DIGITAL TWIN-BASED PROGNOSTICS AND HEALTH MANAGEMENT FOR FATIGUE-PRONE COMPONENTS AND SYSTEMS USING METAMODEL-BASED APPROACHES

HUANG CHAO

PhD

The Hong Kong Polytechnic University

2025

The Hong Kong Polytechnic University

Department of Electrical and Electronic Engineering

Digital Twin-Based Prognostics and Health Management for Fatigue-Prone Components and Systems Using Metamodel-Based Approaches

Huang Chao

A thesis submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy

February 2025

CERTIFICATE OF ORIGINALITY

I hereby declare that this thesis is my own work and that, to the best of my knowledge and belief, it reproduces no material previously published or written, nor material that has been accepted for the award of any other degree or diploma, except where due acknowledgment has been made in the text.

_____(Signed)

HUANG Chao (Name of student)

Abstract

Prognostics and health management (PHM) is an interdisciplinary field that addresses challenges in reliability, availability, maintainability, and safety arising from design, manufacturing, environmental, and operational complexities. The primary objective of PHM is to ensure asset integrity while mitigating risks to mission reliability and safety. Enabled by the industrial internet-of-things and model-based systems engineering, digital twins (DTs) are transforming industries by creating accurate virtual replicas of physical assets. The DT technology significantly enhances PHM implementation by enabling reliable life reliability and health management, reducing costs, and improving system availability. To address challenges in life reliability prediction, metamodeling is used to support DTs. Traditional approaches to life reliability prediction rely on quantifying uncertainties and applying Monte Carlo simulations to finite element models. While effective, these methods are computationally intensive and time-consuming. In contrast, metamodeling improves prediction efficiency without sacrificing accuracy, as demonstrated in case studies involving turbine blisks and fan shafts. With the metamodel-based DT, the DT-PHM architecture enables life reliability prediction without requiring run-to-failure data, supporting cost-effective and availability-driven health management strategies, such as warranty design and predictive maintenance planning. The developed DT-PHM architecture is applicable to the health management of both single-unit components and

multi-unit systems. For multi-unit systems, an opportunistic predictive maintenance has been designed, incorporating economic dependencies to optimize maintenance strategies effectively.

Acknowledgments

First and foremost, I would like to express my deepest gratitude to my chief supervisor, Professor Siqi Bu. As a leading scholar in the field of power system operation and risk assessment, it has been a profound honor to pursue my Ph.D. under his guidance, which has thoughtfully and diligently set me on the path of academic research. His dedication and enthusiasm for scientific research have consistently served as an exemplary model for me to follow. From academic inquiry to personal development, he has provided invaluable guidance in shaping my critical thinking and writing skills. His research integrity, academic mentorship, and unwavering support throughout my Ph.D. journey have been invaluable assets in my life. Additionally, I extend my heartfelt gratitude to my co-supervisors, Professor Edward Chung and Professor Hwa-Yam Tam from The Hong Kong Polytechnic University, for their insightful contributions to my research. In addition, I extended my sincere thanks to my thesis committee members for their valuable time spent reviewing my thesis and providing constructive suggestions for improvement. I sincerely acknowledge the support of the Centre for Advances in Reliability and Safety (CAiRS) for offering me a research associate position, which greatly facilitated my Ph.D. studies. This support was instrumental in making my research possible.

I would also like to extend my heartfelt appreciation to my family, Rainbow Lee, Howard Chan, Justin Kong, Jiawei Duan, Yun Li, Chao Lei, and Qifan Chen, for their unwavering support and patience throughout my Ph.D. journey. Their invaluable advice

and practical assistance, both in my personal life and research, have been instrumental to my success. I will always treasure this remarkable research journey in Hong Kong.

Last but not least, I would like to thank the support from The Hong Kong Polytechnic University.

Table of Contents

Lists of Figures

Lists of Tables

Lists of Abbreviations

ACO	Ant colony optimization
AGAN	As-good-as-new
AK	Adaptive kriging
BiLSTM	Bidirectional long short term memory
BW	Base warranty
CBM	Condition-based maintenance
CDF	Cumulative distribution function
CI	Confidence interval
CM	Corrective maintenance
CNN	Convolutional neural network
DBN	Deep belief network
DoE	Design of experiment
DT	Digital twin
ECDF	Empirical cumulative distribution function
EGCNN	Ensemble generalized constraint neural network
ELM	Extreme learning machine
EW	Extended warranty
FGD	Flue gas desulphurization
FEM	Finite element model
GCNN	Generalized constraint neural network
GIS	Gas-insulated switchgear
GRNN	Generalized regression neural network
GRU	Gated recurrent unit
HCF	High cycle fatigue
IPM	Imperfect preventive maintenance

IU	Improved U-learning
KCNN	Knowledge constraint neural network
LCF	Low cycle fatigue
LHS	Latin hypercube sampling
LLF	Log-likelihood function
LSF	Limit state function
LSTM	Long short term memory
MAE	Mean absolute error
MCS	Monte Carlo simulation
MLE	Maximum likelihood estimation
MR	Minimal repair
NFMW	Non-renewable free minimal-repair warranty
NFRW	Non-renewable free replacement warranty
NRBFNN	Normalized radial basis function neural network
NRMSE	Normalized root mean squared error
O&M	Operational and maintenance
OFV	Objective function value
OM	Opportunistic maintenance
OPdM	Opportunistic predictive maintenance
PCE	Polynomial chaos expansion
PED	Positive economic dependency
PdM	Predictive maintenance
PHM	Prognostics and health management
PM	Preventive maintenance
PoF	Physics-of-failure
PSO	Particle swarm optimization
R2F	Run-to-failure
RSM	Response surface method
RUL	Remaining useful life

SACO-PSO	Sequential ant colony optimization-particle swarm optimization
SVM	Support vector machine

Chapter 1

Introduction

1.1 Background

Prognostics and health management (PHM) is a multidisciplinary field that combines diagnostics and prognostics to evaluate and predict the health of components and systems. The framework supports reliability management, informs the development of maintenance strategies, and minimizes downtime and costs associated with unexpected failures [1]. PHM has been initially developed by the U.S. Department of Defense [1][2][3], with a notable example being its integration into the F-35 Lightning II Joint Strike Fighter program, enabling autonomous logistics support [4][5]. Over time, PHM has been widely adopted across various industrial sectors, such as manufacturing, energy, and electronics [6]. In the manufacturing sector, Nidec and IBM have collaborated to develop PHM services by leveraging data collected from Nidec's machinery [7]. In the energy sector, X. Lei and P. Sandborn have applied PHM to wind turbines, facilitating the prediction of remaining useful life (RUL) and achieving significant reductions in maintenance costs and downtime [8]. In the electronics industry, M.-H. Chang *et al.* have implemented PHM for LED lighting systems in street lighting applications [9]. Similarly, K. Feldman *et al.* have developed a PHM module to monitor the health of an electronic line-replaceable unit in Boeing 737 aircraft systems [10].

The PHM framework consists of five core elements: Data acquisition, condition assessment, diagnostics, prognostics, and decision-making, as shown in Fig. 1.1. Data acquisition involves collecting information related to asset operation, material degradation, and environmental loads affecting components and systems. Condition assessment focuses on evaluating the operational state of components or systems. Diagnostics extracts health indicators from sensor data to provide insights into failure modes, mechanisms, and the extent of damage. Prognostics estimates the RUL of assets within specified confidence intervals (CIs). Decision-making optimizes health management strategies, such as warranty and maintenance planning, based on health assessments and predictions [1]. While data acquisition, condition assessment, and diagnostics are foundational components, this thesis primarily focuses on prognostics and decision-making. These two processes are highly dependent on the quality and quantity of the run-to-failure (R2F) data. Insufficient or inadequate R2F data can significantly undermine their effectiveness.

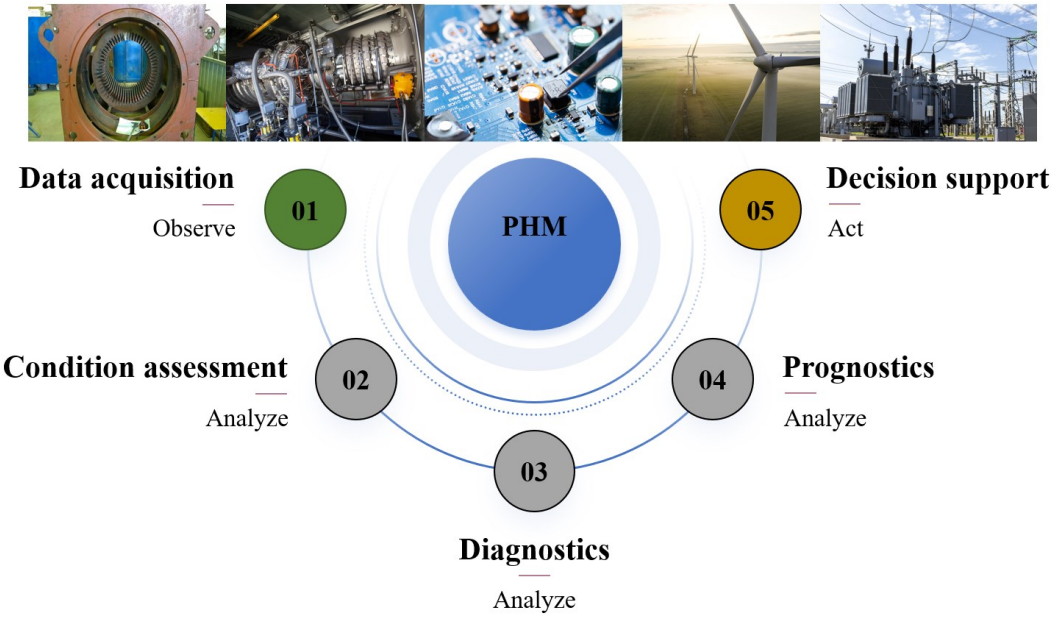


Fig. 1.1 The PHM framework.

Despite advantages of the PHM, implementing PHM in industrial settings poses three major data-related challenges: (i) Lack of critical physical parameters, such as stresses and strains, due to sensor limitations, (ii) Poor data quality, particularly the scarcity of R2F data necessary for RUL estimation, and (iii) Limited data availability resulting from the narrow operational conditions of in-service assets.

The concept of the digital twin (DT) has gained significant attention due to its ability to address data limitations in PHM through advanced modeling and simulation techniques. The DT has been first proposed by M. Grieves via creating virtual replicas for physical assets [11]. The DT integrates physical assets and virtual replicas, which interact and co-evolve through bidirectional communication [13][15]. By representing physical assets through virtual replicas, the DT enables prognosis, warranty design, and predictive maintenance (PdM) planning, thereby enhancing enterprise productivity and efficiency [14][15]. M. Grieves has outlined a three-dimensional DT architecture: (i) physical assets in real space, (ii) virtual replicas in virtual space, and (iii) Data connections linking physical assets and virtual replicas [11]. To address challenges about increasing data and service demands across industries, F. Tao *et al.* have proposed a five-dimensional DT architecture by adding DT data and services, as illustrated in Fig. 1.2.

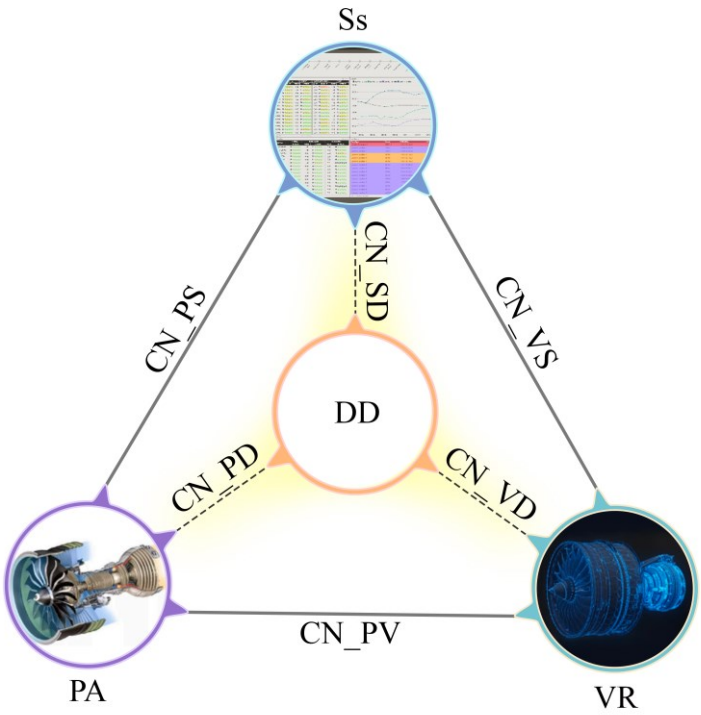


Fig. 1.2 The five-dimensional DT architecture.

Building upon the five-dimensional DT architecture, the proposed DT-PHM framework is also five-dimensional, encompassing physical assets, virtual replicas, services, data, and connection, as illustrated in Fig. 1.3. The proposed DT focuses on: (1) Real-time data exchange between physical assets and virtual replicas for dynamic condition monitoring, (2) Simulation of unobservable parameters (e.g., stresses and strains) for enhanced damage prediction, and (3) Probabilistic prognosis that incorporates uncertainties from manufacturing errors, material variations, and operational disturbances [22][23][24][25]. By adopting this specific formulation of the DT, the thesis effectively narrows its scope to focus on applications oriented toward PHM. Z. Liu *et al.* have used DTs for probabilistic prediction of high cycle fatigue (HCF) life in wind turbine blades [26]. Y. Huang *et al.* have developed an uncertainty-aware DT to predict low cycle fatigue (LCF) life reliability of aeroengine turbine rotors [27]. V. Zaccaria *et al.* have utilized DTs to monitor aircraft engine performance under

wear conditions [14].

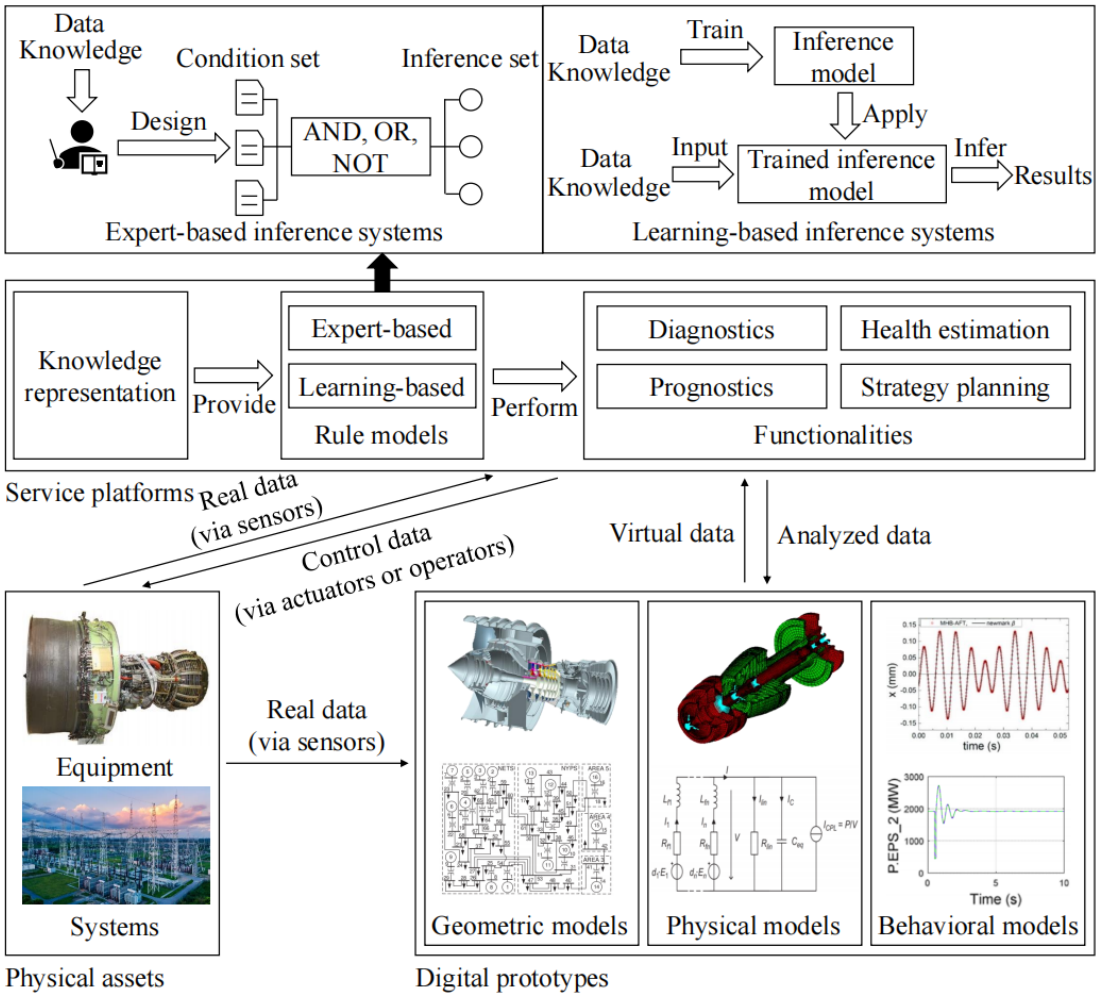


Fig. 1.3 The DT-PHM framework.

The DT-based probabilistic prognosis supports both the design and service phases of the asset lifecycle as shown in Fig. 1.4 [28]. In the design phase, the DT assists in structural design or material selection using reliability-based design optimization methods. In the service phase, the DT can simulate future usage scenarios to predict life distribution, enabling optimal health management strategy design [29][30]. Traditionally, health management strategy design relies on assumed life distributions due to the lack of R2F data. The DT can solve this issue with the life reliability prediction function, allowing practitioners evaluate, compare, and validate health management strategies such as warranty and PdM services without costly physical

accelerated life tests. Furthermore, compared to assumed life-based health management strategy design, the proposed DT-PHM framework enables the evaluation of the impact of operational conditions on health management strategies.

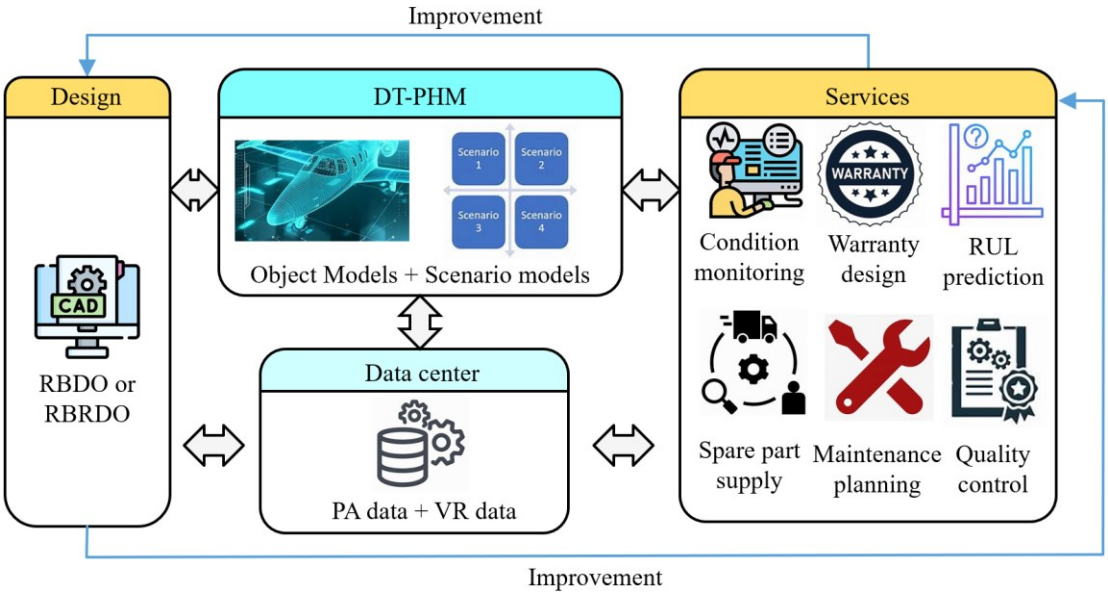


Fig. 1.4 The application of the DT-PHM in design and service phases.

1.2 Research Motivations

DT-PHM offers a transformative framework to enhance the reliability, availability, and cost-efficiency of complex components and systems. Unlike traditional data-driven PHM, which heavily depends on extensive field R2F data, DT-PHM provides solutions in data-scarce scenarios by creating virtual replicas of physical assets. These virtual replicas enable the monitoring of loads, operational conditions, and historical information, facilitating simulations, life predictions, and optimized health management strategies such as maintenance design.

However, despite its potential, several research gaps in existing literature hinder the effective application of DT-PHM:

1. Life reliability prediction

Accurate life reliability prediction is critical for cost-efficient health management but faces several challenges:

(1) Field R2F data is often scarce due to limited sensor availability and the early replacement of components, making life prediction difficult, particularly for components like steel with cumulative, hard-to-measure damage.

(2) Accelerated life tests can provide probabilistic life estimations but are expensive, especially for multi-unit systems.

(3) FEM-based DTs, while effective for simulating degradation, are computationally expensive due to nonlinear iterations, heterogeneous uncertainties (e.g., material variations, load fluctuations), and stochastic factors. These challenges necessitate the development of metamodel-based DTs, which approximate asset life with significantly reduced computational cost while maintaining prediction accuracy.

2. Warranty design

Warranty design is a key health management task but suffers from the following limitations:

(1) New assets often lack sufficiently reliability data, making it difficult to predict warranty claims.

(2) Traditional warranty designs rely on assumed life distributions and fail to reflect the dynamic operational conditions of products.

(3) Operational reliability constraints are often overlooked in warranty planning.

Addressing these gaps requires DT-based warranty design approaches that simulate degradation processes, predict warranty claims under diverse conditions, and incorporate imperfect preventive maintenance (IPM) strategies to sustain operational reliability.

3. Single-unit predictive maintenance scheduling

PdM strategies aim to reduce costs and improve availability, but current approaches face the following challenges:

(1) Traditional maintenance strategies rely on static scheduling, leading to inefficiencies and higher costs.

(2) Existing PdM strategies often prioritize cost minimization without adequately addressing availability concerns.

The integration of metamodel-based DTs into PdM scheduling enables real-time condition assessment and degradation simulations, allowing dynamic maintenance adjustments to optimize cost and availability.

4. Multi-unit predictive maintenance scheduling

PdM for multi-unit systems introduces additional complexities due to interdependencies among units. Current research gaps include:

(1) Single-unit PdM policies often fail to account for system-level interactions, leading to cost inefficiencies.

(2) Fixed maintenance thresholds in prior studies limit flexibility and efficiency.

DT-based opportunistic PdM (OPdM) models address these gaps by optimizing system reliability predictions and incorporating economic dependencies, enabling cost-

effective maintenance strategies for multi-unit systems.

By addressing these research gaps, this thesis aims to advance DT-PHM framework and its applications in life reliability prediction, warranty design, and PdM scheduling for both single-unit and multi-unit systems.

1.3 Primary Contributions

The proposed DT-PHM framework tackles the identified research gaps through the following key contributions:

1. Life reliability prediction. Metamodel-based DTs enhance life reliability prediction by overcoming the computational limitations of FEM-based DTs. The metamodel-based DT enables real-time damage monitoring and probabilistic life predictions, providing a cost-efficient and accurate solution for reliability estimation.
2. Warranty design. A DT-based warranty design approach enables the simulation of life reliability for new products, even with limited field R2F data. This method optimizes warranty costs and policies while incorporating IPM strategies to ensure operational reliability.
3. Single-unit PdM scheduling. The DT-PdM framework allows dynamic and adaptive maintenance scheduling for individual components, reducing costs and downtime by predicting failure times and analyzing operational variability.
4. Multi-unit PdM scheduling. At the system level, DTs model unit interdependencies and optimize opportunistic maintenance thresholds. The

proposed OPdM model reduces maintenance costs and downtime by leveraging shared setup costs.

1.4 Thesis Layout

This thesis investigates DT-based probabilistic life prediction and its applications in warranty design and PdM scheduling for complex components and systems. Chapter 2 provides a review of the literature on DTs for life reliability prediction, warranty service design, and PdM strategy scheduling. Chapter 3 highlights metamodeling as a key enabler for DT construction, emphasizing its role in efficient and accurate probabilistic life prediction. Chapter 4 examines the application of DTs to warranty design, while Chapter 5 focuses on PdM optimization using DTs, with an emphasis on cost minimization. Chapter 6 addresses DT-based PdM scheduling, incorporating hybrid performance metrics that consider both cost and system availability, rather than focusing solely on cost. Chapter 7 examines opportunistic PdM (OPdM) for multi-unit systems, highlighting how DTs leverage component interdependencies and opportunistic thresholds to reduce costs and minimize downtime. Finally, Chapter 8 concludes by summarizing the key findings and proposing future research directions for the DT-PHM framework. The organizational structure of this thesis is outlined in Fig. 1.5 to aid in understanding its layout.

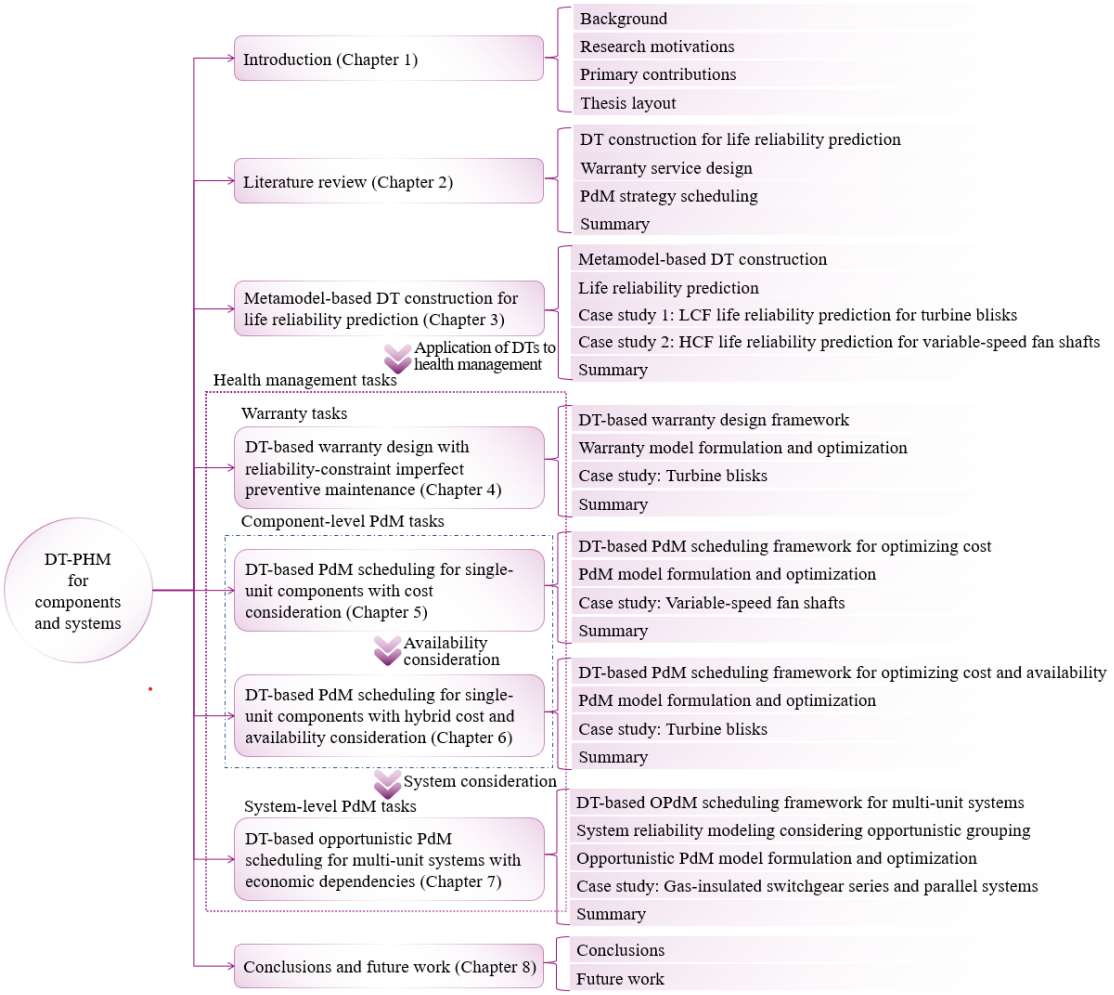


Fig. 1.5 Thesis layout.

Chapter 2

Literature Review

This chapter presents a comprehensive review of the DT construction for probabilistic life prediction, warranty service design, and PdM strategy scheduling, establishing the foundational knowledge for this thesis.

2.1 Digital Twin Construction for Life Reliability Prediction

Traditionally, asset life has been estimated using field R2F data combined with data-driven modeling approaches. For instance, J. Zeng and Z. Liang have proposed a deep Gaussian process model to predict the life distribution of turbofans, while J. Lee and M. Mitici have utilized a convolution neural network with Monte Carlo dropout for dynamic probabilistic life prediction [31][32]. Similarly, C. Chen *et al.* have employed a 1D- convolution neural network-BiLSTM-KDE model to predict the probability density function (PDF) of battery degradation states for uncertainty management, validating its effectiveness using NASA battery datasets [33]. However, in industrial applications, R2F data is often unavailable because assets are frequently subjected to proactive repairs before failure, resulting in a lack of R2F data. Additionally, practical sensor implementation may not always be feasible, and the monitoring data collected can have limited applicability. Although accelerated life tests conducted in laboratory settings can generate R2F data, probabilistic life tests remain prohibitively expensive.

DTs offer an advanced alternative by simulating life and degradation trends to

eliminate the need for costly probabilistic life tests in the lab. The concept of the DT has been first introduced by M. Grieves in 2003 during his Product Lifecycle Management course at the University of Michigan [11]. M. Grieves has proposed the creation of virtual replicas for physical assets. Although various definitions have emerged, NASA's widely accepted definition describes the DT as: *Digital twin is an integrated multi-physics, multi-scale, probabilistic simulation of an as-built vehicle or system that uses the best available physical models, sensor updates, fleet history, etc., to mirror the life of its corresponding flying twin* [12]. The DT integrates physical assets and virtual replicas, which interact and co-evolve through bidirectional communication [13][15]. By representing physical assets through virtual replicas, the DT enables prognosis, warranty design, and maintenance planning, thereby enhancing enterprise productivity and efficiency [14][15]. When integrated with digital representations of operators, it can also support maintenance training. Unlike traditional simulation platforms, DTs serve as dynamic tools applicable in both the design and service phases, offering real-time analysis and management. Recognized as a top technological trend by Gartner, DT has ranked fifth in 2017, fourth in both 2018 and 2019, and has been listed as an emerging technology for the next 5-10 years in 2020 [15][16]. 5-10 years. Lockheed Martin has identified DT as one of six game-changing technologies in defense [17].

Initially adopted by NASA and AFRL for military and aerospace applications, DT has proven its value in these domains [11]. In 2012, NASA has created virtual replicas of spacecraft to perform multi-scenario simulations, enabling the identification and

mitigation of potential issues before they have occurred [18]. The DT concept has since expanded to other industries. Leading enterprises like GE, Dassault, and Tesla have pioneered its adoption. GE Digital has also introduced its Digital Twin software, which leverages data-driven models for asset modeling, simulation, real-time visualization, and predictive analysis [19]. Dassault 3D Experience Platform has been used to create 'Digital Twin Singapore' for urban planning and services, as well as the 'Living Heart Project' to model the human heart [16][20]. Tesla has aimed to develop a DT for each electric vehicle to enable real-time data transfer between vehicles and manufacturing plants [17].

In 2014, M. Grieves has published a white paper on DTs, outlining a three-dimensional architecture: (i) Physical assets in real space, (ii) Virtual replicas in virtual space, and (iii) Data connections linking physical assets and virtual replicas [11]. While this three-dimensional architecture remains widely used, expanding application fields have created increasing data and service demands across industries. To address this, F. Tao et al. have extended Grieves model by adding DT data and services and proposing a five-dimensional DT architecture including physical assets, virtual replicas, services, DT data, and connections. The five-dimensional DT architecture is illustrated in Fig. 1.2.

In the DT architecture, physical assets form the foundation and are classified into three levels: unit-level, system-level, and system-of-systems-level. Virtual replicas are central to the DT, requiring high-fidelity models that accurately replicate the geometries, properties, behaviors, and rules of the physical assets [17]. The virtual replicas can be

modeled as (G_v, P_v, B_v, R_v), where G_v, P_v, B_v, and R_v stand for the geometry model, physics model, behavior model, and rule model. G_v is constructed as a three-dimensional solid model using CAD software, P_v simulates physical properties of the physical asset through FEMs, B_v describes the behaviour of the physical asset as influenced by driving or disturbing factors, while R_v includes rules of constraints, associations, and deductions. With services, the DT provides users with application services related to simulation, optimization, monitoring, diagnosis, prognosis, and health management, etc. DT data is a critical driver of the DT. DT data includes five parts $(D_p, D_v, D_s, D_k, D_f)$,where D_p is the data from the physical asset, D_v is the data from the virtual replica, D_s is the data from the services, D_k is the domain knowledge, and D_f is the fused data of D_p, D_v, D_s, and D_k. Connection includes six parts $(\mathrm{CN\_SD}, \mathrm{CN\_PD}, \mathrm{CN\_VD}, \mathrm{CN\_PS}, \mathrm{CN\_VS}, \mathrm{CN\_PV})$, where $\mathrm{CN\_SD}$, $\mathrm{CN\_PD}$, $\mathrm{CN\_VD}$, $\mathrm{CN\_PS}$, $\mathrm{CN\_VS}$, and $\mathrm{CN\_PV}$ denote the connection between services and DT data, physical assets and DT data, virtual replicas and DT data, physical assets and services, virtual replicas and services, physical assets and virtual replicas, respectively [21].

In the five-dimensional DT architecture, virtual replicas are critical, incorporating material science, physics, system theory, and control science, all of which are essential for DT modelling. Virtual replicas involve four steps: geometric, physical, behavioral, and rule modeling as described bellows.

1. Geometric modeling defines shapes and assembly relationships of physical assets, ranging from wireframe to solid models, with tools like AutoCAD, CATIA, and SolidWorks commonly used in industrial machinery [33].

2. Physical modelling focuses on FEMs that govern physical assets using tools such as ANSYS, ABAQUS, and COMSOL, enabling the representation of physical properties and constraints. Y. Wei *et al.* have developed a FEM-based real-time wear simulation to evaluate performance degradation in CNC machine tools using DT models [34].
3. Behavioral modeling depicts dynamic responses of physical assets to internal and external factors. By integrating kinematic and dynamic analysis, behaviors are represented through differential equations, allowing static physical models to exhibit dynamic behaviors.
4. Rule modeling enhances DT intelligence by extracting and evolving rules from historical data and expert knowledge. Rule modeling involves processes such as rule extraction, description, association, and evolution, with machine learning playing a key role in improving virtual replica performance by reconstructing knowledge bases. For example, the motion control system of CNC machine tools can be designed using CoDeSys, a soft PLC platform [17]. Industrial platforms like PTC ThingWorx can automatically learn operating patterns of pumps, while academic efforts, such as those by Lermer and Reich, have used fuzzy set theory to develop DT models based on production data [17][35].

FEMs are widely utilized for the geometric, physical, and behavioral modeling of physical assets to facilitate life prediction. However, due to uncertainties in manufacturing geometries, material properties, load variations, and other uncontrollable random factors, probabilistic FEM simulations are essential to account

for these uncertainties in life reliability analysis as even small uncertainties can significantly affect the life prediction. However, FEM-based probabilistic simulations are computationally expensive, especially for large-scale tasks, due to thousands of nonlinear iterative calculations.

To tackle this challenge, metamodeling approaches, also referred to as surrogate modeling, have been developed as efficient alternatives to FEM for probabilistic life prediction. Metamodel-based rule modeling improves the efficiency of life reliability analysis by considering heterogeneous uncertainties, such as manufacturing geometries, material properties, and load variations. Metamodeling enables the estimation of unobservable quantities, such as fatigue life, from observable data, supporting damage prediction and decision-making [36]. Various design-of-experiments (DoE) strategies have been employed to establish and adapt metamodels for probabilistic life prediction, including support vector machines (SVM), polynomial chaos expansion (PCE), Kriging, and artificial neural networks, as outlined in Table 2.1 [25][37][38][39]. Metamodel-based DTs differ from PoF-based PHM. While PoF models are grounded in a detailed understanding of physical and material-level failure mechanisms, metamodel-based DTs employ data-driven or surrogate approaches, such as machine learning, to approximate system behavior. By prioritizing computational efficiency, metamodels are particularly suited for applications where high-fidelity physics-based models are computationally infeasible.

Table 2.1 Concepts, characteristics, and reliability prediction applications of metamodels.

Metamodels	Concepts	Characteristics	Applications	References
PCE	PCE represents random variables as polynomial functions, enabling efficient statistical computation in uncertain systems.	Although the PCE is computationally efficient for uncertainty quantification, it can become computationally demanding when dealing with high-dimensional systems or the underlying relationships are highly nonlinear.	Wind turbine (input: wind direction, wind speed, turbulence, and wind shear; output: HCF life)	[40]
SVM	SVM can find an optimal hyperplane in a high-dimensional space by fitting a nonlinear decision boundary using kernel functions or maximizing the margin between different classes.	SVM can handle high-dimensional data and are effective for modeling nonlinear relationships using kernel functions. However, SVM can be sensitive to the choice of kernels, and their training can be computationally demanding for large datasets.	Extrusion container (input: Material, output: fatigue life)	[41]
Kriging	Kriging is a geostatistical interpolation technique that estimates unknown values at unsampled locations by incorporating spatial correlation information from nearby observations, providing a spatially continuous representation of the variable of interest.	Kriging is effective for modeling nonlinear relationships and handling noisy data. However, it can become computationally expensive for large datasets.	Offshore wind turbine (input: wind speed; output: high cycle fatigue life)	[42]
			Welded joints (input: fatigue load; output: crack size)	[43]
			Single-edge notched specimens (input: tension loads; output: crack size)	[44]
Artificial neural network	Artificial neural network can approximate complex functions with multiple hidden layers.	Artificial neural network is capable of capturing nonlinear relationships. However, artificial neural network is computationally intensive to train.	Turbine discs (input: temperature and rotary speed; output: low cycle fatigue life)	[45]
			Wind turbine tower (input: wind speed; output: cumulative fatigue damage)	[46]
			Cylinder head (input: assembly loads, gas force loads, and temperature; output: cumulative fatigue damage)	[47]

The metamodeling workflow for probabilistic life prediction typically involves two stages. First, offline training of the metamodel is conducted using support points and output responses generated from FEMs and physics-of-failure (PoF) models. Next, the trained metamodel is deployed for online monitoring of life consumption ratios and remaining useful life (RUL) prediction in in-service assets, as illustrated in Fig. 2.1. Metamodel-based DTs have demonstrated significant potential in damage and life prediction, especially in scenarios lacking R2F data. For example, D'Urso et al. have

developed a dynamic reliability DT to generate synthetic training data [48], while X. Lai et al. proposed a metamodel-based DT for structural health monitoring of aircraft wings [39]. Their modeling workflow includes collecting measurement and computational data, enhancing simulation accuracy using metamodels, applying a rainflow counting algorithm to assess degradation states, and creating a graphical user interface for visualization. Luo et al. have combined FEM-based wear models with metamodels to predict the RUL of CNC machine tools [49], and M. Xiong et al. have mapped aeroengine simulation data to metamodel parameters, enabling life prediction across various scenarios [50]. F. Tao et al. further detailed the implementation of metamodel-based DTs for PHM in wind turbines [51].

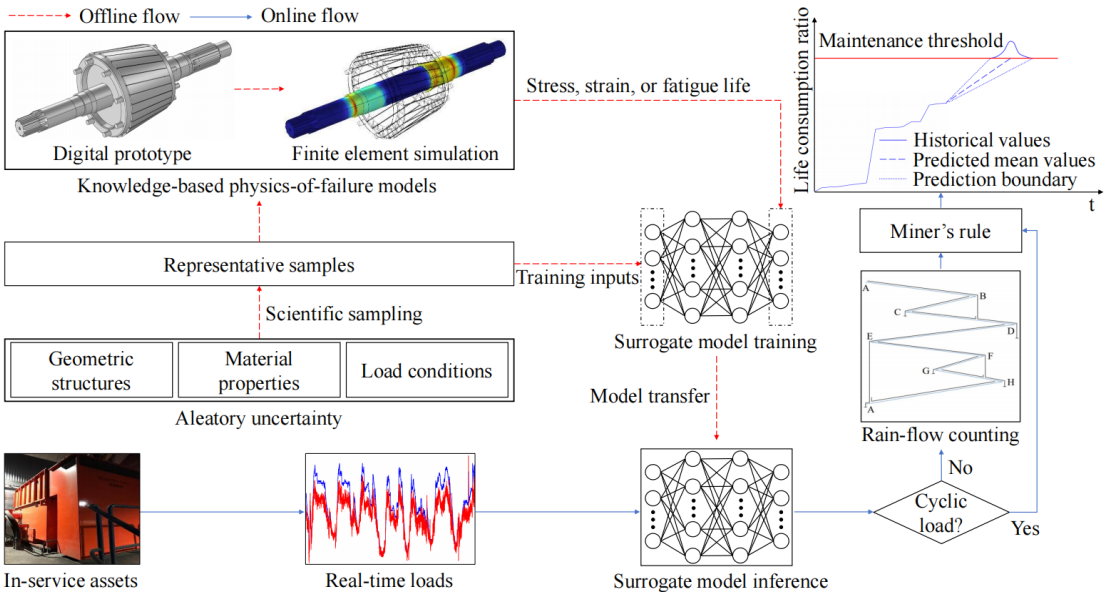


Fig. 2.1 The schematic diagram of the metamodel-based DT for life consumption monitoring.

Beyond degradation failures, shock failures must also be considered, as multiple failure modes—including degradation and shock failures—can occur simultaneously [52]. While degradation failures result from mechanical stresses such as wear, erosion, and fatigue [53], shock failures are caused by random events and can be modeled using

discrete probability functions, such as geometric distributions or Poisson processes. These failure modes can be identified using fault tree analysis, and the reliability engineering community has increasingly focused on modeling multi-failure behaviors for accurate reliability prediction. To address this, competing failure models have been developed, integrating degradation and random shock models to assess failure risks and optimize maintenance strategies [54][55][56]. For instance, Caballe et al. have proposed a maintenance model for systems experiencing degradation and external shocks [57], while Wang et al. utilized a competing failure model to optimize inspection intervals and shock thresholds [58][59]. These models, combined with metamodel-based DTs, offer a robust framework for life reliability prediction and maintenance optimization, enabling efficient and accurate assessment of asset health under complex multi-failure scenarios.

2.2 Warranty Service Design

Although PHM has traditionally focused on low-volume critical systems, there is a significant opportunity to apply it in warranty analysis for high-volume commercial products. Specifically: (1) Scalability of PHM/RUL Techniques: The principles of PHM/RUL, such as failure prediction and maintenance optimization, are applicable to high-volume systems. Adapting these methods can reduce warranty costs and enhance customer satisfaction. (2) Critical Subsystems in High-Volume Applications: High-volume products often include critical subsystems (e.g., EV batteries) where PHM/RUL techniques are essential, aligning warranty analysis with PHM/RUL

methodologies. (3) Economic Benefits: Implementing PHM/RUL in high-volume systems enables proactive failure management, reducing warranty claims and repair costs while demonstrating the value of these techniques beyond low-volume domains.

A warranty is a typical health management service offered by manufacturers for products, regardless of whether they are repairable or replaceable at the point of sale. A repairable asset refers to a product that, in the event of failure, can be restored to its functional condition through repair actions, whereas a replaceable asset is a product that is replaced entirely upon failure. The warranty is a contractual agreement between the manufacturer and the customer, requiring the manufacturer to resolve any failures that arise in the specified contract period. In the event of an asset's premature failure, it outlines the manufacturer's liability [60]. The warranty serves as both protective measures and promotional tools. As a protective measure, manufacturers repair failures during the warranty period, incurring costs ranging from 2% to 15% of revenues, depending on warranty terms and product reliability [61]. Warranties can also protect against unreasonable claims. Simultaneously, as a promotional tool, an appealing warranty policy is often presented to signify product quality and enhance market sales [62]. Consequently, warranties have emerged as widely utilized marketing tools in today's highly competitive marketplace. For example, the warranty period for airplanes has increased from 1,000 flight hours in the early 1980s to between 5,000 and 10,000 flight hours by 2013 [63]. Although warranties can serve as a tool to enhance asset sales, they also pose a potential cost to manufacturers. For instance, Apple has incurred warranty claims amounting to US$ 1.25 billion in a single quarter during the latter half

of 2015 [64]. Similarly, in 2016, automakers worldwide have faced warranty claim expenses totaling US$ 48.0 billion [65].

Warranties are categorized by sales method into base warranties (BWs) and extended warranties (EWs). A BW is a fundamental aspect of asset sales, with its cost incorporated into the sale price. A longer BW can increase customer willingness to purchase a product but may also lead to significant costs for the manufacturer. To balance customer appeal with warranty expenses, careful consideration must be given to designing the BW duration [66]. Alongside the warranty period, pricing is also recognized as a powerful marketing strategy. Therefore, manufacturers must thoughtfully determine both the selling price and the BW duration to achieve an ideal balance between profitability and market share [67]. Glickman and Berger have modeled demand as a log-linear function of price and BW duration, providing a basis for price-BW optimization [68]. Lin and Shue have identified optimal selling prices and BW lengths for various product life distributions [69], while Zhou et al. have proposed profit maximization models for dynamic BW adjustments during the product lifecycle [70]. Huang et al. have identified product reliability, BW duration, and product price as three key decision variables for maximizing the manufacturer's profit [71]. Some researchers have incorporated production rates or spare part inventory into the BW optimization model. Wu et al. have determined the product price, BW period, and production rate via maximizing the profit [72]. Kim and Park have optimized the product price, BW duration, and spare part inventory in an optimal way [73]. Different from BWs, EWs are optional purchases that can extend coverage beyond the BW period.

EW can meet customer demands for greater protection and generate significant manufacturer profits, especially given declining product profit margins. For example, in 2003, EWs have delivered profit margins of 50–60% for Best Buy and Circuit City, nearly 18 times higher than those on products [67][74][75].

For EWs, decision variables also include contract length and price. Bian et al. and Wang et al. have optimized EW periods and pricing, with Wang et al. proposing an EW menu with multiple options using a multinomial logit choice model [76]. Joint optimization of BW and EW is also possible. Wang et al. have considered preventive maintenance during BW and EW periods, proposing models to optimize product price, EW price, and warranty durations. Mitra et al. have developed a model to jointly optimize product price, BW and EW periods, and EW price, while Afsahi and Safiee have proposed a stochastic simulation-optimization model integrating warranty lengths, pricing, repair strategies, and spare part inventory management [76]. In practice, warranty modeling requires consideration of market scenarios. In monopolized markets, manufacturers have the autonomy to set warranty price, whereas in competitive markets, warranty price is influenced by market forces. Shang et al. have proposed a condition-based warranty policy, optimizing the BW period, price, and replacement threshold in both monopolized and competitive markets [30]. Fang has developed a non-cooperative game model to determine the optimal price and warranty period in a competitive market [77]. The warranty can be further classified by renewal type into non-renewable and renewable warranties, and by cost-sharing into free-replacement warranties (FRWs), where maintenance costs are fully covered, and pro-rata warranties, where costs can be

shared [78].

Many manufacturers provide preventive maintenance (PM) during warranty periods to ensure asset reliability, reduce operating costs, and improve customer satisfaction. Yeh et al. have found that PM can significantly reduce manufacturers' costs via failure rate reduction models [79]. In practice, it is common to associate PM with a warranty when the maintenance requires specialized skills, such as for aircraft and automobiles [80]. For example, Nissan has launched a 6615 program in Taiwan, which offers warranty contracts for six years or 150,000 kilometers, contingent upon PM every six months or 10,000 kilometers [81]. Periodic PM is commonly implemented since it simplifies asset management for both manufacturers and customers [82]. A key decision for manufacturers offering PM services is to determine the most effective PM interval. While frequent PM increases costs, it lowers the likelihood of product failures, making it crucial to balance PM frequency and warranty costs for an efficient warranty strategy [83]. The impact of a maintenance action on asset reliability can be perfect, minimal, or imperfect. In many practical situations, an imperfect PM policy is utilized, restoring the asset to a condition that falls between minimal repair and perfect repair. Considering maintenance costs and time, manufacturers can preset the rejuvenation level. Nakagawa and Mizutani have summarized IPM for a finite-horizon maintenance problem [84]. Extensive research has explored optimal PM policies integrated with warranty strategies. For a comprehensive review of maintenance models under warranty, see Shafiee and Chukova or Park and Pham [63][85]. Shafiee et al. have examined optimal PM-integrated warranty strategies by determining key maintenance

policy variables [86]. Yeh and Fang have optimized sales price, BW duration, and PM schedules for products under NFRWs [87]. Su and Wang have proposed customized PM strategies for warranties, and have concluded that offering customized PM can reduce the warranty cost of the manufacturer [88]. Tao and Zhang have determined the optimal length of warranty period for assets subject to PM, such that the manufacturer's profit can be maximized [89].

Warranty costs are higher for assets with low reliability and lower for those with high reliability. Therefore, manufacturers should evaluate the deterioration process of their assets to accurately estimate warranty costs [74]. However, during the design phase, prior to an asset's market introduction, manufacturers often encounter difficulties in precisely estimating its lifespan, which directly influences the determination of warranty durations for marketing strategies [90]. Accelerated life tests can evaluate asset reliability under different stress levels through extrapolation; however, these experiments are costly, particularly when testing multiple units. To reduce hardware costs, DTs can be employed to predict life reliability in virtual environments that incorporate heterogeneous uncertainties. The reliability result produced by the DT can be utilized to forecast warranty costs, thereby improving the warranty design process. With the DT, degradation processes under various operating conditions and failure modes can be simulated, enabling reliability and warranty cost prediction without field failure data. Once the DT is established, the warranty decision-maker can accurately analyze warranty costs across various product operating environments.

2.3 Predictive Maintenance Strategy Scheduling

The growing prevalence of asset purchasing and leasing in various industries has intensified the emphasis on maintenance strategy planning, largely due to the corresponding increase in O&M costs [91]. In manufacturing systems, maintenance expenses constitute approximately 15–70% of total O&M costs [92]. Likewise, in the energy sector, maintenance accounts for roughly 20–25% of the total levelized cost per kWh for wind energy assets [93]. Maintenance planning is a multidisciplinary field encompassing technical, economic, and temporal complexities, posing significant challenges to its effective implementation in industrial assets. Over recent decades, companies and researchers have developed a variety of maintenance techniques aimed at minimizing costs, reducing downtime, and enhancing asset reliability. Corrective maintenance (CM), time-based maintenance, and condition-based maintenance (CBM) are three traditional maintenance strategies, as depicted in Fig. 2.2. CM involves restoring PAs after failures, often incurring high corrective costs. In contrast, time-based maintenance does not account for the actual operating conditions of assets over time, which can result in unnecessary maintenance and increased expenses. CBM, however, has demonstrated greater effectiveness in minimizing economic losses by monitoring asset health conditions [94]. R. Ahmad and S. Kamaruddin have reviewed time-based maintenance and CBM policies, concluding that CBM offers significant advantages over time-based maintenance in practical applications [95]. However, CBM is limited by its lack of prognostic capabilities, restricting its ability to accurately predict future asset performance [96]. With the growing demand for reliability,

availability, maintainability, and safety, traditional maintenance strategies are increasingly deemed inadequate, leading to significant penalties for enterprises [97]. Evolving from CBM, PdM, also referred to as CBM+, integrates prognostic modeling to optimize the balance between CM and preventive maintenance (PM), enabling maintenance at optimal thresholds and ensuring the efficient procurement of spare parts [98]. To distinguish PHM, CBM, and CBM+, Table 2.2 presents a detailed taxonomy among PHM, CBM, and CBM+. PdM has originated with the Rio Grande Company in the late 1940s and has been later adopted by the U.S. Army as a core maintenance strategy for military equipment [99]. In 2003, ISO 13374 has introduced the Open System Architecture for Condition-Based Maintenance (OSA-CBM), developed by the Machinery Information Management Open Systems Alliance (MIMOSA), to provide a standardized framework for the design and implementation of PdM [99]. PdM can optimize O&M by extending asset lifespan, minimizing costs, and maximizing asset reliability and availability. B. S. Rashmi has identified several key benefits of PdM, including enhanced productivity, cost savings in materials and labor, risk mitigation, and improved inventory optimization [1]. In the absence of PdM, enterprises face significant risks, including financial penalties and reputational damage. For instance, Amazon has reportedly incurred losses of approximately $138,000 per hour due to unplanned data center downtime [101]. PdM has been widely utilized in both academia and industry to improve economic efficiency and asset availability.

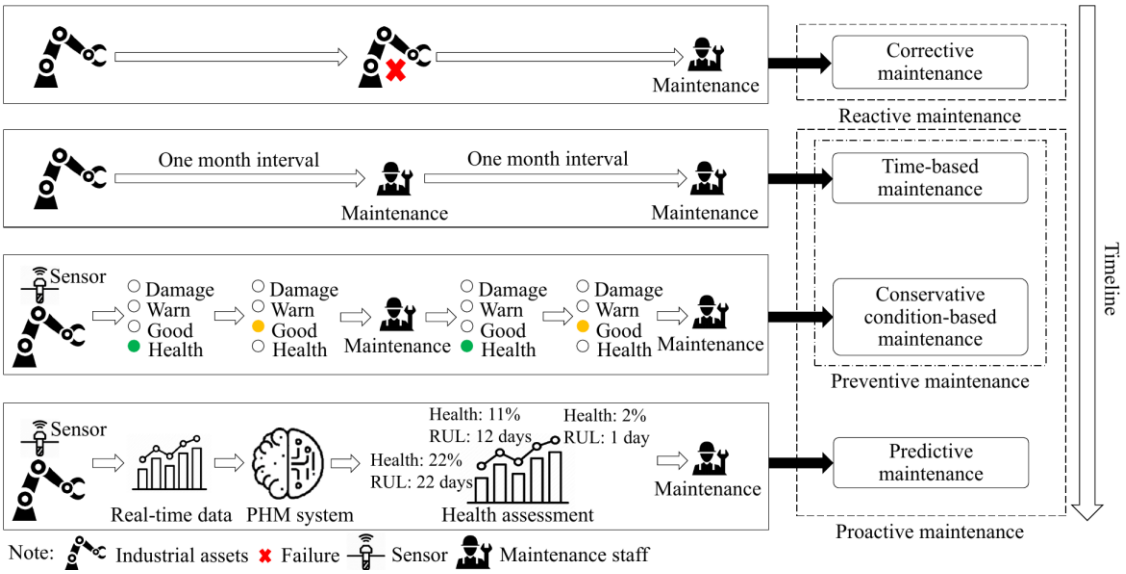


Fig. 2.2 Evolutions and relationships of maintenance schemes.

Table 2.2 Taxonomy of PHM, CBM, and CBM+.

Aspect	PHM	CBM	CBM+
Definition	A framework for monitoring, diagnosing, and predicting system health to optimize maintenance and reliability	Maintenance strategy that monitors the real-time condition of equipment to inform maintenance decisions	An advanced version of CBM that integrates predictive analytics
Modeling approach	Combines physical models, data-driven approaches, and hybrid techniques to assess system health and predict RUL	Uses sensor-based condition monitoring techniques, e.g., vibration, to assess system health	Combines sensor data for real-time monitoring and prediction
Prognostic capability	Focuses on predicting failures and estimating RUL with uncertainty	Limited to current condition monitoring and no RUL estimation	Integrates RUL estimation and predictive analytics for maintenance
Computational demand	High, as it requires significant computational resources for modeling, simulation, and data analysis	Moderates, as it uses condition-based assessments	High, given the integration of AI and predictive analytics

Numerous studies in academia have contributed to the advancement of PdM strategies. Compare and Zio have used a risk-sensitive particle filter and periodic inspection to monitor crack sizes, updating the Paris-Erdogan fatigue crack growth model to reduce unavailability [102]. Blancke et al. have applied a Petri Net propagation model combining physical degradation and expert knowledge, validated on a Hydro-Quebec hydro-generator [103]. Bai et al. have developed a grey relation model to assess wind turbine blade health using availability, reliability, and fatigue life, tested on 3MW turbine blades [98]. L. Zhuang et al. have proposed a two-stage PdM

framework: a Bayesian BiLSTM model predicts probabilistic RUL, while maintenance scheduling and spare part ordering are dynamically updated for turbofan engines [104]. Chen et al. have employed a 1D-convolution neural network-BiLSTM-QR-KDE method to predict battery capacity degradation with uncertainty management, optimizing replacement cost per cycle [33]. In the industry, companies like IBM, Siemens, GE, and Delta Air Lines have developed PdM solutions to minimize costs and maximize availability [105]. In the railway sector, IBM has partnered with Downer to implement PdM using the IBM Maximo Suite, increasing train reliability by 51% [106]. In manufacturing, an automotive company has adopted Siemens Senseye PdM, reducing production downtime by 50% and achieving ROI whitin three months. Similarly, an aluminum manufacturer has deployed Siemens Senseye across 10,000+ machines, reducing unplanned downtime by 20% and achieving ROI in 4–6 months [107]. GE Asset Performance Management solution has improved pipeline availability to 93% and reduced unplanned outages for a major producer [108]. Delta Air Lines has developed a PdM system to monitor the B737 air conditioning pack, enhancing aviation maintenance [109].

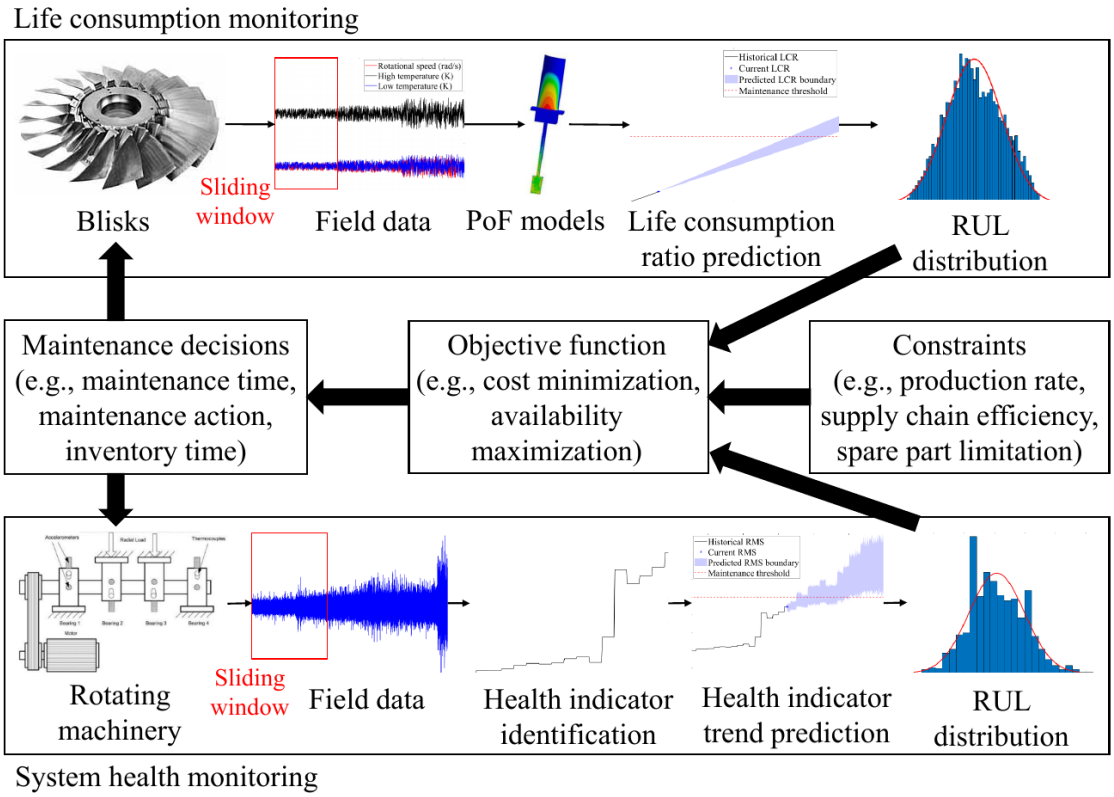


Fig. 2.3 The flowchart of life reliability prediction for PdM scheduling.

The design of an optimal PdM strategy involves considering key optimization objectives, including cost minimization, availability maximization, and multi-objective optimization. The flowchart of life reliability prediction for PdM scheduling is shown in Fig. 2.3. Cost minimization focuses on developing cost-effective PdM schemes using prognostic and cost models, which generally include CM cost, PM cost, and penalty cost. The optimization objective is expressed as in Eq. (2.1).

$$\begin{aligned} &\min_d C(d,p) \\ &\text{s.t.} \begin{cases} L_i(d,p) \le 0 \\ d^L \le d \le d^U \\ p^L \le p \le p^U \end{cases} \end{aligned} \tag{2.1}$$

where $C(\cdot)$ is the total cost, d is the design variable set, p is the parameter set, $L_i(\cdot)$ is the constraint related to d and p, d^L and d^U are lower and upper bound vectors of d, p^L and p^U are lower and upper bound vectors of p.

To minimize operational risk or cost, optimal PdM strategies have been developed

through the formulation of cost models. Mohammadhasani et al. have evaluated risks associated with a nuclear power plant system and implemented a risk-based maintenance management approach [110]. Liao et al. have proposed a novel predictive model to evaluate mechanical performance, determine the optimal maintenance threshold, and establish the required number of maintenance cycles to minimize average costs [111]. He et al. have developed a maintenance decision model for a multistate manufacturing system, taking into account maintenance cost and time constraints [112]. M. Zhang et al. have developed a PdM policy and utilized a cost rate function to dynamically optimize replacement timing [113]. K. T. P. Nguyen and K. Medjaher have optimized maintenance and spare-part ordering decisions to minimize total costs, including holding inventory and out-of-stock costs, based on data-driven prognostic models [114]. Some studies have integrated production planning with PdM for total cost minimization. H. D. Shoorkand et al. have developed an integrated model for production and PdM planning aimed at cost minimization [115]. In practice, in addition to CM costs, PM costs, and penalty costs, contractors may also incur rental costs for crews and heavy equipment required for maintenance. Incorporating theses costs provides a more realistic estimation of maintenance expenses. Yildirim et al. have integrated crew deployment costs into the objective function, combining maintainability constraints for maintenance optimization [116].

Assets experiencing progressive degradation and accidental failures are used as examples to validate the effectiveness of cost-based PdM strategies. Progressive degradation, driven by mechanisms such as fatigue, corrosion, creep, or crack growth,

is modeled using a gamma process, while accidental failures, caused by extreme events (e.g., natural disasters or improper operations), follow a Bernoulli distribution. Initial asset health is represented by a uniform distribution. Life reliability and cost models are developed to evaluate PdM strategies under an infinite-horizon age replacement policy. The optimal maintenance time is optimized for cost minimization with a reinforcement learning algorithm, the off-policy Monte Carlo control algorithm. Q-value analysis (Fig. 2.4) indicates that non-replacement is cost-effective before the 18th year, while replacement becomes economically advantageous thereafter. Details regarding the modeling and optimization of cost-based PdM policies can be found in [117].

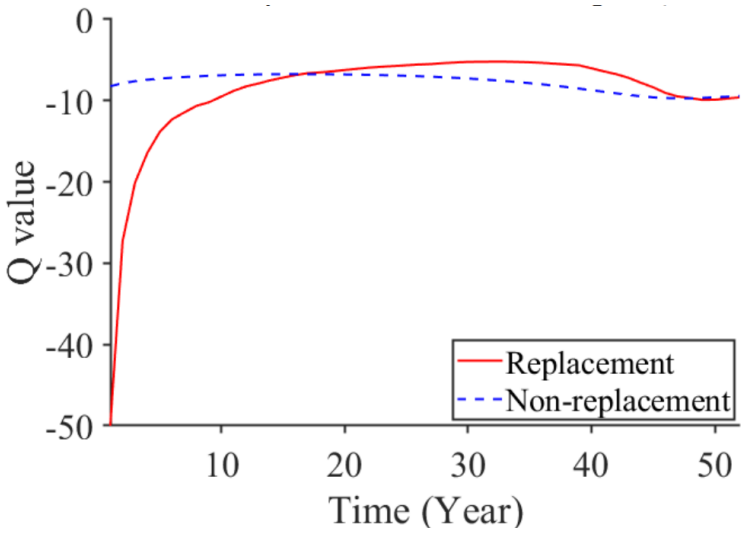


Fig. 2.4 Q values of replacement and do-nothing actions.

In addition to costs, availability is a critical metric for assessing the effectiveness of a PdM policy, as defined in Eq. (2.2). Low availability (i.e., extended downtime) can result in penalties and reputational losses for enterprises.

$$A = \frac{\Delta T_{MTBM}}{\Delta T_{MTBM} + \Delta T_M} \tag{2.2}$$

where ΔT_{MTBM} is the mean time between maintenance and ΔT_M is the mean

maintenance time. Based on A, the optimization objective can be derived from Eq. (2.3).

$$\begin{aligned} &max_d A(d,p) \\ &s.t. \begin{cases} L_i(d,p) \le 0 \\ d^L \le d \le d^U \\ p^L \le p \le p^U \end{cases} \end{aligned} \tag{2.3}$$

Single-objective optimization models often fail to meet multiple metrics in assets. For example, minimizing maintenance costs C in a multi-unit system may cause the availability of certain units to drop below acceptable levels due to unit heterogeneity, varying degradation, and maintenance costs. Multi-objective optimization addresses this by balancing tradeoffs among objectives. The goal is to determine optimal decision variables that considers multiple objectives, as formulated in Eq. (2.4).

$$\begin{aligned} &max_d f = \{f_1(d,p), f_2(d,p), \dots, f_i(d,p), \dots, f_k(d,q)\} \\ &s.t. \begin{cases} L_i(d,p) \le 0 \\ d^L \le d \le d^U \\ p^L \le p \le p^U \end{cases} \end{aligned} \tag{2.4}$$

where k means the total number of objectives and f_i indicates the i-th objective function. The weighted-sum format $f = \sum_{i=1}^{k} w_i f_i (\sum_{i=1}^{k} w_i = 1)$ is a commonly used technique for multi-objective optimization models. Y. Xiang et al. have proposed a joint burn-in and maintenance model to minimize cost rates and maximize average availability, optimizing burn-in duration, inspection intervals, and maintenance thresholds using the weighted sum method [53]. Z. Liang and A. K. Parlikad have proposed a PdM model that utilizes a fault propagation-integrated Markov chain to optimize the mean time between inspections and maintenance thresholds, considering availability and operating costs as performance criteria [119]. Y. Xiang et al. have employed multi-objective optimization to minimize costs and maximize availability in

the context of joint planning for burn-in and imperfect maintenance [120]. L. Lin et al. have developed a multi-objective PdM optimization model that considers cost minimization and availability maximization for maintaining fatigue-induced structures [121].

While PdM strategies enhance asset reliability and reduce expenditure, overly frequent maintenance can result in excessive costs. Therefore, determining the optimal maintenance time is crucial to achieving a balance between benefits and costs. In addition to maintenance time, the health threshold is also a key design variable in the development of a PdM strategy. Degradation does not necessarily lead to failure but significantly increases the likelihood of its occurrence [121]. In previous PdM research, it is assumed that an asset fails when its degradation level surpasses a preset failure threshold. However, maintenance costs are generally positively correlated with the maintenance threshold. Therefore, the maintenance threshold (set below the failure threshold) can be designed to ensure safety while minimizing costs. D. D'Urso et al. have utilized a dynamic reliability model to control the maintenance threshold (mean cumulative damage) for initiating maintenance interventions [48]. M. Zhang et al. have designed a health threshold for cost minimization of the maintenance policy to screen out weak assets [113]. T. Luan et al. have employed a multi-dimensional search method to determine the optimal PdM threshold that minimizes the cost in the case of carrier-based aircraft [122]. In addition to setting maintenance thresholds, procuring spare parts in advance is essential for enhancing asset availability. However, spare parts incur inventory costs; therefore, the timing of their procurement should be optimized to

balance increased asset availability with reduced inventory costs. K. T. P. Nguyen and K. Medjaher have utilized failure probability to design spare part ordering [114]. L. Zhuang et al. have employed prognostic uncertainties to inform spare part ordering decisions [104]. Q. Feng et al. have optimized the integration of PdM and production processes by minimizing total costs while considering the constraints of spare part inventories [123]. When faults are not easily detectable by sensors (i.e., non-self-announcing faults), inspection policies have received increased attention. The purpose of inspection is to assess the degradation levels of assets. Periodic inspections are commonly employed to reduce the likelihood of failure due to their ease of implementation. However, excessively frequent inspections can lead to increased costs. Consequently, determining an optimal inspection interval is essential to balance inspection costs and failure costs. Besnard and Bertling have optimized the inspection interval for wind turbine blades [124]. Liu et al. have adopted a periodic inspection policy to minimize the cost rate for both known and unknown degradation distributions [94]. Zhang et al. have developed an inspection policy to address assets with heterogeneous quality and prevent wear-out failures of weak assets [113]. In practice, maintenance duration is a critical factor that must not be overlooked. Typically, longer maintenance duration correlates with a higher likelihood of successful maintenance outcomes.

Complex systems, comprising numerous interdependent units, pose challenges for PdM planning, making effective system-level policies essential [22]. The advantage of the PdM policy can be enhanced by optimizing maintenance solutions for the multi-

unit system instead of mono-unit components [125]. However, PdM policies for mono-unit components cannot be directly applied to multi-unit systems, as unit dependencies may result in cost-ineffective maintenance [126][127].

OPdM provides an effective approach for making maintenance decisions in multi-unit systems by accounting for dependencies between units. OPdM focuses on performing OM on units during opportunities created by the CM or PM of other units [128]. Unlike unit-level maintenance, OM accounts for the economic interdependence of maintenance costs, offering greater cost efficiency. Compared with PdM, OPdM can reduce O&M costs by sharing setup costs. For example, maintenance staff and equipment are often far from assets, leading to high setup costs (e.g., transportation and maintenance equipment rental costs). Coordinating these maintenance activities can effectively reduce these costs [129][130]. Einabadi et al. have designed a dynamic maintenance planning for multi-machine, multi-component systems with opportunistic grouping and validated it using an automotive case [131]. The OPdM policy typically involves designing an opportunistic threshold. Thus, in the OPdM policy, there are two types of thresholds: the failure threshold for CM or PM, and the opportunistic threshold for OM. Thus, in the OPdM policy, there are two types of thresholds: When PM or CM occurs, other units are considered for OM if their degradation level surpasses the opportunistic threshold. Researchers have extensively studied the optimization of opportunistic threshold. Dinh et al. have developed a multi-level OPdM strategy for multi-unit systems with economic dependence by optimizing opportunistic thresholds [132]. Zhu et al. have optimized opportunistic threshold for the joint optimization of

OPdM and spare part supply [133]. Tian et al. have developed an OPdM strategy for optimizing the opportunistic threshold using dynamic neural network-based prognostics [134]. Zhu et al. have developed a failure intensity limit control theory to determine the optimal opportunistic threshold for CNC machine tools [135].

Maintenance models for multi-unit systems are more accurate than single-unit models as they account for dependencies between units. The unit dependency in the OPdM policy include stochastic (i.e., the failure models of one unit can influence those of other units), structural (i.e., maintenance of a failed unit requires dismantling other units), and economic dependencies (i.e., the cost of OPdM differs from the total cost of maintaining each unit individually) [136]. While these dependencies are crucial for OPdM optimization, they complicate the model. Compared to structural and stochastic dependencies, economic dependencies receive more attention for their advantages in minimizing the expected cost rate [137]. Economic dependencies include positive and negative effects. Positive economic dependence (PED) occurs when group maintenance is cheaper than maintaining units separately, and vice versa for negative economic dependence (NED). To achieve PED between units, group maintenance has been proposed to share setup costs. While these dependencies are crucial for OPdM optimization, they complicate the model. Zhang et al. and Erguido et al. have considered the economic dependencies between different wind turbines and provided an OPdM plan for the wind farms [138][139]. Sedghi et al. have grouped nearby track segments into a single maintenance event to save setup costs and optimize railway expenses [140]. Zhao et al. have optimized an OM policy for a series system with

accelerated damage in shock environments [141]. Wang et al. have investigated the OM for a performance-balanced system operating in a shock environment [142].

2.4 Summary

This chapter reviews the concepts and knowledge related to the metamodel-based DT for life reliability prediction, warranty service design, and PdM strategy planning, laying a solid knowledge foundation for the research presented in the following chapters.

Chapter 3

Metamodel-Based Digital Twin Construction for Life Reliability Prediction

To address the challenge of limited run-to-failure (R2F) data caused by sensor unavailability and the high costs of accelerated life testing, a metamodel-based digital twin (DT) offers a promising solution by facilitating life reliability prediction in virtual environments that account for heterogeneous uncertainties. The development of the metamodel-based DT begins with the analysis of finite element models (FEMs) of physical assets, followed by the construction of metamodels. After the metamodel is trained and validated, probabilistic analysis can be conducted using the metamodel. However, probabilistic life prediction is closely related to the operational behavior of the asset. For cycle-based operational profiles, the life distribution can be directly estimated. In contrast, for non-cycle-based operational profiles, the rainflow counting algorithm should be integrated to enable life prediction. Case studies on turbine blisks and fan shafts illustrate cycle-based and non-cycle-based operational profiles, respectively, showcasing the effectiveness of the proposed metamodel-based DT approach across different operational scenarios.

3.1 Metamodel-Based Digital Twin Construction

As outlined in Chapter 2, virtual replicas play a pivotal role in the DT, as it accurately replicates the geometry, properties, behaviors, and governing rules of

physical assets, modeled as follows.

$$\mathrm{VR} = (G_v, P_v, B_v, R_v) \tag{3.1}$$

where G_v, P_v, B_v, and R_v represent the geometry, physical, behavior, and rule models, respectively. In the DT-PHM architecture, FEM-PoF models capture the geometric G_v, physical P_v, and behavioral B_v properties of the physical asset. To overcome the computational challenges of FEM-PoF and MCSs in probabilistic degradation modeling, metamodeling is used as the rule model R_v for efficient prediction.

In metamodel-based DTs, FEMs combined with PoF principles serve as high-fidelity simulators. These simulations provide accurate representations of the geometry, physical properties, and behaviors of physical assets. Such models are essential for degradation prediction from a physical perspective, even in the absence of R2F data [143].

$$g(X) = \hat{g}(X) + e(X) \tag{3.2}$$

where X is the input variable set, $g(X)$ is the output response of the FEM, $\hat{g}(X)$ is the output response of the metamodel, and $e(X)$ is the metamodeling error [144].

Fig. 3.1 illustrates the process of the DT modeling using metamodels. The metamodel-based DT modeling primarily consists of the following six steps [145].

Step 1: Physical data and information are gathered to construct VRs.

Step 2: A preliminary DoE is established to select representative samples for the simulation of the VR [146].

Step 3: Structural responses, such as stress and strain at each node, can be calculated

by simulating FEM.

Step 4: The relationship between inputs and outputs are trained with metamodels.

Step 5: Real-time sensor data is collected as inputs for the trained metamodel to efficiently predict stress and strain information [146][147].

Step 6: The deviation between the metamodel-predicted value and the actual value is monitored. If the deviation exceeds a predefined threshold, the cause of the abnormality is identified.

Step 7: If the deviation originates from the metamodel, a model update strategy (e.g., sequential sampling strategy) is implemented to correct the metamodel. The detailed metamodel update policy is described below [148].

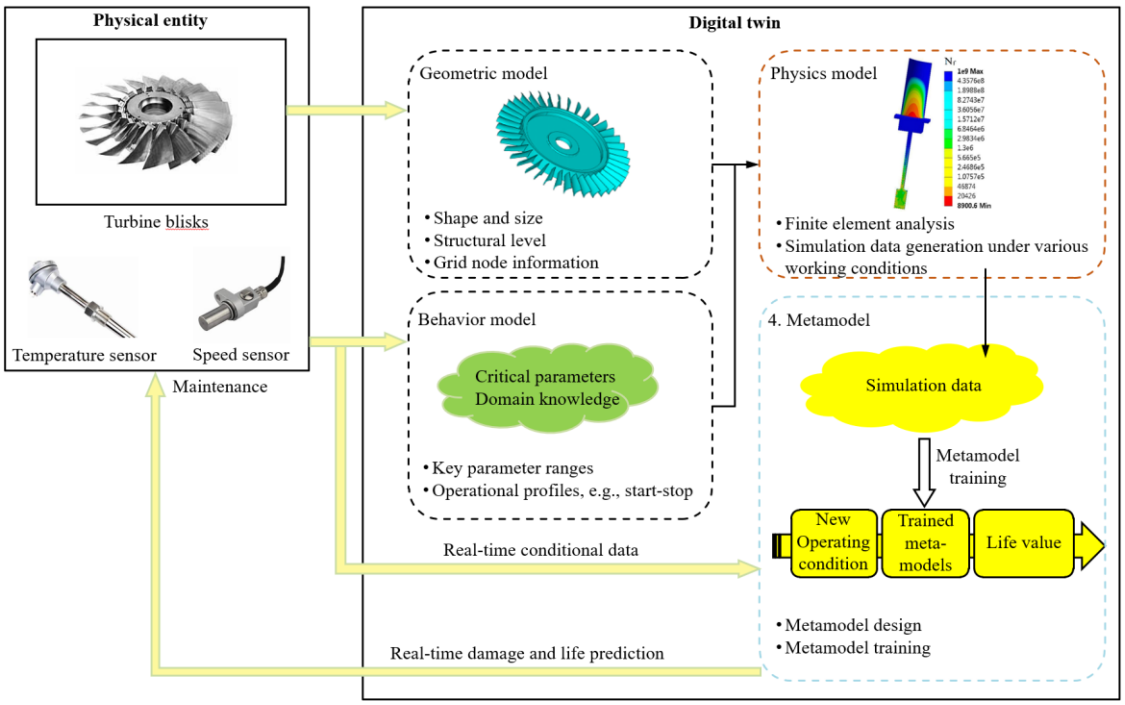


Fig. 3.1 The framework of metamodel-based DT.

A metamodel is a simplified model that serves as a substitute for a complex model, providing a faster means of analysis and optimization. As shown in Fig. 3.2, the key steps in metamodeling are to prepare training samples through sampling techniques and

establish a data-driven model. The establishment of an adaptive metamodel is to start with a small training set and then gradually increase its size by sequentially adding samples to improve the accuracy of the metamodel based on requirements. Fig. 3.2 illustrates the flowchart of the adaptive metamodel establishment. Selecting effective training samples is crucial to train the metamodel that successfully meets specified accuracy requirements. With the surrogating-error-guided strategy, one can carefully decide the optimal training set by employing the following formula.

$$E_r = \frac{\sum_{x \in S} |\mu_{\hat{g}^{(i-1)}}(x) - \mu_{\hat{g}^{(i)}}(x)|}{|\max_{x \in S} \mu_{\hat{g}^{(i)}}(x) - \min_{x \in S} \mu_{\hat{g}^{(i)}}(x)|} \tag{3.3}$$

where S is the candidate set, $\mu_{\hat{g}^{(i-1)}}(x)$ is the prediction mean of the trained metamodel at the $(i-1)$-th iteration, and $\mu_{\hat{g}^{(i)}}(x)$ is the prediction mean of the trained metamodel at the i-th iteration. If $E_r < \varepsilon_r$ (ε_r is the specified accuracy requirement), it can be inferred that the existing training samples are adequate to construct a metamodel with a satisfactory level of accuracy. In addition, there exist other sequential sampling strategies that can effectively guide the attainment of the desired accuracy for the metamodel.

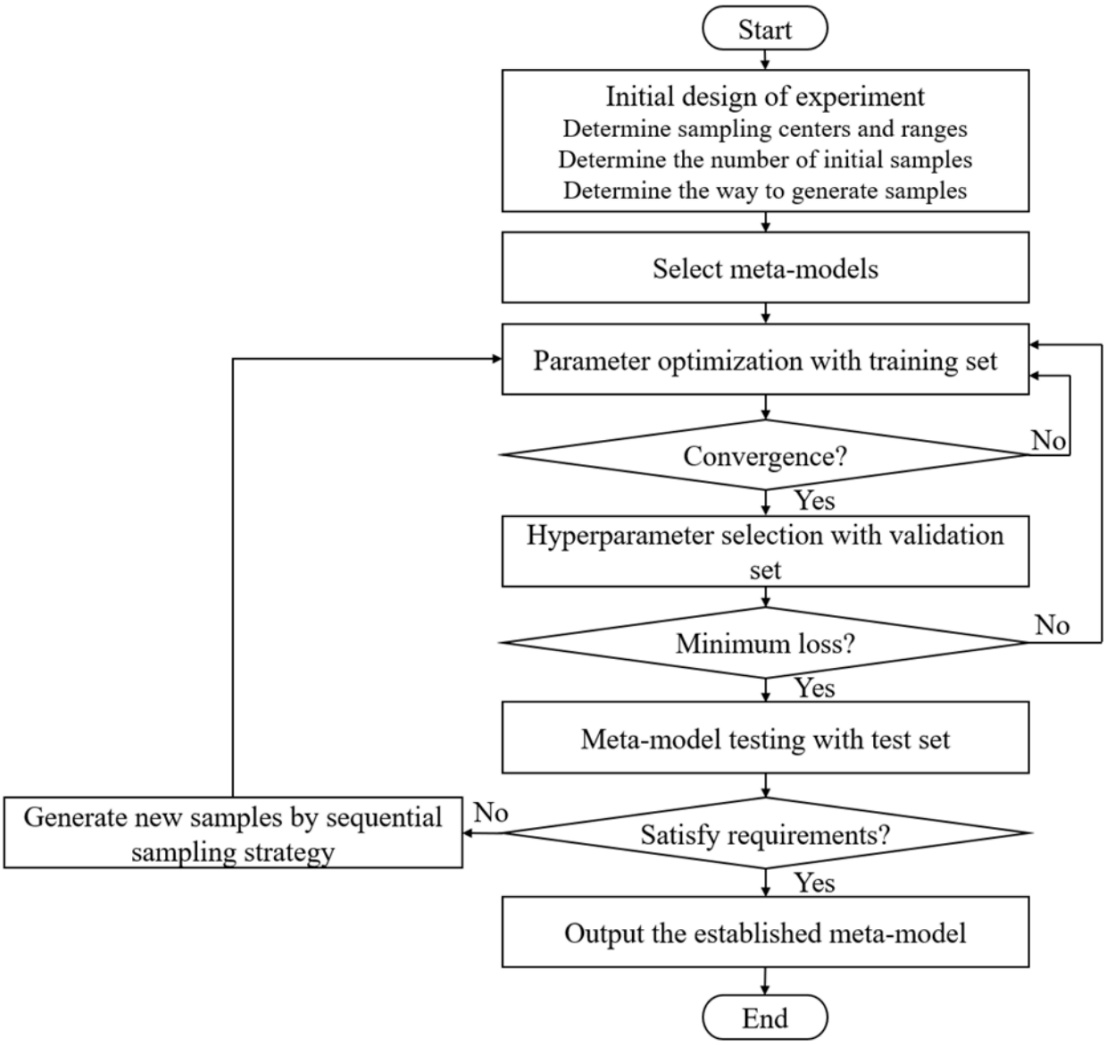


Fig. 3.2 The flowchart of metamodeling.

Operational scenarios of the PA may evolve progressively, represented as $\mathcal{D}_X^{(0)} \subset \mathcal{D}_X^{(1)} \subset \cdots \subset \mathcal{D}_X^{(n)}$, with operational data for each region k described by the PDF $f_X(X; \hat{\theta}^{(k)})$, where $k = 0,1, \ldots, n$. New operational scenarios can lead to test data being distributed far from training data, a phenomenon known as out-of-distribution (OOD). Failure to incorporate OOD data into metamodel training significantly degrades metamodel performance. To address this, it is essential to detect OOD conditions and update the metamodel to reflect new scenarios. This requires an update mechanism with both event-trigger and time-trigger approaches. The event-trigger mechanism uses a

quantitative metric to initiate updates, while the time-trigger mechanism updates the model at regular intervals [149]. For instance, adaptive kriging with a U-learning function can detect new operational domains, as kriging shows high predictive uncertainty for OOD data. Fig. 3.3 illustrates the metamodel update policy.

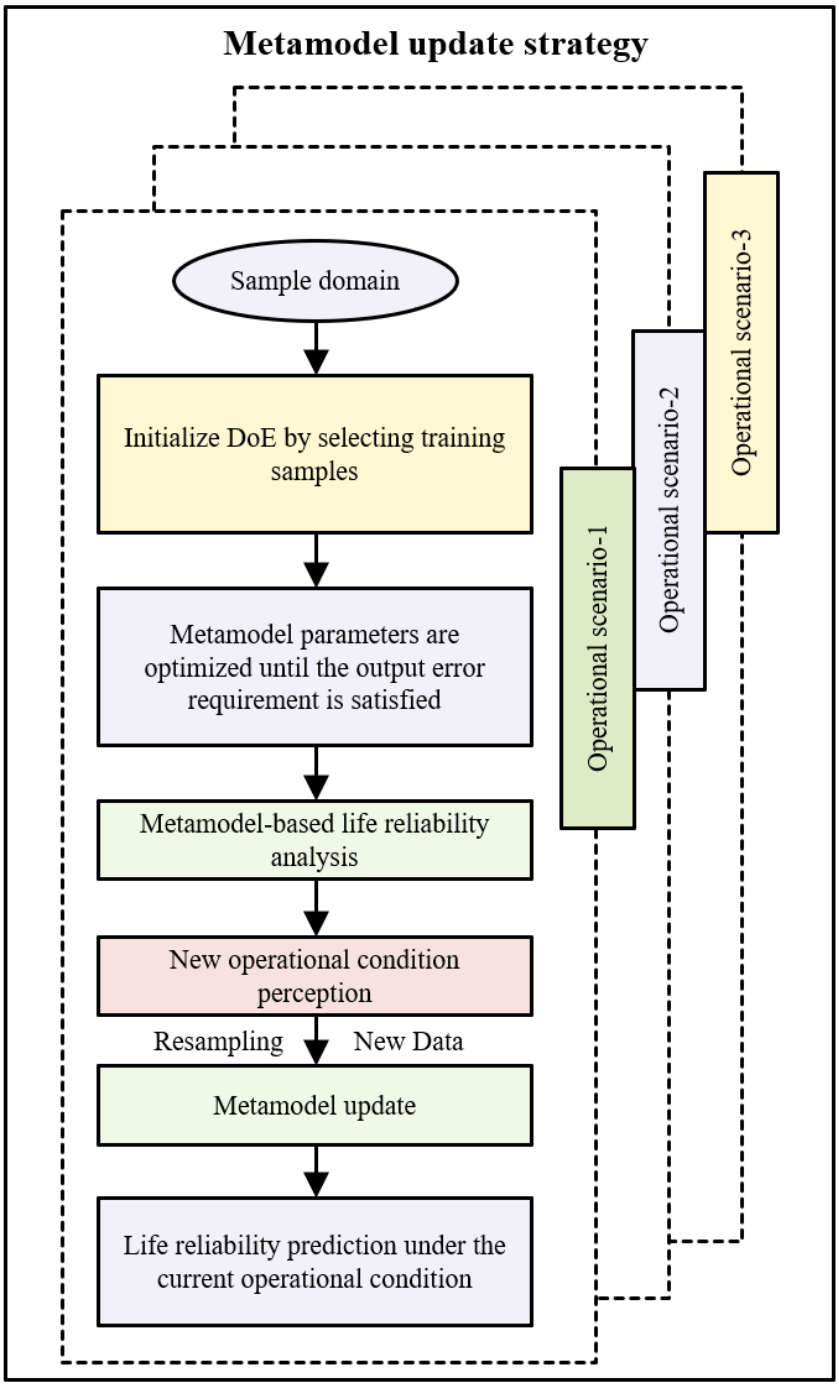


Fig. 3.3 The metamodel update strategy.

3.2 Life Reliability Prediction

3.2.1 Low Cycle Fatigue Life Prediction

The Manson-Coffin model is taken to predict the LCF life of the aeroengine blisks, as shown in

$$\frac{\Delta\varepsilon_t}{2} = \frac{\Delta\varepsilon_e}{2} + \frac{\Delta\varepsilon_p}{2} = \frac{\sigma_f'}{E}\left(2\tau_f\right)^b + \varepsilon_f'\left(2\tau_f\right)^c \tag{3.4}$$

where $\Delta\varepsilon_t$ is the strain range, $\Delta\varepsilon_e$ is the elastic strain range, $\Delta\varepsilon_p$ is the plastic strain range, σ_f' is the fatigue coefficient, E is the modulus of elasticity, τ_f is the LCF life, b is the fatigue strength index, ε_f' is the fatigue ductility coefficient, and c is the fatigue ductility index.

J. Morrow has introduced the effect of mean stress σ_m to the LCF life, which modifies Eq. (3.4) as follows:

$$\frac{\Delta\varepsilon_t}{2} = \frac{\sigma_f'-\sigma_m}{E}\left(2\tau_f\right)^b + \varepsilon_f'\left(2\tau_f\right)^c \tag{3.5}$$

where σ_m is the mean stress and the other parameters are the same as Eq. (3.4).

3.2.2 High Cycle Fatigue Life Prediction

From the HCF failure, the S-N curve, also called Wohler curve, can be considered a HCF lifetime model to characterize material performance. It describes the magnitude of cyclic stresses plotted against the logarithmic scale of cycles to failure. The empirical S-N curve can be generalized using the Basquin model, as shown in Eq. (3.6) [150].

$$S_a = S_f'(2\tau_f)^b \tag{3.6}$$

where S_a is the cyclic stress amplitude, S_f' is the fatigue strength coefficient, τ_f is the HCF life, and b is the fatigue strength index. The numerical values of S_f' and b are

often given in handbooks or standards.

The S-N curve is typically derived using a zero mean cycle stress. However, the S-N curve is known to be strongly dependent on the mean stress level [151]. Mean stress correction models may therefore be used to adjust the stress amplitude prior to damage calculation from S-N curves using:

$$\frac{S_a}{S_e} + \left(\frac{S_m}{S_u}\right)^Z = 1, |S_m| \leq S_u \tag{3.7}$$

where S_a is the cyclic stress amplitude at non-zero mean stress, S_e is the equivalent cyclic stress amplitude at zero mean stress, S_m is the mean stress, and S_u is the ultimate tensile strength. For $Z = 1$ and $Z = 2$, the Goodman and Gerber correction are applied, respectively, while the Soderberg correction is obtained by using $Z = 1$ and replacing S_u with the yield stress. Eq. (3.6) makes it possible to transform each cycle of non-zero mean stress S_m into an equivalent stress cycle of zero mean S_e with the alternating stress S_a.

This study utilizes the Goodman correction; therefore, the Basquin model incorporating the Goodman correction is presented in Eq. (3.8).

$$S_a = \left(1 - \frac{S_m}{S_u}\right) S_f'\left(2N_f\right)^b \tag{3.8}$$

3.2.3 Fatigue Damage Estimation

Different from the cycle-based failure mechanism, if the load is oscillatory, the resulting stress will exhibit oscillatory behavior. Consequently, these oscillatory stresses induce material fatigue, which can lead to crack formation. To predict fatigue life, the stress history should be decomposed into periodic cycles, as shown in Eq. (3.9).

$$S = S_m + \sum_{i=1}^{n_s} S_{a,i} sin\ (w_i t) \tag{3.9}$$

where S is the time-series stress history, S_m is the mean stress, n_s is the number of stress levels, $S_{a,i}$ is the amplitude of the ith periodic stress cycle, and w_i is the frequency of the ith periodic stress cycle.

To estimate fatigue damage under oscillatory stresses, fatigue cycles in a random stress amplitude series must be identified using a cycle counting algorithm. The rainflow counting algorithm, introduced by Matsuishi and Endo in 1968 [150], is widely regarded as the most accurate method for fatigue analysis. It identifies cycles based on the hysteresis loop characteristics of the stress-strain curve, as shown in Fig. 3.4, with its rules outlined below.

(1) The starting point of the rainflow is at each peak or valley, called "roof".

(2) The rainflow continues to the next peak or valley until it reaches a larger peak or a smaller valley.

(3) If the rainflow reaches a larger peak or a smaller valley, it stops and creates a cycle.

(4) Based on the starting and ending points, each cycle is extracted, and their amplitudes and mean values are recorded.

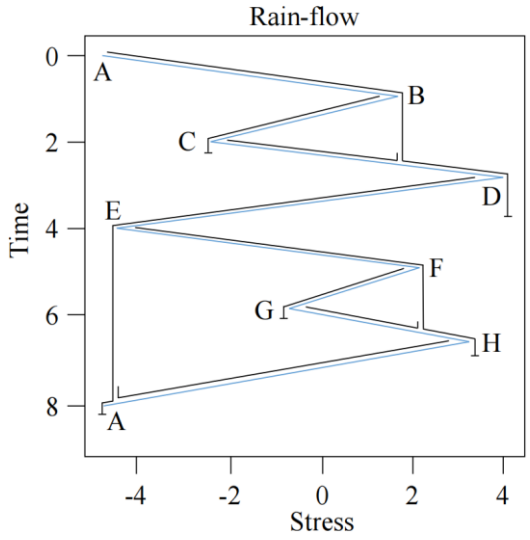


Fig. 3.4 Graphical illustration of rainflow counting.

The rainflow counting algorithm, developed for fatigue analysis, accurately identifies stress cycles over time by segmenting and concatenating cycles to maintain consistency with the original stress series. MATLAB codes by A. Nieslony are utilized for this study [152].

3.2.4 Reliability and Sensitivity Analysis

Although there are some criteria to estimate the reliability under different levels of cyclic loads, Miner's cumulative damage theory is one of the most popular criteria, which is expressed by

$$D = \sum_{i=1}^{n} \frac{\tau_i}{\tau_{fi}} \tag{3.10}$$

where D is the structural damage, n is the number of load levels, τ_i is the operational cycles, and τ_{fi} is the LCF life under the ith load level.

Based on the structural damage, the structural response can be defined as

$$Y = 1 - D \tag{3.11}$$

where Y is the structural response. $Y \leq 0$ mean that the blisks are in the unsafe domain while $Y > 0$ mean that the blisks are in the safe domain.

Based on the previous structural response, the structural failure probability and reliability can be obtained with

$$FP = \frac{1}{N} \sum_{i=1}^{N} I_F[g_i(X_i)] = \frac{N_f}{N} \tag{3.12}$$

$$R = 1 - FP \tag{3.13}$$

$$I_F[g_i(X_i)] = \begin{cases} 1, & g_i(X_i) \leq 0 \\ 0, & g_i(X_i) > 0 \end{cases} \tag{3.14}$$

where FP is the failure probability, R is the reliability indicator, $I_F[S_i(X_i)]$ is the

indicator function in the failure domain, N_f is the number of samples in the failure domain, and N is the number of total samples.

More specifically, life reliability represents the probability that a PA remains functional over a specified time interval and can be expressed as:

$$R(t) = Pr\{t < \tau_f\} = Pr\{max(H(u)) < D, 0 \leq u \leq t\} \quad (3.15)$$

where τ_f represents the life of PAs, $H(u)$ denotes the HI at time u, and D is the failure threshold [22].

Sensitivity reflects the influence level of input random variables to output responses, which is used to identify the main effecting factors of system failures. The sensitivity S_j of the j-th input variable x_j to the output response can be calculated by

$$S_j = E\left(\frac{I_f[g(x_j)](x_j - E[x_j])}{Var[x_j]}\right) \quad (3.16)$$

where $I_f[g(x_j)]$ is the failure indicator, $I_f[g(x_j)] = 0$ indicates that PAs are secure and $I_f[g(x_j)] = 1$ reveals PAs fail.

3.3 Case Study 1: Low Cycle Fatigue Life Reliability Prediction for Turbine Blisks

3.3.1 Finite Element Modeling for Deterministic Low Cycle Fatigue Life Prediction

This case study utilizes aeroengine blisks as a representative example. To reduce the computational burden of FEM, a 1/40th segment of the circumferential blisk, leveraging its symmetry, is selected for reliability assessment (Fig. 3.5). The FEM model comprises 31,380 nodes and 17,111 elements. The blisk, made of GH4133-a material with high-temperature and high-pressure tolerance-is evaluated under stable

operating conditions. Since aerodynamic forces have minimal impact compared to thermal expansion and centrifugal force, this study focuses on their combined effects on the stress, strain, and LCF life distribution. Key parameters include material properties such as density ρ and modulus of elasticity E, thermal parameters like thermal conductivity λ, blisk-root temperature T_{root}, blisk-tip temperature T_{tip}, and rotational speed ω in the centrifugal force. The LCF of aeroengine blisks is primarily influenced by two physical fields: the centrifugal and temperature fields. The centrifugal field is simulated with a rotational speed of 1,168 $rad \cdot s^{-1}$, while the temperature field is modeled using the radial temperature distribution between the blisk root and tip, derived from solid heat transfer principles and defined by Eq. (3.17). Additionally, the load cycle, comprising "take-off, cruise, and landing," is analyzed, with stresses and strains primarily occurring during the take-off and landing phases.

$$T = T_{root} + (T_{tip} - T_{root}) \frac{R^m - R_{root}^m}{R_{tip}^m - R_{root}^m} \tag{3.17}$$

where T is the radial temperature distribution, R is the radius of the blisk at different positions, R_{root} is the radius of blisk-root, R_{tip} is the radius of blisk-tip, and m is the temperature distribution coefficient of the blisk GH4133.

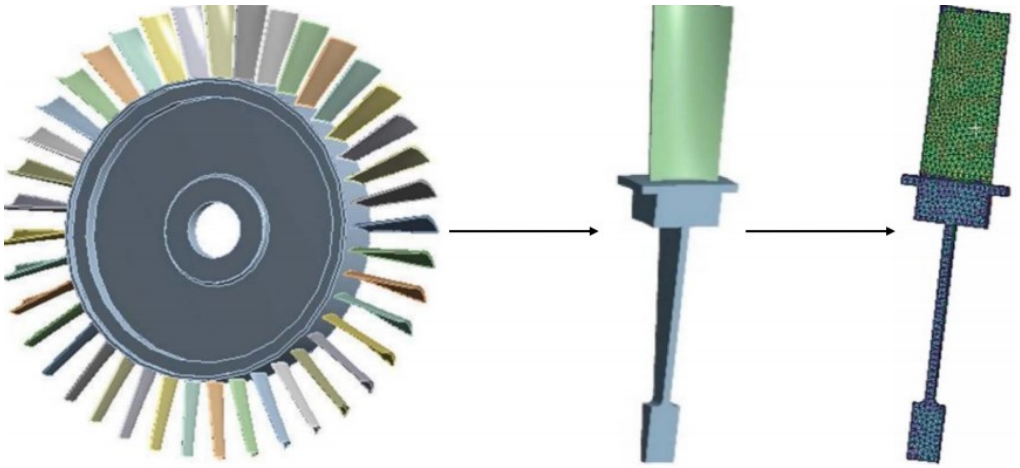


Fig. 3.5 The structure and FEM of the blisk.

The FEM simulation of the blisk, performed in the Workbench 16.0 environment, considers the combined effects of thermal expansion and centrifugal force. Conducted on a system with an Intel Xeon E5-2630V3 CPU and 64 GB RAM, the simulation calculates the radial temperature distribution using Eq. (3.15), shown in Fig. 3.6(a). The resulting stress and strain distributions, influenced by temperature, thermal expansion, and centrifugal force, are depicted in Figs. 3.6(b) and 3.6(c). The maximum stress reaches 1,057.7 MPa, while the maximum strain at the blisk-root is 8.1427×10^{-3}, identifying the blisk root as the critical point for reliability assessment and inspection priority. Using the improved Manson–Coffin model and Miner's rule, the minimum LCF life is calculated at 8,900.6 cycles, with a double-safety criterion setting it at 4,450 cycles.

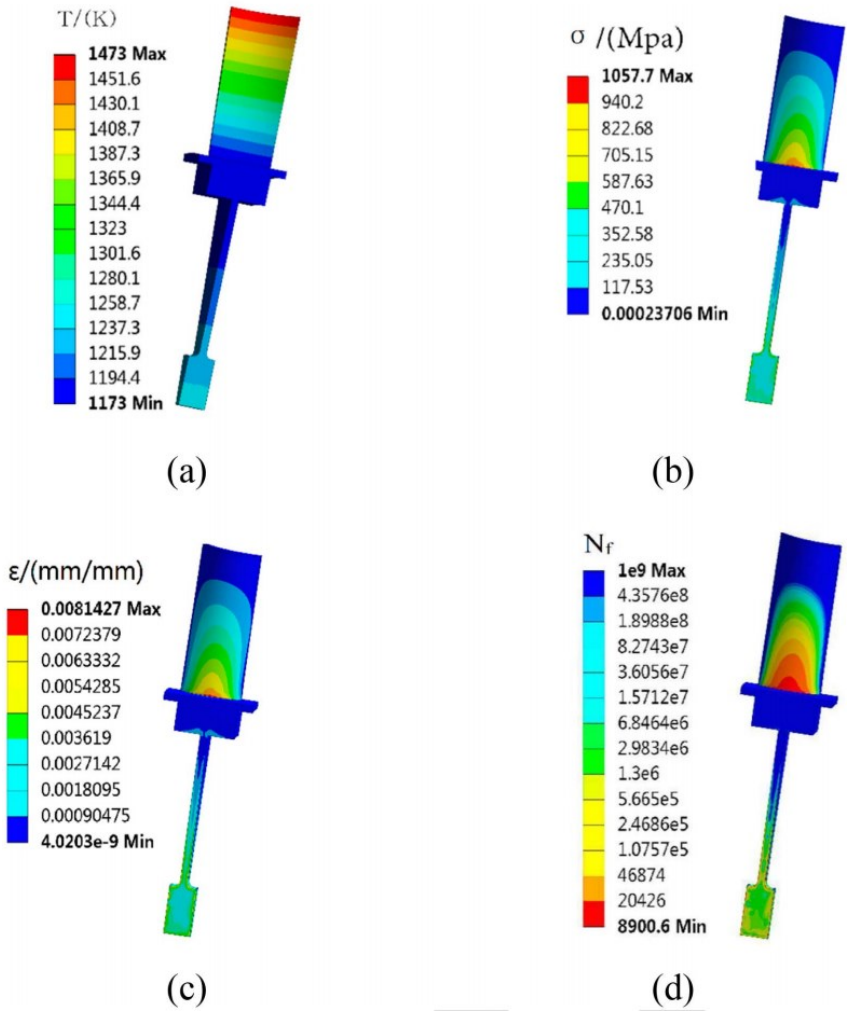


Fig. 3.6 Contour plots illustrating the distributions of temperature, stress, strain, and LCF life responses. (a) Temperature distribution. (b) Stress distribution. (c) Strain distribution. (d) LCF life distribution.

3.3.2 Metamodeling of Ensemble Generalized-Constraint Neural Networks for Low Cycle Fatigue Life Reliability Prediction

Generalized regression neural networks (GRNNs), consisting of input, pattern, summation, and output layers, is a typical neural network model. The pattern layer is critical in the GRNN, which uses sampling techniques to memorize representative centers. These centers aim to group new samples based on their distance from the cluster centers. Traditional GRNN employs sample-wise Gaussian kernels for regression, which makes it computationally efficient. However, the GRNN, considered

a black-box model, still suffers from weak interpretability. The interpretability of metamodels is also important, as the analysis results can be more robust if the metamodels are more transparent. (1) Incorporating prior information into the metamodels before training and (2) Extracting knowledge from the trained metamodels. It has been recognized that the incorporation of prior information can improve computing efficiency and generalization performance. Thus, this study will focus on the first strategy to build up interpretable neural network-based metamodels. The first strategy can be achieved in the following ways: (1) Probabilistic models, which employ data distributions as Bayesian priors, (2) Initialization of model parameters which is performed with pre-training or matrix completion algorithms, (3) Theory-based regularization, which imposes knowledge constraints by adding some penalties to the objective functions, and (4) Constraint optimization which mainly applies constraints on the optimization techniques such as deterministic optimization or stochastic searching. This study employs the third way to impose knowledge constraints on the GRNN model, which adds penalties to the objective function to guide the optimization direction. A generalized constraint neural network (GCNN), which incorporates prior information or knowledge of the GRNN, is developed. The GCNN consists of two sub-models: One is constructed by GRNN to approximate unknown parts of the responses, and the other one is formed from a partially known relationship to impose constraints on the whole model. In this study, the knowledge obtained from previous statistical research is taken as the partial knowledge relationship (LCF life distribution obeys lognormal distribution) to train the GCNN model. However, the performance of GCNN

will also be affected by the selection of Latin hypercube sampling (LHS) training sets. To improve the generalization performance, ensemble learning is introduced to form ensemble GCNN (EGCNN) considering multiple LHS training sets.

GRNN is an effective tool for approximating nonlinear functions. It establishes the relationship between inputs and outputs using a PDF derived from the observed data, as illustrated below.

$$E[Y|X] = \frac{\int_{-\infty}^{\infty} Yf(X,Y)dY}{\int_{-\infty}^{\infty} f(X,Y)dY} \tag{3.18}$$

where X and Y are input and output vectors, respectively, $\mathrm{E}[Y|X]$ represents the expected value of the output vectors given the input vectors, while $f(X,Y)$ denotes the joint continuous PDF of X and Y. The GRNN is composed of four distinct layers: (1) the input layer, (2) the pattern layer, (3) the summation layer, and (4) the output layer. In the GRNN, the pattern layer is critical for memorizing patterns. Conventionally, the Gaussian kernel is applied on a sample-wise basis, as demonstrated in Eqs. (3.19) and (3.20).

$$o_i^{pat} = e^{\frac{-D_i^2}{2\sigma^2}} \tag{3.19}$$

$$D_i^2 = (x^{new} - x_i^{center})^T (x^{new} - x_i^{center}) \tag{3.20}$$

In GRNN, o_i^{pat} represents the output of the ith memorized cluster center, D_i is the Euclidean distance between the new sample and the ith cluster center, σ is the Gaussian kernel parameter, x^{new} is the input vector of the new sample, and x_i^{center} is the vector of the ith cluster center. The Gaussian kernel parameter σ is critical, as higher values enhance generalization and smoothing, while lower values weaken generalization. Thus, optimizing σ is essential for GRNN performance. The GRNN output is computed

through the summation and output layers, as shown below:

$$\hat{y} = \frac{\sum_{i=1}^{n} y_i \times e^{\frac{-D_i^2}{2\sigma^2}}}{\sum_{i=1}^{n} e^{\frac{-D_i^2}{2\sigma^2}}} \tag{3.21}$$

In GRNN, y_i denotes the target value of the ith memorized cluster center. Traditional GRNNs have limited nonlinear fitting capabilities. To address this, this study replaces sample-wise Gaussian kernels in the pattern layer with feature-wise Gaussian kernels, applying distinct kernels to each input feature individually, as shown in Fig. 3.7. The pattern layer functionality is redefined by Eqs. (3.21) and (3.22).

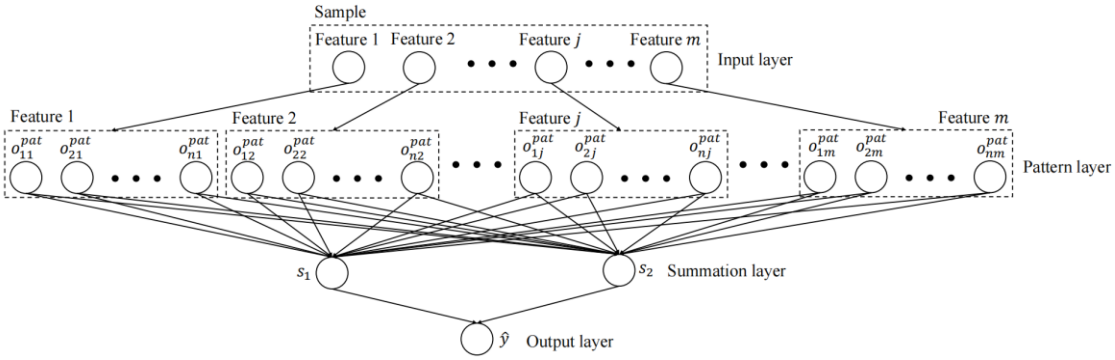


Fig. 3.7 The improved GRNN with feature-wise Gaussian kernels.

$$o_{ij}^{pat} = e^{\frac{-D_{ij}^2}{2\sigma_j^2}} \tag{3.22}$$

$$D_{ij}^2 = \left(x_j^{new} - x_{ij}^{new}\right)^T \left(x_j^{new} - x_{ij}^{new}\right) \tag{3.23}$$

In GRNN, o_{ij}^{pat} represents the output of the jth feature in the ith memorized cluster, D_{ij} is the Euclidean distance between the jth feature of the new sample and the jth feature of the ith cluster center, x_j^{new} is the jth feature of the new sample, and x_{ij}^{center} is the jth feature of the ith cluster center. To address the challenge of neural network interpretability, GCNN is introduced to enhance IGRNN, as shown in Fig. 3.8. GCNN comprises two submodels: one uses IGRNN to approximate unknown system components, while the other enforces knowledge constraints. Numerical experiments

demonstrate that GCNN improves generalization, interpretability, and learning speed, meeting surrogate model requirements. The core of GCNN involves defining the PKR submodel and coupling it with the IGRNN submodel. In structural reliability analysis, the LCF life is modeled as a lognormal distribution, supported by MCS-based studies and literature. To ensure this, a constraint is applied to the IGRNN submodel, enforcing lognormal probabilistic LCF life estimation. The coupling is achieved by adding a log-likelihood-based penalty term to the cost function.

$$E = E_1 + E_2 = \frac{\sqrt{(1/t)\sum(\hat{y}-y)^2}}{\bar{y}} + ln\left[f(\hat{y}|\theta)\right] \tag{3.24}$$

where E_1 is the normalized root mean square error (NRMSE) utilized to measure the distance between estimated and target LCF life values, E_2 is the log-likelihood function (LLF) employed to measure the similarity between estimated LCF life distribution and lognormal distribution, t is the number of input samples, $\hat{y}$ is estimated LCF life, y is target LCF life, $\bar{y}$ is expectation of target LCF life, $f(\cdot)$ is the PDF of estimated LCF life, and θ is the feature-wise Gaussian kernel parameters to be optimized.

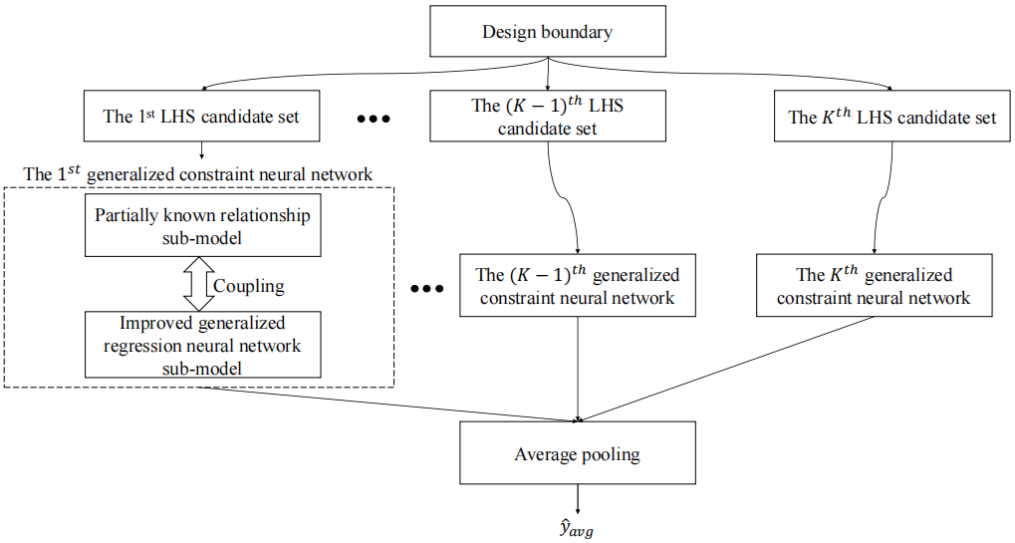


Fig. 3.8 The EGCNN model.

To address uncertainty in LHS candidate sets and improve generalization, an average-pooling ensemble approach is used to build EGCNN, as shown in Fig. 3.8. Each GCNN model's input layer includes m neurons, corresponding to the kth LHS candidate set and the number of input features.

$$s_1 = \sum_{j=1}^{m} \sum_{i=1}^{n} y_i o_{ij}^{pat} \tag{3.25}$$

$$s_2 = \sum_{j=1}^{m} \sum_{i=1}^{n} o_{ij}^{pat} \tag{3.26}$$

where y_i is the target LCF life value of the ith memorized neuron and o_{ij}^{pat} is the output of the jth feature in the ith memorized sample.

Average pooling is employed in the ensemble learning process, enabling the final estimation of LCF life as shown below:

$$\hat{y}_{avg} = \frac{\hat{y}_1 + \hat{y}_2 + \cdots + \hat{y}_k + \cdots + \hat{y}_K}{K} \tag{3.27}$$

where K is the number of LHS candidate sets. The overall forward calculation process is listed in Algorithm 3.1.

Algorithm 3.1: Ensemble GCNN.

Input:
New samples $\mathrm{X}^{\mathrm{new}}$, the κth candidate set taken by LHS $\mathrm{X}_{\kappa}^{\mathrm{LHS}}$, the number of representative samples n in each LHS candidate set, the number of input features m, and the number of LHS candidate sets K.
Output:
Estimated average LCF life $\hat{y}_{\mathrm{avg}}$.
1: Initialization:
2: Structure of the pattern layer in the κth GCNN based on $\mathrm{X}_{\kappa}^{\mathrm{LHS}}$ and feature-wise Gaussian kernels σ_j.
3: for κ←1 to K do
4: for $i \leftarrow 1$ to n do
5: Calculate the Euclidean distance between $\mathrm{X}^{\mathrm{new}}$ and $\mathrm{X}_{\kappa}^{\mathrm{LHS}}$ with Eq. (3.21).
6: Calculate the output of the pattern layer with Eq. (3.20).
7: Output the estimated LCF life in theκth GCNN with $\hat{y} = s_1/s_2$.
8: Predict the estimated average LCF life $\hat{y}_{\mathrm{avg}}$ with Eq. (3.25).

The training process of EGCNN is reformulated to focus on determining the optimal feature-wise Gaussian kernels, σ_j, that minimize the cost function, as presented in Eq.

(3.27).

$$\sigma_j^{k*} = argmin_{\sigma_j^k} E(X, \sigma_j^k) \tag{3.28}$$

where σ_j^k is the jth Gaussian kernel in the kth GCNN and X is the input vector.

The optimization of the EGCNN model can be achieved through deterministic techniques (e.g., gradient descent) and stochastic searching (e.g., meta-heuristics). The gradient descent can only find near-optimal solutions. The meta-heuristics, on the other hand, possess the advantages of strong global searching capability and model independence. Among the meta-heuristics, ant colony optimization (ACO), which is inspired by the foraging activities of ants, has been applied to a wide range of NP problems. The ACO demonstrates robust optimization ability in different applications. Furthermore, its distributed computing power can avoid premature convergence. Thus, this chapter employs the ACO technique to train the EGCNN model. ACO is employed to optimize feature-wise Gaussian kernels due to its robustness and resistance to premature convergence. The detailed ACO process is shown in Appendix A.3.1. Using a pheromone-based model with state transition and updating rules, it probabilistically explores the solution space. The detailed algorithm is outlined in Algorithm 3.2.

Algorithm 3.2: Ant Colony Optimization.

Input:
Upper and lower limits of the j th feature-wise Gaussian kernel in the k th GCNN $\upsilon_{jk}^{\text{upper}}$
$\upsilon_{jk}^{\text{lower}}$, number of grids N, convergence thresholdε, number of ants A, iteration number N_{cmax},
evaporation coefficient ρ, cost function of NRMSE E_1, cost function of LLF E_2, pheromone intensity
Output:
Optimal feature-wise Gaussian kernels of EGCNN σ_j^{k*}.
1: Initialization:
2: Grid interval h_{jk} of the jth featurewise Gaussian kernel in the kth GCNN, pheromone level of the jth feature-

wise Gaussian kernel in the kth GCNN carried by the ath ant τ_{jak}, and transition probability of the jth feature-
wise Gaussian kernel in the kth GCNN carried by the ath ant.
3: while maximum grid interval max(h_{jk}) >ε do
4: Place the ants on any grid in the range ($\upsilon_{jk}^{\mathrm{upper}}$, $\upsilon_{jk}^{\mathrm{lower}}$) randomly.
5: Input the feature-wise Gaussian kernels carried by each ant into the EGCNN.
6: Compute costs of NRMSE and LLF E_1 and E_2 based on selected feature-wise Gaussian kernels with(7).
7: for $N_{\mathrm{c}} \leftarrow 1$ to N_{cmax} do
8: Calculate the pheromone level carried by the ath
ant τ_a with (13).
9: Calculate the transition probability P_a carried by the ath ant with (14).
10: Find the ant with the maximum transition probability and narrow [$\upsilon_{jk}^{\mathrm{upper}}$, $\upsilon_{jk}^{\mathrm{lower}}$] with (16) and(17).
11: Calculate the grid interval with the updated range of [$\upsilon_{jk}^{\mathrm{upper}}$, $\upsilon_{jk}^{\mathrm{lower}}$].
12: Obtain optimal feature-wise Gaussian kernels with (18).

To obtain EGCNN metamodel, the following process should be performed.

(1) Initial EGCNN model establishment by randomly selecting K LHS candidate sets and feature-wise Gaussian kernels.

(2) Optimization of K feature-wise Gaussian kernels with the training set and ACO. If the cost function E satisfies convergence criteria, the next step is performed. If not, the current step is repeated.

(3) Find the optimal ant colony hyperparameters with fivefold cross-validation and grid search until the best hyperparameters are found.

(4) Test the established EGCNN model in the test set.

For clearer expression, Fig. 3.9 shows the process of training and testing the EGCNN with training and test sets.

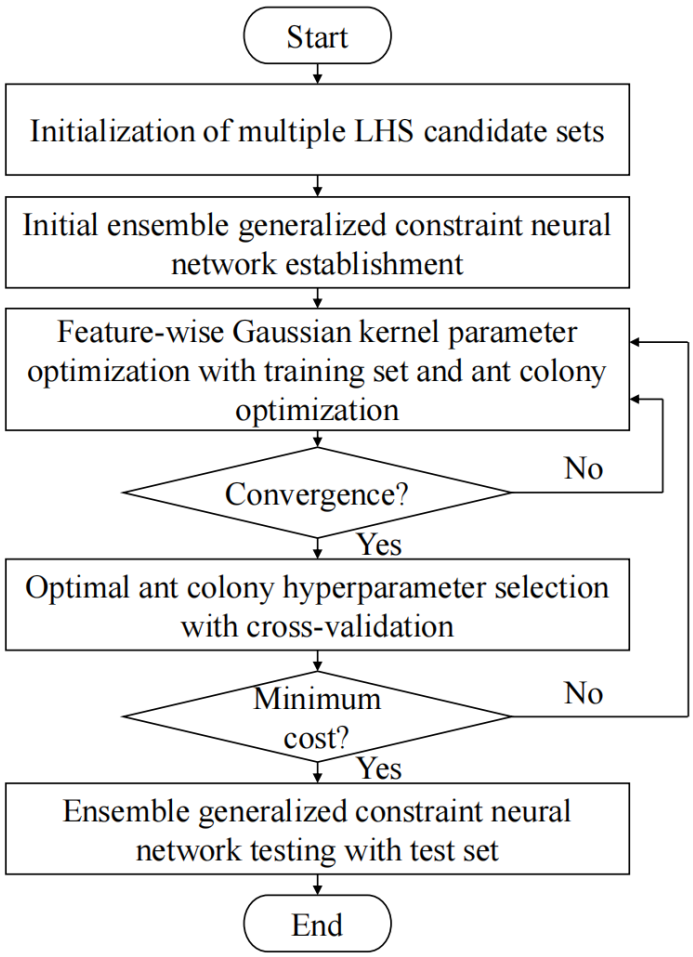


Fig. 3.9 The process of establishing EGCNN.

Herein, the essential process of reliability assessment with the EGCNN model is summarized as follows.

(1) Establish the structure and FEM of aeroengine blisks in a workbench environment.

(2) Select material and operational parameters that have a significant impact on the LCF life of the blisks.

(3) Conduct deterministic LCF life prediction with the established FEM, thermal–structure interaction, and stress–strain–life relationship.

(4) Perform LHS K times to obtain K candidate sets based on uncertainty characteristics of material and operational parameters.

(5) Train and test the EGCNN model with the candidate sets and ACO techniques.

(6) Check the validity of the established EGCNN by comparing with other established surrogate models. If unacceptable, return to (5); if acceptable, conduct (7).

(7) Employ the established EGCNN model to estimate the reliability trend and obtain sensitivity degree of the blisks under different design boundaries or confidence intervals (CIs) based on application requirements.

The reliability assessment framework using the EGCNN surrogate model is shown in Fig. 3.10. By integrating the ACO-GCNN for improved fitting accuracy and an ensemble mechanism for enhanced generalization, the EGCNN surrogate model increases the efficiency and accuracy of reliability assessments. Its robustness, demonstrated in Section III, enables adaptive application across various industries and dynamic updates to address changing criteria or model degradation.

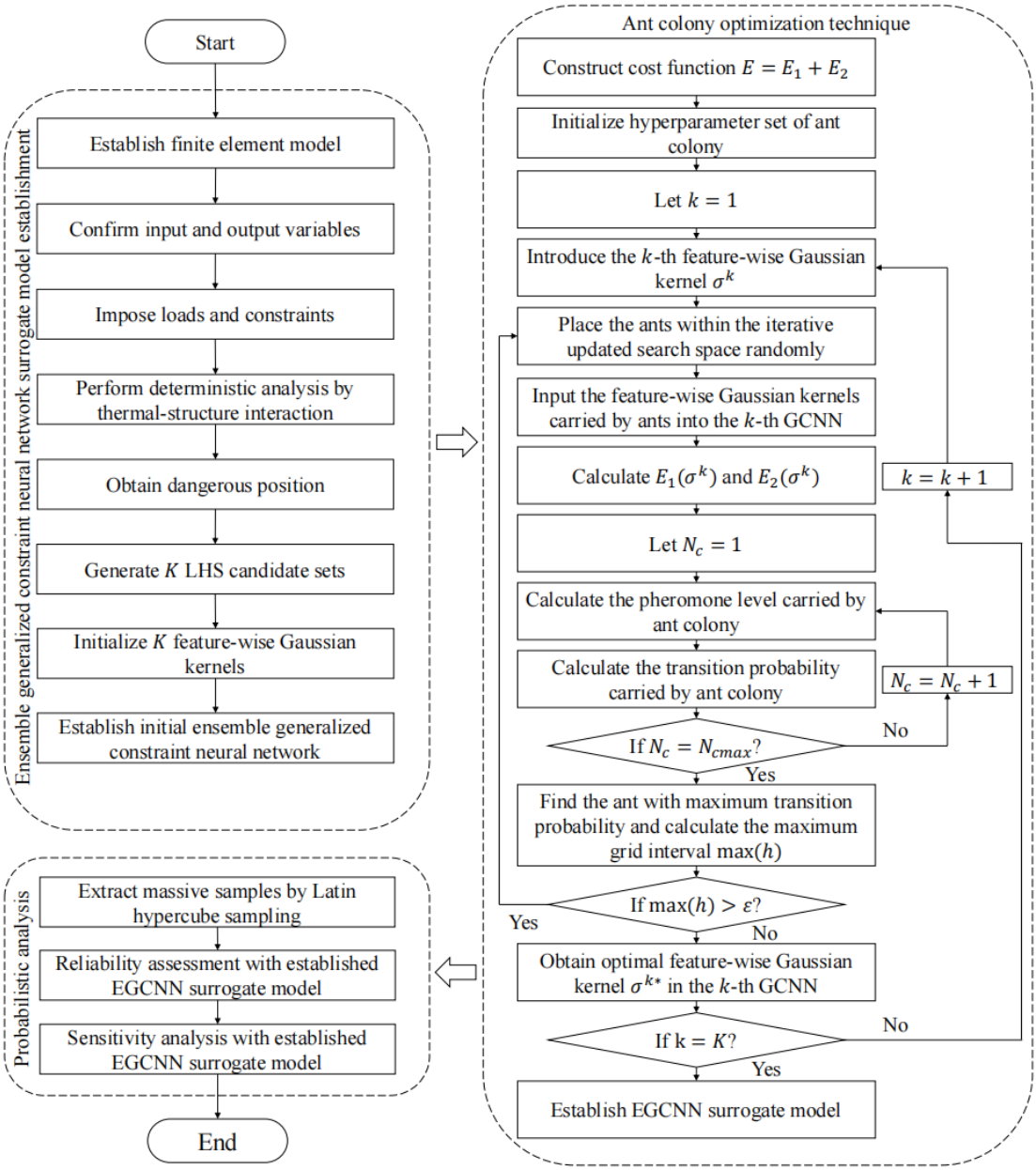


Fig. 3.10 The flowchart of reliability assessment with EGCNN.

Subsection 3.3.1 discussed the deterministic analysis of the blisk, focusing on predicting its LCF life under rated conditions. However, uncertainties in material and load parameters arise from manufacturing and operational variations. Assuming the critical parameters are mutually independent, three design boundaries are selected to evaluate the robustness of the EGCNN-based reliability assessment method, as shown in Table 3.1.

Table 3.1 Statistical characteristics of critical parameters under different design boundaries.

Parameter	ρ	E	λ	T_{root}	T_{tip}	ω	b	c	σ'_f	ε'_f
Unit	$kg{\cdot}m^{-3}$	MPa	$W{\cdot}m^{-1}{\cdot}°C^{-1}$	K	K	$rad{\cdot}s^{-1}$	/	/	MPa	/
Mean	8210	163000	23	1173.15	1473.15	1168	-0.1	-0.84	1419	50.5
Design boundary I (Std. Dev.)	262.72	3912	0.004	28.16	37.6	28	0.004	0.0336	34	1.224
Design boundary 2 (Std. Dev.)	295.56	4401	0.0045	31.68	42.3	31.5	0.0045	0.0378	38.25	1.377
Design boundary 3 (Std. Dev.)	328.4	4890	0.005	35.2	47	35	0.005	0.042	42.5	1.53
Distribution	Normal	Normal	Normal	Normal	Normal	Normal	Normal	Normal	Normal	Lognormal

The EGCNN structure is initialized using multiple LHS candidate sets and feature-wise Gaussian kernels, then trained with the ACO technique (comprising 50 GCNN models). Convergence plots for different design boundaries, shown in Fig. 3.11(a)-(c), indicate that as training iterations increase, NRMSE, LLF, and total error steadily decrease and converge. This highlights the optimization efficiency of the ACO technique and the robustness of the ACO-EGCNN model.

The convergence capability of the ACO-EGCNN model is evaluated using five-fold cross-validation. Fig. 3.12(a)-(c) show the expected NRMSE on training and validation sets under various design boundaries. The low NRMSE in validation sets confirms the robust generalization performance of the EGCNN model.

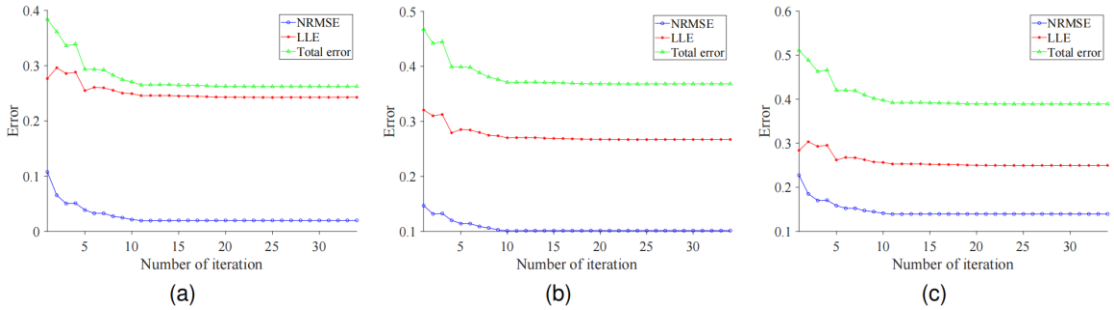


Fig. 3.11 The training convergence plot of EGCNN under different design boundaries. (a) Design boundary 1. (b) Design boundary 2. (c) Design boundary 3.

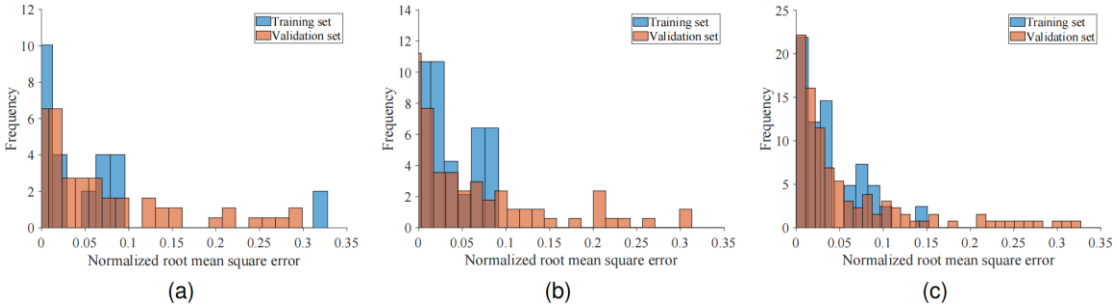


Fig. 3.12 Expected NRMSE histograms of EGCNN in the training and validation sets. (a) Design boundary 1. (b) Design boundary 2. (c) Design boundary 3.

K-fold cross-validation evaluates the generalization performance and aids in optimizing hyperparameters, such as the number of ants A, evaporation coefficient ρ, and pheromone intensity ratio $\frac{Q_1}{Q_2}$, using grid search. Other hyperparameters are set to empirical constants (search space: v^{lower}: [0.0, ...,1.0], v^{upper}: [0.2, ...,2.0]; grids: N =10; convergence threshold: ε =0.000001; max iterations: N_{cmax}=50). Fivefold cross-validation determines optimal values (A^*, ρ^*, $\frac{Q_1}{Q_2}^*$). Fig. 3.13(a)–(c) show expected values and standard deviations of EGCNN validation errors for varying A^*, ρ^*, and $\frac{Q_1}{Q_2}^*$.

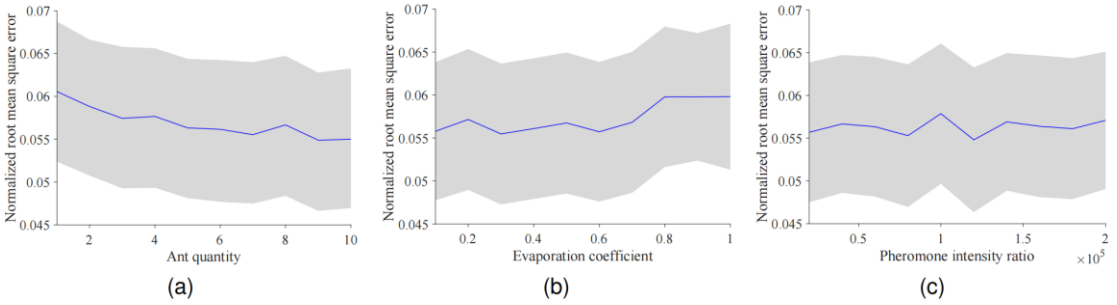


Fig. 3.13 Expectations and standard deviations of NRMSE in terms of ant quantity, evaporation coefficient, and pheromone intensity ratio. (a) Ant quantity. (b) Evaporation coefficient. (c) Pheromone intensity ratio.

As shown in Fig. 3.13(a), NRMSE decreases with more ants due to a higher likelihood of finding the global optimum, but training time increases. Considering resource constraints, ten ants are selected to optimize the Gaussian kernels. Fig. 3.13(b) shows that NRMSE increases with the evaporation coefficient ρ, indicating better accuracy with lower ρ. However, smaller ρ reduces training efficiency. To balance

accuracy and efficiency, ρ is set to 0.1.

To assess robustness, the EGCNN learning capability was tested with 10,000 simulations under varying design boundaries. Fig. 3.14(a)–(c) show the PDF, empirical cumulative distribution function (ECDF), and cumulative distribution function (CDF) of the estimated LCF life fitted to a lognormal distribution. One-sample K-S tests confirm that the LCF life distributions follow a lognormal distribution across all design boundaries.

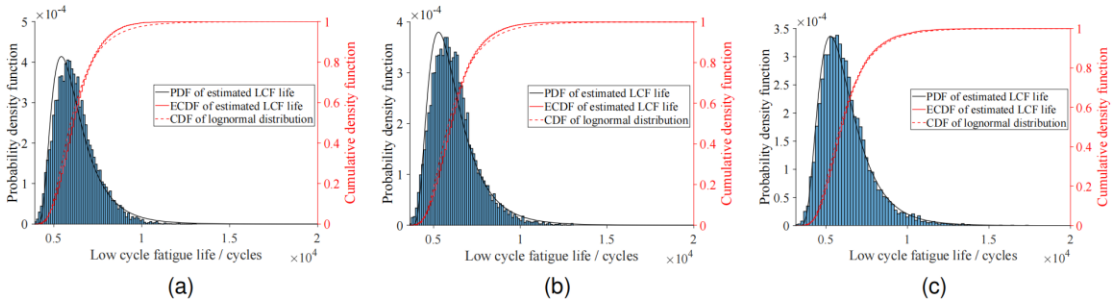


Fig. 3.14 The histogram and ECDF of probabilistic LCF life under different design boundaries. (a) Design boundary 1. (b) Design boundary 2. (c) Design boundary 3.

Table 3.2 shows that the lognormal distribution has higher LLF values than exponential, Weibull, and normal distributions, based on PDF discrepancies across design boundaries. This validates the robustness and effective knowledge-learning capability of the PKR sub-model.

Table 3.2 Statistical characteristics of LCF life estimated by EGCNN under different design boundaries.

Distribution	Design boundary 1		Design boundary 2			Design boundary 3	
Mean (Std. Err.)	Std. Dev. (Std. Err.)	LLF	Mean (Std. Err.)	Std. Dev. (Std. Err.)	LLF	Mean (Std. Err.)	Std. Dev. (Std. Err.)
6244.58 (62.45)	/	-97279.9	6173.34 (61.73)	/	-97318.8	6197.39 (61.97)	/
6729.5 (13.56)	5.27 (0.04)	-86427.8	6699.46 (15.29)	4.66 (0.03)	-87807.1	6781.05 (17.98)	4.01 (0.03)
6244.58 (11.41)	1140.98 (8.07)	-85689.6	6173.34 (12.74)	1274.19 (9.01)	-87158.2	6197.39 (14.76)	1475.76 (10.44)
8.72 (0.002)	0.18 (0.001)	-84967	8.71 (0.002)	0.20 (0.001)	-86165.4	8.71 (0.002)	0.22 (0.002)

After evaluating its capabilities, the EGCNN model is compared to MCS, ERSM,

GRNN, IGRNN, and GCNN on the same test set to validate its advantages as a reliability surrogate model. Table 3.3 presents its accuracy (NRMSE) and computational efficiency (time).

Table 3.3 Comparisons of accuracy and computing efficiency.

Metrics	Algorithm					
	MCS	ERSM	GRNN	IGRNN	GCNN	EGCNN
NRMSE	/	9.107	8.971	8.272	7.978	5.647
Computational time (s)	432,000	1.681	1.311	1.589	1.597	1.601

The results show that while the EGCNN model is slightly less efficient due to the IGRNN sub-model complexity and ensemble learning costs, it achieves superior accuracy. Its enhanced interpretability, incorporating LCF life distribution constraints, and improved accuracy through knowledge penetration make it ideal for aeroengine blisk reliability assessment.

After training and testing, the EGCNN model is used for reliability assessment and sensitivity analysis of aeroengine blisks. Material properties and operational loads are assumed independent, with input parameter mean and variance following Eqs. (3.29) and (3.30).

$$E(X) = \mu_X\left(\mu_\rho, \mu_\lambda, \mu_E, \mu_{\sigma_f'}, \mu_{\varepsilon_f'}, \mu_b, \mu_c, \mu_\omega, \mu_{T_{root}}, \mu_{T_{tip}}\right) \tag{3.29}$$

$$V(X) = \sigma_X^2\left(\sigma_\rho^2, \sigma_\lambda^2, \sigma_E^2, \sigma_{\sigma_f'}^2, \sigma_{\varepsilon_f'}', \sigma_b^2, \sigma_c^2, \sigma_\omega^2, \sigma_{T_{root}}^2, \sigma_{T_{tip}}^2\right) \tag{3.30}$$

where $E(X)$ is the mean function and $V(X)$ is the variance function.

The mean and variance of the reliability obey the following equations, respectively:

$$\begin{aligned} E(R) = \mu_R^2(&\mu_\rho, \mu_\lambda, \mu_E, \mu_{\sigma_f'}, \mu_{\varepsilon_f'}, \mu_b, \mu_c, \mu_\omega, \mu_{T_{root}} \\ &\mu_{T_{tip}}, \sigma_\rho^2, \sigma_\lambda^2, \sigma_E^2, \sigma_{\sigma_f'}^2, \sigma_{\varepsilon_f'}', \sigma_b^2, \sigma_c^2, \sigma_\omega^2, \sigma_{T_{root}}^2, \sigma_{T_{tip}}^2) \end{aligned} \tag{3.31}$$

$$V(R) = \sigma_R^2(\mu_\rho, \mu_\lambda, \mu_E, \mu_{\sigma_f'}, \mu_{\varepsilon_f'}, \mu_b, \mu_c, \mu_\omega, \mu_{T_{root}} \mu_{T_{tip}}, \sigma_\rho^2, \sigma_\lambda^2, \sigma_E^2, \sigma_{\sigma_f'}^2, \sigma_{\varepsilon_f'}^2, \sigma_b^2, \sigma_c^2, \sigma_\omega^2, \sigma_{T_{root}}^2, \sigma_{T_{tip}}^2) \tag{3.32}$$

Thus, the sensitivities of the uncertain parameters X with respect to reliability indicator R can be obtained by

$$\begin{cases} \dfrac{\partial R}{\partial \mu_X} = \sum_{i=1}^{N} \left(\dfrac{\partial R}{\partial \mu_{Y_f}} \dfrac{\partial \mu_{Y_f}}{\partial \mu_{X_i}} + \dfrac{\partial R}{\partial \sigma_{Y_f}} \dfrac{\partial \sigma_{Y_f}}{\partial \mu_{X_i}} \right) \\ \dfrac{\partial R}{\partial \sigma_X} = \sum_{i=1}^{N} \left(\dfrac{\partial R}{\partial \mu_{Y_f}} \dfrac{\partial \mu_{Y_f}}{\partial \sigma_{X_i}} + \dfrac{\partial R}{\partial \sigma_{Y_f}} \dfrac{\partial \sigma_{Y_f}}{\partial \sigma_{X_i}} \right) \end{cases} \tag{3.33}$$

This study analyzes the impact of CIs (0.8, 0.9, and 0.95) on the reliability and sensitivity, enabling clearer comparisons.

This study explores whether reliability assessment is affected by dataset-wide CIs or feature-specific CIs. Global CIs apply to the entire dataset, analyzing their overall impact, while local CIs assign a predefined CI to one feature, keeping others fixed at 0.95.

MCS are performed using the EGCNN model under global CIs of 0.8, 0.9, and 0.95, with structural responses calculated via Eqs. (3.13) and (3.14) and shown in Figs. 3.15(a)–(c). Results indicate consistently decreasing structural response values across all CIs, reflecting increased damage to the blisks.

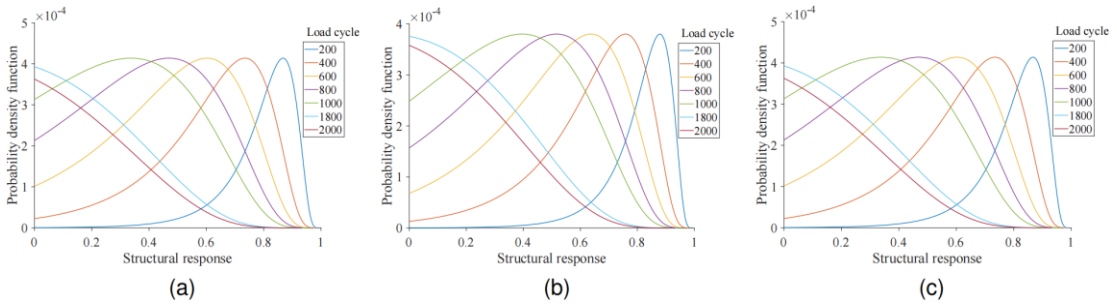


Fig. 3.15. Correlations between structural response and PDF under varying global CIs. (a) 0.8 global CI. (b) 0.9 global CI. (c) 0.95 global CI.

Fig. 3.16 shows the reliability declining steadily with increasing load cycles, with

the sharpest drop between 5,000 and 10,000 cycles. Higher global CIs further reduce the reliability for a given load cycle.

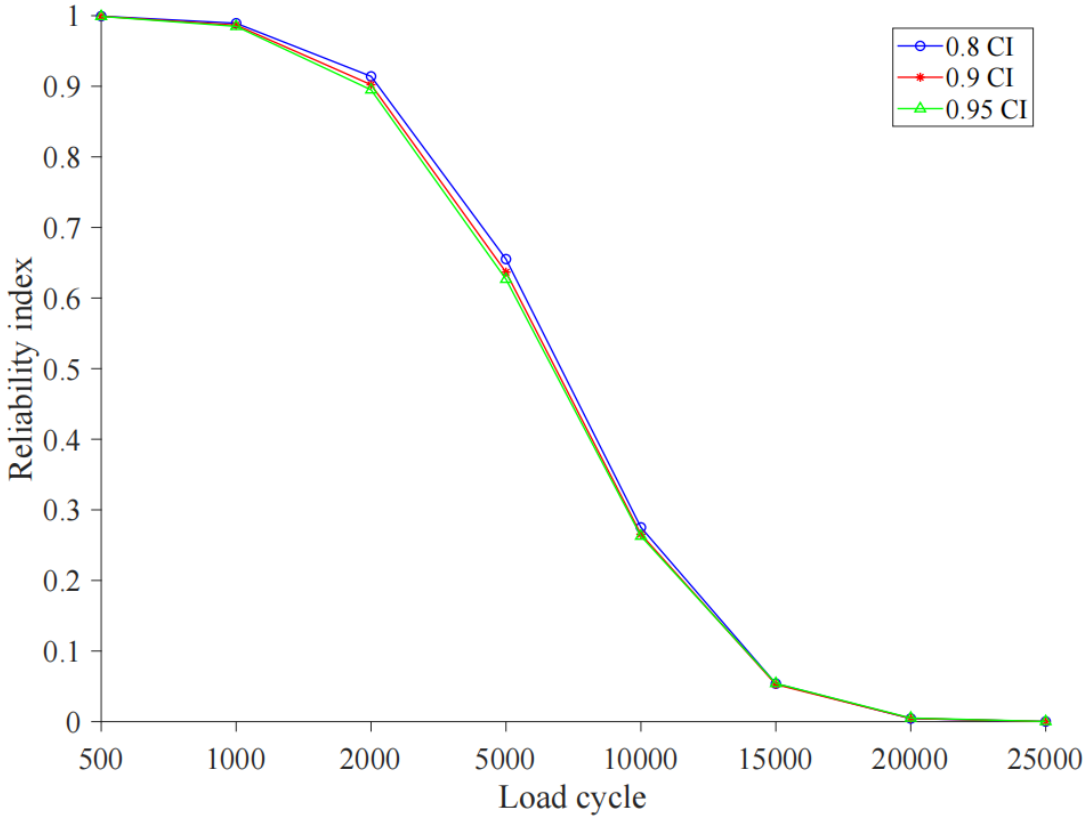


Fig. 3.16 Reliability trends of the blisk under varying global CIs.

Fig. 3.16 presents the blisk reliability for different global CIs, showing a steady decline with increasing load cycles, with the sharpest drop between 5,000 and 10,000 cycles. Higher global CIs further reduce the reliability for the same load cycle.

Local CIs also affect the reliability, with numerical experiments evaluating their impact on input features (Table 3.4). Results show positive correlations between the reliability (R) and the local CIs of density ρ, rotational speed ω, fatigue strength coefficient σ_f', and fatigue strength index b, while the local CI of blisk-root temperature T_{root} shows a negative correlation. Larger sampling ranges for these critical features are essential to better account for uncertainties during material testing and improve blisk reliability assessments. Sensitivity analysis follows to identify key material and

operational factors influencing reliability. Sensitivity degrees quantify the impact of input features, with positive values indicating a direct relationship with the reliability and negative values indicating an inverse relationship. Sensitivity degrees under various global CIs are summarized in Table 3.5.

Table 3.4 Reliability trends under different local CIs of input parameters.

Parameter	Local CI	Load cycle							
		500	1000	2000	5000	10000	15000	20000	25000
/	0.95	0.999	0.986	0.898	0.631	0.265	0.054	0.005	0
ρ	0.9	0.999	0.985	0.893	0.623	0.260	0.053	0.005	0
	0.8	0.999	0.984	0.893	0.622	0.259	0.052	0.005	0
E	0.9	0.999	0.985	0.895	0.626	0.262	0.054	0.005	0
	0.8	0.999	0.985	0.896	0.627	0.263	0.054	0.005	0
λ	0.9	0.999	0.985	0.894	0.624	0.260	0.052	0.005	0
	0.8	0.999	0.985	0.895	0.627	0.263	0.054	0.005	0
T_{root}	0.9	0.999	0.986	0.898	0.632	0.265	0.054	0.005	0
	0.8	0.999	0.987	0.903	0.640	0.271	0.056	0.005	0
T_{tip}	0.9	0.999	0.985	0.894	0.625	0.261	0.053	0.005	0
	0.8	0.999	0.985	0.895	0.626	0.260	0.052	0.005	0
ω	0.9	0.999	0.985	0.895	0.626	0.262	0.054	0.005	0
	0.8	0.999	0.984	0.893	0.622	0.258	0.052	0.005	0
σ'_f	0.9	0.999	0.985	0.896	0.628	0.262	0.053	0.005	0
	0.8	0.999	0.985	0.896	0.627	0.261	0.052	0.005	0
ε'_f	0.9	0.999	0.985	0.896	0.627	0.261	0.052	0.005	0
	0.8	0.999	0.987	0.903	0.641	0.273	0.057	0.005	0
b	0.9	0.999	0.985	0.896	0.628	0.262	0.053	0.005	0
	0.8	0.999	0.985	0.896	0.626	0.260	0.052	0.004	0
c	0.9	0.999	0.985	0.896	0.628	0.262	0.053	0.005	0
	0.8	0.999	0.986	0.899	0.633	0.265	0.054	0.005	0

Table 3.5 identifies blade-tip temperature (T_{tip}), blade-root temperature (T_{root}), and fatigue ductility index (c) as the most influential parameters, with sensitivities of −2.0125, 1.2700, and 1.6279, respectively. These factors should be prioritized and carefully controlled in blisk design, as other parameters have minimal impact on reliability.

Table 3.5 Sensitivity analysis of aeroengine blisks under different global CIs.

Parameter	Sensitivity degree under different global CIs		
	0.8	0.9	0.95
ρ	-0.2827	-0.2636	-0.2197
ω	-0.5312	-0.5158	-0.4772
T_{tip}	-2.0705	-2.0125	-1.8265
T_{root}	1.2700	1.0075	0.9590
E	0.1896	0.1953	0.1936
λ	0.0186	0.0169	0.0169
b	0.5233	0.4955	0.4524
c	1.6302	1.6279	1.5527
σ'_f	1.0845	1.0931	1.0904
ε'_f	0.1636	0.1661	0.1269

Density ρ, rotational speed ω, and blade-tip temperature T_{tip} negatively correlate with reliability, while other parameters show positive correlations, consistent with engineering practices. As global CIs increase, the sensitivity of T_{tip}, T_{root}, λ, b, and ε'_f decreases, whereas the sensitivity of the fatigue ductility index c increases. Parameters like density ρ, modulus E, and σ'_f exhibit minimal correlation with global CIs.

3.3.3 Metamodeling of Knowledge-Constraint Neural Networks for Low Cycle Fatigue Life Reliability Prediction

Normalized radial basis function neural networks (NRBFNNs) have garnered significant attention in the field of metamodels due to their notable attributes, including robust nonlinear fitting capabilities and the ability to distribute computation across multiple nodes. Considering that the knowledge presented in this study pertains to input-output probabilistic relationship, a regularization technique is employed to effectively incorporate prior knowledge into the metamodeling process by introducing penalty terms in the loss function. As a result, a knowledge-constraint neural network (KCNN) is devised as a metamodel, merging the prior probabilistic relationship with

the NRBFNN. The tailored metamodel ensures the seamless integration of prior knowledge and thus facilitates trustworthy reliability prediction.

This study develops a KCNN metamodel for efficient reliability modeling by integrating prior probabilistic relations into an NRBFNN. The NRBFNN excels at curve-fitting in high-dimensional spaces and offers better generalization than the RBFNN, as normalization emphasizes relevant data features. The Gaussian RBF, commonly used in NRBFNNs, is defined in Eq. (3.34).

$$\Psi\big(\boldsymbol{x}(i), \boldsymbol{c}(j)\big) = e^{-\theta_j^{(1)}||x(i)-c(j)||^2} \tag{3.34}$$

where $\theta_j^{(1)}$ is the jth Gaussian kernel to be optimized, $c(j)$ is the center vector for the j th hidden neuron, $||x(i) - c(j)||$ represents the Euclidean distance from the sample to the center point, and $x(i)$ denotes the i-th input of the network, representing the product's material and load vector in this study. After normalizing the network, the output of the NRBFNN is shown in Eq. (3.35).

$$\hat{g}(\boldsymbol{x}) = \sum_{j=1}^{n} \theta_j^{(2)} \mathcal{U}(\|\boldsymbol{x} - \boldsymbol{c}(j)\|) \tag{3.35}$$

where $\theta_j^{(2)}$ is the output weight of the jth hidden neuron and $\mathcal{U}(\|\boldsymbol{x} - \boldsymbol{c}(j)\|)$ is the NRBF, which can be calculated with Eq. (3.36).

$$\mathcal{U}(\|x - c(j)\|) = \frac{\Psi(\boldsymbol{x}, \boldsymbol{c}(j))}{\sum_{i=1}^{n} \Psi(\boldsymbol{x}, \boldsymbol{c}(j))} \tag{3.36}$$

To enhance the trustworthiness of estimated reliability, additional knowledge regularizer is incorporated into the NRBFNN in the form of probabilistic relations, as shown in Eq. (3.37).

$$\hat{\theta} = \underset{\theta}{argmin} \underbrace{\zeta_l \mathcal{L}_l\big(\hat{g}(\boldsymbol{x}|\boldsymbol{\theta}), g(\boldsymbol{x})\big)}_{label-based} + \underbrace{\zeta_k \mathcal{L}_k(\hat{g}(\boldsymbol{x}|\boldsymbol{\theta}), \boldsymbol{x})}_{knowledge-based} \tag{3.37}$$

where $\mathcal{L}_l$ is the label-based loss and $\mathcal{L}_k$ is the penalty loss of given prior knowledge

relations. Parameters ζ_l and ζ_k denote the weight of the term. The knowledge-based loss $\mathcal{L}_k$ should be sufficiently small to benchmark the model output against prior knowledge.

This chapter utilizes the normalized root mean square error as the label-based loss, as demonstrated in Eq. (3.38).

$$\mathcal{L}_l\left(\hat{g}\left(\boldsymbol{x}|\hat{\boldsymbol{\theta}}\right), g(\boldsymbol{x})\right) = \sqrt{\frac{\frac{1}{n}\sum_{i=1}^{n}\left(\hat{g}\left(\boldsymbol{x}(i)|\boldsymbol{\theta}\right)-g(x(i))\right)^2}{\bar{g}(\boldsymbol{x})}} \tag{3.38}$$

where n is the number of training samples, $\hat{g}(\boldsymbol{x}(i)|\boldsymbol{\theta})$ is the ith output of the KCNN, $g(\boldsymbol{x})$ is the target from the FEM-PoF model, and $\bar{g}(\boldsymbol{x})$ is the average target.

The optimization of the KCNN, detailed in Appendix A.3.2, uses a meta-heuristic approach for $\theta^{(1)}$ instead of gradient descent, leveraging global search capability and model independence. Since single meta-heuristic algorithms are suboptimal [153], a sequential ACO-PSO (SACO-PSO) is developed. SACO-PSO combines the discrete search efficiency of ACO with the continuous search capability of PSO, outperforming either algorithm alone [154][155].

After training the KCNN metamodel for failure time prediction, Miner's law estimates incremental damage and life reliability. The modeling process follows these steps:

(1) Evaluations of material properties and usage conditions x. Assigning pre-determined PDF $f_x(x)$ to x, which can be estimated using historical data and maximum likelihood estimation.

(2) Experimental design for metamodeling. Determining initial training samples with

$f_x(x)$ ($x \in [x_l, x_u]$) and LHS technique.

(3) Metamodeling of the KCNN. Training the KCNN using initial samples and optimization algorithms, monitoring loss convergence to ensure effectiveness. Evaluate the model's accuracy, efficiency, and trustworthiness against requirements. If met, the KCNN is used for reliability modeling; otherwise, new samples are generated via sequential sampling.

(4) Reliability prediction. After validating the KCNN metamodel, the KCNN can be utilized to efficiently predict asset reliability w.r.t. operational time.

Using the random properties of blisk materials and loads (Table 3.6), 150 candidates are generated via LHS. After iterative FEM simulations, the LCF life is used to train the KCNN with the SACO-PSO algorithm. The dataset is split into training and test sets, with the training set used for SACO-PSO-KCNN training and the test set for validation. The optimal KCNN parameters, θ, are then determined as shown below.

Table 3.6 Parameters used in the case study.

Parameter (Unit)	Distribution	Mean	Standard deviation
Density ρ,(kg·m^{-3})	Normal	8,210	328.4
Thermal conductivity λ,(W ·m^{-1}.°C^{-1})	Normal	23	0.005
Young's modulus E,(MPa)	Normal	163,000	4,890
Fatigue strength coefficient σ'_f,(MPa)	Normal	1,419	42.5
Fatigue ductility coefficient ε'_f	Normal	50.5	1.53
Fatigue strength index b	Normal	-0.1	0.005
Fatigue ductility index c	Lognormal	-0.84	0.042
Rotational speed ω,(rad·s^{-1})	Normal	1,168	35
Blisk-root temperature T_{root},(K)	Normal	1,173.15	35.2
Blisk-tip temperature T_{tip},(K)	Normal	1,473.15	47

$$\begin{cases} \boldsymbol{\theta}^{(1)} = (0.2548 \quad 0.5004 \quad \ldots. \quad 1.1163)_{1\times 10} \\ \boldsymbol{\theta}^{(2)} = (-1.3157 \quad 4.8250 \quad \ldots \quad 6.2161)_{1\times 120} \end{cases} \tag{3.39}$$

The effectiveness of SACO-PSO-KCNN is evaluated through model-fitting and simulation performance. Its fitting accuracy, measured by NRMSE with FEM output as a reference, is compared against RSM, SVM, ELM, GRNN, NRBFNN, ACO-KCNN, and PSO-KCNN. The MCS performance is assessed using FEM-PoF as the benchmark. All evaluations, using 120 input groups for model construction and 30 samples for generalization testing, are conducted in the same environment. Table 3.7 shows SACO-PSO-KCNN $\mathrm{NRMSE}_{\mathrm{test}}$ of 0.0066, demonstrating superior point-wise estimation. It ranks second in fitting accuracy, slightly behind NRBFNN, due to the additional probabilistic relationships embedded in SACO-PSO-KCNN.

MCS is used to evaluate the probabilistic prediction performance of FEM-PoF and the metamodels under identical computational conditions, with FEM-PoF-MCS results serving as the reference. Simulations are conducted with 100, 1,000, and 10,000 samples, with metrics such as precision and computing time summarized in Table 3.7. FEM-PoF becomes computationally impractical for simulations exceeding 10,000 samples, while metamodels are significantly more efficient. As simulations increase, computing time rises across all models. Precision is assessed using KL-divergence, where SACO-PSO-KCNN achieves the smallest KL-divergence, closely matching FEM-PoF predictions. This highlights SACO-PSO-KCNN's superior probabilistic prediction accuracy, attributed to its integration of probabilistic prior knowledge.

Table 3.7. Metamodel fitting and simulation performance.

Model	Modeling accuracy		Simulation precision and efficiency		
	$\mathrm{NRMSE}_{\mathrm{train}}$	$\mathrm{NRMSE}_{\mathrm{test}}$	Number of simulations		
			100	1,000	10,000
			KL divergence (Simulation time, s)		
FEM-PoF	/	/	/ (5,400)	/ (14,400)	/ (432,000)
RSM	0.1346	0.1944	1.4579 (1.249)	1.8509 (1.266)	1.7916 (1.681)
SVM	0.2088	0.1927	12.4743 (1.249)	12.8429 (1.265)	12.9709(1.687)
ELM	0.1004	0.0637	14.3467 (1.250)	14.4817 (1.282)	14.4688 (1.823)
GRNN	0.1326	0.1915	1.4362 (1.121)	1.8234 (1.151)	1.7650 (1.311)
NRBFNN	0.0030	0.0016	1.2088 (1.184)	1.2858(1.200)	1.2424 (1.594)
ACO-KCNN	0.0344	0.0348	0.7091 (1.187)	0.8658 (1.203)	1.0076 (1.597)
PSO-KCNN	0.0227	0.0230	0.2992(1.148)	0.2897(1.164)	0.3365(1.545)
SACO-PSO-KCNN	0.0092	0.0066	0.0909 (1.168)	0.0487 (1.184)	0.0288 (1.572)

To evaluate the trustworthiness of SACO-PSO-KCNN, the prior knowledge embedded in the KCNN is tested. Since the turbine blisk LCF life typically follows a lognormal distribution, a constraint is imposed to ensure the predicted probabilistic LCF life adheres to this distribution [156][157]. Fig. 3.17a shows the statistical characteristics of the predicted life, including the histogram, PDF, ECDF, and lognormal CDF. The alignment between the predicted life ECDF and the lognormal CDF demonstrates successful knowledge learning. However, graphical representation cannot quantitatively evaluate the fit, so the one-sample K-S test is employed to assess the goodness of fit. The test compares the maximum absolute difference, $D = \sup_x |F_1(x) - F_2(x)|$, between two CDFs, $F_1(x)$ and $F_2(x)$, with the critical value D_α for a given confidence level α. If $D \leq D_\alpha$, the null hypothesis is accepted. Results for sample sizes of 1,000 and 10,000 (Table 3.8) show $D<D_{0.01}$and $D_{0.01}$, confirming the LCF life follows a lognormal distribution with 99% confidence. In summary, SACO-PSO-KCNN is efficient, accurate, and trustworthy, validating its reliability for life

prediction.

$$D_\alpha = \begin{cases} \frac{1.63}{\sqrt{n}}, & \alpha = 0.001 \\ \frac{1.63}{\sqrt{n}}, & \alpha = 0.005 \end{cases} \tag{3.40}$$

Table 3.8 K-S test results.

Number of samples (n)	Confidence level		D
	$D_{0.001}$	$D_{0.005}$	
1,000	0.0515	0.0430	0.0211
10,000	0.0163	0.0136	0.0120

The SACO-PSO-KCNN effectively derives the limit state function (LSF) for efficient reliability prediction. Fig. 3.17b shows the life histogram, illustrating the relationship between the PDF of LCF life and the LSF under varying loads. As cycles increase, both the mean and standard deviation of the LSF rise, indicating reduced asset reliability.

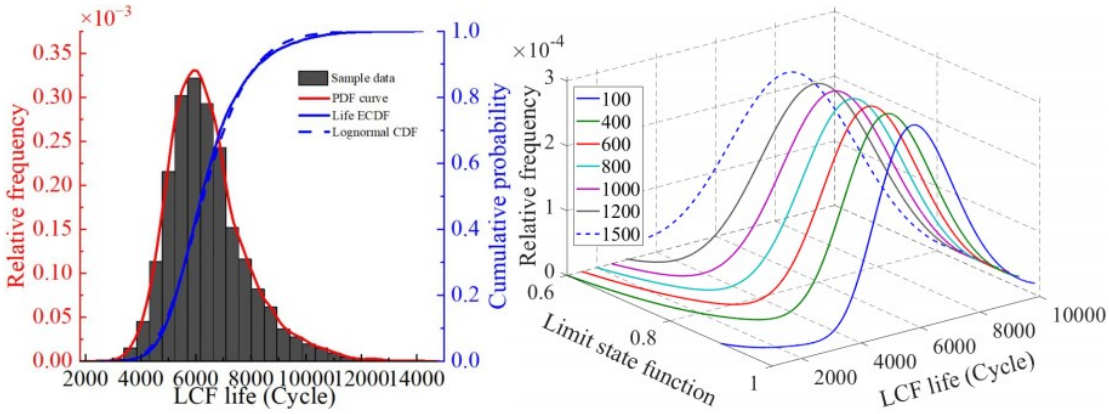


Fig. 3.17 Probabilistic prediction with KCNN. (a) Fitted life PDF, life ECDF, and lognormal CDF and (b) Relationship among life, LSF, and relative frequency.

3.3.4 Metamodeling of Adaptive Kriging for Low Cycle Fatigue Life Reliability Prediction

The kriging model is trained using an initial DoE selected via LHS. To automate sample selection, the U-learning function identifies samples near the limit state surface or with high prediction variance. Here, the kriging output $\hat{g}(x)$ represents life, with the

limit state surface at zero, making accuracy near zero critical. However, U-learning does not account for distance, and closely spaced samples add little new information. To enhance efficiency, an improved U-learning function $\mathcal{U}(x)$ is used, as shown in Eq. (3.41).

$$X_c^* = arg \min_{X \in \mathcal{D}_X} \mathcal{U}(X) = arg \min_{X \in \mathcal{D}_X} \frac{\mu_{\hat{g}}(X)}{\sigma_{\hat{g}}(X) d(X)} \tag{3.41}$$

where X_c^* is the new candidate, $\hat{g}(\cdot)$ is the kriging model replacing the FEM-PoF response $g(\cdot)$, $\mu_{\hat{g}}(\cdot)$ is the expectation of the kriging output, $\sigma_{\hat{g}}(\cdot)$ is the standard deviation of the kriging output, and $d(\cdot)$ is the distance calculated in the normalized space, ranging from zero to one, as formulated in Eq. (3.42).

$$d(X) = \frac{\min_{X(p) \in X_c} \|X - X(p)\|}{\max_{X(p), X(q) \in X_c} \|X(p) - X(q)\|} \tag{3.42}$$

where X_c is the current candidate set, $X(p)$ and $X(q)$ are the pth and qth candidate in the candidate set X_c, and $||\cdot||$ is the Euclidean operator. The improved U-learning function selects new candidates by considering three aspects: (1) The kriging output is close to zero (with small $\mu_{\hat{g}}(X)$), as it estimates the asset life, where accuracy is particularly critical for smaller lifetime. (2) The prediction exhibits large uncertainties (with large $\sigma_{\hat{g}}(X)$). (3) The sample is far away from the existing candidates (with large $d(X)$).

Kriging is chosen as the metamodel for replacing complex FEM-PoF simulators due to its effectiveness in handling highly nonlinear and high-dimensional problems. It functions as a single-layer neural network with infinite units while providing variance, making it ideal for probabilistic analysis and design optimization. The kriging prediction $\hat{g}(\cdot)$ at an untried point follows a normal distribution characterized by its

mean $\mu_{\hat{g}}$ and variance $\sigma_{\hat{g}}^2$, i.e., $\mathcal{G}\sim\mathcal{N}(\mu_{\hat{g}},\sigma_{\hat{g}}^2)$ [158][159][160]. The kriging has the following regression.

$$\hat{g}(X) = h(X) + n(X) \tag{3.43}$$

where $h(X)$ is the basis function as shown in Eq. (3.43) (quadradic polynomial is employed in this study since the quadradic trend produces the most favorable results) while $n(X)$ accounts for the realization of predictive uncertainties as shown in Eq. (3.45).

$$h(X) = \beta_1 + \beta_2 X + X^T \beta_3 X \tag{3.44}$$

where β_1, β_2, and β_3 are undetermined weights.

$$\begin{cases} E[n(X)] = 0 \\ Var\,[n(X)] = \sigma^2 \\ Cov\,[n(X(p), X(q))] = \sigma^2 r(\nu, X(p), X(q)) \end{cases} \tag{3.45}$$

where $\mathrm{X(p)}$ and $\mathrm{X(q)}$ are the pth and qth candidates $(\mathrm{p}, \mathrm{q} = 1,2,\dots,\mathrm{N_c}$, with $\mathrm{N_c}$ being the number of candidates), ν is the correlation parameter vector, and $\mathrm{r}(\cdot)$ is the distance between $\mathrm{X(p)}$ and $\mathrm{X(q)}$ satisfying the following Eq. (3.46).

$$\begin{aligned} r(\nu, X(p), X(q)) &= \prod_{j=1}^{m} r\left(\nu_j \left| x_j(p) - x_j(q) \right|\right) \\ &= e^{-\sum_{j=1}^{m} \nu_j \left| x_j(p) - x_j(q) \right|^2} \end{aligned} \tag{3.46}$$

where $r\left(\nu_j \left| x_j(p) - x_j(q) \right|\right) = e^{-\nu_j \left| x_j(p) - x_j(q) \right|^2}$, $(j = 1,2,\dots,m)$.

A gradient descent algorithm is employed to optimize the undetermined parameters $\Theta = (\beta_1, \beta_2, \beta_3, \sigma^2, \nu)$ and the optimization process is detailed in Appendix A.3.3 [161]. The target value used to train the AK model is the life of the asset, simulated with the FEM-PoF. With Miner's law, the incremental degradation and reliability of turbine blisks can be predicted.

The step-by-step implementation of the AK-based DT for probabilistic degradation

simulation is outlined as follows:

(1) Based on the IU function, training samples are generated until the minimum IU learning function output exceeds two. Using adaptive sampling, the AK model $\hat{g}(X;\widehat{\Theta}^{(0)})$ is modeled.

(2) Once the AK model is trained, it is deployed for real-time probabilistic degradation simulation, continuing until new operational scenarios are detected by the IU function.

(3) After detecting new operational scenarios, data distribution $f_X(X|\hat{\theta})$ is updated.

(4) Preserving historical training samples, new training samples are adaptively selected with the IU function, and the AK is updated with new training samples.

(5) If no new operational scenarios are observed, the latest AK model $\hat{g}(X;\widehat{\Theta}^{(k+1)})$ is used to probabilistically model degradation.

In this case study, the input load statistics for the blisk across different load conditions are presented in Table 3.9. In addition to the operational load, information on material properties is required for FEM simulation. The material parameters used in the experiment are presented in Table 3.10.

Table 3.9 Load statistics under different load conditions.

Variable	Distribution	Expected value	Standard deviation		
			Stage 0	Stage 1	Stage 2
ω	Normal	1,168	5	15	25
T_{root}	Normal	1,173	5	15	25
T_{tip}	Normal	1,473	17	27	37

Table 3.10 Material parameter values in the case.

Parameter, unit	ρ, kg/m³	E,MPa	λ,W/(m·°C)	b	c	σ'_f,MPa	ε'_f

Value	8,210	163,000	23	-0.1	-0.84	1,419	50.5

The fitting and simulation performance of the AK emulator are evaluated using operational stage 2, as its larger domain is representative. The IU-AK model is validated by comparing it with LHS-RSM, LHS-OK, LHS-SVM, LHS-PCE, and U-AK emulators for FEM-based blisk LCF prediction. Fitting accuracy and efficiency are tested under identical computing conditions. For LHS, 120 candidates construct the emulators, while adaptive sampling uses 10 initial LHS samples, followed by enrichment with U, or IU criteria, and 30 additional samples for performance evaluation. Model-fitting results are summarized in Table 3.11, with accuracy measured by average relative error (Eq. (3.47)) and efficiency by training time.

$$\bar{e}_r = \frac{1}{n}\left(\frac{|\hat{g}(X)-g(X)|}{g(X)}\right) \tag{3.47}$$

where n is the number of test samples, $\hat{g}(X)$ is the response of the model and $g(X)$ is the true response of the FEM.

As shown in Table 3.11, the IU-AK achieves the lowest average relative error (1.430%), highlighting its superior fitting accuracy and efficiency compared to other models. Beyond fitting, simulation performance is evaluated using MCS to predict the LCF life distribution with eight models, including FEM and seven emulators. Numerical characteristics of input variables (Table 3.12) are used, with computations conducted under identical conditions. FEM-MCS serves as the reference for simulation performance, using 100 and 1,000 samples. Table 3.12 shows that IU-AK provides the most accurate LCF life distribution, closely matching FEM, due to the IU function. Additionally, K-L divergence decreases with more simulations. Table 3.12 also highlights IU-AK's superior simulation efficiency, achieving the highest accuracy and

efficiency among all models for both 100 and 1,000 simulations.

Table 3.11 Model-fitting performance of metamodels.

Emulator	Fitting accuracy	Improved accuracy, %	Fitting time, s	Improved efficiency, %
LHS-RSM	3.080	—	1.628	—
LHS-SVM	3.475	-12.82%	0.638	60.81%
LHS-PCE	2.345	23.86%	0.504	69.04%
LHS-OK	1.519	50.68%	0.874	46.31%
U-AK	1.431	53.54%	0.765	53.01%
IU-AK	1.430	53.57%	0.459	71.81%

Table 3.12 Comparisons of simulation precision and time for different metamodels.

Model	KL divergence (Improved precision, %)		Simulation time, s (Improved efficiency, %)	
	100 samples	1,000 samples	100 samples	1,000 samples
FEM	—	—	5,400	14,400
LHS-RSM	1.4579 (-)	1.8509(-)	1.365 (-)	3.760(-)
LHS-SVM	1.3826 (5.16%)	1.8181 (1.77%)	0.935(31.50%)	2.246 (40.27%)
LHS-PCE	1.3839 (5.08%)	1.8198(1.68%)	0.825 (39.56%)	1.531 (59.28%)
LHS-OK	1.3812 (5.26%)	1.8163 (1.87%)	0.660 (51.65%)	1.225 (67.42%)
U-AK	1.3491 (7.46%)	1.7744 (4.13%)	0.578 (57.66%)	1.072 (71.49%)
IU-AK	1.3490 (7.47%)	1.7739 (4.16%)	0.346 (74.65%)	0.643 (82.90%)

After training and validating the IU-AK model, Fig. 3.18 shows changes in blisk degradation and reliability with applied cycles across operational stages. In Fig. 3.18(a), degradation increases and reliability decreases with more cycles and higher load boundaries. Fig. 3.18(b) shows reliability declines as the degradation threshold lowers.

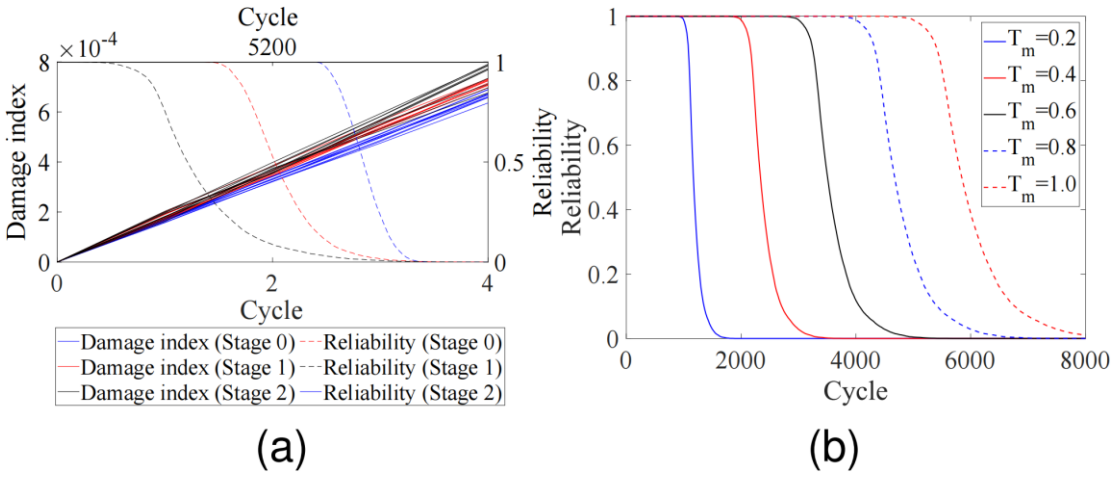


Fig. 3.18 LCF reliability analysis with the AK metamodel. (a) Probabilistic damage indexes and reliability trends under different operational scenarios. (b) Reliability trends under different degradation thresholds.

3.4 Case Study 2: High Cycle Fatigue Life Reliability Prediction for Variable-Speed Fan Shafts

3.4.1 Finite Element Modeling for Deterministic Stress Prediction

A fan shaft from an FGD system in a thermal power plant is used as a case study to demonstrate the metamodel-based HCF life reliability prediction method. The FGD system removes sulfur compounds, promoting cleaner production. Continuous operation and fluctuating output power cause frequent speed changes, leading to fatigue damage in the fan shaft. As the shaft is not exposed to high thermal environments, the analysis focuses on HCF.

The metamodel is trained using real-time rotational speed data of the FGD fan shaft and deterministic material parameters (Table 3.13). Stress sequences are generated using COMSOL software with an axisymmetric shaft model subjected to centrifugal forces under varying speeds, leading to HCF. To ensure FEM accuracy, boundary conditions include (i) axial displacement constraints and (ii) centrifugal forces from rotor and shaft rotation. The collected speed data serve as input for the FEM, identifying the critical point with maximum stress, as shown in Fig. 3.19. Using the FEM, HCF stress responses are obtained (Fig. 3.20), with the critical point experiencing maximum stress, potentially leading to HCF failure. This location is used for reliability evaluation, with FEM results highlighting the highest stress magnitude (black line in Fig. 3.19).

Table 3.13 Material parameters for stress estimation of FGD fan shafts.

Parameter	Unit	Value
Density ρ	kg/m^3	7,850

Modulus of elasticity E	M Pa	200,000
Poisson ratio ν	/	0.3
Ultimate tensile strength S_u	MPa	245
Fatigue strength coefficient S'_f	kPa	100
Fatigue strength index b	/	-0.4

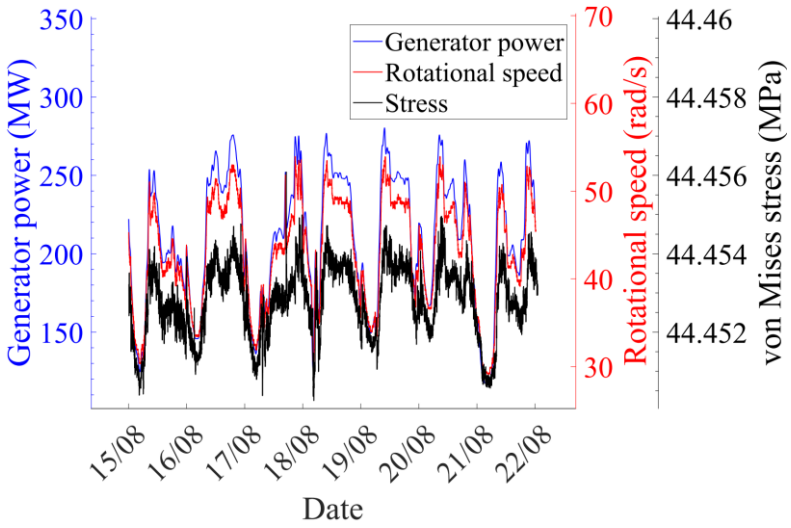


Fig. 3.19 Time-series generator power, fan speed, and simulated von Mises stress of the critical point in the fan shaft.

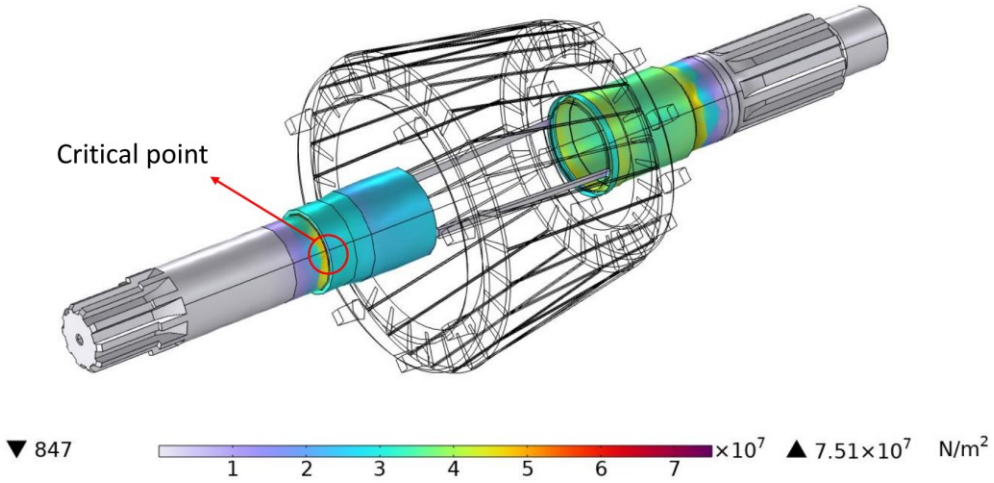


Fig. 3.20 Von Mises stress distribution of the FGD fan shaft.

3.4.2 Metamodeling of Bayesian BiLSTM for High Cycle Fatigue Life Reliability Prediction

Traditionally, auto-regressive models like ARMA and ARIMA are used for RUL prediction, with ARMA suited for stationary processes and ARIMA for non-stationary

ones [162]. However, these models are limited to linear processes. For nonlinear processes, nonlinear auto-regressive (NAR) models, such as NARX and NARMAX, integrated with neural networks, are applied [149][163]. Yet, NAR models, as feed-forward neural networks, lack internal memory to handle long-sequential patterns efficiently, requiring larger models as complexity increases. Recurrent neural networks (RNNs) address this issue by using a recursive mechanism to memorize time-series data without increasing model size. However, training vanilla RNNs is challenging due to gradient exploding or vanishing problems. To overcome these issues, gated recurrent unit (GRU) and long short term memory (LSTM) models were developed. GRU models are computationally efficient due to their lower complexity, while LSTMs are more accurate but computationally intensive. LSTMs have been widely applied in RUL prediction. Wu et al. used LSTMs with time-step labeling for RUL prediction and fed features into LSTMs for sequential sampling [164][165]. Miao et al. proposed dual-task deep LSTMs for aeroengine degradation assessment and RUL prediction, validated using the C-MAPSS dataset [166].

Mainstream models typically provide point estimates for RUL without prediction intervals. Bayesian techniques are commonly used to quantify model uncertainty, and recent studies have incorporated this into prognostics using Bayesian DL. Zhu et al. proposed a Bayesian deep active learning method with a recursive training approach to reduce training data size [167]. Kim and Liu introduced a Bayesian DL framework to generate interval estimates for RUL [168]. Zhang et al. have combined LSTM with

Lebesgue sampling and MCS to address uncertainty in battery capacity estimation [169].

The BiLSTM model consists of forward and backward components. The forward and backward components process sequential data to extract time-dependent features and capture short-term dependencies, which is illustrated in Fig. 3.21 [170][171]. Given the training input $X = \{x(i)\}_{i=1}^{n}$, each training sample $x(i)$ consists of l time steps as $x(i) \in \mathbb{R}^{1\times m}$, $k = 1,2,\dots,l$. With the forward network, $x^k(i)$ is processed as follows.

$$f_{fw}^k = sigm\left(\mathcal{W}_{fw}^f x^k(i) + \mathcal{V}_{fw}^f q_{fw}^{k-1} + \mathcal{B}_{fw}^f\right) \tag{3.48}$$

$$i_{fw}^k = sigm\left(\mathcal{W}_{fw}^i x^k(i) + \mathcal{V}_{fw}^i q_{fw}^{k-1} + \mathcal{B}_{fw}^i\right) \tag{3.49}$$

$$c_{fw}^k = f_{fw}^k \otimes c_{fw}^{k-1} + i_{fw}^k \otimes tanh\left(\mathcal{W}_{fw}^c x^k(i) + \mathcal{V}_{fw}^c q_{fw}^{k-1} + \mathcal{B}_{fw}^c\right) \tag{3.50}$$

$$o_{fw}^k = sigm\left(\mathcal{W}_{fw^o}^o x^k(i) + \mathcal{V}_{fw}^o q_{fw}^{k-1} + \mathcal{B}_{fw}^o\right) \tag{3.51}$$

$$q_{fw}^k = o_{fw}^k \otimes tanh\left(c_{fw}^k\right) \tag{3.52}$$

where $\mathcal{W}_{\text{fw}}$, $\mathcal{V}_{\text{fw}}$, and $\mathcal{B}_{\text{fw}}$ with different superscripts are weight vectors and bias vectors in different functions, sigm(·) and tanh (·) are sigmoid and hyperbolic tangent functions, $\otimes$ is a point-wise multiplication, f_{fw}^k and i_{fw}^k are forget and input gates, which jointly control the information stored in the memory unit c_{fw}^k. o_{fw}^k is an output gate and controls how much information in c_{fw}^k being output as q_{fw}^k.

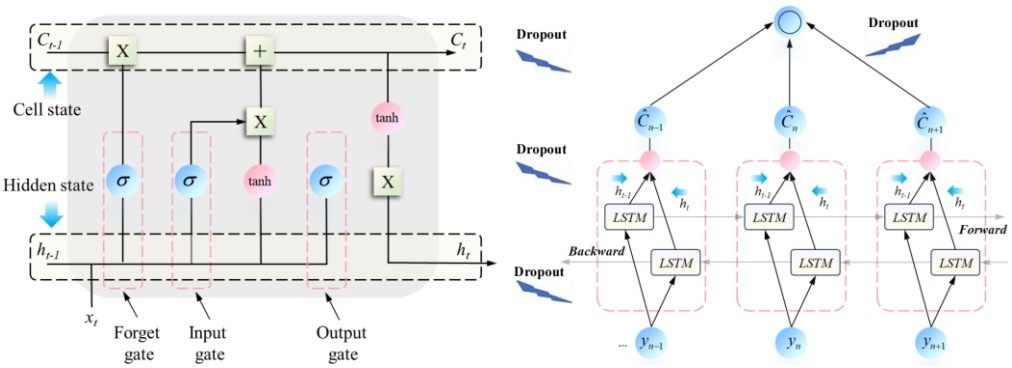


Fig. 3.21 Bayesian BiLSTM model.

In the backward network, $x^k(i)$ is processed similarly. Major differences are that c_{bw}^{k+1} and q_{bw}^{k+1} associated with $x^{k+1}(i)$ are used in the backward network rather than c_{fw}^{k-1} and q_{fw}^{k-1}, and the related weight and bias vectors are different. The output sequences $q_{\mathrm{fw}}^{1:l}$ and $q_{\mathrm{bw}}^{1:l}$ from the BiLSTM are further concatenated as $q(i) = [q_{\mathrm{fw}}^{1:l}, q_{\mathrm{fw}}^{1:l}]$. Then, the concatenated vector $q(i)$ is further processed by a fully-connected network.

Due to limitations in units and durations, the data may be insufficient for validation. If the test data is limited, relying on the estimated parameters for maintenance planning may pose risks. In this context, Bayesian methods can be incorporated into the BiLSTM model, thereby enhancing the reliability of the results. The parameter $\Theta = (\mathcal{W}_{\mathrm{fw}}, \mathcal{V}_{\mathrm{fw}}, \mathcal{B}_{\mathrm{fw}}, \mathcal{W}_{\mathrm{bw}}, \mathcal{V}_{\mathrm{bw}}, \mathcal{B}_{\mathrm{bw}})$ of the BiLSTM model is treated as a random variable represented by probability distributions, supporting risk-informed analysis and decision-making. Bayesian theorem is used to update the probability distributions of the model parameters based on the training data. Let $X = \{x(i)\}_{i=1}^n$ and $S = \{s(i)\}_{i=1}^n$ denote the training inputs and targets, respectively, and let n represent the number of training samples. A Bayesian BiLSTM model can be formulated as $\hat{S} = \hat{g}(X; \Theta)$, with a prior distribution over the model parameter space $p(\Theta)$ and a likelihood function for Bayesian regression. $\mathcal{L}(S; X, \Theta) = \prod_{i=1}^n l(s(i); \hat{g}(x(i); \Theta))$. Here, the Gaussian distribution is employed for the prior $p(\Theta)$ [172]. The Bayesian BiLSTM model can be trained by processing the joint combination of X and S. The posterior distribution of model parameters is obtained with Eq. (3.53).

$$p(\Theta; X, S) \propto p(\Theta) \cdot \mathcal{L}(\Theta; X, S) = \frac{p(\Theta) \prod_{i=1}^n l(s(i); \hat{g}(x(i); \Theta))}{\int p(\Theta) \prod_{i=1}^n l(s(i); \hat{g}(x(i); \Theta)) d\Theta} \quad (3.53)$$

Based on $p(\Theta; X, S)$, the Bayesian BiLSTM $\hat{g}(\cdot)$ can then be used for inference with uncertainty quantification. Suppose newly observed sample $\hat{x}$ is available, the output $\hat{s}$ can be predicted through $p(\hat{s}; \hat{x}, X, S) = \int p(\hat{s}; \hat{x}, \Theta) \cdot p(\Theta; X, S) d\Theta$. Note that the posterior distribution $p(\Theta; X, S)$ is generally intractable, but it can be approximated using variational inference, as detailed in Appendix A.3.4.

After building the Bayesian BiLSTM model, it is necessary to analyze the performance of this model using a loss function, as shown in Eq. (3.54).

$$\mathcal{L}(X, S) = \frac{1}{2n} \sum_{i=1}^{n} \left\| s(i) - \hat{g}_{h(\varepsilon, \Phi)}(x(i)) \right\|^2 \tag{3.54}$$

where $\hat{g}_{h(\varepsilon,\Phi)}(x(i))$ represents the BiLSTM model with dropouts, Φ is the model parameter set to be learned, and ε is a binary vector sampled from Bernoulli distributions.

After training, the Bayesian BiLSTM predicts $\hat{S}$ for new inputs $\hat{X}$, enabling probabilistic predictions. The pseudocode is provided in Algorithm 3.3 and can be implemented in PyTorch. The model uses Euclidean loss (Eq. (3.53)) and stochastic gradient descent, with hyperparameters optimized via grid search to minimize validation loss. Once trained, the optimal parameters are used for probabilistic predictions. After estimating the stress series $\hat{S}$, $\widehat{D}(t)$ is calculated using the rainflow counting algorithm and Eqs. (3.7) and (3.9).

Algorithm 3.3: Bayesian BiLSTM model.

Input:

Training inputs X=$\{x(i)\}_{i=1}^{n}$ and targets S=$\{s(i)\}_{i=1}^{n}$,new input $\hat{x}$,learning rate η ,dropout rate for $\mathcal{B}$ (ε), initial model parameters Φ.

Output: p($\hat{s}|\hat{x}$,X,S).

1: Repeat

2: Sample n random points $\varepsilon_i \sim \mathcal{B}(\varepsilon)$ i = 1,2,…,n.

3: Sample a random subset of {1,2,…,n } with size n_b.

4: Calculate derivative with respect to Φ :

5: $\Delta\Phi = -\frac{1}{n_b} \sum_{i \in b} \frac{\partial}{\partial \Phi} \log \left(l\left(s(i); \hat{g}_{h(\Phi, \varepsilon_i)} x(i) \right) \right)$.

6: UpdateΦ: $\Phi=\Phi+\eta\triangle\Phi$.
7: Until Φ has converged to be Φ^*
8: Sample η_k random points $\varepsilon_k \sim \mathcal{B}(\varepsilon)$, k= 1,2,…,η_k.
9: For k = 1 to η_k do
10: $\hat{s}_k = \hat{g}h\,(\Phi^*, \varepsilon_k)(\hat{x})$.
11: End for

In this case study, the Bayesian BiLSTM-based DT estimates the RUL distribution by integrating the rainflow counting algorithm, Basquin model, and Miner's rule. The model is implemented on a system with 16 GB RAM, an Intel i9-12900H CPU, an Nvidia 3060Ti GPU, and developed using PyTorch. A sliding time window technique augments the training set. Fig. 3.22a shows consistent reductions in training and validation mean squared error (MSE), demonstrating the model's effectiveness. Optimal hyperparameters, including optimizer, sample size, batch size, epochs, hidden layers, learning rate, and dropout probability, are determined via grid search and cross-validation. Fig. 3.22b compares training performance across optimizers (AdaGrad, Adam, Adamax, Nadam, Radam, and RMSProp), with Adam showing the fastest convergence and lowest loss, making it the selected optimizer.

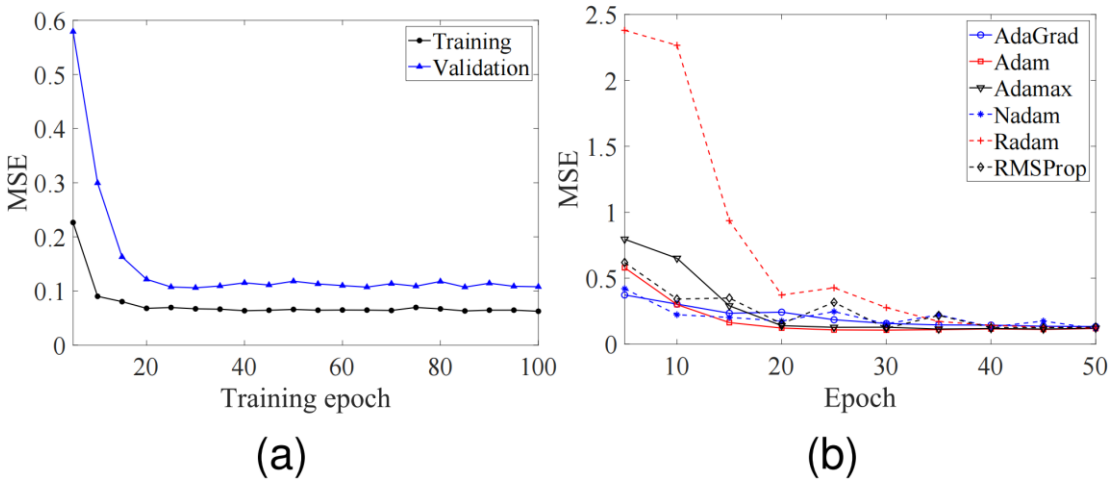


Fig. 3.22 MSE during training processes. (a) Training and validation losses across iterations. (b) Training process with different optimizers.

In addition to the optimizer, other hyperparameters like batch size, training epochs, hidden layers, and sequence length are optimized (Fig. 3.23). To address variability from data sampling and model initialization, 50 repeated experiments are conducted.

(1) Batch size (Fig. 3.23a): Batch sizes of 128 and 256 yield lower mean MSE and quantiles, but smaller batch sizes significantly increase training time. Hence, a batch size of 256 is selected.

(2) Training Epochs (Fig. 3.23b): Increasing epochs reduces mean and quantile MSE, with stable values achieved at 20 and 25 epochs. Considering training efficiency, 20 epochs are chosen.

(3) Hidden Layers (Fig. 3.23c): Two hidden layers achieve the lowest MSE. While deeper networks capture complex patterns, they risk overfitting.

(4) Sequence Length (Fig. 3.23d): MSE decreases as sequence length increases, leveraging more information. However, beyond a length of five, MSE rises, likely due to the model learning incorrect patterns or trends.

The optimized Bayesian BiLSTM architecture includes two hidden BiLSTM layers and one fully connected layer, trained using the Adam optimizer. Selected hyperparameters are summarized in Table 3.14.

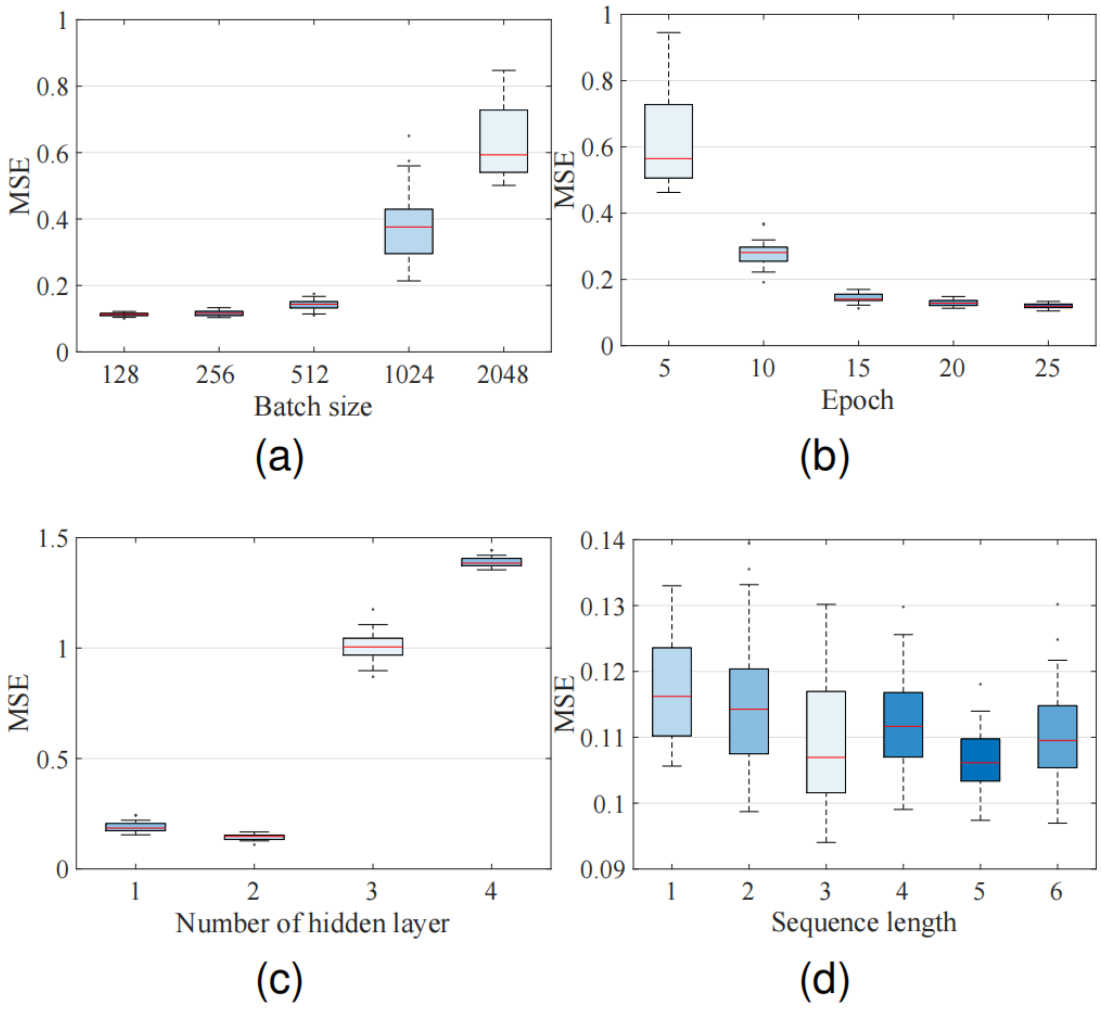


Fig. 3.23 Hyperparameter tuning results. (a) Batch sizes. (b) Epochs. (c) Number of hidden layers. (d) Sequence length.

Table 3.14 Bayesian BiLSTM model hyperparameters.

Hyperparameter	Value
Training sample size	5
Epoch	20
Batch size	256
Learning rate	0.01
Number of hidden layer	2
Loss function	MSE
Dropout probability	0.5
Optimizer	Adam

The Bayesian BiLSTM model performs 1,000 stochastic forward passes to estimate mean stress. Table 3.15 compares stress point estimates from the proposed model and state-of-the-art models, including ARIMA, NAR, 1-D CNN, DBN, GRU, LSTM, and BiLSTM. For a fair comparison, the Bayesian BiLSTM and BiLSTM share identical

architectures. Model performance is evaluated using MSE, mean absolute error (MAE), and R^2, calculated as follows.

$$MSE = \frac{\sum_{i=1}^{n} |\hat{s}(i) - s(i)|^2}{n} \tag{3.55}$$

$$MAE = \frac{\sum_{i=1}^{n} |\hat{s}(i) - s(i)|}{n} \tag{3.56}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n} (\hat{s}(i) - s(i))^2}{\sum_{i=1}^{n} (\bar{s} - s(i))^2} \tag{3.57}$$

where $s(i)$, $\hat{s}(i)$, $\bar{s}$, and n are true stress, estimated stress, mean stress, and n is the number of test samples.

Table 3.15 compares the MSE, MAE, and R^2 of the proposed Bayesian BiLSTM with other state-of-the-art models, showing it achieves the MSE (1.962), MAE (1.403), and R^2 (0.965). While the 1-D CNN achieves a slightly lower MSE, lower MAE, and higher R^2 than the Bayesian BiLSTM for stress estimation, the Bayesian BiLSTM is intentionally chosen due to its advantages in handling prognostic uncertainty. Unlike point estimates, probabilistic predictions offer richer insights, with wider prediction intervals indicating greater uncertainty and encouraging safer maintenance planning. Future work will explore Bayesian 1-D CNN, Bayesian GRU, and Bayesian LSTM models to further compare their performance with the Bayesian BiLSTM model.

Table 3.15 Comparisons of point estimates for stress estimation.

Model	MSE	MAE	R^2
ARIMA	2.152	1.422	0.950
NAR	2.157	1.421	0.949
1-D CNN	1.860	1.394	0.966
DBN	2.192	1.484	0.951
GRU	2.127	1.490	0.953
LSTM	2.139	1.418	0.951

BiLSTM	2.003	1.407	0.964
Bayesian BiLSTM	1.962	1.403	0.965

Fig. 3.24 shows the probabilistic stress predictions from the Bayesian BiLSTM model, with prediction intervals defined between the 0.025 and 0.975 quantiles. Fig. 3.24b depicts the PDF of predicted stress at the 4,500th minute, where the peak indicates the most likely stress value. The results confirm the model ability to provide accurate probabilistic stress predictions.

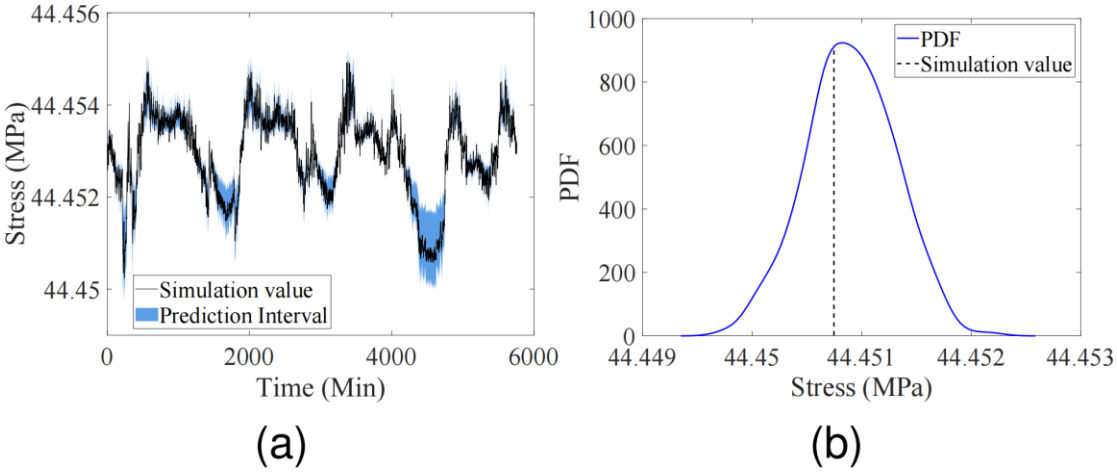


Fig. 3.24 Probabilistic stress prediction results with the Bayesian BiLSTM model. (a) Stress prediction of FGD fan shafts across 2.5% to 97.5% quantiles. (b) PDF of predicted stresses at the 4,500th minute.

The HCF life is estimated using the Rainflow counting algorithm and Basquin model based on stress predictions from the Bayesian BiLSTM. The Rainflow algorithm, implemented in MATLAB, extracts full and half cycles from the FEM-derived load history. The Basquin model, with Goodman correction, incorporates mean stress and stress range to account for mean load effects. Fig. 3.25a shows the stress range and mean stress distribution at the shaft critical location, with most rainflow counts concentrated in small alternating stresses. While low-to-medium damage cycles are more frequent, the highest stress cycles contribute most to damage. By integrating

probabilistic stress predictions, the Rainflow algorithm, the Basquin model, and Miner's rule, the Bayesian BiLSTM provides prognostics with uncertainty quantification. Fig. 3.25b shows the RUL distribution evolving over time, improving accuracy as more sensor data becomes available. Table 3.16 presents RUL interval estimates, with narrower 95% CIs indicating greater confidence. Both Fig. 3.25b and Table 3.16 show that RUL point estimates align closely with actual values. A dispersed RUL distribution reflects low confidence, highlighting the need for more informative training samples via active learning. Operators can use the RUL distribution to assess failure probability and enable dynamic PdM.

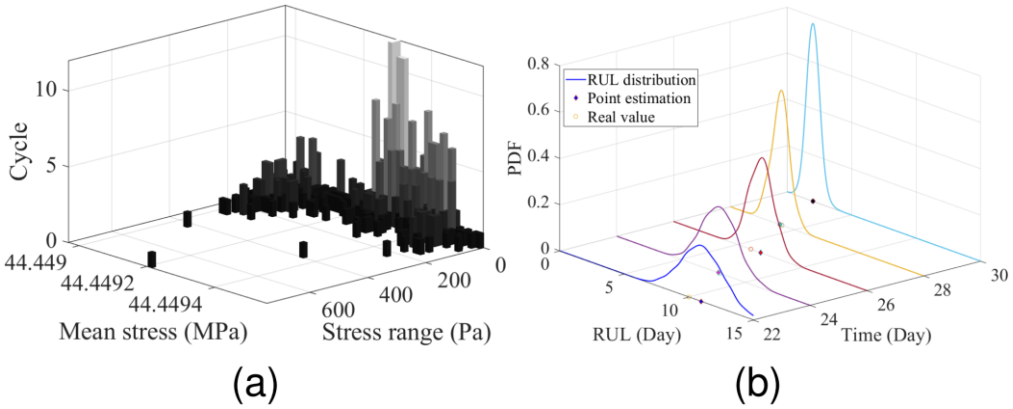


Fig. 3.25 Probabilistic HCF life analysis. (a) Stress range-mean stress cycle histogram for a one-day FGD fan shaft dataset. (b) Prognostics with uncertainty quantification for FGD fan shafts.

Table 3.16 Prognostics with uncertainty quantification for FGD fan shafts.

Time (Day)	Real value (Day)	Mean value (Day)	95% CI (Day)
22	10	11.1	[8, 14]
23	9	9.7	[7, 13]
24	8	8.0	[6, 11]
25	7	7.3	[5, 9]
26	6	6.8	[5, 8]
27	5	4.9	[4, 6]

3.5 Summary

This chapter develops a metamodel-based DT designed to probabilistically predict the life of assets lacking R2F data, with the goal of improving computational efficiency while maintaining accuracy in life reliability assessments. The proposed metamodel-based DT enables efficient, accurate, and robust reliability evaluations with strong nonlinear fitting capabilities. The development of the metamodel-based DT begins with modeling the FEM of the asset, followed by applying multi-physics fields to perform deterministic life predictions. Once the FEM simulations are completed, the metamodel is trained and validated using appropriate DoEs. The validated metamodel is then employed for life reliability assessments and sensitivity analyses of the asset. The metamodel-based DT provides a valuable tool for life reliability analysis, supporting warranty and maintenance decision-making. Although the developed DT-based life reliability prediction approach is specifically applied to assets such as turbine blisks and fan shafts, and the fatigue failure mechanism is investigated, the approach is versatile and can be extended to other types of assets and failure mechanisms. The effectiveness of this approach relies on understanding failure mechanisms and accurate PoF models. When these are unavailable, data-driven techniques can model degradation directly from operational data, or hybrid models combining partial physical insights with data-driven methods can approximate system behavior under uncertainty.

Chapter 4

Digital Twin-Based Warranty Design with Reliability-Constraint Imperfect Preventive Maintenance

For new products, insufficient reliability information often limits the ability to accurately predict warranty claims, posing a significant challenge for manufacturers. In addition to reliability modeling, effective warranty designs should incorporate operational reliability constraints and market scenarios to enhance customer satisfaction and increase net profits. To address these issues, this chapter proposes a digital twin (DT)-based warranty design method for new products with imperfect preventive maintenance (IPM). This method aims to maximize net profits while meeting reliability constraints in both monopolized and competitive markets. The proposed approach integrates finite element models (FEMs) and metamodels within the DT framework to enable efficient life reliability prediction. Design variables, including product price, base warranty (BW) duration, and preventive maintenance (PM) interval, are optimized using degradation data generated by DTs. Material properties and usage conditions are treated as random parameters. By incorporating IPM, the method ensures acceptable product reliability and helps decision-makers optimize non-renewable free minimal-repair warranties (NFMW) for monopolized and competitive markets. A case study on turbine blisks employing a knowledge-constraint neural network (KCNN)-based DT demonstrates the effectiveness of the proposed method, supported by

sensitivity analyses and policy comparisons to assess the impact of various factors on the optimization results.

4.1 Digital Twin-Based Warranty Design Framework

Manufacturers offering BWs should predict warranty claims using product reliability, $R_\tau(\tau)$, with minimal repair (MR) restoring defective items. An IPM strategy is required to maintain the specified reliability level, R^*. Warranty optimization is analyzed in two market conditions: (1) monopolized and (2) competitive markets. The manufacturer must address the following objectives:

(1) In a monopolized market, where the manufacturer has control over the product price, the goal is to determine the optimal product price p_m^*, the ideal BW duration w_m^*, and the most suitable PM interval δ_m^*.

(2) In a competitive market, where the product price is set by market forces, the objective is to identify the optimal BW duration w_c^* and the ideal PM interval δ_c^*.

This chapter seeks to answer the above questions by proposing a DT-based warranty prediction approach that incorporates five design variables. The objective is to maximize the manufacturer's discounted net profits while ensuring compliance with the operational reliability constraints R^* across both monopolized and competitive markets. Fig. 4.1 illustrates the framework of the DT-based warranty optimization, highlighting the connections between DTs, product reliability, IPM strategies, and warranty policies. The development of DTs includes the following key components: (i) uncertainties in product materials and loads, represented by $f_{\boldsymbol{x}}(\boldsymbol{x})$, where $\boldsymbol{x} \in [\boldsymbol{x}_l, \boldsymbol{x}_u]$,

(ii) the FEM-PoF model $g(\boldsymbol{x})$, and (iii) metamodels $\hat{g}(\boldsymbol{x})$. By leveraging the DT, product reliability $R_\tau(\tau)$ can be predicted with both accuracy and efficiency. Warranty costs are composed of three main elements: (i)MR costs C_{mr}, (ii) failure costs C_f, and (iii) PM costs C_{pm}. When a failure occurs, MR is performed, resulting in both the failure cost C_f and the MR cost C_{mr}. However, during each PM interval, only the PM cost C_{pm} is incurred. It is assumed that the manufacturer is responsible for covering the failure cost, MR cost, and PM cost. Additionally, the manufacturer must also account for the combined manufacturing and burn-in cost C_{mbi} associated with each product sold. The manufacturer's revenue function is driven by the sales volume, which depends on the product price p_m and the warranty period w over the course of the market lifecycle l_c.

Products are assumed to follow a monotonic deterioration process, such as fatigue. To address variations in material properties between units and differences in field usage conditions, represented by $f_x(x)$, DTs are utilized for degradation modeling and reliability prediction. This approach allows for accurate warranty cost estimation without relying on field failure data. After undergoing a burn-in period τ_{bi}, the product is introduced to the market. The burn-in test serves to identify and eliminate low-quality products by subjecting them to field-like operating conditions. As a result, the product's age at the time of sale begins at a value greater than zero. Within the framework, the focus is placed on performing MR for each failure occurrence, with the repair time considered insignificant in comparison to the duration of the warranty period w. MR restores a failed unit to its operational state as it was just before the failure, without affecting the failure rate $\lambda_\tau(\tau)$. During the warranty period, manufacturers also carry

out a periodic IPM program, which relies on reliability as a conditional indicator. The IPM model employs an age reduction approach, utilizing an age reduction factor ϑ. The

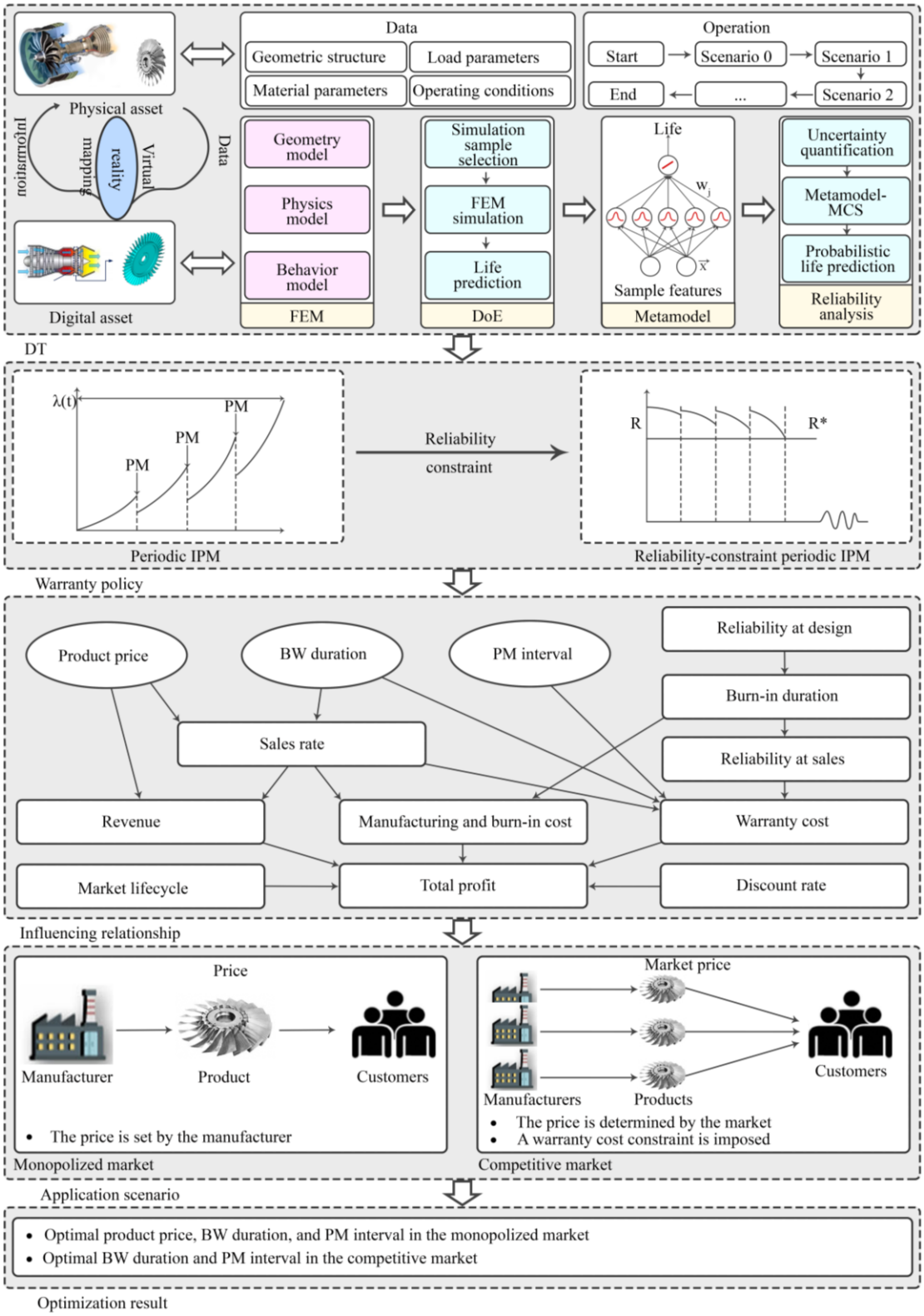


Fig. 4.1 The framework of the DT-based warranty prediction.

product lifecycle is illustrated in Fig. 4.2. A manufacturer offers new products accompanied by BWs, with a market lifecycle of l_{lc}. After this period, the product is withdrawn from the market. The product lifecycle is generally limited, as advancements in technology may render it obsolete or lead to the release of a next-generation version. The NFMW is one of the most widely used warranty policies for products, providing coverage for each unit sold over a defined warranty period w. If a product fails during the warranty period w, the manufacturer bears the cost of repairing the defective item. For instance, if a customer purchases the product at $t = 0$ and it fails at a time $t \in [0, w]$, the manufacturer provides a free repair, with the remaining warranty period reduced to $w - t$. Non-renewable warranties are typically used for high-cost products [172]. The time value of money reflects the concept of discounting future costs, mathematically expressed through the continuous time discounting function $e^{-\eta_D t}$ [87].

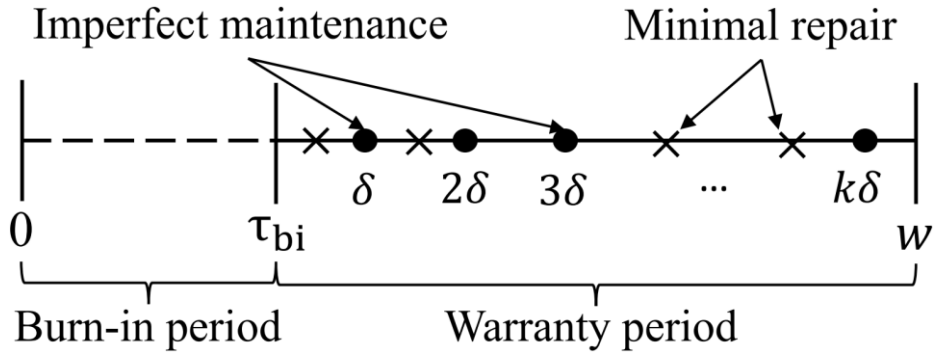


Fig. 4.2 The product lifecycle involving MR at failure and planned IPM.

The warranty prediction framework based on DTs is built on a series of assumptions centered around product reliability and IPM, outlined as follows:

1. The product undergoes deterioration driven by mission cycles and fails once the accumulated damage surpasses a specified threshold.

2. Parameters such as product material properties, load conditions, key PoF mechanisms, costs, and demands can be reasonably estimated by reliability engineers and market retailers.
3. The product is repairable.
4. Degradation failures are identified instantly, and both repair time after failure and preventive maintenance are insignificant when compared to the product's overall lifespan.
5. All warranty claims during the coverage period are assumed to be genuine, with no intentional fabrication of failures by customers to secure repairs.

4.2 Warranty Model Formulation and Optimization

4.2.1 Warranty Cost Modeling with Reliability-Constraint Imperfect Preventive Maintenance

The burn-in period, which addresses production defects and reduces infant mortality prior to product release, is represented by the product age in the customer possession, denoted as τ_{bi}. Consequently, the product reliability at the time of sale, $\tilde{R}_{\tau}(\tau)$, and the failure rate at the time of sale, $\tilde{\lambda}_{\tau}(\tau)$, are adjusted as outlined in Eqs. (4.1) and (4.2).

$$\tilde{R}_{\tau}(\tau) = 1 - \frac{\int_0^{\tau+\tau_{bi}} f_{\tau_f}(u)du - \int_0^{\tau_{bi}} f_{\tau_f}(u)du}{R_{\tau}(\tau_{bi})} = \frac{R_{\tau}(\tau+\tau_{bi})}{R_{\tau}(\tau_{bi})} \tag{4.1}$$

$$\tilde{\lambda}_{\tau}(\tau) = \frac{\tilde{f}_{\tau_f}(\tau)}{\tilde{R}_{\tau}(\tau)} = \frac{f_{\tau_f}(\tau+\tau_{bi})}{[R_{\tau}(\tau_{bi})]^2} \tag{4.2}$$

Warranty policies, such as renewable, non-renewable, pro-rata, and free warranties, directly impact the number of failures and associated warranty costs. Among these, non-renewable free warranties are widely adopted for products, allowing customers to

make claims at no expense, which helps improve customer satisfaction. Suppose in the non-renewable warranty with a duration of w, a product fails at time $\tau \in [0, w]$, where $w \in \{w_m, w_c\}$ denotes the warranty period set. In this scenario, the manufacturer provides a free repair or replacement, including a remaining warranty period of $w - \tau$. To begin, we examine the costs related to two types of non-renewable free warranties: Non-renewable free replacement warranty (NFRW) and NFMW. Assuming the product is both identical and repairable, the time required for repair or replacement is considered negligible compared to the warranty period. Let $\bar{\Lambda}(w)$ represent the expected number of failures per product sold throughout a warranty period of w.

In the case of an FRW, failures over time are modeled as a renewal process defined by $\tilde{R}_\tau(\tau)$ and $f_{\tau_f}(\tau)$. As a result, the expected number of failures within the interval $[0, w]$ can be determined using the following renewal function.

$$\bar{\Lambda}_{NFRW}(w) = \left(1 - \tilde{R}_\tau(w)\right) + \int_0^w \bar{\Lambda}(w-u)\tilde{f}_{\tau_f}(u)du \qquad (4.3)$$

Unlike the replacement policy, MR restores a failed unit to working condition without altering its failure rate. Assuming the repair time is negligible in comparison to the warranty period, the number of failures occurring within the interval $[0, w]$ is determined by the rate of occurrence of failures (ROCOF), represented as $\tilde{\lambda}_\tau(\tau)$. Accordingly, the expected number of failures is expressed as $\bar{\Lambda}_{\mathrm{NFMW}}(w) = \int_0^w \tilde{\lambda}_\tau(u)du$.

Let C_r represent the cost of replacement, C_{mr} denote the cost of MR, and C_f signify the penalty cost associated with failure, where typically $C_{\mathrm{mr}} < C_r$. Let $\bar{C}_w(w)$ symbolize the manufacturer's expected warranty cost per item sold over the warranty

period w. The expression for $\bar{C}_w(w)$ for both NFRW and NFMW is provided in Eq. (4.4) [175].

$$\bar{C}_w(w) = \begin{cases} \left(C_r + C_f\right)\left[\left(1 - \tilde{R}_\tau(w)\right) + \int_0^w \bar{\Lambda}(w-u)\, \tilde{f}_{\tau_f}(u)du\right], NFRW \\ \left(C_{mr} + C_f\right)\int_0^w \tilde{\lambda}_\tau\,(u)du, NFMW \end{cases} \quad (4.4)$$

Product failures negatively affect customer satisfaction, ultimately impacting the manufacturer's reputation. To address this, a PM policy is introduced to reduce the probability of product failures. This PM service provides advantages for both customers and manufacturers: (i) For customers, PM services can be costly, so having them covered under a warranty allows customers to save money; (ii) For manufacturers, implementing PM ensures the product reliability throughout the contract period. An effective PM strategy is essential for manufacturers of repairable products. Consider a product with a warranty period of w, during which periodic PM is performed at time intervals $\tau_k = k\delta$, where $k = 1,2, \ldots, \lfloor w/\delta \rfloor$, and $\lfloor w/\delta \rfloor$ represents the greatest integer less than or equal to w/δ. Here, δ is defined as $\{\delta_m, \delta_c\}$. For repairable products, PM is often imperfect, and its effects can be categorized based on the failure rate into three types: (i) reduction of the failure rate, (ii) reduction of the product's effective age, and (iii) slowing down the deterioration process. In this study, we propose an IPM model by reducing the product's effective age using the GRP I model introduced by Kijima [176]. The effect of each IPM action is characterized by the age reduction factor $\vartheta \in [0,1)$, where $\vartheta = 0$ represents MR and $\vartheta = 1$ corresponds to preventive replacement. Following the framework proposed by Martorell et al., the virtual age of the product immediately before and after the kth IPM is denoted as $\tau_-^{(k)}$ and $\tau_+^{(k)}$, respectively [177]. Hence, the virtual age of the product before the kth IPM is given by $\tau_-^{(k)} = \tau_+^{(k-1)} +$

$\delta = k\delta - (k-1)\vartheta\delta = [k-(k-1)\vartheta]\delta$. Similarly, the virtual age after the kth IPM is expressed as $\tau_+^{(k)} = \tau_-^{(k)} - \vartheta\delta = [k-(k-1)\vartheta]\delta - \vartheta\delta = (1-\vartheta)k\delta$, where $\delta = \{\delta_m, \delta_c\}$ represents the set of PM intervals. The expected number of failures, considering periodic IPM combined with MR, is formulated in Eq. (4.5).

$$\begin{aligned}\bar{\Lambda}(w,\delta,\vartheta) &= \int_0^{\tau_-^{(1)}} \tilde{\lambda}_\tau(u)du + \int_{\tau_+^{(1)}}^{\tau_-^{(2)}} \tilde{\lambda}_\tau(u)du \\ &+ \cdots + \int_{\tau_+^{(\lfloor w/\delta \rfloor - 1)}}^{\tau_-^{(\lfloor w/\delta \rfloor)}} \tilde{\lambda}_\tau(u)du + \int_{\tau_+^{(\lfloor w/\delta \rfloor)}}^{\tau_+^{(\lfloor w/\delta \rfloor)+(w-\lfloor w/\delta \rfloor\delta)}} \tilde{\lambda}_\tau(u)du \\ &= \sum_{k=0}^{\lfloor w/\delta \rfloor - 1} \int_{k\delta}^{(k+1)\delta} \tilde{\lambda}_\tau(u - k\vartheta\delta)du + \int_{\lfloor w/\delta \rfloor\delta}^{w} \tilde{\lambda}_\tau(u - \vartheta\lfloor w/\delta \rfloor\delta)du\end{aligned} \tag{4.5}$$

where $\sum_{k=0}^{\lfloor w/\delta \rfloor - 1} \int_{k\delta}^{(k+1)\delta} \tilde{\lambda}_\tau(u - k\vartheta\delta)du$ is the expected number of failures within $[0, \lfloor w/\delta \rfloor]$, and $\int_{\lfloor w/\delta \rfloor\delta}^{w} \tilde{\lambda}_\tau(u - \vartheta\lfloor w/\delta \rfloor\delta)du$ is the surplus of the expected number of failures after the $\lfloor w/\delta \rfloor$'s PM [87].

Proposition 4.1. *Given a known δ, the expected number of failures $\bar{\Lambda}(w|\delta)$ increases with the warranty period w.*

Proposition 4.1, derived from Eq. (4.5), is straightforward, and its proof is provided in Appendix A.4.1. Fig. 4.3 illustrates the IPM scheme in relation to the failure rate. Due to the complexity of the failure time distribution $f_{\tau_f}(\cdot)$, a closed form solution for $\bar{\Lambda}(\cdot)$ does not exist for NFRW, NFMW, and IPM-NFMW. Consequently, Monte Carlo simulation (MCS) may be necessary to compute $\bar{\Lambda}(\cdot)$. In the simulation, parameters such as $[\boldsymbol{x}_l, \boldsymbol{x}_u]$, $f_x(\boldsymbol{x})$, and R^* can be configured flexibly.

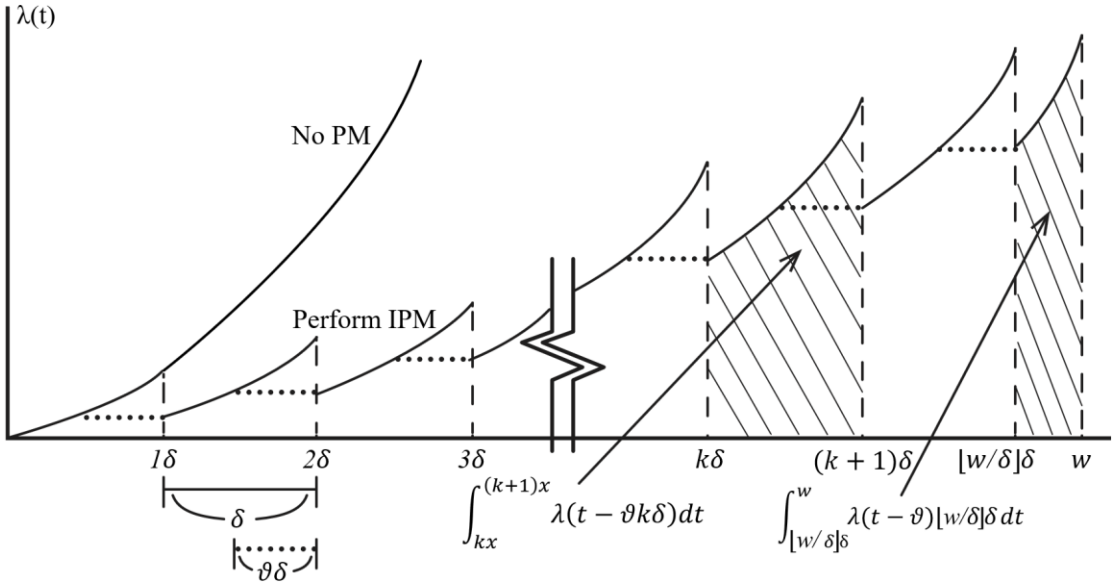


Fig. 4.3 The imperfect preventive maintenance scheme.

Using the cumulative failure rate $\int_{k\delta}^{(k+1)\delta} \tilde{\lambda}_\tau(u-k\vartheta\boldsymbol{\delta})du\,, k=0,1,\ldots,\frac{\boldsymbol{w}}{\boldsymbol{\delta}}-1$ (or $\int_{[\frac{w}{\delta}]\delta}^{w} \tilde{\lambda}_\tau(u-\vartheta\lfloor w/\delta\rfloor\delta)du$), the reliability of the product following the kth IPM is expressed as follows.

$$\tilde{R}_\tau^{(k)} = \begin{cases} e^{-\int_{k\delta}^{(k+1)\delta} \tilde{\lambda}_\tau(u-k\vartheta\delta)du}, k=0,1,\ldots,\lfloor w/\delta\rfloor-1 \\ e^{-\int_{\lfloor w/\delta\rfloor\delta}^{w} \tilde{\lambda}_\tau(u-\vartheta\lfloor w/\delta\rfloor\delta)du}, k=\lfloor w/\delta\rfloor \end{cases} \tag{4.6}$$

Note that for $k=0,1,\ldots,\lfloor w/\delta\rfloor-1$, the reliability under each interval will decrease as the PM interval remains constant, i.e., $\tilde{R}_\tau^{(0)} > \tilde{R}_\tau^{(1)} > \cdots > \tilde{R}_\tau^{(\lfloor w/\delta\rfloor-1)}$ [82][178]. Thus, the reliability level during the warranty duration w is $\tilde{R}_\tau(w,\delta) = \min\{e^{-\int_{(\lfloor w/\delta\rfloor-1)\delta}^{\lfloor w/\delta\rfloor\delta} \tilde{\lambda}_\tau(u-(\lfloor w/\delta\rfloor-1)\vartheta\delta)du}, e^{-\int_{\lfloor w/\delta\rfloor\delta}^{w} \tilde{\lambda}_\tau(u-\lfloor w/\delta\rfloor\vartheta\delta)du}\}$. Given that the required reliability level of the product is R^*, the condition for the existence of w^* and p^* is $\tilde{R}_\tau(w,\delta) \geq R^*$.

Remark: *IPM may introduce new failure mechanisms due to unintended side effects, such as improper handling, inadequate component replacement, or procedural errors. To address this, an adjustment factor can be incorporated into the failure rate to*

account for the likelihood of additional failures caused by maintenance actions, which will be explored in future research.

In the context of the periodic IPM model combined with an MR strategy, a warranty cost model is developed to evaluate the expected warranty cost throughout the NFRW period. As outlined earlier, a periodic PM program is implemented with a fixed PM interval δ, resulting in the number of PM services being represented as $\lfloor w/\delta \rfloor$. The warranty cost includes three components: MR costs (C_{mr}), failure costs (C_f), and PM costs (C_{pm}). It is assumed that the PM cost (C_{pm}) is less than the combined cost of MR and failure ($C_{\mathrm{mr}} + C_f$), i.e., $C_{\mathrm{pm}} < C_{\mathrm{mr}} + C_f$. Consequently, the expected warranty cost over the warranty period $(0, w]$ is expressed in Eq. (4.7).

$$\begin{aligned} \bar{C}_w(w,\delta) &= \left(C_{mr} + C_f\right)\bar{\Lambda}(w,\delta) + C_{pm}\lfloor w/\delta \rfloor \\ &= \left(C_{mr} + C_f\right)\begin{bmatrix} \sum_{k=0}^{\lfloor w/\delta \rfloor - 1} \int_{k\delta}^{(k+1)\delta} \tilde{\lambda}_\tau(u - k\vartheta\delta)du \\ + \int_{\lfloor w/\delta \rfloor \delta}^{w} \tilde{\lambda}_\tau(u - \vartheta\lfloor w/\delta \rfloor \delta)du \end{bmatrix} \\ &\quad + C_{pm}\lfloor w/\delta \rfloor \end{aligned} \tag{4.7}$$

According to Proposition 4.1, for a given δ, as w increases, $\overline{\Lambda}(w|\delta)$ grows, and $\lfloor w/\delta \rfloor$ either increases or remains unchanged. Consequently, $\bar{C}_w(w|\delta)$ also increases with w. However, as w grows, product sales will increase as well. This implies that an optimal value of w^* exists, balancing warranty costs and sales volume to maximize profit. The analysis of the existence and uniqueness of w is provided in the next section. In contrast to w, one could also identify the optimal PM interval δ^* that minimizes $\bar{C}_w(\cdot)$ to achieve profit maximization. To guarantee the existence of the optimal δ^* that minimizes $\bar{C}_w(\cdot)$, the condition $\bar{C}_w(\delta_1) > \bar{C}_w(\delta_2)$ and $\bar{C}_w(\delta_2) \leq \bar{C}_w(\delta_3)$, where $\delta_1 >$

$\delta_2 > \delta_3$. Additionally, the operational reliability constraint must be taken into account. Based on this, Theorem 4.1 is presented below.

Theorem 4.1. *Given* w, $\exists \delta_1, \delta_2, \delta_3$ *satisfy the following conditions,*

$$\begin{cases} \bar{\Lambda}(\delta_2) - \bar{\Lambda}(\delta_3) \leq \dfrac{C_{pm}}{C_{mr} + C_f} \\ \bar{\Lambda}(\delta_1) - \bar{\Lambda}(\delta_2) > \dfrac{C_{pm}}{C_{mr} + C_f} \\ \tilde{R}_\tau(\delta_1) \geq R^* \\ \delta_1 > \delta_2 > \delta_3 \end{cases}$$

Then there exists a finite $\delta^* = \delta_2$.

It is important to note that estimating δ^* is challenging. Therefore, search methods are utilized to determine its value. The detailed solution algorithm is provided in Subsection 4.3.

4.2.2 Warranty Profit Modeling in Monopolized and Competitive Markets

Markets serve as a critical platform for facilitating trading interactions between manufacturers and customers. Therefore, it is essential to consider warranty optimization within the context of the current market environment, with a particular focus on differentiating between monopolized and competitive market conditions.

4.2.2.1 Monopolized Markets

In a monopolized market with a single manufacturer, the product is non-substitutable. Given the manufacturer's monopolistic position, it has the ability to set the product price, p_m, independently. Consider a scenario where a manufacturer offers a new product bundled with an IPM-NFMW in such a monopolized market. The market lifecycle of the product, denoted by l_c, represents the period after which the product

will no longer be available for sale. This lifecycle is typically finite, as the technology used in the product may become outdated or the product may be replaced by a newer generation in the future [174]. Let $Q(t_1, t_2) = \int_{t_1}^{t_2} q(u)du$ represent the cumulative number of sales over a specific time interval, $[t_1, t_2]$. Since the product will no longer be sold after the end of its market lifecycle, sales will be truncated at l_c, as illustrated in Eq. (4.8).

$$Q(t) = \begin{cases} \int_0^t q(u)du, t \leq l_c \\ \int_0^{l_c} q(u)du, t > l_c \end{cases} \tag{4.8}$$

The potential market, Q_m, is anticipated to grow monotonically as the product price decreases or the warranty period increases. Based on the work of Glickman and Berger, Q_m represented as a log-linear function of both the product price (p_m) and the warranty period (w_m), as demonstrated below [68].

$$Q_m = \kappa_1 (p_m)^{-\alpha_1} (w_m + \kappa_2)^{\alpha_2} \tag{4.9}$$

where κ_1 is the amplitude factor, κ_2 is the constant of time displacement which allows for the possibility of non-zero demand when w_m is zero, α_1 is the price elasticity, and α_2 is the warranty duration elasticity ($\kappa_1 > 0$, $\kappa_2 > 0$, $\alpha_1 > 0$, and $0 < \alpha_2 < 1$).

The demand function presented in Eq. (4.9) is a static model that overlooks the time value of money. Given that sales volume fluctuates over different periods of the marketing timeline, a depreciation effect arises. The Bass model, a widely acknowledged first-purchase diffusion model, is commonly used to describe the diffusion of new products in the market [179]. According to the Bass model, the market consists of two types of purchasers: innovators and imitators [65][76][175].

Let $q(t)$ denote the product's sales rate at time t, where $t > 0$. Accordingly, the sales rate is defined in Eq. (4.10).

$$\begin{aligned} q(t) &= Q_m \left[1 - \frac{Q(t)}{Q_m}\right]\left[\psi + \frac{Q(t)}{Q_m}\right] \\ &= \kappa_1 (p_m)^{-\alpha_1} (w_m + \kappa_2)^{\alpha_2} \left[1 - \frac{Q(t)}{Q_m}\right]\left[\psi + \frac{Q(t)}{Q_m}\right] \end{aligned} \tag{4.10}$$

where $Q(t)$ is the cumulative sales volume in $[0, t]$, $\psi > 0$ reflects the relative influence of innovators to imitators, and $q(t) = \frac{dQ(t)}{dt}$. As in Eq. (4.10), $q(t)$ can be viewed as a form of separable demand functions $q(t) = Q_m(p_m, w_m)G(Q(t))$ [180].

Assume that the product price (p_m), warranty period (w_m), and PM interval (δ_m) remain unchanged over the entire lifecycle (l_c). The objective function is designed to maximize the manufacturer's net profit from product sales and is expressed in Eq. (4.11).

$$\begin{aligned} J_m &= \kappa_1 (p_m)^{-\alpha_1} (w_m + \kappa_2)^{\alpha_2} [p_m - \bar{C}_{mbi} - \bar{C}_w(w_m, \delta_m)] \\ &\int_0^{l_c} G\big(Q(t)\big) e^{-\eta_D t} dt \end{aligned} \tag{4.11}$$

Here, $e^{-\eta_D t}$ represents the continuous time discounting factor, where η_D denotes the discount rate. Eq. (4.12) calculates the expected combined cost of manufacturing and burn-in per product, denoted as $\bar{C}_{\text{mbi}}$.

$$\bar{C}_{mbi} = \frac{C_m + \bar{C}_{bi}}{R_\tau(\tau_{bi})} = \frac{C_m + C_{bif} + C_{bit} \int_0^{\tau_{bi}} R_\tau(u) du}{R_\tau(\tau_{bi})} \tag{4.12}$$

where C_m is the manufacturing cost, $\bar{C}_{\text{bi}}$ is the expected burn-in cost, C_{bif} is the fixed set-up cost of the burn-in, C_{bit} is the time-dependent cost of the burn-in, and τ_{bi} is the burn-in duration [181].

With the various components of the manufacturer's net profit function defined across different time intervals, the optimization model is structured as follows.

$$\begin{aligned} \max_{p_m, w_m, \delta_m} J_m &= \kappa_1 (p_m)^{-\alpha_1} (w_m + \kappa_2)^{\alpha_2} [p_m - \bar{C}_{mbi} - \bar{C}_w(w_m, \delta_m)] \\ &\int_0^{l_c} G\big(Q(t)\big) e^{-\eta_D t} dt \end{aligned} \tag{4.13}$$

$$\text{s.t.}\begin{cases}\tilde{R}_\tau(w_m,\delta_m)\geq R^*\\ \boldsymbol{x}_l\leq\boldsymbol{x}\leq\boldsymbol{x}_u\\ 0\leq\vartheta\leq 1\\ p_m\geq 0\\ w_m>\delta_m\\ \delta_m\geq 0\\ Q(0)=Q_0\end{cases}$$

The optimization model is designed to maximize the manufacturer's net profit generated from selling products over l_c time intervals. The first constraint ensures that the product's reliability under IPM measures exceeds R^*. The second constraint specifies that the product's reliability is assessed within the material and load uncertainty range $[x_l, x_u]$. Constraints three through six set the value limits for the imperfect repair level, product price, warranty period, and PM interval, respectively. Lastly, the seventh constraint establishes the initial sales volume condition.

The manufacturer's task is to identify the optimal product price (p_m^*), warranty duration (w_m^*), and PM interval (δ_m^*) that maximize the expected discounted profit (J_m^*). However, due to the complexity of joint optimization, we establish the existence and uniqueness of these optimal values. From Eq. (4.13), it is evident that, for given values of p_m and w_m, J_m^* and δ_m^* can be determined by minimizing $\bar{C}_w(\delta_m)$, such that $\delta_m^* = \text{argmin}_{\delta_m}\bar{C}_w(\delta_m)$. The existence condition for δ_m is outlined in Theorem 4.1. As for p_m, it can be readily shown that when w_m and δ_m are fixed, an optimal p_m^* exists to maximize J_m, even if the specific form of the failure rate $\tilde{\lambda}_\tau(\cdot)$ is unknown [182]. Using Eq. (4.13), we propose Proposition 4.2, with its proof provided in Appendix A.4.2.

Proposition 4.2. *If p_m is the decision variable and both w_m and δ_m are known, an optimal price exists that maximizes the expected profit in a monopolized market.*

$$p_m^* = \frac{\alpha_1\left(\bar{C}_{mbi} + \bar{C}_w(w_m, \delta_m)\right)}{\alpha_1 - 1}$$

Eq. (4.13) is clearly continuously differentiable with respect to w_m over the interval $w_m \in [0, \infty)$. To establish the existence and uniqueness of w_m^*, it is necessary to verify both the first-order conditions ($\lim_{w_m \to 0} \frac{\partial J_m}{\partial w_m} > 0$ and $\lim_{w_m \to \infty} \frac{\partial J_m}{\partial w_m} < 0$) and second-order conditions ($\frac{\partial^2 J_m}{\partial (w_m)^2} < 0$). If these conditions are satisfied, it confirms that J_m is a concave function with respect to w_m, derived from setting $\frac{\partial J_m}{\partial w_m} = 0$. The profit model incorporates the constraint $\tilde{R}_\tau(w_m) \geq R^*$, and the optimization problem under this constraint has been analyzed in [30].

By differentiating Eq. (4.13) with respect to w_m, while keeping p_m and δ_m constant, the first-order derivative is obtained as follows.

$$\begin{aligned} H(w_m) \quad &= \frac{\partial J_m}{\partial w_m} = A_{mw}\Big[\alpha_2 (w_m + \kappa_2)^{\alpha_2 - 1}\left(p_m - \bar{C}_{mbi} - \bar{C}_w(w_m, \delta_m)\right) \\ &\quad -(w_m + \kappa_2)^{\alpha_2} \frac{\partial \bar{C}_w(w_m, \delta_m)}{\partial w_m}\Big] \end{aligned} \tag{4.14}$$

where $A_{\mathrm{mw}} = \kappa_1 (p_m)^{-\alpha_1} \int_0^{l_c} G(Q(t)) e^{-\eta_D t} dt$.

Let $S_{m,1} = \{w_m \geq 0 | \tilde{R}_\tau(w_m) = R^*\}$ and $S_{m,2} = \{w_m \geq 0 | \tilde{R}_\tau(w_m) > R^*\} \cap \{w_m \geq 0 | H(w_m)\}$, then we present the following Theorem 4.2.

Theorem 4.2. *Given p_m and δ_m, the following holds:*

1. *When $\frac{\partial H(w_m)}{\partial w_m} < 0$ for $w_m \geq 0$, the following holds:*

(a) *When $H(0) < 0$ or $H(\infty) > 0$, there exists a unique finite $w_m^* \in S_{m,1}$.*

(b) *When $H(0) > 0$ and $H(\infty) < 0$, there exists a unique finite $w_m^* \in S_{m,2}$.*

2. *When $\frac{\partial H(w_m)}{\partial w_m} > 0$ for $w_m \geq 0$, the following holds:*

(a) *When $H(0) > 0$ or $H(\infty) < 0$, there exists a unique finite $w_m^* \in S_{m,1}$.*

(b) *When $H(0) < 0$ and $H(\infty) > 0$, there exists a finite $w_m^* \in S_{m,2}$.*

3. *When* $H(w_m)$ *is non-decreasing (non-increasing) w.r.t.* w_m *and* $H(0) > 0$ *(*$H(0) < 0$*) or* $H(\infty) < 0$ *(*$H(\infty) > 0$*), the following holds:*

(a) *When* $S_{m,2} \neq \emptyset$*, there exists a finite* $w_m^* \in S_{m,2}$.

(b) *When* $S_{m,2} = \emptyset$*,* w_m^* *does not exist.*

We can make the following remark based on Theorem 4.2.

Remark 4.1. *In the monopolized market, we have*

1. *An optimal warranty duration may not exist if* $w_m^* = 0$ *or* $\tilde{R}_\tau(w_m|H(w_m) = 0) > R^*$.
2. *If* w_m^* *exists, it can be obtained either under* $H(w_m) = 0$ *or* $\tilde{R}_\tau(w_m) = R^*$.

4.2.2.2 Competitive Markets

In a competitive market where multiple manufacturers offer similar products, each must account for the competition posed by others. From the customers' perspective, the sales price plays a key role in their purchasing decisions. To remain competitive, the manufacturer chooses not to raise the sales price but instead focuses on providing enhanced services as a competitive strategy. The competitive market can be viewed as a specific case of a monopolized market, particularly when $\alpha_1 = 0$, as the product price becomes less significant to customers in this scenario. However, adopting these service-oriented strategies increases the manufacturer's total costs, leading to reduced profit margins at the current sales price. To resolve this conflict, the manufacturer's total cost is expressed as a fraction of the sales price [30].

It is assumed that the product price, p_c, is a predetermined constant, and the BW cost is represented as a fraction of $p_c - \bar{C}_{\text{mbi}}$, such that $\bar{C}_w(w_c) \leq \rho(p_c - \bar{C}_{\text{mbi}})$, where

$0 < \rho < 1$. The objective function in the competitive market is comparable to that of Eq. (4.13), as outlined below.

$$\max_{w_c,\delta_c} J_c = \kappa_1(w_c + \kappa_2)^{\alpha_2}[p_c - \bar{C}_{mbi} - \bar{C}_w(w_c,\delta_c)] \int_0^{l_c} G\big(Q(t)\big)e^{-\eta_D t}dt$$
$$s.t. \begin{cases} \bar{C}_w(w_c) \le \rho(p_c - \bar{C}_{mbi}) \\ 0 < \rho < 1 \\ \tilde{R}_\tau(w_c,\delta_c) \ge R^* \\ x_l \le x \le x_u \\ 0 \le \vartheta \le 1 \\ w_c > \delta_c \\ \delta_c \ge 0 \\ Q(0) = Q_0 \end{cases} \tag{4.15}$$

Unlike Eq. (4.13), in Eq. (4.15), p_c is treated as a parameter rather than a decision variable, and $Q_c = \kappa_1(w_c + \kappa_2)^{\alpha_2}$ is independent of p_c. Moreover, Eq. (4.15) introduces two additional constraints (the first and second constraints) that define the relationship between the product price and warranty service costs. The existence condition for δ_c^* is provided in Theorem 4.1, while the necessary condition for w_c^* follows the same structure as in Theorem 4.2, with the exception that additional constraints must be satisfied.

By differentiating Eq. (4.15) w.r.t. w_c, the first-order derivative is obtained as follows.

$$\begin{aligned} K(w_c) \quad &= A_{cw}[\alpha_2(w_c + \kappa_2)^{\alpha_2 - 1}\big(p_c - \bar{C}_{mbi} - \bar{C}_w(w_c,\delta_c)\big) \\ &\quad -(w_c + \kappa_2)^{\alpha_2}\frac{\partial \bar{C}_w(w_c)}{\partial w_c} \end{aligned} \tag{4.16}$$

where $A_{\mathrm{cw}} = \kappa_1 \int_0^{l_c} G(Q(t))e^{-\eta_D t}dt$.

Let $S_{c,1} = \{w_c \geq 0 | \bar{C}_w(w_c) = \rho(p_c - \bar{C}_{\text{mbi}})\} \cap \{w_c \geq 0 | \tilde{R}_\tau(w_c) = R^*\}$ and $S_{c,2} = \{w_c \geq 0 | \bar{C}_w(w_c) < \rho(p_c - \bar{C}_{\text{mbi}})\} \cap \{w_c \geq 0 | \tilde{R}_\tau(w_c) > R^*\} \cap \{w_c \geq 0 | K(w_c) = 0\}$, then we have the following Theorem 4.3.

Theorem 4.3. *Given δ_c, the following results hold.*

1. *When $\frac{\partial K(w_c)}{\partial w_c} < 0$ for $w_c \geq 0$, the following holds:*

(a) When $K(0) < 0$ or $K(\infty) > 0$, there exists a unique finite $w_c^* \in S_{c,1}$.

(b) When $K(0) > 0$ and $K(\infty) < 0$, there exists a unique finite $w_c^* \in S_{c,2}$.

2. *When $\frac{\partial K(w_c)}{\partial w_c} > 0$ for $w_c \geq 0$, the following holds:*

(a) When $K(0) > 0$ or $K(\infty) < 0$, there exists a unique finite $w_c^* \in S_{c,1}$.

(b) When $K(0) < 0$ and $K(\infty) > 0$, there exists a finite $w_c^* \in S_{c,2}$.

3. *When $K(w_c)$ is non-decreasing (non-increasing) w.r.t. w_c and $K(0) > 0$ ($K(0) < 0$) or $K(\infty) < 0$ ($K(\infty) > 0$), the following holds:*

(a) When $S_{c,2} \neq \emptyset$, there exists a finite $w_c^* \in S_{c,2}$.

(b) When $S_{c,2} = \emptyset$, w_c^* does not exist.

Similar to Remark 4.1, the following remark can also be concluded.

Remark 4.2. *In the competitive market, we have*

1. *An optimal warranty duration may not exist if $w_c^* = 0$ or $\tilde{R}_\tau(w_c | K(w_c) = 0) > R^*$ or $\bar{C}_w(w_c | K(w_c) = 0) > \rho(p_c - \bar{C}_{mbi})$.*
2. *If w_c^* exists, it can be obtained either under $K(w_c) = 0$ or $\{\tilde{R}_\tau(w_c) = R^*\} \cap \{\bar{C}_w(w_c) > \rho(p_c - \bar{C}_{mbi})\}$.*

4.2.3 Optimization Solutions

Building on the mathematical analysis, the deduction process serves as a basis for formulating the optimization solution for the warranty model [182]. Here, p_m is treated as a temporary variable. For a given warranty period w_m and PM interval δ_m, the optimal price p_m^* can be determined using Proposition 4.2. Subsequently, an enumeration algorithm is applied to compute w_m^* and δ_m^*. The optimization solution is illustrated in Fig. 4.4.

As illustrated in Fig. 4.4, the process begins by assigning initial values: the candidate number for the PM interval is set to $i = 1$, the warranty duration is initialized to $j = 1$. Next, the expected number of failures $\overline{\Lambda}(w_m^{(j)}, \delta_m^{(i,j)})$, the operational minimum reliability, $\tilde{R}_\tau(w_m^{(j)}, \delta_m^{(i,j)})$, and expected warranty costs $\bar{C}_w(w_m^{(j)}, \delta_m^{(i,j)})$, are computed using Eqs. (4.5), (4.6), and (4.7), respectively, for all $i = 1,2, \dots, n$ and $j = 1,2, \dots, m$. This step identifies the optimal PM interval i^* that satisfies the reliability constraint $\tilde{R}_\tau(w_m^{(j)}, \delta_m^{(i^*,j)}) \geq R^*$. Third, calculate $p_m^{(i^*,j)}$ and substitute $p_m^{(i^*,j)}$, w_m^j, and $\delta_m^{(i^*,j)}$ are substituted into the objective function $J_m(\cdot)$ to determine the optimal warranty duration j^*. In the competitive market, where p_c is predetermined, the additional constraint $\bar{C}_w(w_c^{(i^*,j)}) \leq \rho(p_c - \bar{C}_{\text{mbi}})$ must be verified to finalize the selection of j^*.

In particular, the following algorithm is developed to numerically identify the optimal PM interval $\delta_m^{(i^*,j)}$:

1. Initialize $i = 1$ and $j = 1$.

2. Compute $\tilde{R}_\tau(w_m^{(j)}, \delta_m^{(i,j)})$ and compare it with R^*. If $\tilde{R}_\tau(w_m^{(j)}, \delta_m^{(i,j)}) \geq R^*$, proceed to Step 3. Otherwise, increment i by 1 and repeat Step 2.
3. Compute the expected warranty cost $\bar{C}_w(w_m^{(j)}, \delta_m^{(i,j)})$ and store the result.
4. If $\Delta\bar{C}_w\left(w_m^{(j)}, \delta_m^{(i,j)}\right) = \bar{C}_w\left(w_m^{(j)}, \delta_m^{(i,j)}\right) - \bar{C}_w\left(w_m^{(j)}, \delta_m^{(i-1,j)}\right) < 0$, then the optimal i^* is determined as $i^* = i - 1$. Otherwise, increment i by 1 and go back to Step 2.
5. If $i = j$, set $i = 1, j = j + 1$ and return to Step 2.

It is important to note that a smaller δ_m results in higher PM costs, despite reducing the MR cost. On the other hand, a larger δ_m increases the expected MR cost, leading to reduced profits. Therefore, an optimal PM interval must exist.

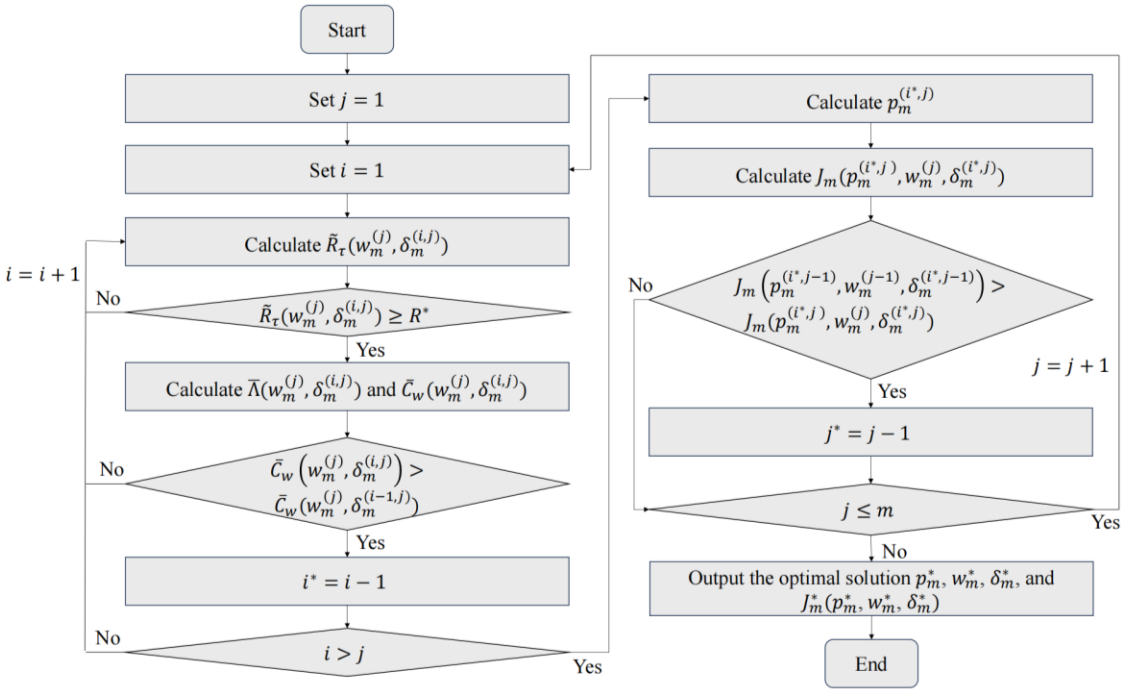


Fig. 4.4 The flowchart of the solution algorithm.

4.3 Case Study: Turbine Blisks

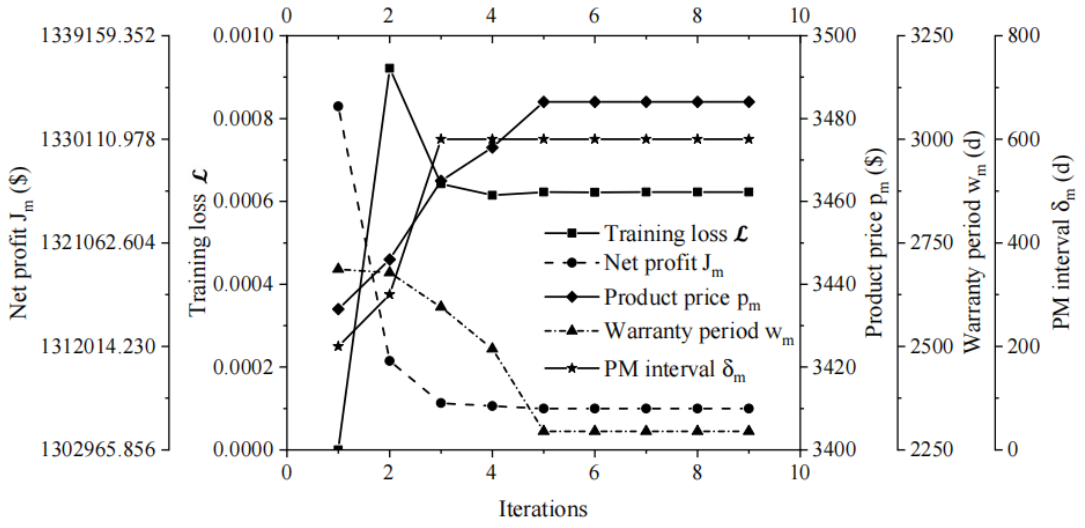
This section presents a turbine blisk case study to demonstrate the DT-based warranty prediction method. In the design phase, warranty costs are assessed by analyzing blisk reliability under predefined conditions. Since reliability data for new products is scarce, DT is used to generate degradation data. Warranty designers optimize profits by determining the blisk price p_m, warranty duration $w = \{w_m, w_c\}$, and PM interval $\delta = \{\delta_m, \delta_c\}$ for monopolized and competitive markets. The study uses data on materials, loads, costs, and market conditions to develop an optimal warranty strategy. Blisk material and load are modeled as random variables, while markets and costs are treated as fixed parameters affecting economic outcomes. Table 4.1 has outlined all input parameters used in the case study, including material, operational, cost, and market parameters.

Table 4.1 Parameters used in the case study.

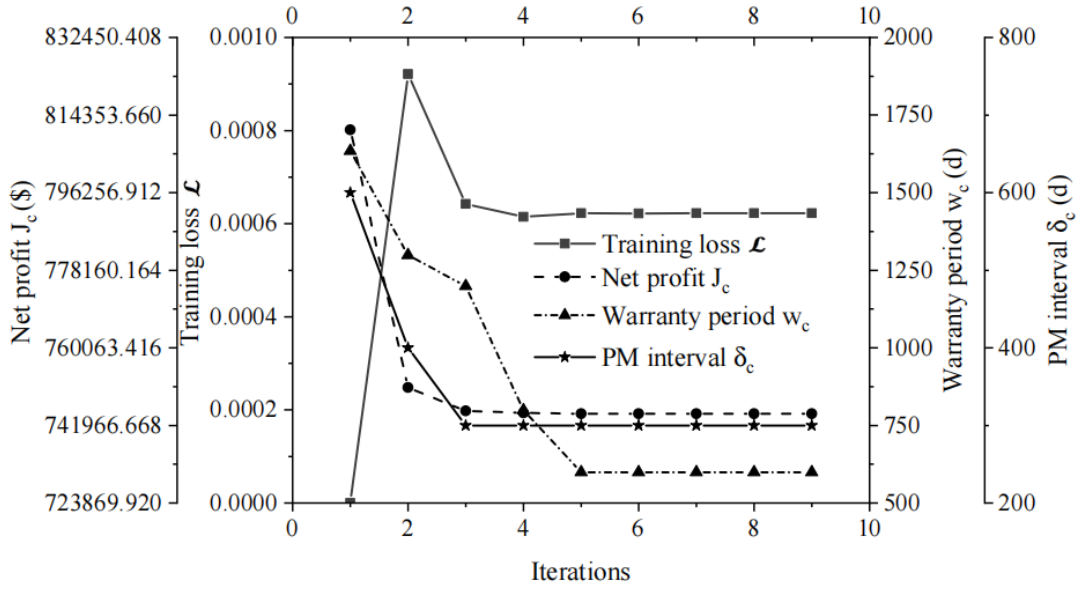
Parameter (Unit)	Distribution	Mean	Standard deviation
Density ρ, $(kg \cdot m^{-3})$	Normal	8,210	328.4
Thermal conductivity λ, $(W \cdot m^{-1} \cdot ℃^{-1})$	Normal	23	0.005
Young's modulus, $E, (MPa)$	Normal	163,000	4,890
Fatigue strength coefficient σ_f', (MPa)	Normal	1,419	42.5
Fatigue ductility coefficient ε_f'	Normal	50.5	1.53
Fatigue strength index b	Normal	-0.1	0.005
Fatigue ductility index c	Lognormal	-0.84	0.042
Rotational speed $\omega, (rad \cdot s^{-1})$	Normal	1,168	35
Blisk-root temperature $T_{\text{root}}, (K)$	Normal	1,173.15	35.2
Blisk-tip temperature $T_{\text{tip}}, (K)$	Normal	1,473.15	47
Required reliability level R^*	Deterministic	0.975	-
Age reduction factor ϑ	Deterministic	0.5	-

Burn-in duration τ_{bi}, (d)	Deterministic	1	-
MR cost C_{mr}, (\$)	Deterministic	5	-
Failure cost C_f, (\$)	Deterministic	1,500	-
PM cost C_{pm}, (\$)	Deterministic	10	-
Manufacturing cost C_m (\$)	Deterministic	700	-
Fixed burn-in cost C_{bif}, (\$)	Deterministic	0.1	-
Time-dependent set-up cost C_{bit} (\$)	Deterministic	0.01	-
Market lifecycle l_c (d)	Deterministic	10,000	-
Discounted rate η_D	Deterministic	0.001%	-
Amplitude factor κ_1	Deterministic	1,000	-
Time displacement constant κ_2, (d)	Deterministic	1	-
Price elasticity α_1	Deterministic	1.3	-
Warranty period elasticity α_2	Deterministic	0.1	-
Market price p_c (\$)	Deterministic	800	-
Fraction ratio ρ_f	Deterministic	0.9	-

To address the DT-based warranty optimization for monopolized and competitive markets, a SACO-PSO-KCNN metamodel is integrated with a numerical search algorithm. Fig. 4.5 shows convergence curves for training loss, net profits, and design variables, demonstrating the method's efficiency. Figs. 4.5a and 4.5b reveal a steady decrease in training loss with stability in net profits and design variables by the 5th epoch, confirming the model's effectiveness with minimal iterations. The SACO-PSO-KCNN configuration and performance, detailed in Subsection 3.4, are not reiterated here. This study uses only the LCF life distribution and reliability from SACO-PSO-KCNN for warranty optimization.



(a)



(b)

Fig. 4.5 Iteration curves of training loss, net profit, and design variables in the (a) monopolized market and (b) competitive market.

4.3.1 Optimization Results

The SACO-PSO-KCNN metamodel enables precise blisk reliability evaluation. Combining this with reliability constraints, costs, and market parameters allows optimal BW design for competitive and monopolized markets. In a monopolized market, design variables include blisk price, BW period, and PM interval. The optimal policy, determined via a numerical search (Subsection 4.2), yields p_m^* =3,484, w_m^*=2,296, and

δ_m^*=600, with a maximum profit of J_m^*=1,306,585.2. Figs. 4.6a–4.7c show these values as the globally optimal solution.

Fig. 4.6a, with δ_m=600, analyzes the impact of p_m and w_m on net profit in the monopolized market, showing a concave, non-monotonic trend as p_m and w_m increase. Similarly, Figs. 4.6b and 4.6c, with w_m=2,296 and p_m=3,484, exhibit the same concave behavior. Fig. 4.7 further confirms the existence and uniqueness of the optimal p_m, w_m, and δ_m.

In Fig. 4.7a, with w_m=2,296 and δ_m=600, net profit J_m as a function of p_m confirms a unique, globally optimal p_m. Fig. 4.7b, with p_m=3,484 and δ_m=600, shows a uniquely optimal warranty period. Fig. 4.7c illustrates a concave relationship between net profit and δ_m, confirming a unique profit-maximizing PM interval.

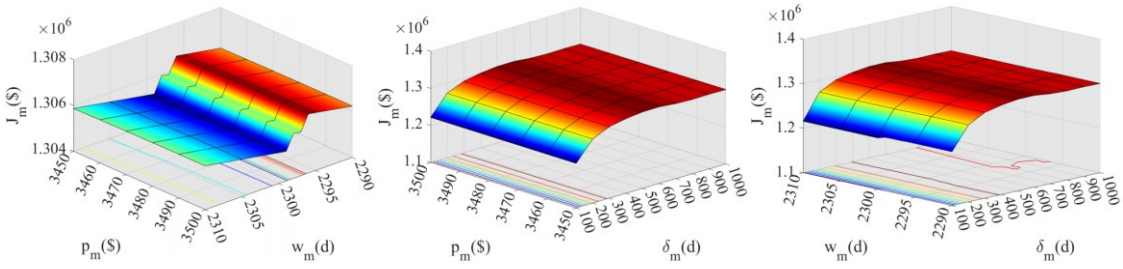


Fig. 4.6 J_m under varying (a) p_m and w_m, (b) p_m and δ_m, and (c) w_m and δ_m.

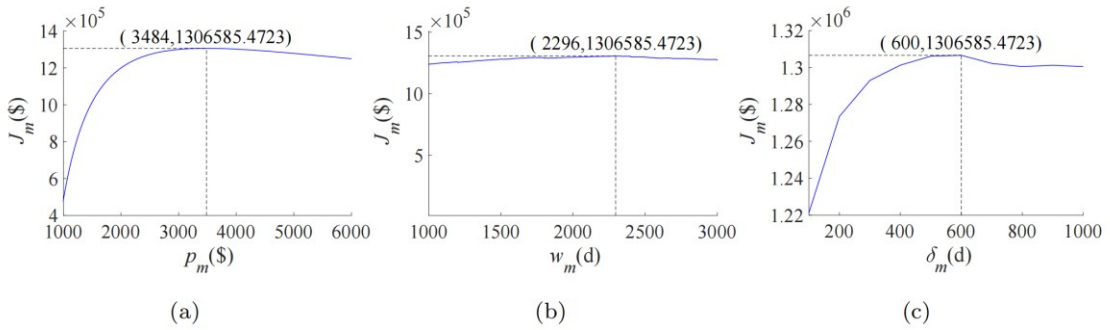


Fig. 4.7 J_m under varying (a) p_m, (b) w_m, and (c) δ_m.

Unlike the monopolized market, the product price p_c remains fixed in the competitive market, reducing the design variables to only w_c and δ_c. To maintain

profitability in the competitive market, it is crucial to satisfy the condition $\bar{C}_w(w_c) \leq \rho(p_c - \bar{C}_{\text{mbi}})$, where ρ represents the fraction ratio set by the manufacturer. Fig. 4.8 depicts the relationship between J_c, w_c, and δ_c.

Fig. 4.8a evaluates the influence of w_c and δ_c on net profit in the competitive market, highlighting its concave nature. The results demonstrate non-monotonic behavior and confirm the existence of a unique optimal solution. The net profit reaches its maximum value of 744,912.8 when w_c=599 and δ_c=300. However, there are regions where $J_c = 0$, caused by violations of the constraints $\tilde{R}_\tau(w_c, \delta_c) \geq R^*$, $w_c > \delta_c$, and $\bar{C}_w(w_c) \leq \rho(p_c - \bar{C}_{\text{mbi}})$. With δ_c=300, net profit J_c is plotted as a function of w_c in Fig. 4.8b. The figure reveals that the net profit forms a jagged curve for different warranty periods, with a unique globally optimal warranty period of w_c^*=599, corresponding to J_c^*=744,912.8. When w_c< 300, the condition w_c>δ_c is violated, resulting in J_c=0. Similarly, when w_c>1,800, net profit also drops to zero due to warranty costs exceeding acceptable limits, thereby violating the constraint $\bar{C}_w(w_c) \leq \rho(p_c - \bar{C}_{\text{mbi}})$. Likewise, with w_c=599, Fig. 4.8c plots J_c as a function of δ_c. The optimal PM interval is found to be δ_c^*=300. For δ_c>600, net profit becomes zero because an excessively large PM interval, with a fixed warranty period, violates the reliability constraint.

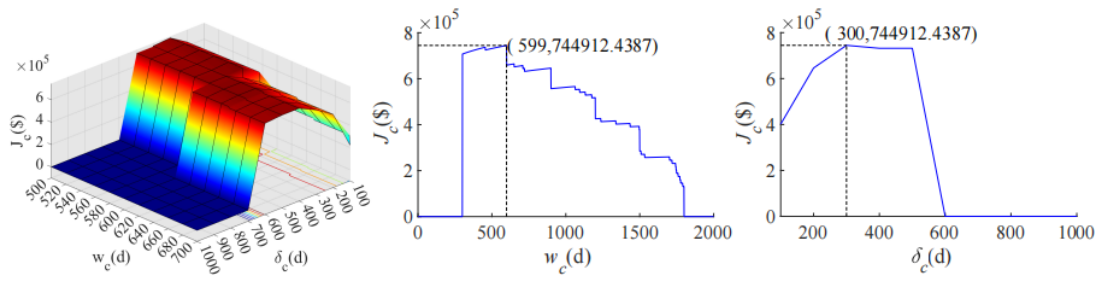


Fig. 4.8 J_c under varying (a) w_c and δ_c, (b) w_c, and (c) δ_c.

4.3.2 Sensitivity Analysis

A sensitivity analysis is performed to assess the impact of reliability, cost, and market factors on optimal decisions, providing meaningful managerial insights for decision-makers.

Once the blisk enters service, unpredictable usage changes may require warranty reassessment. DT technology accounts for variations in parameters like speed and temperature, which depend on environmental conditions and mission profiles. To assess their impact on reliability and warranty strategies, mean values are fixed while adjusting the load range within the 95% CI. Three usage scenarios (Table 4.2) are analyzed: Scenario I has the widest load range (harshest environment), while Scenario III has the narrowest (least demanding). Simulated using the SACO-PSO-KCNN metamodel, Fig. 4.9 shows reliability trends across cycles, with Scenario I having the lowest reliability and Scenario III the highest, indicating improved reliability as load range narrows.

Table 4.2 Statistical characteristics of load parameters across different usage scenarios.

Load parameter	Usage scenario (Load range, distribution, 95% CI)		
	I	II	III
ω	[1,063,1,273], Normal	[1,093,1,243], Normal	[1,123,1,213], Normal
T_{root}	[1,068,1,278], Normal	[1,098,1,248], Normal	[1,128,1,218], Normal
T_{tip}	[1,332,1,614], Normal	[1,362,1,584], Normal	[1,392,1,554], Normal

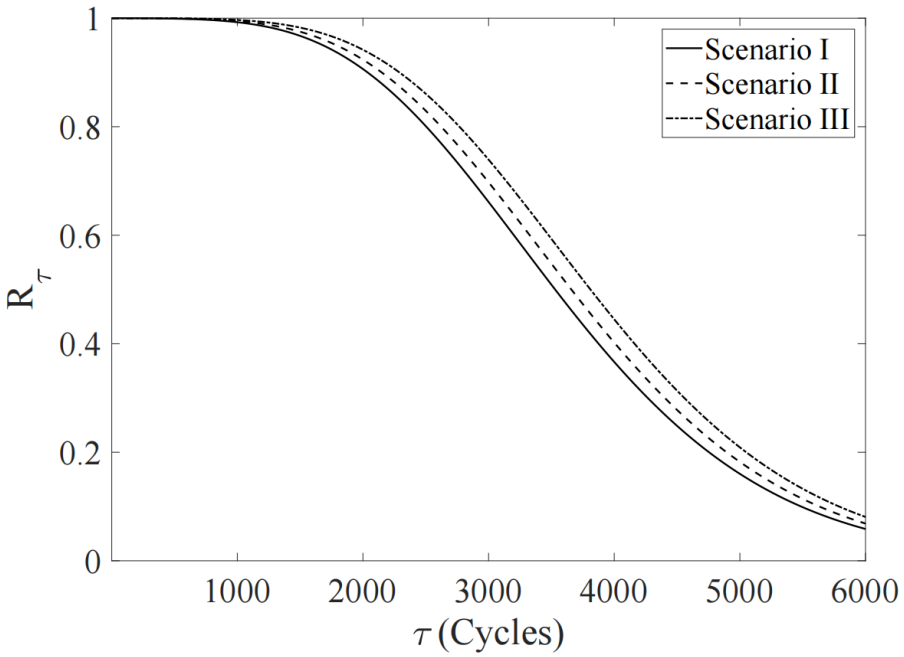


Fig. 4.9 Blisk reliability with applied cycles τ under different operational scenarios.

Table 4.3 summarizes p_m^*, w_m^*, δ_m^*, J_m^*, w_c^*, δ_c^*, and J_c^* across usage scenarios, R^*, and ϑ for monopolized and competitive markets. Table 4.2 and Fig. 4.9 show that reliability improves under milder conditions. Table 4.3 reveals that as usage becomes less severe, p_m^* decreases, w_m^* remains constant, and other values increase. Milder conditions reduce PM needs, lowering warranty costs and boosting profits, while harsher conditions require cautious PM strategies. In competitive markets, severe scenarios shorten w_c^* and increase PM frequency to mitigate failure risks, enhancing profits. Products with lower reliability are given shorter warranties to control costs. In monopolized markets, higher R^* reduces p_m^*, w_m^*, δ_m^*, and J_m^*, while competitive market values remain stable. Increased R^* shortens PM intervals, reducing warranty periods and prices to sustain sales. Higher ϑ raises p_m^*, w_m^*, J_m^*, w_c^*, δ_c^*, and J_c^*, while reducing δ_m^*. Improved PM recovery slows deterioration, extends warranty periods, boosts sales, and increases profits.

Table 4.3 Optimal net profits and design variables across different cost parameters in monopolized and competitive markets.

Monopolized market

			C_f =1,000				C_f =1,500				C_f =2,000			
C_m	C_{mr}	C_{pm}	P_m^*	w_m^*	δ_m^*	J_m^*	P_m^*	w_m^*	δ_m^*	J_m^*	P_m^*	w_m^*	δ_m^*	J_m^*
700	2	5	3,351	2,402	500	1,327,885.1	3,400	2,296	500	1,316,077.0	3,494	2,296	500	1,305,445.1
		10	3,420	2,399	600	1,319,578.7	3,483	2,296	600	1,306,657.6	3,580	2,296	500	1,295,881.0
		15	3,481	2,402	900	1,312,783.4	3,548	2,296	600	1,299,428.0	3,654	2,296	600	1,287,981.8
	5	5	3,351	2,402	500	1,327,803.7	3,401	2,296	500	1,316,013.6	3,494	2,296	500	1,305,381.8
		10	3,421	2,399	600	1,319,488.2	3,484	2,296	600	1,306,585.2	3,581	2,296	500	1,295,826.7
		15	3,482	2,402	900	1,312,683.8	3,549	2,296	600	1,299,355.6	3,655	2,296	600	1,287,918.4
	8	5	3,352	2,402	500	1,327,722.2	3,402	2,296	500	1,315,941.3	3,495	2,296	500	1,305,318.5
		10	3,422	2,399	600	1,319,397.7	3,484	2,296	600	1,306,512.8	3,581	2,296	500	1,295,763.4
		15	3,483	2,402	900	1,312,575.3	3,549	2,296	600	1,299,292.2	3,655	2,296	600	1,287,855.1
750	2	5	3,567	2,402	500	1,303,155.9	3,617	2,296	500	1,291,908.8	3,710	2,296	500	1,282,0913
		10	3,637	2,399	600	1,295,482.9	3,700	2,296	600	1,283,213.3	3,797	2,296	500	1,273,242.0
		15	3,698	2,402	900	1,289,221.4	3,765	2,296	600	1,276,526.5	3,871	2,296	600	1,265,921.8
	5	5	3,568	2,402	500	1,303,074.5	3,618	2,296	500	1,291,854.5	3,711	2,296	500	1,282,037.0
		10	3,638	2,399	600	1,295,401.4	3,700	2,296	600	1,283,149.9	3,798	2,296	500	1,273,187.7
		15	3,699	2,402	900	1,289,121.9	3,765	2,296	600	1,276,463.2	3,871	2,296	600	1,265,858.5
	8	5	3,569	2,402	500	1,303,002.1	3,618	2,296	500	1,291,791.1	3,711	2,296	500	1,281,973.7
		10	3,638	2,399	600	1,295,320.0	3,701	2,296	600	1,283,077.6	3,798	2,296	500	1,273,124.3
		15	3,699	2,402	900	1,289,022.3	3,766	2,296	600	1,276,390.8	3,872	2,296	600	1,265,795.2

Competitive market

			C_f =1,000				C_f =1,500				C_f =2,000			
C_m	C_{mr}	C_{pm}	P_c	w_c^*	δ_c^*	J_c^*	P_c	w_c^*	δ_c^*	J_c^*	P_c	w_c^*	δ_c^*	J_c^*
700	2	5	800	999	500	801,993.6	800	599	300	787,851.0	800	599	300	779,273.1
		10	800	999	500	756,851.3	800	599	300	744,961.7	800	599	300	736,383.8
		15	800	999	500	711,718.0	800	599	300	702,081.4	800	599	300	693,503.6
	5	5	800	999	500	801,830.7	800	599	300	787,796.7	800	599	300	779,218.8
		10	800	999	500	756,688.4	800	599	300	744,916.5	800	599	300	736,338.6
		15	800	999	500	711,555.1	800	599	300	702,027.2	800	599	300	693,449.3
	8	5	800	999	500	801,667.9	800	599	300	787,751.5	800	599	300	779,173.6
		10	800	999	500	756,525.5	800	599	300	744,862.2	800	599	300	736,284.3
		15	800	999	500	711,392.2	800	599	300	701,972.9	800	599	300	693,395.0
750	2	5	800	599	300	367,535.9	800	455	300	361,781.1	800	455	300	357,600.8
		10	800	599	300	324,646.6	800	455	300	320,050.0	800	272	200	315,951.1
		15	800	455	300	282,499.3	800	455	300	278,328.0	800	272	200	276,319.3
	5	5	800	599	300	367,490.7	800	455	300	361,754.0	800	455	300	357,582.7
		10	800	599	300	324,601.4	800	455	300	320,022.9	800	272	200	315,951.1
		15	800	455	300	282,472.1	800	455	300	278,300.8	800	272	200	276,319.3
	8	5	800	599	300	367,436.4	800	455	300	361,726.9	800	455	300	357,555.6
		10	800	599	300	324,547.1	800	455	300	320,004.8	800	272	200	315,951.1
		15	800	455	300	282,445.0	800	455	300	278,273.7	800	272	200	276,319.3

Table 4.4 summarizes optimal design variables and net profits for monopolized and competitive markets under varying PM, MR, manufacturing, and failure costs. In monopolized markets, as PM, MR, manufacturing, or failure costs rise (with others constant), p_m^* increases while J_m^* declines. PM, MR, and manufacturing costs do not

affect w_m^*, but higher failure costs shorten w_m^*, increasing warranty burdens. Rising PM costs lengthen δ_m^*, while higher failure costs shorten it. MR and manufacturing costs have minimal impact on δ_m^*. In competitive markets, PM, manufacturing, and failure costs reduce w_c^*, while MR cost has no effect. PM, MR, and manufacturing costs have little impact on δ_c^*, but failure costs shorten it. J_c^*, like J_m^*, consistently declines as costs rise, reducing marginal profit and shortening w_c^*. Manufacturing costs have the greatest impact on profitability, emphasizing the need to minimize them during production. Table 4.4 explores how market parameters α_1 (price elasticity) and α_2 (warranty elasticity) affect design variables and profits in monopolized markets. Higher α_1 reduces optimal price, shortens warranty periods, and lengthens PM intervals to optimize costs and revenue. Higher α_2 increases warranty periods and prices while shortening PM intervals to boost reliability and sales. Lower α_2 makes longer warranties less advantageous due to higher costs. Amplifying α_2 through targeted advertising could significantly increase profits.

Table 4.4 Optimal net profits and design variables in a monopolized market under different α_1 and α_2.

α_1	α_2=0.1				α_2=0.2				α_2=0.3			
	P_m^*	w_m^*	δ_m^*	J_m^*	P_m^*	w_m^*	δ_m^*	J_m^*	P_m^*	w_m^*	δ_m^*	J_m^*
1.2	4,914	2,399	600	3,000,440.9	5,575	2,990	500	6,656,888.9	5,593	2,999	500	14,824,856.3
1.3	3,484	2,296	600	1,306,585.2	4,014	2,978	500	2,860,191.1	4,027	2,990	500	6,366,436.1
1.4	2,814	2,296	600	584,525.0	2,866	2,399	600	1,269,486.9	3,242	2,978	500	2,807,710.5
1.5	2,246	1,701	600	267,831.9	2,412	2,296	600	577,286.3	2,779	2,978	500	1,261,343.4
1.6	1,990	1,672	1,000	124,867.6	2,144	2,296	600	266,927.0	2,184	2,399	600	579,095.9

Table 4.5 highlights the effect of p_c, ρ, and α_2 on design variables and net profits in the competitive market. When ρ becomes too small, the constraint $\bar{C}_w(w_c) \leq \rho(p_c - \bar{C}_{\text{mbi}})$ cannot be met, resulting in zero net profit. The findings indicate that

slight increases in ρ result in small improvements in the warranty period, PM interval, and net profits. However, p_c and α_2 have a more pronounced influence on both design variables and net profits.

Table 4.5 Optimal net profits and design variables in a competitive market under different α_2, p_c, and ρ.

P_c	ρ	α_2=0.1			α_2=0.2			α_2=0.3		
		w_c^*	δ_c^*	J_c^*	w_c^*	δ_c^*	J_c^*	w_c^*	δ_c^*	J_c^*
750	0.1	0	0	0	0	0	0	0	300	0
	0.3	455	300	320,312.4	599	300	599,002.4	599	300	1,135,571.0
	0.5	455	300	320,312.4	599	300	599,002.4	599	300	1,135,571.0
	0.7	455	300	320,312.4	599	300	599,002.4	599	300	1,135,571.0
	0.9	455	300	320,312.4	599	300	599,002.4	599	300	1,135,571.0
800	0.1	0	0	0	0	0	0	0	0	0
	0.3	599	300	744,681.2	999	500	1,455,883.4	999	500	2,904,528.1
	0.5	599	300	744,681.2	999	500	1,455,883.4	999	500	2,904,528.1
	0.7	599	300	744,681.2	999	500	1,455,883.4	999	500	2,904,528.1
	0.9	599	300	744,681.2	999	500	1,455,883.4	999	500	2,904,528.1
850	0.1	599	300	1,173,574.1	650	400	2,236,758.1	650	400	4,276,261.6
	0.3	999	500	1,180,812.8	1,063	600	2,358,006.3	1,063	600	4,735,014.2
	0.5	999	500	1,180,812.8	1,063	600	2,358,006.3	1,063	600	4,735,014.2
	0.7	999	500	1,180,812.8 1	1,063	600	2,358,006.3	1,063	600	4,735,014.2
	0.9	999	500	1,180,812.8	1,063	600	2,358,006.3	1,063	600	4,735,014.2

4.3.3 Impacts of DTs on Optimization Results

Subsection 3.4 outlines eight metamodels used to estimate blisk reliability for warranty optimization. Their accuracy heavily impacts the warranty solution. Fig. 4.10 compares the objectives and design variables from the SACO-PSO-KCNN metamodel with other metamodels, using normalization to ensure consistency. As shown, SACO-PSO-KCNN achieves the highest precision compared to other models, using FEM-PoF as the benchmark, confirming its effectiveness in optimizing warranty strategies.

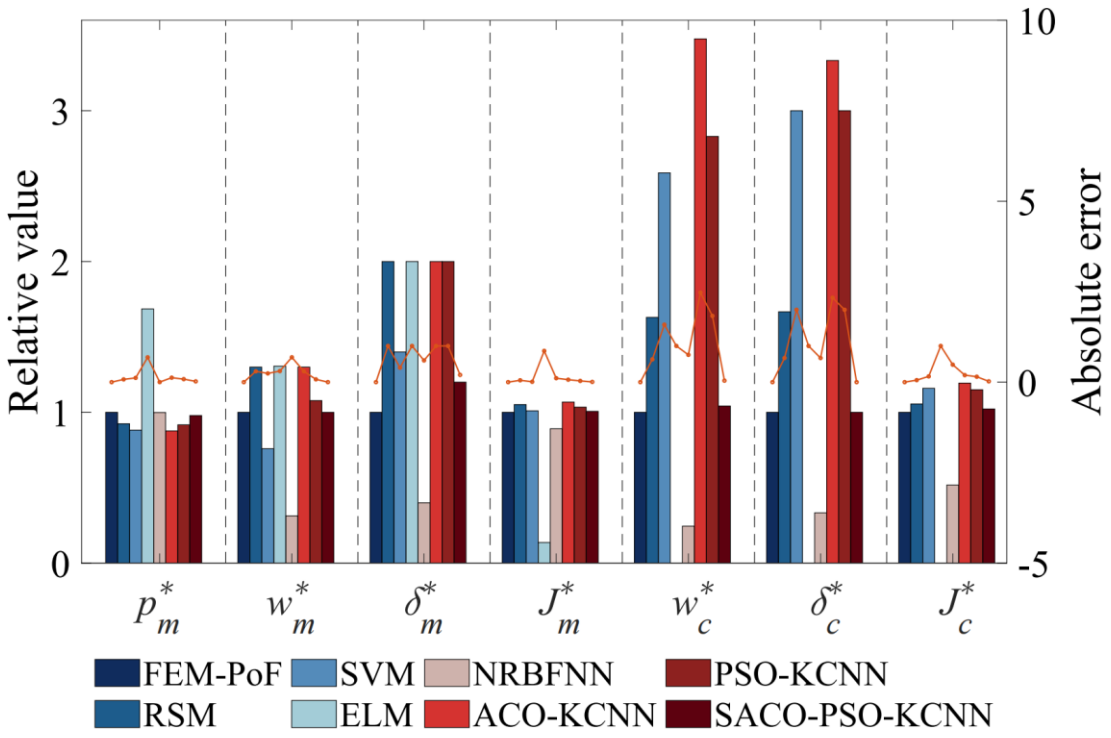


Fig. 4.10 The prediction precision of warranty design variables and net profits using different metamodels.

4.3.4 Warranty Policy Comparison

Fig. 4.11 shows the optimal results for NFRW, NFMW, and IPM-NFMW policies (Subsection 4.2), with normalization applied for consistency. The figure compares expected failures, warranty costs, and net profits for each policy in monopolized and competitive markets. In the monopolized market, the maximum profit difference between policies, such as NFMW and IPM-NFMW, is 2.74%, significant for manufacturers. The highest net profit, \$1,306,530.9, is achieved under IPM-NFMW, despite higher failure numbers and warranty costs, as its extended warranty and increased sales boost profits. In the competitive market, warranty constraints and PM costs reduce IPM-NFMW profits. The highest net profit, \$814,444.2, is achieved under NFMW. Managers can use Fig. 4.11 to choose the most suitable policy by balancing profit and practical factors.

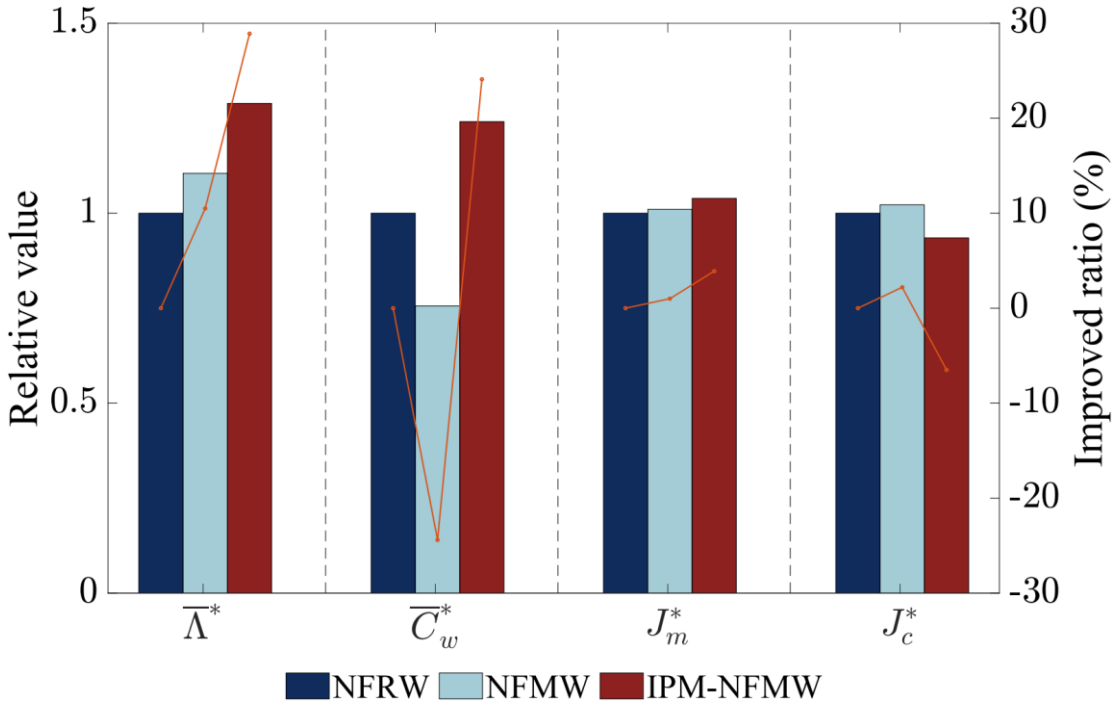


Fig. 4.11 Comparisons between NFRW, NFMW, and IPM-NFMW.

4.3.5 Extended Study: Joint Design of Product Materials and Warranty Services

Product-service systems are innovative business models where manufacturers offer integrated products and services. Using DT-PHM, manufacturers can maximize net profits. Product-service systems represent a shift from selling products to providing combined product-service packages [183]. Beyond warranty services, DT also supports joint decisions on product materials and warranties.

Designers face a tradeoff between reliability and manufacturing costs, as higher-reliability materials are more expensive. Reliability assessments help manufacturers design cost-effective warranty solutions. Longer warranties signal better quality but increase costs, prompting manufacturers to adopt joint decision-making for warranty and maintenance. In DT-based design, the goal is to maximize profit by optimizing materials, warranty durations, and maintenance cycles [184]. After DT-based reliability assessment, material choice, PM intervals, and maintenance visits are optimized for

profit. The optimization results are shown in Fig. 4.12. For material A, the maximum profit of $33,269 is achieved with 16 maintenance visits and 80 cycles. For material B, the profit is $6,155 with 9 visits and 160 cycles. Comparing profits, material A is preferred for turbine blisks, with optimal results at 16 visits and 80 cycles.

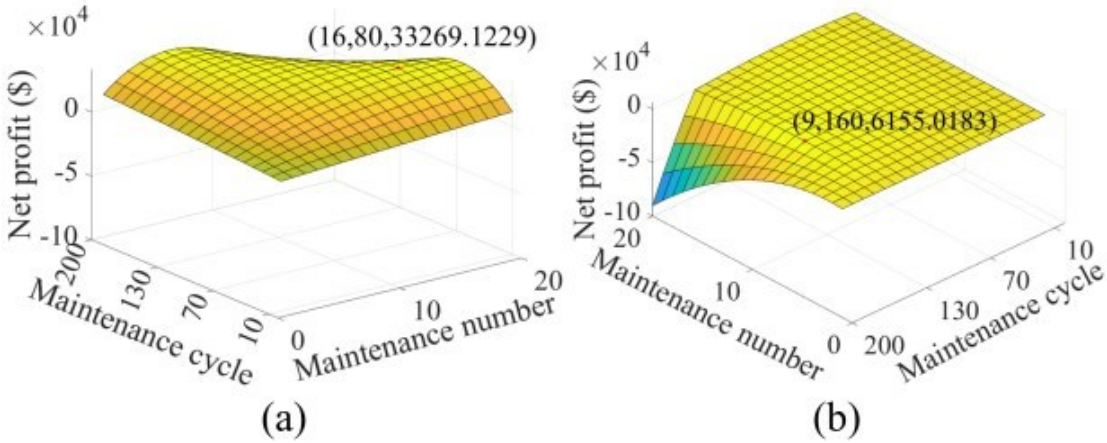


Fig. 4.12 Net profits w.r.t. maintenance cycle and maintenance number under product materials A and B. (a) Material A. (b) Material B.

4.4 Summary

To design warranties for new products lacking field failure data, this chapter proposes a novel DT-based warranty prediction method. The method aims to optimize product pricing, BW duration, and PM intervals by maximizing discounted profits, while accounting for IPM and varying market scenarios. Unlike prior studies using generic failure models (e.g., gamma processes or NHPP), this approach employs metamodel-based DTs for efficient degradation simulations and life reliability prediction. Using the predicted reliability, a mathematical model for IPM-NFMW policy profit is formulated for monopolized and competitive markets. Profit optimization is performed via a simulation-based enumeration algorithm, with market-specific optimality conditions analyzed. A case study on turbine blisks utilizing a

KCNN-based DT demonstrates the effectiveness of the proposed DT-based warranty design approach. Comparative analysis of metamodels confirms that the KCNN outperforms alternatives in both precision and efficiency, establishing it as a suitable method for reliability assessment and warranty cost prediction. The analysis highlights the substantial influence of product reliability, reliability constraints, costs, sales parameters, and market types on the formulation of warranty policies. Lower product reliability necessitates reduced prices, shorter warranty durations, and decreased preventive maintenance (PM) intervals due to accelerated degradation. In contrast, higher reliability justifies increased prices, longer warranty periods, and longer PM intervals. As failure risks increase, warranty durations should be shortened to mitigate costs. For price-sensitive markets, lowering prices enhances sales volume. These insights offer valuable guidance for manufacturers in optimizing warranty policies.

Chapter 5

Digital Twin-Based Predictive Maintenance Scheduling for Single-Unit Components with Cost Consideration

To ensure the cost-effective operation of components under varying operational conditions, dynamic predictive maintenance (PdM) strategies, which integrate prognostic modeling with decision-making processes, should be implemented. One major challenge in developing PdM strategies is the limited availability of run-to-failure (R2F) data, which is essential for data-driven prognostic modeling. Digital twins (DTs) offer a promising solution to this issue by generating synthetic R2F data and facilitating the development of accurate prognostic models. In this context, this chapter proposes a dynamic framework for DT-PdM designed to optimize an age replacement policy that minimizes operational and maintenance (O&M) costs. The framework is particularly suited for fatigue-prone components that lack R2F data, leveraging a DT for life prediction and employing a rolling horizon technique for dynamic maintenance optimization. The effectiveness of the proposed DT-PdM method is demonstrated through a cost-sensitive case study on a fan shaft, utilizing a Bayesian Bidirectional Long Short-Term Memory (BiLSTM) model-based DT. Furthermore, a comparison with other maintenance policies highlights the advantages of the age replacement policy.

5.1 Digital Twin-Based Predictive Maintenance Scheduling Framework for Optimizing Cost

The proposed framework uses a DT for PdM, even without R2F data, by employing an uncertainty-aware metamodel to predict real-time reliability under fatigue degradation across rolling horizons. A dynamic age replacement policy optimizes replacement time to minimize costs under varying inspection times.

The PdM framework illustrates the connection between the metamodel-based DT, cost models, and decision variables, as shown in Fig. 5.1. The metamodel is trained offline using historical data and FEM responses. Once trained, it probabilistically predicts future stresses. Combining fatigue life prediction with the rainflow counting algorithm enables estimation of current fatigue damage and future trends across rolling horizons. In the age replacement policy, if the physical asset's age exceeds the threshold, preventive replacement occurs at a cost denoted as C_R. Unexpected failures before the threshold result in corrective replacement at a higher cost, denoted as C_F, due to additional failure-induced costs. The framework predicts real-time reliability to optimize the maintenance threshold dynamically.

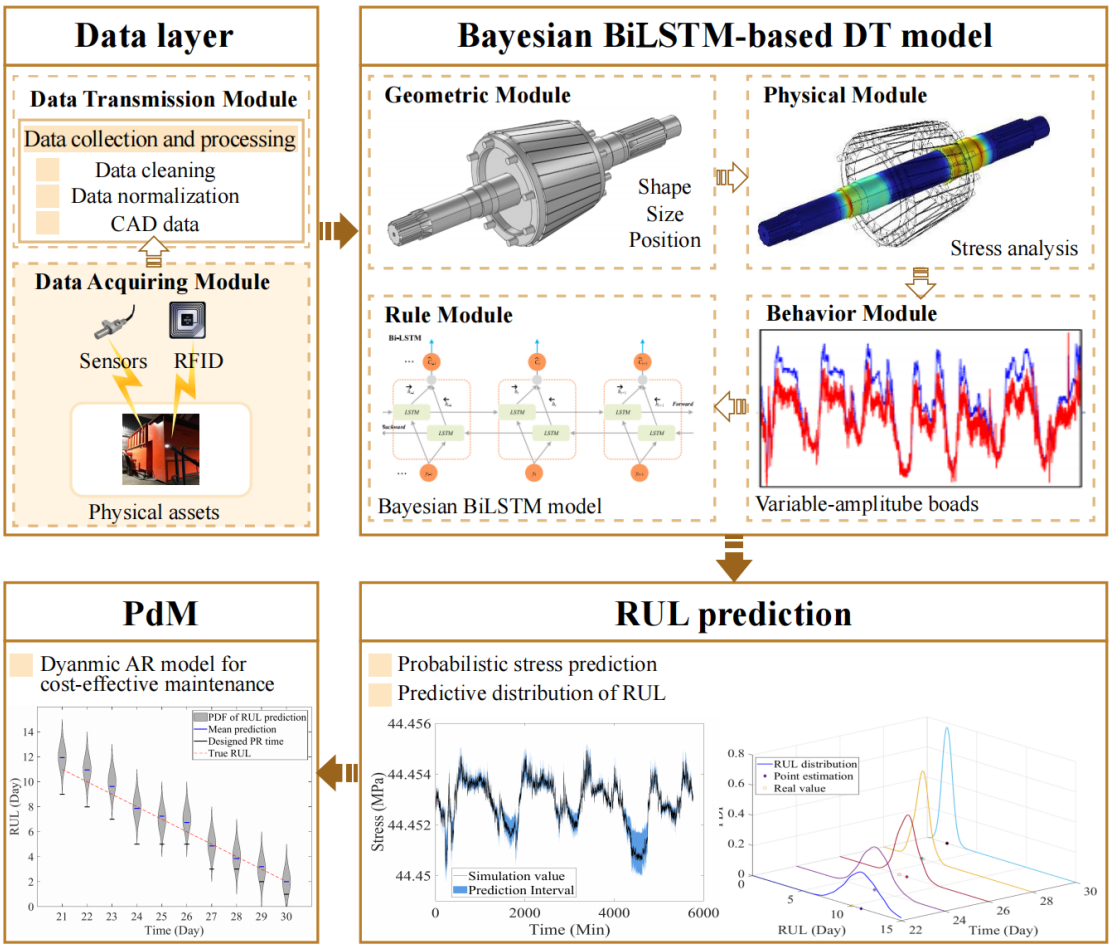


Fig. 5.1 The framework of the proposed DT-based dynamic maintenance.

To achieve the developed DT-PdM framework, the following assumptions are made.

1) The PA's dominant degradation mechanism is fatigue, and the load it experiences will vary. Fatigue degradation exhibits an upward, monotonically increasing trend.

2) Failures can be observed instantaneously.

3) Replacement time is negligible compared to the PA's expected life.

4) Preventive replacement and corrective replacement can restore the PA to an as-goo-as-new (AGAN) condition.

5) The maintenance analysis assumes an AGAN initial condition, indicating that the system begins the analysis in a brand-new state.

5.2 Predictive Maintenance Model Formulation and Optimization

5.2.1 Dynamic Maintenance Scheduling with Rolling Horizon Approach

The proposed dynamic PdM integrates RUL estimation with model-based maintenance planning. Using a rolling horizon, maintenance decisions are updated over time, enabling short-term planning while extending the horizon. Fig. 5.2 illustrates the operational mechanism of the rolling horizon-based dynamic planning. This approach reduces computational complexity in long-term scheduling with minimal error.

PAs can be continuously monitored over time, with the interval between successive inspections assumed to be constant, denoted as Δt_r. In the maintenance setting, the PA's failure probabilities at successive time $t_0, t_1, t_2,\ldots$ (where $t_1 - t_0 = t_2 - t_1 = \Delta t_r$ and $t_0 = 0$) can be updated based on the newly collected data. As the rolling horizon advances, the current health $D(t)$ at time step t is estimated using the trained model and historical sensor data $X_{0:t}$, while the epistemic uncertainty-aware metamodel dynamically predicts the RUL distribution. The only exception occurs during the first rolling horizon, when sensor data is unavailable initially. At the current time t, the PDF of the RUL $f_{\tau_f}^{(t)}(\tau) = \Pr\{\tau_f^{(t)}; X_{0:t}\}$, for any future time can be derived with the framework, where $\tau_f^{(t)} = \inf\{\tau \geq 0; D(t+\tau) \geq T_d\}$ and $D(t+\tau)$ is the damage at time step $t+\tau$. Using the RUL $f_{\tau_f}^{(t)}(\tau)$, the operator can determine the optimal maintenance time $\tau_m^{(t)*}$ at time step t to minimize the cost rate. Let $R^{(t)}(\tau)$ and $\lambda^{(t)}(\tau)$ denote the estimated reliability and failure rate given $X_{0:t}$, formally,

$$\begin{aligned} R^{(t)}(\tau) &= Pr\{D(t+\tau) < T_d; X_{0:t}\} \\ &= Pr\left\{\tau < \tau_f^{(t)}; X_{0:t}\right\} = 1 - \int_{\tau}^{\infty} f_{\tau_f}^{(t)}(\tau) d\tau \end{aligned} \tag{5.1}$$

$$\lambda^{(t)}(\tau) = \frac{f_{\tau_f}^{(t)}(\tau)}{R^{(t)}(\tau)} \tag{5.2}$$

Based on $R^{(t)}(\tau)$ and $\lambda^{(t)}(\tau)$ at time step t, we determine whether to schedule a maintenance action at time step $t + \tau$ or to only collect sensor data for the next Δt_r. With $R^{(t)}(\tau)$, the expected RUL at time t is:

$$E\left[\tau_f^{(t)}\right] = \int_0^{\infty} R^{(t)}(\tau) d\tau \tag{5.3}$$

The maintenance optimization model can incorporate probabilistic RUL during the forecasting stage. Maintenance actions are executed based on the total cost of maintenance planning, with plans developed for each rolling horizon. When new data becomes available at each inspection time, the information is refreshed, and updated PdM decisions are made. The maintenance time for the asset is updated every Δt_r cycles, i.e., every Δt_r cycles, we should determine when to replace the PA in the future. At each time step t, decisions $\tau_m^{(t)*}$ are made for future periods for minimize costs [185]. The algorithm for the rolling horizon approach to dynamic PdM planning is presented in Algorithm 5.1.

Algorithm 5.1 Rolling horizon approach for dynamic maintenance planning.

Input:
Rolling horizon length Δt_r
Output:
Optimal maintenance time $\tau_m^{(t)*}$ at time t.
1: For $t = 0$ to $t = t_n$ do
2: Set the current decision window as t, $t + 1$, …, $t + \Delta t_r$.
3: Update the current health status $D(t)$ and RUL $\tau_f^{(t)}$.
4: Utilize the health state $D(t)$ and RUL $\tau_f^{(t)}$ to minimize the objective function and determine the optimal maintenance time $\tau_m^{(t)*}$
5: Implement the maintenance decision for the current time step.
6: $t = t + 1$
7: End for

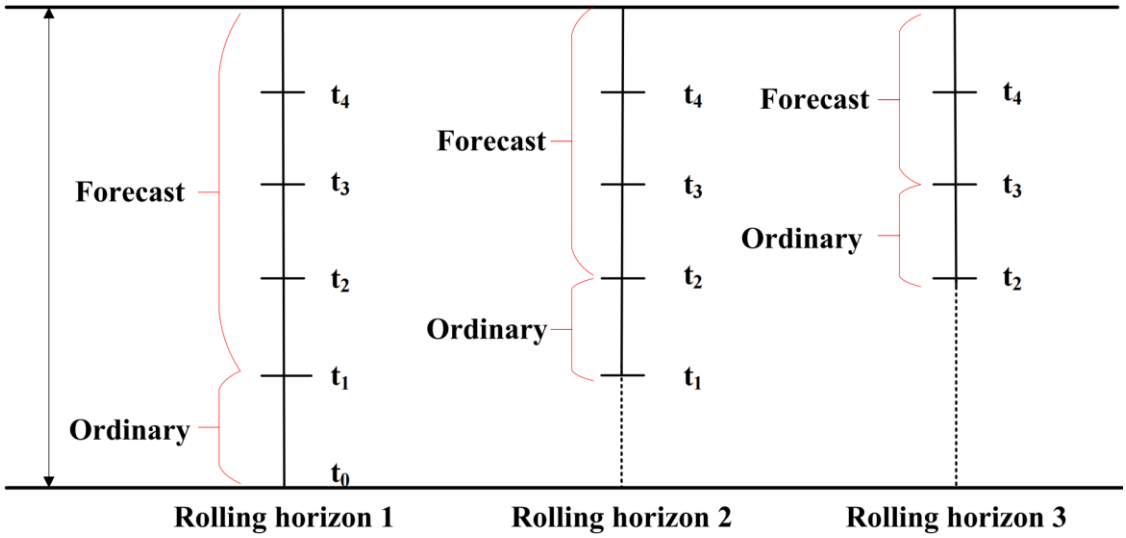


Fig. 5.2 Rolling horizon-based dynamic planning.

5.2.2 Cost Modeling under Age Replacement Policy

The supervisory control and data acquisition system monitors PA degradation loads X, triggering maintenance if levels exceed T_d. Frequent failures from excessive loading increase operational costs, making cost minimization crucial. This study develops an AR policy for cost-effective maintenance, detailed below.

The AR policy entails corrective replacement upon failure or preventive replacement at the age $\tau_m^{(t)}$, whichever occurs first. Thus, the expected total cost during the AR period $\tau_m^{(t)}$ is calculated as follows.

$$\begin{aligned} E\left[C^{(t)}(\tau)\right] &= C_F\, Pr\left\{D\left(t+\tau_m^{(t)}\right) \geq T_d\right\} + C_R Pr\left\{D\left(t+\tau_m^{(t)}\right) < T_d\right\} \\ &= C_F\left(1 - R^{(t)}\left(\tau_m^{(t)}\right)\right) + C_R R^{(t)}\left(\tau_m^{(t)}\right) \end{aligned} \tag{5.4}$$

Let the random variable $\tau_f^{(t)}$ represent the occurrence time of corrective replacement before the replacement age, and let $\tau_m^{(t)}$ denote the AR age. The length of a renewal cycle is $\tau_{\mathrm{pr}}^{(t)} = \min\left(\tau_f^{(t)}, \tau_m^{(t)}\right)$. The expected duration of the replacement time can be expressed as follows.

$$
\begin{aligned}
E\left[\tau_{pr}^{(t)}\right] &= \int_0^{\tau_m^{(t)}} ud\, Pr\{D(t+\tau) > T_d\} + \tau_m^{(t)} Pr\{D(t+\tau) \le T_d\} \\
&= \int_0^{\tau_m^{(t)}} ud\left(1 - R^{(t)}(u)\right) + \tau_m^{(t)} R^{(t)}\left(\tau_m^{(t)}\right) \\
&= \int_0^{\tau_m^{(t)}} R^{(t)}(u)du
\end{aligned} \tag{5.5}
$$

According to Eqs. (5.4) and (5.5), the expected maintenance cost per unit time can be derived, as shown in Eq. (5.6).

$$
E\left[C_{cr}^{(t)}\right] = \frac{C_R R^{(t)}\left(\tau_m^{(t)}\right) + C_F\left[1 - R^{(t)}\left(\tau_m^{(t)}\right)\right]}{\int_0^{\tau_m^{(t)}} R^{(t)}(u)du + t} \tag{5.6}
$$

Proposition 5.1. *The AR has a unique optimal solution* $\tau_m^{(t)*}$ *under historical data* $X_{0:t}$ *that satisfies*

$$
\lambda^{(t)}\left(\tau_m^{(t)}\right)\int_0^{\tau_m^{(t)}} R^{(t)}(u)du - \left(1 - R^{(t)}\left(\tau_m^{(t)}\right)\right) = \frac{C_R}{C_F - C_R}
$$

The proof is given in the Appendix A.5.1. According to Proposition 5.1, given $X_{0:t}$, the AR has a unique optimal solution.

Proposition 5.2. *If* $\lambda^{(t)}(X_{0:t}^1) \le \lambda^{(t)}(X_{0:t}^2)$*, then* $\mathrm{E}[C_{cr}^{(t)}](X_{0:t}^1) \le \mathrm{E}[C_{cr}^{(t)}](X_{0:t}^2)$.

Proposition 5.2 indicates that the minimum maintenance cost rate increases as failure rates $\lambda^{(t)}(\cdot)$ rise under given $X_{0:t}$. The proof is shown in Appendix A.5.2.

To obtain the optimal solution for the objective function defined by Eq. (5.6), we employ the MCS method, which is the most widely used approach for solving complex nonlinear functions. Obtaining an analytical solution is nearly intractable due to the nonlinear complexity of $\lambda^{(t)}(\tau)$. The process of DT-based dynamic maintenance decision process is shown in Fig. 5.3. The MCS relies on generating random events within a computer model, which is repeated multiple times to count the occurrences of specific events. In this study, PA failures are simulated with MCS under both the AR

policy. The pseudocode for the two models for expected cost rate prediction is provided in Algorithm 5.2. For different rolling horizons, it is necessary to adjust the reliability and recompute the aforementioned algorithms. The solution $\tau_m^{(t)*}$ suggests that to reduce costs, we should maintain the PA until the optimal time $t + \tau_m^{(t)*}$. Thus, the maintenance policy will evolve dynamically as deterioration progresses. In this study, we adopt a maintenance when $\tau_m^{(t)*} = 0$, as predictions during the later period are generally more accurate.

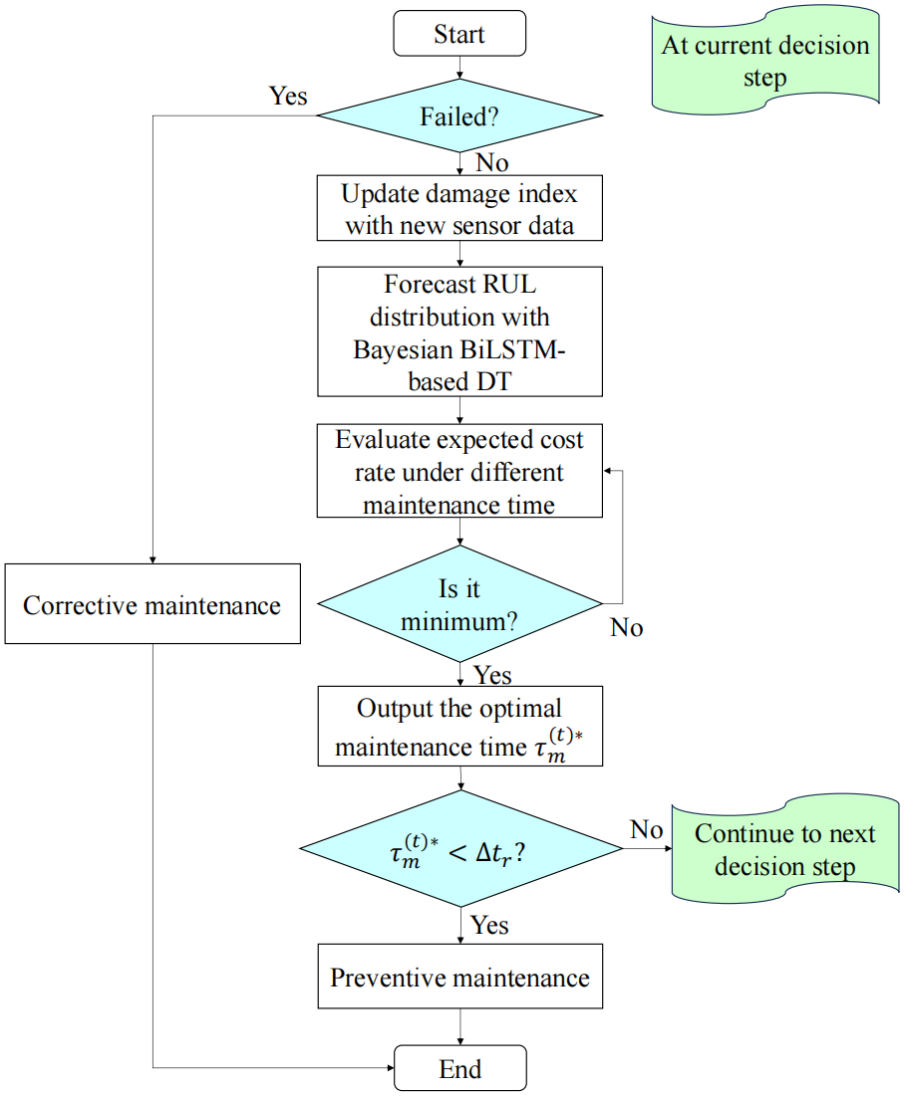


Fig. 5.3 DT-based dynamic maintenance decision process.

Algorithm 5.2 MCS for estimation of the expected cost rate under the AR policy at time t.

Input:
Asset reliability $R^{(t)}(\tau)$ estimated with the Bayesian BiLSTM model at time t, failure cost C_F, replacement cost C_R, and episode number n_e.
Output:
Expected cost rate at time t $E[C_{cr}^{(t)}](\tau_m^{(t)})$.
1: Generate a uniformly distributed $\tau_m^{(t)}$ between $\left[0, \tau_{m,u}^{(t)}\right]$.
2: $i = 0$, $EC_t = 0$.
3: While $i \leq n_e$ Do
4: $\tau = 1$ and $C_{sum}\left(\tau_m^{(t)}\right) = 0$
5: While $\tau \leq \tau_m^{(t)}$ Do
6: Generate a uniformly distributed number $r \in [0,1]$.
7: If $r > R^{(t)}(\tau)/R^{(t)}(\tau - 1)$ Do
8: $C_{sum}\left(\tau_m^{(t)}\right) = C_{sum}\left(\tau_m^{(t)}\right) + C_F$
9: $\tau = 1$
10: Break while
11: Else
12: $\tau = \tau + 1$
13: End if
14: End while
15: $EC_t\left(\tau_m^{(t)}\right) = EC_t\left(\tau_m^{(t)}\right) + \left[C_{sum}\left(\tau_m^{(t)}\right) + C_R\right]/\tau$
16: End while
17: $E\left[C_{cr}^{(t)}\right]\left(\tau_m^{(t)}\right) = EC_t(\tau_m^{(t)})/n_e$

5.3 Case Study: Variable-Speed Fan Shafts

5.3.1 Case Descriptions

A fan shaft from a flue gas desulfurization (FGD) system is used to demonstrate the proposed DT-PdM method. The FGD system, implemented in coal-fired power plants to reduce sulfur emissions, ensures cleaner production. Continuous operation results in fatigue damage to the fan shaft, and as the fan lacks redundancy, malfunctions can lead to environmental penalties. Figure 5.4 illustrates the FGD system and fan location.

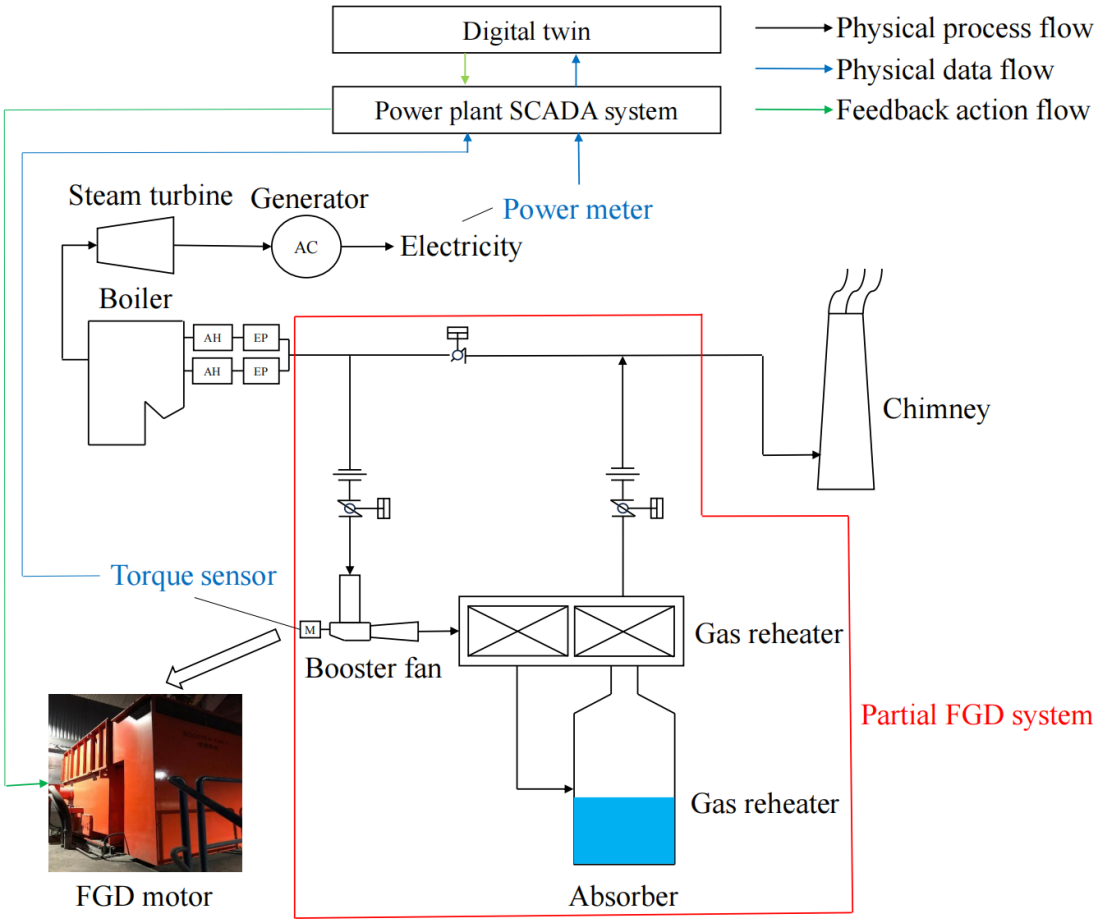


Fig. 5.4 The FGD system and the position of the FGD fan.

5.3.2 Maintenance Optimization with Bayesian BiLSTM-Based Digital Twins

The dynamic PdM scheduling is demonstrated using the predicted RUL distribution over a ten-day horizon. Decisions are made at intervals of two days, using updated predictions to optimize the age replacement policy. Corrective maintenance is costlier than preventive maintenance due to unplanned downtime, making a cost-sensitive approach essential.

Dynamic RUL predictions allow operators to adjust maintenance schedules using new monitoring data. Fig. 5.5 shows the decision process at steps 8 and 9. After 24 days of operation (step 8), the Bayesian BiLSTM model predicts zero failure probability for the next three days (Fig. 5.5a), so no replacement is scheduled. At step 9, updated sensor

data shows a 4% failure probability within three days (Fig. 5.5b), prompting a replacement on day 30. The true RUL is five days, meaning the dynamic AR policy schedules replacement two days before failure.

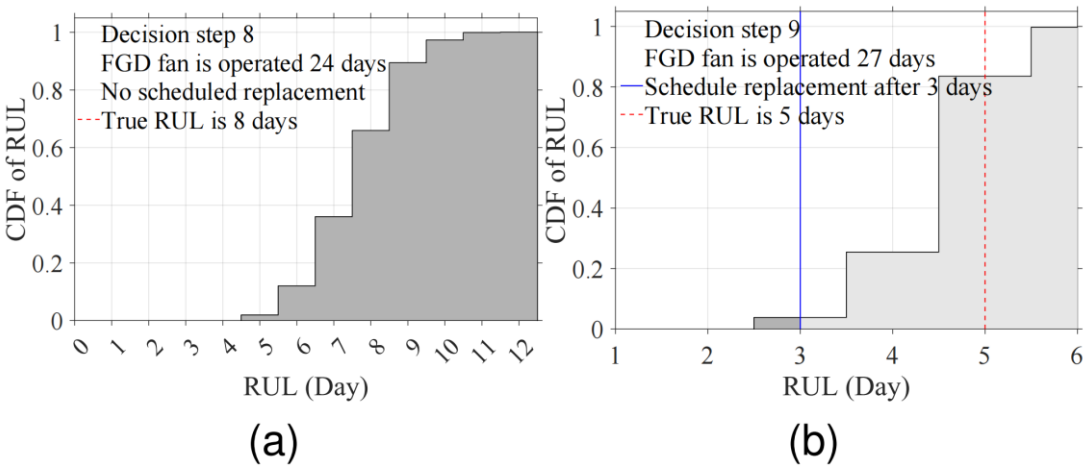


Fig. 5.5. RUL distribution estimation and AR schedules for the FGD fan shaft across two consecutive decision steps (Rolling horizon Δt_r=3 days).

Fig. 5.6 shows how the estimated RUL distribution evolves with daily updates to the dynamic AR policy as new sensor data arrives, adjusting until τ_f=1 is met. In Fig. 5.6, the dashed red line shows the true RUL, solid blue lines represent mean maintenance time, and solid black lines indicate the optimal maintenance time by the AR policy. With replacement cost C_R = 150 and failure cost C_F=250, the optimal thresholds are set lower than the mean, as failures cost twice as much. The dynamic AR policy schedules maintenance just before the true RUL, minimizing failure risks and enhancing cost-efficiency.

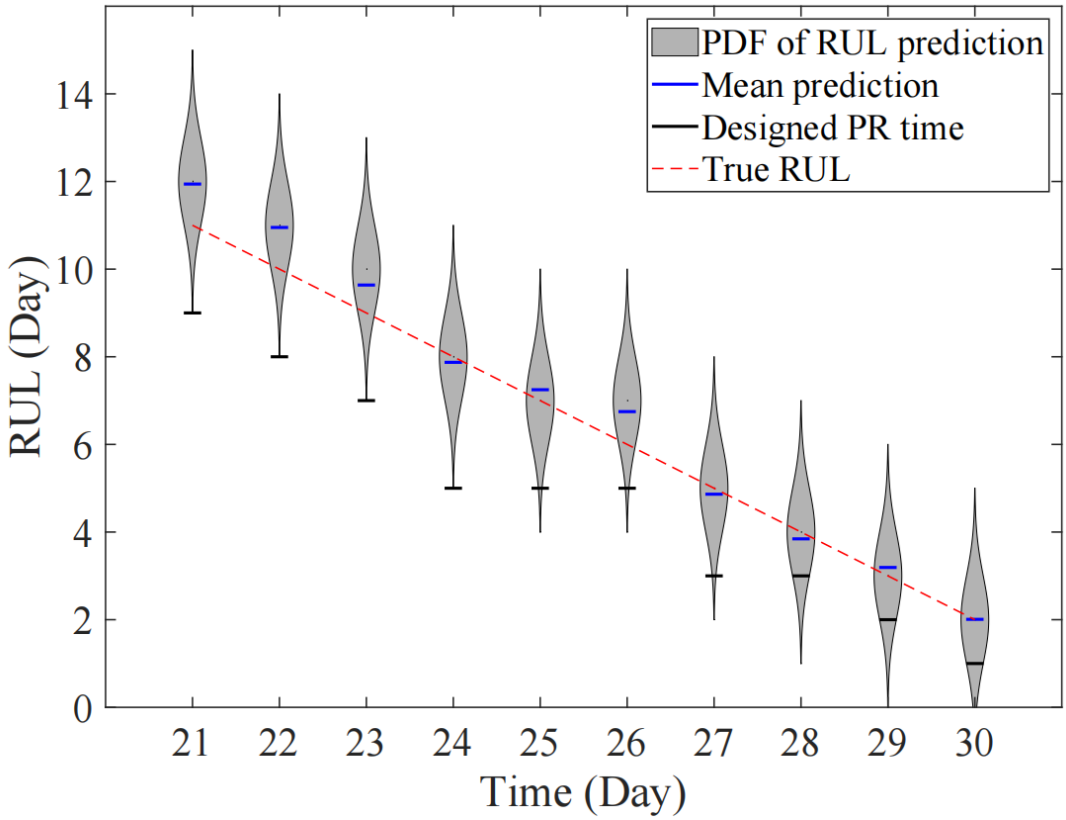


Fig. 5.6 Dynamic AR policy for the FGD fan shaft utilizing the Bayesian BiLSTM model.

5.3.3 Sensitivity Analysis

The proposed PdM method supports maintenance decisions by enabling real-time evaluation of C_R, C_F, and T_d, offering flexibility to adapt strategies in dynamic conditions. Fig. 5.7 highlights decision flexibility with parameter changes. In Fig. 5.7a, the expected cost rate $\mathrm{E}[C_{\mathrm{cr}}]$ under the dynamic AR policy is more robust to C_F and T_d, but more sensitive to C_R. Figs. 5.7a and 5.7b show that increasing C_R raises $\mathrm{E}[C_{\mathrm{cr}}]$ and τ_m^*, while increasing C_F raises $\mathrm{E}[C_{\mathrm{cr}}]$ but lowers τ_m^*. Higher T_d reduces $\mathrm{E}[C_{\mathrm{cr}}]$ and increases τ_m^*. CM is advised when C_F is low, but as C_F increases, earlier maintenance is recommended, aligning with the principle that higher failure costs justify earlier replacements.

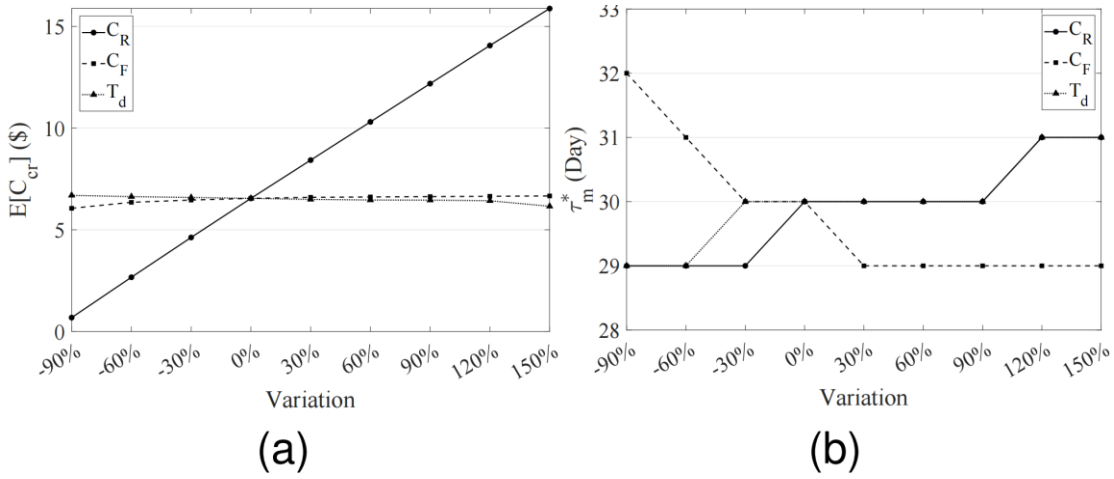


Fig. 5.7 Impacts of the replacement cost, the failure cost, and the maintenance threshold on the expected cost rate and the optimal maintenance time.

5.3.4 Maintenance Policy Comparison

This section evaluates the proposed dynamic age replacement policy by comparing it to other strategies:

(1) IM policy (Policy I): This policy operates on the assumption that the FGD fan shaft is replaced exactly at the end of its true RUL, thereby avoiding any unplanned maintenance activities. The associated cost rate is as follows:

$$C_{cr} = \frac{C_R}{\tau_f^{(0)}} \tag{5.7}$$

(2) AR Policy with Mean RUL (Policy II): This policy schedules maintenance for the FGD fan shaft based on the mean RUL, $E[\hat{\tau}_f]$, as predicted by the BiLSTM model. The objective is to evaluate the advantages of leveraging uncertainty-aware RUL predictions compared to point-based RUL predictions. The cost rate for the r-th renewal cycle is calculated as follows:

$$C_{cr} = \begin{cases} \frac{C_R}{E\left[\hat{\tau}_f^{(t)}\right]+t}, & E\left[\hat{\tau}_f^{(t)}\right] < \tau_f^{(t)} \\ \frac{C_F}{\tau_f^{(t)}+t}, & E\left[\hat{\tau}_f^{(t)}\right] \geq \tau_f^{(t)} \end{cases} \tag{5.8}$$

(3) AR Policy with RUL Distribution (Policy III): Comprehensive details regarding this policy can be found in Subsection 6.2.

(4) MR-AR Policy with RUL Distribution (Policy IV): Under this policy, the asset is replaced at a predefined replacement age, $\tau_m(t)$. If the asset fails before reaching $\tau_m(t)$, it undergoes MR. Let $\mathrm{E}[\Lambda^{(t)}(\tau_1, \tau_2)]$ denote the expected number of MR events within the interval $[\tau_1, \tau_2]$ as predicted at time t. The expected cost incurred before $\tau_m(t)$ is calculated as follows:

$$\begin{aligned} E\left[C^{(t)}(\tau)\right] &= (C_F + C_M)E\left[\Lambda^{(t)}(0,\tau)\right] \\ &= (C_F + C_M)\int_0^{\tau} \lambda^{(t)}(u)du, \left(\tau < \tau_m^{(t)}\right) \end{aligned} \tag{5.9}$$

As the replacement time is not taken into account, the expected length of the renewal cycle corresponds to the anticipated duration of the replacement time, $\tau_m^{(t)}$. Consequently, the expected cost rate for the MR-AR policy, $\mathrm{E}[C_{\mathrm{cr}}^{(t)}]$, can be determined as expressed in Eq. (5.10).

$$\begin{aligned} E\left[C_{cr}^{(t)}\right] &= \frac{(C_F+C_M)E\left[\Lambda^{(t)}\left(0,\tau_m^{(t)}\right)\right]+C_R}{\tau_m^{(t)}+t} \\ &= \frac{(C_F+C_M)\int_0^{\tau_m^{(t)}} \lambda^{(t)}(u)du+C_R}{\tau_m^{(t)}+t} \end{aligned} \tag{5.10}$$

where C_F denotes the failure cost, C_M denotes the MR cost, and C_R denotes the replacement cost.

The optimal maintenance time $\tau_m^{(t)*}$ that minimizes the expected cost rate $\mathrm{E}[C_{\mathrm{cr}}^{(t)}]$ can be identified via differentiating $\mathrm{E}[C_{\mathrm{cr}}^{(t)}]$ w.r.t. $\tau_m^{(t)}$ and set $\frac{\partial \mathrm{E}[C_{\mathrm{cr}}^{(t)}]}{\partial \tau_m^{(t)}} = 0$. Then, we have

$$\mathcal{J}\left(\tau_m^{(t)}\right) = \frac{C_R}{C_F+C_M} \tag{5.11}$$

where $\mathcal{J}\left(\tau_m^{(t)}\right) = \tau_m^{(t)}\lambda^{(t)}\left(\tau_m^{(t)}\right) - \int_0^{\tau_m^{(t)}} \lambda^{(t)}(u)du$. The following proposition can be formulated concerning the optimal $\tau_m^{(t)*}$ [185].

Proposition 5.3. *If*

$$\lim_{\tau_m^{(t)} \to \tau_{m,u}^{(t)}} \mathcal{J}\left(\tau_m^{(t)}\right) < \frac{C_R}{C_F + C_M}$$

then there exists a unique solution $\tau_m^{(t)*}$ *to Eq. (5.11). Otherwise,*

$$infE\left[C_{cr}^{(t)}\left(\tau_m^{(t)}\right)\right] = \lim_{\tau_m^{(t)} \to \tau_{m,u}^{(t)}} E\left[C_{cr}^{(t)}\left(\tau_m^{(t)}\right)\right]$$

$\lim_{\tau_m^{(t)} \to \tau_{m,u}^{(t)}} \mathcal{J}\left(\tau_m^{(t)}\right) < \frac{C_R}{C_F + C_M}$ guarantees the uniqueness of the solution $\tau_m^{(t)*}$ to Eq. (5.11). To satisfy the condition, $\frac{C_R}{C_F + C_M}$ should be relatively large, indicating that the replacement cost should be larger than the failure and MR costs.

Algorithm 5.3 MCS for estimation of the expected cost rate under the MR-AR policy at time t.

Input:

Asset reliability $R^{(t)}(\tau)$ estimated with the Bayesian BiLSTM model at time t, failure cost C_F, MR cost C_M, replacement cost C_R, and episode number n_e.

Output:

Expected cost rate at time t $\mathrm{E}[C_{\mathrm{cr}}^{(t)}](\tau_m^{(t)})$.

1: Generate a uniformly distributed $\tau_m^{(t)}$, between $[0, \tau_{m,u}^{(t)}]$.
2: $i = 1$, $\mathrm{EC}_t = 0$.
3: While $i \leq n_e$ Do
4: $\tau = 1$, $C_{sum}\left(\tau_m^{(t)}\right) = 0$.
5: While $\tau \leq \tau_m^{(t)}$ Do
6: Generate a uniformly distributed number $r \in [0,1]$.
7: If $r > R^{(t)}(\tau)/R^{(t)}(\tau - 1)$ Do
8: $C_{sum}\left(\tau_m^{(t)}\right) = C_{sum}\left(\tau_m^{(t)}\right) + C_F + C_M$
9: $\tau = \tau + 1$
10: Else
11: $\tau = \tau + 1$
12: End if
13: End while
14: $EC_t\left(\tau_m^{(t)}\right) = EC_t\left(\tau_m^{(t)}\right) + \left[C_{sum}\left(\tau_m^{(t)}\right) + C_R\right]/\tau_m^{(t)}$
15: End while
16: $E\left[C_{cr}^{(t)}\right]\left(\tau_m^{(t)}\right) = EC_t(\tau_m^{(t)})/n_e$

(5) CM policy (Policy V). This approach involves replacing the FGD fan shaft immediately upon failure. The corresponding cost rate is given by:

$$C_{cr} = \frac{C_F}{\tau_f^{(0)}} \tag{5.12}$$

Fig. 5.8 provides an overview of the performance of various maintenance policies, assessed based on $\mathrm{E}[C_{\mathrm{cr}}]$, the number of maintenance actions, and the total maintenance cost over a specified time period. As illustrated in Fig. 5.8a, Policy III achieves the lowest expected cost rate among all policies, with the exception of Policy I. Figs 5.8b and 5.8c present a comparative analysis of the number of maintenance actions and total maintenance costs. It is evident that Policy III results in the lowest total cost compared to all other policies, except Policy I. Furthermore, the number of CM actions under Policy III is fewer than those under Policies II and V. Although the total number of replacements under Policy III is slightly higher—since the FGD fan shaft is replaced earlier to avoid unplanned maintenance—this increase is compensated by substantial economic efficiency and improved maintenance reliability.

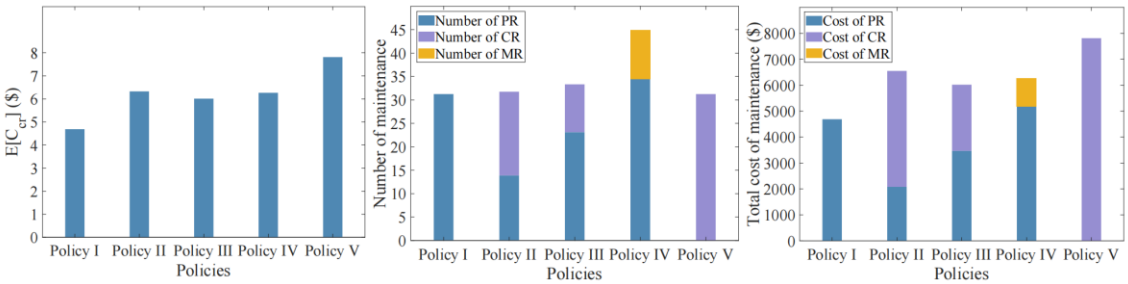


Fig. 5.8 Performance of policies I, II, III, IV, and V in terms of expected cost rate, maintenance frequency, and total maintenance costs.

Fig. 5.9 compares maintenance policies based on $\mathrm{E}[C_{\mathrm{cr}}]$, maintenance actions, and total costs over a given period. In Fig. 5.9a, Policy III achieves the lowest expected cost rate, except for Policy I. Figs. 5.9b and 5.9c show that Policy III has the lowest total cost and fewer CM actions compared to Policies II and V. While Policy III involves slightly more replacements due to earlier replacements to avoid unplanned maintenance, this is offset by greater economic efficiency and improved reliability.

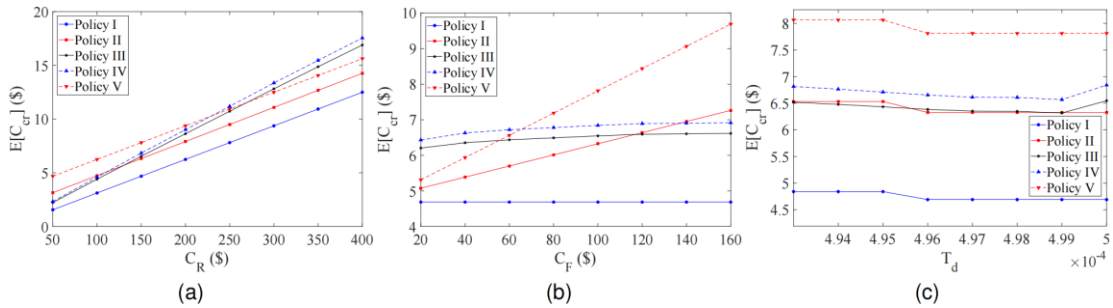


Fig. 5.9 Comparisons of expected cost rate for different maintenance policies: Policy I, Policy II, Policy III, Policy IV, and Policy V.

5.4 Summary

This chapter addresses the dynamic PdM problem for components using an epistemic uncertainty-aware metamodel-based DT approach. The metamodel can predict interval-based RUL rather than point-wise RUL, enabling optimal dynamic AR decisions at different rolling horizons. Validation with fan shafts shows the effectiveness of the proposed DT-PdM method, which employs Bayesian BiLSTM models as metamodels in the DT. Numerical experiments reveal that the Bayesian BiLSTM model outperforms point-estimate RUL models. Furthermore, policy comparisons highlight the superiority of the AR policy with estimated RUL distribution. A contradistinctive analysis is also conducted to support maintenance staff in selecting the optimal maintenance policy under varying conditions.

Chapter 6

Digital Twin-Based Predictive Maintenance Scheduling for Single-Unit Components with Hybrid Cost and Availability Consideration

In addition to cost, availability is a critical metric for evaluating the effectiveness of a predictive maintenance (PdM) policy. To address this, this chapter develops a performance index that integrates both cost and availability to guide the design of an effective PdM policy. Unlike the PdM model presented in Chapter 5, which assumes a predefined degradation threshold, this chapter treats the degradation threshold as a design variable, acknowledging the positive correlation between maintenance costs and the degradation threshold. Moreover, this chapter considers a more realistic industrial scenario by optimizing spare-part procurement time, rather than maintenance scheduling, to account for spare-part lead time and inventory costs. This ensures better alignment with practical industry requirements. To enhance the effectiveness of the PdM model, competing failure modes and maintenance durations are explicitly incorporated, as multiple failure modes are common in practice and should be accounted for. Additionally, maintenance duration, which directly impacts availability and downtime costs, is included in the model. In summary, this chapter proposes a digital twin-based predictive maintenance (DT-PdM) method for components, leveraging DT technology to model fatigue degradation. The DT predicts life reliability dynamically and supports a dynamic PdM model that optimizes degradation thresholds,

shock thresholds, and spare-part procurement times, while considering availability, competing failure modes, and maintenance duration. A simulation-based optimization algorithm is employed to identify optimal decision variables that enhance availability and reduce operational costs. Validation of the proposed method is achieved using turbine blisks through the implementation of an adaptive kriging-based DT. Additionally, the effects of operational conditions, availability, maintenance rate, and cost-related parameters on the PdM solution are evaluated.

6.1 Digital Twin-Based Predictive Maintenance Scheduling Framework for Optimizing Cost and Availability

The developed DT-PdM architecture can predict asset reliability and inform CBM decisions adaptively under progressive operational scenarios. Reliability information $R(\tau)$ is crucial for effectively planning maintenance activities. Due to the lack of R2F data, a metamodel-based DT $\hat{g}(X)$ is employed to replace the time-consuming FEM-PoF analysis $g(X)$ for probabilistic degradation simulation. With the metamodel updating mechanism presented in Subsection 3.2, the asset reliability can be updated adaptively under new operational scenarios $\mathcal{D}_X$, which allows for the design of the CBM policy. Taking into account the incomplete maintenance effect, the optimal degradation threshold T_m^*, shock threshold H^*, and timing for spare part purchases τ_p^* can be determined to increase average availability $\bar{A}$ and decrease costs $C_{\mathrm{cr}}\delta_{\mathrm{lc}}$ with $[0, \delta_{\mathrm{lc}}]$ in the CBM policy. The DT-CBM is characterized by its real-time capability, predictability, dynamism, and interactivity. Real-time capability, predictability, and

dynamically replaces the time-consuming FEM-PoF simulator for predicting life distribution. Interactivity is attained through the CBM model.

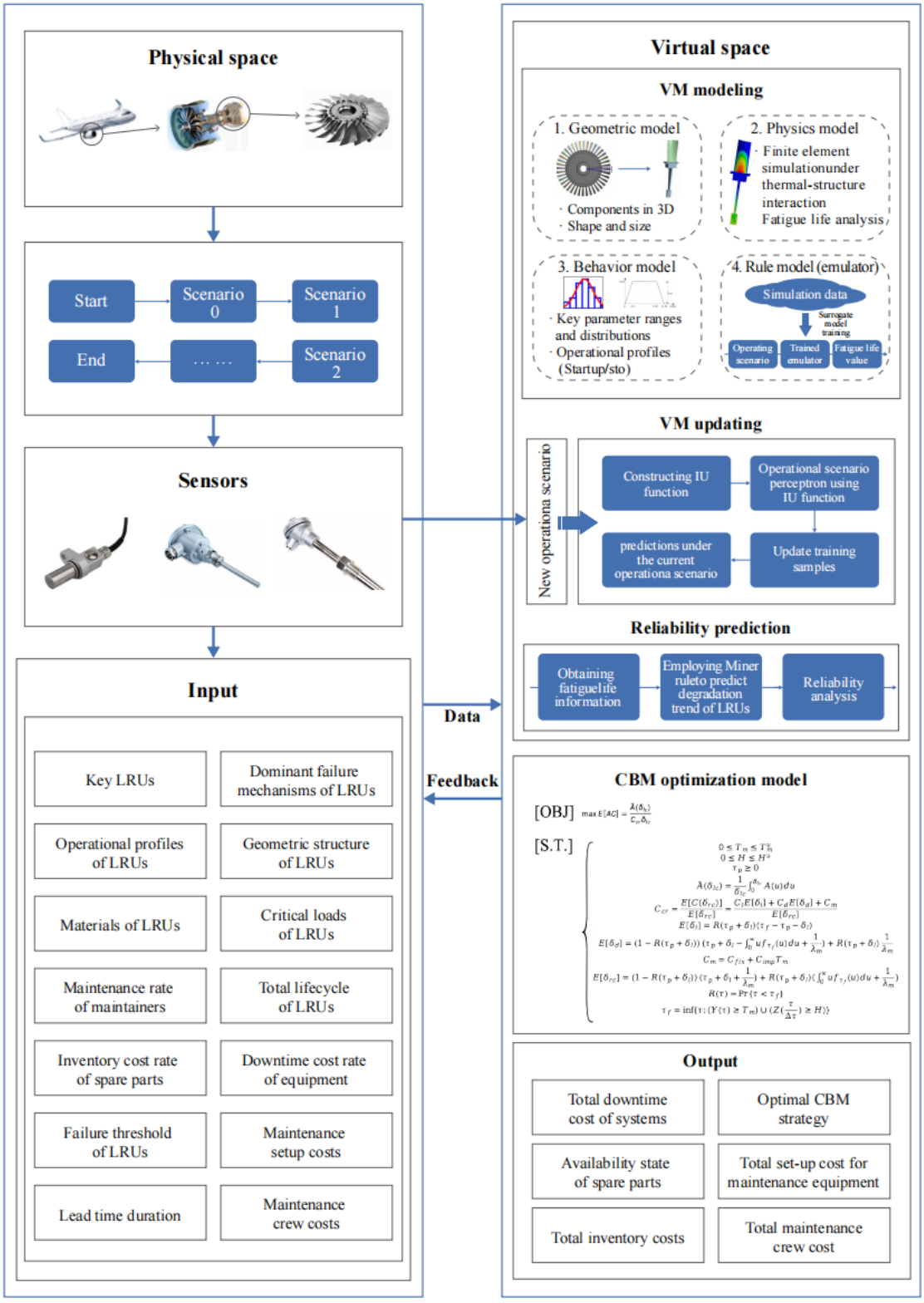


Fig. 6.1 The proposed DT-based proactive maintenance architecture.

The DT-CBM architecture is illustrated in Fig. 6.1. In the architecture, the interaction between PAs and their VMs is crucial [21]. A sensor system is employed to continuously monitor the PA and collect degradation-related data. Initially, data $X \in \mathcal{D}_X$ are collected and preprocessed to determine the data distribution $f_X(X)$ and to create the training set X_c for the metamodel construction. The target value is determined using the FEM-PoF $g(X)$. If new operational scenarios are detected, the metamodel is updated with new training samples. In addition to degradation-induced failures, instantaneous failures caused by shocks Z can also affect the CBM policy. Thus, competing failure models are employed for assessing the asset reliability. The cost structure in the CBM model includes inventory cost rate C_i, downtime cost rate C_d, and maintenance cost C_m. The maintenance cost C_m is related to the degradation threshold T_m, so the setting of the degradation threshold T_m impacts the economy of the CBM. Due to the incomplete maintenance effect and downtime caused by delayed spare part delivery, average availability $\bar{A}$ will be impacted. Considering availability and costs, a simulation algorithm is used to determine the optimal decision variables for the CBM model. The implementation flowchart of the DT-PdM is summarized as follows:

1) Modeling and calibration of the metamodel. Initially, the metamodel $\hat{g}(X)$ is constructed and trained using historical data distribution $f_X(X)$ and FEM-PoF simulators $g(X)$. If new operational scenarios arise for the PA, the metamodel can be updated with new training samples. To facilitate the metamodel updating, real-time communication between the PA and the VM should be maintained.
2) Simulation of the trained metamodel. The distribution $f_X(X)$ is estimated using

maximum likelihood estimation (MLE), and the metamodel is then employed to probabilistically generate the life distribution based on $f_X(X)$ and MCS.

3) Maintenance model optimization. As the PA operates, its reliability decreases due to normal degradation and intermittent faults. Normal degradation is assessed by the metamodel, while intermittent failures are modeled through statistic analysis. Using the reliability of the PA, CBM can be designed to decrease operational costs while increasing availability.

Given the inherent complexities of real industrial processes, certain assumptions should be made to simplify elements or processes that are less significant. Assumptions underlying the modeling and optimization of the DT-CBM are outlined as follows:

1) The behavior model of the PA is founded on a cyclic mission, with fatigue being its primary degradation mechanism. In addition to degradation mechanisms, shocks can occur due to environmental disturbances or faults. If the shock value exceeds the established threshold, maintenance actions should be initiated.
2) The geometric and material properties of the PA are considered deterministic. Measurement uncertainties are also not considered in the analysis.
3) Maintenance actions are perfect, i.e., each maintenance action restores the PA to an AGAN condition.
4) Maintenance costs are related to the rejuvenation level. As the rejuvenation level increases, the maintenance cost also rises.

6.2 Predictive Maintenance Model Formulation and Optimization

6.2.1 Reliability Prediction Considering Degradation and Shocks

Shocks in the monitoring parameter can signal potential faults in the PA, prompting maintenance actions. However, these shocks might also stem from environmental noise, leading to unnecessary use of maintenance resources. As a result, the shock threshold H plays a vital role in developing a cost-effective CBM strategy. It is assumed that the probability of shock occurrence follows a geometric distribution P_s, while the magnitude of shocks at time τ is modeled as independent and identically distributed (i.i.d.) random variables following a normal distribution $Z(\left\lfloor\frac{\tau}{\Delta\tau}\right\rfloor)\sim\mathcal{N}(\mu_Z,\sigma_Z^2)$, where $\Delta\tau$ is the inspection interval and $\lfloor\cdot\rfloor$ is the floor operator [55]. The expectation μ_Z represents the average magnitude of critical loads, while the standard deviation σ_Z indicates the extent of random variation in critical loads [58]. It can be assumed that the PA experiences disturbance when the critical load surpasses the threshold H. Consequently, the failure probability at time τ is given by:

$$Pr\left\{Z\left(\left\lfloor\frac{\tau}{\Delta\tau}\right\rfloor\right)>H\right\}=1-\Phi(\frac{H-\mu_Z}{\sigma_Z}) \tag{6.1}$$

where $\Phi(\cdot)$ is the CDF of the standard normally distributed variable.

By modeling degradation using the metamodel and representing shocks through a compound geometric distribution, the reliability of the PA can be estimated. Failures in the PA can result from either the degradation process or the occurrence of shocks. Reliability is defined as the likelihood that the PA remains operational under shocks up to time τ (i.e., $Z(1)<H,\ldots,Z\left(\left\lfloor\frac{\tau}{\Delta\tau}\right\rfloor\right)<H$ and that the degradation level at τ is less

than the threshold, i.e., $Y(\tau) < T_m$). The reliability of the LRU at time τ can be derived as

$$\begin{aligned} R(\tau) &= Pr\left\{Y(\tau) < T_m, Z(1) < H, \dots, Z\left(\left\lfloor\frac{\tau}{\Delta\tau}\right\rfloor\right) < H\right\} \\ &= Pr\left\{Y(\tau) < T_m\right\} \cdot \prod_{i=1}^{\left\lfloor\frac{\tau}{\Delta\tau}\right\rfloor} Pr\left\{Z(i) < H\right\} \\ &= \int_0^{T_m} f_Y(Y(\tau))dY(\tau) \cdot \left[\Phi\left(\frac{H-\mu_Z}{\sigma_Z}\right)\right]^{\left\lfloor\frac{\tau}{\Delta\tau}\right\rfloor} \\ &= Pr\left\{\tau < \tau_f\right\} \end{aligned} \tag{6.2}$$

where $\tau_f = \inf\{\tau: (Y(\tau) \geq T_m) \cup (Z(\left\lfloor\frac{\tau}{\Delta\tau}\right\rfloor) \geq H)\}$ is the lifetime [30].

Proposition 6.1 outlines the relationship between the reliability $R(\tau)$, the expected number of failures E[Λ] over a fixed duration, and the degradation threshold T_m.

Proposition 6.1. *(1) For a fixed τ and H, the reliability $R(\tau)$ improves as the degradation threshold T_m increases. (2) For a fixed H, the expected number of failures $E[\Lambda]$ over a specified duration δ_{lc} decreases as the degradation threshold T_m rises.*

The proof of Proposition 6.1 is presented in Appendix A.6.1. Using Eq. (6.2), the PDF of the lifetime, $f_{\tau_f}(\tau)$, can be represented as follows.

$$f_{\tau_f}(\tau) = \lim_{\Delta\tau \to 0^+} \frac{Pr\left\{\tau \leq \tau_f < \tau + \Delta\tau\right\}}{\Delta\tau} = -\frac{dR(\tau)}{d\tau} \tag{6.3}$$

Thus, the mean life can be obtained as

$$E\left[\tau_f\right] = \int_0^{\infty} \tau f_{\tau_f}(\tau)d\tau = \int_0^{\infty} R(\tau)d\tau \tag{6.4}$$

Based on the reliability $R(\tau)$ and the failure rate $f_{\tau_f}(\tau)$, Remark 5.1 can be stated as follows.

Remark 1. *If the degradation level at time t $Y(t) < T_m$ and $Z(1) < H, \dots, Z\left(\left\lfloor\frac{\tau}{\Delta\tau}\right\rfloor\right) < H$, the reliability and mean residual life at time t can be obtained as*

$$R(\tau; t) = \frac{R(t+\tau)}{R(t)}$$

$$E[\tau_f; t] = \int_0^{\infty} \int_0^{T_m} R(\tau; t) f_Y(Y(t)) dY(t) d\tau$$
$$= \int_0^{\infty} \int_0^{T_m} e^{-\int_t^{t+\tau} \lambda(u) du} f_Y(Y(t)) dY(t) d\tau$$

where $f_Y(Y(t))$ is the PDF of degradation level by time t and $\lambda(t)$ is the failure rate.

6.2.2 Maintainability and Availability Modeling

To account for the impact of incomplete maintenance, the maintenance duration δ_m is modeled using an exponential distribution with a rate parameter λ_m. Unlike reliability $R(\tau)$, maintainability refers to the likelihood that the PA is successfully restored through maintenance activities [59]. The maintainability function, representing the probability of successfully restoring the PA within the interval $(0, \tau]$, is derived as shown in Eq. (6.5).

$$M(\tau, \lambda_m) = 1 - e^{-\lambda_m \tau} \tag{6.5}$$

Therefore, the maintenance PDF, which represents the probability of successfully restoring the PA within a unit of time at any given maintenance time τ, is expressed in Eq. (6.6).

$$m(\tau, \lambda_m) = \frac{dM(\tau, \lambda_m)}{d\tau} = \lambda_m e^{-\lambda_m \tau} \tag{6.6}$$

Once reliability and maintainability have been modeled, availability can be evaluated, incorporating both the failure rate and maintenance rate over a renewal cycle. Instantaneous point availability, $A(t)$, is defined as the probability that the PA is operational at time t, expressed as $A(t) = \Pr[W(t) = 1]$, where $W(t)$ is the functional indicator. Specifically, $W(t) = 1$ if the PA is operational, and $W(t) = 0$ otherwise [186]. At time t, the PA may remain operational under two conditions: either

it has not failed since the beginning, or the last renewal occurred at time u, and the PA has continued to operate since that point [55]. Hence:

$$A(t) = R(t) + \int_0^t R(t-u) r_e(u) du \tag{6.7}$$

where $r_e(t)$ is the probability that a renewal occurs in the interval $[t, t+\Delta t]$.

By applying Laplace transforms and renewal theory, instantaneous point availability can be represented in the Laplace domain, as outlined in Eq. (6.8). The detailed derivation process for instantaneous point availability $A(t)$ is provided in Appendix A.6.2.

$$A^*(s) = R^*(s)[1 + r_e^*(s)] = \frac{1-\delta_u^*(s)}{s[1-\delta_u^*(s)\delta_d^*(s)]} \tag{6.8}$$

where $\delta_u(t)$ is the PDF of uptime, $\delta_d(t)$ is the PDF of downtime, $A^*(s)$, $R^*(s)$, $r_e^*(s)$, $\delta_u^*(s)$, and $\delta_d^*(s)$ denote the Laplace transform of $A(t)$, $R(t)$, $r_e(t)$, $\delta_u(t)$, and $\delta_d(t)$, respectively, and s is the complex number.

Thus, the infinite-horizon and finite-horizon availability can be obtained with Eqs. (6.9) and (6.10), respectively.

$$A(\infty) = \lim_{t\to\infty} A(t) = \lim_{s\to 0} sA(s) = \frac{E[\delta_u]}{E[\delta_u]+E[\delta_d]} \tag{6.9}$$

$$\bar{A}(\delta_{lc}) = \frac{1}{\delta_{lc}} \int_0^{\delta_{lc}} A(u) du \tag{6.10}$$

where δ_{lc} is the finite asset lifecycle.

6.2.3 Cost Modeling

In addition to reliability, maintainability, and availability, cost plays a critical role in CBM planning. This study considers costs related to spare part inventory, downtime, and repairs. To ensure the economical operation of the PA, minimizing total costs is essential. The lifecycle of the PA is represented as δ_{lc}, after which the PA will be

upgraded to newer technology when $t > \delta_{\mathrm{lc}}$. Given the extended lifespan of the PA, δ_{lc}, it becomes difficult to account for all maintenance activities comprehensively. To address this, the lifespan δ_{lc} is divided into a finite number of renewal cycles, and an asymptotic evaluation of the cost criterion, referred to as the expected cost rate, is employed. This evaluation is based on the renewal-reward theorem described by Barlow and Proshan. According to this theorem, the expected long-term cost rate, C_{cr}, can be expressed as follows.

$$C_{cr} = \lim_{\delta_{lc}\to\infty} \frac{C(\delta_{lc})}{\delta_{lc}} = \frac{E[C(\delta_{rc})]}{E[\delta_{rc}]} \tag{6.11}$$

where $C(\delta_{\mathrm{lc}})$ is the accumulative cost until asset lifecycle δ_{lc}, $C(\delta_{\mathrm{rc}})$ is the cost in a renewal cycle, and δ_{rc} is the length of a renewal cycle. As a result, the study's time frame is simplified from the long-term lifespan to a single expected renewal cycle.

The lifecycle of the PA, denoted as δ_{rc}, is defined as the duration from its installation to its replacement. Based on this definition, the expected total cost over the renewal cycle δ_{rc} is the combined sum of the expected inventory cost $C_i\mathrm{E}[\delta_i]$, downtime cost $C_d\mathrm{E}[\delta_d]$, and repair cost C_m. In other words:

$$E[C(\delta_{rc})] = C_i E[\delta_i] + C_d E[\delta_d] + C_m \tag{6.12}$$

where C_i is the inventory cost rate, δ_i is the inventory reservation duration, C_d is the downtime cost rate, δ_d is the downtime duration, and C_m is the repair activity cost.

To reduce downtime, maintenance resources such as tools and crews must be prepared in advance. Conversely, to lower inventory costs ($C_i\mathrm{E}[\delta_i]$), spare parts are ordered only when required. It is assumed that a spare part is pre-ordered at τ_p and arrives precisely at $\tau_p + \delta_l$, where the lead time δ_l for an order is considered to be equal

to or greater than $\Delta\tau$. The inventory period can be categorized into two scenarios: (1) $\delta_i = 0$, when $\tau_f < \tau_p + \delta_l$, and (2) $\delta_i = \tau_f - \tau_p - \delta_l$, when $\tau_f \geq \tau_p + \delta_l$. Based on these conditions, the expected inventory cost ($C_i \mathrm{E}[\delta_i]$) over a renewal cycle is calculated as follows.

$$\begin{aligned} C_i E[\delta_i] &= C_i E\left[\tau_f - \tau_p - \delta_l\right] \\ &= C_i \cdot Pr\left\{\tau_f \geq \tau_p + \delta_l\right\} \cdot \left(\tau_f - \tau_p - \delta_l\right) \\ &= C_i R\left(\tau_p + \delta_l\right)\left(\tau_f - \tau_p - \delta_l\right) \end{aligned} \tag{6.13}$$

where τ_p is the spare-part order time. Downtime occurs when a spare part is unavailable, preventing maintenance from being carried out. Therefore, the timing of spare-part orders must be carefully planned based on the life distribution.

The downtime period can be divided into two scenarios: (1) $\delta_d = \tau_p + \delta_l - \tau_f + \delta_m$, when $\tau_p + \delta_l \geq \tau_f$, and (2) $\delta_d = \delta_m$, when $\tau_p + \delta_l < \tau_f$. Consequently, the expected downtime cost during a renewal cycle is expressed as follows.

$$\begin{aligned} C_d E[\delta_d] &= C_d \left(1 - R\left(\tau_p + \delta_l\right)\right)\left(\tau_p + \delta_l - \int_0^\infty u f_{\tau_f}(u)du + \frac{1}{\lambda_m}\right) \\ &+ C_d R\left(\tau_p + \delta_l\right)\frac{1}{\lambda_m} \end{aligned} \tag{6.14}$$

Maintenance costs (C_m) are expected to rise as the degradation thresholds (T_m) increase, as demonstrated in Eq. (5.15) [30][187][188].

$$C_m = C_{fix} + C_{imp} T_m \tag{6.15}$$

where C_{fix} is the setup cost for maintenance and C_{imp} is a proportional constant. By Eq. (6.16), the expected total cost during the lifecycle can be obtained as

$$\begin{aligned} E[C(\delta_{rc})] &= C_i E[\delta_i] + C_d E[\delta_d] + C_m \\ &= C_i R\left(\tau_p + \delta_l\right)\left(\tau_f - \tau_p - \delta_l\right) + C_d \left(1 - R\left(\tau_p + \delta_l\right)\right) \\ &\left(\tau_p + \delta_l - \int_0^\infty u f_{\tau_f}(u)du + \frac{1}{\lambda_m}\right) + C_d R\left(\tau_p + \delta_l\right)\frac{1}{\lambda_m} + C_{fix} + C_{imp} T_m \end{aligned} \tag{6.16}$$

Let random variable δ_{rc} denote the length of a renewal cycle. In every renewal cycle, one of the following two exclusive scenarios occurs [113].

1) S_1: $\tau_p + \delta_l > \tau_f$: An asset fails before spare-part order time $\tau_p + \delta_l$, $\delta_{\mathrm{rc}} = \tau_p + \delta_l + \delta_m$.
2) S_2: $\tau_p + \delta_l \leq \tau_f$: An asset fails after spare-part order time $\tau_p + \delta_l$, $\delta_{\mathrm{rc}} = \tau_f + \delta_m$.

After each renewal, the renewal cycle can be formulated to be

$$\delta_{rc} = \left(\tau_p + \delta_l + \delta_m\right)\mathbb{I}(S_1) + (\tau_f + \delta_m)\mathbb{I}(S_2) \tag{6.17}$$

where $\mathbb{I}(\cdot)$ is the indicator function which equals one if the scenario is true and zero otherwise. Consequently, the expected length of a renewal cycle can be derived as follows.

$$\begin{aligned} E[\delta_{rc}] \quad &= \left(1 - R\left(\tau_p + \delta_l\right)\right)\left(\tau_p + \delta_l + \frac{1}{\lambda_m}\right) \\ &+R\left(\tau_p + \delta_l\right)\left(\int_0^\infty uf_{\tau_f}(u)du + \frac{1}{\lambda_m}\right) \end{aligned} \tag{6.18}$$

The expected total cost $E[C(\delta_{\mathrm{rc}})]$ during the renewal cycle and the expected length $E[\delta_{\mathrm{rc}}]$ are derived in Eqs. (5.16) and (5.18), respectively. Based on the renewal-reward theorem in Barlow and Proshan, the expected cost rate can be calculated in Eq. (6.19).

$$C_{cr} = \frac{\begin{array}{c} C_i R(\tau_p+\delta_l)(\tau_f-\tau_p-\delta_l)+C_d\left(1-R(\tau_p+\delta_l)\right)\left(\tau_p+\delta_l-\int_0^\infty uf_{\tau_f}(u)du+\frac{1}{\lambda_m}\right) \\ +C_d R(\tau_p+\delta_l)\frac{1}{\lambda_m}+C_{fix}+C_{imp}T_m \end{array}}{\left(1-R(\tau_p+\delta_l)\right)\left(\tau_p+\delta_l+\frac{1}{\lambda_m}\right)+R(\tau_p+\delta_l)\left(\int_0^\infty uf_{\tau_f}(u)du+\frac{1}{\lambda_m}\right)} \tag{6.19}$$

This chapter incorporates the inflation rate to adjust future costs – such as inventory, downtime, and maintenance costs-to reflect rising prices over time. The inflation rate η_I represents the expected percentage increase in costs due to economic factors. Future costs are adjusted using the following formula:

$$C(t) = C(0)(1 + \eta_I)^t \tag{6.20}$$

where η_I is the inflation rate and $C(0)$ is the initial cost [67].

Thus, the future values of the cost parameters are calculated using the following equations.

$$\begin{cases} C_i(t) = C_i(0)(1+\eta_I)^t, t \in \{1,2,\dots,\delta_{lc}\} \\ C_d(t) = C_d(0)(1+\eta_I)^t, t \in \{1,2,\dots,\delta_{lc}\} \\ C_m(t) = C_m(0)(1+\eta_I)^t, t \in \{1,2,\dots,\delta_{lc}\} \end{cases} \tag{6.21}$$

The inflation rate reflects economic growth and price increases, distinguishing it from the discount rate. Table 6.1 provides a clear comparison between the two to highlight their differences.

Table 6.1 Comparison between inflation rate and discount rate.

Aspect	Inflation rate η_I	Discount rate η_D
Definition	Represent the annual percentage increase in costs due to rising prices	Reflect the time value of money
Purpose	Adjust future costs to reflect expected price increases	Discount future costs to their current value
Formula	$C(t) = C(0)(1+\eta_I)^t$	$P_v = \frac{C(t)}{(1+\eta_D)^t}$
Application	Adjust inventory, downtime, and maintenance costs	Ensure accurate present value calculations for cost evaluation

6.2.4 Optimization Model Formulation

The performance of CBM for the PA is assessed using the average availability per unit of maintenance cost, as defined in Eq. (5.21).

$$\begin{aligned} E[AC] &= \frac{\bar{A}(\delta_{lc})}{C_{cr}\delta_{lc}} \\ \text{s.t.} &\begin{cases} 0 \le T_m \le T_m^u \\ 0 \le H \le H^u \\ \tau_p \ge 0 \end{cases} \end{aligned} \tag{6.22}$$

The average availability during long-term operation, $\hat{A}(\delta_{\mathrm{lc}})$, is calculated using Eq. (6.10). Constraints define bounds for the degradation threshold, disturbance threshold, and spare-part order time. To ensure safety, T_m^u and H^u are applied. A higher AC

indicates a more effective CBM policy, improving availability while reducing costs. The optimal values (T_m, τ_p, H), maximizing E[AC], are determined as $\left(T_m^*, \tau_p^*, H^*\right) = \text{argmax}_{T_m,\tau_p,H}\text{E[AC]}$. This policy adapts to changing operational scenarios. Historical data trains the degradation metamodel, updated with real-time data for new scenarios. The operator dynamically adjusts the degradation threshold $T_m^{(k)*}$, disturbance threshold $H^{(k)*}$, and spare-part order time $\tau_p^{(k)*}$ to optimize $\text{AC}^{(k)}$ for each stage $k = 0,1,\ldots,n$.

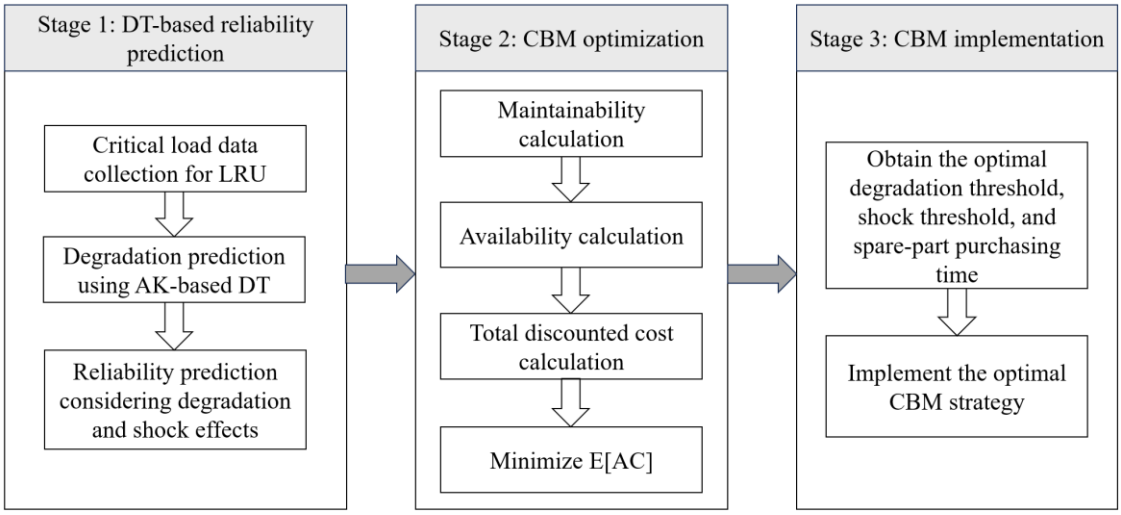


Fig. 6.2 The process of determining the optimal proactive maintenance policy with the DT.

Proposition 6.1 states that for given δ_{lc}, τ_p, and H, the expected failures E[Λ] decrease as the degradation threshold T_m increases, reducing failures but raising maintenance costs C_m. Thus, an optimal T_m^* balances failure and maintenance costs. For τ_p, earlier ordering reduces downtime costs but raises inventory costs, leading to an optimal τ_p^* that maximizes AC. Similarly, increasing H limits maintenance but keeps AC^* unchanged if H^* exceeds a certain limit. In such cases, the minimum H^* is chosen to maintain reliability and avoid intermittent faults.

The expression for AC is highly complex, requiring numerical methods instead of analytical solutions for estimation [66]. MCS is used to calculate AC, suitable for degradation processes lacking analytical lifespan distributions, as outlined in Algorithm 6.1. Fig. 6.2 shows the process for determining optimal T_m, H, and τ_p. In stage I, metamodel parameters are optimized to minimize prediction errors for degradation and lifespan estimation. In Stage II, the optimized metamodel evaluates reliability, costs, and availability under the CBM framework. Optimal thresholds and spare-part order time are identified via a numerical search within a predefined range using the enumeration algorithm.

Algorithm 6.1 MCS to obtain E[AC].

Input:
Inventory cost rate C_i, downtime cost rate C_d, maintenance cost C_m, episode number N_{sim}.
Output:
E[AC](T_m, H, τ_p).
1: Generate a uniformly distributed $T_m \in [0,1], H \in [H^l, H^u]$, and $\tau_p \in \left[0, \tau_p^u\right]$.
2: Calculate the asset reliability $R(\tau; T_m, H) = \int_0^{T_m} f_Y(Y(\tau))dY(\tau) \cdot \left[\Phi\left(\frac{H-\mu_Z}{\sigma_Z}\right)\right]^{\left\lfloor\frac{\tau}{\Delta\tau}\right\rfloor}$.
3: For iteration $i = 1,2, \dots, N_{sim}$ Do
4: $C_{sum}(T_m, H, \tau_p) = 0$, $\delta_d(T_m, H, \tau_p) = 0$, $EAC_t(T_m, H, \tau_p) = 0$.
5: Generate a sequence of uniformly distributed number $r \in [0,1]$.
6: If $r(\tau) > R(\tau + 1; T_m, H, \tau_p)/R(\tau; T_m, H, \tau_p)$ Do
7: If $\tau < \tau_p$ Do
8: Estimate cumulative costs due to downtime $C_{sum}(T_m, H, \tau_p) = C_{sum}(T_m, H, \tau_p) + \sum_{i=\tau}^{\tau_p} C_d(i) + C_{imp}(\tau)T_m$.
9: Estimate downtime duration $\delta_d(T_m, H, \tau_p) = \delta_d(T_m, H, \tau_p) + (\tau_p - \tau)$.
10: $\tau = \tau_p$.
11: Else
12: Estimate cumulative costs due to inventory management $C_{sum}\ (T_m, H, \tau_p)$= $C_{sum}\ (T_m, H, \tau_p)$+ $\sum_{i=\tau_p}^{\tau} C_i(i) + C_0(\tau)T_m$.
13: End if
14: End if
15: $AC_t(T_m, H, \tau_p) = EAC_t(T_m, H, \tau_p)$ +[$\tau - \delta_d(T_m, H, \tau_p)]/C_{sum}(T_m, H, \tau_p)$.
16: End while
17: $E[AC](T_m, H, \tau_p) = EAC_t(T_m, H, \tau_p)/N_{sim}$.

6.3 Case Study: Turbine Blisks

6.3.1 Case Descriptions

The numerical experiments focus on aeroengine blisks, using data from blisks, manufacturers, and airlines to develop a DT-CBM model. Operational data includes rotational speed ω, blisk-tip temperature T_{tip}, blisk-root temperature T_{root}, and shock amplitude Z. A shock event occurs when T_{tip} exceeds 1,600 K, requiring repairs. Over time, operational scenarios expand, classified into three stages: Stage 0 ([0, 11,000) cycles), Stage 1 ([11,000, 22,000) cycles), and Stage 2 ([22,000, 33,000) cycles).

Degradation predictions for the blisk across stages are detailed in Subsection 3.5. Along with reliability data, inputs include maintenance costs and thresholds from manufacturers, and lifecycle, inventory costs, downtime rates, and inspection intervals from airlines (Table 6.2). This information allows optimization and updates of degradation thresholds, shock thresholds, and spare-part procurement timing for different stages.

Table 6.2 Maintenance parameter values.

Parameter, unit	$\Delta\tau$, Cycle	T_m^u	H^u, K	δ_{lc}	$C_{\mathrm{imp}}(0)$, \$	$C_d(0)$, \$	η	λ_m	δ_l, Cycle	$C_i(0)$, \$	$C_{\mathrm{fix}}(0)$, \$
Value	22	1	1,650	33,000	10	12	0.01 %	0.5	22	6	5

6.3.2 Maintenance Optimization with Adaptive Kriging-Based Digital Twins

Using the AK-based DT in Subsection 3.4, LCF degradation is probabilistically simulated for adaptive reliability prediction. After predicting blisk reliability, CBM scheduling is enabled, allowing airlines to adjust T_m^*, H^*, and τ_p^* at each stage. Optimal values across stages are determined via an enumeration algorithm: T_m^* = {0.48, 0.49,

0.54}, H^*= {1,650, 1,650, 1,650}, and τ_p^*= {2,838, 2,838, 3,058}, with E[AC]*= {0.0010, 0.0013, 0.0025}. T_m^* and τ_p^* increase with advancing stages, while H^* remains constant, simplifying CBM implementation. These adjustments reflect the increasing degradation rate from Stage 0 to Stage 2, raising renewal frequency and lifecycle costs. To minimize renewals, T_m^* increases, extending τ_p^* to control inventory costs. H^* remains unaffected as shock amplitude statistics stay consistent. While higher T_m^* reduces renewals, it raises maintenance costs C_{cr}. Advancing stages reduce lifecycle δ_{lc}, decreasing overall costs $C_{\mathrm{cr}}\delta_{\mathrm{lc}}$ and increasing E[AC]*.

Fig. 6.3 evaluates the impact of T_m, H, τ_p, and operational stages on E[AC], showing unique optimal values for T_m^*, H^*, and τ_p^*. As scenarios progress, T_m^*, τ_p^*, and E[AC]* increase, while H^* remains constant. Figs. 6.3a and 6.3c show a non-monotonic, concave relationship between E[AC] and T_m and τ_p across stages, while E[AC] increases monotonically with advancing stages. Fig. 6.3b shows that higher H consistently raises E[AC] by reducing repairs, improving availability, and cutting costs, with E[AC] peaking when H^* reaches its upper limit H^u.

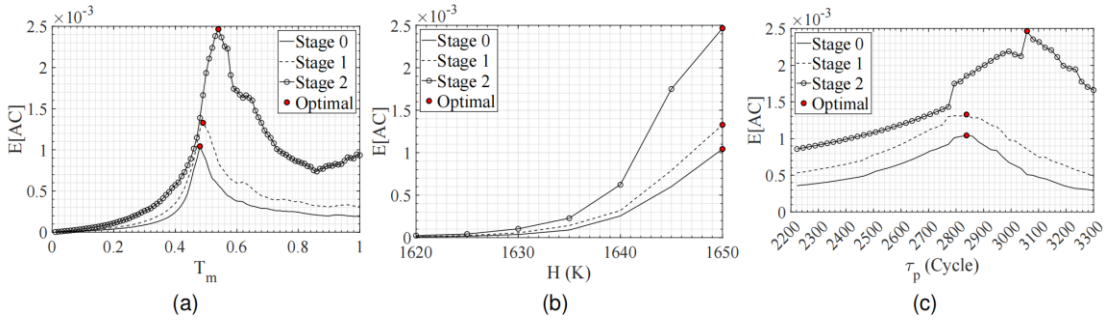


Fig. 6.3 Optimal maintenance policies. (a) E[AC] curves under varying T_m and operational stages. (b) E[AC] curves under varying H and operational stages. (c) E[AC] curves under varying τ_p and operational stages.

6.3.3 Sensitivity Analysis

The shock probability P_s increases from 0.001 to 0.005 in 0.002 increments, with optimal design variables and objective function values (OFVs) summarized in Table 6.3. As P_s rises, the optimal degradation threshold, spare-part purchasing time, and OFV decrease, while the shock threshold remains unchanged across stages. Higher P_s leads to more frequent maintenance, raising costs and lowering $\mathrm{E[AC]}^*$. The degradation threshold decreases, and spare-part purchasing time shortens to minimize downtime costs from shocks. To offset higher inventory costs from earlier purchases, the degradation threshold is further reduced. Table 6.3 shows similar trends when the mean μ_Z and variance σ_Z of shock amplitudes increase. Higher μ_Z and σ_Z raise maintenance demands, requiring earlier spare-part purchases and lower degradation thresholds to manage costs, ultimately reducing $\mathrm{E[AC]}^*$.

Table 6.3 Optimal design variables and OFVs across operational stages for various P_s, μ_Z, and σ_Z.

P_s	μ_Z	σ_Z	Stage 0				Stage 1				Stage 2			
			T_m^*	H*	T_p^*	E[AC]*	T_m^*	H*	T_p^*	E[AC]*	T_m^*	H*	T_p^*	E[AC]*
0.001	1,620	10	0.48	1,650	2,838	0.0010	0.49	1,650	2,838	0.0013	0.64	1,650	3,278	0.0025
0.003	—	—	0.48	1,650	2,838	0.0008	0.42	1,650	2,420	0.0010	0.57	1,650	3,168	0.0020
0.005	—	—	0.39	1,650	2,332	0.0006	0.42	1,650	2,376	0.0009	0.55	1,650	3,102	0.0019
0.001	1,620	10	0.48	1,650	2,838	0.0010	0.49	1,650	2,838	0.0013	0.55	1,650	3,102	0.0025
—	1,625	—	0.42	1,650	2,332	0.0006	0.42	1,650	2,420	0.0009	0.55	1,650	3,102	0.0019
—	1,630	—	0.41	1,650	2,310	0.0003	0.39	1,650	2,222	0.0004	0.45	1,650	2,508	0.0007
0.001	1,620	10	0.48	1,650	2,838	0.0010	0.49	1,650	2,838	0.0013	0.54	1,650	3,058	0.0025
—	—	15	0.41	1,650	2,354	0.0003	0.39	1,650	2,222	0.0004	0.45	1,650	2,508	0.0007
—	—	20	0.41	1,650	2,354	0.0001	0.39	1.650	2,222	0.0002	0.45	1,650	2,222	0.0003

CBM integrates shock- and availability-related parameters, including maintenance rate λ_m and lead time δ_l, to optimize strategies. A sensitivity analysis (Table 6.4) shows that as λ_m decreases, T_m^* and τ_p^* increase, $\mathrm{E[AC]}^*$ decreases, and H^* remains constant across stages. Lower λ_m leads to longer maintenance downtime, reducing $\mathrm{E[AC]}^*$ and favoring fewer maintenance actions. To limit maintenance, T_m^* increases, requiring

longer τ_p^* to manage inventory costs. Table 6.4 also shows that δ_l primarily impacts τ_p^*, with a negative effect, as δ_l and τ_p^* together balance inventory and downtime costs.

Table 6.4 Optimal design variables and OFVs for varying λ_m and δ_l.

λ_m	Stage	$\delta_l = 1$				$\delta_l = 5$				$\delta_l = 9$			
		T_m^*	H*	τ_p^*	E[AC]*	T_m^*	H*	τ_p^*	E[AC]*	T_m^*	H*	τ_p^*	E[AC]*
1/8	0	).69	650	,092	0006	).69	,650	,004	)006	).69	1,650	,916	).0006
	1	0.72	1,650	4,180	0.0008	0.72	1,650	4,092	0.0008	0.72	1,650	4,004	0.0008
	2	0.74	1,650	4,224	0.0016	0.74	1,650	4,136	0.0016	0.74	1,650	4,048	0.0016
1/5	0	0.51	1,650	3,058	0.0007	0.51	1.650	2,970	0.0007	0.51	1,650	2,882	0.0007
	1	0.57	1,650	3,278	0.0010	0.57	1,650	3,190	0.0010	0.57	1,650	3,102	0.0010
	2	0.60	1,650	3,388	0.0021	0.60	1,650	3,300	0.0021	0.60	1,650	3,212	0.0021
1/2	0	0.48	1,650	2,838	0.0010	0.48	1,650	2,750	0.0010	0.48	1,650	2,662	0.0010
	1	0.49	1,650	2,838	0.0013	0.49	1,650	2,750	0.0013	0.49	1,650	2,662	0.0013
	2	0.54	1,650	3,058	0.0025	0.54	1.650	2,970	0.0025	0.54	1,650	2,882	0.0025

Table 6.5 lists T_m^*, H^*, τ_p^*, and E[AC]* for varying cost parameters (C_i, C_d, C_{fix}, and C_{imp}) across stages. Increasing any cost parameter reduces E[AC]*, while T_m^* and τ_p^* rise and H^* remains constant. When C_i increases, spare-part purchases are delayed to lower inventory costs, requiring a higher T_m^* to offset downtime costs. With higher C_d, reducing renewals by raising T_m^* and extending τ_p^* becomes optimal. For C_{fix} and C_{imp}, linked to maintenance costs, fewer maintenance actions are required, achieved by raising T_m^* to lower renewal frequency, while extending τ_p^* optimizes inventory costs.

Table 6.5 Optimal design variables and OFVs across stages for various C_i, C_d, C_{fix}, and C_{imp}.

C_i	C_d	C_{fix}	C_{imp}	Stage 0				Stage 1				Stage 2			
				T_m^*	H*	τ_p^*	E[AC]*	T_m^*	H*	τ_p^*	E[AC]*	T_m^*	H*	τ_p^*	E[AC]*
6	12	5	10	0.48	1,650	2,838	0.0010	0.49	1,650	2,838	0.0013	0.54	1,650	3,058	0.0025
26	—	—	—	0.51	1,650	3,124	0.0008	0.51	1,650	2,728	0.0008	0.52	1,650	3,080	0.0013
46	—	—	—	0.51	1,650	3,124	0.0007	0.50	1,650	3,146	0.0007	0.52	1,650	3,234	0.0011
6	8	5	10	0.48	1,650	2,860	0.0014	0.48	1,650	2,816	0.0017	0.54	1,650	3,058	0.0030

—	48	—	—	0.53	1,650	3,014	0.0004	0.61	1,650	3,080	0.0006	0.64	1,650	3,256	0.0012
—	88	—	—	0.53	1,650	3,014	0.0002	0.61	1,650	3,080	0.0004	0.64	1,650	3,190	0.0008
6	12	5	10	0.48	1,650	3,014	0.0014	0.48	1,650	2,816	0.0017	0.54	1,650	3,058	0.0030
—	—	15	—	0.55	1,650	3,278	0.0012	0.48	1,650	2,816	0.0015	0.54	1,650	3,058	0.0027
—	—	25	—	0.55	1,650	3,278	0.0011	0.57	1,650	3,300	0.0013	0.54	1,650	3,058	0.0025
6	12	5	10	0.48	1,650	2,860	0.0014	0.48	1,650	2,816	0.0017	0.54	1,650	3,058	0.0030
—	—	—	50	0.48	1,650	2,860	0.0010	0.48	1,650	2,816	0.0013	0.53	1,650	3,058	0.0024
—	—	—	90	0.55	1,650	3,278	0.0008	0.48	1,650	2,816	0.0011	0.53	1,650	3,058	0.0021

6.4 Summary

In this chapter, we investigate a dynamic PdM optimization model integrated with DT technologies under varying operational conditions. To achieve the aim, a metamodel-based DT is used to accurately and efficiently predict life reliability, while dynamically updating the metamodel with learning function under new operational conditions. Using the reliability prediction from the metamodel, we can design the optimal degradation threshold, shock threshold, and inventory purchasing time to increase availability while decreasing operational costs. Experiments on turbine blisks using an AK-based DT demonstrate the effectiveness of the proposed method. Numerical analyses highlight the accuracy and efficiency of the AK model, with results indicating that operational conditions significantly influence the component life distribution. The proposed DT-PdM method is broadly applicable to components subjected to cumulative fatigue damage. The developed method provides a roadmap for DT-PdM design, including operational condition analysis, DT-based reliability estimation, and PdM planning.

Chapter 7

Digital Twin-Based Opportunistic Predictive Maintenance Scheduling for Multi-Unit Systems with Economic Dependencies

Chapters 5 and 6 focus on predictive maintenance (PdM) scheduling for single-unit components; however, industrial systems often comprise multiple interconnected units, where interdependencies introduce additional complexities. To address this limitation, this chapter proposes a digital twin-based opportunistic predictive maintenance (DT-OPdM) model, offering a practical framework for optimizing maintenance decisions in multi-unit systems by explicitly accounting for economic dependencies. A key innovation of the OPdM model is the treatment of the opportunistic degradation threshold as a design variable, enabling more effective and flexible maintenance planning. The model addresses challenges such as: (1) Non-self-announcing faults, which are failures undetectable by sensors. (2) Incomplete maintenance cost models, which often overlook setup costs (e.g., rental expenses for crews and equipment). DTs are integrated into the model to improve failure modeling and maintenance cost estimation. By considering multiple failure modes, undetectable faults, and economic dependencies, the DT-enabled OPdM model minimizes the expected cost rate by optimizing four decision variables: the opportunistic degradation threshold, the shock spatial threshold, the shock temporal threshold, and the inspection interval. The implementation of the proposed model depends on the system configuration, with

comparative analyses performed on series and parallel systems. Due to the complexity of deriving analytical solutions, Monte Carlo simulation (MCS) is employed to determine the optimal decision variables. To validate the proposed method, a case study is conducted using gas-insulated switchgear (GIS) in both series and parallel systems. Validation includes sensitivity analyses and policy comparisons to explore the impact of reliability parameters, opportunistic policies, unit quantity, and cost parameters on the cost rates and decision variables in different system configurations.

7.1 Digital Twin-Based Opportunistic Predictive Maintenance Scheduling Framework for Multi-Unit Systems

This chapter optimizes an OPdM model for multi-unit systems with competing failure mechanisms by leveraging positive economic dependencies (PEDs) at the system level. The framework of the proposed OM is shown in Fig. 7.1. In a system of n binary units, each unit i is functional or failed, with its degradation state represented by a scalar random variable $X_i(t)$ and a shock process $S_i(t)$. A unit fails if its degradation exceeds the critical threshold T_d, its shock amplitude surpasses the spatial threshold H, or the time between shocks is below the temporal threshold δ. The system is coherent, with reliability improving as individual unit reliability increases. To minimize downtime and ensure restoration, periodic inspections are required to detect faults undetectable by online monitoring. Maintenance actions restore units to an AGAN condition, and servicing multiple units simultaneously reduces setup costs, reflecting PEDs.

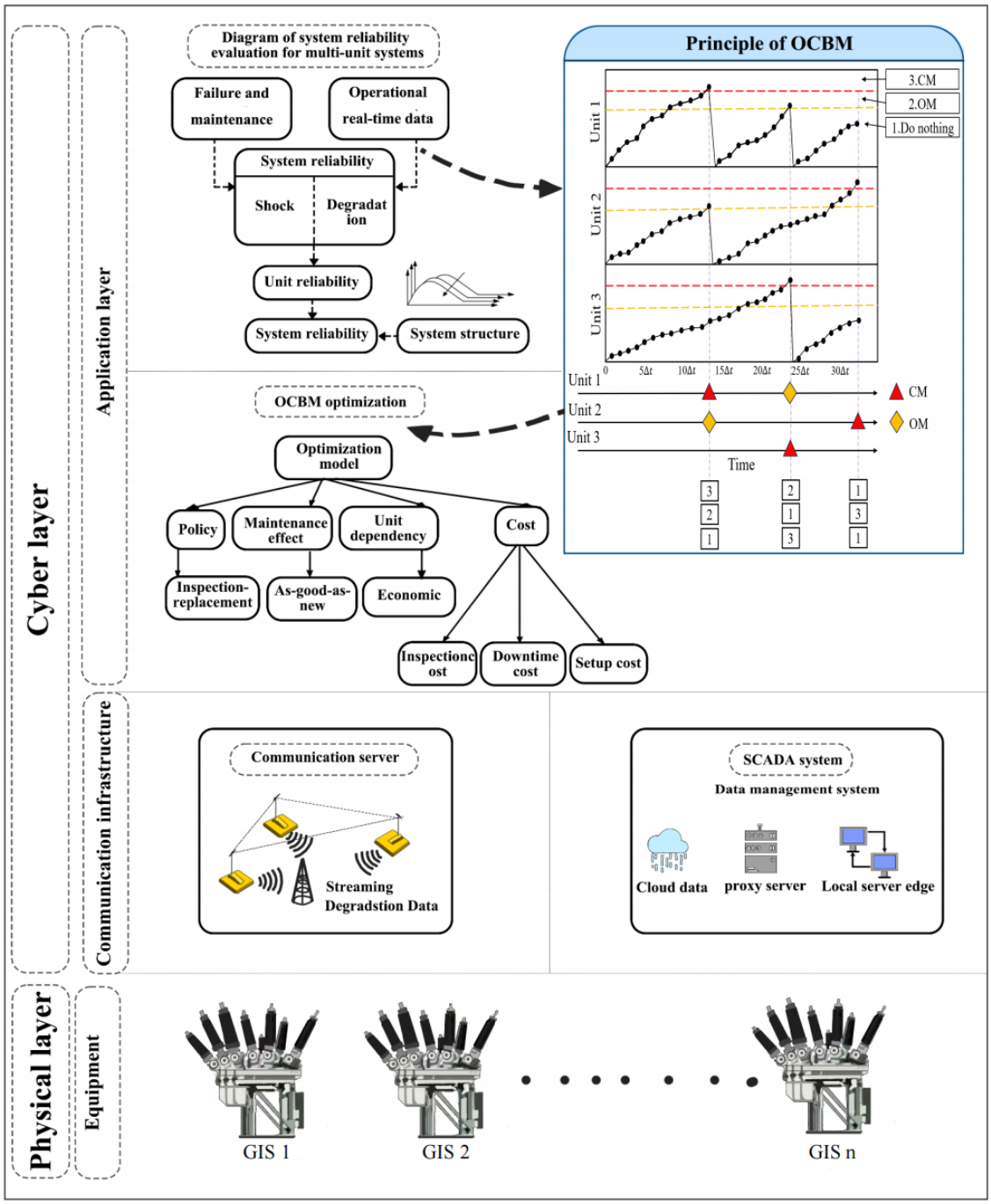


Fig. 7.1 The diagram of the OPdM for multi-unit systems incorporating degradation and shock models.

General assumptions are outlined below [58][113].

(1) The components of the system are governed by competing failure models. It is assumed that units remain non-degrading when not in operation and behave stochastically independently. Furthermore, failures are not self-announcing.

(2) The degradation level of a functional unit is assumed to be observable only during periodic inspections conducted at $t_k = k\Delta t$. These inspections are assumed to be flawless, instantaneous, and non-intrusive, providing an exact measurement of the degradation level.

(3) A minimum of one maintenance team is always on hand to carry out inspections and necessary maintenance activities. Maintenance is restricted to scheduled inspection times, and its duration is considered insignificant compared to the system's expected lifespan. Following maintenance, the unit is restored to an AGAN condition.

(4) Setup costs can be distributed across multiple units within the same system, whereas spare part costs scale directly with the quantity of items being maintained.

7.2 System Reliability Modeling Considering Opportunistic Grouping

7.2.1 Unit Reliability Prediction Considering Degradation and Shocks

Units can fail through multiple independent failure modes, and the competing failure model assumes: (i) failure modes are independent, (ii) a unit fails if any mode reaches its failure state, and (iii) each mode has a unique life distribution [55]. Failures may result from gradual degradation or random shocks.

Degradation, caused by mechanisms like wear and fatigue, is modeled using stochastic processes such as the Wiener, inverse Gaussian, or gamma processes. The Wiener process suits non-monotonic degradation, while monotonic degradation is modeled using inverse Gaussian or gamma processes. This study uses a univariate

gamma process, with non-negative, independent increments following a gamma distribution, to represent monotonic deterioration (e.g., wear, fatigue, corrosion) [59][189]. Let $X_i(t)$, where $i = 1,2, \dots, n$, denote the degradation level of unit i at age t, expressed as:

$$X_i(t_2) = X_i(t_1) + \Delta X_i(t_2 - t_1) \tag{7.1}$$

where $\Delta X_i(t_2 - t_1) = X_i(t_2) - X_i(t_1)$, $t_1 < t_2$. The degradation process $\{X_i(t), t \geq 0\}$ is defined as a gamma process if it meets the following criteria:

(1) $X_i(0) = 0$ with probability one.

(2) $X_i(t)$ is monotonic and has independent increments, i.e., for $t_1 < t_2 < t_3 < t_4$, $\Delta X_i(t_4 - t_3)$ is independent of $\Delta X_i(t_2 - t_1)$.

(3) $\Delta X_i(t_2 - t_1)$ is subject to the gamma distribution for $t_1 < t_2$.

The degradation increment ΔX_i is governed by a gamma distribution characterized by a shape parameter α_i and a scale parameter β_i, where $\alpha_i > 0$ and $\beta_i > 0$. The expected degradation level is reflected by $\frac{\alpha_i}{\beta_i}$, while the variance is given by $\frac{\alpha_i}{\beta_i^2}$. The marginal density function of $X_i(t)$ is defined as:

$$g_i(x \mid \alpha_i t, \beta_i) = Pr\{X_i(t) = x\} = \frac{\beta_i^{\alpha_i t} x^{\alpha_i t - 1} e^{-\beta_i x}}{\Gamma(\alpha_i t)} \tag{7.2}$$

where $\Gamma(\alpha_i t) = \int_0^\infty u^{\alpha_i t - 1} e^{-u} du$ denotes the gamma function. Let T_d denote the constant representing the degradation threshold of the gamma process. We define the first passage time to failure as $t_{f,i}^d = \inf\{t: X_i(t) \geq T_d\}$, and the CDF of $t_{f,i}^d$ can be expressed as follows:

$$
\begin{aligned}
F_i^d(t \mid \alpha_i, \beta_i, T_d) &= Pr\left\{t_{f,i}^d \le t\right\} = Pr\left\{X_i(t) \ge T_d\right\} \\
&= \int_{T_d}^{\infty} \frac{\beta_i^{\alpha_i t} x^{\alpha_i t-1} e^{-\beta_i x}}{\Gamma(\alpha_i t)} dx = \int_{\beta_i T_d}^{\infty} \frac{x^{\alpha_i t-1} e^{-x}}{\Gamma(\alpha_i t)} dx \\
&= \frac{\Gamma(\alpha_i t, \beta_i T_d)}{\Gamma(\alpha_i t)}
\end{aligned} \tag{7.3}
$$

where $\Gamma(\alpha_i t, \beta_i T_d)$ is the upper incomplete gamma function. The PDF of the first passage time to failure can thus be determined as follows:

$$
\begin{aligned}
f_i^d(t \mid \alpha_i, \beta_i, T_d) &= \frac{\partial F_i^d(t|\alpha_i, \beta_i, T_d)}{\partial t} \\
&= \frac{\alpha_i}{\Gamma(\alpha_i t)} \int_{\beta_i T_d}^{\infty} [log\,(x) - \phi(\alpha_i t)] x^{\alpha_i t-1} e^{-x} dx
\end{aligned} \tag{7.4}
$$

where $\phi(\alpha_i t) = \frac{\Gamma'(\alpha_i t)}{\Gamma(\alpha_i t)}$ is the digamma function.

Given the monotonic nature of the gamma process, the PDF of the first passage time to failure can be expressed as $f_i^d(t|\alpha_i, \beta_i, T_d) = \Pr\{X_i(t) = T_d\} = g_i(T_d|\alpha_i t, \beta_i)$. Consequently, we have $g_i(x|\alpha_i t, \beta_i) = \frac{\partial F_i^d(t|\alpha_i, \beta_i, x)}{\partial t}$. The parameters α_i and β_i of the gamma process can be estimated using historical data. This study assumes that the parameters of the gamma process for each unit are predetermined.

Random shocks result from intermittent variations. For instance, in the traction power supply system (TPSS) of high-speed railways (HSR), these shocks are caused by traction loads. Shocks can be characterized by two dimensions: temporal and spatial. The temporal dimension represents the stochastic timing of shock occurrences, while the spatial dimension reflects the random magnitude of the shocks. To capture the randomness in both dimensions, a Compound Homogeneous Poisson Process is employed. The occurrence of random shocks in the ith unit is modeled using an HPP, $\{S_i(t), t \ge 0\}, i = 1,2, \dots, n$, where the time interval between consecutive shocks follows an exponential distribution. The probability of j shocks occurring within the

interval $[0, t]$ is given by $\Pr\{S_i(t) = j\} = \frac{e^{-\lambda_i t}(\lambda_i t)^j}{j!}, j = 0,1, \ldots$, where λ_i is the mean arrival rate of shock arrivals. As λ_i increases, the frequency of shocks per unit time also grows. The magnitude of the jth shock in the ith unit is assumed to be an independent and identically distributed (i.i.d.) random variable following a normal distribution: $Y_{i,j} \sim \mathcal{N}\left(\mu_{Y_i}, \sigma_{Y_i}^2\right)$ [189]. The parameters of the Compound Homogeneous Poisson Process can be estimated using the MLE method as $\widehat{\Theta}_i = \left(\hat{\lambda}_i, \hat{\mu}_{Y_i}, \hat{\sigma}_{Y_i}\right) = \operatorname{argmax}_{\lambda_i, \mu_{Y_i}, \sigma_{Y_i}} \{\mathcal{L}(\lambda_i, \mu_{Y_i}, \sigma_{Y_i} | S_i, Y_i)\}$. With the estimated parameters $\widehat{\Theta}$, shocks can be predicted based on the load data S_i and Y_i. After estimating the parameters, a goodness-of-fit test is performed to validate their accuracy. The Kolmogorov-Smirnov test is a widely used method for evaluating the acceptability of the estimated parameters [190].

To account for the temporal and spatial randomness of shocks, the δ-shock and extreme shock models are introduced to describe shock-induced failures. In the δ-shock model, failure occurs when the time interval between the jth and $(j+1)$th shocks, denoted as $\Delta t_{S_{i,j}}$, is shorter than a predefined temporal threshold δ. Since $\Delta t_{S_{i,j}}$ follows an exponential distribution with rate parameter λ_i, the probability of failure is given by $\Pr\left\{\Delta t_{S_{i,j}}\right\} \leq \delta = 1 - e^{-\lambda_i \delta}$. In the extreme shock model, failure is assumed to occur if the magnitude of a shock exceeds a critical threshold H. Accordingly, the failure probability is expressed as $\Pr\left\{Y_{i,j} \geq H\right\} = 1 - \Phi(\frac{H - \mu_{Y_i}}{\sigma_{Y_i}})$, where $\Phi(\cdot)$ represents the CDF of the standard normal distribution. By combining the δ-shock and extreme models, the survival probability of a unit after the jth shock within the interval $[0, t]$ can be formulated as follows.

$$\bar{F}_i^s(t \mid S_i(t) = m) = Pr\left\{\cap_{j=1}^{m} Y_{i,j} < H\right\}.$$
$$Pr\left\{\cap_{j=1}^{m} \Delta t_{S_{i,j}} > \delta\right\} = \left[\Phi\left(\frac{H-\mu_{Y_i}}{\sigma_{Y_i}}\right)\right]^m \cdot e^{-m\lambda_i\delta} \tag{7.5}$$

Failure of the ith unit can occur due to either of the competing failure processes: degradation or shocks. The reliability of the ith unit at time t, denoted as $\bar{F}_i(t)$, is defined as the probability that the degradation level $X_i(t)$ remains below the threshold T_d and the unit successfully withstands all shocks $S_i(t)$ up to time t [58].

Let $t_{f,i}^d$ and $t_{f,i}^s$ represent the failure times of the ith unit caused by degradation and shocks, respectively. The overall lifetime of the unit is then given by:

$$\begin{aligned}\bar{F}_i(t) &= \sum_{m=0}^{\infty} Pr\left\{X_i(t) < T_d, \cap_{j=0}^{m} Y_{i,j} < H, \cap_{j=0}^{m} \Delta t_{S_{i,j}} > \delta, S_i(t) = m\right\} \\ &= \sum_{m=0}^{\infty} Pr\{X_i(t) < T_d\} \cdot Pr\left\{\cap_{j=0}^{m} Y_{i,j} < H, \cap_{j=0}^{m} \Delta t_{S_{i,j}} > \delta \mid S_i(t) = m\right\} \\ &\cdot Pr\{S_i(t) = m\} \\ &= \frac{\Gamma(\alpha_i t) - \Gamma(\alpha_i t, \beta_i T_d)}{\Gamma(\alpha_i t)} e^{-\lambda_i t} + \sum_{m=1}^{\infty} \frac{\Gamma(\alpha_i t) - \Gamma(\alpha_i t, \beta_i T_d)}{\Gamma(\alpha_i t)} \\ &\cdot \left[\Phi\left(\frac{H-\mu_{Y_i}}{\sigma_{Y_i}}\right)\right]^m \cdot e^{-m\lambda_i\delta} \cdot \frac{e^{-\lambda_i t}(\lambda_i t)^m}{m!}\end{aligned} \tag{7.6}$$

Proposition 7.1. *Given T_d and t, the unit reliability $\bar{F}_i(t)$ increases an H increases or δ decreases.*

A proof of Proposition 7.1 is provided in Appendix A.7.1. Let $f_i(t)$ represent the failure PDF and $h_i(t)$ denote the hazard rate of the ith unit at time t. These can be estimated using the following expressions:

$$f_i(t) = -\frac{\partial \bar{F}_i(t)}{\partial t} \tag{7.7}$$

$$h_i(t) = \frac{f_i(t)}{\bar{F}_i(t)} \tag{7.8}$$

7.2.2 System Reliability Modeling Considering Opportunistic Threshold

In a multi-unit system with n binary units (functional or failed), some units require maintenance after each renewal cycle to restore functionality or improve system

reliability. Unit-level maintenance can increase downtime and O&M costs. To address this, OM allows maintenance on additional units during failures if certain conditions (e.g., exceeding an opportunistic threshold T_o) are met. OM during scheduled events is more cost-efficient. For a single unit, maintenance strategies are classified based on the opportunistic threshold T_o and degradation threshold T_d as follows:

1. No Action Zone: When $0 \leq X_i(t) < T_o$, no maintenance action is required.
2. Opportunistic Maintenance Zone: When $T_o \leq X_i(t) < T_d$, the unit qualifies for OM.
3. Corrective Maintenance Zone: When $X_i(t) \geq T_d$, the unit enters the CM zone.

The trinary state of a unit at age t, based on the threshold T_o, is expressed in Eq. (7.9).

$$\begin{cases} \text{failure, if } X_i(t) \geq T_d \\ \text{unhealthy, if } T_o \leq X_i(t) < T_d \\ \text{healthy, if } 0 \leq X_i(t) < T_o \end{cases} \tag{7.9}$$

Eq. (7.9) indicates that a unit is classified as failed when its degradation level exceeds the predefined failure threshold, T_d. The differentiation between healthy and unhealthy states is determined by the opportunistic threshold, T_o. This threshold serves as a decision variable and plays a crucial role in defining the OM policy.

The main objective of the OM is to determine a set of units eligible for OM whenever CM is required for certain units during each inspection interval. The maintenance decision-making process at the kth inspection time, t_k, involves two key steps:

1. CM Selection. When units fail during the interval between t_{k-1} and t_k, the failed units are replaced at time t_k. If no CM actions are required, step 2 is skipped.

However, if CM is necessary for one or more units, it creates an opportunity to perform maintenance on other operational units. In such cases, the units requiring CM are grouped into a set, denoted as G_{CM}, and maintained according to step 1.

2. OM Selection. At time t_k, surviving units are considered for maintenance if their degradation level satisfies the condition $X_i(t_k) \geq T_o$, where T_o is the opportunistic threshold and serves as a decision variable requiring optimization. After this step, the selected units form a group referred to as G_{OM}. Consequently, the combined maintenance group, G_{COM}, is defined as the union of units selected in steps 1 and 2, i.e., $G_{\mathrm{COM}} = G_{\mathrm{CM}} \cup G_{\mathrm{OM}}$.

Units are interconnected through interdependencies, which are represented using a reliability block diagram. In a series system consisting of n units arranged in a series configuration, the failure of any single unit leads to the failure of the entire system. Assuming there are n independent units in the series, each having a reliability of one at time $t = 0$, the failure time of each unit, denoted as $t_{f,i}$, follows the distribution $F_i(t)$. The failure time of the entire system, t_f, is defined as the minimum of $\{t_{f,1}, t_{f,2}, \dots, t_{f,n}$, i.e., $t_f = \min\{t_{f,i}, i = 1,2, \dots, n\}$. Based on the hazard rate of the ith unit, the survival function of t_f is expressed as follows:

$$\bar{F}(t) = Pr\{t < t_f\} = \prod_{i=1}^{n} \bar{F}_i(t) = e^{-\int_0^t \sum_{i=1}^{n} h_i(u)du} = e^{-\int_0^t h(u)du} \quad (7.10)$$

where $h(t) = \sum_{i=1}^{n} h_i(t)$ is the system hazard rate.

For complex systems, criticality analysis is essential to assess the operational interdependence among units within a renewal cycle. However, in the case of a series system, this step can be skipped since every unit is inherently considered critical [191].

At the conclusion of each renewal cycle, appropriate reliability-updating actions are applied to the series system. Let $\bar{F}_i^-$ represent the reliability of unit i before the maintenance action. The system reliability after the maintenance action, denoted as $\bar{F}^+$, is calculated both with and without considering opportunistic factors, as expressed in Eq. (7.11).

$$\begin{cases} \bar{F}^{+,I} = \prod_{i\in\bar{G}_{CM}} \bar{F}_i^-, CBM \\ \bar{F}^{+,II} = \prod_{i\in\bar{\bar{G}}_{COM}} \bar{F}_i^-, OCBM \end{cases} \tag{7.11}$$

where $\bar{F}^{+,I}$ and $\bar{F}^{+,II}$ are system reliability functions without and with OM. Since $\bar{G}_{COM} \subseteq \bar{G}_{CM}$, it follows that when $\bar{F}_i^-$ is given, $\bar{F}^{+,II}$ is greater than $\bar{F}^{+,I}$. Due to the monotonic decreasing property of $\bar{F}(t)$, $\bar{F}^{II}(t) > \bar{F}^{I}(t)$ in the next renewal cycle.

7.3 Predictive Maintenance Model Formulation and Optimization

7.3.1 Cost Modeling Considering Periodic Inspections and Economic Dependencies

It is assumed that the unit is non-repairable, with degradation not immediately apparent and only detectable during scheduled inspections. This necessitates the implementation of an inspection-replacement policy. The system undergoes regular inspections at intervals of Δt, a design variable, while load shocks are continuously monitored through online monitoring systems [94].

At each inspection time $t_k = k\Delta t$, where $k = 1,2, \ldots$, a maintenance action is initiated immediately after the inspection if any of the following conditions are met: the degradation level exceeds the threshold T_d, or historical shocks surpass the thresholds H or δ. If none of these conditions occur, no maintenance is performed. Assuming the

system is still operational at the kth inspection, the probability of the system continuing to function until the next inspection at $t_{k+1} = (k+1)\Delta t$ is given by:

$$Pr\left\{t_f > (k+1)\Delta t \mid t_f > k\Delta t\right\} = e^{-\int_{k\Delta t}^{(k+1)\Delta t} h(u)du} \tag{7.12}$$

Proposition 7.2. *Given the degradation level* $X(t) = \{X_1(t), \ldots, X_n(t)\}$, $k\Delta t \le t \le (k+1)\Delta t$, *if* $h'(t) > 0$, *the cumulative hazard function* $\int_{k\Delta t}^{(k+1)\Delta t} h(u)du$ *between two consecutive inspections increases in* k.

The proof for Proposition 7.2 is provided in Appendix A.7.2. From Proposition 7.2, it can be inferred that the system's conditional reliability declines as the inspection index k increases.

The extended operational period is generally segmented into a finite number of renewal cycles, each representing the time interval between consecutive maintenance actions. Based on the renewal-reward theorem, the expected cost rate is expressed as follows:

$$E[C_{cr}] = \lim_{t\to\infty} \frac{C(t)}{t} = \frac{E[C(l_{rc})]}{E(l_{rc})} \tag{7.13}$$

Here, $C(t)$ represents the cumulative cost up to time t, l_{rc} denotes the length of a renewal cycle, and $C(l_{\mathrm{rc}})$ corresponds to the cost incurred within a single renewal cycle. This approach effectively reduces the time scale of the study from the extended operational period to the average duration of a renewal cycle.

As maintenance is restricted to inspection points, any system failure occurring before the next inspection will result in downtime costs accumulating until the subsequent inspection. The expected cost over a renewal cycle (l_{rc}) is calculated as the

sum of the inspection cost ($C_I \mathrm{E}[N_I]$), the system downtime cost ($C_D \mathrm{E}[\Delta t_D]$), and the maintenance cost (C_M), which can be expressed as follows:

$$E[C(l_{rc})] = C_I E[N_I] + C_D E[\Delta t_D] + C_M \tag{7.14}$$

Here, C_I represents the cost of each inspection, N_I denotes the total number of inspections, C_D is the downtime cost rate, Δt_D refers to the duration of system downtime caused by failure, and C_M is the cost associated with performing maintenance activities.

Since failures can only be identified during inspections, the number of inspections within a renewal cycle, represented by N_I, is associated with the PDF of system failure, as illustrated below.

$$\begin{aligned} E[N_I] &= \sum_{k=1}^{l} kPr\left\{(k-1)\Delta t < t_f \le k\Delta t\right\} \\ &= \sum_{k=1}^{l} k \cdot \int_{(k-1)\Delta t}^{k\Delta t} f^{II}(t)dt \end{aligned} \tag{7.15}$$

Here, l represents the total number of renewal cycles and the failure PDF of the system under the OM policy is given by $f^{II}(t) = -\frac{\partial \bar{F}^{II}(t)}{\partial t}$.

For system downtime, attention is directed to the interval $((N_I - 1)\Delta t, N_I \Delta t]$. The downtime, Δt_D, represents the time elapsed from the moment of system failure (t_f) to the next scheduled inspection, expressed as $\Delta t_D = N_I \Delta t - t_f$. The expected system downtime during a renewal cycle is determined as follows:

$$\begin{aligned} E[\Delta t_D] &= \sum_{k=1}^{l} E\left[N_I \Delta t - t_f \mid N_I = k\right] Pr\{N_I = k\} \\ &= \sum_{k=1}^{l} \int_0^{\Delta t} (k\Delta t - t) f^{II}(t)dt \cdot \int_{(k-1)\Delta t}^{k\Delta t} f^{II}(t)dt \end{aligned} \tag{7.16}$$

Maintenance cost plays a vital role in OM modeling. The total cost of performing maintenance on unit i consists of setup costs, replacement costs, and shutdown costs. The setup cost ($C_{S,i}$) includes expenses related to acquiring maintenance tools and

hiring experts, with this cost being shared when multiple units are maintained simultaneously. The replacement cost ($C_{R,i}$) accounts for the price of purchasing spare parts. Shutdown costs arise whenever a unit or system is halted for maintenance and are tied to the process of restarting the unit or system. It is worth noting that these costs may still apply even if the maintenance duration for the unit or system is ignored. Let $C_{U,i}$ and $C_{U,s}$ denote the shutdown costs for the unit and the system, respectively. Consequently, the maintenance cost for unit i can be expressed as follows:

$$C_{M,i} = C_{S,i} + C_{R,i} + (1 - \mathbb{I}_i)C_{U,i} + \mathbb{I}_i C_{U,s} \tag{7.17}$$

Here, $\mathbb{I}_i$ represents a criticality function, where $\mathbb{I}_i = 1$ if unit i is critical, and $\mathbb{I}_i = 0$ otherwise. This cost model is especially relevant for complex systems, particularly in the case of series systems (where $\mathbb{I}_i = 1$ for all i). As a result, the maintenance cost can be easily determined using the following expression:

$$C_{M,i} = C_{S,i} + C_{R,i} + C_{U,s} \tag{7.18}$$

In practice, the total maintenance cost for servicing multiple units simultaneously may differ from the cost of maintaining each unit individually. Let $C_{M,G_{COM}}$ denote the total maintenance cost for the group G_{COM}, generally expressed as $C_{M,G_{COM}} \neq \sum_{i \in G_{COM}} C_{M,i}$. The group G_{COM} is classified as PED if $C_{M,G_{COM}} < \sum_{i \in G_{COM}} C_{M,i}$, and as negative economic dependency otherwise. PED typically occurs due to shared maintenance setup costs, setup time, and shutdown costs [130]. This study focuses solely on economic dependencies; as a result, maintenance setup time, which is influenced by structural dependencies, will not be covered in detail. By sharing setup

and shutdown costs, economic dependencies among units are established, thereby improving group performance.

The maintenance setup cost for the group G_{COM} is influenced by the suppliers of the units. Let $f_{C_S}(G_{\mathrm{COM}})$ represent a function that defines the maintenance setup cost for the group G_{COM}, satisfying the condition $max_{i\in G_{COM}} C_{S,i} \leq f_{C_S}(G_{COM}) \leq \sum_{i\in G_{COM}} C_{S,i}$. Consequently, the maintenance setup cost savings can be determined as $\sum_{i\in G_{\mathrm{COM}}} C_{S,i} - f_{C_S}(G_{\mathrm{COM}}) > 0$. and if this condition holds, the group G_{COM} is deemed cost-effective. In this study, it is assumed that $f_{C_S}(G_{COM}) = C_S$, where $C_{S,i} = C_S$ for all $i = 1,2,\dots,n$, indicating that performing maintenance incurs a single setup cost. Under this assumption, the setup cost savings are given by $\Delta C_{S,G_{COM}} = (|G_{COM}| - 1)C_S$, where $|G_{\mathrm{COM}}|$ represents the number of units in the group G_{COM} [192]. Additionally, shutdown costs can generally be reduced by grouping units, as demonstrated in Eq. (7.19).

$$\Delta C_{U,G_{COM}} = \left[\sum_{i\in G_{COM}} \mathbb{I}_i - \mathbb{I}_{G_{COM}}\right] \cdot C_{U,s} \tag{7.19}$$

The reduction in shutdown cost, $\Delta C_{U,G_{\mathrm{COM}}}$, depends on the structure of the system. A PED occurs when the group G_{COM} contains more than one critical unit. On the other hand, it is classified as negative economic dependencies if G_{COM} is critical but consists solely of non-critical units [193][194]. Since the system under consideration operates in a series configuration, it results in $\Delta C_{U,G_{COM}} = 0$.

Considering the economic dependence in both setup and shutdown costs, the maintenance cost for G_{COM} over a renewal cycle can be expressed as follows:

$$\begin{aligned} C_{M,G_{COM}} &= \sum_{i\in G_{COM}} C_{M,i} - \Delta C_{S,G_{COM}} - \Delta C_{U,G_{COM}} \\ &= \sum_{i\in G_{COM}} C_{M,i} - (|G_{COM}| - 1)C_S - \left[\sum_{i\in G_{COM}} \mathbb{I}_i - \mathbb{I}_{G_{COM}}\right] \cdot C_{U,s} \end{aligned} \tag{7.20}$$

The length of a renewal cycle depends on the number of inspections performed during that cycle. The expected duration of a renewal cycle can be represented as:

$$E[l_{rc}] = \sum_{k=1}^{l} E[l_{rc} \mid N_I = k] Pr\{N_I = k\} = \sum_{k=1}^{l} k\Delta t \cdot \int_{(k-1)\Delta t}^{k\Delta t} f^{II}(t)dt \quad (7.21)$$

From Eqs. (7.13) and (7.14), it is clear that $\mathrm{E}[C_{\mathrm{cr}}]$ is mainly composed of three expected values: $\mathrm{E}[N_I]$, $E[\Delta t_D]$, and $E[l_{\mathrm{rc}}]$. These expected values can be calculated using MCS.

Therefore, the expected long-term cost rate is calculated as shown in Eq. (7.22).

$$E[C_{cr}] = \frac{C_I \sum_{k=1}^{l} k \cdot \int_{(k-1)\Delta t}^{k\Delta t} f^{II}(t)dt + C_D \sum_{k=1}^{l} \int_{0}^{\Delta t} (k\Delta t - t) f^{II}(t)dt \cdot \int_{(k-1)\Delta t}^{k\Delta t} f^{II}(t)dt + \sum_{i \in G_{COM}} C_{M,i} - (|G_{COM} - 1|) C_S - [\sum_{i \in G_{COM}} \mathbb{I}_i - \mathbb{I}_{COM}] \cdot C_{U,s}}{\sum_{k=1}^{l} k\Delta t \cdot \int_{(k-1)\Delta t}^{k\Delta t} f^{II}(t)dt} \quad (7.22)$$

The system's lifecycle, l_{lc}, is defined as its long-term operational duration. The goal is to minimize the expected cost over the lifecycle l_{lc} by optimally determining the inspection interval Δt, the opportunistic threshold T_o, the shock spatial threshold H, and the shock temporal threshold δ. The optimization model is formulated as follows:

$$\begin{aligned} & minE[C(l_{lc})] = E[C_{cr}] \cdot l_{lc} \\ & \text{s.t.} \begin{cases} \eta_O T_d \le T_o \le T_d \\ 0 \le H \le H^u \\ \delta \ge \delta^l \\ \Delta t > 0 \end{cases} \end{aligned} \quad (7.23)$$

The first constraint establishes that the lower limit of the opportunistic threshold is $\eta_O T_d$, where η_O represents the opportunistic coefficient [133]. Meanwhile, the second and third constraints utilize H^u and δ^l to guarantee operational safety.

The expected inspections, $\mathrm{E}[N_I]$, and downtime, $E[\Delta t_D]$, depend on Δt, T_o, H, and δ. A very small Δt increases inspection costs, while a large Δt raises failure risks. Setting T_o too low leads to premature replacements, wasting remaining useful life, whereas a high T_o increases failure risks post-inspection. Raising H (or lowering δ)

reduces maintenance frequency, but $\mathrm{E}[C_{\mathrm{cr}}]$ stabilizes once H (or δ) exceeds a threshold. Thus, the optimal H^* (or δ^*) balances reliability and maintenance efficiency.

Eq. (7.23) optimizes the inspection interval Δt, opportunistic threshold T_o, spatial shock threshold H, and temporal shock threshold δ to minimize O&M costs over $[0, l_{\mathrm{lc}}]$. Each combination of $(\Delta t, T_o, H, \delta)$ defines an OCBM policy. While semi-regenerative process theory can theoretically provide closed-form solutions, it applies only to single-unit systems. Deriving expectations like $\mathrm{E}[N_I]$, $\mathrm{E}[\Delta t_D]$, and $\mathrm{E}[l_{rc}]$ in Eq. (7.22) is highly complex, so an MCS approach is used to simulate and evaluate $C(l_{\mathrm{lc}}|\Delta t, T_o, H, \delta)$.

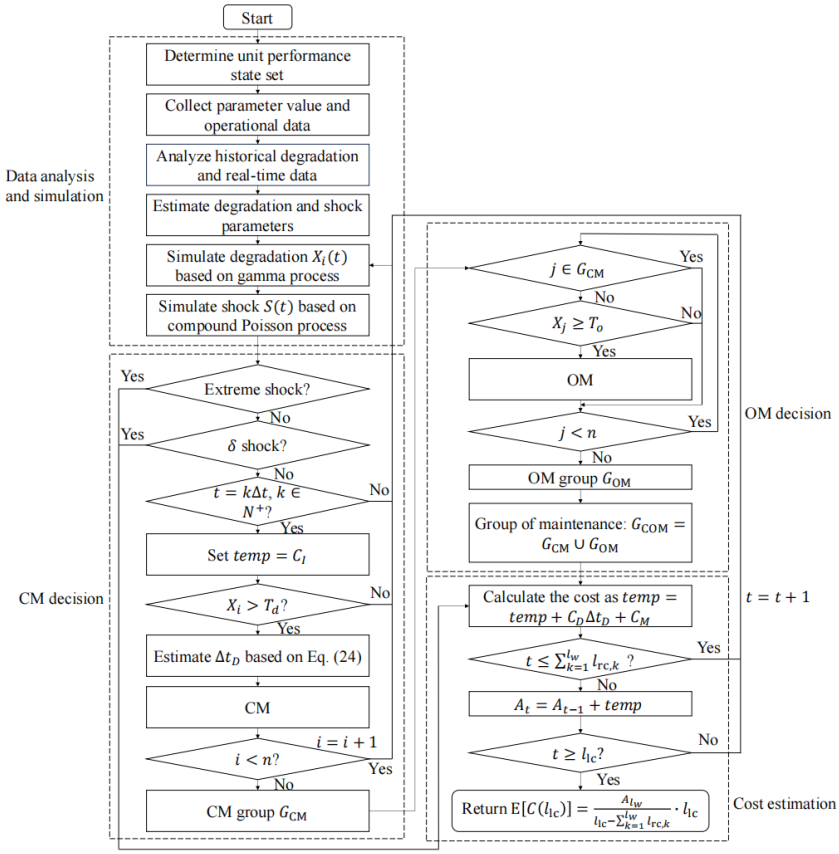


Fig. 7.2 The simulation process of the OPdM model.

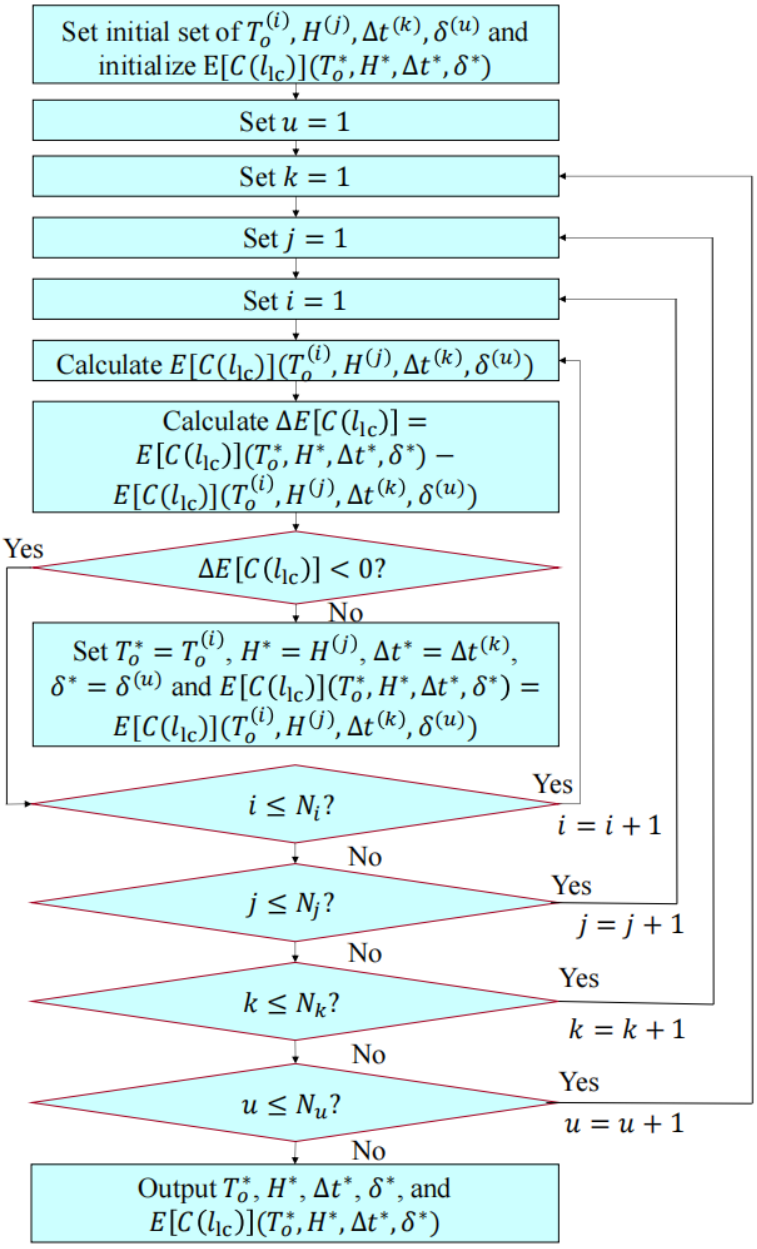


Fig. 7.3 The optimization process of the OPdM model.

A MATLAB program was developed to create an MCS environment for calculating $C(l_{lc}|\Delta t, T_o, H, \delta)$. The simulation and optimization processes of the OM are shown in Figs. 7.2 and 7.3, respectively. MCS evaluates $C(l_{lc})$ for each simulation episode by simulating the O&M process under policy $(\Delta t, T_o, H, \delta)$ until the lifecycle l_{lc} ends. All units start as new, and the initial l_w renewal cycles are treated as warm-up periods, excluded from final results. The O&M cost is given by $C(l_{lc}) = \frac{A_{l_w}}{l_{lc} - \sum_{k=1}^{l_w} l_{rc,k}} \cdot l_{lc}$, where

A_{l_w} is the total O&M cost during $[\sum_{k=1}^{l_w} l_{\mathrm{rc},k}, l_{\mathrm{lc}}]$, and the steady-state cost rate is $\frac{A_{l_w}}{l_{lc}-\sum_{k=1}^{l_w} l_{\mathrm{rc},k}}$. For each policy, N_{sim} trials are conducted, and the average O&M cost is computed. Convergence of $\mathrm{E}[C(l_{\mathrm{lc}})]$ is ensured using a large N_{sim} and a tolerance parameter ρ, satisfying $\Delta E[C(l_{\mathrm{lc}})] \leq \rho$.

Inspection and maintenance costs are easily calculated from the degradation state $X_i(t_k)$ at inspection time t_k. Downtime costs, however, are estimated based on the system's overall degradation level at each inspection, expressed as $X(t_k) = \max\{X_i(t_k), i = 1,2,\dots,n\}$. Let $X^-(t_k)$ and $X^+(t_k)$ denote the system's degradation levels at the start and end of the kth inspection interval, respectively. The system downtime during this interval is then expressed as:

$$\Delta t_D = \begin{cases} 0, \text{ if } X^+(t_k) < T_d \\ \frac{\Delta t(X^+(t_k)-T_d)}{X^+(t_k)-X^-(t_k)}, \text{ if } X^-(t_k) < T_d \& X^+(t_k) \geq T_d \end{cases} \tag{7.24}$$

The downtime cost within a renewal cycle, $C_D \Delta t_D$, can be estimated through simulation [133]. Using $X_i^+(t_k)$, the condition of the units in the system—classified as failed, unhealthy, or healthy—can be determined based on Eq. (7.9). This allows for identifying the maintenance group G_{COM}, from which the maintenance cost $C_{M,G_{\mathrm{COM}}}$ can be calculated. By accumulating the inspection cost, downtime cost, and maintenance cost over an episode, the total cost $C(l_{\mathrm{lc}})$ can then be determined.

7.3.2 Performance Comparison Metrics

A comparative analysis is conducted between CM (policy I) and OM (policy). To facilitate this, four quantitative criteria are introduced: Reliability improvement factor, renewal frequency, maintenance cost rate, and O&M cost rate [52].

(1) Health improvement rate: To evaluate the health improvement potential of the OM, the health improvement rate, denoted as $\%\epsilon_1$, defined as follows.

$$\%\epsilon_1 = \begin{cases} \frac{E[X^{II*}]-E[X^{I*}]}{E[X^{I*}]} \times 100\%, \text{ if } E[X^{I*}] < E[X^{II*}] \\ 0, \text{otherwise} \end{cases} \tag{7.25}$$

(2) Renewal frequency. For a fixed operational lifecycle l_{lc}, a decrease in renewal frequency results in an increase in renewal duration, which subsequently reduces maintenance expenses, such as the rental costs of maintenance vessels. The percentage increase in renewal duration, denoted as $\%\epsilon_2$, is determined as follows:

$$\%\epsilon_2 = \begin{cases} \left(\frac{E[l_r^{I*}]}{E\left[l_c^{+[}\right]} - 1\right) \times 100\%, \text{ if } E[l_{rc}^{I}] < E[l_{rc}^{II*}] \\ 0, \text{otherwise} \end{cases} \tag{7.26}$$

(3) Maintenance cost rate. Let ϵ_3^{I*} and ϵ_3^{II*} denote the maintenance cost rate of the CBM and OCBM, respectively. Then

$$\epsilon_3^{I*} = \frac{\sum_{i \in G_{CM}} C_{M,i}}{\sum_{k=1}^{l} k\Delta t^{I*} \cdot \int_{(k-1)\Delta t^{I*}}^{k\Delta t^{I*}} f^{I*}(t)dt} \tag{7.27}$$

$$\epsilon_3^{II*} = \frac{C_{M,G_{COM}}}{\sum_{k=1}^{l} k\Delta t^{II*} \cdot \int_{(k-1)\Delta t^{II*}}^{k-II*} (t)dt} \tag{7.28}$$

Let $\%\epsilon_3$ represent the percentage reduction in the maintenance cost achieved through the OCBM policy, thus

$$\%\epsilon_3^{I*} = \frac{\sum_{i \in G_{CM}} C_{M,i}}{\sum_{k=1}^{l} k\Delta t^{I*} \cdot \int_{(k-1)\Delta t^{I*}}^{k\Delta t^{I*}} f^{I*}(t)dt} \tag{7.29}$$

(4) O&M cost rate. Let Δt^{I}, H^{I}, and δ^{I} represent the decision variables for the OM policy. Thus, the expected cost rate $\mathrm{E}[C_{cr}^{\mathrm{I}}]$ is expressed by:

$$E[C_{cr}^{I*}](\Delta t^{I*}, H^{I*}, \delta^{I*}) = inf\{E[C_{cr}^{I}](\Delta t^{I}, H^{I}, \delta^{I})\} \tag{7.30}$$

Let $\%\epsilon_4$ represent the percentage reduction in O&M costs achieved by OCBM

compared to CBM. Then, we have:

$$\%\epsilon_4 = \begin{cases} \left(1 - \frac{E[C_{Cr}^{II*}]}{E[C_r rr]}\right) \times 100\%, \text{ if } E[C_{cr}^{I*}] > E[C_{cr}^{II*}] \\ 0, \text{ otherwise} \end{cases} \tag{7.31}$$

$$E[C_{cr}^{I}] = \frac{\begin{matrix} C_I \sum_{k=1}^{l} k \cdot \int_{(k-1)\Delta t^I}^{k\Delta I^I} f^I(t)dt \\ +C_D \sum_{k=1}^{l} \int_0^{\Delta t^I} (k\Delta t^I - t) f^I(t)dt \cdot \int_{(k-1)\Delta t^I}^{k\Delta t^I} f^I(t)dt + \sum_{i \in G_{CM}} C_{M,i} \end{matrix}}{\sum_{k=1}^{l} k\Delta t^I \cdot \int_{(k-1)\Delta t^I}^{k\Delta t^I} f^I(t)dt} \tag{7.32}$$

7.4 Case Study: Gas-Insulated Switchgear Series and Parallel Systems

7.4.1 Case Descriptions

This section demonstrates the proposed OM policy for GIS maintenance in series and parallel systems. GIS units (Model XGN-2×27.5/2500-31.5) from the Xinlechang traction substation in the TPSS of the Wuhan-Guangzhou HSR, China, are used, as in [58]. GIS performance deteriorates due to wear, fatigue, or insulation aging, while frequent high-amplitude shocks from traction loads can cause failures. A supervisory control and data acquisition system monitors load data to detect shocks, defined as currents exceeding the 2,500 A rating per IEC 62271-200 [195]. Shock model parameters are estimated via MLE and validated with the K-S test, while other parameters are provided by GIS manufacturers and the railway bureau. Table 7.1 lists all parameter values.

Table 7.1 Parameters in the case study.

Parameters	Values	Sources
λ	6.1033	MLE method (95% CI)

μ_Y	2703.8019	MLE method (95% CI)
σ_Y	88.2643	MLE method (95% CI)
α	1/12	Manufacturers
β	1/6	Manufacturers
T_d	15	Manufacturers
η	1/15	Manufacturers
H^u	3,143 A	Manufacturers
δ^l	20 h	Manufacturers
l_{lc}	250,000 h	Manufacturers & Railway Bureau
C_I	10 $	Manufacturers & Railway Bureau
C_D	200 $	Manufacturers & Railway Bureau
C_R	20 $	Manufacturers & Railway Bureau
C_S	20 $	Manufacturers & Railway Bureau
C_U	10 $	Manufacturers & Railway Bureau

7.4.2 Optimization Results

To validate the convergence of O&M costs via MCS, the OM policy is evaluated with parameters $\Delta t = 500h$, $T_o = 10$, $H = 3{,}143A$, and $\delta = 20h$. The simulation runs over an extended duration to ensure cost convergence. Fig. 7.4 shows the convergence of C_{cr} as the number of planned periods increases, considering varying initial degradation levels. Initially, C_{cr} fluctuates due to adjustments to initial conditions but gradually stabilizes, demonstrating convergence. Trajectories from six simulations in Fig. 7.4 confirm that, despite differing initial degradation levels, the cost rate converges to similar stable values, highlighting the model's robustness. After about 50,000 planned periods, the cost rate stabilizes around $2.

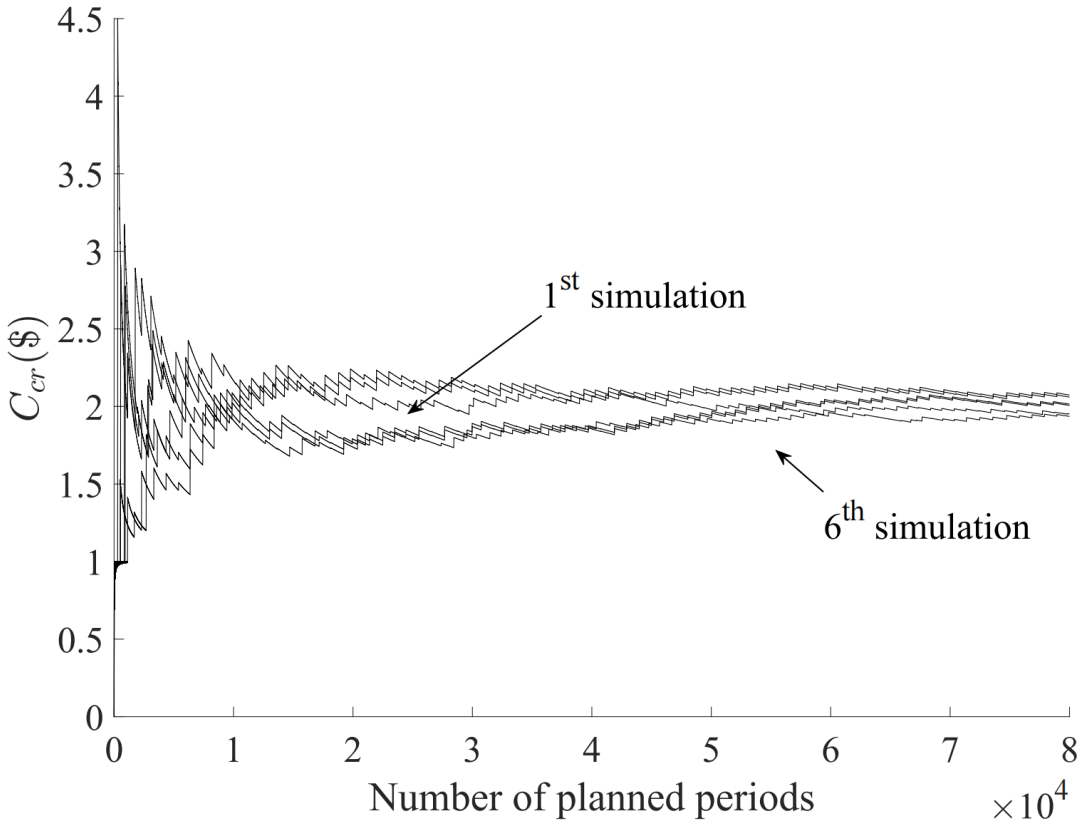


Fig. 7.4 The convergence of the cost rate under $\Delta t = 500h$, $T_o = 10$, $H = 3{,}143A$, $\delta = 20h$, and different initial degradation levels.

After 1,000 MCS repetitions, the optimization results are shown in Figs. 7.5 and 7.6. Although only for the series system, the parallel system can be optimized similarly. Fig. 7.5 explores how $\mathrm{E}[C(l_{\mathrm{lc}})]$ varies with two variables from $\{\Delta t, T_o, H, \delta\}$ while keeping the others constant. Fig. 7.6 includes four 3D surface plots showing $\mathrm{E}[C(l_{\mathrm{lc}})]$ variations with decision variables:

1. **Fig. 6a**: $\mathrm{E}[C(l_{\mathrm{lc}})]$ vs. Δt and T_o forms a convex function, indicating unique optimal values for inspection intervals and opportunistic thresholds.
2. **Fig. 6b**: $\mathrm{E}[C(l_{\mathrm{lc}})]$ vs. Δt and H shows a sharp increase in cost as H decreases, suggesting H should be maximized. With $H^u = 3{,}143A$, the optimal $H^* = H^u = 3{,}143$.

3. **Fig. 6c**: E$[C(l_{\mathrm{lc}})]$ vs. T_o and δ indicates δ is positively correlated with cost. Minimizing E$[C(l_{\mathrm{lc}})]$ requires $\delta^* = \delta^l = 20h$.

4. **Fig. 6d**: E$[C(l_{\mathrm{lc}})]$ vs. H and δ shows lower costs with larger H and smaller δ, while costs rise significantly as H decreases and δ increases.

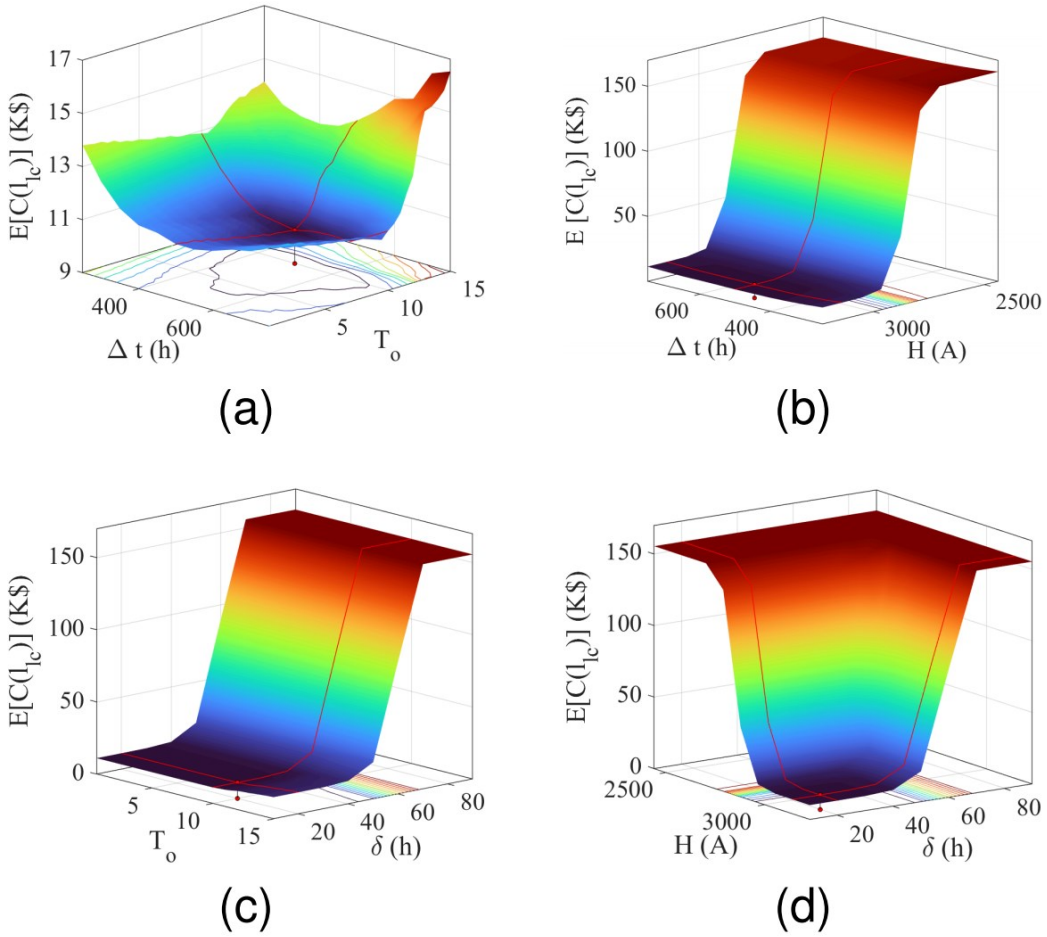


Fig. 7.5 Surfaces of the expected total cost E$[C(l_{\mathrm{lc}})]$ concerning decision variables $\{\Delta t, T_o, H, \delta\}$. (a) E$[C(l_{\mathrm{lc}})]$ as a function of Δt and T_o. (b) $E[C(l_{\mathrm{lc}})]$ as a function of Δt and H. (c) E$[C(l_{\mathrm{lc}})]$ as a function of T_o and δ. (d) E$[C(l_{\mathrm{lc}})]$ as a function of H and δ.

The convex shape of the expected cost surfaces confirms the existence and uniqueness of optimal decision variables. In Fig. 7.6, orthogonal curves converge at the minimum E$[C(l_{\mathrm{lc}})]$ point, showing its variation with each decision variable:

1. **Fig. 7.6a**: E$[C(l_{\mathrm{lc}})]$ forms a U-shaped curve with **Δt**, reaching its minimum (optimal inspection interval) at Δt = 500 h.

2. **Fig. 7.6b**: A U-shaped curve with T_o shows the minimum (optimal opportunistic threshold) at T_o=10.
3. **Fig. 7.6c**: E$[C(l_{\mathrm{lc}})]$ decreases as H increases, stabilizing at the optimal $H^* = H^u = 3{,}143A$.
4. **Fig. 7.6d:** E$[C(l_{\mathrm{lc}})]$ rises with δ, confirming the optimal $\delta^* = \delta^l = 20h$.

The convex shape of the expected cost surfaces confirms the existence and uniqueness of optimal values for the decision variables. As depicted in Fig 7.6, the orthogonal curves converge at the minimum point of E$[C(l_{\mathrm{lc}})]$. The figure illustrates how E$[C(l_{\mathrm{lc}})]$ changes as it varies with each decision variable. Fig 7.6a demonstrates that E$[C(l_{\mathrm{lc}})]$ initially decreases and then increases as Δt grows, forming a U-shaped curve. The minimum cost, representing the optimal inspection interval, occurs at $\Delta t = 500h$. Similarly, Fig 7.6b reveals that E$[C(l_{\mathrm{lc}})]$ also follows a U-shaped curve with respect to T_o, reaching its minimum at $T_o = 10$, which corresponds to the optimal opportunistic threshold. In Fig 7.6c, E$[C(l_{\mathrm{lc}})]$ decreases as H increases and eventually stabilizes, indicating that the optimal value is $H^* = H^u = 3{,}143A$. Meanwhile, Fig 7.6d shows that E$[C(l_{\mathrm{lc}})]$ rises as δ increases, confirming that the optimal value is $\delta^* = \delta^l = 20h$. Together, these subfigures highlight the existence and uniqueness of the optimal decision variables required to minimize the expected cost.

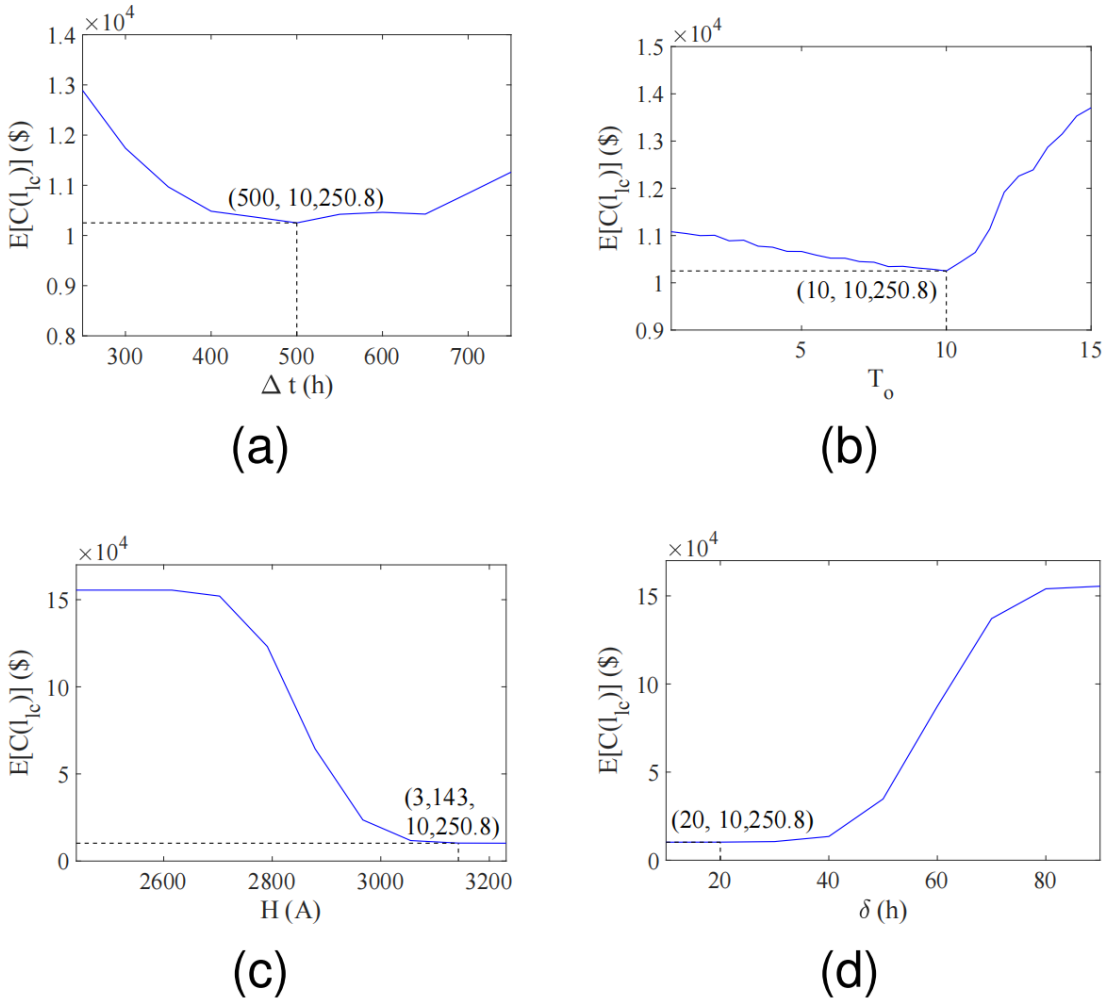


Fig. 7.6 Curves of the expected total cost $\mathrm{E}[C(l_{\mathrm{lc}})]$ concerning decision variables $\{\Delta t, T_o, H, \delta\}$. (a) $\mathrm{E}[C(l_{\mathrm{lc}})]$ as a function of Δt. (b) $\mathrm{E}[C(l_{\mathrm{lc}})]$ as a function of T_o. (c) $\mathrm{E}[C(l_{\mathrm{lc}})]$ as a function of H. (d) $\mathrm{E}[C(l_{\mathrm{lc}})]$ as a function of δ.

7.4.3 Sensitivity Analysis

In practice, OM models are built using factors like load data, GIS equipment, system structure, and cost parameters. Sensitivity analysis is crucial to evaluate how parameters (e.g., reliability, cost, and structure) affect outcomes. Twelve parameters, grouped into shock, degradation, cost models, and GIS unit count, are analyzed. Results for series and parallel systems are presented in Figs. 7.7–7.8 and Tables 7.2–7.3.

OM models rely on factors like load data, GIS equipment, system structure, and costs. Sensitivity analysis evaluates the impact of 9 parameters across degradation,

costs, and GIS unit count. Results for series and parallel systems appear in Figs. 7.7–7.8 and Tables 7.2–7.3.

Fig. 7.7 shows the impact of degradation parameters on series systems:

1. **Fig. 7.7a**: As α rises, reliability decreases, causing $\mathrm{E}[C(l_{\mathrm{lc}})]^*$ to increase and Δt^* to decrease, with little effect on T_o^*, H^*, and δ^*. Reduced reliability shortens lifespan, increasing costs and requiring shorter inspection intervals.
2. **Fig. 7.7b**: Increasing β extends lifespan, lowering $\mathrm{E}[C(l_{\mathrm{lc}})]^*$ and increasing Δt^*, while other variables remain stable. Fewer renewals reduce costs and extend inspection intervals.
3. **Fig. 7.7c**: Higher T_d improves reliability, lowering $\mathrm{E}[C(l_{\mathrm{lc}})]^*$ while increasing the inspection interval and opportunistic threshold to reduce wasted life.

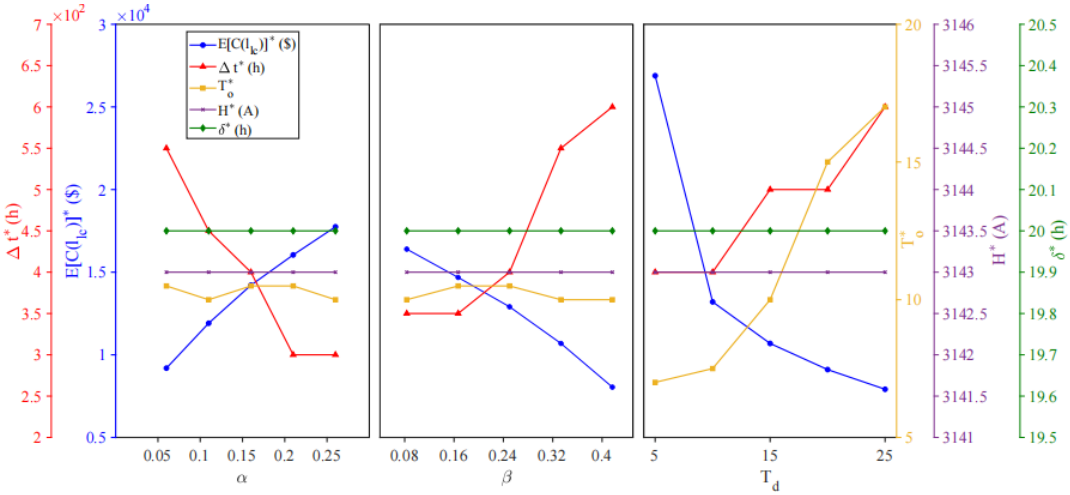


Fig 7.7 The sensitivity analysis of unit degradation parameters on series systems.

The effects of α, β, and T_d on parallel systems are similar to series systems (Fig. 7.8):

1. α increases $\mathrm{E}[C(l_{\mathrm{lc}})]^*$ and reduces Δt^*, *while* β *reduces* $\mathrm{E}[C(l_{\mathrm{lc}})]^*$ and increases Δt^*.
2. Higher T_d lowers $\mathrm{E}[C(l_{\mathrm{lc}})]^*$ and increases Δt^* and T_o^*.
3. Unlike series systems, α negatively correlates with T_o^* in parallel systems due

to system structure.

In series systems, a unit failure halts the entire system, preserving other units' degradation levels. In parallel systems, failed units don't stop operation, leading to higher average degradation at inspections and reducing T_o^*.

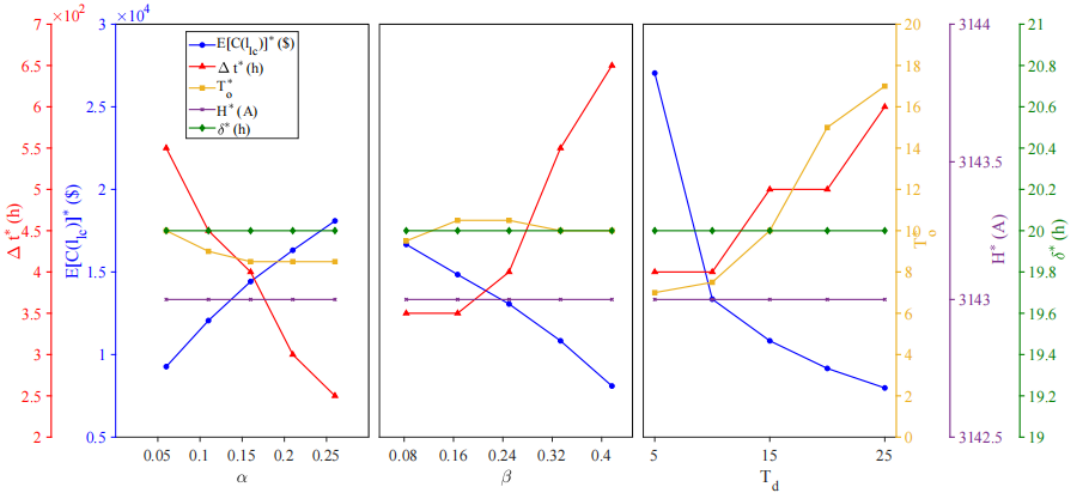


Fig. 7.8 The sensitivity analysis of unit degradation parameters on parallel systems.

Table 7.2 summarizes the effects of cost parameters on expected cost and design variables for series and parallel systems:

1. Increasing C_I, C_D, C_S, C_U, or C_R raises $\mathrm{E}[C(l_{\mathrm{lc}})]^*$, with C_I having the largest impact.
2. In series systems:
   (1) Higher C_I increases Δt^* to reduce inspection frequency, while other variables remain unaffected.
   (2) Higher C_D shortens Δt^* to minimize downtime costs.
   (3) C_S and C_U minimally affect design variables but raise $\mathrm{E}[C(l_{\mathrm{lc}})]^*$.
   (4) Higher C_R increases T_o^* to reduce wasted life during opportunistic maintenance.

3. In parallel systems, cost parameters have similar effects, except C_U, which increases T_o^*. This adjustment minimizes unit shutdown costs during opportunistic maintenance.

The number of GIS units also affects OCBM policies for series and parallel systems, as shown in Table 7.3:

1. Series Systems
   (1) The number of GIS units does not impact Δt^*, H^*, or δ^*.
   (2) T_o^* decreases with more GIS units, improving reliability and controlling downtime costs.
   (3) $\mathrm{E}[C(l_{\mathrm{lc}})]^*$ increases with more units, rising from \$10,261 to \$11,202 due to higher O&M costs.
2. Parallel Systems
   (1) Δt^* decreases as GIS units increase (from 500 to 400 hours) to reduce penalty costs from concurrent unit failures.
   (2) T_o^* decreases, while H^* and δ^* remain unchanged.
   (3) $\mathrm{E}[C(l_{\mathrm{lc}})]^*$ rises with more units, increasing from \$10,332 to \$11,564.

Table 7.2 Sensitivity analysis of cost parameters in series and parallel systems.

C_I	C_D	C_S	C_U	C_R	Series					Parallel				
					Δt^*	T_0^*	H^*	δ^*	$E[C(l_{lc})]^*$	Δt^*	T_0^*	H^*	δ^*	$E[C(l_{lc})]^*$
10	200	20	10	20	500	10	3,143	20	11,087	500	10	3,143	20	10,834
20	—	—	—	—	650	11	3,143	20	15,170	700	11	3,143	20	14,969
30	—	—	—	—	700	10	3,143	20	18,597	800	10	3,143	20	18,390
10	100	20	10	20	550	10	3,143	20	8,489	700	11	3,143	20	7,988
—	200	—	—	—	500	10	3,143	20	11,087	500	10	3,143	20	10,834
—	300	—	—	—	350	11	3,143	20	13,158	400	11	3,143	20	12,893
10	200	10	10	20	500	10	3,143	20	11,035	500	10	3,143	20	10,781
—	—	20	—	—	500	10	3,143	20	11,087	500	10	3,143	20	10,834
—	—	30	—	—	500	10	3,143	20	11,139	500	10	3,143	20	10,886
10	200	20	10	20	500	10	3,143	20	11,087	500	10	3,143	20	10,834
—	—	—	20	—	500	10	3,143	20	11,139	500	10.5	3,143	20	10,934

—	—	—	30	—	500	10	3,143	20	11,191	500	11	3,143	20	11,035
10	200	20	10	10	500	9.5	3,143	20	10,986	500	9.5	3,143	20	10,733
—	—	—	—	20	500	10	3,143	20	11,087	500	10	3,143	20	10,834
—	—	—	—	30	500	10.5	3,143	20	11,188	500	10.5	3,143	20	10,934

Table 7.3 Effects of GIS number on design variables and cost rates in series and parallel systems.

Structure	Variables and OFVs	Number of GIS				
		2	3	4	5	6
Series	$\triangle t^*$ (h)	500	500	500	500	500
	T_0^*	10	9.5	8.0	7.5	7.5
	H^* (A)	3,143	3,143	3,143	3,143	3,143
	δ^*(h)	20	20	20	20	20
	$E[C_{lc}]^*$ ($)	10,261	10,649	10,771	11,141	11,202
Parallel	$\triangle t^*$ (h)	500	450	450	400	400
	T_0^*	10	8.5	8.5	8	8
	H^* (A)	3,143	3,143	3,143	3,143	3,143
	δ^*(h)	20	20	20	20	20
	$E[C_{lc}]^*$ ($)	10,332	10,784	11,155	11,421	11,564

7.4.4 Policy Comparison

To highlight the effectiveness of the proposed OM policy, a comparison is made with two related maintenance strategies.

1. CM: Each GIS unit is replaced individually as soon as its degradation level surpasses a predetermined threshold.
2. OM without shared setup costs: OM is performed when the degradation levels of certain units exceed the opportunistic threshold. However, this approach does not consider economic dependence, meaning the setup costs are not shared.

For the CM policy without OM, the optimal design uses T_o = 15. Fig. 7.9 compares optimal maintenance under CM and OM policies for series and parallel systems:

1. Series systems: The CM policy causes four failures, over twice as many as the

OM policy.

2. Parallel systems: Failures under the CM policy are also significantly higher than with the OM policy.

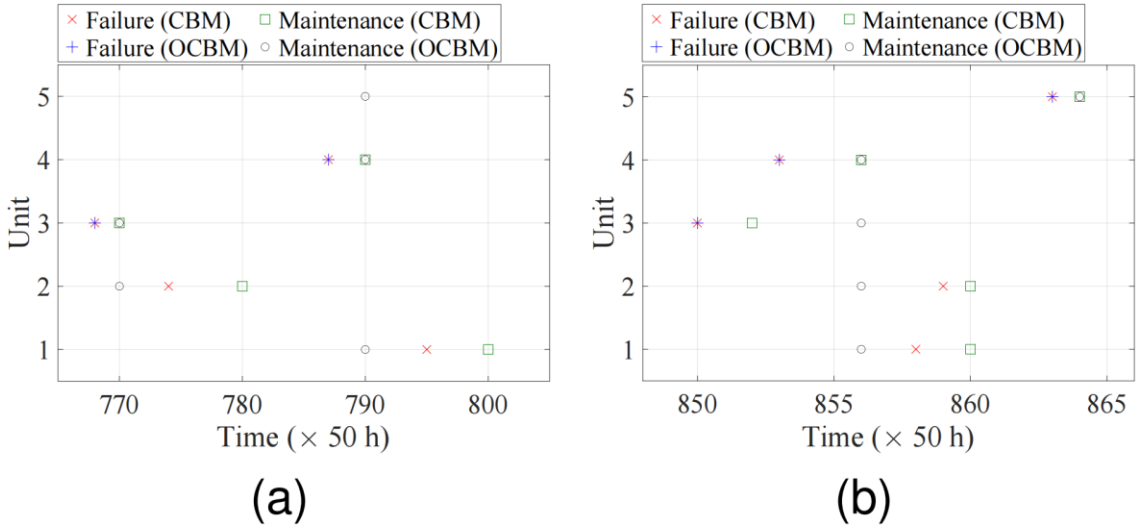


Fig. 7.9 OM policy versus CM policy. (a) Series systems. (b) Parallel systems.

To evaluate the OM policy, a state duration sampling technique is used to design optimal CM, OM without shared setup costs, and OM with shared setup costs for GIS series and parallel systems. Table 7.4 compares performance metrics, including average degradation level, renewal frequency, maintenance cost rate, and O&M cost rate:

1. Series systems:

   (1) Average Degradation Level: OM without shared setup costs reduces it by 29.6%–39% compared to CM, unchanged after shared setup costs.

   (2) Renewal Frequency: OM without shared setup cuts it by 45.5%–80.4%, consistent with shared setup costs.

   (3) Maintenance Cost Rate: OM without shared setup reduces it by 2.2%–3.3%, increasing to 24.2%–41.3% with shared setup costs.

   (4) O&M Cost Rate: OM without shared setup lowers it by 24.7%–62.1%,

improving to 25.4%–63.8% with shared setup costs.

2. Parallel systems:

   (1) Average Degradation Level: No change under OM (with or without shared setup costs).

   (2) Renewal Frequency: Significant reductions remain consistent with shared setup costs.

   (3) Maintenance and O&M Cost Rates: Reductions are greater with shared setup costs.

Table 7.4 shows that the OCBM policy with shared setup costs achieves greater reductions in average degradation level, renewal frequency, maintenance cost rate, and O&M cost rate as the number of GIS units increases, except for the average degradation level in parallel systems.

The distinction arises because:

1. In series systems, degradation is driven by the unit with the highest degradation.
2. In parallel systems, it depends on the unit with the lowest degradation.

Additionally, the OCBM policy without shared setup costs has a lower maintenance cost rate than CBM, as OM incurs extra setup costs.

Table 7.4 Performance analysis of the OM policy with shared setup costs compared to CM and OM policies without shared setup costs in series and parallel systems.

Criteria	Number of GIS	CBM	OCBM (without shared setup costs)		OCBM (with shared setup costs)	
			Value	% Reduction	Value	% Reduction
Series						
Average degradation level	2	9.92	7.72	%29.6↓	7.72	%29.6↓
	3	11.16	7.86	%29.6↓	7.86	%29.6↓
	4	12.00	7.78	%35.2↓	7.78	%35.24
	5	12.45	7.82	%37.2↓	7.82	%37.24
	6	12.86	7.84	%39.0↓	7.84	%39.04
Renewal frequency	2	1,826	996	%45.5↓	996	%45.5↓
	3	2,731	1,044	%61.8↓	1,044	%61.8↓

	4	3,616	1,026	%71.6↓	1,026	%71.6↓
	5	4,503	1,044	%76.8↓	1,044	%76.8↓
	6	5,367	1,052	%80.4↓	1,052	%80.4↓
Maintenance cost rate	2	0.091	0.088	%3.3↓	0.069	%24.2↓
	3	0.137	0.131	%4.4↓	0.092	%32.8↓
	4	0.181	0.174	%3.9↓	0.113	%37.6↓
	5	0.225	0.219	%2.7↓	0.136	%39.6↓
	6	0.269	0.263	%2.2↓	0.158	%41.3↓
O&M cost rate	2	2.751	2.071	%24.7↓	2.052	%25.4↓
	3	3.597	2.169	%39.7↓	2.130	%40.8↓
	4	4.448	2.240	%49.6↓	2.178	%51.0↓
	5	5.266	2.312	%56.1↓	2.228	%57.7↓
	6	6.191	2.345	%62.1↓	2.240	%63.8↓
Parallel						
Average degradation level	2	7.71	7.71	0	7.71	0
	3	7.73	7.73	0	7.73	0
	4	7.83	7.83	0	7.83	0
	5	7.82	7.82	0	7.82	0
	6	7.87	7.87	0	7.87	0
Renewal frequency	2	1,828	994	%45.6↓	994	%45.64
	3	2,736	1,010	%63.1↓	1,010	%63.1↓
	4	3,634	1,038	%71.4↓	1,038	%71.4↓
	5	4,525	1,048	%76.8↓	1,048	%76.8↓
	6	5,367	1,060	%80.2↓	1,060	%80.24
Maintenance cost rate	2	0.092	0.097	0	0.078	%15.24
	3	0.138	0.151	0	0.111	%19.6↓
	4	0.183	0.205	0	0.144	%21.3↓
	5	0.230	0.261	0	0.178	%22.6↓
	6	0.272	0.315	0	0.210	%22.8↓
O&M cost rate	2	2.737	2.085	%23.8↓	2.066	%24.5↓
	3	3.461	2.197	%36.5↓	2.157	%37.7↓
	4	4.218	2.292	%45.7↓	2.231	%47.1↓
	5	4.733	2.368	%50.04	2.284	%51.7↓
	6	6.227	2.418	%61.2↓	2.313	%62.9↓

7.5 Summary

This chapter presents a DT-OPdM model for system-level maintenance scheduling, incorporating economic dependencies. Its effectiveness is demonstrated via a case study on GIS series and parallel systems. Units are modeled with a DT of competing failure mechanisms combining shock effects and degradation. System reliability is based on unit reliability, connection topology, and opportunistic thresholds. The model minimizes cost rates by considering PED and maintenance costs. Key parameters, including opportunistic threshold, inspection interval, and shock thresholds, are optimized using MCS to enhance OPdM performance in series and parallel systems.

Chapter 8

Conclusions and Future Work

8.1 Conclusions

DT-PHM represents an innovative and transformative architecture for enhancing system reliability, optimizing maintenance strategies, and extending the operational life of components. By leveraging industrial Internet-of-Things and model-based systems engineering, DT-PHM bridges the gap between PAs and their VRs. This integration allows for precise monitoring of system health, accurate prediction of failures, and informed decision-making for maintenance and operational strategies. The adaptability of DT-PHM makes it invaluable for addressing challenges posed by complex engineering systems, helping organizations minimize downtime, reduce costs, and improve safety and performance.

The adoption of metamodel-based DTs for life reliability prediction has further enhanced the applicability of DT-PHM. Metamodels, such as Bayesian BiLSTM and AK, provide computationally efficient and accurate predictions of component degradation and RUL. These models are especially well-suited for scenarios involving cumulative fatigue damage and other complex degradation mechanisms. Numerical experiments have demonstrated the superiority of metamodel-based approaches over FEMs, offering greater predictive efficiency and robustness under varying operational conditions.

The application of DTs in warranty design has revolutionized how manufacturers

optimize warranty policies. By utilizing DTs to analyze component degradation and failure risks, manufacturers can tailor warranty durations, pricing, and PM intervals to specific reliability levels. The incorporation of DT insights allows manufacturers to balance warranty costs with customer satisfaction, while also providing a competitive edge in the marketplace. The DT-based warranty design ensures that warranty policies align with real-world performance, reducing risks and fostering greater trust between manufacturers and customers.

At the unit level, DT-based PdM enables precise monitoring and optimization of individual components. By creating VRs of critical assets, the DT facilitates real-time tracking of operational conditions and degradation patterns. This allows maintenance teams to implement maintenance strategies, where interventions are performed only when necessary, minimizing unnecessary repairs and maximizing asset utilization. The application of the DT at the unit level is particularly effective for components with high failure risks or those operating under harsh conditions, ensuring timely maintenance actions and reducing the likelihood of unexpected failures.

At the system level, the DT extends its capabilities to manage interconnected systems comprising multiple components. DT-OPdM integrates data from individual units to provide a holistic view of system health and operational performance by considering economic dependencies. This method accounts for interactions between components, enabling maintenance strategies that minimize overall costs rather than focusing solely on individual parts. By leveraging system-level insights, organizations can achieve more reliable and efficient operations, ultimately enhancing productivity and reducing

lifecycle costs.

In conclusion, DT-PHM represents a paradigm shift in maintenance and reliability management. The framework enables decision-making, improve operational efficiency, and optimize lifecycle costs across various industries. The integration of DT technologies continues to evolve, offering new opportunities for innovation and further enhancing the reliability and performance of engineered systems.

8.2 Future Work

The significant contribution of this thesis lies in the seamless integration of a PHM system with a DT, showcasing the potential of DT technology in enhancing warranty management and PdM. However, certain deviations from the expected actual responses observed during the study can be attributed to the incomplete consideration of several relevant factors, such as the influence of complex environmental conditions, stochastic operational variations, and the interactions among system components. These limitations highlight opportunities for further refinement and development. To enhance the robustness and applicability of the proposed framework, future research should focus on addressing these aspects in the broader context of DT technology, including improved modeling techniques, the incorporation of more comprehensive datasets, and the integration of real-time adaptive mechanisms. Additionally, the study can be extended by exploring optimized warranty strategies that leverage DT insights to dynamically adjust warranty terms based on real-time system health. Furthermore, maintenance policies can be refined by considering more sophisticated decision-

making frameworks that account for uncertainties in degradation processes and operational conditions. Such advancements would not only improve the accuracy and reliability of DT-PHM systems but also expand their utility in practical applications, ultimately contributing to more effective maintenance and warranty management strategies. The following aspects can be further enhanced in relation to DT technology, warranty strategies, and PdM policies.

1. DT

    a) Failure mechanisms, such as creep and combined cycle fatigue, can be incorporated to enhance the life reliability analysis by providing a more comprehensive understanding of degradation processes under complex operating conditions. This integration would allow for more accurate modeling of component behavior, enabling the development of advanced health management strategie. Additionally, accounting for these mechanisms can improve the precision of DT models, ensuring they reflect real-world conditions and support more effective decision-making for lifecycle management and reliability optimization.

    b) Alternative metamodels, such as physics-informed neural networks, can be utilized to enhance the physical interpretability of metamodels by embedding domain-specific knowledge and governing physical laws into the model architecture. This approach not only improves the accuracy of predictions but also ensures that the results are consistent with underlying physical principles, making them more reliable and explainable. Furthermore, the incorporation of

physics-informed neural networks can reduce the reliance on large datasets typically required for purely data-driven models, enabling robust performance even in data-scarce scenarios. By integrating these advanced metamodels, DT frameworks can achieve greater fidelity and provide actionable insights for complex engineering systems.

c) The application could potentially be extended to multi-state systems by incorporating Markov processes to model their behavior, enabling a more accurate representation of systems that transition between various performance or degradation states over time. This approach would allow for the prediction of state probabilities, facilitating a deeper understanding of system dynamics and failure mechanisms. By integrating Markov processes, the framework could capture the stochastic nature of multi-state systems, improve the estimation of RUL, and optimize maintenance strategies accordingly. Additionally, such an extension could support the development of more sophisticated reliability models, risk assessments, and decision-making tools tailored to complex, multi-component systems operating under uncertain conditions.

2. Warranty strategies

a) The developed warranty is designed from the manufacturer's perspective. A promising direction for future research would be to investigate warranty decisions from the customers' perspective, focusing on how factors such as perceived value, risk tolerance, usage patterns, and maintenance preferences

influence customer decision-making. This could involve exploring personalized or usage-based warranty schemes that align more closely with individual customer needs, thereby improving customer satisfaction and loyalty. Additionally, incorporating customer behavior modeling into warranty design could enable manufacturers to develop adaptive warranty strategies that dynamically adjust coverage based on real-time feedback or predicted customer actions. Such research would not only bridge the gap between manufacturer and customer priorities but also create opportunities for co-optimized warranty agreements that balance cost, reliability, and perceived value for both parties.

b) To improve the applicability of the warranty model, it can be extended to address 2D or even 3D warranty problems, which consider multiple dimensions such as time, usage, and environmental factors simultaneously. This extension would enable the warranty model to capture more complex scenarios, such as systems operating under varying usage intensities or environmental conditions, and provide a more realistic representation of failure risks. By incorporating these additional dimensions, the model can better account for interactions between different factors that influence reliability, allowing for the development of more tailored and flexible warranty policies. Furthermore, such a framework could support dynamic and condition-based warranty schemes, offering manufacturers and customers a more adaptive approach that aligns with real-world operating conditions and

maintenance practices.

c) Production and spare parts management represent promising areas for further research, particularly in optimizing inventory levels, reducing lead times, and improving cost efficiency while ensuring high system availability and reliability. Future studies could explore advanced techniques such as predictive analytics, machine learning, and DT technology to forecast demand for spare parts based on real-time operational data and failure trends. Additionally, integrating just-in-time manufacturing with maintenance strategies could streamline production processes and minimize waste. Research could also address the challenges of managing spare parts for complex systems with long lifecycles, focusing on the reuse, remanufacturing, and recycling of components to promote sustainability and circular economy principles. By advancing these areas, organizations could achieve more resilient supply chains and better align production with maintenance needs.

d) A discrete option-based model is developed for the joint optimization of product materials and warranty services. Future research could focus on developing a continuous decision-making framework for optimizing product materials and warranty services, enabling more precise and dynamic adjustments to account for varying customer requirements, operating conditions, and material performance over time. Such a framework could incorporate real-time data from IoT-enabled devices, advanced predictive models, and machine learning algorithms to continuously evaluate and

optimize the trade-offs between material costs, durability, and warranty coverage. Additionally, a continuous approach could better address uncertainties in demand, usage patterns, and environmental impacts, supporting more flexible and personalized warranty strategies. This would not only enhance the efficiency and sustainability of product lifecycle management but also strengthen customer satisfaction by aligning product and warranty offerings more closely with individual needs and preferences.

e) Exploring investment strategies for warranty reserve funds is worthwhile. While manufacturers collect reserves to cover warranty claims, these funds also have earning potential when strategically invested, impacting the total reserve required. Future research could develop financial models integrating reserve collection with investment decisions, balance liquidity needs with maximizing returns, and explore dynamic allocation strategies based on warranty trends and market conditions. Incorporating investments into warranty models could optimize reserve use, reduce financial burdens, and improve profitability.

3. PdM policies

a) Future research could investigate the incorporation of random thresholds or random improvement factors, which would allow for a more realistic representation of uncertainties in system performance, degradation, or maintenance effects, thereby enhancing the robustness of the model. Additionally, the scope of the study could be extended by integrating

production planning, enabling a holistic approach that considers not only reliability and maintenance but also the synchronization of production schedules, inventory levels, and resource allocation. This integration could explore the interplay between production constraints and maintenance requirements, optimizing decisions such as downtime scheduling, spare parts availability, and production capacity utilization. Moreover, combining production planning with random improvement factors could open new avenues for adaptive strategies, ensuring system performance and production efficiency in highly dynamic and uncertain environments. Such extensions would provide a comprehensive framework for managing complex systems across their entire lifecycle, bridging maintenance, production, and operational objectives.

b) The study presents a system-level maintenance considering economic dependencies. Future research can include structural dependencies, geometrical dependencies, or stochastic dependencies to capture a broader range of interactions and relationships between system components. Incorporating structural dependencies would allow for the modeling of physical connections or functional hierarchies wherein the failure or maintenance of one component directly affects others. Geometrical dependencies could account for spatial arrangements of components, where proximity or alignment impacts maintenance actions, wear rates, or failure propagation. Stochastic dependencies would enable a more sophisticated

analysis of random or probabilistic relationships between components, such as shared failure risks or environmental influences. By integrating these additional dependency types, future studies could provide a more comprehensive and realistic maintenance framework, improving decision-making for complex systems with interdependent components and dynamic operational conditions. This would enhance predictive maintenance strategies, optimize resource allocation, and reduce downtime, ultimately improving system reliability and cost efficiency.

c) Outcome-based contracts are promising future research, which align system providers' financial incentives with customers' desired performance outcomes, such as availability levels. These contracts can utilize DT tools to track performance for risk allocation. Integrating outcome-based contracts into PdM policies will incentivize reliability and improve customer satisfaction. Future work can focus on optimizing outcome-based contract terms and evaluating cost implications.

d) A critical direction for future research involves exploring challenges associated with limited maintenance resource availability, particularly in contexts such as offshore wind farms or remote facilities. Constraints like accessibility, weather, and resource availability can delay maintenance and increase failure risks. Future studies could develop predictive models for optimizing resource allocation, dynamic scheduling frameworks to prioritize tasks based on criticality and accessibility, and resilient strategies to ensure

system functionality during delays. Stochastic models, such as Markov processes, could simulate inaccessible conditions, predict delay impacts, and support preemptive actions. These advancements would enhance reliability and cost efficiency in systems with constrained maintenance conditions.

e) Future research should focus on analyzing multi-unit systems with varied unit types and configurations, moving beyond the current emphasis on series and parallel combinations of identical units. Real-world systems often include heterogeneous units or k-out-of-n configurations, which pose unique challenges. Thus, developing maintenance models for systems with distinct failure mechanisms and maintenance needs, along with optimizing k-out-of-n reliability strategies, is essential.

f) The application of distributed robust optimization to opportunistic maintenance represents a promising direction for future research. This approach enables decentralized decision-making, particularly advantageous for large-scale systems with multiple components or geographically dispersed assets. By addressing uncertainties in degradation, operational conditions, and resource availability, distributed robust optimization enhances the coordination of maintenance policies. Furthermore, it facilitates real-time decision-making by utilizing local information while ensuring global system performance, making it well-suited for dynamic environments.

Appendix

A.3.1 ACO Algorithm

In the ACO, the state updating rule is designed to guide the ants toward generating subsets with minimal prediction errors. Notably, this rule is applied only after all ants have completed the evaluation of their solutions. The state updating rule is defined as follows:

$$\tau_{jak}^{new} = (1-\rho)\tau_{jak}^{old} + \frac{Q_1}{E_1} + \frac{Q_2}{E_2}$$

In this context, τ_{jak} represents the pheromone level associated with the jth feature-wise Gaussian kernel in the kth GCNN carried by the ath ant. Here, ρ denotes the evaporation coefficient, while Q_1 and Q_2 signify the pheromone intensities for NRMSE and LLF, respectively. Similarly, E_1 and E_2 correspond to the cost functions of NRMSE and LLF. The primary goal of the state updating rule is to allocate a greater amount of pheromones to solution sets that yield fewer prediction errors. Once the pheromone levels for each feature-wise Gaussian kernel in each GCNN are determined by the ants, they are directed toward regions with higher pheromone concentrations using the state transition rule, as outlined below.

$$P_{jak} = \frac{\tau_{jak}}{\sum_{a=1}^{A} \tau_{jak}}$$

Here, P_{jak} represents the transition probability for the jth feature-wise Gaussian kernel in the kth GCNN carried by the ath ant.

The main steps of the ACO are listed as follows.

(1) Initially, certain parameters of the ACO are set. Subsequently, the grid interval for each feature-wise Gaussian kernel in each GCNN is divided into N grids, which are computed using the following formula.

$$h_{jk} = \frac{v_{jk}^{upper} - v_{jk}^{lower}}{N}, (j = 1,2, \dots, m; k = 1,2, \dots, s)$$

where v_{jk}^{upper} and v_{jk}^{lower} are upper and lower limits of the jth feature-wise Gaussian kernel in the kth GCNN, respectively.

(2) All ants randomly choose their starting points. Next, they determine the combination of Gaussian kernels in each GCNN based on the state transition rule. Once the LCF life is estimated using the selected Gaussian kernel parameters, the ants assess the pheromone levels and narrow down the range of the Gaussian kernel in each GCNN using the following approach.

$$v_{jk}^{lower} = v_{jk}^{upper} + (m_{jk} - \Delta)h_{jk}$$

$$v_{jk}^{lower} = v_{jk}^{upper} - (N - m_{jk} - \Delta)h_{jk}$$

(3) The above steps are repeated until the grid interval h_{jk} becomes smaller than the predefined precision ε. While a smaller ε can yield more precise results, it also increases the training time. To balance solution accuracy with computational efficiency, ε is set to 0.000001 during the training process. Ultimately, the optimal feature-wise Gaussian kernels for each GCNN are determined as follows.

$$v_{jk}^{*} = \frac{v_{jk}^{lower} + v_{jk}^{upper}}{2}, (j = 1,2, \dots, m; k = 1,2, \dots, s)$$

A.3.2 Optimization Solutions of the KCNN

Trainable parameters in the KCNN include Gaussian kernel vector $\theta^{(1)}$ and output weight vector $\theta^{(2)}$. Regarding $\theta^{(1)}$, it is optimized through the developed SACO-PSO algorithm. The pseudocode is presented in Algorithm 4.3. Regarding $\theta^{(2)}$, it can be solved by evaluating $\psi\big(c(i), c(j)\big) = \mathcal{U}(||c(j) - c(i)||)$ shown as follows.

$$\hat{\theta}^{(2)} = \Psi^{-1} g(X)$$

where

$$\boldsymbol{\Psi} = \begin{bmatrix} \Psi\big(\boldsymbol{c}(1), \boldsymbol{c}(1)\big) & \Psi\big(\boldsymbol{c}(1), \boldsymbol{c}(2)\big) & \cdots & \Psi(\boldsymbol{c}(1), \boldsymbol{c}(n)) \\ \Psi\big(\boldsymbol{c}(2), \boldsymbol{c}(1)\big) & \Psi\big(\boldsymbol{c}(2), \boldsymbol{c}(2)\big) & \cdots & \Psi(\boldsymbol{c}(2), \boldsymbol{c}(n)) \\ \vdots & & \ddots & \vdots \\ \Psi\big(\boldsymbol{c}(n), \boldsymbol{c}(1)\big) & \Psi\big(\boldsymbol{c}(n), \boldsymbol{c}(2)\big) & \cdots & \Psi(\boldsymbol{c}(n), \boldsymbol{c}(n)) \end{bmatrix}$$

Algorithm 1 SACO-PSO algorithm.

Input:
ACO: Number of grids $\mathcal{N}_{\text{grid}}$, number of ants $\mathcal{N}_{\text{ant}}$, number of iterations $\mathcal{N}_{\text{cmax}}$, eva poration coefficient ρ_{ara}, pherom-one intensity Q, grid coefficient Δ, and convergen ce threshold ε.
PSO: Number of particles $\mathcal{N}_{\text{particle}}$, number of maximum iterations $\mathcal{N}_{\text{MaxIter}}$, inerti a parameter ω, cognitive coefficient c_1, and social coefficient c_2.
Output:
Optimal Gaussian kernels of the KCNN $\hat{\theta}^{(1)}$.
1: Step 1:
2: While Maximum ranges of weights>ε, do
3: Place the ants on any grid randomly.
4: Input the weight $\theta^{(1)}$carried by each ant into the KCNN to estimate the output with Eq. (2).
5: Calculate the loss function of each an t based on carried weight with Eq. (4).
6: for $\mathcal{N}_{\text{c}} \leftarrow 1$ to $\mathcal{N}_{\text{cmax}}$ do
7: Calculate the pheromone level with τ^{new}=(1-$\mathcal{N}\rho_{\text{vma}}$) τ^{old}+Q/($\zeta_{\text{l}}\zeta_{\text{l}}$+($\zeta_{\text{k}}\zeta_{\text{k}}$).
8: Calculate the transition probability of each an t with $P_a = \tau_a / (\sum_{a=1}^{N_{\text{ant}}} \tau_a)$.
9: Find the ant with the maximum transition probability and narrow the search ran Ge.
10: Calculate the grid interval with the updated search range.
11: Step 2:
12: Initialize particles of PSO randomly within the optimal search range obtained fr om ACO.
13: While iteration <$\mathcal{N}_{\text{MaxIter}}$ do
14: Estimate the output of the KCNN as in line 4.
15: Calculate the loss function of particles as in line 5.

16: Find out the personal best position P* and global best position G* in the PSO.
17: Update the velocity of the i-th particle with
$v_i^{new} = \omega\, v_i^{old} + c1r1(P_i^* - X_i) + c2r2(G^* - X_i)$, where X_i is the position of i th particle.
18: Update the position of each particle with the new velocity.

A.3.3 Optimization Processes of the Kriging

To estimate the unknown parameters β_1, β_2, β_3, σ^2, and υ, a gradient descent optimizer is employed to maximize the MLE w.r.t. υ, as illustrated follows.

$$\mathcal{L}(\upsilon) = -N_c\, ln(\hat{\sigma}^2) - ln\,[det\,(r(\upsilon, X_c))]$$

where N_c is the number of candidates, X_c is the candidate set, and $\hat{\sigma}^2$ is the estimated variance determined by:

$$\hat{\sigma}^2 = \frac{1}{N_c}\big(\hat{g}(X_c) - \beta\mathcal{K}(X_c)\big)^T \\ r(\nu, X_c)^{-1}\big(\hat{g}(X_c) - \beta\mathcal{K}(X_c)\big)$$

where $\hat{g}(X_c)$ is the estimated vector corresponding to N_c candidates, $\beta = (\beta_1, \beta_2, \beta_3)$, and $\mathcal{K}(X_c)$ denotes the $N_c \times 3$ matrix of the basis function $k(X) = (1, X, X^T X)$. Based on the least square solution concerning $r(\upsilon, X_c)$, the coefficient β is given as:

$$\hat{\beta} = [\mathcal{K}(X)^T r(\nu, X)^{-1}\mathcal{K}(X)]^{-1} \\ \mathcal{K}(X)^T r(\nu, X)^{-1}\hat{g}(X)$$

The predictive uncertainty $n(X)$ at a new point X is calculated follows.

$$n(X) = \mathcal{R}(X)^T r(\nu, X)^{-1}(\hat{g}(X) - \nu\mathcal{K}(X))$$

where $\mathcal{R}(X)$ is the correlation vector between candidates and the new point denoted as

$$\mathcal{R}(X) = [r(\nu, X, X_c(1)), r(\nu, X, X_c(2)), \dots, \\ r\big(\nu, X, X_c(N_c)\big)]$$

A.3.4 Bayesian Neural Network Optimization with Variational Inference

Variational inference is introduced to optimize the Bayesian neural network model by approximating $p(\Theta; X, S)$. $V_\Phi(\Theta)$ is chosen as the variational distribution family. Then, $V_{\Phi^*}(\Theta)$ is obtained by minimizing the KL divergence between $V_\Phi(\Theta)$ and $p(\Theta; X, S)$ as formulated follows.

$$KL(V_\Phi(\Theta) \parallel p(\Theta; X, S)) = \int V_\Phi(\Theta)\, log\, \frac{V_\Phi(\Theta)}{p(\Theta; X, S)} d\Theta$$

Then, we have

$$\begin{aligned} KL\,(V_\Phi(\Theta) \parallel p(\Theta; X, S)) \quad &= \int V_\Phi(\Theta) log\, \frac{V_\Phi(\Theta)}{p(\Theta)} d\Theta \\ &- \sum_{i=1}^{n} \int V_\Phi(\Theta)\, log\, (l(s(i); \hat{g}(x(i); \Theta)))d\Theta \\ &+ log \left(\int p(\Theta) \prod_{i=1}^{n} l(s(i); \hat{g}(x(i); \Theta))d\Theta \right) \end{aligned}$$

The first term is the KL divergence between $V_\Phi(\Theta)$ and $p(\Theta)$ as KL $(V_\Phi(\Theta) \parallel p(\Theta))$, the second term is referred to the expected log likelihood w.r.t. $V_\Phi(\Theta)$, and the last term can be replaced by a constant since it does not depend on Φ. To maximize $\sum_{i=1}^{n} \int V_\Phi(\Theta) \log (l(s(i); \hat{g}(x(i); \Theta)))d\Theta$, a data subsampling strategy and a reparameterization trick are adopted. The data subsampling strategy is used to overcome the huge computation over all n samples through an approximation $\mathrm{E}[\widehat{\Omega}(\Phi, \varepsilon, b)]$. Here, b is a random subset of $\{1,2, \ldots, n\}$ with size n_b. $\widehat{\Omega}$ is given as:

$$\begin{aligned} \hat{\Omega}(\Phi, \varepsilon, b) \quad &= \int V_\Phi(\Theta)\, log\, \frac{V_\Phi(\Theta)}{p(\Theta)} d\Theta \\ &- \frac{n}{n_b} \sum_{i \in b} log\, \left(l\big(s(i); \hat{g}_{h(\varepsilon, \Phi)}(x(i))\big) \right) \end{aligned}$$

A stochastic optimizer can be used to obtain an optimum Φ^* for minimizing $\hat{\Omega}(\Phi, \varepsilon, b)$ and thus $\mathrm{KL}\left(V_\Phi(\Theta) \parallel p(\Theta; X, S)\right)$. By using $V_{\Phi^*}(\Theta)$ to approximate $p(\Theta; X, S)$, the prediction of $\hat{s}$ given $\hat{x}$ through the learned Bayesian BiLSTM can be implemented as:

$$p(\hat{s}; \hat{x}, X, S) = \frac{1}{n_k} \sum_{k=1}^{n_k} l\left(s; \hat{g}(\hat{x}; \Theta_k^*)\right)$$

where $\Theta_k^* \sim V_{\Phi^*}(\Theta)$.

A.4.1 Proof of Proposition 4.1

Given δ, if $w_1 < w_2 < (\left\lfloor \frac{w_1}{\delta} \right\rfloor + 1)\delta$, then

$$\begin{aligned} \bar{\Lambda}(w_2) - \bar{\Lambda}(w_1) &= \int_{\lfloor w_1/\delta \rfloor \delta}^{w_2} \tilde{\lambda}_\tau(u) du - \int_{\lfloor w_1/\delta \rfloor \delta}^{w_1} \tilde{\lambda}_\tau(u) du \\ &= \int_{w_1}^{w_2} \tilde{\lambda}_\tau(u) du > 0 \end{aligned}$$

If $w_2 \geq (\lfloor w_1/\delta \rfloor + 1)\delta$, then

$$\begin{aligned} \bar{\Lambda}(w_2) - \bar{\Lambda}(w_1) &= \int_{\lfloor w_1/\delta \rfloor \delta}^{(\lfloor w_1/\delta \rfloor + 1)\delta} \tilde{\lambda}(u) du + \int_{(\lfloor w_1/\delta \rfloor + 1)\delta}^{w_2} \tilde{\lambda}(u) du - \int_{\lfloor w_1/\delta \rfloor \delta}^{w_1} \tilde{\lambda}(u) du \\ &= \int_{w_1}^{(\lfloor w_1/\delta \rfloor + 1)\delta} \tilde{\lambda}(u) du + \int_{(\lfloor w_1/\delta \rfloor + 1)\delta}^{w_2} \tilde{\lambda}(u) du \end{aligned}$$

As $w_2 \geq (\lfloor w_1/\delta \rfloor + 1)\delta > w_1$, $\overline{\Lambda}(w_2) - \overline{\Lambda}(w_1) > 0$. The proof is completed.

A.4.2 Proof of Proposition 4.2

To validate the optimality conditions of p_m^*, it is essential to examine the continuity and differentiability of J_m w.r.t. p_m given w_m and δ_m. It is straightforward to observe that for $\forall p_m > 0$

$$\begin{cases} \lim\limits_{x \to p_m^-} J_m(x) = \lim\limits_{x \to p_m^+} J_m(x) \\ \lim\limits_{x \to p_m^-} \dfrac{\partial J_m(x)}{\partial x} = \lim\limits_{x \to p_m^+} \dfrac{\partial J_m(x)}{\partial x} \end{cases}$$

The necessary condition for the existence of p_m^* is given as $\frac{\partial J_m}{\partial p_m} = 0$. Given w_m and δ_m, by differentiating Eq. (4.9) w.r.t. p_m, the first-order derivative $\frac{\partial J_m}{\partial p_m}$ can be expressed as follows.

$$\frac{\partial J_m}{\partial p_m} = A_{mp}(p_m)^{-\alpha_1 - 1}\left[(1-\alpha_1)p_m + \alpha_1\left(\bar{C}_{mbi} + \bar{C}_w(w_m, \delta_m)\right)\right]$$

where $A_{\mathrm{mp}} = \kappa_1(w_m + \kappa_2)^{\alpha_2} \int_0^{l_c} G(Q(t))dt$

and the second-order derivative of p_m is

$$\frac{\partial^2 J_m}{\partial (p_m)^2} = -\alpha_1 A_{mp}\left[(1-\alpha_1)p_m + (\alpha_1+1)\left(\bar{C}_{mbi} + \bar{C}_w(w_m, \delta_m)\right)\right]$$

To validate the uniqueness of p_m^*, it is necessary to verify that $\frac{\partial^2 J_m}{\partial (p_m)^2} < 0$, i.e., $p_m < \frac{(\alpha_1+1)\left(\bar{C}_{mbi}+\bar{C}_w(w_m,\delta_m)\right)}{\alpha_1 - 1}$. Let $\frac{\partial J_m}{\partial p_m} = 0$, we can obtain $p_m^* = \frac{\alpha_1\left(\bar{C}_{\mathrm{mbi}}+\bar{C}_w(w_m,\delta_m)\right)}{\alpha_1 - 1}$. As $p_m^* < \frac{(\alpha_1+1)\left(\bar{C}_{mbi}+\bar{C}_w(w_m,\delta_m)\right)}{\alpha_1 - 1}$, the uniqueness of p_m^* can be validated.

A.5.1 Proof of Proposition 5.1

Given operational scenarios X, the necessary condition for minimizing $E[C_{cr}^{(t)}]$ is $\mathcal{J}\left(\tau_m^{(t)}\right) = \frac{\partial E[C_{cr}^{(t)}]}{\partial \tau_m^{(t)}} = 0$. To demonstrate the existence of a unique solution $\tau_m^{(t)}$ that satisfies the equality, it is necessary to prove that $\mathcal{J}\left(\tau_m^{(t)}\right)$ is continuous and strictly increasing to infinity, which can be proved with $\frac{\partial \mathcal{J}\left(\tau_m^{(t)}\right)}{\partial \tau_m^{(t)}} > 0$, i.e.,

$$\frac{\partial \mathcal{J}\left(\tau_m^{(t)}\right)}{\partial \tau_m^{(t)}} = \frac{\partial \lambda^{(t)}\left(\tau_m^{(t)}\right)}{\partial \tau_m^{(t)}} \int_0^{\tau_m^{(t)}} R^{(t)}(u)du$$

As the hazard rate $\lambda^{(t)}\left(\tau_m^{(t)}\right)$ is an increasing function of $\tau_m^{(t)}$, i.e., $\frac{\partial \lambda^{(t)}\left(\tau_m^{(t)}\right)}{\partial \tau_m^{(t)}} > 0$.

Thus, $\frac{\partial \mathcal{J}\left(\tau_m^{(t)}\right)}{\partial \tau_m^{(t)}} > 0$.

A.5.2 Proof of Proposition 5.2

Based on Proposition 6.1, with Eq. (6.6),

$$(C_F - C_R)\lambda^{(t)}\left(X_{0:t}; \tau_m^{(t)*}\right) = E\left[C_{cr}^{(t)}\left(X_{0:t}; \tau_m^{(t)*}\right)\right]$$

The above equation indicates that if $\lambda^{(t)}(X_{0:t}^1; \tau_m^{(t)*}) \leq \lambda^{(t)}(X_{0:t}^2; \tau_m^{(t)*})$ and consequently $\mathrm{E}\left[C_{cr}^{(t)}\right]\left(X_{0:t}^1; \tau_m^{(t)*}\right) \leq \mathrm{E}\left[C_{cr}^{(t)}\right]\left(X_{0:t}^2; \tau_m^{(t)*}\right)$.

A.5.3 Proof of Proposition 5.3

As the hazard rate at time t $\lambda^{(t)}(\tau)$ is an increasing function of τ, i.e., $\frac{\partial \lambda^{(t)}(\tau)}{\partial \tau} > 0$, we have

$$\frac{\partial \mathcal{J}\left(\tau_m^{(t)}\right)}{\partial \tau_m^{(t)}} = \tau_m^{(t)} \cdot \frac{\partial \lambda^{(t)}\left(\tau_m^{(t)}\right)}{\partial \tau_m^{(t)}}$$

First, it is evident that $\mathcal{J}\left(\tau_m^{(t)}\right)$ is continuous on $[0, +\infty)$. Let $\tau_{m,1}^{(t)} < \tau_{m,2}^{(t)} < \cdots < \tau_{m,i}^{(t)}$ and $lim_{i\to\infty}\lambda^{(t)}\left(\tau_{m,i}^{(t)}\right) = \infty$. Such a sequence exists as $\lambda^{(t)}(\tau)$ is strictly increasing. Then, it is evident that

$$\begin{aligned}
\mathcal{J}\left(\tau_{m,i}^{(t)}\right) &= \lambda^{(t)}\left(\tau_{m,i}^{(t)}\right) \cdot \tau_{m,i}^{(t)} - \int_0^{\tau_{m,i}^{(t)}} \lambda^{(t)}(u)du \\
&> \lambda^{(t)}\left(\tau_{m,1}^{(t)}\right) \cdot \tau_{m,1}^{(t)} - \int_0^{\tau_{m,1}^{(t)}} \lambda^{(t)}(u)du \to \infty \\
&as\ i \to \infty
\end{aligned}$$

Thus, $\mathcal{J}\left(\tau_m^{(t)}\right)$ is increasing w.r.t. $\tau_m^{(t)}$ and tends to ∞ as $\tau_m^{(t)} \to \infty$. Therefore, $\tau_m^{(t)*}$ is unique and finite when $\lim_{\tau_m^{(t)} \to \tau_{m,u}^{(t)}} \mathcal{J}\left(\tau_m^{(t)}\right) < \frac{C_R}{C_F + C_M}$ is met. Otherwise, $\mathrm{infE}\left[C_{\mathrm{cr}}^{(t)}\left(\tau_m^{(t)}\right)\right] = \lim_{\tau_m^{(t)} \to \tau_{m,u}^{(t)}} \mathrm{E}\left[C_{\mathrm{cr}}^{(t)}\left(\tau_m^{(t)}\right)\right]$.

A.6.1 Proof of Proposition 6.1

Assume that $Y(\tau)$ is the degradation level at time τ, estimated with Miner's law, with two thresholds $T_{m,1}$ and $T_{m,2}$, where $0 < T_{m,1} < T_{m,2} < 1$. Obviously, the inequality $\Pr\{Y(\tau) < T_{m,1}\} < \Pr\{Y(\tau) < T_{m,2}\}$ holds for the same τ. This implies that the reliability $R(\tau; T_{m,1}, H)$ w.r.t. $T_{m,1}$ is less than the reliability $R(\tau; T_{m,2}, H)$ w.r.t. $T_{m,2}$. Thus, the reliability function $R(\tau; T_m, H)$ is increasing as T_m increases. Thus, given a fixed duration δ_{lc}, the expected number of failures $\mathrm{E}[\Lambda]$ decreases as T_m increases. The proof is completed.

A.6.2 Derivation of Instantaneous Point Availability

Assume U is the duration of uptime with failure PDF $\delta_u(t)$ and D is the duration of downtime with repair PDF $\delta_d(t)$, then the PDF of the renewal process $r(t)$ is given with

$$r(\tau) = \int_0^t \delta_u(u)\delta_d(t-u)du$$

Taking Laplace transform, we have

$$r^*(s) = \delta_u^*(s)\delta_d^*(s)$$

Given $r_e^*(s) = \frac{r^*(s)}{1-r^*(s)}$, then we can derive

$$r_e^*(s) = \frac{\delta_u^*(s)\delta_d^*(s)}{1-\delta_u^*(s)\delta_d^*(s)}$$

Taking $r_e^*(s)$ into $A^*(s) = R^*(s)[1 + r_e^*(s)]$, we can obtain

$$
\begin{aligned}
A^*(s) &= R^*(s)\left[1+\frac{\delta_u^*(s)\delta_d^*(s)}{1-\delta_u^*(s)\delta_d^*(s)}\right] \\
&= \frac{R^*(s)}{1-\delta_u^*(s)\delta_d^*(s)} = \frac{\frac{1}{s}-\mathcal{L}\left[\int_0^\tau \delta_u(d)du\right]}{1-\delta_u^*(s)\delta_d^*(s)} \\
&= \frac{1-\delta_u^*(s)}{s[1-\delta_u^*(s)\delta_d^*(s)]}
\end{aligned}
$$

Thus,

$$A(t) = \mathcal{L}^{-1}(A^*(s))$$

A.7.1 Proof of Proposition 7.1

We assume that $Y_{i,j}$ denote the jth shock amplitude of the ith unit in a Compound Homogeneous Poisson Process with two spatial thresholds, H_1 and H_2, where $H_1 < H_2$. The inequality $\Pr\{Y_{i,j} < H_1\} < \Pr\{Y_{i,j} < H_2\}$ holds for the same t, indicating that $\bar{F}_i(t|H_1)$ is less than $\bar{F}_i(t|H_2)$. Thus, the survival probability $\bar{F}_i(t)$ increases as H increases. The relationship between $\bar{F}_i(t)$ and δ can also be analyzed in a similar manner.

A.7.2 Proof of Proposition 7.2

Denote $H(t) = \int_0^t h(u)du$. Since $\frac{\partial H(t)}{\partial t} = h(t) > 0$ and $\frac{\partial^2 H(t)}{\partial t^2} = h'(t) > 0$, we can conclude that $H(t)$ is a convex function in t. Based on the Jensen inequality, we have:

$$H(t_1) + H(t_3) > 2H(t_2)(t_1 < t_2 < t_3)$$

Since the cumulative hazard rate $H(\cdot)$ between two consecutive inspections can be written as

$$\int_{k\Delta t}^{(k+1)\Delta t} h(u)du = H((k+1)\Delta t) - H(k\Delta t)$$

Readily based on Jensen inequality, we can obtain

$$\int_{k\Delta t}^{(k+1)\Delta t} h(u)du = H((k+1)\Delta t) - H(k\Delta t)$$
$$> H(k\Delta t) - H((k-1)\Delta t) = \int_{(k-1)\Delta t}^{k\Delta t} h(u)du$$

The proof is completed.

References

[1] M. Pecht and M. Kang, *Prognostics and Health Management of Electronics: Fundamentals, Machine Learning, and the Internet of Things*, Second edition, Newark: Wiley, 2018.

[2] Z. S. Chen, Y. M. Yang, and Zheng Hu, "A technical framework and roadmap of embedded diagnostics and prognostics for complex mechanical systems in prognostics and health management systems," *IEEE Transactions on Reliability*, vol. 61, no. 2, pp. 314-322, 2012.

[3] Y. Hu, X. Miao, Y. Si, E. Pan, and E. Zio, "Prognostics and health management: A review from the perspectives of design, development and decision," *Reliability Engineering & System Safety*, vol. 217, pp. 108063-, 2022.

[4] G. Smith, J. B. Schroeder, S. Navarro, and D. Haldeman, "Development of a prognostics and health management capability for the Joint Strike Fighter," in *1997 IEEE Autotestcon Proceedings AUTOTESTCON '97. IEEE Systems Readiness Technology Conference. Systems Readiness Supporting Global Needs and Awareness in the 21st Century*, IEEE, 1997, pp. 676-682.

[5] A. Hess, G. Calvello, and T. Dabney, "PHM a key enabler for the JSF autonomic logistics support concept," in *2004 IEEE Aerospace Conference Proceedings (IEEE Cat. No.04TH8720)*, IEEE, 2004, pp. 3543-3550 Vol.6.

[6] N. Omri, Z. Al Masry, N. Mairot, S. Giampiccolo, and N. Zerhouni, "Industrial data management strategy towards an SME-oriented PHM," *Journal of Manufacturing Systems*, vol. 56, pp. 23-36, 2020.

[7] K. Daeil, M. R. Hodkiewicz, J. Fan, T. Shibutani, and M. Pecht, "IoT-based prognostics and systems health management for industrial applications," *IEEE Access*, vol. 4, pp. 3659–3670, 2016.

[8] X. Lei and P. A. Sandborn, “Maintenance scheduling based on remaining useful life predictions for wind farms managed using power purchase agreements,” *Renewable Energy*, vol. 116, pp. 188-198, 2018.

[9] M.-H. Chang, P. Sandborn, M. Pecht, W. K. C. Yung, and W. Wang, “A return on investment analysis of applying health monitoring to LED lighting systems,” *Microelectronics Reliability*, vol. 55, no. 3–4, pp. 527-537, 2015.

[10] K. Feldman, P. Sandborn, and T. Jazouli, “The analysis of return on investment for PHM applied to electronic systems,” in *2008 International Conference on Prognostics and Health Management*, IEEE, 2008, pp. 1-9.

[11] M. Grieves, “Digital twin: Manufacturing excellence through virtual factory replication,” White paper. Melbourne, FL: Florida Institute of Technology; 2014.

[12] M. Shafto, M. Conroy, R. Doyle, E. Glaessgen, C. Kemp, J. LeMoigne, *et al.*, “Modeling, simulation, information technology & processing roadmap,” National Aeronautics and Space Administration; 2012. Available: https://www.nasa.gov/sites/default/files/501321main_TA11-ID_rev4_NRC-wTASR.pdf, Accessed: 2019- 09-24.

[13] F. Tao, J. Cheng, Q. Qi, M. Zhang, H. Zhang, and F. Sui, “Digital twin-driven product design, manufacturing and service with big data,” *International Journal of Advanced Manufacturing Technology*, vol. 94, no. 9-12, pp. 3563-3576, 2018.

[14] V. Zaccaria, M. Stenfelt, I. Aslanidou, and K. Kyprianidis, “Fleet monitoring and diagnostics framework based on digital twin of aero-engines,” in *Proceedings of the ASME Turbo Expo*, 2018.

[15] Dassault Systems. Meet Virtual Singapore, the city’s 3D digital twin. Available: https://govinsider.asia/digital-gov/meet-virtual-singapore-citys-3d-digital-twin /Accessed: 2019-09-24. 2019.

[16] 5 Trends Drive the Gartner Hype Cycle for Emerging Technologies. Available: 〈https://www.gartner.com/smarterwithgartner/5-trends-drive-the-gartner-hype-cycle-for-emerging-technologies-2020.

[17] Q. Qi *et al.*, "Enabling technologies and tools for digital twin," *Journal of Manufacturing Systems*, vol. 58, pp. 3-21, 2021.

[18] E. H. Glaessgen and D. S. Stargel, "The digital twin paradigm for future NASA and U.S. air force vehicles," Langley Research Center, 2012.

[19] GE Renewable Energy. Digital Wind Farm-The Next Evolution of Wind Energy. Available: https://www.ge.com/content/dam/gepower-renewables/global/en_US/downloads/brochures/digital-wind-farm-solutions-gea31821b-r2.pdf, Accessed: 2019-09-24. 2019.

[20] The Living Heart Project. 2019. Available: https://www.3ds.com/products-services/simulia/solutions/life-sciences/the-living-heart-project/, Accessed: 2019-09-24.

[21] F. Tao, M. Zhang, Y. Liu, and A. Y. C. Nee, "Digital twin driven prognostics and health management for complex equipment," *CIRP Annals*, vol. 67, no. 1, pp. 169-172, 2018.

[22] T. Zhu, Y. Ran, X. Zhou, and Y. Wen, "A survey of predictive maintenance: Systems, purposes and approaches," *arXiv.org*, 2024.

[23] F. Tao and Q. Qi, "Make more digital twins," *Nature (London)*, vol. 573, no. 7775, pp. 490-491, 2019.

[24] F. Tao, Q. Qi, L. Wang, and A. Y. C. Nee, "Digital twins and cyber-physical systems toward smart manufacturing and Industry 4.0: Correlation and comparison," *Engineering (Beijing, China)*, vol. 5, no. 4, pp. 653-661, 2019.

[25] C. Ling, W. Kuo, and M. Xie, "An overview of adaptive-surrogate-model-assisted methods for reliability-based design optimization," *IEEE Transactions on Reliability*, vol. 72, no. 3, pp. 1-22, 2023.

[26] Z. Liu, Z. He, L. Tu, X. Liu, H. Liu, and J. Liang, "A fatigue reliability assessment approach for wind turbine blades based on continuous time Bayesian network and FEA," *Quality and Reliability Engineering International*, vol. 39, no. 5, pp. 1603-1621, 2023.

[27] Y. Huang, J.-G. Zhang, L.-K. Song, X.-Q. Li, and G.-C. Bai, "A unified reliability evaluation framework for aircraft turbine rotor considering multi-site failure correlation," *Structural and Multidisciplinary Optimization*, vol. 66, no. 7, pp. 171-, 2023.

[28] F. Tao *et al.*, "Digital twin-driven product design framework," *International Journal of Production Research*, vol. 57, no. 12, pp. 3935-3953, 2019.

[29] S. Adams, M. Malinowski, G. Heddy, B. Choo, and P. A. Beling, "The WEAR methodology for prognostics and health management implementation in manufacturing," *Journal of Manufacturing Systems*, vol. 45, pp. 82-96, 2017.

[30] L. Shang, S. Si, S. Sun, and T. Jin, "Optimal warranty design and post-warranty maintenance for products subject to stochastic degradation," *IISE Transactions*, vol. 50, no. 10, pp. 913-927, 2018.

[31] J. Zeng and Z. Liang, "A deep Gaussian process approach for predictive maintenance," *IEEE Transactions on Reliability*, vol. 72, no. 3, pp. 1-18, 2023.

[32] J. Lee and M. Mitici, "Deep reinforcement learning for predictive aircraft maintenance using probabilistic Remaining-Useful-Life prognostics," *Reliability Engineering & System Safety*, vol. 230, pp. 108908-, 2023.

[33] C. Chen, G. Tao, J. Shi, M. Shen, and Z. H. Zhu, "A lithium-ion battery degradation prediction model with uncertainty quantification for its predictive maintenance," *IEEE Transactions on Industrial Electronics*, vol. 71, no. 4, pp. 3650-3659, 2024.

[34] Y. Wei, T. Hu, T. Zhou, Y. Ye, and W. Luo, "Consistency retention method for CNC machine tool digital twin model," *Journal of Manufacturing Systems*, vol. 58, pp. 313-322, 2021.

[35] M. Lermer and C Reich, "Creation of digital twins by combining fuzzy rules with artificial neural networks," in *Proceedings of the 2019 45th Annual Conference of the IEEE Industrial Electronics Society*.

[36] S. S. Afshari, F. Enayatollahi, X. Xu, and X. Liang, “Machine learning-based methods in structural reliability analysis: A review,” *Reliability Engineering & System Safety*, vol. 219, pp. 108223-, 2022.

[37] Q. Yao, M. Zhang, Y. Liu, and Q. Guo, “Life reliability assessment of twin-web disk using the active learning kriging model,” *Structural and Multidisciplinary Optimization*, vol. 61, no. 3, pp. 1229-1251, 2020.

[38] H. Zhan, H. Liu, and N.-C. Xiao, “Structural expected lifetime estimation for systems with multiple failure modes based on adaptive learning kriging models,” *IEEE Transactions on Reliability*, vol. 73, no. 1, pp. 549-559, 2024.

[39] X. Lai, L. Yang, X. He, Y. Pang, X. Song, and W. Sun, “Digital twin-based structural health monitoring by combining measurement and computational data: An aircraft wing example,” *Journal of Manufacturing Systems*, vol. 69, pp. 76-90, 2023.

[40] S.-P. Zhu, H.-Z. Huang, W. Peng, H.-K. Wang, and S. Mahadevan, “Probabilistic Physics of Failure-based framework for fatigue life prediction of aircraft gas turbine discs under uncertainty,” *Reliability Engineering & System Safety*, vol. 146, pp. 1-12, 2016.

[41] W. Du, S. Li, and Y. Luo, “A novel method for structure’s fatigue life scatter simulation under material variability,” *International Journal of Fatigue*, vol. 149, pp. 106296-, 2021.

[42] R. Teixeira, M. Nogal, A. O’Connor, J. Nichols, and A. Dumas, “Stress-cycle fatigue design with Kriging applied to offshore wind turbines,” *International Journal of Fatigue*, vol. 125, pp. 454-467, 2019.

[43] Y. Dong, A. P. Teixeira, and C. Guedes Soares, “Application of adaptive surrogate models in time-variant fatigue reliability assessment of welded joints with surface cracks,” *Reliability Engineering & System Safety*, vol. 195, pp. 106730-, 2020.

[44] P. E. Leser *et al.*, “Probabilistic fatigue damage prognosis using surrogate models trained via three-dimensional finite element analysis,” *Structural Health Monitoring*, vol. 16, no. 3, pp. 291-308, 2017.

[45] L.-K. Song, G.-C. Bai, and C.-W. Fei, “Probabilistic LCF life assessment for turbine discs with DC strategy-based wavelet neural network regression,” *International Journal of Fatigue*, vol. 119, pp. 204-219, 2019.

[46] H. Bai, L. Shi, Y. Aoues, C. Huang, and D. Lemosse, “Estimation of probability distribution of long-term fatigue damage on wind turbine tower using residual neural network,” *Mechanical Systems and Signal Processing*, vol. 190, pp. 110101-, 2023.

[47] G. Jing *et al.*, “Research on fatigue reliability assessment of engine cylinder head based on neural network,” *International Journal of Fatigue*, vol. 175, pp. 107800-, 2023.

[48] D. D’Urso, F. Chiacchio, S. Cavalieri, S. Gambadoro, and S. M. Khodayee, “Predictive maintenance of standalone steel industrial components powered by a dynamic reliability digital twin model with artificial intelligence,” *Reliability Engineering & System Safety*, vol. 243, pp. 109859-, 2024.

[49] W. Luo, T. Hu, Y. Ye, C. Zhang, and Y. Wei, “A hybrid predictive maintenance approach for CNC machine tool driven by Digital Twin,” *Robotics and Computer-Integrated Manufacturing*, vol. 65, pp. 101974-, 2020.

[50] M. Xiong, H. Wang, Q. Fu, and Y. Xu, “Digital twin–driven aero-engine intelligent predictive maintenance,” *International Journal of Advanced Manufacturing Technology*, vol. 114, no. 11-12, pp. 3751–3761, 2021.

[51] F. Tao, M. Zhang, Y. Liu, and A. Y. C. Nee, “Digital twin driven prognostics and health management for complex equipment,” *CIRP Annals*, vol. 67, no. 1, pp. 169-172, 2018.

[52] M. Shafiee, M. Finkelstein, and C. Bérenguer, “An opportunistic condition-based maintenance policy for offshore wind turbine blades subjected to degradation and environmental shocks,” *Reliability Engineering & System Safety*, vol. 142, pp. 463–471, 2015.

[53] Z.-S. Ye and M. Xie, “Stochastic modelling and analysis of degradation for highly reliable products,” *Applied Stochastic Models in Business and Industry*, vol. 31, no. 1, pp. 16–32, 2015.

[54] D. Tang, J. Yu, X. Chen, and V. Makis, “An optimal condition-based maintenance policy for a degrading system subject to the competing risks of soft and hard failure,” *Computers & Industrial Engineering*, vol. 83, pp. 100–110, 2015.

[55] E. A. Elsayed, *Reliability Engineering*, 2nd ed. Hoboken, N.J: John Wiley & Sons, Inc., 2012.

[56] K. T. Huynh, I. T. Castro, A. Barros, and C. Bérenguer, “Modeling age-based maintenance strategies with minimal repairs for systems subject to competing failure modes due to degradation and shocks,” *European Journal of Operational Research*, vol. 218, no. 1, pp. 140–151, 2012,

[57] N. C. Caballé, I. T. Castro, C. J. Pérez, and J. M. Lanza-Gutiérrez, “A condition-based maintenance of a dependent degradation-threshold-shock model in a system with multiple degradation processes,” *Reliability Engineering & System Safety*, vol. 134, pp. 98–109, 2015.

[58] Q. Wang, Z. He, S. Lin, and Z. Li, “Failure modeling and maintenance decision for GIS equipment subject to degradation and shocks,” *IEEE Transactions on Power Delivery*, vol. 32, no. 2, pp. 1079–1088, 2017.

[59] Q. Wang, Z. He, S. Lin, and Y. Liu, “Availability and maintenance modeling for GIS equipment served in high-speed railway under incomplete maintenance,” *IEEE Transactions on Power Delivery*, vol. 33, no. 5, pp. 2143–2151, 2018.

[60] W. R. Blischke and D. N. P. Murthy, “Product warranty management — I: A taxonomy for warranty policies,” *European Journal of Operational Research*, vol. 62, no. 2, pp. 127–148, 1992.

[61] W. Xie and Z.-S. Ye, “Aggregate discounted warranty cost forecast for a new product considering stochastic sales,” *IEEE Transactions on Reliability*, vol. 65, no. 1, pp. 486–497, 2016.

[62] M. Park, K. M. Jung, and D. H. Park, "A generalized age replacement policy for systems under renewing repair-replacement warranty," *IEEE Transactions on Reliability*, vol. 65, no. 2, pp. 604–612, 2016.

[63] M. Shafiee and S. Chukova, "Maintenance models in warranty: A literature review," *European Journal of Operational Research*, vol. 229, no. 3, pp. 561–572, 2013.

[64] M. Luo and S. Wu, "A comprehensive analysis of warranty claims and optimal policies," *European Journal of Operational Research*, vol. 276, no. 1, pp. 144–159, 2019.

[65] K. F. Cheong, C. Zhang, and Y. Zhang, "Joint dynamic optimization of price and two-dimensional warranty policy," *IEEE Transactions on Engineering Management*, vol. 68, no. 2, pp. 528–547, 2021.

[66] Y.-S. Huang, J.-W. Ho, J.-W. Hung, and T.-L. (Bill) Tseng, "A customized warranty model by considering multi-usage levels for the leasing industry," *Reliability Engineering & System Safety*, vol. 215, pp. 107769-, 2021.

[67] M. Afsahi and M. Shafiee, "A stochastic simulation-optimization model for base-warranty and extended-warranty decision-making of under- and out-of-warranty products," *Reliability Engineering & System Safety*, vol. 197, pp. 106772–18, 2020.

[68] T. S. Glickman and P. D. Berger, "Optimal price and protection period decisions for a product under warranty," *Management Science*, vol. 22, no. 12, pp. 1381–1390, 1976.

[69] P.-C. Lin and L.-Y. Shue, "Application of optimal control theory to product pricing and warranty with free replacement under the influence of basic lifetime distributions," *Computers & Industrial Engineering*, vol. 48, no. 1, pp. 69–82, 2005.

[70] Z. Zhou, Y. Li, and K. Tang, "Dynamic pricing and warranty policies for products with fixed lifetime," *European Journal of Operational Research*, vol. 196, no. 3, pp. 940–948, 2009.

[71] H.-Z. Huang, Z.-J. Liu, and D. N. P. Murthy, "Optimal reliability, warranty and price for new products," *IIE Transactions*, vol. 39, no. 8, pp. 819–827, 2007.

[72] C.-C. Wu, C.-Y. Chou, and C. Huang, "Optimal price, warranty length and production rate for free replacement policy in the static demand market," *Omega (Oxford)*, vol. 37, no. 1, pp. 29–39, 2009.

[73] B. Kim and S. Park, "Optimal pricing, EOL (end of life) warranty, and spare parts manufacturing strategy amid product transition," *European Journal of Operational Research*, vol. 188, no. 3, pp. 723–745, 2008.

[74] Y.-S. Huang, C.-C. Fang, C.-M. Lu, and T.-L. (Bill) Tseng, "Optimal warranty policy for consumer electronics with dependent competing failure processes," *Reliability Engineering & System Safety*, vol. 222, pp. 108418-, 2022.

[75] X. Wang, X. Zhao, and B. Liu, "Design and pricing of extended warranty menus based on the multinomial logit choice model," *European Journal of Operational Research*, vol. 287, no. 1, pp. 237–250, 2020.

[76] Z. Zhang, S. He, Z. He, D. Wang, and F. Dong, "A systematic warranty-reliability-price decision model for two-dimensional warranted products with heterogeneous usage rates," *Computers & Industrial Engineering*, vol. 163, pp. 107820-, 2022.

[77] C.-C. Fang, "Optimal price and warranty decision for durable products in a competitive duopoly market," *Reliability Engineering & System Safety*, vol. 203, pp. 107068-, 2020.

[78] L. Shang, B. Liu, Q. Qiu, and L. Yang, "Three-dimensional warranty and post-warranty maintenance of products with monitored mission cycles," *Reliability Engineering & System Safety*, vol. 239, pp. 109506-, 2023.

[79] R. H. Yeh, K.-C. Kao, and W. L. Chang, "Optimal preventive maintenance policy for leased equipment using failure rate reduction," *Computers & Industrial Engineering*, vol. 57, no. 1, pp. 304–309, 2009.

[80] Y.-H. Chien, "The optimal preventive-maintenance policy for a NHPBP repairable system under free-repair warranty," *Reliability Engineering & System Safety*, vol. 188, pp. 444–453, 2019.

[81] Y.-S. Huang, C.-D. Huang, and J.-W. Ho, "A customized two-dimensional extended warranty with preventive maintenance," *European Journal of Operational Research*, vol. 257, no. 3, pp. 971–978, 2017.

[82] Y.-S. Huang, E. Chen, and J.-W. Ho, "Two-dimensional warranty with reliability-based preventive maintenance," *IEEE Transactions on Reliability*, vol. 62, no. 4, pp. 898–907, 2013.

[83] Y.-H. Chien, Z. G. Zhang, and X. Yin, "On optimal preventive-maintenance policy for generalized polya process repairable products under free-repair warranty," *European Journal of Operational Research*, vol. 279, no. 1, pp. 68–78, 2019.

[84] T. Nakagawa and S. Mizutani, "A summary of maintenance policies for a finite interval," *Reliability Engineering & System Safety*, vol. 94, no. 1, pp. 89–96, 2009.

[85] M. Park, K. M. Jung, and D. H. Park, "Optimal maintenance strategy under renewable warranty with repair time threshold," *Applied Mathematical Modelling*, vol. 43, pp. 498–508, 2017.

[86] M. Shafiee, M. Finkelstein, and M. J. Zuo, "Optimal burn-in and preventive maintenance warranty strategies with time-dependent maintenance costs," *IIE Transactions*, vol. 45, no. 9, pp. 1024–1033, 2013.

[87] C.-W. Yeh and C.-C. Fang, "Optimal decision for warranty with consideration of marketing and production capacity," *International Journal of Production Research*, vol. 53, no. 18, pp. 5456–5471, 2015.

[88] C. Su and X. Wang, "A two-stage preventive maintenance optimization model incorporating two-dimensional extended warranty," *Reliability Engineering & System Safety*, vol. 155, pp. 169–178, 2016.

[89] N. Tao and S. Zhang, "The optimal extended warranty length of durable-goods-based preventive maintenance behavior," *Systems Science & Control Engineering*, vol. 3, no. 1, pp. 472–477, 2015.

[90] N.-J. Hsu, S.-T. Tseng, and M.-W. Chen, "Adaptive warranty prediction for highly reliable products," *IEEE Transactions on Reliability*, vol. 64, no. 3, pp. 1057–1067, 2015.

[91] T. Xia, G. Si, D. Wang, E. Pan, and L. Xi, “Progressive opportunistic maintenance policies for service-outsourcing network with prognostic updating and dynamical optimization,” *IEEE Transactions on Reliability*, vol. 71, no. 3, pp. 1340–1354, 2022.

[92] D. S. Thomas, The costs and benefits of advanced maintenance in manufacturing. US Department of Commerce, National Institute of Standards and Technology; 2018.

[93] I. El-Thalji and J. P. Liyanage, “On the operation and maintenance practices of wind power asset: A status review and observations,” *Journal of Quality in Maintenance Engineering*, vol. 18, no. 3, pp. 232–266, 2012.

[94] B. Liu, Z. Liang, A. K. Parlikad, M. Xie, and W. Kuo, “Condition-based maintenance for systems with aging and cumulative damage based on proportional hazards model,” *Reliability Engineering & System Safety*, vol. 168, pp. 200–209, 2017.

[95] R. Ahmad and S. Kamaruddin, “An overview of time-based and condition-based maintenance in industrial application,” *Computers & Industrial Engineering*, vol. 63, no. 1, pp. 135–149, 2012.

[96] X. Jin, B. A. Weiss, D. Siegel, and J. Lee, “Present status and future growth of advanced maintenance technology and strategy in US manufacturing,” *International Journal of Prognostics and Health Management*, vol. 7, no. 3, pp. 012–012, 2016.

[97] J. Wang, L. Zhang, L. Duan, and R. X. Gao, “A new paradigm of cloud-based predictive maintenance for intelligent manufacturing,” *Journal of Intelligent Manufacturing*, vol. 28, no. 5, pp. 1125–1137, 2017.

[98] X. Bai, Z. An, Y. Hou, and Q. Ma, “Health assessment and management of wind turbine blade based on the fatigue test data,” *Microelectronics and Reliability*, vol. 75, pp. 205–214, 2017.

[99] A. Prajapati, J. Bechtel, and S. Ganesan, “Condition based maintenance: a survey,” *Journal of Quality in Maintenance Engineering*, vol. 18, no. 4, pp. 384–400, 2012.

[100] MIMOSA OSA-CBM. Available: http://www.mimosa.org/mimosa-osa-cbm/ [Online].

[101] Cost of data center. https://www.futurefacilities.com.

[102] M. Compare and E. Zio, “Predictive maintenance by risk sensitive particle filtering,” *IEEE Transactions on Reliability*, vol. 63, no. 1, pp. 134–143, 2014.

[103] O. Blancke *et al.*, “Predictive maintenance approach for complex equipment based on a failure mechanism propagation model,” *International Journal of Prognostics and Health Management*, vol. 10, no. 1, 2019.

[104] L. Zhuang, A. Xu, and X.-L. Wang, “A prognostic driven predictive maintenance framework based on Bayesian deep learning,” *Reliability Engineering & System Safety*, vol. 234, pp. 109181-, 2023.

[105] V. Negandhi *et al.*, IBM predictive maintenance and quality 2.0 technical overview. IBM Redbooks, 2015.

[106] IBM. [Online]. https://www.ibm.com/case-studies/downer.

[107] Siemens. [Online]. https://assets.new.siemens.com/siemens/assets.

[108] 'GE.' [Online]. https://www.ge.com/digital/sites/default/files.

[109] 'Delta.' [Online]. https://www.iata.org/contentassets.

[110] F. Mohammadhasani and A. Pirouzmand, “Multi-state risk-based maintenance analysis of redundant safety systems using the Markov model and fault tree method,” *Frontiers in Energy Research*, vol. 9, 2021.

[111] W. Liao, Y. Wang, and E. Pan, “Single-machine-based predictive maintenance model considering intelligent machinery prognostics,” *International Journal of Advanced Manufacturing Technology*, vol. 63, no. 1–4, pp. 51–63, 2012.

[112] Y. He, C. Gu, Z. Chen, and X. Han, “Integrated predictive maintenance strategy for manufacturing systems by combining quality control and mission

reliability analysis," *International Journal of Production Research*, vol. 55, no. 19, pp. 5841–5862, 2017.

[113] M. Zhang, Z. Ye, and M. Xie, "A condition-based maintenance strategy for heterogeneous populations," *Computers & Industrial Engineering*, vol. 77, pp. 103–114, 2014.

[114] K. T. P. Nguyen and K. Medjaher, "A new dynamic predictive maintenance framework using deep learning for failure prognostics," *Reliability Engineering & System Safety*, vol. 188, pp. 251–262, 2019.

[115] H. D. Shoorkand, M. Nourelfath, and A. Hajji, "A deep learning approach for integrated production planning and predictive maintenance," *International Journal of Production Research*, vol. 61, no. 23, pp. 7972–7991, 2023.

[116] M. Yildirim, N. Z. Gebraeel, and X. A. Sun, "Integrated predictive analytics and optimization for opportunistic maintenance and operations in wind farms," *IEEE Transactions on Power Systems*, vol. 32, no. 6, pp. 4319–4328, 2017.

[117] C. Huang and S. Bu, "Strategic electrical asset maintenance by reinforcement learning," in *12th IET International Conference on Advances in Power System Control, Operation and Management (APSCOM 2022)*, Hybrid Conference, Hong Kong, China, 2022, pp. 48-52.

[118] Y. Xiang, D. W. Coit, and Z. Zhu, "A multi-objective joint burn-in and imperfect condition-based maintenance model for degradation-based heterogeneous populations," *Quality and Reliability Engineering International*, vol. 32, no. 8, pp. 2739–2750, 2016.

[119] Z. Liang and A. K. Parlikad, "A condition-based maintenance model for assets with accelerated deterioration due to fault propagation," *IEEE Transactions on Reliability*, vol. 64, no. 3, pp. 972–982, 2015.

[120] Y. Xiang, D. W. Coit, and Z. Zhu, "A multi-objective joint burn-in and imperfect condition-based maintenance model for degradation-based heterogeneous populations," *Quality and Reliability Engineering International*, vol. 32, no. 8, pp. 2739–2750, 2016.

[121] L. Lin, B. Luo, and S. Zhong, “Multi-objective decision-making model based on CBM for an aircraft fleet with reliability constraint,” *International Journal of Production Research*, vol. 56, no. 14, pp. 4831–4848, 2018.

[122] T. Luan, M. Sun, B. You, H. Yao, and Q. Fu, “Life prediction of carrier-based aircraft replaceable units with time-varying drift and optimization strategy for imperfect maintenance,” *IEEE Transactions on Reliability*, vol. 73, no. 2, pp. 952–966, 2024.

[123] Q. Feng *et al.*, “Multi-level predictive maintenance of smart manufacturing systems driven by digital twin: A matheuristics approach,” *Journal of Manufacturing Systems*, vol. 68, pp. 443–454, 2023.

[124] F. Besnard and L. Bertling, “An approach for condition-based maintenance optimization applied to wind turbine blades,” *IEEE Transactions on Sustainable Energy*, vol. 1, no. 2, pp. 77–83, 2010.

[125] Z. Liang and A. K. Parlikad, “Predictive group maintenance for multi-system multi-component networks,” *Reliability Engineering & System Safety*, vol. 195, pp. 106704-, 2020.

[126] L. C. Thomas, “A survey of maintenance and replacement models for maintainability and reliability of multi-item systems,” *Reliability Engineering*, vol. 16, no. 4, pp. 297–309, 1986.

[127] D. I. Cho and M. Parlar, “A survey of maintenance models for multi-unit systems,” *European Journal of Operational Research*, vol. 51, no. 1, pp. 1–23, 1991.

[128] X. Yang, Y. He, R. Liao, Y. Cai, and W. Dai, “Mission reliability-centered opportunistic maintenance approach for multistate manufacturing systems,” *Reliability Engineering & System Safety*, vol. 241, pp. 109693-, 2024.

[129] P. Zhou and P. T. Yin, “An opportunistic condition-based maintenance strategy for offshore wind farm based on predictive analytics,” *Renewable & Sustainable Energy Reviews*, vol. 109, pp. 1–9, 2019.

[130] B. Iung, P. Do, E. Levrat, and A. Voisin, “Opportunistic maintenance based on multi-dependent components of manufacturing system,” *CIRP Annals*, vol. 65, no. 1, pp. 401–404, 2016.

[131] B. Einabadi, M. Mahmoodjanloo, A. Baboli, and E. Rother, “Dynamic predictive and preventive maintenance planning with failure risk and opportunistic grouping considerations: A case study in the automotive industry,” *Journal of Manufacturing Systems*, vol. 69, pp. 292–310, 2023.

[132] D.-H. Dinh, P. Do, and B. Iung, “Multi-level opportunistic predictive maintenance for multi-component systems with economic dependence and assembly/disassembly impacts,” *Reliability Engineering & System Safety*, vol. 217, pp. 108055-, 2022.

[133] X. Zhu, J. Wang, and D. W. Coit, “Joint optimization of spare part supply and opportunistic condition-based maintenance for onshore wind farms considering maintenance route,” *IEEE Transactions on Engineering Management*, vol. 71, pp. 1086–1102, 2024.

[134] Z. Tian, T. Jin, B. Wu, and F. Ding, “Condition based maintenance optimization for wind power generation systems under continuous monitoring,” *Renewable Energy*, vol. 36, no. 5, pp. 1502–1509, 2011.

[135] J. Sun, Z. Sun, C. Chen, C. Yan, T. Jin, and Y. Zhong, “Group maintenance strategy of CNC machine tools considering three kinds of maintenance dependence and its optimization,” *International Journal of Advanced Manufacturing Technology*, vol. 124, no. 11–12, pp. 3749–3760, 2023.

[136] H.-C. Vu, P. Do, and A. Barros, “A study on the impacts of maintenance duration on dynamic grouping modeling and optimization of multicomponent systems,” *IEEE Transactions on Reliability*, vol. 67, no. 3, pp. 1377–1392, 2018.

[137] B. Castanier, A. Grall, and C. Bérenguer, “A condition-based maintenance policy with non-periodic inspections for a two-unit series system,” *Reliability Engineering & System Safety*, vol. 87, no. 1, pp. 109–120, 2005.

[138] C. Zhang, W. Gao, T. Yang, and S. Guo, “Opportunistic maintenance strategy for wind turbines considering weather conditions and spare parts inventory management,” *Renewable Energy*, vol. 133, pp. 703–711, 2019.

[139] A. Erguido, A. Crespo Márquez, E. Castellano, and J. F. Gómez Fernández, “A dynamic opportunistic maintenance model to maximize energy-based availability while reducing the life cycle cost of wind farms,” *Renewable Energy*, vol. 114, pp. 843–856, 2017.

[140] M. Sedghi, B. Bergquist, E. Vanhatalo, and A. Migdalas, “Data-driven maintenance planning and scheduling based on predicted railway track condition,” *Quality and Reliability Engineering International*, vol. 38, no. 7, pp. 3689–3709, 2022.

[141] X. Zhao, Z. Lv, Z. He, and W. Wang, “Reliability and opportunistic maintenance for a series system with multi-stage accelerated damage in shock environments,” *Computers & Industrial Engineering*, vol. 137, pp. 106029-, 2019.

[142] X. Wang, X. Zhao, S. Wang, and L. Sun, “Reliability and maintenance for performance-balanced systems operating in a shock environment,” *Reliability Engineering & System Safety*, vol. 195, pp. 106705-, 2020.

[143] X. Lai *et al.*, “A scalable digital twin framework based on a novel adaptive ensemble surrogate model,” Journal of Mechanical Design, vol. 145, no. 2, 2023.

[144] J. Pan *et al.*, “Digital twin method for real-time stress prediction based on surrogate model,” in *Advanced Manufacturing and Automation XIII*, Singapore: Springer, 2024.

[145] X. Liu, X. Han, H. Wang, and G. Liu, “A modelling and updating approach of digital twin based on surrogate model to rapidly evaluate product performance,” *International Journal of Advanced Manufacturing Technology*, vol. 129, no. 11–12, pp. 5059–5074, 2023.

[146] D. Alaez, M. Prieto, J. Villadangos, and J. J. Astrain, “On constructing efficient UAV aerodynamic surrogate models for digital twins,” *IEEE Transactions on Industrial Informatics*, pp. 1–9, 2024.

[147] X. Lai, X. He, S. Wang, X. Wang, W. Sun, and X. Song, “Building a lightweight digital twin of a crane boom for structural safety monitoring based on a multi-fidelity surrogate model,” *Journal of Mechanical Design*, vol. 144, no. 6, 2022.

[148] X. Liu, X. Han, H. Wang, and G. Liu, “A modelling and updating approach of digital twin based on surrogate model to rapidly evaluate product performance,” *International Journal of Advanced Manufacturing Technology*, vol. 129, no. 11–12, pp. 5059–5074, 2023.

[149] A. Thelen *et al.*, “A comprehensive review of digital twin — part 1: modeling and twinning enabling technologies,” *Structural and Multidisciplinary Optimization*, vol. 65, no. 12, 2022.

[150] C. Guellec, C. Doudard, B. Levieil, L. Jian, A. Ezanno, and S. Calloch, “Parametric method for the assessment of fatigue damage for marine shaft lines,” *Marine Structures*, vol. 87, pp. 103325-, 2023.

[151] J. Ren, Y. Qie, H. Xie, F. Yin, and X. Cui, “Fatigue analysis of 75 kg/m-12 heavy-haul railway frog based on finite element simulation,” *Engineering Failure Analysis*, vol. 117, pp. 104799-, 2020.

[152] A. Niesiony, “Determination of fragments of multiaxial service loading strongly influencing the fatigue of machine components,” *Mechanical Systems and Signal Processing*, vol. 23, no. 8, pp. 2712–2721, 2009.

[153] C. Lu, J. Zhong, Y. Xue, L. Feng, and J. Zhang, “Ant colony system with sorting-based local search for coverage-based test case prioritization,” *IEEE Transactions on Reliability*, vol. 69, no. 3, pp. 1004–1020, 2020.

[154] M. Du and Y.-F. Li, “An investigation of new local search strategies in memetic algorithm for redundancy allocation in multi-state series-parallel systems,” *Reliability Engineering & System Safety*, vol. 195, pp. 106703-, 2020.

[155] Y.-F. Li, N. Pedroni, and E. Zio, “A memetic evolutionary multi-objective optimization method for environmental power unit commitment,” *IEEE Transactions on Power Systems*, vol. 28, no. 3, pp. 2660–2669, 2013.

[156] H. Gao, A. Wang, E. Zio, and G. Bai, “An integrated reliability approach with improved importance sampling for low-cycle fatigue damage prediction of turbine disks,” *Reliability Engineering & System Safety*, vol. 199, pp. 106819–19, 2020.

[157] C. Lu, H. Li, L. Han, B. Keshtegar, and C.-W. Fei, “Bi-iterative moving enhanced model for probability-based transient LCF life prediction of turbine blisk,” *Aerospace Science and Technology*, vol. 132, pp. 107998-, 2023.

[158] J. Zeng and Z. Liang, “A deep Gaussian process approach for predictive maintenance,” *IEEE Transactions on Reliability*, vol. 72, no. 3, pp. 1–18, 2023.

[159] H.-F. Gao, E. Zio, A. Wang, G.-C. Bai, and C.-W. Fei, “Probabilistic-based combined high and low cycle fatigue assessment for turbine blades using a substructure-based kriging surrogate model,” *Aerospace Science and Technology*, vol. 104, pp. 105957-, 2020.

[160] J. Zhang, L. Gao, M. Xiao, S. Lee, and A. T. Eshghi, “An active learning Kriging-assisted method for reliability-based design optimization under distributional probability-box model,” *Structural and Multidisciplinary Optimization*, vol. 62, no. 5, pp. 2341–2356, 2020.

[161] C. Lu, Y.-W. Feng, C.-W. Fei, and S.-Q. Bu, “Improved decomposed-coordinated kriging modeling strategy for dynamic probabilistic analysis of multicomponent structures,” *IEEE Transactions on Reliability*, vol. 69, no. 2, pp. 440–457, 2020.

[162] Y. Zhou and M. Huang, “Lithium-ion batteries remaining useful life prediction based on a mixture of empirical mode decomposition and ARIMA model,” *Microelectronics and Reliability*, vol. 65, pp. 265–273, 2016.

[163] O. Djedidi, M. A. Djeziri, and S. Benmoussa, “Remaining useful life prediction in embedded systems using an online auto-updated machine learning based modeling,” *Microelectronics and Reliability*, vol. 119, pp. 114071-, 2021.

[164] Y. Wu, M. Yuan, S. Dong, L. Lin, and Y. Liu, “Remaining useful life estimation of engineered systems using vanilla LSTM neural networks,” *Neurocomputing*, vol. 275, pp. 167–179, 2018.

[165] Q. Wu, K. Ding, and B. Huang, “Approach for fault prognosis using recurrent neural network,” *Journal of Intelligent Manufacturing*, vol. 31, no. 7, pp. 1621–1633, 2020.

[166] H. Miao, B. Li, C. Sun, and J. Liu, “Joint learning of degradation assessment and RUL prediction for aeroengines via dual-task deep LSTM networks,” *IEEE Transactions on Industrial Informatics*, vol. 15, no. 9, pp. 5023–5032, 2019.

[167] R. Zhu, Y. Chen, W. Peng, and Z.-S. Ye, “Bayesian deep-learning for RUL prediction: An active learning perspective,” *Reliability Engineering & System Safety*, vol. 228, pp. 108758-, 2022.

[168] M. Kim and K. Liu, “A Bayesian deep learning framework for interval estimation of remaining useful life in complex systems by incorporating general degradation characteristics,” *IISE Transactions*, vol. 53, no. 3, pp. 326–340, 2020.

[169] H. Zhang, G. Niu, B. Zhang, and Q. Miao, “Cost-effective lebesgue sampling long short-term memory networks for lithium-ion batteries diagnosis and prognosis,” *IEEE Transactions on Industrial Electronics*, vol. 69, no. 2, pp. 1958–1967, 2022.

[170] W. Peng, Z.-S. Ye, and N. Chen, “Bayesian deep-learning-based health prognostics toward prognostics uncertainty,” *IEEE Transactions on Industrial Electronics*, vol. 67, no. 3, pp. 2283–2293, 2020.

[171] M. Kim and K. Liu, “A Bayesian deep learning framework for interval estimation of remaining useful life in complex systems by incorporating general degradation characteristics,” *IISE Transactions*, vol. 53, no. 3, pp. 326–340, 2020.

[172] W. Peng, Z.-S. Ye, and N. Chen, “Bayesian deep-learning-based health prognostics toward prognostics uncertainty,” *IEEE Transactions on Industrial Electronics*, vol. 67, no. 3, pp. 2283–2293, 2020.

[173] P. Liu and G. Wang, “Generalized non-renewing replacement warranty policy and an age-based post-warranty maintenance strategy,” *European Journal of Operational Research*, vol. 311, no. 2, pp. 567–580, 2023.

[174] W. Xie and Z.-S. Ye, “Aggregate discounted warranty cost forecast for a new product considering stochastic sales,” *IEEE Transactions on Reliability*, vol. 65, no. 1, pp. 486–497, 2016.

[175] X. Zhu, C. Jiao, and T. Yuan, “Optimal decisions on product reliability, sales and promotion under nonrenewable warranties,” *Reliability Engineering & System Safety*, vol. 192, pp. 106268-, 2019.

[176] M. Kijima, “Some results for repairable systems with general repair,” *Journal of Applied Probability*, vol. 26, no. 1, pp. 89–102, 1989.

[177] S. Martorell, A. Sanchez, and V. Serradell, “Age-dependent reliability model considering effects of maintenance and working conditions,” *Reliability Engineering & System Safety*, vol. 64, no. 1, pp. 19–31, 1999.

[178] Y. Zhu *et al.*, “Optimal maintenance service strategy of service-oriented aviation manufacturers for two-stage leased system under capacity limits,” *IEEE Transactions on Reliability*, vol. 73, no. 2, pp. 1353–1367, 2024.

[179] F. M. Bass, “A new product growth for model consumer durables,” *Management Science*, vol. 15, no. 5, pp. 215–227, 1969.

[180] J.-T. Teng and G. L. Thompson, “Optimal strategies for general price-quality decision models of new products with learning production costs,” *European Journal of Operational Research*, vol. 93, no. 3, pp. 476–489, 1996.

[181] M. N. Darghouth, D. Ait-Kadi, and A. Chelbi, “Joint reliability-based design and periodic preventive maintenance policy for systems sold with warranty,” *Journal of Quality in Maintenance Engineering*, vol. 22, no. 1, pp. 2–17, 2016.

[182] J. Pongpech and D. N. P. Murthy, “Optimal periodic preventive maintenance policy for leased equipment,” *Reliability Engineering & System Safety*, vol. 91, no. 7, pp. 772–777, 2006.

[183] O. K. Mont, “Clarifying the concept of product–service system,” *Journal of Cleaner Production*, vol. 10, no. 3, pp. 237–245, 2002.

[184] C. Huang, S. Bu, S. W. Kong, and H. H. Lee, “A meta-model-based life consumption monitoring method for efficient decision-making of PSSs,” in *2024 IEEE International Conference on Prognostics and Health Management (ICPHM)*, Spokane, WA, USA, 2024, pp. 74-83.

[185] X. Wang, H. Zhou, A. K. Parlikad, and M. Xie, “Imperfect preventive maintenance policies with unpunctual execution,” *IEEE Transactions on Reliability*, vol. 69, no. 4, pp. 1480–1492, 2020.

[186] M. Compare, L. Bellani, and E. Zio, “Availability model of a PHM-equipped component,” *IEEE Transactions on Reliability*, vol. 66, no. 2, pp. 487–501, 2017.

[187] X. Zhao, S. He, and M. Xie, “Utilizing experimental degradation data for warranty cost optimization under imperfect repair,” *Reliability Engineering & System Safety*, vol. 177, pp. 108–119, 2018.

[188] T. Li, S. He, and X. Zhao, “Optimal warranty policy design for deteriorating products with random failure threshold,” *Reliability Engineering & System Safety*, vol. 218, pp. 108142-, 2022.

[189] Q. Wang, S. Bu, and Z. He, “Achieving predictive and proactive maintenance for high-speed railway power equipment with LSTM-RNN,” *IEEE Transactions on Industrial Informatics*, vol. 16, no. 10, pp. 6509–6517, 2020.

[190] R. Zheng, X. Qian, and L. Gu, “Group maintenance for numerical control machine tools: A case study,” *IEEE Transactions on Reliability*, vol. 72, no. 4, pp. 1407–1419, 2023.

[191] H. C. Vu, P. Do, and A. Barros, “A stationary grouping maintenance strategy using mean residual life and the Birnbaum importance measure for complex structures,” *IEEE Transactions on Reliability*, vol. 65, no. 1, pp. 217–234, 2016.

[192] D.-H. Dinh, P. Do, and B. Iung, “Multi-level opportunistic predictive maintenance for multi-component systems with economic dependence and assembly/disassembly impacts,” *Reliability Engineering & System Safety*, vol. 217, pp. 108055-, 2022.

[193] H.-C. Vu, P. Do, and A. Barros, “A study on the impacts of maintenance duration on dynamic grouping modeling and optimization of multicomponent systems,” *IEEE Transactions on Reliability*, vol. 67, no. 3, pp. 1377–1392, 2018.

[194] K.-A. Nguyen, P. Do, and A. Grall, “Multi-level predictive maintenance for multi-component systems,” *Reliability Engineering & System Safety*, vol. 144, pp. 83–94, 2015.

[195] High-Voltage Switchgear and Controlgear - Part 200: AC Meta-Enclosed Switchgear And Controlgear for Rated Voltages Above 1 KV and up to and Including 52 KV, 1st edition, IEC 62271-200, 2003.

List of Publications

Journal paper

1. **C. Huang**, S. Bu, H. H. Lee, K. W. Chan, and W. K. C. Yung, "Prognostics and Health Management for Induction Machines: A Comprehensive Review," *Journal of Intelligent Manufacturing*, vol. 35, no. 3, pp. 937–962, 2024. (Chapter 1) (IF: 6.4, Rank: 39/197, Quartile: Q1)

2. **C. Huang**, S. Bu, H. H. Lee, C. H. Chan, S. W. Kong, and W. K. C. Yung, "Prognostics and Health Management for Predictive Maintenance: A Review," *Journal of Manufacturing Systems*, vol. 75, pp. 78–101, 2024. (Chapters 1 and 2) (IF: 11.3, Rank: 1/106, Quartile: Q1)

3. **C. Huang**, S. Bu, C.-W. Fei, N. Lee, and S. W. Kong, "Reliability Assessment for Aeroengine Blisks Under Low Cycle Fatigue With Ensemble Generalized Constraint Neural Network," *IEEE Transactions on Reliability*, vol. 73, no. 2, pp. 922–936, 2024. (Chapter 3) (IF: 5.2, Rank: 13/132, Quartile: Q1)

4. **C. Huang**, S. Bu, X. Liu, H. H. Lee, C. H. Chan, and S. W. Kong, "Digital Twin-Based Reliability Prediction for Warranty Design Under Reliability-Constraint Imperfect Preventive Maintenance", *Computers & Industrial Engineering*. (Chapters 3 and 4) (IF: 6.7, Rank: 21/170, Q1) (Under review)

5. **C. Huang**, S. Bu, H. H. Lee, C. H. Chan, S. W. Kong, and W. L. Yeung, "Dynamic Predictive Maintenance Planning with Bayesian Deep Learning-Based Digital Twins", *Computers in Industry*. (Chapters 3 and 5) (IF: 9.3, Rank: 11/170, Quartile: Q1) (Under review)

6. **C. Huang**, S. Bu, C. Lei, H. H. Lee, S. W. Kong, and C. H. Chan, "Digital Twin-Based Condition-Based Maintenance Planning for Line Replaceable Units with Adaptive Kriging Emulators", *IEEE Transactions on Reliability*. (Chapters 3 and 6) (IF: 5.2, Rank: 13/132, Quartile: Q1) (Under review)

7. **C. Huang**, S. Bu, Q. Wang, H. H. Lee, C. Lei, S. W. Kong, and C. H. Chan, “Opportunistic Condition-Based Maintenance for Multi-Unit Systems Considering Competing Failure Models and Economic Dependencies”, *IEEE Transactions on Reliability*. (Chapter 7) (IF: 5.2, Rank: 13/132, Quartile: Q1) (Under review)

8. **C. Huang**, S. Bu, Q. Chen, and H. H. Lee, “Meta-Power: Next-Generation Smart Grid,” *Power Generation Technology*, vol. 43, no. 2, pp. 287–304, 2022. (Chapter 1) (IF: 2.14) (Outstanding paper)

9. **C. Huang**, S. Bu, and H. H. Lee. “Meta-Power: Digitalized Power Systems Driven by Metaverse.” *Journal of Power Generation Technology*, vol. 1, no. 2, pp. 20–49, 2024. (Chapter 1)

Conference paper

1. **C. Huang** and S. Bu, "Strategic Electrical Asset Maintenance by Reinforcement Learning," *12th IET International Conference on Advances in Power System Control, Operation and Management (APSCOM 2022)*, Hong Kong, China, 2022, pp. 48-52. (Chapter 2) (Best paper award)

2. **C. Huang**, S. Bu, S. W. Kong and H. H. Lee, "A Meta-Model-Based Life Consumption Monitoring Method for Efficient Decision-Making of PSSs," *2024 IEEE International Conference on Prognostics and Health Management (ICPHM)*, Spokane, WA, USA, 2024, pp. 74-83. (Chapter 4)