) from the equation E.2 is obtained as a function of time and space as equation E.3 :

$$u(x,t) \approx \sum_{k=1}^{N} \alpha_k(t) \phi^k(x) \tag{E.3}$$

Reduced-order models from the full-order model are generated by capturing the dominant modes (Liang et al. 2002). In other words, ROM is about reducing the numbers (*N*) of basis functions (ϕ) to represent the response variable. For the above case, the reduced count of basis after truncation let's say is $K <<< N$, and ROM is represented as equation .

$$\hat{u} \approx \sum_{k=1}^{K} \hat{\phi}^k \hat{\alpha}_k \tag{E.4}$$

With coefficient vector $\hat{\alpha}$ and basis functions $\hat{\phi}$ of count $k = 1, 2, 3, \ldots K$ to approximate the response variables ($\hat{u}$).

E.0.2 Methods for ROM building

One of the widely used methods for parametric ROM building is the Proper Orthogonal Decomposition (POD) method (Chatterjee 2000). POD generates the optimally ordered orthonormal basis for a given set of experimental or computational data. POD is mostly achieved with Singular-Value-Decomposition (SVD) method and is also sometimes known so (Liang et al. 2002). Though the generated ROM is supposed to be grey-box, it only requires solutions from the full-order model as snapshots of the solution. The dependency upon only the simulation snapshots generated by varying input parameters offline makes it not rely upon the operator (solver) matrix from the full-order model.

This ability to generate the reduced model using only the system input-output data make SVD algorithm especially suited for industrial applications (Rozza et al. 2018). For this, the data are sampled at a predefined set of surface positions (count 'n') and sample parameter combinations (count 'm') are considered. Generally, the altered parameters can be any combination of material properties and/or boundary conditions. Each u_j is then stored inside a rectangular snapshot matrix X of dimension $n * m$.

$$X = \begin{bmatrix} \vdots & \vdots & \ddots & \vdots \\ u_1 & u_2 & \dots & u_m \\ \vdots & \vdots & \ddots & \vdots \end{bmatrix}, X \epsilon C^{n*m} \tag{E.5}$$

The snapshot is obtained from the numerical solver of the system (FEM or BEM), or from actual empirical data. SVD capitalises on the correlation between the known direct problem and the sought-after solution.

$$C = X^T \cdot X \tag{E.6}$$

$$\tag{E.7}$$

where, C represents the covariance matrix.

Now eigen decomposition of the covariance matrix is performed which is done by SVD and the outcomes result to:

$$C \cdot V = A \cdot V \tag{E.8}$$

where, A represents the diagonal matrix storing the eigen values $\lambda_1, \lambda_2, \dots$ of C. Similarly, V represents the eigenvectors of the covariance matrix C.

Now, based on the eigenvalues (only dominant are kept) of the matrix, the eigenvector matrix can be truncated (cut-off). The resulting POD basis $\hat{\phi}$ referred to as the truncated

POD basis, consists of $K <<< M$ vectors and obtained from $\hat{V}$, i.e., truncation of the eigen-vector matrix (V).

Parametric model are desired in SHM which is why parameterised ROM will be anticipated, and the parameterisation for ROM is usually done with the interpolating function. For this job, an interpolating function such as Radial Basis Functions (RBF) has been used as it has got good approximation and smoothing properties (Buhmann 2000). An adoption of RBF for interpolation together with POD can be found in Rogers et al. (2012). When parameterised, the response variables for the arbitrary parameter set 'k' will be obtained from ROM as,

$$X^a(k) \approx \hat{\phi} B F^a(k) \tag{E.9}$$

where, $X^a(k)$, denotes a snapshot corresponding to an arbitrary parameters vector k.

Appendix F - Analysis for CP surrogate model updating requirement with time

To create the surrogate, the focus will be on two input parameters (CA_0 and CB_0) using the approach detailed in Chapter 7. The surrogate will be built using simulation snapshots taken at time t_0, which are depicted in Figure F.1.

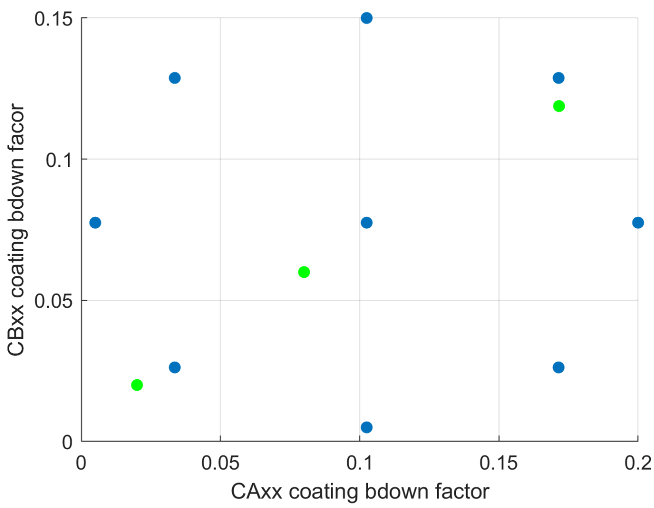


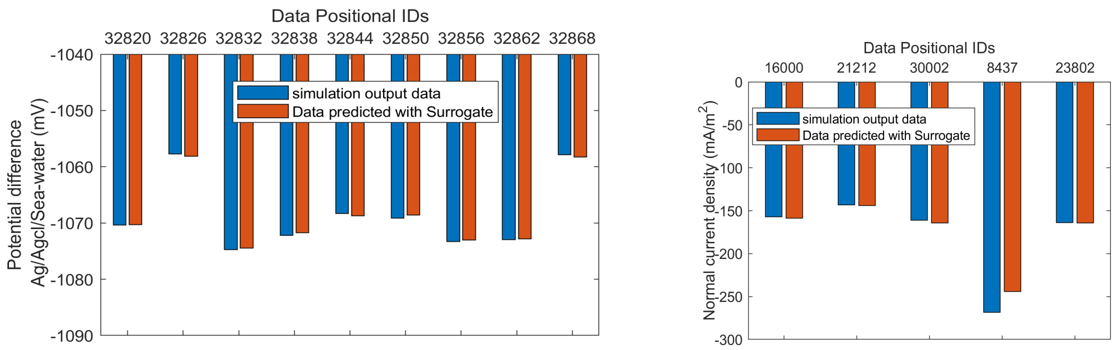
Figure F.1: Sample points from CCD for two selected parameters for surrogate building (in blue), including parameter values for performance testing of the surrogate model (in green)

Next, the 2^{nd}-order polynomial fit approach detailed in Chapter 7 is used to construct the surrogate model based on the corresponding snapshots of input parameters and response data. The surrogate model is specifically built for the initial phase ($t = 0$), where no anodes have been consumed and the average anode consumption factor is equal to 0.

Now, the performance of the surrogate model will be analysed against the against simulation data obtained at different stages of anode consumption. To this end, the time step models for simulation run are generated following the assumed pattern of changing parameters (i.e., CA_t and CB_t). The 3 testing samples points in ascending order of parameter values from Figure F.1, are considered for year 0, 5 and 10. The time-step simulation is used to obtain the anode consumption factor for the given time. Table F.1 presents the parameters value for two different time-step considered and computed average anode consumption state for those corresponding years.

Table F.1: Parameter related values considered and computed for different time-steps.

Years	CA_{xx}	CB_{xx}	**Avr. Anode cons.**
0	0.02	0.02	0
5	0.08	0.06	0.0753
10	0.1716	0.1188	0.2345

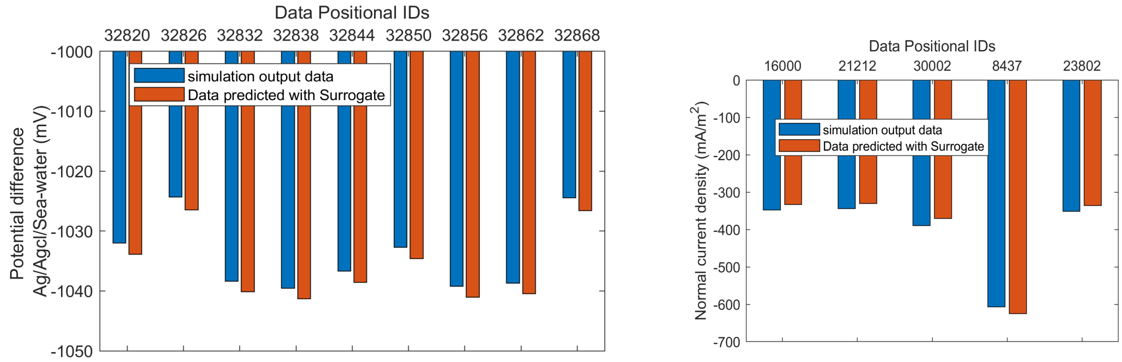


(a) (b)

Figure F.2: Surrogate (built at t = Year 0) performance compared to anode-related data updated simulation model for time (t = Year 0), for 2 different data-types a) Surface Potential, b) Normal Current density.

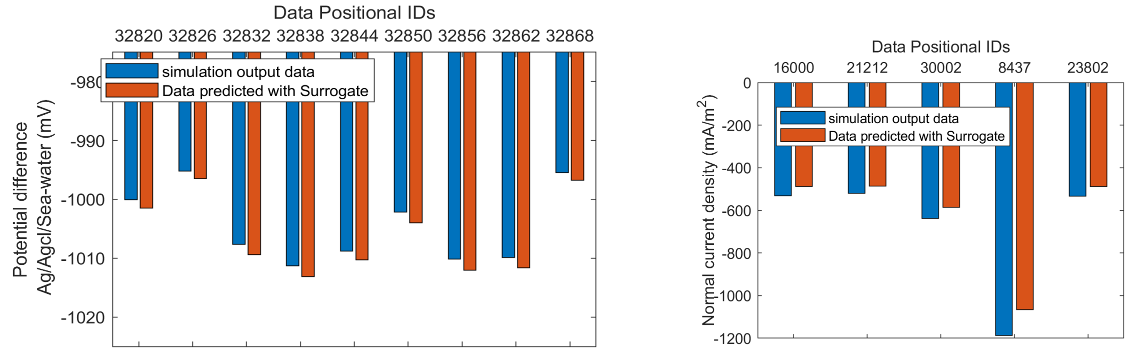
The performance of the surrogate model is evaluated by comparing its predictions with response data from the simulation run (graphical and NMSE measurement). Table F.2 shows the Normalised Mean Square Error (NMSE) between the output of the initial surrogate model, which did not account for anode status, and the output of the simulation that considers changes in anode status over time.

The performance drift of the surrogate over time in this case is primarily due to the unaccounted aspects, i.e., the anode consumption status during surrogate construction. Based on the analysis, one can identify the need to update the surrogate model. Additionally, if the design variables (parameters) of the system have reached the values beyond the the surrogate sample space, the surrogate updating is essentially required.



(a) (b)

Figure F.3: Surrogate (built at t = Year 0) performance compared to anode-related data updated simulation model for time (t = Year 5), for 2 different data-types a) Surface Potential, b) Normal Current density.



(a) (b)

Figure F.4: Surrogate (built at t = Year 0) performance compared to anode-related data updated simulation model for time (t = Year 10), for 2 different data-types a) Surface Potential, b) Normal Current density.

Table F.2: Initial-built Surrogate Performance drift with time.

Years	Avr. Anode cons.	NMSE
0	0	0.0051
5	0.0753	0.0068
10	0.2345	0.0130