Risk-based maintenance decision-making approaches for drilling and measurement tools

Jinlong Kang

Thèse de doctorat de l'établissement Université Bourgogne Franche-Comté
preparée à l'Université de Franche-Comté

École doctorale n° 37
Sciences Pour l'Ingénieur et Microtechniques

Doctorat en Automatique

par

Jinlong Kang

Approches décisionnelles de la maintenance basées sur le risque pour les outils de forage et de mesure

Thèse soutenue publiquement le 07/03/2024, à Besançon,
devant le jury composé de:

Abderrahman EL MHAMEDI	Professeur, Université Paris 8	Président
Labib TERISSA	Professeur, Université de Biskra	Rapporteur
Sofiene DELLAGI	Maître de Conférences, Université de Lorraine	Rapporteur
Mingjian ZUO	Professeur, University of Alberta	Examinateur
Noureddine ZERHOUNI	Professeur, SUPMICROTECH-ENSMM	Directeur de thèse
Christophe VARNIER	Maître de Conférences, SUPMICROTECH-ENSMM	Codirecteur de thèse
Ahmed MOSALLAM	Responsable du Hub Data Science & AI - EUR, SLB	Industrial supervisor

Ph.D. thesis of the University Bourgogne Franche-Comté
prepared at the university of Franche-Comté

Doctoral school n° 37
Engineering Sciences and Microtechnologies

Ph.D. in Automatic

by

Jinlong Kang

Risk-based maintenance decision-making approaches for drilling and measurement tools

Thesis defended publicly on 07/03/2024, in Besançon,
composition of jury:

Abderrahman EL MHAMEDI	Professor, Université Paris 8	President
Labib TERISSA	Professor, Université de Biskra	Reporter
Sofiene DELLAGI	Associate Professor, Université de Lorraine	Reporter
Mingjian ZUO	Professor, University of Alberta	Examiner
Noureddine ZERHOUNI	Professor, SUPMICROTECH-ENSMM	Supervisor
Christophe VARNIER	Associate Professor, SUPMICROTECH-ENSMM	Co-supervisor
Ahmed MOSALLAM	Data Science & AI Hub Manager - EUR, SLB	Industrial supervisor

Acknowledgment

I would like to express my heartfelt gratitude to the following individuals and organizations for their invaluable support and contributions throughout my journey in pursuit of this PhD:

First and foremost, I am deeply thankful to my family for their unwavering support, understanding, and encouragement, which sustained me during the demanding and challenging phases of my doctoral research.

I extend my sincere appreciation to my manager, Ahmed, who not only played a vital role in my professional development but also served as my industry supervisor, offering guidance and insights that enriched my research.

I am deeply indebted to SRPC, SLB for providing me with the incredible opportunity to embark on this CIFRE PhD position, which has been a pivotal step in my academic and professional journey.

I would like to extend my gratitude to my colleagues at SLB, particularly Nannan and Fares, whose collaboration and assistance in problem formulation were instrumental in shaping the direction of my research.

In addition, I am thankful to my colleagues in the Data Science & AI Team, including Karolina, Alejandro, Nahieli, and others, for their camaraderie and valuable insights, which greatly enriched my professional experience.

I am also appreciative of the support I received from the DFMS digital team, specifically Guillaume and Matthieu, as well as the D&M tool SMEs at SLB, who provided expertise and resources that were integral to my research.

Special thanks go to the Department of AS2M at FEMTO-ST for their support in enabling my participation in three international conferences, which expanded my horizons and facilitated academic networking.

I am profoundly grateful to my two supervisors, Christophe and Noureddine, at the research institute for their unwavering guidance, mentorship, and belief in my research capabilities.

Last but not least, I want to express my gratitude to my colleagues in Besançon, including Ning, Zeina, Jean-Marc, Kejun, Meesam, Yucely, Nahashon, and many others, whose camaraderie and friendship brought joy and a sense of belonging during my stay.

Your support, guidance, and companionship have been essential in making this journey possible. I am deeply appreciative of each of you and am inspired by the collective effort that has led to the successful completion of my PhD.

Contents

List of Figures

List of Tables

Acronyms

AIC Akaike Information Criterion
BHA Bottom Hole Assembly
BIC Bayesian Information Criterion
CBM Condition-based Maintenance
CM Corrective Maintenance
CMMS Computerized Maintenance Management System
CPU Central Processing Unit
D&M Drilling and Measurement
DFMS Digital Fleet Management System
EM Expectation Maximization
HMI Human-Machine Interface
HMM Hidden Markov Model
IEEE Institute of Electrical and Electronics Engineers
IoT Internet of Things
ISO International Organization for Standardization
KPI Key Performance Indicator
LWD Logging-While-Drilling
MeTTF Median Time To Failure
MTTF Mean Time To Failure
MWD Measurement-While-Drilling
PCBA Printed Circuit Board Assembly
PdM Predictive Maintenance
PHM Prognostics and Health Management
PM Preventive Maintenance
QR Quantile Regression
RBM Risk-based Maintenance
RSS Rotary Steering System
RUL Remaining Useful Life
SME Subject Matter Expert
SRA Society for Risk Analysis

Chapter I
General introduction

I.1 The context of the study

The oil and gas industry plays an indispensable role in modern society, profoundly impacting our daily lives in multifaceted ways. Beyond its primary function as an energy source, oil and gas are fundamental constituents in producing various chemical products that permeate our everyday existence. The hydrocarbons derived from oil and gas are omnipresent, from the fuel that powers our vehicles to the plastics used in countless consumer goods. This industry underpins transportation, manufacturing, and countless other sectors, serving as a linchpin in the global economy.

In recent years, the increasing urgency of addressing climate warming and the energy crises resulting from local conflicts have prompted numerous countries worldwide to undertake an energy transition. This transition involves a shift towards developing sustainable energy sources, notably wind and solar energy, to reduce dependence on fossil fuels. However, despite concerted efforts to diversify energy resources and reduce the dependence on fossil fuels, oil and gas continue to dominate the energy landscape. Figure I.1 shows their market share in the energy sector unmistakably reveals their enduring significance [1]. While the pursuit of alternative energy sources is commendable, it is an indisputable fact that, for the foreseeable future, oil and gas will remain the primary energy resources upon which our societies rely.

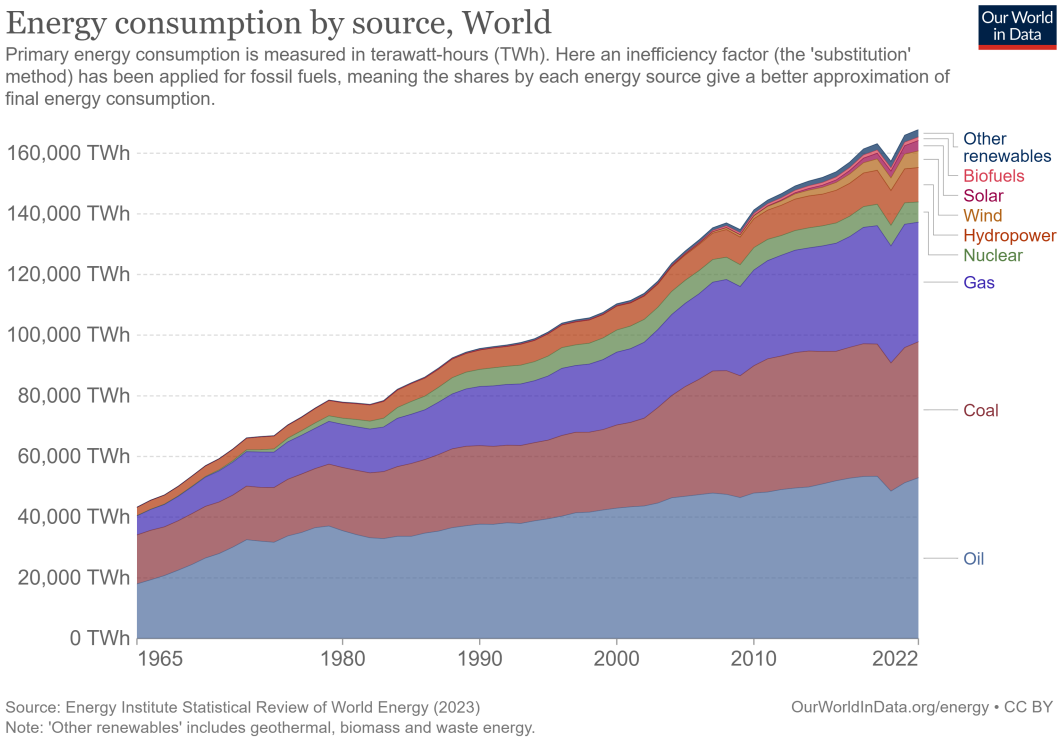


Figure I.1 – Global energy consumption by source [1]

The process of oil and gas traveling from underground reservoirs to our everyday use comprises three main stages illustrated in Figure I.2: upstream, midstream, and downstream. Upstream pertains to the initial phase of oil and gas production, encompassing exploration, drilling, and extraction activities. Meanwhile, midstream primarily focuses on the processing, storing, and transporting of oil and gas products. Lastly, downstream signifies the concluding phase involving the refining and distributing of petroleum products. This study concentrates specifically on the drilling process within the upstream stage.

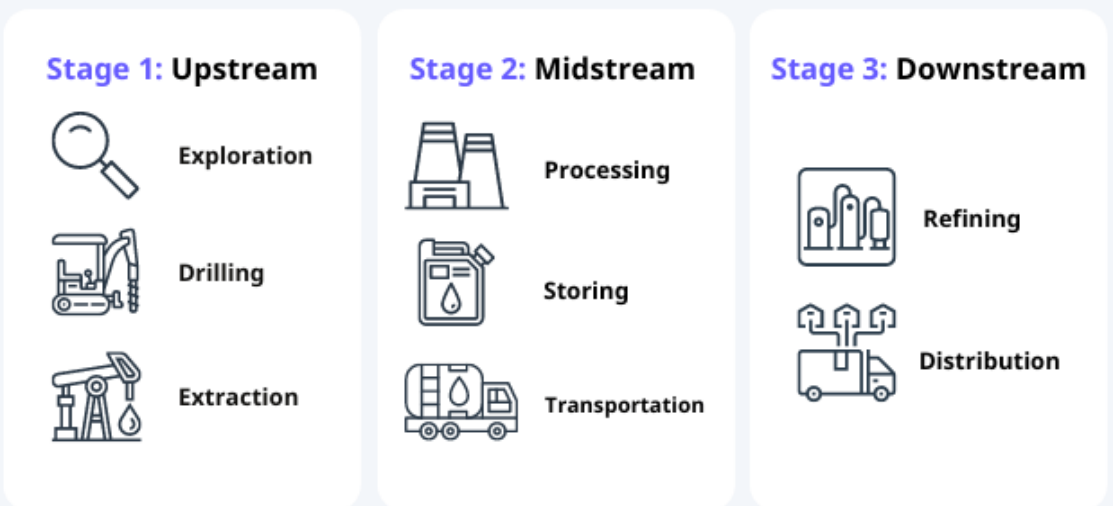


Figure I.2 – Oil and gas production process [2]

The drilling process relies on the drilling system as shown in Figure I.3. The system usually consisting of a drilling rig, drillpipe, a Bottom Hole Assembly (BHA) and a drill bit. The BHA is an important part of the drilling system because it must provide power for the bit to rotate and break through the rock, survive a harsh operational environment, and provide accurate directional control of the well [5]. The BHA is configured based on drilling operation requirements; thus, different drilling jobs could have different BHA configurations. Nevertheless, the BHA frequently includes Measurement-While-Drilling (MWD) tool(s), Logging-While-Drilling (LWD) tool(s), and Rotary Steering System (RSS) as shown in Figure I.4. The MWD tool on the top of the BHA is responsible for delivering real-time data to the surface, powering and transmitting data from multiple LWD tools, and determining the position and orientation of the drillstring [6]. The LWD combines a complete set of functions, including formation evaluation, well placement, and drilling optimization measurements into a single collar [7]. The RSS at the bottom of the BHA is designed to rotate the drill bit in the desired direction; thereby, control the well path [8]. MWD, LWD, and RSS are collectively termed Drilling and Measurement (D&M) tools or technologies.

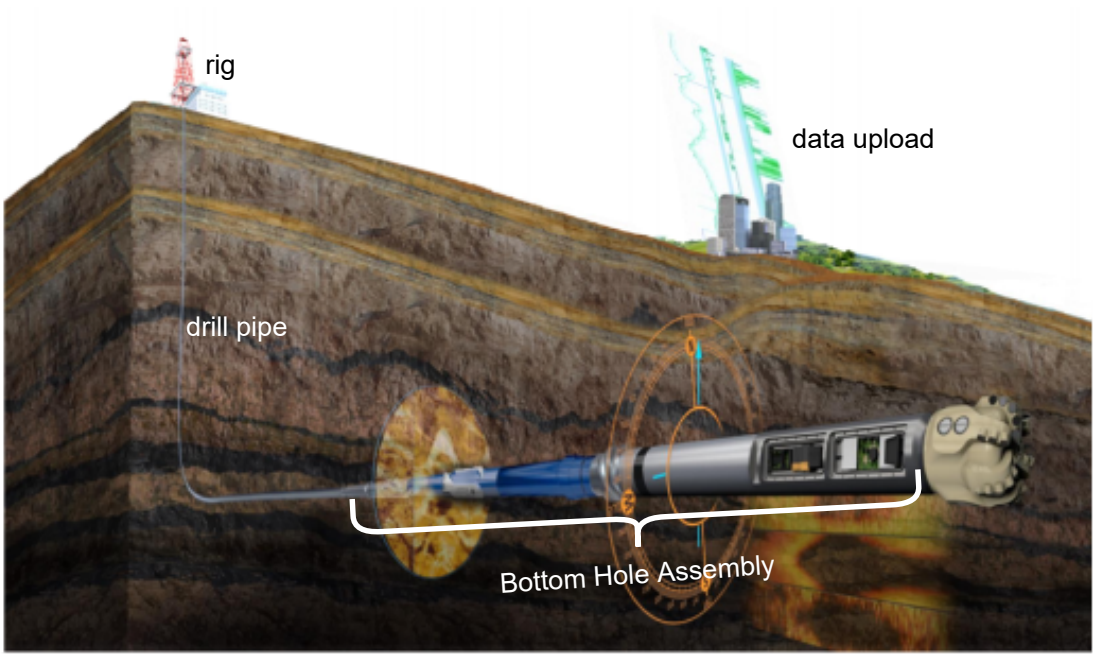


Figure I.3 – Drilling system schematic

Figure I.4 – Bottom hole assembly schematic

The functional descriptions of these tools undeniably underscore the crucial role their reliability plays in ensuring the efficiency and success of drilling operations. A fundamental approach to ensuring this reliability is implementing proactive maintenance measures before potential failures occur.

This PhD study's primary focus is positioned within this context, delving deep into the domain of maintenance decision-making for D&M tools. It acknowledges that making well-informed decisions about maintenance is critical for upholding the reliability and longevity of these D&M tools. These decisions result in reduced downtime, decreased unnecessary maintenance activities, and an overall enhancement in the success and efficiency of drilling operations. Furthermore, these endeavors play an important role in reducing the carbon footprint of D&M tools and drilling operations in the oil and gas industry, aligning with the broader goals of the ongoing energy transition and the urgent need to combat global warming.

I.2 Positioning of the research

Maintenance encompasses a wide range of activities, including tests, measurements, replacements, adjustments, and repairs, all of which are carried out to either preserve the functional condition of a unit or return it to a specified state in which it can effectively perform its required functions [9].

In the area of maintenance decision-making, several strategies exist designed to ensure the optimal performance and longevity of equipment. These strategies encompass a spectrum of approaches, each tailored to specific situations and equipment types. The five primary maintenance strategies are corrective, preventive, condition-based, predictive, and risk-based maintenance [10, 11].

Corrective Maintenance (CM) strategy addresses maintenance needs passively, primarily when equipment failures occur. Preventive Maintenance (PM) relies on scheduled time-based or use-based inspections and maintenance tasks performed on the equipment. Condition-based Maintenance (CBM) performs maintenance activities based on the actual condition or performance of the equipment. Maintenance activities are usually initiated when incipient failure or light degradation is detected. Predictive Maintenance (PdM) involves planning maintenance activities in advance, leveraging the capabilities of failure prognostics or predicting the Remaining Useful Life (RUL) of the equipment. Risk-based Maintenance (RBM) is a maintenance methodology that prioritizes maintenance resources based on risk assessments. More details about these five maintenance strategies will presented in the next chapter.

The choice of a maintenance strategy hinges on several critical factors, including the criticality of the equipment, the potential costs associated with equipment failure, and the availability of maintenance resources. The selection process may also consider hybrid maintenance strategies that combine elements of these approaches [12, 13].

This PhD study exclusively centers on the risk-based maintenance strategy. Further elaboration on why this study focus on the risk-based maintenance strategy will be presented in the next chapter. The risk-based maintenance strategy involves developing risk-based maintenance decision-making approaches for D&M tools. These approaches balance addressing potential drilling operation failures and optimizing resource allocation by prioritizing D&M tool maintenance activities based on risk assessments of D&M tool failures and drilling operational environments. The study explores the applications, benefits, and implications of risk-based maintenance strategy within the oil and gas industry context. Through a comprehensive examination of this maintenance strategy, the study aims to provide valuable insights into maintenance decisions to enhance the reliability of D&M tools, contributing to the overall success and improved service quality of oil well drilling operations.

I.3 Research questions

Risk-based maintenance involves two key phases: risk assessment and decision-making [14]. In addition, the oilfield operational environment and the reliability of D&M tools are key factors that influence the efficiency and success of drilling operations. Last, the risk estimation model used in this study relies heavily on data-driven models where data quality is critical. Unfortunately, the data collected from fields often suffer from various quality issues, including noise and missing values. Therefore, several key research questions must be addressed to successfully implement risk-based maintenance decisions for D&M tools in the oil and gas industry.

- **Operational Environment Risk Characterization:** The first question involves understanding and characterizing the risks of different oilfield operational environments. That is, how can we assess and quantify the risks posed by the operational environment in oil fields (e.g., temperature, vibration, shock, and pressure)?
- **Tool Risk Assessment:** The second question relates to assessing the failure risk of D&M tools and ensuring that the risk evolves as they are used.
- **Optimal Maintenance Decisions:** Once the risks associated with operational environments and D&M tools are assessed, how can we make informed maintenance decisions at the individual tool and fleet level?
- **Data Quality Management:** Recognizing that risk estimates depend on data-driven models using measurement data collected during drilling operations in oil fields, how can we manage the quality of these data? This task includes addressing challenges such as data noise and missing values. What methods can be employed to assess the data quality and determine if the data quality is adequate and if not, what measures can be implemented to improve the data quality?

I.4 Organization and overview of the thesis

This thesis is structured to provide a comprehensive look at risk-based decision-making approaches for D&M tools in the oil and gas industry. The organization and overview of this thesis are shown in Figure I.5. Except for Chapter I of the general introduction, brief descriptions of the remaining chapters are given below.

Chapter II: From PHM to risk

This chapter provides an in-depth look at the existing maintenance decision-making strategies. A critical assessment of each strategy's limitations lays the foundation for transitioning to risk-based decision-making. This chapter clarifies the distinction between Prognostics and Health Management (PHM) and risk, further setting the research motivation for a risk-based maintenance strategy.

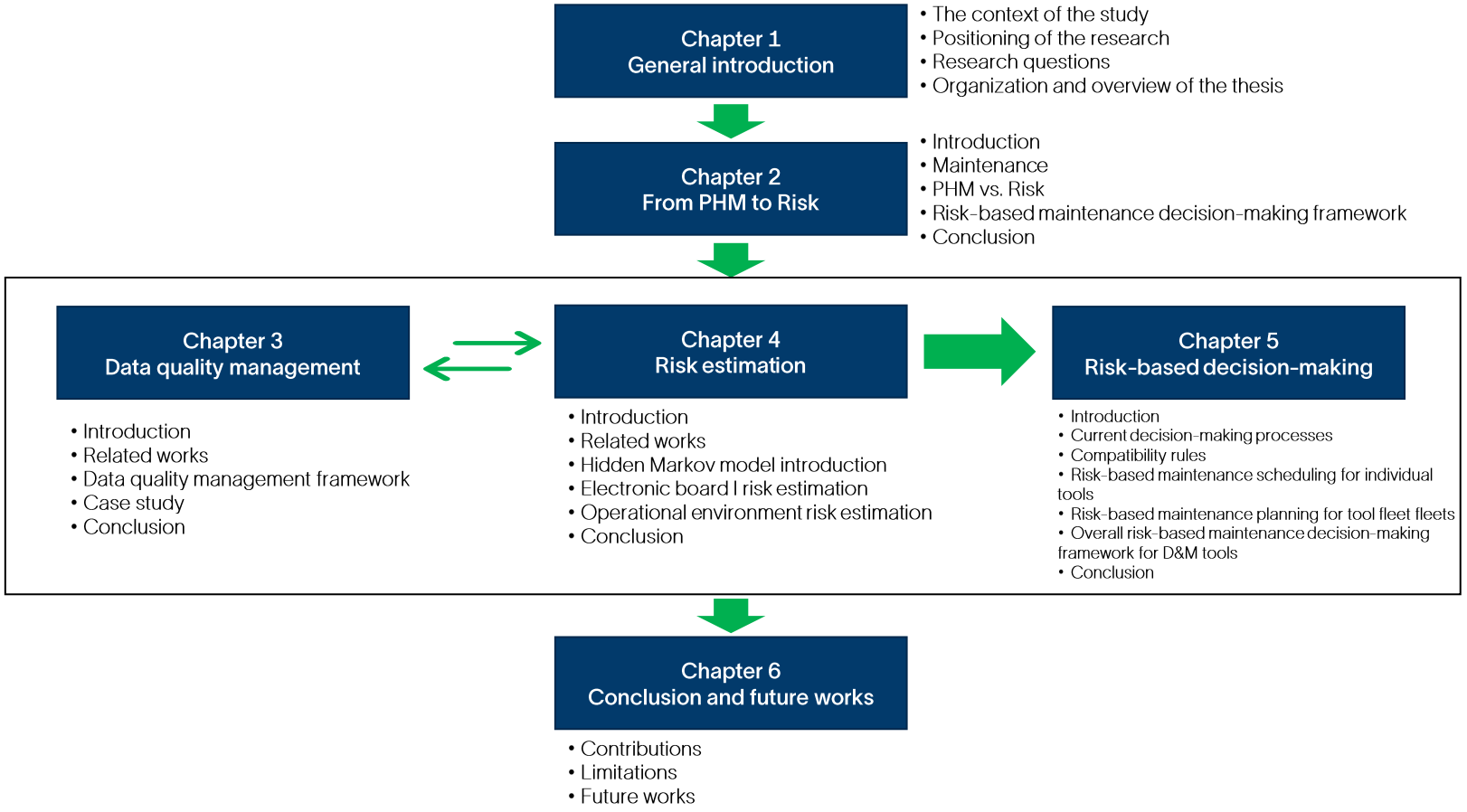


Figure I.5 – The organization and overview of the thesis

Chapter III: Data quality management

Recognizing the foundational role of data in risk-based decision-making, this chapter provides an in-depth look at data quality management. We examine why data quality is critical, especially in the context of forecast uncertainty. Through an extensive literature review of data quality metrics, particularly in engineering asset management, we develop a robust framework for data quality management. The application of the framework is demonstrated through a case study using data from two D&M tools.

Chapter IV: Risk estimation

This chapter discusses the core elements of risk assessment, focusing on the critical impact of tool health and oilfield operational environment on drilling operation success and efficiency. A comprehensive review of related work lays the foundation for introducing Hidden Markov Model (HMM)-based electronic board risk estimation. In addition, the chapter provides an overview of the operational environment risk assessment process, which lays the foundation for subsequent risk-based decision-making.

Chapter V: Risk-based decision-making

After establishing robust risk estimates, Chapter V transitions to practical decision-making. We analyze the current decision-making process for D&M tool selection and fleet maintenance planning commonly used in the oil and gas industry and then introduce compatibility rules. The chapter further explores risk-based tool selection and maintenance decision approaches and risk-based fleet maintenance planning decision approaches. It also presents real-world use cases to illustrate the application of the proposed risk-based decision approaches.

Chapter VI: Conclusion and future works

This chapter summarizes this research's contributions, acknowledges the study's limitations, and sets the stage for future work. By describing the specific research questions addressed throughout this thesis, we aim to provide valuable insights into risk-based decision-making approaches for D&M tools in the oil and gas industry.

Chapter II
From PHM to risk

II.1 Introduction

In the field of maintenance strategies, it is crucial to have a deep understanding of the strengths and limitations of each strategy, especially when applying them to the maintenance of D&M tools in the oil and gas industry. This chapter provides a comprehensive examination of five widely employed maintenance strategies: corrective maintenance, preventive maintenance, condition-based maintenance, predictive maintenance, and risk-based maintenance. These strategies were briefly introduced in the General Introduction chapter, and this chapter delves into their respective advantages and disadvantages. Through this study, we aim to illuminate the constraints and challenges encountered when implementing the first four strategies in D&M tool maintenance. It's important to note that condition-based and predictive maintenance strategies heavily rely on PHM models, which encompass fault detection, diagnosis, and prediction models, utilizing condition monitoring data. In contrast, risk-based maintenance is firmly rooted in the concept of risk. Thus, we will conduct a thorough analysis to differentiate between PHM and risk, elucidate the limitations of PHM in addressing specific issues, and provide a rationale for embracing risk-based decision-making. Finally, this chapter will introduce a framework for risk-based maintenance decision-making to complement existing mainstream maintenance strategies.

II.2 Maintenance

Maintenance is a crucial factor in industry and accounts for a significant portion of production costs, ranging from 15% to 70%, depending on the specific industry [15]. However, despite the importance of maintenance, maintenance costs have been on the rise, with about one-third of these costs being wasted because of planning uncertainties and inefficiencies [16]. As a response to optimizing maintenance, maintenance strategies have evolved through the tireless efforts of researchers and experts. At present, there are five commonly used maintenance strategies.

II.2.1 Corrective maintenance

CM is the most primitive maintenance strategy. It is also known as breakdown maintenance, reactive maintenance, and run-to-failure. As the name suggests, this strategy addresses maintenance needs passively, primarily when equipment failures occur, as depicted in Figure II.1. Thus, this strategy is straightforward and needs minimal planning and scheduling. In addition, it maximizes the utilization of the equipment service life [17]. However, this reactive strategy requires maintaining extensive spare parts inventories that include spare machines or at least all critical components for the equipment. The alternative is to rely on equipment vendors that can provide immediate delivery of all required spare parts [16]. Moreover,

this strategy can result in unexpected equipment failures and unplanned downtime, disrupting the operation of other interconnected systems. Considering these advantages and disadvantages, CM is typically deployed in situations where the consequences of equipment failure are relatively minor, and the costs associated with immediate corrective actions are low [18].

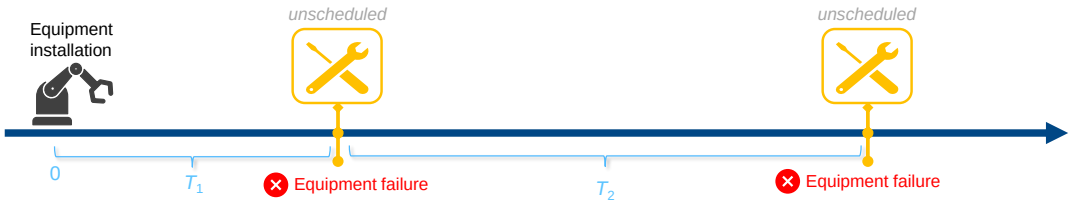


Figure II.1 – Illustration of corrective maintenance strategy

II.2.2 Preventive maintenance

Periodic maintenance, scheduled maintenance, time-based maintenance, age-based maintenance, and usage-based maintenance are other terms for PM. Unlike CM, PM relies on scheduled inspection and maintenance tasks performed on equipment that are either time-based (as shown in Figure II.2, at certain intervals T) or usage-based (e.g., at a certain number of duty cycles, runs, or jobs) [16]. One of the advantages of PM is that it is easy to implement. It also follows predetermined time or usage intervals, and the maintenance schedules are consistent and predictable. Finally, PM maintenance tasks are relatively simple and do not require extensive training. However, PM can result in parts being replaced too early or too late. Premature replacement increases maintenance costs because it replaces parts that can perform more duty cycles and are in good condition [16]. On the other hand, delayed replacement may lead to severe consequences because CM operations must be performed [19]. Therefore, PM is usually used when the potential cost of equipment failure is high, making proactive maintenance a cost-effective strategy. This approach is often applied in industries such as aerospace, where safety and reliability are paramount, or where maintenance programs can have a positive financial effect due to the need to procure scarce replacement parts in advance [20].

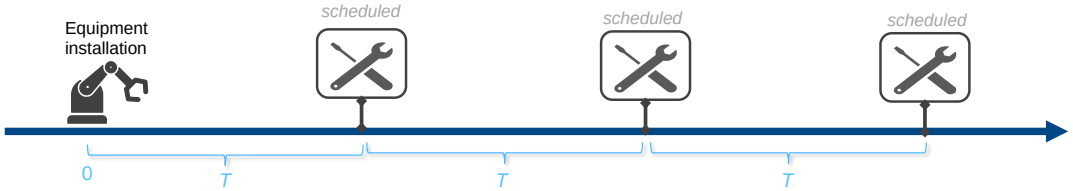


Figure II.2 – Illustration of preventive maintenance strategy

II.2.3 Condition-based maintenance

CBM strategy arose with the technical advancements in Industry 4.0 and the growth of the Internet of Things (IoT) [21]. CBM leverages sensor technologies to gather real-time condition monitoring information on equipment, including forces, vibrations, temperatures, current, voltage, auditory signals, and more. After that, data analytic techniques are applied to these signals in order to assess the equipment's overall health. As shown in Figure II.3, when early failure or light degradation is identified, maintenance operations are started, enabling prompt interventions and possibly prolonging the life of the equipment [22, 23]. Therefore, compared to CM and PM, CBM is more cost-effective as it targets specific components for repair when needed, thereby avoiding unnecessary maintenance actions. Moreover, CBM enhances system reliability and availability by proactively addressing issues, thereby reducing unplanned downtime. However, CBM entails additional investment in sensor installation. It also relies on diagnostic results on real-time data, which can result in variable maintenance schedules. Additionally, developing effective equipment health assessment models requires significant effort. Nevertheless, CBM is appropriate for systems when the benefits of preventing unscheduled shutdowns due to malfunctions outweigh the expenses associated with the monitoring system and the advancement of instruments for detection and diagnosis [24].

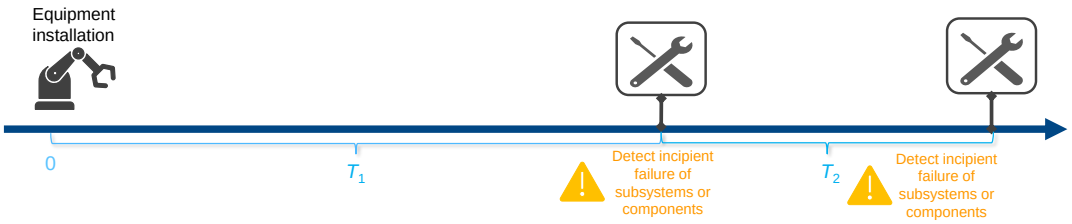


Figure II.3 – Illustration of condition-based maintenance strategy

II.2.4 Predictive maintenance

PdM is an extension of CBM, it processes further the condition monitoring information for prognostics [25]. Specifically, PdM is a forward-looking strategy that employs various forecasting methods to predict the remaining useful life of equipment. By forecasting when equipment is likely to reach the end of its serviceable life, maintenance activities can be scheduled in advance, accordingly, as shown in Figure II.4. As with CBM, PdM prevents unnecessary maintenance actions, maximizes equipment utilization, and minimizes downtime [26, 27]. PdM schedules are more predictable than CBM schedules because maintenance activities can be planned in advance. However, PdM also requires additional investment in condition monitoring systems, including sensor installation, communication, and signal processing. Additionally, effective prognostic models are essential for PdM. PdM

is suitable for systems that can benefit from the same advantages of CBM, but can also benefit from advance planning, such as systems that require special spare parts that need to be ordered in advance [20].

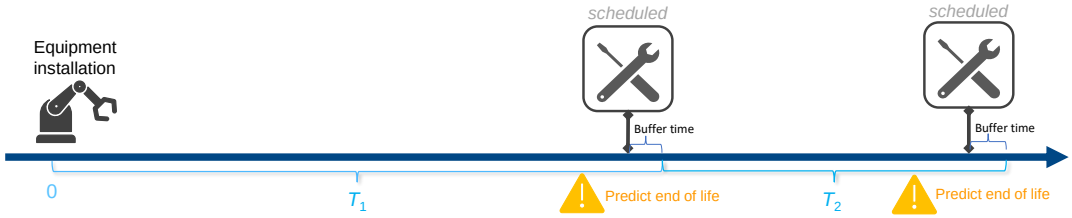


Figure II.4 – Illustration of predictive maintenance strategy

II.2.5 Risk-based maintenance

RBM is not an extension of predictive maintenance; instead, it is an integrated approach that incorporates maintenance and safety and recognizes that the two functions are not mutually exclusive [14]. The emergence and increasing popularity of RBM has occurred since 2000. The goal of RBM is to minimize the probability of system failure and its consequences by prioritizing maintenance activities based on risk assessment [28, 29]. Risk assessment includes hazard analysis to identify failure scenarios, likelihood assessment to calculate the incidence of unforeseen events, and consequence assessment to quantify potential consequences [14]. In some aspects, RBM is similar to prescriptive maintenance because both approaches recommend maintenance actions based on an analysis of potential scenarios. RBM emphasizes potential hazard scenarios, while prescriptive maintenance centers around the impact of potential maintenance activities [11, 20]. The main benefit of RBM is to improve the safety and reliability of the system. In addition, it allows limited resources to be allocated where they are most needed. However, RBM is more complex and requires a significant upfront investment compared to CM and PM. RBM is suitable for systems with the potential for serious consequences of failure, including safety hazards, environmental impacts or significant economic losses. It is also applicable to systems with different mission profiles or hazardous scenarios that lead to different failure probabilities and consequences.

II.2.6 Why towards risk-based maintenance for D&M tools

At present, the first four maintenance strategies mentioned above are employed within the context of D&M tool maintenance. Both CM and PM carry the risk of unexpected tool failures, resulting in financial losses and reputational damage. In addition, PM often leads to unnecessary maintenance interventions, incurring avoidable costs. In contrast, CBM and PdM prove superior to CM and PM in

terms of conserving maintenance costs. Nevertheless, their applicability is confined to systems equipped with a thorough comprehension of failure mechanisms, observable symptoms, a functioning condition monitoring system, and effective PHM models encompassing fault detection, diagnosis, degradation assessment, and RUL prediction. These prerequisites, however, remain unmet for D&M tools due to the following factors:

1. The intricate nature of failure mechanisms in electronic boards, that is, the critical subsystems of D&M tools, poses challenges for comprehensive understanding and investigation of electronics failures.
2. The difficulty in observing and identifying failure symptoms for electronic boards, a point that will be expounded upon in the subsequent section.
3. Inadequate or limited provision of condition monitoring information by existing monitoring systems in D&M tools. Specifically,
   - The compact nature of D&M tools presents challenges in accommodating numerous sensors to collect sufficient condition-monitoring information.
   - Operational constraints in downhole environments impede the transmission of copious information to the surface due to bandwidth limitations in communication, coupled with limited storage capacity in built-in memory chips.
   - For most of the electronic boards, only operational environment data (e.g., temperature, shocks, and vibration) are measured. Moreover, these data are measured at the up level (i.e., tool or system) instead of down levels (i.e., board level or electronic component level).

Given the aforementioned constraints, constructing effective PHM models becomes either implausible or exceedingly challenging. Moreover, the dynamic and extreme operational environments in downhole settings, coupled with the diverse fields and drilling missions encountered by D&M tools, underscore the applicability of the risk concept. It is crucial to note that these constraints apply not only to D&M tools, but also to other electronic-rich systems, such as unmanned aerial vehicles and electric vehicles.

Building upon the above analysis, the present study delves into risk-based maintenance for D&M tools within the oil and gas industry. It is essential to emphasize that risk-based maintenance is not envisaged as a replacement for existing strategies but rather as a complementary approach. For the subsystems which have effective PHM models, CBM and PdM are still preferred. For example, CBM and PdM are used for the neutron generator subsystem in LWD tools because it has effective fault detection [7], fault diagnostic [30], and degradation modeling [31] models.

II.3 PHM vs. risk

PHM serves as the foundation for CBM and PdM, whereas risk forms the basis for RBM. Thus, in order to gain a deeper understanding of the motivation behind this study's exploration of RBM, this section will elucidate the distinction between PHM and the broader concept of risk. Through this elucidation, we further draw out the need and relevance of studying RBM in the context of D&M tool maintenance.

II.3.1 Definitions and standards

II.3.1.a PHM

The absence of a standardized definition for PHM is evident, with scholars offering diverse interpretations, as highlighted in the excerpts from published works presented in Table II.1. Despite the variability in PHM definitions, a well-established PHM cycle exists. Figure II.5 depicts the widely accepted PHM cycle, comprising seven layers: data acquisition, data processing, condition assessment, diagnostics, prognostics, decision support, and Human-Machine Interface (HMI). These layers can be further categorized into three stages: (1) observe, (2) analyze, and (3) act [3][32].

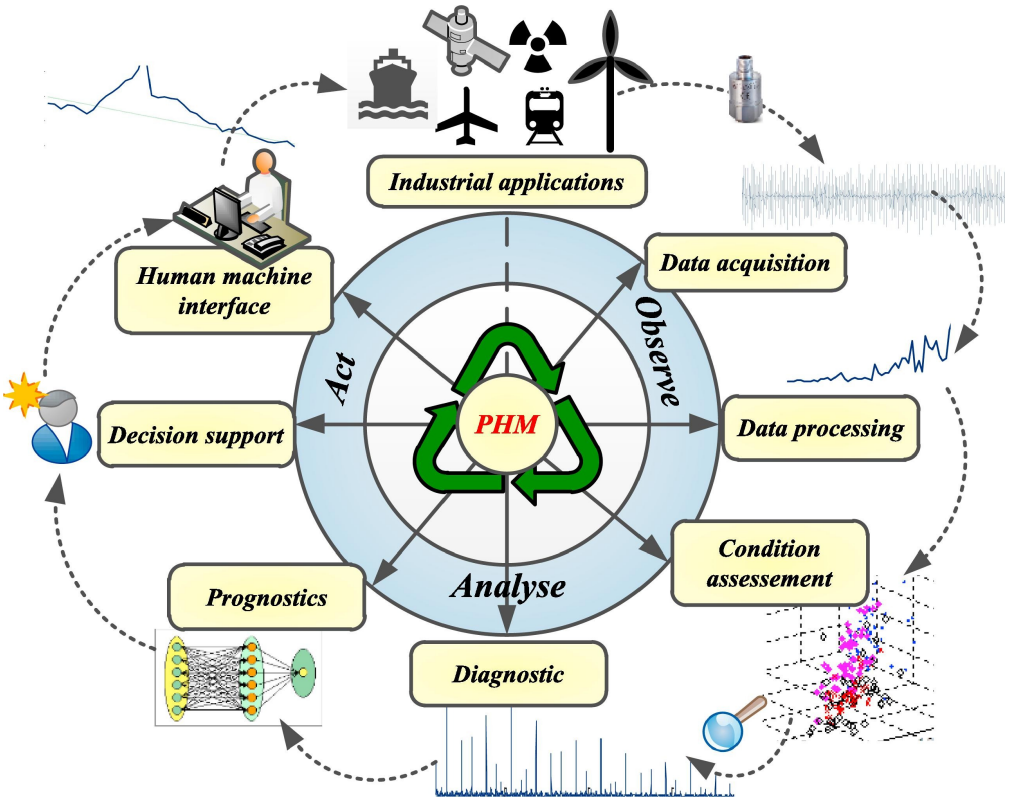


Figure II.5 – PHM cycle [3]

Table II.1 – PHM Definitions

Reference	Definition
Das et al. [33]	PHM systems are designed to predict impending faults and to determine remaining useful life of machinery.
Biggio and Kastanis [34]	PHM systems aim at detecting whether an industrial component has deviated from its normal operating condition or predicting when a fault will occur.
CALCE, University of Maryland [35]	Prognostics is the process of monitoring the health of a product and predicting its remaining useful life by assessing the extent of deviation or degradation from its expected state of health in its expected usage conditions. Health Management utilizes prognostic information to make decisions related to safety, condition-based maintenance, ensuring adequate inventory, and product life extension.
Calabrese et al. [36]	PHM is a maintenance policy aimed at predicting the occurrence of a failure in components and consequently minimizing unexpected downtimes of complex systems
Sutharssan et al.[37]	PHM are used to detect anomalies, diagnose faults and predict remaining useful lifetime
Brahimi et al. [38]	PHM aims to afford solutions for asset monitoring, algorithms for health assessment, fault diagnostics, and failure prognostics, as well as decision optimization and Human Machine Interface development.
Medjaher et al. [39]	PHM aims to monitor systems, assess their health state, detect and diagnose their faults, anticipate the time to failures by calculating the RUL and take appropriate decisions accordingly.

While PHM itself remains a multifaceted term, specific domains have instituted standards related to it. Noteworthy examples include the Institute of Electrical and Electronics Engineers (IEEE) standard framework for PHM of electronic systems [4], standards tailored for PHM within the manufacturing sphere [40], and International Organization for Standardization (ISO) standards centered on the condition monitoring and diagnostics of machines [41].

II.3.1.b Risk

In contrast, risk is subject to well-defined descriptions found in two influential references: scientific-technical publications from the Society for Risk Analysis (SRA) and the ISO 31000 standard series. These references, outlined in Table II.2, offer comprehensive insights into the concept of risk.

The definition provided clarifies that PHM focuses on forecasting and health management of engineering assets, consistent with its terminology. In contrast, risk focuses on managing uncertainty and mitigating adverse consequences. In

Table II.2 – Risk definitions

Risk definitions from Society for Risk Analysis [42]
We consider a future activity [interpreted in a wide sense to also cover, for example, natural phenomena], for example, the operation of a system, and define risk in relation to the consequences (effects, implications) of this activity with respect to something that humans value. The consequences are often seen in relation to some reference values (planned values, objectives, etc.), and the focus is often on negative, undesirable consequences. There is always at least one outcome that is considered negative or undesirable.
Overall qualitative definitions:

1. Risk is the possibility of an unfortunate occurrence
2. Risk is the potential for realization of unwanted, negative consequences of an event
3. Risk is exposure to a proposition (e.g., the occurrence of a loss) of which one is uncertain
4. Risk is the consequences of the activity and associated uncertainties
5. Risk is uncertainty about and severity of the consequences of an activity with respect to something that humans value
6. Risk is the occurrences of some specified consequences of the activity and associated uncertainties
7. Risk is the deviation from a reference value and associated uncertainties

Risk definition from ISO 31000 [43]
risk: effect of uncertainty on objectives
Notes:

- An effect is a deviation from the expected positive and/or negative.
- Objectives can have different aspects (such as financial, health and safety, and environmental goals) and can apply at different levels (such as strategic, organization-wide, project, product, and process).
- Risk is often characterized by reference to potential events and consequences or a combination of these.
- Risk is often expressed in terms of a combination of the consequences of an event (including changes in circumstances) and the associated likelihood of occurrence.
- Uncertainty is the state, even partial, of deficiency of information related to, understanding or knowledge of an event, its consequence, or likelihood.

addition, risk has a broader scope and can be applied not only in the field of engineering asset management but also in other areas, such as healthcare [44, 45], economic management [46, 47] , and environmental science [48, 49].

Furthermore, Table II.3 encapsulates the disparities between PHM and risk in the context of health management for D&M tools, spanning data, main tasks, maintenance strategies, and examples.

Table II.3 – Differences between PHM and risk in terms of data, task, maintenance strategy and examples

	PHM	**Risk**
Data	Condition monitoring information (e.g., current and voltage)	Operational environment data (e.g., temperature, vibration, and shock)
Main tasks	• Fault detection • Fault diagnosis • Degradation assessment • RUL prediction	Risk (level) estimation
Maintenance strategies	• Condition-based maintenance • Predictive maintenance	Risk-based maintenance
Examples	• Fault detection models [7, 50, 51] of subsystems in LWD tools • Fault diagnosis model for power supply boards in LWD tools[30] • Degradation assessment model for neutron generator subsystems in LWD tools[31]	• Risk level estimation model of electronic boards in D&M tools [52] • Risk-based D&M tool selection and maintenance [53] • Risk-based D&M fleet maintenance planning [54]

II.3.2 Degradation state vs. risk level

PHM includes degradation assessment, while risk involves the estimation of risk levels. In the specific area of state estimation, one may argue that the level of risk has similarities to the state of degradation, but the two are fundamentally different in nature. The rationale for this difference is elaborated below.

1. A degraded state of a machine means a gradual decline in performance over time, which in turn leads to malfunctions and failures [55]. This performance degradation can be reflected in a graph of the calculated health indicator values over time. However, such tangible changes are often difficult to detect when considering risk level estimations of electronic boards, that is, the critical subsystems of D&M tools.
2. The modeling of degradation traditionally finds application in mechanical systems or structures, where condition monitoring signals undergo variations in response to tangible alterations, such as wear and tear on material surfaces. In contrast, electronic systems (or boards) may sustain functionality even with elevated risk levels, with observable physical changes on the system's surface being infrequent. In contrast, microscopic changes may occur within the system but are challenging to detect.
3. Risk can be applied to characterize the operational environment, a capability that degradation assessment lacks. Operational environment characterization is critical for ensuring the reliability of systems deployed across diverse mission profiles and locations, where varying operational environments exert distinct influences on system reliability. As a more comprehensive concept, risk facilitates quantifying the hazardous degree associated with different operational environments.

Building upon the descriptions and analyses provided in the preceding two sections, it can be inferred that risk estimation and risk-based maintenance are better suited for systems, such as D&M tools, possessing the following characteristics:

— Containing electronics and electronic subsystems are critical

— Insufficiency of condition monitoring information or effective PHM models

— Utilization in diverse missions across different locations, implying distinct operational environments

In the subsequent section, we will propose a risk-based maintenance decision framework for systems that have the abovementioned features.

II.4 Risk-based maintenance decision-making framework

Similar to the widely accepted PHM cycle shown in Figure II.5, this section introduces a generic framework of risk-based maintenance decision-making. This framework comprises four pivotal layers: data, risk estimation, decision, and HMI, as shown in Figure II.6. These layers are described in detail in the following section.

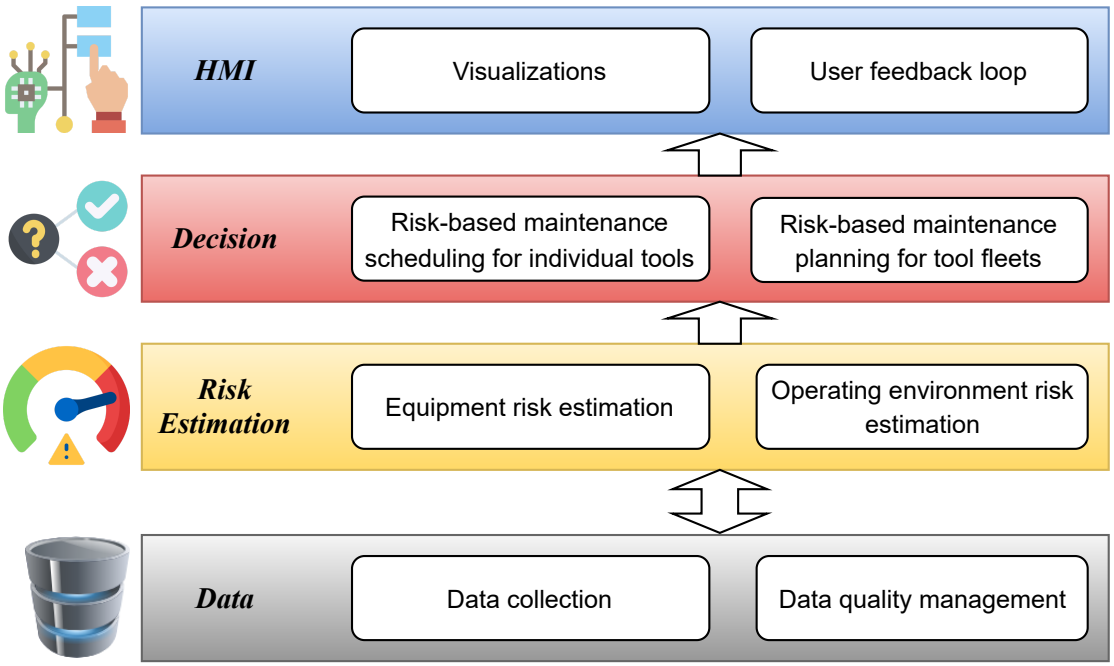


Figure II.6 – Risk-based maintenance decision-making framework

II.4.1 Data

This risk-based maintenance decision-making framework is data-driven. Thus, the data layer is the basis for all subsequent risk assessment and maintenance decisions. The data layer consists of two key aspects: data collection and data quality management.

— **Data Collection:** This process is similar to the data acquisition layer in the traditional PHM cycle shown in Figure II.5, aiming to gather comprehensive data aligning with the scope of risk estimation and risk-based decisions. It involves identifying and seamlessly integrating data from diverse sources, including tool or equipment measurement data, historical maintenance records, asset information, and domain expert knowledge. Table II.4 succinctly outlines the primary data sources utilized in this study, offering insights into the richness and diversity of the data landscape.

— **Data Quality Management:** As explained in the General introduction chapter, measurement data collected from equipment operations in the field often encounter various data quality issues such as noise, missing values, and inconsistencies. Unlike the traditional PHM cycle shown in Figure II.5, this risk-based maintenance decision-making framework incorporates data quality management. The data quality management is dedicated to assessing data quality, determining data quality requirements, and improving data quality. Data quality management is closely related to risk estimation models. Data quality requirements are dependent on the performance assessment and requirements of risk estimation models. On the other hand, data quality

affects the choice of risk estimation models, and improving data quality improves the performance of risk estimation models. The bi-directional arrow between the data and risk estimation layers in the framework describe this interrelationship. Data quality management will be further elaborated in the next chapter.

II.4.2 Risk estimation

The risk estimation layer is positioned atop the data layer. For equipment or systems deployed in diverse missions across different locations, the causes of mission or operation failures typically stem from two primary sources. The first is the risk of equipment or system failure, and the second is the risk imposed by the operational environment. Therefore, the risk estimation layer is divided into two parts: equipment risk estimation and operational environment risk estimation.

— **Equipment Risk Estimation:** This process aims to estimate equipment risk, in other words, the likelihood of equipment failure, using historical measurements collected by the data layer. The process involves identifying critical components or subsystems of the equipment. In systems laden with electronics, these critical subsystems typically include electronic boards. Thus, it is crucial to assess the failure risk of these electronic boards based on historical operational environment measurement data, encompassing factors such as temperature, shocks, and vibrations.

— **Operational Environment Risk Estimation:** Quantifying operational environment risks is imperative, as the operational environment has a significant effect on the efficiency and success of a mission or operation. The operational environment risk can be perceived as the likelihood of exposure to a harsh or hazardous environment that could lead to mission or equipment operation failure. For example, drilling operations are run in diverse geographical locations, each characterized by distinct operational environments. Drilling operations are more prone to fail in harsh operational environments. Thus, characterizing the likelihood of exposure to a harsh or hazardous environment becomes essential for precautionary measures, such as employing a more reliable tool or upgrading existing tools.

II.4.3 Decision

Scheduling is the systematic process of aligning jobs with resources and determining their sequence for execution at a specific time. Conversely, planning involves proactively identifying the elements necessary to perform a task before its initiation [56][57]. The effectiveness of planning and scheduling significantly contributes to the reduction of maintenance costs, improved utilization of maintenance resources, and enhanced quality of maintenance work [57]. Consequently, the decision layer in this framework encompasses two key aspects: scheduling and planning.

Table II.4 – Data sources used in this study

Data source	Data sets	Data formats	Short description	Quantity	Quality
CMMS	Measurement data	• Time series • Tabular	• Measured operational environment data by built-in sensors in D&M tools, e.g., temperature, shock and vibration • Measured time information by timer or clock, e.g., pumping time, datetime stamp • Other information, e.g., where the data measured	High	Medium or low
CMMS	Asset information	Tabular	• Asset configuration, e.g., equipment id, subsystems'or components' identity information, revisions • Asset run history	Medium	High
CMMS	Maintenance log	Tabular	• Maintenance workorders, e.g., workorder id, equipment id, creator, date • Failure description and analysis by maintenance technicians (free text)	Medium	Medium
SME	Domain knowledge	Tabular	• Compatibility rule tables • Preset parameters or thresholds	Low	High

CMMS: Computerized Maintenance Management System
SME: Subject Matter Expert

Maintenance scheduling is applied at the level of a single tool, ensuring the tool is optimally utilized for a given task. In contrast, maintenance planning focuses on clusters of tools or tool fleets, with parts purchasing plans developed to meet the demand of tools for jobs in the coming months or the following year. Both scheduling and planning are based on valuable insights gained from the risk estimation layer and use risk-based approaches to optimize maintenance activities.

— **Risk-Based Maintenance Scheduling for Individual Tools:** Tools vary in configuration, reliability, and compatibility with different risk levels in operational environments. Highly reliable tools are versatile across multiple environments but often entail high manufacturing costs. Jobs are allocated to diverse locations, each presenting varying operational environments. Risk-based maintenance scheduling aims to select the most suitable tool for the specific operational environment resulting from a given job. In cases where no suitable tools are available, tool upgrade is tailored based on the associated risk of the operational environment. This systematic approach enhances tool availability by ensuring an appropriate match between tools and operational environments.

— **Risk-Based Maintenance Planning for Tool Fleets:** Regularly procuring new parts is essential to restore failed tools within a fleet. Moreover, this procurement process must consider the risks associated with anticipated operational environments in future jobs. The objective is to guarantee that adequate compatible tools are available to handle potential risks related to the operational environment. As a result, risk-based fleet maintenance planning implements fleet maintenance activities grounded in operational environment risk assessments and the compatibility of the fleet with the operational environment. This strategic approach ensures that maintenance activities are conducted strategically, facilitating resource allocation optimization.

II.4.4 Human-machine interface

The HMI layer serves as a crucial link between the analytical insights generated by the framework and the human operators responsible for making maintenance decisions. This layer also includes two components: visualizations and a user feedback loop.

— **Visualizations:** This component involves the development of intuitive and informative visualizations, such as interactive dashboards and user-friendly web applications. These visualizations are pivotal in enabling operators and decision-makers to comprehend risk estimates and maintenance recommendations quickly. The emphasis is on providing a user-friendly interface that facilitates informed and timely actions based on the insights generated by the framework.

— **User Feedback Loop:** Establishing a feedback loop through which operator expertise and domain knowledge can directly inform and refine the framework's algorithms and models. This iterative feedback mechanism fosters a continuous improvement cycle, enhancing the framework's adaptability and efficacy.

II.5 Conclusion

In this chapter, we first introduced five commonly used maintenance strategies: CM, PM, CBM, PdM, and RBM. While each strategy has its advantages, there are also disadvantages. Then, the limitations of current maintenance strategies for D&M tools, i.e., the first four maintenance strategies, are analyzed. These limitations mainly include the intricate nature of electronic board failure mechanisms and failure symptoms and insufficient condition monitoring information, which leads to the absence of effective PHM models. These shortcomings underscore the imperative need for RBM.

PHM is the basis for CBM and PdM. To further explore the need for and rationalization of RBM, we also delve into the distinction between PHM and the broader concept of risk. Drawing on various published definitions, we emphasize the multifaceted nature of PHM and contrast it with more precisely defined risk concepts from the SRA and the ISO 31000 series of standards. Notably, we emphasize that while risk levels may be similar to degradation states, they are fundamentally different, with degradation focusing on a gradual decline in performance. In contrast, risk encompasses a broader range of uncertainties. Furthermore, we highlighted the applicability of risk in characterizing the likelihood of a high-risk operational environment, a capability not shared by degradation states.

Following a comprehensive analysis of maintenance strategies and comparisons between PHM and risk, we proposed a RBM decision-making framework tailored for systems (e.g., D&M tools) characterized by an abundance of electronics, insufficient condition-monitoring information, absence of effective PHM models, and utilization in various operational environments. The framework comprises four layers: data, risk estimation, decision, and HMI. This holistic framework aims to harness the capabilities of data-driven risk estimation and decision support to enhance maintenance strategies, reduce downtime, and ultimately elevate operational efficiency and success. Subsequent chapters will delve into each of these layers (except the HMI, which is more relevant to software development) in greater detail, elucidating their roles and functions, and exploring their application within the context of D&M tool maintenance in the oil and gas industry.

Chapter III
Data quality management

III.1 Introduction

As previously discussed in the preceding chapter, the data layer within the proposed risk-based maintenance decision-making framework incorporates data quality management. This chapter will provide an in-depth exploration of data quality management concerning equipment risk estimation.

In equipment risk estimation, risk is defined as the likelihood of an unwanted event or failure in an industrial system that can lead to costly consequences such as downtime, maintenance costs, and even safety hazards [42, 43]. By quantifying the risk of failure associated with industrial equipment, organizations can prioritize maintenance tasks, reduce unplanned downtime, and extend the life of critical assets [58]. Therefore, accurate risk estimation is critical to making informed maintenance decisions.

With the rapid advancement of IoT technology, computer science, and artificial intelligence, failure risk estimation for industrial equipment has increasingly relied on data-driven models powered by machine learning algorithms. For instance, Mazumder et al. conducted a comparative study of eight machine-learning algorithms to estimate the failure risk in steel oil and gas pipelines [59]. Betz et al. introduced an innovative risk estimation model for building equipment, leveraging condition inspection data and a neural network algorithm [60]. Wang et al. utilized historical failure mode and effect analysis data to predict component or product failure risk using various machine learning classifiers [61]. In machine learning applications, three primary sources of prediction uncertainty emerge scope compliance, data quality, and model fitting [62].

Scope compliance-related uncertainty arises because of disparities between the context in which the model is developed and the real-world application context. Data quality-related uncertainty stems from limitations in data quality when applying the model, encompassing issues like missing values and noisy data. Lastly, model fitting-related uncertainty is a consequence of the inherent limitations of the learned model.

In the context of equipment risk estimation, scope non-compliance is often deemed impossible as the scope is clearly defined – it is to estimate the risk of equipment failures. Thus, two predominant factors significantly influence failure risk estimation accuracy: data quality and model fitting. However, much of the research effort in academia has been focused on solving problems related to model fit, and little research has been done on data quality. In fact, in practical machine learning applications, a significant amount of time and resources invested are devoted to data collection, cleaning, and preparation [63, 64].

The framework aims to address key issues related to data quality in industrial equipment risk estimation:

1. How can data quality from industrial equipment be assessed? This question explores indicators for assessing the quality of data collected from industrial equipment, such as data volume, completeness, accuracy, and consistency.

2. How to measure the effect of data quality on the performance of equipment risk estimation models given actual risk are partially known? Understanding the relationship between data quality and model performance is critical. This question explores techniques to quantify the effect of data quality on the accuracy of risk estimates.
3. How can it be determined whether data are sufficient for modeling risk estimates? This question addresses the process of determining whether the data collected meets the data quality required for modeling.
4. If the data are sufficient, which model is the best for modeling risk estimates? This question addresses the decision-making process of model selection.
5. If the data are substandard and insufficient for modeling risk estimates, what measures are in place to improve the quality of the data? This question explores the data quality improvement techniques including data preprocessing (e.g., missing value imputation, and outlier detection), and other data management practices.

The rest of the this chapter is organized as follows: Section 2 reviews the existing data quality management related works. Section 3 presents a comprehensive framework for managing data quality in equipment risk assessment, including data quality assessment, data quality requirement decision-making, data quality improvement, and risk estimation model development. Section 4 presents a case study to validate the effectiveness of the proposed framework, utilizing field data collected from real-world drilling operations conducted globally. Section 5 concludes this chapter and introduces the topic for the next chapter.

III.2 Related works

III.2.1 Data quality definition

Data quality has attracted significant attention and research in various domains, including, but not limited to, information management, IoT, digital manufacturing, and PHM. Many definitions of data quality have emerged in these domains. Table III.1 briefly summarizes these different definitions.

In addition to the data quality definitions found in research articles, the international standard ISO 8000 series defines data quality as the degree to which a set of inherent data characteristics fulfills requirements [69]. Another international standard [70] defines data quality as the degree to which the characteristics of data satisfy stated and implied needs when used under specified conditions. Moreover, the company IBM defines that data quality measures how well a dataset meets criteria for accuracy, completeness, validity, consistency, uniqueness, timeliness, and fitness for purpose [71].

Table III.1 – Data quality definitions

Domain	Reference	Definition
Information management	[65]	Data that are fit for use by data consumers.
IoT	[66]	How suitable the gathered data (from the smart things) are for providing ubiquitous services for IoT users.
Digital manufacturing	[67]	A measure of the agreement between the data views presented by an information system and that same data in the real world.
PHM	[68]	The data quality should reflect the suitability of data to satisfy the modeling for purposes of failure detection, diagnosis and prediction.

Although there are various definitions of data quality in different domains, a common consensus is that data is considered high quality when it is well suited for the intended purposes in the application context.

III.2.2 Data quality metrics

In the research literature, the term "data quality metrics" is synonymous with a variety of terms, including "data quality dimensions" [65], "data quality indicators" [72], and "data quality characteristics" [73]. The data quality metrics are a set of data quality attributes that represent a single aspect or construct of data quality [65]. There are two widely recognized international standards related to data quality. The first one is the standard ISO/IEC25012. In this international standard, the data quality metrics are classified into two categories: inherent and system-dependent [70].

— Inherent data quality refers to the degree to which quality characteristics of data have the intrinsic potential to satisfy stated and implied needs when data is used under specified conditions.

— System-dependent data quality refers to the degree to which data quality is reached and preserved within a computer system when data is used under specified conditions.

For inherent data quality, five data quality metrics are defined: accuracy, completeness, consistency, credibility, and currentness. On the other hand, system-dependent data quality is assessed using three metrics: availability, portability, and recoverability. Additionally, there are seven metrics that jointly consider inherent and system-dependent data quality.

The other international standard is ISO 8000-8:2015 [74]. This standard identifies three categories to measure data quality, which are

— syntactic quality: degree to which data conforms to its specified syntax;
— Semantic quality: degree to which data corresponds to what it represents;
— pragmatic quality: degree to which data is found suitable and worthwhile for a particular purpose.

In addition to the two international standards, researchers have identified various data quality dimensions from different perspectives or application contexts. For instance, In the sensor data streaming environment, Klein and Lehner adopted a set of five metrics to represent data quality. These metrics include accuracy, confidence, completeness, data volume, and timeliness [75]. Rekatsinas et al. used metrics such as coverage, accuracy, timeliness, and position bias to assess the quality of data sources, which are well suited for applications involving data integration and fusion [76]. Purnomoadi et al. applied six dimensions of data quality for asset health indexing [77]. The six dimensions are accuracy, completeness, validity, consistency, uniqueness, and timeliness. Iturry et al. found that in health records, most data quality problems are related with completeness, followed by consistency, correctness and accuracy [78]. Xu et al. defined five data quality metrics for imbalanced data in multiple products manufacturing process: free of error, appropriate amount of data, ease of manipulation, relevance, and imbalance level [79]. Merino et al. proposed the three data quality characteristics for assessing the levels of data quality-in-use in big data projects: contextual adequacy, operational adequacy and temporal adequacy [80]. From a product perspective, Wang et al. proposed a comprehensive framework consisting of four categories of data quality metrics: intrinsic data quality, contextual data quality, representational data quality, and accessibility data quality, each of which includes several subcriteria [81]. The framework provides a detailed taxonomy for assessing data quality from different perspectives.

Given the wide variety of data quality indicators, careful consideration must be given to selecting these indicators, because not all have uniform applicability or relevance in different contexts. Similar to the contextual relevance of data quality, the selection of data quality metrics should be based on the precise requirements and objectives inherent in the intended application.

Table III.2 summarizes definitions of a few widely used data quality metrics. It is noteworthy to acknowledge that the enumeration of metrics presented herein is not comprehensive; rather, it serves to provide a representative overview of the research endeavors undertaken in this domain.

III.2.3 Data quality related works in engineering asset management

While the field of data quality management has seen significant developments, it is worth noting that the majority of research efforts have been directed towards data quality measurements and monitoring within databases or information

Table III.2 – Definitions of widely used data quality metrics

Metric	Definition
Accuracy	The degree to which the data values reflect the actual event state in a specific context of use [70].
Completeness	The ratio of complete elements. An element can refer to any data unit, e.g., an attribute, a record, or a table [82].
Data volume	The number of raw data items (values) available for use to compute a result data item [66].
Consistency	The degree to which the data's format and value conform to the predefined schema [83].
Timeliness	The probability that an attribute value is still up-to-date [84].
Currency	Time difference between when data are stored in the system and when data are updated in the real world [82].

management contexts [82, 85]. In contrast, the domain of engineering asset management, which deals with critical industrial systems, has received relatively less attention in terms of data quality management research.

Existing research work about data quality management in engineering assets management mainly focuses on different data quality issues in various engineering asset management tasks, including failure detection, fault diagnosis, degradation assessment, maintenance decision-making, and structural health monitoring. Table III.3 provides a concise overview of the explored data quality issues of these data quality-related works.

The existing body of work in the literature has primarily concentrated on discrete facets of data quality management, often lacking discussions on holistic frameworks that encompass the complete data lifecycle—from data generation to the development of data-driven models. Additionally, there has been a noticeable scarcity of emphasis on data quality within the specific context of equipment risk estimation. Consequently, there is a discernible need for a comprehensive data quality management framework tailored specifically for equipment risk estimates. The subsequent section will introduce a novel data quality management framework designed to address this specific requirement.

Table III.3 – Summary of data quality related works in engineering asset management

Reference	Task	Explored data quality issue
[86]	Fault diagnosis	Data quality requirement
[87]	Fault diagnosis	Data quality assessment
[68]	Fault diagnosis	Data quality assessment and improvement
[87, 88]	Fault detection, fault diagnosis, and degradation assessment	Data quality assessment
[89]	Degradation assessment	Data quality assessment and improvement
[90]	Maintenance decision-making	Data quality improvement
[91]	Maintenance decision-making	Data quality assessment
[92]	Maintenance decision-making	Investment decision-making for data quality improvement
[93]	Structural health monitoring	Data quality metric selection

III.3 Proposed data quality management framework

The proposed data quality management framework for equipment risk estimation is illustrated in Figure III.1. This comprehensive framework is composed of five phases: data development, data quality assessment, data quality requirement, data quality improvement, and risk estimation model development. Each of these phases is explained in detail in the following paragraphs.

III.3.1 Data development

As mentioned in the Introduction section of this chapter, the cornerstone of building data-driven models for estimating equipment failure risk lies in the data, because the efficacy and usefulness of these models depend heavily on the data quality. This section on data development outlines the four key steps required to formulate and process the data, that is, data collection, data preprocessing, feature extraction, and data labeling.

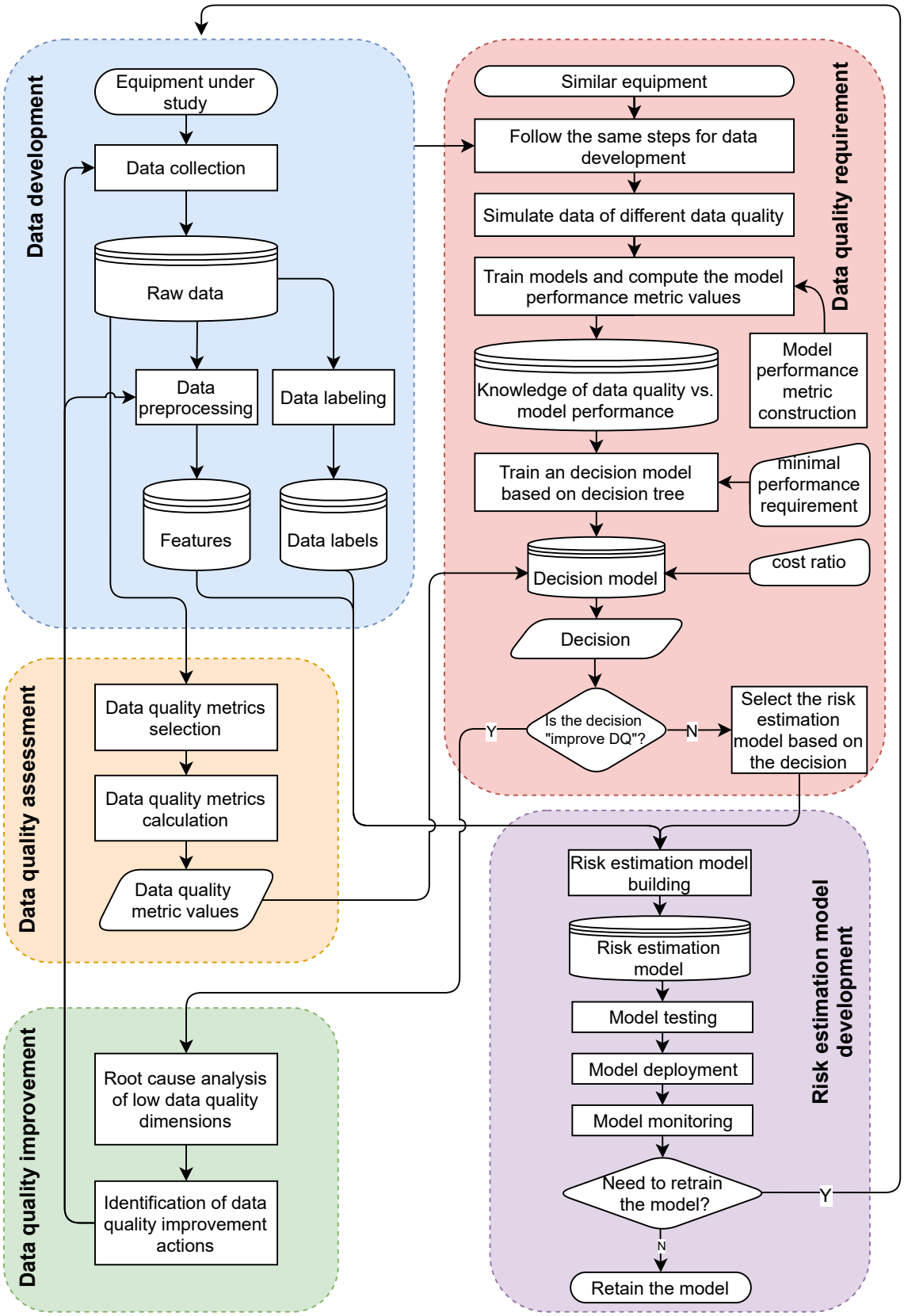


Figure III.1 – The data quality management framework

III.3.1.a Data collection

In the risk-based maintenance decision-making framework outlined in the previous chapter, the data layer includes data collection that extends beyond equipment risk estimation to include operational environment risk estimation and data for the decision layer. In the specific context of data quality management discussed in this chapter, the first step is also data collection. However, the scope of this data collection is smaller, which can be consider as a subset of the data collected of the data collection in the data layer. Typically, these data come from the Computerized Maintenance Management System (CMMS) associated with the equipment. A CMMS embodies software infrastructure tailored to centralize maintenance intelligence and streamline the orchestration of maintenance activities. It is designed to optimize the utilization and availability of physical equipment such as machinery, vehicles, transportation infrastructure, plant facilities, and associated resources [94].

An extensive spectrum of information is stored within the CMMS database, encompassing details about equipment identity, operation data, work orders, materials inventory, and more. Equipment operation data includes readings taken from various sensors mounted on the equipment. Some of these sensors are attuned to multiple aspects of the operational environment, including temperature sensors employed to gauge equipment temperature and accelerometers used to quantify equipment vibration. Additionally, specific sensors, such as transmitters and receivers, are integral to the equipment's function. The scope of work orders include various categories, including maintenance, equipment order, shipment, and related operational tasks. To provide a more straightforward overview of the primary data stored in a CMMS, Table III.4 presents a summary of the key elements.

As evident from the preceding description, the CMMS database is abundantly stocked with a profusion of data. However, utilizing all these data to estimate the risk of equipment failure is neither practical nor necessary. In practice, it is common to selectively extract specific subsets of data from the CMMS database. For equipment failure risk estimation, it is often required to collect equipment identity information, equipment run history, operational environment data (e.g., temperature, vibration) in equipment measurement data, and associated maintenance work orders. This careful selection of data elements emphasizes a pragmatic approach to identifying the data needed for accurate equipment risk assessment.

III.3.1.b Data preprocessing

Data preprocessing is a process of refining and reshaping raw data into a format suitable for subsequent risk estimation model training. Traditionally, this process requires significant engineering effort and is characterized by iterative improvement through rigorous trial and error handling [95].

As shown in Figure III.2, data preprocessing consists of four key steps, namely data cleaning, feature extraction, feature transformation, and feature reduction, each of which plays a different role.

Table III.4 – CMMS data summary

Data group	Sub-group	Characteristics
Equipment data	Equipment identity information (e.g., equipment id, revision number, equipment type, status, location)	structured
	Equipment run history (e.g., equipment id, run id, run start and end datetime, run duration)	structured
	Equipment measurement data (e.g., run id, equipment id, temperature, vibration, pressure, current, voltage, and other measured parameters)	unstructured or semi-structure, time series
	Equipment manuals (e.g., operation manual, repair manual)	unstructured, text, images, videos
Workorder data	Equipment order (e.g., order id, client, order date, equipment id, cost)	structured
	Equipment shipment (e.g., shipment id, order id, delivery company)	structured
	Equipment maintenance workorder (maintenance workorder id, equipment id, failure description, shop analysis, failure consequence)	structured, text
Inventory data	inventory records (e.g., material type, amount, cost)	structured

1. **Data cleaning** involves carefully identifying and correcting errors and inconsistencies in the dataset, such as missing values, outliers, and duplicates. In Section III.3.4, the authors will describe missing-value handling, outlier detection, and data deduplication in detail; the readers can refer to that section for more information on data-cleaning methods. It is important to note that industrial equipment operational data cleansing typically requires additional steps beyond regular data cleansing. These extra steps usually rely on the knowledge of Subject Matter Expert (SME)s, who have developed superior expertise and experience in the equipment. SMEs may belong to diverse fields, including reliability, electrical, and physics engineering, and bring a wealth of specialized knowledge to bear on the task. For example, it is often recommended that operational data during equipment startup and

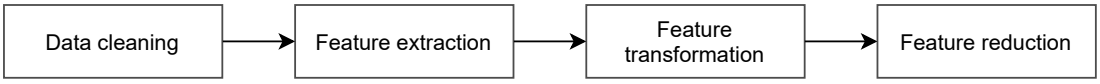


Figure III.2 – Data preprocessing steps

shutdown be deleted, because data recorded during these periods usually contains too much noise due to unstable operation. Therefore, it is necessary to rely on SME knowledge to determine the stable operation phase of the equipment[30, 50]. Additionally, leveraging unsupervised methods can aid SMEs in effectively exploring data within a specific domain [96].

2. **Feature extraction** aims to extract discriminative features from the raw data that can be consumed by failure risk estimation models. Many studies have been conducted on general-purpose equipment such as gearboxes and motors. Common statistical features can be extracted for these types of equipment based on existing methods such as time-domain, frequency-domain, and time-frequency-domain analysis methods [97]. However, for specialized equipment, such as D&M tools in the oil and gas industry, the feature extraction process usually requires SMEs to guide the feature extraction. Even though there are numerous published works of equipment prognostics based on deep learning, these methods can automatically learn features through deep networks. However, the transfer of deep learning to real industrial applications is limited because of its weak interpretability and computational complexity [98]. In many real-world industrial artificial intelligence applications, conventional feature extraction techniques are still favored to ensure interpretability of results and reduce sensitive information [95].
3. **Feature transformation** involves transforming original features into new representations to improve model performance. Commonly used feature transformation methods include normalization and standardization. Normalization scales the data to a common range like [0,1] or [-1,1], while standardization converts the data to zero mean and unit variance. Both methods aim to promote uniformity of data magnitude across attributes, ensure that all features contribute equally to the model, and avoid the dominance of features with larger values. Other techniques include Box-Cox transformation for transforming skewed data and generating new features by multiplication between features.
4. **Feature reduction** includes two approaches: feature selection and dimensionality reduction.

    - **Feature selection** is a process that entails the choice of a subset of input features from a dataset. Feature selection methods can be classified into two categories: unsupervised and supervised [99].
        - **Unsupervised Feature Selection:** Unsupervised methods do not rely on the target variable (e.g., the failure risk in the context of equipment risk estimation) for selection. Instead, these methods eliminate redundant features based on correlations among features. The goal is to retain only the most informative features while removing highly correlated ones, which can lead to multicollinearity.

— **Supervised Feature Selection:** Supervised methods, in contrast, employ target variables in the selection process. These methods can be further categorized into three subtypes: wrapper methods, filtering methods, and embeded methods [100].

- **Wrapper Methods** iteratively search for a subset of features that yield best model performance. Examples include backward elimination, where features are removed step by step, and forward selection, which adds features incrementally.
- **Filtering Methods** select a subset of features based on their relationship with the target variable. This relationship is quantified using various metrics like Pearson's correlation coefficient or Spearman's rank correlation, depending on the nature of the variables involved (continuous or categorical). Features are ranked or scored, and a predetermined threshold is applied to select the best subset.
- **Embedded Methods** integrate feature selection into the model training process. Features are automatically chosen during model training based on their importance in predicting the target variable. Examples of embedded feature selection methods include decision trees, lasso regression, and ridge regression.

The choice of which feature selection method to implement depends on the specific problem, the dataset's characteristics, and the goals of the analysis. Each method has advantages and limitations, and selecting the most suitable approach is essential to optimize model performance and interpretability.

■ **Dimensionality reduction** is another aspect of feature reduction that aims to transform high-dimensional features into a low-dimensional space while retaining salient information. One of the most widely used methods of dimensionality reduction is principal component analysis.

III.3.1.c Data labeling

Data labeling is an essential process in data development. It involves attaching one or more meaningful and informative labels to raw data, usually time series sensor data, in the context of industrial equipment failure risk estimation. These labels typically convey information regarding the equipment's health status (or fault mode) and the underlying failure mechanisms, enabling data practitioners (e.g., data scientists) to select correct data for model training.

Maintenance work orders can also be initiated in response to suspected equipment failures; notably, maintenance technicians rather than maintenance experts often record the failure description and shop analysis provided in maintenance work order data. As a result, there can be uncertainty regarding the accuracy of the failure reports in the maintenance work order and the identification of the failure root cause.

Furthermore, unlike more common annotation tasks, such as image or text annotation, which can be assigned to individuals with general expertise, labeling data derived from industrial equipment sensor readings is a more complex and costly endeavor. This complexity arises primarily from the need for an in-depth understanding of equipment operations, maintenance protocols, and the underlying mechanisms of failure. Consequently, the responsibility for labeling industrial equipment sensor data is usually entrusted to SMEs. SMEs play a critical role in reviewing the sensor data, along with the failure descriptions and shop analyses contained in the associated maintenance work order data to validate the failure's occurrence and its root cause. To gain a deeper understanding of the failure and its contributing factors, in some instances, failed equipment may return to the technology center to undergo a more extensive investigation process, where a detailed analysis is conducted.

III.3.2 Data quality assessment

Data quality assessment is one of the five key phases in data quality management. It aims to evaluate the suitability of a dataset for its intended purpose. Section III.2.2 described widely used quantitative data quality metrics. These metrics enable data practitioners to calculate values that offer insights into the data's fitness for use. The process of data quality assessment consists of the following two key steps:

1. **Data Quality Metric Selection**: This initial step involves identifying and selecting pertinent data quality metrics. The choice of metrics should align with the specific application context and the data characteristics. Heinrich et al. proposed five requirements for data quality metrics, namely, the existence of minimum and maximum metric values (R1), the interval scaling of the metric values (R2), the quality of the configuration parameters, and the determination of the metric values (R3), the sound aggregation of the metric values (R4), and the economic efficiency of the metric (R5) [101]. These criteria support both decision-making under uncertainty and economically oriented data quality management.
2. **Data Quality Metric Calculation**: Once the relevant data quality metrics are chosen, the next step is to apply these metrics to the dataset. This step involves calculating the metric values based on the dataset's characteristics and the definitions of the selected metrics. For example, completeness metrics involve calculating the percentage of missing values, while data volume metrics assess the number of samples.

III.3.3 Data quality requirement

Once the data quality assessment has been completed, it is necessary to determine whether the data quality meets the data quality requirements. A straightforward approach is establishing thresholds for data quality metrics, but determining these thresholds remains challenging.

This chapter proposes a new method for determining if the data quality meet specific requirements. The method is based on the relationship between model performance and data quality, and a decision tree model. In addition, model performance is evaluated based on the average maintenance cost. Thus, this method for determining data quality requirements takes costs into account. This section will first present an indicator for assessing the performance of the equipment risk estimation model. Subsequently, it will describe how to acquire knowledge of the relationship between model performance and data quality. Finally, this section will outline the process of constructing a decision model to determine data quality requirements.

III.3.3.a Risk estimation model performance metric

When assessing the risk of equipment failure in an industrial environment, it is often challenging to determine the actual risk of failure over time, which is especially true for complex equipment with particularly complicated failure mechanisms. In such cases, available data usually only provide information on when equipment failures occurred.

On the other hand, cost factors play a pivotal role in maintenance decisions guided by risk estimation. Consider the scenario where a critical component (or subsystem) is the primary driver of the equipment failures, and this component cannot be repaired but only replaced. In this case, the risk-based maintenance decision-making process involves three principal costs: component replacement cost, cost associated with undetected failures, and cost associated with premature component replacement. The first of these costs is often deterministic, while the latter two depend on the accuracy of the risk assessment model.

Specifically, suppose the risk estimation model can predict the component failures accurately. In that case, the components can be replaced at the optimal time, thus avoiding any losses due to failures or premature replacement. However, if the model predicts the failure too late, it can incur costs associated with equipment failures. Conversely, if the model predicts failures too early, it may lead to unnecessary component replacement costs.

Given the above analysis, this chapter proposes a loss function for assessing the performance of a risk estimation model as defined in Equation (III.1).

$$\ell = \frac{\sum_{i=1}^{N} \left[c_1 \mathbb{I}(\hat{T}_i \geq T_i) + c_2 (T_i - \hat{T}_i) \mathbb{I}(\hat{T}_i < T_i) \right]}{N}, \tag{III.1}$$

where

— N: the number of components.

— $\hat{T}_i$: the time when the component i is replaced based on failure risk estimate, assuming all components' lives start at time 0; specifically, the component i is replaced when the failure risk estimate reaches a certain level.

— T_i: actual life of component i, i.e., the time when the component actually failed.

— c_1: unit failure cost, i.e., the cost caused by one undetected failure.

— c_2: premature replacement cost per unit of time.

— $\mathbb{I}$ is an indicator function. In other words, $\hat{T}_i \geq T_i$ means the component i is replaced too late, which incurs failure cost, while $\hat{T}_i < T_i$ means the component i is replaced too early, which incurs premature replacement cost.

Different components could have different unit failure cost and premature replacement cost. In order to capture the effect of these cost differences on the loss function and reduce the number of cost parameters, a new parameter called cost ratio (denoted as $r = c_1/c_2$) is introduced; then, Equation (III.1) can be written as Equation (III.2).

$$\ell = c_2 \left[\frac{c_1}{c_2} \times \frac{\sum_{i=1}^{N} \mathbb{I}(\hat{T}_i \geq T_i)}{N} + \frac{\sum_{i=1}^{N} (T_i - \hat{T}_i)\mathbb{I}(\hat{T}_i < T_i)}{N} \right]. \tag{III.2}$$

Since the cost parameters c_1 and c_2 are constants for the component, Equation (III.2) can be further reformulated as Equation (III.3) by substituting r for c_1/c_2.

$$\ell \propto r \times \underbrace{\frac{\sum_{i=1}^{N} \mathbb{I}(\hat{T}_i \geq T_i)}{N}}_{\text{term 1}} + \underbrace{\frac{\sum_{i=1}^{N} (T_i - \hat{T}_i)\mathbb{I}(\hat{T}_i < T_i)}{N}}_{\text{term 2}}. \tag{III.3}$$

This new expression consisting of two different terms as shown in Equation (III.3). The first term can be interpreted as the average undetected failures, while the second term can be interpreted as the average premature replacement time. Both these terms are influenced by the data quality, while the parameter r is inherent to the component itself. The inclusion of r is essential as different components exhibit distinct values for ratio between unit failure cost and premature replacement cost per unit of time.

III.3.3.b Knowledge of data quality vs. model performance

Failure risk estimation of industrial equipment is inherently contextual. Different types of equipment have different characteristics and monitoring parameters and, therefore, cannot be cross applied. For example, it would be unwise to attempt to use a model trained on gearbox data for electronic boards, because the fundamental nature of these equipment types and their associated data varies widely. Consequently, the knowledge gained through simulation studies on publicly available datasets, which are often not correlated with the equipment under study, cannot reflect the true relationship between the quality of the data from the equipment under study and the performance of the model.

Adding to this challenge is that obtaining large amounts of high-quality data for newly implemented or infrequently used equipment can be daunting. This chapter proposes a new approach for indirectly acquiring knowledge about the relationship between data quality and model performance, as illustrated in Figure

III.1, that is, using data from similar equipment, which have more data than the equipment under study. Based on the data from similar equipment, knowledge about the relationship between data quality and model performance can be obtained through simulation studies. In this chapter, this knowledge is succinctly represented as $\mathcal{K}(\mathbf{Q}, \mathbf{\Omega}, r, \ell)$, where $\mathbf{Q}$ denotes a vector containing data quality metrics, $\mathbf{\Omega}$ represents risk estimation models, r is the cost ratio, and ℓ corresponds to the previously defined loss function in Equation (III.3).

The simulation studies are based on carefully processing data from similar equipment. By selectively removing or modifying data segments, data with different levels of data quality can be simulated. These synthetic datasets can be used to train risk estimation models, thus effectively exploring the relationship between data quality and model performance. Subsequently, these trained models are tested on the same test dataset and compared by the loss function values. For more robust assessments of model performances, it is recommended that cross-validation techniques are used.

III.3.3.c Decision model

Once the knowledge $\mathcal{K}(\mathbf{Q}, \mathbf{\Omega}, r, \ell)$ is obtained, along with the minimum performance requirement, it becomes feasible to develop a decision model using the decision tree algorithm. The choice of the decision tree algorithm in this chapter is motivated by its simplicity, ease of understanding, and the capacity to visualize the decision model. The decision model can be mathematically expressed as in Equation (III.4). The developed decision model can then determine whether the data quality of the equipment under study should be improved and, if not, which risk estimation model is the best.

$$D = g(C, \mathcal{K}(\mathbf{Q}, \mathbf{\Omega}, r, \ell)), \tag{III.4}$$

where C is the minimum performance requirement. It can be determined based on the average cost requirement thanks to the definition of the loss function. D is the decision predicted by the decision model.

A detailed case study will be presented in Section III.4 to illustrate the practical application of the method to acquire the knowledge $\mathcal{K}(\mathbf{Q}, \mathbf{\Omega}, r, \ell)$ and make decisions based on the decision model.

III.3.4 Data quality improvement

Data quality improvement involves two key steps: analyzing the root causes of low data quality dimensions and identifying data quality improvement actions.

1. **Root Cause Analysis of Low Data Quality Dimensions**: The first step in the data quality improvement process is to examine the dimensions of low data quality to discover the root causes of low data quality, which requires a comprehensive understanding of the data generation mechanisms and collection process. The root causes can be grouped into three categories:

hardware-related issues, software-related issues, and human factors. Hardware problems may include insufficient sensor accuracy, capacity limitations and physical damage of memory storage boards, and communication errors during data transfer from lower to upper computer systems. These problems are rooted in the design and infrastructure of the hardware system. Software issues include data loss or inconsistency due to CMMS system migration or limitations. In addition, data loss issues may occur during the data collection process, which often stems from human error, including inadvertent data overwriting and field engineers neglecting to upload data to the server.

2. **Identification of Data Quality Improvement Actions**: After thoroughly analyzing the root causes of data quality deficiencies, the next step is to develop corresponding data quality improvement measures. Improvement measures may include a variety of strategies and initiatives, each aligned with the specific causes and dimensions of data quality that need to be improved. Based on the above analysis of the root causes of low data quality, the data quality of industrial equipment can be improved in the following three areas:

   - **System Upgrades:** This area focuses on improving data quality from a hardware and software technology perspective. For example, deploying enhanced sensors with higher accuracy and upgrading communication systems can reduce measurement errors and inaccuracies, directly affecting data quality. Choosing a powerful, stable, and mature CMMS can also help manage equipment data, avoiding data loss or inconsistency caused by frequent CMMS migrations.
   - **Management Improvement:** Management improvement centers on optimizing data quality from a management and human factor perspective. Many data quality issues often stem from human error or negligence. As analyzed above, data loss can be due to field engineers forgetting to transfer the data from the built-in memory storage boards to the server. To minimize such issues, increased training and introduction of Key Performance Indicator (KPI)s, such as data collection ratios, for field engineers can increase their awareness of the importance and responsibility of data collection.

     Both technical and management improvements require a significant investment of time and resources. In addition, they require sustainable commitment and ongoing efforts to significantly improve data quality. They are essential for continuous data quality improvement but may not produce immediate results and/or business value.
   - **Data Preprocessing Improvements:** In contrast, data preprocessing improvements offer an immediate and practical approach to improving data quality. One can quickly resolve some data quality issues by employing data preprocessing techniques such as deduplication, outlier detection, and missing value imputation.

— **Data deduplication** aims to compress data through removing duplicated data items and replacing them with a pointer to the unique remaining copy [66]. Intrinsically, it reduces the amount of data and affects the data quality of data volume.

— **Outlier detection** focuses on finding observations significantly different from most data [102]. Outlier detection methods can be categorized into four groups: statistical-based, distance-based, density-based, and clustering-based methods [103]. Statistical-based methods rely on statistical techniques to identify outliers. Distance-based methods assess the dissimilarity or distance between data points to determine outliers. Density-based methods focus on the density distribution of data points. They identify outliers as data points existing in regions of low data density. Clustering-based methods seek to partition data into clusters, with outliers being data points that do not conform to any cluster or belong to small, isolated clusters. For more details on outlier detection methods, readers can refer to two review articles: [103] and [104].

— **Missing value imputation** attempts to replace missing data with estimated values. Missing value estimation methods can be categorized into two types: statistical-based and machine learning-based method [105, 106]. Statistical-based methods rely on statistical measures and patterns within the data to impute missing values. Widely used statistical-based techniques include expectation maximization, linear regression, and imputation using the mean or mode of the available data. Machine learning leverages algorithms and models to predict and impute missing values. Typical machine learning-based techniques for missing value imputation include regression trees, random forests, support vector regression, and k-nearest neighbor. For more details on missing value estimation, readers can refer to two review articles: [105] and [106].

III.3.5 Risk estimation model development

Once it has been determined that the data quality meets the minimum requirements, developing a risk assessment model can begin. This phase involves well-defined steps, including model building, testing, deployment, and monitoring. These critical steps in the model development phase are described in detail below.

1. **Model Building**: The foundation of risk estimation lies in building the model. In this initial step, data practitioners use the features and data labels from the data development phase to create a robust, accurate model that effectively captures the relationship between input data and risk estimates. The risk estimation model can be formulated as in Equation (III.5).

$$Risk = \mathbf{\Omega}(\mathbf{X}, \mathbf{y}), \tag{III.5}$$

where **X** are the extracted features from the data development phase, and **y** are the data labels. It is important to note that the data labels here are not the equipment's failure risk, because the actual risk is often difficult to access, as described earlier. The data labels here are more of failure modes or failure mechanism analysis for each device, which helps data practitioners select the right data for training risk estimation models.

2. **Model Testing**: Rigorous testing is critical to assess the performance and reliability of the risk estimation model. In this step, the model built above is delivered for field testing to a small group of users who apply the model to new unseen data. Testing helps to identify any issues, such as over- or under-fitting, and ensures that the model generalizes risk estimates well to the new data.
3. **Model Deployment**: Once the model has performed satisfactorily during testing, the machine learning engineer or software engineer can deploy it into a production environment, including integrating the model into a system or application for real-time or batch processing of risk estimates. Factors to consider when deploying the model include scalability, reliability, and ensuring the model is synchronized with the latest data.
4. **Model Monitoring**: Model monitoring is a continuous process to ensure that models continue to perform accurately and reliably in the production environment, including tracking KPIs, detecting deviations from expected behavior, and initiating corrective action where necessary. Monitoring may also include periodically retraining the model with new data to adapt to changing patterns and maintain its predictive accuracy. In addition, monitoring the model can help identify and mitigate potential biases that may arise.

In summary, developing a risk estimation model is a comprehensive process involving multiple players, starting with sufficiently good-quality data and then building, testing, deploying, and monitoring the model. Each of these steps is critical to ensure that the model performs well in the initial stages and maintains its validity and fairness in real-world applications.

III.4 Case study

III.4.1 Data development

The raw data used in this case study were collected during drilling operations in multiple fields worldwide. These drilling operations varied in terms of duration and operating environment. Specifically, operating environment data (i.e., temperature, shocks, and vibration) of 554 failed Central Processing Unit (CPU) boards of a type of RSS (similar equipment, as shown in Figure III.4) were collected. The lifetimes of these boards span a range from 500 to 3000 hours. For the CPU board

of a type of LWD tool (equipment under study, as shown in Figure III.3), data were gathered from 18 failed boards. These boards had lifetimes ranging from 700 to 3700 hours. Among these, 12 boards were utilized as training data, while the remaining six boards, which had nearly complete data, were designated for use as test data. All the boards mentioned have been confirmed as failed by SMEs, and their raw data are stored in the CMMS.

The data preprocessing steps are adapted from the authors' previous work in [52]. However, due to space limitations and data preprocessing is not the core of this chapter, the authors will not delve into the details of data preprocessing here. Interested readers can refer to [52] or the next chapter for a comprehensive understanding of how features are extracted from the raw time series data of temperature, shock, and vibration for the CPU board.

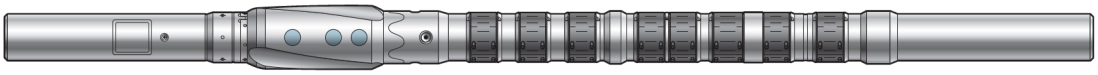


Figure III.3 – LWD tool

Figure III.4 – RSS

III.4.2 Data quality assessment

As mentioned earlier, the selection of data quality metrics should align with the specific application context and the characteristics of the data. In this case study, the equipment failure risk estimation is conducted offline, meaning it occurs after the tool has been pulled up from the oil well and is not during drilling operations. Therefore, metrics related to timeliness and currency are not crucial. Additionally, the sensors in the tool are assumed to be robust, and the readings are considered correct. As a result, there is no need to specifically evaluate the accuracy of the raw data. Instead, the operational environment data of the drilling tool may have missing values. Moreover, data volume is typically considered important for data-driven models. Hence, in this case study, the selected data quality metrics are completeness (denoted as $Comp$) and data volume (denoted as n). Data volume is the number of failed CPU boards, while completeness is the life cycle data coverage of the board.

Based on these two metrics, the data quality of the training data for the equipment under study is calculated to be $Q = [12, 0.76]$. Similarly, the data quality of the test data for the equipment under study is calculated to be $Q = [6, 1]$, where $\mathbf{Q} = [n, Comp]$.

III.4.3 Data quality requirement

III.4.3.a Acquire the knowledge

The procedures for acquiring the knowledge $\mathcal{K}(\mathbf{Q}, \mathbf{\Omega}, r, \ell)$ are shown in Algorithm 1. In this case study,

— the number of simulations $m = 60$,

— the data volume sequence $\mathbf{L}_1 = [2, 3, 4, \ldots, 30]$,

— the completeness sequence $\mathbf{L}_2 = [0.50, 0.55, 0.60, \ldots, 1.00]$,

— the cost ratio sequence $\mathbf{L}_3 = [50, 100, 200, 400, 500, 1000, 1500, 2000, 3000, 4000, 5000, 6000]$,

— the number of test boards $n_{test} = 50$.

Four models are compared in this case study, namely, Mean Time To Failure (MTTF), Median Time To Failure (MeTTF), Quantile Regression (QR), and HMM. When calculate the loss function values for MTTF, the test CPU board is replaced when it reaches the mean time to failure of training boards/data; for MeTTF, the test CPU board is replaced when it reaches the median time to failure of training boards; for QR, the test CPU board is replaced when the predicted remaining lifetime is less than 200 hours; for HMM, the test CPU board is replaced when the estimated risk level reaches the highest risk level.

Table III.5 gives an excerpt of the acquired knowledge $\mathcal{K}(\mathbf{Q}, \mathbf{\Omega}, r, \ell)$. One can also store the average undetected failure, i.e., first term in Equation III.3, and premature replacement time, i.e., second term in Equation III.3 to preserve the knowledge and then calculate the loss using Equation III.3. This way, one can save storage space.

III.4.3.b Build the decision model

Based on the acquired knowledge, one can decide which model is best by comparing the loss function values if the data volume, completeness, and cost ratio are known. For example, based on the acquired knowledge shown in the first row of Table III.5, it can be inferred that the best model is MTTF when the data volume, the completeness, and the cost ratio are 5, 0.5, and 400, respectively, because MTTF obtains the minimal loss function value among the four models. Figure III.5 shows the best model under different data volumes, completeness, and cost ratios. The figure indicates that the QR model tends to perform best when the data volume or the cost ratio is high, the HMM model excels when the completeness is high, and MTTF is favored when the completeness and the cost ratio are low. MeTTF emerges as the top choice in very few scenarios. To enhance precision in model selection, a decision tree is constructed on top of the information presented in Figure III.5. The decision model for determining the best model without minimal performance requirement is shown in Figure III.6. The term "unused" in the legend indicates the presence of the corresponding category in the training data used to train the decision tree model. However, since their number is very small,

Algorithm 1 Knowledge $\mathcal{K}(\mathbf{Q}, \mathbf{\Omega}, r, \ell)$ acquisition

Input: entire dataset of the CPU in the RSS tool, the number of simulations m, sequence $\mathbf{L}_1$ containing the numbers of training CPU boards, sequence $\mathbf{L}_2$ containing the completeness of each training CPU board, sequence $\mathbf{L}_3$ containing the cost ratios, the number of test boards n_{test}

Output: $\mathcal{K}(\mathbf{Q}, \mathbf{\Omega}, r, \ell)$

1: **for** $i \in \{1, 2, 3, \ldots, m\}$ **do**
2: sampling observations of n_{test} boards from the entire dataset without replacement as the test data
3: **for** n in $\mathbf{L}_1$ **do**
4: sampling observations of n boards from the remaining data without replacement as temporary dataset $Temp$
5: **for** $Comp$ in $\mathbf{L}_2$ **do**
6: remove some observations from $Temp$ to make the completeness of each board equal to $Comp$, use the data after removal operation as the training data
7: train the four candidate risk estimation models $\mathbf{\Omega}$ using the training data
8: predict the replacement time of test boards using the models $\mathbf{\Omega}$
9: **for** r in $\mathbf{L}_3$ **do**
10: calculate the loss function and store the result in $\mathcal{K}_i(\mathbf{Q}, \mathbf{\Omega}, r, \ell)$
11: **end for**
12: **end for**
13: **end for**
14: **end for**
15: calculate the average loss over m simulations, i.e., $\mathcal{K}(\mathbf{Q}, \mathbf{\Omega}, r, \ell) = \frac{1}{m} \sum_{i=1}^{m} \mathcal{K}_i(\mathbf{Q}, \mathbf{\Omega}, r, \ell)$

Table III.5 – Excerpt of the acquired knowledge $\mathcal{K}(\mathbf{Q}, \mathbf{\Omega}, r, \ell)$

n	$Comp$	r	ℓ_{HMM}	ℓ_{MTTF}	ℓ_{MeTTF}	ℓ_{QR}
5.00	0.50	400.00	618.04	479.89	507.74	783.38
5.00	0.50	500.00	652.42	538.09	559.49	798.96
5.00	0.50	1000.00	824.29	829.14	818.24	876.88
5.00	0.50	1500.00	996.17	1120.18	1076.99	954.79
5.00	0.50	2000.00	1168.04	1411.22	1335.74	1032.71
5.00	0.50	3000.00	1511.79	1993.30	1853.24	1188.54
5.00	0.50	4000.00	1855.54	2575.39	2370.74	1344.38
5.00	0.50	5000.00	2199.29	3157.47	2888.24	1500.21
5.00	0.50	6000.00	2543.04	3739.55	3405.74	1656.04
5.00	0.55	50.00	455.13	276.16	326.61	656.34
5.00	0.55	100.00	472.76	305.26	352.49	667.06
5.00	0.55	200.00	508.01	363.47	404.24	688.52
5.00	0.55	400.00	578.51	479.89	507.74	731.44
5.00	0.55	500.00	613.76	538.09	559.49	752.90
5.00	0.55	1000.00	790.01	829.14	818.24	860.19
5.00	0.55	1500.00	966.26	1120.18	1076.99	967.48
5.00	0.55	2000.00	1142.51	1411.22	1335.74	1074.77
5.00	0.55	3000.00	1495.01	1993.30	1853.24	1289.36
5.00	0.55	4000.00	1847.51	2575.39	2370.74	1503.94
5.00	0.55	5000.00	2200.01	3157.47	2888.24	1718.52
5.00	0.55	6000.00	2552.51	3739.55	3405.74	1933.11
5.00	0.60	50.00	412.46	276.16	326.61	617.42
5.00	0.60	100.00	432.17	305.26	352.49	628.96
5.00	0.60	200.00	471.59	363.47	404.24	652.04
5.00	0.60	400.00	550.42	479.89	507.74	698.21
5.00	0.60	500.00	589.84	538.09	559.49	721.29

their impact on the estimation of the decision tree model parameters is negligible. A node on a split indicates that the corresponding category is the main category on a particular split in the decision tree. In other words, it indicates that most of the data points on that split point belong to that particular category. The decision process follows the left branch if the condition on the node is true, and the right branch otherwise. The node on the leaf (i.e., the most bottom) represents the final decision in the decision tree model. These implications also apply to subsequent visualizations of the decision tree.

Furthermore, if the minimum model performance requirement is also given, e.g., a requirement that the loss must be not greater than C, then the data quality must be improved if the $\ell^* > C$ (ℓ^* is the loss function value of the best model). Otherwise, the best model can be determined. For instance, if the minimum model performance requirement C is set at 500, then referring to the information in the initial row of Table III.5, it can be deduced that MTTF is the best model when the

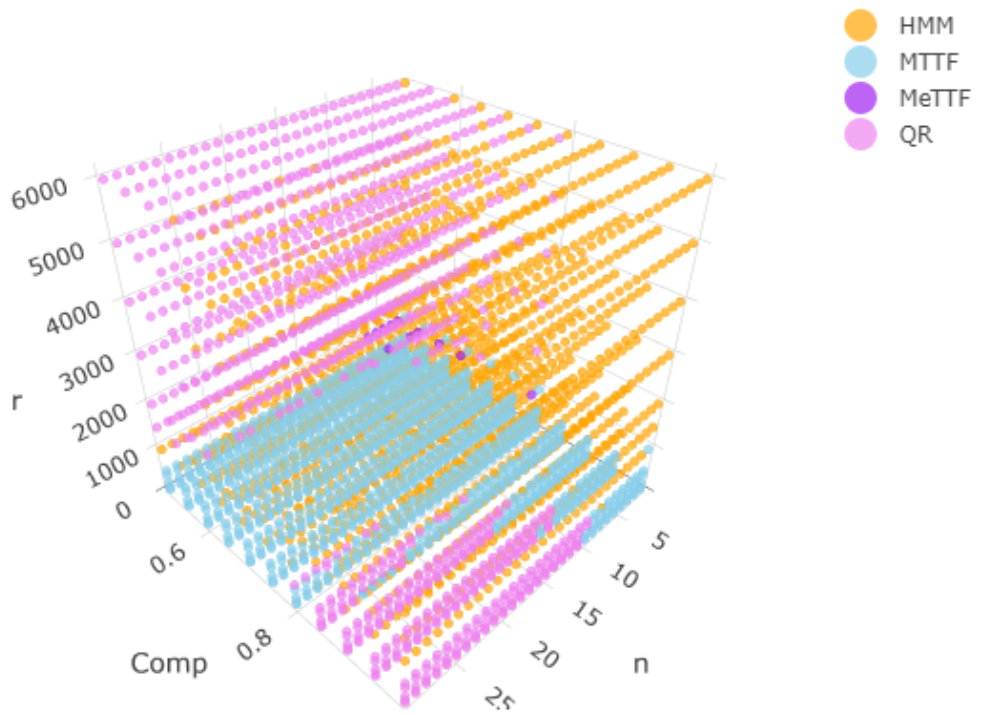


Figure III.5 – Best model under different data volume, completeness, and cost ratio

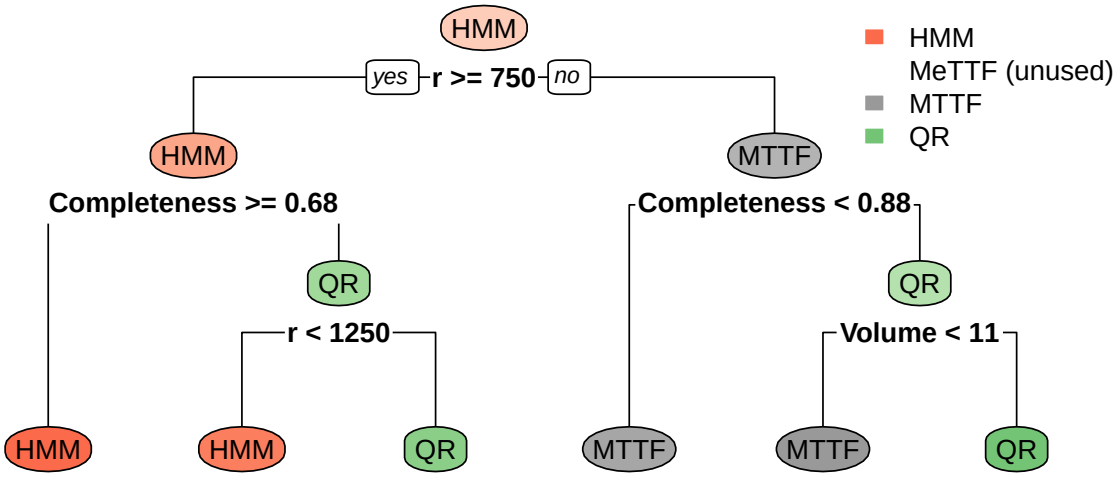


Figure III.6 – The decision model for determining the best model without minimal performance requirement

data volume, the completeness, and the cost ratio are 5, 0.5, and 400, respectively. This is because MTTF achieves the minimum loss function value among the four models, and its loss is below 500. However, in cases where the data volume, the

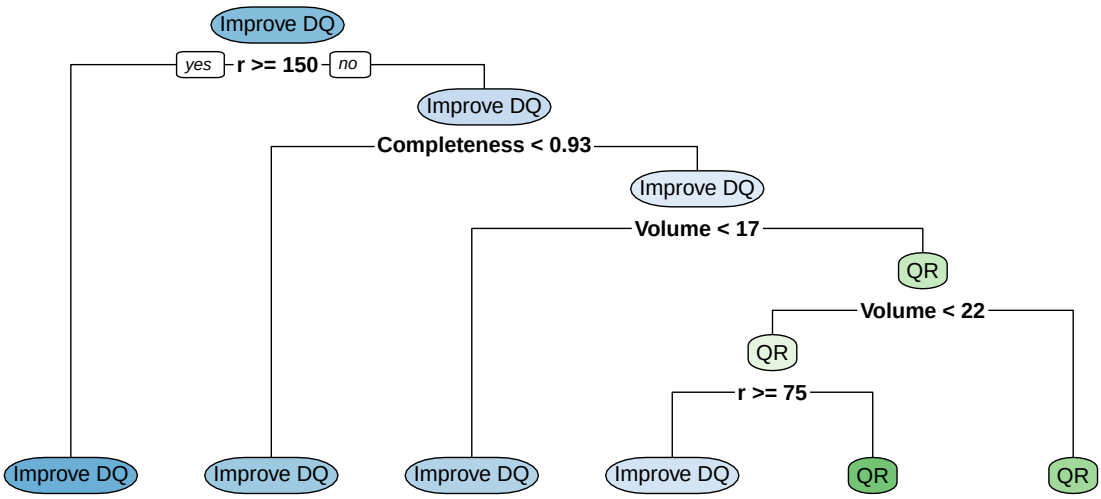


Figure III.7 – The decision model under $C = 250$

completeness, and the cost ratio are 5, 0.5, and 500, respectively, as indicated in the second row of the table, then the decision is to improve the data quality. This is because MTTF obtains the minimum loss function value among the four models, but its loss exceeds the specified threshold C of 500.

The decision models for several values of C are illustrated in Figure III.7 through III.11. In these figures, the notation “Improve DQ" signifies “the need to improve data quality." It is evident that the predominant decision leans towards “improve DQ" for small values of C. Conversely, as the value of C becomes significantly larger, the decision model tends to align with the decision tree model that does not specify a minimum performance requirement. This observation is logical because when the value of C (i.e., the minimum model loss) is small, there is a higher performance requirement for the model, which typically results in all four models failing to meet the requirement. On the contrary, when the value of C is large, the performance requirement of the model is lower, and at least one of the four models meets the specified requirement.

III.4.4 Make the decision and act

To facilitate a comparison of model performance before and after data quality improvement, 10 boards are utilized as training data before data quality enhancement. The average completeness score for these 10 boards is 0.76. Then, all 12 boards are used as training data after data quality improvement by filling in the missing data using mean imputation. Consequently, the data quality metrics before data quality improvement can be represented as $\mathbf{Q} = [10, 0.76]$. In contrast, the six test boards remain unchanged, ensuring the models are tested on the same dataset.

As previously mentioned, the decisions made by the decision tree-based model depend on the knowledge $\mathcal{K}(\mathbf{Q}, \mathbf{\Omega}, r, \ell)$ and the minimum performance requirement C. To rigorously validate the proposed framework, confusion matrices are gener-

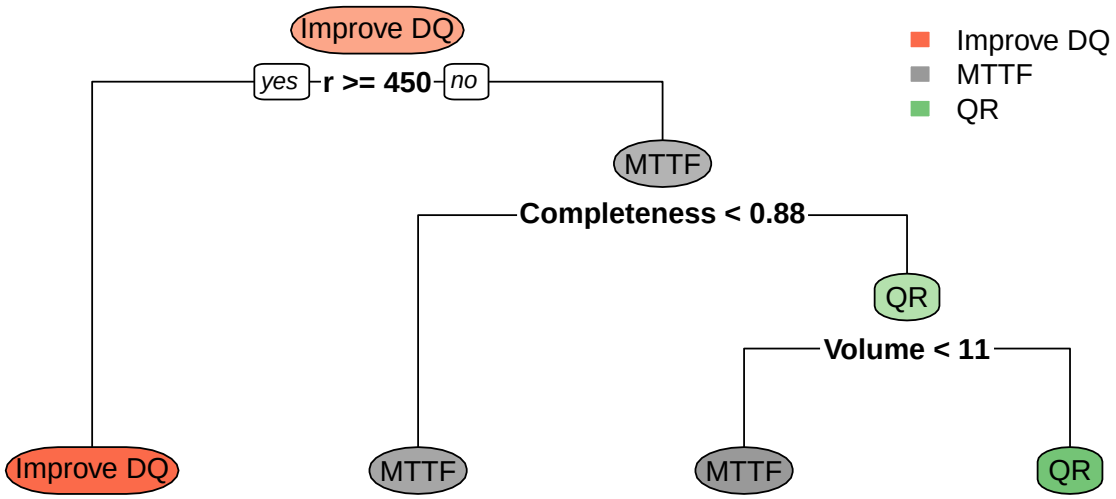


Figure III.8 – The decision model under $C = 500$

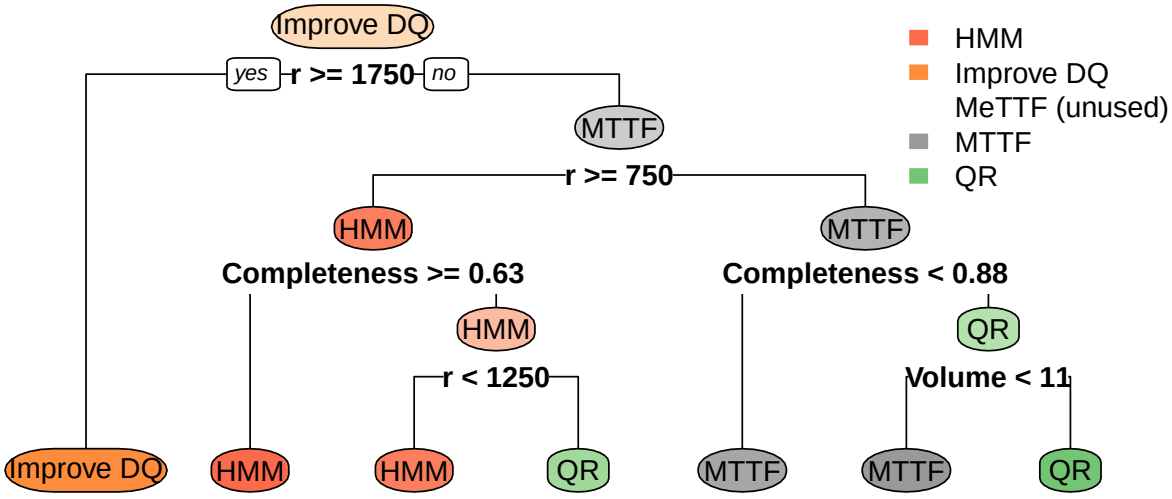


Figure III.9 – The decision model under $C = 1000$

ated to compare the predicted decisions derived from the decision models with the actual decisions inferred from the test results on the test data (i.e., The four trained models are applied to the test data, and the model losses are calculated and compared to the minimum model performance requirement C. If the loss of all four models is greater than C, the actual decision is "Improve DQ", otherwise, the corresponding model with the smallest loss is the best model.). These comparisons are made under varying cost ratios r and C. The cost ratios used in this subsection are the same as defined above in the sequence $\mathbf{L}_3$, while the values of C are consistent with the decision models shown in Figure III.7 through III.11

The result of the confusion metrics under different C is shown in Table III.6. From the table, the average decision accuracy can be computed, that is,

$$\frac{10+9+10+9+9+10}{12\times 6}\times 100\% = 79.17\%, \tag{III.6}$$

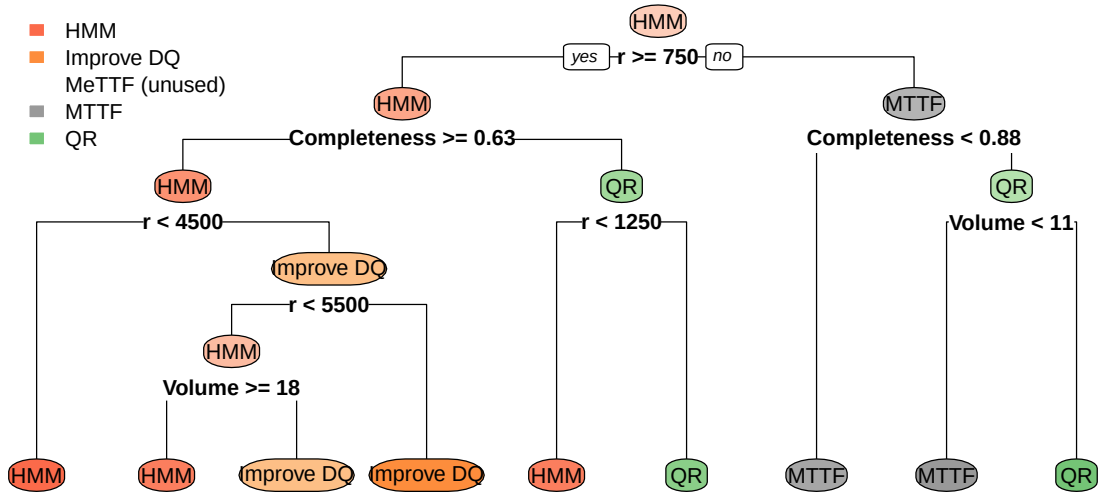


Figure III.10 – The decision model under $C = 2000$

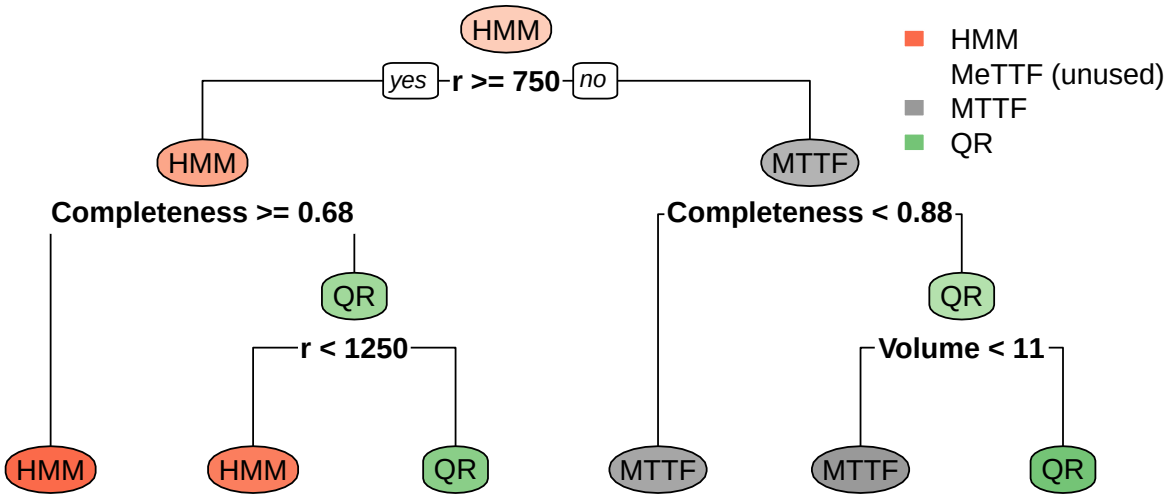


Figure III.11 – The decision model under $C = 3000$

which proves the effectiveness of the framework.

Given the cost ratio and minimum model performance requirement, the need to improve data quality can be determined with the help of the corresponding decision model. As an example, suppose the minimum model performance requirement C is fixed at 1000. The authors examine two scenarios involving the cost ratio r of the equipment under study.

In the first scenario, when the cost ratio r is set at 1000, the decision model depicted in Figure III.9 suggests that the data quality (i.e., $\mathbf{Q} = [10, 0.76]$) does not require improvement, and as a result, the HMM is chosen to construct the risk estimation model.

In the second scenario, with the cost ratio is r set at 2000, the decision model indicates that the data quality needs enhancement. The decision derivation process for both scenarios is illustrated in Figure III.12.

Table III.6 – Confusion metrics under no minimal performance requirement and different C

No minimum performance requirement					
		Actual decision			
		Improve DQ	HMM	MeTTF	MTTF
Predicted decision	HMM		7		
	MeTTF				
	MTTF			2	3
C=250					
		Actual decision			
		Improve DQ	HMM	MeTTF	MTTF
Predicted decision	Improve DQ	9			3
	HMM				
	MeTTF				
	MTTF				
C=500					
		Actual decision			
		Improve DQ	HMM	MeTTF	MTTF
Predicted decision	Improve DQ	7			
	HMM				
	MeTTF				
	MTTF			1	3
C=1000					
		Actual decision			
		Improve DQ	HMM	MeTTF	MTTF
Predicted decision	Improve DQ	4	1		
	HMM		2		
	MeTTF				
	MTTF			2	3
C=2000					
		Actual decision			
		Improve DQ	HMM	MeTTF	MTTF
Predicted decision	Improve DQ	1	1		
	HMM		5		
	MeTTF				
	MTTF			2	3
C=3000					
		Actual decision			
		Improve DQ	HMM	MeTTF	MTTF
Predicted decision	Improve DQ				
	HMM		7		
	MeTTF				
	MTTF			2	3

Additionally, Figure III.13 showcases the model's performance before and after the data quality improvement, achieved by including two additional failed boards and addressing missing values.The data quality after data quality improvement is $\mathbf{Q}' = [12, 1]$. From the figure, it can be seen that the data quality improvement leads to a reduction in losses, especially when the cost ratios are large.

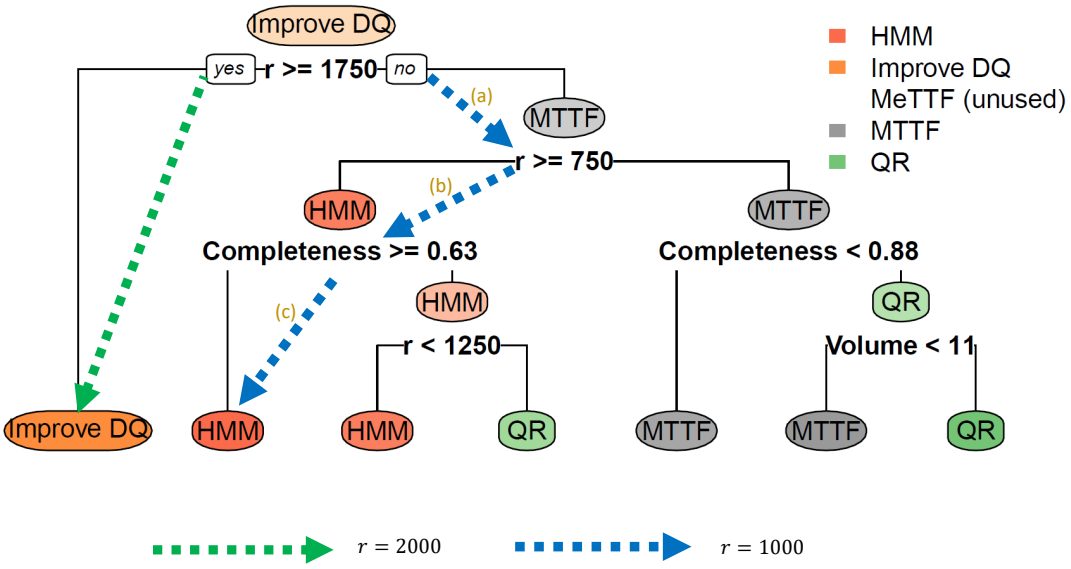


Figure III.12 – Decision derivation process for two scenarios ($r = 2000$ and $r = 1000$) with C fixed at 1000 and [Volume, Completeness] $= [10, 0.76]$

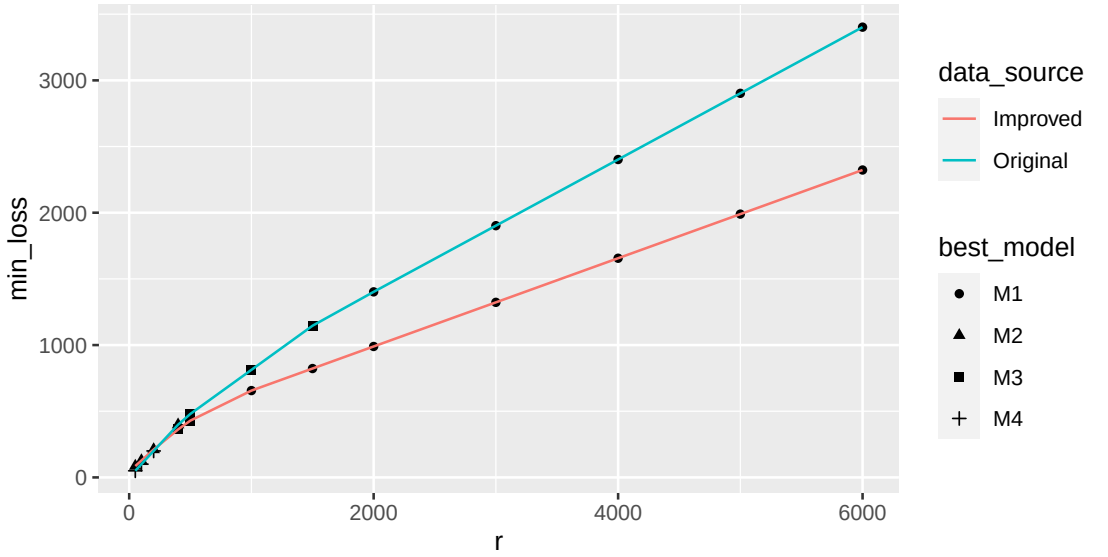


Figure III.13 – Comparisons of the best model and loss before and after data quality improvement under different cost ratios

III.5 Conclusion

This chapter introduces an innovative framework for managing data quality, which is specifically tailored for estimating the risk of industrial equipment failure. This comprehensive framework encompasses data development, data quality assessment, decision-making for data quality requirements, data quality enhancement, and model development. It furnishes valuable guidance for data practitioners seeking to manage the data quality for risk estimation. Noteworthy advancements in this framework include incorporating a decision tree-based model for evaluating data quality compliance and selecting the best risk estimation model. Additionally, it introduces an improved loss function featuring a "cost ratio" parameter, enabling the model to accommodate equipment with varying failure costs versus early replacement costs.

The efficacy of this framework is exemplified through a case study utilizing actual data of D&M tools from actual oil well drilling operations. The proposed framework's practical utility is showcased by comparing four risk estimation models, including baseline and machine learning. The framework's validation employs a confusion matrix across different cost ratios and minimum performance requirements, revealing an average decision accuracy of 79.17%, confirming the effectiveness of the decision-making approach in this framework. In real-world situations, the decision-making model aids data practitioners in making informed decisions, enabling them to determine if data quality meets specific criteria and if not, guiding improvements. The presented case study results underscore the tangible advantages of enhancing data quality, especially when the cost ratio is substantial.

In summary, this chapter establishes a comprehensive framework for data quality management, featuring a decision model that empowers data practitioners in engineering asset management to make informed decisions regarding the need for data quality improvement and the best selection of a risk estimation model. This framework plays a critical role within the data layer of the risk-based maintenance decision-making framework proposed in Chapter II. The subsequent chapter will delve into the subsequent layer of the proposed risk-based maintenance decision-making framework, that is, risk estimation.

Chapter IV
Risk estimation

IV.1 Introduction

The previous chapter presented a data quality management framework for equipment risk estimation, which is an important component of the data layer in the proposed risk-based maintenance decision-making framework in Chapter II. This chapter will further explore the subsequent layer, namely, the risk estimation layer. The risk estimation layer consists of two key aspects: equipment risk estimation and operational environment risk estimation. Specifically, this chapter will present an electronic board risk estimation method. In other words, it is how to build the equipment risk estimation model in detail if the data meets the data quality requirements. In addition, a method of operational environment risk estimation will be introduced. Both methods are only applicable to D&M tools in the oil and gas industry. It is worth noting that the equipment risk estimation is closely related to the data quality management framework presented in the previous chapter. However, the data quality management framework presented in the previous chapter is generic and can be applied to a wide range types of equipment, whereas the methods presented in this chapter are only applicable to D&M tools.

The drilling process has undergone a remarkable transformation in the oil and gas industry, evolving into a complex and sophisticated endeavor. This increased complexity stems directly from the necessity of accessing and extracting valuable resources hidden deep within the earth's crust. In order to accomplish this daunting task, the industry relies heavily on D&M tools, which are the technological cornerstones underpinning these operations. D&M tools represent exceedingly intricate systems enriched with electronics, comprising a multitude of electronic boards, each meticulously designed to fulfill specialized functions that are of paramount significance to the success of drilling operations. These electronic boards function as the central hubs of technological operations, assuming responsibilities encompassing data acquisition, signal processing, management of control systems, and the facilitation of seamless communication. Thus, the reliability and performance of these electronic boards are inexorably linked to the overall effectiveness of drilling endeavors. However, the harsh operating conditions encountered downhole, including elevated temperatures, dynamic vibrations, and substantial shocks, render these boards susceptible to complex failure modes, potentially resulting in drilling operation failures.

Furthermore, as expounded upon in both Chapter I and Chapter II, the secondary significant factor influencing the efficiency and success of drilling operations is the field environment. This factor substantially impacts the reliability of the D&M tools deployed. Given that drilling operations are executed across diverse geographical regions, each characterized by varying degrees of environmental harshness, assessing the probability of encountering adverse field environments (i.e., operational environment risk) is imperative. This assessment is critical in guiding optimal tool maintenance scheduling and fleet maintenance planning strategies.

The remainder of this chapter is structured as follows: Section 2 reviews existing related works concerning electronics health assessment, failure prognostics, and risk estimation. Section 3 provides a brief introduction to the theory of HMM, which forms the foundation of the electronic board risk estimation method. Section 4 details the electronic board risk estimation method along with a case study. Section 5 outlines the operational environment risk estimation method and presents a case study. Finally, Section 6 concludes this chapter and introduces the topic for the next chapter.

IV.2 Related works

IEEE Standard 1856-2017 [4] provides a comprehensive framework for categorizing electronics into six distinct levels, ranging from individual devices and components to larger assemblies, subsystems, systems, and even complex systems of systems, as outlined in Table IV.1. A substantial body of research has delved into the domain of health prognostics and condition monitoring of electronics, as evidenced by the documentation in [107, 108, 109]. However, it is worth noting that the majority of these research endeavors have primarily centered on assessing the health or predicting the failure of electronic components, such as capacitors [110, 111, 112] and transistors [113, 114, 115]. There exists a conspicuous dearth of research initiatives aimed at comprehensively assessing the health and predicting the failure of electronic assemblies, exemplified by electronic boards or Printed Circuit Board Assembly (PCBA)s, which comprise an amalgamation of multiple integrated components. Furthermore, a considerable proportion of these studies has limited their scope to a singular environmental loading or stress, primarily focusing on either temperature [116, 117] or vibration [118, 119], with scant attention dedicated to elucidating the combined impact of multiple environmental factors on electronics health.

The oil and gas well electronics industry has a low volume compared to other electronic industries [120]. Electronics within D&M tools constitute a subcategory of oil and gas well electronics specifically engineered for the challenging conditions of oil and gas well drilling and measurements. Consequently, most of the research on health assessment and failure prediction of D&M tool electronics has been led by prominent oil and gas service companies. For example, for the fault detection and diagnosis of D&M tools, Mosallam et al. introduced a fault detection method for the neutron generator subsystem in multifunction LWD services, leveraging the empirical mode decomposition algorithm [7]. Bhatnagar et al. proposed a data-driven fault detection approach tailored to electronic boards within intelligent remote dual-valve systems [121]. Gupta et al. presented an automatic fault detection method based on support vector machine for resistivity subystems in LWD tools [50]. Sobczak et al. introduced a data-driven fault detection for transmitter subsystems in LWD tools [51]. Mosallam et al. proposed a data-driven fault diagnostics for three power supply boards in LWD tools [30].

Table IV.1 – Definitions and examples of six levels of electronics by IEEE Standard 1856-2017 [4]

Level	Definition	Examples
Device level (Level 0)	The device level represents the lowest level of complexity that makes up an overall system. For electronics applications, the device level would include the chip and on-chip sites.	chips, circuits, die, and metallization
Component level (Level 1)	The component level represents the lowest level of a packaged product comprising the lower level elements within the product. Component-level electronics include packaged semiconductor devices, and memory storage devices, and other packaged circuit elements.	diodes, transistors, integrated circuits, inductors, resistors, and capacitors
Assembly level (Level 2)	The assembly level includes the printed circuit board and various board-specific features. The assembly includes the components as well as the support structures and interconnects that provide mechanical stability and at the same time provide electrical and functional connectivity.	PCBA
Sub-system level (Level 3)	The sub-system level for electronics can be defined on the basis of the functionality of a group of assemblies. A sub-system may include the enclosure, PCBAs, and interconnections between PCBAs and associated electro-mechanical parts.	hard disk drives, line replaceable unit boxes, transmitters, receivers, and power supplies
System level (Level 4)	The system level includes a more integrated and standalone functioning product. The system could be made up of many sub-systems or could be a functional combination of PCBAs and associated electro-mechanical components within an external enclosure.	radar installations, computers, and televisions
System-of-systems level (Level 5)	System-of-systems is defined as a set or arrangement of systems that results when independent and useful systems are integrated into a larger system that delivers unique capabilities. The system-of-systems level incorporates many systems into a complex overarching product.	aircraft, trains, automobiles, or seagoing vessels; aircraft communication, navigation, and identification systems

For the failure prediction or risk estimation of electronic boards in D&M tools, Kale et al. presented a probabilistic method for estimating the risk associated with D&M tool electronics, relying on reliability functions such as Weibull, log-normal, and exponential distributions in conjunction with stress levels [122]. Zhan et al. introduced a cumulative damage model incorporating the Weibull distribution [123, 124], which shares similarities with Kale's work but with slight variations in stress levels. These methods presuppose a predetermined form for the reliability function based on prior knowledge, and estimating the function's parameters often necessitates a significant volume of failure data. Drawing on the resemblance in cumulative stress paths between different tools, Garvey et al. introduced a pattern recognition-based approach for estimating the remaining useful life of D&M tools [125]. However, this method relies on a strong assumption that the stress path is linear, which may not always hold in practice. Additionally, it does not provide a quantification of the prognostic result's uncertainty.

In addition to the limitations highlighted in the existing methods described above, the failure prognostics or risk estimation for electronic boards in D&M tools presents a unique and intricate set of challenges. These challenges include but are not limited to insufficient condition monitoring information, with operational environment data typically measured at the tool or system level rather than at the granularity of individual electronic boards or electronic components, as detailed in subsection II.2.6. Moreover, D&M tools operate under dynamic and extreme environmental conditions. For instance, temperature variations during a single drilling operation can range from 30 °C to 150 °C in an erratic and irregular manner. Furthermore, the electronic boards in these tools are affected by a combination of environmental factors, such as temperature, shock, and vibration, rather than just a single environmental factor.

The limitations and challenges described above underscore the need for more robust and adaptable approaches to evaluate the performance and health of electronic boards within D&M tools, particularly given the complexity and criticality of these tools in the oil and gas industry.

To the best of our knowledge, there are no published works specifically addressing the operational environment risk estimation of oil fields. This gap in the literature may be attributed to the highly specialized nature of oil well drilling and the unique environmental challenges associated with it. Existing research in related areas often focuses on environmental and natural resource protection [126, 127, 128] rather than the specific environmental risks tied to operating conditions of D&M tools in the oil and gas industry.

IV.3 Hidden Markov model introduction

HMM, initially proposed by Baum and Petrie [129], has become a foundational statistical model with widespread applications in diverse fields, including speech recognition [130, 131, 132, 133], bioinformatics [134, 135, 136, 137], and natu-

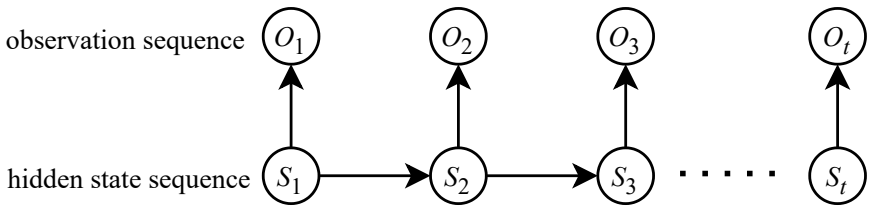


Figure IV.1 – Illustration of HMM

ral language processing [138, 139, 140, 141]. HMMs excel at capturing complex temporal dependencies in data by representing systems that evolve through a sequence of hidden states, generating observable outputs [142]. This unique ability to model latent variables makes HMMs indispensable for tasks such as equipment risk modeling, where understanding sequential relationships is critical.

An HMM consists of two parts, including an unobservable or hidden state sequence and a state-dependent observation sequence, as shown in Figure IV.1. The unobservable state sequence $S_t : t = 1, 2, \ldots$ satisfies the Markov property; i.e., that the conditional probability of the state at time t (i.e., S_t) on previous states is equivalent to the conditional probability of S_t only on the most recent state S_{t-1}. Mathematically, this condition is expressed as $Pr(S_t|S_{t-1}, S_{t-2}, \ldots, S_1) = Pr(S_t|S_{t-1}), t = 2, 3, \ldots$. The observation sequence $\{O_t : t = 1, 2, \ldots\}$ is state-dependent. In other words, the distribution of O_t depends only on the current state S_t and not on historical states or observations. This can be expressed as $Pr(O_t|S_1, S_2, \ldots, S_t, O_1, O_2, \ldots, O_{t-1}) = Pr(O_t|S_t)$ [143].

The HMM parameters can be characterized by a 3-tuple, $\mathbf{\Omega} = (\mathbf{M}, \mathbf{\Phi}, \boldsymbol{\pi})$, where $\mathbf{M}$ is the state transition probability matrix, $\mathbf{\Phi}$ is the observation probability distribution, and $\boldsymbol{\pi}$ is the state initial distribution.

Two problems can be formulated for HMM when applying the method to risk level estimation in this study.

IV.3.1 The learning problem

Given T consecutive observations $\mathbf{O} = \{O_1 = o_1, O_2 = o_2, \ldots, O_T = o_T\}$, the learning problem of HMM aims to find the model parameter $\mathbf{\Omega}$ that maximize the likelihood L shown in Equation (IV.1).

$$\begin{aligned} L &= Pr\left(\mathbf{O}\right) = Pr\left(O_1 = o_1, O_2 = o_2, \ldots, O_T = o_T\right) \\ &= \sum_{s_1, s_2 \ldots, s_T = 1}^{m} Pr(O_1 = o_1, O_2 = o_2, \ldots, O_T = o_T, S_1 = s_1, S_2 = s_2, \ldots, s_T = S_T) \\ &= \boldsymbol{\pi}\mathbf{M}\mathbf{\Phi}(o_1)\mathbf{M}\mathbf{\Phi}(o_2)\cdots\mathbf{M}\mathbf{\Phi}(o_T)\mathbf{1}' \end{aligned} \tag{IV.1}$$

The Baum-Welch algorithm is commonly used to solve this problem [131].

IV.3.2 The decoding problem

Given the HMM model parameter $\mathbf{\Omega}$ and T consecutive observations, $\mathbf{O} = \{O_1 = o_1, O_2 = o_2, \ldots, O_T = o_T\}$, as well the decoding problem of HMM is to determine the hidden state sequence $\mathbf{S} = \{S_1 = s_1, S_2 = s_2, \ldots, S_T = s_T\}$ that are most likely to generate the observation sequence, which is mathematically expressed as follows:

$$\max Pr(S_1 = s_1, S_2 = s_2, \ldots, S_T = s_T | O_1 = o_1, O_2 = o_2, \ldots, O_T = o_T) \quad \text{(IV.2)}$$

This problem can be solved by the Viterbi algorithm [144].

IV.4 Electronic board risk estimation

As mentioned in the chapter introduction, the electronic boards in D&M tools are exposed to high temperatures, dynamic vibrations, and extensive shocks when the tool is drilling or measuring formation properties. These environmental conditions are the critical factors that cause the declined reliability or even the failure of the electronic boards. Thus, sensors are installed on the D&M tool to collect environmental data such as temperature, vibration, and shock during drilling operations. One could estimate the risk level of the electronic board through analyzing these environmental data.

IV.4.1 Assumptions

In order to implement the HMM for estimating the risk levels of electronic boards using operational environment data, the proposed method relies on three key assumptions as follows.

1. **Gaussian Distribution of Observations:** The observations, representing the final features used to train the HMM, are assumed to follow a Gaussian distribution. Additionally, it is assumed that these features are mutually independent. This assumption simplifies the modeling process and facilitates using HMM for risk estimation.
2. **Four Risk Levels:** The risk associated with the electronic boards is considered unobservable, but the model assumes four distinct risk levels: Level 1, Level 2, Level 3, and Level 4. These levels are used to categorize the risk, with Level 1 indicating the lowest risk and Level 4 representing the highest risk. Furthermore, electronic boards are assumed to start in a functional state at the beginning of the monitoring process. In other words, the HMM's initial distribution of risk levels is $\boldsymbol{\pi} = [1, 0, 0, 0]$, signifying that the boards begin in Level 1.

3. **Transition Between Risk Levels:** The model assumes that the risk level of electronic boards at the next time step can either transition to the next risk level in sequence or remain at the same risk level as the current time step. This implies a left-to-right HMM structure, where the transition between risk levels follows a sequential order [145]. This assumption reflects the notion that electronic boards may gradually deteriorate or experience changes in risk over time.

IV.4.2 Proposed method

Based on the previously stated assumptions, the framework of the proposed method is shown in Figure IV.2. The framework includes four parts, namely, data collection, data preprocessing, learning, and decoding.

IV.4.2.a Data collection

The data collection process here is the same as the data collection process in the data quality management framework detailed in Chapter III. This process involves gathering the tool measurement data from the CMMS database. To offer a comprehensive view of how the tool measurement data are acquired, Figure IV.3 presents a graphical representation of the data collection process for tool measurements. In drilling operations, each D&M tool within the BHA is tasked with measuring a substantial volume of analog signals via its integrated sensors. These analog signals are subsequently converted into digital format through signal acquisition cards or analog-to-digital converters. However, the transmission of data to the surface rig is constrained by bandwidth limitations. Most of the acquired data undergoes real-time processing within the tool, accomplished by a field programmable gate array or a digital signal processor. Subsequently, the processed information is directed to a CPU, which contributes to the automated control of the tool's functions or is stored on a memory board for subsequent offline analysis. Upon the conclusion of the drilling operation, the tool is extracted from the well. At this juncture, the data stored on the memory board is transferred to the hard disk and subsequently uploaded to a data cloud, contingent upon the availability of an internet connection. The raw data in the cloud can then be retrieved and harnessed to construct the risk estimation model of the electronic boards housed within the D&M tool.

IV.4.2.b Data preprocessing

The data preprocessing part consists of five steps; i.e., channel selection, histogram feature extraction, construction of cumulative exposure data sequence, data enrichment, and feature aggregation. These five steps are described as follows.

— **Channel Selection:** The channel selection is to choose the relevant signal channels for the risk level estimate. Because the D&M tool acquires an enormous number of channels of information during drilling operations, many

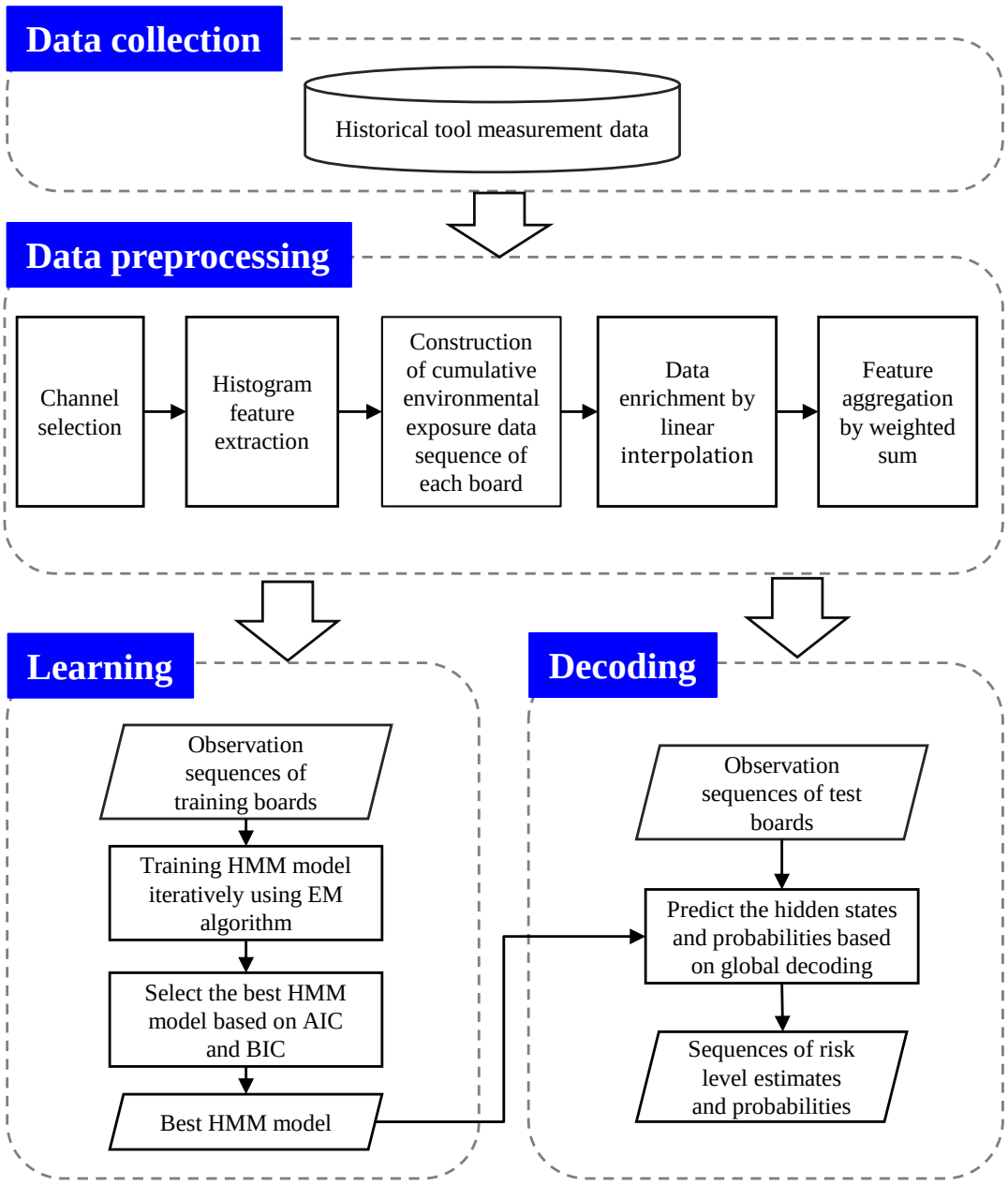


Figure IV.2 – Proposed risk level estimate framework

channels, however, do not contain information concerning failure risk of the electronic boards. Removing these channels does not only reduce the computation complexity in the following analysis, but also improves the estimation performance. Therefore, only the most important channels are selected for the risk-level estimate. In general, channels containing temperature, vibration, and shock data are selected because they are believed to have significant impacts on the electronic board lifetime [107, 108].

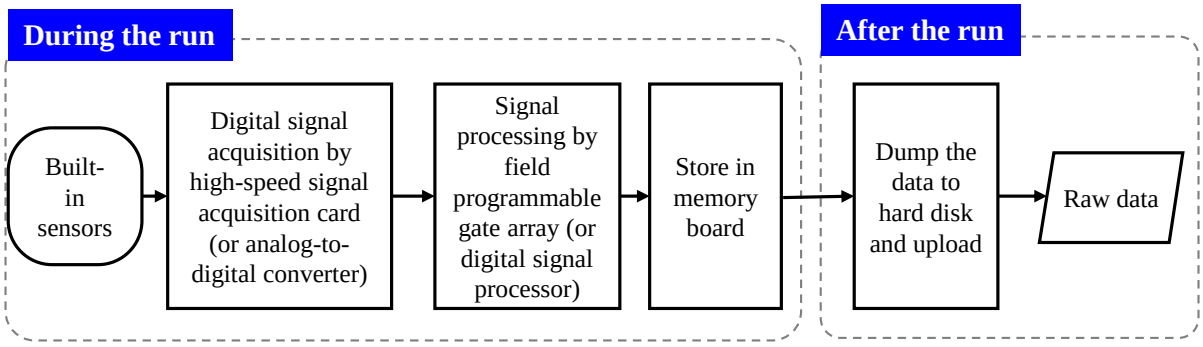


Figure IV.3 – Tool measurement data acquisition steps

While selecting channels, we opted not to include channels related to electrical signatures, such as those for current and voltage. This decision was driven by the fact that D&M tools incorporate specialized circuitry designed to regulate electrical signals and maintain consistent currents and voltages. Furthermore, we limited the measurement of electrical signals to specific boards, with most boards not requiring such measurements.

Additionally, we did not account for potential factors such as dust, humidity, chemicals, and radiation. This omission is because D&M tools do not typically measure these parameters for electronic boards. The physical arrangement of electronic boards within these tools inherently protects against exposure to dust, humidity, radiation, and chemicals that may be present in the wellbore. These tools are typically enclosed within robust steel tubing, shielding internal electronics from direct contact with these environmental factors. Moreover, before deployment, field engineers frequently introduce nitrogen into these tools, reducing the likelihood of exposure to potentially harmful substances. As a result of these protective measures and practices, the risk of electronic board damage due to dust, humidity, radiation, and chemical exposure is significantly mitigated.

— **Feature Extraction:** The histogram feature extraction computes exposure time under different environmental levels (also referred to as stress levels) using the data from the selected channels of each run. Environmental levels are equivalent to histogram bins, and exposure time of these levels corresponds to the histogram frequencies multiplied by the data recording rate. The histogram feature of board b under environmental level i of channel j of run k is mathematically expressed as follows:

$$H_{ijk}^{(b)} = Frequency_{ijk}^{(b)} * RecordingRate \tag{IV.3}$$

— **Construction of Cumulative Environmental Exposure:** Electronic board failures are usually caused by accumulative effects of environmental exposure during the numerous downhole drilling runs. Thus, environmental exposure data from a single run is not sufficient for determining electronic board risk. To use cumulative environmental exposure data, multiple drilling

runs are required. This step aims to construct cumulative environmental exposure data over runs based on the extracted histogram features. Specifically, the cumulative environment exposure data of board b under environmental level i of channel j after run K is denoted as follows:

$$C_{ijK}^{(b)} = \sum_{k=1}^{K} H_{ijk}^{(b)} \tag{IV.4}$$

— **Data Enrichment:** Training an HMM typically requires substantial data or observations. However, for each electronic board, only one observation can be obtained for each run after collecting cumulative exposure data. Moreover, obtaining complete life-cycle data for electronic boards in a D&M tool is challenging. This difficulty arises because the tool's measurement data retrieved from cloud storage often contains missing values for some runs due to various reasons, including tool loss in the borehole, memory board damage, and oversight by field engineers in data retrieval, as discussed in subsection III.3.4 in the last chapter. Therefore, to augment the data volume for training the HMM, a linear interpolation method is adopted to generate synthetic data. The linear interpolation method is simple. Specifically, given two cumulative exposure data points from board b under environmental level i of channel j at time t_0 (end time of run $K-1$) and t_1 (end time of run K), the interpolated cumulative exposure data point at time t between t_0 and t_1 is given by:

$$C_{ijt}^{(b)} = C_{ij,K-1}^{(b)} + (t - t_0)\frac{C_{ijK}^{(b)} - C_{ij,K-1}^{(b)}}{t_1 - t_0} \tag{IV.5}$$

In this chapter, an equally spaced time sequence from time zero to the life end time of the board is first generated, and then new cumulative exposure data points are interpolated using Equation (IV.5). It is important to note that the data enrichment step is optional if a large amount of data is available.

— **Feature Aggregation** The D&M tool is less likely to be exposed to high-temperature, high-vibration, and high-shock peak environments, which makes cumulative environmental exposure data at those levels not change significantly with time. Analyzing HMM parameters would fail if directly using these data. The reason is it would be impossible to estimate the state-dependent Gaussian distribution parameters when the observations are the same. In addition, it is difficult to determine which histogram features should be used. Nevertheless, the model complexity would substantially increase if all of the histogram features are used for HMM learning. Therefore, to simplify the model and achieve successful HMM parameter learning, we use weight sum to aggregate features from the same channel. The aggregated feature of board b of channel j at time t is formulized as follows:

$$AggF_{jt}^{(b)} = \sum_{i=1}^{n_j} w_{ij} C_{ijt}^{(b)} \tag{IV.6}$$

where n_j represents the number of levels of channel j, and w_{ij} denotes the feature weight of level i of channel j.

IV.4.2.c Learning and decoding

After data preprocessing, the aggregated features can be used to build an observation of board b at time t as shown in Equation (IV.7). Then, it becomes possible to construct the observation sequence for each board. The boards are further divided into training boards and test boards. The observation sequences of the training boards are used to learn the HMM model parameter $\mathbf{\Omega}$. Because the Baum-Welch algorithm is a type of Expectation Maximization (EM) algorithm, the solution of EM is not unique. It is possible to train the HMM model many times and select the best model according to the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) as shown in Equation (IV.8). The less are the AIC and BIC, the better the model. With the trained HMM model and observation sequence of the test boards, the sequences of risk level and probabilities of those boards can be estimated. It should be noted that one can also estimate the same information for training boards.

$$O_t^{(b)} = [AggF_{1t}^{(b)}, AggF_{2t}^{(b)}, \ldots, AggF_{Nt}^{(b)}]^T \tag{IV.7}$$

where N means the number of channels.

$$\begin{aligned} AIC &= -2\log(L) + 2p \\ BIC &= -2\log(L) + p\log(T) \end{aligned} \tag{IV.8}$$

where L is the log likelihood of the HMM shown in Equation (IV.1), p is the number of model parameters, and T denotes the number of observations.

IV.4.3 Case study

In this case study, historical tool measurement data obtained from D&M tools are employed to validate the effectiveness of the proposed methodology. These data were gathered during actual oil well drilling operations, focusing on a specific type of RSS. This tool is critical in directing the drilling direction while simultaneously rotating the drill string. Various types of electronic boards are integrated into this tool to support its multifaceted functions. In this case study, particular attention is given to the CPU board.

The dataset used in this case study comprises historical tool measurement data obtained from a total of 21 CPU boards, all of which experienced failures associated with thermal and mechanical loading. As previously discussed in the *Data Enrichment* part, collecting complete life-cycle environmental exposure data from these failed boards proved challenging. Thus, some of the boards in the dataset exhibit missing data.

For further clarity, Table IV.2 provides a detailed description of the 21 CPU boards included in the study. These boards serve as the foundation for the subsequent analysis and validation of the proposed risk level estimation method.

Table IV.2 – Data description of the 21 electronic boards

Board ID	Runs	Missing Runs[a]	Data Coverage[b]	Total Pumping Time (hr)[c]
x1005	18	4	92%	1104.5
x4028	20	2	87%	862.6
x1014	11	0	100%	714.4
x6054	19	1	93%	742.2
x9031	18	2	89%	1355.9
x9046	33	9	83%	1328.1
x0048	15	2	95%	1037.2
x5005	19	3	92%	1186.2
x5050	18	1	97%	1110.1
x5010	22	3	87%	1403.7
x0405	17	1	95%	1210.2
x1012	3	0	100%	125.7
x9045	43	8	80%	2067.2
x3007	11	1	94%	963.2
x9035	28	10	72%	1176.5
x7043	25	8	73%	1371.3
x7079	23	7	71%	1608.2
x9003	29	5	84%	1691.2
x9049	22	4	84%	1239.3
x6004	11	0	100%	450.3
x1260	20	2	90%	1324.2

[a] number of runs that have missing data.
[b] total pumping time of runs with data/total pumping time of all runs.
[c] total pumping times of all runs.

Additionally, four channels of tool measurement data, specifically, temperature, lateral vibration, axial vibration, and lateral shock peak were selected for this case study. The number of levels for the four channels and the corresponding weights for feature aggregation are summarized in Table IV.3. The time allowed for generating the time sequence for the Data Enrichment step is set to 10 hours.

The tool measurement data from the 21 boards were processed following the steps described in the *Data Preprocessing* part. Then, the first 11 boards listed in Table IV.2 were selected as training boards, and the remaining 10 boards as test boards.

The software used for the HMM learning and decoding in this paper is the *depmixS4* [146] R package. The HMM model was trained 100 times using the training boards data. The best HMM model has an AIC of 52887.48 and a BIC of 53124.96. The learned state transition probability matrix of the HMM is as

Table IV.3 – Description of environmental levels and weights

Channel Name	Number of Levels	Level 1 Weight	Level 2 Weight	Level 3 Weight	Level 4 Weight
Temperature	4	0.5	1.5	2	10
Lateral vibration	3	0.5	1.5	10	–
Axial vibration	3	0.5	1.5	10	–
Lateral shock peak	3	0.5	1.5	10	–

follows:

$$\mathbf{M} = \begin{bmatrix} 0.95 & 0.05 & 0 & 0 \\ 0 & 0.963 & 0.037 & 0 \\ 0 & 0 & 0.967 & 0.033 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{IV.9}$$

The learned observation probability distribution (Gaussian distribution) parameters are shown in Table IV.4, where μ and σ denote the mean and standard deviation of the Gaussian distribution, respectively.

Table IV.4 – Learned observation probability distribution parameters

State (Risk Level)	Aggregated Temperature		Aggregated Lateral Vibration		Aggregated Axial Vibration		Aggregated Lateral Shock Peak	
–	μ	σ	μ	σ	μ	σ	μ	σ
Level 1	67.4	42.6	57.2	33.4	58.8	35.4	52.6	30.0
Level 2	288.5	127.9	182.7	46.7	184.0	46.4	68.0	41.1
Level 3	557.7	235.8	30.9	56.1	330.1	53.2	308.8	46.8
Level 4	803.2	299.6	546.6	116.0	533.1	102.5	491.1	81.1

Using the trained HMM model, the risk level and its probability can be estimated on the condition of observing the sequence of each run. It should be noted that the risk estimation for the runs with missing data is not included. The sequences of risk level estimates and probabilities of training boards are shown in Table IV.5.

The sequences of risk level estimates and probabilities of test boards are shown in Table IV.6.

Table IV.5 – Sequences of risk level estimates and probabilities of training boards

Board ID	Risk Level Sequence	Probability Sequence
x1005	1-2-2-2-2-3-3-4-4-4-4-4-4-4-4	1-1-1-1-0.999-1-0.999-0.06-1-1-1-1-1-1
x4028	1-1-2-2-2-2-2-3-3-3-3-3-3-3-3-4-4	1-1-1-1-1-1-1-0.597-1-1-1-1-1-0.993-0.793-1
x1014	1-1-1-2-2-2-2-2-2-3-3	1-1-1-0.838-1-1-1-1-0.998-1-1
x6054	1-1-1-2-2-2-2-2-2-2-2-3-3-4-4-4-4	1-1-1-0.825-1-1-1-1-0.999-0.999-0.996-0.991-0.999-0.169-1-1-1
x9031	1-2-2-3-3-3-3-3-3-3-3-4-4-4-4-4-4	1-1-1-0.044-0.125-0.579-0.974-1-1-1-0.626-1-1-1-1-1
x9046	1-1-2-2-2-2-3-3-3-3-3-3-3-4-4-4-4-4-4-4-4-4-4-4-4	1-0.991-0.741-1-1-0.998-0.955-1-1-1-1-0.979-0.993-1-1-1-1-1-1-1-1-1-1
x0048	1-1-1-2-2-2-2-3-3-3-4-4	1-1-1-1-1-1-1-0.956-1-1-1-1
x5005	1-1-1-2-2-3-3-3-3-3-3-3-3-4-4-4-4	1-1-1-1-1-0.062-1-1-1-1-1-0.992-1-1-1-1
x5050	1-1-2-2-3-3-3-3-3-3-4-4-4-4-4-4-4-4	1-0.999-1-1-0.114-1-1-1-0.999-0.997-1-1-1-1-1-1-1
x5010	1-1-1-1-2-2-2-2-2-3-3-3-3-3-3-4-4-4-4-4-4	1-1-1-0.987-1-1-1-1-1-1-1-1-1-0.996-1-1-1-1-1
x0405	1-1-2-2-2-2-3-3-3-3-4-4-4-4-4-4-4	1-1-0.915-1-1-1-1-1-1-0.991-1-1-1-1-1-1

From these results, almost all of the boards (10/11 of training boards, 8/10 of test boards) are predicted to be at risk of Level 4 in their last run, which is consistent with the actual situation because all of the boards failed. The three boards whose risk level in the last run is not estimated as Level 4 have much less pumping time (714.5 hr, 125.7 hr, and 450.3 hr) than the other boards according to the pumping time in Table IV.2. This result implies that the proposed method might not be suitable for boards with minor failures.

Considering that most boards do not have full-life cycle coverage, the proposed method is still capable of achieving acceptable results. Thus, the proposed method is effective for estimating the risk level of electronic boards.

Table IV.6 – Sequences of risk level estimates and probabilities of test boards

Board ID	Risk Level Sequence	Probability Sequence
x1012	1-1-1	1-1-1
x9045	1-1-1-1-2-2-2-3-3-3-3-4-4-4-4-4-4-4-4-4-4-4-4-4-4-4-4-4-4-4-4-4-4-4-4-4-4	1-1-1-1-1-1-1-1-1-1-0.99-1-1-1-1-1-1-1-1-1-1-1-1-1-1-1-1-1-1-1-1-1-1-1-1-1
x3007	1-2-2-2-2-2-3-3-3-4	1-1-1-1-1-1-1-1-1-0.21
x9035	1-1-1-1-2-2-2-3-3-3-3-4-4-4-4-4-4-4	1-1-1-1-1-1-0.98-1-1-1-1-0.05-1-1-1-1-1-1
x7043	1-2-2-2-2-2-3-3-3-4-4-4-4-4-4-4	0.97-1-1-1-1-1-1-1-0.99-1-1-1-1-1-1-1
x7079	1-1-2-2-2-3-3-3-4-4-4-4-4-4-4	1-0.99-1-1-1-1-1-1-0.08-1-1-1-1-1-1
x9003	1-1-1-1-2-2-2-3-4-4-4-4-4-4-4-4-4-4-4-4-4-4-4-4-4	1-1-1-1-1-1-1-1-1-1-1-1-1-1-1-1-1-1-1-1-1-1-1-1-1
x9049	1-1-2-2-3-3-3-3-3-4-4-4-4-4-4-4-4-4-4	1-1-0.98-1-0.04-1-1-1-1-1-1-1-1-1-1-1-1-1-1
x6004	1-1-1-1-1-1-2-2-2-2	1-1-1-1-1-0.95-1-1-1-1
x1260	1-1-2-2-2-2-3-3-3-3-3-4-4-4-4-4	1-1-0.06-1-1-0.99-1-1-1-1-1-0.2-1-1-1-1

IV.5 Operational environment risk estimation

IV.5.1 Proposed method

As previously discussed, the environmental risk is crucial as the conditions where a D&M tool is operated can significantly affect its reliability, thus impacting drilling job efficiency and success. In light of this, properly assessing environmental risks is imperative when making an investment plan for tool upgrades. To address this issue, we present a novel method for environmental risk estimation, as shown in Figure IV.4. The method includes four main steps: environmental level definition, exposure time computation, risk level assignment, and risk index calculation, which are elaborated as follows:

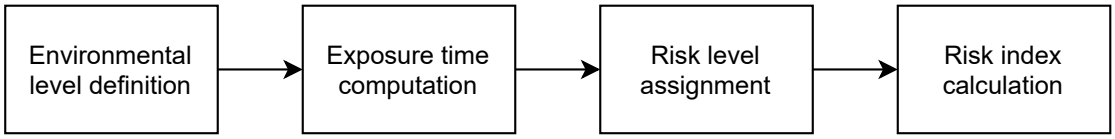


Figure IV.4 – Flowchart of operational environment risk estimation

1. **Environmental Level Definition:** In the initial phase, we define three distinct environment levels, denoted as Level 1, Level 2, and Level 3, for each environmental category that affects the efficiency and success of drilling operations. For instance, if we consider temperature, then these three levels can be defined as follows: (a) temperatures below 100°C, (b) temperatures ranging from 100°C to 130°C, and (c) temperatures exceeding 130°C. The determination of these risk levels and environmental categories is guided by the technical specifications and requirements of the specific drilling tool in use.
2. **Exposure Time Computation:** For each historical drilling operation (or run), we extract precise information concerning the exposure time endured under each environmental level within every environmental category. This valuable data is obtained from the historical tool measurement database in CMMS, where meticulous records of drilling environmental conditions have been diligently maintained.
3. **Risk Level Assignment:** Building upon the extracted exposure time, we assign a risk level (i.e., low, medium, high) to each environment category for each historical drilling operation. This assignment is based on predefined time thresholds for each category and the computed exposure time from the previous step. The decision-making process for determining the risk level of an environment category is detailed in Figure IV.5. Here, ET_{g3} and ET_{g2} represent the exposure time for Level 3 and Level 2 of environment category g. At the same time, Th_{g3} and Th_{g2} are the corresponding time thresholds. These thresholds are also subject to adjustment in accordance with the drilling tool's technical specifications.

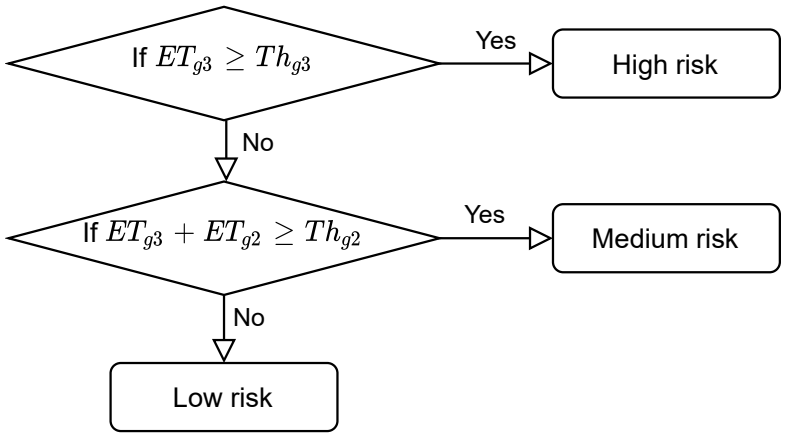


Figure IV.5 – Flowchart for determining the risk level of environment category g of a drilling operation

4. **Risk Index Calculation:** We proceed to group drilling operations by field (location or geo unit) and calculate the percentage of operations falling into each risk level for every environment category. It's worth noting that a geo

unit encompasses several locations, and a location can comprise multiple fields. Subsequently, a risk index for each environment category within a field (location or geo unit) is computed as expressed in Equation (IV.10):

$$r_g = W \times P_{g2} + P_{g3} \tag{IV.10}$$

where r_g denotes the risk index of environment category g, while P_{g2} and P_{g3} represent the percentages of jobs in the medium-risk and high-risk levels, respectively. The weight W is determined by domain experts to combine the percentages of medium-risk and high-risk jobs into a single value, typically ranging from 0.1 to 0.5.

By applying Equation (IV.10) to all fields, locations, or geo units, we can calculate comprehensive environmental risk indices for each category within every field, location, or geo unit. This method offers a systematic approach to assessing environmental risks, facilitating data-driven decisions on tool upgrades by considering the nuanced influence of environmental conditions on tool reliability.

IV.5.2 Case study

In order to illustrate the practical application of the proposed environmental risk estimation method, this case study employs data collected from ten drilling operations involving a specific LWD tool as a demonstrative example. These ten drilling operations were conducted within a specific geographical location in a particular country.

In this case study, we initiate by defining three distinct environmental temperature levels: <100°C, 100-130°C, and >130°C. Subsequently, we compute the exposure duration for these three temperature levels across the ten drilling operations, as depicted in Figure 1. Following this, we establish the two critical thresholds for exposure time: one for high risk (Th_{g3}), set at 5 hours, and the other for medium risk (Th_{g2}), set at 20 hours. These threshold values are also visually represented in Figure IV.6.

Utilizing the computed exposure durations and the two thresholds, we assign risk levels to the ten drilling operations, as delineated in Table IV.7. Finally, we calculate the risk index based on Equation (IV.10) as follows:

$$r_g = 0.3 \times \frac{8}{10} + \frac{1}{10} = 0.34 \tag{IV.11}$$

where W is defined as 0.3. This calculated risk index serves as a quantitative characterization of the likelihood of encountering adverse temperature conditions during a drilling operation conducted within this specified location.

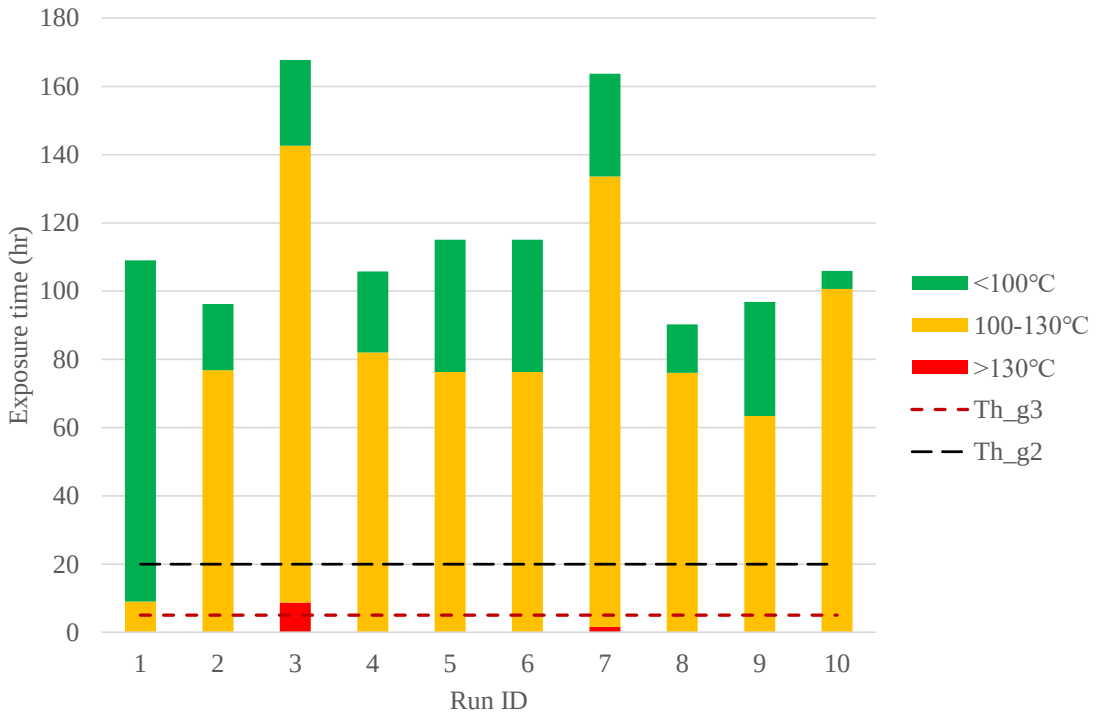


Figure IV.6 – Computed exposure time of the three temperature levels and the two thresholds

Table IV.7 – Risk level assignments for the ten drilling runs

Run ID	1	2	3	4	5
Risk Level	Low	Medium	High	Medium	Medium
Run ID	6	7	8	9	10
Risk Level	Medium	Medium	Medium	Medium	Medium

IV.6 Conclusion

This chapter underscores the fundamental importance of risk estimates in complex drilling operations within the oil and gas industry. As the drilling process has evolved into a complex endeavor heavily reliant on sophisticated tools, the reliability of the electronic boards in these tools has become a significant concern. A review of relevant research in health assessment, failure prognostics, and risk estimation of electronic equipment reveals two significant research gaps. Firstly, existing research has predominantly focused on individual electronic components, neglecting the examination of electronic assemblies such as circuit boards, which are prevalent in D&M tools. Secondly, many studies have limited their investigations to a single environmental category, primarily focusing on factors such as

temperature or vibration, without exploring the combined effects of multiple environmental variables on electronics health. However, D&M tools are exposed to various environmental combinations, including high temperatures, wide-ranging shocks, and dynamic vibrations.

To address these gaps, we introduce an innovative method that utilizes HMM to estimate the risk of electronic boards. This method considers cumulative environmental exposures to temperature, shock, and vibration during multiple drilling processes, providing a holistic and comprehensive understanding of electronic board health. A Case study conducted to validate this method demonstrates its effectiveness. The sequence of risk level estimates for the training and test boards closely matched the actual conditions. Notably, most of the failed boards were correctly identified as high-risk. This encouraging result highlights the practical applicability and validity of the proposed method even when full life-cycle data are unavailable.

Furthermore, understanding and estimating operational environment risks are essential to guide decision-making for tool maintenance scheduling and fleet maintenance planning. Therefore, we introduce a systematic method for assessing the likelihood of encountering unfavorable operating conditions during drilling operations in a given geographic location. The method involves defining environmental risk levels, calculating exposure times, assigning risk levels, and calculating a risk index. We demonstrate how to implement this method in a case study. In the next chapter, we will show how to make risk-based maintenance decisions based on the risk estimation results in this chapter. Risk-based decisions constitute the next layer of the proposed risk-based maintenance decision-making framework, the decision layer.

Chapter V
Risk-based decision-making

V.1 Introduction

In the previous chapter, we outlined the process of estimating risks associated with electronic boards of D&M tools and oilfield operational environments. In this chapter, we will detail how to make maintenance decisions for D&M tools based on the risk estimation results from the preceding chapter. The decision approaches presented in this chapter align with the decision layer in the proposed risk-based maintenance decision-making framework introduced in Chapter II. Specifically, the decisions cover two aspects: maintenance scheduling for individual tools and maintenance planning for tool fleets. As highlighted in Chapter II, maintenance scheduling aims to select the most suitable D&M tool for the specific operational environment resulting from a given drilling job. In instances where no suitable tools are available, tool upgrades are tailored based on the associated risk of the operational environment. On the other hand, maintenance planning aims to ensure that a sufficient number of suitable D&M tools are available to address potential risks related to the operational environments of upcoming drilling jobs in the next year.

In the earlier discussed Chapter I of the General Introduction, we established the fundamental role of the BHA in drilling operations and highlighted its configuration variability based on specific job requirements. The BHA, a vital part of a drilling system, comprises various D&M tools, each representing a complex electronic system. Through decades of development, the built-in electronic boards in these D&M tools have undergone various design revisions, resulting in variations in electronic board reliability. This diversity in reliability, in turn, affects the overall tool reliability. Meanwhile, the reliability of D&M tools plays a vital role in drilling operation [122].

Another main factor affecting the success and efficiency of drilling operations is the oilfield operational environment because operational environment could have a great impact on tool reliability. For example, elevated vibration can cause mechanical structure damage, and high temperature and humidity can cause electronic malfunctions [108]. For an upcoming drilling job, if the D&M tools that make up the BHA are not compatible with the operational environment, this would cause the job to fail and/or tool failures, resulting in a substantial economic loss. From the perspective of field engineers, they definitely would prefer to use the most reliable D&M tools to configure the BHA. This way, the BHA reliability can be maximized, but this is unrealistic and non-optimal. On the one hand, the availability of tools of high reliability is limited. If these tools were used for drilling operations of low-risk operational environments, then there might not be suitable tools for drilling jobs of high-risk operational environments. On the other hand, tools of low reliability might not be compatible with harsh environments but could be used for drilling operations in moderate environments. In addition, tools of higher reliability or configuration generally mean added manufacturing costs. Therefore, selecting the right D&M tool for the right field environment is of great significance for tool maintenance scheduling and cost savings.

However, the current tool selection process is primarily based on the "first come, first serve" principle. Specifically, once a contract is agreed upon between the client and the service provider company and a tentative shipment date is established, the field engineer will allocate available tools based on specific requirements of the oil well parameters, such as hole size. Importantly, this allocation is made without considering the tools' operational history and their parts' revisions. In addition, the tool selection usually requires the field engineer to manually check many data sheets, which is labor-intensive and inefficient. Moreover, this decision-making process neglected the impact of environmental factors on the drilling operations and tool reliability. Operational environment risks can significantly affect the reliability of D&M tools and result in the need for high-reliability tools in drilling operations with high-risk operational environments[52]. A more detailed examination of the current tool selection decision-making process will be provided in the following section. In light of these challenges, we propose a novel indicator designed to characterize the compatibility or fitness of tools with different field operational environments. This indicator can facilitate more informed decisions regarding tool selection and upgrades, i.e., maintenance scheduling for individual tools.

The process of recovering failed D&M tools is a protracted one and is contingent upon the lead time for delivering new parts. This delivery time can vary considerably, spanning from a few weeks to several months, influenced by factors like the complexity of part manufacturing and the efficiency of the delivery company. Thus, maintenance bases have endeavored to proactively plan for their D&M tool fleets to mitigate these delays. This decision-making process involves forecasting drilling operations and maintaining an inventory of tools based on past failures, ensuring they always have tools available for new jobs. However, once again, this decision-making process relies on manual analysis of extensive data sheets and subjective expert opinions, which is both time-consuming and subject to subjectivity. A more detailed examination of the current fleet maintenance planning decision-making process will be provided in the following section.

In response to the challenges posed by the current maintenance planning of D&M tool fleets, we propose a novel D&M tool fleet maintenance planning solution that centers around data-driven decision-making. This solution systematically analyzes multiple existing data sources, encompassing historical operational environment data, part reliability levels, instances of part failures, and cost considerations. It aims to provide decision-makers with straightforward insights to purchase new parts for D&M tool fleets, aiming to enhance the overall reliability for handling upcoming drilling jobs in the next year.

The remainder of this chapter is structured into six sections. Section 2 delves into the current decision-making processes for D&M tools in the oil and gas industry. Section 3 elucidates compatibility rules and offers an example for clarity. Section 4 introduces an indicator called tool compatibility index for facilitating risk-based tool selection and upgrade decision-making. Section 5 unveils a novel solution tailored for risk-based fleet maintenance planning. Section 6 illustrates

the application of the two decision frameworks through various use cases. Finally, Section 7 serves as the conclusion, summarizing key insights and underscoring the significance of data-driven risk-based decision-making in the oil and gas industry.

V.2 Current decision-making processes

In this section, a concise overview of the Digital Fleet Management System (DFMS) utilized in the current decision-making process for maintenance scheduling and planning is provided. Subsequently, a comprehensive description of the existing D&M tool selection processes and their limitations is presented. Finally, this section details the current D&M fleet maintenance planning decision-making processes and their associated limitations.

V.2.1 Brief introduction of DFMS

The DFMS is a commercialized business information dashboard to help field engineers choose the most reliable D&M tool for a given drilling job. As mentioned, the revision design of D&M tool electronic boards develops with time. Even with the same design, the electronic parts can slightly differ from batch to batch. Indeed, no two tools today have the same hardware configuration or reliability. The DFMS is designed to extract necessary CMMS data for all parts of D&M tools, and compute a reliability score based on the following information:

- Manufacturing process changes
- Supplier traceability
- Design changes

The reliability score has three levels; i.e., Level 1 meaning the least reliable and Level 3 indicating the most reliable.

The DFMS output materializes as an intuitive dashboard, offering a comprehensive overview of active D&M tools for each location. The dashboard encompasses vital configurations, including equipment hierarchy, part status (e.g., active, junked, lost in hole), and detailed part identity information (e.g., part number and serial number). Additionally, the dashboard provides insights into parts revision and assigns reliability scores to offer a quick and concise assessment of tool reliability.

V.2.2 Current tool selection process and limitations

Currently, the tool selection decision-making involves the following three steps as shown in Figure V.1:

1. The field engineer obtains some general parameters (e.g., geographical coordinate, temperature, flow rate, and hole size) of the planned well where the upcoming drilling job will take place.

2. The field engineer refers to the DFMS where information is available about equipment revision design, equipment status and parts reliability levels of tools in the location.
3. According to the DFMS output and experience, the field engineer decides which tool to deploy.

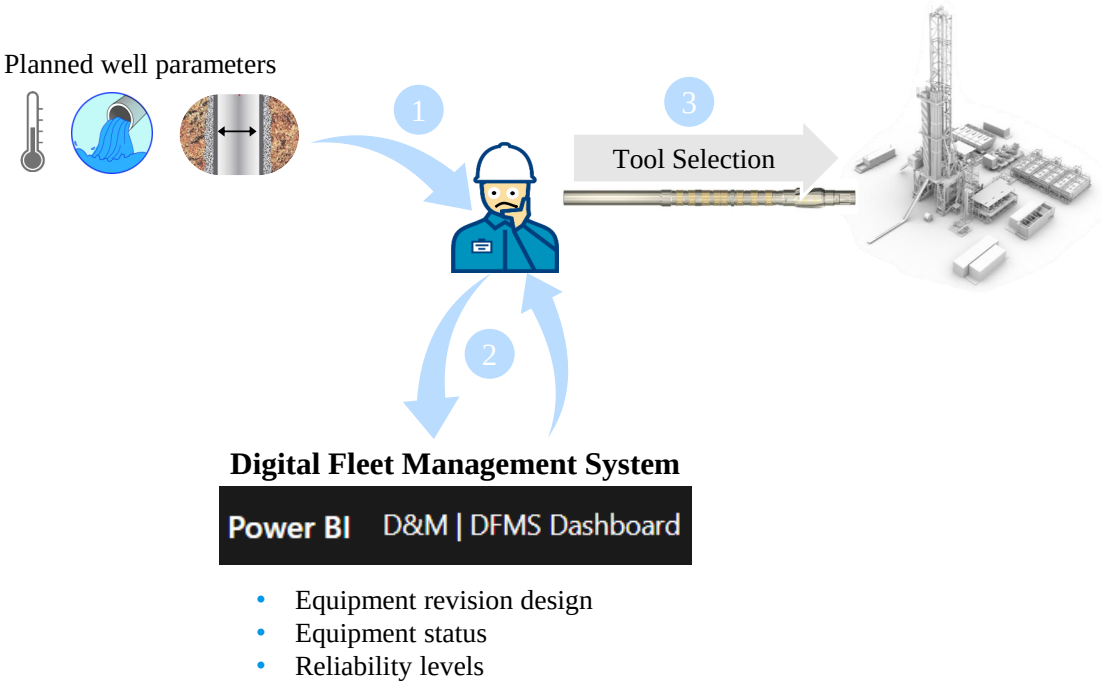


Figure V.1 – Illustration of current tool selection decision-making process

Despite the richness of the DFMS data, a critical nuance comes to the fore. The parts' reliability levels, while valuable, do not encompass environmental definitions or systematic criticality information. In other words, they fail to consider the dynamic oilfield operational environment and the relative importance of parts when determining reliability scores. Consequently, the current DFMS output informs which tool or part boasts superior reliability but falls short in quantifying the tool's compatibility with the operational environment. Consequently, tool selection continues to rely heavily on empirical knowledge.

As highlighted in Section V.1, this process is not only reliant on empirical wisdom but is also labor-intensive. In order to achieve objective, effective, and efficient tool selection decision making, it was decided to develop a tool compatibility computation method, which can provide the field engineer a score range from 0% to 100%. This score serves as an indicator of the tools' suitability under a specific field environment (e.g., high temperature, medium vibration, and shocks) where the upcoming job will demonstrate. This innovative solution is poised to improve tool selection and upgrade decision-making, infusing it with precision and data-driven prowess.

V.2.3 Current fleet maintenance planning decision-making process and limitations

The current fleet maintenance planning decision-making process, as depicted in Figure V.2, involves the following steps:

1. **Year-End Fleet Examination:** At the conclusion of each year, maintenance and operation bases conduct a comprehensive assessment of their tool fleet. During this assessment, they determine which parts or tools should be discarded, and then they update this information within the CMMS system.
2. **DFMS Dashboard Update:** The updated information from the CMMS is integrated into the DFMS dashboard, effectively refreshing the dashboard with the latest fleet status.
3. **Drilling Operations Forecast:** The marketing team provides forecasts for the drilling operations expected in the upcoming year. These forecasts are then delivered to the decision maker, who is typically the location or geounit manager.
4. **Data Collection:** The decision maker collects additional data, including historical fleet utilization statistics and failure statistics from the CMMS. Furthermore, they review summary information of fleet status from the DFMS dashboard.
5. **Decision-making:** Armed with this collected information, the decision maker proceeds to make crucial decisions regarding tool and part orders from the manufacturing center.

It is worth noting that this process exhibits several limitations. Specifically, it involves a labor-intensive manual examination of various data sheets. Furthermore, it does not account for the potential impact of risks of operational environments of future drilling jobs. These shortcomings align with the issues identified in the tool selection decision-making process discussed previously.

V.3 Compatibility rules

V.3.1 Definition

The compatibility rules are defined by SMEs. Different tools have different rules. The rules contain parts criticality information and compatible environment definitions, which overcome the disadvantages of the DFMS reliability levels.

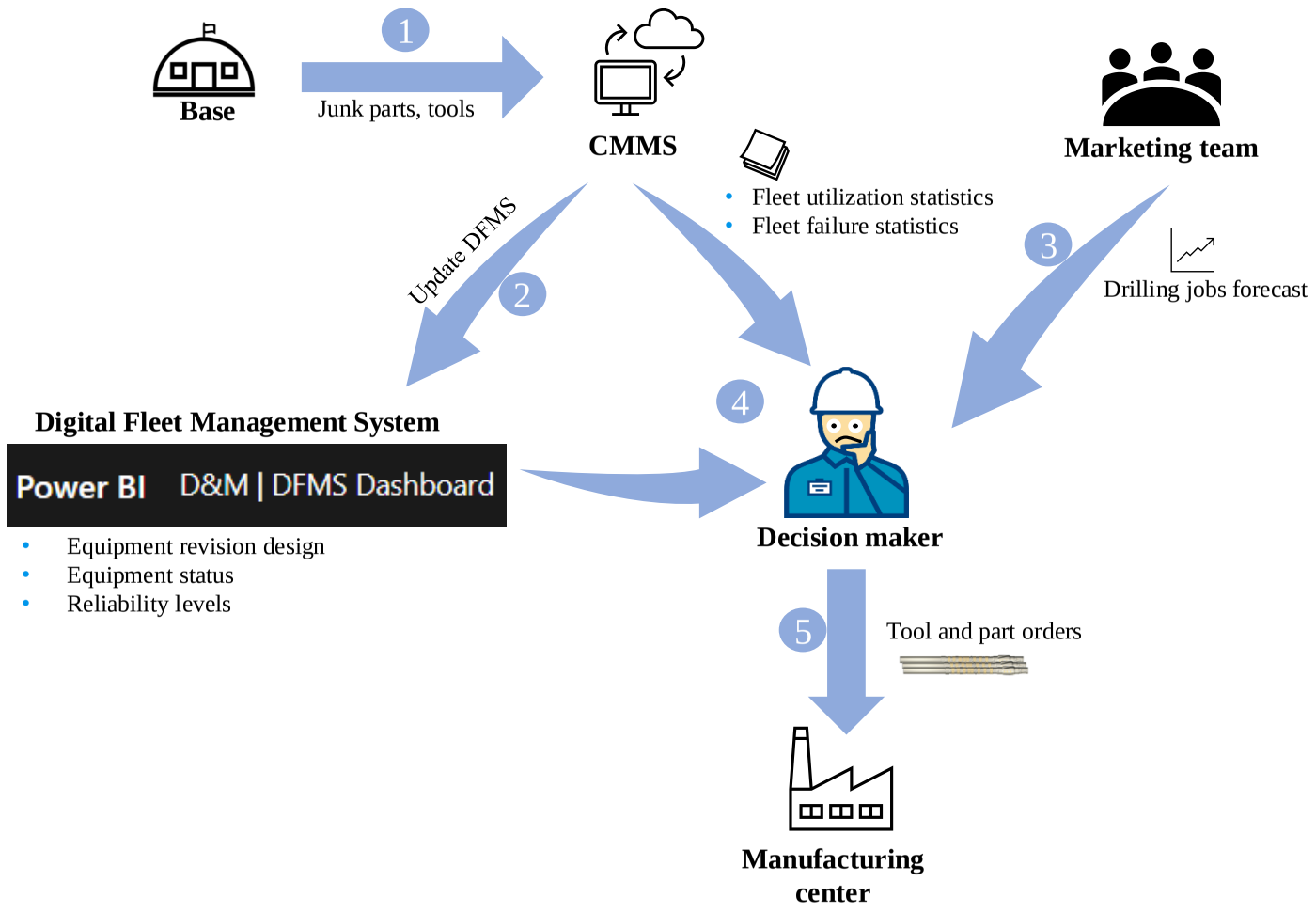


Figure V.2 – Illustration of current fleet maintenance planning decision-making process

Table V.1 – An excerpt from the compatibility rules of a specific LWD tool

Part Name	DFMS Reliability Level	Temperature	Lateral Vibration	Lateral Shock	Criticality
X215	1	NA	NA	NA	4
X215	2	1,2	1,2	1,2	4
X215	3	1,2,3	1,2,3	1,2,3	4
X105	1	NA	NA	NA	3
X105	2	1,2	1,2	1,2	3
X105	3	1,2,3	1,2,3	1,2,3	3
SX207	1	1	1	1	2
X207	2	1,2	1,2	1,2	2
X207	3	1,2,3	1,2,3	1,2,3	2
X117	1	1	1	1	1
X117	2	1,2	1,2	1,2	1
X117	3	1,2,3	1,2,3	1,2,3	1

Not all parts exhibit the same level of impact in the event of a failure, and their failure rates vary. Therefore, SMEs define the significance of each part based on historical service quality statistics, specifically, the occurrence of failure events. These statistics enable SMEs to determine the number of failure events associated with each part and classify their criticality using straightforward binning methods. For instance, parts with five or fewer failure events may be assigned a criticality rating of 1, while those with six to ten failure events receive a rating of 2. Furthermore, SMEs assist in establishing rules that map the reliability levels of parts to critical operational environments. These critical environments pertain to the environment categories (e.g., temperature, vibration, shocks) capable of causing tool failure, which may differ among various D&M tools.

Table V.1 offers a segment of compatibility rules for a specific LWD tool, as defined by the corresponding tool SME. In the table, "NA" signifies that the part is not suitable for any environment, rendering it obsolete. Numerical values, such as 1, 2, and 3 in the Temperature, Lateral Vibration, and Lateral Shock columns, denote low, medium, and high levels of these environment categories, respectively. The Criticality column shows the importance of parts. The larger the value, the more important or critical is the part. For example, the first row of the table indicates the X215 part with reliability level 1 should not used, the X215 part has criticality value of 4. The fifth row suggests that the X105 part with reliability level 2 can be used for a drilling operation of low and medium temperature, low and medium lateral vibration, low and medium lateral shock. the X215 part has criticality values of 3.

V.3.2 Effects of electronic board risk estimates on compatibility rules

The results of electronic board failure risk estimation can have an impact on compatibility rules. However, this influence is confined to specific types of electronic boards where sufficient failure data is available, allowing the construction of effective risk estimation models for estimating the failure risk of corresponding electronic boards. In addition, whether the data is sufficient or not can be determined by the data quality management method proposed in Chapter III.

We can modify the compatibility rules for the electronic boards with effective risk estimation models. For instance, in a four-level risk estimation model like the one introduced in Chapter IV, if the electronic board is estimated as risk level 4 (the highest risk level), it necessitates replacement (resulting in a forced incompatibility with any environment, i.e., changing the compatible environment to "NA"). Conversely, if the board is assessed as risk level 1 (the lowest risk level), the corresponding compatibility rules remain unchanged. In other cases (estimated as levels 2 and 3), the board is designated for use only in jobs of low-level environments (e.g., low temperature, low vibration, and low shocks) in the compatibility rules. In other words, its compatible environment is forced to "1".

V.4 Risk-based maintenance scheduling for individual tools

V.4.1 Proposed solution

This section will commence by introducing a method for calculating the tool compatibility index. Subsequently, we will conclude this section by presenting an overview of the framework for the proposed solution for tool maintenance scheduling.

V.4.1.a Tool compatibility index

To account for part criticality and the suitability of different environments in the decision-making processes for risk-based maintenance scheduling for individual tools, we introduce a novel metric called tool compatibility index. Denoted as I, this index is formally expressed in Equation (V.1). This indicator leverages a straightforward weighted average method, which ensures ease of comprehension and straightforward interpretation of the results. The values of this indicator are confined within a range of 0% to 100%.

$$I = \frac{\sum_{i=1}^{N} w_i x_i}{\sum_{i=1}^{N} w_i} \times 100\%, \tag{V.1}$$

where

— N is the number of parts in the tool,

— w_i is the criticality value of part i in the tool,

— x_i is 1 if part i can be used for the specific environment; otherwise, x_i is 0. We can infer the w_i and x_i according to the compatibility rules defined by SMEs.

To illustrate how the tool compatibility index is calculated, we provide an example using an LWD tool, as depicted in Figure V.3. Let consider that the upcoming drilling job will be conducted in an oil field characterized by a medium temperature, high lateral vibration, and high lateral shock environment.

Using the information from the DFMS output and the predefined compatibility rules, we can deduce the compatibility values x_i for the components of the tool, as shown on the right-hand side of the figure. For instance, the X105 part in the fifth row is suitable for use in a medium-temperature environment but is not recommended for high lateral vibration or high lateral shock environments, resulting in a corresponding x_i value of 0.

After obtaining the x_i values for all parts within this tool, we incorporate both the criticality values w_i and x_i into Equation (V.1). This calculation reveals that the tool compatibility index for this particular tool, when operating in the specified environment, is 60%. It's important to note that the zero values of x_i highlight the non-compatible parts of this tool, indicating that these parts would need upgrades if the field engineer intends to use this tool in the specified environment.

V.4.1.b Framework of the proposed solution

The framework for the new solution for tool maintenance scheduling, based on the compatibility rules and the tool compatibility index calculation method, is depicted in Figure V.4. The framework encompasses the following steps:

1. SMEs define compatibility rules for D&M tools. These rules are established by mapping parts to specific environmental conditions, taking historical service quality statistics into account, and incorporating DFMS reliability levels and electronic board risk estimates if applicable.
2. Using the information provided by the DFMS, the tool configurations are determined. These configurations are then combined with the field environment parameters selected by the field engineer. Subsequently, this data is fed into the Tool Compatibility Analyzer, which calculates the tool compatibility index based on Equation (V.1).
3. The tool compatibility index for the tool is generated as an output. In addition, information about non-compatible parts and the associated upgrade costs can be obtained.
4. By applying the steps mentioned in 2 and 3 to the entire tool fleet and all possible field environment combinations (e.g., considering all permutations of three environmental categories, each with three levels, resulting in $3^3 = 27$ combinations), compatibility indices, non-compatible parts and associated

upgrading recommendations and costs for the entire tool fleet under each environment combination are obtained. These outputs can be used to create a dashboard for decision-makers.

This framework enhances tool maintenance scheduling by providing a systematic and data-driven approach that considers both part criticality and potential field operational environment. It facilitates informed decision-making while optimizing the reliability and performance of D&M tools in various drilling environments.

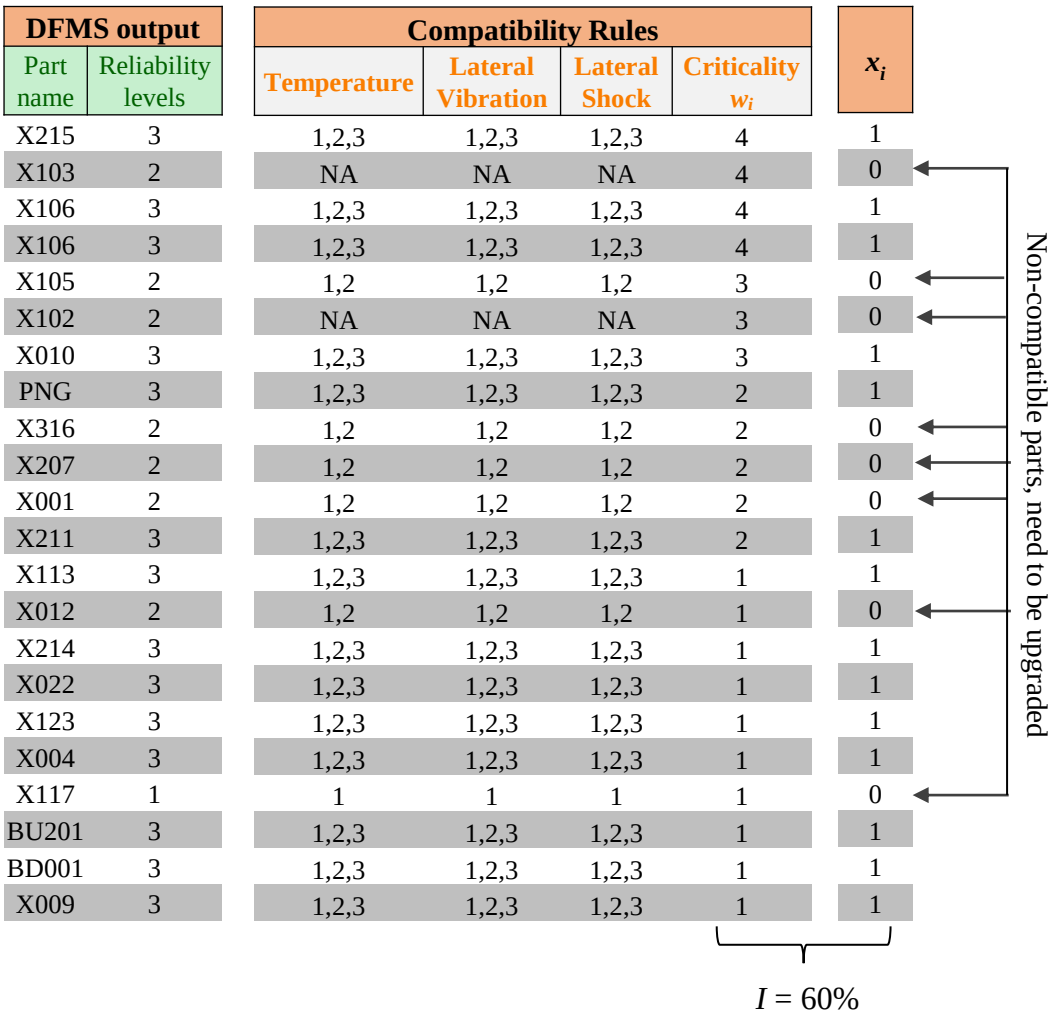
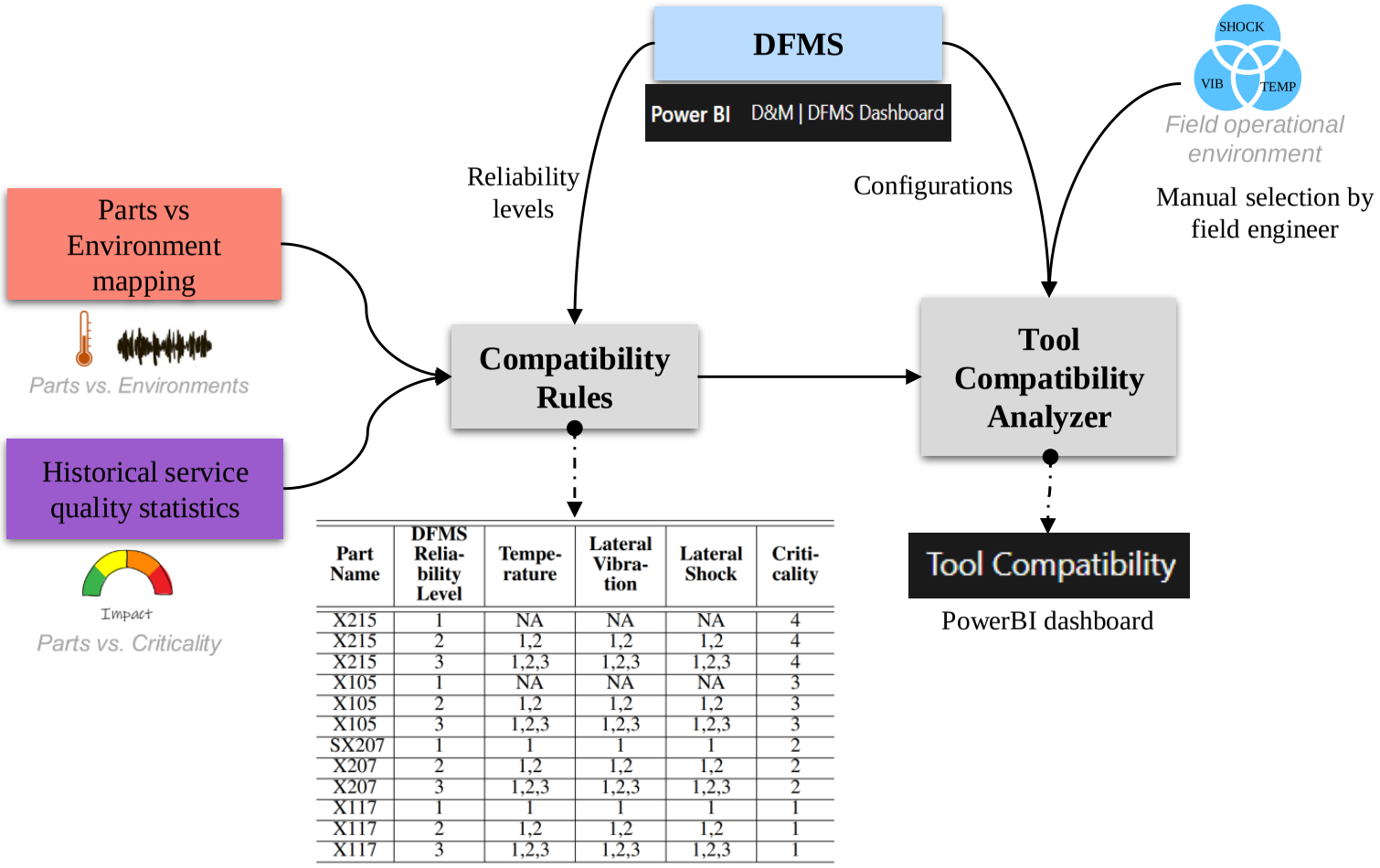


Figure V.3 – Tool compatibility index computation example



Part Name	DFMS Reliability Level	Temperature	Lateral Vibration	Lateral Shock	Criticality
X215	1	NA	NA	NA	4
X215	2	1,2	1,2	1,2	4
X215	3	1,2,3	1,2,3	1,2,3	4
X105	1	NA	NA	NA	3
X105	2	1,2	1,2	1,2	3
X105	3	1,2,3	1,2,3	1,2,3	3
SX207	1	1	1	1	2
X207	2	1,2	1,2	1,2	2
X207	3	1,2,3	1,2,3	1,2,3	2
X117	1	1	1	1	1
X117	2	1,2	1,2	1,2	1
X117	3	1,2,3	1,2,3	1,2,3	1

Figure V.4 – Framework of the proposed solution for tool maintenance scheduling

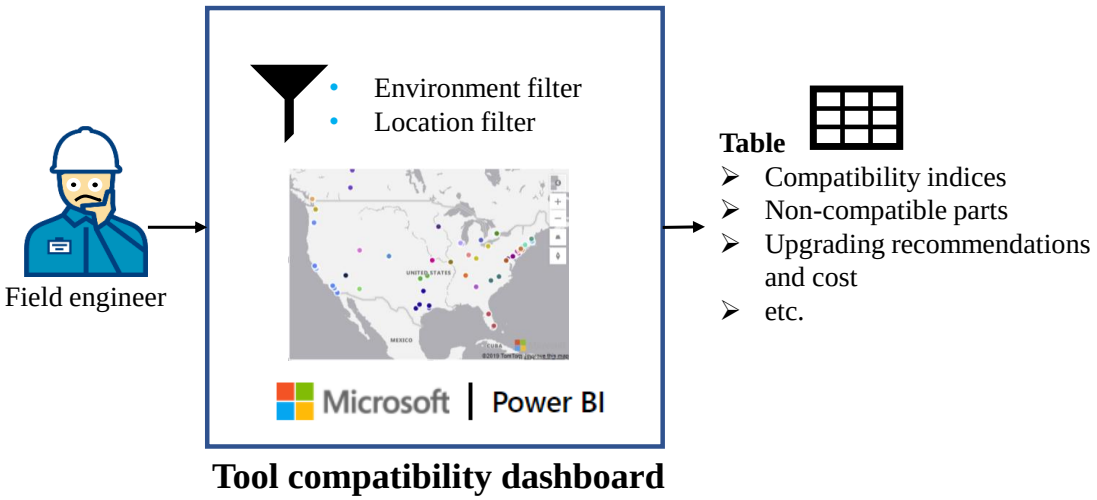


Figure V.5 – Use case diagram for tool compatibility dashboard

V.4.2 Use Cases

The proposed solution delivers a dashboard as its output, which provides compatibility indices for the tool fleet across various environmental conditions. In this subsection, we showcase the use case diagram for the dashboard and two application scenarios to illustrate the practical value of our solution. It should be noted that certain details have been omitted to ensure confidentiality.

V.4.2.a Use case diagram

A use case diagram is a visual representation that illustrates how a system interacts with external entities, such as users or other systems, by depicting various use cases. It helps understand the high-level functionalities and the relationships between external entities and use cases in a system [147].

Figure V.5 depicts the use case diagram. The users of the dashboard are field engineers tasked with tool operation responsibilities. In this diagram, the field engineer initiates actions such as selecting the field environment for an upcoming job and specifying the maintenance base location by employing the environment filter and location filter functionalities. Subsequently, the dashboard responds by providing a table that contains compatibility index information regarding the tool fleet in the selected location. Additionally, it offers suggestions for upgrades and the associated costs for non-compatible parts of each tool.

V.4.2.b Application scenarios

This subsection presents two scenarios, as depicted in Figure V.6, to illustrate the proposed solution.

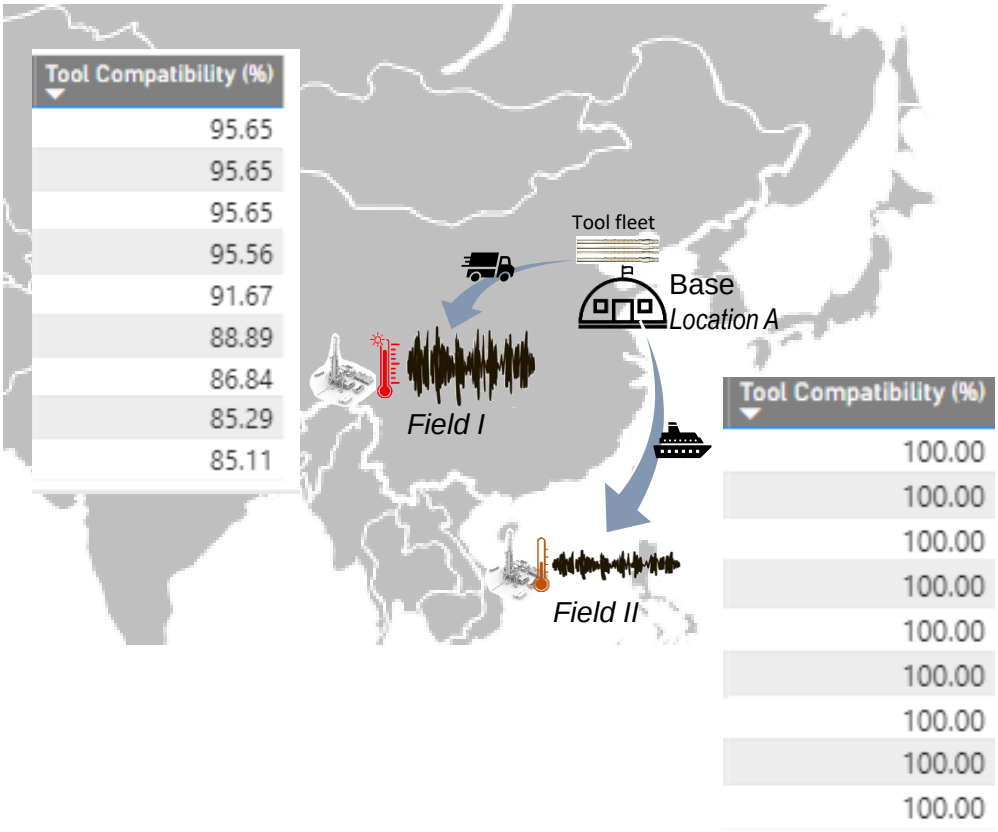


Figure V.6 – Two application scenarios for tool compatibility index dashboard

— **Scenario I:** In this scenario, a new drilling operation is scheduled for *Field I*, an environment characterized by *high temperature, high lateral vibration, and high shock*. The job is designated for the *Location A* base. The corresponding compatibility indices of the tool fleet for this specific environment are displayed alongside "Field I" in Figure V.6.

— **Scenario II:** In the second scenario, a new drilling job is planned for *Field II*, featuring a *medium temperature, low lateral vibration, and low shock* environment. Once again, the job is assigned to the *Location A* base. The compatibility indices of the tool fleet for this environment are presented adjacent to "Field II" in Figure V.6.

The comparison of compatibility indices for these two scenarios reveals that the same tool fleet exhibits distinct tool compatibility indices under varying environmental conditions. The tool compatibility indices for the challenging environment (Scenario I) are lower than those for the milder environment (Scenario II). By establishing a compatibility index threshold of 90% as the tool selection criterion, it becomes evident that only the first five tools are compatible with the harsh environment (Scenario I). Consequently, these tools can be designated for use in the Scenario I drilling job. Conversely, all of the tools can be selected for the Scenario II drilling job because all of their compatibility indices exceed the 90% threshold.

V.5 Risk-based maintenance planning for tool fleets

V.5.1 Proposed solution

The proposed solution for fleet maintenance planning hinges on the tool compatibility index and operational environment risk estimation introduced in Section IV.5. This solution comprises three essential components: fleet compatibility assessment, operational environment risk estimation, and maintenance planning optimization modeling. Since the preceding chapter thoroughly covered operational environment risk estimation, it will not be rehashed here. Instead, this section will begin by introducing a method for evaluating fleet compatibility, building upon the tool compatibility index detailed in the previous section. Following that, we will formulate a mathematical optimization model to address the D&M tool fleet maintenance planning problem. Finally, we will conclude this section by presenting an overview of the framework of the proposed solution, which integrates these three core components.

V.5.1.a Fleet compatibility assessment

By applying the tool compatibility calculation to a fleet of D&M tools, one can get the compatibility indices for the entire fleet. Suppose we define a value I_{th} as the compatibility index threshold. In other words, if the tool compatibility index of tool j is greater than I_{th}, then this tool is considered compatible with the environment; otherwise, it is deemed incompatible. Then, we can assess what percentage of tools in this fleet are compatible with the given environment, which is mathematically expressed as in Equation (V.2).

$$Prt = \frac{\sum_{j=1}^{M} \mathbb{I}(I_j > I_{th})}{M}, \tag{V.2}$$

where M denotes number of tools in the fleet, I_j represents the tool compatibility index of tool j, and $\mathbb{I}$ is an indicator function.

V.5.1.b Fleet maintenance planning optimization modeling

As mentioned in Section IV.5, the operational environment risk index serves as a statistical representation of historical drilling jobs potentially exposed to hazardous conditions, such as elevated temperatures, shocks, and vibrations. This index can therefore be regarded as a probability or likelihood that a future drilling job will encounter a risky environment.

It is imperative for the field (location or geounit) to sustain a sufficient number of tools compatible with harsh environments to ensure the readiness for future drilling jobs that pose operational environment risks such as high temperature, shock, and vibration. The percentage of compatible tools is recommended to be not less than the risk index plus an additional buffer within the range of 10% to

30%. This additional buffer serves as a precautionary measure to guarantee the continued availability of tools even in the event of unavailability caused by tool failures and maintenance needs.

Based on the fleet compatibility assessment, we can obtain the percentage of tools compatible with harsh or risky environments for the field (location or geounit). Suppose the field (location or geounit) does not have the number of compatible tools as recommended. In that case, a fleet maintenance plan is necessary to determine the most cost-effective replacement of parts to ensure the adequate availability of compatible tools. This chapter will focus on fleet maintenance planning at the geounit level. The fleet maintenance planning at the field level and the location level are similar.

Suppose the geounit has M tools and each tool has N parts, and the geounit does not have adequate compatible tools. A part will be compatible with the harsh environment if it is replaced with one of the latest revision design. Then, based on the above analysis, the fleet maintenance planning optimization problem is formulized as follows:

— **Decision Variables**

$$X_{ij} \in \{0, 1\}, i = 1, 2, \ldots, N; j = 1, 2, \ldots, M \tag{V.3}$$

where X_{ij} denotes if the part i in tool j is replaced, with 1 indicating replaced, and 0 meaning not.

— **Objective Function**

$$\min \sum_{j=1}^{M} \sum_{i=1}^{N} X_{ij} c_i, \tag{V.4}$$

where c_i denotes the replacement cost of part i.

— **Constraints**

$$X_{ij} = 0 \textbf{ if } z_{ij} = 1 \tag{V.5}$$

$$z'_{ij} = X_{ij} + z_{ij} \tag{V.6}$$

$$Prt' \geq r + b \tag{V.7}$$

- The first constraint means that if part i in tool j is compatible with the environment, then there is no need for replacement.
- In the second constraint, z'_{ij} is the compatible flag of part i in tool j after part replacement. In other words, if part i in tool j is replaced with a new part, then it is compatible with the environment; otherwise, its compatible flag remains unchanged.
- In the third constraint, r is the operational environment risk index of the geounit; b is the buffer size; Prt' is the percentage of compatible tools after part replacement, which can be computed via Equation (V.1) and Equation (V.2).

The previously discussed optimization problem is a simple integer programming problem that can be solved using various methods, including conventional methods like cutting plane, branch and bound [148], or evolutionary algorithms [149]. Additionally, due to the binary nature of the decision variables, a brute-force search method can also be used to solve this problem, which is relatively simple to understand and implement. This chapter will solve the optimization problem based on the brute-force search method, and the solution involves the following two steps:

1. Determine the minimum cost for replacing parts of each incompatible tool to reach the compatibility threshold. This cost is obtained by enumerating all possible solutions of the decision variables of the tool. Furthermore, sort these minimum costs in ascending order.
2. Based on the required number of compatible tools (i.e., $M \times (r+b)$) and the available number of compatible tools, one can determine the number of tools that need part replacements, referred to as m. Then the tools corresponding to the top m costs in Step 1 are the tools that require part replacement, and the sum of these m costs is the minimum value for the objective function.

By implementing the above modeling process for each environment category, one can obtain fleet maintenance planning solutions for all environment categories.

V.5.1.c Framework of the proposed solution

The proposed solution integrates fleet compatibility and operational environment risk estimates into D&M tool fleet maintenance planning. The framework of the proposed solution consists of three layers: data, model, and decision, as shown in Figure V.7.

— **Data Layer:** This layer serves as the input and incorporates three primary sources of information:
  - **Tool Measurement Data:** This includes historical operational environment information such as temperature, shocks, and vibration measurements.
  - **Asset Information:** This source provides details about tool identity, composed parts, and their locations.
  - **SME Knowledge:** SMEs defines the compatibility rules based on their knowledge.

— **Model Layer:** The model layer comprises three distinct models corresponding to the three core components mentioned earlier. These models are:
  - **Operational Environment Risk Estimation Model:** This model, using tool measurement data, computes risk indices for all environment categories within a geounit.
  - **Fleet Compatibility Assessment Model:** The second model evaluates the compatibility indices of all tools within the fleet and determines the percentage of tools that are compatible with the specified environmental requirements.

- **Fleet Maintenance Planning Optimization Model:** The third model integrates the requirements for compatible tools, taking into account the calculated risk indices and the compatibility data obtained from the fleet compatibility assessment. This information feeds into an optimization model that forms the basis for decision-making.

— **Decision Layer:** Serving as the output, this layer takes the results generated by the optimization model, which is typically solved using the brute-force search method. It determines which parts should be replaced and calculates the associated costs. These decisions guide the fleet maintenance planning process, ensuring the D&M tool fleet is adequately equipped to handle operational environment risks.

V.5.2 Use cases

Like the risk-based tool maintenance scheduling discussed in the previous section, the output of the proposed solution for fleet maintenance planning is presented as a dashboard that displays maintenance plans for various geounits. In order to showcase the effectiveness of this solution, this section provides a use case diagram and two real-life scenarios. It is important to note that certain information is omitted due to confidentiality constraints.

V.5.2.a Use case diagram

The use case diagram presented in Figure V.8 illustrates the interaction between geounit managers, who are the users of the dashboard, and the dashboard of fleet maintenance planning. These managers are responsible for making fleet maintenance planning decisions regarding the purchase of new tools or parts to for their respective geounits. The dashboard serves as a valuable tool for providing them with a holistic view of relevant information. This information aids them in making well-informed decisions. The dashboard offers several key features, including:

— **Operational Environment Risk Indices on a Map:** This feature displays operational environment risk indices of various locations within the selected geounit on a map.

— **Compatibility Information:** Geounit managers can access compatibility information for the entire fleet in the geounit. This information is presented in a tabular format.

— **Required Percentage of Compatible Tools:** The dashboard provides a bar chart that indicates the required percentage of compatible tools based on the risk index and buffer.

— **New Part Order Suggestions:** Users can receive suggestions for new part orders to replace the non-compatible parts, including associated costs, specifically tailored to the selected environment categories.

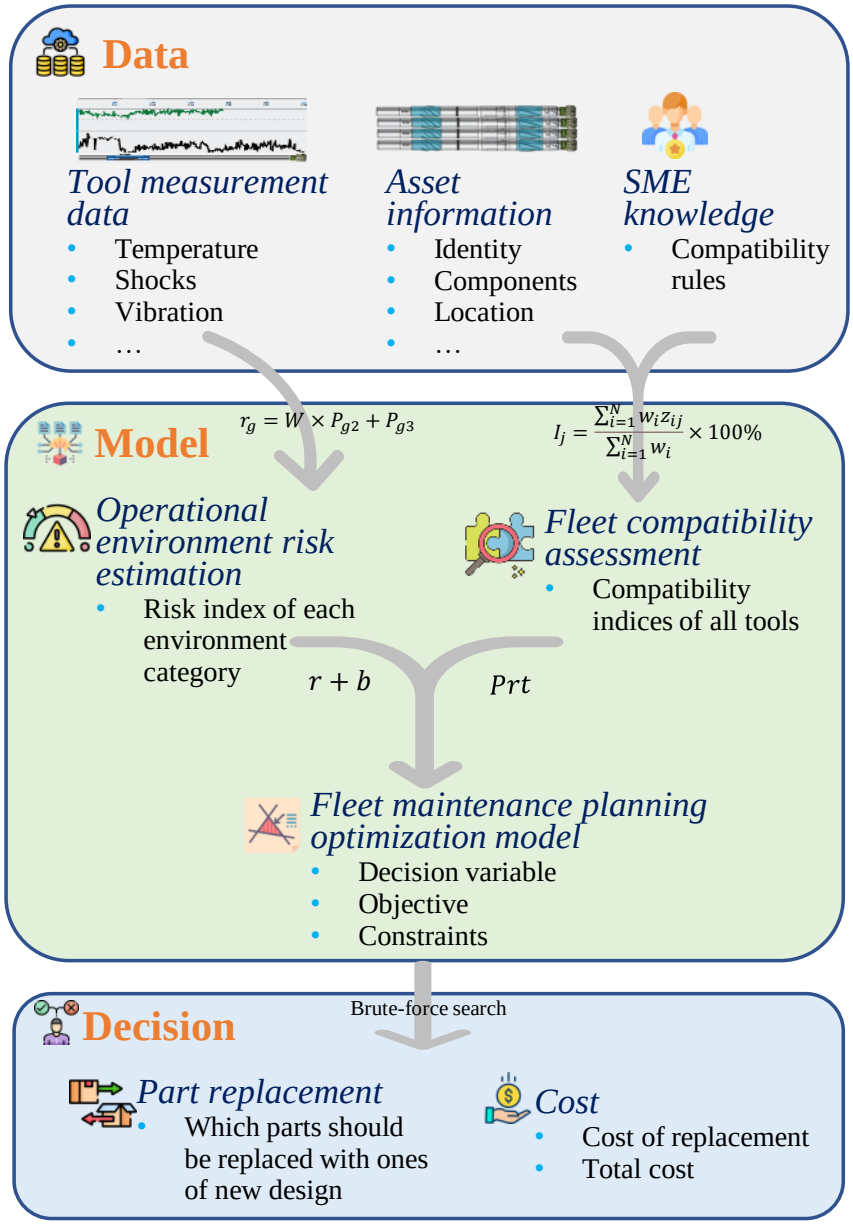


Figure V.7 – Framework of proposed solution for D&M tool fleet maintenance planning

This comprehensive dashboard equips geounit managers with the data and insights they need to make informed fleet maintenance planning decisions, optimizing the maintenance and readiness of their tool fleet.

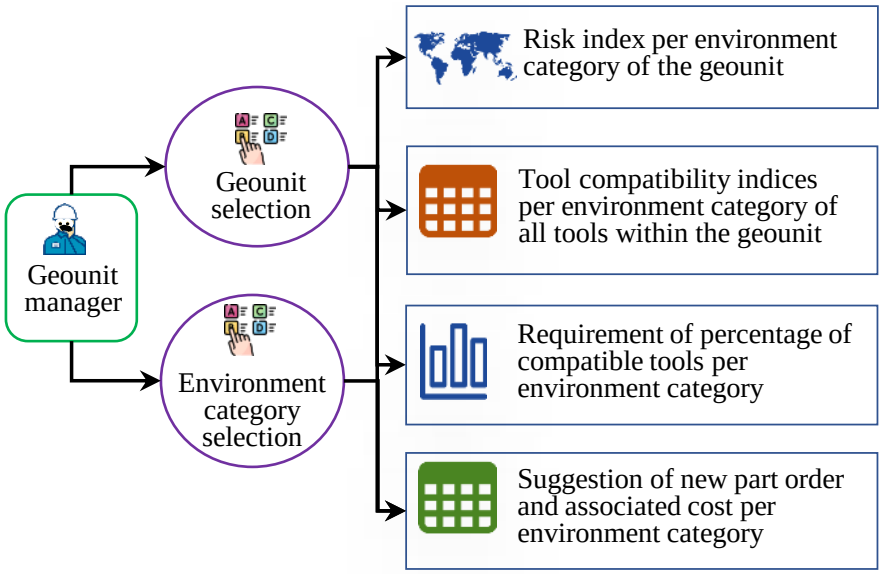


Figure V.8 – Use case diagram for fleet maintenance planning dashboard

V.5.2.b Application scenarios

Two scenarios (see Figure V.9 and Figure V.10) are presented in this subsection to demonstrate the solution. The compatibility threshold and buffer are set in both scenarios to 93% and 20%, respectively. The colored circle in the environmental risk index diagram indicates the average operational environment risk indices for the two environment categories, with darker colors indicating higher risk and lighter colors indicating lower risk.

— **Scenario 1:** The first scenario is about a geounit with a high-risk environment but low-percentage compatible tools. In this scenario, the environment risk index indicates that the environment is harsh. The compatibility information of the entire fleet shows that the current percentage of compatible tools is lower than the required percentage of compatible tools based on the risk index and buffer. The dashboard suggests several part replacements, and associated costs ensure the required number of compatible tools. The geounit manager can then make an informed decision based on the suggested part replacements and their costs and take the necessary action to invest in new parts to improve the compatibility of the fleet.

— **Scenario 2:** The second scenario is about a geounit with a low-risk environment. In this scenario, the environmental indices of the locations within the geounit indicate that the environment is relatively mild. The compatibility information of the entire fleet shows that the current percentage of compatible tools is already higher than the required percentage of compatible tools based on the risk index and buffer. The dashboard suggests no

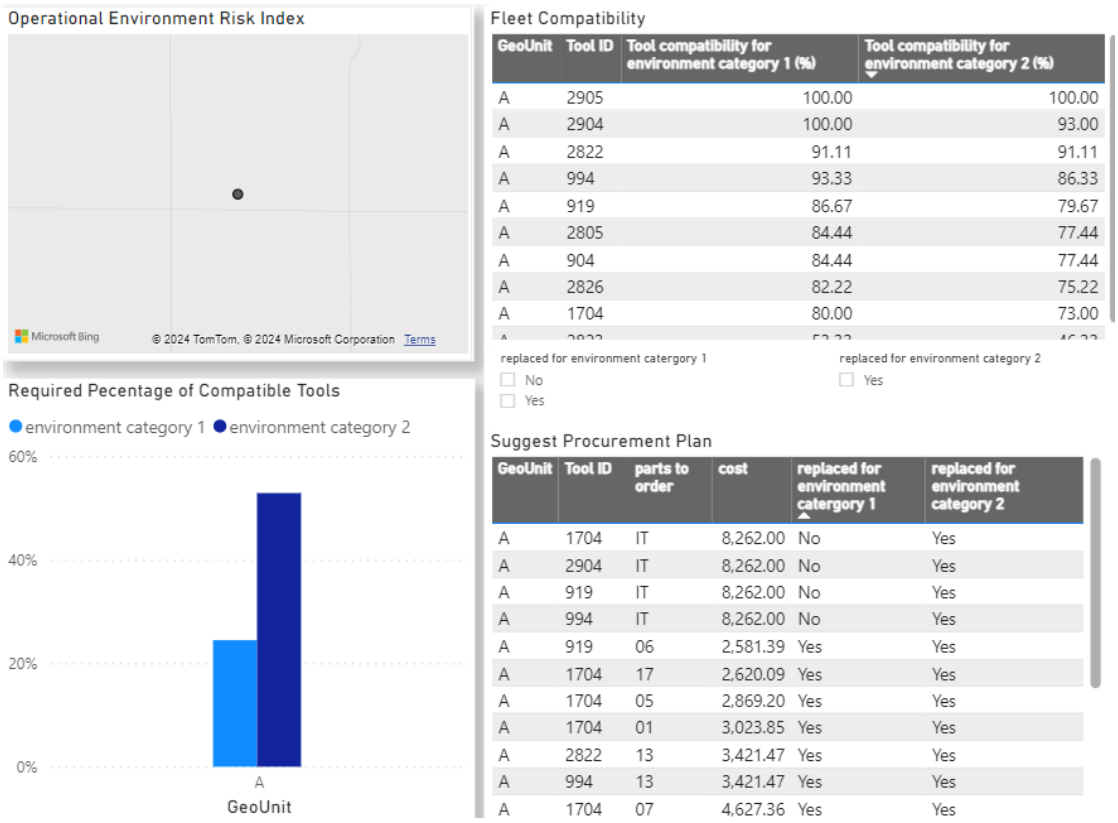


Figure V.9 – Application scenario 1 for fleet maintenance planning dashboard

part replacements, indicating that the current fleet is sufficient to handle the environmental conditions. The geounit manager can save investment costs and allocate resources to other areas.

These scenarios demonstrate the effectiveness of the proposed solution in providing actionable information to the geounit managers for fleet maintenance planning, taking into account both the operational environment risk estimates and the compatibility of the fleet. The solution enables managers to make informed decisions about which parts to replace and how much to invest, leading to improved reliability and cost savings for D&M tools compared to the existing fleet maintenance planning decision-making process.

V.6 Overall risk-based maintenance decision-making framework for D&M tools

The detailed overall framework for risk-based decision-making for D&M tools in this study, as depicted in Figure V.11, builds upon the concepts and research findings from previous chapters. It represents an enriched version of the framework introduced in Chapter II, offering a more intricate understanding of the underlying

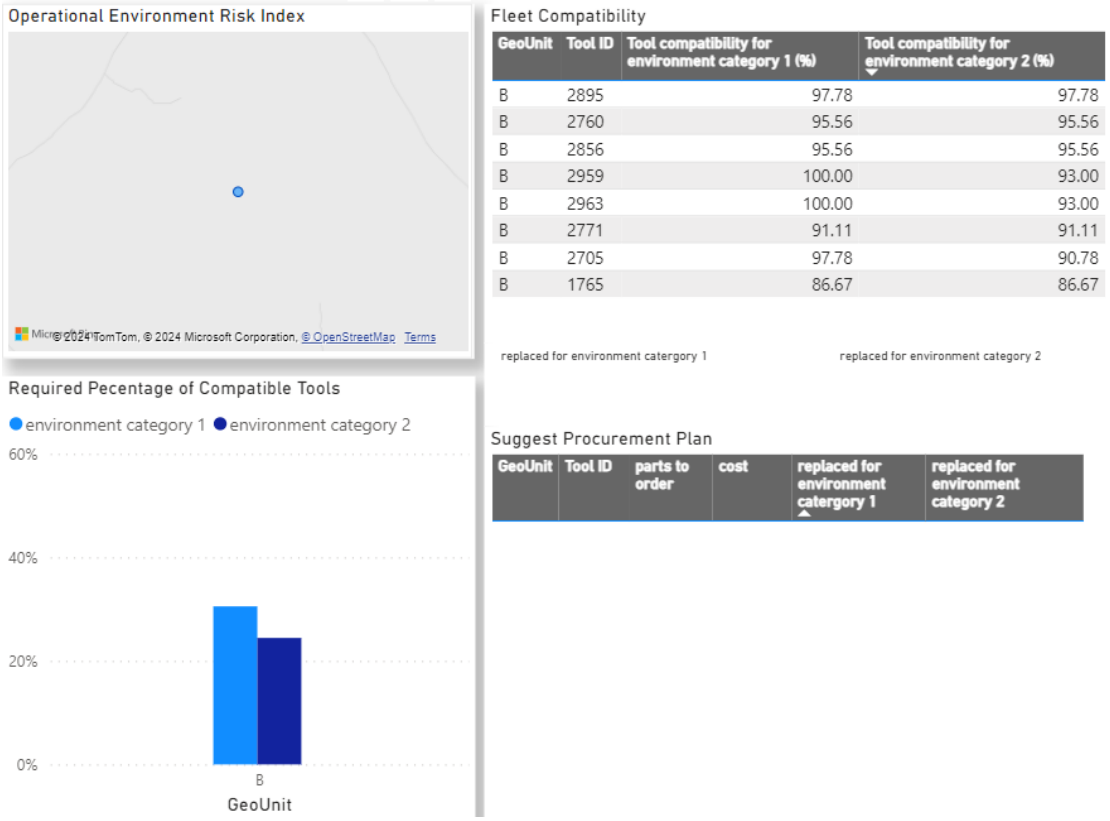


Figure V.10 – Application scenario 2 for fleet maintenance planning dashboard

processes specified for D&M tools. This framework is particularly detailed in its approach to data utilization and flow. It begins with the data layer, where all information is gathered from the CMMS. Within this data layer, compatibility rules and DFMS reliability levels are derived from the available data, while data quality management methods (as discussed in Chapter III) are employed to evaluate the data's adequacy for training the electronic board failure risk estimation model.

Moving to the second layer, we enter the risk estimation phase, which encompasses two critical dimensions: operational environment risk estimation and electronic board failure risk estimation. The results of electronic board failure risk estimation can, in turn, have an impact on compatibility rules and DFMS reliability levels, which was detailed in the subsection V.3.2.

The third layer is the decision-making phase, which involves computing the tool compatibility index based on compatibility rules and asset information. Additionally, the dashed line connecting the operational environment risk estimation and the tool compatibility analyzer suggests that operational environment risk estimation can also assist in determining the operational environment risk level for upcoming drilling jobs. This is achieved by applying two risk index thresholds to the operational environment risk estimation result (i.e., operational environment risk index r_g), resulting in the classification of the operational environment

risk index into three levels. With the tool compatibility analyzer and operational environment risk estimation method in place, the fleet maintenance planning optimization model can be developed.

The final layer encompasses the HMI. The outcomes from the tool compatibility analyzer and the fleet maintenance planning optimization model contribute to the creation of two dashboards, forming the HMI system designed for accessible and actionable decision-making.

V.7 Conclusion

In conclusion, this chapter addresses critical challenges related to tool maintenance scheduling and fleet maintenance planning of D&M tools in the oil and gas industry. It begins by highlighting the prevalent "first come, first serve" approach in tool selection, largely neglecting tool operation history and operational environment risks. Similarly, the existing fleet maintenance planning process relies heavily on manual analysis and lacks consideration for environmental factors. These limitations underscore the need for data-driven and risk-based solutions.

The chapter introduces two innovative solutions to address these challenges. The first solution, the tool compatibility index, optimizes tool maintenance scheduling by incorporating operational environment risks and part criticality. This solution enhances objectivity and efficiency in tool selection and upgrades decision-making through practical implementation in Microsoft Power BI and field testing.

The second solution leverages the tool compatibility index and operational environment risk estimation methods to provide cost-effective fleet maintenance planning decisions. This solution offers a more comprehensive and objective approach to decision-making, eliminating the pitfalls of experience-based judgments. Use case diagrams and application scenarios validate the effectiveness of this solution in providing valuable insights to decision-makers.

Furthermore, the detailed framework of risk-based maintenance decision-making for D&M tools, presented in the end of this chapter, builds upon earlier concepts and research outcomes. It offers a clearer global view of the decision-making processes, emphasizing data utilization and flow compared with the generic framework in Chapter II.

In summary, this chapter presents innovative solutions for tool maintenance scheduling and fleet maintenance planning, and a comprehensive framework that integrates data, risk estimation, and decision-making. These solutions are poised to transform the industry's decision-making processes, ensuring optimized maintenance decision-making for D&M tools.

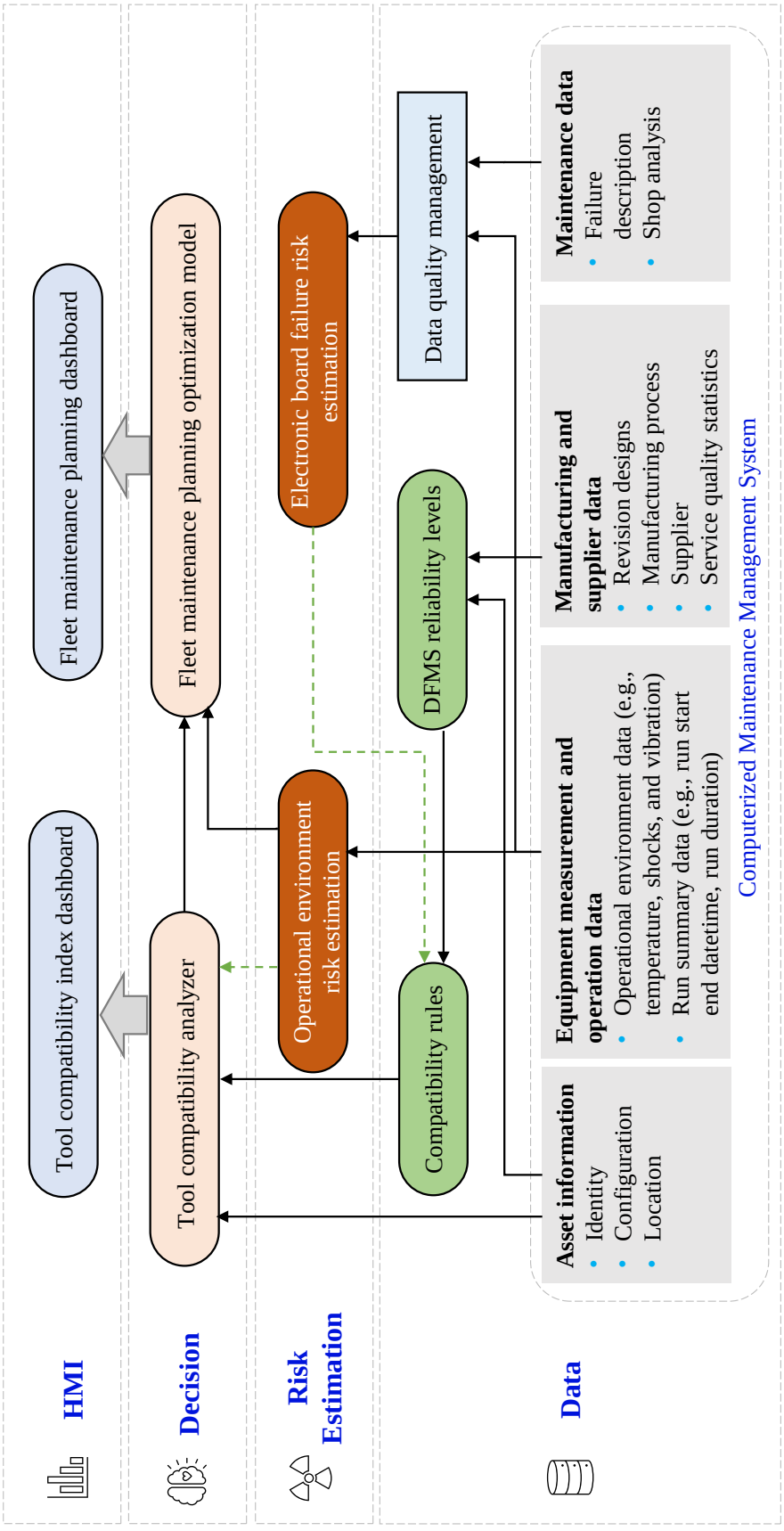


Figure V.11 – Detailed overall risk-based decision-making framework for D&M tools

Chapter VI
Conclusion and future works

VI.1 Conclusion

D&M tools are integral to drilling operations in the oil and gas industry. These tools perform essential functions, including communicating with the surface, measuring formation properties, and controlling drilling direction. Given the importance of these functions, proactive maintenance measures must be implemented to address potential failures before they occur. Proactive maintenance strategies include approaches such as PM, CBM, PdM, and RBM.

PM can lead to unnecessary maintenance activities and avoidable costs. CBM and PdM strategies involve continuous monitoring of the tool and initiating maintenance actions when early failures are detected, or significant degradation of subsystems or components is predicted. CBM and PdM are expected to prevent major failures, maintain tool integrity, and reduce downtime. The effectiveness of CBM and PdM relies on a comprehensive understanding of failure mechanisms, observable failure symptoms, and sufficient condition monitoring information to construct effective PHM models. However, D&M tools are intricate systems housing numerous electronic components, posing challenges in comprehending failure mechanisms, and observing symptoms. Moreover, for most electronic boards in D&M tools, only operational environment data (e.g., temperature, shock, and vibration) are measured, and these measurements are taken at the tool level rather than at the electronic assembly (e.g., electronic board) or component level. Additionally, D&M tools are deployed in diverse missions across different locations, introducing variations in operational environments.

Considering the limitations of these strategies, this study delves into RBM for D&M tools in the oil and gas industry. RBM is not intended to replace existing strategies but to complement them. We introduce a specialized risk-based decision-making framework explicitly addressing drilling and measurement tool management in the oil and gas industry. The framework comprises four layers: data, risk estimation, decision, and HMI. Based on this framework, we explored various aspects, including data quality management, risk estimation of electronic boards (key components of D&M tools), operational environment risk estimation, risk-based maintenance scheduling for individual tools, and risk-based maintenance planning for tool fleets.

VI.1.1 Contributions

The aim of this thesis is to comprehensively explore risk-based maintenance decision-making approaches for D&M tools, providing a fresh perspective on how RBM can improve the current maintenance workflow for D&M tools in the oil and gas industry. The contributions of this study can be summarized as follows:

1. **Risk-Based Decision Framework:** We introduced a novel risk-based maintenance decision-making framework consisting of four interconnected layers: Data, Risk Estimation, Decision, and HMI). This is a versatile framework that can be applied not only to D&M tools but also to other systems with similar characteristics as D&M tools.
2. **Data Quality Management:** An innovative framework for data quality management was presented, emphasizing the pivotal role of data in equipment risk estimation. This framework comprises five components: data development, data quality assessment, data quality requirement, data quality improvement, and risk estimation model development. It provides valuable guidance for data practitioners aiming to manage data quality for industrial equipment risk estimation.
3. **Electronic Board Risk Estimation:** We proposed a groundbreaking method using HMM for electronic board risk estimation, providing a holistic understanding of board health. This approach addresses the shortcomings of traditional models and aligns with real-world scenarios.
4. **Operational Environment Risk Estimation:** A systematic method for assessing risks of operational environments during drilling operations in specific geographical locations was introduced. This method offers valuable insights into the likelihood of encountering adverse conditions, a critical aspect of RBM.
5. **Tool Compatibility Index:** We introduced the tool compatibility index, a novel solution that improves tool selection and upgrade decisions by incorporating operational environment risks and part criticality. This method enhances objectivity and efficiency in tool selection.
6. **Risk-Based Fleet Maintenance Planning:** A comprehensive solution for fleet maintenance planning was presented, leveraging the tool compatibility index and operational environment risk estimates to provide cost-effective decisions. This approach eliminates the pitfalls of experience-based judgments.

In light of these contributions, we have published or submitted a series of papers as follows:

1. J. Kang, C. Varnier, A. Mosallam, N. Zerhouni, F. B. Youssef and N. Shen, "Risk Level Estimation for Electronics Boards in Drilling and Measurement Tools Based on the Hidden Markov Model," 2022 Prognostics and Health Management Conference (PHM-2022 London), London, United Kingdom, 2022, pp. 495-500. (Best Paper Reward)
2. J. Kang, C. Varnier, A. Mosallam, N. Zerhouni, F. B. Youssef and N. Shen, "Tool Compatibility Index: Indicator Enables Improved Tool Selection for Well Construction". PHM Society European Conference, Turin, Italy, 2022, pp. 239–244.

3. J. Kang, N. Shen, F. B. Youssef, A. Mosallam, C. Varnier and N. Zerhouni, "Integrating Fleet Compatibility and Environmental Risk in Downhole Tool Investment Planning," 2023 Prognostics and Health Management Conference (PHM), Paris, France, 2023, pp. 112-116.
4. J. Kang, Z. A. Masry, C. Varnier, A. Mosallam, N. Zerhouni, “Data Management Framework for Risk Estimate of Electronic Boards in Drilling and Measurement Tools". IFAC World Congress 2023, 2023, Yokohama, JAPAN.
5. J. Kang, Z. A. Masry, C. Varnier, A. Mosallam, N. Zerhouni, “A data quality management framework for equipment risk estimate: application to oil and gas industry". Engineering Applications of Artificial Intelligence (2nd round review).

In addition to the publications mentioned, I have collaborated with colleagues on other papers related to maintenance in the oil and gas industry, as listed below.

1. K. Sobczak-Oramus, A. Mosallam, C. Basci, and J. Kang. "Data-Driven Fault Detection for Transmitter in Logging-While-Drilling Tool". PHM Society European Conference, Turin, Italy, 2022, pp. 458–465.
2. G. Cao, X. Wang, J. Kang, Q. Liang, Y. Xie, P. Williams, G. Kartoatmodjo, R. Henson, F. Florez, J. Cui, A. Fendt, X. Mu, L. Wang, W. Yang, B. Zhang, J. He, and H. Zheng, "A Novel Approach to Building an Autonomous System that Enables Predictive Maintenance for Compression Systems". ADIPEC, Abu Dhabi, UAE, 2022.
3. V. Gupta, J. Kang, A. Mosallam, N. Shen, F. B. Youssef and L. Laval, "Automatic Fault Detection for Resistivity Systems in Logging-While-Drilling Tools," 2023 Prognostics and Health Management Conference (PHM), Paris, France, 2023, pp. 128-132.
4. A. Mosallam, J. Kang, F. B. Youssef, L. Laval and J. Fulton, "Data-Driven Fault Diagnostics for Neutron Generator Systems in Multifunction Logging-While-Drilling Service," 2023 Prognostics and Health Management Conference (PHM), Paris, France, 2023, pp. 171-176.
5. A. O. Gonzalez, W. C. Tartari, J. Kang, F. B. Youssef and A. Mosallam, "Predictive Maintenance Planning for Mechanical Components of Bottomhole Assemblies," 2023 Prognostics and Health Management Conference (PHM), Paris, France, 2023, pp. 194-199.
6. A. Mosallam, F. B. Youssef, K. Sobczak-Oramus, J. Kang, V. Gupta, N. Shen, L. Laval, "Data-Driven Degradation Modeling Approach for Neutron Generators in Multifunction Logging-While-Drilling Service," 2023 Prognostics and Health Management Conference (PHM), Paris, France, 2023, pp. 101-106.

Additionally, the successful industrialization and deployment of the two risk-based maintenance solutions within our company’s platform have generated substantial business value. When applied to the top seven types of D&M tools used by the company, these solutions have the potential to yield savings of up to 300

million dollars compared to the full upgrade approach. This significant cost reduction underscores the practical impact and relevance of this PhD research in the oil and gas industry.

VI.1.2 Limitations

While this thesis has made significant strides in advancing the understanding and application of risk-based maintenance decision-making for D&M tools in the oil and gas industry, it is essential to acknowledge its limitations. The limitations are summarized as follows:

1. **Specificity to D&M Tools:** The proposed risk estimation and risk-based maintenance decision-making approaches are tailored specifically for D&M tools and cannot be easily extended to other equipment types due to the extensive domain knowledge required for feature extraction, threshold definition, and compatibility rules creation.
2. **Data Requirements:** The electronic board failure risk estimation method requires full-life cycle data, which can be challenging to obtain due to missing data. The weights for feature aggregation are experience-based and lack theoretical support.

VI.2 Future works

To build on the foundations laid in this thesis and further advance the field of risk-based maintenance decision-making for D&M tools in the oil and gas industry, several avenues for future research are proposed:

1. **Incorporating economic considerations**: One key direction for improvement is the integration of cost and gain considerations related to data quality enhancement within the decision-making model. Understanding the economic implications of data quality improvements can lead to the development of more cost-effective data quality management strategies. This approach can help organizations make informed decisions about allocating resources for data improvement.
2. **Assessing label quality**: The quality of data labels, especially those related to failure modes and causes, significantly impacts the accuracy of risk estimation. Investigating the influence of label quality on model performance is essential. In practical engineering asset management scenarios, obtaining accurate labels can be a complex and resource-intensive process, especially when dealing with complex equipment. Addressing this challenge is crucial for more realistic data quality management. One promising research direction for addressing this problem could involve leveraging large language models to analyze the failure description and analysis text in maintenance data. This approach has the potential to significantly reduce the burden on SMEs by automating or assisting in the labeling of data.

3. **Diverse data formats**: Real-world data quality management often involves dealing with diverse data formats, including time series, tables, text, images, and more. Expanding our understanding of methodologies and frameworks for assessing data quality across these heterogeneous formats is vital. Each format poses unique challenges, and future research should aim to develop versatile quality assessment techniques to handle this diversity effectively.
4. **Deep learning-based data augmentation:** In addition to the three data preprocessing approaches mentioned—outlier detection, missing value handling, and data deduplication—deep learning methods, such as generative adversarial network [150], can be employed to generate synthetic data and in turn improve the data quality. These methods can further enhance data quality and expand the repertoire of data quality improvement strategies.
5. **Feature extraction improvement**: Feature extraction remains a critical aspect of the risk estimation model, profoundly impacting model performance. When a risk estimation model performs poorly, prioritizing the refinement of feature extraction methods should be considered. Improving the model by extracting more informative features can lead to more meaningful insights. Consequently, focusing on feature extraction enhancement may take precedence over data quality management efforts in some cases.
6. **Hybrid Model for Electronic Board Risk Estimation:** Throughout the product life cycle of electronic boards in D&M tools, much data remains untouched to build the risk estimation model. For example, mechanical modeling results (e.g., finite element model simulation results), verification and validation (V&V) testing results, and accelerated aging testing results for specific electronic components during design and failure investigation phases. Investigating how to integrate these laboratory or computer simulation-derived data with tool measurement data obtained during actual drilling operations to develop hybrid models that combine physics-based and data-driven approaches for electronic board risk estimation is a worthwhile avenue for exploration.
7. **Optimizing Weighted Models:** In the HMM-based electronic board failure risk estimation model, the weights for feature aggregation are defined based on experiences, which lacks theoretical support. The weights can be further optimized based on accelerated factors learned from highly accelerated life tests.
8. **Considering More Constraints in Fleet Maintenance Planning:** Considering tool transfer between geo units and part swapping between tools, incorporating client importance and budget limitations as additional constraints in the optimization problem for enhanced fleet maintenance planning.
9. **Hybrid Maintenance Strategy:** Extending the maintenance decision-making framework to include condition-based and predictive maintenance techniques to improve drilling operations' reliability further and minimize downtime.

The future research outlined in these areas aims to address the limitations and further enhance the practicality and effectiveness of the risk-based maintenance decision-making framework developed in this thesis. By delving into these directions, we can contribute to continuously improving maintenance for D&M tools in the oil and gas industry, ultimately leading to more reliable and efficient drilling operations.

Bibliography

[1] H. Ritchie, M. Roser, and P. Rosado, "Energy Mix," *Our World in Data*, oct 2022. [Online]. Available: https://ourworldindata.org/energy-mix

[2] SafetyCulture, "A Simple Guide to Oil and Gas Production," 2023. [Online]. Available: https://safetyculture.com/topics/oil-and-gas-production/

[3] L. Saidi, J. Ben Ali, M. Benbouzid, and E. Bechhofer, "An integrated wind turbine failures prognostic approach implementing kalman smoother with confidence bounds," *Applied Acoustics*, vol. 138, pp. 199–208, 2018.

[4] IEEE Reliability Society, "IEEE Standard Framework for Prognostics and Health Management of Electronic Systems," *IEEE Std 1856-2017*, pp. 1–31, 2017.

[5] Schlumberger. (2023) BHA. [Online]. Available: https://glossary.oilfield.slb.com/en/terms/b/bha

[6] Schlumberger. (2023) measurements-while-drilling. [Online]. Available: https://glossary.oilfield.slb.com/en/terms/m/measurements-while-drilling

[7] A. Mosallam, L. Laval, F. B. Youssef, J. Fulton, and D. Viassolo, "Data-driven fault detection for neutron generator subsystem in multifunction logging-while-drilling service," in *PHM Society European Conference*, 2018.

[8] L. Kirschbaum, D. Roman, G. Singh, J. Bruns, V. Robu, and D. Flynn, "AI-driven maintenance support for downhole tools and electronics operated in dynamic drilling environments," *IEEE Access*, vol. 8, pp. 78 683–78 701, 2020.

[9] Wikipedia, "Maintenance," Aug. 2023. [Online]. Available: https://en.wikipedia.org/w/index.php?title=Maintenance&oldid=1170548522#cite_note-1037C-5

[10] R. Abbassi, E. Arzaghi, M. Yazdi, V. Aryai, V. Garaniya, and P. Rahnamayiezekavat, "Risk-based and predictive maintenance planning of engineering infrastructure: Existing quantitative techniques and future directions," *Process Safety and Environmental Protection*, vol. 165, pp. 776–790, 2022.

[11] P. Nunes, J. Santos, and E. Rocha, "Challenges in predictive maintenance – a review," *CIRP Journal of Manufacturing Science and Technology*, vol. 40, pp. 53–67, 2023.

[12] P. Gackowiec, "General overview of maintenance strategies – concepts and approaches," *Multidisciplinary Aspects of Production Engineering*, vol. 2, no. 1, pp. 126–139, 2019.

[13] A. Patil, G. Soni, A. Prakash, and K. Karwasra, "Maintenance strategy selection: a comprehensive review of current paradigms and solution approaches," *International Journal of Quality & Reliability Management*, vol. 39, no. 3, pp. 675–703, 2021.

[14] N. Arunraj and J. Maiti, "Risk-based maintenance—techniques and applications," *Journal of Hazardous Materials*, vol. 142, no. 3, pp. 653–661, 2007.

[15] M. Ilangkumaran and S. Kumanan, "Application of hybrid vikor model in selection of maintenance strategy," *International Journal of Information Systems and Supply Chain Management*, vol. 5, pp. 59–81, 2012.

[16] R. K. Mobley, "1 - impact of maintenance," in *An Introduction to Predictive Maintenance (Second Edition)*, 2nd ed., ser. Plant Engineering, R. K. Mobley, Ed. Burlington: Butterworth-Heinemann, 2002, pp. 1–22.

[17] A. Mosallam, "Remaining useful life estimation of critical components based on bayesian approaches." Ph.D. dissertation, Université de Franche-Comté, 12 2014.

[18] P. Tchakoua, R. Wamkeue, M. Ouhrouche, F. Slaoui-Hasnaoui, T. A. Tameghe, and G. Ekemb, "Wind turbine condition monitoring: State-of-the-art review, new trends, and future challenges," *Energies*, vol. 7, no. 4, pp. 2595–2630, 2014.

[19] Q. Hao, Y. Xue, W. Shen, B. Jones, and J. Zhu, *A Decision Support System for Integrating Corrective Maintenance, Preventive Maintenance, and Condition-Based Maintenance*. ASCE Press, 2010, pp. 470–479.

[20] L. Pinciroli, P. Baraldi, and E. Zio, "Maintenance optimization in industry 4.0," *Reliability Engineering & System Safety*, vol. 234, p. 109204, 2023.

[21] J. Lee, H.-A. Kao, and S. Yang, "Service innovation and smart analytics for industry 4.0 and big data environment," *Procedia CIRP*, vol. 16, pp. 3–8, 2014, product Services Systems and Value Creation. Proceedings of the 6th CIRP Conference on Industrial Product-Service Systems.

[22] E. Quatrini, F. Costantino, G. Di Gravio, and R. Patriarca, "Condition-based maintenance—an extensive literature review," *Machines*, vol. 8, no. 2, 2020.

[23] R. Ahmad and S. Kamaruddin, "An overview of time-based and condition-based maintenance in industrial application," *Computers and Industrial Engineering*, vol. 63, no. 1, pp. 135–149, 2012.

[24] L. Pinciroli, P. Baraldi, A. Shokry, E. Zio, R. Seraoui, and C. Mai, "A semi-supervised method for the characterization of degradation of nuclear power plants steam generators," *Progress in Nuclear Energy*, vol. 131, p. 103580, 2021.

[25] 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, p. 108063, 2022.

[26] M. Achouch, M. Dimitrova, K. Ziane, S. Sattarpanah Karganroudi, R. Dhouib, H. Ibrahim, and M. Adda, "On predictive maintenance in industry 4.0: Overview, models, and challenges," *Applied Sciences*, vol. 12, no. 16, 2022.

[27] O. O. Ersöz, A. F. İnal, A. Aktepe, A. K. Türker, and S. Ersöz, "A systematic literature review of the predictive maintenance from transportation systems aspect," *Sustainability*, vol. 14, no. 21, 2022.

[28] F. I. Khan and M. Haddara, "Risk-based maintenance (RBM): A new approach for process plant inspection and maintenance," *Process Safety Progress*, vol. 23, no. 4, pp. 252–265, 2004.

[29] F. I. Khan and M. M. Haddara, "Risk-based maintenance (rbm): a quantitative approach for maintenance/inspection scheduling and planning," *Journal of Loss Prevention in the Process Industries*, vol. 16, no. 6, pp. 561–573, 2003.

[30] A. Mosallam, J. Kang, F. B. Youssef, L. Laval, and J. Fulton, "Data-Driven Fault Diagnostics for Neutron Generator Systems in Multifunction Logging-While-Drilling Service," in *2023 Prognostics and Health Management Conference (PHM)*, May 2023, pp. 171–176.

[31] A. Mosallam, F. B. Youssef, K. Sobczak-Oramus, J. Kang, V. Gupta, N. Shen, and L. Laval, "Data-Driven Degradation Modeling Approach for Neutron Generators in Multifunction Logging-While-Drilling Service," in *2023 Prognostics and Health Management Conference (PHM)*, May 2023, pp. 101–106.

[32] K. Javed, R. Gouriveau, N. Zerhouni, and P. Nectoux, "Enabling health monitoring approach based on vibration data for accurate prognostics," *IEEE Transactions on Industrial Electronics*, vol. 62, 05 2014.

[33] S. Das, R. Hall, S. Herzog, G. Harrison, M. Bodkin, and L. Martin, "Essential steps in prognostic health management," in *2011 IEEE Conference on Prognostics and Health Management*, 2011, pp. 1–9.

[34] L. Biggio and I. Kastanis, "Prognostics and Health Management of Industrial Assets: Current Progress and Road Ahead," *Frontiers in Artificial Intelligence*, vol. 3, 2020.

[35] CLACE, University of Maryland. (2023) Prognostics and Health Management. [Online]. Available: https://calce.umd.edu/prognostics-and-health-management

[36] F. Calabrese, A. Regattieri, L. Botti, and F. G. Galizia, "Prognostic health management of production systems. new proposed approach and experimental evidences," *Procedia Manufacturing*, vol. 39, pp. 260–269, 2019.

[37] T. Sutharssan, S. Stoyanov, C. Bailey, and C. Yin, "Prognostic and health management for engineering systems: a review of the data-driven approach and algorithms," *The Journal of Engineering*, vol. 2015, no. 7, pp. 215–222, 2015.

[38] M. Brahimi, K. Medjaher, M. Leouatni, and N. Zerhouni, "Prognostics and Health Management for an Overhead Contact Line System - A Review," *International Journal of Prognostics and Health Management*, vol. 8, no. 3, 2017.

[39] K. Medjaher, H. Skima, and N. Zerhouni, "Condition assessment and fault prognostics of microelectromechanical systems," *Microelectronics Reliability*, vol. 54, no. 1, pp. 143–151, 2014.

[40] G. W. Vogl, B. A. Weiss, and M. A. Donmez, "Standards for Prognostics and Health Management (PHM) Techniques within Manufacturing Operations," in *Annual Conference of the PHM Society*, vol. 6, 2014.

[41] ISO 18436, "Condition monitoring and diagnostics of machines," International Organization for Standardization, Geneva, Switherland, Standard, 2018.

[42] Society for Risk Analysis. (2023) Society for Risk Analysis Glossary. [Online]. Available: https://www.sra.org/wp-content/uploads/2020/04/SRA-Glossary-FINAL.pdf

[43] ISO 31000:2018, "Risk management — Guidelines," International Organization for Standardization, Geneva, CH, Standard, December 2018.

[44] M. J. L. Peters, D. P. M. Symmons, D. McCarey, B. A. C. Dijkmans, P. Nicola, T. K. Kvien, I. B. McInnes, H. Haentzschel, M. A. Gonzalez-Gay, S. Provan, A. Semb, P. Sidiropoulos, G. Kitas, Y. M. Smulders, M. Soubrier, Z. Szekanecz, N. Sattar, and M. T. Nurmohamed, "Eular evidence-based recommendations for cardiovascular risk management in patients with rheumatoid arthritis and other forms of inflammatory arthritis," *Annals of the Rheumatic Diseases*, vol. 69, no. 2, pp. 325–331, 2010.

[45] A. D. Association, "9. Cardiovascular Disease and Risk Management: Standards of Medical Care in Diabetes—2018," *Diabetes Care*, vol. 41, no. Supplement 1, pp. S86–S104, 11 2017.

[46] C. S. Tang, "Perspectives in supply chain risk management," *International Journal of Production Economics*, vol. 103, no. 2, p. 451 – 488, 2006.

[47] M. M. Cornett, J. J. McNutt, P. E. Strahan, and H. Tehranian, "Liquidity risk management and credit supply in the financial crisis," *Journal of Financial Economics*, vol. 101, no. 2, p. 297 – 312, 2011.

[48] L. Marchi, M. Borga, E. Preciso, and E. Gaume, "Characterisation of selected extreme flash floods in europe and implications for flood risk management," *Journal of Hydrology*, vol. 394, no. 1, pp. 118–133, 2010, flash Floods: Observations and Analysis of Hydrometeorological Controls.

[49] E. Koks, B. Jongman, T. Husby, and W. Botzen, "Combining hazard, exposure and social vulnerability to provide lessons for flood risk management," *Environmental Science & Policy*, vol. 47, pp. 42–52, 2015.

[50] V. Gupta, J. Kang, A. Mosallam, N. Shen, F. B. Youssef, and L. Laval, "Automatic Fault Detection for Resistivity Systems in Logging-While-Drilling Tools," in *2023 Prognostics and Health Management Conference (PHM)*, Jun. 2023, pp. 128–132.

[51] K. Sobczak-Oramus, A. Mosallam, C. Basci, and J. Kang, "Data-Driven Fault Detection for Transmitter in Logging-While-Drilling Tool," in *PHM Society European Conference*, vol. 7, Jun. 2022, pp. 458–465.

[52] J. Kang, C. Varnier, A. Mosallam, N. Zerhouni, F. B. Youssef, and N. Shen, "Risk level estimation for electronics boards in drilling and measurement tools based on the hidden markov model," in *2022 Prognostics and Health Management Conference (PHM-2022 London)*, 2022, pp. 495–500.

[53] ——, "Tool compatibility index: indicator enables improved tool selection for well construction," in *PHM Society European Conference*, Jun. 2022, pp. 239–244.

[54] J. Kang, N. Shen, F. B. Youssef, A. Mosallam, C. Varnier, and N. Zerhouni, "Integrating fleet compatibility and environmental risk in downhole tool investment planning," in *2023 Prognostics and Health Management Conference (PHM)*, 2023, pp. 112–116.

[55] V. T. Tran, H. Thom Pham, B.-S. Yang, and T. Tien Nguyen, "Machine performance degradation assessment and remaining useful life prediction using proportional hazard model and support vector machine," *Mechanical Systems and Signal Processing*, vol. 32, pp. 320–330, 2012.

[56] R. S. Velmurugan and T. Dhingra, "Maintenance strategy selection and its impact in maintenance function," *International Journal of Operations & Production Management*, vol. 35, no. 12, pp. 1622–1661, 2015.

[57] S. O. Duffuaa and A. Raouf, *Maintenance Planning and Scheduling*. Cham: Springer International Publishing, 2015, pp. 155–186.

[58] P. Martínez-Galán Fernández, A. J. Guillén López, A. C. Márquez, J. F. Gomez Fernández, and J. A. Marcos, "Dynamic risk assessment for cbm-based adaptation of maintenance planning," *Reliability Engineering & System Safety*, vol. 223, p. 108359, 2022.

[59] R. K. Mazumder, A. M. Salman, and Y. Li, "Failure risk analysis of pipelines using data-driven machine learning algorithms," *Structural Safety*, vol. 89, p. 102047, 2021.

[60] T. S. Betz, K. El-Rayes, M. N. Grussing, K. E. Landers, and L. B. Bartels, "Parametric estimation of equipment failure risk with machine learning and constrained optimization," *Journal of Performance of Constructed Facilities*, vol. 37, no. 1, 2023.

[61] Z. Wang, H. Du, L. Tao, and S. A. Javed, "Risk assessment in machine learning enhanced failure mode and effects analysis," *Data Technologies and Applications*, vol. ahead-of-print, no. ahead-of-print, Jan. 2023.

[62] M. Kläs and A. M. Vollmer, "Uncertainty in machine learning applications: A practice-driven classification of uncertainty," in *Computer Safety, Reliability, and Security*, B. Gallina, A. Skavhaug, E. Schoitsch, and F. Bitsch, Eds. Cham: Springer International Publishing, 2018, pp. 431–438.

[63] G. Press, "Cleaning big data: Most time-consuming, least enjoyable data science task, survey says," Sep 2023. [Online]. Available: https://www.forbes.com/sites/gilpress/2016/03/23/data-preparation-most-time-consuming-least-enjoyable-data-science-task-survey-says/?sh=687198b46f63

[64] N. Gupta, S. Mujumdar, H. Patel, S. Masuda, N. Panwar, S. Bandyopadhyay, S. Mehta, S. Guttula, S. Afzal, R. Sharma Mittal, and V. Munigala, "Data quality for machine learning tasks," in *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining*, ser. KDD '21. New York, NY, USA: Association for Computing Machinery, 2021, p. 4040–4041.

[65] R. Y. Wang and D. M. Strong, "Beyond accuracy: what data quality means to data consumers," *Journal of Management Information Systems*, vol. 12, no. 4, pp. 5–33, 1996.

[66] A. Karkouch, H. Mousannif, H. Al Moatassime, and T. Noel, "Data quality in internet of things: A state-of-the-art survey," *Journal of Network and Computer Applications*, vol. 73, pp. 57–81, 2016.

[67] K. Q. Wang, S. R. Tong, L. Roucoules, and B. Eynard, "Analysis of data quality and information quality problems in digital manufacturing," in *2008 4th IEEE International Conference on Management of Innovation and Technology*, 2008, pp. 439–443.

[68] Y. Chen, F. Zhu, and J. Lee, "Data quality evaluation and improvement for prognostic modeling using visual assessment based data partitioning method," *Computers in Industry*, vol. 64, no. 3, pp. 214–225, Apr. 2013.

[69] "Data quality — part 2: Vocabulary," International Organization for Standardization, Geneva, CH, Standard, December 2022.

[70] ISO/IEC 25012:2008, "Software engineering — software product quality requirements and evaluation (square) — data quality model," International Organization for Standardization, Geneva, CH, Standard, December 2008.

[71] IBM, "What is data quality?" 2024. [Online]. Available: https://www.ibm.com/topics/data-quality

[72] Z. Wang, Y. Fu, C. Song, W. Ge, L. Qiao, and H. Zhang, "A data quality improvement method based on the greedy algorithm," in *Machine Learning and Intelligent Communications*, X. B. Zhai, B. Chen, and K. Zhu, Eds. Cham: Springer International Publishing, 2019, pp. 256–266.

[73] F. Gualo, M. Rodriguez, J. Verdugo, I. Caballero, and M. Piattini, "Data quality certification using iso/iec 25012: Industrial experiences," *Journal of Systems and Software*, vol. 176, p. 110938, 2021.

[74] "Data quality - part 8: Information and data quality: Concepts and measuring," International Organization for Standardization, Geneva, CH, Standard, December 2015.

[75] A. Klein and W. Lehner, "Representing data quality in sensor data streaming environments," *Journal of Data and Information Quality*, vol. 1, no. 2, sep 2009.

[76] T. Rekatsinas, X. L. Dong, L. Getoor, and D. Srivastava, "Finding quality in quantity: The challenge of discovering valuable sources for integration," in *7th Biennial Conference on Innovative Data Systems Research (CIDR '15)*. Citeseer, 2015.

[77] A. Purnomoadi, I. M. Sari, J. Anna Maria, D. B. Fiddiansyah, N. E. Saputro, and M. Sofan Hadi, "A method to quantify data quality in asset health indices model," in *2023 4th International Conference on High Voltage Engineering and Power Systems (ICHVEPS)*, 2023, pp. 16–20.

[78] M. Díaz Iturry, S. N. Alves-Souza, M. Ito, and S. A. da Silva, "Data quality in health records: A literature review," in *2021 16th Iberian Conference on Information Systems and Technologies (CISTI)*, 2021, pp. 1–6.

[79] D. Xu, Z. Zhang, and J. Shi, "A data quality assessment and control method in multiple products manufacturing process," in *2022 5th International Conference on Data Science and Information Technology (DSIT)*, 2022, pp. 1–5.

[80] J. Merino, I. Caballero, B. Rivas, M. Serrano, and M. Piattini, "A data quality in use model for big data," *Future Generation Computer Systems*, vol. 63, pp. 123–130, 2016, modeling and Management for Big Data Analytics and Visualization. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0167739X15003817

[81] R. Y. Wang, "A product perspective on total data quality management," *Communications of the ACM*, vol. 41, no. 2, pp. 58–65, 1998.

[82] C. Batini, C. Cappiello, C. Francalanci, and A. Maurino, "Methodologies for data quality assessment and improvement," *ACM Comput. Surv.*, vol. 41, no. 3, jul 2009.

[83] B. Behkamal, M. Kahani, E. Bagheri, and Z. Jeremic, "A metrics-driven approach for quality assessment of linked open data," *Journal of Theoretical and Applied Electronic Commerce Research*, vol. 9, no. 2, pp. 64–79, 2014.

[84] M. Kaiser, M. Klier, and B. Heinrich, "How to measure data quality? - a metric-based approach," in *Proceedings of the 28th International Conferenceon Information Systems (ICIS)*, 2007.

[85] L. Ehrlinger and W. Wöß, "A Survey of Data Quality Measurement and Monitoring Tools," *Frontiers in Big Data*, vol. 5, 2022.

[86] N. Omri, Z. Al Masry, N. Mairot, S. Giampiccolo, and N. Zerhouni, "Towards an adapted phm approach: Data quality requirements methodology for fault detection applications," *Computers in Industry*, vol. 127, p. 103414, 2021.

[87] X. Jia, M. Zhao, Y. Di, Q. Yang, and J. Lee, "Assessment of data suitability for machine prognosis using maximum mean discrepancy," *IEEE Transactions on Industrial Electronics*, vol. 65, no. 7, pp. 5872–5881, Jul. 2018.

[88] X. Jia, D.-Y. Ji, T. Minami, and J. Lee, "Data quality and usability assessment methodology for prognostics and health management: A systematic framework," *IFAC-PapersOnLine*, vol. 55, no. 19, pp. 55–60, 2022, 5th IFAC Workshop on Advanced Maintenance Engineering, Services and Technologies AMEST 2022.

[89] S. Lukens, D. Rousis, T. Baer, M. Lujan, and M. Smith, "A Data Quality Scorecard for Assessing the Suitability of Asset Condition Data for Prognostics Modeling," *Annual Conference of the PHM Society*, vol. 14, no. 1, Oct. 2022, number: 1.

[90] S. Lukens, M. Naik, K. Saetia, and X. Hu, "Best Practices Framework for Improving Maintenance Data Quality to Enable Asset Performance Analytics," *Annual Conference of the PHM Society*, vol. 11, no. 1, Sep. 2019, number: 1.

[91] M. Madhikermi, S. Kubler, J. Robert, A. Buda, and K. Främling, "Data quality assessment of maintenance reporting procedures," *Expert Systems with Applications*, vol. 63, pp. 145–164, 2016.

[92] S. Koziel, P. Hilber, P. Westerlund, and E. Shayesteh, "Investments in data quality: Evaluating impacts of faulty data on asset management in power systems," *Applied Energy*, vol. 281, p. 116057, 2021.

[93] N. Makhoul, "Review of data quality indicators and metrics, and suggestions for indicators and metrics for structural health monitoring," *Advances in Bridge Engineering*, vol. 3, no. 1, p. 17, Nov. 2022.

[94] IBM, "What is a CMMS? Definition, how it works and benefits | IBM," 2023. [Online]. Available: https://www.ibm.com/topics/what-is-a-cmms

[95] D. Zha, Z. P. Bhat, K.-H. Lai, F. Yang, Z. Jiang, S. Zhong, and X. Hu, "Data-centric Artificial Intelligence: A Survey," Jun. 2023.

[96] A. Mosallam, S. Byttner, M. Svensson, and T. Rögnvaldsson, "Nonlinear relation mining for maintenance prediction," in *2011 Aerospace Conference*. IEEE, 2011, pp. 1–9.

[97] V. Atamuradov, K. Medjaher, P. Dersin, B. Lamoureux, and N. Zerhouni, "Prognostics and Health Management for Maintenance Practitioners - Review, Implementation and Tools Evaluation," *International Journal of Prognostics and Health Management*, vol. 8, no. 3, 2017.

[98] O. Fink, Q. Wang, M. Svensén, P. Dersin, W.-J. Lee, and M. Ducoffe, "Potential, challenges and future directions for deep learning in prognostics and health management applications," *Engineering Applications of Artificial Intelligence*, vol. 92, p. 103678, 2020.

[99] J. Cai, J. Luo, S. Wang, and S. Yang, "Feature selection in machine learning: A new perspective," *Neurocomputing*, vol. 300, pp. 70–79, 2018.

[100] S. Meisenbacher, M. Turowski, K. Phipps, M. Rätz, D. Müller, V. Hagenmeyer, and R. Mikut, "Review of automated time series forecasting pipelines," *WIREs Data Mining and Knowledge Discovery*, vol. 12, no. 6, p. e1475, 2022.

[101] B. Heinrich, D. Hristova, M. Klier, A. Schiller, and M. Szubartowicz, "Requirements for data quality metrics," *J. Data and Information Quality*, vol. 9, no. 2, jan 2018.

[102] A. Zimek and E. Schubert, "Outlier detection," in *Encyclopedia of Database Systems*, L. Liu and M. T. Özsu, Eds. New York: Springer New York, 2017, pp. 1–5.

[103] A. Smiti, "A critical overview of outlier detection methods," *Computer Science Review*, vol. 38, p. 100306, 2020.

[104] V. Chandola, A. Banerjee, and V. Kumar, "Anomaly detection: A survey," *ACM Comput. Surv.*, vol. 41, no. 3, jul 2009.

[105] W.-C. Lin and C.-F. Tsai, "Missing value imputation: a review and analysis of the literature (2006–2017)," *Artificial Intelligence Review*, vol. 53, no. 2, pp. 1487–1509, Feb. 2020.

[106] M. K. Hasan, M. A. Alam, S. Roy, A. Dutta, M. T. Jawad, and S. Das, "Missing value imputation affects the performance of machine learning: A review and analysis of the literature (2010–2021)," *Informatics in Medicine Unlocked*, vol. 27, p. 100799, 2021.

[107] M. G. Pecht and M. Kang, *Prognostics and Health Management of Electronics: Fundamentals, Machine Learning, and the Internet of Things*. Wiley-IEEE Press, 2019.

[108] C. Bhargava, P. K. Sharma, M. Senthilkumar, S. Padmanaban, V. K. Ramachandaramurthy, Z. Leonowicz, F. Blaabjerg, and M. Mitolo, "Review of Health Prognostics and Condition Monitoring of Electronic Components," *IEEE Access*, vol. 8, pp. 75 163–75 183, 2020.

[109] D. Bhat, S. Muench, and M. Roellig, "Application of machine learning algorithms in prognostics and health monitoring of electronic systems: A review," *e-Prime - Advances in Electrical Engineering, Electronics and Energy*, vol. 4, p. 100166, 2023.

[110] Q. Sun, L. Yang, H. Li, and G. Sun, "Rul prediction for aecs of power electronic systems based on machine learning and error compensation," *Journal of Intelligent and Fuzzy Systems*, vol. 44, no. 5, p. 7407 – 7417, 2023.

[111] J. Zhao, Y. Zhou, Q. Zhu, Y. Song, Y. Liu, and H. Luo, "A remaining useful life prediction method of aluminum electrolytic capacitor based on wiener process and similarity measurement," *Microelectronics Reliability*, vol. 142, p. 114928, 2023.

[112] A. F. Shahraki, S. Al-Dahidi, A. R. Taleqani, and O. P. Yadav, "Using lstm neural network to predict remaining useful life of electrolytic capacitors in dynamic operating conditions," *Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability*, vol. 237, no. 1, pp. 16–28, 2023.

[113] Q. Lei, D. Luchun, Z. Xiangyu, Z. Wuyu, and S. Hong, "Online correction method of igbt lifetime evaluation based on bonding wire failure monitoring," *IET Power Electronics*, vol. 16, no. 3, p. 347 – 356, 2023.

[114] X. Wang, Z. Zhou, S. He, H. Jia, and Y. Huang, "Igbt performance degradation feature construction and real-time prediction based on machine learning," in *The Proceedings of the 17th Annual Conference of China Electrotechnical Society*, Q. Yang, J. Li, K. Xie, and J. Hu, Eds. Singapore: Springer Nature Singapore, 2023, pp. 743–750.

[115] D. Xiao, C. Qin, J. Ge, P. Xia, Y. Huang, and C. Liu, "Self-attention-based adaptive remaining useful life prediction for igbt with monte carlo dropout," *Knowledge-Based Systems*, vol. 239, p. 107902, 2022.

[116] P. Lall, R. Vaidya, V. More, and K. Goebel, "Assessment of accrued damage and remaining useful life in leadfree electronics subjected to multiple thermal environments of thermal aging and thermal cycling," *IEEE Transactions on Components, Packaging and Manufacturing Technology*, vol. 2, no. 4, pp. 634–649, 2012.

[117] W. Hu, Y. Li, Y. Sun, and A. Mosleh, "A model of bga thermal fatigue life prediction considering load sequence effects," *Materials*, vol. 9, no. 10, p. 860, 2016.

[118] P. Lall, P. Choudhary, S. Gupte, and J. C. Suhling, "Health monitoring for damage initiation and progression during mechanical shock in electronic assemblies," *IEEE Transactions on Components and Packaging Technologies*, vol. 31, no. 1, pp. 173–183, 2008.

[119] J. Gu, D. Barker, and M. Pecht, "Health monitoring and prognostics of electronics subject to vibration load conditions," *IEEE Sensors Journal*, vol. 9, no. 11, pp. 1479–1485, 2009.

[120] S. Tiku, A. Veneruso, R. Etchells, and M. Pecht, "Risk factors in oil and gas well electronics compared to other electronic industries," *Oil & Gas Science and Technology – Rev. IFP*, vol. 60, pp. 721–730, 07 2005.

[121] S. Bhatnagar, M. L. Cassou, Z. A. Masry, and A. Mosallam, "Data-Driven Fault Detection Method for Electronic Boards in Intelligent Remote Dual-Valve System," in *PHM Society European Conference*, Jun. 2021, pp. 1–7.

[122] A. A. Kale, K. Carter-Journet, T. A. Falgout, L. Heuermann-Kuehn, and D. Zurcher, "A Probabilistic Approach for Reliability and Life Prediction of Electronics in Drilling and Evaluation Tools," in *Annual Conference of the PHM Society*, vol. 6, 2014.

[123] S. Zhan and J. Zhao, "Advanced prognostic technique for improving the drilling performance of downhole tools," in *2015 First International Conference on Reliability Systems Engineering (ICRSE)*, 2015, pp. 1–9.

[124] S. ZHAN and I. Ahmad, "Real-time prognostic on downhole printed circuit board assembly of measurement-while-drilling/logging-while-drilling," WO Patent WO2 012 047 860A2, Apr., 2012.

[125] D. R. Garvey, J. Baumann, J. Lehr, and J. W. Hines, "Pattern Recognition-Based Remaining Useful Life Estimation of Bottomhole Assembly Tools," in *SPE/IADC Drilling Conference and Exhibition*. OnePetro, Mar. 2009.

[126] R. Beiras, "Chapter 6 - environmental risk assessment of pharmaceutical and personal care products in estuarine and coastal waters," in *Pharmaceuticals in Marine and Coastal Environments*, ser. Estuarine and Coastal Sciences Series, J. C. Durán-Álvarez and B. Jiménez-Cisneros, Eds. Elsevier, 2021, vol. 1, pp. 195–252.

[127] J. Tarazona and M. Ramos-Peralonso, "Environmental risk assessment, terrestrial," in *Encyclopedia of Toxicology (Third Edition)*, 3rd ed., P. Wexler, Ed. Oxford: Academic Press, 2014, pp. 411–414.

[128] "Chapter 5 - the use of mathematical modelling in lake and reservoir management," in *Lake and Reservoir Management*, ser. Developments in Water Science, S. Jørgensen, H. Löffler, W. Rast, and M. Straškraba, Eds. Elsevier, 2005, vol. 54, pp. 243–314.

[129] L. E. Baum and T. Petrie, "Statistical Inference for Probabilistic Functions of Finite State Markov Chains," *The Annals of Mathematical Statistics*, vol. 37, no. 6, pp. 1554–1563, 1966.

[130] A. Poritz, "Linear predictive hidden markov models and the speech signal," in *ICASSP '82. IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 7, 1982, pp. 1291–1294.

[131] L. Rabiner, "A tutorial on hidden markov models and selected applications in speech recognition," *Proceedings of the IEEE*, vol. 77, no. 2, pp. 257–286, 1989.

[132] M. K. Mustafa, T. Allen, and K. Appiah, "A comparative review of dynamic neural networks and hidden markov model methods for mobile on-device speech recognition," *Neural Computing and Applications*, vol. 31, pp. 891–899, 2019.

[133] R. Kumar, L. S. Videla, S. SivaKumar, A. G. Gupta, and D. Haritha, "Murmured speech recognition using hidden markov model," in *2020 7th International Conference on Smart Structures and Systems (ICSSS)*. IEEE, 2020, pp. 1–5.

[134] M. Soruri, S. Hamid Zahiri, and J. Sadri, "A new approach of training hidden markov model by pso algorithm for gene sequence modeling," in *2013 First Iranian Conference on Pattern Recognition and Image Analysis (PRIA)*, 2013, pp. 1–4.

[135] S. R. Eddy, "What is a hidden Markov model?" *Nature Biotechnology*, vol. 22, no. 10, Oct. 2004.

[136] M. Momenzadeh, M. Sehhati, and H. Rabbani, "Using hidden markov model to predict recurrence of breast cancer based on sequential patterns in gene expression profiles," *Journal of Biomedical Informatics*, vol. 111, p. 103570, 2020.

[137] T. M. Porter and M. Hajibabaei, "Profile hidden markov model sequence analysis can help remove putative pseudogenes from dna barcoding and metabarcoding datasets," *BMC bioinformatics*, vol. 22, no. 1, p. 256, 2021.

[138] A. Anandika, S. P. Mishra, and M. Das, "Review on usage of hidden markov model in natural language processing," in *Intelligent and Cloud Computing*, D. Mishra, R. Buyya, P. Mohapatra, and S. Patnaik, Eds. Singapore: Springer Singapore, 2021, pp. 415–423.

[139] T. Almutiri and F. Nadeem, "Markov models applications in natural language processing: a survey," *Int. J. Inf. Technol. Comput. Sci*, vol. 2, pp. 1–16, 2022.

[140] J. Yousif, "Hidden markov model tagger for applications based arabic text: A review," *Journal of Computation and Applied Sciences IJOCAAS*, vol. 7, no. 1, 2019.

[141] M. D. Drovo, M. Chowdhury, S. I. Uday, and A. K. Das, "Named entity recognition in bengali text using merged hidden markov model and rule base approach," in *2019 7th International Conference on Smart Computing & Communications (ICSCC)*. IEEE, 2019, pp. 1–5.

[142] B. T. McClintock, R. Langrock, O. Gimenez, E. Cam, D. L. Borchers, R. Glennie, and T. A. Patterson, "Uncovering ecological state dynamics with hidden markov models," *Ecology letters*, vol. 23, no. 12, pp. 1878–1903, 2020.

[143] W. Zucchini, I. L. MacDonald, and R. Langrock, *Hidden Markov Models for Time Series: An Introduction Using R*, 2nd ed. CRC Press, 2016.

[144] A. Viterbi, "Error bounds for convolutional codes and an asymptotically optimum decoding algorithm," *IEEE Transactions on Information Theory*, vol. 13, no. 2, pp. 260–269, 1967.

[145] I. Oliver, *Markov Processes for Stochastic Modeling*, 2nd ed. Elsevier, 2013.

[146] I. Visser and M. Speekenbrink, "depmixS4: An R Package for Hidden Markov Models," *Journal of Statistical Software*, vol. 36, pp. 1–21, Aug. 2010.

[147] Visual Paradigm Community Circle, "1. Use Case Diagram," May 2018. [Online]. Available: https://circle.visual-paradigm.com/docs/uml-and-sysml/use-case-diagram/

[148] B. Kolman and R. E. Beck, "4 - Integer Programming," in *Elementary Linear Programming with Applications (Second Edition)*, B. Kolman and R. E. Beck, Eds. San Diego: Academic Press, Jan. 1995, pp. 249–293.

[149] C. Ryan, "Evolutionary Algorithms and Metaheuristics," in *Encyclopedia of Physical Science and Technology (Third Edition)*, R. A. Meyers, Ed. New York: Academic Press, Jan. 2003, pp. 673–685.

[150] X. Li, V. Metsis, H. Wang, and A. H. H. Ngu, "TTS-GAN: A transformer-based time-series generative adversarial network," in *Artificial Intelligence in Medicine*, M. Michalowski, S. S. R. Abidi, and S. Abidi, Eds. Cham: Springer International Publishing, 2022, pp. 133–143.

SPIM

Titre : Approches décisionnelles de la maintenance basées sur le risque pour les outils de forage et de mesure

Mots clefs: estimation des risques, prise de décision, pétrole et gaz, outils de forage, cartes électroniques

Résumé : Les outils de forage et de mesure (F&M) sont essentiels pour assurer l'efficacité et le succès des opérations de forage dans l'industrie pétrolière et gazière. La maintenance proactive est essentielle pour maintenir la fiabilité de ces outils. Cette étude se concentre sur les stratégies de maintenance basées sur le risque.

Le cœur de cette thèse est l'introduction d'un cadre décisionnel complet de maintenance basé sur le risque pour les outils de F&M. Ce cadre est composé de quatre couches interconnectées. Ce cadre se compose de quatre couches interconnectées : Données, Estimation du risque, Décision et Interface homme-machine. Sur la base de ce cadre, nous présentons des solutions innovantes, notamment un cadre de gestion de la qualité des données pour la couche Données, une méthode d'estimation des risques liés à la carte électronique et une estimation des risques liés à l'environnement opérationnel dans la couche Estimation des risques, un indice de compatibilité des outils pour une sélection plus éclairée des outils et des décisions de mise à niveau, et une solution de planification de la maintenance de la flotte basée sur les risques dans la couche Décision. Ces solutions sont conçues pour surmonter les limites des pratiques de maintenance existantes, en fournissant des processus de prise de décision de maintenance plus objectifs et mieux informés. En outre, ces solutions ont été déployées avec succès, ce qui a permis à l'entreprise de réaliser des économies substantielles.

En résumé, cette étude marque une avancée significative dans la maintenance basée sur le risque pour les outils de F&M, offrant des solutions pratiques et efficaces pour améliorer la prise de décision en matière de maintenance dans l'industrie pétrolière et gazière.

Title : Risk-based maintenance decision-making approaches for drilling and measurement tools

Keywords : risk estimate, decision-making, oil and gas, drilling tools, electronic boards

Abstract : Drilling and Measurement (D&M) tools are essential in ensuring the efficiency and success of drilling operations in the oil and gas industry. Proactive maintenance is essential to maintain the reliability of these tools. This study focuses on risk-based maintenance strategies.

The core of this thesis is introducing a comprehensive risk-based maintenance decision framework for D&M tools. This framework consists of four interconnected layers : Data, Risk Estimation, Decision, and Human-Machine Interface. Building upon this framework, we present innovative solutions, including a data quality management framework for the Data layer, an electronic board risk estimation method and an operational environment risk estimation within the Risk Estimation layer, a tool compatibility index for more informed tool selection and upgrade decisions, and a risk-based fleet maintenance planning solution within the Decision layer. These solutions are designed to overcome the limitations of existing maintenance practices, providing more objective and informed maintenance decision-making processes. Furthermore, these solutions have been successfully deployed, resulting in substantial cost savings for the company.

In summary, this study marks a significant advancement in risk-based maintenance for D&M tools, offering practical and effective solutions to enhance maintenance decision-making in the oil and gas industry.

UBFC UNIVERSITÉ BOURGOGNE FRANCHE-COMTÉ

Université Bourgogne Franche-Comté
32, avenue de l'Observatoire
25000 Besançon