Tampere University

Saara Väänänen

PROCESS INDUSTRY DATA REGRESSION ANALYSIS FOR PREDICTIVE MAINTENANCE

Faculty of Engineering and Natural Sciences
Master of Science Thesis
February 2020

ABSTRACT

Saara Väänänen: Process industry data regression analysis for predictive maintenance
Master of Science Thesis
Tampere University
Energy and Biorefining Engineering
February 2020

The goal of this thesis was to explore the possibilities of combining data from different sources in a perspective of predictive maintenance. The data was collected from two pulp and paper industry plants, including vibration monitoring and control loop monitoring key performance indicators, as well as measurement data from process equipment. In addition to local data aggregation and analyzing, the possibilities of cloud-based systems in predictive maintenance were discussed.

Industrial processes rely on the operation of numerous machines. The truth is that every machine eventually breaks down unless it is being maintained. By predictive maintenance, premature breakdowns can be avoided by maintaining the items' normal operating conditions without immoderate overhauls. A reliable and sufficient amount of condition monitoring data is the basis of predictive maintenance. Today, data collection is cheap and well exploited in the process industry. However, one big challenge is the distribution of information. This means that the data is collected and stored to separate systems that are not communicating with each other.

The potential of data aggregation from different sources has not yet been fully exploited in predictive maintenance. In this work, the data aggregation was examined by combining values using timestamps. In the experimental part of the work, relationships between variables were further examined by regression analysis. The used analysis method did not confirm causalities between examined variables, but the study did partially substantiate the hypothesis about the benefits of combining data. The results increased understanding of versatile and efficient use of data in the pulp and paper industry.

Finally, this study lays the groundwork for future research into cloud-based data analysis methods in predictive maintenance. Although this study was based on only two examples, the findings suggested that the data to be transferred to the cloud should be as raw as possible. In the higher-level computational indices, data storage cycles are typically longer, and part of the information has already been lost. Moreover, it is discussed how data aggregation and regression analysis could be utilized to improve Valmet's current maintenance management product.

Keywords: Predictive Maintenance, Condition Monitoring, Vibration Monitoring, Control Loop Monitoring, Key Performance Indicator, KPI, Regression Analysis

TIIVISTELMÄ

Saara Väänänen: Regressioanalyysi prosessiteollisuuden datalle ennakoivassa kunnossapidossa
Diplomityö
Tampereen yliopisto
Energia- ja biojalostustekniikka
Helmikuu 2020

Tämän opinnäytetyön tavoitteena oli tutkia datan yhdistämistä eri lähteistä ennakoivan kunnossapidon näkökulmasta. Analysoitava data kerättiin kahdelta eri sellu- ja paperiteollisuuden laitokselta. Data sisälsi värähtelynvalvonnan ja säätöpiirien valvonnan tunnuslukuja, sekä prosessilaitteiston mittadataa. Paikallisen datan käsittelyn ja analysoinnin lisäksi työssä pohdittiin pilvipohjaisten järjestelmien mahdollisuuksia ennakoivassa kunnossapidossa.

Prosessiteollisuus on riippuvainen lukuisien koneiden toiminnasta. Totuus on, että jokainen kone lopulta rikkoutuu, ellei sen kuntoa ylläpidetä. Koneen ennenaikainen hajoaminen voidaan välttää ennakoivan kunnossapidon avulla, jonka tarkoituksena on ylläpitää koneen normaalia toimintaa ilman ylimääräisiä huoltoja. Ennakoivan kunnossapidon perustana on luotettava ja riittävä määrä kunnosta kertovaa dataa. Datan keruu on tänä päivänä halpaa ja se onkin hyvin hyödynnetty prosessiteollisuudessa. Yksi iso haaste on kuitenkin tiedon hajanaisuus. Tämä tarkoittaa sitä, että data kerätään ja talletetaan erillisiin järjestelmiin, jotka eivät kommunikoi keskenään.

Eri lähteistä saatavan tiedon yhdistämistä ei ole täysin hyödynnetty ennakoivassa kunnossapidossa. Tässä työssä datan yhdistämistä tutkittiin liittämällä muuttujat toisiinsa aikaleiman avulla ja työn kokeellisessa osiossa muuttujien välisten suhteiden tukitkimiseen käytettiin regressioanalyysiä. Vaikka käytetty analyysimenetelmä ei vahvistanut tietoa muuttujien välisistä syyseuraus suhteista, se kuitenkin tuki oletusta datan yhdistämisen hyödyllisyydestä. Tulokset lisäävät ymmärrystä monipuolisesta ja tehokkaasta datan käytöstä prosessiteollisuudessa.

Lopuksi tämä työ luo perustan jatkotutkimukselle koskien pilvipohjaisia data-analyysimenetelmiä ennakoivassa kunnossapidossa. Vaikka suoritettu tutkimus perustuu vain kahteen esimerkkitapaukseen, tulokset viittaavat siihen, että pilveen siirrettävät tiedot tulisi olla mahdollisimman vähän käsiteltyjä. Korkeamman tason laskennallisissa indekseissä talletussyklit ovat tyypillisesti pidempiä, jolloin osa alkuperäisestä tiedosta menetetään. Lisäksi työssä pohditaan mahdollisuutta parantaa Valmetin kunnossapitoon käytettävää tuotetta yhdistämällä regressioanalyysi osaksi nykyistä ratkaisua.

Avainsanat: ennakoiva kunnossapito, kunnonvalvonta, värähtelynvalvonta, säätöpiirien valvonta, avaintunnusluku, regressioanalyysi

PREFACE

This Master's Thesis was done for the Valmet Automation services department. I would like to equally thank all the colleagues in Valmet who have helped me and provided guidance in my work. Special thanks to Ville Hietanen, Tuomas Repo, and Juha Virtanen for providing data for this study, even though I know you all were busy with your own work. I also wish to thank Teemu Kiviniemi and Krister Sällinen for giving me this research topic. You gave me the opportunity to learn about an important industry sector, condition monitoring.

Moreover, I want to express my gratitude to Valmet Info-team, that has the best teammates. Your kindness, sense of humor, and admirable commitment to coffee break times made me feel grateful every day. Besides the support I received at Valmet, I sincerely appreciate the guidance of my supervisor, university lecturer Henrik Tolvanen. He has a valuable ability to perceive the big picture in a short time, which helped me in outlining this work.

I want to owe remarkable thanks to my intelligent friend and workmate, Sofia Koivumäki, for supporting me at work and in Pyynikki jogging paths –regardless of the weather. Finally, I want to thank my family, friends, and fiancé for their endless encouragement. Ville, your presence makes the moments colorful. Special thanks to my diligent grandmother, who has supported me in my studies from the very beginning.

Tampere, 12th February 2020

Saara Väänänen

CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

AWS	Amazon Web Services
CLM	Control Loop Monitoring
CSV	Comma-Separated Values
CTI	Control Travel Index
DCS	Distributed Control System
FFT	Fast Fourier Transform
Hz	Hertz
IAE	Integrated Absolute Error
KPI	Key Performance Indicator
MM	Machine Monitoring
ODS	Operating Deflection Shape
OI	Oscillation Index
PID	Proportional-Integral-Derivative
PP	Peak to Peak
RF	Rotational Frequency
RMS	Root Mean Square
SQL	Structured Query Language
ToPi	Total Performance Index
VI	Variability Index

LIST OF SYMBOLS

Sign	Description	Unit
a	acceleration	m/s^2
d	density	kg/m^3
p	pressure	bar
T	temperature	$^\circ$C
v	velocity	m/s

1 INTRODUCTION

The pulp industry has stayed solid for a few years now, and the total revenue in this industrial sector has been growing. (IBISWorld 2019; Pöyry 2019). Globally, the overall production in the pulp and paper industry is steadily increasing (Finnish forest industries 2017). Even though digitization and internet usage has cut the demand for various paper products, such as newspapers and magazines, the total paper and board production has been increasing. This is mostly due to the growth in packing product and tissue paper consumption. A similar trend can be observed in Finland, where the wood-processing industry remains an important industry sector accounting for 20 % of the total export revenue (Ministery of Agriculture and Forest of Finland 2018).

Technological development in the pulp and paper industry is still evolving. With this development, optimization, automation, and information technology have shown increasing importance. New technologies enable better machinery performance with lower investments, yet still having some challenges. Without a proper maintenance management system, vital signs of reduced machinery condition might be missed. Paper production is, in general, a continuous process, where unexpected equipment failures cause unprofitable downtime for the company owners. The need for accurate maintenance systems and the increased knowledge of customers, who make the investments, have moved the pressure from improving the physical properties of the machine to develop systems with capabilities to predict disturbances. (Mikkonen et al. 2009; Windrock 2018)

A reliable and sufficient amount of condition monitoring data is the basis of process maintenance. Reduced sensor and analyzer prizes have enabled the collection of large amounts of precise data, which is today considered as a fundamental operation in the process industry (Martinsuo and Kärri 2017).

However, one big challenge is the distribution of information. This means that the data is collected and stored to separate systems that are not communicating with each other. Lack of communication is particularly challenging in maintenance management, which is collaboration between different industrial processes. Recent developments in cloud-based systems have overcome this problem to some extent by enabling the data aggregation from separate databases.

Valmet is continually developing its industrial internet platform, which utilizes cloud computing (Valmet 2019b). By applying data analytics to various industrial processes, new services and solutions can be offered to increase customer value. In cloud-based systems, new memory technologies, flexible computational power, as well as the data communication capabilities, have made the data utilization economically viable (Dimecc 2016). When the data is in the same place and the same format, further data analysis can be done. In predictive maintenance and condition monitoring, this could be utilized by developing more efficient fault detection and prediction tools. The potential of data aggregation from different sources has not yet been fully exploited in predictive maintenance. (C.K.M. Lee, Ty Cao and Kam Hung. Ng 2017)

The aim of this thesis is to examine if combining data from different sources benefit predictive maintenance. The data is collected from various databases and transferred to a cloud-based data warehouse. Data transfer to the cloud and its web-based processing has been studied in Valmet for several years. Harmaala (2017) and Nopanen (2017) have introduced Valmet Industrial Internet data structure, data transfer and analysis possibilities in more detail in their master's theses. This work focuses on examining if it is possible to add new value to the customer's process maintenance by combining and analyzing data from different data sources. Following research questions will be discussed:

1. How is the condition of a machine monitored in the pulp and paper industry?
2. What defines the performance of a process device?
3. How to combine process data from different sources?
4. Does data integration from different sources benefit predictive maintenance in the pulp and paper industry?

The first two research questions are discussed in the theoretical part of the work. Chapter 2 presents the background of maintenance strategies and machine monitoring techniques. The main focus is on vibration monitoring, which is a widely used condition monitoring method in paper and board industry. Vibration indices compose the first data source of the work. The second data source that is examined contains control loop monitoring indicators. Chapter 3 attempts to provide a brief summary of the literature relating to control loop performance in the process industry, and some essential control loop monitoring key performance indicators are presented.

Chapter 4 summarizes the data processing and analyzing methods applied in this work. The experimental part of the work consists of two different cases. Case 1 combines machine monitoring and control loop monitoring data, and case 2 combines machine monitoring and equipment data. The current situation in most cases is that the examined data is stored in separate systems. No previous study has investigated if combining these separate data sets gives a better holistic view of the plant's condition compared to the current situation. The dependencies of the parameters from different sources are examined by regression analysis. Results from the analysis are presented and discussed in chapter 5. Chapter 6 proposes future development potential in this field and finally, Chapter 7 summarizes the most important findings.

2 PREDICTIVE MAINTENANCE IN PULP AND PAPER INDUSTRY

Modus operandi and production philosophies of process industry companies have changed considerably during the past few years. Part of this change has been the development of maintenance management. Traditionally companies in pulp and paper industry have competed with the efficiency of production machinery, but along the technical development, machinery features have consolidated. Competitive advantage is searched from other sectors, especially from the area of maintenance. In the field of maintenance, both economic significance and the needed level of competence has increased. (Mikkonen et al. 2009, pp. 27–29)

Manufacturing methods and techniques influence productivity, effectiveness, and profitability of the plant (Rao 1996). In complex process industry systems, several failure patterns can be predicted by maintenance management systems. Unsuitable operating conditions may lead to failure and destroy the healthy condition of machinery. For example, vibration, heat, dust, pressure, humidity, and corrosion may abet failures. The expensive machinery breakdowns and accidents can be reduced by predictive maintenance management.

This work aims to study if it is possible to get advantage in predictive maintenance by combining data from different sources. Figure 2.1 illustrates the three data sources of the study from the perspective of predictive maintenance. The first data source consists of vibration indices, which are calculated from vibration monitoring data. Vibration is an excellent condition indicator, especially for rotating machinery. The second data sources consist of process equipment measurements such as pressure, temperature, density, and flow. Whereas vibration indices are refined data, this second data source includes only raw equipment data. Both of these two data can be utilized in condition monitoring, which is one field of predictive maintenance. Predictive maintenance and its subcategory condition monitoring are discussed in more detail in chapter 2.1.

All industrial processes need adjustment and control to some extent. The adjustment can be manually made by an operator or automatically made by a controller. In automation systems, control loop (KPIs) represent the performance of the controllers. These control loop monitoring KPIs compose the third data source of the work.

The conducted works consisted of two cases. In case 1, it was examined how vibration indices and condition monitoring KPIs could be combined and utilized in predictive

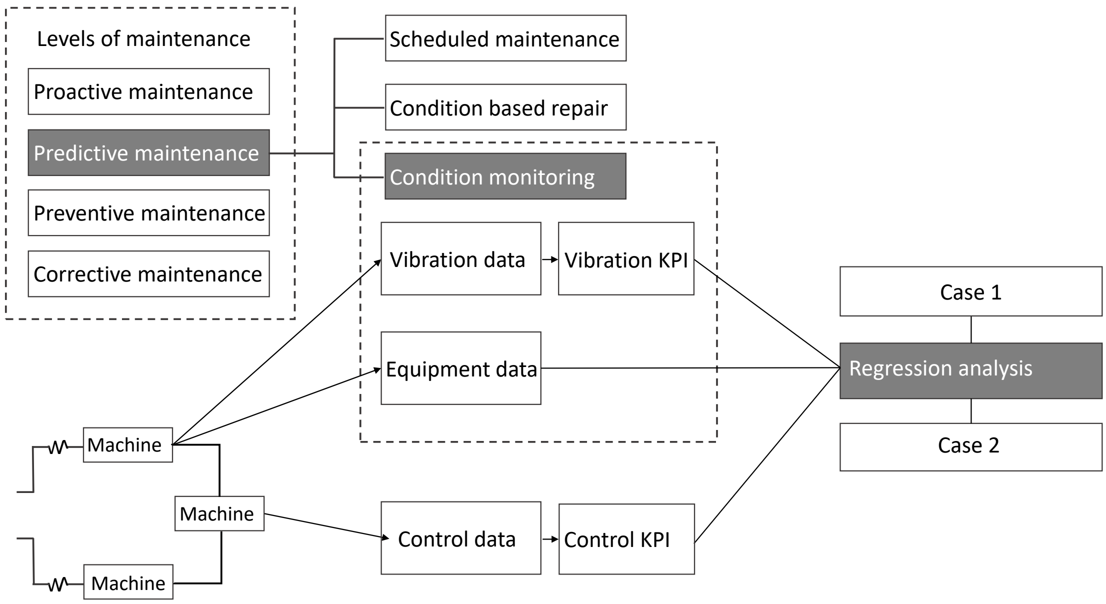


Figure 2.1. *Examined data sources from the perspective of preventive maintenance.*

maintenance. Case 2, in turn, studied how vibration indices and equipment data could be combined. The aim in both cases was to examine relations between data sets and conclude if it is reasonable to combine data into the same system.

2.1 Levels of maintenance management

Maintenance is a combination of technical administrative and managerial actions (PSK Standard Association 2011). The intention of maintenance is to retain or restore an item to a state in which it can perform the required function during its whole life cycle. In pulp and paper industry, the paper machines are supposed to run round the clock, and malfunction of a device may lead to a machine shutdown (Jokinen et al. 2012). Even a short machine stop leads to a web break causing extra costs for the owner. By a well-implemented maintenance management system, these unexpected machine stops can be avoided. Maintenance is not a cost but an essential factor in production to ensure the competitiveness of a plant.

Maintenance management is a field that has been examined a lot. Various literature sources use a bit different terminology, which can be miscellaneous. Therefore it is essential to define which interpretation is used in this work. Two of the machinery maintenance subcategory classifications that complete each other are discussed.

Mikkonen et al. (2009) have presented a widely used concept of four levels of machinery maintenance. These four levels are corrective maintenance, preventive maintenance, predictive maintenance, and proactive maintenance. This classification is illustrated in figure 2.2. Another perspective is to categorize maintenance activities based on the recommended time interval of technical servicing for each machine (Eisenmann Sr. and

Eisenmann Jr. 2005, p. 703). The three categories are reaction-based, time-based, and condition-based maintenance.

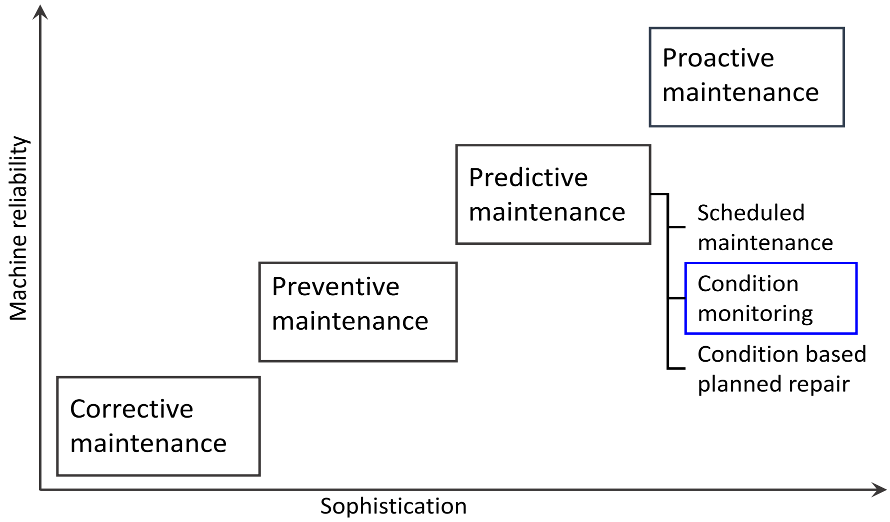


Figure 2.2. *Four levels of maintenance. Adopted from (Mikkonen et al. 2009; PSK Standard Association 2010)*

Reaction-based maintenance corresponds to corrective maintenance, which operation principle is a continuous production until a failure occurs. Corrective maintenance requires little, if any, advanced planning. In process industry, this type of maintenance typically leads to unplanned and unpredictable shutdowns causing extra costs for the owner. When machines are allowed to run until a failure occurs, sudden machinery failures may cause safety risks. Corrective maintenance is suitable only in the facilities where the number of critical machines is low or the operation is not dependent on an individual machine. (Eisenmann Sr. and Eisenmann Jr. 2005; Mikkonen et al. 2009)

In process industry, there are many critical machines to operate. A better option than acting only when a failure occurs is to plan and schedule machinery maintenance ahead. This is called preventive maintenance (Mikkonen et al. 2009, p. 22) or time-based maintenance (Eisenmann Sr. and Eisenmann Jr. 2005, p. 703). Preventive maintenance evolved due to the economic impact and safety unsuitability of corrective maintenance. Within preventive or time-based maintenance category, machinery maintenance is scheduled on a periodic basis when also the mechanical inspections, disassembles and reassembles are done. In most cases, preventive maintenance is more cost-efficient than corrective maintenance, still having its deficiencies.

First of all, unexpected machinery failures may also occur during the scheduled normal operating period (Eisenmann Sr. and Eisenmann Jr. 2005, p. 704). Similarly to corrective maintenance, this leads to unplanned production breaks when the losses can be considerable. Besides that, during the maintenance break, unnecessary machine overhauls are done with an intention to prevent future faults. Replacement of machines in good condition causes redundant extra costs and often do more damage than allowing them to run without overhaul. Mechefske (2005) has introduced "a bathtub curve", which illustrates the frequency of failures. According to the source, a relatively high rate of failures occurs at the beginning of a machine's life. These wear-in failures are caused by assembly mistakes, design errors, manufacturing defects, and installation problems. Therefore, frequent machine replacements increase the total amount of failures. The bathtub curve is presented in figure 2.3.

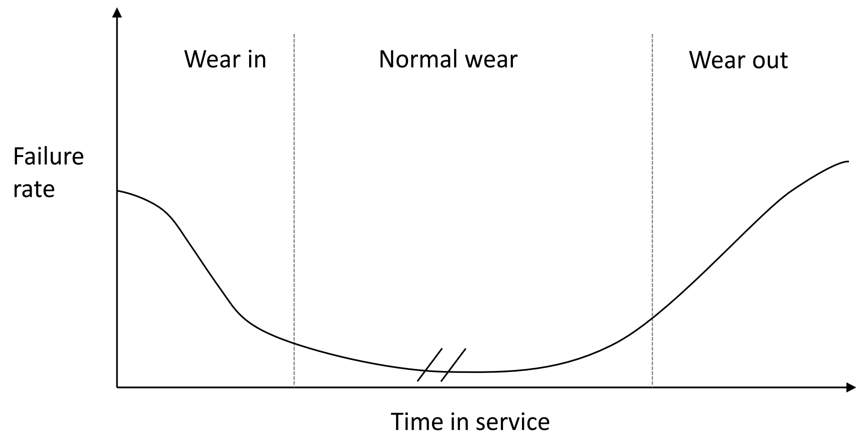


Figure 2.3. *Bathtub curve illustrates the frequency of failures. Adopted from (Mechefske 2005).*

The first two maintenance strategies corrective maintenance ('run to failure') and preventive maintenance ('scheduled maintenance') lack of predictive capabilities. They are therefore unacceptable when it comes to complex process and manufacturing operations (Rao 1996). The most cost-efficient maintenance activity is predictive (Mikkonen et al. 2009, p. 22) or condition-based (Eisenmann Sr. and Eisenmann Jr. 2005, p. 703–705) maintenance. If the machine is not broken, then there is no need to fix it. To achieve a holistic view of a machine's condition without opening it, machine's performance must be evaluated from many aspects. Several parameters are related to a machine's mechanical condition such as vibration, temperature, quality of oil, pressure, and operating speed of the machine (Galar et al. 2011). In PSK Standard 7501 (PSK Standard Association 2010), predictive maintenance is further divided into three subcategories. They are scheduled maintenance, condition monitoring, and condition-based planned repair. In this thesis, the focus is on condition monitoring which is described in more detail in chapter 2.1.1.

The most sophisticated maintenance method is proactive maintenance (Mikkonen et al. 2009). Its purpose is to improve the machine's reliability and maintainability performance without changing its way to operate. This can be done for example with reparations, redesigning, replacing original parts with new ones, or with modernization. In summary, maintenance management methods can be roughly divided into four levels. As seen in figure 2.2, the reliability of the machinery is dependent on the maintenance method. The predictability of failures increases within the method's degree of sophistication. Predictability enables planning and scheduling of machinery maintenance. This work intends to mainly focus on the second-highest maintenance level, predictive or condition-based maintenance.

2.1.1 Machine condition monitoring

Condition monitoring is a complex of technical, administrative, and managerial related actions (PSK Standard Association 2011). Fundamental objectives of condition monitoring are overall equipment effectiveness and high dependability (PSK Standard Association 2010). By condition monitoring, it is possible to detect faults in early stage. Early detection of faults enables to schedule optimal maintenance breaks, reduces unforeseeable shutdowns and extends the lifetime of components in a controlled way. This is important for steady, safe, and reliable production.

There are several condition monitoring techniques for different industrial sectors. These techniques include for example physical parameters related to lubrication and vibration analysis, wear particle monitoring, force, sound, temperature, product quality, odor, and visual inspections. The basic principle of condition monitoring is to measure some parameter of a machine which changes with the health or condition of a machine (Rao 1996). Regular monitoring of condition parameters and observation of changes gives valuable information about the prevailing condition. After the detection of a change, more detailed analysis can be performed.

In this work, the condition of machinery parts is monitored by a vibration monitoring system. Vibration tends to increase as a machine moves away from its smooth running condition. Typically the warning signs appear months before the final break down, starting with increasing or abnormal vibration. If the problem is not solved, noise, warming, or in some cases smoke, may follow the vibration before an emergency stop. An uncontrolled vibration may cause serious reliability, productivity, quality, and safety problems. For rotating machinery parts, vibration monitoring is the most common condition monitoring method. (Goyal and Pabla 2015; *Machine Condition Monitoring Technical Library* 2017)

2.1.2 Vibration monitoring

In industrial environment, almost every machine vibrates (Rao 1996). Vibration origins during the machine's operation from rotating or reciprocating parts such as shafts, bearings, turbines, and gearboxes. Mechanical vibration is periodic motion around the position of equilibrium. The motion can manifest in a structure or components of a machine. In most cases, the vibration is harmful because it causes energy losses, bearing damages, unbalance, miss alignment, resonance, and structural fatigue, which all lead to reduced service lifetime. (Mikkonen et al. 2009)

A mechanical system requires an initial disturbance to deviate from its equilibrium. A periodic excitation force, for example a rotary movement of a shaft, causes a machine's oscillation. When it comes to mechanical systems, vibration is often perceived as free oscillatory motion without applied forces (Goyal and Pabla 2015). A comprehensive basic theory of vibration is introduced in more detail in many sources such as in (Eisenmann Sr. and Eisenmann Jr. 2005) and (Thomson 1981).

Display formats

Vibration monitoring consists of vibration measurements, signal monitoring, signal processing, and data interpretation (Goyal and Pabla 2015). With modern data acquisition technologies, the vibration signals are easy to collect, and the data can be used to evaluate a state of a machine without opening it. In order to draw conclusions from the vibration data, it has to be in an informative format. The best known vibration signal formats are time domain and frequency domain. (Eisenmann Sr. and Eisenmann Jr. 2005; Mechefske 2005)

Time domain format shows the vibration data as a function of time. In practice, the vibration is always measured as a function of time. Time domain vibration data includes a great amount of information which can be used for several analyzes. It provides valuable information for impulsive vibration signal troubleshooting. One of the most used parameter that is determined from time domain signal is amplitude (peak to peak). Amplitude indicates the energy content of the vibration and refers to the problem severity. More vibration indices that are calculated from the time-waveform domain are presented later in the chapter 2.3. Vibration signals consists of many different frequencies which together form a sum wave. A sum wave of three simple theoretical oscillators is illustrated in time domain in figure 2.4. (Mechefske 2005)

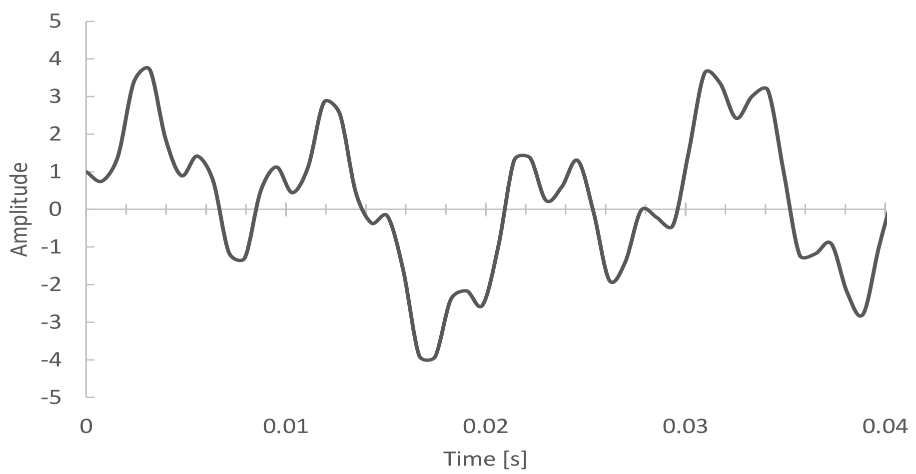


Figure 2.4. *A sum wave of three simple theoretical oscillators as a function of time.*

Sometimes time domain may even consist of too much data for clear fault analysis. To simplify the format, time domain data can be transformed to frequency domain. Frequency domain presents the vibration data as a function of frequency. As seen in figure 2.5, this format shows right away what frequencies the vibration data includes. In frequency domain, the vibration data is separated into discrete contributing frequencies. The size of a peak in a spectrum reveals the energy content of a certain frequency. The time-frequency transformation can be done for cases where the frequency does not vary within time. If the frequency is not time constant, the energy spreads out across the spectrum. Energy spreading happens for example with transient signal components. (Väänänen 1990)

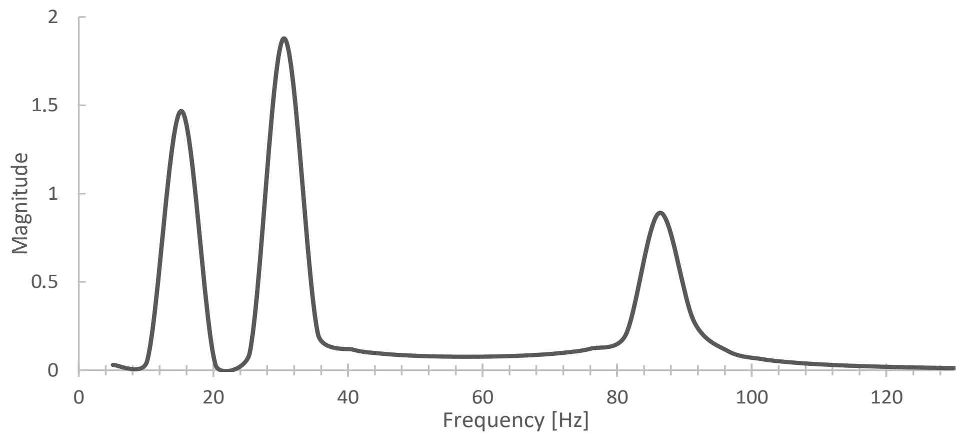


Figure 2.5. *Vibration data presented in frequency domain.*

The most used mathematical method to transform time domain vibration data into frequency domain is Fast Fourier Transform (FFT). FFT is an efficient and fast way to calculate a discrete Fourier transform. Faults generate a specific frequency response, which can be detected as peaks in the spectrum. To avoid wrong conclusions, the selection of the right frequency range and resolution for FFT is crucial. (Mechefske 2005; Väänänen 1990)

As seen in figure 2.5, the vibration data mainly consists of three different frequencies. The FFT transformation was done for the data seen in figure 2.4. In real cases, there are many more oscillators, each of them producing its own vibration curve. The signal profile becomes quite complicated when multiple oscillators with different frequencies and noise in the data are all combined in the same figure. However, the frequency domain does not have information concerning the shape of the vibration spectrum, so it is rather essential to observe the placement and height of the peaks. A growing peak refers to increasing vibration, and some faults can be seen as high harmonics of the natural frequency. For example, typical bearing faults and cavitation manifest as high harmonics of the rotational frequency in the frequency domain spectrum.

Envelope analysis

The rotational frequency and its harmonics, as well as the resonance frequencies together, form a base level vibration of a machine. Some incipient faults have such weak vibration levels that they disappear under the base level vibration (Mikkonen et al. 2009, p. 252). In envelope analysis, recurring low energy components are separated from the base level frequencies. The envelope method is based on the demodulation technique and filtering the amplitude modulated signal.

Amplitude modulation is a non-linear process, where the modulated signal is multiplied by another signal. In condition monitoring, a high-frequency carrier wave is multiplied by a fault frequency, which typically has a much lower frequency. This carrier wave may be, for example, the natural frequency of a roller bearing or structure, gear mesh frequency, or bearing pass frequency. As a result, new frequency components, which cannot be found from the separate signals, appear in the spectrum. These new spectral components are called sidebands. (Institute 2019; Mikkonen et al. 2009)

In envelope analysis, the desired frequency band is filtered from the time domain signal (Mikkonen et al. 2009, p. 221). The filtered signal is demodulated and amplified. In demodulation, the signal is rectified and the carrier wave is filtered away. The rectifying can be performed either by taking an absolute value of a signal or by raising it to a power of two. Further, FFT is performed so that the fault frequencies stand out in the envelope spectrum. The pilot frequency, its multiples, and possibly sideband frequencies can be seen in the generated envelope spectrum.

The envelope spectrum is used to identify different types of low energy faults such as inner and outer race damages (Mikkonen et al. 2009, p. 252). Peaks in the spectrum refer to failures which can be analyzed by determining the fault frequencies of the bearing. As for, increased overall vibration levels in the envelope spectrum without clear peaks is a typical sign of inadequate lubrication. In this work, the focus is on increased overall levels rather than in envelope peaks. Usually, the temperature and pressure of the lubricant are controlled by controllers. Poor control performance in these areas may affect the lubrication and manifest in the envelope spectrum as increased overall vibration levels.

2.2 Pulp and paper process

The wood processing industry remains an important industry sector in Finland, accounting for 20 % of Finland's export revenues (Ministery of Agriculture and Forest of Finland 2018). Two-thirds of the total production value is covered by the pulp and paper industry. In recent years pulp and board production in Finland has been steadily increasing, and the same trend can be perceived globally (Finnish forest industries 2017). While printing and writing paper production has declined, packing paper, board and tissue paper production has slowly increased. The Global paper and board production by grade are illustrated in figure 2.6.

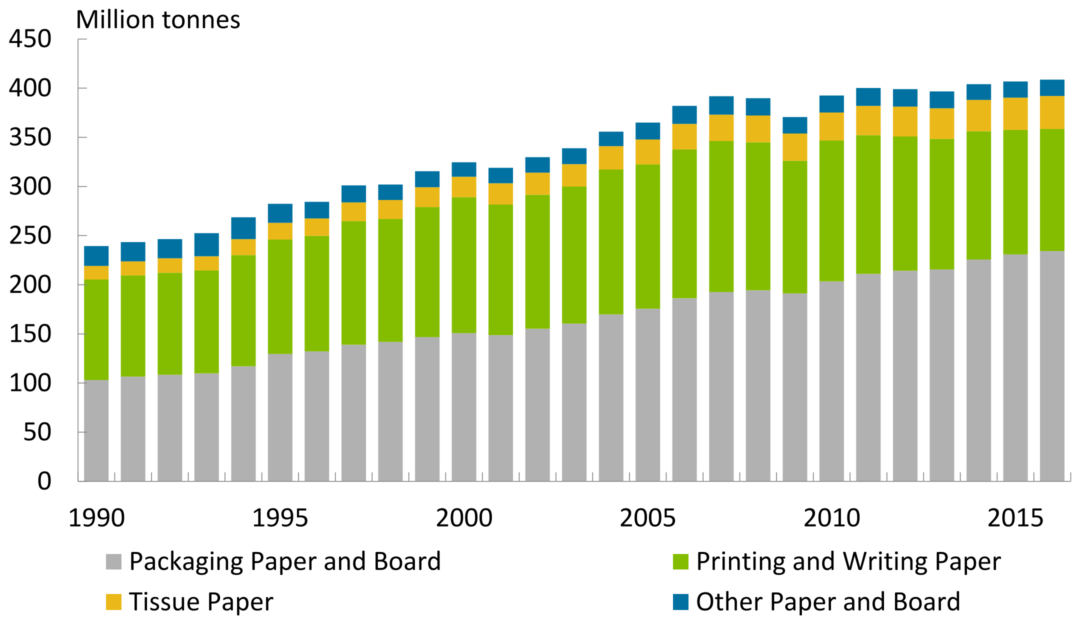


Figure 2.6. *Global paper and board production by grade (Finnish forest industries 2017).*

The demand for packing paper and board is expected to continue growing (Environmental Paper Network 2018). One primary reason for the growth is that with current paper and board coating methods, it is possible to replace plastic as a packaging material. In figure 2.6, also a small increase in tissue paper production can be observed. The tissue paper market growth is mostly attained in China and India, where with urbanization, an economically rising middle class forms a new customer group for hygiene products.

A typical feature of the pulp and paper industry is that the production runs 24/7. Breaks and quality faults in production directly affect the amount and price of the final product (Jokinen et al. 2012). With machine condition monitoring systems and maintenance plans, the operation conditions are maintained at the required level. In this work, both subjects of the study are placed in the pulp and paper industry. The next two chapters shortly describe the operational environment in both cases.

2.2.1 Green liquor pump

The kraft pulping process is a combination of many unit operations. The basic idea of the pulping process is to produce paper pulp by removing the fibre bonding agent – lignin – from lignocellulose. This can be done either chemically or mechanically. In chemical pulping, lignin is dissolved by using heat and chemicals. As for in mechanical pulping, lignin is softened by heat and mechanical stress. The resulting pulp from both processes may be used as raw material for paper and board production. (Andersson 2014; KnowPap 1997–2000; MetsäFibre 2020)

The examined green liquor pump is located in the chemical kraft process. The chemical part of the kraft process contains fibre line and regeneration line (KnowPap 1997–2000).

At the fibre line, the wood chips are cooked with white liquor. Cooking dissolves lignin and part of the hemicellulose, enabling separation of the cellulose for further processing. One by-product of the kraft process is black liquor, which consists of dissolved wood and spent cooking chemicals.

In the chemical recovery line, black liquor is rectified and utilized by burning it in a recovery boiler. The organic matter releases energy, which is used to produce steam and electricity. Sulfate pulp mills that burn black liquor are therefore energy self-sufficient in energy and steam production. The incombustible part of the black liquor (molten inorganic matter) is removed from the bottom of the furnace as smelt. This mixture of molten salts is further dissolved in water and weak white liquor when together they form green liquor. In the examined case, the green liquor is stored in the smelt dissolving tank. The dissolving tank has two identical pipelines to a balancing reservoir. The considered pump is used to pump green liquor from the dissolving tank to the balancing reservoir. (KnowPap 1997–2000)

2.2.2 Board machine

Pulp is a raw material to almost all paper and board products (American Forest Paper Association 2019). After the pulping process, cellulose fibers are mixed with water and possible fillers, adhesives, chemical additives, and coating agents. The formed stock suspension is distributed to a thin web followed by operations which purpose is to remove water from the slurry. The operating principle of a board machine is very close to a paper machine. The most significant difference between paper and board machines is that in the board machine, there are as many headboxes as there are layers in the final product. This makes the board basis weight higher than in paper. The structure of a board machine is illustrated in figure 2.7. (KnowPap 1997–2000)

The cardboard making process proceeds so that the furnish with high water content is first pumped to the headboxes from where it is distributed across the width of the fourdrinier wires. On a wire, the water drains through the fabric, forming a fibre mat called web. In the middle of the wire section, web layers are merged, and the layered wet web travels along the wire to the pick-up roller.

Pick-up roller literally picks up the web from the forming section to the press section, where the wet web is carried between a series of rollers (Valmet 2010). High pressure between rollers presses out the water from the wet web and compresses it. The pressing begins carefully so that fines would not flush away from the web. Too excessive and fast pressing may also crush the fibre network. As the water content decrease, the web travels through several press nips where the nip force is gradually increased. (KnowPap 1997–2000)

After the press section, the felt carries the web to the dryer where water is further removed through evaporation. The purpose is to achieve the properties required for the finished board by drying the wet web as much as possible. In the examined board mill, the final

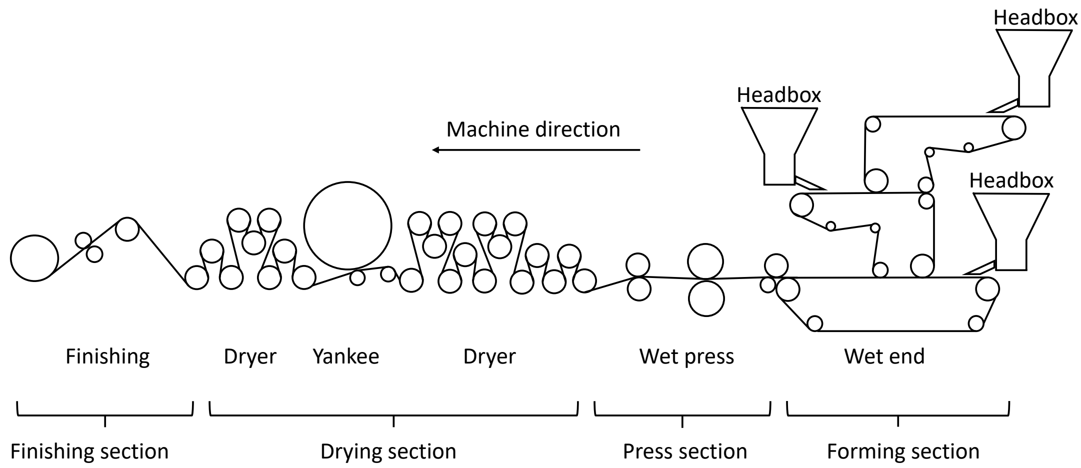


Figure 2.7. *Structure of a fourdrinier board machine.*

product is folding boxboard. Its smoothness and gloss are improved in the drying section with a glazing barrel, also known as Yankee cylinder. This Yankee cylinder is located between the two drying sections, as seen in figure 2.7.

The last step in the board making process is finishing (KnowPap 1997–2000). Finishing methods depends on the grade requirements and produced board type. Finishing methods that are typically done for all grades are winding and reeling. Besides this, coating, surface sizing, calendering, and coating might be performed for the board. The examined board machine has a size press unit, two calenders, four coating units, a winder, and a cutter.

Valmet is especially interested in improving paper and board machine operations that meet their customers' needs because 40 % of the paper and board in the world is produced by Valmet paper machines (Valmet 2019a). The basic industrial internet infrastructure is existing and enables the collection and analysis of customer process data. Via industrial internet, process information concerning intelligent pulp, paper, and energy production machines are available for condition monitoring and predictive maintenance. (Martinsuo and Kärri 2017)

2.2.3 Process data

In the pulp and paper industry, it has been invested in data acquisition and storing for several years. Reduction in sensor and analyzer prices, as well as a growth in database storage sizes, have enabled the collection of large amounts of real-time data. Data that is collected from the process include, for example, pressure, temperature, and flow measurements, quality measurements, and vibration measurements from the maintenance management systems. This data is used for tracking, troubleshooting, and optimizing the process. However, even a large amount of data is not enough as itself to provide support for decision making or application development. What matters is the ability to create value for the customer. Companies that can transfer the data to applications and services that support the customer's operations will succeed in the industrial market. Refining the data to a useful form requires strong data analysis skills. (Janssen, Laflamme-Mayer and P.R. Stuart 2003; Martinsuo and Kärri 2017; Varsinais-Suomen Yrittäjä 2015)

Most of the process industry data is collected in the form of time series. In equipment and process data level, the sampling periods are quite regular. Figure 2.8 represents a typical hierarchical data structure. On the left side in figure 2.8, there are equipment sensor measurements such as pressure, temperature, and vibration. The sampling time is fast, and it can happen in milliseconds. The next level is regularly sampled process control data, which is located in the middle of the figure. This level includes control feedback and control limits. It uses the equipment data to determine control errors and other necessary control indices in order to perform control actions. The sampling time happens in seconds. In this work, operation data have the longest sampling period. This operation data includes irregularly sampled quality metrics and operational level KPIs. Joe Qin has introduced one more hierarchy level in his article (S. Joe Qin 2001), which is the customer feedback data. Its collection can vary from customer service channels to social network complaints and it is not therefore available in the automation system. This thesis considers only the first three levels of process data hierarchy.

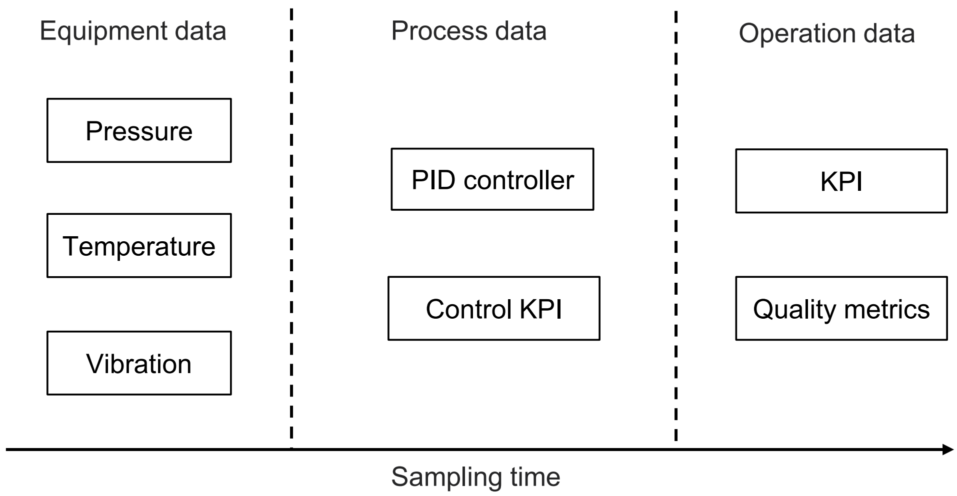


Figure 2.8. *Process data hierarchy. Sampling time grows from left to right. Adopted from (S. Joe Qin 2001).*

The operators of the plant are continually following equipment, process, and operation data. Based on the equipment and process data, operative decisions are made at the plant. The decision situations are related to e.g., sudden failure, disturbance, or process deviation of an individual device. In the operative decision-making process, the most valuable information is parameters such as control and vibration indices that describe the individual actuator's or machine's condition of the moment. (Martinsuo and Kärri 2017)

On tactical level decision making, the production system is observed as one complex. This angle of view looks at the process during a longer period of time when it is not essential to stare minute level data, which only shows the condition of an individual machine. Instead of that, it is vital to recognize bottlenecks and find the improvement potential of the production system. Operation data can be used to support tactical level decision-making.

Raw process data in different databases is not enough at itself to be used in the decision-making process (Martinsuo and Kärri 2017, p. 25). Further calculations are made from the raw data to describe essential process variables that cannot be measured. A conventional way is to generate KPIs, reports, and trend lines from the collected raw data. In recent years besides the traditional analysis, advanced analytical methods have shown their potential. This discipline is entirely new because most of the advanced data analysis methods require large amounts of raw data, which collection and storing have been possible only for a few years. Along with remote systems and cloud-based data warehouses, it has also become easier to access the data and analyze it from a distance.

2.3 Vibration monitoring key performance indicators

Key performance indicators are measured variables that describe progress against the desired results (*What is a KPI* 2019). Different kinds of KPIs are widely used in various business fields. In process industry, KPIs may also differ depending on the level of the examination. Typically operators are interested in process specific KPIs that support the operative decision making. As for a higher management level, decision making is more tactical and process specific values will not give as much value to them. The higher management level benefits from KPIs that show the direction of the production or the development potential of production. A good KPI is meaningful for the business and it inspires its interpreter for further actions.

Maintenance management KPIs can be applied to individual equipment, sub-processes, and whole plants. They can be used to describe the performance of individual equipment or the economical losses/wins of a process. Most importantly, KPIs give information, how well the maintenance targets are achieved (PSK Standard Association 2011). Maintenance management KPIs are considered as process or operational level data depending on the calculation period. This chapter introduces condition monitoring KPIs that are universally used to supervise vibrating machinery parts.

Vibration of a machine can be measured as displacement, velocity, acceleration, or derivatives of acceleration (PSK Standard Association 2015). In most cases, the frequency of a monitored object determines the used method. Frequencies between 10-1000 Hz are typically measured with velocity transducers. For higher frequencies, over 1000 Hz, an acceleration measurement is recommended. Accelerometers are the most popular sensor type for condition monitoring because they can be used to measure absolute frequencies over a wide measurement range. (PSK Standard Association 2018)

Low-frequency vibrations below 10 Hz are typically measured as displacement. Displacement transducers are preferred in journal bearing equipped rotating machinery measurements for detecting shaft vibrations and for monitoring the shaft position inside the bearing. However, displacement transducers have a limited value in the study of mechanical vibrations. (Goyal and Pabla 2015)

Besides trending overall vibration levels, an efficient tool for machinery condition monitoring is to calculate indices that characterize the total vibration signals. Typical indices that describe the vibration signal are: RMS (*root mean square*), PP (*peak to peak*), Crest factor and Kurtosis. (Väänänen 1990)

When vibration indices are determined, it is essential to define a correct sampling frame size. One frame consists of N pieces of uniformly pitched data samples that have been acquired during the period $[0, T]$. Sampling size is especially important in time to frequency domain transformation. FFT requires the number of samples to be a power of two. Further, statistical indicators that describe the intensity of the vibration and the shape of the signal are calculated from the sample block. (Väänänen 1990)

One of the most used variable in machine vibration evaluations is RMS (*root mean square*) (Väänänen 1990). RMS is proportional to the energy content of vibration, and it is used to evaluate the overall condition of the components. It can be calculated as an area under a half-wave, which in pure sinusoidal case corresponds to the peak amplitude divided by $\sqrt{2}$ (*The Difference Between RMS, Peak and Peak to Peak Amplitudes* 2019). For velocity vibration signal the vRMS value is calculated as follow:

$$v_{RMS} = \sqrt{\left[\frac{1}{T}\int_0^T v^2(t)\,dt\right]}, \tag{2.1}$$

where T is the integration time (frame size) and v is the value of the velocity signal. For the acceleration sensor, aRMS value can be calculated similarly to vRMS. In condition monitoring, it is also typical to calculate envelope RMS. Envelope signal is firsts generated by filtering and demodulating velocity signal followed by RMS calculation. Envelope RMS refers to the energy of the demodulated signal which may reveal faulty lubrication or incipient bearing faults (Mikkonen et al. 2009, p. 252).

Other common indices that are used for vibration analysis are peak value, peak to peak value and *Crest factor* (Vecer, Kreidl and Smid 2005, p. 37). Peak value is the maximum value of a signal within a frame and peak to peak value is simply the distance between the maximum positive and the maximum negative peak value. Within a certain sampling time, peak to peak value x_{pp} is defined as follow:

$$x_{pp} = max(x(t)) - min(x(t)), 0 <= t <= T, \tag{2.2}$$

where $x(t)$ is the signal and t is time within the selected time period T.

Crest factor is defined as the peak value divided by the *RMS* value of a signal. According to Vecer, Kreidl and Smid (2005), *Crest factor* indicates the damage in the early stage. The index reveals especially tiny surface damages in gears. If only one tooth is damaged, *RMS* value will not change but the peak value will increase. This will also increase the numerical value of *Crest factor*.

In rolling element bearing monitoring, *Kurtosis* index represent the excess of a signal. Damaged components cause spikes to time domain vibration signal which manifests as a sharper distribution function. The equation 2.3 can be used for calculating *Kurtosis* variable.

$$Kurt = \frac{1}{T}\int_{0}^{T} \frac{(x(t) - \overline{x})^4}{(x_{RMS})^4}\, dt, \tag{2.3}$$

where $\overline{x}$ is the average of the signal. Peaked amplitude distributions with relatively sharp spikes have a *Kurtosis* value higher than 3. *Kurtosis* value close to 3 reminds Gaussian-like signal and *Kurtosis* value less than 3 refers to flat distribution.

In summary, performance indicators are used to represent and predict the condition of machinery. Too little maintenance causes unplanned stops and production losses when damaged machinery parts crack up. On the other hand, too often performed maintenance work becomes expensive for the company. Vibration KPIs give information about the machine's condition enabling the planning of optimal maintenance breaks timing. Fundamental goals of condition monitoring KPIs are to extend machinery lifetime, lower maintenance costs, reduce unplanned shutdowns, and improve safety in operation. Condition monitoring KPIs are used to identify poor machine performance and in an ideal case, KPIs reveal the improvement potential of the process. (Juuso and Lahdelma 2013)

2.4 Valmet Machine Monitoring application

Valmet's Machine Monitoring (MM) application is used for condition monitoring in process industry. The application supervises the state of mechanical equipment and observes running stability. It is principally used for paper, board and tissue machines. (Valmet Automation Inc. 2019)

Machine Monitoring is an application that enables the monitoring of generic production KPIs online. Monitoring is based on vibration measurements and calculated vibration indices. It gathers process data from the customer's automation systems and provides information on production, economy, efficiency, energy, water, and quality. The application includes characteristic limit values for the indices, which exceeding generate warnings and alarms.

For a user, MM offers analysis tools for investigating vibration signals. History trends provide information about vibration values and calculated characteristics. With MM it is possible to examine fault mechanisms and severity of the mechanical faults. However, its usage requires professional skills and knowledge. Alarms may tell which vibration values are off-limit, but for the fault diagnosis, experience and knowledge are necessities. (Valmet Automation Inc. 2019)

This work examines whether predictive predictive maintenance could be facilitated by combining MM data to other data sources. The original idea was to bring these KPIs from different systems together on the same dashboard and discover whether a change in one KPI can be predicted from other variables. Tämän tiedon ydistämiseksi tarvitaan kuitenkin valistunut arvaus KPI kombinaatiosta, joka raportilla tulisi näyttää.

It would be a great advantage to have an analysis tool that finds abnormal vibration factors automatically. That information could be further used to assist the work of condition monitoring specialists in detecting failure mechanisms.

3 CONTROL LOOP PERFORMANCE IN PULP AND PAPER INDUSTRY

At the level of basic automation, process automation comprises the management of various unit operations. At the level of factory automation, process automation encompasses the management of entire production. The demands for the existing process automation systems have been expanded to include also process equipment condition monitoring and product quality control. The aim of this work is to examine if the existing process automation data could be utilized in condition monitoring and through that in predictive maintenance. The development trend of process automation, in turn, seems to contribute to the entire raw material supply chain and correspondingly the supply chain of products from the factory to the customer. (Oulun yliopisto n.d.)

The process automation system continuously receives measurement data from the process via measuring devices connected to it. Measuring instruments consist of sensors and transmitters. The transmitter converts the measuring signal from a sensor into a more transferable and easily accessible format for the other equipment. Actuators receive these control signals and affect the process in the desired way. An actuator consists of a motor element and a controlling element. (Jelali 2013; Oulun yliopisto n.d.)

By effective process automation, it is possible to reduce emissions and the primary consumption of energy and raw materials (Mary and Raynaud 2016). The modern industry could not either adequately meet the environmental standards without process automation. The development of process automation is often the first and the most cost-effective way to reduce the plant's environmental load because the changes to process equipment or techno-environmental structures require significantly higher financial investments than changes to automation.

3.1 Feedback control

A Control system is an interconnection of components influencing the behavior of the system to achieve desired system response (Dorf and Bishop 2010; Jelali 2013). The functioning of a control system is based on a cause-effect relationship where the cause is an input signal for the controller. As an effect, the controller sends a command to an actuator to make a change in a process. Depending on if the actual output signal from the process is compared to the desired output value, the control loop systems are categorized into two groups: open-loop and closed-loop control systems. Figure 3.1 presents an open-loop control system which is independent of the output.

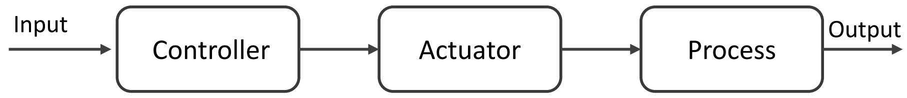


Figure 3.1. *An open-loop control system. Adopted from (Dorf and Bishop 2010, p.2).*

An open control loop consists of a controller, an actuator, and a process. In a control loop, after the controller, there is an actuator that produces motion (Reac n.d.). It utilizes the control signal from the controller to, for example, move a valve in a process to change the flow in a pipe. An open-loop does not use feedback to determine if its output achieved the set point or not.

Typically the output signal differs from the desired reference value. In a closed-loop control system, the output value (feedback signal) is measured and sent back to the controller (Dorf and Bishop 2010, p. 3). The closed loop control system is also known as a feedback control system. It is the most central principle of process automation. A Closed loop control system is presented in figure 3.2.

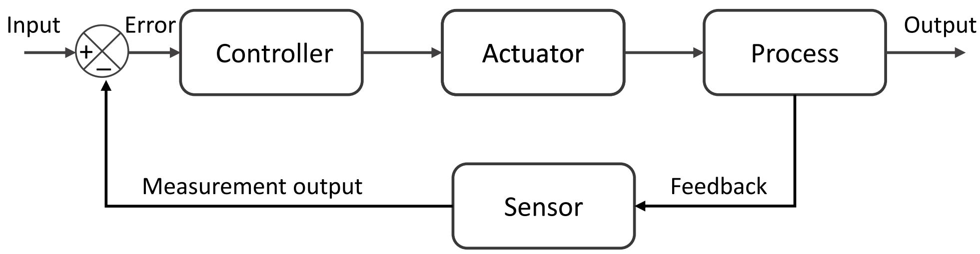


Figure 3.2. *A closed-loop control system. Adopted from (Dorf and Bishop 2010, p. 3).*

The operating principle of feedback control is to continually measure the value of the controlled variable and transfer the measured value to a computer which compares it to the desired setpoint value. The difference between the setpoint (spa) and the measured output value (me) is, in general, equal to the error. Based on the error, the computer generates a control variable for the actuator. Depending on the controller's type, different calculations are performed in order to achieve new control output (con) closer to the setpoint value. The function of a controller is simply to keep the measured value as close to the setpoint value as possible in all situations. After the calculations, the controller adjusts the actuator to modulate the process. A widely used controller type in process industry is a Proportional-Integral-Derivative (PID) controller. (Dorf and Bishop 2010)

3.2 Control loop performance monitoring

The performance of control loops directly affects the quality of the product, profitability, and operability of the plant (Knierim-Dierz, Hanel and Lehner 2012). In pulp and paper plants, there are typically hundreds to thousands of control loops that cannot be analyzed at the same time by a single person. With an automatic and continuous monitoring system, poorly performing control loops can be identified. A monitoring system enables to proceed corrective actions when poor performance is detected in order to maintain optimum plant performance.

One basic function of control loop monitoring (CLM) is to collect measurement data from the controllers. Similarly to all data-analytic methods, the accuracy of the results can only be as good as the quality of the data. In CLM systems, the data is collected in high frequency; a typical collection cycle is one second.

Control loops may perform poorly for two reasons. The disturbance may originate outside or inside the control loop. The key challenge in the control loop assessment is to distinguish these two situations from each other. Problem diagnosis for disturbances inside the loop can be, for example, a controller output running into its limits due to sluggish tuning. If an internal fault causes the malfunction, the controller should be re-tuned to achieve improved loop performance. An example of an outside disturbance is oscillation caused by valve friction. External malfunctions are handled within the framework of plant inspection and maintenance. In the process industry, friction in control valves is the most common external factor of poor control performance. (Bauer et al. 2019; Jelali 2013)

To detect poorly performing control loops, they need to be monitored either online or offline (Knierim-Dierz, Hanel and Lehner 2012). Online monitoring analyzes the data in real-time and offline monitoring performs the evaluation after the data collection. Online monitoring differs from offline monitoring in a way that online CLM evaluation is not synchronized. The relationships between the different control loop events are hard to identify since the evaluations are performed one after another. In offline monitoring, the loops are evaluated synchronously, and therefore it is more suitable for detecting relationships between different events. However, one disadvantage of offline monitoring is that it does not provide real-time events such as alarms.

Fault detection is used to distinguish control loop malfunctions. Alarm limits are specified for each control loop to generate alarms when a process value falls outside its limits. These limits should be set carefully because too many false alerts reduce the reliability of the CLM system. (Jelali 2013)

3.3 Control loop key performance indicators

In control loop monitoring, several different indicators are used to describe the state of a control circuit (Knierim-Dierz, Hanel and Lehner 2012). They illustrate the performance of a control loop from different points of view, providing clear information for the operator. An unambiguous index is an objective way to examine how well the controllers are performing and what is possibly causing poor performance. Performance indices expedite the fault detection enabling to perform corrective actions with less delay.

Besides the ability to describe the performance of the moment, CLM indices are a useful tool when the control loops are being tuned during the commissioning (Jämsä-Jounela et al. 2003). Performance indices can be utilized as reference values, which makes the tuning process more efficient and uniform from the start. During the operation, poorly performing control loops are detected by comparing dimensionless indices to the corresponding reference values during the normal operation. (Lahti 2018)

By key performance index monitoring, quality and product losses can be reduced. At the organizational level, it is though more useful to aggregate individual loops together and represent an overview of the control. Some KPIs that represent the plant's state of control are listed below:

- The automation utility index,
- The number of oscillating control loops,
- The saturation level of control,
- The settling time after a set-point change.

The automation utility index represents the number of control loops in manual mode. It often refers to an underlying problem where the controller is not acting in the desired way. For some reason, the operator does not trust the controller and it is switched from automatic to manual mode. The saturation level is used to identify insufficient actuator sizing, poor tuning, and wind-ups of PID controllers. If the controller output is constantly at its maximum or minimum, it often refers to saturation. (Jämsä-Jounela et al. 2003)

There are several causes for oscillation, and it not always easy to prevent them. In this work, the case 1 concerns control data from a board mill, which is very similar to a paper mill. In paper mills, it is indicated that about 30 percent of the control loop oscillation is caused by valve problems (Hägglund 1995). Other examples that cause oscillation are bad tuning, oscillating load disturbances, and oscillation induced by neighbor control loops. Oscillation may increase energy consumption, waste of raw materials, and quality deviation. (Bauer et al. 2019)

Settling time represents the time that it takes for the controller output value to reach the desired value after a set-point change (Lahti 2018). The faster the controller can react to the process changes, the smaller the settling time is. Settling time, control error, and oscillation are good single dimension measurements of control performance. However, they only detect a small portion of poorly performing control circuits (Friman et al. 2004).

3.3.1 Common indicators

In this chapter, basic calculation principles are presented for four control loop key performance indicators. The data, collected in the experimental part of the work, includes these four variables. In Valmet's application, the total performance index is further calculated from these four indices. Besides these indices, there are many other ways to determine control loop performance. For example, Jämsä-Jounela et al. (2003) have provided information about common control indicators, and Jelali (2012) has presented more control performance evaluation methods. Control loop key performance index calculations utilize the equipment and process-level data. The following four indices are evaluated:

- Integral Absolute Error IAE
- Variability Index VI
- Control Travel Index CTI
- Oscillation Index OI

A prior art method for control loop performance monitoring is to measure the absolute magnitude of the error (Dorf and Bishop 2010). Integrated Absolute Error (IAE) reflects the difference between the setpoint value and the process measurement. When the process measurement equals the setpoint value, the error is zero. IAE index is calculated with the following equation 3.1.

$$IAE = \int_0^T |e(t)|\, dt, \tag{3.1}$$

where $e(t)$ is the error as a function of time and T is the total sampling time.

Variability Index (VI) illustrates the variance between maximum and minimum error $e(k)$ measured during the sampling time period (Friman et al. 2004). The simplest way to calculate VI is by using equation 3.2.

$$VI = max(e(k)) - min(e(k)), k = 1....N, \tag{3.2}$$

where N is the number of samples collected during the sampling period. For example, if the sample interval is one second and the sampling time is five minutes $N = 5 * 60 = 300$. (Friman et al. 2004)

Control Travel Index (CTI) represents the amount of work made by a controller. The value of CTI grows in portion to the change of a control signal. A growing CTI value may correspond to oscillating control loop or an increasing amount of process disturbances.

CTI index for the period which sampling size is N is calculated as follow:

$$CTI = \frac{1}{N}\sum_{k=1}^{N} \left| con(k) - con(k-1) \right|, k = 1...N, \tag{3.3}$$

where con is the controller output value. PID controller function block generates its control output (con) value based on the setpoint and the measured process value.

Oscillation Index (OI) indicates the control circuit oscillation (Jämsä-Jounela et al. 2003). OI calculation is based on the predefined limit value IAE_{lim}. If the IAE_i value exceeds the predefined limit IAE_{lim}, the exceeding will be saved, and the variable $DIST_i$ will be reset to zero. An exceeding of the limit refers to a load disturbance. A sequence of registered single disturbances forms an oscillating control circuit, which is calculated by using the forgetting factor γ as follow:

$$OSC_i = \gamma OSC_{i-1} + (1-\gamma)DIST_i, \tag{3.4}$$

where

$$DIST_i = \begin{cases} 1, & IAE_i >= IAE_{lim}, \\ 0, & IAE_i < IAE_{lim}. \end{cases} \tag{3.5}$$

Forgetting factor gamma is calculated by using the equation 3.6.

$$\gamma = 1 - \frac{1}{5\tau}, \tag{3.6}$$

where τ is an estimate of process time constant.

These all four indices IAE, VI, CTI, and OI describe the performance of a controller from a bit different point of view. In Valmet's application, these indices are further utilized to calculate Total Performance Index (ToPi).

3.3.2 Total performance index

In process industry, the number of control loops is typically very high. Tracking hundreds of different KPIs gets rather complicated for the person who is analyzing controllers' performance in the control room. Following numerous parameters and changing them towards the desired outcome without influencing the other loops, forms a quite challenging task. The total performance index represents the control loop's overall performance by combining several individual control indices together.

The total performance index is defined on a scale from 1 to 100. Number 1 reflects a bad performing control loop and number 100 indicates a good performing control loop. In Valmet's CLM application, four indices IAE, VI, CTI, and OI are used to determine the total performance index.

Scaling and fuzzy logic

According to Russel and Norvig (p. 550–551, 2010), "Fuzzy control is a methodology for constructing control systems in which the mapping between real-valued input and output parameters is represented by fuzzy rules". Before adding fuzzy logic, the control indices IAE, VI, CTI, and OI indices are being scaled by using the index specific nominal values. The nominal values are defined by hand, based on the trend data during the normal operating conditions. The scaling harmonizes the indices in a way that they can fit into the same fuzzy logic database. When evaluating the ToPi index, fuzzy membership functions presented in figure 3.3 are being used. Based on the four functions, the scaled indices get four truth values that describe the strength of the membership to the classes "high", "normal", "small" and "zero". (Hooda and Raich 2017)

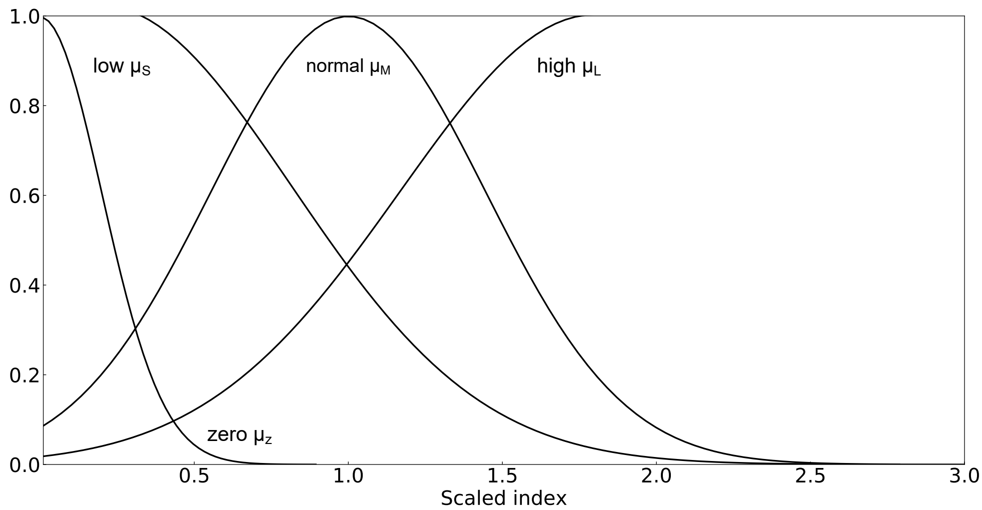


Figure 3.3. *Membership functions for scaled indices. Adopted from (Friman et al. 2004).*

By default, a scaled index gets a value 1. This means that the index gets a value that is exactly the same as the predefined nominal value. A scaled index which value is 1 belongs to the class "normal". Depending on how much the scaled value differs from its nominal, the degree of membership (truth value) changes. The degree of membership is defined as a numerical value varying between 0 and 1. For example a scaled index with a value 1.2 has a degree of membership 0.35 for class "low", 0.6 for class "high", 0.95 for class "normal" and 0 for class "zero". (Friman et al. 2004)

As a result of fuzzy logic, each normalized index gets four values μ_Z, μ_S, μ_M, and μ_L illustrating the degrees of membership to each class. Different membership combinations based on the knowledge and experiments are defined into the database. Some example combinations are listed in table 3.1. (Friman et al. 2004)

In table 3.1, each reference state has its own letter combination. The calculated and normalized indices (IAE, VI, CTI, and OI) are used to compute through a correlation with each example state. As a result, the calculation gives an interpretation concerning the current situation, for example, "load disturbance." Load disturbance obeys the conditions

VI	CTI	IAE	OI	Cotrol loop state
M	M	M	S	TRUE OK/ Normal situation
M	M	S	S	TRUE OK/ Control travel particularly small
-	-	S	L	TRUE Post-oscillation stage
S	L	S	S	TRUE Noisy measurement, control good
L	S	S	S	TRUE Disturbance peak, control good
S	M	S	S	TRUE OK/small control travel & variability
-	M	L	S	FALSE Load disturbance
L	L	S	-	FALSE Disturbance peak + measurement noise
-	Z	L	-	FALSE Control saturated
M	L	M	-	FALSE Measurement noise
-	-	L	L	FALSE Circuit oscillates
L	M	M	S	FALSE Disturbance peak

Table 3.1. *Fuzzy logic membership table for control loop performance indices. The letters correspond to membership functions L = High, M = Normal, S = low and Z = Zero. Adopted from Friman et al. (2004)*

where IAE index is high, CTI is normal, and OI is small.

Based on the data in index combination table 3.1, a numerical output value T_i is defined. T_i express how well the measured state of an index combination correlates with the example state given in the table. The truth value T_i for each row i in table 3.1 is calculated as follow (Friman et al. 2004; Russell and Norvig 2010):

$$T_i = \prod_c \mu_{(i,c)} I_c, \tag{3.7}$$

where $\mu_{(i,c)}$ is the degree of membership calculated for each performance index in column c. I_c is the scaled value of the index in column c. The higher the number T_i is, the better the measured state correlates to the example state.

Definition of total performance state

The total performance of an individual control loop is divided into five categories:

1. Performance "Good"
2. Performance "Bad"
3. Performance "Unrecognized"
4. Controller in manual mode
5. Controller locked

If the controller has been in auto mode and it is not locked, the best correlating example state (table 3.1) is selected to be the control loop state (Friman et al. 2004). A threshold value for T_i is 0.5. If the truth value is smaller than that, the correlation between the example state and the measured state is "unrecognized". When T_i gets a value that is higher than 0.5, the controller's state is categorized to either "Good" or "Bad" category. In the fuzzy logic membership table 3.1 all example states are defined with TRUE or FALSE status. TRUE value corresponds to "Good" performance and FALSE value corresponds to "Bad" performance.

The performance index represents the performance state distribution within a certain period of time, typically within one day, week or month. In figure 3.4, a typical control situation is illustrated where within one day the state of control performance has been good 75,49 % of the time, unrecognized 24,44 % of the time and the controller has been on manual mode 0,07 % of the time.

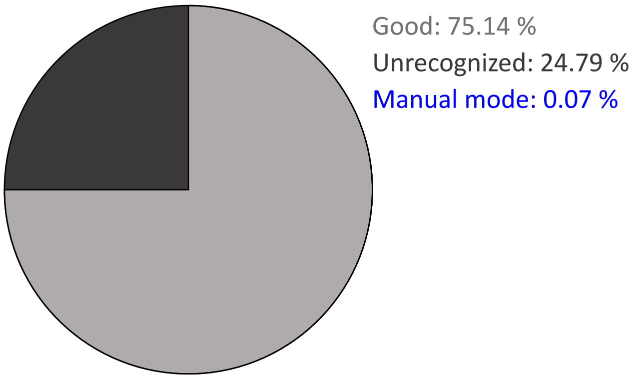


Figure 3.4. *The state of control loop performance as percentage of the time period.*

If the controller has been in manual or forced mode, the state is classified into class "controller in manual mode" or "controller locked". Performance state cannot be evaluated for controllers in manual mode.

ToPi Index

Performance index reflects the overall performance state within a certain period of time. However, it does not reveal when the possible failure has occurred. ToPi shows the loop performance as a function of time, whereas the failures are easier to detect.

To define ToPi, the best correlating "Good" performance state (G_{max}) and the best correlating "Bad" performance state (B_{max}) are selected from the fuzzy logic table. The suitability to states "Good" and "Bad" are scaled from 0 to 100. Number 100 means that the example state is perfectly suitable for the measured state and number 0 means that the state is not correlating at all. The value for ToPi can be calculated with the following equation 3.8.

$$ToPI = \frac{100 + G_{max} + B_{max}}{2}. \tag{3.8}$$

The behavior of total performance index is illustrated as a function of time in figure 3.5. In figure 3.5 total performance of the controller has decreased dramatically.

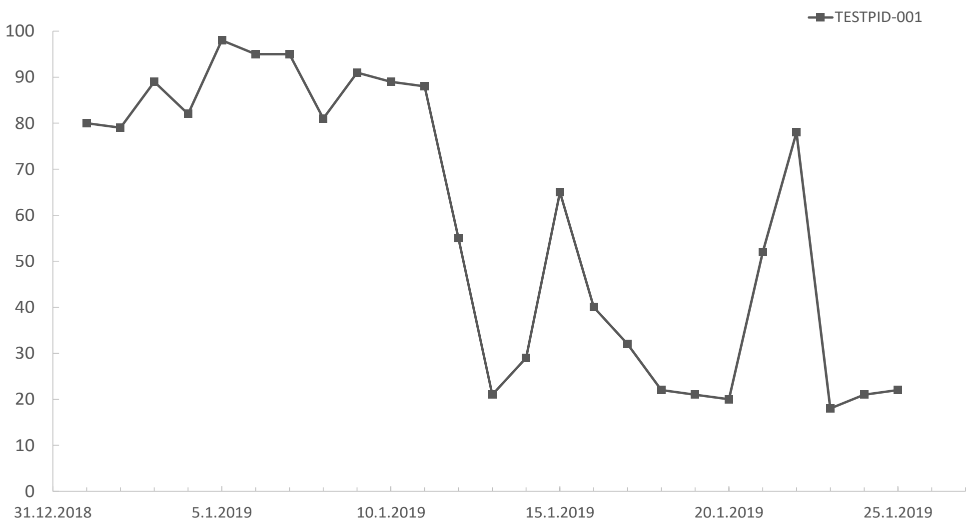


Figure 3.5. *Total performance index plotted as a function of time.*

ToPi has a longer sampling time than the lower level CLM indicators. It is used to describe the performance of an individual loop within a long operating time window. Some oscillation in performance may happen because of the changing operation conditions, but a longer time trend line shows the direction of the condition. It is also typical to aggregate individual ToPi values together within one process area and present them as treemaps. This operational level data helps to get a more holistic view of the process control. (Valmet 2019c)

3.4 Data structure of Valmet Control Performance application

Valmet has its own application for control loop monitoring. The product name for the application is Control Performance. This application was used in case 1, and therefore it is essential to know a little about the application's data structure. Data collection for the analysis part was done remotely from the local computer, so it was fundamental to know from which database the data should be collected.

The Control Performance application is built on top of the Valmet DNAsystem. It utilizes the DNAHistorian database, but the index calculations are performed in a separate loop monitoring software. This allows the integration of the control loop monitoring application with third-party automation systems through open platform communications (OPC) gateway. The loop monitoring system architecture is presented in figure 3.6.

Data from the supervised PID-controllers is collected to a distributed control system (DCS) that gathers continuous data from the process. High-frequency data collection provides measurements approximately once per second. The data is transferred to the

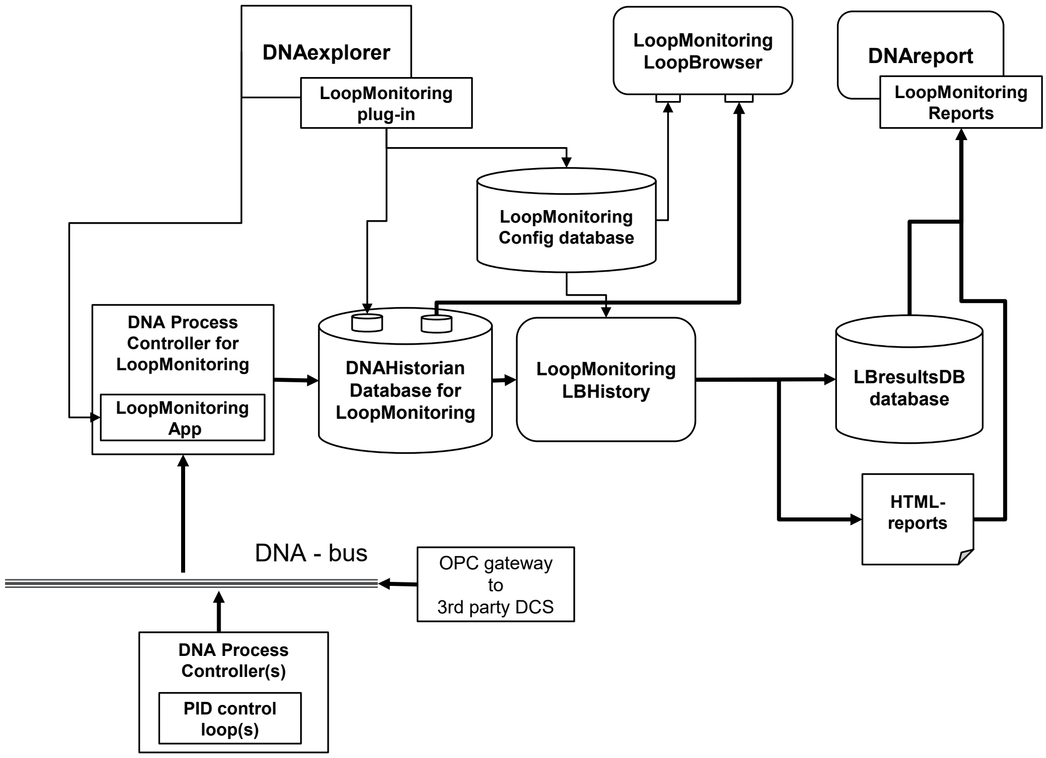


Figure 3.6. *Control loop monitoring system architecture.*

Loop Monitoring Application via DNA-bus where the index calculations are performed within desired cycles. PID parameters and calculated indices (CTI, OI, IAE, and CTI) are stored to DNAHistorian database. ToPi index is further calculated in a separate process control system named Loop Monitoring LBHistory. This Loop Monitoring LB History uses data from the DNAHistorian database to calculate the index.

Analyzed ToPi indices with control state status are stored in LBresultDB database. This database uses Microsoft Structured Query Language (SQL). SQL is a language that allows the user to access data by making queries. The results can be interpreted online via Loop Monitoring browser or via DNA report portal. Reports that summarize the performance of the controllers are generated in Loop Monitoring LB History.

3.5 Combining data in predictive maintenance

Among the growth of process data volume, the demand for bigger databases has increased. The prevailing direction of development is moving from local databases towards cloud-based data warehouses. A cloud-based storage is a scalable choice to store large amounts of data in a quite cheap prize (Nopanen 2017). Combining data from different sources and accessing it at any time have been difficult before the cloud-based servers.

Ideally, all the raw data from the plant would be sent to central cloud storage, which in Valmet's case is AWS (*Amazon Web Server*). When the data is on the same platform, advanced data analysis methods are possible to implement.

From the point of view of predictive maintenance, cloud-based causality analysis, and predictive modeling would give a great advantage. With enough process data, certain behavior patterns can be recognized and further modeled (Nopanen 2017). These models help to estimate how a process should behave in certain operating conditions. Cloud-based computing also offers effective data processing, which would quickly become too much for a single computer. (C.K.M. Lee, Ty Cao and Kam Hung. Ng 2017)

The person who operates the plant or a machine gets information from the process automation system. The data needs to be reliable and in clear format so that the operator can make the right decision based on the information. For the condition monitoring specialist, data aggregation from separate systems would help to form an overall picture and speed up the decision making process.

In examined cases, vibration monitoring, control loop monitoring, and equipment data, as well as their reports, were all in separate systems. When the data is scattered, it is relatively difficult and time-consuming to go through all possible scenarios. This work aims to examine if there is a link between vibration and control loop monitoring or a link between vibration monitoring and equipment data. If certain parameters correlate, does it give new valuable information for the user to present these indices together in the same report? It is also studied how certain KPI values link together by their behavior.

According to Vintilescu (2010): "Dynamic data acquisition has always been at the heart of every sound and vibration application. However, it is not enough to simply be able to acquire data, you also have to be able to analyze, process, and interpret the raw data into meaningful content". The challenge today is that there is a vast amount of process data available, but its efficient usage has not been fully exploited yet. In this work, before scaling up to cloud-based analysis, it is first examined locally on a small scale to see if the chosen data analysis methods could be exploited in predictive maintenance.

4 MATERIALS AND METHODS

The experimental part of this thesis consisted of two different cases. It was examined if combining data from different sources benefit predictive maintenance in pulp and paper industry. In case 1, it was examined if vibration monitoring data together with control loop monitoring data would bring new insights concerning the condition of the machinery. Both data sources (I and III) included only KPIs which were calculated from the raw data. In case 2, the examined data sources were vibration monitoring KPIs and raw equipment data (I and II). Figure 4.1 illustrates the nature of data sources.

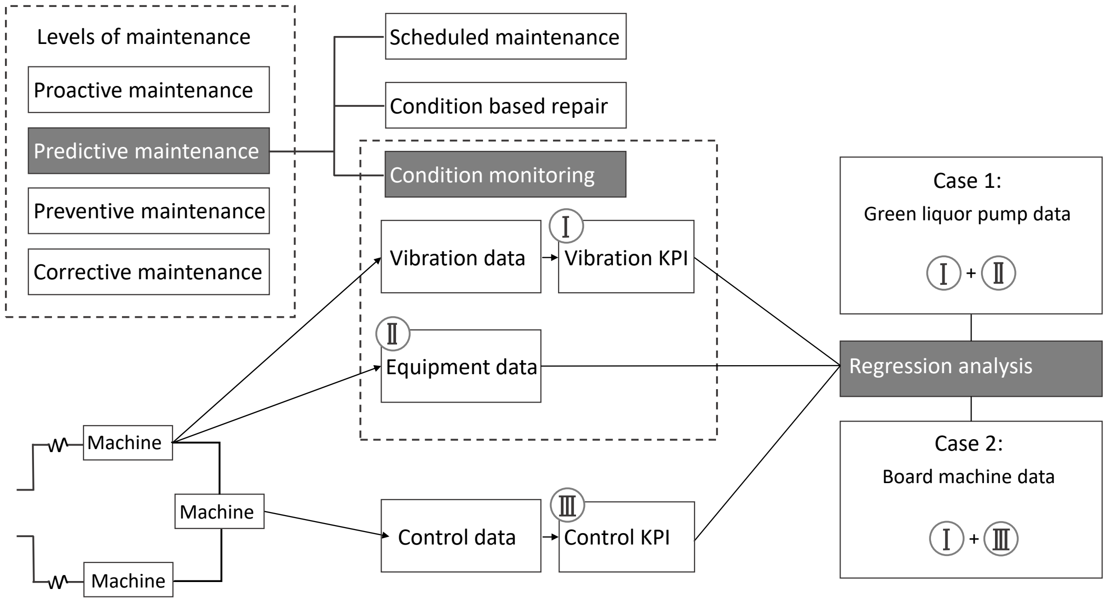


Figure 4.1. *Examined data sources from the perspective of preventive maintenance.*

Data in case 1 was collected from a board machine. The prediction was that some faults in control loops performance could be seen as a change in the machine's vibration monitoring indices. Conversely, it was examined whether excessive vibration could change the behaviour of control loops According to condition monitoring specialists, machine monitoring system is used to detect vibration frequencies that are much higher than the fault frequencies in the control loop systems. The smallest vibration frequencies that MM system can detect are in Hertz (1 per second) as for oscillating control loop fault frequencies are typically in minutes. Case 1 is represented in more detail in chapter 4.1.

Case 2 included vibration monitoring data (I) from a green liquor pump and the relevant equipment data (II) was collected from the surrounding process area. In case 2, there were an existing vibration problem that delimited the subject of the study and operating time window. Case 2 is represented in more detail in chapter 4.2.

In both cases 1 and 2 the examined data was stored in separate history databases. Vibration monitoring KPIs, control loop monitoring KPIs and equipment data had all their own physical servers on the plant. Prior to data analysis, the first step was to collect and transform the data into a same platform. Workflow of the conducted work is depicted in figure 4.2.

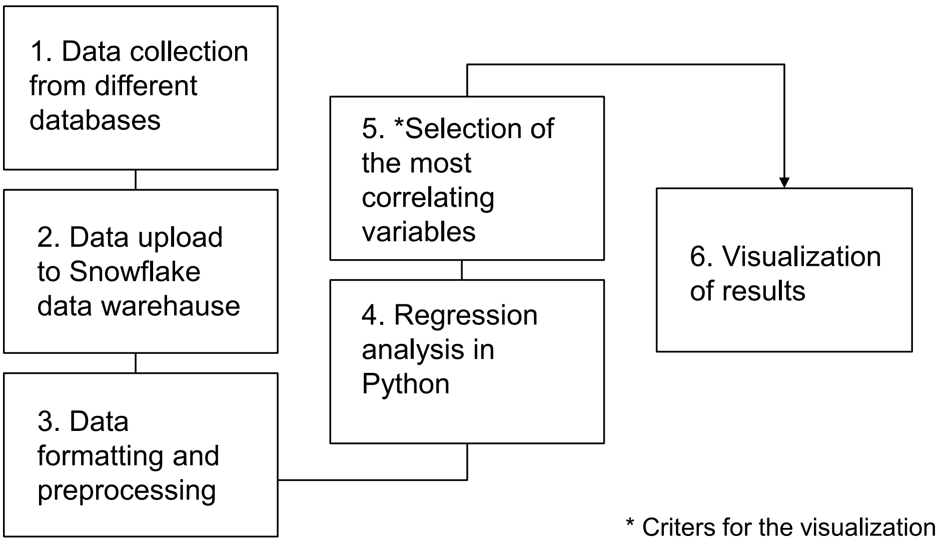


Figure 4.2. *Data handling procedure.*

Roughly, the experimental part of this work can be divided into three main phases. The first part consisted of data gathering and processing. The collected data was transferred to a web based cloud server from where it was easy to access. Data processing was the most time consuming part of the work in which data formatting had to be compromised. This phase is explained in more detail in chapter 4.3.

The second part consisted of data analysis. An open source cross-platform Spyder was used to examine the relationships between data sources by regression analysis. Spyder was chosen because of its ability to integrate numerous Python packages that can be used for data analysis. Data analysis and regression methods are described in chapter 4.4.

After the analysis part, relevant variables were selected for the third stage. In the third stage essential variables were chosen for the visualization. Results from the regression analysis were plotted in Python and further discussed in chapter 5. Visualization tool Tableau was used for future visualization ideas which are presented in chapter 6. Tableau is in general used in Valmet industrial internet applications.

4.1 Case 1: Board machine

Case 1 contained data from a Finnish board mill. In the mill, there were one board machine producing food service board and two kinds of coated folding boxboard. The final product consisted of three layers: back, middle and surface layers. To produce three layers of cardboard, three wire sections with three headboxes are required.

In the examined boar mill, online vibration measurements started from the wet end as seen in figure 4.3. Process areas with online Machine Monitoring (MM) system are marked with blue. In the middle of the wire section, the three layers are merged and the wet web travels along the wire to the pick-up roller. According to the condition monitoring specialist who serviced the machine, pick-up roller was one interesting machinery part to be examined. Faulty pick-up roller causes web breaks and in some situations when the root cause of the fault is not clear, the pick-up roller might be changed "just in case". In vain removed pick-up rollers cause extra costs because the felt need to be cut off and the roller has to be overhauled before reinstallation. Better fault analysis tools could reduce unnecessary roll changes.

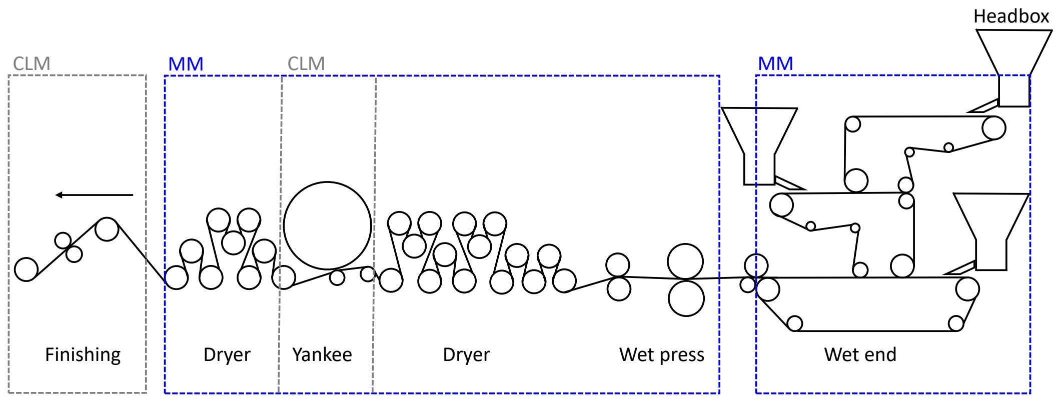


Figure 4.3. *A schematic picture of examined board machine. Control Loop Monitoring (CLM) and Machine Monitoring (MM) are used to monitor different parts of the machine.*

The pick-up roller picks the wet web to the press section. Press section was another interesting section to to be studied because it affects strongly to the paper characteristic. Board strength properties, stiffness, profile, and smoothness can be controlled at this phase of the process. A desired machine speed and efficiency must be attained with low vibrations to achieve good quality and high dryness.

After the press section, the felt carries the web to the dryer where water is further removed through evaporation. Most of the online vibration measurements were collected from the drying section. Between the drying sections, there is a Yankee cylinder that had some control loops among CLM system. Otherwise, the monitored control loops were located before the headboxes at stock and water systems and after the drying section where the board surface is finished. Machinery parts that had CLM system are marked with grey in figure 4.3. The exact locations of the PID controllers were not shared.

Vibration levels of the machine are monitored both online and offline. The online monitoring is primarily used for machinery parts that have difficult working conditions and cannot be measured manually. Vibrating parts in the forming section (wet end), press section and drying section were monitored with MM system. The data of these sections were available in the plant's history database and it contained vibration measurements of felts, rollers, cylinders and gearboxes. Besides this, condition monitoring experts perform manual measurements for the other vibrating parts of the machine. These measurements which are done by hand were not available in the history database. Vibration monitoring and control loop performance monitoring systems were mainly used for different machinery parts in the process.

4.2 Case 2: Green liquor pump

Case 2 contained data from a bioproduct plant. The examined pump was located in recovery part of the kraft process, where the green liquor was pumped from the dissolving tank to a balancing reservoir. The dissolving tank had two identical pipelines and as for comparison, both lines were examined. The location of the green liquor pump is illustrated in figure 4.4.

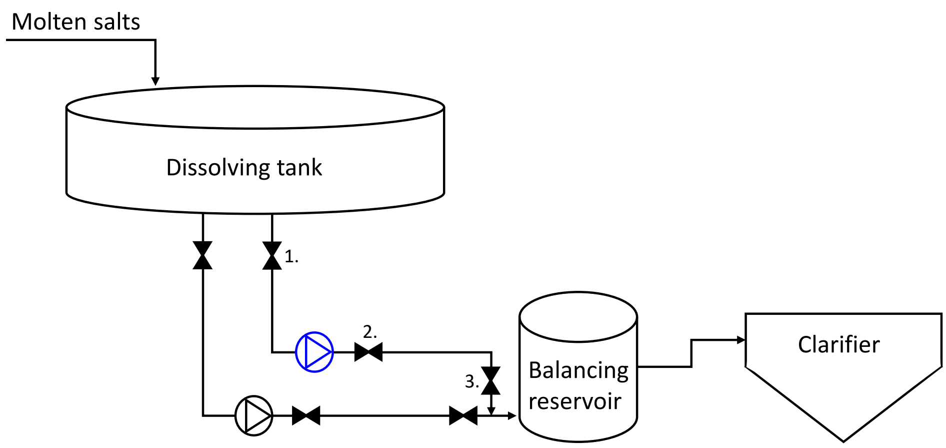


Figure 4.4. *Illustrative graph of the green liquor pump location. Examined pump is located between valves 1 and 2.*

In a certain situation, when the pump's speed of revolution is high, the vibration levels of the pump started to increase excessively. Excessive vibrations should be avoided due to harmful consequences. Vibration may, for example, lose the bolts which hold down the pump or damage the pump's components such as wear rings, bushings or impellers. The assumption was that this specific pump started to cavitate under certain operating conditions. Cavitation means that the pumped fluid begins to vaporize inside the impeller inlet and disturbs the pump's operation. Cavitation causes erosion and in the worst case, destroys the pump.

The aim was to examine through regression analysis which process variables contribute to the pump's vibration. Cavitation in a pump is mainly due to low suction pressure. This would happen for example in a situation where the level of the dissolving tank lowers too much. Another suggestive scenario was rotating stall which may occur when the pump is operated in low flow rate conditions (Ulrik et al. 2006). In stall phenomenon, a preswirl at the impeller inlet blocks a smooth flow into the impeller chamber causing a situation where the flow separates from the blades. This phenomenon induces vibration and hydraulic loss.

4.3 Data acquisition and processing

In both cases, measurement data from process automation was stored in a local history database. Besides this, in Valmet DNA system there are two separate applications to monitor machine vibrations (MM) and control loops (CLM). Both applications have their own DCS where the system specific KPIs are calculated from the measured signals. The first step was to access the equipment, CLM, and MM data in order to gather information from the desired time period. The collected data was saved as Comma-Separated Values (CSV) files to external storage and further loaded into Snowflake.

Snowflake is a cloud-based data warehouse that also offers analytical services. The advantage of Snowflake database was that the data could be access from any computer and there were no need to log in to a local server to preview the data. Some data formatting was already done in Snowflake, but its capabilities were limited. Further data analysis and effective processing were done locally by using Python.

In this thesis, the focus was to examine what kind of benefits it would bring to combine different data sources. The nature of the thesis was a proof of concept and therefore it was not relevant to build a live connection to the plants. In Valmet industrial internet solutions, a typical practice for cloud computing is to send process data in packed CSV files to the cloud server in regular cycles. This is necessary especially when the amount of data is large. In this work, the data was imported manually from the local server and transferred manually to the cloud.

The data was further accessed from the cloud by using a Snowflake connector. This was done by utilizing Mikko Ritala's complimentary Python code which had been originally developed for anomaly detection software. With Python, the data was formatted in a way that the timestamp values matched between different measurements and calculated indices. At this phase, compromises which presumably affected the results were done.

All the measurements and calculated signals had a unique tag name. It is relevant to clarify what "a tag name" means in this context. The tag name separates variables from each other and it is used to find certain values from the database. Index calculations use more than one signal producing different KPI values. In the case of MM, nine indices were calculated for one machinery part and all indices were saved with a different tag name. CLM application calculated four indices for each control loop and they as well had all a unique tag name identifier.

4.3.1 Case 1

Machine monitoring system om case 1 contained 628 vibration measurement points. For each measured data point, nine different indices were calculated. These indices were aRMS, acceleration peak, vRMS, vRMS 1–2 x RF (1–2 multiples of rotational frequency), vRMS 5–20 x RF (5–20 multiples of rotational frequency), vRMS HL x RF (higher multiples of rotational frequency), envelope RMS, envelope peak and rotational frequency. In addition to that also bias of the sensor was measured. Bias (offset error) represents the transducer's performance and it does not relate to the vibration levels. In this work, bias was only used at the validation part to verify transducer's normal operating capability. Bias values and data from poorly performing sensors were excluded from the regression analysis. All in all, there were 7617 different machine monitoring tags of which 6280 were included into the analysis.

The period of the collected MM data was ten months. Nine indices were calculated twice in an hour for each measurement point. Because of the large number of measurement points, the evaluation could not be performed synchronously. Nine indices were calculated for three measurement points at the same time. As a result, three measurement points multiplied by nine indices produced 21 vibration KPIs with the same timestamp value. Each KPI had its own tag name. The next 21 tags were evaluated after the first calculation was done and they got a slightly different timestamp. The calculation method is illustrated in figure 4.5. This led to a situation, where within an hour, there were numerous different timestamp values.

Control loop monitoring was performed for 614 control loops. For each loop, four non-scaled indices OI, CTI, IAE, and VI were calculated from raw data in every minute. In total 2456 tags with 60 values in one hour were stored in local DNAHistorian database. The data was collected from DNAHistorian database and therefore it did not include ToPi values. To reduce the size of the data set, CLM index values were collected at five-minute intervals and saved to an external data source. At this phase, 80 percent of the original data was already left outside the analysis. Similar to MM data, the time period of data collection was ten months. Only the time period that was common for both data sets was examined.

A common factor was needed to combine CLM and MM data, which in this case was the time stamp value. Within the examined data set, MM indices had two actual values in

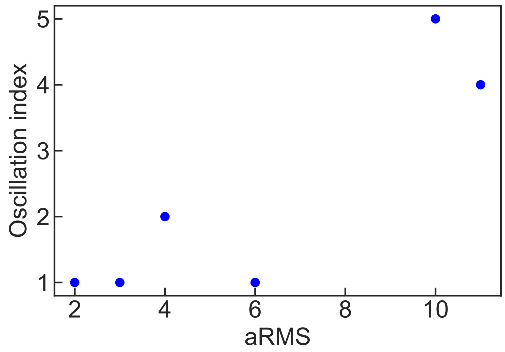
TIME STAMP	TAG1	TAG2	TAG3	TAG4	TAG5	TAG6	TAG207	TAG208	TAG209	TAG210
1.1.2019 4:20:14	0.092135	nan	nan	nan	nan	nan	0.058807	nan	nan	nan
1.1.2019 4:20:15	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan
1.1.2019 4:20:16	nan	0.044477	nan	nan	nan	nan	nan	0.03601	nan	nan
1.1.2019 4:20:17	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan
1.1.2019 4:20:18	nan	nan	0.056245	nan	nan	nan	nan	nan	0.024103	nan
1.1.2019 4:20:19	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan
1.1.2019 4:20:20	nan	nan	nan	0.080033	nan	nan	nan	nan	nan	0.037979
1.1.2019 4:20:21	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan
1.1.2019 4:20:22	nan	nan	nan	nan	0.037979	nan	nan	nan	nan	nan
1.1.2019 4:20:23	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan
1.1.2019 4:20:24	nan	nan	nan	nan	nan	0.075088	nan	nan	nan	nan
1.1.2019 4:20:25	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan
1.1.2019 4:20:26	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan
1.1.2019 4:20:27	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan

Figure 4.5. *Original data format.*

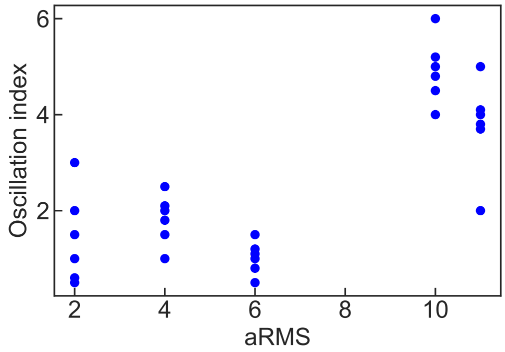
one hour whereas CLM indices had 12 actual values within one hour. Data sets were combined into the same Pandas DataFrame in Python where missing time stamp values were handled by propagating the last valid observation forward until the next valid observation. This method is called "fillna". The data table which "nan" values are filled with the last valid observation is illustrated in figure 4.6. This method was chosen because it uses the last valid information and it won't iterate any values.

TIME STAMP	TAG1	TAG2	TAG3	TAG4	TAG5	TAG6	TAG207	TAG208	TAG209	TAG210
1.1.2019 4:20:14	0.092135	nan	nan	nan	nan	nan	0.058807	nan	nan	nan
1.1.2019 4:20:15	0.092135	nan	nan	nan	nan	nan	0.058807	nan	nan	nan
1.1.2019 4:20:16	0.092135	0.044477	nan	nan	nan	nan	0.058807	0.03601	nan	nan
1.1.2019 4:20:17	0.092135	0.044477	nan	nan	nan	nan	0.058807	0.03601	nan	nan
1.1.2019 4:20:18	0.092135	0.044477	0.056245	nan	nan	nan	0.058807	0.03601	0.024103	nan
1.1.2019 4:20:19	0.092135	0.044477	0.056245	nan	nan	nan	0.058807	0.03601	0.024103	nan
1.1.2019 4:20:20	0.092135	0.044477	0.056245	0.080033	nan	nan	0.058807	0.03601	0.024103	0.037979
1.1.2019 4:20:21	0.092135	0.044477	0.056245	0.080033	nan	nan	0.058807	0.03601	0.024103	0.037979
1.1.2019 4:20:22	0.092135	0.044477	0.056245	0.080033	0.037979	nan	0.058807	0.03601	0.024103	0.037979
1.1.2019 4:20:23	0.092135	0.044477	0.056245	0.080033	0.037979	nan	0.058807	0.03601	0.024103	0.037979
1.1.2019 4:20:24	0.092135	0.044477	0.056245	0.080033	0.037979	0.075088	0.058807	0.03601	0.024103	0.037979
1.1.2019 4:20:25	0.092135	0.044477	0.056245	0.080033	0.037979	0.075088	0.058807	0.03601	0.024103	0.037979
1.1.2019 4:20:26	0.092135	0.044477	0.056245	0.080033	0.037979	0.075088	0.058807	0.03601	0.024103	0.037979
1.1.2019 4:20:27	0.092135	0.044477	0.056245	0.080033	0.037979	0.075088	0.058807	0.03601	0.024103	0.037979

Figure 4.6. *Data format after "nan" values are filled with last valid observation.*



(a) A hypothetical situation where one CLM index value corresponds to one MM index value.



(b) In this work six CLM index values corresponded to one MM index value.

Figure 4.7. *The effect of the used method to fill missing data points.*

From the perspective of the analysis, replacement of the nan values increased the total amount of data points. This phenomenon is illustrated in figure 4.7 where one MM index is drawn as a function of one CLM index. In an ideal case, one CLM data point would correspond to one MM data point as seen in figure (a). In this situation, both indices would have had the same timestamp value. Diagram (b) illustrates the real situation, where six CLM index values corresponded to one MM index value. The influence of this method to the regression analysis is represented in chapter 4.4.

Changes in CLM indices happen faster than changes in MM indices. Controllers need to react in seconds to the process changes, which make the changes in the CLM indices also fast. Therefore, taking only one CLM index every half an hour, as in diagram (a) in figure 4.7, and comparing it to one MM value would have discarded too much of the information. Vibration issues typically develop within days or weeks so instead of tracking fast changes, a longer period trend line shows the direction of vibration levels. Because changes in vibration levels are quite slow, it was reasonable to use the last valid value for the analysis, as in the diagram (b) in figure 4.7. The most notable changes in vibration levels within a short period of time were due to the changes in the machine's running speed. The speed of the board machine had an effect on overall vibration levels along with the rotating speeds of the shafts.

The quality of the data is essential from the point of view of the regression analysis. The analysis can only be as good as the input data. The different timestamp values and different amounts of data points affected the reliability of regression analysis. In this case, delays were not taken into account, which probably affected to the results. The length of the board machine is several dozens of meters so if a controller at the end of the process would have reacted to a process change in the beginning of the process, there would have been a delay. The running speed of the machine was not known so the exact delays could not be determined.

4.3.2 Case 2

Case 2 examined the reasons behind the green liquor pump's anomalous vibrations. The kraft process had two similar pump lines from the dissolving tank to the balancing reservoir so that for comparison, the examined data was collected from both pump lines.

Green liquor pumps had four different vibration measurement points. Similar to the MM system in case 1, nine different vibration indices were calculated for each measurement point. In total there were 32 machine monitoring indices for each pump and 28 of them were examined. MM index vRMS HL x RF (higher multiples of rotational frequency) was not defined for the pumps and bias was not included in the analysis.

The vibration indices were calculated once an hour and the values were saved to a local database. The data was imported from the database to an external data source. In addition to that, minute level data was collected from the surrounding equipment. This data contained rotational speed of the pump, current of the pump's motor, pressure, temperature, flow, density, tank level measurements and information about valve positions. The equipment measurements and calculated vibration indices that were selected for the analysis are represented in table 4.1.

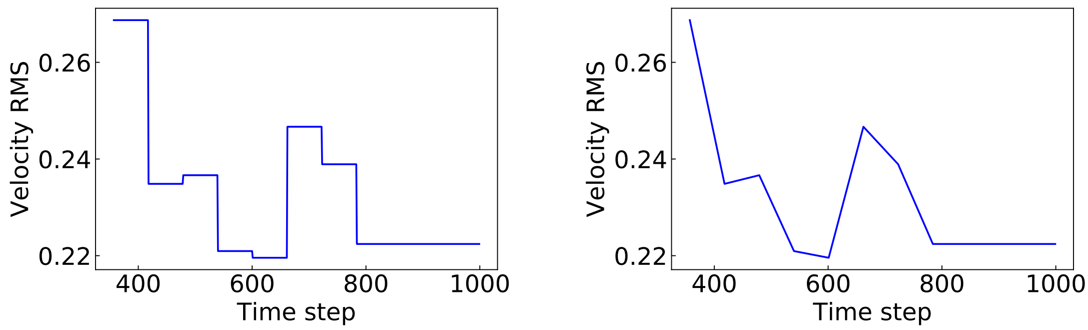
Vibration monitoring data	Equipment data
Velocity RMS	Pressure
Velocity RMS 1-2 x rotational frequency	Temperature
Velocity RMS 5-20 x rotational frequency	Flow
Rotational frequency	Tank level
Acceleration RMS	Valve positions
Acceleration peak	Density
Envelope RMS	Current of the pump's motor
Envelope peak	Rotational speed of the pump

Table 4.1. *Vibration monitoring indices and chosen equipment data points for the regression analysis.*

As seen in figure 4.4 there were three valves in each pipeline. The first valve was located before the pump, the second one was after the pump and the third one was right before the balancing reservoir. Pressure, temperature, flow, and density were all measured after the pump. Tank levels were expressed as a percentage of the maximum capacity.

Both equipment data and MM data were collected from four days time period. The increased vibration levels occurred within this chosen period. It was a remarkable advantage to have this short time period compared to the case 1. Data collection, processing and analyzing took much less time when the subject of the study was marked off. The amount of data was small enough so that there were no need to upload it to the cloud. It was transferred to Excel, formatted and converted to a CSV format. Further processing and analysis were performed by using Python.

In case 2, equipment data was collected as one minute cycles. Within one hour there were 60 values for each equipment measurement tag and one value for each MM tag. Similarly to case 1, to combine these two data sets they needed a common factor. Two methods were used and the first one was similar to the one used in case 1. Timestamp was used as a common factor and missing vibration index values were filled with the last valid observation. This way each index got the last valid value for 60 times in an hour. Another used method was to take an average MM value for each minute. This method reminds about a line, drawn from one data point to another. These both methods are illustrated in figure 4.8.



(a) Last valid observation is propagated forward to next valid backfill.

(b) The minute average is calculated based on the actual values.

Figure 4.8. *Two different methods are used to handle missing data points. The aim is to have a value for every minute.*

4.4 Regression analysis

Relationships between variables were examined by regression analysis. By regression analysis, it was possible to search for parameters that were related to each other. The goal was to search for clear correlations and examine, if these could be used to predict operational behavior patterns. In this work regression analysis was used to study:

- Case 1: The impact of vibration monitoring KPIs on control loop monitoring KPIs.
- Case 1: The impact of control loop monitoring KPIs on vibration monitoring KPIs.
- Case 2: The impact of process variables on vibration monitoring KPIs.

For the analysis part, two different regression methods were chosen. The first method was called Pearson correlation which is a widely used method to measure linear correlation between two variables. This method gives a Pearson correlation coefficient as a result from the analysis. Pearson correlation coefficient gets a numerical value between -1 to 1, where number -1 refers to a strong negative correlation, 0 means that there is no correlation at all and +1 refers to a strong positive correlation. Figure 4.9 represents different values of correlation coefficient.

In figure 4.9 (a), Pearson correlation coefficient r gets a value of 0.92. This refers to a strong positive correlation where one variable increases with the other. In figure 4.9

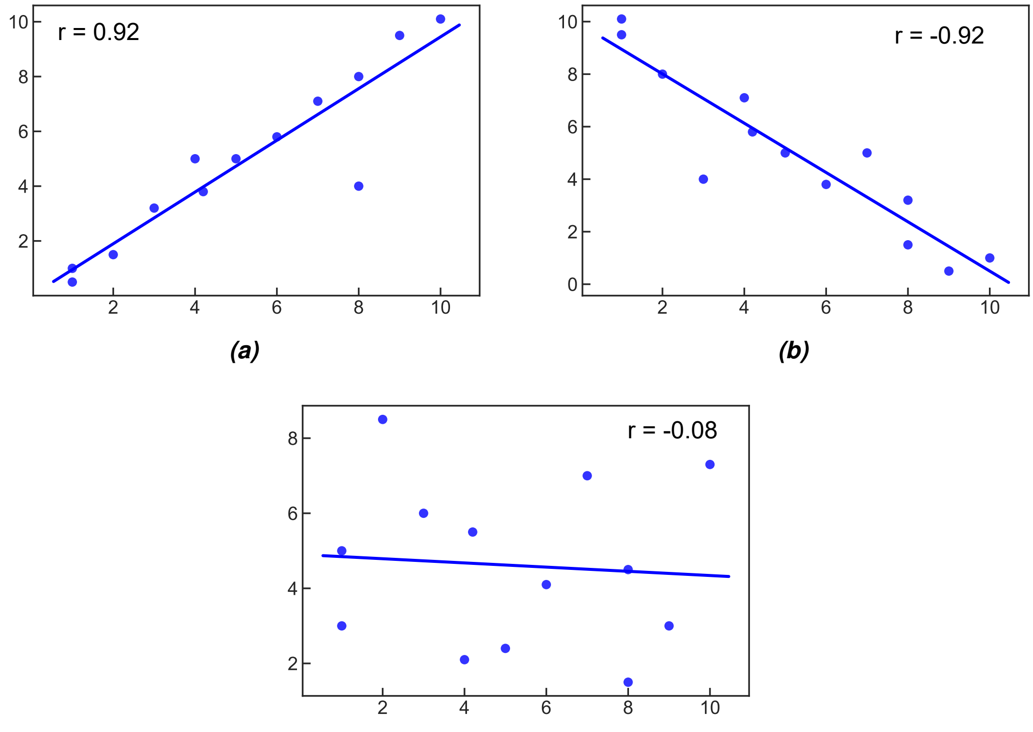


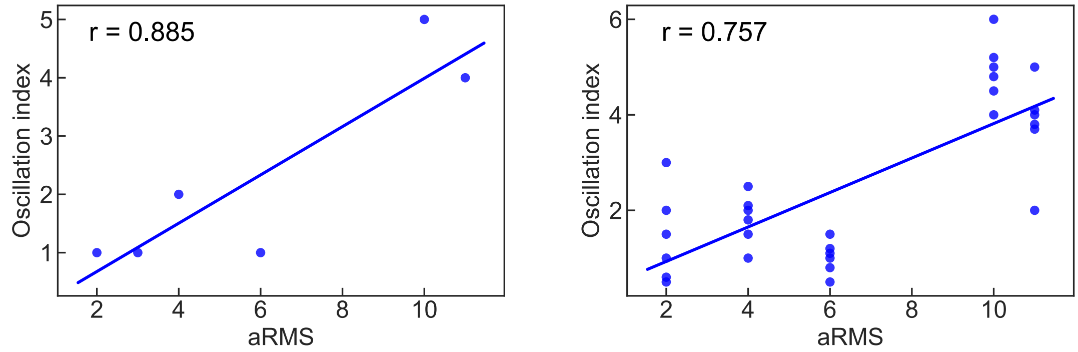
Figure 4.9. *Pearson correlation coefficient r for three different data sets.*

(b), the correlation is strongly negative, and r gets a value of -0.92. When the linear regression is negative, one variable decreases as the other variable increase. In the last figure (c), there is no linear correlation between the data points and r is close to zero.

At the data processing stage, missing values were handled with a method called 'fillna'. This method filled the missing data points with the last valid observation affecting the value of Pearson correlation coefficient. This effect is illustrated in figure 4.10. Pearson correlation coefficient r for the data set in figure (a) is 0.885. In figure (b), missing data points are filled with lats valid observation and Pearson correlation coefficient gets a value of r = 0.757. This means that data (a) has a stronger linear correlation than data (b). The presented phenomenon concerns case 1 and it is taken into account when interpreting the results.

The other used regression analysis method was called Spearman correlation. Spearman correlation measures the nonparametric association between two ordinal variables. This method is suitable for cases where the variables are related but not linearly. Spearman correlation coefficient does not react to parameters' deviation as strongly as Pearson correlation coefficient because instead of linear relationships, it measures monotonic relationship between two variables.

In order to calculate Spearman correlation coefficient, the parameters had to be ranked. In both data sets that were compared, the smallest parameter got the highest ranking and the highest parameter got the smallest ranking. After the parameters were ranked,



(a) Pearson correlation coefficient is calculated for the situation where one CLM index value corresponds to one MM index value.

(b) Pearson correlation coefficient is calculated for the situation where six CLM index values correspond to one MM index value.

Figure 4.10. *Data processing affected to Pearson correlation coefficient. There is a stronger linear regression between data points in figure (a).*

the differences between ranked numbers were calculated and utilized in correlation calculation. Spearman correlation coefficient ρ versus Pearson correlation coefficient r for nonlinear data set is illustrated in figure 4.11.

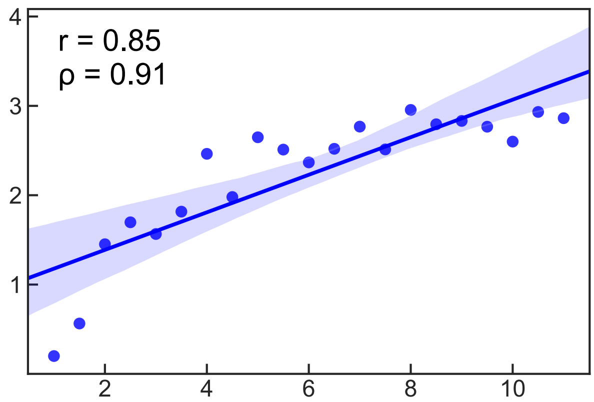


Figure 4.11. *Spearman correlation versus Pearson correlation for nonlinear dataset.*

Similar to Pearson correlation, Spearman correlation coefficient gets a value varying from -1 to 1. As seen in figure 4.11, the relationship between the variables is not completely linear but the relation is evident. For this case, Pearson correlation coefficient gets a value of r = 0.85 and Spearman correlation coefficient gets a value of rho = 0.91. The light blue area is a 95 % confidence interval for that regression line. It can be seen in the figure 4.11 that in the middle of the data set, linear Pearson correlation fits better than with small and high values, where the confidence band is wider. (seaborn 2012–2020)

Both Spearman and Pearson correlations were used for the examined data sets. The disadvantage of Spearman correlation was that it seemed to be much slower than Pearson correlation. Spearman correlation for Pandas data frame ranked the values again on each loop which was pretty time-consuming. An attempt was made to speed up the calculation process by ranking the values beforehand and performing the regression analysis for the preranked values. However, if the data set included any nan values, this way of working gave slightly different results than Pandas Spearman correlation function. Since all data sets contained nan values, it was decided to use the same Pandas Spearman correlation for all calculation rounds. On local computing, this meant that it was desirable to keep the size of one computed data set relatively small. In case 1, a reasonable time period for Spearman correlation calculation was one day. The amount of data in case 2 was much smaller and for that case, there were no limitations.

In an ideal case, the analysis would have been done on cloud. However, the nature of the conducted work was a proof of concept, so the computing was done locally. The creation of cloud-based data analytics would have required more familiarization with Snowflake and security issues. Cloud computing is included in further development potential in chapter 6.

5 RESULTS AND DISCUSSION

In this chapter, results from the regression analysis are be presented and discussed. Regression analysis was performed separately for cases 1 and 2. The goal was to find correlating parameters and examine if these variables together bring advantage in predictive maintenance. In regression analysis, data from different sources were first combined, and the correlations were observed for variables one by one. Most correlating variables were collected together in table format and plotted with Python as 2D scatter diagrams.

By correlation analysis, it was possible to determine which variables correlate, but it did not answer to the question *"Why the variables correlate"*? For the correlating parameters, it was further examined if there were a causality between them. Causalities could be used to predict the behavior of condition monitoring variables. In chapter 6, it is discussed how the correlation analysis could be utilized as a part of the existing machine monitoring application.

5.1 Results of the regression analysis

Machine monitoring, control loop monitoring, and equipment data regression analysis was performed by using Pearson and Spearman correlations. Both methods were used to calculate relationships between two variables at a time. As a result, correlation matrices for each case ware generated. The correlation matrices had as many rows and columns as there were examined parameters. For case 1, this meant hundreds of rows.

Regression analysis was performed for different time periods, each of which produced one correlation matrix. The correlation matrices were filtered so that only variables with a bigger positive correlation than 0.85 or lower negative correlation than -0.85 were saved. From these filtered matrices, in case 1 tags related to press section or pickup roll were chosen. In case 2, the most correlating variables were selected for further analysis. Correlating variables were plotted as a function of time and together as 2D scatter diagrams. From scatter diagrams it was easy to conclude whether the regression was linear or non-linear.

Because the work's nature was proof of concept, the available computing power was limited. It was not sensible to perform Spearman correlation analysis for a longer time period than one day. In case 2, the pump's time of deterioration was known, so the selection of time periods for the analysis was clear. As for in case 1, it was challenging to choose suitable periods, and the selection affected strongly to the results.

5.1.1 Case 1

One of the research questions was: *"Does data integration from different sources benefit predictive maintenance in the pulp and paper industry"*? In case 1, the data contained vibration monitoring and control loop monitoring key performance indicators. These KPI values were calculated from raw data in separate systems. Only the KPI values were available for the analysis.

The number of variables from both data sources was considerable, so the first step was to determine which variables should be combined. The correlations between variables from both systems were investigated by regression analysis.

The data was collected from ten months time period. Within the available calculation capacity, it was not possible to analyze the whole time period. When choosing the time periods for the analysis, high vibration levels concerning the press section and pick up roll were examined. After the definition of examined time intervals, data processing and regression analysis were performed. Regression analysis produced correlation matrices where each machine monitoring KPI was compared to control loop monitoring KPI one by one.

One time period for regression analysis was selected based on the relatively big change in the pick-up roll aRMS values. Regression analysis was performed for all vibration aRMS values and control KPIs. The investigation revealed that the pick-up roll aRMS value had the most significant correlation (0.944) with IAE value of a controller that is used to control temperature of the oil lubrication system. Other pick-up roll correlations to control loop monitoring KPIs were minor.

The pick-up roll is located at the end of the wet section, where it picks up the web from the forming fabric to the press section. The pick-up roll is in other words, located between the wet end and the press section. Since the pick-up roll did not have other notable correlations besides the temperature controller IAE value, IAE value was further examined. It had high correlations to other vibration KPIs measured close to pick-up roll. Five most correlating KPIs were located either at the wet end or press section. These correlating variables are presented in table 5.1.

	Oil lubrication system temperature (IAE)	
Vibration measurement KPI (aRMS)	Pearson	Spearman
Pick-up	0.944	0.930
Drawing roll secondary gear	0.954	0.927
Drawing roll primary gear	0.951	0.919
Press section primary gear	0.953	0.847
Felt cloth leading roll	0.943	0.883

Table 5.1. *Most correlating vibration KPIs with the oil lubrication system temperature controller IAE value.*

As seen in table 5.1 Pearson and Spearman correlations gave slightly different values. Pearson correlation coefficient got a bit higher values than Spearman correlation coefficient in every case. It may be concluded from the table 5.1 that the biggest correlations were all linear and positive. Correlation between the pick-up aRMS values and the oil lubrication system temperature controller IAE values are presented as a scatter diagram in figure 5.1. This strong positive correlation means that the temperature IAE value has been growing among the growing aRMS value.

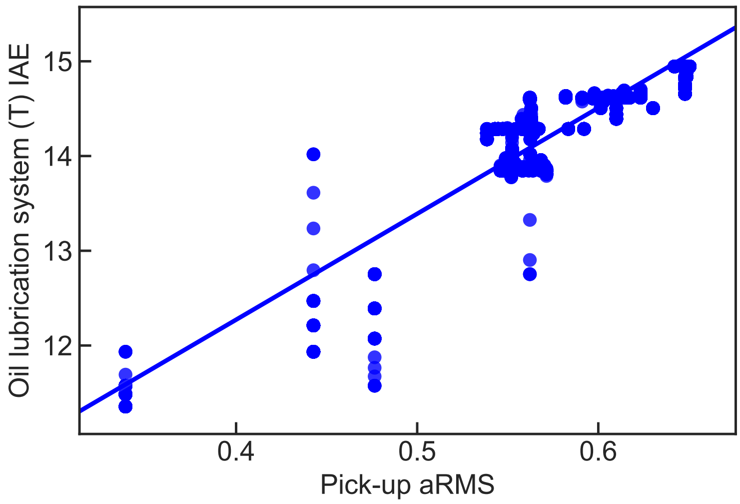


Figure 5.1. *Oil lubrication system temperature controller IAE value as a function of pick-up roll aRMS value.*

The line in the figure 5.1 illustrates linear Pearson correlation. Vibration KPI aRMS is proportional to the energy content of the vibration. The bigger the value is, the more pick-up roll has been vibrating. In the analysis period, aRMS values have changed but still stayed on a relatively low scale. If there were a critical fault situation, aRMS values would have increased more.

The control IAE values reflect the difference between the setpoint value and the process measurement. When the setpoint changes, it takes time for the controller to achieve the desired state. How fast a controller reaches the setpoint value depends on its tuning parameters. If the setpoint is changing several times within the examination period, IAE value increases. Bad tuning parameters may also cause a situation where the controller never achieves the setpoint, and IAE value stays continuously high. The situation in figure 5.1 looks like the running speed of the machine has changed. The pick-up aRMs value has been close to zero, which means that its rotating frequency has also been close to zero. Pick-up roll aRMs value and oil lubrication system temperature IAE are presented as a function of time in figure 5.2

It can be interpreted from figure 5.2 that there has been a drop in both KPI values at the same time. The aRMS value has dropped a bit before the IAE value, so presumably the production speed of the machine has changed. The effect of the used 'fillna' method can also be seen in figure 5.2. In the time series chart, the vibration KPI changes incremen-

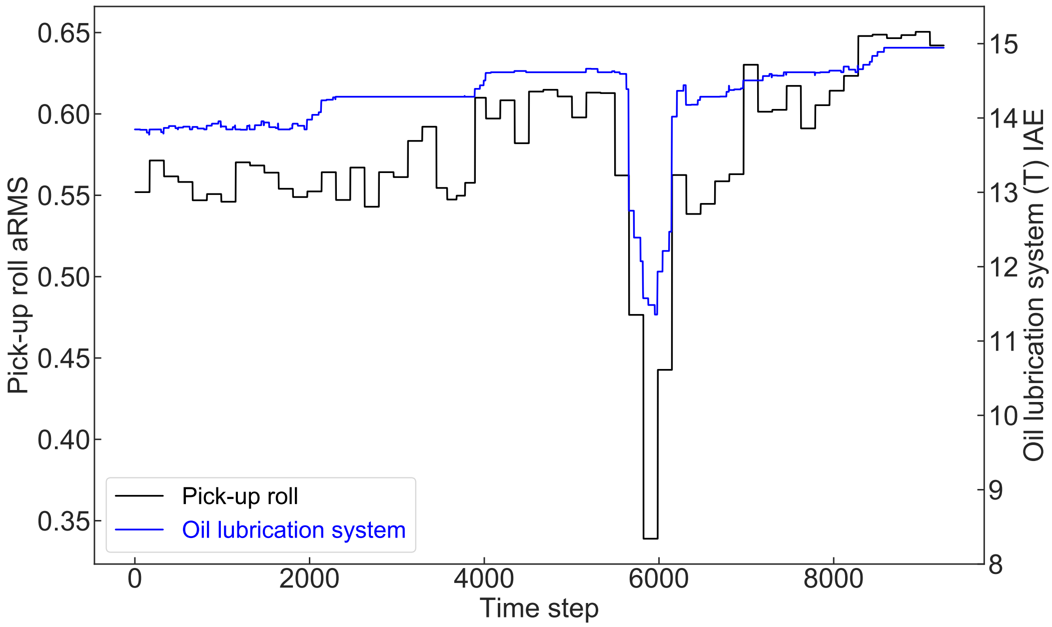


Figure 5.2. *The oil lubrication system temperature controller IAE and the pick-up roll aRMS as a function of time.*

tally, forming an angular line. The control KPI had more frequent sampling time, which appears as a smoother line.

Besides the pick-up roll, there were three vibration KPIs related to gears among the top five correlating measurements. The oil lubrication system temperature controller had a strong linear correlation with the drawing roll gears and the gears at the press section. This behavior is consequential since the circulating oil lubrication system is used to lube gearboxes. The lubrication system at the beginning of the board making process is typically used for both wet end and press section. The behavior of two drawing roll gear KPIs and oil lubrication system temperature controller KPI are presented in figure 5.3.

Similarly to figure 5.2, a definite drop in each KPI can be observed at the same time. In both cases, a correlation coefficient does not tell the origin of the correlation. In figures 5.2 and 5.3, KPIs have reacted similarly to the process change, but it is impossible to say if one of them has caused the other KPIs' behavior. Even though KPIs do have a strong correlation within the chosen time period, conclusions concerning causalities cannot be made based on the performed regression analysis.

In case 1, regression analysis was performed nine more times for different time periods. The choosing of periods was not only performed based on the vibration levels in the press section and the pick-up roll but also the other way around based on the CLM data. Information on index-specific limits was not included in the collected data, so pure values of CLM indices did not tell much about the performance rate. It would have been useful to have either the limits or data concerning the total performance index. ToPi index refers directly to the performance of a control loop, which would have been used to identify bad performance. However, the data collection was performed only for DNAHistorian

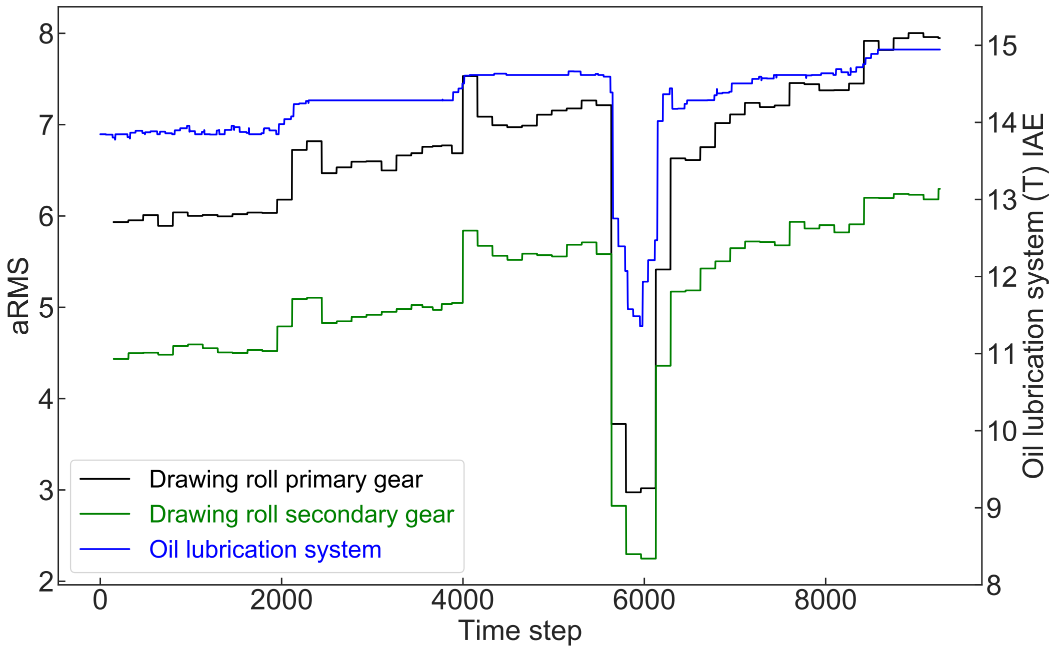


Figure 5.3. *The oil lubrication system temperature controller IAE and the drawing roll primary and secondary gear aRMS as a function of time.*

database, whereas the ToPi index values would have been located in LoopMonitoring LBHistory database. Without information about ToPi index, exceptional control loop performance was searched by tracing relatively big changes in IAE, CTI, OI, and VI values.

All nine analyzes gave similar results to the first one. In total, there were 6280 vibration monitoring KPIs and 2456 CLM KPIs. Among these indices, some correlations were found in every analysis. When the variables were examined as a function of time, their behaviour remind the cases in figures 5.3 and 5.2. If the KPIs were correlating with the running speed of the machine, they probably correlated with each other as well. In Future, information regarding the running speed of the machine should definitely be added to the analysis.

Because the time delay was not taken into account in this work, some correlations might have been missed. If a poorly performing controller at the beginning of the process causes vibrations later in the process, there would have been a time delay. The KPIs at the beginning of the process reacts faster than the KPIs at the end of the process. Therefore, they have different timestamp values. Data from different sources was, in particular, combined using timestamp values. Information regarding the running speed of the machine and location of the measuring points would have been needed to estimate the delay.

It is still rare that anomalous vibrations or badly performing control loops at the end of the process could be tracked by using KPIs at the beginning of the process. In a board machine, there are hundreds of moving parts whose vibration will cover the fault signals under numerous strong natural frequencies. Even if a faulty control loop would cause some vibration changes later in the process, the exact frequency should be known so that the vibration signal could be filtered just right.

Based on the analysis, relevant variables which combination would benefit predictive maintenance in case 1 are:

- Control KPIs related to the oil lubrication system.
- Vibration KPIs related to machinery parts, such as gears, which are lubricated by the oil lubrication system.

Quality of the lubrication is strongly related to the vibration levels. A lubricant reduces friction between machinery parts, whereby its feeding pressure and temperature affect the condition of a machine. Lowering the temperature of the oil increases its viscosity, whereas too high temperature may overheat the lubricated machinery parts. Pressure, in turn, affects the amount of flowing lubricant. It would possibly be useful to present lubrication system control KPIs with vibration KPIs related to machinery parts that are being lubricated. Relationships between these KPIs should be further examined.

5.1.2 Case 2

The aim of case 2 was to study reasons behind the green liquor pump's vibrations. The initial hypothesis for the excessive vibration was cavitation. One typical factor that causes cavitation is a low suction pressure at the impeller inlet. This may happen if the level of the dissolving tank lowers too much.

By regression analysis, it was examined if the cavitation claim was valid. The analysis was performed for the equipment data and vibration KPIs. As a result, the five most correlating vibration KPIs and equipment parameters are presented in table 5.2. Parameter T is temperature, p is pressure, and d is density of green liquor after the pump. RF is the rotational frequency of the pump. Level of the dissolving tank was not among the five most correlating parameters.

Vibration KPI	Method	Avg. flow	T	p	RF	d
Acceleration RMS	Pearson	0.77	0.92	-0.83	-0.86	0.92
	Spearman	0.71	0.79	-0.67	-0.69	0.80
vRMS 1–2 x RF	Pearson	-0.65	-0.91	0.88	0.94	-0.90
	Spearman	-0.58	-0.77	0.69	0.73	-0.77
vRMS 5–20 x RF	Pearson	-0.80	-0.95	0.86	0.88	-0.96
	Spearman	-0.68	-0.73	0.70	0.75	-0.75
Envelope peak	Pearson	0.83	0.95	-0.84	-0.88	0.96
	Spearman	0.77	0.79	-0.67	-0.71	0.80
Envelope RMS	Pearson	0.82	0.96	-0.85	-0.89	0.96
	Spearman	0.76	0.79	-0.68	-0.71	0.80

Table 5.2. *Five most correlating parameters in case 2.*

Vibration KPIs envelope peak, envelope RMS and vRMS 5–20 x RF had the most remarkable correlations. Both envelope KPIs had a strong linear relationship with temperature and density. Peaks in envelope spectrum refer to bearing faults and increased overall vibration levels without clear peaks are a sign of inadequate lubrication (Mikkonen et al. 2009). Envelope KPIs are presented as a function of time in figure 5.4.

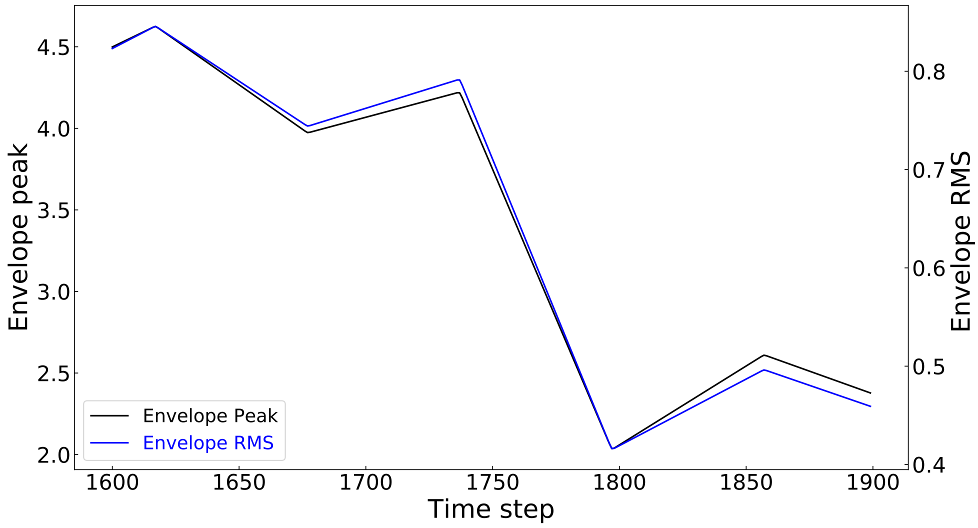


Figure 5.4. *Green liquor pump envelope peak and envelope RMS as a function of time.*

As seen in figure 5.4, both envelope KPIs have decreased. Therefore, bearing faults and lubrication problems were discarded from fault analysis. Since the Pearson and Spearman correlations with temperature and density were positive, these parameters must have decreased among the envelope KPIs. Correspondingly correlations with pressure were negative from where it may be concluded that pressure after the pump has increased.

Besides the envelope KPIs, the third most correlating parameter was vRMS 5–20 x RF. This KPI observe vibration changes in 5–20 multiples of rotational frequency. Variable vRMS 5–20 x RF correlated especially with temperature, pressure, and rotational frequency of the pump. As mentioned above, pressure and temperature after the pump had decreased. Negative correlation with these parameters indicates increased vibration levels in 5–20 multiples of rotational frequency. This certain vibration KPI vRMS 5–20 x RF was studied in more detail. A closer examination revealed that vRMS 5–20 x RF had the most significant overall growth.

Vibration KPI vRMS 5–20 x RF is presented as a function of time with the most correlating parameters pressure, temperature, and density in figure 5.5. In figure 5.5 the equipment data variables are normalized because they all had different scales. Normalization eliminates the measurement units enabling examination of parameters' relative changes in the same graph. Normalization was done by scaling each variable to have values between 0 and 1.

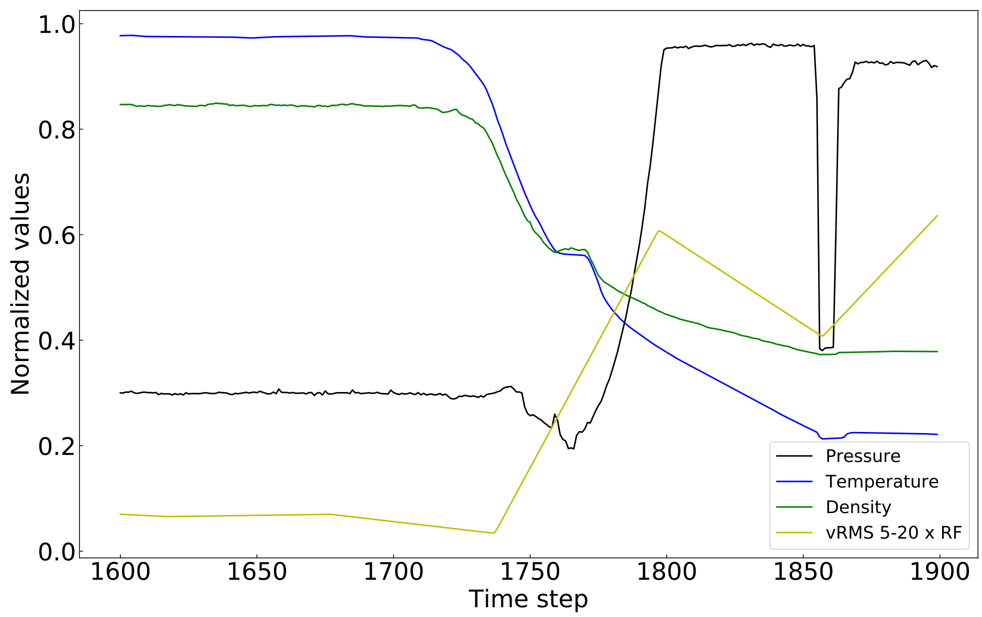


Figure 5.5. *Normalized pressure, temperature, density, and vibration KPI vRMS 5–20 x RF as a function of time.*

As assumed, it seemed that temperature and density had decreased whereas the pressure had increased. If the pump were cavitating, pressure after the pump should have decreased. Cavitation disturbs the pumping efficiency, which means that with the same rotational speed, the pump's outlet pressure remains lower. There is also a simultaneous drop in the pressure and vRMS 5–20 x RF at time step 1860. Vibration KPI is clearly following the pressure after the pump.

When vibration KPI vRMS 5–20 x RF was examined in frequency domain, the dominant frequencies were seen as peaks in the spectrum. In this case, the highest peak was at 80 Hz, which was the fifth multiple of prevailing rotation frequency 16 Hz. The examined pump had five blades so this fifth multiple of the rotational frequency is also known as blade passing frequency. Vibration KPI vRMS 5–20 x RF is presented in frequency domain in figure 5.6. In frequency domain spectrum, also second (160Hz) and third (240 Hz) harmonics of the blade passing frequency can be distinguished.

Typically, increased vibration levels at blade passing frequency and its harmonics refer to cavitation. Another typical sign of cavitation is a high-frequency random vibration spectrum, which is generated by imploding bubbles (Mikkonen et al. 2009). As seen in figure 5.6, these broadband frequencies are missing. According to Mikkonen et al. (2009), the reason behind the vibration could be rather something else. This kind of vibration spectrum is a typical sign of structural damages, radius of the rotor or a flow hindrance, such as an obstacle inside the pipe.

Increased vibration levels manifested only for a short period of time, so structural changes at the impeller were improbable. As the temperature decreased before the vibration levels rose, the physical properties of the green liquor may have changed. The green liquor is heated before pumping because of its high viscosity. Green liquor in low temperature and

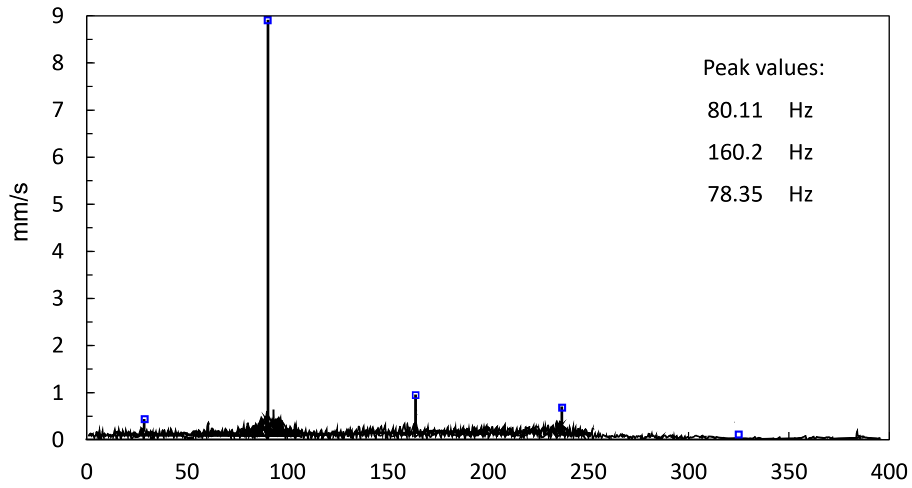


Figure 5.6. *Vibration KPI vRMS 5-20 x RF in the frequency domain. Peak values can be distinguished at 80.11 Hz and 160.2 Hz.*

high viscosity could possibly cause a blockage in a pipe. A sudden increase in pressure supports the theory of flow hindrance in the pipe. Before further research concerning green liquor viscosity, other alternatives were considered.

The examined pipeline had three valves. Correlation analysis took only one of them into account, which was the valve 2 in figure 4.4. The opening of this valve was expressed as a percentage, in which case the pumping could happen against different valve opening values. Valves 1 and 3 had only positions open (1) or close (0). The cause for the pump's vibration found from valve 3. While valves 1 and 2 stayed open, valve 3 was closed. It can be perceived from the figure 5.7 that pumping against a closed valve caused the pressure gain after the pump. A closed valve corresponds to an obstacle inside a pipe, causing a similar frequency-domain vibration spectrum at the pump's blade passing frequency as cavitation. The difference to cavitation is that the broadband frequencies were missing from the high-frequency spectrum.

In case 2, there were a clear causality between a closed valve, pressure, temperature, density, and vibration of the pump. However, the causality could not be proved only via regression analysis because the binary valve positions 1 and 0 did not correlate with the other parameters. When the correlation analysis was performed for normalized data, and only for a short time period, there were strong negative correlation between the valve 3 and vRMS 5-20 x RF. It is hard to tell if the regression analysis would have detected a closing valve if its position information had been expressed as percentage. For further development, a checkpoint that checks the valve positions should be added to the analysis. When the pump's motor is running, valves 1, 2, and 3 should be open.

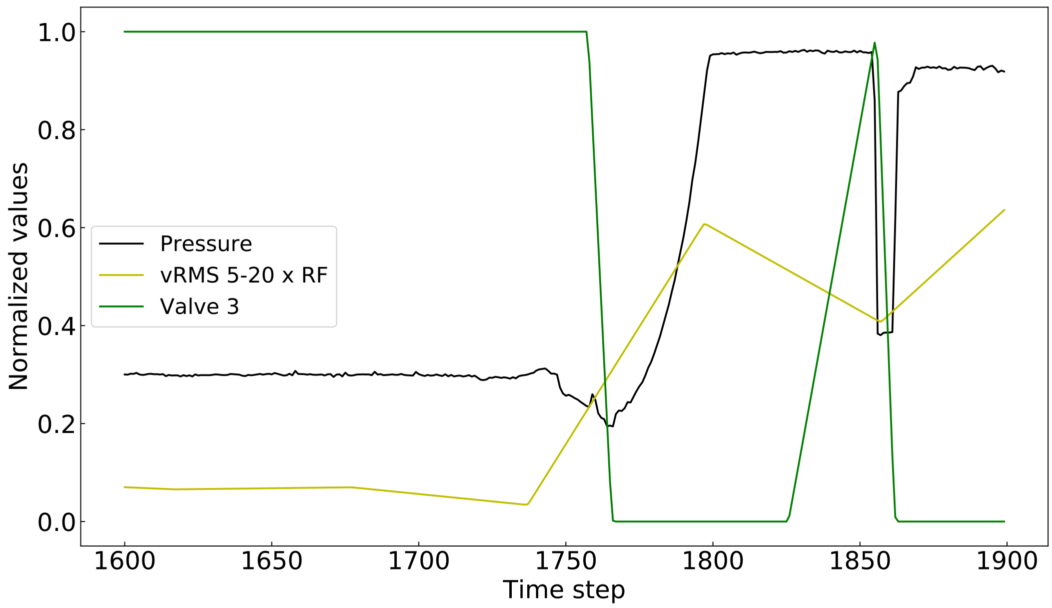


Figure 5.7. *Pressure in the pipeline and vibrations of the pump have increased after the valve 3 was closed.*

The regression analysis intended to find, which changes in process variables had led to increased vibration levels. By regression analysis, it was possible to determine vibration KPIs that correlated with certain process variables. Based on the analysis, the number of variables was delimited for further examination, and eventually, the fault was limited to concern a flow hindrance in the pipe. Unfortunately, the reason for the blockage could not be determined with the used data analysis methods, it was rather found manually by hand.

Pulp and paper industry process automation systems are mainly custom-made for each plant. The processes are sophisticated, including hundreds to thousands of control loops. The logic has not been built between every process device, and an estimation is that similar situations to case 2 can be found from several plants and industry sectors. The simplest way to avoid the problem is to define a feedback system, where a closing valve is a signal for the pump to stop operating. Witch such a feedback system, harmful vibrations that may decrease the pump's condition could be avoided.

6 FUTURE DEVELOPMENT

During the work, many development areas were identified. One fruitful area for further work could be adding analytical capabilities to existing monitoring applications. This chapter describes how regression analysis could be integrated to Valmet's existing machine monitoring application. It is also discussed what kind of reporting opportunities exist for the future and how the new reports would benefit predictive maintenance.

In case 1, one of the biggest challenge was to choose a suitable time period for the regression analysis. Under normal operating conditions, there is no need to perform regression analysis. The fault situations should be somehow identified from the data so that the selection of reasonable time periods for the analysis could be systematically automated.

One way of choosing suitable time periods for case 1 could have been anomaly detection. By anomaly detection, it is possible to recognize unexpected items or events in data sets (Pushe et al. 2017). Anomaly detection could be utilized in predictive maintenance by identifying situations where machinery condition deviates from "normal" operation. In this work, anomaly detection could have been used first to detect the fault situations, and after that, regression analysis could have been performed for the desired time period.

Besides anomaly detection, another way to find fault situations would have been to connect alarm data to the regression analysis. In both vibration monitoring and control loop monitoring applications, alarm limits are already set for the KPIs at the implementation phase. An exceeding of the limit causes an alarm and the number of alarms refers to the severity of the problem. Regression analysis could be performed based on this information for the most alarming measurements. Analysis would help to limit the causes of the alarms and possibly reveal failure chains that have not been recognized earlier.

Cloud-based computing enables the utilization of advanced data analytical methods in predictive maintenance applications. Valmet has recently developed a cloud-based machine monitoring application that combines alarm data and vibration monitoring KPIs. Data from both sources are shown in the same Tableau dashboard, but the solution does not contain any data analytics. It is the responsibility of the condition monitoring specialist to find the factors that have led to the alert. Previewing of data from various databases often requires logging on to different systems. In cloud-based solutions, there is potential to combine multiple data sources together. With Valmet Industrial Internet solutions, control loop monitoring data is possible to combine with the existing machine monitoring application.

In figures 6.1 and 6.2 it is illustrated how the regression analysis could be integrated into the current machine monitoring application. The current cloud-based application has six dashboards that are used to visualize the vibration KPI values and alarms related to the machine monitoring system. Regression analysis could be integrated into one of them, which concerns about most alarming measurement points and vibration KPIs over time. Figure 6.1 is a modified version of the existing dashboard, and it is not part of Valmet's product offering. Green liquor pump data from case 2 is used to generate these reports.

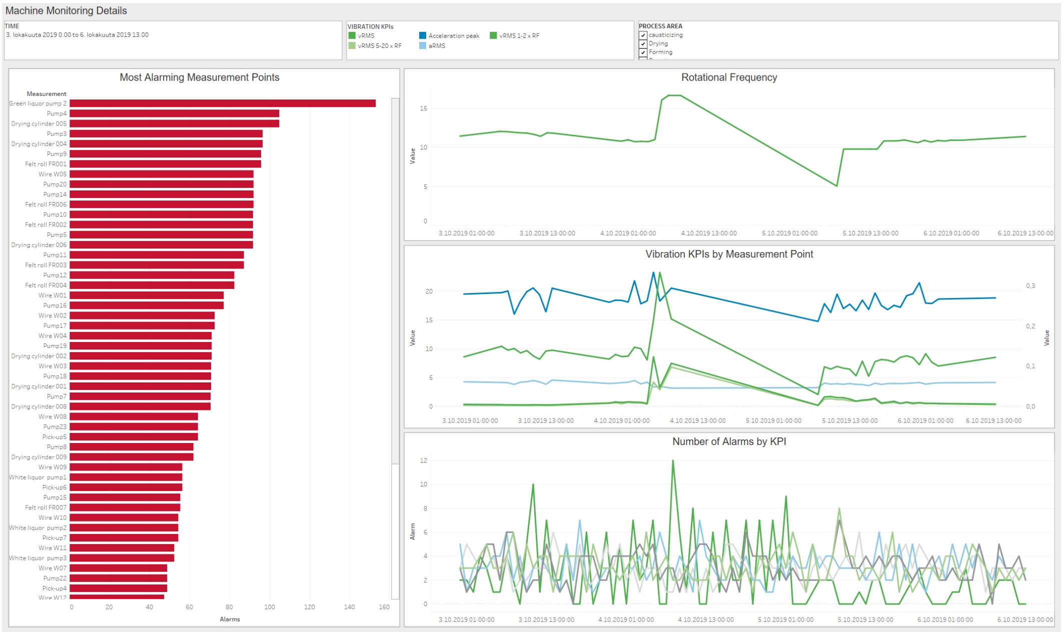


Figure 6.1. *"Machine Monitoring Details" -dashboard.*

In figure 6.1, the graph on the left shows the most alarming measurement points. Each bar refers to one measurement point, and the size of a bar is defined by a cumulative number of alarms within the reporting time period. The time period can be chosen from the filter section on the top of the dashboard. By clicking a bar in a bar chart of most alarming measurement points, other graphs on the dashboard are updated. Graphs on the right show vibration KPIs and number of alarms related to the chosen measurement point.

The graph on the top right shows the rotational frequency of the examined machinery part. In this case, the chart shows the rotational frequency of the green liquor pump. As for the chart below represents calculated vibration KPIs related to the chosen measurement point. Each KPI has a slightly different scale, so putting them into the same graph could be misleading. KPIs in the chart are normalized so that the relative changes stand out better. The normalization was done by scaling each variable to have values between 0 and 1. The highest value within the time period gets the value of 1 and the smallest value within the time period gets the value of 0. By clicking this graph, it is possible to navigate to another sheet where nonscaled values are presented.

The last graph on the bottom right in figure 6.1 shows the number of alarms by KPI as a function of time. This graph shows which vibration KPI has generated the most alerts. It also reveals the time period when most of the alarms are generated. This information can be used to for example filter the reporting time.

Another representative dashboard 6.2 shows the results of the regression analysis for the chosen vibration KPI. The selection can be done from the filter panel on top of the dashboard or by using the most alarming measurement point chart in previous dashboard 6.1. The graph on top left shows the cumulative amount of alarms by each vibration KPI. By clicking a KPI bar in this graph, the data on the other charts are updated.

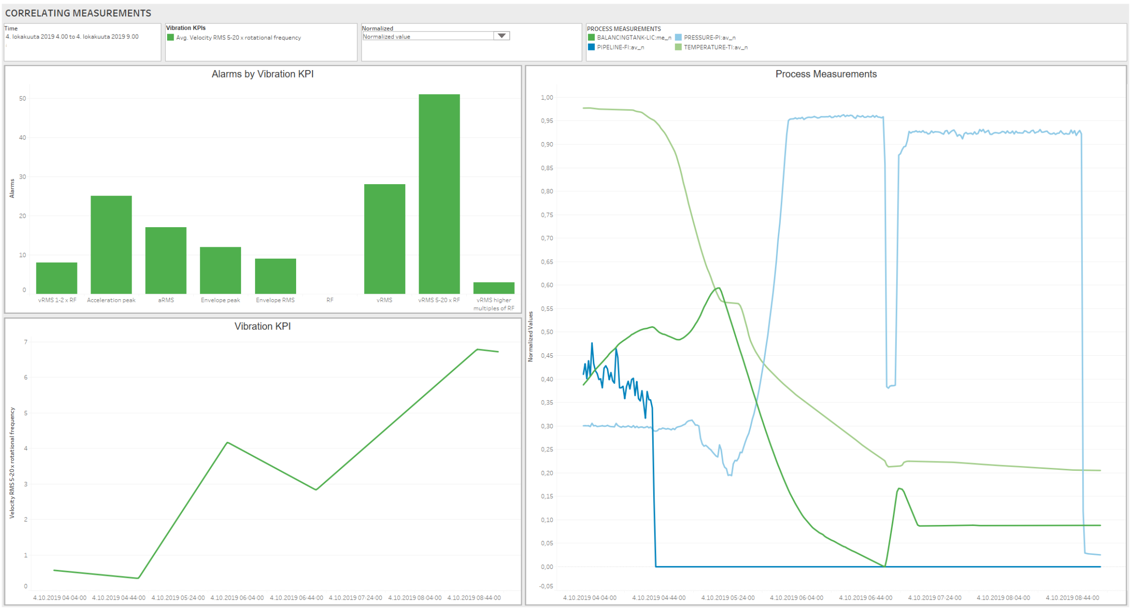


Figure 6.2. *"Correlating Measurements" -dashboard.*

A graph below the bar chart shows the trend line of the chosen KPI. In the example figure, the most alarming KPI was vRMS 5–20 x RF. In the trend line chart, clear growth in this KPI value can be seen within the selected time period. The future development challenge is to perform cloud-based regression analysis for the chosen KPI. From the analysis, most correlating process measurements are shown in the graph on the right. The purpose of this graph is to provide supporting information in troubleshooting for the operators or condition monitoring specialists.

These two dashboards 6.1 and 6.2 are built on top of the existing machine monitoring solution and both of them are illustrative. They represent how regression analysis could bring new value in predictive maintenance. Potential of other more advanced analytics capabilities, such as prediction, in predictive maintenance should be further examined. A prospective research question could be: *"How can the behaviour of a condition monitoring variable be predicted"*? In maintenance management, an answer to this question would lead the predictive maintenance to the next – proactive maintenance level.

7 CONCLUSION

The main objective of this thesis was to aggregate data from various sources and examine if the data integration benefits predictive maintenance in pulp and paper industry. At the industrial plants under the study, the information was stored in separate systems that were not intercommunicating. The data consisted of process equipment data, vibration monitoring, and control loop monitoring key performance indicators. In the experimental part of the work, relationships between variables were studied by regression analysis.

In predictive maintenance, it is essential to measure machine parameters that describe the condition of a machine. Monitoring of these parameters enables to identify failure patterns in an early stage. For rotating machinery, typical condition parameters are vibration signals. The signals are further used to calculate vibration key performance indicators. Vibration indicators refer to the state of a machine, but if the values change in an alarming direction, it falls to the maintenance engineer's responsibility to determine the cause behind the change. This thesis investigated whether the indices from different systems together could provide sophisticated information about the state of a machine or the causes of the deterioration.

The conducted work consisted of two cases. The first case combined vibration monitoring and control loop monitoring data. As a result of the regression analysis, some correlating parameters were distinguished. Still, these results cannot be utilized in failure pattern identification without the event data and information concerning the running speed of the board machine. Multiple regression analysis revealed that variables that correlated with the prevailing operating conditions, most likely correlated with each other as well. Combining and monitoring these variables together would not bring significant advantage in predictive maintenance. The most remarkable subject of further investigation in case 1 was the circulating oil lubrication system. Potential connections between control loop and vibration monitoring indicators at this process area should be examined in a fault situation.

The second subject of study concerned vibration monitoring and process equipment data. The green liquor pump vibrations unexpectedly increased at the blade passing frequency in certain operating conditions. Cavitation was indeed initially estimated to be the reason for pump vibrations, but the applicability of regression analysis to this situation was investigated. The results showed that instead of cavitation, the pumping happened against a closed valve causing similar vibrations as an obstacle in a pipe. A suggestion to avoid such situations is a simple feedback system where the pump cannot run when the valve after it is closed. In general, the amount of data to be examined in case 2 was much smaller than in case 1, which made the data management much easier. It seemed that combining vibration and equipment data would benefit predictive maintenance more than combining vibration and control loop data.

As the work proceeded, further subjects of research were distinguished. Combining the control loop monitoring data with the vibration monitoring data did not produce the desired results, but instead of that, aggregation of control loop monitoring and field device management data should be further studied. The field device management provides configuration and maintenance capabilities for intelligent field equipment. For example, a control valve is typically adjusted by compressed air. In a common failure situation, the used air is not clean enough, and the pilot valve gets dirty. Malfunction of a field device affects to the process in a way that the controller cannot adjust the valve as desired. The fault may manifest in both control indices and field device management values. Further research is needed to examine the link between these two data sources more closely.

Taken together, these results increase the understanding of the rapidly expanding field of industrial internet solutions. Cloud-based data warehouses and analysis tools have brought new possibilities for predictive maintenance. Although the current study is based on only two examples, the findings suggest that the data to be transferred to the cloud should be as raw as possible. In the higher-level computational indices, data storage cycles are typically longer, when some data has already been lost. The quality of regression analysis and other data analysis methods are strongly related to the quality of input data. The advantage of cloud-based solutions is that analyzes can be particularly performed on raw data – even though it requires a lot of computing capacity.

This study lays the groundwork for future research into versatile and efficient use of data in process industry. Combining data from different sources opens new possibilities for maintenance management strategies. Advanced data analytics are playing an increasingly important role in industrial internet solutions offering new customer value through different types of services and solutions.

REFERENCES

American Forest Paper Association (2019). *Pulp*. URL: https://www.afandpa.org/our-products/pulp. (accessed: 27.01.2020).

Andersson, P. (2014). *A dynamic Na/S balance of a kraft pulp mill. Modeling and simulation of a kraft pulp mill using WinGEMS*. URL: https://www.diva-portal.org/smash/get/diva2:704466/FULLTEXT01.pdf. (accessed: 23.01.2020).

Bauer, M., Horch, A., Xie, L., Jelali, M. and Thornhill, N. (2019). The current state of control loop performance monitoring - A survey of application industry. *Journal of Process control* 38, 1–10.

C.K.M. Lee, Ty Cao and Kam Hung. Ng (2017). *Big Data Analytics for Predictive Maintenance Strategies*. URL: https://www.researchgate.net/publication/312004126_Big_Data_Analytics_for_Predictive_Maintenance_Strategies. (accessed: 31.10.2019).

Dimecc (2016). *S4Fleet – Service solutions for fleet management*. URL: https://www.dimecc.com/dimecc-services/s4fleet/. (accessed: 28.01.2020).

Dorf, R. C. and Bishop, R. H. (2010). *Modern control systems, 12th ed.* New Jersey: Pearson Education, 2–5.

Eisenmann Sr., R. and Eisenmann Jr., R. (2005). *Machinery Malfunction Diagnosis and Correction. Vibration Analysis and Troubleshooting for the Process Industries*. Pearson Education, Inc. ISBN: 0-13-240946-1.

Environmental Paper Network (2018). *The state of the global paper industry*. URL: https://environmentalpaper.org/wp-content/uploads/2018/04/StateOfTheGlobalPaperIndustry201 FullReport-Final-1.pdf. (accessed: 23.01.2020).

Finnish forest industries (2017). *Pulp and paper industry*. URL: https://www.forestindustries.fi/statistics/pulp-and-paper-industry/. (accessed: 23.01.2020).

Friman, M., Harju, T., Marttinen, A. and Ketonen, M. (2004). *Method and system from analysing control circuit performance in industrial process. Patent Application Publication*. URL: https://patentimages.storage.googleapis.com/8b/83/ba/894ef23f7b6ab5/US20040199360A1.pdf. (accessed: 24.07.2019).

Galar, D., Parida, A., Gustafson, A. and Berges, L. (2011). Use of Vibration Monitoring and Analysis as KPIs in Paper Industry. *Condition Monitoring and Diagnostics Engineering Management International Congress*, 135–147.

Goyal, D. and Pabla, B. (2015). The Vibration Monitoring Methods and Signal Processing Techniques for Structural Health Monitoring: A Review. *Archives of Computational Methods in Engineering* 23.4, 585–594.

Hägglund, T. (1995). A Control Loop Performance Monitor. *Control Eng. Practice* 3.11, 1543–1551.

Hooda, D. and Raich, V. (2017). *Fuzzy Logic Models and Fuzzy Control. An Introduction*. International series of monographs on physics. Alpha Science International LTd. ISBN: 978-1-78332-281-7.

IBISWorld (2019). *Global Paper & Pulp Mills Industry – Market Research Report*. URL: https://www.ibisworld.com/global/market-research-reports/global-paper-pulp-mills-industry/. (accessed: 28.01.2020).

Institute, M. (2019). *Vibration Analysis Dictionary*. URL: https://www.mobiusinstitute.com/site2/item.asp?LinkID=2002. (accessed: 24.09.2019).

Jämsä-Jounela, S., Poikonen, R., Vatanski, N. and Rantala, A. (2003). Evaluation of control performance: methods, monitoring tool and applications in a flotation plant. *Mineral Engineering* 16, 1069–1074.

Janssen, M., Laflamme-Mayer, M. and P.R. Stuart (2003). Survey of data management system used in the pulp and paper industry. *Proceedings Foundations of Computer-Aided Process Operations (FOCAPO2003)*. URL: https://www.researchgate.net/publication/266450544_Survey_of_data_management_systems_used_in_the_pulp_and_paper_industry. (accessed: 22.1.2020).

Jelali, M. (2013). *Control Performance Management in Industrial Automation: Assessment, Diagnosis and Improvement of Control Loop Performance*. London: Springer-Verlag. ISBN: 978-1-4471-4545-5.

Jokinen, J., Postelnicu, C., Zhang, B., Camp, R., Suhonen, M., Karhumäki, O. and J. L. Martinez Lastra (2012). Condition Monitoring for Predictive Maintenance in the Pulp Paper Industry: Two Implementations. *IEEE International COnference on Systems, Man, and Cybernetics*. URL: https://ieeexplore-ieee-org.libproxy.tuni.fi/stamp/stamp.jsp?tp=&arnumber=6377971. (accessed: 22.1.2020).

Juuso, E. K. and Lahdelma, S. (2013). Intelligent performance measures for condition-based maintenance. *Journal of Quality in Maintenance Engineering* 38, 278–294.

Knierim-Dierz, N., Hanel, L. and Lehner, J. (2012). Definition and Verification of the Control Loop Performance of Different Power Plant Types.

KnowPap (1997–2000). *Paper Technology*. URL: http://www.knowpap.com/extranet/english/knowpap_system/user_interfaces/tuotantoprosessit/tuotantoprosessit.htm. (accessed: 23.01.2020).

Lahti, L. (2018). *PERFORMANCE INDEX FOR CLOSED-LOOP CONTROLS*. URL: https://trepo.tuni.fi/bitstream/handle/123456789/25888/Lahti.pdf?sequence=4&isAllowed=y. (accessed: 5.12.2019).

Machine Condition Monitoring Technical Library (2017). URL: http://www.ni.com/example/6511/en/. (accessed: 15.07.2019).

Martinsuo, M. and Kärri, T. (2017). *Teollinen internet uudistaa palveluliiketoimintaa ja kunnossapitoa*. Savon Kirjapaino Oy, Kerava. ISBN: 978-952-68687-0-7.

Mary, S. and Raynaud, J. (2016). Green Impact Screener. Thematic Impact Investing. *Kepler Cheuvreux Transition Research*. URL: https://www.longfinance.net/media/documents/KC_impact_screener_scouting_2_C_opportunities_.pdf. (accessed: 22.1.2020).

Mechefske, C. K. (2005). *Vibration and Shock Handbook. Chapter 25 Machine Condition Monitoring and Fault Diagnostics*. URL: https://pdfs.semanticscholar.org/5d86/6f9570c01c30371324656272c790b20cd299.pdf. (accessed: 16.07.2019).

MetsäFibre (2020). *Sellunvalmistus*. URL: https://www.metsafibre.com/fi/Sellu/Pages/Sellunvalmistus.aspx. (accessed: 23.01.2020).

Mikkonen, H., Miettinen, J., Leinonen, P., Jantunen, E., Kokko, V., Riutta, E., Sulo, P., Komonen, K., Lumme, V. E., Kautto, J., Heinonen, K., Lakka, S. and Mäkelänen, R. (2009). *Kuntoon perustuva KUNNOSSAPITO käsikirja*. Helsinki: KP-Media Oy.

Ministery of Agriculture and Forest of Finland (2018). *Forest industry in Finland*. URL: https://mmm.fi/en/en/forests/use-of-wood/forest-industry. (accessed: 24.01.2020).

Nopanen, A. (2017). *Utilizing the industrial internet for power plants*. URL: https://dspace.cc.tut.fi/dpub/bitstream/handle/123456789/25242/nopanen.pdf?sequence=1&isAllowed=y. (accessed: 30.10.2019).

Oulun yliopisto (n.d.). *Prosessitekniikan perusta: Automaatiotekniikka*. (accessed: 6.12.2019).

Pöyry (2019). *How future will reshape global pulp and recovered paper demand and supply*. URL: https://www.poyry.com/news/articles/how-future-will-reshape-global-pulp-and-recovered-paper-demand-and-supply. (accessed: 29.01.2020).

PSK Standard Association (2010). *PSK 7501, Key performance indicators of maintenance for use in process industry*. Version 2. URL: https://www.psk-standardisointi.fi/Standard/Ryhma75/PSK7501_2p.pdf. (accessed: 25.06.2019).

– (2011). *PSK 6201, Maintenance. Terms and definitions*. Version 3. URL: https://www.psk-standardisointi.fi/Standard/Ryhma62/PSK6201.pdf. (accessed: 25.06.2019).

– (2015). *PSK 5706, Vibration Measurements in Condition Monitoring. Monitoring Methods.* Version 4. URL: https://www.psk-standardisointi.fi/Standard/Ryhma57/PSK5706-4p.pdf. (accessed: 25.06.2019).

– (2018). *PSK 5703, Vibration Measurement in Condition Monitoring. Selection and Mounting of Sensor, Connection and Cable*. Version 5. URL: https://www.psk-standardisointi.fi/Standard/Ryhma57/PSK5703_5p.pdf. (accessed: 08.08.2019).

Pushe, Z., Masaru, K., Junichi, T., Tojiro, N., Shigeyoshi, C. and Tadashi, S. (2017). Advanced Correlation-Based Anomaly Detection Method for Predictive Maintenance. *2017 IEEE International Conference on Prognostics and Health Management (ICPHM)*. URL: https://ieeexplore-ieee-org.libproxy.tuni.fi/stamp/stamp.jsp?tp=&arnumber=7998309. (accessed: 29.11.2019).

Rao, B. (1996). *Handbook of Condition Monitoring. First Edition*. Elsevier Advanced Technology. ISBN: 1-85617-234-1.

Reac (n.d.). *What is an actuator*. URL: https://www.reac.se/facts/actuators/what-is-an-actuator/. (accessed: 13.06.2019).

Russell, S. and Norvig, P. (2010). *Artificial Intelligence. A Modern Approach*. Pearson Education, Inc., 550–552. ISBN: 978-0-13-604259-4.

S. Joe Qin (2001). Process Data Analytics in the Era of Big Data. *AIChE Journal* 60.9, 3092–3100. URL: https://aiche-onlinelibrary-wiley-com.libproxy.tuni.fi/doi/epdf/10.1002/aic.14523. (accessed: 15.10.2019).

seaborn (2012–2020). *Visualizing linear relationships*. URL: https://seaborn.pydata.org/tutorial/regression.html. (accessed: 29.01.2020).

The Difference Between RMS, Peak and Peak to Peak Amplitudes (2019). URL: http://www.vibrationschool.com/mans/Plots/Plots03.htm. 2019-6-25.

Thomson, W. T. (1981). *Theory of vibration with applications*. Englewood Cliffs, NJ:Prentice-Hall. ISBN: 0-13-91423-0.

Ulrik, U., Jennifer, W., Oguz, D., Ali, E., Azzeddine, S., Robert, P. and Hemant, K. (2006). Prediction of rotating stall within an impeller of a centrifugal pump based on spectral analysis of pressure and velocity data. *Journal of Physics* 52, 36–45. URL: https://iopscience.iop.org/article/10.1088/1742-6596/52/1/004. (accessed: 26.11.2019).

Väänänen, P. (1990). *TrendVib konnonvalvontajärjestelmän signaalianalyysiohjelmisto*. International series of monographs on physics. Helsingin Yliopisto.

Valmet (2010). *Roll Modernizations: Suction Rolls*. URL: https://www.valmet.com/globalassets/media/downloads/white-papers/roll-services/wpr_rollmodsuction.pdf. (accessed: 27.01.2020).

– (2019a). *Board and paper machines*. URL: https://www.valmet.com/board-and-paper/board-and-paper-machines/. (accessed: 28.01.2020).

– (2019b). *Improved performance with Industrial Internet services and applications*. URL: https://www.valmet.com/about-us/industrial-internet/. (accessed: 29.01.2020).

– (2019c). Valmet Control Performance. Asset reliability optimization – Internal.

Valmet Automation Inc. (2019). *Valmet DNA Machine Monitoring Operator Manual*. Version Rev.1 / C2018.

Varsinais-Suomen Yrittäjä (2015). *Valmet muuttaa tiedonjyvät liiketoiminnaksi*. URL: https://www.y-lehti.fi/uutiset/nayta/12504?uutiset/nayta/12504. (accessed: 28.01.2020).

Vecer, P., Kreidl, M. and Smid, R. (2005). Condition Indicators for Gearbox Condition Monitoring System. *Acta Polytechnica* 45.6, 35–43.

What is a KPI (2019). *Measure your performance against key business objectives*. URL: https://www.klipfolio.com/resources/articles/what-is-a-key-performance-indicator. (accessed: 11.07.2019).

Windrock (2018). *Why Condition Monitoring Is Important*. URL: https://windrock.com/blog/why-condition-monitoring-is-important/. (accessed: 10.12.2019).