Development of an advanced artificial intelligent reliability analysis tool to enhance ship operations and maintenance activities

By

OIKONOMOU STYLIANOS

Submitted for the degree of Master of Philosophy

Department of Naval Architecture, Ocean and Marine Engineering

University of Strathclyde

April 2021

Declaration of Authenticity and Author's Rights

This thesis is the result of the author's original research. It has been composed by the author and has not been previously submitted for examination which has led to the award of a degree.

Signed:

Date: April 2021

Acknowledgements

I would like to express my gratitude and sincere thanks to my supervisor Dr Iraklis Lazakis of the Department of Naval Architecture, Ocean and Marine Engineering at the University of Strathclyde. He always keeps providing me with the opportunity to research this study and for supervising my work throughout my thesis. In addition, I would like to thank AP&A, the industrial partner to my Knowledge Transfer Partnership (KTP), for giving me the opportunity to develop this thesis while working with them. Furthermore, this partnership received partial financial support from KTP programme. KTP aims to help businesses to improve their competitiveness and productivity through the better use of knowledge, technology and skills that reside within the UK knowledge base. This successful Knowledge Transfer Partnership project, funded by UK Research and Innovation through Innovate UK, is part of the government's Industrial Strategy.

Stylianos Oikonomou

Nomenclature

y_0: Reliability of the machinery when the first deterioration of the machinery is detected

y_n: Reliability of the machinery when the n^{th} deterioration of the machinery is detected

x_0: Time of first deterioration of the machinery

x_n: Time of n^{th} deterioration of the machinery

t_{F_0}: First prediction of the time of failure of the machinery when first deterioration of the machinery is detected

t_{F_n}: Prediction of the time of failure of the machinery at the n^{th} time deterioration has been detected

a: universal factor

Table of Contents

1. Introduction & Project Definition

1.1 Chapter Outline

In this Chapter, the impact of shipping in the world economy will be showcased as well as the importance of maintenance for the operation of a ship in the ever-increasing globalized market. In addition, the structure of the MPhil will be discussed together with a summary of each chapter. Finally, the research question that this MPhil is trying to address will be presented as well as the main aim and objectives to answer this question.

1.2 Shipping and Maintenance

The globalization of the world economy has led to increased competition in the maritime transport market.

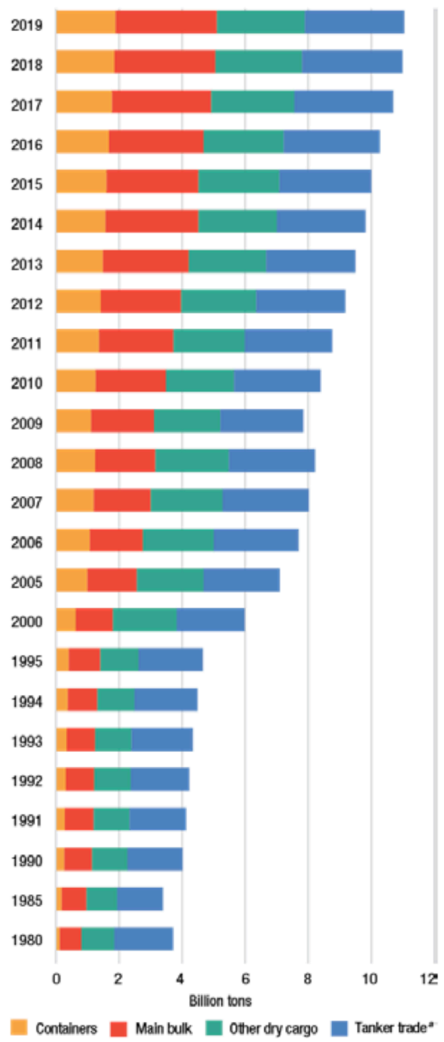


Figure 1: Development of international maritime trade by cargo type, selected years (Billion tons loaded) (UNCTAD, 2020)

A look in Figure 1 (UNCTAD, 2020) at how the various market segments have evolved since 1990 shows that growth in maritime trade over the past three decades has been sustained by bullish trends in containerized trade volumes starting in the 2000s, coinciding with the wave of hyper-globalization. In addition, shipping is a vital facilitator of the world trade since it carries most of the value of world trade and it boosts the world GDP (Gross Domestic Product). Figure 2 shows some data concerning the economic value of shipping (OECS, 2016). From Figure 2, IMO (International Marine Organization) estimated that 90% of the world's trade is carried by ships while shipping also helped increasing the world's GDP by an impressive 73% in a two-decade period. Furthermore, seaborne trade has increased 112% over the same period and contributed 3.5 trillion euro in the EU's trade value with the rest of the world in 2015.

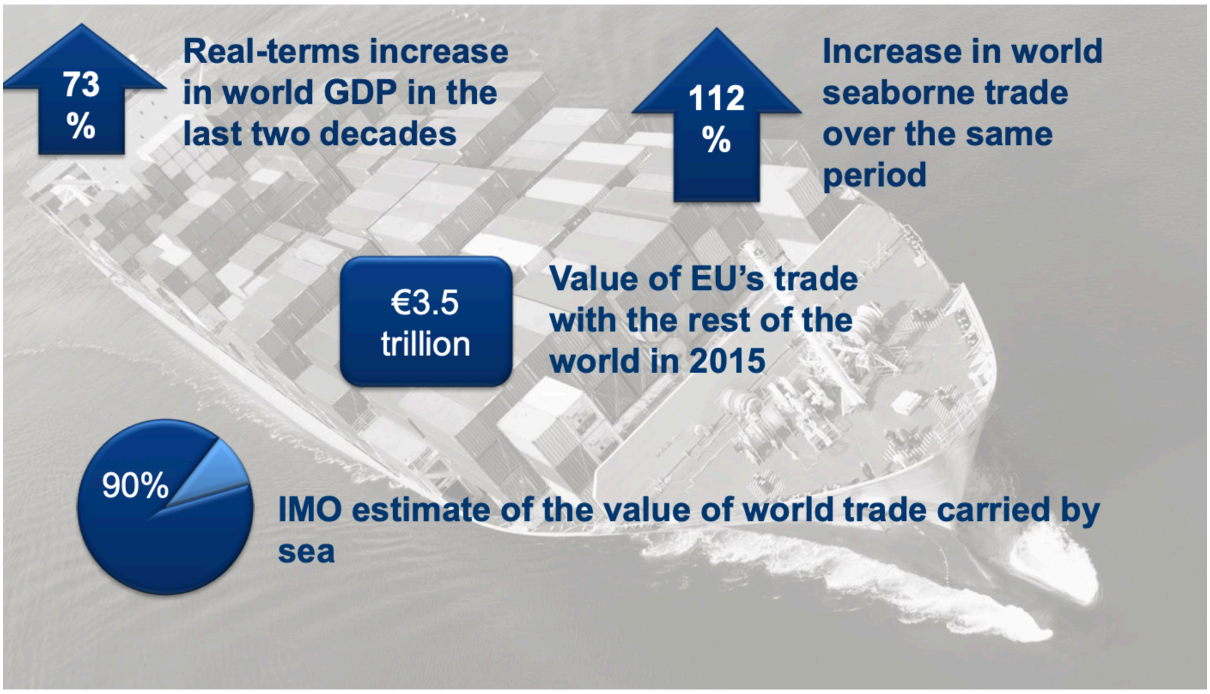


Figure 2: The economic value of shipping (OECS, 2016)

Since shipping is the moving power behind the world's economy it is very important to assure that ships are operated correctly and efficiently. It is not by chance that in recent years, the management cost of the life cycle of a ship has become the main management tool of shipping companies. One of the key cost items that determine the competitiveness of the vessel is the cost of maintenance and repair. Maintenance and repair costs are part of a bigger family of operational costs called fixed operational costs that together with the running operational costs are the total operational costs of a ship. Fixed operational costs are those whose objective is to maintain the ship in seaworthy conditions to offer transport services, even though the vessel can be laid-up, in contrast of the running operational that depend on every specific voyage and, especially, on the ports of call, distance crossed, cargo handling operations, the possible need of passing some channels and other (Gerardo, 2012). Concerning the fixed operational costs, from OpCost 2018 database (Moore, 2018) in Figure 3 it can be seen the five fixed operational costs for different types of ships.

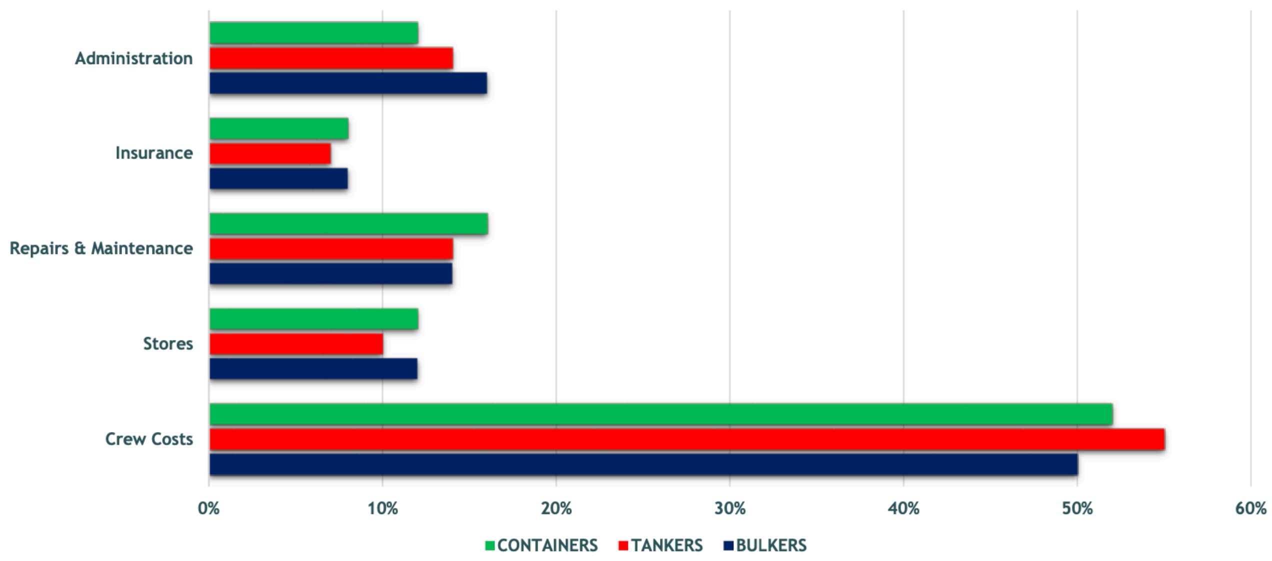


Figure 3: Fixed Operational Costs Structure for different types of ships (Moore, 2018)

OpCost is a unique ship operating costs benchmarking tool, and it includes all major sectors and currently covers operating costs for 26 vessel types. Most of the costs in Figure 3 are not easy to reduce, for example even though the biggest portion of the operational costs is the crew cost, it is not a flexible cost (Bayer, 2016). Insurance also has a very limited chance to negotiate for cost. Except company staff salaries, all other administration aspects (office rent, transportation, and communication, stationary) are in the procurement and purchase. The stores and spare & repair cost together account for more than 20% of the operational expenses. Stopford (2008) also agrees that maintenance amounts to 16% of the shipping company fixed operating costs. Other studies have shown that the contribution of maintenance can be as big as 40% of the overall fixed operational costs (Alhouli et al., 2009) thus highlighting even more how important is maintenance in the everyday operation of ships. These data together with the fact that maintenance costs are more flexible to reduce (Dekker, 1996), is the reason why maintenance has become an area of active research for shipping companies.

However, maritime industry faces unique challenges in the execution of scheduled and unscheduled maintenance. Ships spend significant periods at sea far from the logistically supported areas (Rustenburg et al., 2001) and a ship at sea is isolated from onshore repair and maintenance facilities, and, if a failure occurs during the passage, the required replacement parts may not be available on board. The rising cost of ship operation is a problem, since the failure of a vital piece of equipment can be very expensive and may put the safety of the whole ship at risk. Added to this is the cost of out of service (off hire), when the ship is in the downtime period because of the failure. Additionally, it may cause increases of the cargo expenses and may pose danger to the environment as well (Rothblum, 2000). Finally, in many cases serious accidents and interruptions are related to poor maintenance, or, more specifically, to poor maintenance planning (Reason and Hobbs, 2003)

To sum up, shipping is a major contributor in the world's economy and therefore the operation of the ships is of utmost importance. Ship maintenance is a crucial factor in a ship's performance and, in turn, can heavily affect the shipping company's revenue. There should be an optimal maintenance level for all the equipment on ships and therefore a balance between maintenance cost and over-maintenance. Over-maintenance is the term used for the case of excessive use of maintenance activities more than required level. Thus, establishing an appropriate framework with which to measure maintenance performance to have an optimal level and to plan maintenance policy has vital importance for the shipping organization.

1.3 Structure of MPhil

This thesis is consisting of seven chapters and the flowchart in Figure 4 depicts the names of these chapters and how they are connected.

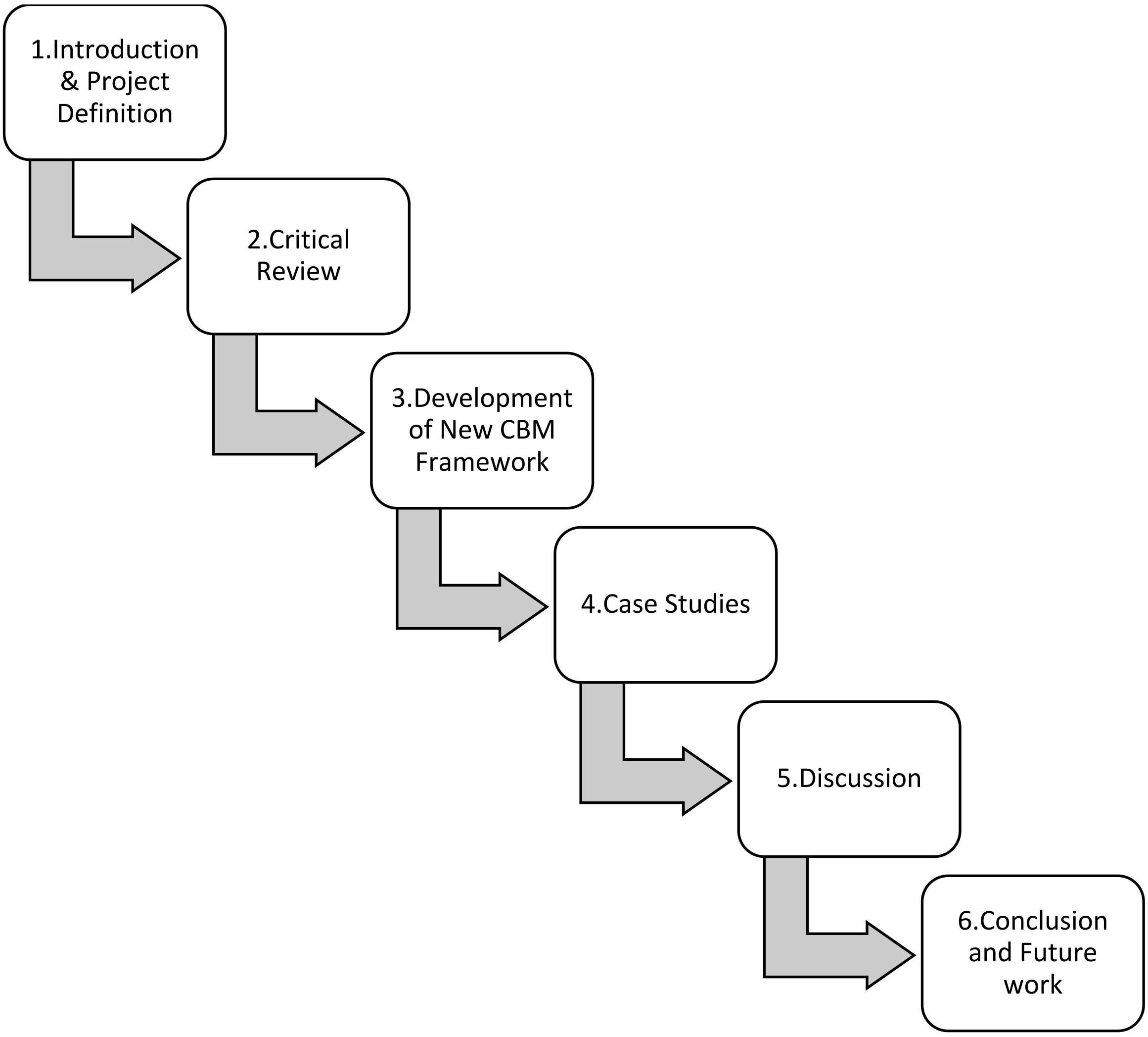


Figure 4: Structure of MPhil

Following Figure 4 and in more detail: Chapter one contains an introduction that gives a brief description of the impact of shipping for the global economy and the importance of maintenance for the operation of ships. In addition, the structure of the MPhil is shown as well as the details of each chapter. Chapter two gives the project definition which includes the research question as well as the list of objectives for answering it.

Chapter three gives a thorough critical review of all the different maintenance approaches and their advantages and disadvantages. More focus will be given in describing the predictive maintenance strategy and specifically the data-driven approach for condition-based maintenance. All the algorithms and techniques used for data-driven approaches are also being shown. Finally, a review of the database types used previously for data-driven solutions in shipping is given.

Chapter four introduces the purposed condition-based maintenance framework which consists of an object-oriented database, a semi-supervised machine learning algorithm and a novel failure model. Firstly, the scheme of the object-oriented database is shown together with an example for better understanding of how values are stored and utilised. Then the semi-supervised machine learning algorithm is explained and how it detects possible deterioration. Finally, the novel failure model is analysed mathematically and the way it predicts failure through the use of temporal detection of deteriorations of a machinery.

Chapter five contains all the case studies for the evaluation of different parts of the proposed maintenance framework as well as a real-world application of the whole system. In more detail, the capability of the machine learning algorithm to understand when a machinery has deteriorated is evaluated by using a database containing data of degradation of a turbine and compressor of a propulsion system of a Navy ship. Next, the new failure model is evaluated with a dataset of failure data for bearings of a motor. Finally, the last case studies are evaluating the whole condition-based maintenance framework with the prediction of the time of failure of three diesel generators of a tanker. Chapter six provides a conclusion to the MPhil as well as the future steps.

Chapter six is a discussion on the entire MPhil thesis, summarising what it has been done overall and for answering the main aim and objectives of the thesis

Chapter seven gives the conclusion to this thesis and proposes future work to be done in the research field.

1.4 Research Question

As it was shown in chapter one, shipping is the “engine” of global economy and maintenance is an important cost of the operation of the ships, thus making it a very important area for research. Therefore, now is more necessary than ever to answer to the question:

Why there is a need for condition-based maintenance framework in the shipping industry?

The research question will be answered through the main aim and objectives of this MPhil. Concerning the former, the main aim is to develop a novel condition-based maintenance framework for application within the marine industry and suggest solutions for the best ship maintenance strategy. The objectives that will facilitate the main aim are:

- Perform a thorough critical review on various methods of fault detection, reliability tools, maintenance processes and databases used and consider their advantages and disadvantages.
- Develop a novel condition-based maintenance framework, including: A condition monitoring database, a machine learning method to detect deterioration and a new modelling theory for system failure.
- Application and validation of the selected machine learning algorithm through the case of a Navy ship propulsion plant system and examining the decay of its main compulsory engine consisting of a compressor and a turbine.
- Application and validation of the developed failure modelling theory by using failure data of bearings of an AC motor.
- Application and validation of the condition-based maintenance framework by using real data from a tanker ship to predict the failure of its Diesel Generators (DG).

1.5 Chapter Summary

This Chapter showed that shipping is very important for the global economy and facilitates the development of countries individually (GDP increase). In addition, the operation of these ships is governed by different costs, one of which is maintenance. Even though maintenance is not the highest of the operational costs it is a flexible one and can be reduced with adoption of new technologies and strategies. However, maintenance is a complex problem and in order to be optimized other aspects need to be considered before, like safety. Finally, this thesis contains six chapters named: Introduction, Project Definition, Critical Review, Development of New CBM Framework, Case Studies and Conclusion and Future Work. Finally, the research question was presented as well as the main aim and objectives of this MPhil that will be used to answer it. The main aim is to introduce a novel condition-based maintenance framework for application within the marine industry and suggest solutions for the best ship maintenance strategy. The objectives are ranging from thorough critical review to numerous applications of the new condition-based maintenance framework.

2. Literature and Critical review

2.1 Chapter Outline

In this chapter, a critical review of different maintenance methods used for ship maintenance will be discussed. Then the predictive maintenance will be analysed further and most importantly the Condition Based Maintenance (CBM), approaches for diagnosis and prognosis in CBM and applications in shipping. Finally, the most common types of databases used for CBM will be reviewed.

2.2 Maintenance approaches

In the previous chapters it was shown that maintenance is a very important cost for shipping operation and various methodologies have been used in the past to optimize it and increase its effectiveness. A critical review of these methods will follow. Maintenance approaches in shipping can be divided in three major approaches (Lazakis et al. 2009), namely: Corrective, Predictive and Preventive maintenance.

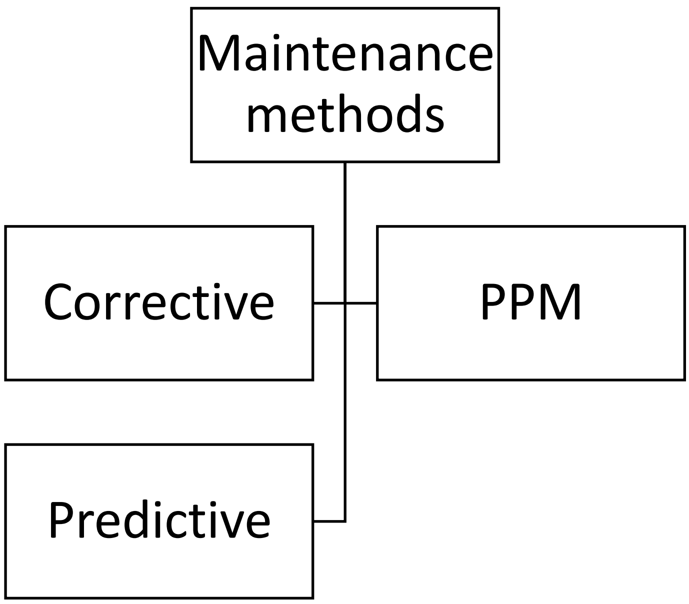


Figure 5: Different Maintenance Approaches (Lazakis et al. 2009)

From Figure 5 it can be seen that the maintenance methods are divided into four categories, namely: Corrective Maintenance, Planned Preventive Maintenance (PPM) and Predictive Maintenance. The next sections will analyse each of these maintenance methods in details.

2.3 Corrective Maintenance

Starting with corrective maintenance, it is one of the three main maintenance methods in shipping also called run-to-failure, hard- time or reactive maintenance (Wilson et al., 2014; Zaal, 2016). Corrective maintenance refers to performing maintenance when a failure or breakdown occurs (Basim et al., 2003). The International Association of Classification Societies (IACS) has identified in Rec 74 (2018) the steps to be followed when corrective maintenance is being used in shipping. Figure 6 shows that steps.

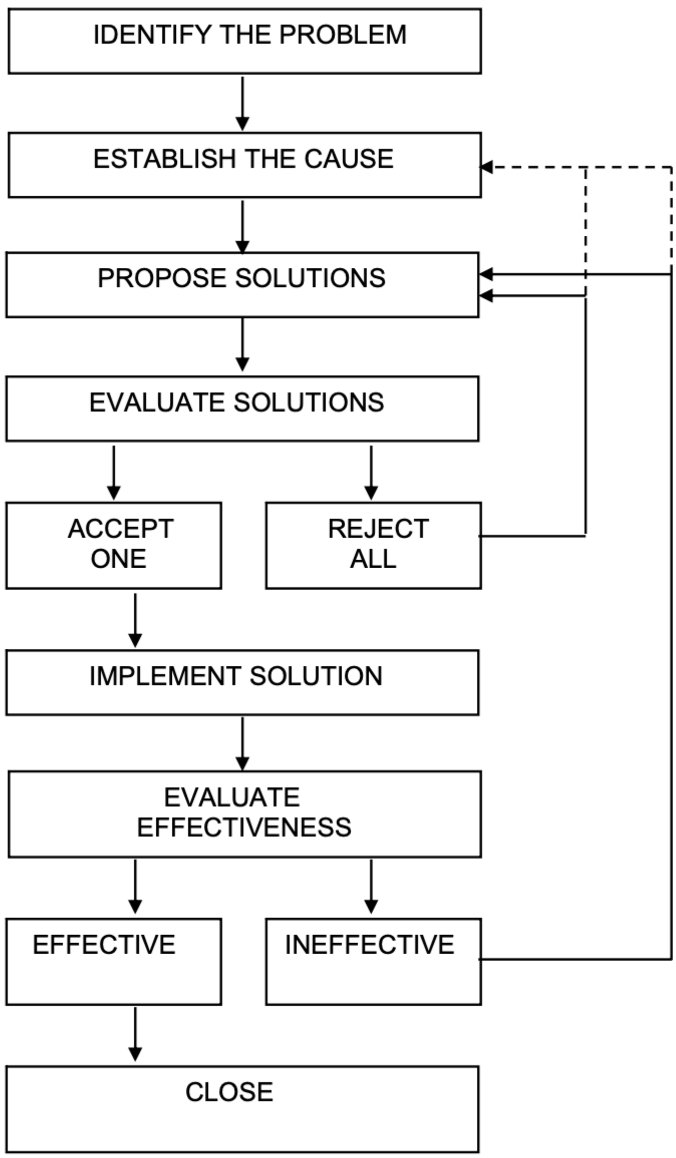


Figure 6: The corrective action plan (IACS, 2018)

According to Figure 6 the first steps are to identify the problem and understand the cause of it. Then solutions are being purposed and evaluated until one is accepted. After that the solution is implemented and its effectiveness is evaluated. If it is effective, then the problem is closed otherwise more solutions need to be proposed and the previous steps are followed again. One benefit of this approach is the simplicity of the planning as seen also in Figure 6. In addition, for purely random failures and low failure costs this may be a cost-effective method (Pintelon et al., 1992).

On the other hand, this maintenance approach is not considered effective mainly because ships are forced to react to problems rather than anticipating and planning in advance (Jimenez et al. 2020). In addition, there is always the possibility that a fault can go unnoticed leading to unpredictable consequences for the crew and the ship. These consequences can vary from expensive damages to dangerous situations involving explosions (Starr et al. 2010). In addition, the uncertainty of the failure requires a breakdown crew to be available on standby together with a large spares inventory which is almost always costly (Lazakis et al. 2009). Finally, corrective maintenance does not focus on the root cause of the equipment failure and therefore the mean time before failure will be much lower than proactive maintenance (Bai et al. 2016).

Recently, various studies on corrective maintenance have been conducted. Nachimuthu et al. (2019) suggested a decision-making model for corrective maintenance of offshore wind turbines considering uncertainties. A mathematical model is proposed to assist wind farm stakeholders in making critical resource- related decisions for corrective maintenance at offshore wind farms (OWFs), considering uncertainties in turbine failure information. The finding shown that the model could lead to more than 80% cost savings in comparison with traditional practises.

Wang et al. (2014) proposed a corrective maintenance scheme for engineering equipment. Firstly, the Failure propagation graph (FPG) is built to represent failure mechanism with the extended FMECA. Then, the FPG updating, and fault diagnosis process are proposed based on the constructed FPG. Finally, a binary decision tree is originally built to determine the failure ascertainment order for corrective maintenance and the proposed scheme is implemented on a boring machine tool and proves to be valid and practical. From the results from the case study the proposed corrective maintenance scheme is easily implemented for engineering applications and proves to be powerful in the corrective maintenance for engineering equipment. Shabrina et al. (2018) extracted knowledge and experience from operators based on knowledge transformation and creates e-Learning content for correct corrective maintenance activities that are fixing the bearings on the machine spindle. Erkoyuncu et al. (2017) presented an approach focusing on building a process for understanding the trade-offs within corrective maintenance activities at the equipment type level. The main benefit of the presented approach is a better prediction of the maintenance system performance and therefore an enhanced cost and availability estimation becomes available. Table 1 sums up all the advantages and disadvantages of corrective maintenance.

Table 1: Advantages and Disadvantages of Corrective Maintenance

Advantages	**Disadvantages**
•Planning is simple – the organisation need only adapt to match the failure rate.	•A component fault may go unnoticed, leading to expensive consequential damage
•Work is not scheduled until it is really needed.	•Dangerous and/or expensive failure consequences should be expected. •No data are available regarding the past, present and possible future state of the machine.
	•A large breakdown crew may need to be available on standby.
	•A large spares inventory is necessary to ensure quick repair.
	•Failures exceeding the capacity of the repair team lead to “fire-fighting”.

•Failure can, and probably will, occur at an inconvenient time, e.g., when the ship is at full load, or while it is starting.

•Does not focus on the root cause of damage

2.4 Planned Preventive Maintenance (PPM)

Planned preventive maintenance generally has the form of scheduled inspections (Figure 7), which are performed to assess whether a component or equipment can still operate satisfactorily or determine the item's deterioration (Mobley, 2002). According to Marquez (2007), the goal of preventive maintenance is to reduce the possibility of failure due to equipment degradation.

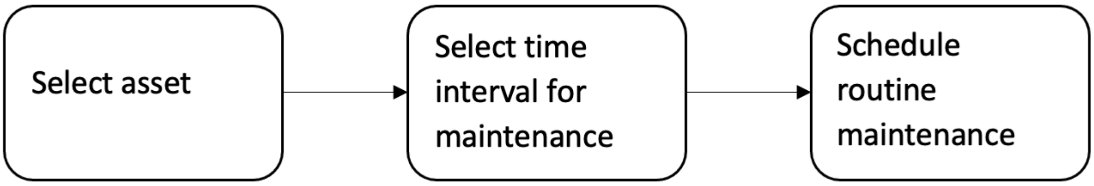


Figure 7: General form of planned preventive maintenance (Mobley, 2002)

Since preventive maintenance tries to determine a series of checks, replacements and/or component revisions with a frequency related to the failure rate, it is effective in overcoming the problems associated with the wearing of components (Dagkinis et al. 2013). In addition, preventive maintenance can reduce the number of spares required since the time of maintenance is known beforehand. Furthermore, more effective use of time is accomplished since fixed time intervals help reduction of downtime (Lazakis et al. 2009).

On the other hand, random failures which are the most common in shipping (see MSDP (1982) and Alen (2001)) cannot be prevented with fixed time maintenance schedule. In addition, for the case of using planned preventive maintenance for helicopter gearboxes, it was found that almost half of the units were removed for overhauling even though they were in a satisfactory operating condition (Lee et al. 2008). This unnecessary strip down and bearing changes may cause problems like a phenomenon called infant mortality and is shown in Figure 8 (Verma et al. 2018). Figure 8 shows the failure rate of a machinery against the time of operation indicating that infant mortality happens when the machinery has just started to operate after maintenance due to human error.

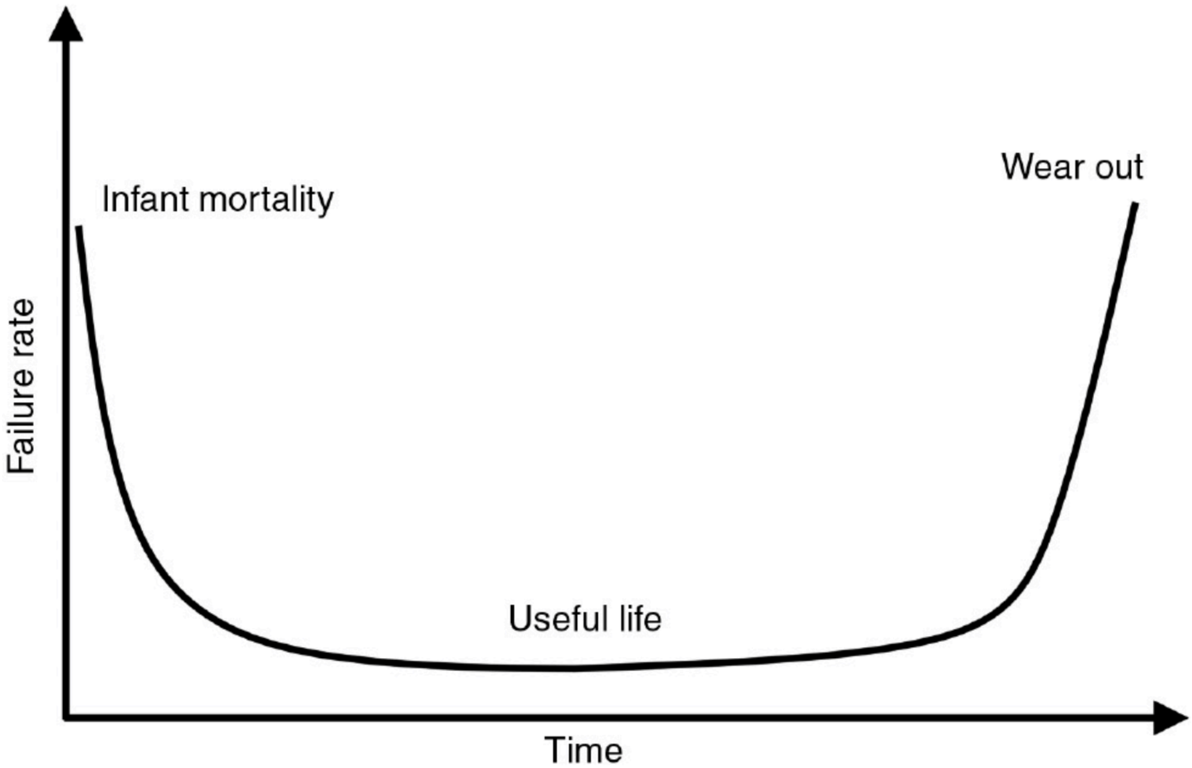


Figure 8: Probability of failure over the life of equipment (Verma et al. 2018)

Concerning the marine industry, preventive maintenance is the most common technique for minimizing the full cost occurred onboard such as the cost of investigation and repair and component downtime. The marine industry is regulated by schedules enforced through classification societies (e.g., ABS, DNV, etc.) and supported by standards recognised by insurance companies (e.g., P&I club). It is normal for a vessel to be taken out of service for a total survey after a predetermined time (DNV, 2015). This will include dry docking of the ship, cleaning and inspecting the hull and attending to all of the below water equipment and because this is a hugely expensive all other internal surveys and refits are scheduled to coincide. Because of the cost and logistics of this process, operators strive for longer and longer periods between dry dockings. However, the internal machinery usually requires maintenance between dry dockings and the challenge is always to keep the vessel in service. Therefore, a considerable amount of redundancy is built into the shipboard systems with reserve or duplicate equipment, spares carried on board or stocked at ports along the ship's itinerary (Starr et al., 2010).

Since planned preventive maintenance was suggested research has suggested different variations of it in order to improve it. A variant of planned preventive maintenance is condition-based preventive maintenance, which was developed by Mann et al. (1995) and uses sensors for monitoring the machine conditions to predict when equipment failures may appear. The paper showed that conventional preventive maintenance policies that have the same time interval time may easily neglect system's reliability. This is because the system deteriorates with increased usage and age (Liao et al. 2009). In addition, Liu et al. (2014) investigates dynamic preventive maintenance policies that the maintenance strategies are performed from a value perspective and component values are modelled as a function of the reliability distribution unlike the traditional cost-based preventive maintenance policies. This maintenance system is implemented when the stability falls below a certain threshold.

Furthermore, Alsyouf et al. (2016) calculated the optimal replacement time for critical components and then reduced the overall cost by suggesting an actual preventive maintenance scheme. The author also developed a control chart to monitor the time between failures based on the calculated failure rate. Huang et al. (2015) takes a two-step approach that simultaneously considers the time and use of a repairable product and considers periodic preventive maintenance to develop a two-dimensional warranty policy for the repairable product. Sheu et al. (2015) proposes optimal preventive maintenance for multi-state systems. This study proposes a recursive approach to efficiently calculate the time-dependent distribution of a multi-state system and finds the optimal PM schedule that minimizes the average cost rate for each type of repair. Table 2 shows the advantages and disadvantages of Planed Preventive Maintenance.

Table 2: Advantages and Disadvantages of Planned Preventive Maintenance

Advantages	**Disadvantages**
•A more effective use of time.	•The machinery may not fail according to a fixed time period (calendar or run hours).

•Spares are only ordered as required.	•Random failures may still occur.
•Improved safety and quality	•The method depends on statistical analysis; in many cases suitable and correct failure data are not present.
•Reduction of downtime	•The asset may not need maintaining, spares and labour are used unnecessarily, and the asset is unavailable during maintenance.
	• Unnecessary strip down and bearing changes may cause problems.

2.5 Predictive Maintenance

In predictive maintenance the regular monitoring of the actual condition, operating efficiency and other indicators of the operating condition of machinery will provide the data required to ensure the maximum interval between repairs and minimize the cost of unscheduled maintenance created by machine failures (Mobley, 2002). According to Fedele (2011) predictive maintenance can be defined as on-condition assessment of assets, employing real time programming by avoiding unnecessary downtime, inspections and reactive failures due to human mistakes. Figure 9 shows the different types of predictive maintenance, namely: Risk Cantered Maintenance (RCM), Risk Based Inspection (RBI) and Computerized Maintenance Management System (CMMS).

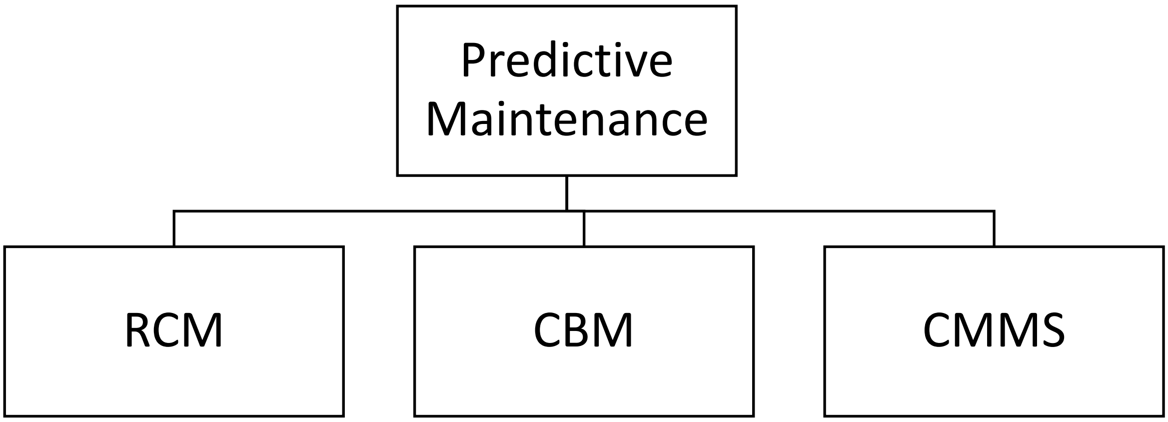


Figure 9: Types of Preventive Maintenance

Predictive maintenance has the strong advantage of being able to plan maintenance according to the actual condition of the machinery through sensor data. This has the benefit of reduction in maintenance costs by 25%–30% (You et al., 2010). Those numbers are backed up also by a survey of 500 plants that have implemented predictive maintenance methods and indicates substantial improvements in reliability, availability and operating costs such as: actual costs normally associated with the maintenance operation reduced by more than 50 per cent and reduced number of catastrophic, unexpected machine failures by an average of 55 per cent (Mobely, 2001). In addition, preventive maintenance minimizes equipment failure rates, improve equipment condition, prolong the life of the equipment and reduce the maintenance costs better than the previous mentioned maintenance techniques (Jezzini et al. 2013).

On the other hand, predictive maintenance requires that the data are collected with proper techniques and interpret correctly (Tomlinson, 2016). In addition, one of the biggest challenges in applying predictive models is the validation of the method (Tinga et al. 2017) because in order to check the accuracy, model predictions will have to be compared with real failures. However, this is only feasible when the complete usage history of the system or component is available, and the service life or the decrease in condition of the component can be assessed on a regular basis. Furthermore, concerning shipping, many systems today contain embedded monitoring systems, but data is often not stored, and the total time history of operation is unavailable. In addition, the sampling frequency is often different from system to system and sometimes can be inadequate- e.g., hourly measurements are useless for rapidly changing diesel engine parameters such as speed or load (Kwon, 2018). Table 3 shows the main advantages and disadvantages of Predictive Maintenance.

Table 3: Advantages of using Predictive Maintenance

Advantages	**Disadvantages**
•Providing actual data for planning the repair activities	•Not adequate frequency of data gathering.
•Early detection of potential failure modes	•The correct interpretation of vast amounts of data has led to several initiatives not being successful.
•Minimisation of unscheduled repairs	•The reliance on technically complex monitoring systems.
•Maximisation of the availability and operability of the system	•Random Failures cannot be detected.

In the bellow sub-sections, the different types of predictive maintenance will be analysed further.

2.5.1 Reliability Centred Maintenance

Reliability Centred Maintenance (RCM) is a highly structured method for maintenance planning, developed firstly for the airline industry (Rausand et al. 2008) and later introduced in different industrial fields (Moubray 2001) such as: transportation (Carretero et al, 2003), the nuclear (Kadak & Matsuo, 2007), defence (MoD, 2006) and the offshore sector (Conachey & Montgomery, 2003). According to the standard IEC 60300-3-11, RCM is an approach to recognize efficient preventive maintenance operations on items and set maintenance work interval according to the specific procedures (IEC, 2010).

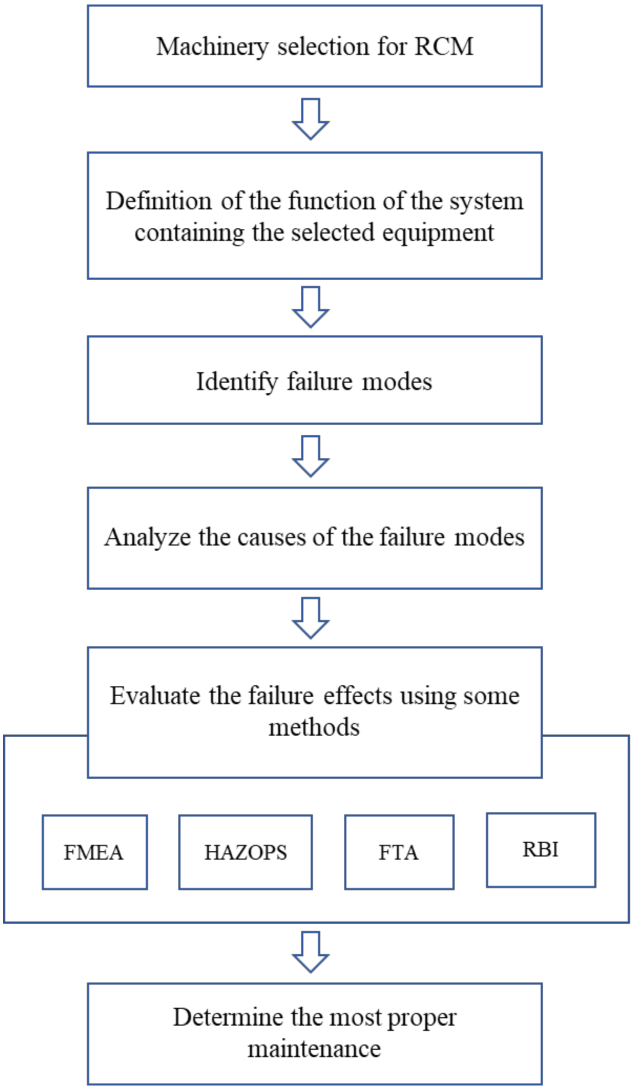


Figure 10: The steps of performing RCM (IEC, 2010)

As a risk-based approach, RBI provides means to evaluate the consequences and likelihood of component failure from specific degradation mechanisms and develop inspection approaches that will effectively reduce the associated risk of failure (ABS, 2018). Figure 10 shows the steps of performing RCM, namely: For the selected machinery the function of the system which it is part of, is defined. Then by using this definition the failure modes need to be identifies. Then, there is an analysis of the causes of the failure modes and the effects that these failures have. The effect is evaluated by one of four methodologies, namely: FMEA (Failure modes and effects analysis), HAZOPS (A hazard and operability study), FTA (Fault tree analysis) and RBI (Risk-based inspection).

FMEA is an engineering technique used to define, identify, and eliminate known and/or potential problems, errors, and so on from the system, design, process, and/or service before they reach the customer (Omdahl, 1988; ASQC, 1983).

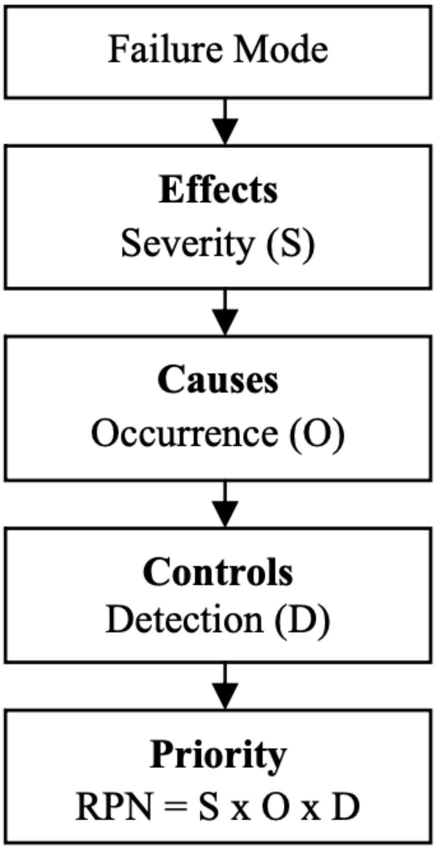


Figure 11: FMEA Tasks (Ben-Daya, 2009)

In Figure 11 the tasks of FMEA are shown. The term severity refers to the consequence of the failure when it happens. The term occurrence refers to the probability or frequency of the failure occurring, and detection is the probability of the failure being detected before the impact of the effect is realized (Ben-Daya, 2009). These three factors are then multiplied in order to produce the risk priority number (RPN) to reflect the priority of the failure modes identified.

On the other hand, HAZOP emerged with the purpose of recognizing potential hazards in establishments that operate using extremely harmful materials. The major care was to abolish every source that can probably lead to a serious accident, such as explosions, fires and toxic release (Swann et al., 1995). Nevertheless, with the passage of time, HAZOP's handling was expanded to various other kinds of services because of its capability, not only to recognize hazards, but also to identify functional deviations from the preferred state (Marhavilas et al., 2020).

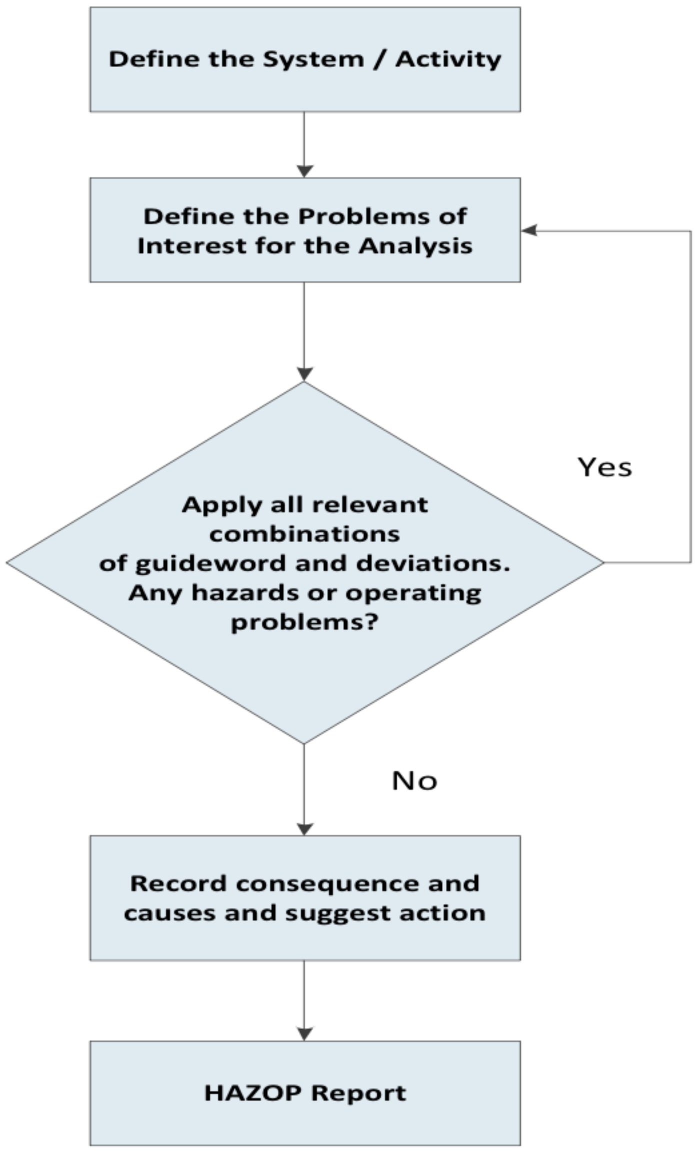


Figure 12: HAZOP Procedure (Silvianita et al., 2014)

Figure 12 shows the steps for performing HAZOP (Silvianita et al., 2014). The first step is to identify and examine the system/activity that is going to be analysed. The second step is to define the potential hazards and significant impact on the system by using guideword and deviation parameters. The third step is to analyse the findings that focus on the critical hazards as well as critical operational problems. The final step is to record the results of HAZOP using special spreadsheets which generally include the guideword, deviation, possible causes, outcome, safeguards and suggestion action.

Moving on to FTA, the development of FTA is an outgrowth of the systems safety approach (Recht, 1965) initiated in the space industry in the late 1950's and early 1960's (Wood et al. 1979). It is a systematic safety analysis tool that proceeds deductively from the occurrence of an undesired event (accident) to the identification of the root causes of that event (Goodman, 1988). Figure 13 shows the six basic steps used to develop a fault tree analysis (Dhillon, 2008).

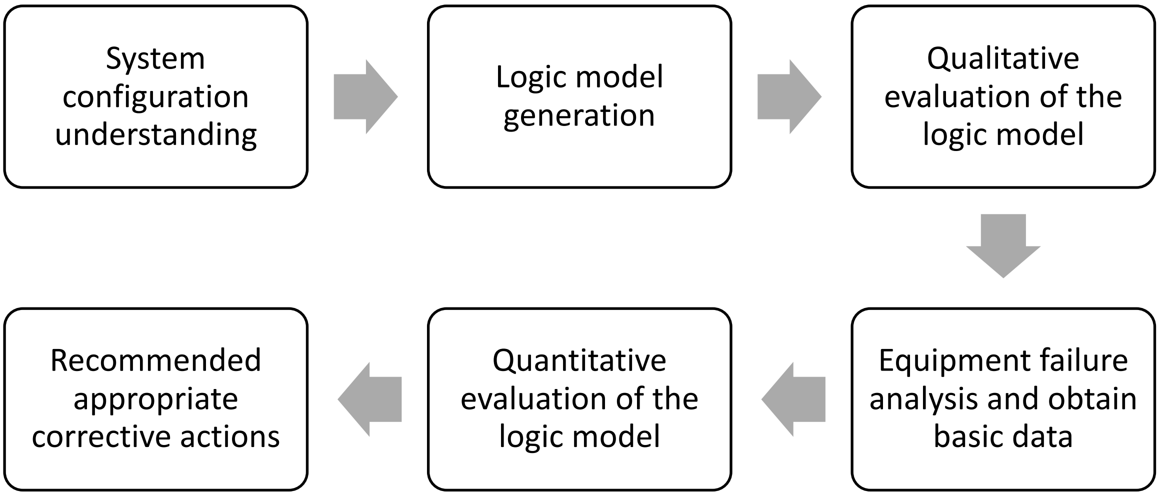


Figure 13: Steps for FTA (Dhillon, 2008)

Firstly, the system definition needs to be defined as well as the system's undesirable event. Then the logic model will be generated by constructing the fault tree using logic and other symbols. The third step is to evaluate the fault tree qualitatively. The fourth step is to obtain basic data (e.g., elementary parts' failure rate and failure occurrence probabilities). The fifth step is to evaluate the fault tree quantitatively. Finally, the last stop is to recommend appropriate corrective actions.

Risk Based Inspection (RBI) provides a modelling process for organizations to control its reliability, safety and health aspects, ensure maintenance compliance and to iteratively improve the technical performance and cost of projects (Sutton, 2015). It requires structured and coherent information management in order to maximize its integration in the target computerized management information system for maintenance and inspection. RBI can be performed in qualitative, quantitative and semi-quantitative way. The results of each method are almost the same, but with the qualitative method results can be evaluated quickly.

Quantitative method involves more detail and calculation, but with more accuracy. Semi-quantitative method uses qualitative speed and quantitative accuracy. In Figure 14 the RBI planning process can be seen (API, 2008). Both the probability of failure and their possible consequences can be previously estimated, quantitatively and qualitatively, depending on the considered system characteristics and the data available for evaluation (Soares et al., 2015).

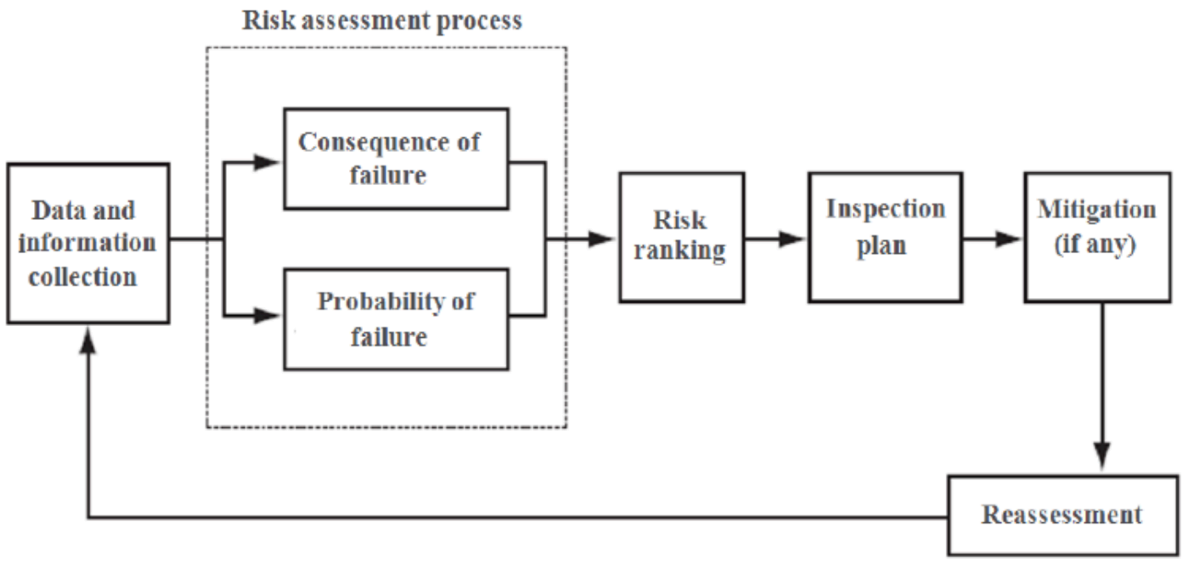


Figure 14: RBI Planning Process (API, 2008)

Then the risk is ranked with the help of an RBI Matrix like the example in Figure 15 where categories of consequence and probability are organized in such a way that the greatest risk position is toward the upper right corner (Zhaoyang et al., 2010). Different array sizes can be used but regardless of the selected matrix, the categories of consequence and probability should provide sufficient discrimination between the evaluated items (Goyet, 2001). Once the equipment risk is known and an acceptable risk criterion established, risk management has to be conducted. Some risks could be identified as low, and no mitigation measure will be required.

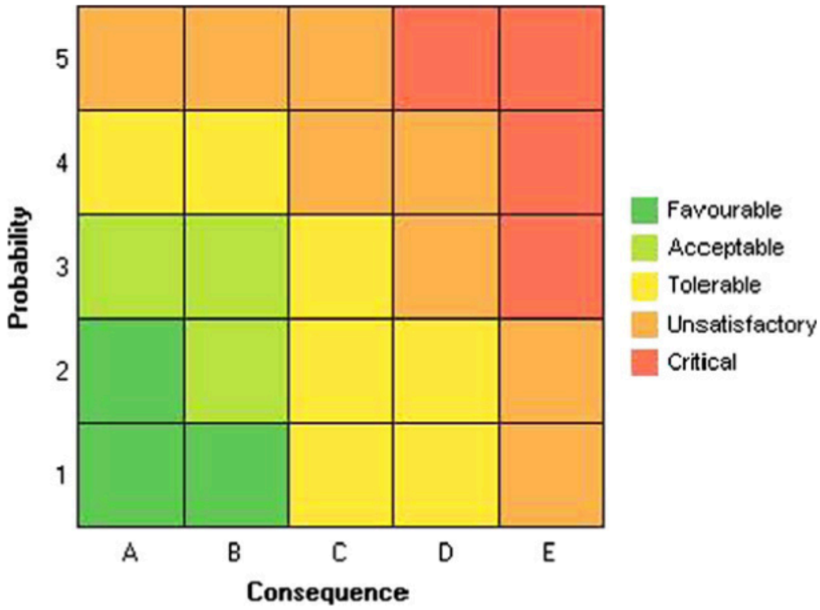


Figure 15: RBI Matrix (Zhaoyang et al., 2010)

Back to the steps of RCM, after the evaluation of the failure effects with some of the above methods (FMEA, HAZOPS, FTA and RBI) the most appropriate maintenance can be determined.

RBI has been used to describe the structural components of a ship or offshore installation (Serratella et al, 2007, Ku et al, 2004 and Faber, 2002). RBI is an interesting maintenance perspective with an incremental stepwise procedure used to examine sensitive equipment such as pressure vessels, heat exchangers and piping in complex industrial plants El-Reedy (2012). Vessels that are designed to be moored on station for the duration of their service life such as Offshore Floating Structures such as FPSO and Floating Liquefied Natural Gas (FLNG) are impractical to return periodically to dry dock for inspection and maintenance. Risk-based Inspection (RBI), in the case of these vessels, provides a flexible approach to help operators mitigate risk and improve reliability leading to a longer time on station, by optimising maintenance schedules and reducing unnecessary inspection activities (Ozguc, 2020). On the other hand, utilizing RBI requires very robust planning and strong management as well as an initial investment before the shipowner can see the benefits of this methodology. In addition, Gabbar et al. (2003) combined RCM with a CMMS in the case of a water-feed process of a nuclear power plant and Rausand and Vatn (2008) illustrated an RCM application in the railway sector. Fonseca and Knapp (2000) demonstrated the combination of RCM with a software package in the chemical process industry. In the shipping industry, RCM is related to the machinery equipment of the vessel (Lazakis et. al 2009).

From the above someone can understand that RCM is a widely applied methodology. However, it may become challenging to implement into complex systems (e.g., ship) because of the need of extensive use of resources and it requires company's top management support during RCM employment (Starr et al. 2010). Moreover, RCM it does not make full provision for the use of condition monitoring techniques, so that the development of potential failures is not followed until just before failure. It is this last remark which highlights a significant RCM shortcoming; that is the lack of an overall maintenance management system which will be flexible enough to suit each specific company/ship in the maritime domain. Table 5 shows the advantages and disadvantages of RCM.

Table 4: Advantages and Disadvantages of RCM

Advantages	**Disadvantages**
•Good audit trail	•Failure data is not easy to obtain because equipment and components are usually replaced before failures to avoid high consequential costs especially in the process and chemical industries.
•Consistent decision-making	•Reliability may not be the main focus – manufacturing plants typically focus on availability.
•Step- by-step procedure	•The RCM structure is not concerned with the outcome of monitoring.
	•Resource intensive

2.5.3 Computerised Maintenance Management System (CMMS)

As machinery and equipment become more complex there was the need for implementation of automated maintenance management systems enhanced by computerized, flexible tools for managing critical assets (Dikis et al. 2014). Computerised Maintenance Management System (CMMS) can manage maintenance information effectively and efficiently (Liu et al. 2010). CMMSs are software programs based on a computer for adjustment and connection which is employed to manage resource usage and work actions and to control extensive data on the labour force, inventory, restoration programs (Cholasuke et al., 2004). CMMS converts maintenance information into appropriate data for decision-making. In this regard, a computerized management system can substantially enhance the opportuneness and correctness of storing and retrieving the necessary data. Figure 16 shows the steps involved in a CMMS plan (Bagadia, 2006).

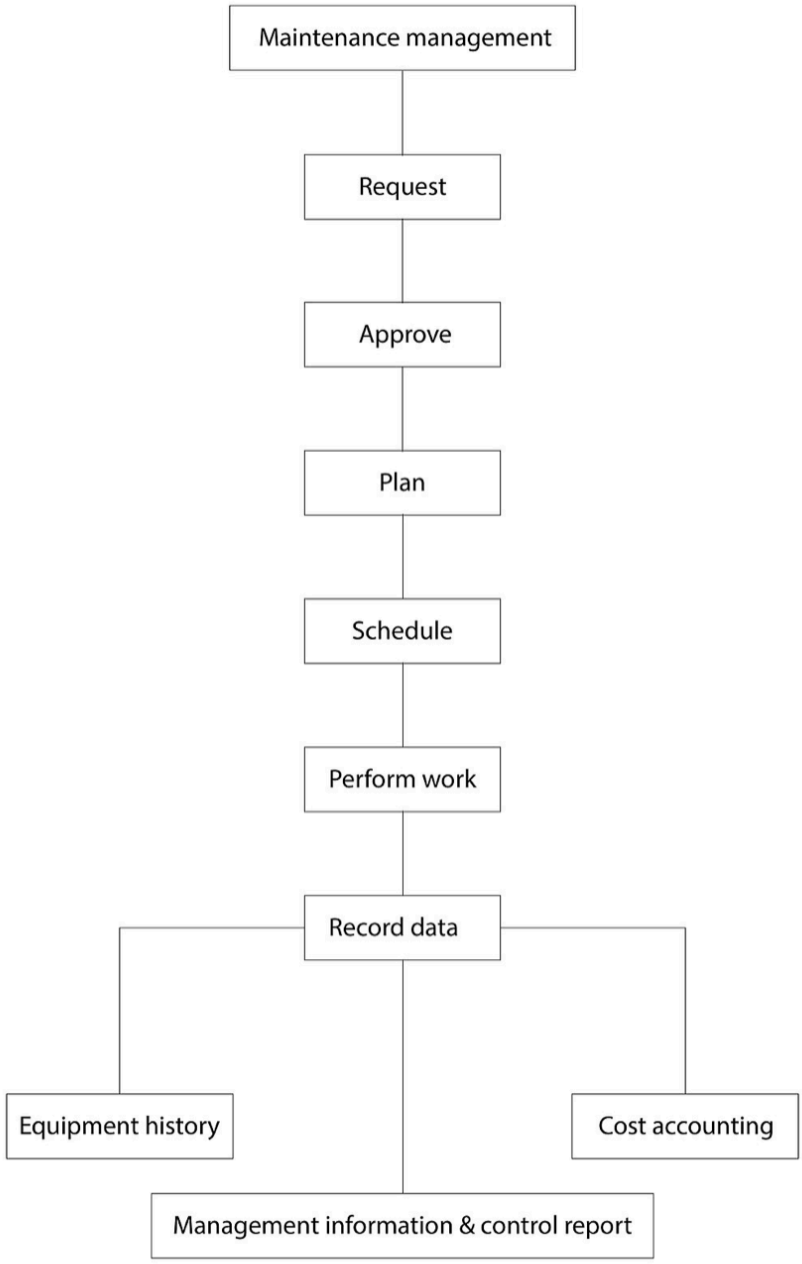


Figure 16: Steps in a CMMS Plan (Bagadia, 2006)

From Figure 16, in more detail: First a computer software program assists in generating, planning and reporting of work orders. Then the request to perform maintenance work is being approved by maintenance supervisors or several levels of management (if large expenditures are required). The CMMS also includes scheduling/planning of work and labour tracking. The data recording varies from simple lists of hours of work to comprehensive records containing equipment history and cost accounting. Concerning the former, complete maintenance history on equipment which can help making decisions regarding maintenance and/or replacing it.

Concerning the latter, every time work is done on an equipment the cost is computed and recorded. This information is important for performing a replacement analysis. Finally, as management information is developed, control reports are summarised covering performance, equipment data, etc. which help plant managers with the decision-making process.

CMMS is coming more and more essential in the industry area (Raouf et al., 1993). It intends to diminish the total time out and frequency of machines breakdown as enhancing the performance and efficiency of maintenance operations by delivering precise information which is important in making intelligent (O'Hanlon, 2004). In addition, it helps gaining information from raw data and enhancing decision making by automating existing processes (Fernandez et al. 2003).

On the other hand, the installation of CMMS may cause opposition to adjust and requires education for those who handle delicate system. Also, due to the high degree of equipment dependency, a failure can interrupt the entire process (Marquez et al., 2004). In addition, a CMM requires an effective control mechanism for restoring and managing information (Swanson, 2003). Also, virtually no commercially available CMMS offers decision support, and this can be a serious drawback as the key to systematic and effective maintenance is managerial decision-making that is appropriate to the particular circumstances of the machine, plant or organisation (Labib, 2008). This decision-making process is made all the more difficult if the CMMS package can only offer an analysis of recorded data. As an example, when a certain preventive maintenance schedule is input into a CMMS, (e.g., to change the oil filter every month) the system will simply produce a monthly instruction to change the oil filter and is thus no more than a diary. Table 5 shows the facilities offered by commercially available CMMS, showcasing that the decision analysis is a "black hole", meaning systems that are hungry for data and resources and provide the decision-maker with information that they already know.

Table 5: Facilities offered by Commercially Available CMMS Packages (Labib, 2008)

Price range	**£ 1,000 +**	**£ 10,000 +**	**£ 30,000 +**	**£ 40,000 +**
Data collection	✓	✓	✓	✓
Data analysis		✓	✓	✓
Realtime			✓	✓
Network				✓
Decision analysis		A "black hole"		

In shipping CMMS integrate all the necessary information (e.g., planned and unplanned maintenance events, machinery monitoring, inventory/spare parts lists) in one database which connects different departments of the shipping company with the ship (Lazakis et al. 2009). This gives valuable time for the ship crew to manage an increasing daily workload much more effectively and perform their duties better. On the other hand, CMMS requires more skilled personnel and thus additional training for handling new technologies. Finally, there is a significant number of interdependent equipment when using CMMS which increases the probability of failures that can disrupt the correct operation of the system.

A system that improves the RCM process integrated with CMMS was proposed by Gabbar et al. (2003). The proposed solution was integrated with design and operational systems and consolidated some successful maintainability approaches to formulate an effective solution for optimized plant maintenance. A case study was used to show the effectiveness of the proposed RCM-based CMMS solution in optimizing plant maintenance over the traditional approaches. Moreover, in the study of Labib (2004) proposed an intelligent model that can be linked to CMMSs to add value to data collected in the form of provision of decision support capabilities and thus overcome the "black hole" problem that was mentioned before. That model was based on combining the analytic hierarchy process (AHP) with fuzzy logic control to render a "Decision Making Grid". Carnero and Novés (2006) developed the evaluation system for selecting computerized maintenance management software in an industrial plant using multi-criteria approaches. This system facilitates the activities of the planner of the system, as well as avoids the problems that an erroneous selection of the software brings, since the CMMS will adapt to the conditions and utilities needed by the industrial plant in these computer packages. Table 6 shows some advantages and disadvantages of CMMS.

Table 6: Advantages and Disadvantages of CMMS

Advantages	Disadvantages
•Helps to cope with the complexity of the work on a daily operational routine while keeping the cost of such a system in reasonable levels.	•Requires education for those who handle delicate system.
•Gaining information from raw data and enhancing decision making by automating existing processes.	•Due to the high degree of equipment dependency, a failure can interrupt the entire process.
•Technical consideration of the asset in question, improved asset reliability, cost-effectiveness.	• “Black hole” when it comes to decision analysis.
	•A number of participants needed substantial time for implementation

2.5.4 Condition Based Maintenance

Finally, the last method of predictive maintenance techniques is called Condition Based Maintenance (CBM). is the most modern and popular maintenance technique discussed in the literature (Dieulle et al., 2001, Han and Song, 2003, Moya, 2004). The scope of Condition Based Maintenance is to detect the upcoming failures before even taking place (Mechefske, 2005). In other words, CBM has as a goal to understand the risks and predeterminate strategic actions, leading to reliability and operational cost reduction (SKF, 2012a). In a more technical aspect, Butcher (2000) defined CBM as a set of maintenance actions based on real-time or near real-time assessment of equipment condition, which is obtained from embedded sensors and/or external tests & measurements taken by portable equipment. The motivation of CBM is that 99% of equipment failures are preceded by certain signs, conditions, or indications that a failure is going to occur (Bloch & Geitner, 1983). Figure 17 shows the procedure for the MIMOSA Open Standard Architecture Condition Based Maintenance (OSA-CBM).

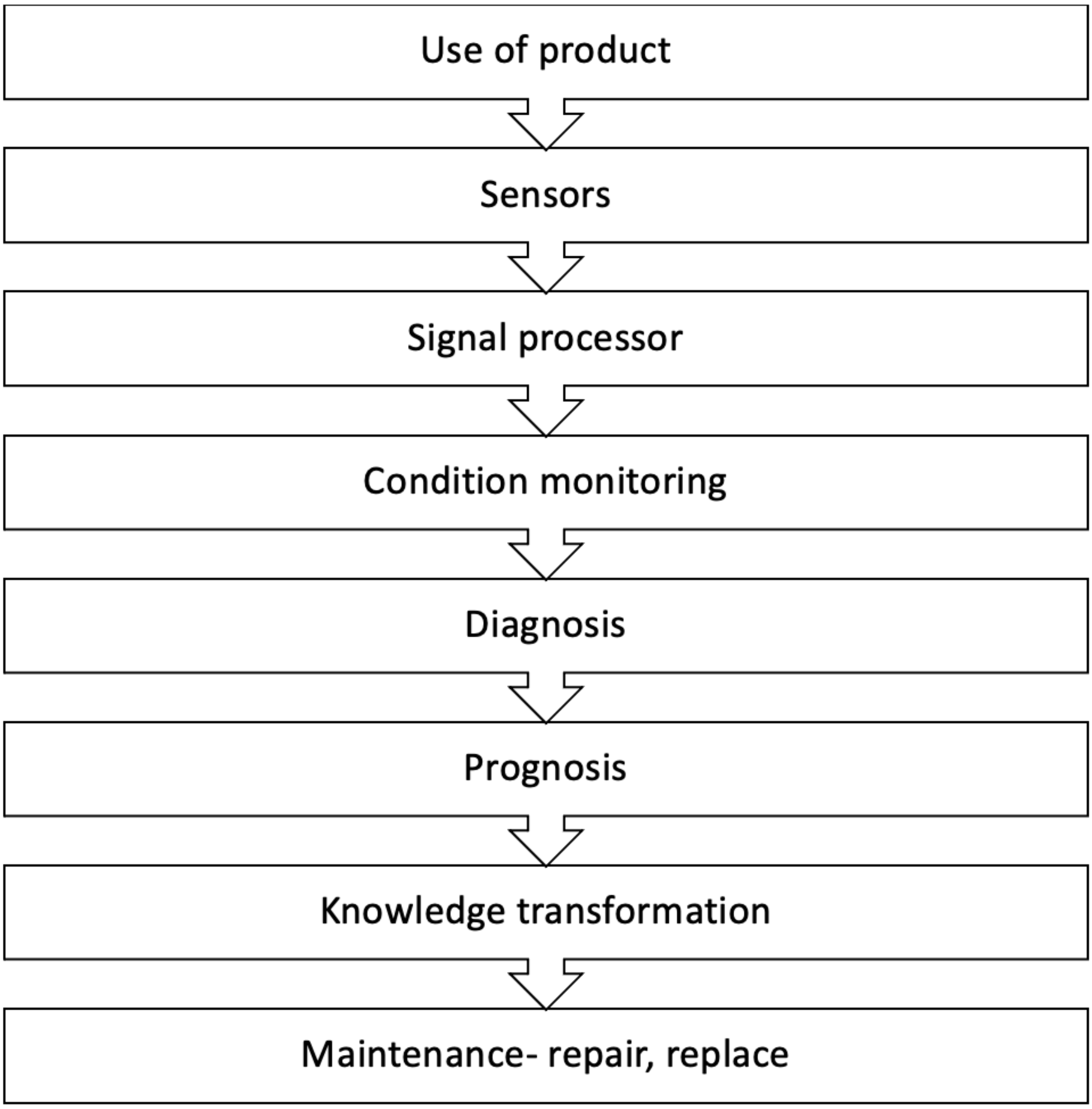


Figure 17:Procedure for CBM Approach (Prajapati et al., 2012)

OSA-CBM is designed by MIMOSA which is an organization involved in the development of the standards for CBM and is a standard for information flow to help realize an end-to-end CBM system (Prajapati et al., 2012). First the use of the asset under question is being determined as well as the data to be collected related to this usage. In the data gathering level, large amount of field data is collected by various data acquisition methods using sensors, wired and wireless techniques, and stored in a database. Before data gathering, it is necessary to identify which data should be gathered during asset usage period for CBM. For the analysing, it is required to develop an algorithm that assesses the behaviour and degrading level of an asset and predicts its remaining lifetime. Analysing has two parts in CBM: diagnostics and prognostics. Diagnostics consists of fault detection, fault isolation determining the location of the fault, and fault identification determining the fault mode (Gruber et al., 2013). Diagnostics requires data pre/post-processing, data interpretation, data fusion, and several statistics methods with asset specific knowledge. On the other hand, prognostics corresponds to the estimation of the time to failure and the risk for one or more existing and future failure modes based on anticipated future usage (Mejia et al., 2012). Prognostics deals with estimation of system health index and predictions of Remaining Useful Life (RUL) (Chen et al., 2012) and is a word recently coined by the scientific community to address the combination of diagnosis and prognosis (Audisio et al., 2004). Finally, the final step is the knowledge transformation which refers to advisory generation for maintenance (repair or replace).

After presenting the steps of CBM it would be useful to review the international standards of a CBM system. There are several international standards related to CBM approach as it can be seen in Table 7.

Table 7: CBM International Standards

Standards	Subject
IEEE 1451	Smart transducer interface for sensors and actuators
IEEE 1232	Artificial Intelligence exchange and service tie to all test environments
ISO 13372	Condition monitoring and diagnostics of machines- Vocabulary
ISO 13373-1	Condition monitoring and diagnostics of machines- Vibration condition monitoring-Part 1. General procedures
ISO 13373-2	Condition monitoring and diagnostics of machines- Vibration condition monitoring-Part 2. Processing, analysis and presentation of vibration data
ISO 13374	MIMOSA OSA-CBM formats and methods for communicating, presenting and displaying relevant information and data
ISO 13380	Condition monitoring and diagnostics of machines- General guidelines on using performance parameters
ISO 13381-1	Condition monitoring and diagnostics of machines- Prognostics, general guidelines

ISO 14224	Petroleum, petrochemical and natural gas industries-collection and exchange of reliability and maintenance data for equipment
ISO 17359	Condition monitoring and diagnostics of machines-General guidelines
ISO 18435	MIMOSA OSA-EAI diagnostic and maintenance applications integration
ISO 55000	Asset management

From Table 7, some of the international standards are the condition monitoring and diagnostics standards for machinery industry, such as: ISO 13372, ISO 13373, ISO 13380, and ISO13381. In addition, ISO 13374 addresses the MIMOSA OSA-CBM representing formats and methods for communicating, presenting, and displaying relevant information and data. There are also standards related to the issues of integration and data sharing among manufacturing facilities for CBM like ISO 18435 (MIMOSA OSA-EAI). Recently, not only machinery industry but also plant engineering industry, e.g., petroleum, petrochemical and natural gas industry, starts to have more interest in the CBM policy, as it can be seen in ISO 14224.

Moving on to the advantages of this method, one great benefit of CBM is that minimizes the failures occurred by human error during maintenance operations, also called infant mortality since less invasive maintenance needs to be performed (Dhillon and Liu, 2006). In addition, CBM allows to perform better planned maintenance, reduce or eliminate unnecessary inspections, and decrease time-based maintenance intervals with confidence (Chen et al. 2012). Furthermore, the application of CBM systems in industry has been reported to be one way of decreasing maintenance budgets (Bengtsson, 2004; Horner et al., 1997). Also, CBM can discover the root cause of failure and that means that the cause can be eliminated/ engineered out (Yam et al., 2001). On the other hand, according to Hashemian and Bean (2011), a 30% of industrial equipment does not benefit from CBM. That is mostly because of the investment cost for CBM which is usually high. In addition, to implement the CBM, it is prerequisite to install and use monitoring equipment and to develop some level of modelling or decision-making strategy (Ellis, 2009). Also, to implement the CBM, not only investment of hardware but also training on staff is required which will cause expensive cost. In addition, the technologies and technical methods for the CBM approach are still in their infancy which means that there are some limitations in ensuring the accuracy of diagnostics and prognostics (Shin et al. 2015). Finally, CBM is not applicable to all retained assets and should only be applied if condition monitoring techniques are useful and cost-effective (Horner et al., 1997).

Concerning applications of CBM, Trutt, Sottile, and Kohler (2002) presented a CBM method for induction motor windings based on a voltage mismatch technique. The proposed method demonstrated the robust nature of the monitoring process not only under conditions of power supply imbalance but also in situations where motor construction imperfections exist, and mechanical loads are unpredictable (Ahmad et al., 2012). Yang, Mathew, and Ma (2005) monitored and diagnosed rolling bearing defects at inner and outer race faults based on vibration signals. A new basis pursuit method was applied in the extraction of features from signals collected. Shonet (2003) noted that in order to effectively implement CBM, performance metrics for components and systems should be developed. Liu and Wang (2006) presented a CBM process that integrated data collection and vibration signal analysis to assess equipment conditions and maintain the operational performance of hydropower turbine units. Mehta et al. (2015) suggests a way to prevent fatal errors by combining information from two or more sensors and the intelligence in the CBM system is described using the Bayesian probabilistic decision framework and the data generated during validation. Table 8 shows the advantages and disadvantages of CBM.

Table 8: Advantages and Disadvantages of CBM

Advantages	**Disadvantages**
•Enhance machine's availability, reliability, efficiency and safety.	•CBM is not applicable to all retained assets and should only be applied if condition monitoring techniques are useful and cost-effective.
•Discovers the root cause of failure.	• Expensive initial investment of hardware and training on staff is required.
•Minimizes infant mortality.	•Is a rather new approach and thus has some limitations.
•Reduced maintenance budget through controlled spare part inventories.	

2.6 Approaches for Diagnostics and Prognostics in CBM

As it was mentioned before, there are two major Condition Based Maintenance (CBM) approaches (Manno et al., 2014):

- Data driven approach
- Model driven approach

These two approaches will be analysed further in the next sub-sections.

2.6.1 Data Driven Approach

The data driven approach is also known as the data mining approach or the machine learning approach and is a rather new approach which uses historical data to automatically learn a model of system behaviour (Schwabacher, 2005). Data-driven approaches use real data obtained from data acquisition system and track features revealing the components degradation and to forecast the global behaviour of a system (Marton et al., 2013). Figure 18 shows a high-level representation of a data driven CBM framework (DNV, 2014).

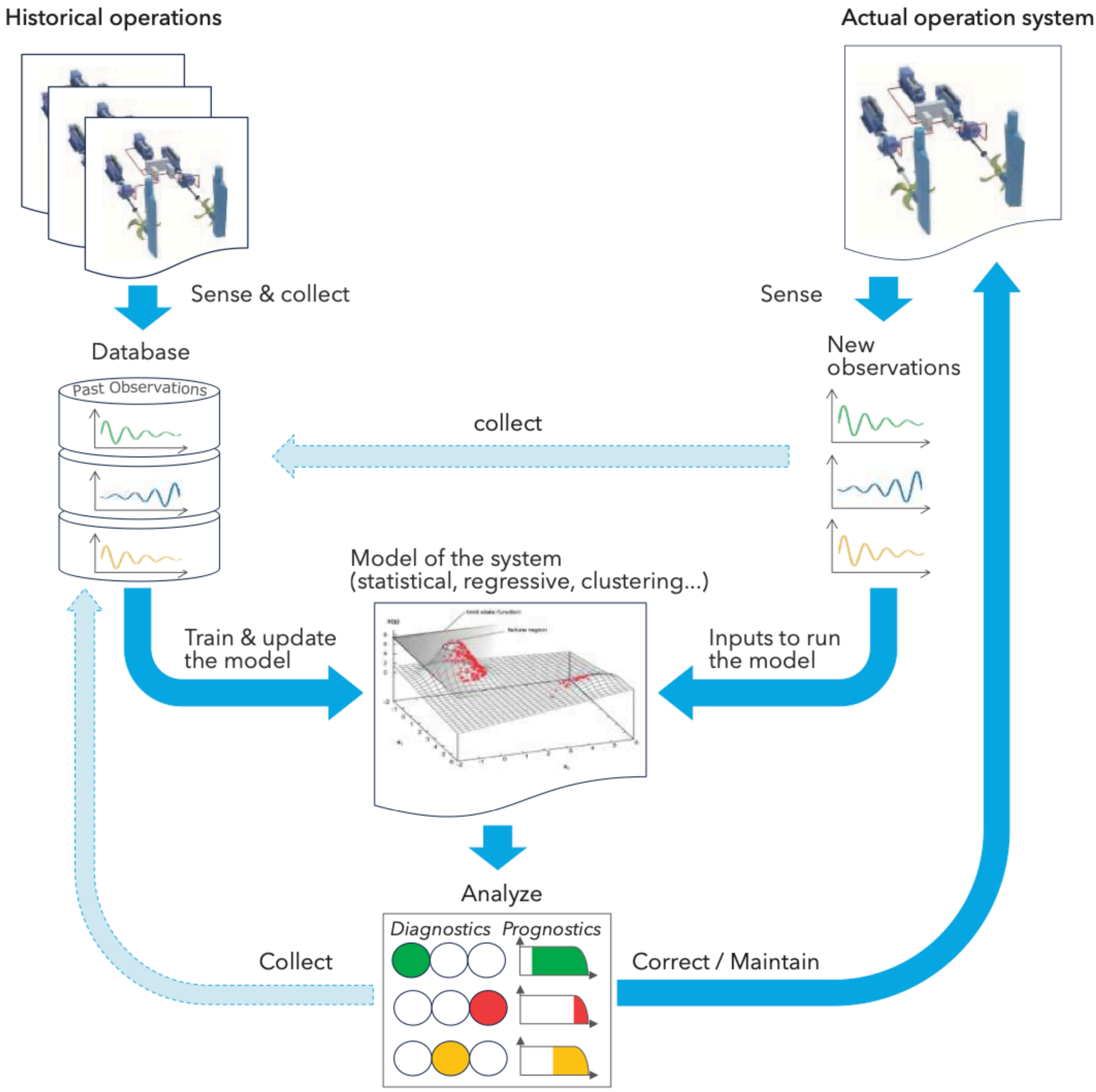


Figure 18: Data-Driven CBM framework (DNV, 2014)

According to Figure 18 the data driven approach uses systematically a database to store data referring to: historical operational values of the system, new operational values of the system and the output of the model of the system (diagnostics and prognostics). In addition, the database also has the usage of feeding new data to update the model that was trained with the historical data with the new observations and thus increase the accuracy long term. The models used in data-driven approaches can be divided into two categories: artificial intelligence (AI) techniques (neural networks, fuzzy systems, decision trees, etc.), and statistical techniques (multivariate statistical methods, linear and quadratic discriminators, partial least squares, etc.) (Dragomir et al., 2009; Lam et al., 2010).

By using data driven CBM, maintenance can be predicted before failure occurs to ensure continuous production, product quality, and enterprise efficiency (Morant et al. 2016). In addition, a data driven approach can process a wide variety of data types and exploit the nuances in the data that cannot be discovered by rule-based systems (Peng et al., 2010). Furthermore, data-driven approaches have the ability to transform high-dimensional noisy data into lower dimensional information and thus reduce a complex problem to a solvable solution (Luo et al., 2003). Finally, data driven CBM facilitate model building via identification of dynamic relationships among data elements (Poongodai et al., 2013). On the other hand, data driven CBM system's efficacy is highly dependent on the quantity and quality of system operational data (Niu, 2017). In addition, performance is limited by the availability/quality of the database (Bengtsson, 2003). Table 9 summarises the advantages and disadvantages of the data driven approach.

Table 9: Advantages and Disadvantages of Data Driven Approach

AI Approach	Description	Classification	Advantages	Disadvantages
Data Driven	•Automatically fit a model of system behaviour to historical data, rather than hand-coding a model.	•Conventional numerical algorithms: Linear regression Kalman filters •Machine learning algorithms: Neural networks Decision trees Support vector machines.	• Can process a wide variety of data types. •Very adaptable. •Ability to transform the high-dimensional data into lower-dimensional data •Facilitate model building via identification of dynamic relationships among data elements.	•Their efficacy is highly dependent on the quantity and quality of system operational data. •Performance is limited by the availability/quality of the database.

2.6.2 Model Driven Approach

In contrast to the data-driven approach, the model driven approach incorporates a physical understanding of the targeted system (Shin et al. 2015). It includes more classical AI techniques such as rule based expert systems, finite-state machines and qualitative reasoning (Sugeno et al. 1993). Figure 19 shows a high-level representation of a model driven CBM framework (DNV, 2014).

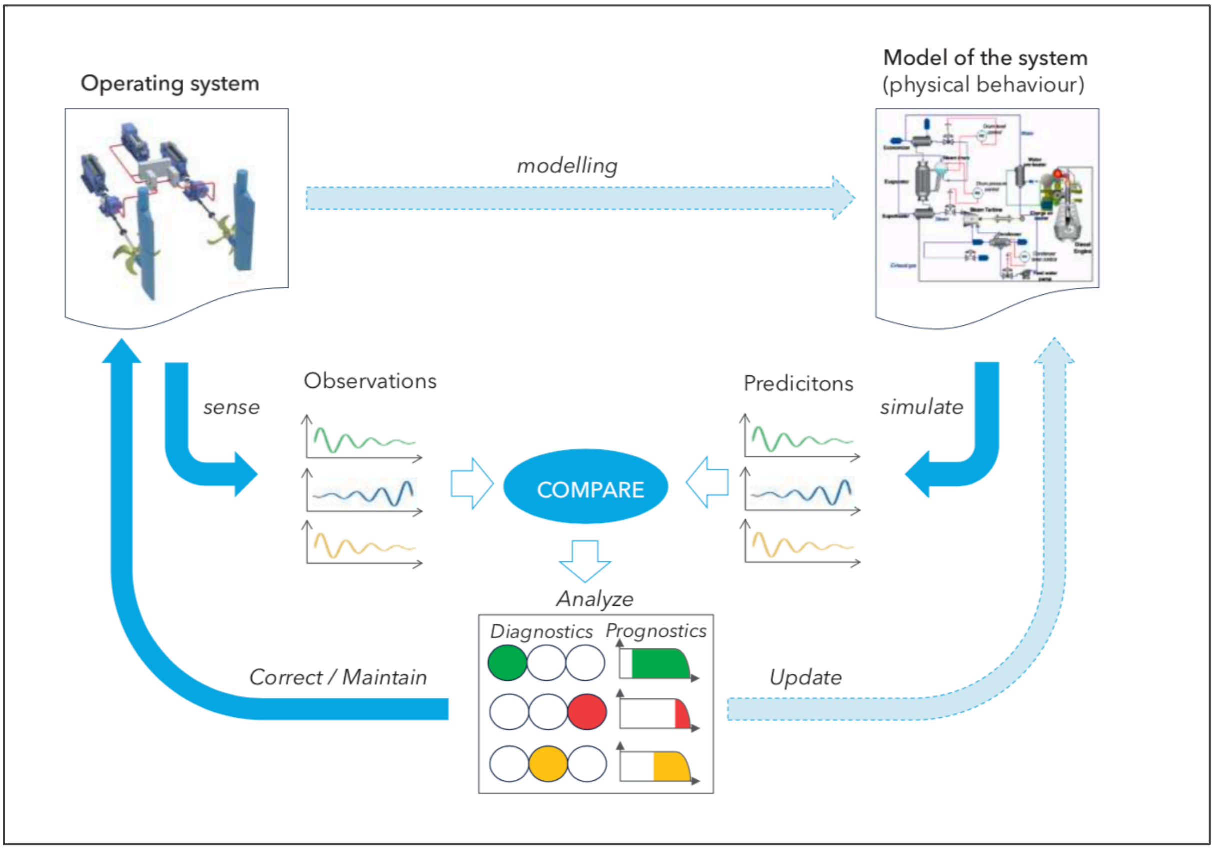


Figure 19: Model-Driven CBM framework (DNV, 2014)

From Figure 19 in the model driven CBM the diagnostics/prognostics come after comparing the observations obtained by sensors with the predictions of different system conditions from the simulation of the system by a mathematical model. Model based algorithms encode human knowledge via a hand-coded representation of the system and can be either physics based, or AI based (Schwabacher et al., 2007). Hand-coded model uses qualitative, rather than numerical, variables to describe the physics of the system (Weld & de Kleer, 1990) while model-based AI techniques include rule-based expert systems, finite-state machines and qualitative reasoning (Williams & Nayak, 1996; Kurien & Nayak, 2000).

Model-driven approaches can have high precision and give guide efficient diagnostic procedures for specific situations (Atamuradov et al., 2017). In addition, in many situations, the changes in feature vector are closely related to model parameters (Chelidze et al., 2002). Thus, it can also establish a functional mapping between the drifting parameters and the selected prognostic features can be established (Luo et al., 2003).

On the other hand, model-driven approaches are known for their Inability to generalize (they are specific to one system), to deal with new conditions and to learn from their mistakes

(Medjaher et al. 2013). Furthermore, the model-based approach requires very specific knowledge which may not be available for a specific problem and thus a good model is very difficult to build up (Caesarendra et al., 2010). Table 10 summarises the advantages and disadvantages of the model driven approach.

Table 10: Advantages and Disadvantages of Model Driven Approach

AI Approach	Description	Classification	Advantages	Disadvantages
Model Driven	•Encode human knowledge via a hand-coded representation of the system.	•Physics based: System of differential application. •Classical AI Techniques: Rule-based expert systems Finite-state machines Qualitative Reasoning.	•Can guide efficient diagnostic procedures for specific situations. •Can establish a functional mapping between parameters and selected prognostic features	•Inability to generalize. •Inability of the model to adjust in new conditions. •Inability of the model to learn from its errors. •Require very specific knowledge.

2.7 Steps of Data-Driven CBM

This thesis is more concerned about data-driven approaches rather than the traditional model-driven approaches. A data-driven Condition Based Maintenance System (CBMS) generally has the following steps (Lee et al., 2004):

- Data acquisition step
- Data processing step
- Maintenance decision-making step

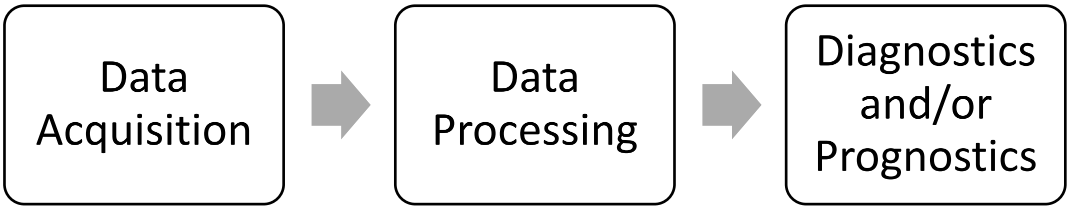


Figure 20: Process of a Data-Driven CBMS (Lee et al., 2004)

The first step is the data acquisition step (information collecting), to obtain data relevant to system health. The second step is the data processing step (information handling), to handle and analyse the data or signals collected in first step for better understanding and interpretation of the data. Finally, the third step is the maintenance decision-making step (decision-making), to recommend efficient maintenance policies. In the following sections the three steps of the CBM approach will be analysed, namely: data acquisition, data processing and diagnostics and/or prognostics.

2.7.1 Data Acquisition

As it was shown in Figure 20, the first step of a CBM system is to acquire data to perform the analysis. Acquired data can be split into two categories (Jardine et al., 2006):

- Event Data
- Condition Monitoring Data

The first category refers to events such as overhauls, installations, replacements and the second one refers to data used to assess the health of the machinery such as pressures, temperatures, flows and other. Table 11 shows an example of the event data from a maintenance database (Robles et al., 2012).

Table 11: Example of Recorded Events from a Maintenance Database (Robles et al., 2012)

Name	Date	Ope.	Cd	IT	N	Code
Dupond	11/01/2007	Lubrication	PM	20	1	9
Dupond	11/01/2007	Lubrication	PM	20	2	9
Dupond	12/01/2007	Lubrication	SEC	30	3	5
Dupond	12/01/2007	Lubrication	PM	30	4	5
Dupond	13/01/2007	Padlock	PM	10	5	6
Dupond	13/01/2007	Padlock	NTR	30	6	5
Dupond	13/01/2007	Padlock	NTR	30	7	5
Dupond	16/01/2007	Lubrication	SP	90	8	1
Dupond	19/01/2007	Padlock	OT	10	9	3

On the other hand, Figure 21 shows an example of condition monitoring data coming from two sensors and then plotted in a scatter plot against each other (Hiruta et al., 2019).

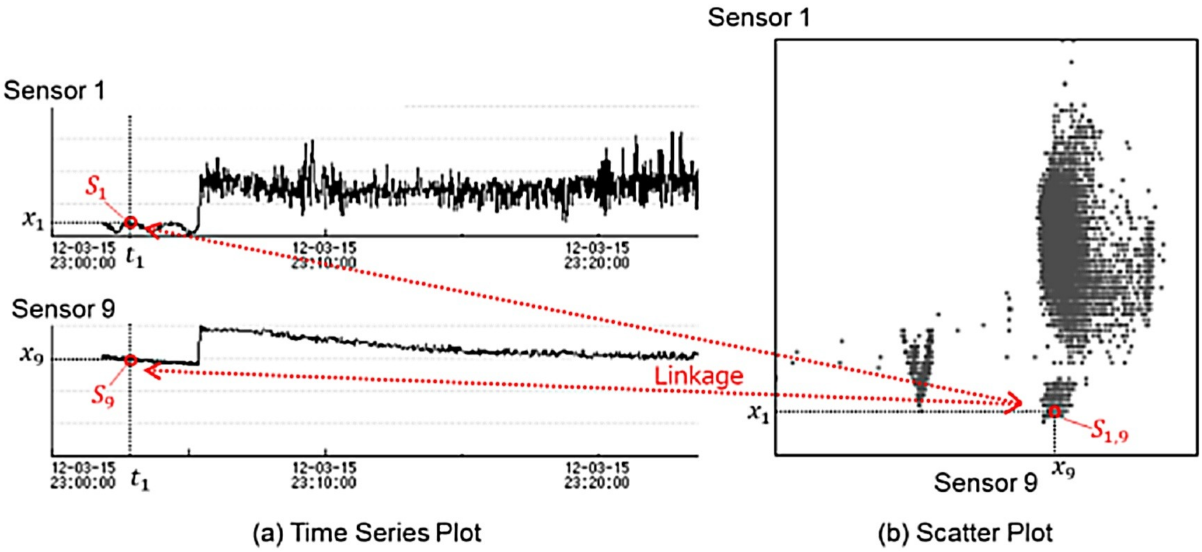


Figure 21: Example of Condition Monitoring Data (Hiruta et al., 2019)

Both of the above categories are important for a successful CBMS. In that case, the sampling of the data can be implemented in two ways namely, continuous condition monitoring and periodic condition monitoring (Cocconcelli et al., 2018). Regarding the former, sensors are recorded continuously. This sampling policy is recommended for those critical components with a high impact on the costs and a short time-to-failure. Regarding the latter, sensors are recorded at scheduled time intervals. This policy is particularly suitable for components with a medium–high time-to-failure.

2.7.2 Data Processing

The very first step of data processing is always data cleaning and this needs to be done for both event data and condition monitoring data. The event data are usually added manually so a lot of mistakes are always made. On the other hand, in condition monitoring data the error may come from sensor faults and in that case a sensor fault isolation is usually the common practise (Xu et al., 2003). Performing data cleaning in condition monitoring especially is of paramount importance since it reduces the probability of having a CBMS with 'garbage in-garbage out' situation, where invalid data values determine the algorithm's diagnostic and prognostic behaviour (Sabari et al., 2014). Usually, the number of variables in the raw data is big and there can be complex correlation patterns hidden (Guatum et al., 2015). In that case, is better to perform multivariate analysis techniques such as principal component analysis (PCA), random projection (RP) with Johnson-Lindenstrauss lemma, independent component analysis (ICA), and Self-Organized Maps (SOM) in order to reduce the dimensions of the problem and eliminate unnecessary correlations between the features (Thrun et al., 2020).

The PCA method finds a new coordinate system that is obtained from the old one by translation and rotation only and it moves the centre of the coordinate system with the centre of the data (Jolliffe et al. 2016). This transformation implies a dimensionality reduction of the original data so, a few of these components are sufficient to adequately represent the hidden sources of variability in the process (Wold et al., 1987). It moves the x-axis into the principal axis of variation where you see the most variation relative to all the data points and it moves further axis down the road into an orthogonal less important directors of variation.

PCA returns an importance value (measures the amount of spread) which is much larger for the first axis of variation and very small for the second axis of variation and always gives independent features. This method is heavily used with unsupervised learning in order to provide to the machine learning algorithms data with only the important independent features (Zhong et al. 2016). Figure 22 shows an illustration of the PCA method (Niculescu et al., 2016).

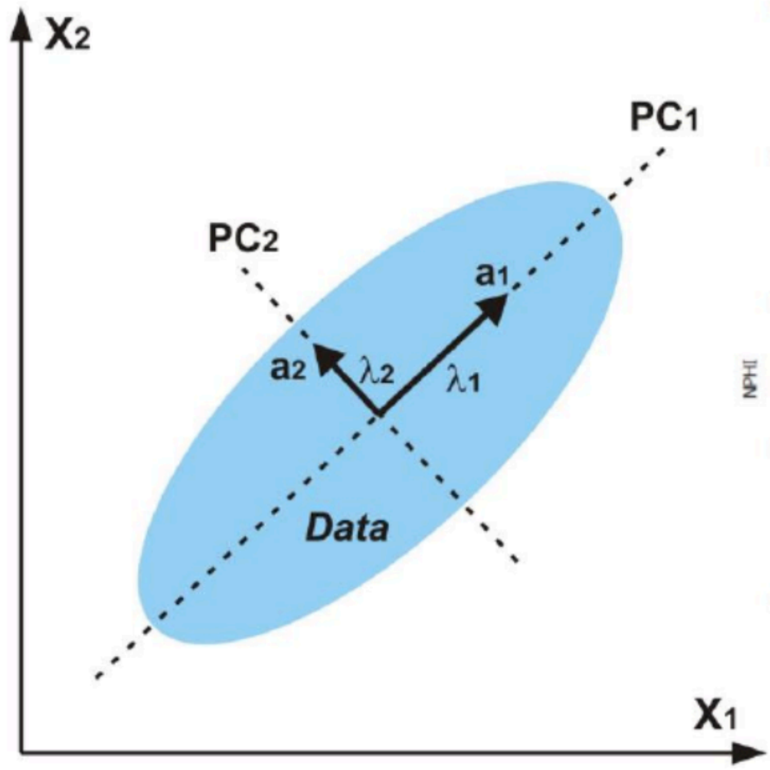


Figure 22: Illustration of the PCA Method (Niculescu et al., 2016)

PCA finds the main variability directions in the data and defines a new coordinate system, using optimal rotations. The axes of this system are defined by the eigenvectors a1 and a2. The eigenvalues λ1 and λ2 correspond to the data variance in the newly defined coordinate system. PCA's key advantages are its low noise sensitivity, the decreased requirements for capacity and memory, and increased efficiency given the processes taking place in a smaller dimension (Phillips et al., 2005; Asadi et al., 2010; Karamizadeh et al., 2013).

On the other hand, the covariance matrix is difficult to be evaluated in an accurate manner and even the simplest invariance could not be captured by the PCA unless the training data explicitly provides this information (Diao et al., 2008). In addition, PCA cannot handle non-linear data (Juvonen et al., 2014). In the various real-world problems, the PCA is frequently used in data sets with some intrinsic complexity (Tuncer et al., 2008; Horenko et al., 2006; Rothenberger et al., 2003; Atsma & Hodgson, 1999; Barbieri et al., 1999).

The PCA is applied as a cluster analysis tool to form machine groups and part families simultaneously (Hachicha et al., 2006). Application of PCA has been in representing the data using a smaller number of variables (Wall et al., 2003). For example, PCA have been used to represent face images efficiently (Sirovich et al., 1987; Turk et al., 1991). Hernandez et al.

(2018) also used PCA in an application of Image Compression- an original image was taken and compressed by using different principal components. Finally, Wang (2018) used PCA, categorical regression tree and back propagation network for the prediction of engine failure time and experimental and analytical comparisons showed that the proposed method provided significantly improved prediction accuracy to the previous used method.

Moving on to the RP: It is also a dimensionality reduction method. It is computationally more efficient than PCA but with less quality (Schu et al, 2016). The difference from PCA is that RP just uses a random line for projection from d dimensions to k dimensions (Linial et al., 1995; Johnson et al., 1984; Achlioptas, 2003). RP uses the Johnson-Lindenstrauss lemma (Johnson et al., 1984) where dataset of N points in high-dimensional space can be mapped down to a much lower dimension in a way that preserves the distance between the points to a large degree (Vempela, 2005). Figure 23 shows a Johnson-Lindenstrauss lemma illustration (Fejri, 2017).

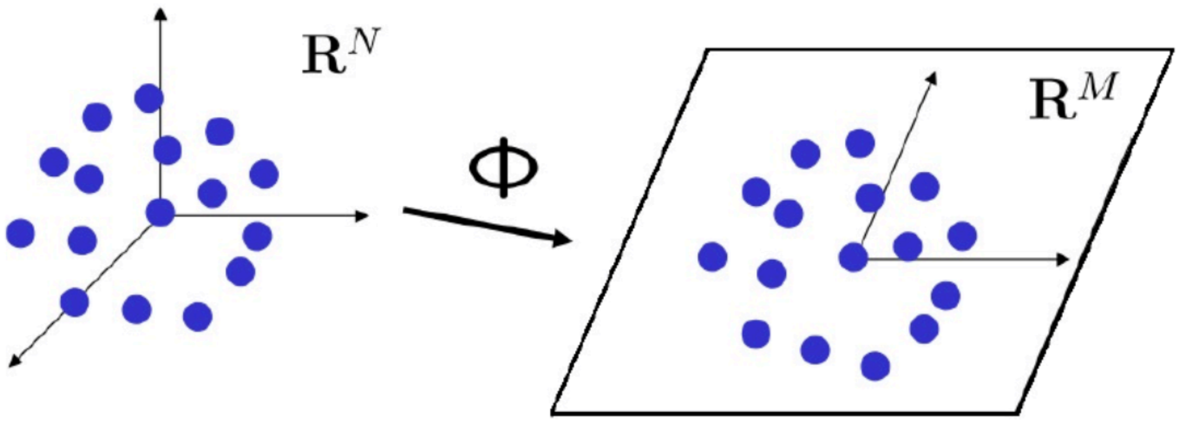


Figure 23: Johnson-Lindenstrauss Lemma Illustration (Fejri, 2017)

From Figure 23 the linear application Φ, provided by Johnson-Lindenstrauss lemma, is simply a way to project the dataset points in another space, without moving too much the distance of each pairwise of points. Concerning the benefits of choosing the RP method: Firstly, it is simple and efficient to use (Lin et al., 2003). In addition, PC can be used also to reduce the dimension of a mixture of Gaussians (Dasgupta 2000). On the other hand, a drawback of RP is that it is highly unstable – different random projections may lead to radically different clustering results (Fern et al., 2003). In addition, it offers less accuracy than the PCA method (Deegalla et al., 2007).

Concerning application of the RP: Zhou et al. (2015) used RP and k-Nearest Neighbor Rule for fault detection in semiconductor manufacturing processes. The authors concluded that the proposed method not only could reduce the computational complexity and storage space, but also approximately guarantee the advantages of kNN rule in dealing with the problems of multimode batch trajectories and nonlinearity that often coexist in semiconductor processes.

Papadimitriou et al. (1998) use RP in the pre-processing of textual data. Kurimo (1999) applies RP to the indexing of audio documents. In Damasevicius et al. (2016) the authors proposed a method that does feature extraction and feature dimensionality reduction by using computationally efficient RP and can recognize of daily human activities. Kleinberg (1997) and Indyk and Motwani (1998) use RP in nearest-neighbor search in a high dimensional Euclidean space, and also present theoretical insights. In Liu et al. (2006) the authors introduced data perturbation technique using RP transformation where some noise is added to the data before being sent to the cloud server.

The next data processing technique is the ICA. In contrast with PCA which maximize variance, the ICA assumes that the features are a mixture of statistical independent sources and it tries to isolate the independent sources (Hyvarinen et al. 2000) by assuming they have non-Gaussian distributions (Ge et al. 2007). Figure 24 illustrates the differences between the PCA and ICA (Lubo-Robles, 2018).

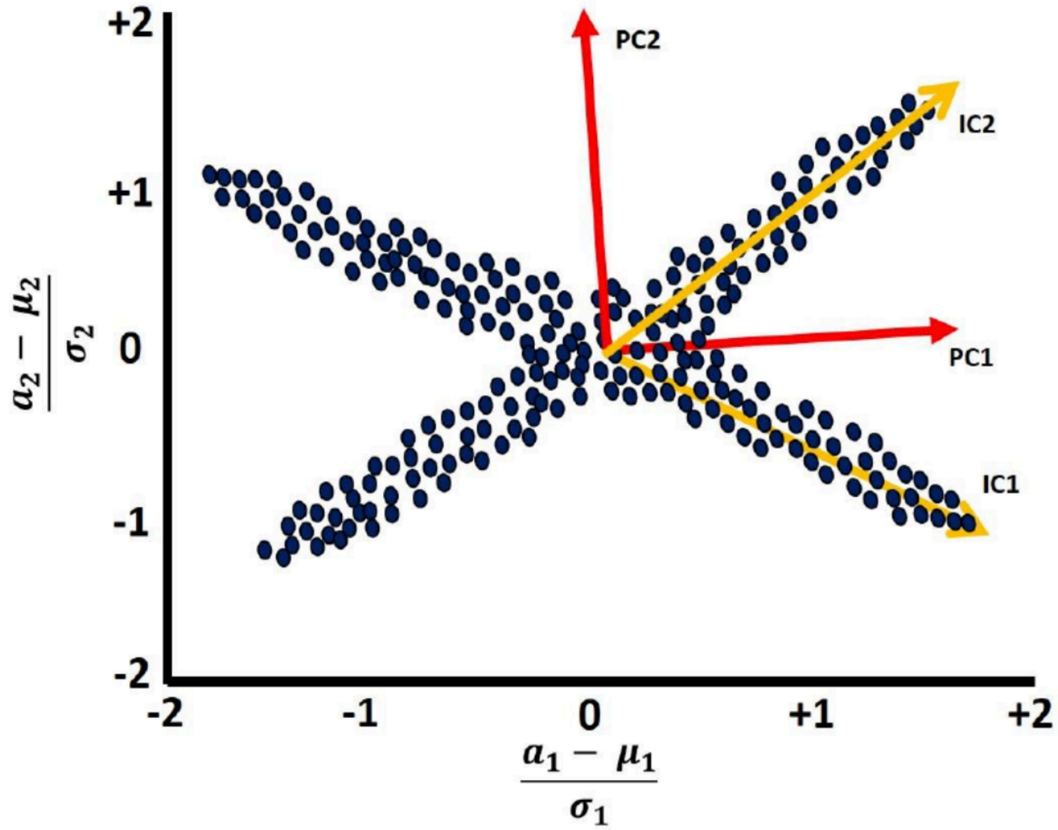


Figure 24: Differences Between PCA and ICA (Lubo-Robles, 2018)

In Figure 24, attributes a1 and a2 are scaled by their means and standard deviations. The first eigenvector v1 is a line that least-squares fits the data cloud and best represent the variance of the data. PC1 is a projection of each data point onto v1. The second eigenvector v2 is a perpendicular to v1 and for two dimensions these two eigenvectors best represent the data. In contrast, the independent components IC1 and IC2 are latent variables whose order is undefined, and they are not orthogonal between each other (Hyvarinen et al. 2000; Tibaduiza et al., 2012). One advantage of using ICA is that it is generally more precise than the PCA method. On the other hand, ICA attempts to find statistically independent new dimensions. In contrast to PCA, which only requires that the different dimensions be uncorrelated, this is a stronger constraint (Edelman, et al., 1997). In addition, concerning algorithm speed, ICA is computationally more intensive than PCA, as it is a more complex algorithm (Jung et al., 1998).

Stefatos et al. (2010) proposed a dynamic ICA approach for fault detection and diagnosis. The authors concluded that the proposed approach is able to accurately detect and isolate the root causes for each individual fault. ICA finds application in many areas such as: separation of mixed voices or images (Dagher et al., 2006; Kwak et al. 2008), analysis of several types of data (Jutten et al., 1991), feature extraction (Delfosse et al., 1995), speech and image recognition (Cardoso, 1997), data communication (Oja et al., 1992), sensor signal processing (Cvejic et al., 2007; Cardoso et al., 1996), system identification (Yang et al., 2007; Shifeng et al., 2007) and biomedical signal processing (Van Dun et al., 2007, Waldert, 2007; Tan, 2001; Cichocki, 1996).

Finally, SOM maps input data to neurons in such a way that the distance relationships between input signals are mostly preserved (Kohonen 2013). Every data item is mapped into one point in the map and the distances of the items in the map reflect similarities between the items (Kohonen 1998). The SOM can be interpreted as a topology for preserving mapping from input space onto the two-dimensional grid of the map (Raptodimos et al., 2018). Figure 25 shows the structure of the SOM network (Ghaseminezhad et al., 2011).

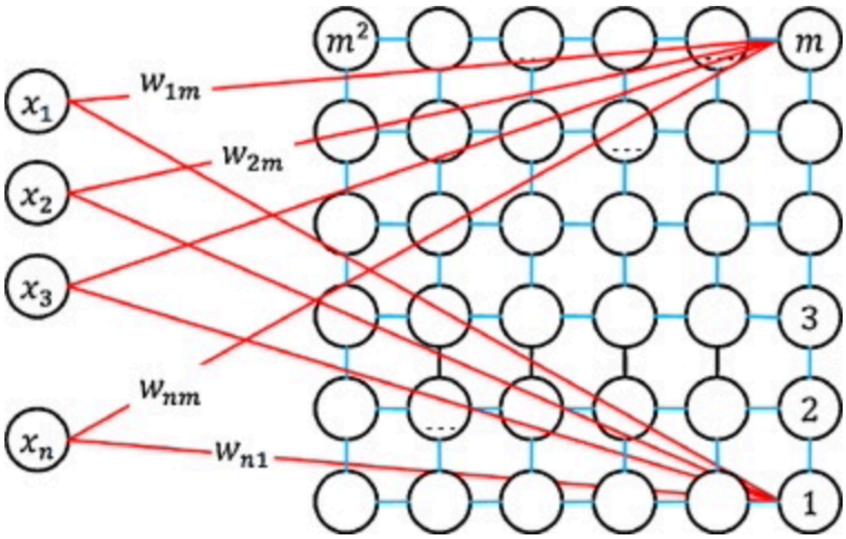


Figure 25: The Structure of the SOM Network (Ghaseminezhad et al., 2011)

As shown in Figure 25, the SOM network is an array of M = m × m processing neurons. The n components of the input vector x are connected to each neuron in the array. A synaptic weight wij is defined for a connection from the ith component of the input vector to the jth neuron. Therefore, an n-dimensional vector wj of synaptic weights is associated with each neuron j. Figure 26 shows a schematic representation of SOM (Kind et al., 2014).

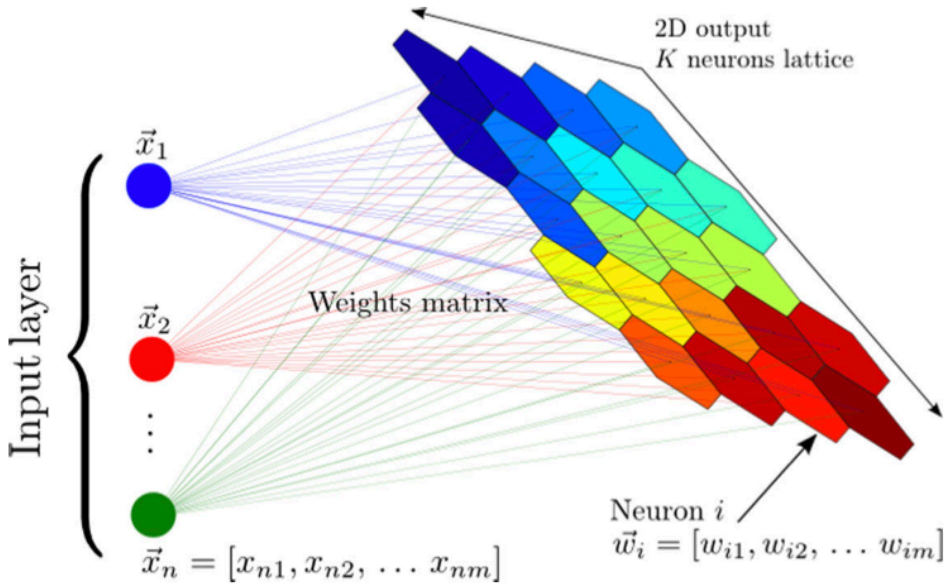


Figure 26: A schematic representation of a SOM (Kind et al., 2014)

In Figure 26 the colour of the map encodes the organization of groups of objects with similar properties. The. main benefit of SOM maps is the detection of nonlinear relationships between variables (Lawrence et al., 1999) compared to the other techniques that require linear relationships. On the other hand, SOM requires a number of parameters that to be set beforehand. That means that the size and topology of the map needs to be determined, as well as the values for the training parameters and one must spend time on the optimization of the mapping (Wehrens, 2009).

In Birgelen et al. (2018) SOM were used for anomaly localization and predictive maintenance. and concluded that the anomaly localization provides information to better locate the origin of the degradation giving experts a head-start on the analysis. In addition, Schwart al. (2020) presented a fault mode identification methodology based on SOM and concluded that without prior knowledge on the faults, the proposed algorithm was able to identify the number of operational modes as well as the fault mode number for the five datasets. Table 12 shows a summary of the processing techniques discussed in this section.

Table 12: Summary of Processing Techniques

Dimensionality reduction techniques	**Advantages**	**Disadvantages**
PCA	•Visualize high-dimensional data. •Reduce noise. •Independent features. •The number of components is determined by variance criteria.	•Computationally intensive. •Assumes linearity between principal components and original features. •Difficulty to express the results of the principal component projection in a straightforward manner.
Random Projection	•Fast. •Low computational power.	•Less quality than PCA. •Uses a random line of projection.
ICA	•Better overall results.	•Assumes statistically independency of the features. • Computationally expensive.
SOM	•Detection of nonlinear relationships between variables.	•Subjective nature of defining clusters and establishing relationships between variables.

2.8 Diagnostics

Diagnostics can be performed either with unsupervised learning, supervised learning or semi-supervised learning. In contrast to supervised learning that usually makes use of human-labelled data, unsupervised learning allows for modelling of probability densities over inputs

(Hinton et al., 1999). Semi-supervised learning is using large amount of unlabelled data, together with some labelled data (Zhu, 2005) and thus falls between unsupervised learning and supervised learning. Figure 27 shows examples of real-life problems in the context of supervised and unsupervised learning tasks (Yip et al., 2013).

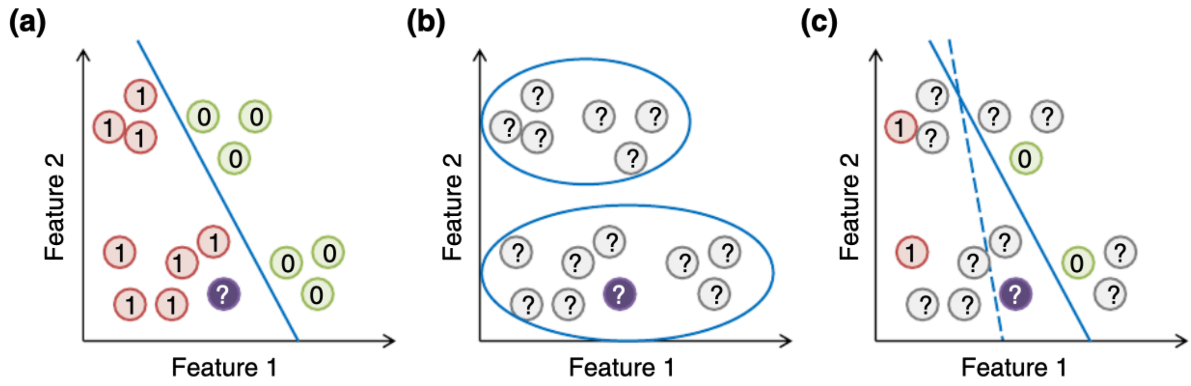


Figure 27: Examples of Supervised, Unsupervised and Semi-Supervised Learning (Yip et al., 2013)

In Figure 27, the first example is supervised learning where the model (blue line) is learned based on the positive and negative training examples, and the genomic region without a known class label (purple circle) is classed as positive according to the model. In unsupervised learning, all examples are unlabelled, and they are grouped according to the data distribution. In semi-supervised learning, information of both labelled and unlabelled examples is used to learn the parameters of the model. In this illustration, a purely supervised model (dashed blue line) classifies the purple object as negative, while a semi-supervised model that avoids cutting at regions with a high density of genomic regions (solid blue line) classifies it as positive. In the sub sections bellow an overview of the most important algorithms in each category for the purpose of diagnostics will be presented.

2.8.1 Unsupervised Learning

Starting with the unsupervised learning algorithms, there are a number of them that can be used for fault detection (Amruthnath et al., 2018), namely: K-Means, Gaussian Mixture Models, Hierarchical Clustering and Density-based spatial clustering of applications with noise (DBSCAN).

K-means is the most used algorithm for clustering (Fahim et al., 2006). The K-means algorithm classifies the data into K different independent clusters through an iterative, converging process (Hossain et al., 2019). Figure 28 shows the steps of the k-means algorithm (Wang et al., 2011).

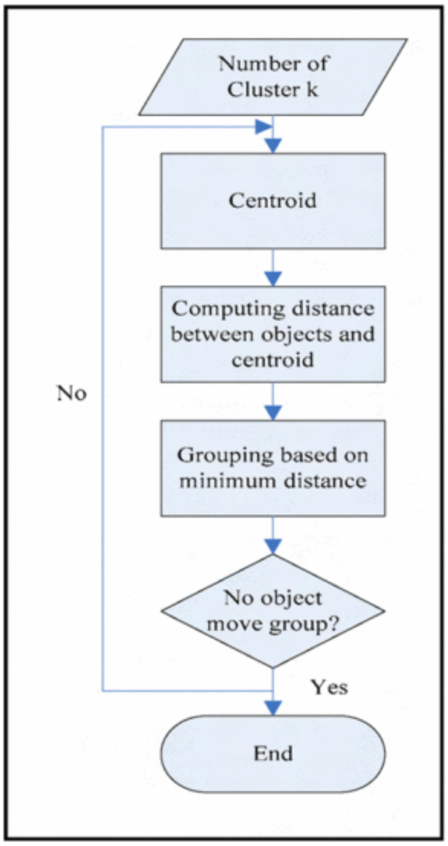


Figure 28: Steps of k-Means Algorithm (Wang et al., 2011)

Firstly, a K value is selected by the user where K is the number of clusters to be formed. First k data objects are selected as initial cluster centres. Then the Euclidean distance between each data point is compared to the nearest centre of a K centroid (Fahim et al., 2006). Finally, all the data points are used to create some group and this process will be continuing until minimum. Figure 29 shows an illustration of k-means clustering (Xu et al., 2011).

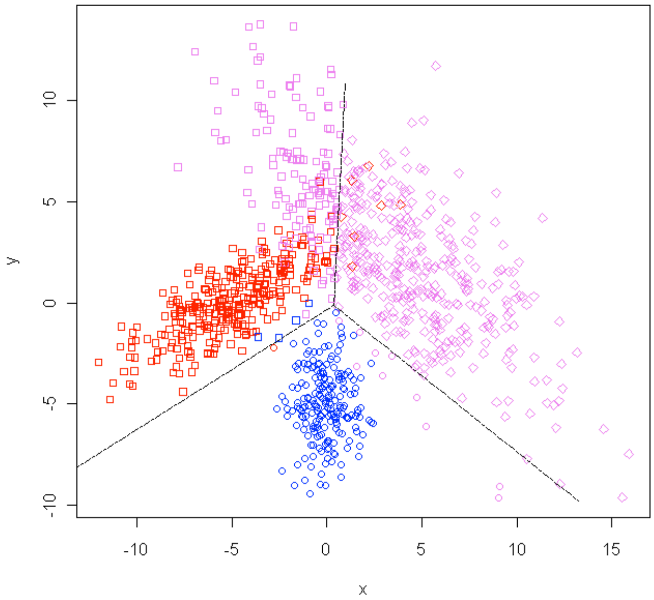


Figure 29: Example of K-Means Clustering (Xu et al., 2011)

From Figure 29 it can be seen that the number of clusters has been selected to be 3 and thus the algorithm has produced the decision lines accordingly to separate the data into three clusters. K-means algorithm has many advantages such as simple mathematical ideas, fast convergence, and easy implementation (Li et al., 2017). A fundamental problem of the k-means algorithm is that requires the number of clusters to be defined beforehand, which is responsible for different cluster shapes and outlier effect (Ahmed et al., 2020). In addition, k-means performs poor if the data are not uniform (circular or spherical) (Schelling et al., 2020). Furthermore, the k-means algorithm does not guarantee finding the optimal solution because it converges to the local minima (Hill-climbing algorithm) instead of the global minima (Komarasamy et al., 2013). Finally, k-means require the user to specify the number of clusters beforehand (Jung et al., 2003).

One example of k-means used for fault detection is for Rolling element bearing fault detection (Yiakopoulos et al. 2011). Wu et al. (2010) presents the development of an algorithm based on K-Means clustering and probabilistic neural network for classifying the industrial system fault. According to the authors, the proposed algorithm not only provides an accepted degree of accuracy in fault classification under different fault conditions and the result is also reliable. Majid et al. (2012) developed a fault detection and diagnosis system of complex processes usually involve large volumes of highly correlated data. he results of applying the clustering technique on real data sets show that the boundary of each class of faults can be identified.

Moving on to hierarchical clustering, a hierarchical clustering algorithm applied to a data set produces a series of nested partitions, usually designated by hierarch (Sousa et al., 2014). A hierarchy is a complex and difficult structure to interpret, so that, it is usual to post-process a hierarchy to find the best partition in it. The post-processing consists in cutting off the dendrogram through horizontal lines at determined levels. Figure 30 shows a dendrogram showing the result of the hierarchical clustering (Tseng et al., 2013).

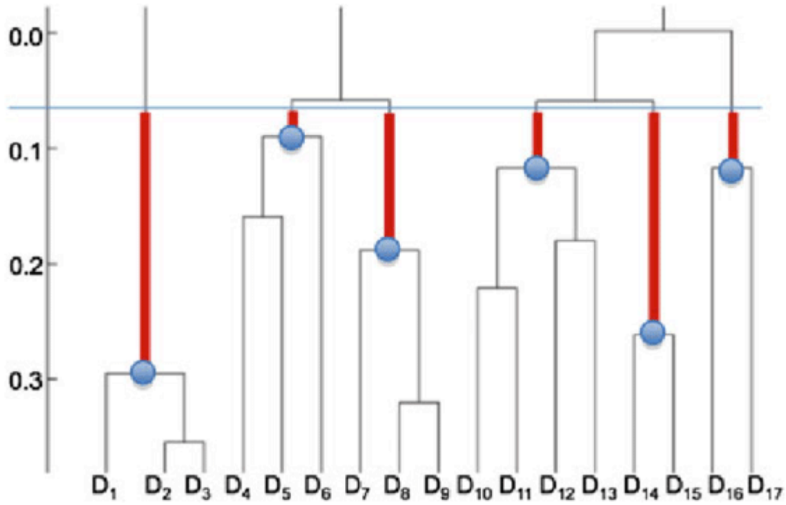


Figure 30: A Dendrogram (Tseng et al., 2013)

In Figure 30, when a threshold (e.g., 0.07) is set and then the dendrogram can be split into several clusters as marked by the dots at their roots. There are three aggregation methods, namely, single-linkage (SL), complete-linkage (CL) and average-linkage (AL). Hierarchical clustering (single-link) starts by using each point as a cluster then calculates the distances between the points (the two closer points between the clusters) and connects the ones closer to each other. Then it checks the distance between clusters and remaining points. Finally, by looking the dendrogram that has been created we can understand which is the best number of clusters. The Complete link starts exactly the same as the single link, but it uses as distances the furthest points of the clusters (Murtagh et al., 2011). It is considered better than the single link as it gives us compact clusters. The average link takes the distances for every point between the two clusters and take the average. Finally, the ward's method takes the distance from the centre of each clusters (Lee et al., 2014).

One advantage of hierarchical clustering is that there is no need to define number of clusters in advance in contrast with k-means (Bhagat et al., 2016). In addition, dendrograms provide an additional ability to visualize (Bisson et al., 2012). On the other hand, hierarchical clustering scales poorly in both memory and computing time with increasing (Embrechts et al., 2013). Also, hierarchical clustering is sensitive to noise and outliers (Ros et al., 2019).

Fault detection of fault component in wide area backup protection system can be performed successfully by hierarchical cluster analysis and calculation (Zhang et al., 2009). Zhang et al. (2012) presented sensor fault detection for industrial systems using a hierarchical clustering-based graphical user interface. The authors shown through use of real-time operational data, that in operation sensor faults can be detected and identified by the hierarchical clustering-based graphical user interface.

Finally, Gaussian mixture models (GMM) is a parametric probability density function represented as a weighted sum of Gaussian component densities (Reynolds, 2009). Instead of hard assigning a data point to a particular Gaussian component, it assigns probability of a Gaussian component belonging to a data point. GMM parameters are estimated from training data using the iterative Expectation-Maximization (EM) algorithm or Maximum A Posteriori (MAP) estimation from a well-trained prior model (McLachlan et al., 2008; Shi et al., 2011). EM is the most popular technique used to determine the parameters of a mixture with a priori given number of components (Maretic et al., 2014). EM is an iterative technique for maximum likelihood estimation where the maximum likelihood of the data increases with each subsequent iteration, meaning it is guaran-teed to converge (Dempster et al., 1977). Figure 31 shows the steps of EM (Kumar, 2018).

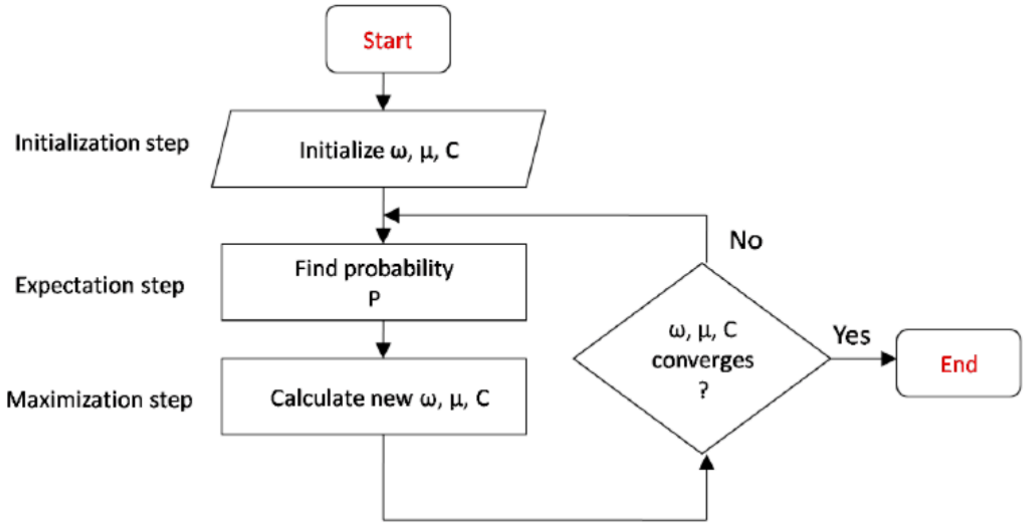


Figure 31: Expectation Maximization Algorithm (Kumar, 2018)

From Figure 31 it can be seen that EM consists of two steps, namely: Expectation step and Maximization step. After each step the parameters of the Gaussian function are calculated, such that the likelihood of the model approaches a local maximum. Figure 32 shows an example of GMM with 4 components (Takagi et al., 2009).

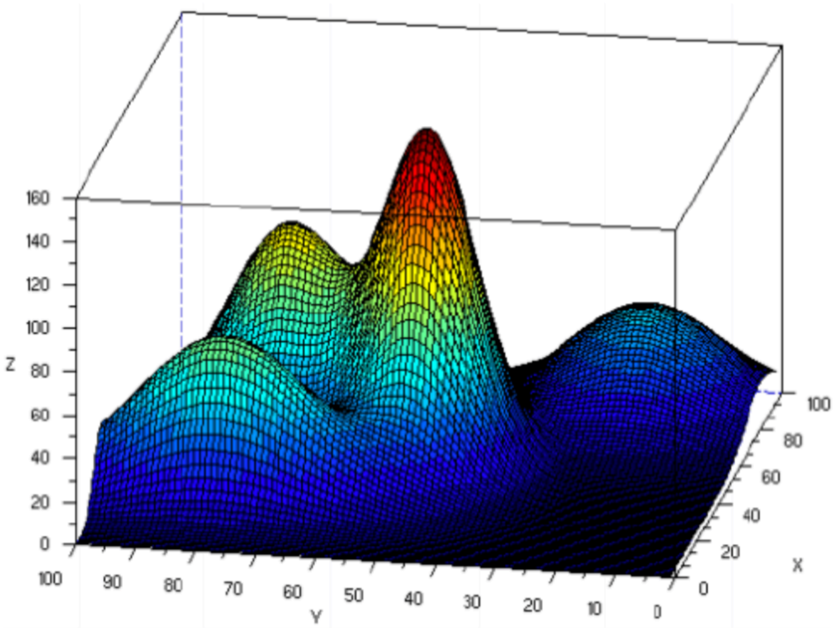


Figure 32: Example Gaussian Mixture Model (Takagi et al., 2009)

One advantage of using GMM is that it is a soft clustering technique, meaning that points can be part of more than one cluster and thus increases the flexibility of the model (Enami et al., 2012). In addition, GMM provides more cluster shape flexibility, and this increases the accuracy of the model (Greve et al., 2015). On the other hand, GMM is sensitive to initialization values (Li et al., 2018). Furthermore, GMM is possible to converge to a local optimum. (Zhang et al., 2008). Finally, GMM has a slow convergence rate. (Park et al., 2009).

The effectiveness of the GMM framework can be seen on an industrial fault simulator of rotary machine (Zhou et al. 2017). Torres et al. (2014) suggested gaussian mixture models approach for multiple fault detection. The author a presented a case study using the DAMADICS benchmark, where this approach was validated. Yu et al. (2009) proposed a machine fault diagnosis based on Gaussian mixture model. The authors concluded that experimental results based on the application on bearing fault diagnosis have shown that GMM can reliably diagnose not only the type of bearing faults, but also the degree of fault severity that are associated with incipient faults, moderate faults, and severe faults. Table 13 shows a summary of the unsupervised learning algorithms.

Table 13: Summary of Unsupervised Learning Algorithms

Clustering Models	Advantages	Disadvantages
K-Means	•Most commonly used algorithm. •Performs better when uniform data (spherical). •Performs good with a large number of clusters. •Performs good if the number of clusters is known.	•Local minima (Hill-climbing algorithm). •Performs poor if the data are not uniform (circular or spherical). •Need to specify number of clusters.
Hierarchical Clustering	•Resulting Hierarchical representation can be very informative. •Provides an additional ability to visualize. •Don't need to specify number of clusters.	•Sensitive to noise and outliers. •Computationally intensive.
Gaussian Mixture	•Soft clustering. •Cluster shape flexibility.	•Sensitive to initialization values. •Possible to converge to a local optimum. •Slow convergence rate.

2.8.2 Supervised Learning

Moving on, the most used supervised learning algorithms, for fault diagnosis are (Lo et al., 2019): Bayesian Networks (BN), Artificial Neural Network (ANN), Support Vector Machines (SVM), Hidden Markov Models (HMM) and Decision Tress (DT) & Random Forest (RF). Starting with Bayesian Networks, they are a schematic representation of the Bayes theorem of posterior probabilities (Daly et al., 2011). In order to use BN, there is the need to specify the probabilities of each node and the network structure (Ademujimi et al., 2017). In some cases, a Naïve Bayesian Network can be used where there is the assumption that the features are conditionally independent given class (Rish, 2001). Then, the problem can be simplified because the probability of two events happening together is just a product now. Surprisingly enough this method has hold on in many cases (spam e-mails classifier), but it becomes weaker the stronger the dependency between the features is and the complexity of the data (e.g., in the case of machinery data in shipping). Figure 33 shows a Bayesian network for automobile troubleshooting problem (Heckerman et al., 1970).

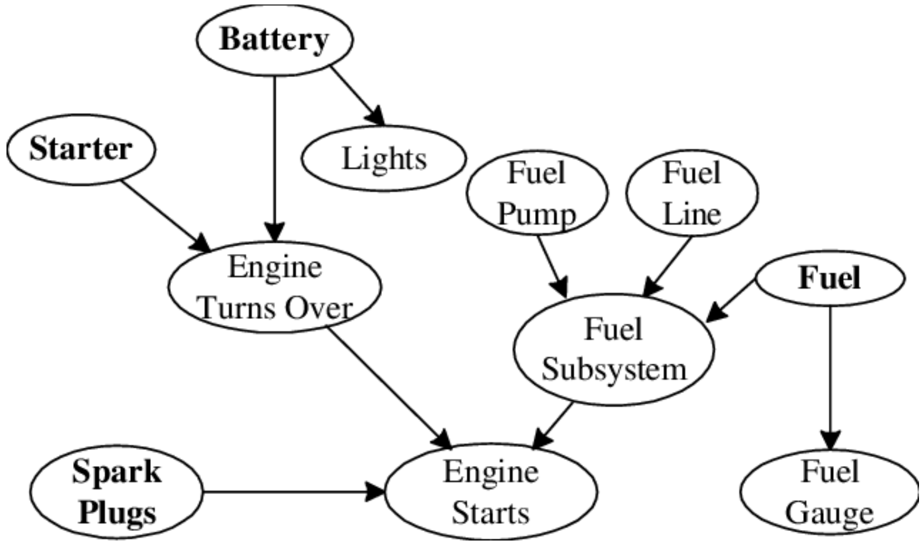


Figure 33: Example of a Bayesian network (Heckerman et al., 1970)

From Figure 5, the component nodes of the network are: “Battery”, “Starter”, “Spark Plugs”, “Fuel Pump”, “Fuel Line” and “Fuel”. The device node is “Engine Starts” and in this network it is decomposed to into the subsystems “Engine Turns Over” and “Fuel Subsystem”. The “Engine Starts” exhibits a noisy max relationship with its parents, where “Engine Starts” is Abnormal, if either subsystem is Abnormal. To show the cause/effect relationship between nodes different authors choose different approach such as: failure mode and effect analysis (FMEA), cause and effect diagrams, fault- tree analysis and variation sensitivity matrix (De et al, 2010; Pradhan et al, 2007; Nguyen et al, 2016; Liu & Jin, 2013). Data obtained from sensors can then be used to generate the conditional probabilities of the network (Ademujimi et al., 2017). BN provide a method for avoiding overfitting of data (Heckerman, 1995) and can show good prediction accuracy even with rather small sample sizes and they can be easily combined with decision analytic tools to aid management (Kuikka et al., 1999; Jensen, 2001). In addition, they provide a natural way to handle missing data, they allow combination of data with domain knowledge, they facilitate learning about causal relationships between variables (Uusitalo, 2007). On the other hand, the main challenge of training a BN is in the construction of the tree structure and several methods including expert opinion have been proposed to mitigate this challenge (Gheisari, 2016). In addition, their ability to deal with continuous data is limited (Jensen, 2001), and such data generally needs to be discretized, which may cause certain difficulties.

Zhang et al. (2017) explored fault detection and diagnosis using Bayesian-network inference. The authors showed three cases to verify the method. The maximum probability fault can be found according to the way of BN reference results, which demonstrates the feasibility of the developed method. Verron et al. (2008) proposed a fault detection with Bayesian Network. The detection is viewed as a classification task like the discriminant analysis, which can be transposed in a Bayesian network. The authors provided an application on the Tennessee Eastman Process is given in order to demonstrate the approach.

Moving on to ANNs, they are artificial adaptive systems that are inspired by the functioning processes of the human brain (McCulloc et al., 1943). Figure 34 shows biological neuron in comparison to an artificial neural network: (a) human neuron;(b) artificial neuron; (c) biological synapse; and (d) ANN synapses (Suzuki, 2013).

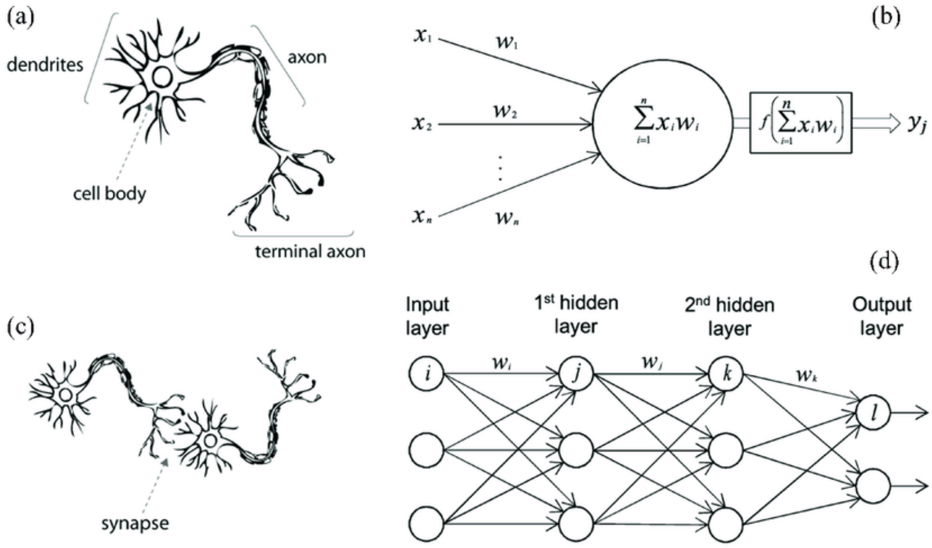


Figure 34: A Biological Neuron in Comparison to an Artificial Neural Network (Suzuki, 2013)

From Figure 34, the base elements of the ANN are the nodes, also called processing elements (PE), and the connections. Each node has its own input, from which it receives communications from other nodes and/or from the environment and its own output, from which it communicates with other nodes or with the environment (Grossi et al., 2008). Finally, each node has a function through which it transforms its own global input into output. Each connection is characterized by the strength with which pairs of nodes are excited or inhibited. Positive values indicate excitatory connections, the negative one's inhibitory connections (McClelland, 1986; Anderson et al., 1988). The significance of each input is multiplied by a weight and together with bias they get to cell body, the processing element. In the first step the multiplied inputs are summed by summation function and in the second step they are propagated by transfer function to an output (Krenek et al., 2016).

ANNs reveal the hidden rules of a problem (Von der Malsburg, 1973; Willshaw et al., 1976). This makes ANNs particularly useful in solving a problem for which it is unknown how the data are related to one another. In addition, ANNs are adaptive and dynamically discover the fuzzy rules that connect various sets (Kohonen et al., 1984; Carpenter, 1988). This means that if ANNs receive after training new and different data, ANNs will adjust their rules in accordance, integrating the old data with the new, and they can do this without any external instruction. On the other hand, work with artificial neural networks is always connected with processing of high volume of data which requires advanced commercial and open-source software tools (Krenek et al., 2014). In addition, when initialize the ANNs is difficult to know how many neurons and layers are necessary (Sidda et al., 2017). Finally, ANNs may be considered as 'black boxes' and thus it is not easy to understand and to explain its process to the users (Idri et al., 2002).

Artificial neural networks have shown promising results as a robust tool for evaluating data in order to support predictive maintenance activities. There exist a lot of papers focused on application of ANN in maintenance. Mainly multi-layer perceptrons (MLP) are used for fault diagnosis of bearings, induction motors, non-destructive evaluation of check valve performance and degradation and in robotic systems (Meireles et al., 2003). Some examples are: Sun et al. (2013) has developed self-organising map for monitoring multi-equipment health management system with fault detection and a real-time monitoring. In Raptodimos et al. (2018) artificial neural networks were used to predict the main engine cylinders exhaust gas temperature of a Panamax size container ship.

Moving on, a normal classification linear line just splits the data points while the Support Vector Machine (SVM) algorithm tries to find a boundary line that separates the data and forces this line to be as far away from the points as possible (Yue et al. 2003). So, two equidistant parallel lines are used to the main line and try to maximize the distance between these two lines or else the margin. In SVM points inside the margin are consider misclassified and add up to the classification error (Hamel, 2009). Figure 35 shows an illustration of SVM (Manjrekar et al., 2019).

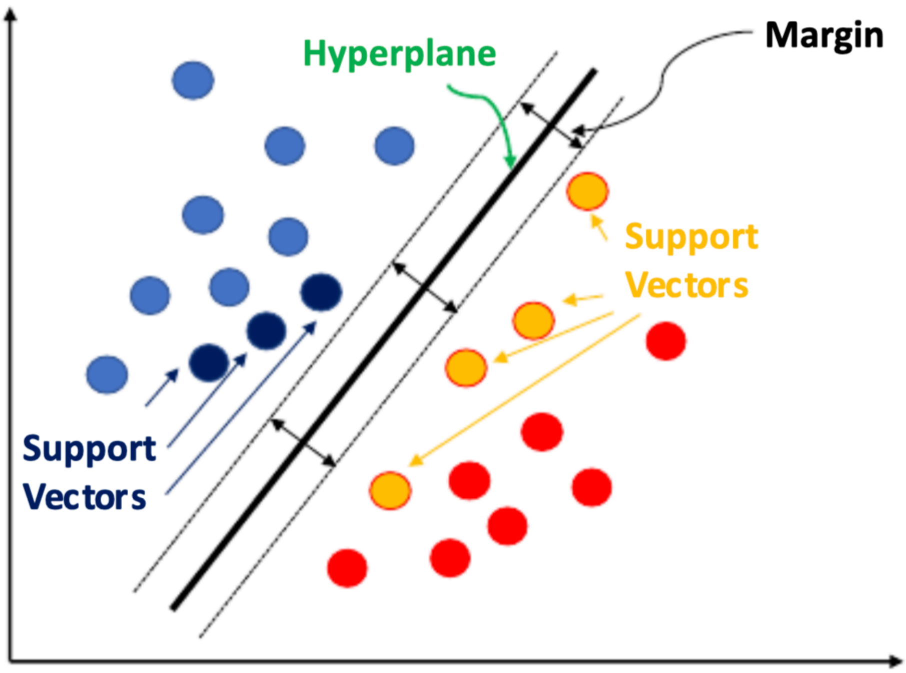


Figure 35: SVM Visualization (Manjrekar et al., 2019)

Consider an example of data clusters in Figure 35, a hyperplane can be drawn to separate the two clusters and maximize margin width. Finding the optimum location of the hyperplane that maximizes margin width is a complex optimization problem. The points close to the margin are called support vectors and these points are critical in the determination of the hyperplane. When a line is not enough to separate points the Kernel trick is used in SVM: Some dimensions to the data are being added in order to find a high dimensional surface, project it down and get curves to separate the data (Tian et al. 2014). Figure 36 shows the non-linear SVM classifier with the kernel trick (Moreira, 2011).

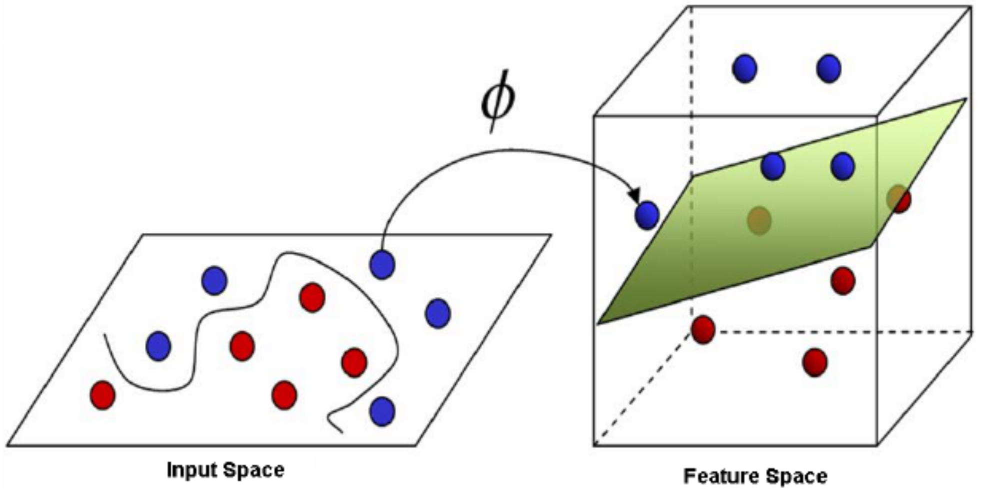


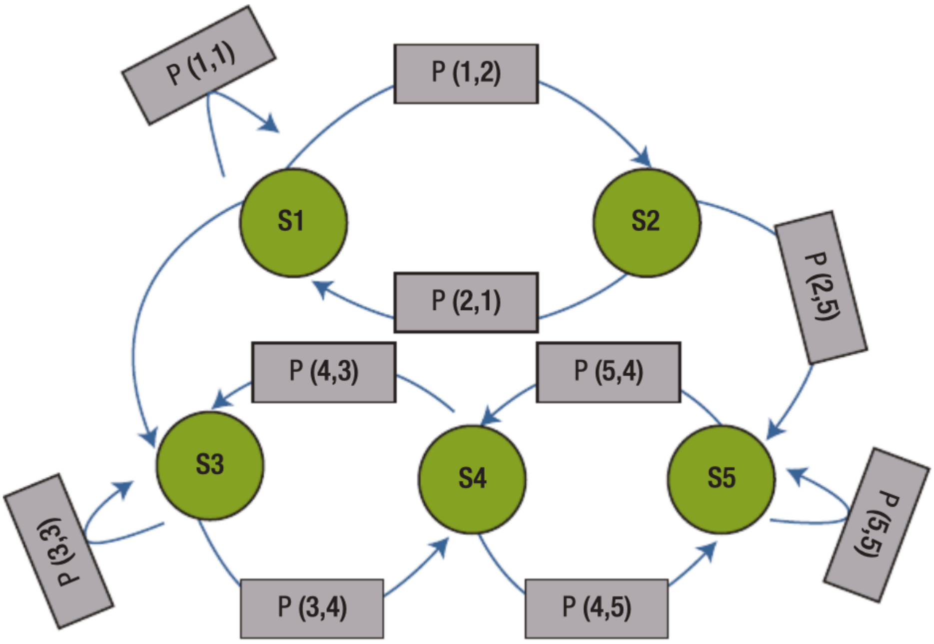
Figure 36: The Kernel Trick (Moreira, 2011)

With a polynomial kernel the curves can be circles, hyperbolas, parabolas and in general more functions can be used to build complex boundaries. There is also the Radial Basis Functions (RBF) Kernel which uses radial basis functions to help separate the data points.

Linear SVMs are easy to implement even in massive data sets and they are easy to interpret and understand (Doumpos et al., 2007). In addition, a further characteristic of SVM is the efficient learning with small amount of data, due to their simpler, more effective architecture and learning procedure (Pantazi et al., 2016). Furthermore, due to its strong theoretical foundation, good generalization capability, low sensitivity to the curse of dimensionality (Hughes, 1968) and ability to find global classification solutions, SVMs is usually preferred by many researchers over other classification paradigms (Fei et al., 2020). On the other hand, the selection of the kernel function parameters is challenging. (Tharwat et al., 2019).

Application of SVM exist in fault localization, although it is not as common as BN and ANN (Widodo et al., 2007). Some examples of using one-class SVM for decay detection are: detection of decay of marine propulsion plant (Tian et al. 2019) and estimate ship systems condition with noon-report data (Lazakis et al. 2018). Batur et al. (2002) used support vector machines for fault detection. The authors used a conventional heat exchanger dynamic to illustrate the technique. Also, Yin et al. (2014) did a study on Support Vector Machine-Based Fault Detection in Tennessee Eastman Process. By comparing the indices of detection performance, the SVM technique showed superior fault detection ability to the PLS algorithm.

Next, the HMMs incorporates the Markov principal saying that any future state of a system depends only on the current state and not the past events (Beal et al. 2002). It estimates the probability distributions of state transitions and that of the measurement outputs in a dynamic process, given unobservable states of the process (Ademujimi et al., 2017). Figure 37 shows an example of five states HMM (Awad et al., 2015).



(Transition Matrix) P =

P(1,1)	P(1,2)	P(1,3)	P(1,4)	P(1,5)
P(2,1)	P(2,2)	P(2,3)	P(2,4)	P(2,5)
P(3,1)	P(3,2)	P(3,3)	P(3,4)	P(3,5)
P(4,1)	P(4,2)	P(4,3)	P(4,4)	P(4,5)
P(5,1)	P(5,2)	P(5,3)	P(5,4)	P(5,5)

Figure 37: Example of five states HMM (Awad et al., 2015)

From Figure 37, a system may be described at any time as being in one of the states S1, S2, S3, S4 and S5. When the system undergoes a change from state Si to Sj at regular time intervals with a certain probability pij, this can be described by a simple stochastic process, in which the distribution of future states depends only on the present state and not on how the system arrived at the present state. The matrix P, with elements pij, is called the transition probability matrix of the Markov chain.

HMM has the ability to efficiently model any kind of data that contains spatial-temporal relations (Dadgar, 2015). HMM are very popular to use because of their efficiency in estimating parameter and doing inferences (Alghamdi, 2016). On the other hand, training process of HMM is usually computationally intensive (Zhang et al., 2006).

Chen et al. (2012) used HMM for health condition monitoring on a bearing dataset. Furthermore, Prakash et al. (2017) for deterioration of low-speed rolling elements. Smyth (1994) proposed a HMMs for fault detection in dynamic systems. The model was validated on a real-world fault diagnosis problem and it was shown that Markov modelling in this context offers significant practical benefits. Wang et al. (2016) used HMM based fault detection approach for a multimode process. A numerical simulation example and the Tennessee Eastman chemical process was utilized to show that the authors proposed approach is effective.

Table 14: Summary of Supervised Learning Techniques

Technique	Advantages	Disadvantages
BN	•Avoid overfitting of data. •Show good prediction accuracy even with rather small sample sizes. •Easily combined with decision analytic tools to aid management. •They provide a natural way to handle missing data, •They allow combination of data with domain knowledge. •They facilitate learning about causal relationships between variables.	•The main challenge of training a BN is in the construction of the tree structure. •Their ability to deal with continuous data is limited.
ANN	•ANNs reveal the hidden rules of a problem. •Are adaptive and dynamically discover the fuzzy rules that connect various sets.	•Requires advanced commercial and open-source software tools. •Difficulty to know how many neurons and layers are necessary. •It is not easy to understand and to explain its process to the users.
SVM	•Easy to implement even in massive data sets. •Easy to interpret and understand.	•Selection of the kernel function parameters is challenging.

	•Efficient learning with small amount of data. •Strong theoretical foundation. •Good generalization capability. •Low sensitivity to the curse of dimensionality. •Ability to find global classification solutions.	
HMM	•Efficiently model any kind of data that contains spatial-temporal relations •Efficiency in estimating parameter and doing inferences.	•Training process is usually computationally intensive.

2.8.3 Semi-Supervised Learning

From all the approaches, semi-supervised learning has several advantages (Heras et al., 2014), namely:

- Engineers are in full control to specify what should be consider as nominal.
- Repeated anomalies can be detected since they are not in the nominal set.
- Since no assumptions are made about the possible behaviour of the anomalies, any anomalous behaviour can be potentially detected.

Some of the most used semi-supervised approaches for fault detection are: Local Outlier Factor (LOF) and one-class support vector machine (OC-SVM).

Local Outlier Factor (LOF) is an algorithm first presented in 2000 by Breunig et al. (2000) and its purpose was to identify local outliers in a set of data. The algorithm doesn't just categorize the data in a binary way namely, outliers or not but rather assigns to each data point a local outlier factor (LOF) score which shows the degree of a data point being an outlier, something similar to density-based clustering (Breunig et al., 2000). The authors first found the smallest hypersphere cantered at the given samples that contained the k-nearest neighbours and the LOF was calculated by dividing k by the volume of the hypersphere (Song et al., 2017). LOF creates a decision frontier from the decision function learned from a training dataset which consists of ‘normal’ data and then assess if a new data-point is a novelty or not (scikit-learn, 2019). The desired decision boundary can be obtained by varying the LOF threshold (Song et al., 2017). Figure 38 shows an example of the local density between data points.

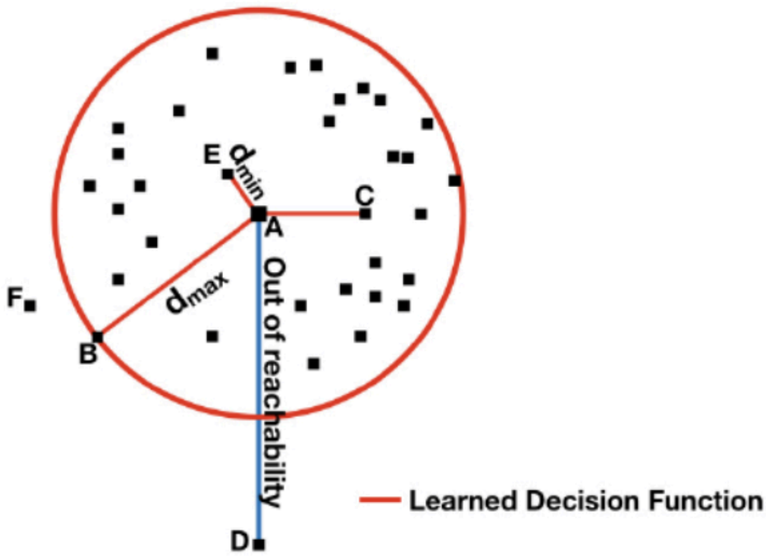


Figure 38: Local Density (Song et al., 2017)

As shown in Figure 38, point D has a lower local density than point A. If the red circle is the learned decision function from the training data, the distance between point A and B is the maximum distance. Thus, point D is out of the reachability distance and considered as an outlier. The main advantage of the LOF algorithm is that it does not need an assumption for the data distribution and can be applied to different data types. In addition, LOF works very well in many cases, while often outperforming the competitors, for example in network intrusion detection (Lazarevic et al., 2005) and on processed classification benchmark data (Campos et al., 2015). On the other hand, the performance of LOF depends on the parameter k which is defined as the least number of the nearest neighbours in the neighbourhood of an object (Papadimitriou et al., 2003). However, in LOF, the value of k is determined based on the average density estimate of the neighbourhood, which is statistically vulnerable to the presence of an outlier. Hence, it is hard to determine an appropriate value of this parameter to ensure the acceptable performance in the complex and large databases (Gao et al., 2011).

Ding et al. (2018) proposed a local outlier factor-based fault detection and evaluation of photovoltaic system. The results of experiments reveal that the LOF has good performance in fault detection and fault degree evaluation in different scales of the PV systems. Zhao et al. (2014) suggested fault experiments in a commercial-scale PV laboratory and fault detection using local outlier factor. The authors concluded that in experimental results, the proposed method demonstrates several advantages over traditional PV monitoring systems, such as simplicity, quick response, easy implementation and no requirement of weather information.

On the other hand, OC-SVM was introduced by Schölkopf et al. (1999) as a support vector method for novelty detection. The idea behind the OC-SVM is to describe the target class by a function that maps most part of it to a region where the function is nonzero. To this end, the origin is treated as the only available member of the non-target class (outlier), and then the problem is solved by finding a hyperplane with maximum margin separation from the origin (Munoz et al., 2010). Figure 39 shows a. example of the learned frontier for OC-SVM (Bara et al., 2014).

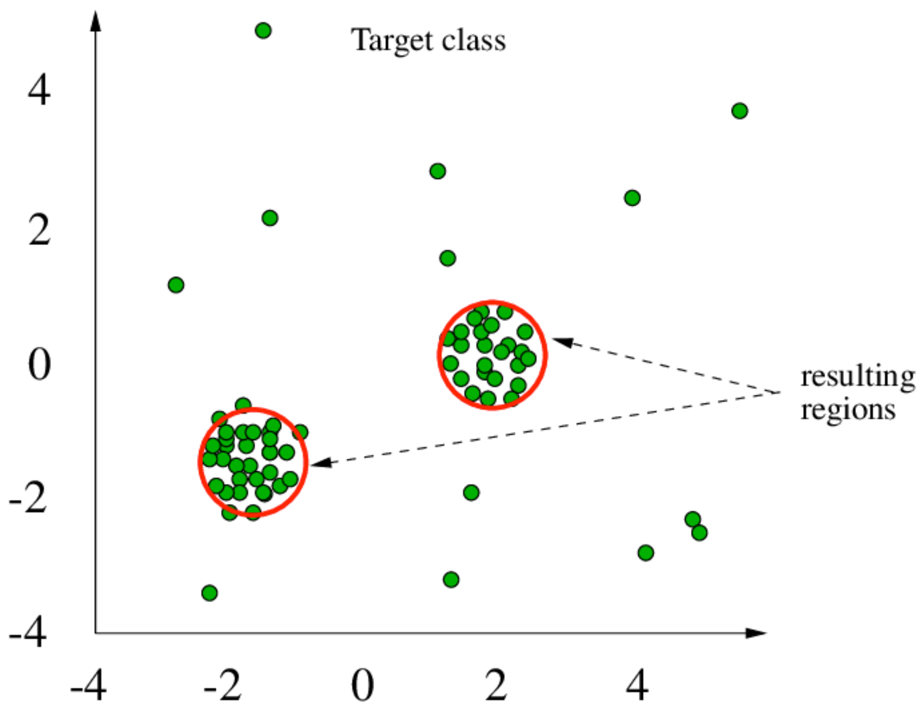


Figure 39: Learned Frontier for OC-SVM (Bara et al., 2014)

In Figure 39 it can be seen the learned frontiers of a set of two-dimensional data points. These two regions are the result of mapping a higher dimension hyperplane generated using a Radial Basis Function (RBF) kernel back to a two-dimensional representation. Although all of the points are part of the same class (normal) only the ones inside the frontier will be classified as normal, the rest being abnormal.

OC-SVMS have a simple geometric representation, that is, they determine a few parameters of the normal data model only. For example, if the normal data is enclosed in a sphere, then one-class SVM tends to determine the centre and radius of the sphere only (Shahid et al., 2013). In addition, OC-SVM, like LOF does not make any assumptions about the data distribution (Liu et al., 2019). On the other hand, OC-SVM does not perform well with reduces samples. The inefficient use of the OC-SVM classifier for reduced samples is due to the hard threshold used in the decision function for accepting a sample, which should be higher than zero (Guerbai et al., 2014). In addition, OC-SVM is sensitive to outliers included in the training data set (Yin et al., 2014). Table 15 shows a summary of the Semi-Supervised techniques.

Table 15: Summary of Semi-Supervised Learning Techniques

Technique	**Advantages**	**Disadvantages**
LOF	• Does not need an assumption for the data distribution. • Works very well in many cases, while often outperforming the competitors	•It is hard to determine an appropriate value of the parameter k
OC-SVM	• Have a simple geometric representation • Does not need an assumption for the data distribution.	•Poor performance when reduced sample. •Sensitive to outliers

2.9 Prognostics

Compared to diagnostics, the literature of prognostics is much smaller. Prognostics is defined as the process of predicting the Remaining Useful Time (RUL) at which a component will no longer perform a particular function (Okoh et al., 2014). Prognostics results are used to support proactive decision making. ISO definition for fault prognostics can be found in (ISO 13381-1:2005, 2015). Figure 40 shows an example of prognostics (Atamuradov et al., 2017).

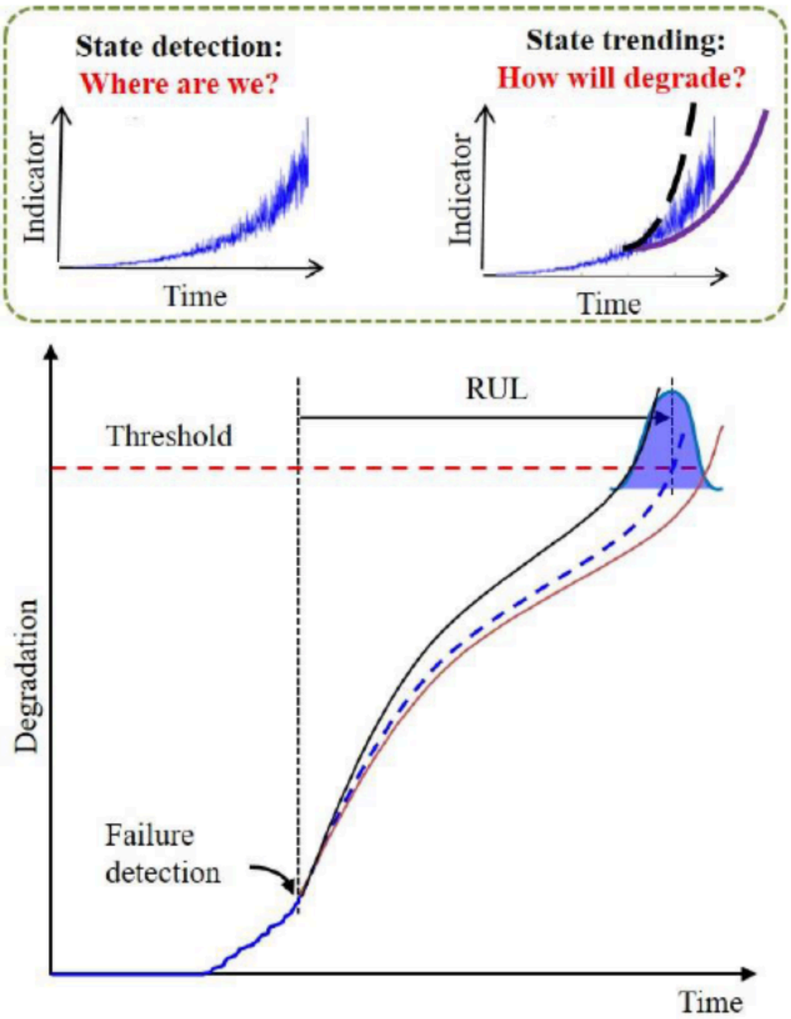


Figure 40: Prognostics (Atamuradov et al., 2017)

Within the field of maintenance problems, Artificial Neural Networks (ANNs) and neuro-fuzzy systems (NFs) have successfully been used to support the detection, diagnostic and prediction processes, and research works emphasize on the interest of using it (Dragomir et al., 2009). Zhang and Ganesan (1997) used SOM for multivariable trending of the fault development to estimate the residual life of a bearing system. Wang and Vachtsevanos (2001) applied dynamic NN to predict the fault propagation process and estimate the RUL as the time left before the fault reaches a given value. Yam et al. (2001) applied a recurrent NN for predicting the machine condition trend. Dong et al. (2004) utilised a grey model and a Back Propagation NN to predict machine condition. Wang et al. (2004) compared the results of applying recurrent NN and neural–fuzzy inference systems to predict the fault damage propagation trend. Chinnam and Baruah (2004) presented a neural–fuzzy approach to estimating RUL for the situation where no failure data and no specific failure definition model are available, but domain experts with strong experiential knowledge are available. Tian (2009) used an artificial neural network (ANN) based method for achieving more accurate remaining useful life prediction of equipment subject to condition monitoring.

Other approaches are Bayesian prediction method and Support Vector Machines (SVM) which makes use statistical estimates of condition for limited samples to define predictive learning base (Yang et al., 2012). Okoh et al. (2014) stated that the study so far confirms that data collected from sensors can be translated through ANN to predict RUL of an asset.

Hoa (2017) used PCA using data from the using the data of the PHM'08 Challenge Problem in order to calculate the RUL. Hsu et al. (2018) used deep learning in order to predict the RUL of aero-propulsion engines and concluded that the evaluated results are compared with those of related methods, namely the methods using multi-layer perceptron (MLP), support vector regression (SVR), relevance vector regression (RVR) and convolutional neural network (CNN) and found it to be superior.

2.10 Database Types for Data-Driven CBM

Before finishing the discussion about the data driven CBM systems there is one important element to taggle, which is the database types. There are two types of databases to consider, namely: Relational databases and Object-Oriented Databases (OOD). Up until now the Relational Database has been the most commonly used data model and the vast majority of current database systems for condition monitoring are based on this model (Atzeni et al. 1999, Silberschatz et al. 2011). This model was proposed by Edgar F. Codd (1970). In a relational database the data are grouped into relations, called also entities, that are often perceived as tables. Each relation is made up of attributes (columns) and tuples (rows). Each tuple contains a set of interconnected atomic data that describes the properties of an object or an event from the real world (Elmasri & Navathe 2010). Each tuple in a given entity is uniquely characterized by one attribute (or a combination of several attributes) called a primary key which determines the corresponding tuple. The values of the primary key in a single relation are unique. The relationship between two relations is accomplished by a common attribute, which in the first relation represents the primary key, and in the second foreign key. The number of value quotes in the second relation determines also the type of relationship between the two relations, for in-stance: one-to-one, one-to-many and many-to-many (Kraleva et al., 2018). The latter type of relationship is realized through an associative relation in which the primary keys of the relationship are encountered (Date, 2012). Figure 41 illustrates an Entity Relation-ship diagram (ER-diagram) and its concrete tables within a relational database management system (RDBMS) (Sint et al., 2009).

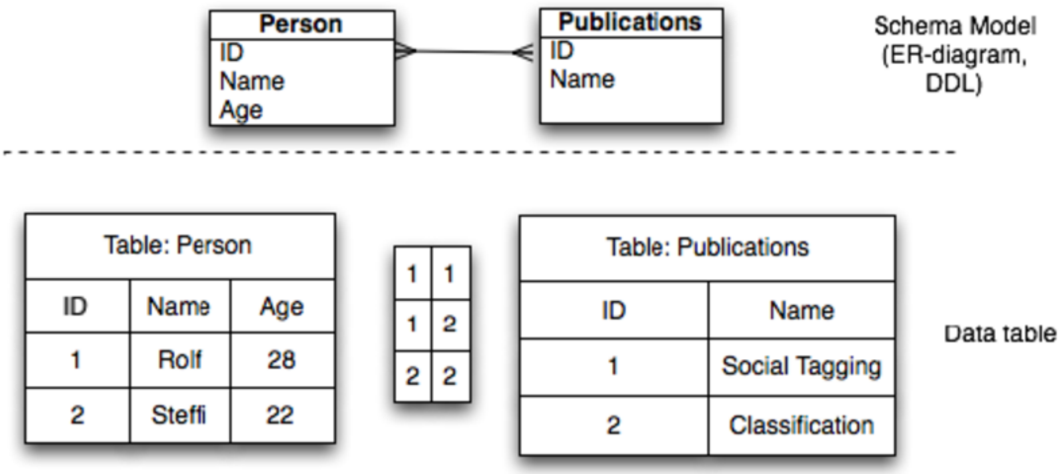


Figure 41: Sample Table in a Relational Database System (Sint et al., 2009)

An advantage of relational databases is that they have a strong foundation and well documented literature about SQL which is the only data manipulation language that all relation databases use (Barbierato et al., 2014). In addition, relational databases offer stronger consistency with the strict schema (Singh, 2016).

On the other hand, one downside of using relational databases is scalability because it depends on the vertical scalability (by adding more hardware resources like RAM, CPU, etc.) which is very costly and actually impractical for the reason of hardware limitation (Mohamed et al., 2014). In addition, the best relational databases are proprietary and therefore, require great amounts of investment from organizations and individuals that want to benefit from their advanced features (Phiri & Kunda 2017). Additional hardware for upgrades also adds other additional costs. This makes Relational Databases to be an expensive approach to data storage (Zaki, 2014). Also, Relational Databases usually suffer from single point of failure even for very powerful servers (Moniruzzaman & Hossain, 2013). Furthermore, relational databases require much more time to process information making them slow (Okman et al., 2013). Moreover, relational databases create complex data in circumstances where data to be stored by users is difficult to convert into tables (Abourezq & Idrissi, 2016). Finally, machine learning algorithms are written in an Object-Oriented language and there is a commonly known problem between Relational Databases and Object-Oriented programming called Object-Relational Impedance Mismatch (Ireland & Bowers, 2015). This mismatch can occur when an Object-Oriented program uses a Relational Database for persistence.

Moving on to OOD, these databases are a type of NoSQL databases and store data in a non-tabular way, different from a relational database (Li, 2018). Historically, OODBs developed first as an approach to add persistence seamlessly into object-oriented programming languages (OOPLs) (Dietrich & Urban, 2011). A fundamental concept in OOD is an object. An object consists of a private data structure, with a public interface. Objects are organized into classes, which can contain within them methods to carry out the operations on the objects (Garvey & Jackson, 1989). An object is an abstract concept, generally representing an entity of interest in the enterprise to be modelled by a database application. An object has state and behaviour. The state of an object describes the internal structure of the object where the internal structure refers to descriptive properties of the object.

Viewing a person as an object, the state of the object might contain descriptive information such as an identifier, a name, and an address. The behaviour of an object is the set of methods that are used to create, access, and manipulate the object (Ogunlere & Idowu, 2015). For example, a person object may have methods to create the object, to modify the object state, and to delete the object. The object may also have methods to relate the object to other objects, such as enrolling a person in a course or assigning a person to the instructor of a course. A method has a signature that describes the name of the method and the names and types of the method parameters. Objects having the same state and behaviour are described by a class. A class essentially defines the type of the object where each object is viewed as an instance of the class (Ajita & Patel, 2009). An object can be an instance of only one class (Banerjee et al., 1988; Beach, 1988) or an instance of several classes (Cluet et al., 1989; Dayal, 1989). A method is a specific implementation of a method signature. Figure 14 shows an example using the relational database approach and the OOD approach (Caseau, 1991).

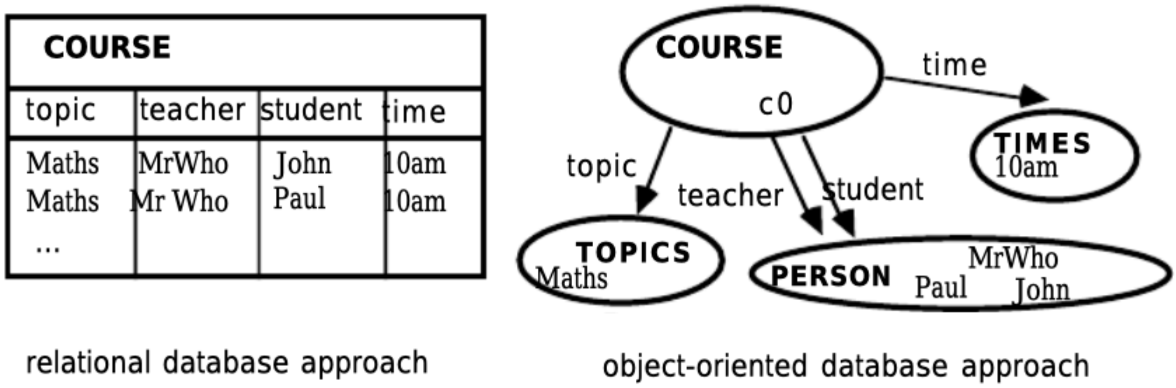


Figure 42: Object-Oriented vs. Relational Database (Caseau, 1991)

In Figure 42 each n-ary relation is usually replaced by a class in the object-oriented system, and as many binary relations as fields in the original relation are created. A well-known example is the COURSE (topic, teacher, student, time) relation that is transformed into a set (COURSE) and four binary relations: topic, teacher, student and time. By doing so, each tuple of the COURSE relation is transformed into an object.

The distributed nature of OOD makes a better choice to provide availability to users all the time even in the presence of hardware failures (Sharma, et al., 2016). In addition, OOD is cheaper as it is open source and support inexpensive upgrade (Kim, 2014). Furthermore, OOD supports both semi structured and unstructured data which is less complex (Kepner et al., 2016). Also, OOD can handle large volumes data especially in big data (Nayak et al., 2013). Finally, to offer scalability, OOD require the use of commodity server's- scaling horizontally (Singh, 2016; Sharma et al., 2016). Scaling horizontally is not significantly affected by hardware limitations because smaller, cheaper and less powerful server machines can be combined to offer higher levels of scalability instead of having one expensive server. On the other hand, having multiple ways of querying OOD, limits the number queries supported because each implementation must provide its own unique queries (Barbierato, 2014). Also, OOD leaves security to be handled by middleware and is not part of the database (Zaki, 2014).

Table 16: Summary of Database Types

Technique	Advantages	Disadvantages
Relational Databases	• Strong consistency with the strict schema. • Documented literature about SQL. •Data to be stored by users is difficult to convert into tables.	•Expensive approach to data storage. •Suffer from single point of failure. •Require much more time to process information making them slow. •Vertical scalability. •Object-Relational Impedance Mismatch.
OOD	• Availability to users all the time •Open source and support inexpensive upgrade. •Supports both semi structured and unstructured data. •Can handle large volumes data. •Scaling horizontally.	• Not uniform query language. •Leaves security to be handled by middleware.

2.11 CBM in the Marine Sector

In the marine sector data are usually being used for manually monitoring alarms and trends on-board with some cases of simple automated data diagnostic and prognostic procedures (DNV GL, 2014). These simple diagnostic and prognostic systems combined with expert knowledge are part of the first generation of CBM systems. Statistical analysis of condition monitoring data is used by most of the existing commercial CBM systems in shipping and is meant to draw population inferences from a sample and create a mathematical model of the data generation process (Bzdok, 2018). However, failure mechanisms of machinery have been proven not so easy to model since they are mostly random in nature (Nowlen et al., 1978; Broberg, 1973; MSDP, 1982; Allen, 2001). That is why the research community has turned to data-driven techniques in order to solve the problem of unplanned maintenance (e.g., Lazakis et al., 2018; Gkerekos et al., 2017; Raptodimos et al., 2016).

In more recently applications, Raptodimos et al. (2020) developed a Nonlinear Autoregressive with Exogenous Input (NARX) Artificial Neural Network for forecasting future values of the exhaust gas outlet temperature of a marine main engine cylinder however without taking into account the different operational conditions of the ship (they only take into account the different Rounds Per Minute of the main engine). In addition, Cheliotis et al. (2020) presented a novel Machine Learning (ML) and data-driven Fault Detection (FD) methodology, based Expected Behaviour (EB) modelling and Exponentially Weighted Moving Average (EWMA) control charts, and its application on ship systems. This methodology is based in using historical data from healthy condition of a machinery to train an algorithm and then using on-line monitoring to compare the healthy condition data to the current one. This approach again does not take into account the different operational profiles of a ship (healthy data of a machinery are different for different operational condition) and for that is difficult to have a practical value. Furthermore, Lazakis et al. (2016) developed am advanced ship systems condition monitoring for enhanced inspection, maintenance and decision making in ship operations. The CBM system proposed is based in Markov Chains and Bayesian Belief Networks in order to take into account interdependencies between systems and sub-systems. However, this approach requires historical data for all the machinery evolved and does not take into account all the different operational profiles that a ship can be and thus reduces significant the practical value of the CBM. Iraklis et al. (2019) presents a novel methodology for intelligent, system-level engine performance monitoring, utilising noon-report data with minimal data assumptions. Nevertheless, the proposed model requires data of diverse set of load conditions for a machinery in order to train the algorithm to perform CBM, something that is not feasible to do for all machinery on-board a ship and thus making this approach also impractical.

To sum up, the real problem that Data Scientists are facing is that most data-driven techniques proposed for CBM in shipping require huge amounts of historical data, including failure rate data, which are notoriously difficult to find. In addition, a couple of other factors also make a commercial CBM system a difficult task: The existence of different operational profiles of the ship (port manoeuvring, bad weather, cruising, etc.) and the mismatch between relational databases and the algorithms. Concerning the first one, a ship is a moving object moving in a moving environment and thus the values measured from sensors for a machinery can give very different results (e.g., for different payloads) not seen in the training dataset which can confuse a CBM system. All these problems highlight the difficulty in applying CBM in shipping, something that is also reflected by the fact that only 2% of the total percentage of classed ships have condition monitoring scheme in place (Shorten, 2012).

In this thesis, a novel CBM framework will be presented that will be able to perform a fast and reliable detection of the degradation of a machinery through the use of a Condition Monitoring Database (CMD) with an Object-Oriented (NoSQL) nature and the semi-supervised LOF algorithm. The OOD was chosen mainly because of its seaming less integration with programming languages, and thus machine learning, while the LOF algorithm because of the great flexibility that semi-supervised methods provide of not needing historical or failure data. Then with the help of a new failure model theory derived by ellipses, the system will predict the time of failure of the machinery. The proposed CBM framework will require minimum amount of memory, no historical data and no need for prior human knowledge (experts). In addition, crucial parts of the developed CBM framework will be tested namely, storing data in the CMD and then use the data for detecting the degradation of a marine propulsion system with machine learning techniques as well as prediction of failure of bearings. Finally, the whole CBM system will be used to predict maintenance for the diesel generators of a tanker with real-world data.

2.12 Chapter Summary

This Chapter contained the literature and critical review of the maintenance techniques commonly found in shipping. A deeper review of the predictive maintenance techniques and more specifically the condition-based maintenance was done. The data-driven and model-driven approaches of CBM were explained and the data-driven CBMS were further analysed, namely: data pre-processing, diagnostics, prognostics and database systems. Finally, it was concluded with the current situation of CBM in shipping and the challenges its application is facing and how this thesis will tackle these problems.

3. Development of New CBM Framework

3.1 Chapter Outline

Through the critical review performed on chapter 2, it has been shown that existing CBM frameworks in shipping have limited applicability and are mostly based on simple statistical analysis methods due to the lack of failure data for different machinery. In this respect, it is more than ever necessary to propose a CBM framework that will address these problems and incorporate an intelligent database with advanced machine learning techniques and failure modelling in order to offer better diagnostic and prognostic abilities than the existing systems.

3.2 Proposed CBM General Framework

In order to deal with the main problems for implementing CBM in shipping, namely: Requirement of huge amounts of historical data, including failure rate data, which are notoriously difficult to find, the existence of different operational profiles of a ship (port manoeuvring, bad weather, cruising, etc.) and the mismatch between relational databases and the algorithms, an innovate approach is followed. The new CBM framework has a simple architecture, it uses machine learning and an intelligent database with a novel model theory to predict failure. Figure 19 shows the proposed CBM general framework.

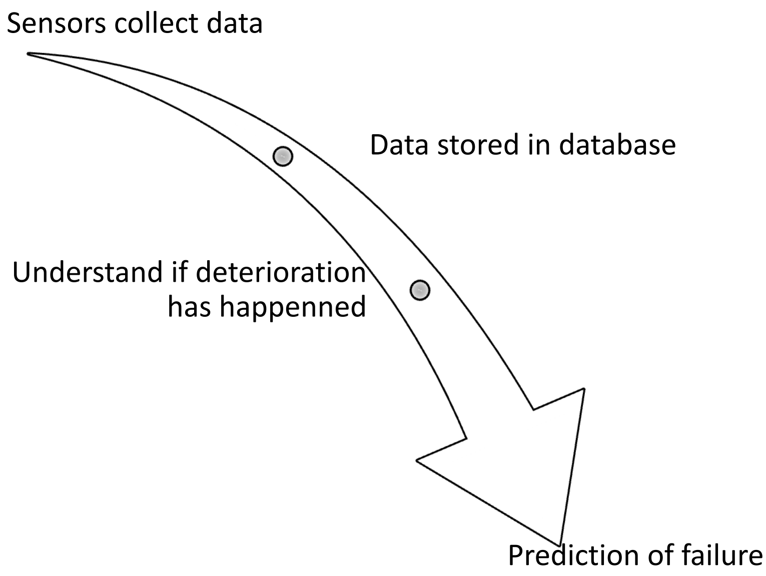


Figure 19: Proposed CBM Framework.

From Figure 19 it can be seen that there are four general steps of the proposed CBM. Firstly, sensors collect data then the data are being stored into a database in order to be used to understand when deterioration has happened and finally the prediction of failure time is estimated. Figure 20 expands these steps in more detail.

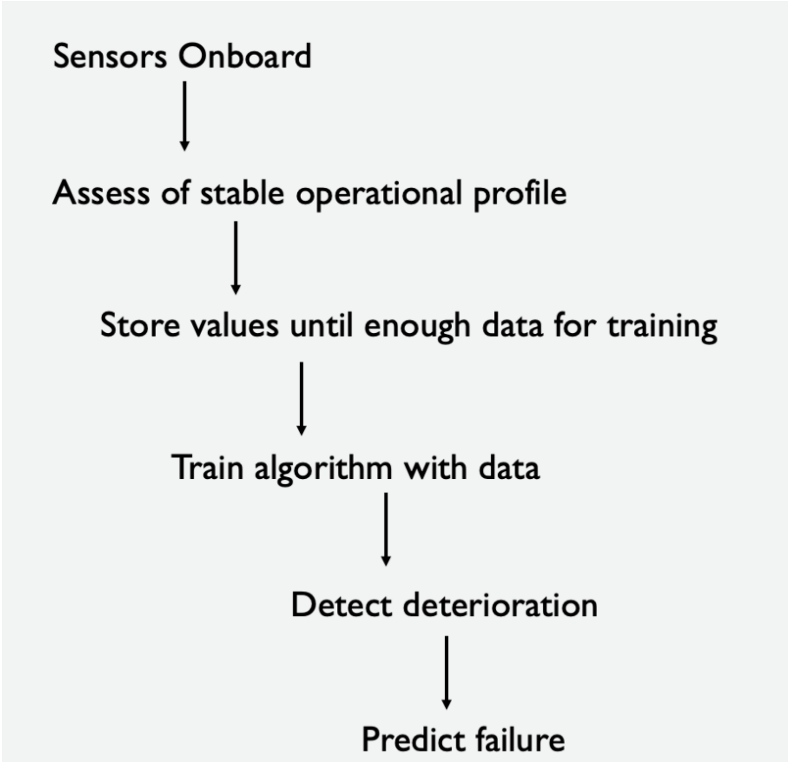


Figure 20: CBM Framework Details

First, sensors onboard the ship are used to collect values of a machinery (e.g., temperature, pressure, etc.). Then, the values are being stored only when the ship's operational profile is stable (more on that in Chapter 5.4). When the stored values are enough, they will be used to train a machine learning algorithm and every new value coming for the same operational profile is being assessed if it represents degradation or normal operation. If the model detects deterioration, the time of deterioration is being fed to the failure model to predict an estimate of the time of failure and the whole process begins again. The following sections will analyse each step of the CBM framework, namely: collection of data and storing in the database, the machine learning algorithm and the failure model with more detail.

3.3 Condition Monitoring Database (CMD)

Following the overview of the new CBM Framework, an Object-Oriented Database (OOD) will be used to intelligently store data in order to minimize any time delays from errors or incompatibility that the traditional relational databases encounter when are used together with machine learning techniques as it was discussed in 3.12.

3.3.1 The Zope Object Database

The CMD database was built entirely in the Python language using the tools provided by the Zope Object Database (ZODB). The Zope Object Database is an open source OOD for transparently and persistently storing Python objects. Because the ZODB is entirely written in Python, no separate language is needed for database operations and there is very little impact on the code to make objects persistent. In addition, there is no database mapper that partially hides the database. Instead of managing relations using different tables with common primary keys, the ZODB let developers use normal Python object references. Thus, the ZODB doesn't require a pre-defined structure of columns and data types for the objects it stores, which means that object attributes can easily change both in quantity and type (ZODB, 2017). This can often be a lot harder when using a relational database for storage.

From all the above, it is more efficient to use an OOD in the new CBM framework since Object-Oriented Databases are designed so they can be directly or effortlessly integrated with software that is developed using Object-Oriented programming languages (Elmasri & Navathe, 2010). The purpose of the database in the proposed CBM framework is to maximize the speed of the machine learning techniques and not to store thousands of unnecessary data. To do that, it intelligently stores data for training and erases them after they have been used to train the algorithm, making space for new training data when they are needed. That is why the new database was named Condition Monitoring Database or else CMD. ZODB offers an efficient way to store objects namely, the BTrees. This structure can hold a large collection of information in an efficient way by having recently, or heavily used objects kept in a memory cache for speed. Also, the whole database can be searched very quickly, because objects are stored in a balanced tree data structure (ZODB, 2017).

3.3.2 CMD Logic

In order to understand better how the CMD understands stable operational profiles and stores data accordingly, Figure 21 shows the logic of the CMD in the case of a data coming from sensors of a Diesel Generator with a frequency of one minute.

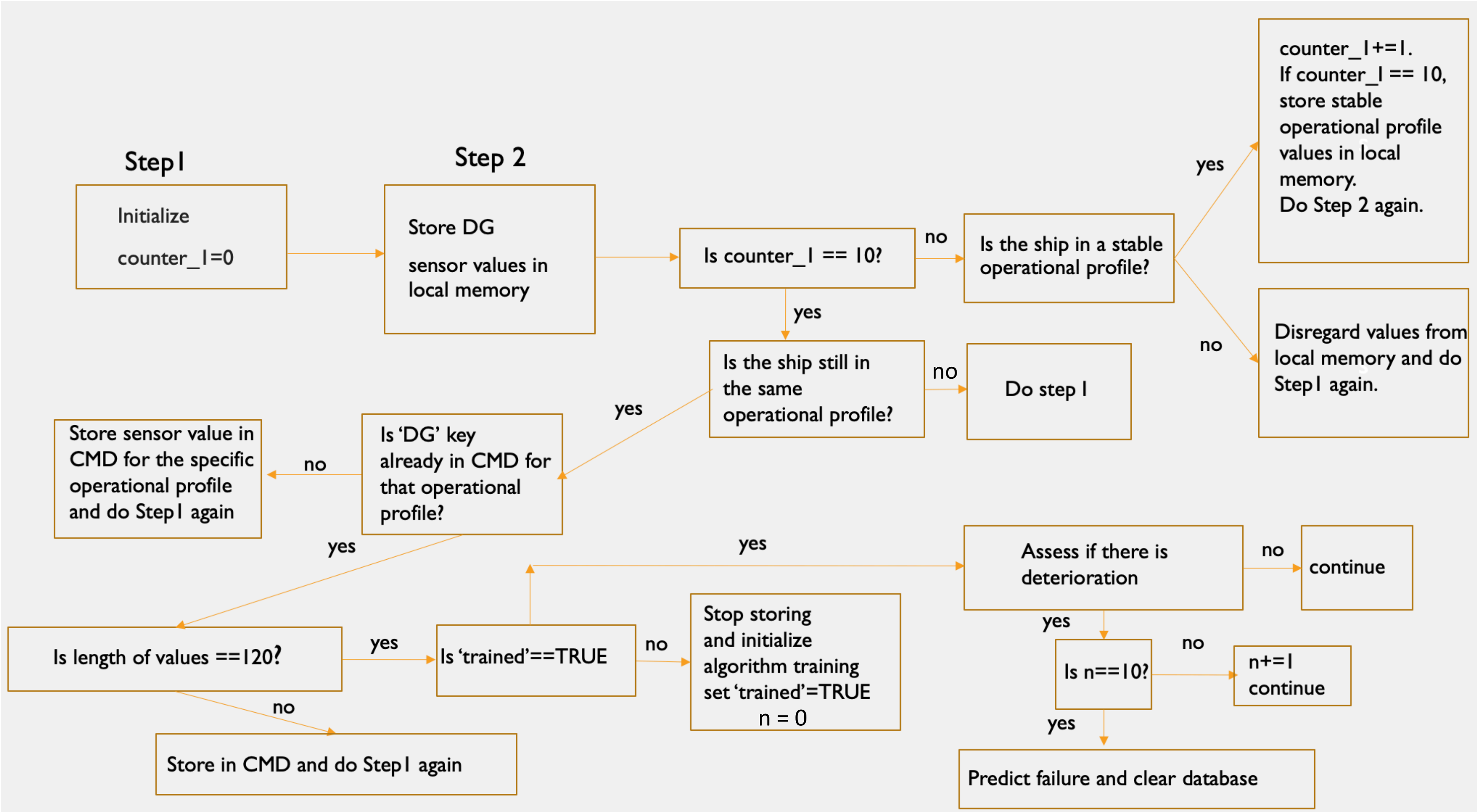


Figure 21: CMD Logic

First, a counter is initialized, counter_1, and data coming from the sensors of the Diesel Generator are temporarily stored to local memory. Only if the ship has been in a stable operational profile for ten consecutive minutes the data storing process will start. These ten minutes are being counted by counter_1, which is incremented every time sensor values come in (since in our example the frequency is 1 minute), and the ship is in a stable operational profile. The ten minutes window was proposed by experts and operators and its significance is that after a ship achieves a stable operational profile the values of the sensors need some time (approximately ten minutes) before they stabilize to their normal values, for that operational profile. In the following sub-section, it will be discussed what conditions need to be met in order for a ship to be in a stable operational profile.

Coming back to the CMD, once the counter_1 is equal to ten, and thus ten minutes with the ship in stable operational profile have passed, the values that represent the stable operational profile are stored in the local memory. If when the next DG values come from the sensors the ship is still in the same operational profile, the DG values will be pushed into the CMD. Then, if values for this operational profile do not already exist in the database, a new key will be created referring to the specific operational profile and the DG values will be stored there. Then as long as the operational profile stays the same the values from the sensors will continue to be stored, otherwise counter_1 will be set to zero and the whole process will start again. Once the amount of data for any specific operational profile reaches a threshold (two hours of data) the training of the algorithm will begin. The two hours of data threshold was proposed by experts and operators suggesting that the values of a stable operational profile do not contain large variance and two hours of data is enough for determining the range of the normal values of a machinery when working in its stable operational profile and assuming the sensor collection frequency is 1 minute. When the training of the algorithm starts the data storing process stops. Every new measurement that comes for the same operational profile is then being assessed by the trained algorithm if it refers to the normal condition or abnormal (deterioration). If the algorithm predicts deterioration for 10 consecutive minutes, then the time of deterioration is being sent to the failure model to predict the time of failure and the database is being cleared by all the data it contains (for all the operational profiles) so the whole process starts again for the new deteriorated state of the machinery. The 10 consecutive minutes threshold again was proposed by experts and operators in order to deal with random spikes of the data caused by the sensors.

3.3.3 Stable Operational Profiles

In order to understand deterioration, one needs to identify the stable operational profiles of the ship. After discussions with experts and operators as well as reviewing manuals of ships' engine (MAN, 2010) and bridge manoeuvring systems (MAN, 2004) it was concluded that a ship's engine speed is controlled by a governor to be within a range of the RPM that have been set from the bridge. That means that a stable speed operational profile can be identified firstly from when the engine speed is stable inside a predefined range. However, even though a stable operational profile requires a stable engine speed it also requires a stable power. The reason is that, for example, a ship can have the same speed for different payloads, but the power required is different and the operational conditions of the machinery are different. For our CBM system it will be required that the power of the engine also needs to be stable between a predefined range in order to have a stable operational profile. In addition, the weather conditions can also have an effect in the operational conditions of the machinery as well as the sea currents. The parameters that need to be kept stable to have a stable operational profile can be seen in Figure 22.

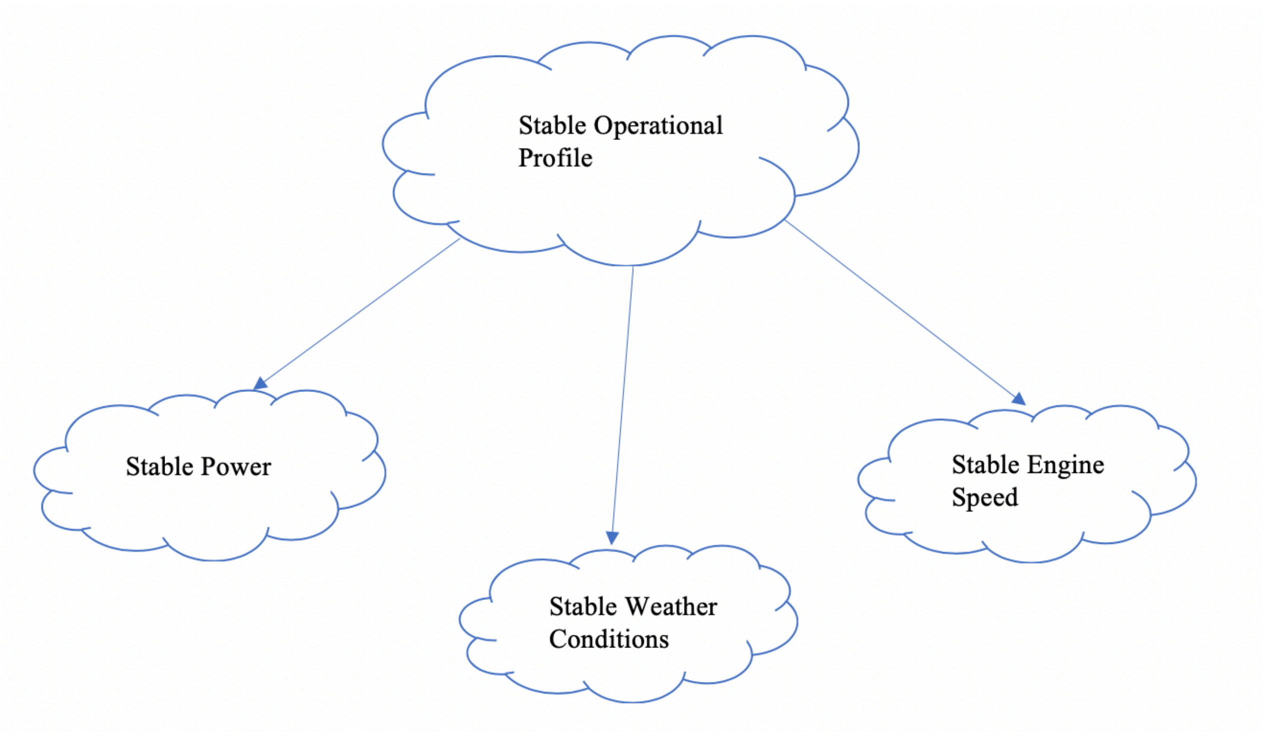


Figure 22: Stable Operational Profiles

The parameters in Figure 22 were selected after discussions with experts and ship operator, and they include: The wind speed, direction, Beaufort, speed over ground and M/E Torque. By keeping the speed over ground and the M/E Torque stable we ensure that sea currents will not affecting the operational conditions (e.g., by increasing resistance to the ship). Overall, all the above parameters need to be kept stable in order to have a stable operational profile as it can be seen in Figure 22.

3.4 Machine learning

As it was shown, the decision for using an OOD for the CMD will significantly boost the speed of interaction of the machine learning techniques with the data accumulated by the system and reduce also the amount memory needed overall. The next step is to choose an algorithm that will be able to detect when the condition of the machinery is deteriorating. Since the CMD contains a small amount of data referring to the current condition of a stable operational condition of the machinery and these data need to be used in order to determine if the machinery's condition is deteriorating, a semi-supervised novelty detection algorithm is needed. From all the semi-supervised algorithms reviewed in the literature review the LOF algorithm was chosen. It's great general applicability as well as it's demonstrated effectiveness compared to other algorithms are the main reasons for that choice. The LOF algorithm will be trained each time a threshold of amount of data is reached, sufficient enough to represent the current condition of the machinery when the ship is on a stable operational profile, and then it will assess the condition of the machinery for every new data point coming from the sensors for that operational profile. Figure 23 shows the process of storing enough data for training.

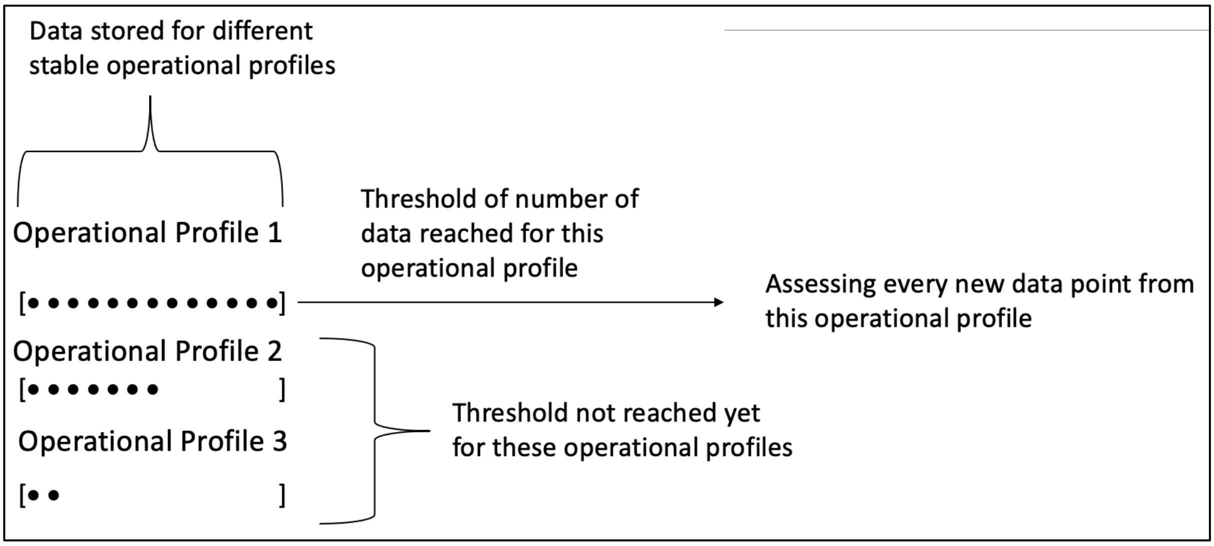


Figure 23: Process of Training the Algorithm

As it can be seen in Figure 23, the CMD collects data for different stable operational profiles but once the amount of the data for any stable operational profile is enough (e.g., for operational profile 1 in Figure 23), the algorithm is trained with the data and assess any new data coming from this operational profile for deterioration. It is important to note that until deterioration has been detected in any operational profile, the data gathering process continues for all the operational profiles that the threshold amount has not be reached. Figure 24 illustrated how the detection of deterioration is being performed.

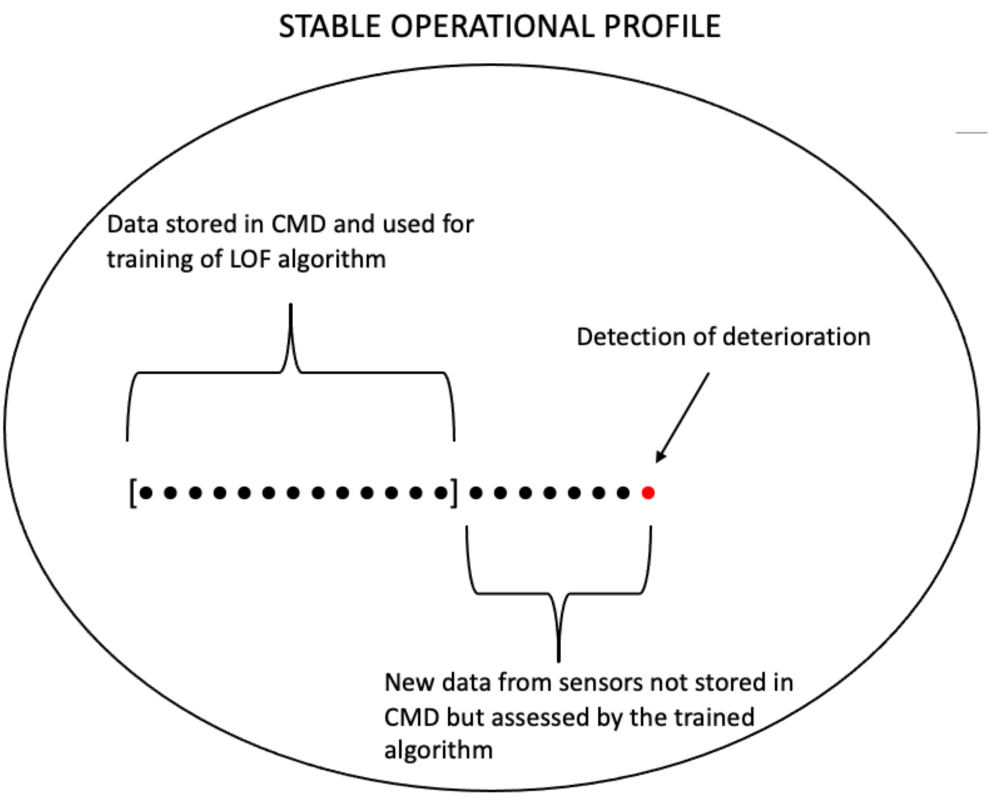


Figure 24: Predicting Deterioration

From Figure 24, once deterioration has been detected for an operational profile the data gathering process stops and the database is completely cleared since the machinery now has entered a new deteriorated state.

3.5 Failure Model

Moving on, after the machine learning algorithm has detected the time that deterioration happened to the machinery a failure model is needed in order to predict an estimation of the time of failure. That failure model will have as an input the time(s) that deterioration happened to the machinery and will give an estimation of the time of failure that will become more and more precise the more data are collected.

3.5.1 P-F Diagrams

In order to understand the new failure model, the Potential-Failure (P-F) diagrams need to be presented. The P-F diagrams illustrate how a condition deteriorates until it reaches a point of functional failure (ABS, 2016). Figure 25 shows how most of the CBM systems currently in place work.

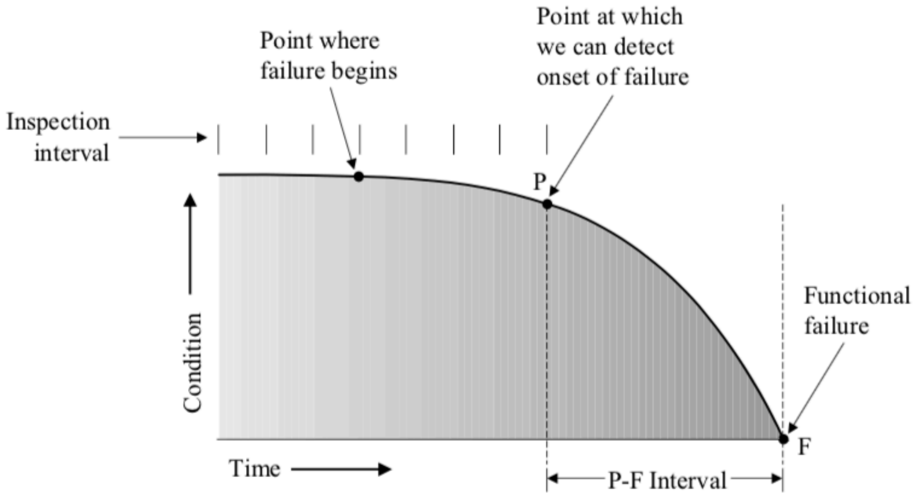


Figure 25: A Potential-Failure diagram (ABS, 2016)

Figure 25 shows how a condition deteriorates to the point at which it can be detected (Point P) and then, if it is not detected and corrected, continues to deteriorate until it reaches the point of functional failure (Point F). A functional failure is a description of how the equipment is unable to perform a specific function to a desired level of performance. The P-F diagrams are used with the 'inspection' interval logic where the system is periodically checked to determine the process of the failure and industry practice is to select an interval of about one-half of the P-F interval. It should be noted that the P-F interval can vary in practice, and in some cases, it can be very inconsistent. In these cases, a task interval should be selected that is substantially less than the shortest of the likely P-F intervals (ABS, 2016). With the proposed CBMS the speculation of a static P-F interval is eliminated by using continuous monitoring and updating the P-F curve accordingly. Here the end goal is to predict and update the functional failure dynamically and treat each machinery individually by their sensor data. In the next sub-section, it will be shown how the times of deterioration caught by the machine learning algorithm can help dynamically predict the failure of a machinery without the need of any historical or failure data.

3.5.2 New General Failure Theory

This thesis suggests an innovate way to model failure by using the intuitive P-F graph and mathematically solve it in order to acquire a general model of the failure mechanism. This general model of failure will be updated dynamically by the machine learning algorithm to depict the actual P-F graph at each deterioration moment for the particular machinery under question. Since the CBM system is only interested from the point where deterioration can be detected and assuming that the condition practically remains the same from the point where deterioration begins till the point that deterioration can be detected, Figure 7 is transformed to Figure 26.

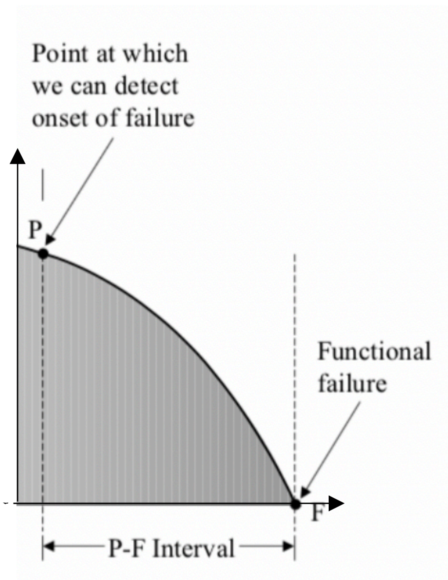


Figure 26: P-F Graph Starting from the Point that the Failure is Detectable (ABS, 2016)

From Figure 26 it can be seen that the shape of the general P-F interval can be modelled as an ellipse starting at the point P. This intuitive understanding of failure will be mathematically analysed in the next sub-session.

3.5.3 Key Concept and Theory

Assuming that at the point P from Figure 8, when the detection of the onset of a failure is possible, the reliability (the probability that the machinery will perform its intended function adequately) of the machinery is equal to 1 (or else 100%), the equation of the ellipse will then be:

$$\frac{x^2}{{t_F}^2} + \frac{y^2}{1^2} = 1 \quad (1)$$

Were,

t_F: the time of failure,

x∈ [0,+∞),

y∈ [0,1]

Solving for the time of failure in (1) will give:

$$t_F = \sqrt{\frac{x^2}{1-y^2}} \qquad (2)$$

So, in a static diagram by knowing a pair of (x,y) the time of failure t_F can be calculated. While the modelling of the failure is a general one the application to a machinery will need to be specific and coming from the on-line data. The idea is that from the on-line data the machine learning algorithm will be able to understand when the machinery is deteriorating through novelty detection. If we assume that the failure is following an ellipse, then the very first detection of changing of condition P_0 in time x_0, will be very close to the beginning of the y axis and the condition is assumed to be almost equal to new as it is seen in Figure 27.

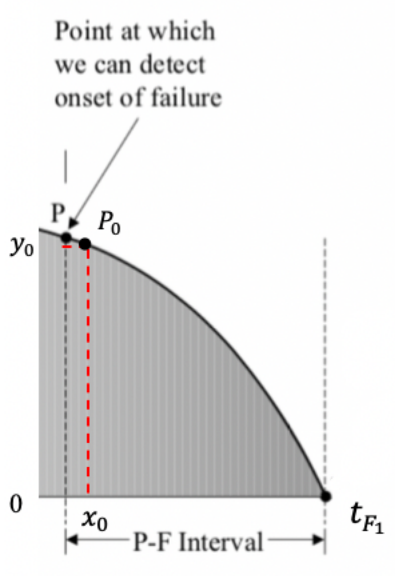


Figure 27: Key Concept of the new CBM System (ABS, 2016)

That means that the time of failure can be calculated as per equation (2) assuming that the reliability y_0 is very close to 100%. The time x_0 is measured from the moment the machinery first put to work or after major maintenance. Equation 3 gives the first prediction of the time of failure from the CBMS.

$$t_{F_0} = \sqrt{\frac{{x_0}^2}{1-{y_o}^2}} \qquad (3)$$

However as new data are coming, and deterioration is happening the P-F diagram needs to be adjusted accordingly. When the second change of condition has been detected the reliability of the machinery would be y_1 as seen in Figure 28.

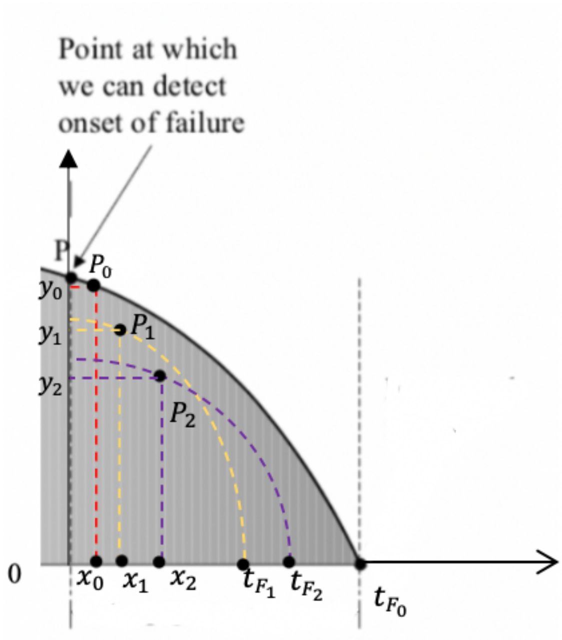


Figure 28: Example of the new CBM System (ABS, 2016)

The loss of reliability it is assumed to be given by equation 4.

$$y_1 = y_0 - a * (x_2/x_1) \qquad (4)$$

Were,

a: universal factor

The new reliability will be calculated by deducting from the previous reliability a percentage of the fraction between the consequent times of deterioration (e.g., x_2/x_1). The fraction is preferred than the deduction of the consequent times of deterioration (x_2-x_1) because it is independent from the time measurement itself (minutes or seconds or days). It is assumed that the percentage factor a is universal and is not connected to the specific machinery or experimental set up (e.g., sensors). This can be generalized for every change of condition as shown in equation 5.

$$y_n = y_{n-1} - a * (x_n/x_{n-1}) \qquad (5)$$

Were,

n∈ [2,+∞),

Then the prediction of the time of failure will be changing as in equation 5.

$$t_{F_n} = \sqrt{\frac{x_n^2}{1-y_n^2}} \qquad (6)$$

This essentially means that different ellipses are drawn every time you have a new change of condition with the starting point as indicated by equation (5) and the predicted time of failure as indicated in equation 6. Figure 28 showed a random example with 3 predictions from 3 different ellipses as the deterioration is progressing. The estimation of time of failure will be, at any time of changing condition, the average of all the predictions until then.

$$\overline{t_{F_n}} = \frac{t_{F_1}+t_{F_2}+t_{F_3}+\cdots t_{F_n}}{n} \qquad (7)$$

Were,

$n \in [1,+\infty)$

From equation 7 it can be seen that the first prediction t_{F_0} is not been regarded in the average because it is the only one that has its condition assumed (close to 100%) and is expected to have larger error than the following ones since most of the machineries on a ship are not brand new even after major maintenance. Nevertheless, if the machinery in the time of starting the CBMS is brand new the first prediction t_{F_0} can be taken into account.

3.6 Chapter Summary

In this Chapter the framework of the proposed CBM system was outlined. First the general steps of the CBM were shown, namely: use of a database to store data, detecting deterioration from data and using a new failure model to predict possible failure of the machinery. Concerning the database, an object-oriented database type was chosen to be used because of the significant advantages compared to the relational type databases. Then, the LOF algorithm was chosen to detect deterioration from data because of its well-known application in detecting anomalies. Finally, the mathematics behind the new failure model were presented as well as how the times of the detected deteriorations will be used with it to predict when a machinery will fail.

4. Case Studies

4.1 Chapter Outline

In this Chapter, three case studies will be presented. The first two case studies will help to evaluate and test the various parts of the proposed CBM framework. The first case study validates the power of the LOF algorithm in understanding deterioration in a Naval combined diesel-electric and gas-turbine propulsion plan. The second case study determines the optimum value of the parameter “a” from equation 4 in 4.4.3 and then validates the proposed failure model by using multiple failure data from bearings where the actual failure time was compared with the predictions of the failure model. Finally, the third case study is a real-world case study and will help to demonstrate the benefits of using the CBM system for diagnosis and prognosis of three Diesel Generators (DGs) of a tanker.

4.2 Case Study 1

4.2.1 Case Study 1 Outline

After explaining the characteristics of the CBM framework the LOF algorithm will be tested for understanding the degradation of a marine propulsion system consisting of a compressor and a turbine. The CMD will also be used to store the data and allow quick access by the LOF algorithm.

4.2.2 Dataset

Coraddu et al. (2016) investigated the problem of performing Condition-Based Maintenance in a Naval combined diesel-electric and gas-turbine propulsion plant. Confidentiality constraints with Navy led to the use of a real data validated simulator and the dataset has been published for free use through the UCI (2016) repository. This is the data set that will be used for the training, validation and testing of the LOF algorithm for novelty detection. The dataset includes 25 features (temperatures, pressures, etc.) and for each array of the 25-array features the decay coefficients are given for the Propeller torque, Hull, GT Compressor and GT Turbine. For the validation of the LOF the decay coefficients of the compressor and turbine will be used.

The ship speed in the dataset has been investigated with sampling the range of feasible speed from 3 knots to 27 knots with a granularity of representation equal to tree knots and the transients between different speeds were not considered (Coraddu et al., 2018). All the measures (25 features) which indirectly represent the state of the system have been acquired and stored in the dataset over all the ship speeds. The data for a ship speed of 18 kn were used for the training, validation and testing of the algorithm but sensitivity analysis of the final model was performed for all the spectrum of different ship speeds and operational profiles.

4.2.3 Data Preparation & CMD

First the data for a specific ship speed (18kn) were gathered. Then these data were further divided to the ones representing perfect condition of the machinery and to the ones representing all the possible degradation conditions. The perfect condition is indicated when the decay coefficient of the machinery equals to one and the different condition is indicated when the decay coefficient of the machinery is not equal to one. The next step is to store the data in the CMD. The data were stored successfully to the new CMD architecture and the machine learning algorithm used the data directly from there with maximum speed since the algorithm and database were both build in the python language.

4.2.4 Logic of the Experiment

Figure 29 shows the steps of the experiment that were followed. The algorithm was trained only in the values representing a specific condition (the perfect one) and then it was validated and tested in both predicting if the condition has changed or if the condition has remained the same. The experiment was done twice, once for the compressor and once for the turbine.

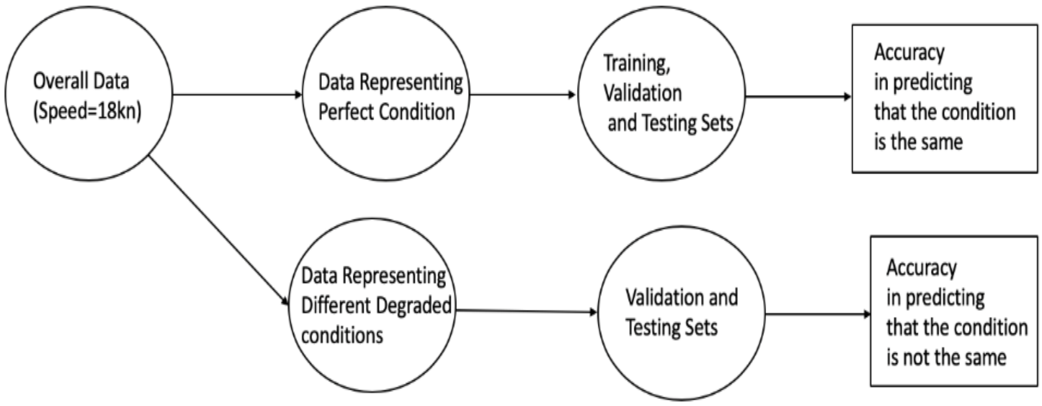


Figure 29: Steps followed for the experiment

4.2.5 Error Metric

The metrics used for determining the performance of the algorithm were the accuracy in predicting when the condition was changed and the accuracy in predicting when the condition was still the same. Accuracy is defined as per equation 8.

$$\text{Accuracy} = \frac{Number\ of\ correct\ predicitons}{Total\ number\ of\ predictions} \qquad (8)$$

The numerator in equation 8 is referring to the number of correct predictions by the algorithm and the denominator to the total amount of predictions (including the false ones).

4.2.6 Training, Validation and Testing Datasets

For the experiment 80% of the data representing degradation were used for validation and the rest of it for testing purposes. Regarding the number of training data, from the perfect condition data pool, 80% were used for training, 10% for validation and 10% for testing. The purpose of using a validation set is to be able to establish the best hyperparameters for the LOF algorithm for highest accuracy. The most important hyperparameter of LOF is the number of neighbours for calculating the local density of a point. Through the validation process this was found to be one, meaning that the local density of a point is calculated by its distance from the nearest data point.

4.2.7 Results

Since the set-up of the experiment was determined, the next steps were to perform the training and validation of the algorithm, establish the best parameters for highest accuracy and use the final algorithm in the testing dataset. In addition, further experiments with lower amount of training data were needed in order to explore the effect that the amount of training data has on the accuracy of the model. Finally, a sensitivity analysis was performed for all the operational profiles of the ship.

4.2.8 Parameters & Accuracy

The best values for the parameters of the algorithm were found to be the same for both the turbine and the compressor experiments showing that these parameters seem not to be affected by the type of machinery. They have a rather general applicability when it comes in understanding the change, or not, of the condition of the machinery.

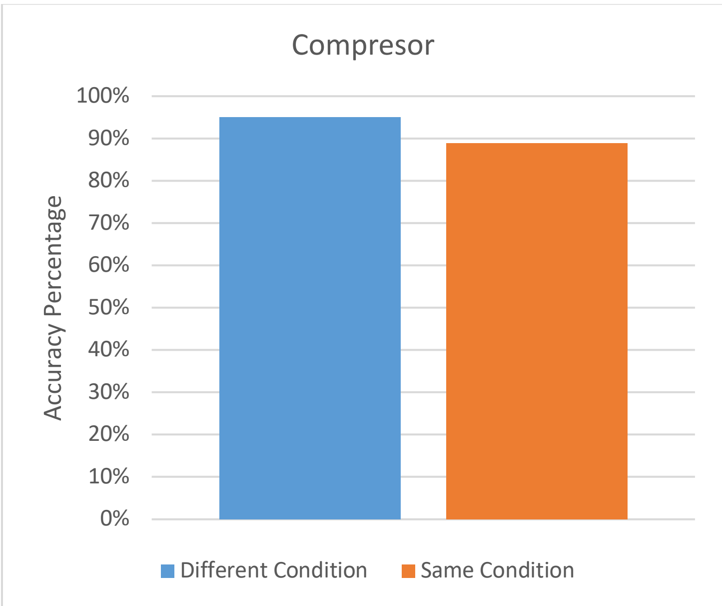


Figure 30: Results of the experiment for the compressor

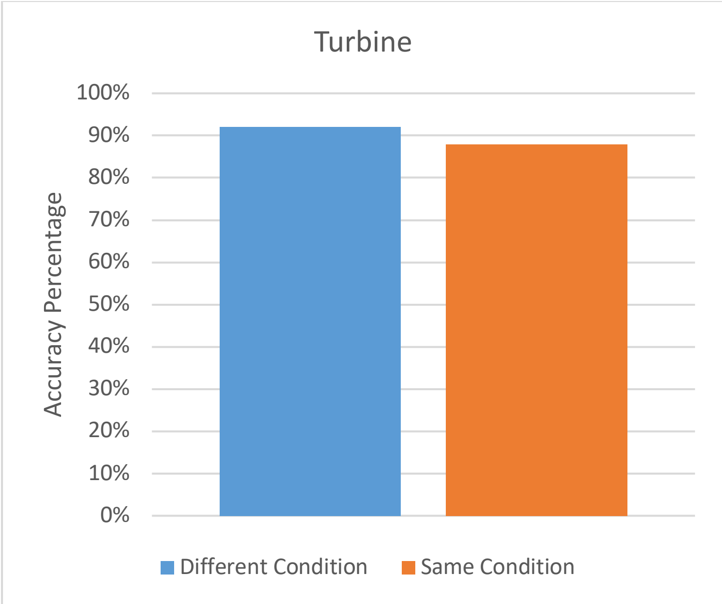


Figure 31: Results of the experiment for the turbine

In Figure 30 the accuracies from the compressor experiment can be seen. The algorithm performs excellently with a mean accuracy for predicting correctly if the condition has changed or not, of 92%. In Figure 31 the accuracies from the turbine experiment can be seen. The algorithm again performs very well with a mean accuracy for predicting correctly if the condition has changed, or not, of 90%. The high accuracies of both experiments reassure us that the algorithm will detect any deviation of the current condition of the machinery successfully.

4.2.9 Amount of Training Data

The accuracies of the experiments refer to a specific number of training data. In the experiments 6550 instances were used approximately for both cases (6550 for compressor 6554 for turbine). It is interesting to explore if the amount of data can be reduced so we can save time in training the algorithm.

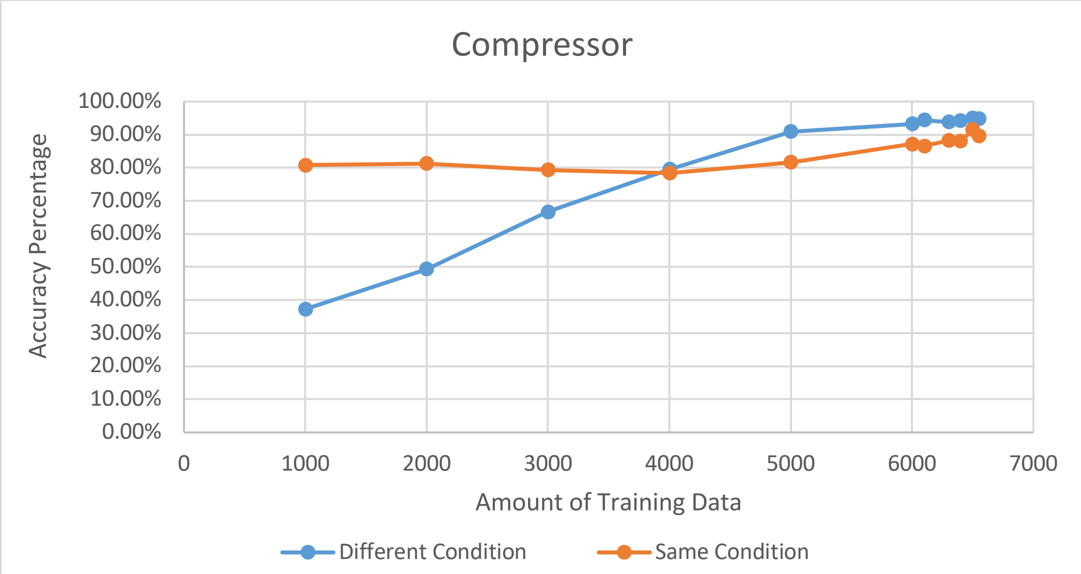


Figure 32: Effect of the training data in the accuracies of the model

As it can be seen in Figure 32, in order to have at least 80% accuracy in both detecting different condition or the same, the algorithm needs at least 5000 training data. Since using 6550 data will give a substantial higher accuracy, with no significant extra time for collecting them, there is no reason to use less training data and compromise the accuracy of the model.

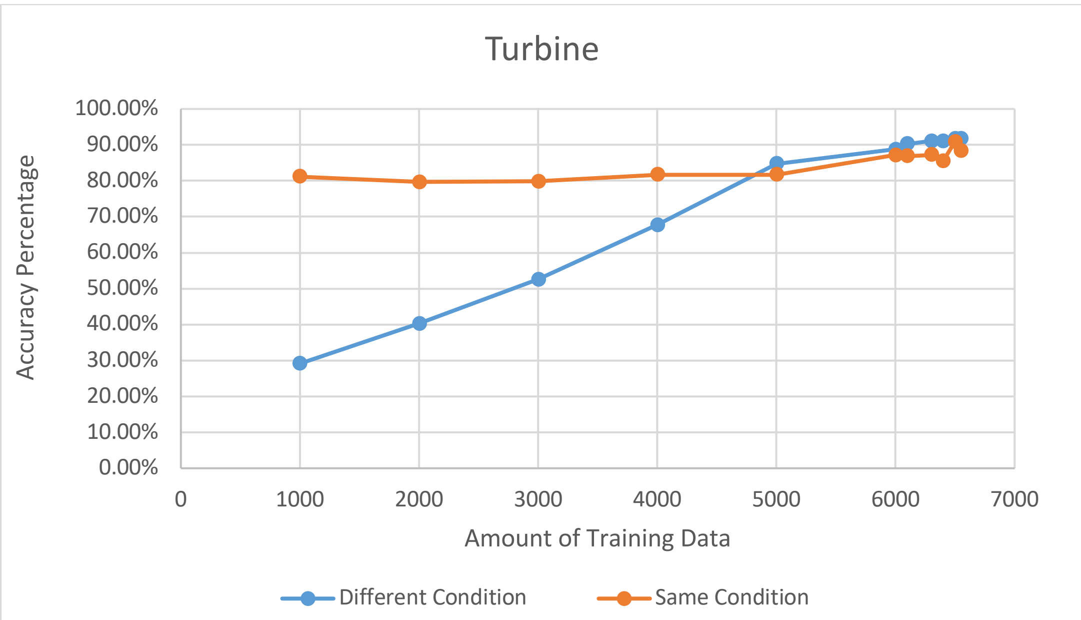


Figure 33: Effect of the training data in the accuracies of the model

Moving on to the turbine, Figure 33 shows similar results with the one's for the compressor. The model seems to require at least 5000 data points in order to achieve 80% accuracy which is also indicating that the threshold of 6550 data points is also applicable in this case. In a real-world scenario, experts and operators argue that the threshold of training data should be much less (equal to 2-hours of data for 1 min frequency of gathering data) because of the small variance of the values in a real stable operational profile which is not only dependant of the ship's speed. The high threshold in these experiments is explained by the way that a stable operational profile was defined, namely only by ship's speed. This assumption contains a lot of error as it will be shown later (in the real-world case studies) since to define a stable operational profile someone needs a lot more information.

4.2.10 Sensitivity Analysis: Ship Speed

Since the ship speed was used as the parameter indicating the operational profile of the ship, it is essential to see how the algorithm performs at different ship speeds. It should be noted that for each operational profile the algorithm is getting trained again, with the parameters that were established in equation 7. That means that there is no validation set in this experiment. The available ship speeds in the database are: 3, 6, 9, 12, 15, 21, 24, 27 kn.

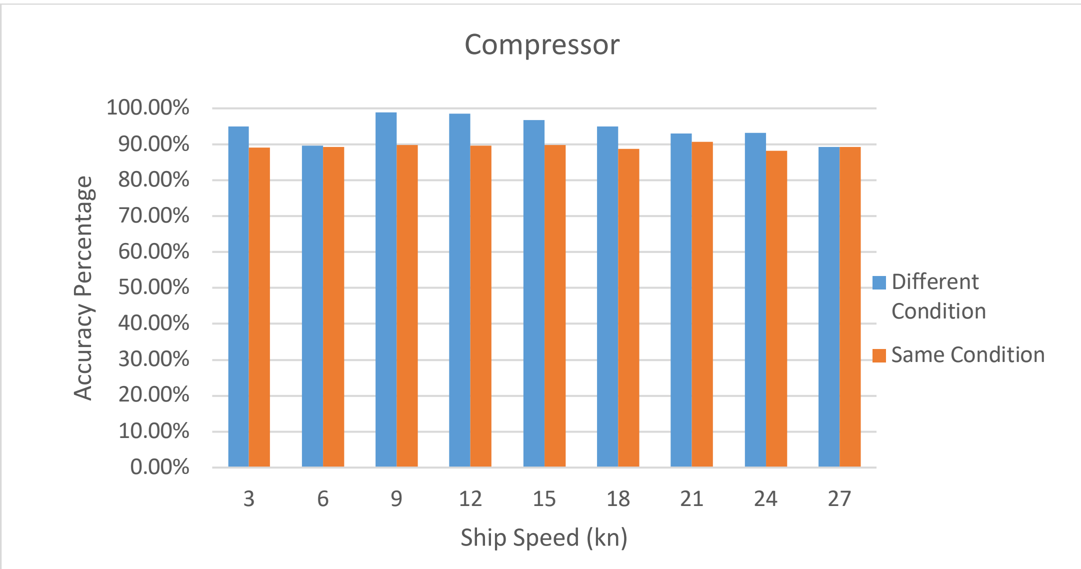


Figure 34: Accuracies for different ship speeds for the compressor

Figure 34 shows the results of the sensitivity analysis for the compressor. The mean accuracies for correctly detecting the change of condition and that the condition is the same, across all speeds, is 94% and 89% respectively. That gives a total of 92% overall accuracy and shows the consistency of the algorithm across different operational profiles.

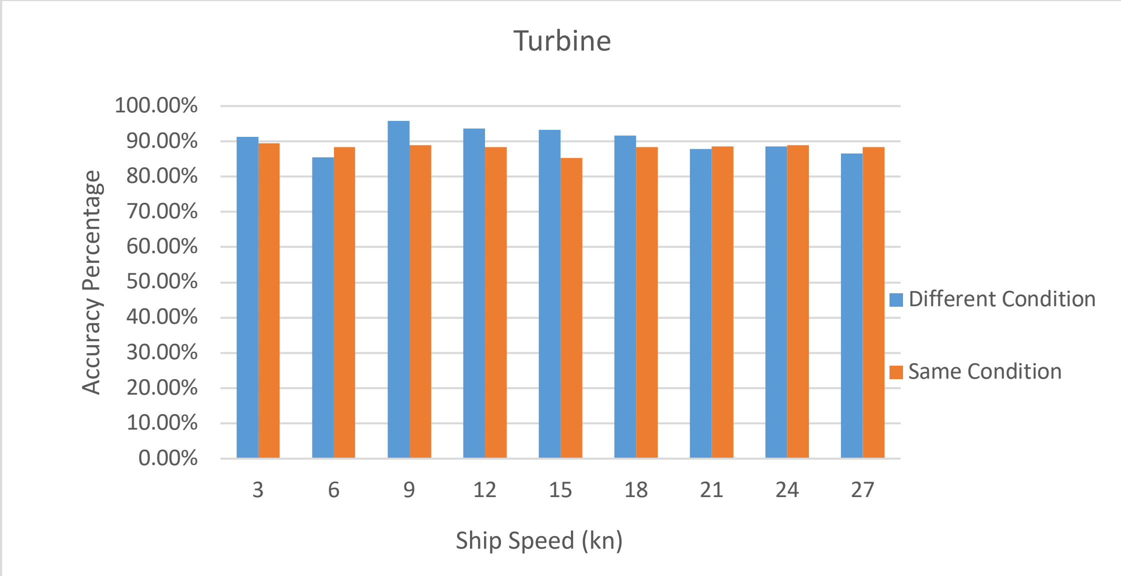


Figure 35: Accuracies for different ship speeds for the turbine.

On the other hand, the results for the turbine can be seen in Figure 35. Here the mean accuracies for detecting a change or not, across all speeds, is 90% and 88% respectively. That gives a total of 89% overall accuracy, indicating again a sufficient amount of accuracy across all operational profiles. Overall, the sensitivity analysis has proven that the algorithm is robust enough to perform with high accuracies across all the operational profiles of the ship and for both the turbine and compressor cases. That means that it can be used to continuously monitor the degradation of the machinery throughout all the operational profiles of a ship and thus increasing the reliability and accuracy of the diagnostic and prognostic system.

4.2.11 Conclusion of Case Study 1

In this chapter the first two parts of the CBM framework were tested through the use of data describing the degradation of a compressor and a turbine of a naval propulsion system. The results of the experiments showed that the algorithm gives high accuracies in both the turbine and compressor cases. In addition, the operational profile doesn't seem to affect the accuracy of the model as it was indicated by the sensitivity analysis performed for both the compressor and turbine for all the available speeds of the ship. That means that the LOF algorithm can be used in continuous monitoring the degradation of a ship going through different stable operational profiles. Finally, it was shown that the algorithm needs a threshold of amount of data to be trained in order to perform the best and it was concluded that this threshold for a real-world case would be much less, as also suggested by experts, since the stable operational profile will be defined in a much stricter sense.

4.3 Case Study 2

4.3.1 Case Study 2 Outline

In order to optimize and validate the proposed failure model, failure data need to be utilized. The purpose of this case study is two-fold: First the parameter "a" from equation (5) will be experimentally established by using a bearing failure dataset and secondly another bearing failure dataset will be used to compare the predictions of the failure model with the actual failure time in order to evaluate its effectiveness.

4.3.2 Dataset

The data was generated by the NSF I/UCR Centre for Intelligent Maintenance Systems (IMS – www.imscenter.net) with support from Rexnord Corp. in Milwaukee, WI (NASA, 2019). The dataset contains three experiments working until failure of bearings and can be summarised as in Appendix.

After carefully examining the dataset there are a couple of problems that needed to be addressed. The second experiment will not be used since the recording duration (7 days) does not start from the beginning of the experiment and so critical information about the failure degradation is lost. Also, the first experiment has missing recordings during the period of the 35 days and overall, 11 days are missing across the dataset. It will be assumed that during these days the condition of the machinery does not change. Overall, the first experiment has a duration of 35 days which is slightly more than the designed life of the bearing (see Appendix), but the third experiment lasts for almost 44 days which is 9 days more than the designed life span of the bearing. It should be also noted that for the first data set two accelerometers for each bearing [x- and y-axes] are used and one accelerometer for each bearing for the third data set.

4.3.3 First Experiment

The first experiment will be used as the training/validation dataset in order to establish the parameter best value for the parameter a in equation 5 and then this parameter will be tested with the third dataset. The estimation of the parameter was based in having the best accuracy in the predictions of the bearing's failure. Table 12 shows the prediction of the failure of the bearing every time that the condition has been detected to change (deterioration) for the optimum value of parameter "a". The prediction of failure starts when the second deterioration has been detected as it was discussed in 4.4.3. Table 17 shows the predictions of the model at the moments that deteriorations have been detected from the selected best value for parameter "a".

Table 17: Time of Failure for Optimal Parameter "a"

Deterioration time (days)	$\boldsymbol{t_F}$ for: (days)	a=0.01	Prediction accuracy (%)
1.08		-	-
1.94		17.19	48.69
11.40		21.58	60.56
14.31		25.06	70.04
14.58		26.92	75.10
19.34		29.60	82.47
26.92		35.03	97.23

From Table 17 it can be seen that the percentage accuracy is increasing with each new detection of deterioration. In fact, after each detection of deterioration the accuracy of the prediction of the failure model increases by an average of 9% starting from a 50% accuracy at only 6% of the time of failure. This is what a machine learning process looks like, the algorithm can perform better and better predictions of the time of failure as time passes and more data are fed into the system. Figure 36 shows exactly this process, namely how the accuracy of predicting the time of failure becomes better and better by time.

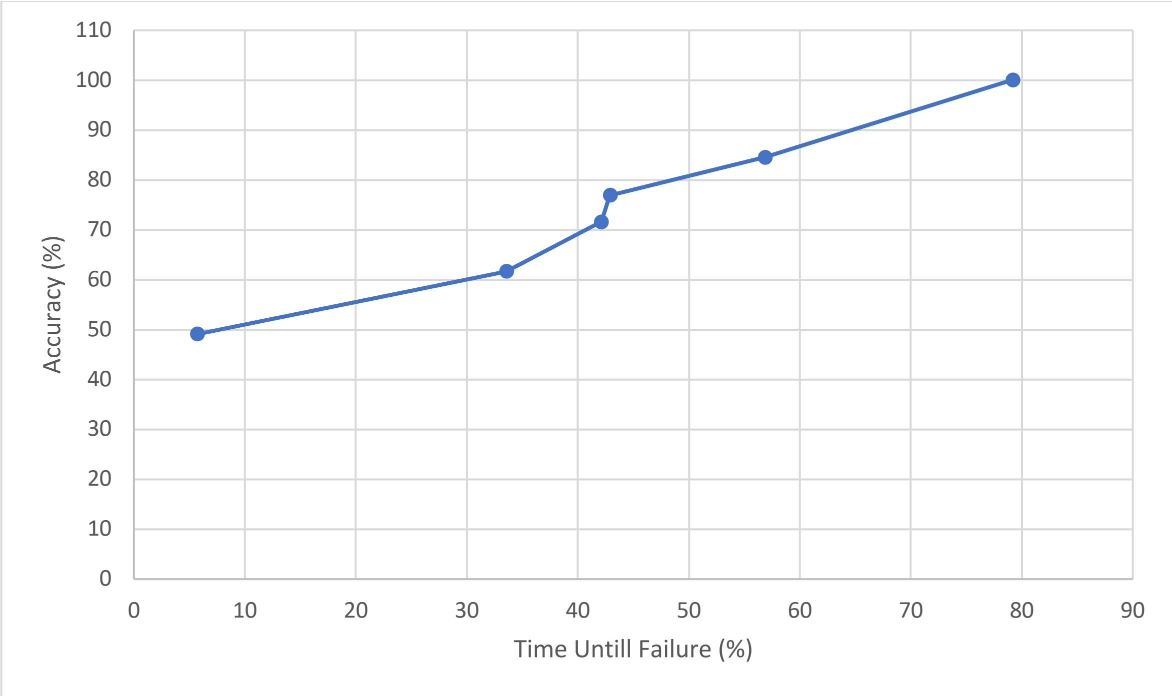


Figure 36: Accuracy of Prediction

It can be seen in Figure 36 that the accuracy of prediction becomes better the more deteriorations are detected, starting from 50% and finishing at 100% accuracy. In more detail, at half of the actual time of failure there is an 85% accuracy of the predicted time of failure of the model. In addition, at 80% of the actual time of failure the predicted time of failure of the model is also the actual time of failure of the bearings. That means that the algorithm can give a very robust estimation of the time of failure early on and the operators can start planning maintenance while the information of the time of failure is pinpointed later on with more data coming on. The experimental established parameter "a" will now be tested with the second experiment.

4.3.4 Second Experiment

Moving on, the second failure experiment will be used for validating the already established parameter "a" from the first experiment. In Table 18 the results of the experiment can be seen.

Table 18: Time that condition was changed and prediction of time of failure

Deterioration time (days)	t_F
0.83	-
6.29	16.89
10.89	21.71
21	30.07
23.92	35.29
24.17	38.16
25.75	40.30
28.63	42.38

As it can be seen from Table 18, similar to the previous experiment, the percentage accuracy of predicting the time of failure is increasing every time a new deterioration is being detected by an average of 9.5% starting from a 42% accuracy at 15% of the time of failure and finishing with 96.5% accuracy at 65% of the time of failure.

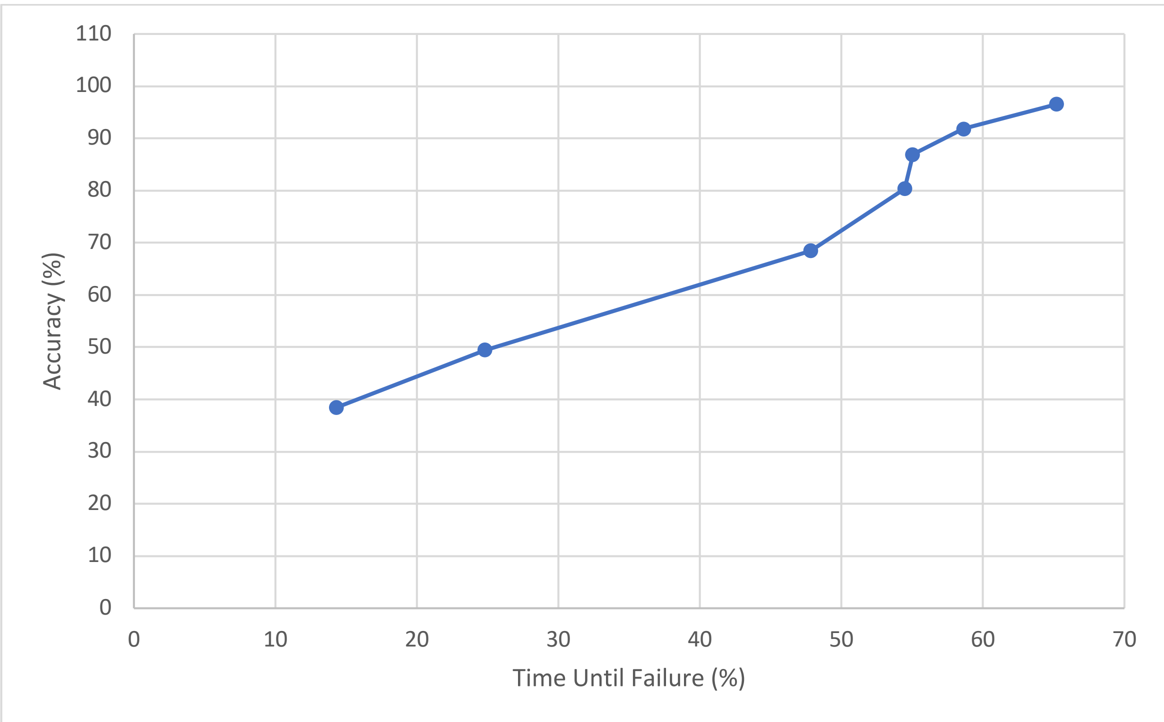


Figure 37: Accuracy of prediction vs time until failure

It can be seen in Figure 37 that the accuracy of prediction becomes better the more changing of conditions are detected starting from 42% and finishing at almost 100% accuracy. In more detail, at almost half of the actual time of failure (54%) there is an 80% accuracy of the predicted time of failure of the bearings. In addition, at 65% of the actual time of failure the predicted time of failure is almost equal the actual time of failure of the bearings (96.5%). Again, the results suggest that the algorithm is able to give a robust estimation of the time of failure early on with high accuracy. The accuracy of the algorithm in the prediction of the time of failure showcase that the general failure model combined with the power of the LOF algorithm can successfully perform diagnosis and prognosis of a machinery.

4.3.5 Conclusion of Case Study 2

This case study demonstrated the effectiveness of the proposed failure modelling theory. It also showed that the failure prediction becomes more and more precise as more data are being fed to the condition-based system. Also, the failure model is able to correctly predict the time of failure in advance before it happens. Figure 38 shows the benefits of using the proposed modelling theory for predicting the time of failure of a machinery.

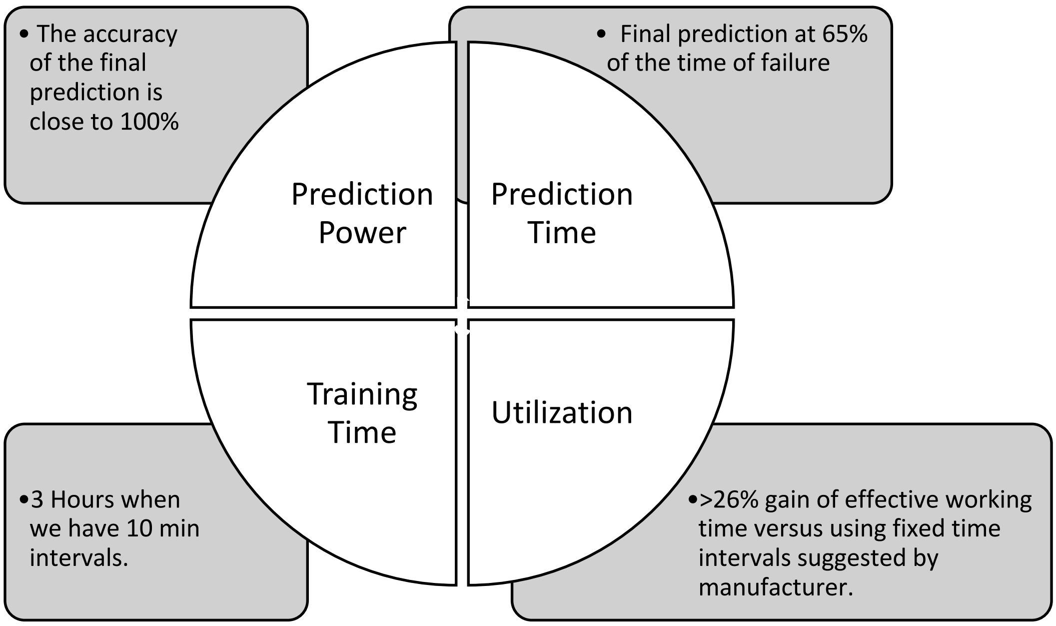


Figure 38: Results of Case Study

From Figure 38, the prediction power of the failure model is shown from the fact that it is able to give with high accuracy the time of failure well in advance of the actual failure. In addition, the training time of the algorithm is fast and equal of 3 hours. Furthermore, by using the proposed CBM system there was a 26% increase of the effective working time of the equipment compared to performing maintenance according to manufacturer's suggestions. This increase of effective working time can become substantial through time when it can accumulate into hours, months or even years of extra working time.

4.4 Case Study 3

4.4.1 Case Study 3 Outline

Moving on to the final case study, the purposed CBM system will be used together with real world data in order to perform diagnostics and prognostics for the maintenance of the cylinders of three diesel generators of a ship. The purpose of these case study is to demonstrate the effectiveness of the proposed CBM system in real world scenarios as well as to explore the benefits that the operators can get from using it.

4.4.2 Dataset

The dataset that was used contains data gathered by sensors onboard a tanker with one minute frequency and for a period of three months, namely: From 03/04/2018 to 03/07/2018. Table 19 shows the basic characteristics of the tanker. The ship under question has an overall length of 183 meters and a displacement of almost 40000 tonnes which makes it a medium range tanker.

Table 19: Characteristics of Ship

Characteristics of Ship	
Overall length (m)	183
Displacement (t)	38396
Main Engine	DMD-MAN B&W 6S50MC-C 9480 KW
Diesel Generators	MAN B&W 6L23/30H

The dataset contains data that can be separated into two datasets: Data that help the CBM system determine a stable operational profile and data that are indicators of the condition of the Diesel Generator's cylinders. The former data to determine a stable operational profile as it was discussed in 4.3.3, are: the main engine's (M/E) rounds per minute (RPM), the M/E's power, the wind speed, the wind's direction, the Beaufort scale, the ship's speed over ground and M/E's Torque. The later data that indicate the condition of the Diesel Generator cylinders are: Exhaust Gases Outlet Temperature, CFW Inlet Pressure, CFW Outlet Temperature, Cooling Air Temperature, LO Inlet Pressure, LO Inlet Temperature and Exhaust Gas Outlet Temperature.

The proposed CBM system uses the first dataset to first identify the stable operational profile that the ship is at a specific time and then store under that specific operational profile the data from the second dataset that indicate the condition of the cylinders in order to assess deterioration. Figure 22 shows the CBM's data collection process.

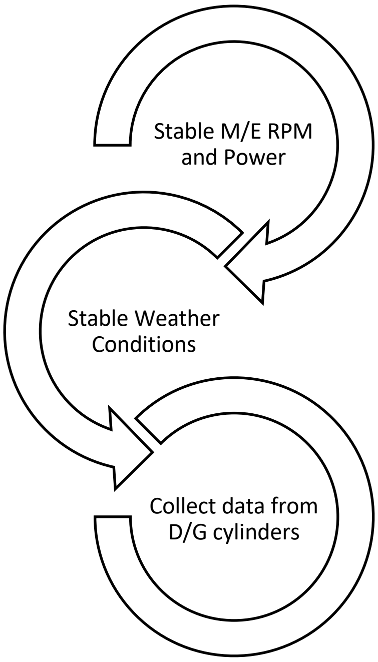


Figure 39: CBM Data-Collection Process

From Figure 39, it can be seen how the new CBM systems collects operational data of the D/G cylinders in a real case scenario and only when a stable operational profile is detected through the use of M/E (RPM and Power) and weather data.

4.4.3 Assumptions

Now that the CBM's data-collection process is explained, the assumptions made for the case studies will be presented. First of all, the reliability (the probability that the machinery will perform its intended function adequately) of the machinery is considered perfect (≈100%) when it is brand new or after major maintenance. Furthermore, after discussing with the ship's operators it is assumed that the ship is at sea 80% of the time in a year. Furthermore, it is assumed that there no deterioration happened between. the last major maintenance and when data collection activity started. However, a sensitivity analysis of the starting reliability of the D/G cylinders will be performed in order to demonstrate how the starting reliability when data gathering activity starts affects the CBM system diagnostics and prognostics. Also, it is assumed that the available sensor parameters for the D/G cylinders are enough to determine their reliability at any time. Finally, as it was mentioned before the stable operational profile is assumed to be stable from M/E and weather data. Table 20 shows the assumptions that were made before the analysis.

Table 20: Assumptions of Case Studies

Condition of machinery is almost perfect after major maintenance.	No deterioration has happened between the last major maintenance and when data collection activity started.	Available parameters show the health of the machinery.	If M/E and weather data are stable the operational profile is assumed stable.	The ship is at sea 80% of the time in a year.

In more detail, concerning the stable operational profiles, it is assumed that there is a stable operational condition when:

- The measurements from the engine speed and engine power are deviating +-5% as suggested in the bridge manoeuvring systems (MAN, 2004)
- The measurements from the weather dataset (the wind speed, wind direction, wind Beaufort scale, speed over ground and M/E Torque) are deviating +-10%. The +-10% instead of +-5% is introduced due to weather's more volatile nature.

By using these assumptions for the stable operational profile, the data were separated automatically into 5 operational profiles represented by their RPM range (but calculated by using all the available parameters above) as it can be seen in Figure 40.

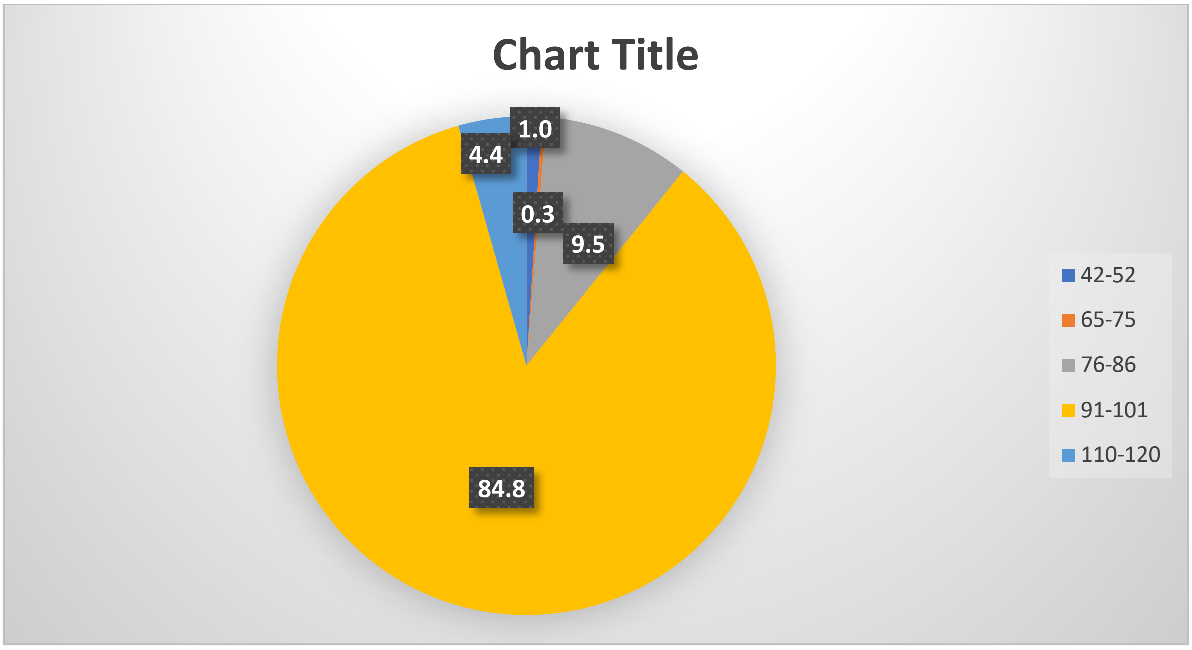


Figure 40: Percentage of RPM Range for Different Stable Operational Profiles

From Figure 40 it can be seen that the ship was operating in a stable operational profile mostly in the 91-101 RPM range at 85% of the time followed by the 76-86 RPM range at only 10% of time. Overall, the range 76-120 RPM contains 99% of the ship's stable operational profiles.

4.4.5 DG1 Maintenance

Because the proposed CBM system requires the machinery to have a starting reliability very close to 100%, the starting reliability of the machinery under evaluation needs to be explored. For that reason, the time of the last major maintenance performed to the DG1 cylinders was searched from the maintenance logs kept onboard the ship. Because the maintenance log does not contain any major overhaul activity for the DG1 cylinders in the period of which the maintenance log is available, it will be assumed that the time of major maintenance was the time that the ship overtaken the last drydocking, namely: 17/06/2017. Table 21 shows the details concerning maintenance of DG1 cylinders.

Table 21: Overhauling Date for DG1 Cylinders

	Assumed Date of Overhauling	**Date that Data Gathering Started**	**Months between Overhauling and Data Gathering Activity**	**Manufacturer suggestion between maintenance**	**Time left until maintenance according to manufacturer**
DG1 Cylinders	17/06/2017	02/04/2018	9 months	2 years	1 year and 1 month

The months between the overhauling and the data gathering activity in column 5 in Table 21 are very important to determine the uncertainty of CBM system's prediction. That is because the longer the period between the previous major maintenance and the time the data gathering activity started, the more probable is that deterioration has already happened without being monitored in the between time. That fact compromises the assumption of the CBM system that the machinery is close to 100% reliable when the data gathering activity started. Finally, in column 6 in Table 21 the time left until maintenance from the date that data gathering activity starts is calculated, from the information from manufacturer's manual that the DG needs maintenance at 1 year and 11 months.

4.4.6 DG1 Detected Deteriorations

The total changes of condition (deteriorations) of the DG1 cylinders for the data gathering period: 03/04/2018-02/07/2018 were found to be four, namely:

- 2018-04-19 09:22:00
- 2018-04-24 17:00:00
- 2018-04-29 04:13:00
- 2018-05-18 22:31:00

Figure 41 shows how deterioration happened for DG1 cylinders during the time under question.

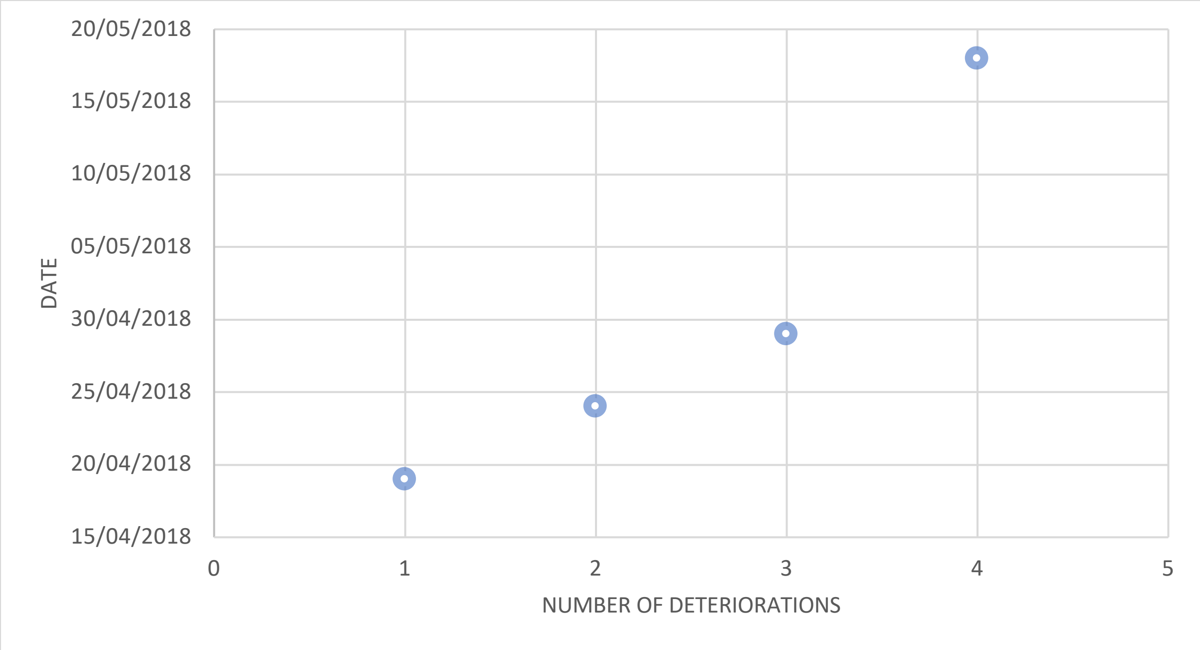


Figure 41: Time Map of Deteriorations-DG1

From Figure 41 it can be seen that most of deterioration happened only during the first two months of the data gathering activity showcasing the randomness of the deterioration process and showcasing why the maintenance problem cannot be solved by fixed time intervals.

4.4.7 DG1 Results

One of the core assumptions of the experiment is the initial reliability of the DG1 cylinders. As it was shown in 5.4.5 there is a gap between the last major maintenance and the data gathering activity which increases the probability that deterioration has already happened in the between time. For that reason, a sensitivity analysis of the assumed initial reliability of the machinery will be performed when calculating the time of failure. The first experiment assumes that the reliability of the machinery is almost as good as new and then the same experiment is repeated for different assumed reliabilities each time reduced by 10% until the initial reliability condition is as low as 10%. Table 22 shows the sensitivity analysis of the assumed initial reliability of for DG1, and the final time of failure predicted by the CBM system.

Table 22: Sensitivity analysis on the Starting Condition of DG1

Experiment	**Initial Reliability (%)**	**Time of failure**
A	99.9%	10 years and 3 months
B	90%	2 years and 4 months
C	80%	1 year and 9 months
D	70%	1 year and 6 months
E	60%	1 year and 4 months
F	50%	1 year and 3 months
G	40%	1 year and 2 months
H	30%	1 year and 2 months
I	20%	1 year and 1 months
J	10%	1 year and 1 months

From Table 22 it can be seen that when the reliability of the machinery is assumed to be very high the prediction of the time of failure becomes bigger. That is logical since it is assumed that between the date of previous major maintenance and the time of the data gathering (9 months) the machinery does not deteriorate at all which indicates a very slow deterioration process. As the assumed reliability of the machinery becomes smaller the predictions are not fluctuating that much indicating that this is closer to the reality. For better visualization Figure 42 shows the predictions of the model for different assumptions of the initial reliability of the machinery when the data gathering activity started.

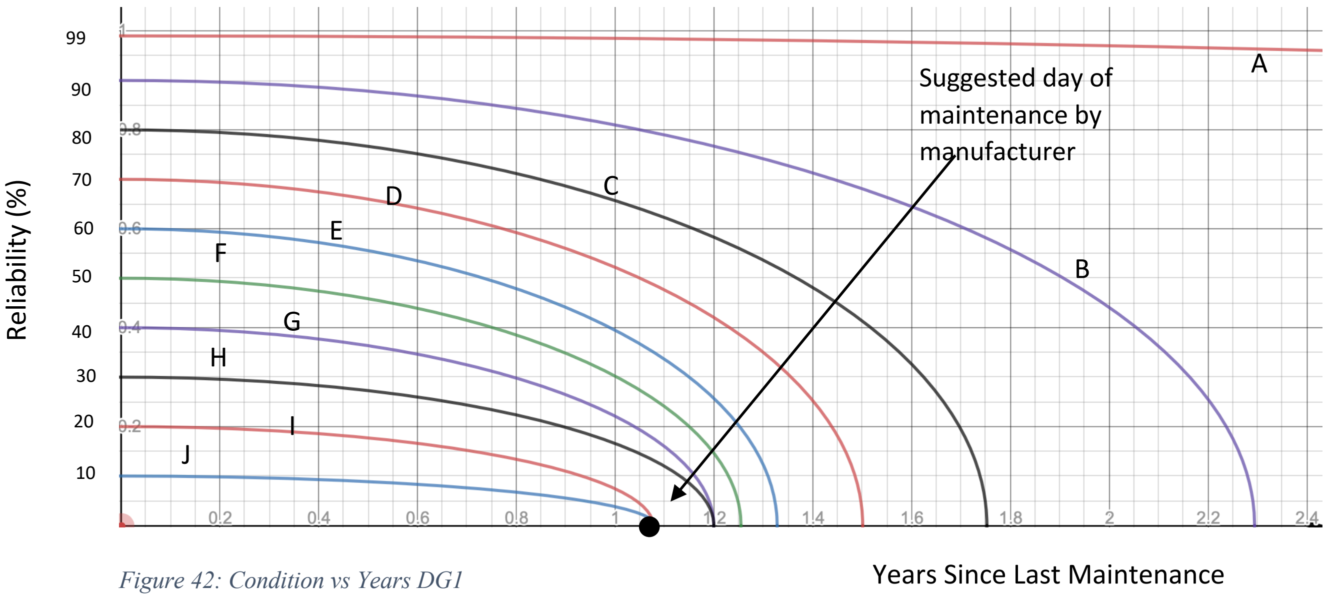


Figure 42: Condition vs Years DG1

From Figure 42 it can be seen that the CBM system suggests maintenance at the time when the manufacturer suggests, if the machinery was at 10-20% reliability when the data gathering activity started. This is very unlike to be true since only 9 months have passed since the major overhaul. That means that the time of maintenance could be increased further of what the manufacturer suggests in the vast majority of the experiments performed. Table 23 shows the time difference between the model's prediction and manufacturer's prediction for the different assumed conditions of the machinery when the data gathering activity started.

Table 23: Time Difference from Manufacturer Maintenance-DG1

Experiment	**Time difference from manufacturer recommendations**
A	9 years and 2 months
B	1 years and 3 months
C	8 months
D	5 months
E	3 months
F	2 months
G	1 months
H	1 months
I	0 months
J	0 months

From Table 23 the results of the corner cases A and J are disregard the average increase in maintenance time is a bit more than 4 months. This is a considerable amount of time and during the lifespan of the cylinders it can accumulate to be even years of extra working time giving great efficiency in the operation of the ship.

4.4.8 DG2 Maintenance

Moving on to the DG2, the maintenance log indicated that major overhauling for the cylinders happened on the 25/08/2017. That means that 7 months had passed until when the data gathering activity started and 1 year and 3 months remaining for overhauling according to the manufacturer's recommendations. Table 24 summarizes the results of DG2 cylinders' maintenance schedule.

Table 24: Overhauling Date for DG2 Cylinders

	Assumed Date of Overhauling	**Date that Data Gathering Started**	**Months between Overhauling and Data Gathering Activity**	**Manufacturer suggestion between maintenance**	**Time left until maintenance according to manufacturer**
DG2 Cylinders	25/08/2017	02/04/2018	7 months	2 years	1 year and 3 months

Again, as it was for DG1, there is a gap between the last major maintenance and the data gathering activity. This problem will be addressed in the sensitivity analysis of the initial reliability of the cylinders later on.

4.4.9 DG2 Detected Deteriorations

The total changes of condition (deteriorations) of the DG2 cylinders for the data gathering period: 03/04/2018-02/07/2018 were found to be four, namely:

- 2018-04-19 08:54:00
- 2018-04-24 16:12:00
- 2018-04-29 02:19:00
- 2018-05-18 22:29:00

Figure 43 shows how deterioration happened for DG2 cylinders during the time under question.

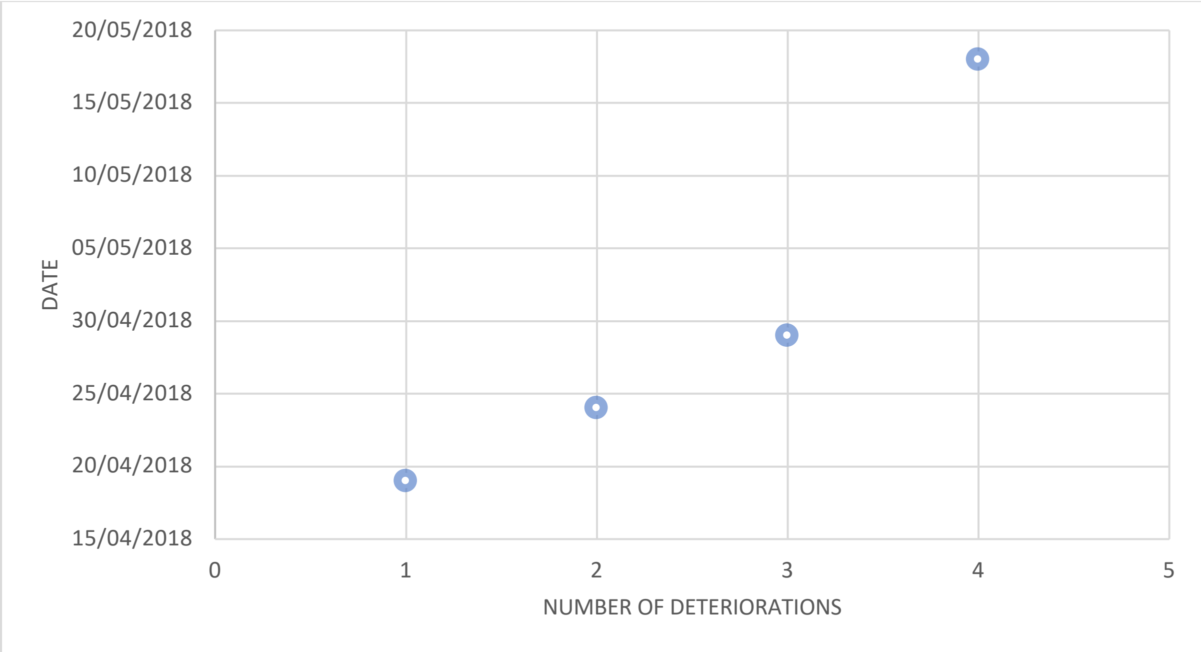


Figure 43: Time Map of Deteriorations-DG2

From Figure 43 it can be seen that most of deterioration happened during the first two months of the data gathering activity at the same days as DG1 but different hours. The deterioration process is expected to be similar since the DGs are the same model and are overhauled in dates close to each other. Nevertheless, as it will be shown in DG3 the deterioration process can be different even for similar machinery.

4.4.10 DG2 Results

Table 25 shows the sensitivity analysis of the assumed initial reliability condition for DG2 and the time of failure predicted by the CBM system as it was done for DG1. The first experiment assumes that the reliability of the machinery is almost as good as new and then the same experiment is repeated for different assumed reliabilities each time reduced by 10% until the initial reliability condition is as low as 10%.

Table 25: Sensitivity Analysis on the Starting Condition of DG2

Index	**Initial Reliability (%)**	**Time of failure**
A	99.9%	8 years
B	90%	1 years and 10 months
C	80%	1 years and 5 months
D	70%	1 years and 2 months
E	60%	1 years and 1 months
F	50%	1 year
G	40%	11 months
H	30%	11 months
I	20%	11 months
J	10%	10 months

From Table 25 it can be seen that when the condition is very high the prediction of the time of failure becomes bigger. As for DG1, this is because it was assumed that between the date of previous maintenance and the time of the data gathering (which is 7 months) the machinery did not deteriorate. The more it is assumed that deterioration has not happened the bigger the time of failure becomes since the model understands that the degradation mechanism is very slow. Figure 44 is a plot of the predictions of the model in years from previous maintenance and for different assumptions of the condition of the machinery when the data gathering activity started.

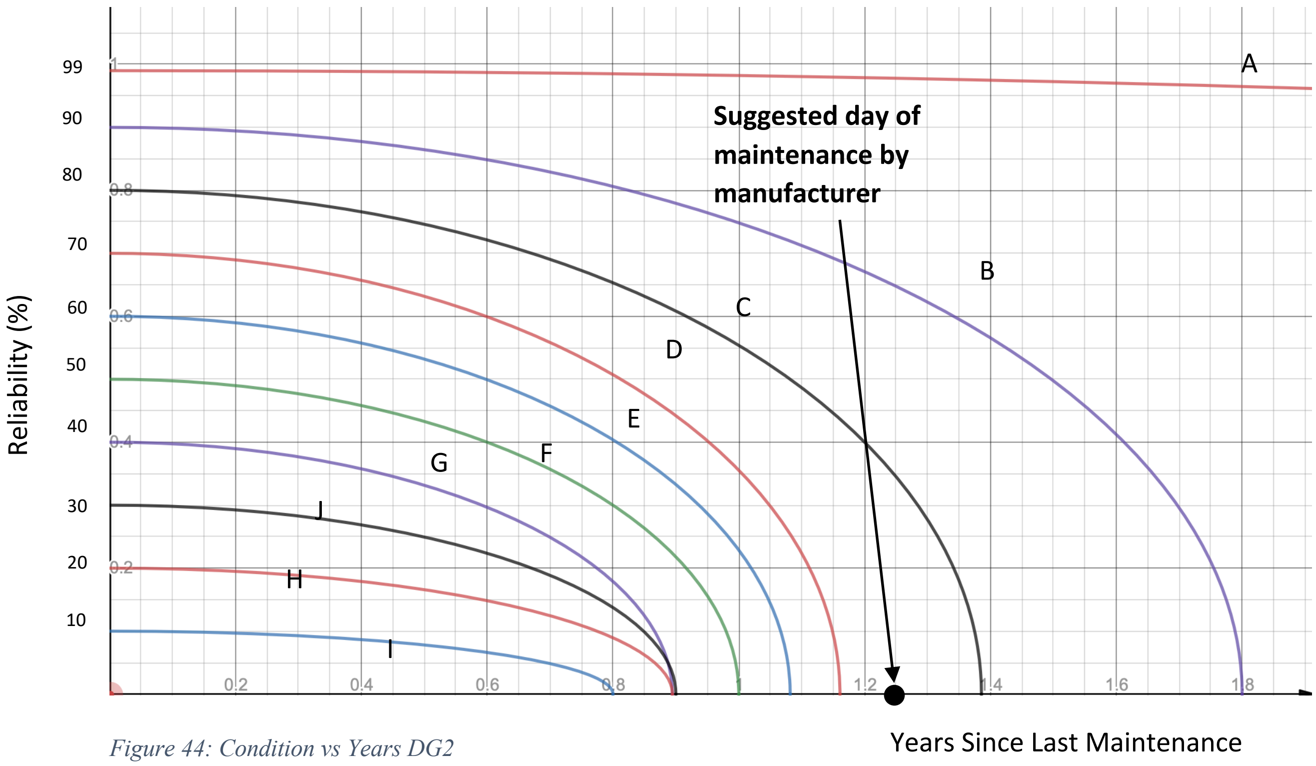


Figure 44: Condition vs Years DG2

From Figure 44 it can be seen that the CBM system suggests maintenance at the time when the manufacturer suggests, if the machinery was between 60% and 70% reliability when the data gathering activity started. This is possible to be true because 7 months have passed since the major overhaul. On the other hand, the time of failure can be lower than the manufacturer suggestion if the machinery bellow 60% reliability when the data gathering activity started which may suggest that the equipment can fail before the next scheduled maintenance. Table 26 shows the time difference between the model's prediction and manufacturer's prediction for the different assumed conditions of the machinery when the data gathering activity started.

Table 26: Time Difference from Manufacturer Maintenance -DG2

Assumed starting condition	**Time difference from manufacturer recommendations**
A	6 years and 7 months
B	7 months
C	2 months
D	-1 months
E	-2 months
F	-3 months
G	-4 months
H	-4 months
I	-4 months
J	-5 months

From Table 26 the results of the corner cases A and J are disregard the average difference from the maintenance time suggested by the manufacturer is one month less. That means that is a big probability that the machinery will fail before the suggested maintenance intervals of the manufacturer. This could potentially cost to the shipowner a lot of costs associated with unplanned maintenance and increased downtime that could be avoided with the use of the proposed CBM system.

4.4.11 DG3 Maintenance

Finally, moving on to DG3, the maintenance log indicated that major overhauling for the cylinders happened on the 11/04/2017 and thus approximately 1 year had passed until when the data gathering activity started and 1 year was remaining for overhauling according to the manufacturer's recommendations. Table 27 summarizes the results of DG2 cylinders' maintenance schedule.

Table 27: Overhauling Date for DG3 Cylinders

	Assumed Date of Overhauling	Date that Data Gathering Started	Months between Overhauling and Data Gathering Activity	Manufacturer suggestion between maintenance	Time left until maintenance according to manufacturer
DG3 Cylinders	11/04/2017	03/04/2018	1 year	2 years	1 year

Again, as it was for DG1 and DG2, there is a gap between the last major maintenance and the data gathering activity. Actually, this period is the highest between the three cases (1 year) and a sensitivity analysis of the initial reliability of the cylinders will help understand how it affects predictions.

4.4.12 DG3 Detected Deteriorations

The total changes of condition (deteriorations) of the DG3 cylinders for the data gathering period: 03/04/2018-02/07/2018 were found to be four, namely:

- 2018-04-18 20:37:00
- 2018-04-21 00:30:00
- 2018-04-28 20:43:00
- 2018-05-03 11:30:00

Figure 45 shows how deterioration happened for DG3 cylinders during the time under question.

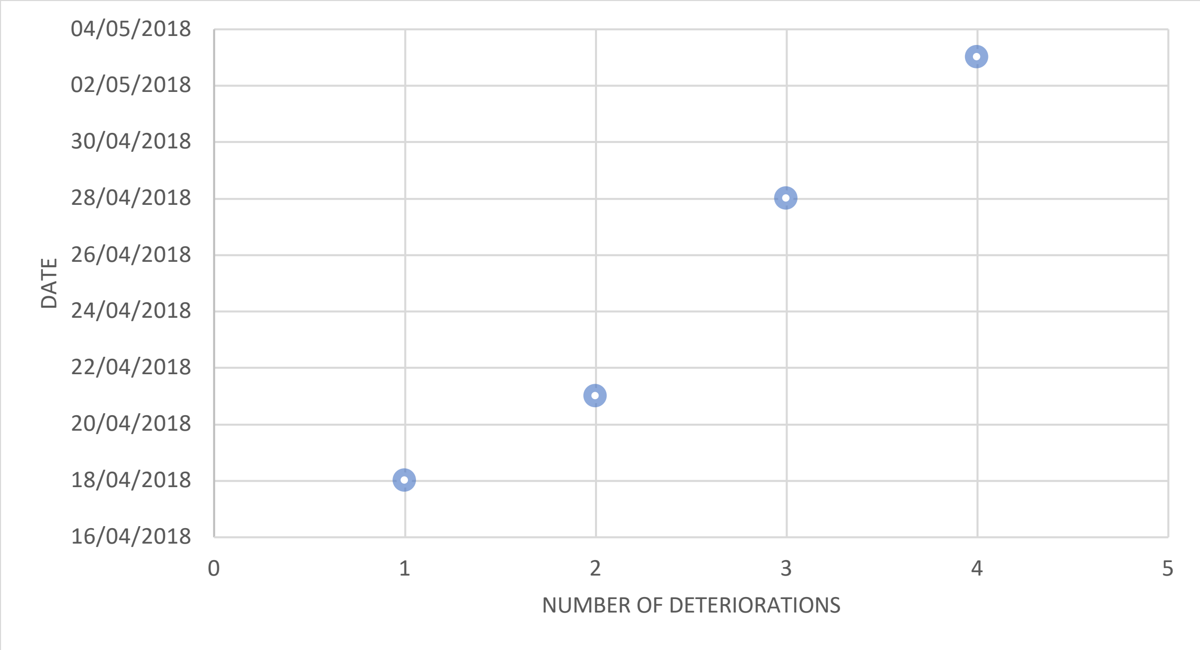


Figure 45: Time Map of Deteriorations-DG3

From Figure 45 it can be seen that most of deterioration happened during the first two months of the data gathering activity but on different days than DG1 or DG2. That means that the deterioration pattern even for identical machines can be different. The traditional scheduled interval maintenance cannot account for these differences but the proposed CBM system, by using live analysis of data, is shown to catch these small differences that can have a big impact in the maintenance schedule and failure time of the machinery.

4.4.13 DG3 Results

Table 28 shows the sensitivity analysis of the assumed initial reliability condition for DG3 and the time of failure predicted by the CBM system as it was done for DG1 and DG2. Again, the first experiment assumes that the reliability of the machinery is almost as good as new and then the same experiment is repeated for different assumed reliabilities each time reduced by 10% until the initial reliability condition is as low as 10%.

Table 28: Sensitivity Analysis on the Starting Condition of DG3

Index	**Initial Reliability (%)**	**Time of failure**
A	99.9%	12 years and 5 months
B	90%	2 years and 10 months
C	80%	2 years and 1 months
D	70%	1 years and 10 months
E	60%	1 years and 7 months
F	50%	1 years and 6 months
G	40%	1 years and 5 months
H	30%	1 years and 4 months
I	20%	1 years and 4 months
J	10%	1 years and 4 months

From Table 28 it can be seen that from all three DGs this one gives the largen time of failure if it is assumed that the condition has remained the same in the period between the date of previous maintenance and the time of the data gathering. That is because this period is the biggest, namely: 1 year. Figure 46 shows a plot of the predictions of the model in years from previous maintenance and for different assumptions of the condition of the machinery when the data gathering activity started.

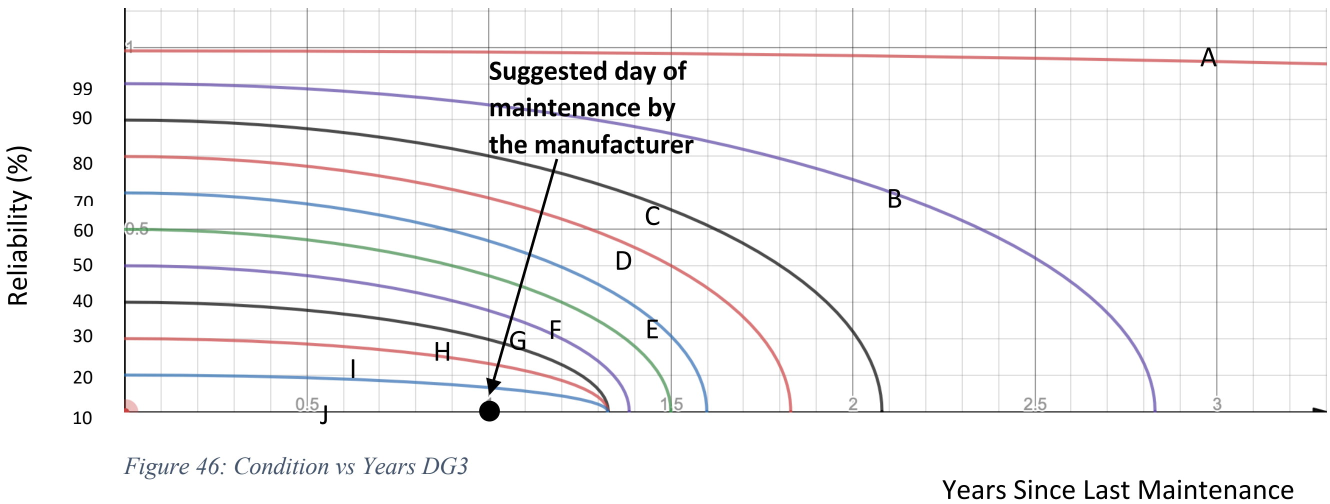


Figure 46: Condition vs Years DG3

From Figure 46 it can be seen that the CBM system suggests that failure will be later than when the manufacturer suggests for any initial reliability assumption. For example, if it is assumed that the machinery has 50% reliability when the data gathering activity started, the gain in time of maintenance will be six months, which is very significant. Table 29 shows the time difference between the model's prediction and manufacturer's prediction for the different assumed conditions of the machinery when the data gathering activity started.

Table 29: Time difference for manufacturer maintenance and model estimate DG3

Assumed starting reliability condition	**Time difference from manufacturer recommendations**
A	11 years and 5 months
B	1 year and 10 months
C	1 year and 1 months
D	10 months
E	7 months
F	6 months
G	5 months
H	4 months
I	4 months
J	4 months

From Table 29, if the results of the corner cases A and J are disregard the average difference from the maintenance time suggested by the manufacturer is over 9 months. The gain in operational time as the assumed reliability is higher can be very significant reaching even more than a year of extra time, (e.g., when the reliability is at 80%) but also in the lowest assumption of reliability (20%) the CBM system gives extra 4 months until maintenance. That means that there is a big probability that the machinery will fail much after the suggested maintenance intervals of the manufacturer which it will result in unnecessary maintenance of the cylinders increasing also the probability of human errors during the maintenance.

4.4.14 Conclusion of Case Study 3

From the results of the final case study in the maintenance problem of the DG1, DG2 and DG3 cylinders it is clear that the proposed CBM system could give an increased operational time from what the manufacturer suggested as well as a warning when the machinery fails beforehand. This shows that the estimations for the maintenance intervals from the manufacturers are either very conservative and do not take into account the real condition of the machinery at each maintenance interval or cannot generalize in cases where there is faster degradation leading to unplanned maintenance and downtime. Thus, the proposed CBM system could greatly benefit the shipowners and operators by giving them the ability to be in charge of the maintenance schedule and either avoid the costs of unplanned maintenance and failure or reduce significantly the number of times that maintenance needs to be performed and increase the operational time of the machinery.

More specifically from the results of Case Study 3:

In the maintenance problem of the DG1 cylinders the results showed that the reliability of the machinery must be very low in order to justify maintenance according to the manufacturer, something that is very unlikely and means that maintenance will be performed too early if the fixed time-intervals of the manufacturer are followed.

For DG2, the results suggested that there is a probability the cylinders can give an early failure leading to unplanned maintenance which will cost considerable to the operators.

In DG3 the CBM system identified that there was no reason to perform maintenance in the cylinders when the manufacturer suggested since even for the lowest assumed reliability the time of failure is months after when the manufacturer suggests.

From all the above results it is shown that the proposed CBM system clearly overcomes the drawbacks of using the most common method of maintenance now in shipping, namely: fixed time intervals. The fixed time intervals suggested by the manufacturers are usually too strict or do not consider cases of increased degradation since there is no real-time assessment of the machinery leading to most of the times to unnecessary maintenance (which increases infant mortality) or even worse to unplanned break down of the machinery. In addition, the proposed

CBM system has numerous advantages compared to other CBM systems proposed in the literature. The main advantage is the elimination of the requirement of large amount of historical data and/or failure data needed for each machinery that the CBM system will monitor. In contrast, the proposed system has a general applicability to all machinery, using data on the fly and calculating the time of failure in a personalised manner. Furthermore, none of the proposed CBM systems in current literature take into account the operational profiles of the ships which mean they have limited applicability in real life scenarios where a ship changes operational profiles constantly through its life (port, manoeuvring, on cruise, different payload, wind conditions, etc.).

5. Discussion

The problem of maintenance for shipping is a very crucial one. Not only it is paramount for the safety of everyone onboard but also for the smooth everyday operation of the ship. From all the maintenance techniques available CBM is one of the newest and has the potential to change the way maintenance has been applied for years. That is because a CBM systems consider the condition of the machinery before suggesting when to perform maintenance rather than relying to fixed time intervals suggested by the manufacturers.

On the other hand, data in the marine sector are usually being used for manually monitoring alarms and trends on-board with some cases of simple automated data diagnostic and prognostic procedures. CBM systems have not been applied extensively because of the poor practical value of the systems being proposed since they require huge amounts of historical data (including failure rate data) which are notoriously difficult to find, they do not take into account the existence of different operational profiles of a ship (port manoeuvring, bad weather, cruising, etc.) and promote the mismatch between relational databases and the algorithms.

The purpose of this thesis was to suggest a novel CBM framework which combines increased practical value and that addresses the above-mentioned shortcomings of the traditional CBM systems until now. To do that, a thorough investigation on every aspect of a CBM system was performed, namely: The database types, the machine learning algorithms and finally the failure prediction model. From that investigation it was concluded that the new proposed CBM system will be consisted of an object-oriented database, a semi-supervised algorithm to detect deterioration and an innovate failure model derived from the P-F diagrams.

The object-oriented database not only solves the mismatch problem between the traditional used relational databases and the machine learning algorithms but also identifies the stable operational profiles of the ship and stores data selectively for minimum storage need and maximum efficiency in predicting deterioration.

The machine learning part of the CBM system relies on the LOF algorithm. From all the semi-supervised algorithms reviewed in the literature review the LOF algorithm was chosen because of its great general applicability as well as its demonstrated effectiveness compared to other algorithms. The LOF algorithm will be trained each time a threshold of amount of data is reached, sufficient enough to represent the current condition of the machinery when the ship is on a stable operational profile, and then it will assess the condition of the machinery for every new data point coming from the sensors for that operational profile.

That failure model of the CBM will have as an input the time(s) that deterioration happened to the machinery (given by the LOF algorithm) and then it will give an estimation of the time of failure that will become more and more precise the more data are collected. This new innovative theory is based on solving mathematically the P-F diagrams that have been used widely for condition monitoring. Furthermore, the innovate failure model does not need any historical data or failure data and thus it increases the practical value of the system exponentially.

All three aspects of the CBM were tested individually and as a whole through the case studies presented in this thesis. Firstly, the machine learning algorithm was evaluated against predicting the degradation of a compressor and a turbine of a naval propulsion system. In this case study. It was shown that the LOF algorithm has high accuracy in predicting if the condition has deteriorated or not for both cases of the turbine and compressor showcasing the general applicability of the algorithm. In addition, the number of training data for both of the machinery was found to be substantial mainly because of the weak assumption that a stable operational profile can be defined only by the ship's speed. Finally, a sensitivity analysis of the ship speed was performed to establish if the algorithm performs well for different ship-speeds. The sensitivity analysis showed that the algorithm is robust enough to perform with high accuracies across all the operational profiles of the ship and for both the turbine and compressor cases. That means that it can be used to continuously monitor the degradation of the machinery throughout all the operational profiles of a ship and thus increasing the reliability and accuracy of the diagnostic and prognostic system.

Next, the failure model was evaluated against predicting correctly the time of failure of bearings. The purpose of this case study was two-fold: The parameter "a" was experimentally established by using one of the bearing failure datasets and secondly to compare the predictions of the failure model with the real failure time in order to evaluate its effectiveness. The parameter "a" was established by using the first dataset and its value was validated with the remaining two datasets. The results of the case study showed that:

- The failure prediction becomes more and more precise as more data are being fed to the condition-based system.
- The failure model is able to correctly predict the time of failure in a good time in advance before it happens (final prediction at 65% of the time of failure).
- 26% gain of effective working time versus traditional maintenance techniques
- Fast training time (3 hours when we have 10 min intervals).

Finally, the database, the machine learning algorithm and the failure model were used in a real-world case study in order to predict the time of failure of three diesel generators of a ship. The 3-month data from multiple sensors on-board a tanker ship was used to evaluate the condition of the cylinders of three Diesel Generators (DGs). In order to be able to perform condition monitoring the Condition monitoring Database (CMD) used weather data in combination with data of ship's speed etc. in order to accurately understand different operational profiles of the ship and accordingly perform correct condition monitoring.

Concerning the DGs cylinders data, there was always a gap between the last major maintenance and the data gathering activity which increased the probability that deterioration has already happened in the between time. For that reason, a sensitivity analysis of the assumed initial reliability of the cylinders was performed for the cylinders of each DGs when calculating the time of failure. For the sensitivity analysis, first it was assumed that the reliability of the DG's cylinders was almost as good as new when the data gathering activity started and then the time of failure was calculated. Then the same experiment is repeated for different assumed starting reliabilities for the cylinders, each time reduced by 10% until the initial reliability condition is as low as 10%.

For the cylinders of DG1 the CBM system suggested maintenance at the time when the manufacturer suggests only when the initial reliability of the cylinders was very low when the data gathering activity started. Because this is very unlikely to be the case, the time of maintenance could be increased further of what the manufacturer suggests in the vast majority of the sensitivity analysis cases.

For DG2 the results suggested that there was a big probability that the cylinders will fail before the suggested maintenance intervals of the manufacturer. This could potentially lead to costs associated with unplanned maintenance and increased downtime.

Finally, for DG3 the results showed that in the majority of the cases from the sensitivity analysis the machinery will fail much after the suggested maintenance intervals of the manufacturer which it will result in unnecessary maintenance of the cylinders.

Overall, the real-world case study results showcased the problems associated with using fixed time intervals for maintenance. More specifically, the manufacturers suggestions are too strict and also do not consider cases of increased degradation since there is no real time assessment of the machinery. That is why the proposed CBM system will provide increased value to shipowners and operators allowing them to monitor machinery in real time and enhancing greatly the decision-making process for maintenance, reducing substantially the costs related with unplanned maintenance and increasing the effective working time of the machinery.

6. Conclusion

In conclusion, this thesis had as an aim to introduce a new practical CBM system that is effective in assessing the condition of a machinery continuously and give predictions of the estimated time of its failure. It does that by using a flexible Object-Oriented database for storing data, a semi-supervised machine learning technique for diagnosis and utilizing a new failure theory for prognostics.

Through research it was shown that in shipping fixed time intervals are still predominantly used for maintenance while CBM systems have not yet been able to penetrate the market and change the maintenance status quote mainly because of the requirement of historical data for each machinery to be monitored and/or the inability to continuously monitor the machinery through different operational profiles (different payloads, weather conditions, ship speeds, etc.).

Based on the analysis of real-world data collected after three months of operation of a tanker ship for the prediction of maintenance of the cylinders of the ship's DGs it was shown that according to the CBM systems predictions fixed time intervals are often too strict leading to loss of hundreds or even thousands of additional operational times as well as introducing high risk of unnecessary infant mortality. In addition, no historical data or failure data were needed for the CBM system to start generating these predictions. On the contrary, the assess of the machinery is happening online and can be applied to other machinery onboard of a ship thanks to the new proposed failure theory utilized for prognostics. Furthermore, it was also demonstrated by the real-word case study that by using a machine-learning friendly database, the proposed CBM system can continuously monitor the health of a machinery even if the ship changes operational profiles.

For future work:

- Further testing of the whole CBM framework will be done with data collected from sensors on-board vessels.
- Real time imputation of missing values.
- Expand the framework to incorporate weather routing and fuel consumption.

- Testing of the proposed CBM system in a real-world implementation on-board a ship including, but not limited: on-line sensor values and stored by the CMD, on-line training of the algorithm by using the CMD data and performing diagnostics and prognostics with the innovative modelling of general machinery failure.

References

Abourezq M. & Idrissi A. (2016). *Database-as-a-Service for Big Data: An Overview*. International Journal of Advanced Computer Science and Applications. 7. 10.14569/IJACSA.2016.070124.

ABS (2016). Guidance notes on equipment condition monitoring techniques.

ABS, (2018). *Guide for Risk-Based Inspection for Floating Offshore Installations*. American Bureau of Shipping Plaza16855 Northchase DriveHouston, TX 77060 USA.

Achlioptas D. (2003). *Database-friendly random projections: Johnson-Lindenstrauss with binary coins*. Journal of Computer and System Sciences 66 (2003) 671–687.

Ademujimi T. & Brundage M. & Prabhu V. (2017). *A Review of Current Machine Learning Techniques Used in Manufacturing Diagnosis*. 407-415. 10.1007/978-3-319-66923-6_48.

Ademujimi T.T., Brundage M.P., Prabhu V.V. (2017). *A Review of Current Machine Learning Techniques Used in Manufacturing Diagnosis*. IFIP Advances in Information and Communication Technology, 513 , pp. 407-415.

Ahmad, R., & Kamaruddin, S. (2012). *An overview of time-based and condition-based maintenance in industrial application*. Computers & Industrial Engineering, 63(1), pp. 135-149.

Ahmed M. & Seraj R. & Islam S.. (2020). *The k-means Algorithm: A Comprehensive Survey and Performance Evaluation*. Electronics. 9. 1295. 10.3390/electronics9081295.

Ajita S. & Patel, Dr R.. (2009). *Use of Object-Oriented Concepts in Database for Effective Mining. International Journal on Computer Science and Engineering*. 1.

Alghamdi R. (2016). *Hidden Markov Models (HMMs) and Security Applications*. International Journal of Advanced Computer Science and Applications (IJACSA), 7(2), 2016. http://dx.doi.org/10.14569/IJACSA.2016.070205.

Alhouli, Y., Ling D., Kirkham, R., Elhag, TM (2009). *On the Factors Affecting Maintenance Planning in the Mercantile Industry*. COMADEM2009, 22nd International Congress on Condition Monitoring and Diagnostic Engineering Management, San Sebastian, Spain.

Allen, T.M, (2001). *US Navy Analysis of Submarine Maintenance Data and the Development of Age and Reliability Profiles*. Submarine Maintenance Engineering, Planning and Procurement (SUBMEPP).

Alsyouf I., Shamsuzzaman M., Abdelrahman G., Al-Taha M. (2016). *Improving reliability of repairable systems using preventive maintenance and time-between-failures monitoring*. European Journal of Industrial Engineering, Inderscience Enterprises Ltd, vol. 10(5), pages 596-617.

Amruthnath and Gupta (2018). *A research study on unsupervised machine learning algorithms for early fault detection in predictive maintenance*. In 2018 5th International Conference on Industrial Engineering and Applications (ICIEA) (pp. 355-361), IEEE (2018).

Anderson, J.A. and Rosenfeld E. (1988). *Neuro computing: Foundation of Research*. Cambridge, MA: The MIT Press.

API (American Petroleum Institute) (2008) Risk-Based Inspection Technology. 2nd Edition.

API RP 581, Washington, DC, EUA, September 2008.

Asadi S.D., Subba Rao D.V., Saikrishna V. (2010). *A comparative study of face recognition with principal component analysis and cross-correlation technique*. Int J Comput Appl 10.

ASQL (1983). *Statistics Division Newsletter*. Volume 4, Number 1, February 1983.

Atamuradov V. & Medjaher K. & Dersin P. & Lamoureux B. & Zerhouni N. (2017). *Prognostics and Health Management for Maintenance Practitioners-Review,* *Implementation and Tools Evaluation*. International Journal of Prognostics and Health Management. 8. 31.

Atamuradov V., Medjaher K., Dersin P., Lamoureux B. (2017). *Prognostics and Health Management for Maintenance Practitioners-Review, Implementation and Tools Evaluation*. International Journal of Prognostics and Health Management 8(Special Issue on Railways & Mass Transportation):31.

Atsma W.J., & Hodgson, A.J. (1999). *Inferring motor plan complexity using a modified principal component analysis*. Engineering in Medicine and Biology, 21st Annual Conference,1, 533.

Atzeni, P., Ceri, S., Paraboschi, S., Torlone, R. (1999). *Database systems : concepts, languages & architectures*. The McGraw-Hill Companies, London ; New York

Audisio R.A., Bozzetti F., Gennari R., Jaklitsch M.T., Koperna T., Longo W.E., Wiggers T., Zbar A.P. (2004). *The surgical management of elderly cancer patients: recommendations of the SIOG surgical task force.* European Journal of Cancer Volume 40, Issue 7, May 2004, Pages 926-938.

Awad M., Khanna R. (2015). *Support Vector Machines for Classification*. In: Efficient Learning Machines. Apress, Berkeley, CA. https://doi.org/10.1007/978-1-4302-5990-9_3.

Bagadia K. (2006). *Computerized Maintenance Management Systems Made Easy: How to Evaluate, Select and Manage CMMS*. McGraw-Hill Education; Illustrated edition.

Bai Y. and Jin W.L. (2016). *Risk-Centered Maintenance.* Marine Structural Design (Second Edition) 2016, Pages 803-825.

Banerjee J., Chou H.T., Garza J., Kim W., Woelk D., Ballou N. and. Kim H. (1987). *Data model issues for object-oriented applications*. ACM TOIS, January 1987.

Bara A. & Niu X. & Luk W. (2014). *A Dataflow System for Anomaly Detection and Analysis*. Proceedings of the 2014 International Conference on Field-Programmable Technology, FPT 2014. 10.1109/FPT.2014.7082793.

Barbierato, E., Gribaudo, M., Iacono, M. (2014). *Performance evaluation of NoSQL big-data applications using multi-formalism models*. Future Gener. Comput. Syst. 37, 345–353.

Barbieri, P., Adami G., Piselli S., Gemiti F., & Reisenhofer E., (2002). *A three-way principal factor analysis for assessing the time variability of freshwaters related to a municipal water supply*. Chemometrics and Intelligent Laboratory Systems, 62(1), 89-100.

Basim A.N., Alsyouf I. (2003). *Selecting the most efficient maintenance approach using fuzzy multiple criteria decision making.* International Journal of Production Economics

Batur C. & Zhou L. & Chan C.-C. (2003). *Support vector machines for fault detection*. Proceedings of the IEEE Conference on Decision and Control. 2. 1355 - 1356 vol.2. 10.1109/CDC.2002.1184704.

Bayer S., (2016). *A Study on Cost Optimization in the Ship Management*. The Second Global Conference on Innovation in Marine Technology and the Future of Maritime Transportation, 24-25 October 2016, Bodrum, Muğla, TURKIYE.

Beal M. & Rasmussen C. (2002). *The Infinite Hidden Markov Model*.

Beech D. (1988). *A foundation for evolution from relational to object databases*. In: Schmidt J.W., Ceri S., Missikoff M. (eds) Advances in Database Technology—EDBT '88. EDBT 1988. Lecture Notes in Computer Science, vol 303. Springer, Berlin, Heidelberg. https://doi.org/10.1007/3-540-19074-0_57.

Ben-Daya M. (2009). *Failure Mode and Effect Analysis*. In: Ben-Daya M., Duffuaa S., Raouf A., Knezevic J., Ait-Kadi D. (eds) Handbook of Maintenance Management and Engineering. Springer, London. https://doi.org/10.1007/978-1-84882-472-0_4.

Bengtsson M. (2004). *Condition Based Maintenance System Technology – Where is Development Heading?* Conference: International Conference of Euromaintenance 2004 At: Barcelona, Spain Volume: 17th.

Bengtsson M. (2004). *Condition Based Maintenance Systems an Investigation of Technical Constituents and Organizational Aspects*.

Bengtsson M. (2007). *On Condition Based Maintenance and its Implementation in Industrial Settings.* Malardalen University Press Dissertations No.48.

Bhagat A. & Kshirsagar N. & Khodke P. & Dongre K. & Ali S. (2016). *Penalty Parameter Selection for Hierarchical Data Stream Clustering.* Procedia Computer Science. 79. 24-31. 10.1016/j.procs.2016.03.005.

Birgelen v. A. & Buratti D. & Mager J. & Niggemann O. (2018). *Self-Organizing Maps for Anomaly Localization and Predictive Maintenance in Cyber-Physical Production Systems*. Procedia CIRP. 72. 480-485. 10.1016/j.procir.2018.03.150.

Bisson G. & Blanch R. (2012). *Improving Visualization of Large Hierarchical Clustering*. Proceedings of the International Conference on Information Visualisation. 220-228. 10.1109/IV.2012.45.

Bloch H., Geitner F. (1983). *Machinery failure analysis and trouble shooting*. Gulf, Houston, TX.

Breunig M.M., Kriegel, H.P., Raymond T.Ng. and Sander, J. (2000). *LOF: Identifying Density-Based Local Outliers.* Proc. ACM SIGMOD 2000 Int. Conf. On Management of Data, Dalles, TX, 2000.

Broberg (1973). *Broberg's report cited in Failure Diagnosis & Performance Monitoring*. Vol. 11 edited by L.F. Pau. Marcel-Dekker, 1981.

Butcher SW (2000). *Assessment of Condition-Based Maintenance in the Department of Defense*. Technical Report; 2000.

Bzdok, D. (2018). *Statistics versus Machine Learning*. Nature Methods-April 2018.

Caesarendra W., (2010). *Model-based and Data-Driven Approach for Machine Prognostics*. Master Thesis, Pukyoung National University.

Caesarendra W., Niu G., Yang B.S. (2010). *Machine condition prognosis based on sequential Monte Carlo method.* Expert Systems with Applications Volume 37, Issue 3, 15 March 2010, Pages 2412-2420.

Campos G.O., Zimek A., Sander J. and Campello R.J. (2015). *On the Evaluation of unsupervised outlier detection: measures, datasets and an empirical study*. Data Mining and Knowledge Discovery.

Cardoso J.-F. (1997). *Infomax and maximum likelihood for blind source separation*. IEEE Signal Processing Letters, 4(4), 112–114.

Cardoso, J.-F., & Laheld, B. (1996). *Equivariant adaptive source separation*. IEEE Trans. on S.P., 45(2), 434–444.

Carnero M. C. & Novés J. L. (2006). *Selection of computerised maintenance management system by means of multicriteria methods*. Production Planning & Control, 17:4, 335-354, DOI: 10.1080/09537280600704085.

Carpenter G.A., Grossberg S. (1988). *Self-Organizing Neural Network Architectures for Real-Time Adaptive Pattern Recognition*. In: Haken H. (eds) Neural and Synergetic Computers. Springer Series in Synergetics, vol 42. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-74119-7_4.

Carretero, J. et al. (2003). *Applying RCM in large scale systems: a case study with railway networks.* Reliability engineering & system safety, 82, 257-273.

Caseau Y. (1991). *Constraints in an object-oriented deductive database*. In: Delobel C., Kifer M., Masunaga Y. (eds) Deductive and Object-Oriented Databases. DOOD 1991. Lecture Notes in Computer Science, vol 566. Springer, Berlin, Heidelberg. https://doi.org/10.1007/3-540-55015-1_16.

Cheliotis M., Lazakis I., Theotokatos G. (2020). *Machine learning and data-driven fault detection for ship systems operations*. Ocean Engineering Volume 216, 15 November 2020, 107968.

Chelidze, D., J.P. Cusumano, and Charterjee, A. (2002). *A dynamical systems approach to damage evolution tracking, part 1: The experimental method*. Jour. of Vib. and Acoustics, 124, 250–257.

Chen N. and Tsui K.L. (2013). *Condition monitoring and remaining useful life prediction using degradation signals: revisited.* IE Transactions, 45:9, 939-952, DOI: 10.1080/0740817X.2012.706376.

Chen X. & Eversole A. & Li G. & Yu D. & Seide F. (2012). *Pipelined Back-Propagation for Context-Dependent Deep Neural Networks*. 1.

Chen Z.S., Yang Y.M., Hu Z. (2012). *A technical framework and roadmap of embedded diagnostics and prognostics for complex mechanical systems in prognostics and health management system*.

Chinnam R.B., Baruah P. (2004). *A neuro-fuzzy approach for estimating mean residual life in condition-based maintenance systems*. International Journal of Materials and Product Technology 20 (2004) 166–179.

Cholasuke, C., Bhardwa, R., & Antony, J. (2004). *The status of maintenance management in UK manufacturing organisations: results from a pilot survey*. Journal of Quality in Maintenance Engineering, 10(1), 5–15.

Cichocki A., Unbehauen R. (1996). *Robust neural networks with on-line learning for blind identification and blind separation of sources*. IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications.

Cluet S., Delobel C., Lecluse C., Richard P. (1989). *Reloop, an algebra based query language for an object-oriented database system*. Presented at the first International Conference on Deductive and Object-Oriented Databases, Kyoto (1989).

Cocconcelli M., Capelli L., Molano J.C.C., Borghi D. (2018). Development of a Methodology for Condition-Based Maintenance in a Large-Scale Application Field. Machines 2018,6, 17; doi:10.3390/machines6020017.

Codd E.F. (1970). *A Relational Model of Data for Large Shared Data Banks*. Volume13/Number 6/June,1970 Communications of the ACM.

Conachey, R.M. & Montgomery, R.L. (2003). *Application of Reliability-Centered Maintenance Techniques to the Marine Industry*. SNAME conference, 8th April, Texas, Houston.

Coraddu A., Oneto L., Ghio A., Savio S., Anguita D. and Figari M. (2016). *Machine learning approaches for improving condition-based maintenance of naval propulsion plants*. - Journal: Proceedings of the Institution of Mechanical Engineers Part M: Journal of Engineering for the Maritime Environment - Number: 1 - Pages: 136-153 - Volume: 230 - Year: 2016.

Cvejic N. & Bull D. & Canagarajah N. (2007). *Improving Fusion of Surveillance Images in Sensor Networks Using Independent Component Analysis*. Consumer Electronics, IEEE Transactions on. 53. 1029 - 1035. 10.1109/TCE.2007.4341582.

Dadgar, S.A. (2015). *Survey: Hidden Markov Model-Based Approaches for Hand Gesture Recognition*.

Dagher I. (2006). *Incremental PCA-LDA Algorithm.* International Journal of Biometrics and Bioinformatics (IJBB), Volume (4): Issue (2).

Dagkinis I. & Nikitakos N. (2013). *Application of Analytic Hierarchy Process & TOPSIS methodology on ships' maintenance strategies*. University of the Aegean, Dept. of Shipping Trade and Transport, Korai 2a,Chios, GREECE

Daly R. & Shen Q. & Aitken S. (2011). *Learning Bayesian networks: Approaches and issues*. Knowledge Eng. Review. 26. 99-157. 10.1017/S0269888910000251.

Damaševičius R., Vasiljevas M., Šalkevičius J., Woźniak M. (2016). *Human activity recognition in AAL environments using random projections*. Comput. Math. Methods Med. 2016:4073584. 10.1155/2016/4073584.

Dasgupta, S. (2000). *Experiments with random projection*. Uncertainty in Artificial Intelligence: Proceedings of the Sixteenth Conference (UAI-2000) (pp. 143–151). Morgan Kaufmann.

Date C.J. (2012). Database Design and Relational Theory. O'Reilly Media, Inc.

Davies C., Greenough R.M. (2000). *The use of information systems in fault diagnosis*. Proceedings of the 16th National Conference on Manufacturing Research, University of East London, UK.

Dayal, U. (1989). *Active database management systems*. SIGMOD Rec. 18, 3, 150–169.

De S., Das A., Sureka A., (2010). *Product failure root cause analysis during warranty analysis for integrated product design and quality improvement for early results in downturn economy*. International Journal of Product Development 12(3-4), 235-253.

Deegalla S. & Boström H. (2007). *Reducing High-Dimensional Data by Principal Component Analysis vs. Random Projection for Nearest Neighbor Classification*. icmla. 245 - 250. 10.1109/ICMLA.2006.43.

Dekker R. (1996). *Applications of Maintenance Optimization Models: A Reviewand Analysis*. Reliability Engineering & System Safety 51(3): 229-240.

Delfosse N. and Loubaton P. (1995). *Adaptive blind separation of independent sources: a deflation approach*. Signal Processing, 45 (1995), pp. 59-83.

Dempster A. P., Laird N. M., & Rubin D. B. (1977). *Maximum likelihood from incomplete data via the EM algorithm*. Journal of the Royal Statistical Society, 39(1), 1–38.

Dhillon B.S. (2008). *Mining equipment reliability, maintainability, and safety*. Springer Series in Reliability Engineering. Springer, London. https://doi.org/10.1007/978-1-84800-288-3_4.

Dhillon, B. S. & Liu, Y. (2006). *Human error in maintenance: a review*. Journal of Quality in Maintenance Engineering, 12, 21-36.

Diao, J., Allister, E.M., Koshkin, V., Lee, S.C., Bhattacharjee, A., Tang, C., Giacca, A., Chan, C.B., and Wheeler, M.B. (2008). *UCP2 is highly expressed in pancreatic alpha-cells and influences secretion and survival*. Proc. Natl. Acad. Sci. USA 105, 12057–12062.

Dietrich S. W. and Prentice H. (2001). *Understanding Relational Database Query Languages.*

Dieulle L., Berenguer C., Grall A., & Roussignol M. (2001). *Continuous time predictive maintenance scheduling for a deteriorating system*. In Annual reliability and maintainability symposium, Philadelphia, PA, USA (pp. 150- 155).

Dikis, K., Lazakis, I., & Turan, O. (2014). *Probabilistic risk assessment of condition monitoring of marine diesel engines*. Paper presented at ICMT 2014, Glasgow, United Kingdom.

Ding H.& Ding K. & Zhang J. & Wang Y. & Gao L. & Li Y. & Chen F. & Shao Z. & Lai W. (2018). *Local outlier factor-based fault detection and evaluation of photovoltaic system*. Solar Energy. 164. 10.1016/j.solener.2018.01.049.

DNV GL, (2014). *Beyond Condition Monitoring in the Maritime Industry*. DNV GL Strategic Research & Innovation Position Paper 6-2014.

DNV GL, (2015). *Rules for Classification: Part 7 Fleet in service*. DNV GL Edition October 2015.

Dong Y.-L., Gu Y.J., Yang K., Zhang W.K. (2004). *A combining condition prediction model and its application in power plant*. Proceedings of the 2004 International Conference on Machine Learning and Cybernetics, vol. 6, Shanghai, China, 2004, pp. 3474–3478.

Doumpos M. & Zopounidis C. & Golfinopoulou V. (2007). *Additive Support Vector Machines for Pattern Classification*. IEEE transactions on systems, man, and cybernetics. Part B, Cybernetics : a publication of the IEEE Systems, Man, and Cybernetics Society. 37. 540-50. 10.1109/TSMCB.2006.887427.

Dragomir, O. E., Gouriveau, R., Dragomir, F., Minca, E., Zerhouni, N., (2009). *Review of prognostic problem in condition-based maintenance*. IFAC and in collaboration with the IEEE Control Systems Society. European Control Conference, ECC'09., Aug 2009, Budapest, Hun- gary. sur CD ROM, pp.1585-1592, 2009.

Edelman S. and Intrator N. (1997). *Learning as Extraction of Low-Dimensional Representations*. Psychology of Learning and Motivation Volume 36, 1997, Pages 353-380.

El-Reedy M. (2012). *Risk-based inspection technique*. Gulf Professional Publishing [Chapter 8] ASIN: B00DGSWO4S

Ellis B. (2009). The Challenges of Condition Based Maintenance. TJP, June 19, 2009, pp. 1-4.

Elmasri R. and S. Navathe, B. (2010). *Fundamentals of Database Systems*. Addison Wesley, 6th edition.

Embrechts M. & Gatti C. & Linton J. & Roysam B. (2013). *Hierarchical Clustering for Large Data Sets*. 10.1007/978-3-642-28696-4_8.

Enami, D. & Zhu, Faqiang & Yamamoto, K. & Nakagawa, S. (2012). *Soft-clustering technique for training data in Age-and gender-independent speech recognition*. 1-4.

Erkoyuncu J.A., Khan S., Eiroa A.L., Butler N., Rushton K., Brocklebank S. (2017). *Perspectives on trading cost and availability for corrective maintenance at the equipment type level.* Reliability Engineering & System Safety Volume 168, December 2017, Pages 53-69.

Faber M. (2002). *Risk-based inspection: the framework*. Structural engineering international, 3, 186-194.

Fahim A. & Salem A.-B.M. & Torkey, F. & Ramadan M. (2006). *Efficient enhanced k-means clustering algorithm*. Journal of Zhejiang University SCIENCE A. 7. 1626-1633. 10.1631/jzus.2006.A1626.

Fedele, L. (2011). *Methodologies and Techniques for Advanced Maintenance*. Springer.

Fei D.-Y., Almasiri O., & Rafig A. (2020). *Skin Cancer Detection Using Support Vector Machine Learning Classification based on Particle Swarm Optimization Capabilities*. Transactions on Machine Learning and Artificial Intelligence, 8(4), 01-13. https://doi.org/10.14738/tmlai.84.8415.

Fejri F. (2017). Matrix approximation methods and compressed sensing via random operators. PIR ENAC.

Fern X. & Brodley C. (2003). *Random Projection for High Dimensional Data Clustering: A Cluster Ensemble Approach*. Proc 20th Int'l Conf Machine Learning. 186-193.

Fernandez, O., Labib, A. W., Walmsley, R. & Petty, D. J. (2003). *A decision support maintenance management system: Development and implementation*. International Journal of Quality & Reliability Management, 20, 965- 979.

Fernandez, O., Walmsley, R., Petty, D.J. and Labib, A.W. (2003). *A decision support maintenance management system development and implementation*. International Journal of Quality and Reliability Management, Vol.20, No.8, pp. 965-979.

Fonseca, D.J. and Knapp, G.M (2000). *An expert system for reliability centered maintenance in the chemical industry*. Expert Systems with Applications, 2000, 19, 45–57.

Gabbar, H.A, Yamashita H., Suzuki K. and Shimada Y. (2003). *Computer-aided RCM-based plant maintenance management system*. Robotics and Computer Integrated Manufacturing, 2003, 19, 449-458.

Gao J. & Hu W. & Zhang Z. & Zhang X. & Wu O. (2011). *RKOF: Robust Kernel-Based Local Outlier Detection*. 6635. 270-283. 10.1007/978-3-642-20847-8_23.

Garvey M. & Jackson M. (1989). *Introduction to object-oriented databases*. Information and Software Technology. 31. 521-528. 10.1016/0950-5849(89)90173-0.

Gautam R., Vanga S., Ariese F. Umapathy S. (2015). *Review of multidimensional data processing approaches for Raman and infrared spectroscopy*. EPJ Techn Instrum 2, 8 (2015). https://doi.org/10.1140/epjti/s40485-015-0018-6.

Ge Z. and Song Z. (2007). *Process Monitoring Based on Independent Component Analysis−Principal Component Analysis (ICA−PCA) and Similarity Factors*. Industrial & Engineering Chemistry Research 46(7).

Gerardo, P. (2012). *On maritime transport costs, evolution, and forecast*. Ship Science & Technology - Vol. 5 - n.° 10 - (19-31) January 2012 - Cartagena (Colombia).

Geron, A. (2017). *Hands-On Machine Learning with Scikit-learn & TensorFlow*. O'Reily Media.

Ghaseminezhad M. & Karami, A. (2011). *A novel self-organizing map (SOM) neural network for discrete groups of data clustering*. Appl. Soft Comput.. 11. 3771-3778. 10.1016/j.asoc.2011.02.009.

Gheisari, S. & Meybodi, M.R., (2016). *Bayesian Network structure Training based on a Game of Learning Automata.* International Journal of Machine Learning and Cybernetics.

Gheisari, S., Meybodi, M.R., Dehghan, M. et al. (2016). *BNC-VLA: bayesian network structure learning using a team of variable-action set learning automata*. Appl Intell 45, 135–151 (2016). https://doi.org/10.1007/s10489-015-0743-1.

Gkerekos, C., Lazakis, I., & Theotokatos, G. (2017). *Ship machinery condition monitoring using performance data through supervised learning*. In Proceedings of the 2017 Smart Ship Technology Conference (pp. 105-111). London: Royal Institution of Naval Architects.

Goodman G.V.R. (1988). *An assessment of coal mine escapeway reliability using fault tree analysis*. Mining Science and Technology. 1988;7(2):205–215.

Goyet J. (2001). *Integrated Approach for RBI of Offshore Installations*. Proceedings of the International Workshop Risk Based Inspection and Maintenance Planning, Zürich, Swizerland, December 14-15, 2000, Vol. 1, pp.117-127, (2001).

Greve B. & Pigeot I. & Huybrechts I. & Pala V. & Börnhorst C. (2015). *A comparison of heuristic and model-based clustering methods for dietary pattern analysis*. Public health nutrition. 19. 1-10. 10.1017/S1368980014003243.

Grossi E. & Buscema M. (2008). *Introduction to artificial neural networks*. European journal of gastroenterology & hepatology. 19. 1046-54. 10.1097/MEG.0b013e3282f198a0.
Gruber A., Yanovski S., Ben-gal I. (2013). *Condition-based maintenance via simulation and a targeted baysian network meta model*. Quality Engineering, 25 (4) (2013), pp. 370-384.

Guerbai Y.e & Chibani Y. & Hadjadji B. (2014). *Writer-independent Handwritten Signature Verification based on One-Class SVM classifier*. Proceedings of the International Joint Conference on Neural Networks. 327-331. 10.1109/IJCNN.2014.6889416.

Hachicha W, Masmoudi F, & Haddar M. (2006). *A correlation analysis approach of cell formation in cellular manufacturing system with incorporated production data*. International Journal of Manufacturing Research, 1(3), 332 – 353.

Hamel L. (2009). *Knowledge Discovery with Support Vector Machines*. 231-235. 10.1002/9780470503065.refs.

Han Y., and Song Y.H. (2003). *Condition monitoring techniques for electrical equipment: A literature survey*. IEEE Transactions on Power Delivery, 18(1), 4-13.

Hashemian H.M., Bean W.C. (2011). *State-of-the-art predictive maintenance techniques IEEE Transactions on Instrumentation and Measurement*. 60 (10) (2011), pp. 3480-349

Heckerman D. & Breese J. & Rommelse, K. (1970). *Troubleshooting under Uncertainty*.

Heckerman, D., Geiger, D. & Chickering, D.M. (1995). *Learning Bayesian Networks: The Combination of Knowledge and Statistical Data*. Machine Learning 20, 197–243 (1995). https://doi.org/10.1023/A:1022623210503.

Heras J.M., Donati A. (2014). *Enhanced Telemetry Monitoring with Novelty Detection.* Association for the Advancement of Artificial Intelligence. ISSN 0738-4602.

Hernández-Cumplido J., Giusti M. & Zhou Y., Kyryczenko-Roth V., Chen Y., Rodriguez-Saona C. (2018). *Testing the 'plant domestication-reduced defense' hypothesis in blueberries: the role of herbivore identity*. Arthropod-Plant Interactions. 12. 10.1007/s11829-018-9605-1.

Hinton G., Sejnowski T. (1999). *Unsupervised Learning: Foundations of Neural Computation*. MIT Press. ISBN 978-0262581684.

Hiruta T., Uchida T., Yuda S.,Umeda Y. (2019). *A design method of data analytics process for condition based maintenance*. CIRP Annals, Volume 68/1, 2019, 145-148.

Hoa N.D. (2017). A Data-Driven Framework for Remaining Useful Life Estimation. Vietnam Journal of Science and Technology 55 (5) (2017) 557-571.

Horenko I., Schmidt-Ehrenberg J., Schütte C. (2006). *Set-Oriented Dimension Reduction: Localizing Principal Component Analysis Via Hidden Markov Models*. 4216. 74-85. 10.1007/11875741_8.

Horner, R. M. W., El-Haram, M. A., & Munns, A. K. (1997). *Building maintenance strategy: a new management approach*. Journal of quality in maintenance engineering, 3(4), 273-280.

Hossain, Md & Akhtar, Md.N. & Ahmad, R.B. & Rahman, M. (2019). *A dynamic K-means clustering for data mining*. Indonesian Journal of Electrical Engineering and Computer Science. 13. 521. 10.11591/ijeecs.v13.i2.pp521-526.

Hsu C-S., Jiang J-R., (2018). *Remaining Useful Life Estimation Using Long Short-Term Memory Deep Learning*. Proceedings of IEEE International Conference on Applied System Innovation 2018 IEEE ICASI 2018- Meen, Prior & Lam (Eds).

Hsu C. and Jiang J. (2018). *Remaining useful life estimation using long short-term memory deep learning*. 2018 IEEE International Conference on Applied System Invention (ICASI), Chiba, Japan, 2018, pp. 58-61, doi: 10.1109/ICASI.2018.8394326.

Huang, Y. S., Gau, W. Y., & Ho, J. W. (2015). *Cost analysis of two-dimensional warranty for products with periodic preventive maintenance*. Reliability Engineering & System.

Hughes G. F. (1968). *On the Mean Accuracy of Statistical Pattern Recognizers*. IEEE Transactions on Information Theory, IT-14:55-63.

Hyvärinen A. and Oja E. (2000). *Independent Component Analysis: Algorithms and Applications*. Neural Networks, 13(4-5):411-430, 2000.

IACS (2018). *A Guide to Managing Maintenance in Accordance with the Requirments of the ISM code*. Rec. 2001/Rev.2 2018. No. 74. (cont).

Idri A. & Khoshgoftaar T.M. & Abran, A. (2002). *Can neural networks be easily interpreted in software cost estimation*? 1162 - 1167. 10.1109/FUZZ.2002.1006668.

IEC (2010). *Geneva: International Electrotechnical Commision*. IEC60300-3-11.

IEEE (2012). *Transactions on Reliability*. pp. 314-322 61 (2) (2012).

Indyk P. and Motwani R. (1998). *Approximate nearest neighbors: towards removing the curse of dimensionality*. STOC '98: Proceedings of the thirtieth annual ACM symposium on Theory of computing.

Ireland C. and Bowers D. (2015). *Exposing the myth: object-relational impedance mismatch is a wicked problem*. In: DBKDA 2015, The Seventh International Conference on Advances in Databases, Knowledge, and Data Applications, IARIA XPS Press pp. 21–26.

Ireland C. and Bowers D. (2015). *Exposing the Myth: Object-Relational Impedance Mismatch is a Wicked Problem*.

Jardine A.K.S, Lin D, Banjevic D. (2006). *A review on machinery diagnostics and prognostics implementing condition-based maintenance*. Mechanical Systems and Signal Processing 2006;20:1483–510.

Jazzini A., Ayache M., Elkhansa L., Makki B. (2013). *Effects of predictive maintenance(PdM), Proactive maintenace(PoM) & Preventive maintenance(PM) on minimizing the faults in medical instruments.* Conference: 2013 2nd International Conference on Advances in Biomedical Engineering (ICABME).

Jensen F.V. (2001). *Bayesian Networks and Decision Graphs*. Statistics for Engineering and Information Science.

Jimenez V.J., Bouhmala N., Gausdal A.H. (2020). *Developing a predictive maintenance model for vessel machinery.* Journal of Ocean Engineering and Science Volume 5, Issue 4, December 2020, Pages 358-386.

Johnson W.B. and Lindenstranss J. (1984). *Extensions of Lipshitz mapping into Hilbert space*. In Conference in modern analysis and probability, volume 26 of Contemporary Mathematics, pages 189-206. Amer. Math. Soc., 1984.

Johnson W.B. and Lindenstrauss J. (1984). *Extensions of lipschitz mappings into a hilbert space*. Contemporary mathematics, 26(189):1, 1984.

Jolliffe I.T. and Cadima, J. (2016). *Principal component analysis: a review and recent developments*. https://doi.org/10.1098/rsta.2015.0202

Jolliffe T. and Cadima J. (2016). *Principal component analysis: a review and recent developments*. https://doi.org/10.1098/rsta.2015.0202.

Jung T-P, Humphries C, Lee T-W, Makeig S, McKeown M, Iragui V, Sejnowski TJ (1998). *Extended ICA removes artifacts from electroencephalographic recordings*. In Advances in neural information processing systems 10, eds Kearns M, Jordan M, Solla S (MIT, Cambridge, MA), pp 894–900.

Jung,Y. & Kang M. & Heo J. (2014). *Clustering performance comparison using K-means and expectation maximization algorithms*. Biotechnology, biotechnological equipment. 28. S44-S48. 10.1080/13102818.2014.949045.

Jutten C., Herault J. (1991). *Blind separation of sources, part I: An adaptive algorithm based on neuromimetic architecture*. Signal Processing Volume 24, Issue 1, July 1991, Pages 1-10.

Kadak A. C. & Matsuo T. (2007). *The nuclear industry's transition to risk-informed regulation and operation in the United States*. Reliability engineering and system safety, 92, 609-618.

Karamizadeh S., Abdullah S., Azizah M., Mazdak Z., Alireza H. (2013). *An Overview of Principal Component Analysis. Journal of Signal and Information Processing*. 10.4236/jsip.2013.43B031.

Kepner J. & Gadepally V. & Hutchison D. & Jananthan H. & Mattson T. & Samsi S.& Reuther A. (2016). *Associative Array Model of SQL, NoSQL, and NewSQL Databases*.

Kim M., & Candan K. (2014). *TensorDB: In-database tensor manipulation with tensor-relational query plans*. In CIKM 2014 - Proceedings of the 2014 ACM International Conference on Information and Knowledge Management (pp. 2039-2041). Association for Computing Machinery, Inc. https://doi.org/10.1145/2661829.2661842.

Kind M.C., Brunner R.J. (2014). *SOMz: photometric redshift PDFs with self-organizing maps and random atlas, Monthly Notices of the Royal Astronomical Society*. Volume 438, Issue 4, 11 March 2014, Pages 3409–3421, https://doi.org/10.1093/mnras/stt2456.

Kleinberg J.M. (1997). Two Algorithms for Nearest-Neighbor Search in High Dimensions.

Knowles et al. (2010). *Reinforcement Learning for Scheduling of Maintenance*. Programme chairs' introduction (pp.409-422).

Knowles, M., Baglee, D., Wermter, S. (2010). *Reinforcement Learning for Scheduling of Maintenance*. International Conference on Innovative Techniques and Applications of Artificial Intelligence SGAI 2010: Research and Development in Intelligent Systems XXVII pp 409-422

Kohonen T. (1998). *The self-organizing map. Neurocomputing*. 21:1–6.

Kohonen T., Mäkisara K.,Saramäki T. (1984). *Phonotopic maps—Insightful representation of phonological features for speech recognition.* Proceedings of the Seventh International Conference on Pattern Recognition, IEEE. Computer Society, Silver Spring, MD (1984), pp. 182-185.

Kohonen, T. (2013). *Essentials of the self-organizing map*. Neural Netw. 37:52–65.

Komarasamy, Dr & Wahi, A. (2012). *An optimized k-means clustering technique using bat algorithm*. European Journal of Scientific Research. 84. 263-273.

Kravela R.S., Kralev V.S., Sinyagina N., Koprinkova-Hristova P., Bocheva N. (2018). *Design and Analysis of a Relational Database for Behavioral Experiments Data Processing*. iJOE – Vol. 14, No. 2, 2018.

Křenek J. & Kuca K. & Blazek P. & Krejcar O. & Jun D. (2016). *Application of Artificial Neural Networks in Condition Based Predictive Maintenance*. 10.1007/978-3-319-31277-4_7.

Křenek J. & Kuca K. & Krejcar O. & Maresova P. & Sobeslav V. & Blazek P. (2014). *Artificial Neural Network Tools for Computerised Data Modeling and Processing*.

Krenek, J., Kuca, K., Krejcar, O., Maresova, P., Sobeslav, V., Blazek, P., (2014). *Artificial neural network tools for computerised data modelling and processing.* In 15th IEEE International symposium on computational intelligence and informatics. IEEE (2014).

Ku A. et al. (2004). *Structural reliability applications in developing risk-based inspection plans for a floating production installation*. 23rd International Conference on Offshore Mechanics and Arctic Engineering, 177-191, 20- 25 June, Vancouver, Canada.

Kuikka, S., Gislason, H., Hansson, S., Hilden, M., Sparholt, H. and Varis, O. (1999). *Enviromnentally Driven Uncertainties in Baltic Cod Management - Modelling by Bayesian Influence Diagrams*. Can. J Fish. Aqua. Sci., 56: 629-641.

Kumar N. (2018). *Enhancing intraoperative neuroimaging by incorporating spatial regularization into semiparametric regression models*.

Kurien J. and Nayak P.P. (2000). *Back to the Future for Consistency-based Trajectory Tracking*. Proceedings of the National Conference on Artificial Intelligence.

Kurimo, M. (1999). Indexing Audio Documents by using Latent Semantic Analysis and SOM.

Kwak N. (2008). *Principal Component Analysis Based on L1-Norm Maximization. IEEE transactions on pattern analysis and machine intelligence*. 30. 1672-80. 10.1109/TPAMI.2008.114.

Kwon D. (2018). *Multi-criteria decision support system for the best ship maintenance strategy*. University of Strathclyde MPhil Thesis.

Labib A.W. (2004). *A decision analysis model for maintenance policy selection using a CMMS*. Journal of Quality in Maintenance Engineering, Vol. 10 No. 3, pp. 191-202. https://doi.org/10.1108/13552510410553244.

Labib A.W. (2008). *Computerised Maintenance Management Systems*. In "Complex Systems Maintenance Handbook", Edited by: K.A.H. Kobbacy and D.N.P. Murthy, Springer, ISBN 978-1-84800-010-0.

Lam H., Klemes J., Friedler F., Kravanja Z., Varbanov P. (2010). *Software tools overview: process integration, modelling and optimisation for energy saving and pollution reduction*. Chemical Engineering Transactions, Vol. 21, 487-491.

Lawrence R. & Almasi G.S. & Rushmeier H. (1999). A Scalable Parallel Algorithm for Self-Organizing Maps with Applications to Sparse Data Mining Problems. Data Mining and Knowledge Discovery. 3. 171-195. 10.1023/A:1009817804059.

Lazakis I., Gkerekos C., Theotokatos G. (2019). *Investigating an SVM-driven, one-class approach to estimating ship systems condition*. Ships and Offshore Structures, 14:5, 432-441, DOI: 10.1080/17445302.2018.1500189

Lazakis I., Gkerekos C., & Theotokatos G. (2018). *Investigating an SVM-driven, one-class approach to estimating ship systems condition*. Ships and Offshore Structures.

Lazakis I., Dikis K., Michala A.L., Theotokatos G. (2016). Advanced ship systems condition monitoring for enhanced inspection, maintenance and decision making in ship operations. Transportation Research Procedia 14 (2016) 1679 – 1688.

Lazakis I., Turan O. , Alkaner S., & Olcer A. (2009). *Effective ship maintenance strategy using a risk and criticality based approach*. Paper presented at 13th International Congress of the International Maritime Association of the Mediterranean (IMAM 2009), Istanbul, Turkey.

Lazarevic A., Kumar V. and Srivastava J. (2005). *Intrusion Detection: A Survey.* Chapter 2.

Lee A. & Willcox B. (2014). *Minkowski Generalizations of Ward's Method in Hierarchical Clustering.* Journal of Classification. 31. 194-218. 10.1007/s00357-014-9157-8.

Lee J., Wang H. (2008). *New Technologies for Maintenance.* In: Complex System Maintenance Handbook. Springer Series in Reliability Engineering. Springer, London. https://doi.org/10.1007/978-1-84800-011-7_3.

Lee, J., R. Abujamra, A.K.S. Jardine, D. Lin, D. Banjevic (2004). *An integrated platform for diagnostics, prognostics and maintenance optimization*. The IMS '2004 International Conference on Advances in Maintenance and in Modeling. Simulation and Intelligent Monitoring of Degradations, Arles, France, 2004.

Li J. and Li J. (2018). *Research on NoSQL Database Technology*. Proceedings of the 2018 2nd International Conference on Management, Education and Social Science (ICMESS 2018).

Li Y. & Chen Y. (2018). *Research on Initialization on EM Algorithm Based on Gaussian Mixture Model. Journal of Applied Mathematics and Physics*. 06. 11-17. 10.4236/jamp.2018.61002.

Li, X., Yu L., Hang L., Tang X (2017). *The parallel implementation and application of an improved k-means algorithm*. J. Univ. Electron. Sci. Technol. China 2017, 46, 61–68.

Lia G., Hu, Y. (2018). *Improved sensor fault detection, diagnosis and estimation for screw chillers using density-based clustering and principal component analysis.* Energy and Buildings, Volume 173, 15 August 2018, Pages 502-515

Liao W., Pan E., Xi L. (2009). *Preventive maintenance scheduling for repairable system with deterioration*. Journal of Intelligent Manufacturing volume 21, pages875–884(2010).
Lin J. & Gunopulos D. (2003). *Dimensionality reduction by random projection and latent semantic indexing*.

Linial N., London E., Rabinovich Y. (1995). *The geometry of graphs and some of its algorithmic applications*. Combinatorica 15 (2) (1995) 215–245.

Liu H.-C. & Liu L. & Liu N. (2013). *Risk Evaluation Approaches in Failure Mode and Effects Analysis: A Literature Review*. Expert Systems with Applications. 40. 828–838. 10.1016/j.eswa.2012.08.010.

Liu K., Kargupta H., Ryan J. (2006). *Random Projection - Based Multiplicative Data Perturbation for Privacy Preserving Distributed Data Mining*. IEEE Transactions on Knowledge and Data Engineering 18(1):92- 106.

Liu S., Wang S. (2006). *Machine Health Monitoring and Prognostication Via Vibration Information.* Conference: Proceedings of the Sixth International Conference on Intelligent Systems Design and Applications (ISDA 2006), October 16-18, 2006, Jinan, China.

Liu Y., Li Z., Zhou C., Jiang Y., Sun J., Wang M. and He X. (2019). *Generative Adversarial Active Learning for Unsupervised Outlier Detection*. IEEE Transactions On Knowledge and Data Engineering, 2019.

Liu, B., Xu, Z., Xie, M., & Kuo, W. (2014). A value-based preventive maintenance policy for multi-component system with continuously degrading components. Reliability Engineering & System Safety, 132, 83-89.

Liu, J., Wang, Golnaraghi. (2010). *An enhanced diagnostic scheme for bearing condition monitoring*. IEEE Transactions on Instrumentation and Measurement. Volume 59, Issue 2, Pages 309–321.

Liu, Y., Jin, S. (2013). *Application of Bayesian networks for diagnostics in the assembly process by considering small measurement data sets*. The International Journal of Advanced Manufacturing Technology 65(9-12), 1229-1237.

Lo N., Flaus J.M., Adrot O. (2019). *Review of Machine Learning Approaches In Fault Diagnosis applied to IoT System*. International Conference on Control, Automation and Diagnosis ICCAD'19, Jul 2019, Grenoble, France. hal-02344344.

Lubo-Rubles D. (2018). *Development of Independent Component Analysis for Reservoir Geomorphology and Unsupervised Seismic Facies Classification in the Taranaki Basin, New Zeeland*. University of Oklahoma Master of Science Thesis.

Luo J., Namburu M., Pattipati K., Qiao L. (2003). *Model-based prognostic techniques [maintenance applications]*. Conference: AUTOTESTCON 2003. IEEE Systems Readiness Technology.

Majid N. A. A., Young B. R., Taylor M. P. and Chen J. J. J. (2012). *K-means clustering pre-analysis for fault diagnosis in an aluminium smelting process*. 2012 4th Conference on Data Mining and Optimization (DMO), Langkawi, Malaysia, 2012, pp. 43-46, doi: 10.1109/DMO.2012.6329796.

MAN B&W (2004). DMS2100i Bridge Manoeuvring System ME/ME-C Engines User Manual.

Manjrekar O. & Duduković M. (2019). *Identification of flow regime in a bubble column reactor with a combination of optical probe data and machine learning technique*. Chemical Engineering Science: X. 2. 100023. 10.1016/j.cesx.2019.100023.

Mann, L., Saxena, A., Knapp, G. (1995). *Statistical-based or condition-based preventive maintenance?* Journal of Quality in Maintenance Engineering 1 (1), 46–59.

Manno, G. and Knutsen, K.E. (2014). *An Importance Measure approach to System level Condition Monitoring of Ship Machinery Systems*. Conference paper-June 2014.

Maretić I.S., Lacković I. (2014). *Application of Gaussian Mixture Models with Expectation Maximization in Bacterial Colonies Image Segmentation for Automated Counting and Identification*. In: Roa Romero L. (eds) XIII Mediterranean Conference on Medical and Biological Engineering and Computing 2013. IFMBE Proceedings, vol 41. Springer, Cham. https://doi.org/10.1007/978-3-319-00846-2_96.

Marhavilas P.K., Filippidis M., Koulinas G.K., Koulouriotis D.E. (2020). *A HAZOP with MCDM Based Risk-Assessment Approach: Focusing on the Deviations with Economic/Health/Environmental Impacts in a Process Industry*. Sustainability. 2020; 12(3):993. https://doi.org/10.3390/su12030993.

Márquez C.A. and Herguedas S.A. (2004). *Learning about failure root causes through maintenance records analysis.* Journal of Quality in Maintenance Engineering, Vol. 10 No. 4, pp. 254-262. https://doi.org/10.1108/13552510410564873.

Marquez, A. C. (2007). *The Maintenance Management Framework: Models and Methods for Complex Systems Maintenance. Springer-Verlag London Limited.*

Marquez, A.C. and Herguedas, A.S. (2004). *Learning about failure root causes through maintenance records analysis*. Journal of Quality in Maintenance Engineering, Vol. 10 No. 4, pp. 254-62.

Marton I., Sanchez A., Carlos S., Martorell S. (2013). *Application of Data Driven Methods for Condition Monitoring Maintenance*. Conference: 4th IEEE Conference on Prognostics and System Health Management (PHM)Volume: 33.

McClelland J.L. and Jeffrey L.E. (1986). *The TRACE Model of Speech Perception*. Cognitive Psychology, 18(1): 1–86. doi:10.1016/0010-0285(86)90015-0.

McCulloch W. S. and Pitts W. (1943). *A logical calculus of the ideas immanent in nervous activity*. The Bulletin of Mathematical Biophysics, 5(4):115–133, 1943.

McLachlan G.J. and Krishnan, T. (2008). *The EM Algorithm and Extensions*. Second Edition. Hoboken, New Jersey: Wiley.

Mechefske, C. K. (2005). *Machine Condition Monitoring and Fault Diagnosis*. Boca Raton, Florida, USA, CRC Press, Taylor & Francis Group.

Mechefske, C. K. (2005). *Machine Condition Monitoring and Fault Diagnosis*. Boca Raton, Florida, USA, CRC Press,Taylor & Francis Group.

Medjaher K., Zerhouni N., Baklouti J. (2013). *Data-Driven prognostics based on health indicator construction: Application to PRONOSTIA's Data*. Conference: Control Conference (ECC), 2013 European.

Medjaher, Kamal & Zerhouni, Noureddine (2013). *Hybrid prognostic method applied to mechatronic systems*. The International Journal of Advanced Manufacturing Technology. 69. 10.1007/s00170-013-5064-0.

Mehta P., Werner A., Mears L. (2013). *Condition based maintenance-systems integration and intelligence using Bayesian classification and sensor fusion*. Journal of Intelligent Manufacturing 26(2).

Meireles M. & Almeida P & Simoes M. (2003). *A comprehensive review for industrial applicability of artificial neural networks*. Industrial Electronics, IEEE Transactions on. 50. 585 - 601. 10.1109/TIE.2003.812470.

Mejia D.A.T., Medjaher K., Zerhouni N. (2012). *A Data-Driven Failure Prognostics Method Based on Mixture of Gaussians Hidden Markov Models.* IEEE Transactions on Reliability 61(2):491-503.

Meng, Zhenzhu & Hu, Yating & Ancey, Christophe. (2020). Using a Data Driven Approach to Predict Waves Generated by Gravity Driven Mass Flows. Water. 12. 10.3390/w12020600.

Mobley R.K. (2001). *Predictive Maintenance.* Plant Engineer's Handbook

Mobley R.K. (2002). *An Introduction to Predictive Maintenance.* 2nd ed. Butterworth-Heinemann.

Mobley, R. K. 2002. *An introduction to predictive maintenance*. 2nd edition, USA: Butterworth-Heinemann.

Mohamed M.A., Altrafi O.G., Ismail M.O. (2014). *Relational vs NoSQL databases: a survey*. Int. J. Comput. Inf. Technol. 03(03), 598–601.

Moniruzzaman, A B M & Hossain, Syed. (2013). *NoSQL Database: New Era of Databases for Big data Analytics - Classification, Characteristics and Comparison*. Int J Database Theor Appl. 6.

Moore S., (2018*). OpCost 2018*. Shipping Industry Group.

Morant A., Larsson-Kraik P.O., Kumar O. (2014). *Data-driven model for maintenance decision support: A case study of railway signalling systems*. Proceedings of the Institution of Mechanical Engineers Part F Journal of Rail and Rapid Transit.

Moreira, C. (2011). *Learning To Rank Academic Experts*.

Moubray J (2001). *Reliability-centered maintenance*. Industrial Press, New York, ISBN0831131462.

Moya C. C. (2004). *The control of the setting up of a predictive maintenance program using a system of indicators*. International Journal of Management Sciences, 32, 57-75.

MSDP Studies (1993). *Age Reliability Analysis Prototype Study*. American Management Systems, U.S. Naval Sea Systems Command Surface Warship Directorate, USA.

MSDP Studies (1993). *Age Reliability Analysis Prototype Study*. American Management Systems, U.S. Naval Sea Systems Command Surface Warship Directorate, USA.

Muñoz-Marí, J., Bovolo, F., Gómez-Chova, L., Bruzzone, L., Camp-Valls, G. (2010). *Semisupervised one-class support vector machines for classification of remote sensing data*. IEEE Trans. Geosci. Remote Sens. 48, 3188–3197.

Murtagh F. and Contreras P. (2011). *Methods of Hierarchical Clustering.* arXiv:1105.0121.

Nachimuthu S., Zuo M.J., Ding Y. (2019). *A Decision-making Model for Corrective Maintenance of Offshore Wind Turbines Considering Uncertainties*. Energies 2019,12, 1408; doi:10.3390/en12081408.

Nayak A. & Poriya A. & Poojary D. (2013). *Type of nosql databases and its comparison with relational databases*. International Journal of Applied Information Systems. 5. 16-19.

Ndeye Lo, Jean-Marie Flaus, Olivier Adrot (2019). *Review of Machine Learning Approaches In Fault Diagnosis applied to IoT System.* International Conference on Control, Automation and Diagnosis ICCAD'19, Jul 2019, Grenoble, France. hal-02344344

Nguyen T.-L. & Shu M.-H. & Hsu B.-M. (2016). *Extended FMEA for Sustainable Manufacturing: An Empirical Study in the Non-Woven Fabrics Industry*. Sustainability. 8. 939. 10.3390/su8090939.

Niculescu B.M., Andrei G. (2016). *Principal Component Analysis as a Tool for Enhanced Well Log Interpretation*. Rev. Roum. GÉOPHYSIQUE, 60, p. 49–61, 2016, Bucureşti.

Niu G. (2017). *Data-Driven Technology for Engineering Systems Health Management*. DOI: 10.1007/978-981-10-2032-2, ISBN: 978-981-10-2031-5.

Nowlan, F.S. and Heap, H.F. (1978). *Reliability-centred maintenance*. United Airlines, San Francisco, 1978.

O'Hanlon T. (2004). *CMMS best practices*. Mainten. J., 17(3): 19–22.

Ogunlere S. & Idowu S. (2015). *Comparison Analysis of Object-Based Databases, Object-Oriented Databases, and Object Relational Databases.*

Oja E. (1992). Principal Components, Minor Components, and Linear Neural Networks. Neural Networks, Vol 5, pp 927-935.

Okman L.& Gal-Oz N. & Gonen Y. & Abramov J. (2011). *Security Issues in NoSQL Databases*. 10.1109/TrustCom.2011.70.

Okoh C., Roy R., J. Redding M.L. (2104). Overview of Remaining Useful Life Prediction Techniques in Through-Life Engineering Services. C. Okoh et al. / Procedia CIRP 16 (2014) 158 – 163.

Okoh, C., Roy, R., Mehnen, J., Redding, L. (2014). *Overview of Remaining Useful Life Prediction Techniques in Through-Life Engineering Services*. Product Services Systems and Value Creation. Proceedings of the 6th CIRP Conference on Industrial Product-Service Systems. Procedia CIRP 16 (2014) 158 – 163

Omdahl TP. (1988). *Reliability, availability, and maintainability dictionary*. ASQC Quality Press, Milwaukee.

Organization for Economic Co-operation and development (OECS) (2016). *The Economic Value of Shipping and Maritime Activity in Europe*. Oxford Economics.

Ozguc, O. (2020). *A new risk-based inspection methodology for offshore floating structures*. Journal of Marine Engineering & Technology, 19:1, 40-55, DOI: 10.1080/20464177.2018.1508804.

Pantazi, X.E. & Moshou, Dimitrios & Alexandridis, Thomas & Whetton, Rebecca & Mouazen, Abdul. (2016). *Wheat yield prediction using machine learning and advanced sensing techniques*. Computers and Electronics in Agriculture. 121. 57-65. 10.1016/j.compag.2015.11.018.

Papadimitriou S. & Kitagawa H. & Gibbons P. & Faloutsos, C. (2003). *LOCI: Fast Outlier Detection Using the Local Correlation Integral*. Proceedings - International Conference on Data Engineering. 315-326. 10.1109/ICDE.2003.1260802.

Park H. & Ozeki T. (2009). *Singularity and Slow Convergence of the EM algorithm for Gaussian Mixtures. Neural Processing Letters*. 29. 45-59. 10.1007/s11063-009-9094-4.

Peng Y., Dong M., Zuo M.J. (2010). *Current status of machine prognostics in condition-based maintenance: a review.* The International Journal of Advanced Manufacturing Technology volume 50, pages297–313(2010).

Peng Y., Dong M., Zuo M.J. (2010). *Current status of machine prognostics in condition-based maintenance: a review*. International Journal of Advanced Manufacturing Technology, 50 (2010), pp. 297-313

Phillips P. J. et al. (2005). *Overview of the face recognition grand challenge*. 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), San Diego, CA, USA, 2005, pp. 947-954 vol. 1, doi: 10.1109/CVPR.2005.268.

Phiri H. & Kunda D. (2017). *A Comparative Study of NoSQL and Relational Database*. Zambia ICT Journal. 1. 1-4. 10.33260/zictjournal.v1i1.8.

Pintelon L., Groenevelt H., Seidmann A. (1992). *Production Lot Sizing with Machine Breakdowns*. Management Science Volume 38, Issue 1 January 1992 Pages 1-156.

Pintelon, L., Gelders, L., (1992). *Maintenance management decision-making*. European Journal of Operational Research 58, 301–317.

Poongodai, A. & Bhuvaneswari S., R., (2013). *AI Technique in Diagnostics and Prognostics. International Journal of Computer Applications*. 2nd National Conference on Future Computing February 2013.

Poongodai, A. & Bhuvaneswari S., R., (2013). *AI Technique in Diagnostics and Prognostics. International Journal of Computer Applications*. 2nd National Conference on Future Computing February 2013.

Pradhan S., Singh R., Kachru K., Narasimhamurthy S. (2007). *A Bayesian network based ap-proach for root-cause-analysis in manufacturing process*. In: 2007 International Conference on Computational Intelligence and Security, pp. 10-14. IEEE, Harbin (2007).

Pradhan, S., Singh, R., Kachru, K., Narasimhamurthy, S. (2007). *A Bayesian network based approach for root-cause-analysis in manufacturing process*. 2007 International Conference on Computational Intelligence and Security, pp. 10-14. IEEE, Harbin.

Prajapati A., Bechtel J. and Ganesan S. (2012). *Condition based maintenance: a survey*. Journal of Quality in Maintenance Engineering, 18(4), 384-400.

Prajapati A., Bechtel J., Ganesan S. (2012). *Condition based maintenance: a survey.* Journal of Quality in Maintenance Engineering. 18 (4) pp. 384-400

Prakash G., Narasimhan S., and Pandey M. D. (2017). *Condition Based Maintenance of Low Speed Rolling Element Bearings using Hidden Markov Model*. International Journal of Prognostics and Health Management, ISSN2153-2648, 2017 006.

Raouf A., Ali Z.,Duffuaa S.O. (1993). *Evaluating a computerized maintenancemanagement system*. International Journal of Operations & Production Management,Vol. 13 No. 3, pp. 38-49.

Raptodimos Y. & Lazakis I. (2020). *Application of NARX neural network for predicting marine engine performance parameters*. Ships and Offshore Structures, 15:4, 443-452, DOI: 10.1080/17445302.2019.1661619.

Raptodimos, Y & Lazakis, I. (2018). *Using artificial neural network- self-organising map for data clustering of marine engine condition monitoring applications*. Ships and Offshore Structures, 13:6, 649-656, DOI: 10.1080/17445302.2018.1443694

Raptodimos, Y., & Lazakis, I. (2016). *An artificial neural network approach for predicting the performance of ship machinery equipment*. In Maritime Safety and Operations 2016 Conference Proceedings (pp. 95-101).

Rausand M. and Vatn J. (2008). *Centred Maintenance. Complex System Maintenance Handbook*. pp. 79-108, Kobbacy KAH and Murthy DNP (eds), London: Springer-Verlag.

Rausand M., Vatn J. (2008). *Reliability Centred Maintenance*. In: Complex System Maintenance Handbook. Springer Series in Reliability Engineering. Springer, London. https://doi.org/10.1007/978-1-84800-011-7_4.

Reason J. T. and Hobbs A. (2003). *Managing Maintenance Error: A Practical Guide.* Ashgate Publishing.

Reason, J., Hobbs, A. (2003). *Managing Maintenance Error: A Practical Guide.* Risk Management © 2004 Palgrave Macmillan Journals.

Recht J.L. (1965). *Systems safety analysis: an introduction. National Safety News.* December pp 37-38.

Reynolds D. (2009). *Gaussian Mixture Models*. In: Li S.Z., Jain A. (eds) Encyclopedia of Biometrics. Springer, Boston, MA. https://doi.org/10.1007/978-0-387-73003-5_196.

Rish I. (2001). *An Empirical Study of the Naïve Bayes Classifier*. IJCAI 2001 Work Empir Methods Artif Intell. 3.

Robles B., Avila M., Duculty F., Vrignat P., Begot S., Kratz F. (2012). *Methods to choose the best Hidden Markov Model topology for improving maintenance policy*. 9th International Conference of Modeling, Optimization and Simulation - MOSIM'12 June 6-8, 2012 - Bordeaux – France "Performance, interoperability and safety for sustainable development".

Ros F. & Guillaume S. (2019). *A hierarchical clustering algorithm and an improvement of the single linkage criterion to deal with noise*. Expert Systems with Applications. 128. 10.1016/j.eswa.2019.03.031.

Rothblum A.M. (2000). *Human Error and Marine Safety*. U.S. Coast Guard,Research & Development Center: 13-14.

Rothblum, A.M. (2000). *Human Error and Marine Safety.* U.S. Coast Guard Research & Development Center.

Rothenberger M.A., Dooley K. J., Kulkarni U.R., & Nada N. (2003). *Strategies for Software Reuse: A Principal Component Analysis of Reuse Practices*. IEEE Transactions on Software Engineering, 29(9), 825-837.

Rustenburg W.D., Van Houtum G. J., Zijm W. H. M. (2001). *Spare Parts Management At Complex Technology-Based Organizations: An agenda forResearch*. International Journal of Production Economics 71(1-3): 177-193.

Rustenburg, W.D., Zijm, W.H.M., Jan van Houtum, G. (2001). *Spare parts management at complex technology-based organizations: An agenda for research.* International Journal of Production Economics 71(1):177-193 DOI: 10.1016/S0925-5273(00)00117-1.

Saberi A, Salmasi F, Najafabadi T. *Sensor fault-tolerant control of wind turbine systems*. In: Proceedings of the 5th Conference on thermal power plants (CTPP); 2014. p. 40–45. http://dx.doi.org/10.1109/CTPP.2014.7040693.

Schelling, B., Plant, C. (2020). *Dataset-Transformation: improving clustering by enhancing the structure with DipScaling and DipTransformation*. Knowl Inf Syst 62, 457–484 (2020). https://doi.org/10.1007/s10115-019-01388-5.

Schölkopf B. & Williamson R. & Smola A. & Shawe-Taylor J. & Platt J. (1999). *Support Vector Method for Novelty Detection*. NIPS. 12. 582-588.

Schu G., Soldera J., Medeiros R., Scharcanski J.. (2016). *Random Projections and Their Applications in Computer Vision*. 10.13140/RG.2.2.15444.40329.

Schwabacher, M. & Goebel, k. (2005). *A Survey of Artificial Intelligence for Prognostics.* NASA Amess Research Center MS 269-3.

Schwartz, S., Montero Jimenez, J.J., Salaün, M. et al. (2020). *A fault mode identification methodology based on self-organizing map*. Neural Comput & Applic 32, 13405–13423 (2020). https://doi.org/10.1007/s00521-019-04692-x.

Scikit-Learn, (2019). *Anomaly detection with Local Outlier Factor (LOF*). https://scikit-learn.org/stable/modules/outlier_detection.html, last accessed: 2019/03/15).

Serratella, C. M. et al. (2007). *Risk-based strategies for the next generation of maintenance and inspection programs*. International symposium on maritime, safety, security and environmental protection (SSE), 20-21 September, Athens, Greece.

Shabrina A.P., Soesanto R.P., Kurniawati A., Kurniawan M.T., Andrawina L. (2018). e-Learning Content Design for Corrective Maintenance of Toshiba BMC 80.5 based on Knowledge Conversion using SECI Method: A Case Study in Aerospace Company. IOP

Conf. Series: Materials Science and Engineeting 012001 doi:10.1088/1757-899X/319/1/012001.

Shahid, N. & Naqvi I. & Qaisar S. (2013). *One-class support vector machines: Analysis of outlier detection for wireless sensor networks in harsh environments*. Artificial Intelligence Review. 43. 10.1007/s10462-013-9395-x.

Sharma A, Bruce KL, Chen B, Gyoneva S, Behrens SH, Bommarius AS, Chernoff YO (2016). *Contributions of the Prion Protein Sequence, Strain, and Environment to the Species Barrier*. J Biol Chem 291(3):1277-88.

Sharma S. & Shandilya R. & Patnaik S. & Mahapatra, A. (2015). *Leading NoSQL models for handling Big Data: A brief review*. International Journal of Business Information Systems. 10.1504/IJBIS.2016.075714.

Sheu, S. H., Chang, C. C., Chen, Y. L., & Zhang, Z. G. (2015). *Optimal preventive maintenance and repair policies for multi-state systems*. Reliability Engineering & System Safety, 140, 78- 87.

Shi L., Tu S., Xu L. (2011). *Learning Gaussian mixture with automatic model selection: a comparative study on three Bayesian related approaches*. Frontiers Electrical Electron Eng China 6(2):215–244.

Shifeng O., Xiaohui Z., Ying G. (2007). *Linear System Identification Employing Independent Component Analysis*. IEEE Int. Conf. on Automation and Logistics, 2007, pp.1336-1340, Aug.2007.

Shin, J-H. & Jun, H-B., (2015). *On condition based maintenance policy*. Journal of Computational Design and Engineering 2, 119-127.

Shina J.H. & Jun H.B. (2015). *On condition based maintenance policy*. Journal of Computational Design and Engineering. Volume 2, Issue 2, April 2015, Pages 119-127.

Shonet I.M. (2003*). Building evaluation methodology for setting maintenance priorities in hospital buildings*. Construction Management and Economics.

Shorten, D.C. (2012). *Marine Machinery Condition Monitoring: Why has the shipping industry been slow to adopt?*. Lloyds Register, London, 2012.

Si X., Wang W., Hua C., Zhou D. (2011). *Remaining useful life estimation: a review on the statistical data driven approaches*. European Journal of Operational Research, 213 , pp. 1-14

Sidda S. & Kiranmayi R. & Nagaraju P. (2017). *A Study on Fuzzy Controller and Neuro-Fuzzy Controller for Speed Control of PMSM Motor*. 10.1109/ICPCSI.2017.8391943.

Silberschatz, A., Korth, F. H. and Sudarshan, S. (2011). *Database System Concepts*. Mc Graw Hill, 6th edition.

Silvianita, Khamidi M.F., Rochani I., Chamelia D.M. (2014). *Hazard and Operability Analysis (HAZOP) of Mobile Mooring System*. Procedia Earth and Planetary Science Volume 14, 2015, Pages 208-212.

Singh G., & Solanki A. (2016). *An algorithm to transform natural language into SQL queries for relational databases*. Selforganizology, 3, 110-126.

Sint R. & Stroka S. & Schaffert S. & Ferstl R. (2009). *Combining Unstructured, Fully Structured and Semi-Structured Information in Semantic Wikis*.

Sirovich L. and Kirby M. (1987). *Low-Dimensional Procedure for Characterization of Human Faces*. Journal Optical Society of America, 4 (1987) 519-524.

SKF (2012a). *Condition-based maintenance must be set up correctly*. Marine Propulsion - Ship lifecyle management.

Smyth, P (1994). *Hidden Markov models for fault detection in dynamic systems*. Pattern Recognition 27(1): 149–164.

Song H., Jiang Z., Men A. and Yang B. (2017). *A Hybrid Semi-Supervised Anomaly Detection Model for High-Dimensional Data*. Comput Intell Neurosci. 2017; 2017: 8501683.

Sousa L. & Gama J. (2014). *The Application of Hierarchical Clustering Algorithms for Recognition Using Biometrics of the Hand*. International Journal of Advanced Engineering Research and Science (IJAERS). ISSN: 2349-6495.

Starr A., Al-Najjar B., Holmberg K., Jantunen E., Bellew J., Albarbar A. (2010). *Maintenance Today and Future Trends*. Holmberg K., Adgar A., Arnaiz A., Jantunen E., Mascolo J., Mekid S. (eds) E-maintenance. Springer, London.

Stefatos George & Hamza A. (2010). *Dynamic independent component analysis approach for fault detection and diagnosis*. Expert Syst. Appl. 37. 8606-8617. 10.1016/j.eswa.2010.06.101.

Stopford, M. (2008). *Maritime Economics*. Routledge.

Sugeno, M. and T. Yasukawa (1993). *A fuzzy-logic- based approach to qualitative modeling*. IEEE Transactions on Fuzzy Systems 1(1), 7–31.

Sun,F.,Gao,L.,Zou,J.,Wu,T.,Li,J., (2013). *Study on Multi-Equipment Failure Prediction Based on System Network. Sensors and Transducers*. 158 (11), 427-435(2013)

Sutton I. (2015). *Maintenance and inspection*. Plant design and operations (2015), pp. 24-45.

Sutton I. (2015). *Process Risk and Reliability Management.* 2nd ed. Elsevier.

Suzuki, K. (2013). *Artificial Neural Networks-Architectures and Applications*. IntechOpen Limited: London, UK, 2013.

Swann, C.D., and Preston, M.L. (1995). *Twenty-five Years of HAZOPs*. J.Loss Prev. Process Ind. 8(6), pp.349-353.

Swanson L. (2003). *An information-processing model of maintenance management*. International Journal of Production Economics, Vol. 83 No. 1, pp. 45-64.

Takagi H. & Pallez D. (2010). *Paired comparison-based Interactive Differential Evolution*. 2009 World Congress on Nature and Biologically Inspired Computing, NABIC 2009 - Proceedings. 475 - 480. 10.1109/NABIC.2009.5393359.

Tan Y. & Jung S.J. & Chung W-Y (2013*). Real time biomedical signal transmission of mixed ECG Signal and patient information using visible light communication*. Conference proceedings : ... Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Conference. 2013. 4791-4794. 10.1109/EMBC.2013.6610619.

Thamara Villegas, María Jesús Fuente and Miguel Rodríguez (2010). *Principal Component Analysis for Fault Detection and Diagnosis. Experience with a pilot plant*. Advances in Computational Intelligence Man-Machine Systems and Cybernetics, ISBN 978-960-474-257-8.

Tharwat A. (2019). *Parameter investigation of support vector machine classifier with kernel functions. Knowledge and Information Systems*. 61. 10.1007/s10115-019-01335-4.

Thrun M.C., Ultsch A. (2020). *Using Projection-Based Clustering to Find Distance- and Density-Based Clusters in High-Dimensional Data*. https://doi.org/10.1007/s00357-020-09373-2.

Tian Y., Ju X., Qi Z. et al. (2014). *Improved twin support vector machine*. Sci. China Math. 57, 417–432 (2014). https://doi.org/10.1007/s11425-013-4718-6.

Tian, Z. (2009). *An artificial neural network method for remaining useful life prediction of equipment subject to condition monitoring*. Conference Paper in Journal of Intelligent Manufacturing 23(2):143 – 148.

Tian, Z. (2012). *An artificial neural network method for remaining useful life prediction of equipment subject to condition monitoring*. J Intell Manuf 23, 227–237 (2012). https://doi.org/10.1007/s10845-009-0356-9.

Tibaduiza D.A., Mujica L.E., Rodellar J. (2012). *Damage classification in structural health monitoring using principal component analysis and self organizing maps*. Structural Control and Health Monitoring 20: 1303–1316.

Tinga, T., Tiddens, W. W., Amoiralis, F., & Politis, M. (2017). *Predictive maintenance of maritime systems: models and challenges*. In 27th European Safety and Reliability Conference (ESREL 2017). Taylor & Francis.

Tomlinson NA. (2016). *What is the ideal maintenance strategy? A look at both MoD and commercial shipping best practice. BMT Defence Services, UK.*

Tomlinson, M. (2016). *What is the ideal maintenance strategy? A look at both MoD and commercial shipping best practice*. bmtdsl.co.uk.

Torres E. & Villarreal-López E. (2014). *Gaussian Mixture Models Approach for Multiple Fault Detection - DAMADICS Benchmark.*

Trutt F. C, Sottile J. and Kohler J.L. (2002). *Online condition monitoring of induction motors*. In IEEE Transactions on Industry Applications, vol. 38, no. 6, pp. 1627-1632, Nov.-Dec. 2002, doi: 10.1109/TIA.2002.804758.

Tseng Y.-H. & Tsay M.Y. (2013). *Journal clustering of Library and Information Science for subfield delineation using the bibliometric analysis toolkit: CATAR*. Scientometrics. 95. 503-528. 10.1007/s11192-013-0964-1.

Tuncer, Y., Tanik, M. M., & Alison, D.B. (2008). *An overview of statistical decomposition techniques applied to complex systems*. Computational Statistics & Data Analysis, 52(5), 2292-2310.

Turk M.A., Pentland A.P. (1991). *Face Recognition using Eigenfaces*. Vision and Modelling Group, The Media Laboratory Massachusetts Institute of Technology.

UCL (2016). *Condition Based Maintenance of Naval Propulsion Systems*. https://sites.google.com/view/cbm/home, last accessed 2019/03/15.

United Nations Conference on Trade and Development (UNCTD) (2020). *Review of Maritime Transport.* Geneva: United Nations Publications.

Uusitalo, L. (2007). *Advantages and challenges of Bayesian networks in environmental modelling, Ecological Modelling*. 203, issue 3, p. 312-318, https://EconPapers.repec.org/RePEc:eee:ecomod:v:203:y:2007:i:3:p:312-318.

Van Dun B. & Wouters J. & Moonen M. (2007). *Improving Auditory Steady-State Response Detection Using Independent Component Analysis on Multichannel EEG Data*. IEEE transactions on bio-medical engineering. 54. 1220-30. 10.1109/TBME.2007.897327.

Vembala, S.S. (2005). *The Random Projection Method.* DIMACS: Series in Discrete Mathematics and Theoretical Computer Science.

Verma A., Narula A., Katyal A., Yadav S.K., Anand P., Jahan A., Pruthi S.K., Sarin N., Gupta R., Singh S. (2018). *Failure rate prediction of equipment: can Weibull distribution be applied to automated hematology analyzers?* De Gruyter, 2018, DOI: https://doi.org/10.1515/cclm-2018-0569.

Verron S. & Tiplica T. & Kobi A. (2008). *Fault Detection with Bayesian Network*. Frontiers in Robotics, Automation and Control. 10.5772/6318. Volume 84, Issue 1, 11 April 2003, Pages 85-100.

Von der Malsburg C. (1973). *Self-organization of orientation sensitive cells in striate cortex*. Biological Cybernetics. 14. 85-100. 10.1007/BF00288907.

Waldert S. & Bensch M.& Bogdan M.& Rosenstiel W. & Schölkopf B. & Lowery C. & Eswaran H. & Preissl H. (2007). *Real-Time Fetal Heart Monitoring in Biomagnetic Measurements Using Adaptive Real-Time ICA*. IEEE transactions on bio-medical engineering. 54. 1867-74. 10.1109/TBME.2007.895749.

Wall M.E., Rechtsteiner A., Rocha L.M. (2003). *A Practical Approach to Microarray Data Analysis*. Kluwer: Norwell, MA, pp. 91-109.

Wang F. & Tan S. & Yang Y. & Shi H. (2016). *Hidden Markov Model-Based Fault Detection Approach for a Multimode Process*. Industrial & Engineering Chemistry Research. 55. 10.1021/acs.iecr.5b04777.

Wang J. and Su X. (2011). *An improved K-Means clustering algorithm*. 2011 IEEE 3rd International Conference on Communication Software and Networks, Xi'an, China, 2011, pp. 44-46, doi: 10.1109/ICCSN.2011.6014384.

Wang P., Vachtsevanos G. (2001). *Fault prognostics using dynamic wavelet neural networks*. AI EDAM-Artificial Intelligence for Engineering Design Analysis and Manufacturing 15 (2001) 349–365.

Wang P.& Vachtsevanos G. (2001). *Fault prognosis using dynamic wavelet neural networks*. AI EDAM. 15. 349-365. 10.1109/AUTEST.2001.949467.

Wang W.Q., Golnaraghi M.F. and Ismail F., (2004). *Prognosis of machine health condition using neuro-fuzzy system*. Mechanical System and Signal Processing, Vol. 18, pp. 813-831.

Wang W.Q., Golnaraghi M.F., Ismail F. (2004). *Prognosis of machine health condition using neuro-fuzzy systems*. Mechanical Systems and Signal Processing 18 (2004) 813–831.

Wang Y., Deng C., Wu J., Wang Y., Xiong Y. (2014). *A corrective maintenance scheme for engineering equipment.* Engineering Failure Analysis Volume 36, January 2014, Pages 269-283.

Wang Y.C. (2018). *Prediction of engine failure time using principal component analysis, categorical regression tree, and back propagation network*. Journal of Ambient Intelligence and Humanized Computing.

Wehrens R. (2009). *Self-Organising Maps for Image Segmentation*. In: Fink A., Lausen B., Seidel W., Ultsch A. (eds) Advances in Data Analysis, Data Handling and Business Intelligence. Studies in Classification, Data Analysis, and Knowledge Organization. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-01044-6_34.

Weld and de Kleer (1990*). Readings in Qualitative Reasoning about Physical Systems.*

Widodo A. & Yang B-S. (2007). *Support vector machine in machine condition monitoring and fault diagnosis*. Mechanical Systems and Signal Processing. Volume 21, Issue 6, August 2007, Pages 2560-2574.

Widodo A. & Yang B.-S. (2007). *Support vector machine in machine condition monitoring and fault diagnosis*. Mechanical Systems and Signal Processing. 21. 2560-2574. 10.1016/j.ymssp.2006.12.007.

Williams B. C. & Nayak P. P. (1996). *Immobile Robots AI in the New Millennium*. AI Magazine, 17(3), 16. https://doi.org/10.1609/aimag.v17i3.1229.

Willshaw D. J. and Von Der Malsburg C. (1976). How patterned neural connections can be set up by self-organization. Proc. R. Soc. Lond. B. 194, 431-445 (1976).

Wilson, G. and McMillan, D. (2014). *Assessing Wind Farm Reliability Using Weather Dependent Failure Rates.* Journal of Physics: Conference Series, Volume 524, The Science of Making Torque from Wind 2014 (TORQUE 2014) 18–20 June 2014, Copenhagen, Denmark

Wold S., Esbensen K., Geladi P. (1987). *Principal component analysis*. Chemometrics and Intelligent Laboratory Systems Volume 2, Issues 1–3, August 1987, Pages 37-52.

Wood R.K., Stephens K.G., Barker B.O. (1979). *Fault Tree Analysis: An emerging methodology for instructional science.* Instr Sci 8, 1–22 (1979). https://doi.org/10.1007/BF00054979.

Wu D.& Yang Q. & Tian F. & Zhang D. (2010). *Fault Diagnosis Based on K-Means Clustering and PNN (PDF)*. Intelligent Networks and Intelligent Systems, International Workshop on. 173-176. 10.1109/ICINIS.2010.169.

Xu Q. & Migault D. & Sénécal S. & Francfort S. (2011). *K-means and adaptive k-means algorithms for clustering DNS traffic*. 281-290. 10.4108/icst.valuetools.2011.245598.

Xu R., Kwan C. (2003). *Robust isolation of sensor failures. Asian Journal of Control*. 5 (2003) 12–23.

Yam R.C.M., Tse P.W., Li L., Tu P. (2001). *Intelligent predictive decision support system for condition-based maintenance*. International Journal of Advanced Manufacturing Technology 17 (2001) 383–391.

Yam, R., Tse, P., Li, L. and Tu, P. (2001). *Intelligent predictive decision support system for condition-based maintenance*. International Journal in Advances Manufacturing Technology, Vol. 17, No. 5, pp.383–391.

Yang H., Mathew J., Ma L. (2005). *Fault diagnosis of rolling element bearings using basis pursuit.* Mechanical Systems and Signal Processing 19(2):341-356.

Yang J., Zhang D. and Yang J. (2007). *Constructing PCA Baseline Algorithms to Reevaluate ICA-Based Face-Recognition Performance*. In IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics), vol. 37, no. 4, pp. 1015-1021, Aug. 2007, doi: 10.1109/TSMCB.2007.891541.

Yang, S., Liu, C., Zhou, X., Liang, W. and Miao, Q (2012). *Investigation on data-driven life prediction methods*. International Conference on Quality, Reliability, Risk, Maintenance, and Safety Engineering (ICQR2MSE),: IEEE; 2012.

Yiakopoulos C. & Gryllias K. & Antoniadis I. (2009). *Rolling Element Bearing Fault Classification Using K-Means Frequency Domain Based Clustering*. 10.1115/DETC2009-87369.

Yiakopoulos C. T., Gryllia K. C., Antoniadis S. I. A. (2011). *Rolling element bearing fault detection in industrial environments based on a K-means clustering approach*. Expert Systems with Applications, vol. 38, no. 3, pp. 2888-2911, March 2011.

Yin S & Zhu X. & Jing C. (2014*). Fault detection based on a robust one class support vector machine*. Neurocomputing. 145. 263–268. 10.1016/j.neucom.2014.05.035.

Yin S. & Zhu X. & Jing C. (2014). *Fault detection based on a robust one class support vector machine*. Neurocomputing. 145. 263–268. 10.1016/j.neucom.2014.05.035.

Yip K. & Cheng C. & Gerstein M. (2013). *Machine learning and genome annotation: A match meant to be?*.Genome biology. 14. 205. 10.1186/gb-2013-14-5-205.

You M.Y., Liu F., Wang W., Meng G. (2010). *Statistically Planned and Individually Improved Predictive Maintenance Management for Continuously Monitored Degrading Systems.* IEEE Transactions on Reliability 59(4):744-753.

Yu G., Li C. & Sun J. (2010). *Machine fault diagnosis based on Gaussian mixture model and its application*. Int J Adv Manuf Technol 48, 205–212 (2010). https://doi.org/10.1007/s00170-009-2283-5.

Yue, S., Li, P. & Hao, P. *SVM classification:Its contents and challenges*. Appl. Math. Chin. Univ. 18, 332–342 (2003). https://doi.org/10.1007/s11766-003-0059-5.

Zaal T. (2016). *Profit-Driven Maintenance for physical assets*. Maj Engineering Publishing.

Zaki A. K. (2014). *NoSQL Databases: New Milleneum Database For Big Data, Big Users, Cloud Computing and Its Security Challenges*. International Journal of Research in Engineering and Technology, vol. 3, no. 3, 2014.

Zhang C. & Bickis M. & Wu F.-X. & Kusalik, A. (2006). *Optimally-connected Hidden Markov models for predicting MHC-binding peptides*. Journal of bioinformatics and computational biology. 4. 959-80. 10.1142/S0219720006002314.

Zhang S., Ganesan R. (1997). *Multivariable trend analysis using neural networks for intelligent diagnostics of rotating machinery*. Transactions of the ASME. Journal of Engineering for Gas Turbines and Power 119 (1997) 378–384.

Zhang Y., Bingham C. M., Gallimore M., Yang Z. and. Chen J. (2012). *Sensor fault detection for industrial systems using a hierarchical clustering-based graphical user interface*. 2012 IEEE International Conference on Multisensor Fusion and Integration for

Intelligent Systems (MFI), Hamburg, Germany, 2012, pp. 389-394, doi: 10.1109/MFI.2012.6343071.

Zhang Y., You L. and Jia C. (2017). *Fault detection and diagnosis using Bayesian-network inference.* IECON 2017 - 43rd Annual Conference of the IEEE Industrial Electronics Society, Beijing, China, 2017, pp. 5049-5053, doi: 10.1109/IECON.2017.8216872.

Zhang Y., Zhang J., Jing MA, Wang Z. (2009). Fault Detection Based on Hierarchical Cluster Analysis in Wide Area Backup Protection System. Energy and Power Engineering, Vol.1 No.1, August 2009.

Zhang Z. & Dai B.T. & Tung, A. (2008). *Estimating local optimums in EM algorithm over Gaussian mixture model*. 1240-1247. 10.1145/1390156.1390312.

Zhang, S. and Ganesan, R. (1997). *Multivariable trend analysis using neural networks for intelligent diagnostics of rotating machinery*. Trans. ASME Journal of Engineering for Gas Turbines and Power, Vol. 119, pp. 378-384.

Zhang, Yagang. (2009). *Fault Detection Based on Hierarchical Cluster Analysis in Wide Area Backup Protection System*. Energy and Power Engineering. 01. 21 27.10.4236/epe.2009.11004.

Zhao Y. & Balboni F. & Arnaud T. & Mosesian J. & Ball R. & Lehman B. (2014). *Fault experiments in a commercial-scale PV laboratory and fault detection using local outlier factor*. 2014 IEEE 40th Photovoltaic Specialist Conference, PVSC 2014. 3398-3403. 10.1109/PVSC.2014.6925661.

Zhaoyang T., Jianfeng L., Zongzhi W., Jianhu Z., Weifeng H. (2011). *An evaluation of maintenance strategy using risk based inspection*. Safety Science Volume 49, Issue 6, July 2011, Pages 852-860.

Zhiqiang Ge and Zhihuan Song (2007). *Process Monitoring Based on Independent Component Analysis-Principal Component Analysis (ICA-PCA) and Similarity Factors*. Ind. Eng. Chem. Res. 2007, 46, 2054-2063.

Zhong G, Wang L.N.,Ling X., Dong J. (2016). *An overview on data representation learning: From traditional feature learning to recent deep learning.* Volume 2, Issue 4, December 2016, Pages 265-278

Zhong Q., Rüschoff J., Guo T. et al. (2016). I*mage-based computational quantification and visualization of genetic alterations and tumour heterogeneity*. Sci Rep 6, 24146 (2016). https://doi.org/10.1038/srep24146.

Zhou B., Qi F. and Tao H. (2017). *Condition-based maintenance modeling for a two-stage deteriorating system with random changes based on stochastic process*. Journal of Quality in Maintenance Engineering, Vol. 23 No. 4, pp. 383-399.

Zhou J., C. K. Pang and W. Yan (2017). *Gaussian mixture model for new fault categories diagnosis*. 22nd IEEE International Conference on Emerging Technologies and Factory Automation (ETFA), Limassol, 2017, pp. 1-6, doi: 10.1109/ETFA.2017.8247640.

Zhou Z., Wen C., Yang C. (2015). *Fault Detection Using Random Projections and k-Nearest Neighbor Rule for Semiconductor Manufacturing Processes*. Semiconductor Manufacturing, IEEE Transactions on. 28. 70-79. 10.1109/TSM.2014.2374339.

Zhu X. (2005). *Semi-Supervised Literature Review*. Technical Report #1530, University of Winsconsin Madison.

Zope Object Database, (2017). *ZODB Book Documentation Release 1*.

https://media.readthedocs.org/pdf/zodb/latest/zodb.pdf, last assessed 2019/03/15.

Appendix

In the beginning, four bearings were installed on a shaft that had a constant rotation speed of 2000 RPM by an AC motor coupled to the shaft via rub belts (NASA, 2019). A radial load of 6000 lbs is applied onto the shaft and bearing by a spring mechanism and all bearings are force lubricated (NASA, 2019).

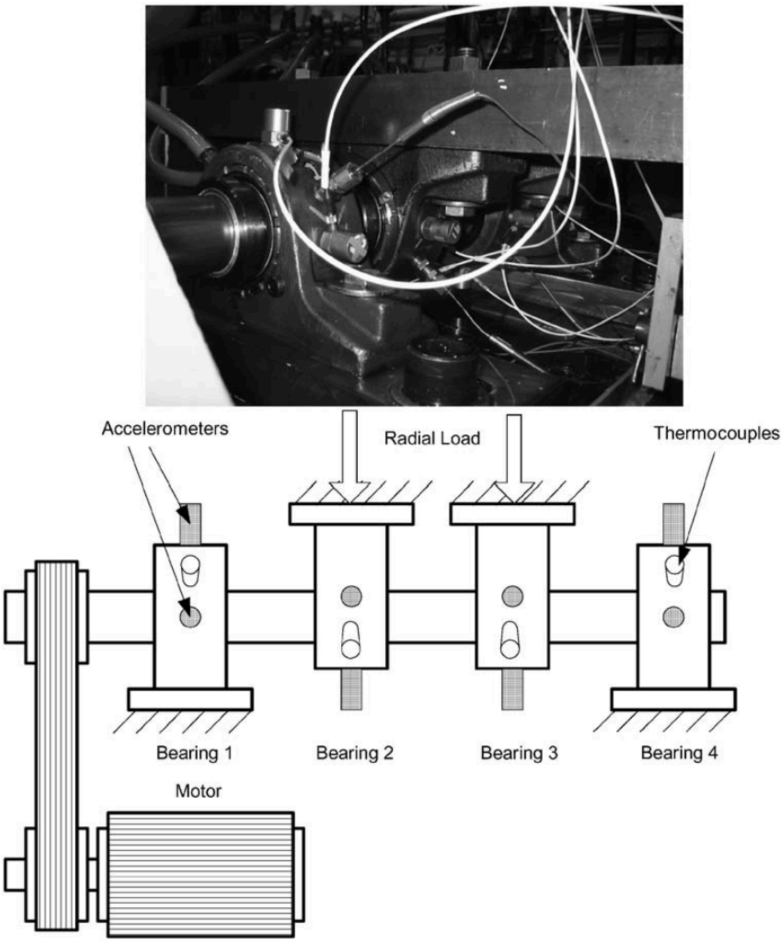


Figure 47: Bearing Test Rig and Sensor Placement Illustration (Qiu et al., 2005)

In Figure 47 it can be seen the experiment set up. Rexnord ZA-2115 double row bearings were installed on the shaft and PCB 353B33 High Sensitivity Quartz ICP accelerometers were installed on the bearing housing (two accelerometers for each bearing [x- and y-axes] for data set 1, one accelerometer for each bearing for data sets 2 and 3). All failures occurred after exceeding designed lifetime of the bearing which is more than 100 million revolutions which divided by the 2000 RPM gives us the designed life equal to 34.7 days. A magnetic plug installed in the oil feedback pipe collects debris from the oil as evidence of bearing degradation and the test will stop when the accumulated debris adhered to the magnetic plug exceeds a certain level and causes an electrical switch to close (Qiu et al., 2005).