

POLITECNICO DI TORINO

Master of Science in Mechatronic Engineering

Master's Degree Thesis

Machine Learning model for tribological data extraction from experimental tests



**Politecnico
di Torino**

**RWTHAACHEN
UNIVERSITY**

ifas Institute for
Fluid Power
Drives and Systems

Academic Supervisors:

Prof. Luigi Mazza

Dr. Achill Holzer

Candidate:

Francesco Lenci

Graduation session October 2022

Academic Year 2021-2022

Abstract

Nowadays, increasing demand for performances and new environmental requirements have led to the need for new materials and lubricants in the fluid power field. The thesis is based on research carried out in the offices at the IFAS institute. Based on existing measurement data of a disc-on-disc tribometer, the shapes of the Stribeck curves were analysed. Furthermore, the essential characteristics of the curves were extracted, such as the minimum and maximum coefficient of friction, the speed at the curve's minimum point, the number of peaks, and the slopes of the curve. Some of the parameters and the experimental setup have the purpose of feeding a machine learning model. After tuning and comparing different models, the Random Forest regression model was chosen. Finally, predictions with new materials and lubricants were performed. In parallel, some significant comparisons between materials and lubricants were investigated. The latter does not need to create new specimens for the test bench, significantly saving time and money. Further research, increasing the sample size, is needed to improve the method and to get more accurate predictions.

Contents

1	Introduction	1
1.1	<i>Motivations.....</i>	<i>1</i>
1.1.1	Velocity and pressure	2
1.1.2	Environment.....	4
1.1.3	Surface finish.....	6
1.2	<i>Goal.....</i>	<i>6</i>
1.3	<i>Contributions.....</i>	<i>7</i>
2	Tribology state of the art	8
2.1	<i>Solid surface topography.....</i>	<i>9</i>
2.1.1	Geometric characteristics of solid surfaces	9
2.1.2	Analysis of surface roughness	9
2.1.3	Adhesion.....	11
2.2	<i>Friction and Wear.....</i>	<i>12</i>
2.2.1	Friction.....	12
2.2.2	Wear.....	22
2.3	<i>Lubrication.....</i>	<i>27</i>
2.3.1	Types of lubricants.....	28
2.3.2	Properties of liquid lubricants	29
2.4	<i>Friction test benches.....</i>	<i>30</i>
2.4.1	Laboratory test bench: tribometer.....	32
3	ML and AI state of the art	34
3.1	<i>Artificial Intelligence and Machine Learning.....</i>	<i>34</i>
3.2	<i>ML and AI in Tribology.....</i>	<i>35</i>
3.3	<i>Condition monitoring.....</i>	<i>37</i>
4	Simulations, graphs and comparisons	40
4.1	<i>Conversion of the dataset</i>	<i>41</i>

4.2	<i>Stribeck curve</i>	42
4.2.1	Boundary lubrication	43
4.2.2	Hydrostatic lubrication	43
4.2.3	Mixed lubrication	43
4.2.4	Hydrodynamic lubrication.....	44
4.2.5	Elastohydrodynamic lubrication.....	45
4.3	<i>Run-in</i>	48
4.4	<i>Saving the data structures</i>	50
4.5	<i>Baseline construction</i>	51
4.5.1	Raman spectroscopy.....	52
4.6	<i>Characteristic extraction</i>	53
4.6.1	Minimum coefficient extraction.....	53
4.6.2	Gradients and slopes	63
4.6.3	Speed in the change point	65
4.6.4	Maximum coefficient of friction	66
4.6.5	Number of peaks	66
4.7	<i>Inputs and labels of the model</i>	75
4.7.1	Density	75
4.7.2	Viscosity.....	76
4.7.3	Pour point	79
4.7.4	Young modulus	80
4.7.5	Poisson ratio	80
4.7.6	Yield strength	81
4.7.7	Hardness.....	82
4.8	<i>Hot encoding and normalisation</i>	84
4.8.1	Hot encoding	84

4.8.2	Normalisation of the dataset	86
4.9	<i>Choice of the machine learning model</i>	87
4.9.1	Types of machine learning regression model	88
4.9.2	Train and test split	93
4.9.3	Scoring metrics	94
4.9.4	Scoring metrics results.....	97
4.10	<i>Random Forest regressor</i>	97
4.10.1	Advantages.....	100
4.10.2	Disadvantages.....	100
4.10.3	Parameters of the Python function	101
4.10.4	Tuning the Random Forest model	102
4.11	<i>Variable importance</i>	107
5	Results and conclusions	110
5.1	<i>Predictions</i>	110
5.2	<i>Results</i>	113
6	Conclusions and outlook	124
6.1	<i>Conclusions</i>	124
6.2	<i>Outlook</i>	125
	Bibliography	126

List of figures

Figure 1.1: Axial pump [4].	4
Figure 1.2: Growth of sales [7].	5
Figure 2.1: Geometric characteristics of solid surfaces [15].	9
Figure 2.2: Comparison of three types of reference line: (a) mean; (b) ten-point average; (c) least squares [15].	11
Figure 2.3: Commercial static friction test bench [1].	13
Figure 2.4: Static friction force and kinetic friction force as a function of time [3].	14
Figure 2.5: Schematic illustration of the transition from static contact to sliding contact for a hard asperity on a soft surface [2].	16
Figure 2.6: Relation between friction and hardness (square: hard steel slider; circle: diamond slider) [18].	18
Figure 2.7: Friction coefficient for the same sliding materials at various normal loads [1].	18
Figure 2.8: Data for a cast Inconel alloy pin sliding unlubricated on an M-10 tool steel disk [1].	20
Figure 2.9: Four possible relationship between friction force and sliding velocity [1].	21
Figure 2.10: Schematic images of four representative wear modes [13].	23
Figure 2.11: Schematic illustration of two break possibilities: at the interface (1) or at the weakest region of one body (2) [3].	24
Figure 2.12: Schematics of abrasive wear processes as a result of plastic deformation [3].	26
Figure 2.13: Types of friction-testing implementations [1].	31
Figure 2.14: Test bench: disc on disc tribometer [21].	33
Figure 2.15: Tribometer scheme [21].	33
Figure 3.1: The types of events preceding machine failure [24].	38
Figure 3.2: The complex nature of a typical lubricated contact [24].	39
Figure 4.1: Stribeck curve and lubrication regimes [25].	42
Figure 4.2: Hydrodynamic pressure generated between two non-parallel surfaces [2].	45
Figure 4.3: Stribeck curve in non-conformal contacts.	46
Figure 4.4: Effect of changing the EHL ratio.	47
Figure 4.5: Effect of varying the film thickness.	47
Figure 4.6: Effect of a solid-like film formation.	47
Figure 4.7: Effect of a change in surface roughness.	47
Figure 4.8: Effect of the running-in on a Stribeck curve.	47
Figure 4.9: Six forms of friction behaviour as a function of time [1].	49
Figure 4.10: Friction and roughness functions during the running-in [3].	50
Figure 4.11: Baseline of a Stribeck curve	52
Figure 4.12: Baseline curve construction by means of the Raman spectroscopy.	53
Figure 4.13: Dynamic programming with min_size=5 and jump=15.	57
Figure 4.14: Binary segmentation with jump=415.	57
Figure 4.15: Bottom-up with jump=915.	58
Figure 4.16: Linearly penalized segmentation (Pelt) with min_size=405 and jump=1265.	59
Figure 4.17: Stribeck curve of v287 (green point = change point detection, red point = minimum detection).	62

Figure 4.18: Stribeck curve of v288 (green point = change point detection, red point = minimum detection).	62
Figure 4.19: Stribeck curve of a specimen with low surface roughness.	67
Figure 4.20: Stribeck curve of a specimen with high surface roughness.	67
Figure 4.21: Types of stress in a contact subject to fretting [26].	69
Figure 4.22: Stribeck with all its peaks highlighted.	71
Figure 4.23: Meaning of the prominence.	72
Figure 4.24: Histogram of comparison for different values of prominence.	72
Figure 4.25: Third quartile or 75th percentile.	73
Figure 4.26: Stribeck with peaks detected setting prominence=0.007 (75th percentile).	74
Figure 4.27: Stribeck with peaks detected setting prominence=0.000125 (95th percentile).	74
Figure 4.28: Schematic representation of the fluid separating two surfaces [2].	76
Figure 4.29: Viscosity as a function of the temperature [2].	78
Figure 4.30: Viscosity as a function of the pressure [27].	79
Figure 4.31: Linear regression.	88
Figure 4.32: Single neuron representation.	90
Figure 4.33: Neural Network.	90
Figure 4.34: Decision tree regression.	91
Figure 4.35: Gaussian Process Regression [29].	92
Figure 4.36: Random Forest regressor [29].	98
Figure 4.37: Feature importance analysis result.	108
Figure 5.1: Interpolation of the predictions with Azolla as lubricant and C27450 as material.	114
Figure 5.2: Interpolation of the predictions with Fuchs as lubricant and EnviB as material.	114
Figure 5.3: Interpolation of the predictions with Tellus as lubricant and OF2228 as material.	115
Figure 5.4: Interpolation of the predictions with Powerflow NZ as lubricant and OF2299 as lubricant.	115
Figure 5.5: Histogram of comparison for different values of surface finishing.	116
Figure 5.6: Comparative costs of different manufacturing processes [3].	117
Figure 5.7: Effect of elastic modulus to static friction [2].	118
Figure 5.8: Histogram of comparison for different values of elastic modulus.	119
Figure 5.9: Histogram of comparison for different values of hardness.	119
Figure 5.10: Static friction as a function of yield strength	121
Figure 5.11: Histogram of comparison for different values of yield strength	121
Figure 5.12: Histogram of comparison for different values of viscosity	123

List of tables

Table 1.1: Comparison of datasheet [16] [17].	2
Table 2.1: Technical specification of the test bench.....	32
Table 4.1: Advantages and disadvantages of dynamic programming.	55
Table 4.2: Advantages and disadvantages of linearly penalized segmentation.	55
Table 4.3: Advantages and disadvantages of binary segmentation.	55
Table 4.4: Advantages and disadvantages of bottom-up segmentation.....	56
Table 4.5: Advantages and disadvantages of window sliding segmentation.	56
Table 4.6: Conversion table for different hardness scales of measurement [28].	83
Table 4.7: Hot encoding example.	85
Table 4.8: Scoring metrics results for the different machine learning methods.	97
Table 5.1: Brass materials specifications.	110
Table 5.2: Lubricants specifications.	110
Table 5.3: New selected brass materials specifications.	111
Table 5.4: New selected lubricants specifications.....	111
Table 5.5: Type of predictions carried out	112
Table 5.6: Young modulus, yield strength and hardness values for the different brass materials.	120

Nomenclature

Acronyms

RMS: “Root mean square”, $\text{RMS} = \sqrt{\frac{1}{n} \sum_i x_i^2}$.

NBA: “National Basketball Association”.

IFAS: “Institute for Fluid Power and Systems”.

ML: “Machine Learning”.

AI: “Artificial Intelligence”.

ANN: “Artificial Neural Network”.

SKF: “Svenska Kullagerfabriken”, it is a Swedish bearing and seal manufacturing company.

GE: “General Electric”, it is an American multinational conglomerate operating in sectors including healthcare, aviation, power, renewable energy.

HD: “Hydrodynamic lubrication”.

EHD: “Elastohydrodynamic lubrication”.

CPD: “Change Point Detection”.

UTC: “Coordinated Universal Time”

RTC: “Real-time clock”.

CPU: “Central Processing Unit”.

HBW: “Hardness Brinell tungsten (Wolfram) carbide”.

SI: “International System of Units”.

HB: “Brinell scale hardness”.

HRA, HRB, HRC: “Rockwell hardness scale A, B and C”.

HV: “Vickers hardness”.

HLD, HLS, HLE: “Leeb hardness test”.

LR: “Linear Regression model”.

SND: “Standard Normal Distribution”.

GPR: “Gaussian Process Regression”.

MSE: “Mean squared error”.

RMSE: “Root-mean-square error”.

MAE: “Mean absolute error”.

TSS: “Total sum of squares”.

RF: “Random Forest”

RAM: “Random access memory”

GS: “Grid Search method”, it is a procedure that thoroughly explores a manually chosen portion of the targeted algorithm's hyperparameter space.

EDA: “Exploratory Data Analysis”

Notations

R_a : Arithmetical mean deviation of the surface profile.

R_q : Root mean squared of the surface profile.

σ_w : shear stress.

λ : film thickness.

η : viscosity.

ν : kinematic viscosity.

β : degree of change in the outcome variable inside a Machine Learning model.

1 Introduction

1.1 Motivations

Fluid power systems are employed in all industrial sectors, and the continuous demand for increased performance puts engineers ahead of new challenges. The latter systems utilise roughly three quadrillions BTUs (British Thermal Units), which are utilised for mobile, industrial, and aeronautical applications. They have a 21 per cent average efficiency [4]. Such hydraulic systems' efficiency can be increased, and energy consumption can be decreased through fluid optimisation. Nevertheless, selecting the proper hydraulic fluid system necessitates that the engineers assess several interrelated criteria, such as the size and kind of equipment and operational parameters like temperature, pressure, and maximum load. In order to attain the best balance, specific attributes must be traded off, such as reliability vs efficiency or mechanical efficiency versus volumetric efficiency.

Knowing which equipment component will break can significantly reduce machine stop time and spare parts cost, and maintenance costs can be highly reduced, leading to substantial savings for the companies. Furthermore, failures or undesired operating modes could lead to significant economic damage for machine and plant operators. For these reasons, condition monitoring (CM) concepts have become one of the primary focuses in science and industry.

Notwithstanding the excellent accessibility of measurement data and the high potential of CM in machine and maintenance management, it has not become popular in fluid power yet. The problems are due to the technology applications and the high effort involved. Since both healthy and damaged components must be considered, obtaining enough data to train a reliable condition monitoring system is challenging. Because it can be costly to damage healthy units and conduct several experiments, this technique is rarely financially viable. In order to overcome this issue, methods from the field of machine learning (ML) started being used. Instead of manually programming rules and parameters as required by conventional rule-based and model-based fault detection methods, ML algorithms identify fault patterns extracted from data sets through algorithmic optimisation.

The challenge in implementing ML-based approaches for automated fault detection is to generate sufficiently large training data sets. Obtaining such data is challenging since the machine must be deliberately driven to an operating point that may lead to severe damage.

Apart from requesting higher performances, the companies also face environmental problems, which lead to the need to find different materials and lubricants.

The following sections analyse the three most important aspects that make this work of paramount interest for companies:

1.1.1 Velocity and pressure

The trend in the last decades has been to increase the value of pressure and speed in the machines in which actuators are employed, e.g. hydraulic or axial pumps. This trend can be easily seen by comparing two pumps, one from the 80s and a modern one.

Pump name	A2V (Series 5)	P2060
Year of manufacture	1983	2012
Rated drive speed	1600 rpm	2800 rpm
Max. Outlet Pressure	250 bar	320 bar

Table 1.1: Comparison of datasheet [16] [17].

It is straightforward to notice the increase from this table constructed using the datasheet of the two pumps.

In parallel, the industrial tendency is to switch from diesel engines to electric ones to satisfy environmental requirements. Apart from that, electric engines allow one to have control of the speed. The electric linear actuator offers precision control and a long lifespan that requires minimal maintenance. Aside from being silent, clean, non-toxic, and energy efficient, electric actuator systems also emit almost no or no noise. These technologies meet the ever-increasing requirements for eco-friendly technology [5]. The awareness observed in the market, politics, and media has caused manufacturers of mobile working machines to engage in electric drive concepts. The key benefit of the electrification idea over the Diesel-Hydraulic driveline is the

improvement in energy efficiency and, subsequently, the decrease in energy consumption and fuel expenses.

Contrary to the original belief, there is not much mass difference between the two. Additionally, electrification offers a variety of alternatives for incorporating an energy storage medium to buffer energy and handle transient power peaks, as well as for intelligently transferring energy. A vital downside of the electrification approach is the high cost needed for the separate components. However, as the technology behind hybrid cars evolves, the pricing for these components will likely fall in the future, making this alternative more and more appealing.

Another important aspect related to the need for measures for a wide range of different speeds and pressures is the different operating modes present in a linear actuator.

A linear actuator is an apparatus that transforms rotational motion into push- or pull-linear motion. It enables the lifting, dropping, sliding, adjusting, and tilting of machinery or materials. Their primary advantages are the safety, cost, ease of installation, and durability of linear actuators. In this work, the focus is drawn on the control plate of the linear actuator. The two surfaces of a control plate pairing are separated from one another while in operation. From a tribological perspective, this behaviour improves performance because there is less friction since the two surfaces are not in contact.

On the other hand, there is less volume flow. Consequently, increasing the hydrodynamic balance decreases the volume flow rate. Therefore, the different operating modes also make a study with different pressures necessary. One exemplification of this issue is present in the control plate. As seen in the figure below, the elements mentioned above are not subjected to uniform pressure.

Despite an accurate choice of pressure, the tribometer cannot perfectly simulate the actual case. On the contrary, in red is highlighted the high-pressure distribution window and in green the low-pressure one. On the same part of the control plate, there is high and low pressure, and the pressure is set to 1MPa to approximate that pressure gradient in the tribometer. This

non-uniform distribution makes the pressure choice in the tribometer very delicate and essential.

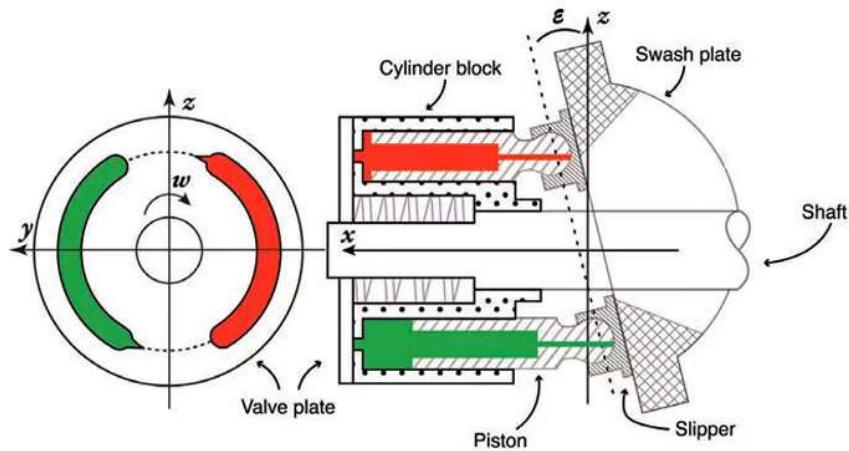


Figure 1.1: Axial pump [4].

1.1.2 Environment

For similar environmental reasons that have led to the transition from diesel engines to electric ones, materials and lubricants used in fluid power applications have to change.

Traditionally the most used material for control plates is lead and its alloys, which allow machines, e.g. pumps, to achieve high-quality performance. Despite this, lead can harm several vital bodily processes, leading to various symptoms. Lead is usually discovered when combined with two or more other elements, generating lead compounds. Processed and lead-containing items can release lead dust, fumes, or vapour. When eating, drinking, without first washing the hands and the face, and when exposed to lead dust, fumes, or vapour, the human body absorbs lead. Humans who consume lead may have various symptoms, including headaches, anaemia, and stomach pains. Long-term, unchecked exposure can result in more significant issues like kidney, nerve, and brain damage, as well as cancer.

Other metals like calcium, iron, and zinc are prevented from performing their usual functions in the body by lead. According to the Chemistry World homepage of the Royal Society of Chemistry, when lead replaces zinc, it leads to cells “around the body being short of oxygen, creating a cascade of linked diseases”. Additionally, it is a crucial concern for kids, especially with their developing nervous systems and brains.

Pb and PbO were recognised as dangerous to the environment and human health due to the Rio Declaration on Environment and Development by the United Nations. Based on these normative, international directives prohibiting using Pb and other harmful materials in consumer items have been developed.

In the past few decades, North America, Europe, and Asia have all seen a rise in the sale of lead-free materials, as highlighted by Figure 1.2. This trend has not stopped growing exponentially. On the contrary, it has become more pronounced.

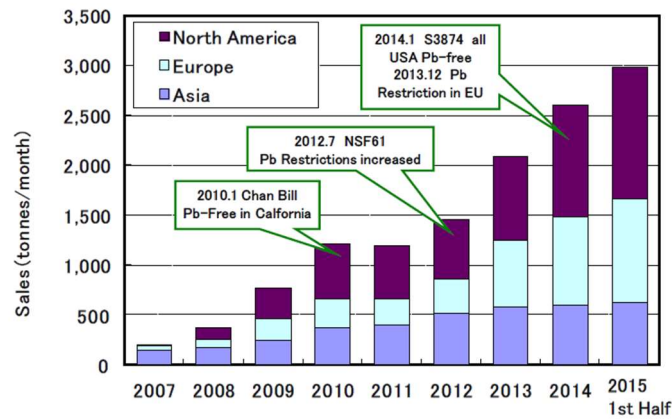


Figure 1.2: Growth of sales [7].

In terms of lubrication, ultra-low viscosity lubricants have been created to reduce shear forces in passenger cars and meet statutory standards for CO₂ emissions. The same reasoning can be applied to fluid power machines, in which saving energy is paramount. Despite this, a drop in the lubricant's viscosity will result in thinner oil films, making it more difficult for the oil to maintain a stable distance between the loaded contacts. The latter could be a result of a switch from a complete film to a mixed film lubrication regime, which could contribute to faster wear rates and localised friction.

ZDDPs are widely used as lubricant additives. Their capacity to minimise friction is mainly due to their better load support, positive mechanical characteristics from preferred shearing, and corrosion protection on the surface. Additives are utilised in high quantities to ensure the creation and endurance of tribo-layers.

When released into the environment, ZDDP is hazardous to aquatic organisms and may have long-lasting consequences if exposed to workers. It can also irritate the eyes and cause skin

corrosion [6]. Furthermore, ZDDP is being phased out or even replaced as a lubricant additive due to restrictions on phosphorus and sulphur in engine oils.

1.1.3 Surface finish

Since it is said to have a significant or little role, this is a controversial topic. Suppose the sliding conditions (applied load and sliding speed) are intense enough to provide a polishing effect during the run-in. In that case, the beginning roughness has no impact—the previously mentioned smoothing effect results in the same equilibrium roughness after a concise sliding duration. Initial roughness has an indirect impact if the sliding conditions are very mild since it may or may not cause transfer phenomena, which change the contact conditions and hence friction. The effect of roughness on friction coefficient, and thus the importance of a surface description in tribology, is described in more detail in Section 2.1.1.

It is not always better to have a more accurate finished surface, for some materials may be convenient to have a rough surface. The measurements carried out on the tribometer consider two types of finishing surfaces: lapped or fine-grinded. Lapping is a type of machining where two surfaces are rubbed together with an abrasive. Fine-grinding is an abrasive machining process that achieves flat and parallel workpiece surfaces by combining the speed and aggressivity of super abrasive wheels with the precision of lapping kinematics. Lapping produces a better surface than grinding on a microscale, yet on a nanoscale, grinding produces a better surface, according to specific research. Since the nanoscale is essential in fluid power applications, a lapping surface ensures better performance.

On the other hand, lapping costs more money than fine-grinding. As it can be seen from the results of this work, the tribological behaviour do not differ a lot between the two types of surface finishing. On the contrary, decreasing too much the surface finish could lead to even worse performance.

1.2 Goal

This thesis aims to build a Machine Learning model, particularly a Random Forest model. The latter has to predict the tribological behaviour of different pairs of materials that can be used in control plate pairings, one of the critical parts of a pump or a compressor. The model can save much time for the test bench (tribometer), doing new experiments directly through

it instead of using new test specimens. The model's inputs are constituted by the experimental setup, the control plate material, and the lubricant employed. From the measures collected through the tribometer, the most significant parameters are extracted. The latter constitutes the expected outputs for the Random Forest model. The tribometer measures the result by considering different cases in which the rotor and stator materials have been changed, the lubricant, surface finishing, and other parameters. The Stribeck curves have been constructed starting from these measurements. The essential features have been extracted using different algorithms, described in the following chapters.

1.3 Contributions

In the course of this thesis, the following contributions are made:

- Chapter 2 describes tribology's state of the art and its related fields. The chapter is divided into sections, in which every aspect of tribology has been treated more in-depth. After a general overview of tribology, the first macro argument to be analysed in section 2.1 is solid surface topography. In the subsequent section, friction, wear, and adhesion are studied. Finally, fluid film lubrication and test bench are explained.
- Chapter 3 shows the state of the art in Artificial Intelligence and Machine Learning. Section 3.2 focuses on the current use of these two sciences in the field of tribology. The last section explains the technique implemented in this thesis: Condition Monitoring.
- Chapter 4 presents the main work carried out in extracting all the features from the measurement data of a disc-on-disc tribometer. Furthermore, the construction and training phase of the machine learning model is reported in detail. Also, the theoretical aspects related to these procedures are indicated.
- Chapter 5 deeply discusses the results of the predictions and the associated conclusions.
- Finally, chapter 6 proposes an outlook concerning the future applicability of this work and other possibly related techniques.

2 Tribology state of the art

Tribology is present in practically all the acts of daily life. When friction occurs while attempting to move a heavy item over the floor, it may be felt as a force that resists the motion. Instead, as the shoe bottoms get thinner and thinner, it is an example of wear. Tribology is the investigation and implementation of the concepts of wear, lubrication, and friction on reciprocal movement between two surfaces. The term tribology originally derives from Greek and means the science of rubbing. The first tribology study was undertaken in ancient Egypt. A painting from approximately 1880 BC uncovered in a cave at El Bersheh portrays a colossal colossus being pulled by massive wooden rollers smeared in animal fat [1]. Tribology is essential for a variety of reasons, not the least of which is that it has the potential to reduce the impact of climate change. Several excellent studies by Kenneth Holmberg and Ali Erdemir demonstrate how tribology research might help us use less energy. According to their research, it is possible to cut the energy consumption of passenger cars, lorries, and buses by up to 60% over the next 20 years. The two scientists have determined that tribology can help reduce our present world energy usage by over 9%.

As stated before, the three main elements studied in tribology are friction, wear, and lubricant. Friction, whose coefficient could be kinetic or static, is the phenomenon that opposes motion. The friction that prevents a body from moving while it is already moving is known as kinetic friction. Instead, static friction is the resistance that keeps a motionless body from beginning to move.

Wear refers to the transfer or loss of material surfaces, and friction is the primary cause of this significant source of energy loss. Friction control may also be used for this purpose, resulting in significant energy savings.

A thin layer of gas, liquid, or solid is placed between the two contact surfaces to reduce the effects of wear and friction. Separating the contact surfaces and avoiding damage are the aims of lubrication. A lubricating layer thickness generally varies from 1 to 100 μm .

2.1 Solid surface topography

One of the critical aspects of tribology is surface description. The breakdown of lubricating layers and subsequent system failure might be brought on by inappropriate surface shape. As a result, surface accuracy standards have grown considerably over time. Typically, it is highly challenging to produce smooth surfaces. Hills and valleys are prominent compared to molecular size, even on carefully polished surfaces. Therefore, the surface comprises low and high spots known as “asperities”. Most solid surfaces, except for noble metals, are chemically reactive in conjunction with surface imperfections. Additionally, they are all composed of many layers with various physicochemical characteristics.

2.1.1 Geometric characteristics of solid surfaces

Figure 2.1 demonstrates how Hamrock, and his research group, divided the geometric features of surfaces, usually known as texture, into three broad groups [15]:

- *Error of form*: manufacturing mistakes produce surfaces that diverge from a specified pattern.
- *Waviness*: while processing, machine tool systems constantly undergo undesired vibration; consequently, rather lengthy waves in the surface profile might be formed.
- *Roughness*: in addition to incorrect shape and waviness, various anomalies are expected during the production processes of cutting and polishing.

Despite this, it is difficult to distinguish between these groups clearly. Usually, roughness is considered, which simply evaluates how long the horizontal gap is between surfaces. Nonetheless, to designate surfaces for tribological investigation, both the wavelength and the amplitude should be quantified.

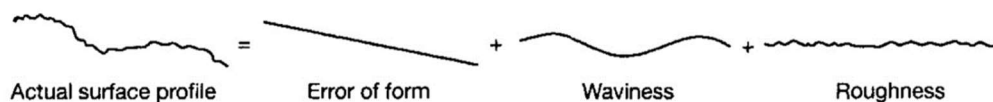


Figure 2.1: Geometric characteristics of solid surfaces [15].

2.1.2 Analysis of surface roughness

The inherent abnormalities in the manufacturing process are referred to as surface roughness (e.g., cutting tool or abrasive grit). It can result from vibrations, material deflections, or

stresses during usage. Surface metrology commonly views roughness as a measured surface's high-frequency, short-wavelength component. In tribology, rough surfaces frequently have more substantial friction coefficients and wear more rapidly than smooth surfaces. Since surface irregularities may serve as initiation locations for cracks or corrosion, roughness is frequently a reliable indicator of how well a mechanical component will operate. When a part is machined, a particle is released during the operation, producing a tiny groove on the component. These grooves cause the roughness that the tool creates as it moves across the component. How the material is separated from the solid substance determines the surface pattern of each groove.

Even though a high surface roughness is typically undesired, it may be challenging and expensive to control its value in production. A component's production cost typically increases as its roughness is reduced, and the latter often results in a trade-off between a product's production cost and application performance.

It is possible to calculate various surface properties using z_i , $i = 1, 2, \dots N$, the surface height, and N , the number of measured points. For example, the most popular average roughness parameters are [15]:

- Mathematical mean:

$$R_a = \frac{1}{N} \sum_{i=1}^N |z_i| \quad (2.1)$$

- Quadratic mean (or RMS):

$$R_q = \sqrt{\frac{1}{N} \sum_{i=1}^N z_i^2} \quad (2.2)$$

- Variation of the maximum peak-to-valley length:

$$R_t = \max(z) - \min(z) \quad (2.3)$$

The figures below depict the three most common ways to compute surface roughness:

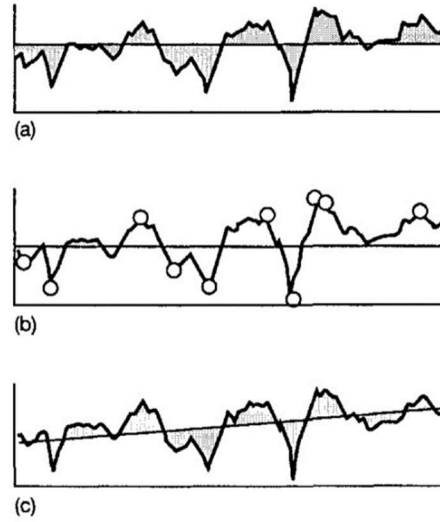


Figure 2.2: Comparison of three types of reference line: (a) mean; (b) ten-point average; (c) least squares [15].

The most used method is, by far, the arithmetic average (R_a). However, it is highly stable statistically and even suitable for assessing random surfaces, such as those created by grinding. On the other hand, R_a is not very useful when applying sealing surfaces or as a discriminator for different types of surfaces because it cannot differentiate between peaks and valleys. The roughness average parameter's most significant advantage is that even the smallest surface finish measurement devices have it because it is the most widely utilised. In practical terms, a patch with a real R_a value is included with most of the reference specimens offered with these instruments. Therefore, it is crucial to select a cut-off (or sampling) length adequate for the procedure while carrying out a R_a measurement. In order to prevent waviness from affecting the measurement, the cut-off length should be as short as possible. Nonetheless, it also should not be so short that only a tiny area of a tool mark is measured. The cut-off should be long enough to accommodate five complete sets of tool marks to get an accurate average roughness measurement.

2.1.3 Adhesion

It is defined as the molecular attraction that holds two different bodies together when they come into touch. Two primary causes of friction have been determined from the classic textbook of Bowden and Tabor. The first is related to the adhesion of the contacting asperities, while the second is related to the asperity or plastic deformation of the bulk surface. When

identical metals slide past each other without surface contamination, adhesion contributes the most to the friction coefficient [14].

A limited normal force, known as the adhesive force, is needed to separate two solid surfaces when they come into contact to prevent adhesion or bonding across the interface. Both solid-solid interactions and two solids separated by liquids or sticky solids exhibit adhesive properties. Of course, adhesion and bonding are often stronger if the surfaces are free of any lubricant coatings. It is a complicated phenomenon that considers various factors, including temperature, roughness, typical load, duration of contact, and chemical interaction between surfaces and the environment. The standard stress distribution W' required for detachment to the orthogonal compression force W initially applied is typically used to define the adhesion coefficient:

$$\mu' = \frac{W'}{W} \quad (2.4)$$

2.2 Friction and Wear

2.2.1 Friction

A group of interfacial mechanisms known as friction generates an exceptional resistance to sliding when they act on relatively moving surfaces. The friction mechanism generates the functioning of a friction process. This physical phenomenon can be analysed on a first-principles level and operates in concert with other phenomena. Furthermore, friction depends on the properties of the materials rubbing against each other and is not an intrinsic quality of the materials.

All mechanical elements in relative motion with one another wear down and give rise to energy loss due to friction. For instance, resisting frictional forces in moving parts requires around 20% of an automobile's engine power. Notwithstanding this, friction is not necessarily a negative phenomenon because walking, picking up, or operating a car would be hard without it.

The human body and its parts both experience frictional phenomena. For example, skin friction plays a significant role in sports. NBA players expressed 2006 dissatisfaction with

basketball's newly adopted composite surfaces, and they were perceived to be both too slippery and too sticky when wet.

When two solid bodies come into contact, a resistive tangential force (F) - whose direction is opposing that of motion - is produced. This force is known as the friction force, a nonconservative force, i.e. the effort to move an object depends on the route followed.

$$F = \mu \cdot W \quad (2.5)$$

Where W : normal load acting between both surfaces.

In other words, the friction coefficient is the ratio between the force preventing tangential motion between two bodies and the force that ordinarily pushes the bodies together [2]. The first of the two common laws concerning friction is this. The second one claims that friction is independent of the location of contact. This statement can be demonstrated through a simple experiment. The latter consists of dragging three components with different shapes, but each of these is of the same material and weight, and in each case, the force applied is almost the same. It turns out that a larger area of contact does not produce more frictional force. The stress between the contact objects reduces as the region of contact grows. The effect of the increase in friction-causing area is precisely neutralised by the reduction in pressure between the two surfaces. Less pressure implies weak bonds between the molecules at the contact surfaces, while more pressure implies the opposite. Hence, friction does not rely on the degree of surface area in contact between the two entities. In most technical applications, these two conventional criteria are followed, but not when sliding occurs at high speeds or under various standard stresses.



Figure 2.3: Commercial static friction test bench [1].

A practical method for determining a material's macroscopic scale static friction coefficient is to place it on a surface and tilt it gradually until relative motion begins. A stiff beam is pivoted at its left end, to which a flat strip of an intriguing material is connected. The inclination angle (θ) is measured using a protractor or other tool. It is possible to derive a different mathematical formula to express the coefficient of friction using this measurement technique.

When $F_s = P$, as motion approaches, the static friction coefficient is defined as follows:

$$\mu_s = \frac{F_s}{N} = \frac{W \cdot \sin \theta_s}{W \cdot \cos \theta_s} = \tan \theta_s \quad (2.6)$$

Under these circumstances, $\alpha = \theta_s$ has termed the angle of friction or rest [1].

2.2.1.1 Types of friction

1. Static friction:

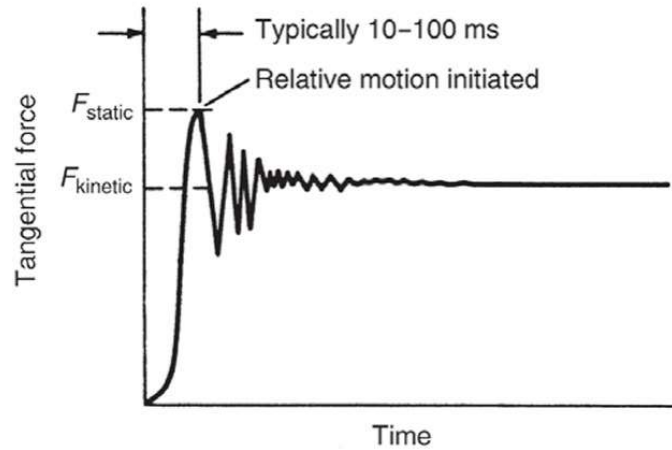


Figure 2.4: Static friction force and kinetic friction force as a function of time [3].

It is the force between surfaces that are at rest against one another. When a force is applied to an object, the frictional force opposes it, keeping the object at rest until the static friction force is overcome. There is static friction between zero and the smallest force necessary to start motion. The standard load W and the static friction force (F_s) are linearly related. Therefore, the static friction coefficient (μ_s) serves as a measure of proportionality:

$$F_s = \mu_s \cdot W \quad (2.7)$$

The coefficient of static friction is evaluated under a specific load, time, and other environmental conditions. If these variables change, then this coefficient will change

as well. For instance, the value μ_0 will be lower in a contaminated environment where adsorbed gases are present as opposed to one with clean surfaces and no reactive gases [3]. Additionally, research shows that the duration of contact between the two bodies impacts the coefficient of static friction. Two distinct scenarios are possible:

- the friction decreases when species with lower shear strengths contaminate the interface;
- the interfacial connection strengthens as a result of the clean contact. In this situation, the friction will probably increase.

2. Sliding friction:

Kinetic friction - also called dynamic friction, sliding friction, or sliding resistance - occurs when objects move against or across a surface. Sliding friction is often weaker than static friction because it is easier to maintain motion over a horizontal surface than starting motion from rest. In this situation, the frictional force indicated as F_k prevents an object from moving. Its formula is:

$$F_k = \mu_k \cdot W \quad (2.8)$$

Where μ_k is the dynamic friction and W is the orthogonal force. Analysing the asperity interactions between rigid and soft materials can help define the transition from static to sliding contact, Figure 2.5. Three different stages of contact could be distinguished: static contact, or the point right before the asperity moves in its entirety; when the tangential force is at its highest level; and unfettered movement of the asperity. Since the tangential force is initially small, both sides of the hard asperity are supported. The asperity dips into the softer material when the force is increased until it reaches the critical value, which causes the flank on which it operates to become unloaded. Thus, the maximum amount of tangential force that resists motion is produced by the growing contact area brought on by the additional depth. The tangential component decreases when the asperity moves further than the threshold of stationary contact, stretching the soft material sufficiently to provide enough stability [2].

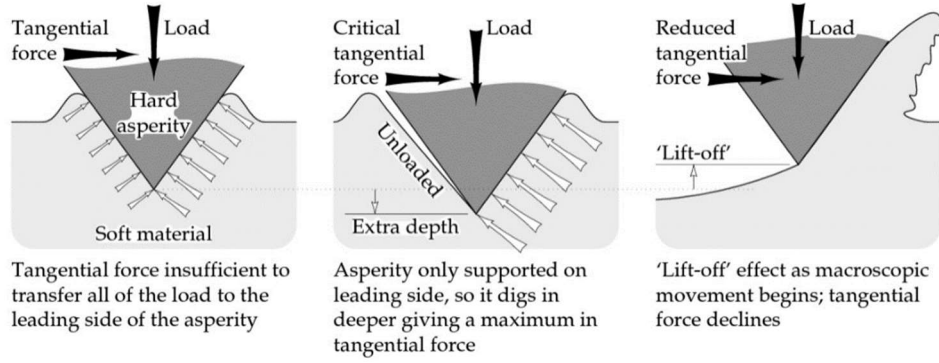


Figure 2.5: Schematic illustration of the transition from static contact to sliding contact for a hard asperity on a soft surface [2].

3. Rolling friction:

The force that prevents a wheel or other circular item from rolling along a surface is known as rolling resistance, sometimes known as rolling friction. Deformations bring on this force in the object or surface. Because it is not the consequence of two surfaces in contact pushing against one another, it is not correctly friction. It is essentially a resistive force caused by non-elastic processes linked to the energy losses incurred during surface deformation and distortion of the rolling object [3]. In other words, when the pressure is removed, not all of the energy needed for the tyre or asphalt pavement to deform (or move) is recovered. Therefore, it has a much lower value than static and sliding friction. Ball bearings are built using the theory of rolling friction:

$$F = C_{rr} \cdot N \quad (2.9)$$

Where F refers to the rotational resistance load, C_{rr} refers to the dimensionless coefficient of rotational resistance, and N refers to the average load. C_{rr} is defined as the amount of force required to propel (or tow) a wheeled vehicle forward at a constant pace on a level floor with no gradient and no air resistance.

4. Fluid friction:

Friction causes fluid layers to move around one another or affects things moving through the fluid. Liquids, gases, and other substances that can flow and assume the shape of their receptacles are collectively referred to as fluids. The forces can appear among the molecules of the liquid or between the substance and another material with which it comes into contact [2]. The internal resistance to flow is named viscous drag,

or more commonly viscosity, which is deeply described in section 4.7.2. The two forms of forces between fluid molecules are as follows:

1. Cohesive forces are the attraction forces that bind molecules of the same material together. Therefore, they are the primary reason for viscosity.
2. Adhesive forces are those that attract molecules that are dissimilar to one another.

An object must use energy to overcome fluid friction as it moves through a fluid. When fluid flows beside an item, its shape significantly affects whether the fluid friction increases or decreases. The dynamic friction factor may be expressed mathematically as:

$$f = \frac{\sigma_w}{\rho \cdot \frac{\bar{u}^2}{2}} \quad (2.10)$$

Where σ_w is the shear stress at the wall, \bar{u} represents the average velocity.

Two primary types of friction are used in engineering applications: dry and fluid. In the first, nothing is placed in between the two dry objects when they come into touch. When liquid is injected between the two surfaces, fluid friction develops; in this circumstance, the friction will be decreased.

2.2.1.2 Phenomena that affect friction

The five essential phenomena affecting friction are hardness, load and contact pressure; temperature and lubrication; kinematics; surface finish. Each of the effects is clearly explained in the following paragraphs.

It is crucial to note that because most friction experiment parameters are interconnected, changing just one at a time can be challenging. For instance, if the average load increases, the wear rate may also rise, changing the surface's roughness and bringing debris into the interface. Therefore, when graphing either frictional resistance or friction coefficients vs a single test variable, the concurrent variations in other parameters may be obscured within those findings. Contrary to what one might anticipate, friction is unrelated to the contact region, as was stated at the beginning of section 2.1.

1 Effect of hardness

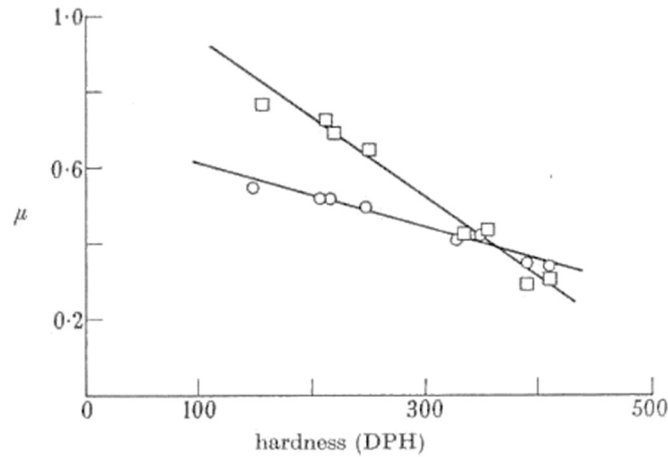


Figure 2.6: Relation between friction and hardness (square: hard steel slider; circle: diamond slider) [18].

The degree of friction is determined by how hard or soft the materials rubbing against one another are. Harder metals generally have less friction; this decrease in friction is linearly correlated with the hardness. The soft metal expands and contracts more plastically at contact points for almost the same load [18]. It is hypothesised that the amount of plastic deformation and, consequently, the substrate hardness determines how easily the oxide coating is disrupted. For softer materials, the corresponding frictional force is more prominent. Additionally, friction is the greatest when the hardness of the contact objects is precisely the same.

2 Effects of load and contact pressure

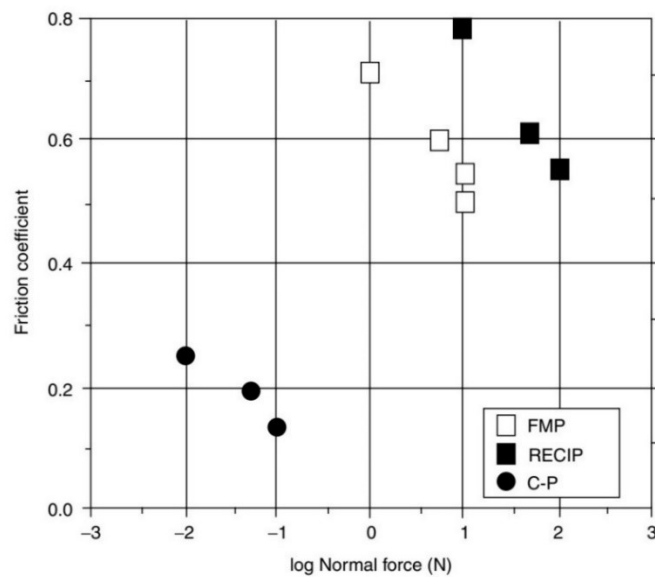


Figure 2.7: Friction coefficient for the same sliding materials at various normal loads [1].

The experiment's results, which comprised rubbing identical materials together at three weights on three separate test rigs, are summarised in the figure above. A total of four orders of magnitude's worth of loads were investigated. The friction coefficient reduces with a higher average load for each set of experiments, although the causes of this behaviour vary for each one [1]. Hertzian contact considerations are relevant in the situation of the lowest load. As a result, the three examples' friction-load dependency slopes are different. This experiment demonstrates that the correlations among friction coefficient, contact pressure, and load cannot be generalised. The energy used for work or to generate heat varies as the load rises and the frictional forces change proportionately. Nevertheless, another approach to varying the energy that a sliding surface holds is to alter the sliding velocity.

3 Effects of lubricant and temperature

➤ Lubrication:

Any material can influence the friction between two moving surfaces. Frictional behaviour is significantly influenced by film surface and chemical environments that result in the formation and replenishment of lubricant films. The equilibrium between two friction-causing mechanisms determined how efficient the lubrication is [15]:

1. Adhesion between the touching surfaces.
2. The tool cutting zone roughens asperities on the workpiece. Because of the instrument's high surface roughness, most ploughing resistance in the total frictional resistance is released.

More details about lubrication are given in section 2.3. Low viscosity oil does not allow the hydrodynamic lubrication layer to form at low speed. As a result, a mixed lubrication regime with increased friction is used for sliding surfaces. Due to their viscosity, medium and high viscous oils can generate hydrodynamic lubricating films at moderate speeds and experience less friction.

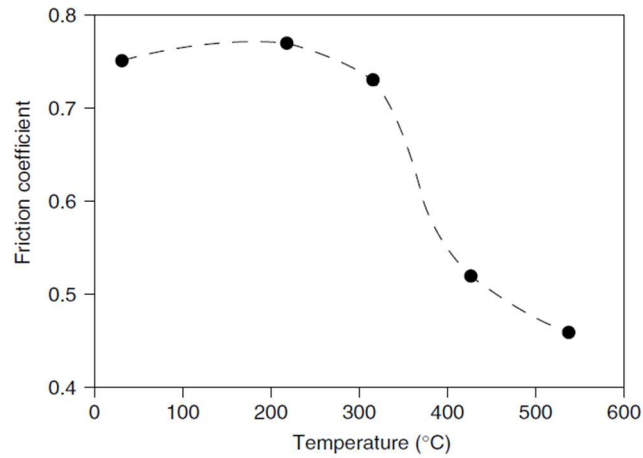
➤ Temperature:

Figure 2.8: Data for a cast Inconel alloy pin sliding unlubricated on an M-10 tool steel disk [1].

The characteristics of the materials at and close to a sliding interface are impacted by thermal energy. In tribological interfaces, thermal energy is primarily derived from internal (from frictional heating) and external (from the environment temperature).

The following are the leading causes of temperature's effects on friction [1]:

- the materials' shear resistance at interfaces
- the viscosity of the lubricant
- lubricants' inclination to alter their chemical elements
- a metal or alloy's ability to alter its composition and properties.

Many technical alloys have sliding friction coefficients that tend to reduce as temperature rises, and this behaviour frequently signifies a decline in the material's shear strength. According to the current theory, thermal expansion causes the intermolecular distance to grow as temperature rises. As a result, the surface becomes smoother, which reduces frictional force.

4 Kinematics of the surface contact

Frictional work is produced by the friction force F operating across a distance x . The amount of distance moved in a given time determines the energy input rate. The ability to calculate things like the temperature rise caused by sliding is made possible by the product of friction force and velocity. In every situation, friction tends to decrease as velocity rises. Since most materials lose some of their tensile strength at

high frictional temperatures, lubrication oxide forms. The melted liquid may moisturise the asperity connections when the layer breaks from rubbing. However, it is also possible to acquire different friction-velocity correlations. In some circumstances, raising the velocity also increases the rate of wear, which in turn causes the surface to become rougher and causes more friction. Therefore, it is advisable to consider the individual components and technologies involved before concluding that friction always reduces as speed increases.

According to the image below, if the speed range is enlarged, there may be a substantial impact (case "b"); on the other hand, there could be no effect within a very narrow range of speed (case "a"). In some cases, when the temperature controls the relative importance of friction-governing processes, μ versus v might undergo complicated modifications [1].

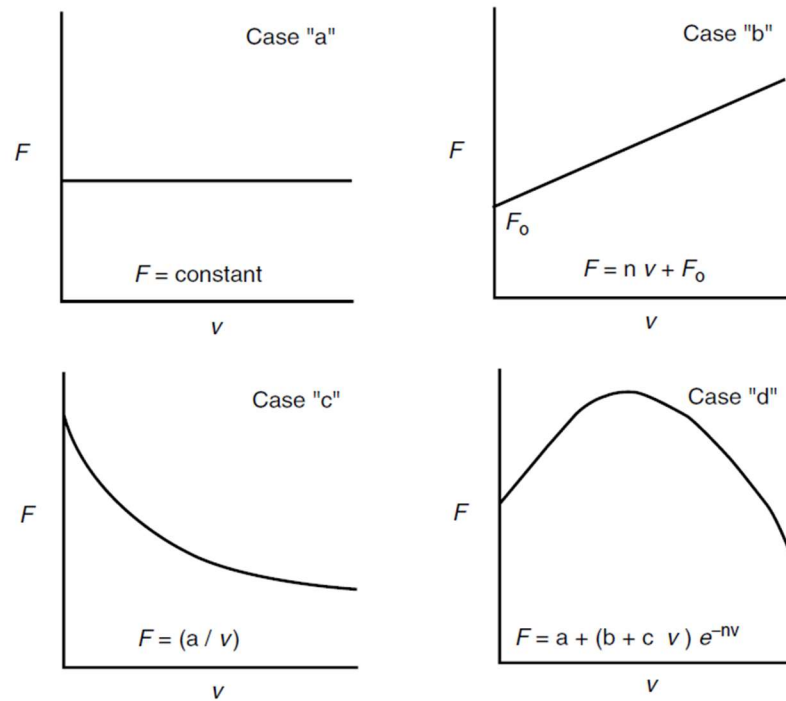


Figure 2.9: Four possible relationship between friction force and sliding velocity [1].

The sliding velocity was expected to match the mating surfaces in the preceding discussion. The implications of directional changes and sense must also be considered because velocity is a vector quantity. When choosing friction testing, it is vital to consider the direction of motion, and choosing a suitable motion is crucial when selecting appropriate materials for specific technical applications.

5 Effect of surface topography on friction

As stated in chapter 1.1.3, the effect of surface finishing on friction is a controversial point. In this context, friction data involving copper moving on copper is frequently used as an example to show how roughness affects the friction coefficient. The experimental results are highly dispersed, but at $Ra < 0.4 \mu\text{m}$, they can be efficiently grouped around $\mu = 1$ (a typical value for copper sliding against itself). The latter consideration is valid despite the roughness value. Only for $Ra > 0.7 \mu\text{m}$, which is a reasonably high degree of roughness, a drop in μ is observed [19]. Given the significant surface plastic deflection or wear processes promoted by the high beginning roughness, it may be hypothesised that this decrease is connected to the complexity for the mated surfaces to attain while slipping, a low stability roughness. If the applied load is mild, the interacting components are reasonably complex, or a thick oil coating is present, the original structure and morphology may be retained for a lengthy sliding time and may continue to affect friction.

2.2.2 Wear

Wear is the interplay of asperities that results in material being destroyed or eliminated from a solid surface. The material is physically eliminated when it melts, dissolves chemically, or fractures at the mating surfaces. The predominant wear pattern may switch from one to the other for several reasons, including modifications in the characteristics of the surface material and interfacial surface reactions driven along by frictional warming and chemical film development. Wear is a framework event, not a material property that must be emphasised. A tribosystem comprises dynamic characteristics, environmental variables, and material properties - even with very slight adjustments, suffers considerable wear change [13]. Future technologies will depend highly on machinery control because all machines experience wear and deterioration over time. Wear coefficient (dimensionless):

$$K = \frac{(\text{Wear volume}) \cdot (\text{Hardness})}{(\text{Normal load}) \cdot (\text{Sliding distance})} = 3 VH NS \quad (2.11)$$

Considerations of intricate alterations during friction are used to describe wear mechanisms. Acknowledging each wear behaviour is crucial since, generally, wear is not caused by a single wear mechanism.

2.2.2.1 Types of wear

The fundamental and significant wear modes are shown in the figure below:

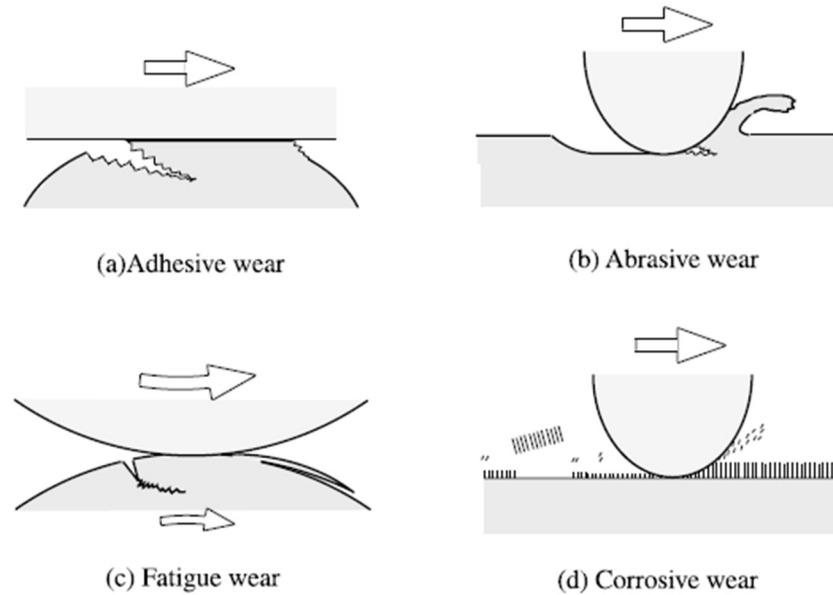


Figure 2.10: Schematic images of four representative wear modes [13].

It is essential to specify that there exists another specific type of wear, known as fretting, associated with a low amplitude oscillatory sliding between bodies. The latter, together with adhesive wear, is the primary cause of peaks in the Stribeck curves. Its causes and consequences are explained in section 4.6.5.1.

(a) Adhesive wear

A strong adhesive bonding reaction is present as long as the sliding of similar materials occurs. Adhesive wear arises when the fracture is believed to have been predominantly produced by solid adhesion at the contact surface. It arises when the strength of the surrounding region is greater than the strength of the atomic bonding forces between the materials. Adhesive wear is one of the most frequently encountered types of wear. Heavy plastic distortion carried on by dislocation is created during this kind of wear in the interface region while being crushed and sheared. Due to such significant distortion in the boundary region, a crack starts and develops in the compound formation of cracks and sheared. A wear particle forms when the fracture approaches

the contact surface and adhesive transfer is finished. Shearing may occur at the elements' lowest point or the original surface. No wear occurs when the interface adhesion strength in the first scenario is lower than the cracking strength of neighbouring local areas, as shown in Figure 2.10 (1). In the alternative situation, a tiny amount of substance inside of one of the two elements cracks, and a small chunk, represented by the dashed line in Figure 2.10 (2), may adhere to the other material as a result of adhesion. The wear particle may either remain at the interface and produce a “transfer film” because of the strong bonding on the contact or become a scattered wear particle. Due to the variation of loads imposed on a component, the latter's formation is plausible.

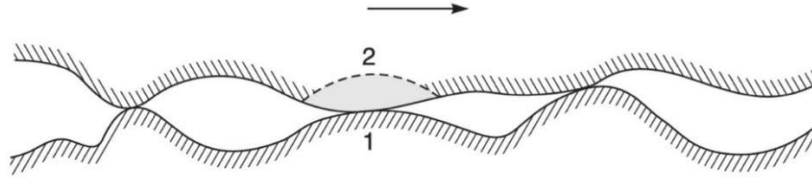


Figure 2.11: Schematic illustration of two break possibilities: at the interface (1) or at the weakest region of one body (2) [3].

An equation based on actual data of various unlubricated materials describes the adhesion wear. For example, the segmented wear volume could be expressed as follows:

$$v = \frac{k \cdot W \cdot x}{H} \quad (2.12)$$

Where the nondimensional wear coefficient (k) is dependent on the cleanliness of the materials in contact. W is the imposed force, x is the moving range, and H is the hardness of the degraded substance [3].

(b) Abrasive wear

Abrasive wear and material deterioration happen when a hard surface or particles rub against a soft surface. Ploughing happens when the interaction border among contact objects interlocks with an inclined or twisted contact. In this instance, chip formation, brittle fracture, or fatigue are the primary causes of material loss. The weaker surface develops an abrasive groove from ploughing, removing a significant quantity of surface material.

This effect occurs when a hard, sharp object and a relatively soft material come into plastic contact. When that occurs, the harder component penetrates, the weaker one. The indented material causes the fracture via micro-cutting, establishing the interlocking contact pattern required for cutting. Besides adhesive wear, abrasive one does not explicitly address fracture mod and adhesive forces.

Two kinds of abrasive wear - two-body and three-body - may be separated. Grit or hard particles extract components from the substrate surface in the two-body. Contrarily, the grains in three-body erosion are unrestrained and may roll and glide over a surface. This type must compete with other mechanisms like adhesive wear, which makes it ten times slower.

Various deformation models may be applied when abrasion removes material from a surface. Figure 2.11 displays the three most typical ones: ploughing, wedge formation, and cutting. A system's operation may involve switching between modes, and multiple of them may operate at once. A series of grooves are made throughout the ploughing process in Figure 2.11(a) due to the softer material's plastic flow. Even though different fatigue processes may arise and some material can be lost after repeated ploughing operations, there is typically minimal material loss as it is relocated from a groove to the sidewalls. The rigid material's tip generates a groove during the wedge creation process, Figure 2.11(b), which is identical to the ploughing mechanism except that a wedge originates toward the front of the groove. Therefore, a piece of the distorted component creates a wedge, whereas the rest is pushed to the side. Figure 2.11(c) shows how the hard material tip removes material as chips or disconnected debris particles by ploughing a groove with a sharp attack angle [3].

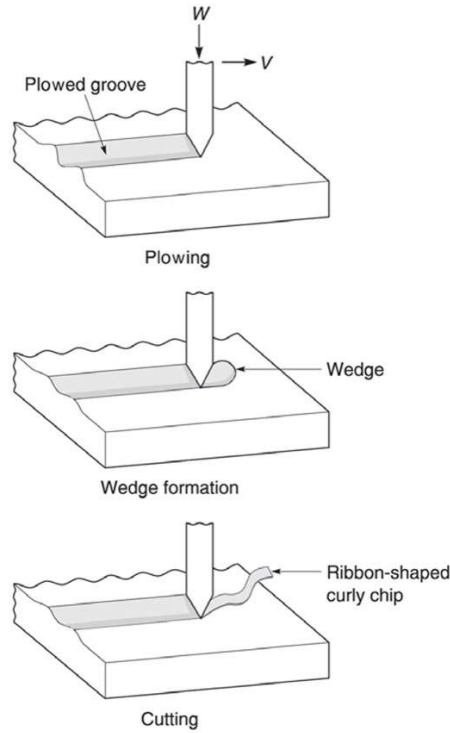


Figure 2.12: Schematics of abrasive wear processes as a result of plastic deformation [3].

(c) Fatigue wear

Several cycles are required to produce debris in the case of fatigue wear. Surface and subsurface cracks may form due to the fatigue process in metals. After a certain number of cycles, they may become severe enough to cause considerable fragments to break off the surface.

Under these circumstances, wear is governed by crack initiation, propagation, and fracture mechanics. Compared to unworn surfaces, worn ones have incredibly high amounts of plastic strain. The material's microstructure is altered due to this strain, which significantly impacts the wear processes. Achieving perfect lubrication, in which interaction between solids is entirely avoided, is challenging or uneconomical. Even if solid contact only happens occasionally, there may be enough surface damage to start contact fatigue.

High-cycle and low-cycle fatigue are two distinct processes of fatigue wear [13]. High-cycle fatigue has an extended component life since many cycles before breakdown occurs. Even if the macroscopic contact is technically in the elastic regime, the fractures in this scenario are caused by pre-existing micro defects in the material close to where the local stress may surpass the yield value. Low-cycle fatigue causes a

component to break quickly with a few cycles before failure. In this condition, plasticity is induced after every cycle, and the wear particle develops after many cycles have been accumulated. In the initial cycles, only shallow grooves caused by plastic deformation are created rather than wear particles.

(d) Corrosive wear

Both wear and corrosion wear mechanisms are involved when a material deteriorates due to corrosion wear. Both wear and corrosion actions can potentially cause significant material loss or damage. When facing both systems simultaneously, the consequences could be more severe. When sliding occurs, especially in corrosive liquids or gases, reaction products are typically generated on the surface through chemical or electrochemical exchanges. The wear process is virtually comparable to the specimen if these reaction products have a high interfacial bonding and behave like bulk material. The reaction products produced when solid materials interact with the corrosive medium cause wear, which typically differs substantially from the composite [13]. The periodic film formation and elimination by corrosive and abrasive action, respectively, are the foundation of the corrosive wear model. The pattern of material loss resembles that of corrosive wear. The tribo-chemical process at the contact surface is sped up when two objects come into touch with corrosive media. The primary corrosive substance in the atmosphere is oxygen, and oxidative wear of metallic materials in the air is the word most usually used to describe this process. The presence of oxygen in the oil is a common cause of corrosive wear. It has been demonstrated that iron oxide makes up most of the wear debris for steel lubricated contacts and that removing oxygen from the oil eliminates wear.

2.3 Lubrication

One of the most significant tribological phenomena is lubrication. It is defined as any substance placed between two surfaces moving relative to one another to minimise friction and wear. Lubricants can transmit power, protect, seal, and remove heat and wear debris depending on their composition. Without lubrication, parts may rub against one another, creating heat and local welding that, if not regulated, will lead to component failure. A

lubricant's primary function is to prevent contact between the moving or sliding surfaces, which reduces friction and the material damage it causes - adding specific additives, which may affect either some physical characteristics or some elemental composition, enhance the quality of the lubricant.

There are primarily two types of fluid films: the first occurs when fluid is injected between mating surfaces, and enough pressure is generated to separate them. Because the two surfaces in this type of lubrication are apart, wear and friction are often relatively minimal. In the other type of lubrication, the lubricant forms a weak layer that may support the load by adhering to the surface.

The distance between bodies is known as the layer width h , and it helps to lessen friction by minimising early mechanical breakdown of the materials. The h_{\min} and the equivalent objects' degree of roughness (σ), which is the geometric mean of the roughness R_a of the two surfaces, are what determine if a film is present [1]:

$$\sigma = [R_{q1}^2 + R_{q2}^2]^{\frac{1}{2}} \quad (2.13)$$

Specific thickness, often known as the lambda ratio λ , is a dimensionless parameter determined from h_{\min} and σ . It is described as the ratio [1]:

$$\lambda = \frac{h_{\min}}{\sigma} \quad (2.14)$$

Here R_{qi} is the standard error of the i -th layer in contact's surface heights. Typically, the lubrication schedules are chosen based on λ as follows:

1. $\lambda \geq 3$: Hydrodynamic lubrication
2. $0.1 < \lambda < 3$: Mixed lubrication
3. $\lambda \leq 0.1$: Boundary lubrication

All the different lubrication regimes are explained in section 4.2 related to the Stribeck curve.

2.3.1 Types of lubricants

Based on their state, lubricants may be divided up into the following categories [15]:

- All liquid lubricants, such as mineral, natural, synthetic, and emulsion oils, are referred to as lubricating oils

- Grease, a form of semi-solid solid lubricant, is frequently comprised of a detergent emulsion with natural resource
- All lubricants in solid forms, such as powders, coatings, and composites, are referred to as solid lubricants (graphite, graphene, molybdenum disulfide)
- Air is the most common type of gaseous lubricant, but any gas can be used.

Liquid lubricant is the most popular form. Lubricating films need to sustain the pressure across the contact surface, keep them apart, and lower the opposition to moving or rolling at the junction. This approach is based on fluid mechanics, such as viscosity connection to temperature and pressure.

2.3.2 Properties of liquid lubricants

- Viscosity: Viscosity is a characteristic of liquids that allows them to oppose their flow.
- Density: a way to gauge how tightly packed the bulk of an object or material is.
- Flash and Fire points: the coldest value at which lubricating oil produces sufficient vapours to burn up when a small flame is introduced close to it is known as the flash point. The fire point, on the other hand, is the coldest value that allows the vapours of the lubricating oil to ignite for a minimum of five seconds when a small flame is present. A suitable lubricant must have a flash point higher than the operating temperature, and the latter prevents the possibility of fire during lubricant use.
- Cloud and Pour point: the cloud point refers to the temperature at which, after being gradually cooled, lubricating oil starts to appear cloudy or hazy. The pour point, instead, is described in section 4.7.3. Low pour points are necessary for lubrication oils used in machinery that operates at lower temperatures; otherwise, the machine will jam up as the lubricant oil solidifies.
- Aniline point: It is described as the lowest equilibrium solution temperature for samples of equal quantities of lubricating oil and aniline. It suggests that the lubricant oil in touch with the rubber sealing package may be degrading.
- Corrosion stability: It is calculated by performing a corrosion test in which a polished copper strip is submerged in lubricating oil for a predetermined time at a predetermined temperature. A good lubricant should not impact the corrosion strip.

The most important properties of liquid lubricants are viscosity, density, and pour point, which are discussed in detail in section 4.7.

One of the most critical challenges of lubricants is using environmentally friendly liquids, as has been more discussed explicitly in chapter 1.1.2.

2.4 Friction test benches

All friction-testing tools share the following four characteristics, they all have a way to (1) sustain or perhaps connect the specimens for which frictional measurements are needed, (2) adjust the entities to one another, (3) impose average load, and (4) quantify or deduce the intensity of the radial friction factor that precludes relative movement.

Listed below are five factors that determine how friction test techniques and technology are considered [1]:

1. Macro-contact stiffness, geometry, and damping
2. The relative motion's type and magnitude
3. How much normal force was exerted on the contact
4. Environmental or temperature control needed
5. Conditioning and material preparation

It is exceedingly challenging to change just one element while keeping everything else constant when designing friction tests. The frictional temperature rises with increasing sliding velocity, possibly promoting tribo-chemical processes. As a result, altering velocity also alters temperature and possibly the kinetics of surface layer production.

The test bench's configuration and working circumstances vary depending on the various types of friction. The test bench for static friction has already been described in section 2.2.

The operating conditions of sliding (kinetic) friction tribometers are much more varied than those of static friction test benches, which measure resistance to approaching motion. These circumstances vary from steady, straight sliding to interchangeable machines, which impose a sophisticated combination of accelerations, decelerations, and direction changes. Six common types of tribometers are listed in the figure below:

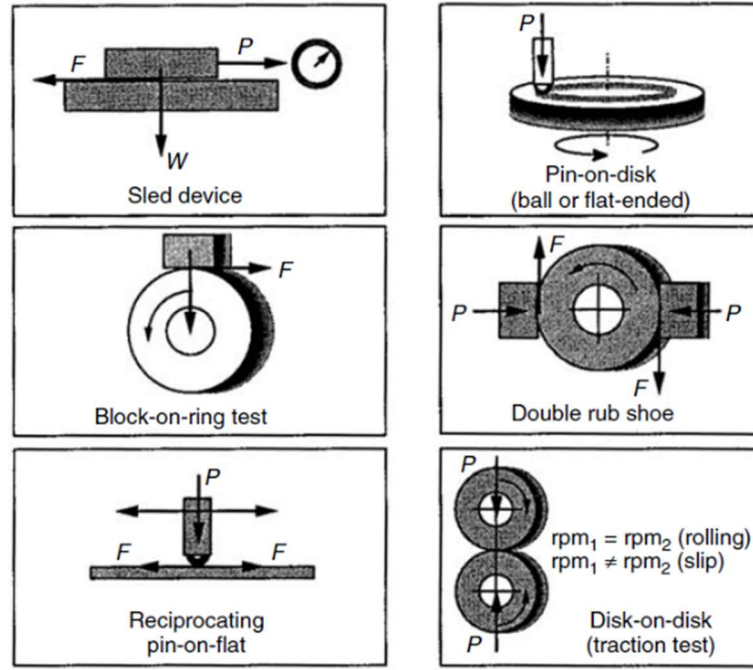


Figure 2.13: Types of friction-testing implementations [1].

Except for the sled test setup, these techniques use non-conformal interaction.

Tribometers can be categorised according to the contact's macro geometric conformity. When two conformal surfaces are fitted together, wear does not significantly alter the visible contact area. Notwithstanding, non-conformal surfaces are ones where the centres of bending are positioned on opposite sides of the interface.

Three advantages of conformal testing include the following: (1) the nominal area of contact is not impacted by wear; (2) the stress distribution is usually more uniform; and (3) the level of lubrication may be better maintained since the film thickness is typically consistent across a broader region. One of the critical drawbacks of conformal contact examination is that coordinating the sample surfaces may sometimes be challenging.

Since the contact is a point of tangency, perfect alignment and coplanar alignment of the fitted or composite surfaces are not needed for non-conformal testing, such as the disk-on-disk test, which is one of their primary benefits. Because rolling friction often involves some degree of the slide, rolling friction devices may range from uncomplicated disk-on-disk machines to more sophisticated multi-axis machines that can modify the quantity and direction of slip and the geometry of the elastic contact patch.

2.4.1 Laboratory test bench: tribometer

Inside the IFAS laboratory, a disc-on-disc tribometer is the test bench utilised for the experiments, as shown in Figure 2.13. When two different discs interact, a hydraulic piston undergoes a conventional force on them. The bottom disc is then placed in movement by a hydraulic system while the top disc is stationary, as illustrated in Figure 2.14. A pump transports the oil between two plates into a heat transfer linked to a refrigeration system. The friction force is examined using a stress detector positioned at a certain distance and detects the tangential force. The friction resistance is multiplied by the distance from the centre to calculate the torque. The technical details are presented in the following table:

Mean friction diameter	54 mm
Wear track size	5 mm
Area of friction	848,23 mm ²
Typical load	10 MPa
Maximum load	25 MPa
Range of friction coefficient, μ	0 – 0,1
Maximum surface speed	6 m/s

Table 2.1: Technical specification of the test bench.



Figure 2.14: Test bench: disc on disc tribometer [21]

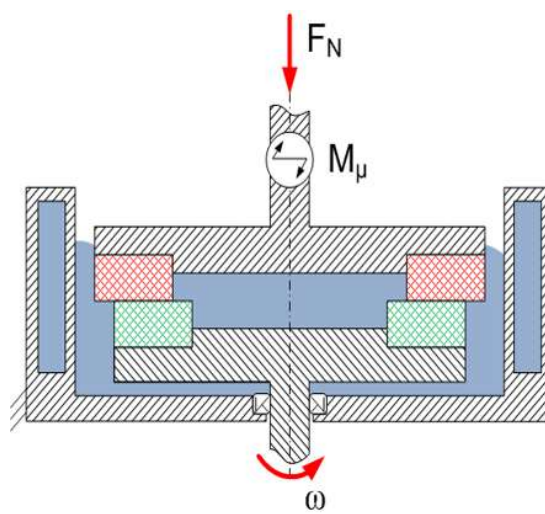


Figure 2.15: Tribometer scheme [21].

3 ML and AI state of the art

3.1 Artificial Intelligence and Machine Learning

As mentioned in the previous chapter, tribology is present in every aspect of our lives. Therefore, gaining an understanding of it can help with technological problems in the future. All of these developments are the result of meticulous tests and sophisticated computer models. Because of this, advanced data processing, analysis, and learning techniques can be created and used to increase the body of current knowledge. In modern machine learning or artificial intelligence applications, opportunities exist for analysing the intricate processes that occur in tribological systems and categorising or quantifying their behaviour in a timely or real-time manner. All modern technology includes artificial intelligence and machine learning. Every branch of science is looking for novel methods, including AI and ML, from biometric data to driverless vehicles. The study of training robots to perform human occupations is known as artificial intelligence [22]. Most historians credit a Dartmouth research project in 1956 that looked into problem-solving and representational methods as the beginning of artificial intelligence. The science of computing techniques that allow software to learn from past experiences effectively is known as machine learning, according to computer scientist Tom M. Mitchell. Supervised, unsupervised, and reinforcement learning are the three machine learning types.

The first learning algorithm uses tagged examples to apply prior knowledge to new data to anticipate future events. The learning technique based on well-known training data can predict an output value. As soon as the system has been adequately trained, it may present goals for any new input. In addition, the model may be checked for flaws and adjusted if necessary by comparing the output with the anticipated one. Unsupervised machine learning techniques are applied when training data lacks either classification or labelling. In machine learning, a programme scans unlabelled data for patterns or trends. The method can never be assured that the result is accurate. Instead, it infers what the outcome should be depending on observations. Reward-maximizing or risk-reducing actions are produced by reinforcement machine learning algorithms, a learning system that interacts with its environment. The agent

in this scenario is a reinforcement learning system that iteratively continually learns from its surroundings [22]. The delayed reward and trial-and-error search are its most important traits. With the help of this technique, machines can automatically decide what actions to take in a given situation to function at their best.

3.2 ML and AI in Tribology

Due to their capacity to forecast tribologically essential characteristics, artificial intelligence techniques have attracted significant attention in the tribological community. By automatically identifying patterns in datasets, these techniques can be used to build models that can forecast uncertain future data or other outcomes of interest. The use of AI and ML in tribology spans several areas. Some of the most significant applications are described here [23]:

1. *Online condition monitoring*

In order to categorise the wear particles in ball-on-disk sliding studies, various experiments have been carried out since 1998 with the aid of an artificial neural network (ANN). The latter accurately predicted the link between the experimental circumstances and the acquired particle features after being trained with data. A few years later, another critical investigation was conducted on the localised fault identification in ball bearings. A multi-layered feed-forward NN that has been trained using both an unsupervised adaptive resonance theory and a supervised error backpropagation method was used to make it operate. For tracking and categorising the wear behaviour of lubricated journal bearings, a more recent work of a similar nature was reported.

2. *Design of material composition*

The sliding contact of sintered steels, produced via powder metallurgy, was the subject of some research. As this type of manufacturing is frequently used in transmission systems and engine parts, friction and wear are typically the causes of failure. Many studies have predicted and optimised the tribological performance of various materials and operating situations using ML and AI methodologies. This experimental validation is demonstrated using three alternative strategies. An ANN, which can handle and

learn nonlinear neural networks, was used in the first method. The second approach, a “fuzzy system”, is efficient with time-dependent functions. The final strategy was based on a fuzzy-neuro system combining both strategies. Besides steel, composite materials have also been put through a variety of experimental processes in order to explore the tribological behaviour of brake and clutch materials. Material composition, sliding speed, and contact pressure are input variables for these tests.

3. *Lubricant Formulations*

In addition to considering the nature of the substance, lubricant formulations can be created using AI and ML techniques. The latter will aid in producing the best lubricant qualities when combining various oil combinations. Two different ANN are employed to forecast the formulation of two lubricants. Algorithms in thermo-hydrodynamically lubricated contacts are learned by combining high-speed data inputs from torque sensors. In most studies, lubrication performance is predicted using an artificial neural network. In the latter, the training dataset is predicted using numerical simulations. It should be mentioned that the sample size and training algorithm significantly influence the accuracy.

4. *Prediction of Lubricant Properties*

In many applications, lubricants include a complicated mixture of two or three different base oils, an additive package, and significant polymers called viscosity modifiers [24]. It is difficult to predict how the lubricant will behave at different temperatures and shear rates. AI and machine learning would be helpful to businesses in this type of issue to increase the effectiveness of their lubricant design process. In order to test their AI models, these businesses are also expected to have reliable viscometrical data on various lubricant formulas. Unfortunately, these methods are likely to be preserved and exploited by lubricant suppliers for their products because the composition of lubricants is sometimes a closely held secret or might change suddenly.

5. *Visualisation and Classification of Wear Particles*

AI and machine learning techniques are especially effective for classifying and identifying images. A community of experts has applied these methods to wear particles. Size, texture, shape, and colour are all characteristics that reveal details on

the location and the nature of wear mechanisms. In order to determine if a wear particle is metallic or oxide, as well as whether it is the result of fatigue or extreme sliding, NN algorithms have been applied.

After examining several applications of AI in tribology, its potential is obvious, and in the future, it may be employed in many more tribological domains. It can be used to activate base oils with modified viscosity and friction to forecast the outcomes of studies carried out under different material compositions and test settings. In addition to predicting the ideal concentration, the development of a tribo-layer will also make use of nanoparticle prediction. Additionally, it is possible to forecast dynamic changes, such as the characterisation and classification of the relevant surface topography. From a more practical perspective, in the future, AI technologies will be used in the design of lubricated components, for instance, determining the best film thickness for machine components.

3.3 Condition monitoring

Sensors are used in condition monitoring to check up on equipment and notify the owner of any changes to operating parameters, including RPMs, temperature, and pressure.

These are the main goals of condition monitoring [24]:

- avoiding malfunctions and equipment breakdowns
- minimising downtime and lowering maintenance costs
- maximise efficiency while cutting down on waste and emissions.

Numerous benefits come with predictive maintenance, including less demand on service workers, less demand for on-site spare parts, decreased energy use, and significantly more reliable output. It starts by connecting the equipment and gathering pertinent data using machine sensors, process systems, and data historians. The following stage is to identify any problems by analysing the vast data using a combination of machine learning and artificial intelligence. In the third step, data insights are shown in simple dashboards that inform and offer advice when an issue materialises. Modern machines are getting more and more sensors to communicate information about their operating circumstances to the machine's owner or the machine's original maker (speeds, loads, and lubricant temperatures). Lubrication sensors can also be used to find wear particles and deteriorated lubricants.

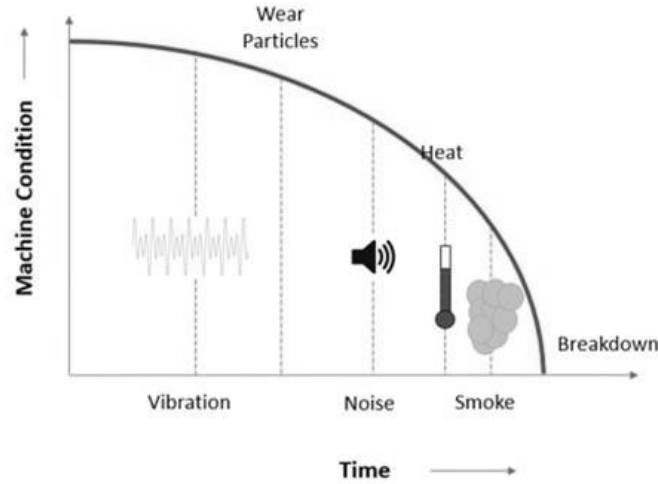


Figure 3.1: The types of events preceding machine failure [24].

By keeping track of numerous devices, it is possible to identify early warning indicators of impending failure based on past failures and their correlation with data from various sensors. Once early failure indications are identified, customers can be urged to service their devices or replace particular components. Even better, they might identify the underlying problems before they manifest as failure symptoms. The schematic in Figure 3.1 illustrates the kinds of indications that can be detected in advance of failure. Several commercial systems are now available that use AI and machine learning to monitor conditions (from companies such as SKF, GE, Siemens, and Bosch). Vibration monitoring and thermography are the two most often utilised methods for condition monitoring. Other specialised methods may also be employed for high-value types of machinery, such as wear particle sensors and infrared lubricant monitoring. These approaches could also be employed to monitor energy use. Consumers can be given recommendations on how to minimise their energy usage, even if the practical approach is to alert customers to equipment that may need maintenance or parts replacement. Many academics have tried forecasting friction and wear using machine learning and artificial intelligence. Since friction and wear are characteristics of the entire system rather than just its separate parts, this is a far more complicated problem. A representation of the complex nature of direct lubricated contact can be seen in the figure below. The running-in process lasts several hours when the contact environment changes.

Additionally, the lubricant creates intricate tribo-films, while its exact composition and other characteristics are unknown. Operating conditions define the type of lubrication to use

(hydrodynamic, mixed, or boundary), and if wear happens, loads and speeds have a significant impact on the type of wear. Much data is required to characterise the entire system, yet, most research publications do not provide enough information for algorithms to be applied. Additionally, the unique nature of lubricant composition means that only the producers of lubricants and additives often have access.

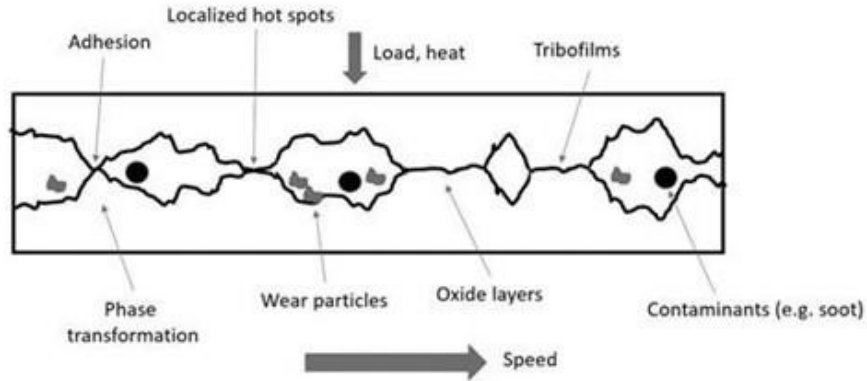
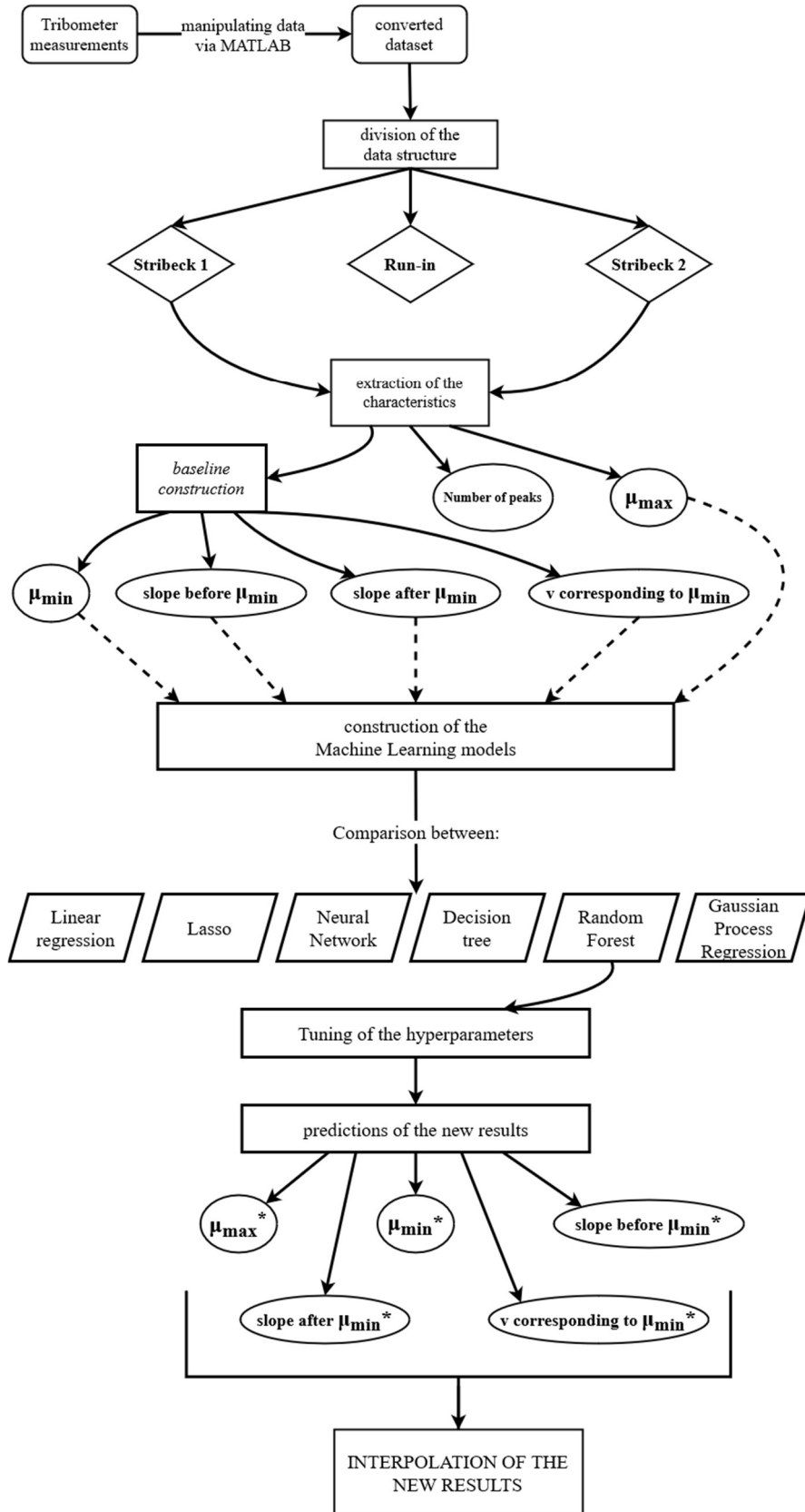


Figure 3.2: The complex nature of a typical lubricated contact [24].

4 Simulations, graphs and comparisons



The procedure which leads to the prediction of the new interpolated Stribeck starts from the tribometer measurements, which are saved by the test bench in a vast *.mat* file. The latter is converted, employing three MATLAB files, into a *.csv* file which can be manipulated with Python. By doing that, the files' variable dimensions are adjusted, making them equal in length. After that, the *.csv* containing the dataset is divided between the first Stribeck, run-in, and second Stribeck curve. After detecting the two Stribeck curves, the essential features were extracted: minimum coefficient of friction, number of peaks, maximum coefficient of friction, slope before μ min, slope after μ min, and speed corresponding to the minimum coefficient of friction. The construction of a baseline curve is needed to extract some of the features mentioned above. The latter is achieved through the Raman spectroscopy procedure. After defining its inputs and outputs, the Machine Learning models have been constructed. The best model is chosen by comparing the scoring metrics of different models. Finally, the Random Forest model is selected, and its hyperparameters are tuned. The outputs of the model are used to interpolate a new Stribeck curve. In parallel, some comparisons are carried out to study how some parameters affect the tribological behaviour of materials and lubricants. Finally, some predictions about new lead-free brass materials and eco-friendly lubricants have been made.

4.1 Conversion of the dataset

The tribometer measurements, for example, the frictional torque and the speed (in rpm), are saved in a *.mat* file. To manipulate the data, construct the curve and extract the features, it has been necessary to convert those files into *.csv* files. This step is performed employing three MATLAB files. The first one recursively looks inside the folders to access the 80 *.mat* files, one file for the measurements of each ID. The second one aims to adjust the dimensions of every variable. The requested size is accomplished by comparing them with the extent of the frictional torque, the variable of most interest since it has been used for the Stribeck curves and the run-in. The third file finally converts the resulting data table into a *.csv* file, which the Python files efficiently manage.

The variables needed to construct the Stribeck curves are obtained from these vast converted datasets. Then, the dataset has to be divided between the first Stribeck, the run-in part, and the second Stribeck. Finally, the variable flag detects the transition between the three phases. Before entering into the details of the parameters extraction, an overview of the Stribeck curve and the run-in is necessary to understand their roles and meaning.

4.2 Stribeck curve

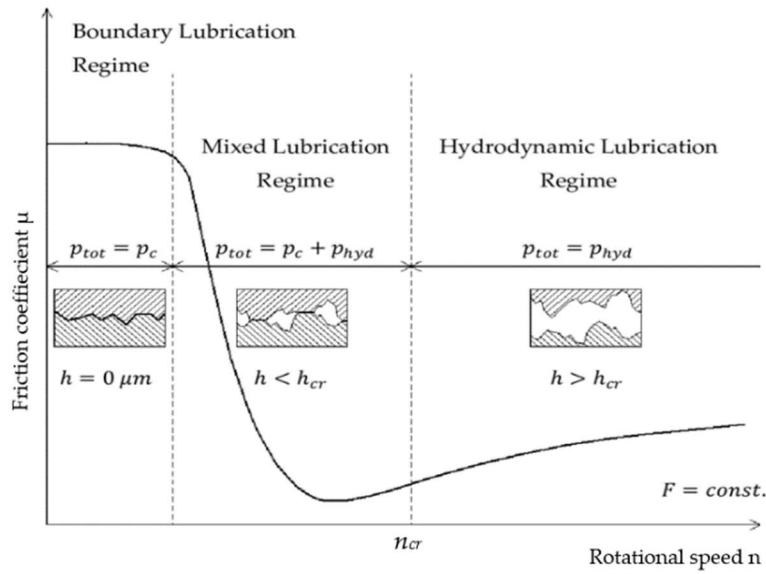


Figure 4.1: Stribeck curve and lubrication regimes [25].

Most sorts of lubrication behaviour have been explained using the so-called Stribeck curve, which is the most essential idea in the science of tribology. The amount of lubricant utilised in the interaction among spinning components significantly impacts how much of a decrease there is. When friction curves are displayed concerning parameters that include viscosity, speed v , and load p , they implement a homogeneous curve. Further study and the development of methods for measuring surface roughness led to the discovery that the friction factor may also be mapped against the fraction of the oil film width to the total quadratic mean of the machined workpiece [25]. A Stribeck curve might also be obtained if all parameters are held constant by graphing the friction coefficient as a function of speed v .

The three lubrication regimes are distinguished based on the lambda ratio, as indicated in section 2.3. However, it should be highlighted that the lambda ratio range listed above is merely a guideline and that it may be significantly different in practical situations.

4.2.1 Boundary lubrication

On the Stribeck curve, the basic level is widely used to describe border lubrication. When the characteristic is very tiny, and the interaction between the solids dictates how much friction occurs in the contact, it is associated with a loss of oil or highly severe circumstances. The latter is because the solid-to-solid connections support the entire load. The friction produced in solid direct communication is often relatively high when correlated to the friction produced by the viscosity tensile stress of the lubricating layers. Anyhow, boundary lubrication offers superior defence against chemical and adhesive wear. In reality, friction can increase significantly when two dry surfaces touch one another.

Additionally, additives are frequently added to lubricants to improve lubrication and the development of fluid films. The lubricant sticks to the surfaces due to fluid molecules being absorbed by the surfaces or chemical reactions. The fluid sheet's thinness means viscosity is typically unaffected by friction and wear. The friction coefficient is generally in the range of $\mu_{BL} \approx 0.08 \div 0.15$ [3].

4.2.2 Hydrostatic lubrication

In contrast to earlier lubrication systems, hydrostatic lubrication maintains loads with the help of a thick fluid layer maintained by a pump. The main advantage of this greasing is that the stress required to divide the two contacts is not created by relative motion. Additionally, slow, start-and-stop actions reduce wear. Much stiffness is also provided by hydrostatic lubrication. However, hydrostatic lubrication devices are always more significant and expensive since the pump takes up more space.

4.2.3 Mixed lubrication

The Mixed Lubrication Regime (ML) is the term used to describe the reduction in friction brought on by better lubrication and softer operating conditions. Boundary lubrication combined with hydrodynamic or elastohydrodynamic lubrication is referred to as the ML Regime. The friction in this situation is made up of friction from lubricant layers and solid-to-solid contact. Because the friction of lubricant layers is often lower than that of solid-to-solid contacts, overall friction is reduced. Certain additives may be added to the lubricant to

prevent adhesion bonds. The following is a description of the mixed lubrication μ_{ML} friction coefficient [3]:

$$\mu_{ML} = \alpha \cdot \mu_{BL} + (1 - \alpha) \cdot \mu_{FF} \quad (4.1)$$

Where α is the ratio of the forces generated by an enhanced interaction to the weight of the lubricant, μ_{BL} is the frictional factor in the boundary domain, and μ_{FF} is the hydrodynamic resistance. This equation indicates that, to a certain extent ($\mu_{BL} \gg \mu_{FF}$), friction lessens as solid-to-solid contact increases.

4.2.4 Hydrodynamic lubrication

As the working conditions improve, the film thickness rises until there is no longer any direct contact and complete film lubrication. Here, the lubricant layers are viscously sheared to create friction. According to the figure above, friction across the layer lubrication phase stops decreasing and grows despite the solid-to-solid contact elimination. The rise in hydrodynamic friction is what is responsible for this behaviour. The problem is that increasing viscosity, speed, or decreasing load is required to improve film thickness. As a result of the friction coefficient, all cause an increase in mechanical tension. For this type of lubrication to occur, two essential requirements must be met:

- There must be relative movement in both the contact objects and a sufficiently high speed for the oil to be capable of bearing loads.
- In order to create the pressure field required to support the requisite weight, surfaces must be slanted in one direction or another [2].

The development of the HD pressure among two sliding and non-parallel objects is graphically shown in Figure 4.2. A lubricant is applied to the lower surface, known as the “runner”, and moves at a specific speed. Concerning the bottom surface, the upper layer is angled in a particular direction. As a result, a pressure field is created due to the bottom surface's motion, which draws the lubricants into the converging wedge. If not, the wedge would lose more lubrication than it would gain. As a result, the entering flow is constrained at the wedge's tip as pressure increases, while the exit flow is boosted as pressure decreases. As a result, the fluid

velocity profile bends inward at the wedge's entrance and outward at its exit due to the pressure gradient, as depicted in the picture below. The pressure created divides the two surfaces and has a limited load capacity [2].

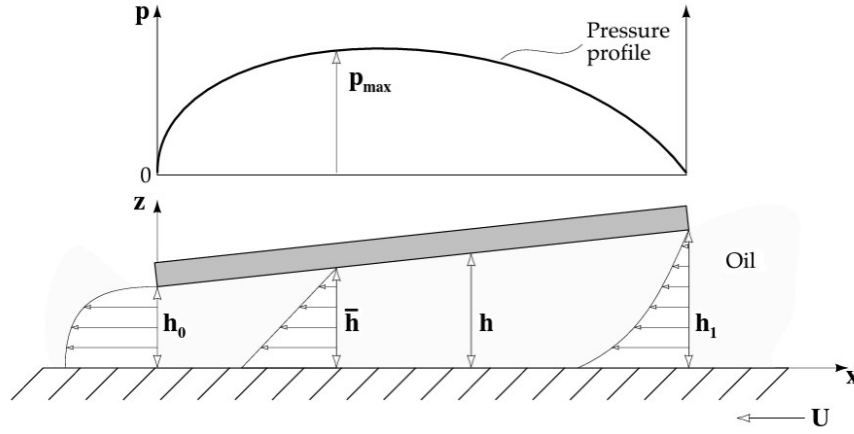


Figure 4.2: Hydrodynamic pressure generated between two non-parallel surfaces [2].

4.2.5 Elastohydrodynamic lubrication

EHD, or elastohydrodynamic lubrication, is a state of lubrication in which a thin layer entirely separates the two surfaces. It is a kind of HL; however, in this instance, the solid bodies experience significant elastic deformations that significantly change the shape and thickness of the lubricant layer in the contact. EHD typically occurs in applications where big loads are applied and high pressures are subsequently created. In this type of lubrication, the fluid layer thickness is thinner (usually 0.5 to 5 μm [3]), and the viscosity's variation with pressure is crucial. Elastic deformation of the contacting surface results from the high hydrodynamic pressure created in the lubricant coating [1]. As a result, the Reynolds formula, linear elastic calculation, and viscosity-pressure equation all impact the behaviour of EHD lubrication.

A conforming contact occurs when two surfaces are similar in form or shape. The hydrodynamic regime decouples the film thickness and friction in non-conforming contacts, such as gear or cam surfaces. The oil viscosity at the surface contact's entrance controls how thick the coating is.

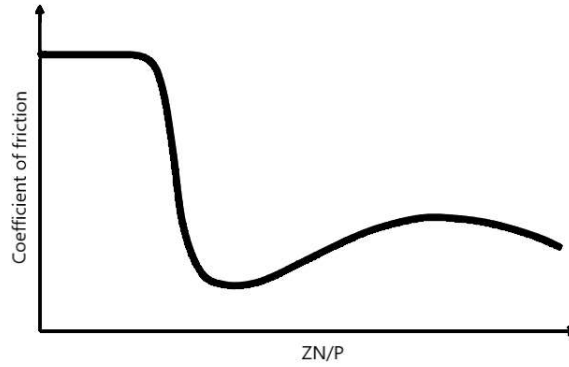


Figure 4.3: Stribeck curve in non-conformal contacts.

Lubricant properties can affect the Stribeck curve shape, giving essential information on how lubricants affect contact surfaces. For example, a first change can be seen in figure 4.4a, obtained by controlling the EHL friction of the base fluid (referred to as the traction coefficient or the internal friction of the lubricant). Being able to drive that friction coefficient down, then in the mixed and hydrodynamic regimes, the friction coefficient can be reduced. Also, controlling film thickness affects the Stribeck curve shape (figure 4.4b). The boundary lubrication starts earlier by increasing the viscosity. That is because, with higher viscosity, more load can be supported. The third one, figure 4.4c, is generated due to the lubricant producing a solid-like film in alignment with the contact surfaces. This film, which resembles a solid, has some shear strength. Therefore, when it comes to boundary lubrication may help lower the coefficient of friction. That can be achieved with surface active additives and friction modifiers. Finally, figure 4.4d shows the effect of changing the surface roughness. Boundary lubrication can be brought earlier by altering high peaks and troughs.

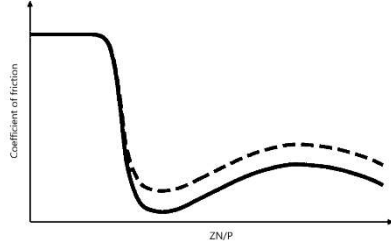


Figure 4.4: Effect of changing the EHL ratio.

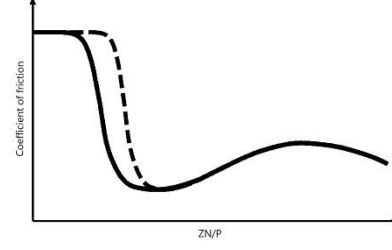


Figure 4.5: Effect of varying the film thickness.

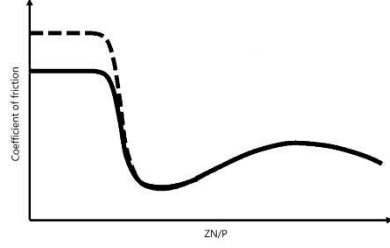


Figure 4.6: Effect of a solid-like film formation.

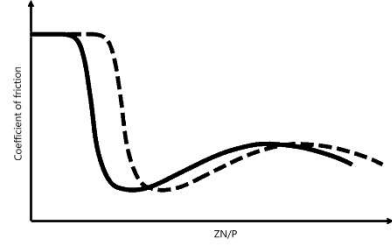


Figure 4.7: Effect of a change in surface roughness.

The original surfaces of the contact surfaces typically wear when machine elements are in use. In steady state wear, this process often starts with a very high wear rate, then the value drops. The early phase, known as the running-in stage, is characterised by high wear loss, friction coefficient change, and surface roughness evolution. At this point, local plastic deformation has worn down the interacting substrate materials. Surface roughness increases as a result of parameter λ growth during running-in. As seen in the image below, this results in a shift in the Stribeck curve. It is interesting to note that, although it does not always happen, the boundary friction coefficient decreases occasionally. The run-in phase and its effects are discussed in the paragraph below.

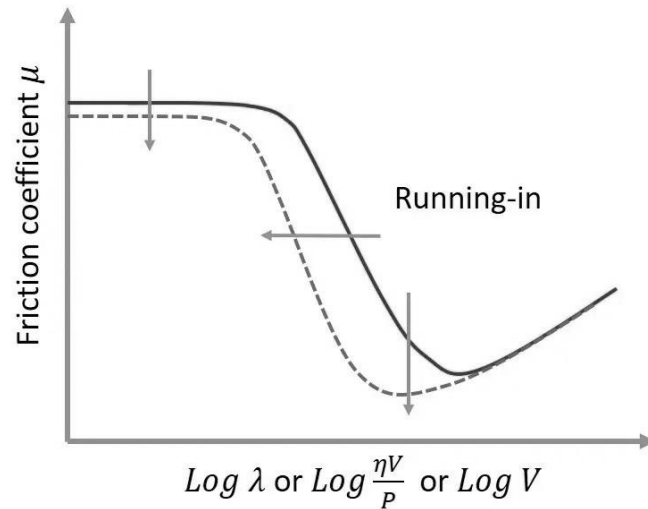


Figure 4.8: Effect of the running-in on a Stribeck curve.

4.3 Run-in

The most crucial phase in a tribology component's existence is the run-in. Historically, it has been linked to modifications to the mating surfaces' micro geometry to improve their conformity. The condition is considered equilibrium as long as the medium dynamic friction and other tribological characteristics have stabilised. The actual interaction area is relatively small when two recently machined surfaces are joined together; only the highest points of the defects make contact [3]. The wear starts very quickly but slows down as the contact area grows. Running-in takes place during the initial phase of a rolling or sliding contact's lifetime in a lubricated system, as schematically seen in Figure 4.9. The tribological contact undergoes conditions that are outside of thermodynamic equilibrium during running-in. Therefore, even little modifications to the existing boundary conditions might cause a catastrophic collapse. When the running-in process swiftly yields low friction and a moderate wear rate, the system, nevertheless, reaches a state of higher tensions and breakdown sensitivity. Specific equipment or elements are specifically built to work under specified run-in procedures after one arrangement or after regular maintenance to achieve the steady state appropriately. During the running-in phase, plastic deformation and light wear are two main mechanisms. The asperities are essentially crushed during the plastic deformation action. Surface temperatures are typically reduced during this time due to a decrease in frictional losses and an increase in contact clearances. As a result, the wear rate drops until it reaches the design contact pairs' typical steady-state wear rate. Even with slight misalignments, the wear rate during running-in is higher than during regular running [20].

The barrier morphology, molecular structure, and grain boundary underwent significant modifications due to running-in. Dissipative structures are frequently created by topography in response to external loads. Third body generation is necessary to obtain minimum wear levels and coefficient of friction. The first and second bodies have an equal impact on how well the tribological system functions in this situation. The device can either quickly reach minimal frictional forces and wear (case I), keep friction and wear constant (case II), or undergo complete collapse, which is distinguished by an overall increase in wear and friction (case III). Scenario I is the preferred regime for a tribological system. Case I may alternatively

be described as a legitimate run-in. Due to the necessity for activation energy in the tribo-chemical processes permitting the synthesis of third bodies, running-in significantly relies on the frictional intensity acting at the beginning of sliding. Excessive tension will bring on a condition of increased wear and friction in the system (case III). Under-stressing, however, can also be harmful. In this circumstance, the process does not acquire sufficient energy to produce the third body, resulting in excessive friction and wear (case II) [20].

Figure 4.7 depicts six typical friction-time curves; during the run-in, friction force can either rise or decrease. Each of the eight forms has eight distinct causes, none of which are specifically linked to a single group of helpful processes. Many interfacial process sets can create curve geometries with high similarity. It is also critical to realise that run-in is a property of the entire tribo-system and not just a material-dependent phenomenon [1].

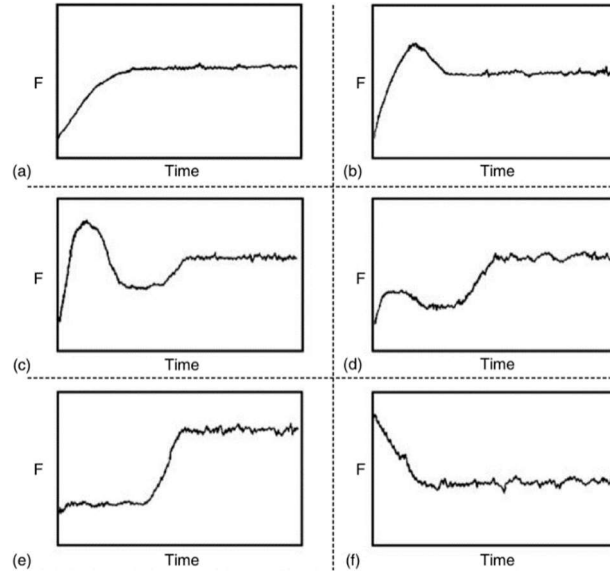


Figure 4.9: Six forms of friction behaviour as a function of time [1].

The terms asperity truncation and elastic shakedown describe the run-in procedure. Asperity truncation refers to the quick truncation of high, acute asperities when two surfaces are brought into contact. The latter modifies the surface's roughness characteristics, often smoothing the surface down, although occasionally, the roughness may even increase.

The full-service conditions can be applied after the running-in time, whose length is invariably dependent on the tribo-system, without any abrupt rise in wear rate. Because the relatively constant wear rate environment is continued for the duration of the planned operational life, the mean kinetic friction and other required properties have been obtained and preserved at a typically constant level.

There are two phases throughout the running-in process, as the image below illustrates. The coefficient of friction significantly drops in Phase I, and the change in surface topography resembles the decline in the centre line average roughness, R_a , value. However, R_a 's lowering and friction coefficient slightly decreased in Phase II. In this phase, minor wear is considered due to eliminating boundary layers created by the interaction between the lubricant's additives and oxygen and the contacting metal surfaces [3].

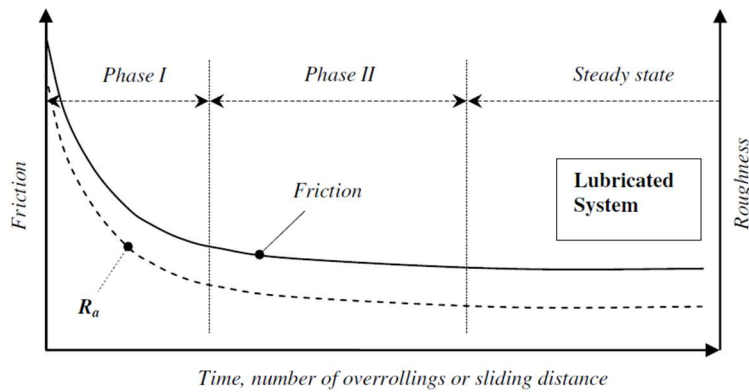


Figure 4.10: Friction and roughness functions during the running-in [3].

4.4 Saving the data structures

The division of the dataset is performed through these lines of code. At first, it was calculated that the coefficient of friction employs the frictional torque. The change from the Stribeck curves and the run-in phase is detected through the variable flag. Finally, the datasets for the Stribeck curves are constructed:

```
# get ID from input and create a csv name
csv_files = ID + '.csv'
# load a csv to a pandas dataframe
folder = "C:\\Users\\Utente\\Desktop\\IFAS\\data"
df = pd.read_csv(folder + "\\\" + csv_files)

df = df[["n_ist_rpm_2", "M_Reib_mittel_Nm_2", "VersuchsID_2", "Flag_2",
"mue_2", "Attempt_2", "FN_ist_N_2"]].copy()
df = df.rename(
    columns={"n_ist_rpm_2": "n_ist_rpm",
            "M_Reib_mittel_Nm_2": "M_Reib_mittel_Nm",
            "VersuchsID_2": "ID", "Flag_2": "Flag", "mue_2": "mue",
            "Attempt_2": "Attempt", "FN_ist_N_2": "FN_ist_N"})
df = df[df.Attempt == 1]
df = df[df.Flag >= 1]
df = df.reset_index(drop=True)
df.FN_ist_N.iloc[:] = 1000

# calculate mue
Radius_mittelReibflaeche = 0.065 / 2
```

```

A_Reib = 1021
P_ist = 1 # N/mm^2
v_mps = df.n_ist_rpm * ((2 * math.pi / 60) * Radius_mittelReibflaeche)
df["v_mps"] = v_mps
df["F_Reib_N"] = df["M_Reib_mittel_Nm"] * (1 / Radius_mittelReibflaeche)
df["FN_ist_N"] = P_ist * A_Reib
df["mue"] = df["F_Reib_N"] / df["FN_ist_N"]

# detect flagchange
strobeckFlag = [0]
for i in range(len(df) - 1):
    if df.Flag[i] != df.Flag[i + 1]:
        strobeckFlag.append(i)
        strobeckFlag.append(len(df))

# create list of strobecktests
strobeck = []
strobeck.append(df[strobeckFlag[0]:strobeckFlag[1]])
strobeck.append(df[strobeckFlag[2] + 1:strobeckFlag[3]])
signal = []
for strobecktest in strobeck:
    # filtering
    strobecktest.v_mps = Filter.filtfilt(strobecktest.v_mps)
    strobecktest.mue = Filter.filtfilt(strobecktest.mue)
    signal.append(strobecktest)

return signal

```

After having detected the two Stribeck curves, the features have been extracted. All the curve characteristics extracted are listed below:

- minimum coefficient of friction;
- number of peaks;
- maximum coefficient of friction;
- gradient before μ_{\min} ;
- gradient after μ_{\min} ;
- speed corresponding to the minimum coefficient of friction.

4.5 Baseline construction

In order to extract some of the features mentioned above, the construction of a baseline curve is needed. Therefore, the Raman spectroscopy procedure is mandatory, since the baseline correction is one of its critical step.

```

v_max = max(filteredStribeckCurve['v_mps'])
v_min = min(filteredStribeckCurve['v_mps'])
y_norm_area = rp.normalise(filteredStribeckCurve['mue'], method="minmax")
roi = np.array([v_min, v_max])
x = np.array(filteredStribeckCurve['v_mps'])
y = np.array(y_norm_area)

```

```
ycalc_poly, base_poly = rampy.baseline(x, y, roi, 'poly',
polynomial_order=4)
```

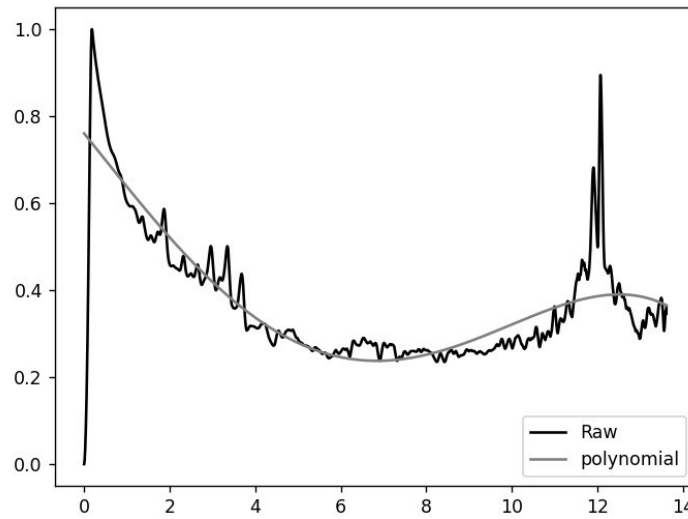


Figure 4.11: Baseline of a Stribeck curve

4.5.1 Raman spectroscopy

A non-destructive technique for chemical analysis, Raman spectroscopy provides comprehensive information on chemical composition and cellular mechanism. It depends on how light and chemical reactions in a material interplay. When a sample is subjected to monochromatic light in the visible range, the sample absorbs the light, and a significant amount of the light is passed through the sample, producing the Raman effect. The sample scatters a small portion of the light in all directions. The frequency of the incident light is distinct. Raman scattering happens when the dispersed light has a frequency different from the incident light. When an electron in the sample interacts with an incident monochromatic light, the electron takes up energy from the photon and rises to a virtual energy state. Energy transfer:

$$E = h \cdot \nu [J] \quad (4.2)$$

Each photon has an energy equal to the Planck constant ($h = 6.626 \cdot 10^{-34}$ J·s), multiplied by the light's frequency ν [Hz].

The electron falls to its starting location, and another photon is emitted if the power wasted is equal to the momentum of the incident radiation ($h\nu_i = h\nu_s$). In this instance, the frequency of the emitted photon equals the frequency of the incident photons (Rayleigh scattering).

When electrons lose energy from the virtual state, they occasionally revert to a lower vibration. As a result, the electron's photon emits energy differently from the incident photon. Raman scattering is created as a result. A Raman spectrum has several peaks that display the strength and location of the Raman scattered light's wavelength. Baseline correction is a required pre-processing method to distinguish genuine spectroscopic signals from interference effects or eliminate background effects brought on by lingering fluorescence or Rayleigh scattering.

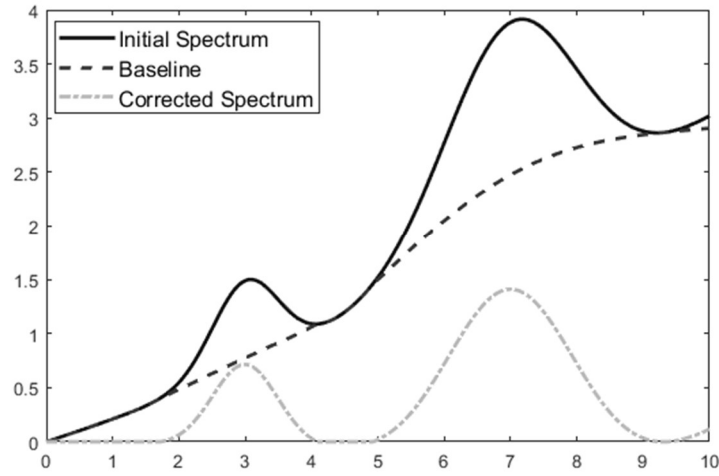


Figure 4.12: Baseline curve construction by means of the Raman spectroscopy.

4.6 Characteristic extraction

4.6.1 Minimum coefficient extraction

The minimum coefficient of friction, also known as elastohydrodynamic lubrication, is probably the most critical parameter in a Stribeck curve. Knowing its value gives information about the amount of energy that can be saved concerning other pairings [15]. In order to calculate the minimum coefficient of friction, two different strategies have been adopted: the change point technique and the research of the curve's minimum.

4.6.1.1 Change point technique

An abrupt change in a time series, such as a shift in underlying trends, frequencies, or probability distributions, is known as a change point. They are typically explained in terms of differences between segments. A change point separates a time series into two parts, each having its statistical properties (e.g., mean, variance). Numerous industries, including

banking, quality assurance in manufacturing, and clinical diagnosis, can benefit from change point detection (CPD). The following are some significant change points:

- A change in mean typically happens when a time series can be separated into distinct, constant parts with distinct mean values.
- Change in variance: although the signal's mean value is unchanged, multiple segments have varying variance values. The latter could be regarded as a spike in signal noise.
- Change in frequency, another name for a change in periodicity, refers to time series with cyclic characteristics. Here, a quick change in frequency causes the change.

CPD is divided into two different categories: offline and online. For statistical analysis, the first type relies on entire time series rather than live streaming data. As a result, offline methods are typically more accurate than online ones since they analyse the entire time series. On the other hand, online change point detection employs on-time series that are streamed live, typically for continuous monitoring or quick anomaly identification. This kind of CPD aims to identify state changes as soon as they occur by processing each data point as it becomes available.

In this thesis, an offline change point detection is used to identify an abrupt change in the gradient of the Stribeck curve. To this aim, the ruptures package of Python has been used. It focuses on offline changepoint detection, where the whole sequence is analysed.

The different techniques to calculate the change point are listed below, together with their advantages and disadvantages:

Dynamic programming: termed this because a dynamic programming approach is used to arrange the search over all feasible segmentations. The user must predetermine the number of changes detected because this is an exact technique.	
ADVANTAGES	DISADVANTAGES
It can calculate the cost of every subsequence of a given signal to determine the precise minimum of the sum of costs.	Cost of computation: the complexity of the order $O(CKn^2)$, where K denotes the number of change points to be detected, n denotes the number of samples, and C is

	the difficulty of invoking the cost function under consideration on a single sub-signal.
--	--

Table 4.1: Advantages and disadvantages of dynamic programming.

<p>Linearly penalized segmentation (Pelt): since it is impossible to enumerate every conceivable partition, this method is precise and relies on a pruning algorithm. In order to significantly lower the computational cost while maintaining the capacity to identify the ideal segmentation, many indices are removed.</p>	
ADVANTAGES	DISADVANTAGES
It is quicker than the previous one, and the complexity of the typical computation is of the order of $O(CKn)$.	It accomplishes this without letting the user choose how many change points to detect.

Table 4.2: Advantages and disadvantages of linearly penalized segmentation.

<p>Binary segmentation (Binseg): it is used to segment signals quickly. A change point in the original signal is found, the series is divided throughout the change point, and the method is repeated on the two resultant sub-signals.</p>	
ADVANTAGES	DISADVANTAGES
It can function whether or not the number of regimes is known in advance, thanks to its low computing complexity of $O(Cn \log n)$.	Finding several change spots by repeatedly splitting the sequence is the goal of the process, which might be time-consuming.

Table 4.3: Advantages and disadvantages of binary segmentation.

<p>Bottom-up segmentation (BottomUp): a sequential technique for quick signal segmentation is called bottom-up change point detection. First, a regular grid breaks the signal into numerous smaller signals. Contiguous segments are then successively combined based on a comparison of their similarity.</p>	
ADVANTAGES	DISADVANTAGES
It begins with a large number of change points and gradually removes the less important ones; low complexity: $O(n \log n)$;	Like the last one, this one is also carried out consecutively.

); it can be extended to detect numerous change points from any single change point detection method; it can work whether the number of regimes is known beforehand or not.	
---	--

Table 4.4: Advantages and disadvantages of bottom-up segmentation.

Window sliding segmentation (Window): this method of approximative search is fairly simple. For the window-based change point detection, two windows move along the data stream. Using a discrepancy measure, the statistical properties of the information within each window are contrasted.	
ADVANTAGES	DISADVANTAGES
low complexity: $O(nw)$; it can function whether or not the number of regimes is known in advance.	The purpose of the last three is to locate several transition points.

Table 4.5: Advantages and disadvantages of window sliding segmentation.

In order to choose the best technique, all the techniques are executed by looping through them. Those have been applied to the baseline. The function argument, *min_size*, and *jump*, iteratively changes for every method. The first argument controls the grid of potential change locations, while the leap regulates the minimum distance between change points. If *jump*=*k*, only changes at *k*, 2*k, 3*k, are considered. Similarly, if *min_size*=5, all change points will be at least five samples apart.

```

if algorithm == "dyn-prog":
    for min_size in range(5, 500, 50):
        for jump in range(15, 1500, 50):
            model = "l2"
            algo = rpt.Dynp(model=model, min_size=min_size,
jump=jump).fit(signal)
            result = algo.predict(n_bkps=1)
            stribeckCurvesChangePointResult.append(result)
            rpt.show.display(signal, result, figsize=(10, 6))
            plt.savefig('Dyn_prog-' + 'min_size' + str(min_size) + 'jump'
+ str(jump) + '.png')
            plt.close('all')

```

The first trial attempted to find the change point directly in the Stribeck curve. As can be seen from the figure below, the change point detected is one corresponds to the high peak at the beginning of the curve. Since this type of detection is misleading, it has been chosen to detect the point in the baseline curve of the Stribeck.

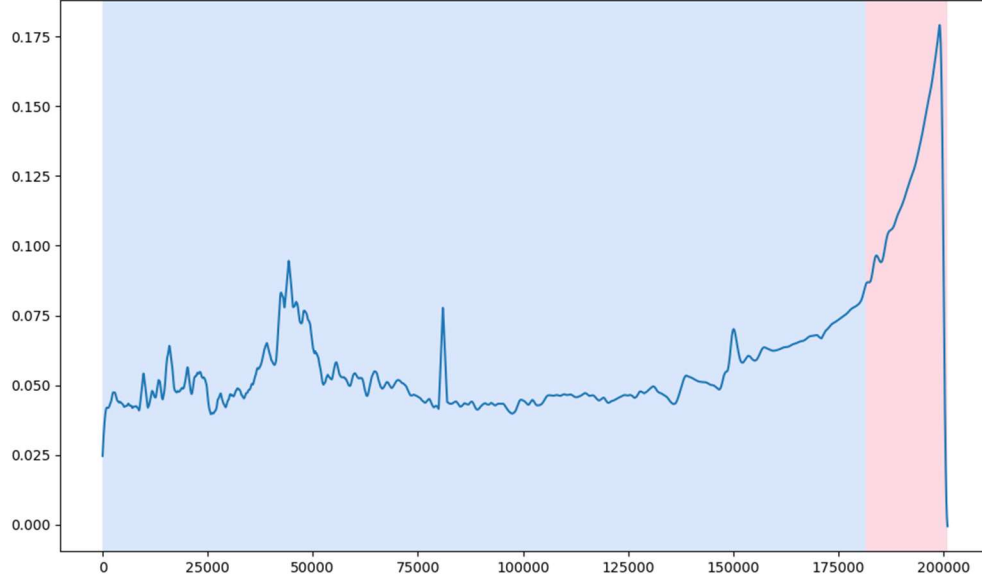


Figure 4.13: Dynamic programming with $\text{min_size}=5$ and $\text{jump}=15$.

The second step for the detection focused on the detection of a single change point inside the baseline curve. Also, this trial has not yielded the desired results; on the contrary, the points were far from the expected results.

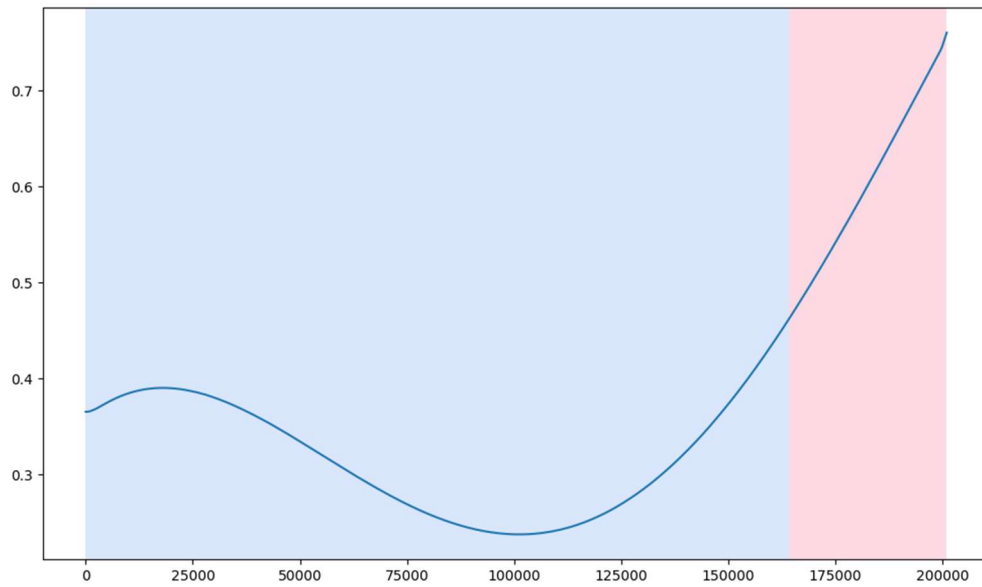


Figure 4.14: Binary segmentation with $\text{jump}=415$.

Finally, the code implemented aims at finding several change points and selecting the one with the minimum value between them.

After saving the figures with the highlighted change points, the latter have been visually analysed. The analysis aims at finding the technique and the parameters through which the closest point to the minimum has been achieved. The code aims at comparing the points detected with the different techniques. The Bottom-up is proven to give the best result, although it detects different change points. The Bottom-up code has been manipulated so that the point selected is the minimum between the points found by the change point technique. After tuning the number of change points detected, the optimal parameters are: *Bottom-up technique with 5 change points detected and jump equal to 915* (graph reported below in Figure 4.15).

```
model = "l2"
algo = rpt.BottomUp(model=model, jump=jump).fit(signal)
result = algo.predict(n_bkps=5)
# calculate the minimum between the change point detected
result = [x - 1 for x in result]
changePointResult = min(signal[result], default=0) [0]

changePointResult = ChangePointClass.changePointLogic(base_poly,
jump=900)
result = changePointResult * (max(filteredStribeckCurve['mue'],
default=0)
- min(filteredStribeckCurve['mue'], default=0)) \
+ min(filteredStribeckCurve['mue'], default=0)
stribeckCurvesChangePointResult.append(result)
```

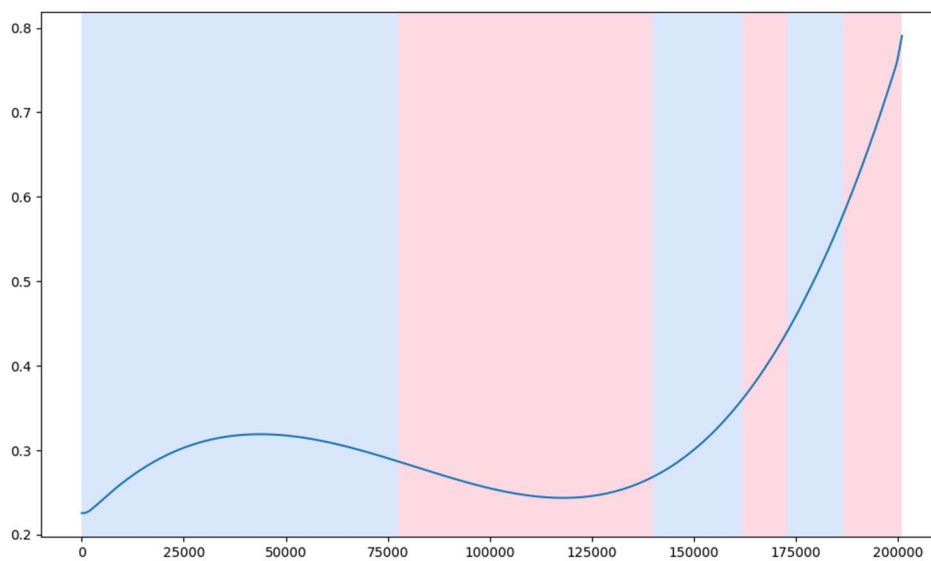


Figure 4.15: Bottom-up with jump=915.

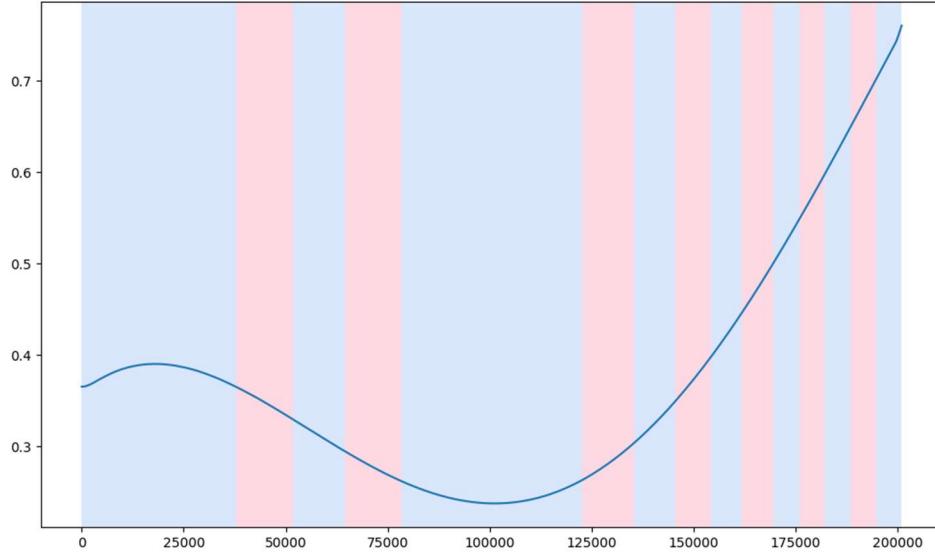


Figure 4.16: Linearly penalized segmentation (Pelt) with $\text{min_size}=405$ and $\text{jump}=1265$.

4.6.1.2 Minimum detection

```

ycalc_poly, base_poly = rampy.baseline(x, y, roi, 'poly',
polynomial_order=4)
MinMueIndex = MueMinClass.min_mue_computation(base_poly)
my_list = map(lambda z: z[0], base_poly)
ser = pd.Series(my_list)
try:
    MinMueResult = ser[MinMueIndex[0][0]]
except:
    MinMueResult = 0

```

Besides the change point technique, the minimum detection also uses the baseline curve. In addition, the script employs the `argrelemin` function from the `scipy` package. The latter part calculates the relative minima of a set of data.

The problem with the first results obtained with this was of two natures:

1. The point detected was the first of the dataset, which is the point at zero speed before the tribometer experiment started, i.e. before the rotor and stator were in contact.
2. The minimum point is at maximum speed since some Stribeck curves have monotonically decreasing behaviour.

After noticing these sources of detection errors visually, the code has been modified accordingly. The solution adopted is to research the point in a restricted range to avoid the detection of these misleading points. The interval of speed taken into consideration was $3 \text{ m/s} < v < 11 \text{ m/s}$. The same expedient has been employed in the change point technique, leading to an improved result.

4.6.1.3 Comparison between the two techniques

The comparison between the two methods adopted relies upon three main aspects: time, stability and efficiency. Now, they are analysed more in-depth.

Simulation speed

The Python function `time()` has been employed to calculate the elapsed time to execute each of the two functions. In a script context, absolute and relative time are separate time categories. The real-world time that the function `time.time()` returns is called absolute time. It is often measured with a resolution of at least 1 second starting from a predetermined point, such as the UNIX epoch of 00:00:00 UTC on January 1, 1970. On most computers, it is maintained by specialised hardware; the RTC (real-time clock) circuit is often battery-powered, so the system maintains real-time between power-ups. This “real-world time” may also change depending on the time of year and where you are. The second type of time is relative time, which is `time.perf_counter` and `time.process_time` both return. Since the relationship depends on the system and implementation, there is no clear relationship between this form of time and actual time. It is primarily used to assess relative performance and can only be used to measure periods. This time, known as the scope in this instance, is generally used to determine if one code is quicker than another. The one used is `time.perf_counter`, which gives the CPU counter's absolute value. The latter function always outputs the time in seconds as a float number. A performance counter, or a clock with the best precision available to measure a brief period, should return its value (in fractional seconds). It is system-wide and includes time spent sleeping. Since the returned value's reference point is unknown, only the difference between the outcomes of subsequent calls may be considered genuine. Advantages of `perf_counter()`:

1. A more accurate number than `time.clock()` function will be provided by `perf_counter()`.
2. Calculating time in seconds and nanoseconds using float and integer data is possible.

The result obtained for the change point detection is:

Elapsed time: 2463.3779978 2.0665198

The elapsed time during the Change Point computation in seconds: 2461.3114779999996

Instead, for the minimum detection, it turned out to be:

Elapsed time: 2391.4145477 1.8801507

The elapsed time during the Change Point computation in seconds: 2389.5343970000004

Therefore, the minimum detection is slightly faster.

Stability

In this context, the aim is to understand if the method can always find a result. For example, the change point technique can always find a point. The minimum technique, instead, has problems handling monotonically decreasing Stribeck curves.

Efficiency

The question is whether the point detected is close to the wanted minimum. In order to answer this question and evaluate the two techniques, it is also necessary to analyse the speed of the minimum point calculated. The comparison has been made visually, plotting the points calculated and the expected minimum points. The first inspection highlighted wrong results for the minimum points, in that the speed of those points was far from the expected minimum points. This visual comparison has led to the modification of the interval of speed considered. The latter has led to the choice of the interval between 3m/s and 11 m/s.

While the μ value for the minimum detection technique could not find a point for the monotonically decreasing curves, the change point technique always returned a result, although in some cases worse than the simple minimum technique.

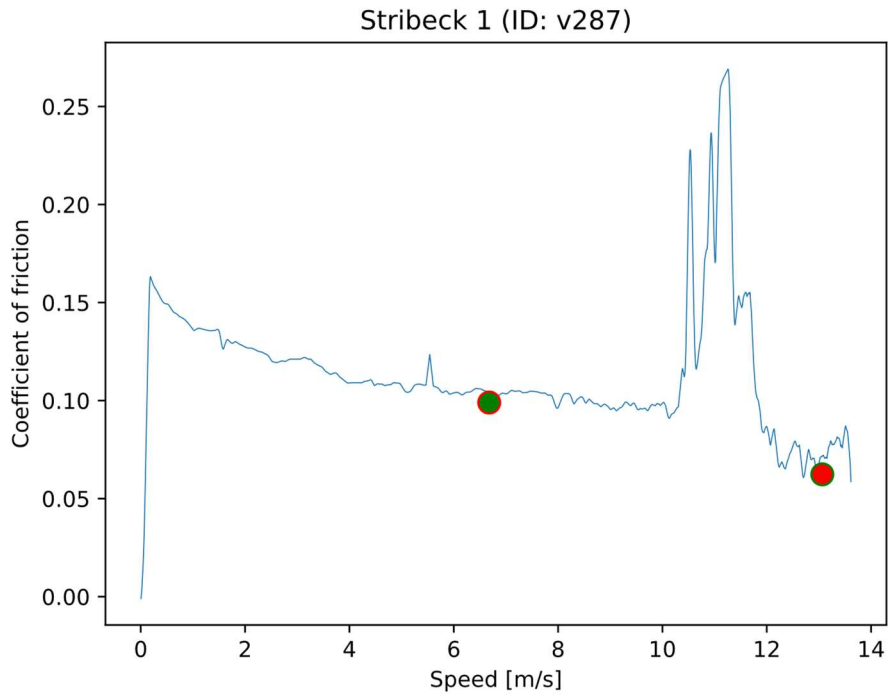


Figure 4.17: Stribeck curve of v287 (green point = change point detection, red point = minimum detection).

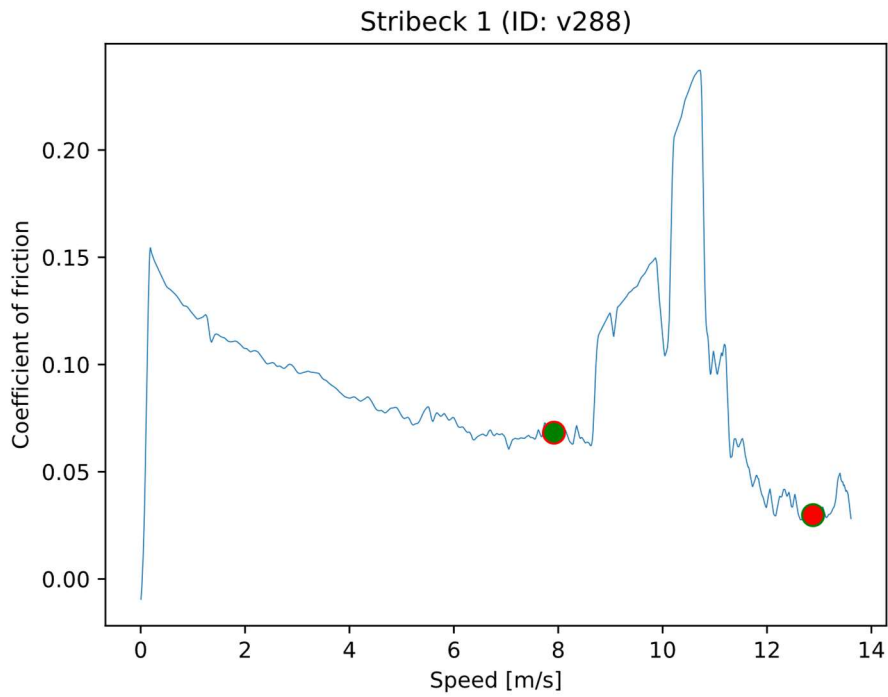


Figure 4.18: Stribeck curve of v288 (green point = change point detection, red point = minimum detection).

A final trade-off between the performance of the two techniques has led to the choice of the change point detection technique. Even though the minimum detection is slightly faster, it

cannot detect any point with monotonically decreasing Stribeck curves, and the point detected is far from the desired one.

4.6.2 Gradients and slopes

```
v_max = max(dataframe['v_mps'])
v_min = min(dataframe['v_mps'])
y_norm_area = rp.normalise(dataframe['mue'], method="minmax")
roi = np.array([[v_min, v_max]])
x = np.array(dataframe['v_mps'])
y = np.array(y_norm_area)
ycalc_poly, base_poly = rampy.baseline(x, y, roi, 'poly',
polynomial_order=4)
my_list = map(lambda z: z[0], base_poly)
ser = pd.Series(my_list)
list_of_gradients = np.gradient(ser, edge_order=2)
dataframe = dataframe[dataframe.v_mps < 11]
index = len(base_poly)-len(dataframe) - 1
# search for the element that corresponds to 3 mps for calculating the
slope there
gradient = list_of_gradients[index]
return gradient
```

The gradients calculated are the two before and after the minimum point. After an inspection of the Stribeck curves, it has been chosen as reference points to calculate the gradients at the speed of 3 m/s and 11 m/s. The gradient is calculated in the baseline curve in conjunction with the minimum detection. The function of the computation is the *gradient* from the NumPy package. The latter function calculates the gradient using either first- or second-order correct one-sided (forward or backward) differences at the boundary and second-order accurate central differences in the interior locations:

$$\delta_h[f](x) = f\left(x + \frac{1}{2}h\right) - f\left(x - \frac{1}{2}h\right) \quad (4.3)$$

The returned gradient has the same shape as the input array, which in the case under study, is the list of elements of the baseline polynomial approximation. From the returned list is selected the element corresponding to the speed of 3 m/s and 10 m/s.

The parameters computed through this technique were useless for constructing the new Stribeck curves with the predicted parameters. As seen from the mathematical relation governing the gradient computation, it is computed using main and first differences. These values can be used neither to extract new points from the calculated ones nor to interpolate the new predicted curve.

For the above reasons, it has been chosen to extract the slopes at the same points. However, as a first step, it is essential to clarify the difference between slope and gradient.

A slope in mathematics is a figure that represents how steep a line is given. Additionally, it aids in defining a line's direction. It can likewise be seen as the gradient of a graph at any point. In mathematics, it refers to how steep a graph is at any given location.

Instead, the gradient is the vector created when an operator is applied to a scalar function at a specific location in a scalar field.

Usually, the word slope is acceptable when there are just two things to consider. First, the curve produced by the function linking the two variables, which forms the slope, is a tangent to or derivative of that curve. It gauges how rapidly a function, $f(x)$, alters in proportion to x .

$$slope = \frac{\partial f}{\partial x} \quad (4.4)$$

The function whose rate of change is of interest, $f = f$, is dependent on many variables, as opposed to the slope in this instance, where it is not (x, y) . Therefore, F has several dimensions. It would be convenient if only one operator took f as input and provided one output. As seen here, the gradient operator:

$$\nabla f = \frac{\partial f}{\partial x} i + \frac{\partial f}{\partial y} j \quad (4.5)$$

Although working on a scalar function, the gradient nevertheless delivers a vector result. Because each gradient component indicates the rate of change concerning a particular dimension, this is significant. Therefore, the slope is a particular application of the gradient operator to a function with a single dimension.

There are three possible methods for completing this task:

1. Python's user-defined function may be used to determine the slope of a given line.
2. Python's SciPy module may be used to determine a line's slope.

```
index_end = 0
index_in = len(base_poly) - 1
dataframe = dataframe[dataframe.v_mps < 3]
index_vel = len(base_poly) - len(dataframe) - 1
a = [x[index_vel], base_poly[index_vel][0]]
b = [x[index_end], base_poly[index_end][0]]
```

```
slope, intercept, r_value, p_value, std_err = stats.linregress(a, b)
```

3. Python's NumPy Module may be used to determine a line's slope.

```
slope, intercept = np.polyfit(a, b, 1)
```

The issue with the techniques that use SciPy and NumPy is that the slope for the regression curve is determined in both instances. In reality, the `linregress()` function is used to compute the linear least-squares regression for two specified one-dimensional arrays of the same length to obtain the slope using the SciPy package. Similar to how the `polyfit()` method can be used to discover and return the slope and intercept of a specified specific line using the NumPy module and a collection of line coordinates supplied as arrays. The problem with the regression methods applied to the Stribeck curve, or its associated baseline approximation, is that the massive amount of data present leads to a misleading computation of the gradient and, thus, to a meaningless result. Therefore, the slope computed with a defined function results in a better result.

The slope detected at 3 m/s is calculated between the points corresponding to the maximum and minimum coefficient of friction. The second slope (at 11 m/s) is computed between the minimum point and the last point of the Stribeck curve, which is at maximum speed. The latter point is extracted from the filtered Stribeck curve dataset and corresponds to the first point of the sequence.

```
slope_low = (min - max) / (vel_mue_min - 0)
vel = dataframe['v_mps'].values
mue = dataframe['mue'].values
slope_high = (mue[0] - min) / (vel[0] - vel_mue_min)
return slope_low, slope_high
```

4.6.3 Speed in the change point

```
model = "l2"
algo = rpt.BottomUp(model=model, jump=jump).fit(signal)
result = algo.predict(n_bkps=5)
# calculate the minimum between the change point detected
result = [x - 1 for x in result]
changepoint = min(signal[result], default=0)[0]
min_index = np.where(signal==changepoint)
index = min_index[0][0]
speed_in_changePointResult = vel[index]
```

Another essential piece of information is the speed at the change point. Knowing at which speed the minimum coefficient of friction is located could be of paramount importance for companies. Since the chosen value for the minimum is the one detected by the change point

technique, this algorithm has the purpose of detecting the speed for that point. The function *where* from the NumPy package returns the index of the minimum coefficient of friction inside the baseline curve list of points. Using that index, the speed at that point is extracted.

4.6.4 Maximum coefficient of friction

```
dataframe = dataframe[dataframe.n_ist_rpm < 100]
mue_zero = max(dataframe.mue, default=0.070)
return mue_zero
```

The maximum coefficient of friction, due to the structure of the Stribeck curve, is at zero speed. Considering how tribometer experiments are carried out, it is clear that the coefficient of friction detected at exact zero velocity is null. For this reason, while constructing the code, it is necessary to search the point considering an interval of low speeds and not only the zero one. It has been chosen to research the point considering speeds smaller than 100 rpm.

4.6.5 Number of peaks

Before entering the details of the python scripts, an overview of the fretting phenomena is needed since it, together with adhesive wear described in section 2.2.2.1, are the causes of the peaks inside the Stribeck curves. It happens on the surface like welding, and to overcome the latter, the coefficient of friction becomes very high toward the point at which it happens. If the surface has an excellent surface finish, the peaks are less but higher (more concentrated) because the surface on which the welding happens is bigger. Instead, if the surface is raw, there are more peaks, but they are less concentrated, less high peaks.

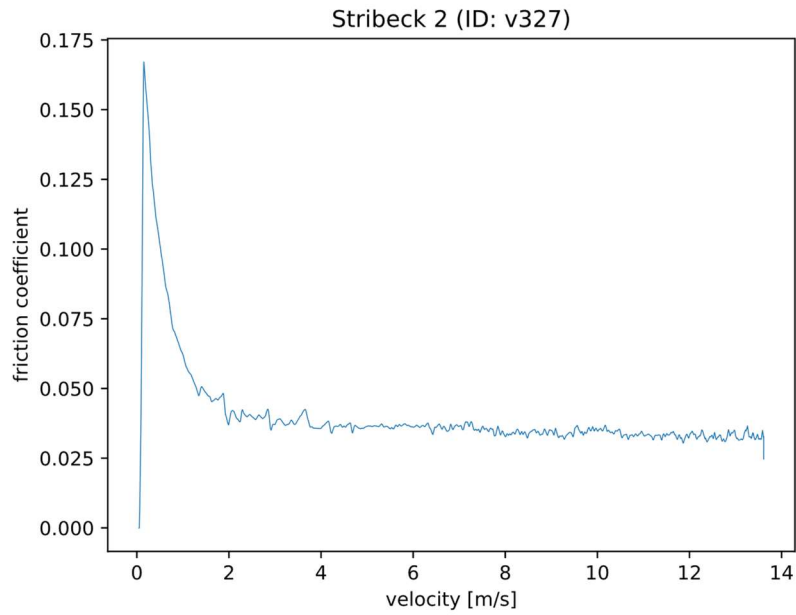


Figure 4.19: Stribeck curve of a specimen with low surface roughness.

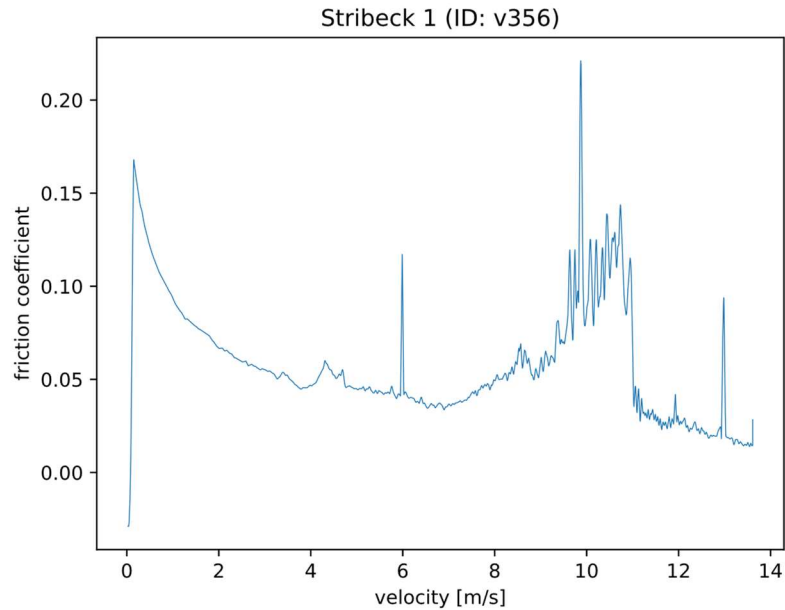


Figure 4.20: Stribeck curve of a specimen with high surface roughness.

4.6.5.1 Fretting

Fretting is an issue brought on by hard particles that become stuck at a bearing surface and start attacking the weaker of the two bearing surfaces. Low magnitude oscillatory sliding between ostensibly at-rest bodies defines this specific wear type. Fretting is caused by the adhesion of asperities on the contact surface. Wear debris is created as a result of the breaking. It often manifests as corrosion products surrounding pits or grooves. Typically, rusting is

present along with this phenomenon (in a corrosive environment). The two types of damage that arise are surface wear and fatigue life degradation. Fretting resistance is not a quality that material or even a material pair has by nature. Several factors influence the fretting behaviour of contact [1]:

- Contact load: normal load modifications typically have an impact on fretting wear. Even though users of the equipment frequently believe that high normal loads will prevent fretting by sufficiently dampening vibration, the increase in contact area leads to additional surface interaction, which tends to counteract this effect. The figure illustrates that rising load or unit pressures produce more significant wear rates.
- Slip amplitude: when there is a little slide, it is impossible to see the nucleation and growth of cracks that result in wear debris. This low wear rate is probably caused by worn debris rolling at that oscillation level. The gross wear rate is produced when hard particles (such as oxide or work-hardened particles) directly abrade the interface at large amplitudes. The fretting wear coefficient is almost identical to unidirectional wear at high oscillation amplitudes.
- Frequency: fretting frequency impacts fretting wear behaviour and wear rate. Frequency is observed to have a geometry-dependent effect; fretting frequency impacts wear rate for contacts with worse conformity.
- Temperature: the pace of chemical reactions, and thus the rate at which materials oxidise, is significantly influenced by temperature. At high heat, the pace of oxidation formation is quick, and the oxidation reaction acts as a safeguarded layer. It avoids the touch of raw materials and the formation of sticky contacts. As a consequence, the wear rate lowers.
- Relative humidity: higher humidity often reduces fretting corrosion. In the case of moisture, oxidation is assumed to proceed more quickly, forming a thick barrier.
- The inertness of materials: stress exhaustion is directly impacted by a material's capacity to react with oxygen.
- Corrosion and the motion-induced contact inadequacy that results from it.

Since fretting typically happens when the asperities of the mating surfaces come into contact, surface roughness is crucial. Due to their ability to decrease friction and prevent oxidation,

lubricants are frequently used to alleviate fretting. However, this may also have the reverse result because more movement might result from a lower coefficient of friction. Therefore, a solution needs to be adequately thought out and evaluated.

More often than not, soft materials are more prone to fretting than rigid materials. The latter is because the fretting wear is influenced by the two sliding materials' hardness ratios. However, softer materials like polymers can exhibit the reverse effect when ingesting hard debris embedded in their bearing surfaces. They then work as an extremely powerful abrasive, removing the stricter metal from contact.

Numerous wear mechanisms, including adhesion, abrasion, oxidation, and fatigue, are involved in fretting wear. Depending on the operating parameters, at least two mechanisms take place concurrently. First, the oxidised wear debris produced by fretting wear quickly increases the amplitude's wear coefficient. Second, reciprocating wear is characterised by wear coefficients that are roughly constant for constant sliding distance and independent of sliding amplitude. The surfaces are ostensibly in static contact for a sphere in a flat container close to the contact's centre, where the orthogonal tension is substantial. However, micro slip occurs at the periphery, with minimal pressure and sufficient radial stress to surmount static friction. Figure 4.21 shows the sticking and slipping regions in a fretting contact.

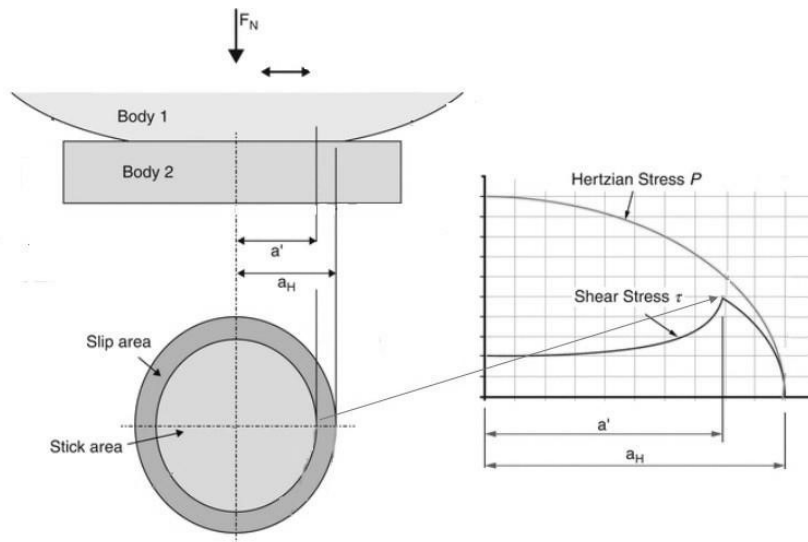


Figure 4.21: Types of stress in a contact subject to fretting [26].

Commonly, there are three distinct phases of fretting [26]:

1. First, metallic contact at an asperity level between the first contact objects. The latter happens when reactive species raw elements come into interaction with a compromised shielding oxide layer. Asperities are a few places where contact occurs.
2. Debris is produced in the second stage, after which it is oxidised. The limited amplitude of motion makes it challenging for particles to leave contact; as a result, they function as abrasives and may cause wear rates to increase.
3. Anxiety weariness is the last stage. Frequent loaded and shear stresses cause the corners of the tension zone to experience more significant stress areas, which leads to fretting. As a result, the fracture begins at the margins of the fretting region and moves within.

Fretting corrosion

Fretting corrosion is a kind of wear that frequently involves fretting and corrosion. Due to material breakdown and debris build-up, the newly formed surface is exposed to the air, which promotes oxidation. Iron oxide is tough and serves as an abrasive in the case of steel. Since the particles cannot escape the contact, they start the process of oxidation and abrasive wear. As a result, wear volumes may be higher than abrasive or adhesive wear alone.

Fretting fatigue

Fretting causes contact surfaces to wear while also subjecting them to cyclic friction stress. Frustration weariness is also known as fretting tiredness. The object's contact area experiences the highest friction stress, rapidly decreasing as the material travels within. Therefore, fretting fatigue crack development behaviour differs significantly from friction-less fatigue crack growth behaviour. With longer cracks, fretting fatigue results in a sharp decline in the growth rate of cracks close to the contact surface. After displaying a minimal value, it then agrees with the growing rate of the plain fatigue crack. The fretting fatigue load capacity will generally be halved or lower than the normal fatigue strength, although the material's compressive modulus is improved. The life of fretting is shorter than the life of simple exhaustion. Even at low-stress amplitudes, where plain fatigue failure might not happen, a fracture can quickly form due to friction stress in fretting fatigue, and as the crack grows, the

material fails. Fretting happens in orthopaedic implants when two materials come into contact, like bone plates or screws.

The function `find_peaks` from the `scipy.signal` package is used to find the number of peaks with Python. Then, using default function parameters, all the peaks in the function, which are a considerable number, are detected.

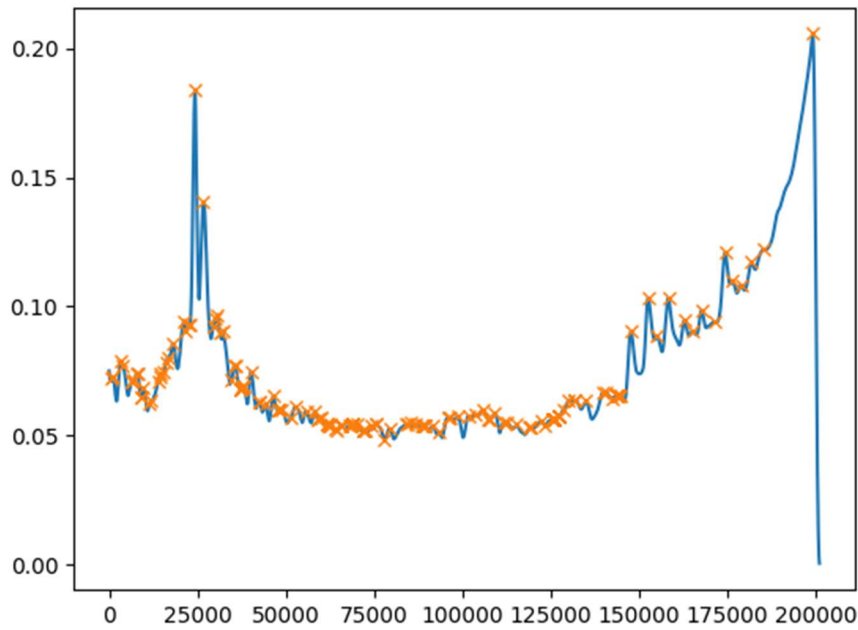


Figure 4.22: Stribeck with all its peaks highlighted.

Aiming at good peak extraction, it is important to understand the meaning of the parameters for the function below: width, threshold, distance, and prominence.

`scipy.signal.find_peaks(x, threshold=None, distance=None, prominence=None, width=None)`

The width option gives samples the required minimum peak width. If the minimum width is too high, the algorithm cannot track extremely near peaks in the high-frequency regions; hence, changing this quantity will not help. The same problem applies to the value of distance. The practical concept of prominence allows for the retention of the desired peaks while eliminating the distracting ones. The lowest point from which to ascend to any higher terrain is the (topographic) prominence. One may determine the prominence or how significantly a peak rises out from the adjacent basis of the signal by determining the vertical displacement between the peak and its bottom boundary line.

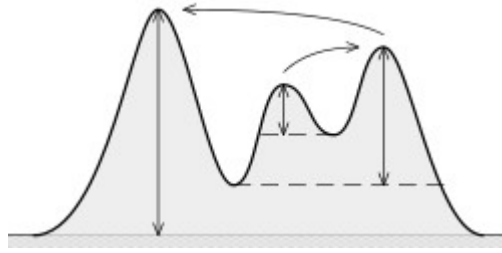


Figure 4.23: Meaning of the prominence.

After selecting the proper parameter to tune the function, it is essential to set its value correctly to find the correct number of peaks. The first trial aims to set an increasing prominence value to search for a value for which it is a plateau of the variance. It was essential to have the highest possible value of the variance for a good distribution. As seen from the histogram below, the problem is that the variance has an exponential increasing behaviour by decreasing the prominence. The reason for such behaviour can be traced back to the signal's noise.

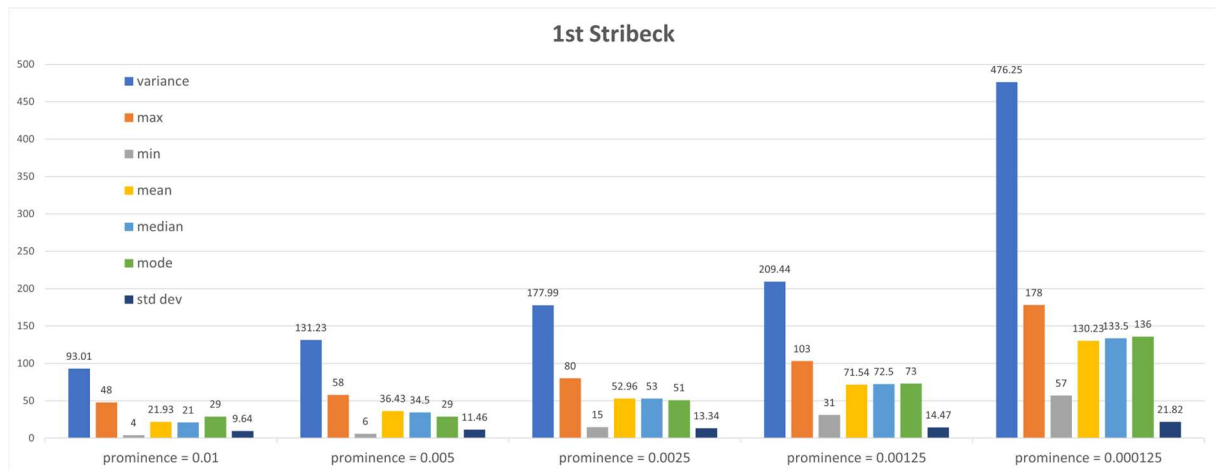


Figure 4.24: Histogram of comparison for different values of prominence.

Since the information about the variance turned out to be useless, another solution has been adopted: the percentile. The latter is a statistician's unit of measurement that shows the point at which a certain percentage of observations in a batch of observations fail. For example, a score below which a specific percentage k of the observations may be found is known as the k -th percentile.

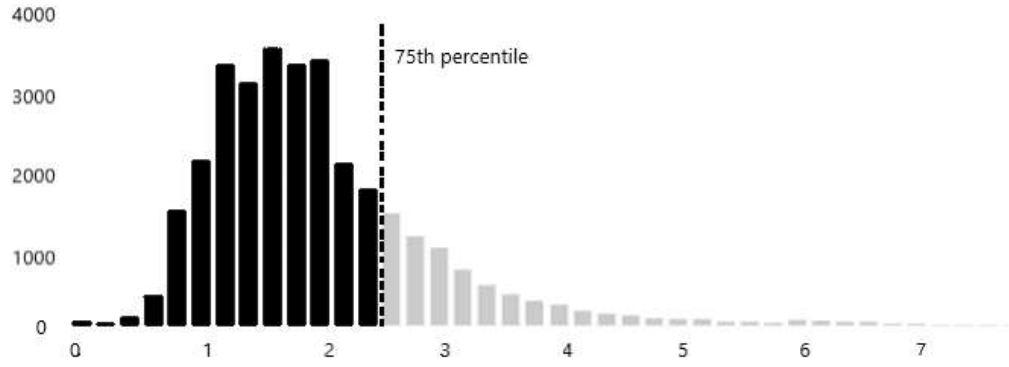


Figure 4.25: Third quartile or 75th percentile.

75th, also known as the third quartile, and 95th percentile have been analysed. The two measures are helpful to highlight the vast majority of measures and not be impacted by outliers. The two percentiles are applied to the entire set of measures to evaluate the best value. First, the prominence of the entire set of peaks is calculated, and the percentile is applied to find the prominence corresponding to a certain percentile. Since the last value is calculated for every ID, these values' mean is computed at the end.

```
peaks, _ = find_peaks(dataframe.mue)
prominences = peak_prominences(dataframe.mue, peaks)[0]
num_peaks = np.percentile(prominences, 75)
```

The results using the two different values of percentile are shown below.

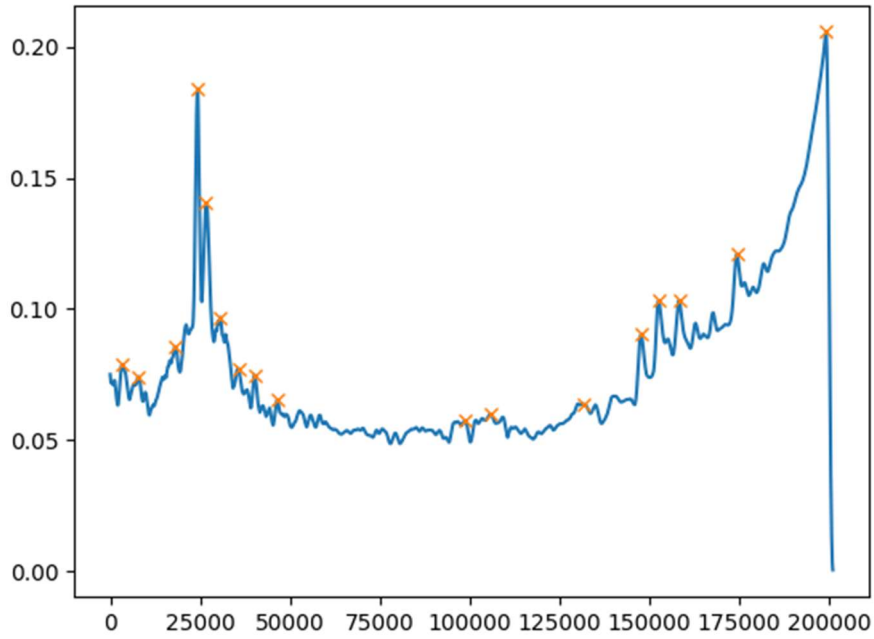


Figure 4.26: *Stribeck* with peaks detected setting $\text{prominence}=0.007$ (75th percentile).

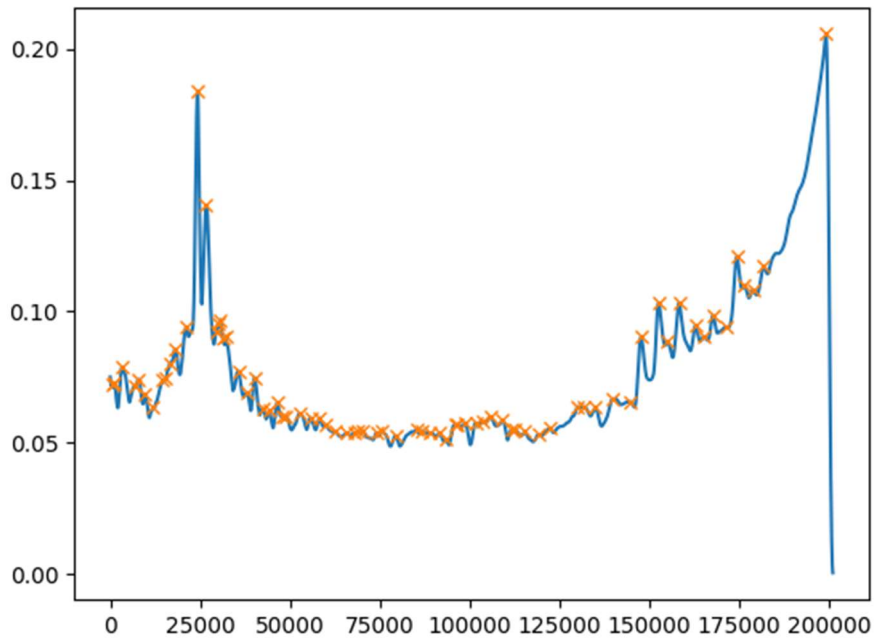


Figure 4.27: *Stribeck* with peaks detected setting $\text{prominence}=0.000125$ (95th percentile).

The two plots show that the 75th percentile returns a more helpful result since the other value gives several peaks too large.

```
peaks = []  
# Find peaks  
peaks, _ = find_peaks(dataframe.mue, prominence=0.007)  
num_peaks = peaks.size
```

For the construction of the Random Forest model, this value is not used as a feature. To include the number of peaks, it is necessary to have as an output of the model the complete Stribeck curve, which was out of the scope of this work. Furthermore, from an engineering point of view, it is more interesting to have information regarding the maximum and minimum friction coefficient rather than all the curve points.

4.7 Inputs and labels of the model

In order to build and then choose the Machine Learning model, its inputs and outputs must be defined. The inputs of the model are listed below:

- The lubricant type and its main characteristics: density, viscosity, and pour point
- Experimental setup: test type, duration, surface pressure, and speed
- Surface topography information: if the material has been fine ground or lapped and if the surface finishing is fine or rough, surface roughness (Ra)
- Brass material employed and its main parameters: Young modulus (E), Poisson ratio, Yield strength, Hardness (Brinell scale HBW)

4.7.1 Density

Both the performance of lubricants and the operation of machines are significantly influenced by density. In addition, a fluid's density impacts pump efficiency because pumps are typically made to move fluids of a particular density.

According to Layman's algorithm, an object's density is determined by how much mass it has in its space. Therefore, according to the following formula, density, mass, and volume are mathematically connected:

$$\rho = \frac{m}{V} \quad (4.6)$$

Where ρ is the density, m corresponds to the mass, and V denotes the volume.

The fluid's erosive potential must be considered because density increases along with it. Therefore, pipes, valves, or any other surface in the fluid's path may erode in areas of a system with significant turbulence or high velocity [3].

The density of a fluid affects impurities like air and water in addition to solid particles. Both pollutants significantly impact density. Additionally, oxidation affects a fluid's density. The density of the oil rises as the oxidation process continues. Even though it has a more negligible effect than other factors, the lubricant density affects the film thickness of the lubricant, resulting in the lubrication regime.

4.7.2 Viscosity

A fluid's most crucial physical characteristic is its viscosity, which describes its resistance to flow and shear. Because of fluid's molecular interactions, the molecules that make up fluid experience friction when it is flowing. It indicates that when the same force is applied, different fluids with different viscosities flow at various rates. It is influenced by several factors, including oil ageing and contamination with water, particulates, or other lubricants. The viscosity or ease with which oil flows at a given temperature is also related to weight. Figure 4.28, which shows two flat surfaces divided by a film with a thickness of h , can help explain it better.

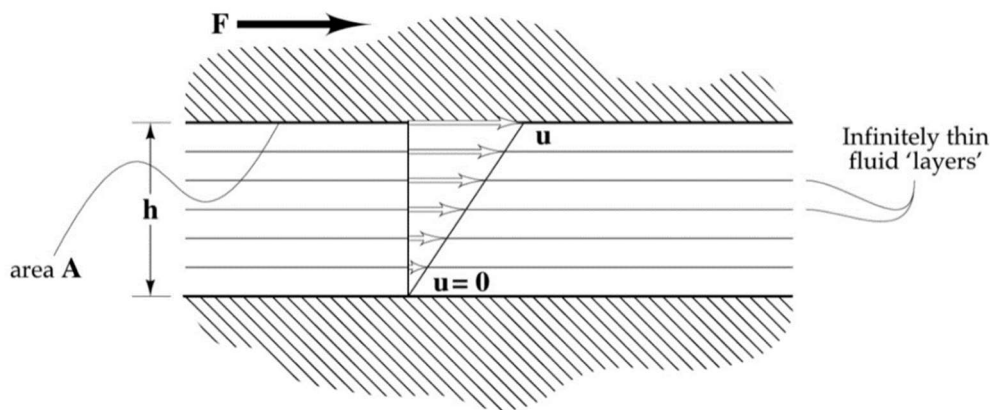


Figure 4.28: Schematic representation of the fluid separating two surfaces [2].

The two types of viscosity are:

1. Dynamic viscosity (μ), also called absolute viscosity, provides details on the amount of pressure required to cause the lubricant to flow. The Pascal-second (Pa s), or the equivalent centipoise, is its physical unit (cP). It is clarified that this is because

tribological calculations consider the lubricating property. Shear stress is calculated as the coefficient of dynamic viscosity times shear rate [1]. The shear stress can be expressed as:

$$\tau = \frac{F}{A} \quad (4.7)$$

Where F is the needed load to displace the top surface and A is the portion of the upper layer that the lubricant has moistened (see Figure 4.28).

For Newtonian fluids, the kinematic viscosity, η , serves as the proportionality coefficient, and the shear stress, τ , is dependent on the velocity gradient, u/h [2]:

$$\tau = \eta \cdot \frac{u}{h} \quad (4.8)$$

2. Kinematic viscosity (ν):

$$\nu = \frac{\eta}{\rho} \quad (4.9)$$

Square millimetre per second (mm^2/s), or its counterpart in centistokes, is its physical unit (cSt) [2]. It indicates the rate of lubricant flow in response to force. In industry, kinematic viscosity is frequently utilised and reported.

A lubricant's viscosity is influenced by temperature; the higher the heat, the lower the viscosity. Depending on the lubricant's composition, the viscosity will change when the temperature increases to a different extent.

The rate at which a lubricant's viscosity changes owing to a temperature change is measured by its viscosity index (VI). The latter is determined by measuring the kinematic viscosity at 40°C and 100°C . The results of two reference oils are compared to these measures after that. For typical mineral oil, the viscosity index varies from 95 to 100. A lubricant's viscosity varies with temperature, pressure, and, in some situations, the rate of shearing. Most liquids have viscosities that increase with pressure and decrease with temperature. The graphic below illustrates how the viscosity of two variables. Most liquids' viscosities rise with pressure and decrease with temperature, respectively. The graphic below illustrates how temperature affects the viscosity of two different lubricants. Because it produces a more stable lubricating coating across a more comprehensive temperature range, a lubricant with a higher viscosity index is preferable.

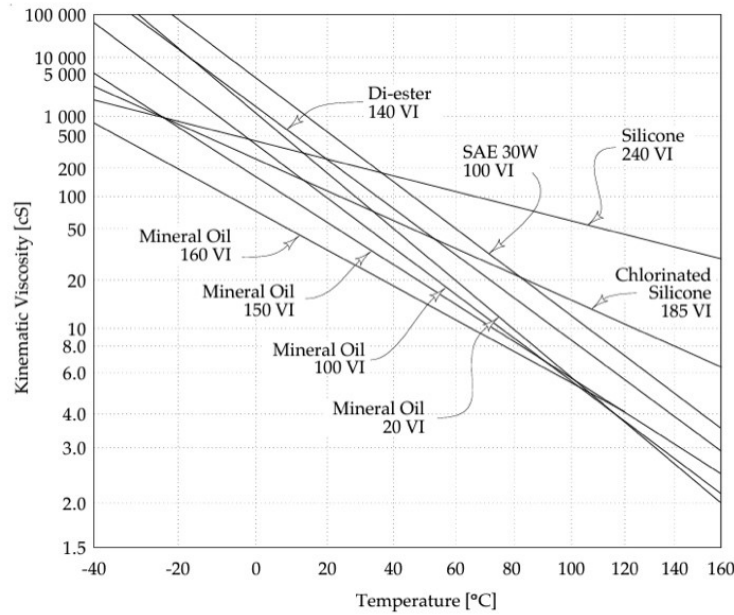


Figure 4.29: Viscosity as a function of the temperature [2].

The Vogel equation is the viscosity-temperature relationship that is most precise [2]:

$$\eta = a \cdot e^{\frac{b}{T-c}} \quad (4.10)$$

Some oils may have the ideal viscosity at some temperatures, but they do not precisely satisfy the requirements at both ends of the temperature range. Remember that even a little temperature differential might result in a significant viscosity shift that could harm the asset. High viscosity is not always desirable; viscous oils result in a thicker layer and greater separation of the mating surfaces (both have good benefits). They also increase the power needed to operate machinery due to shear stress. Therefore, if the viscosity is excessively high, there will be significant power losses and heat production. Greater mechanical friction may happen at excessively low viscosities, resulting in larger weights. Low oil flow might also ensue, leading to oil hunger and dry starts.

Additionally, more significant wear from film loss would be visible. The viscosity index can be made better by adding substances. The latter makes it possible to design a lubricant that meets the original equipment manufacturer's requirements. To choose a suitable lubricant, it is always a good idea to consult the equipment manufacturer's requirements and consider the weather and working environment [27]. The machinery will operate reliably and for the longest possible time with appropriate installation and a precise viscosity index. Additionally, choosing a lubricant with the proper viscosity reduces wear and friction.

Most liquids have viscosities that rise with pressure and, as was previously mentioned, fall with temperature. These changes are significant in most lubricants, including mineral and synthetic oils. The influence of pressure on viscosity (e.g., gears) is particularly significant in the lubrication of severely loaded concentrated contacts [27]. The Barus equation is the formula that is most frequently used to determine lubricant viscosity at pressures close to atmospheric pressure. This model performs pretty well when the pressure is less than 0.5 GPa.

$$\eta = \eta_0 \cdot e^{\alpha \cdot p} \quad (4.11)$$

Where η [Pa · s] is the viscosity at pressure p ; η_0 [Pa · s] the one at atmospheric pressure; α [m²/N] is a lubricant characteristic; p [Pa] is the operating pressure [2].

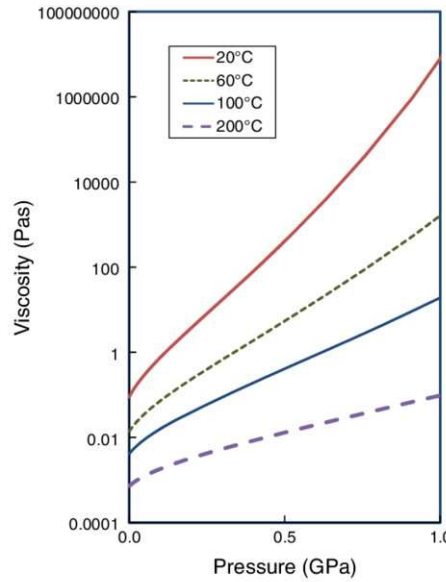


Figure 4.30: Viscosity as a function of the pressure [27].

4.7.3 Pour point

In a particularly scientific test, the coldest temperature at which a specific oil is shown to flow by gravity, unobstructed, is known as the oil's pour point. Specifically, the pour point is 3 °C higher than the temperature at which oil does not move when a specimen is held straight for 5 seconds. The pour point of oil reveals details about its cold-temperature properties. The latter may not impact the lowest temperature at which a hydraulic oil can operate, the lowest temperature at which gears are sufficiently lubricated, or the lowest temperature at which oil will flow to an item of equipment's bearings. Therefore, it cannot be the deciding element when choosing a lubricant [2].

The terms “pour point” and “solidification point” are frequently used interchangeably. The temperature at which oil must solidify entirely is known as the solidification point. The latter is a few degrees Celsius below the pour point.

The pour point can be defined as the flow point or the behaviour of oil when it is cold. The pour point ranges from roughly -65 °C to -10 °C, depending on the type of lubricant. The oil grows viscous until it solidifies into a waxy substance. The paraffin in oil crosslink and crystallise when it freezes; therefore, the oil turns hazy and milky. Because the oil is no longer evenly dispersed to the lubrication spots when it solidifies, the lubricating capabilities degrade. A lubricant's pour point can be lowered by additives like pour point improvers. These postpone the paraffin crystals' contraction.

4.7.4 Young modulus

Young's modulus was first proposed by Euler in 1727, despite being named after the British scientist from the 19th century. Young's modulus was initially applied in tests in the 1780s in its current form. The Latin word *modus*, meaning measure, is the source of the word modulus. Elastic modulus E is a mechanical characteristic that evaluates the flexibility of bulk objects under tensile or compressive loads when the force is delivered longitudinally. It is sometimes referred to as Young's modulus in tension or compression (i.e., negative tension). The connection between strain rate (corresponding displacement) and tensile/compressive normal stress (load per unit surface) in a structure's straight tensile range is described by the formula:

$$E = \frac{\sigma}{\varepsilon} \quad (4.12)$$

Young moduli are frequently so big that their units of measurement are gigapascals rather than pascals (GPa).

4.7.5 Poisson ratio

The axial to transverse strain ratio divided by one is Poisson's ratio. In structural dynamics and materials engineering, the Poisson effect is quantified by Poisson's ratio ν , which represents the distortion of substances in orientations orthogonal to the specific stress distribution. The amount of axial compression divided by the amount of transversal elongation is known as ν for minor values of these changes.

$$\nu = -\frac{d\varepsilon_{trans}}{d\varepsilon_{axial}} = -\frac{d\varepsilon_y}{d\varepsilon_x} = -\frac{d\varepsilon_z}{d\varepsilon_x} \quad (4.13)$$

Where ε_{trans} is the transverse strain and ε_{axial} is the axial one. Positive strain means stretching, whereas negative means contracting.

4.7.6 Yield strength

A measure of the most significant stress that can be applied to a substance without generating plastic deformation is known as yield strength. The elastic limit of a material can be roughly approximated by the stress point at which it permanently deforms. The material will elastically deform before it reaches the yield point, but once the applied stress is removed, it will always return to its initial shape. A tiny portion of the deformation encountered once the yield point is exceeded will become permanent and irreversible.

Yielding is a gradual failing technique that is frequently less disastrous than ultimate failure. The stress corresponding to the point at which a material reaches its yield strength and begins to deform plastically is called yield stress. Since it reflects the most prominent rate of stresses that could be placed on a component without causing irreversible deformation, the yield strength is typically used to calculate the maximum weight imposed on a machine part. However, it may be challenging to determine the precise yield point because non-linear behaviour evolves gradually in some materials, like aluminium. This scenario assumes that the offset yield point is the tension where 0.2% plastic deformation occurs (or proof stress).

When developing and producing components, it is essential to comprehend a material's yield strength since it indicates the maximum load that the material can withstand. Since many materials are produced via pressing, rolling, or forging, yield strength is crucial. The temperature often causes a drop in yield strength, whereas the strain rate causes a rise. When the former occurs, a condition known as “yield strength anomaly”, frequently observed in super-alloys, is considered present in the material. When muscular strength at high temperatures is required, materials in this category are frequently used. When creating structures that may be exposed to unanticipated impact loads, such as earthquakes, the yield strength of a material is particularly crucial. Because the plastic area of the substance contributes a significant portion of the energy that the substance absorbs under these

pressures, it becomes crucial. Therefore, a material's capacity to bear unexpected pressures and loads for a longer duration will enable the implementation of safety measures.

4.7.7 Hardness

The ability of a material to tolerate localised plastic deformation brought on by mechanical indentation or abrasion is known as hardness. It could be particularly crucial when trying to find appropriate material for an environment with tiny particles that might cause material wear. Soft materials are indented instead of hard materials, which resist any shape change. It gauges how well a substance resists localised, continuous deformation. Plastic deformation is another name for permanent distortion. When a material deforms elastically, it does so only when force is applied, whereas when it deforms plastically, it does not change back to its previous shape. Strength, toughness, viscoelasticity, plasticity, strain, ductility, and elastic stiffness are all factors that affect hardness. Although the behaviour of solid materials under stress is complicated, strong intermolecular interactions are referred to as macroscopic hardness. As a result, several methods for determining hardness, such as scratch, indentation, and rebound hardness. Only a specific set of instruments may be used to determine the hardness of each kind. Additionally, depending on the abovementioned types, different hardness levels will apply to the same material.

- *Indentation Hardness:* This hardness category defines a material's resistance to continuous deformation under continual stress. When engineers and metallurgists discuss hardness, they often mean indentation hardness [28]. Establishing its magnitude is crucial since constant stress is alloying' most common form of tension.
- *Scratch Hardness:* This hardness has to do with a substance's ability to resist surface abrasions. The top layer of a surface develops scratches when it comes into touch with a hard, pointy object. Brittle materials like ceramics are frequently subjected to scratch testing since they do not experience much plastic deformation. Since some material applications are susceptible to scoring, scratch hardness is a crucial factor to consider [28]. As an illustration, consider the situation of an engine cylinder lining. Scratching or scoring may occur for a variety of reasons. Numerous metals, including piston rings, foreign materials in fuel, and lubricating oil, come into touch with the surface of the liner. Inadequate liner seats can also be an issue. The abrasive particles

may result in scratches that eventually reduce engine efficiency and increase long-term maintenance, replacement component costs, and fuel consumption. Therefore, excellent metal should be chosen at the design stage, considering the hardness of the contacting materials.

- *Rebound or Dynamic Hardness*: Elastic hardness rather than plastic hardness is more closely related to rebound hardness. The rebound hardness value increases with the distance from the initial dropping height [28]. The substance absorbs the energy from the impact and transfers it back to the indenter, a standard hardness testing device.

The SI unit of hardness is N/mm^2 . Thus, hardness is also measured in units of Pascal, although hardness must not be mistaken for pressure. Several scales of measurement for the various types of hardness are covered above. The units are not appropriate for direct comparison because they are formed from these measuring techniques. However, a conversion table may always be used to compare different hardness technique values. Even if such tables are incorrect, they provide a reasonable generalisation.

Brinell Hardness	Rockwell	Rockwell	Vickers
HB [N/mm^2]	HRC [N/mm^2]	HRB [N/mm^2]	HV [N/mm^2]
469	50	117	505
430	46	115	450
360	39	111	376
250	25	101	255
160	3	84	160

Table 4.6: Conversion table for different hardness scales of measurement [28].

The commonly used units for hardness measurement are [28]:

- Brinell Hardness Number (HB)
- Vickers hardness number (HV)
- Rockwell hardness number (HRA, HRB, HRC)
- Leeb hardness value (HLD, HLS, HLE)

The one used in this work is the HBW hardness. HBW stands for Hardness Brinell Wolfram carbide. The usage of tungsten carbide balls is required by updated Brinell standards, not the (softer) steel balls that were previously used, according to wolfram carbide (also known as tungsten carbide) (HBS). As a result, values will vary as hardness increases.

The labels, instead, are constituted by the parameters extrapolated from the Stribeck curves:

- the minimum coefficient of friction (μ_{min});
- the maximum coefficient of friction (μ_{max});
- gradient before μ_{min} ;
- gradient after μ_{min} ;
- speed corresponding to μ_{min} ;
- weight loss of the materials.

The outputs account for all the parameters extracted from the Stribeck curves employing the techniques described in the previous sections. The weight loss measurements are carried out in the laboratory after the experiments with the tribometer.

4.8 Hot encoding and normalisation

In order to employ and choose the proper machine learning model, it is necessary to hot encode and normalises the datasets.

4.8.1 Hot encoding

Variables with categorical data, often known as “nominal” data, have label values rather than numerical values. The range of potential values is frequently fixed, and each value denotes a distinct category. Many machine learning algorithms cannot work directly with label data; instead, they need all input and output variables to be numbers of functioning. Generally speaking, this imposes more restrictions on the effective use of machine learning algorithms than on the techniques themselves. So, categorical data must be transformed into numerical data. In order to show or apply predictions made by the model in an application, it might also be necessary to transform the categorical variable back into a categorical form if it is an output variable. In one-hot encoding, a new binary variable is inserted instead of the integer

encoded variable for each distinct integer value. In other disciplines, including statistics, the binary variables are frequently called “dummy variables”.

Consider values as several categories that occasionally exhibit a natural ordering. Depending on their implementation, specific machine learning algorithms, such as decision trees, can function directly with categorical data. However, the majority need all input and output variables to have a numerical value. Any categorical data must thus be converted to integers. Data can be transformed utilising one hot encoding to improve forecasting and prepare the data for an algorithm. After constructing a new category column for each categorical value using one-hot, a binary value of 1 or 0 is assigned. A binary vector represents each integer value. The table below demonstrates that all values are zero and that the index is represented by a 1.

ID	Colour	One Hot Encoding →	ID	Colour_Red	Colour_Blue	Colour_Green
1	Red		1	1	0	0
2	Blue		2	0	1	0
3	Green		3	0	0	1
4	Blue		4	0	1	0

Table 4.7: Hot encoding example.

Applying this to a real-world scenario, Take the values red and blue, for example. With one-hot, the colours red and blue are associated with a numerical value of 0 and 1, respectively. It is essential to apply these ideals consistently. The latter enables reversing the encoding and getting the original category data. After assigning numerical values, a binary vector is built to represent those values. Given the two values in this situation, the vector will have a length of 2. As a result, the binary vector $[1,0]$ may represent the red value, whereas $[0,1]$ will represent the blue value.

One hot encoding is valid for variables that are unconnected to one another. Number order is regarded as a crucial attribute for algorithms used in machine learning. In other words, they will see a more significant amount as higher or more meaningful than a smaller number. The lack of ordering for category values in some input data can cause problems with predictions and poor performance, even if this is useful in some ordinal scenarios. A single hot encoding then saves the situation. The training data is improved by a single hot encoding, simplifying rescaling. Utilising numerical numbers makes it simpler to calculate a probability for the

values. Since it offers more precise predictions than single labels, one hot encoding is employed for the output values.

4.8.2 Normalisation of the dataset

In machine learning, feature scaling is one of the most crucial data preprocessing procedures. Distance computation algorithms are skewed towards numerically more significant values if the data is not scaled. On the other hand, the magnitude of the features is unimportant to tree-based algorithms. Additionally, feature scaling promotes faster deep learning, machine learning training, and convergence. Some feature scaling methods, like Normalisation and Standardisation, are the most well-liked and, simultaneously, the most perplexing.

Min-Max or normalisation Features are changed to be on a similar scale by scaling. The new point is determined as follows:

$$X_{new} = \frac{X - X_{min}}{X_{max} - X_{min}} \quad (4.14)$$

The dataset may also scale the range to $[-1, 1]$ if the dataset contains negative values. When the distribution is unknown, it is helpful. The n-dimensional data is geometrically compressed into an n-dimensional unit hypercube via transformation. Since it cannot handle outliers, normalisation is beneficial when there are none.

Z-Score or standardisation: The normalisation process involves transforming features by deducting from the mean and dividing by the standard deviation. Z-score is another name for this.

$$X_{new} = \frac{X - mean}{Std} \quad (4.15)$$

Standardisation may be helpful when the data has a Gaussian or Normal distribution. Nevertheless, this does not need to be the situation at all times. It agrees with the normal standard distribution (SND). It converts data to the linear combination and collapses or expands the points according to whether the normal is 1 or 0. However, because there is no fixed range of modified characteristics, outliers do not significantly impair standardisation.

The dataset can be pre-processed using either StandardScaler or MinMaxScaler after removing the outliers with RobustScaler. Because it is resistant to outliers, RobustScaler employs the interquartile range. It scales features with resilient to outlier statistics. This technique removes

the median; the data is scaled between the 25th and 75th percentile. An interquartile range may also be used to describe this range. In order to use the median and interquartile range with subsequent data using the transform technique, they are then saved. The median and the interquartile range beat the sample mean and variance when outliers are present in the dataset and offer superior results. Hence the following is the formula:

$$\frac{x_i - Q_1(x)}{Q_3(x) - Q_1(x)} \quad (4.16)$$

It was ultimately decided to use normalisation. It can comprehend the decision only to join the necessary tables. A primary external activity may be maintained as duplicate information is removed through standardisation. Generally speaking, this shrinks the amount of the knowledge base. Improved execution is guaranteed, which is related to the first statement. Reaction time and speed are improved when information bases are smaller because sorting through the information is faster and more concentrated. Narrower tables are achievable by modifying standard tables to have smaller sections, which enables plenty of data per page. Fewer files per table ensure quicker support assignments.

The complete dataset has been split between the first and second Stribeck curves before selecting the best machine learning model. The second can be justified using the same logic and results. From this point forward, the initial curve will be used as the reference point for all calculations.

4.9 Choice of the machine learning model

A multi-output regression model must be constructed. Supervised learning is a subset of regression. It builds a model using a training dataset to forecast unknown or upcoming data. Because the desired output value is already known and a part of the training data, the term “supervised” is used to describe the process. Only the output value distinguishes the subcategories Regression and Classification. Regression provides continuous results, whereas classification splits the dataset into classes. Predicting a continuous-valued attribute linked to an item is the goal of this kind of model. The outputs of multi-output regression often rely both on the input and one another. The latter indicates that the outputs are frequently not

independent and that a model that predicts both outputs simultaneously or each output conditionally on the other outputs may be necessary.

4.9.1 Types of machine learning regression model

There exist a considerable number of different models. It has been chosen to compare 6 of the most used and standard machine learning models:

- Linear Regression
- Lasso
- Neural Network Regression
- Decision Tree Regression
- Random Forest
- Gaussian Process Regression

4.9.1.1 Linear regression

One of the models most frequently applied in the actual world can be linear regression. A linear approach depicts the relationship between a numerical output and one or more descriptive variables. According to mathematics, linear regression minimises the sum of residuals between each point and the predicted value [22]. The black line represents the model being solved, the grey points represent the original data, and the residual gap between the point and the red line is illustrated in the image below. Of course, the total amount of residuals must be reduced.

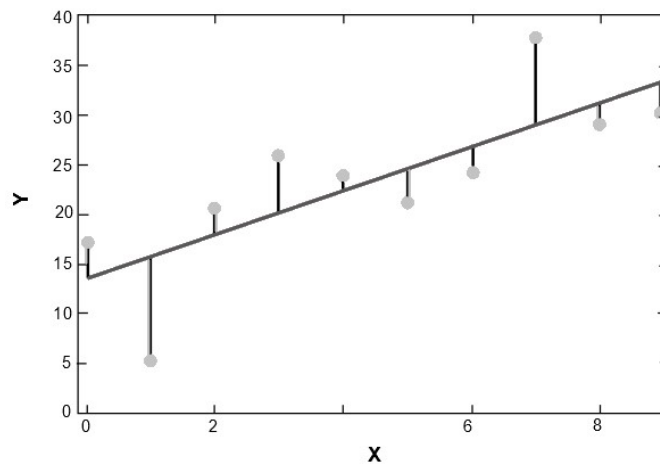


Figure 4.31: Linear regression.

Here is the sample form: The parameter to infer from the data is β , the characteristic or input data (dependent variable) is x , and the independent variable is y . The first equation is represented in vector form in the second equation [29].

$$f(x_i) = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_n x_{in} = x_i^T \beta + \varepsilon \quad (4.17)$$

The assumptions made by linear regression algorithms are that the input and output have a linear relationship.

The python code to use the technique is:

```
lr = LinearRegression()
lr.fit(X_train, y_train)
y_pred = LR.predict(X_test)
```

4.9.1.2 Lasso

Lasso regression, often referred to as the Least Absolute Shrinkage and Selection Operator, is another variation of linear regression. Lasso adjusts the loss function to simplify the model by limiting the combination of the absolute values of the model coefficients (also called the l1-norm) [29].

$$\sum_{i=1}^n (y_i - \sum_j x_{ij} \beta_j)^2 + \lambda \sum_{j=1}^p |\beta_j| \quad (4.18)$$

Where λ represents the degree of shrinking, increasing it results in a decrease of bias, while a decrease implies an increase of variance.

In python:

```
reg = linear_model.Lasso(alpha=0.1)
reg.fit(X_train, y_train)
y_pred = reg.predict(X_test)
```

4.9.1.3 Neural Network

With a sophisticated design resembling the human nervous system, artificial neural networks are computing systems that aim to replicate human learning skills. One input layer and one neuron make up a perceptron. The input layer serves as the dendrites and is in charge of taking in data. The input layer's node count is the same as the input dataset's feature count. The results are mixed after each input is multiplied by a weight (normally initialised with a random number) [22]. Following that, the sum is run through an activation function. The

latter does so and produces an output after processing the data. One can alter the last activation function to convert a neural network into a regression model. This output represents the outcome in the instance of a perceptron. On the other hand, multilayer perceptrons use the output from the neurons in the preceding layer as the input to the neurons in the following layer. Neural network regression ensures non-linearity by mapping a neuron's output to a range of values.

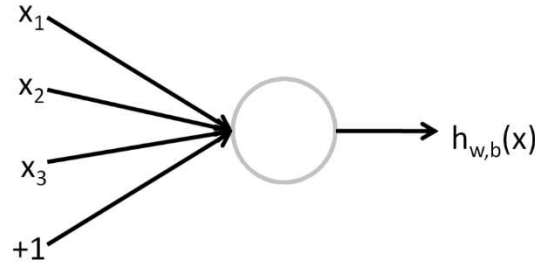


Figure 4.32: Single neuron representation.

$$y = h_{w,b}(x) = f(b + w^T x) = f\left(b + \sum_{i=1}^n w_i x_i\right) \quad (4.19)$$

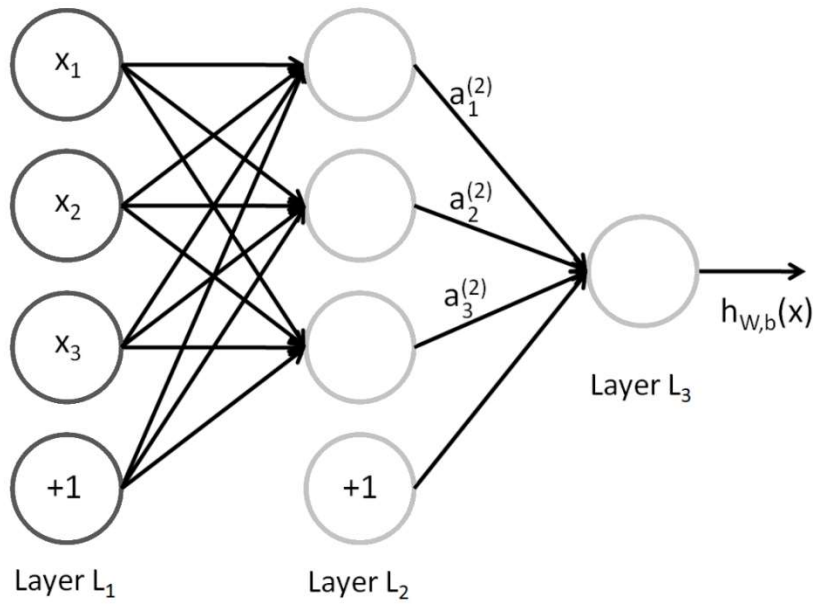


Figure 4.33: Neural Network.

Input $x \in R^n$; $f: R \rightarrow R$ is called the activation function.

The accuracy of single-layer perceptrons drastically declines when dealing with data that cannot be separated linearly [29]. On the other hand, multilayer perceptrons can effectively process non-linearly separable input. A single parameter or a set of parameters can be selected to predict the outcome using neural network regression. The neurons (a neural network's

outputs) are closely coupled, and each neuron has a weight. The interconnected neurons aid in mapping the relationships between dependent and independent factors and future values.

In python:

```
NN = MLPRegressor(max_iter=300, activation="relu",
hidden_layer_sizes=(100, 100))
NN.fit(X_train, y_train)
y_pred = NN.predict(X_test)
```

4.9.1.4 Decision tree

Non-linear regression in machine learning can be carried out via decision tree regression. The dataset must be divided into smaller manageable portions before the decision tree regression method can do its primary work. The dataset subgroups depict the values of all the data points. This algorithm creates a decision tree by splitting the information set into determination and leaf nodes [29]. When the data set has not undergone significant change, ML specialists like this model should be aware that even a slight modification in the data might significantly impact the decision tree's structure. There will not be enough end nodes left to produce the forecast if the decision tree regressors are pruned too much. Therefore, one should not over-prune the decision tree regressors to have many end nodes (regression output values).

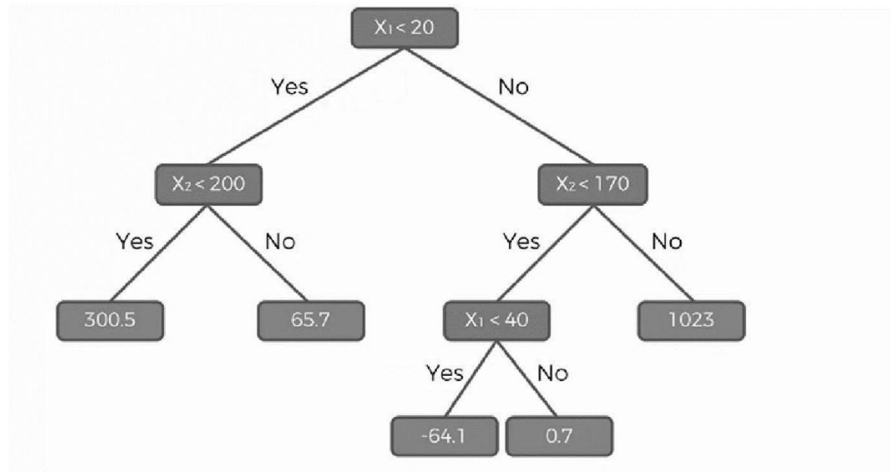


Figure 4.34: Decision tree regression.

In python:

```
reg = DecisionTreeRegressor()
reg.fit(X_train, y_train)
y_pred = reg.predict(X_test)
```

4.9.1.5 Random Forest

The model's accuracy can frequently be significantly improved by combining numerous uncorrelated Decision Trees. Random Forest is the name of this process. As the trees grow, they are affected by several random processes (randomisation). The average of the trees is reflected in the final model. Another popular approach for non-linear regression in machine learning is random forest. A random forest uses many decision trees to forecast the outcome instead of decision tree regression (single tree) [22]. With the help of this approach, a decision tree is constructed using k randomly chosen data points from the given dataset. The value of every new data point is then predicted using several decision trees. A random forest method will anticipate many output values because there are numerous decision trees. The average of all the anticipated values must be calculated to get the final result for a new data point. The sole disadvantage of utilising a random forest approach is that more training data is needed. The latter occurs because this technique maps many decision trees, which use more significant computational resources.

In python:

```
reg = RandomForestRegressor(max_depth=2, random_state=0)
reg.fit(X_train, y_train)
y_pred = reg.predict(X_test)
```

4.9.1.6 Gaussian Process Regression

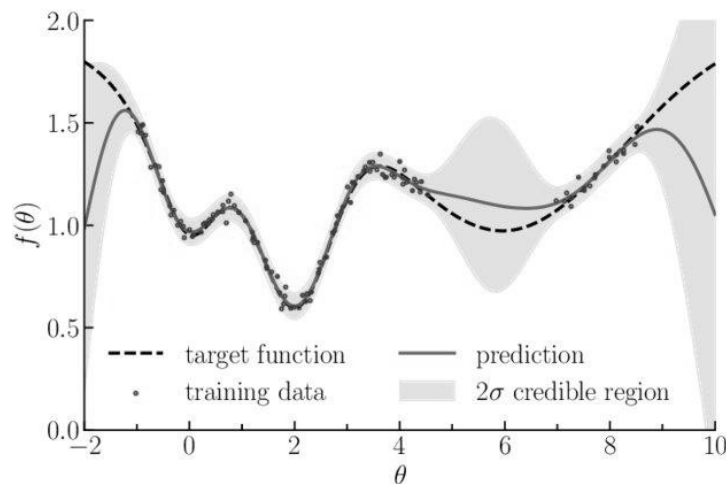


Figure 4.35: Gaussian Process Regression [29].

A nonparametric Bayesian technique is presently revolutionising regression analysis. GPR has various benefits, including quantifying prediction uncertainty and performance on small

datasets. Instead of learning specific values for each parameter of a function, as is the case with many popular supervised machine learning techniques, the Bayesian approach insinuates a probability distribution over all possible values. For example, assume that $y=wx+$ is a linear function. The Bayesian approach works as follows by establishing a prior distribution, $p(w)$, on the attribute, w , and moving probability in line with evidence (i.e., observed data) [29]:

$$p(w|y, X) = \frac{p(y|X, w)p(w)}{p(y|X)} \quad (4.20)$$

As a result, both the prior distribution and the dataset are used to update the distribution, denoted as $p(w|y, X)$ and known as the posterior distribution. Weighing all potential predictions according to their computed posterior distribution will yield predictions at unseen sites of interest, x^* , from which the predictive distribution can be calculated [29]:

$$p(f^*|x^*, y, X) = \int_w p(f^*|x^*, y, X)p(w|y, X)dw \quad (4.21)$$

For the integration to be manageable, it is typically assumed that the prior and likelihood are Gaussian. It may be obtained as a level estimate using the average and a confusion measurement using the variance from a Gaussian distribution by implying the prediction probability under that assumption and calculating for it.

In python:

```
kernel = DotProduct() + WhiteKernel()
gpr = GaussianProcessRegressor(kernel=kernel, random_state=0)
gpr.fit(X_train, y_train)
```

4.9.2 Train and test split

For each of the above programmes, there has to be a separation between the train and test runs. Unsupervised machine learning has several vital components, but one of the most crucial is the assessment and validation of models. When evaluating the predictive power of the model, objectivity is essential. Use `train test split()` from the machine learning package `scikit-learn` to split your dataset into subsets to lessen the chance of bias across the evaluation and validation process. The command used here is:

```
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2)
```

The objective of supervised machine learning is to develop models that consistently connect the given inputs (also known as independent variables or predictors) to the defined outputs (dependent variables or responses).

Depending on the issue, it is necessary to define methods to measure the model's correctness. It is crucial to understand that to apply these metrics effectively, evaluate the model's capability for prediction and verify it, an objective review is usually required.

The latter implies that the exact data for training and assessing a model's prediction performance is not feasible. Instead, the model must be tested with new data it has never seen before.

The dataset must be divided to evaluate prediction performance, typically into three sections at random [29]:

1. The training set is used to train or fit the model.
2. The testing set is used to evaluate models objectively during hyperparameter tuning.
3. An objective assessment of the final model requires the test set. Therefore, it is not recommended for fitting or validation.

Splitting a dataset may also be crucial for identifying if the model is underfitting or overfitting, two extremely prevalent issues [22]:

- Underfitting typically results from a model's inability to capture the relationships between the data. As a result, poorly fitted models are likely to perform poorly on training and test sets.
- Overfitting often occurs when a model learns the current relationships between the data and noise and has an excessively complicated structure. As a result, these models cannot frequently generalise well. As a result, they often perform poorly with test data while having good performance with training data.

4.9.3 Scoring metrics

The performance or competency of a regression model must be revealed as inaccuracy in those estimations. Scoring metrics gauge a scoring setup's or the service's overall performance.

4.9.3.1 Mean Squared Error

Mean squared error, or MSE for short, is a standard error measure for regression problems. In this application, “least squares” means minimising the mean squared error between the predictions and the expected values. The MSE is calculated using the average squared deviations between a dataset's predicted and estimated output values.

$$\frac{1}{n} \sum_{i=1}^n (y_i - f(x_i))^2 \quad (4.22)$$

The two equation variables in the dataset stand for the i-th expected, and i-th projected values, respectively. Therefore, these two integers have a squared difference, eliminating the sign and yielding a favourable estimation error [29].

Significant errors are also exaggerated or magnified as a consequence of the squaring. In other words, the final squared positive error is inversely proportional to the size of the difference between the predicted and expected quantities. Models are, therefore, more severely punished for making more mistakes when MSE is employed as a loss function. When used as a statistic, it also punishes models by increasing their average error scores.

4.9.3.2 Root mean squared error

The RMSE is determined, which is significant since it means that its unit and the goal value that is being forecasted are the same. To compute the RMSE [29]:

$$\sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - f(x_i))^2} \quad (4.23)$$

Where i is the i-th projected value and y_i is the i-th expected value in the collection. The square operation is used by MSE to eliminate the sign from each error value and to penalise significant mistakes. Although it assures that the outcome is still positive, the square root reverses this action.

4.9.3.3 Mean absolute error

Similar to RMSE, the entities of the inaccuracy score reflect the measures of the target result that is being anticipated, making MAE a popular statistic. In contrast to the RMSE, the changes in MAE are continuous and hence evident. In other words, MSE and RMSE increase

or amplify the statistical error score by penalising more significant errors more severely than minor errors. The quadratic of the wrong number is to blame for this. The MAE does not give different types of errors more or less weight; instead, the scores increase linearly as the amount of mistakes increases. When computing the MAE, the difference between an anticipated and forecasted value is compelled to be positive. The MAE may be computed as follows:

$$\frac{1}{n} \sum_{i=1}^n |y_i - f(x_i)| \quad (4.24)$$

Where $f(x_i)$ is the i -th projected value and y_i is the i -th anticipated value in the dataset [29].

4.9.3.4 *R-squared*

R^2 is only a measurement of the degree of correlation between two variables in its most basic form. It is the reciprocal square of the correlation between x and y [29].

$$\sqrt{R^2} = r(x, y) = \frac{\varepsilon\{(x - \bar{x})(y - \bar{y})\}}{\sqrt{\nu\{x\}\nu\{y\}}} = \frac{Cov\{x, y\}}{\sqrt{\nu\{x\}\nu\{y\}}} \quad (4.25)$$

Where ε stands for *expected value of*, ν for *variance of*, and Cov for *covariance of* an operation. R^2 may be considered the ratio of the overall sum of squares (TSS) to the regressed squared sum (RegSS). Because each “sum of squares” value is essentially a rescaled variance, the following calculation shows that R^2 is just the proportion of two variances [29]:

$$R^2 = \frac{RegSS}{TSS} = \frac{\sum_i (\hat{y}_i - \bar{y})^2}{\sum_i (y_i - \bar{y})^2} \quad (4.26)$$

If the numerator and denominator are equal, as in the case of an R^2 value of 1.0, then the predictions, \hat{y}_i , are precisely equal to the initial values, y_i . According to the numerator, the forecasts are essentially identical to a horizontal line or the mean value of y , irrespective of the input signal x . However, for R^2 to equal 0, the integrand must equal zero. When $R^2 = 0$, the poorest prediction for y in this method just outputs the mean of the training data as a whole.

The following section addresses why this score may sometimes be significantly deceptive when modifying the specified machine learning model. Therefore, to pick the most suited model, it is advisable not to use it.

4.9.4 Scoring metrics results

The models are applied to the dataset, and the different models are evaluated with the scoring mentioned above metrics. The code takes 16 random rows from the dataset and uses them to test the model. For example, the MSE value does not refer to the error in predicting a single row (an individual ID) but refers to the prediction of 16 random rows, and the same happens in the other rows of the table. After that, all the values of the scoring metrics were averaged.

	<i>Linear Regression</i>	<i>Lasso</i>	<i>Neural Network</i>	<i>Decision Tree</i>	<i>Random Forest</i>	<i>GPR</i>
MSE	0.0448	0.0477	0.0474	0.0799	0.0383	0.0454
MAE	0.1630	0.1696	0.1667	0.2140	0.1530	0.1641
RMSE	0.2110	0.2175	0.2165	0.2809	0.1922	0.2122
R²	-0.1181	-0.1004	-0.2250	-1.0465	-0.0216	-0.0894

Table 4.8: Scoring metrics results for the different machine learning methods.

4.10 Random Forest regressor

An RF divides on a selection of attributes at each split. The information may be divided into smaller units according to the data's characteristics until it is present in a sufficiently small set of data that only includes data points with a single label. A purity measure determines this split in a decision tree model. In other words, the information acquisition has to be maximised at each node. As part of the regression scenario, they begin at the tree's root and split it based on the outcome variables. They then reach the leaf node and present the results. When dealing with regression problems, the residual sum of squares (RSS) is used, while when dealing with classification problems, it is considered the Gini index or entropy.

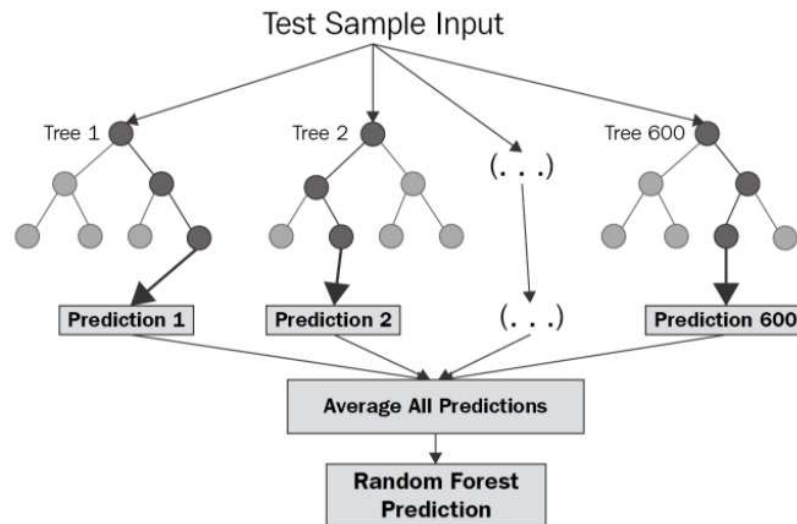


Figure 4.36: Random Forest regressor [29].

The diagram above shows how a Random Forest is structured. It is clear that the trees are parallel to one another and do not interact at all. An RF constructs several decision trees during training and delivers the average of the categories as the prediction of the entire set of trees. The steps of the algorithm are [29]:

1. Take k random data points from the training set and analyse them.
2. Then, create a decision tree using these k data points.
3. To construct N trees, repeat steps 1 and 2.
4. Finally, compute the mean of all predicted values y for every N -tree tree in the model that projects y values for just a new information point.

The best outcomes are obtained when many highly unrelated models collaborate. The poor correlation between models is the key to using Random Forests effectively. Uncorrelated models can produce ensemble forecasts that are more precise than every independent prediction. This remarkable effect is due to the trees' ability to protect one another from another's errors. Many trees will be suitable, while some may be wrong, allowing the group of trees to travel in the proper direction. The following conditions must be met for the random forest to be effective:

- For models created utilising those attributes to perform better than guesswork, there must be some actual signal in those features.
- Low correlations between the separate trees' predictions (and thus the mistakes) are required.

The random forest employs the following two techniques to make sure that each tree's behaviour is not too connected with the performance of any other trees in the model [29]:

1. **Bagging (Bootstrap Aggregation):** In bootstrapping, subsets of a dataset are randomly sampled across a specified number of iterations and variables. These findings are then averaged to provide a more potent outcome. An example of an applied ensemble model is bootstrapping. Ensemble learning uses models trained on identical information and averaging their results to get a more accurate forecast. Bagging selects a random sample from the data collection. Due to this, each model is created utilising the samples (Bootstrap Samples) provided by the Original Data in conjunction with a replacement technique called row sampling. Bootstrap is a method of sampling rows with replacements. There is now a separate training process for each model, which provides results. A majority vote determines the final choice after integrating the results of all the models. Finally, an aggregated result is obtained by taking all the results and, based upon a majority vote obtaining a result.
2. **Future Randomness:** In a standard decision tree, while splitting a node, all feasible characteristics are examined, and select the one that produces the most significant difference between the data in the left and right branches. In contrast, only a random subset of characteristics is available to each tree in a random forest. Ultimately, this leads to less correlation between trees and increased diversity by forcing even more variety across the model's trees.

Let us now develop bagged trees using the decision tree idea and bootstrapping concepts. It may be produced bagged trees by building X decision trees that were initially trained using X bootstrapped training data, also known as bootstrap aggregating. The final anticipated value is the mean of all the X selection trees. Due to its propensity to overfit, a single decision tree is highly varied; however, the variance may be reduced by bagging or merging several weak learners into strong learners.

Random forest improves bagging since it de-correlates the branches by dividing them on a random selection of features. The model only considers a small subset of attributes at every bifurcation. In other terms, a subset of the available characteristics n ($m = \text{square root of } n$) is randomly selected. The latter is essential in order for a variance to be averaged out. Consider

what might occur if the data set had a few accurate predictions. There will be several created trees equal since these forecasters will be repeatedly selected at the greater extent of the trees. The result would be a significant link between the trees as a result. Finally, as was said before, RF is a bagged decision tree system that divides a subset of attributes on each split.

4.10.1 *Advantages*

- **Impressive Versatility:** It can handle binary, categorical, and numerical features. Pre-processing just requires a small amount of effort. There is no need to scale or modify the data.
- **Parallelisable:** They are parallelisable, which enables us to distribute the processing across other processors. As a consequence, computation times are reduced. Contrarily, boosted algorithms are sequential and would require more time to compute.
- **Excellent with High Dimensionality:** Given that data subsets are used in the simulations, random forests perform very well with high dimensional data.
- **Quick Prediction/Training Speed:** The model can easily handle hundreds of features since only a part of the input is utilised, which makes training it faster than using decision trees.
- **Robust to Outliers and Non-linear Data:** Outliers are handled by random forests by effectively binning them. It also does not care about characteristics that are not linear.
- **Handles Unbalanced Data:** For the class population, it has methods for balanced inaccuracy in disparate data sets. The more significant class will have a reduced error margin. In contrast, the lesser class will face a more significant error rate when the data set is uneven because random forest strives to minimise the overall error rate.
- **Low Bias, Moderate Variance:** Each decision tree is somewhat volatile and has a low bias. The random forest produces a system with a slight bias and a significant variance since it combines every tree and the variance.

4.10.2 *Disadvantages*

- **Interpretability of the model:** RF models are essentially opaque and difficult to understand.
- **For large data sets, the dimension of the trees could use a lot of RAM.**

- In addition, the hyperparameters
- should be modified because the model may tend to overfit.

4.10.3 *Parameters of the Python function*

The *RandomForestRegressor* model function's stages are listed below:

1. Determine the independent (x) and dependent (y) variables.
2. Create a training set and a test set from the dataset.
3. Using the entire dataset to train the Random Forest regression model. The model may be trained using the *.fit()* method, modifying the weights by the data values to improve accuracy. The *.predict()* function creates predictions once the model has finished training.
4. Prediction of the test set outcomes.

```
RandomForestRegressor(n_estimators=100, criterion='squared_error',  
                      max_depth=None, min_samples_split=2,  
                      min_samples_leaf=1, max_features=1.0,  
                      bootstrap=True, n_jobs=None, max_samples=None)
```

The model is constructed employing the *RandomForestRegressor* function from the *sklearn* module. Some of the crucial parameters of the function are highlighted below:

- *n_estimators*: how many decision trees are constructed by the algorithm before averaging the predictions. Typically, the more trees, the better the ability to understand the data. However, adding many trees can significantly slow the training process; therefore, a parameter search is required to locate the sweet spot.
- *criterion*: The criteria (loss function) used to determine model results can be chosen using this variable. It may be selected from several loss functions to find splits, including mean squared error (MSE), mean absolute error (MAE), and reduction in Poisson deviation. The default setting is MSE.
- *max_depth*: this determines each tree's maximum depth. The more splits a tree has, the more information it can collect about the data. It accomplishes this on a larger scale by drastically slowing the growth of the Decision Tree.
- *min_samples_split*: there is a minimum requirement for samples required to separate an internal node. Each node may be considered at least once, or every sample may be

considered at each node. Again, the number of trees is limited as this value is raised by considering more data at each node.

- *min_samples_leaf*: There must be a certain number of samples at each leaf node. The leaf of the decision tree is the final element. The only considered split points include at least min samples of leaf training on each left and right branch. The latter might impact softening the model, especially in regression. Similar to min samples split, this setting describes the smallest sampling number at the tree's base or the leaves.
- *max_features*: the most attributes the model can consider when dividing a node. The three main features that may be adjusted to increase the model's prediction ability are *auto/None*, *sqrt*, and *log2*. The first one will only use all the characteristics that each tree's logic requires. Here, it is just not imposing any limitations on the specific tree. For example, the command "sqrt" will calculate the square root of each run's feature count. Another option of this type for max features is *log2*.
- *bootstrap*: the model adheres to bootstrapping principles (defined earlier) since the default setting for this is True.
- *n_jobs*: it instructs the power unit on how many processors it is permitted to use by specifying the number of parallel jobs to be executed.
- *max_samples*: this option presupposes that bootstrapping is set to True; otherwise, it has no effect. This number determines the most significant size of each sample for each tree when True is present. Providing all the data to each decision tree in the Random Forest is not required.

4.10.4 *Tuning the Random Forest model*

The initial setup of the hyperparameters, with *max_depth* set to 2, and *random_state* set to 0, is described in section 4.9.1.5. However, as stated in the preceding justification, the Random Forest's other properties are crucial for having strong performance. Therefore, the first step of *GridSearch* is to make a search space, which is the set of values of the hyperparameters to search to find the best set of values. In order to implement it, a dictionary (in the code below the grid) must be defined.

```
grid = {'n_estimators': [200, 300, 400],  
        'max_features': ['sqrt', 'log2'],  
        'max_depth': [6, 9, 18, 32],  
        }
```

The names of the hyperparameters serve as the dictionary's keys. Additionally, a grid search object must be created, and parameters must be handed in for every piece of information related to the grid search. The model, a Random Forest Regressor model in this instance, is the first parameter. The search space dictionary and several scoring techniques are also present: the negative mean squared error, negative mean absolute error, and the r2 approaches. Finally, the refit is a crucial parameter that should be set to R2 so that the GS returns a model optimised for the r-squared measure.

Additionally, there is the 5-fold cross-validation that has been configured. A resampling strategy is used in the latter method to evaluate the effectiveness of machine learning models with a small sample of data. Data samples are divided into groups based on a single parameter (k). In light of this, k-fold cross-validation is usually referred to as this technique.

Machine learning models are usually tested using cross-validation to determine whether they perform well on untrained data. The model will be evaluated using a small sample to predict data not used during training. In general, what happens is as follows:

1. Randomly shuffle the dataset.
2. Create k groups from the dataset.
3. For every distinct group:
 - a. Use the group as a test or a holdout data set.
 - b. As a training data set, use the remaining groupings.
 - c. The training data fit a model evaluated using the test data.
 - d. In contrast to the model, the evaluation outcome is kept.
4. Finally, summarise the model's proficiency by employing a sample of model evaluation ratings.

The last setting, verbose, controls how much information will be printed.

```
GS = GridSearchCV(estimator=RandomForestRegressor(),  
                  param_grid=grid,  
                  scoring=['neg_mean_squared_error'],
```

```
'neg_mean_absolute_error', 'r2'],  
refit='r2', cv=5, verbose=4)
```

The following are some of the most crucial hyperparameters that were optimised for the Random Forest model:

- Trees in the forest (the number): By generating forests with a considerable number of trees (high number of estimators), a more robust aggregate model with less variation may be developed at the expense of a more significant training time. Most times, the answer here is to analyse the data: how much data is accessible and how many attributes each observation includes. Because of the unpredictability of Random Forest, specific characteristics with high predictive power might be left out of the forest and not be utilised entirely or used very little in the case of little trees or extensive data. The same applies to the data: if many observations are present and each tree is not trained with the complete dataset, if there is a limited number of trees, then some observations might be left out. When all other hyper-parameters are constant, increasing the number of trees decreases model error at the expense of a more extended training period. Because Random Forests seldom overfit, it may be necessary to use many trees to avoid these difficulties and achieve decent results. The central point is that as the number of trees increases, the model variance lowers, and the model error will be near the optimal value. Building a forest with 10,000 trees is a futile strategy.
- The standards by which to divide at each tree node: By calculating which feature and which value of that feature best separates the observations up to that point, decision trees create locally optimum judgments at each node. To do this, they employ a particular metric—MAE or MSE for regression. The general guideline is to use MSE if the data does not contain many outliers because it severely penalises observations that deviate from the mean.
- The Deepest Individual Trees Can Get: The number of possible feature/value pairings rises as the depth of individual trees increases. The more splits a tree has, the more information it contains about the data it uses. Overfitting results when this occurs in a single tree. However, overfitting is more challenging because of how the ensemble is

built in Random Forest, even if it is still possible at substantial depth values. Setting this variable to a reasonable value and making minor adjustments are recommended. Given the number of elements in the tree, this variable should be adjusted to a reasonable value. Neither creates stumps (very shallow trees) nor stupidly enormous trees.

- How many random characteristics are taken into account at each split: This one is one of the most crucial hyperparameters to adjust in the Random Forest ensemble. The most straightforward technique to determine the optimal value of this hyperparameter is by conducting a Grid Search with Cross Validation while taking into consideration the following:
 - A small number will result in fewer characteristics being taken into account when splitting at each node, reducing ensemble variance at the expense of an enormous individual tree (and likely aggregate) bias.
 - This number should be chosen considering noisy characteristics with numerous outliers and how many informative or quality features are present. For example, the value of the number of random characteristics on each split might be pretty minimal if the data collection includes elegant, polished, and high-quality features since all of the aspects that are taken into consideration will be attractive. On the other hand, this value should likely be more significant if the data is noisy, as doing so will enhance the likelihood that the contest will accept a good feature.
 - As there is a more significant likelihood that favourable features will be included, increasing the maximum number of random features examined in a split helps to lower the model's bias; however, this might come at the expense of increased volatility. Furthermore, the training speed also decreases when additional characteristics to test at each node are added.

Cross-validating the potential alternatives and keeping the model that produces the best results while accounting for the earlier factors is the most realistic course of action in this case. The best result obtained after looping over the Logic Table's length, after doing 80 trials, is:

The best hyperparameter values are: {'criterion': 'squared_error', 'max_depth': 18, 'max_features': 'sqrt', 'n_estimators': 200}

Best score according to the metric passed in *refit*: $R^2 = -0.2471368089372133$

The grid search technique was initially run while tweaking only a few hyperparameters and considering the best R^2 score measure. The latter, though, could be deceptive.

R^2 is rarely the appropriate statistic to assess how well it can be predicted a new output, y , from new input, x , for the following two relatively straightforward reasons:

1. The R^2 value is the same if the historical data are reversed, with y becoming x and x becoming y . The input and output of a model must be taken into account when measuring the model's predictive power.
2. Before determining the slope and intercept of the model, it is feasible to predict the R^2 value.

The second statement suggests that R^2 can be determined without fitting a regression model. Therefore, using it to evaluate prediction skills is illogical. The formula is independent of any models and relies on the raw data. Due to the manner that things cancel out or simplify, this is not some kind of mathematical gimmick. It is a simple truth that R^2 measures the degree of connection among two-time series, x and y , as intended. R^2 should only be used to measure the correlation between two sequences.

The new dictionary has been set as follows:

```
grid = {'max_depth': [3, 9, 12, 18, 35, None],
        'max_features': ['auto', 'sqrt', 'log2'],
        'min_samples_leaf': [1, 2, 4, 6],
        'min_samples_split': [2, 5, 10, 15],
        'n_estimators': [100, 200, 300],
        'criterion': ['squared_error', 'absolute_error', 'poisson']}
}
```

The only parameter altered within the *GridSearch* function is *refit*. The latter has been equal to a negative mean squared error. In order to have a more precise tuning of the hyperparameters, the study of the Random Forest variable significance has been carried out. The latter is described correctly in the following section. After that, specific inputs were deleted from the model, and the GridSearch technique was used again.

After removing useless parameters and setting refit to equal negative mean squared error, the final result is:

The best hyperparameter values are: {'criterion': 'poisson', 'max_depth': 35, 'max_features': 'sqrt', 'min_samples_leaf': 2, 'min_samples_split': 5, 'n_estimators': 200}

Best score according to the metric passed in refit:

Negative mean squared error = -0303991474696633

4.11 Variable importance

The relative importance of each characteristic to the prediction may be evaluated quite easily using the random forest method. *Sklearn* provides an excellent method for this. First, it measures the importance of a characteristic by assessing how many tree nodes that use that feature minimise contamination throughout the forest's trees. Then, after training, it automatically determines this score for every feature and modifies the results so that the overall significance equals 1.

By examining the feature significance, one may choose which characteristics to potentially remove from the model since they do not contribute enough—or occasionally at all—to the prediction process. The latter is significant since a common rule in machine learning states that the presence of additional features increases the likelihood that the model will overfit and vice versa. Feature significance refers to techniques that rank input traits in terms of how effectively they can forecast a specific outcome variable.

Possible reasons for a negative performance result (i.e. high MSE):

- Variables must be avoided or added (which is the most important).
- Data wrangling and exploratory data analysis (EDA) were lacking. In the latter phase, patterns and anomalies are discovered utilising data, testing hypotheses, and double-checking assumptions using statistical results and graphical representations.
- There is an inadequacy in the textual or categorical accounting of variables.

The ratings are helpful and may be used for a variety of predictive modelling issues, including:

- A better comprehension of the data.
- A better comprehension of a model.
- Fewer input features are available.

The three methods listed below can be utilised using feature significance scores:

1. The relative scores may reveal which aspects are most likely to be relevant to the target and which outcomes are likely least, providing information about the dataset. A subject matter expert may interpret this and use it as the foundation for collecting more or alternative data.
2. The significance ratings are determined by a prediction model fitted to the dataset, which might shed light on the model. For example, finding out which features are the most and least essential to the model when it makes a prediction may be made by looking at the importance score for that particular model. This kind of model interpretation may be applied to those models that permit it.
3. It may be used to enhance a prediction model by determining which elements to maintain (highest scores) or remove (lowest scores) based on the significance ratings (highest scores). In some cases, this feature extraction can improve the system's effectiveness by simplifying the modelled problem and speeding up the modelling process.

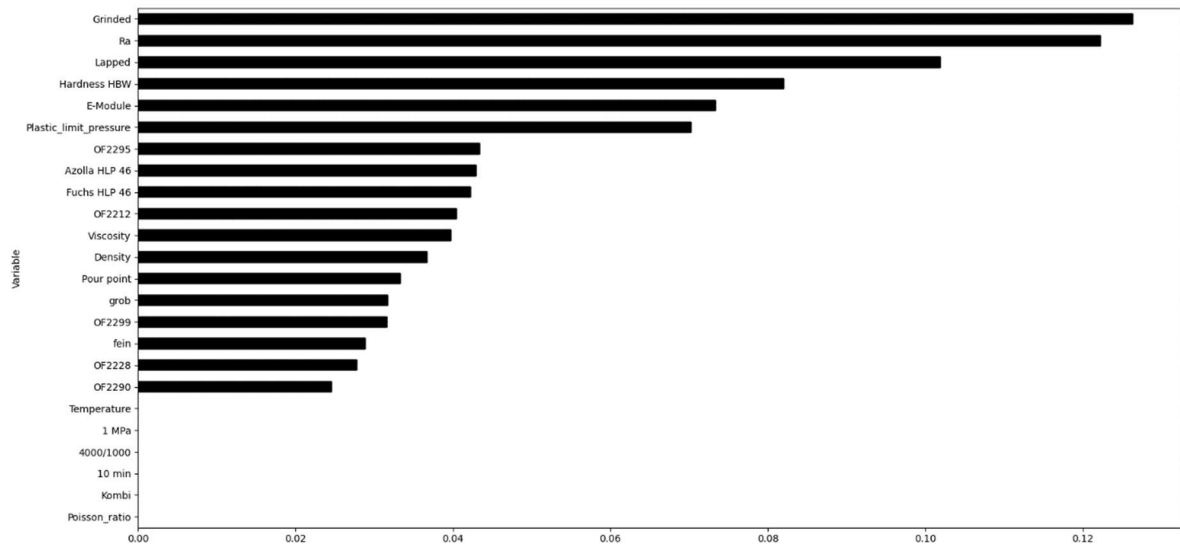


Figure 4.37: Feature importance analysis result.

After having analysed the above histogram, the following input has been removed:

- Temperature
- Pressure
- Speed
- Duration
- Test type
- Poisson ratio

Furthermore, the weight loss has been removed from the list of features since it is not necessary to construct the interpolated Stribeck curve.

5 Results and conclusions

5.1 Predictions

After choosing the model and training it, the goal is to start predicting new values. Three new materials and three new lubricants have been chosen. The features of the brass materials used initially (i.e. the one employed in the tribometer measurements) are reported below:

Brass material code	E module [GPa]	Poisson ratio	Yield strength [MPa]	Hardness HBW
<i>OF2212</i>	90	0.3	300	160
<i>OF2228</i>	95	0.3	340	170
<i>OF2290</i>	105	0.3	570	225
<i>OF2295</i>	110	0.3	390	175
<i>OF2299</i>	110	0.3	310	135

Table 5.1: Brass materials specifications.

The lubricants with which the test bench measurements are performed are:

Lubricant name	Density [kg/m ³]	Viscosity index	Pour point [°C]
<i>Azolla HLP 46</i>	880	100	-27
<i>Fuchs HLP 46</i>	879	150	-45

Table 5.2: Lubricants specifications.

As mentioned in the previous sections, the new materials and lubricants have been selected based on their composition. Concerning the brass materials, they must be lead-free materials. Instead, the lubricants must not contain zinc; therefore, they both are environmentally friendly components.

The three new lead-free brass materials chosen for the new predictions are:

Brass material	E module [GPa]	Poisson ratio	Yield strength [MPa]	Hardness HBW [MPa]
<i>C27450</i>	105	0.3	325	110
<i>ECO Brass</i>	98	0.3	510	164
<i>EnviB ECO C69300</i>	100	0.3	480	180

Table 5.3: New selected brass materials specifications.

The three new environmental friendly lubricants, instead, are:

Lubricant	Density [kg/m ³]	Viscosity index	Pour point [°C]
<i>Tellus S3 M 32</i>	855	105	-33
<i>Powerflow NZ HE Hydraulic Oil</i>	865	168	-51
<i>Q8 Holst 150</i>	888	97	-18

Table 5.4: New selected lubricants specifications

The first set of predictions employs the materials mentioned above and lubricants, comparing them using different combinations of pairings. The following table is reported to explain it better:

Type of predictions	New brass material	New brass material	New brass material	New lubricant	New lubricant	New lubricant
Commercial name	C27450	ECO Brass	EnviB ECO C69300	Tellus S3 M 32	Powerflow NZ HE Hydraulic Oil	Q8 Holst 150
Predictions code	401	404	407	501	504	507
Lubricant or Material	Fuchs HLP 46	Azolla HLP 46	Fuchs HLP 46	OF2228	OF2299	OF2212
Surface finish condition	Grinded Rough	Grinded Fine	Lapped Fine	Grinded Fine	Lapped Fine	Grinded Rough
Predictions code	402	405	408	502	505	508
Lubricant or Material	Azolla HLP 46	Fuchs HLP 46	Azolla HLP 46	OF2228	OF2299	OF2212
Surface finish condition	Lapped Fine	Lapped Rough	Lapped Rough	Lapped Rough	Grinded Fine	Lapped Fine
Predictions code	403	406	409	503	506	509
Lubricant or Material	Azolla HLP 46	Fuchs HLP 46	Fuchs HLP 46	OF2228	OF2299	OF2212
Surface finish condition	Grinded Fine	Grinded Rough	Grinded Rough	Lapped Fine	Lapped Rough	Grinded Rough

Table 5.5: Type of predictions carried out

The purpose is to understand if changing the material or the lubricant in the experimental setup can affect the performance. The results are shown in section 5.2.

Following the same reasoning, other kinds of predictions are performed. The latter has the following goal:

- Studying the performance of new materials. The predictions are carried out with different materials using the same lubricant.
- Comparing different lubricants. Contrary to the previous measurements, the same material is employed, changing the lubricant.
- Analysing the cost-performance relation based on the surface finishing. It is performed using the same lubricant and material for all measurements, gradually increasing the surface roughness (Ra) value.

The output of this set of measurements is also presented in the next section.

5.2 Results

The first results refer to the predictions reported in the table above, with different materials and lubricants. This type of simulation is mainly related to the possibility of extracting the Stribeck curve's characteristics of new materials without using the test bench. The interpolation procedure is performed through the *interp* function of the *NumPy* package. The input functions are the x-coordinates to evaluate the interpolated values, the x and y coordinates of the data points. The function returns the y-values of the interpolated points. In order to have a reliable interpolation function as output, more values of x have to be used. The first points used to interpolate are the maximum coefficient of friction (at zero speed) and the minimum, with the corresponding speed, predicted through the model. The additional points are extracted starting from these two points and using the slope values. The resulting interpolated Stribeck are shown below; only some are displayed to highlight their structure.

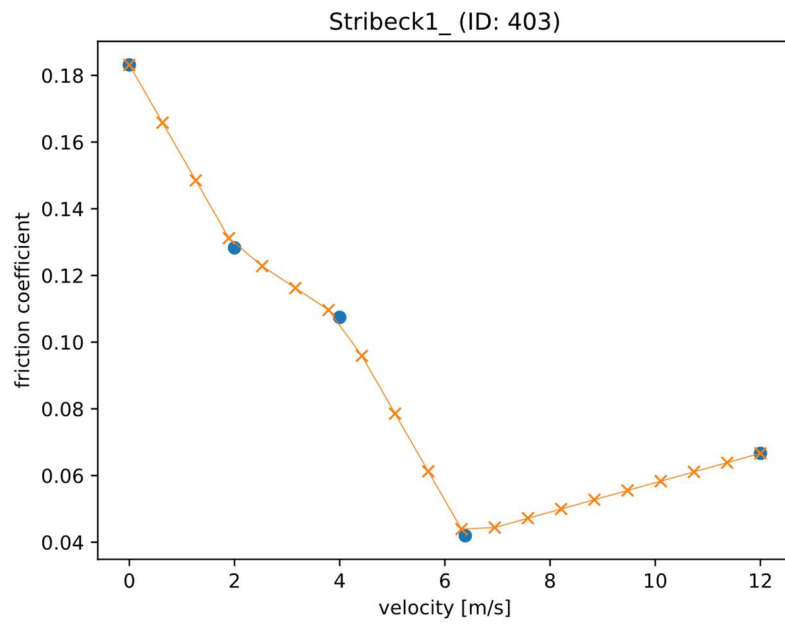


Figure 5.1: Interpolation of the predictions with Azolla as lubricant and C27450 as material.

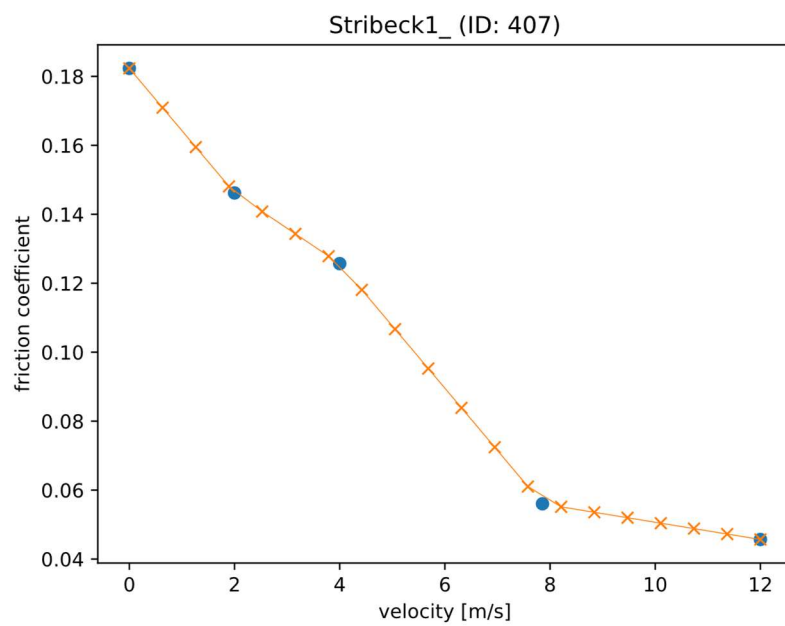


Figure 5.2: Interpolation of the predictions with Fuchs as lubricant and EnviB as material.

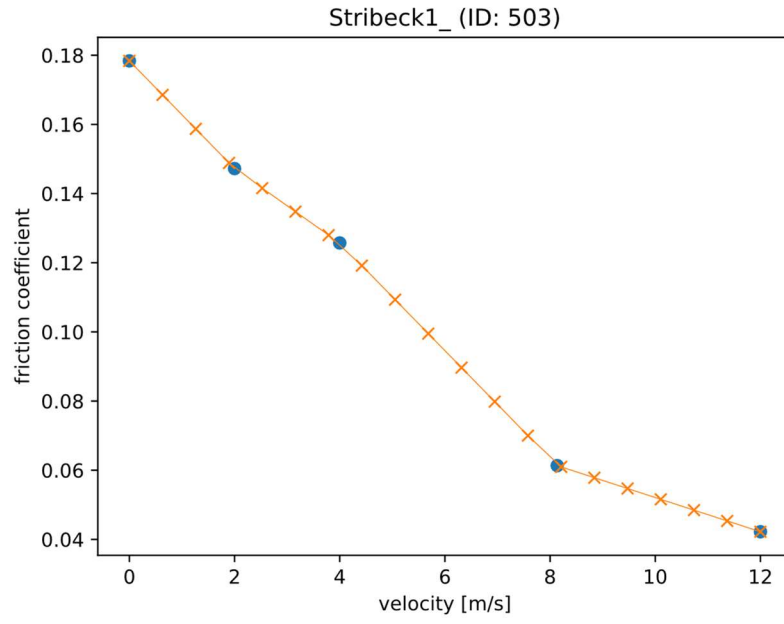


Figure 5.3: Interpolation of the predictions with Tellus as lubricant and OF2228 as material.

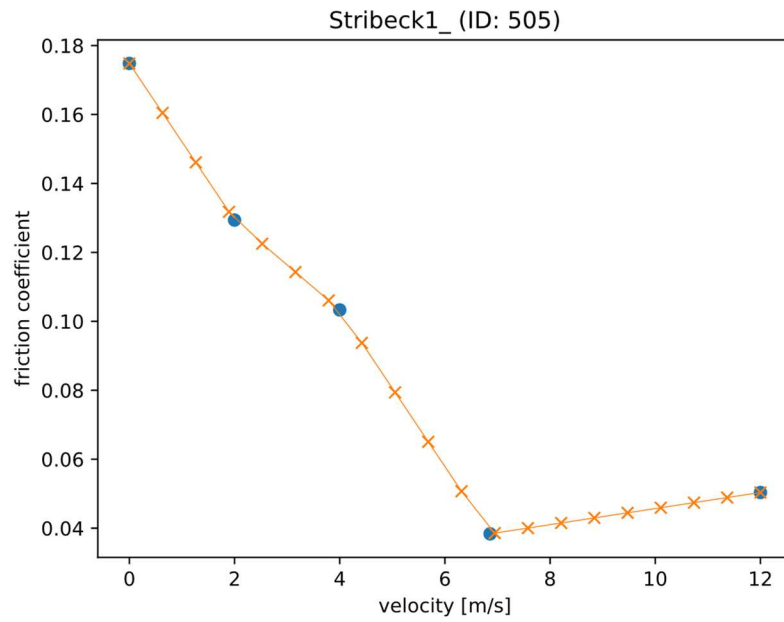


Figure 5.4: Interpolation of the predictions with Powerflow NZ as lubricant and OF2299 as lubricant.

As noted in section 4.6, related to the parameters extraction techniques, some of the Stribeck curves resulting from the experiments are monotonically decreasing. Similarly, some of the predicted Stribeck shapes are that way. The latter is because the model has been trained with those curves.

Nonetheless, these results do not give meaningful and transparent information for comparing different materials or lubricants and their parameters. For this purpose, the subsequent predictions have been performed.

➤ *Different surface finishing but same material and lubricant*

- **Effect of surface finishing**

Surface roughness' impact has previously been covered in section 2.2.1.2. It is important to reiterate that no engineering material is entirely smooth. Now, there is a significant likelihood of abrasion occurring in the presence of notches and weirs, which might result in resistance between the surfaces. Depending on the force exerted, the existence of deformation in both substrates will result in elastic or plastic deformation. Therefore, there will be friction as a result. The development and breaking of connections between the two areas cause friction. As a result of plasticity, very rough surfaces have more significant friction coefficients. Contrarily, as one might expect, exceptionally smooth surfaces have significant friction brought on by adhesion. An inter-anatomical viewpoint is used to explain the latter behaviour. There is a substantial molecular attraction for both the atoms of the two surfaces, which results in significant friction. The latter phenomenon is caused by the atoms' closeness on the two surfaces.

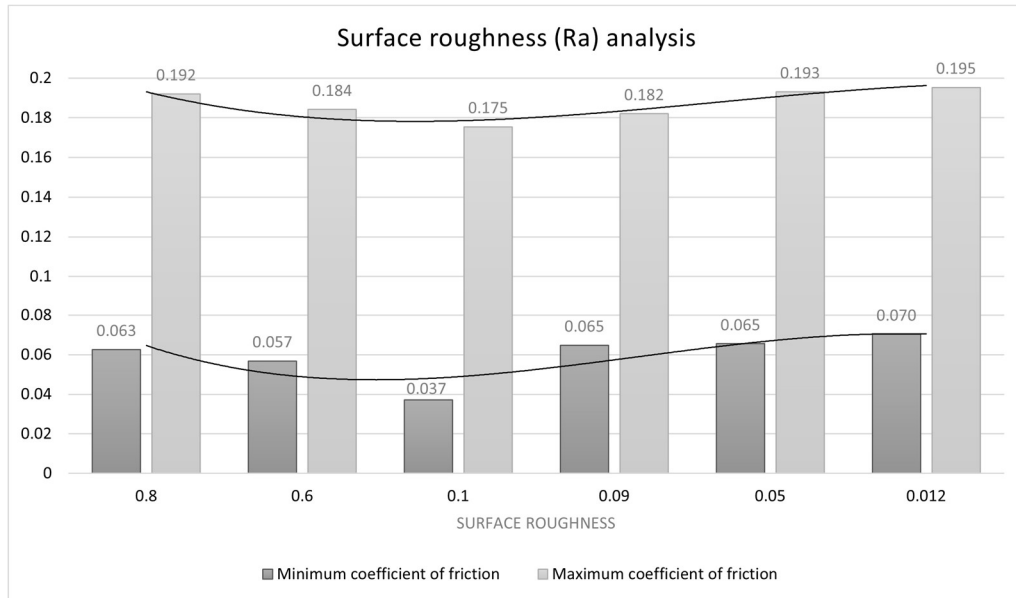


Figure 5.5: Histogram of comparison for different values of surface finishing.

The trend of the predictions shows a significant lowering of μ (in particular concerning μ_{\min}) with a slight decrease in surface roughness. On the other hand, the value μ_{\max}

experiences a decreasing behaviour, but this difference in value is not very enhanced. In particular, μ_{\min} decreases a lot until $R_a=0.1$, but after that value, it increases again. This comparison behaviour suggests that it is unnecessary to decrease the surface roughness, for example, by using a fine-lapped operation. On the contrary, lowering the R_a too much could lead to worse performance. Furthermore, as seen in Figure 5.2, the cost vastly increases with a demand for a more accurate surface roughness operation.

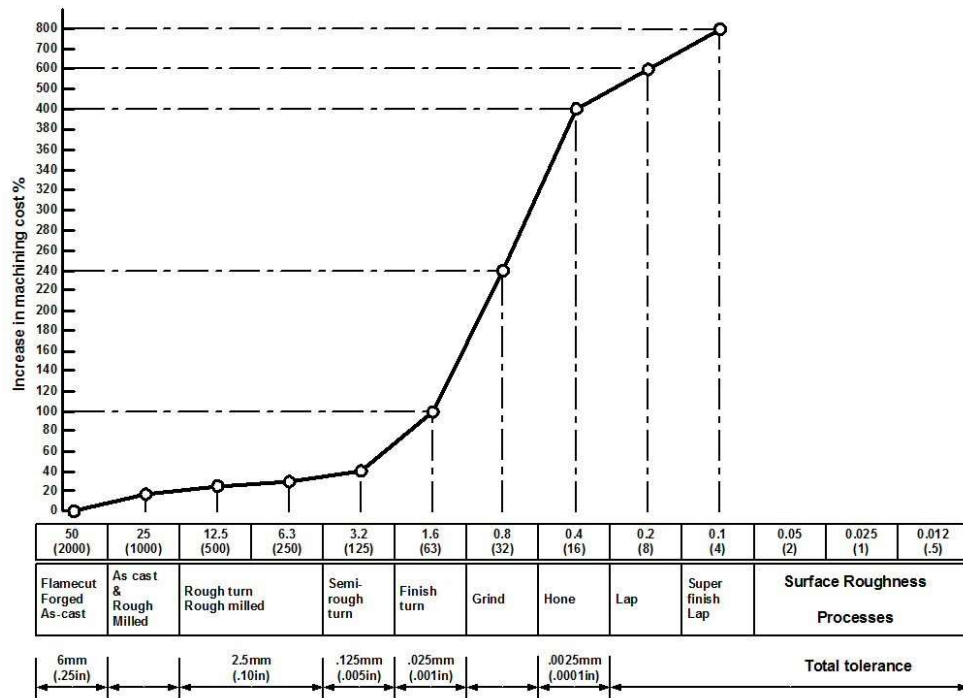


Figure 5.6: Comparative costs of different manufacturing processes [3].

➤ Different material

- Effect of elastic modulus

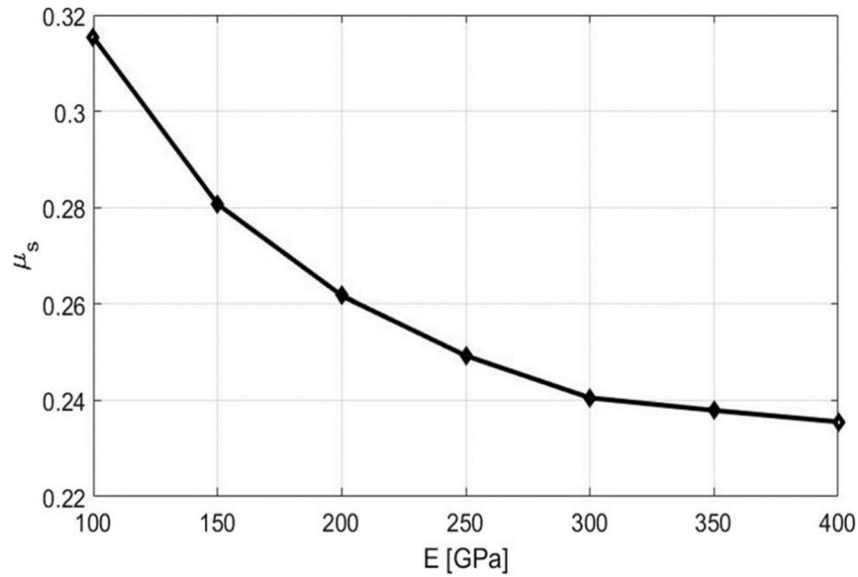


Figure 5.7: Effect of elastic modulus to static friction [2].

Figure 5.3 shows that the curve flattens out as Young's modulus rises. The coefficient of friction curves appears to converge as E increases. The latter is likely due to the increasing deformation and decreasing contact as the elastic modulus declines. The more extensive E contact area has a smaller value when the same contact stress is present, which results in a decreased static friction coefficient for that region.

In this work, this behaviour is not so evident since more or less all the Young modulus are of the same order of magnitude for all the materials. Below is reported the histogram of comparison with all the results.

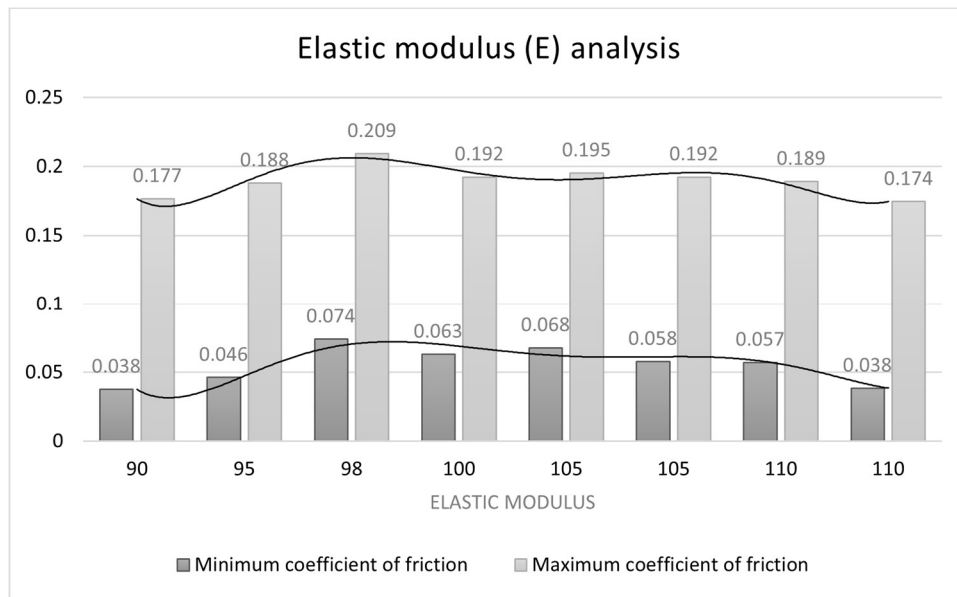


Figure 5.8: Histogram of comparison for different values of elastic modulus.

- **Effect of hardness**

The frictional coefficient experiences a lowering when an increase in hardness material is present. Friction is often lower in harder metals, and this reduction in resistance is directly connected with hardness. The soft metal contracts and expands with more significant deformation at contact sites for practically the same stress.

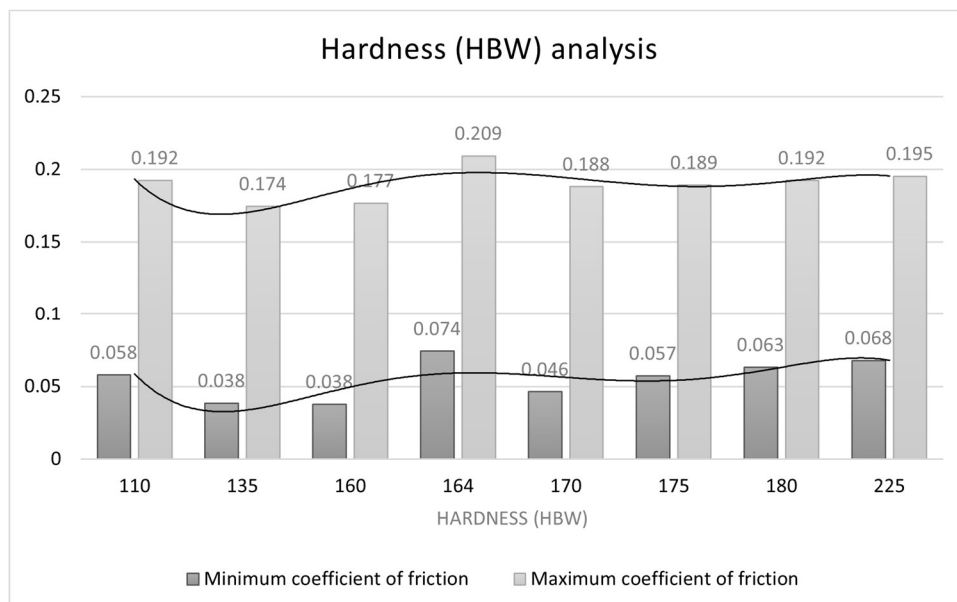


Figure 5.9: Histogram of comparison for different values of hardness.

Both for this parameter and the next one, the behaviour is mixed since they have opposite effects on the coefficient of friction. Besides that, when one of the two parameters also increases in the different materials, the other does the same. Therefore the two effects are deleting their respective contributions. Their values and the elastic modulus are reported in the table below.

	OF2299	OF2212	OF2290	OF2228	ECOB	OF2295	ENVIB	C27450
E [GPa]	110	90	105	95	98	110	100	105
S_y [MPa]	310	300	570	340	510	390	480	325
HBW [MPa]	135	160	225	170	164	175	180	110

Table 5.6: Young modulus, yield strength and hardness values for the different brass materials.

- Effect of yield strength

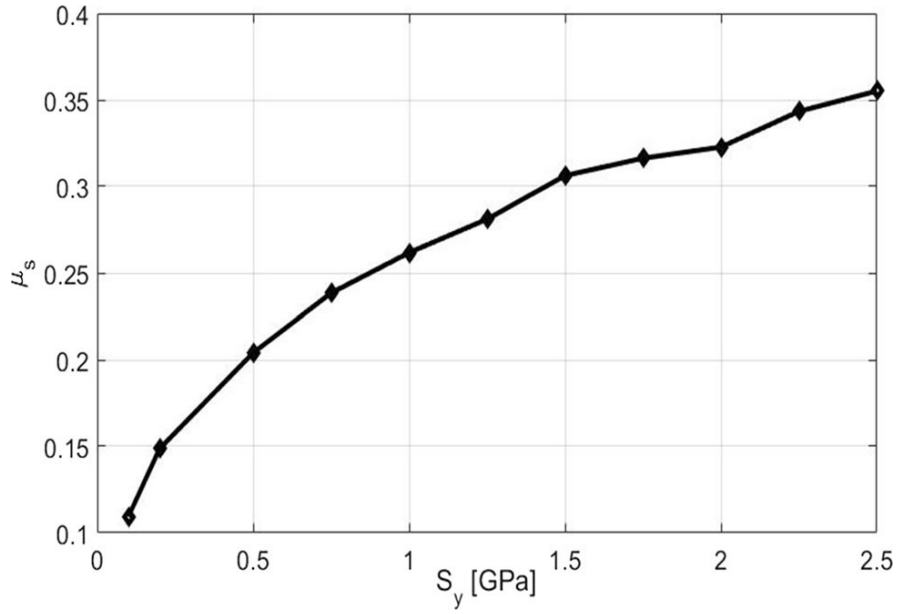


Figure 5.10: Static friction as a function of yield strength

The amplitude of the diagram likewise grows as the yield strength does. The latter is so that the asperity can withstand higher tangential stress before sliding because of the rise in yield strength. The static friction and yield strength are then plotted. The static friction coefficient rises as yield strength increases, as seen in Figure 5.6.

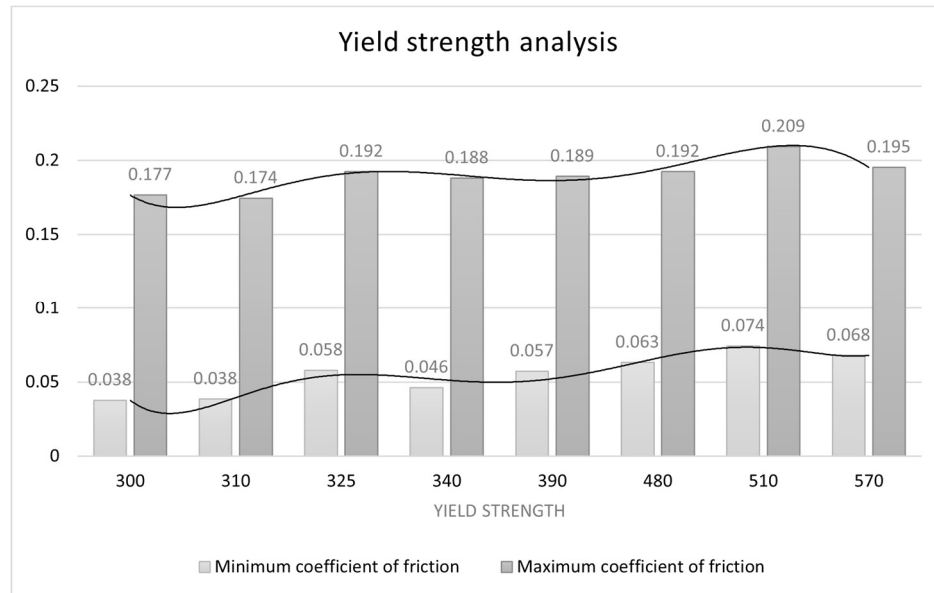


Figure 5.11: Histogram of comparison for different values of yield strength

The above histogram shows the most evident trend between the predictions performed, which denotes that this is the most significant contribution. In addition, the histogram shows the variation putting the value in increasing order of yield strength.

- **Effect of Poisson ratio**

For the sake of completeness, it is worth analysing the effect of Poisson's ratio on static friction. As Poisson's ratio rises, the static friction factor drops almost linearly. However, since the measurements in this work are only performed with brass materials, the Poisson ratio value is unchanged. The dimensionless radial force with the highest Poisson ratio has a more excellent value at each loading stage. The latter is because ν has a smaller x- and y-expansion, a minor surface separation, and a smaller contact area when ν is tiny. As a result, the frictional force increases with greater values of ν .

➤ *Different lubricant*

- **Effect of viscosity**

Because the presence of interstitial liquid creates a capillary force between the grains, wet granular media may withstand more significant standard stress than dry ones. The moist granular matter gains elastic properties as a result. In earlier studies, only low-viscosity liquids like water and low-viscosity synthetic oil were used. Viscosity was demonstrated to impact the dynamic mechanical characteristics of wet granular matter. When the lubricant's viscosity rises, the friction coefficient falls.

This phenomenon occurs because surface tension strengthens capillary bridges, which strengthens elastic modules, helping the connectivity to be more flexible as opposed to viscosity, strengthening the network and making it stiffer and more viscous. In other words, increased surface tension lowers the contact resistance by making the network more elastic. At the same time, higher viscosity raises the frictional force by causing a more rigid network to deform and lose dissipative energy. Therefore, the movement of the slider over the moist grainy matter results in energy losses.

The behaviour described above is visible in the histogram below, where the coefficient of friction has a monotonically decreasing trend when augmenting the lubricant's viscosity.

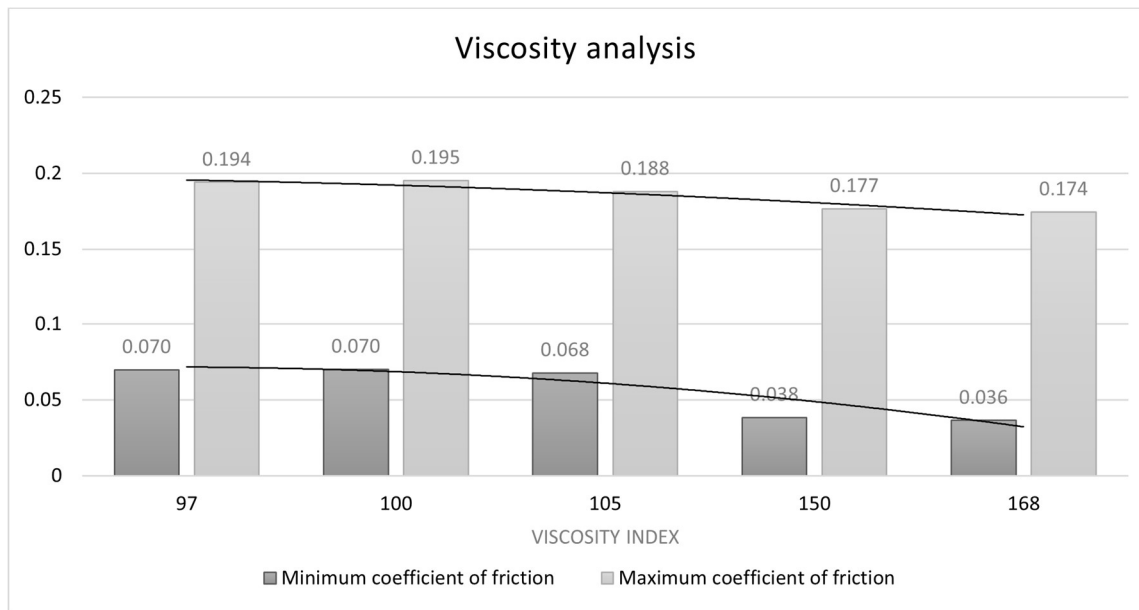


Figure 5.12: Histogram of comparison for different values of viscosity

6 Conclusions and outlook

The study will come to a close in this chapter, which will summarise the significant results regarding the objectives and research questions and their importance and significance. Additionally, it will discuss the study's shortcomings and suggest areas for additional investigation.

6.1 Conclusions

This study aimed to build a Machine Learning model able to predict the tribological behaviour of different materials that can be used in control plate pairings. The findings support this and demonstrate that several of the characteristics have a significant impact on performance.

The output of this research is a model able to overcome the cost limitations of new experiments with the tribometer. The latter allows the extraction of a Stribeck curve from the experimental setup chosen without the need to prepare specimens for the test bench. Furthermore, using a Random Forest regression model allows the user to change material, lubricants, and other conditions to see which parameters return the best performance. This method is a way to have huge savings of money and time. Even if the model does not give the complete curve as output, the interpolated curve is very informative from an engineering perspective. The points (such as μ_{\min} , μ_{\max}) through which it is constructed are the ones that give information about the energy consumed and the life of the components.

Introducing this work, the attention has been focused on the need to find and test many new materials. Environmental and performance requirements force the engineering field in a continuous run to find solutions. Therefore, it is essential to have very accurate models allowing to test lubricants and materials without investing time and money in new experiments.

This work follows the increasing trend which sees a vast introduction of Artificial Intelligence techniques inside the field of tribology.

This research illustrates that Machine Learning and AI can save time analysing new materials and finding new solutions. However, it also raises the need for more collaboration between different research areas to find better and more accurate results.

Based on these conclusions, companies and institutes should consider the possibility of investing money and time in undertaking this direction of working. The power of ML and data analysis, in general, is evident, and putting more effort into building a more extensive company database is of paramount importance.

Future studies could address the origins and causes of the different parameters to understand the implications of these results. A model built in this work can help pursue this goal. The next session explores and suggests two possible new research solutions related to this work.

6.2 Outlook

This work shows significant and good results in predicting tribological outcomes, but the performance can be improved. The first aspect of tackling is the increase of the dimension of the dataset used to train the model. The latter, in general, is a problem related to machine learning and artificial intelligence in many sectors and application areas.

The latest significant data analysis improvement did not ignore tribology, and many tribological problems have already been solved using AI approaches. Different configurations and sensing technologies and tribology's intrinsic multi-scale and statistical nature still pose a significant challenge in addressing ambiguities in the laboratory and operational data sets. Therefore, a more fundamental study is necessary before applying new AI techniques to assure their applicability and dependability in resolving tribological problems. The importance of particular technical expertise cannot be overstated. Because tribology is a multidisciplinary field, the collaboration between physicists, biochemists, materials engineers, mechatronics, and computer scientists is more challenging. Therefore, tribologists should be encouraged to explore new methodologies and develop cross-disciplinary partnerships. A better understanding of tribology can be gained using AI and machine learning algorithms, leading us to a new, more sustainable, and energy-efficient era.

The use of AI in tribology will undoubtedly rise in the next few years because of rapid algorithmic and computational capacity advancements and data's rising accessibility and recyclability. On the other hand, one of the most considerable challenges is the inaccuracy of results caused by experimental settings and sensor systems variations. Such challenges may

be readily overcome with sufficient study for accurate and appropriate data utilisation and cooperation with specialists from other sectors.

All the test bench experiments should be saved for a more extensive set and a more accurate model. However, this requires much time and work; therefore, it is out of the scope of this work. The possible solution to this work, to make it work better, is to create an “institute database” of tribometer measurements. The database mentioned above can include experiments with many different experimental setups. For example, parameters, such as pressure or load, significantly modify the shape of a Stribeck curve. However, in this work, such parameters could not be compared since the measurements available were only related to the same values.

Another way to be followed is to construct a bigger model that allows the extraction of the complete Stribeck curve. That is to say, having as output a considerable number of points that compose the tribological characteristic curve. The latter requires implementing a Neural Network model, particularly a Long short-term memory one. The latter needs to have some decoder inside it since the model has a much larger number of outputs concerning the number of inputs. A model of this type could also be useless from an engineering perspective. However, it is helpful if the user also needs information related to the number of peaks and, therefore, to the fretting wear phenomenon.

Numerous applications in lubrication and tribology have previously made use of artificial intelligence and machine learning techniques. However, the lubrication and tribological industry will need to create strategies for sharing the enormous quantities of information that such models require if it is to advance. Currently, most tribology tests produce only modest amounts of data, making it challenging to analyse the primary material, surface finish, and lubricant characteristics.

Soon, AI and machine learning can provide early warnings of potential issues and faults with machines so that proactive maintenance can be undertaken and customers will avoid potentially costly breakdowns. However, as stated before, the most critical aspect of achieving will be the trustworthy cooperation between researchers and specialists from many areas to increase the effectiveness of ML & AI in tribology.

Bibliography

- [1] Blau, P. J. (2019). Friction Science and Technology: From Concepts to Applications, Second Edition (2nd ed.). CRC Press.
- [2] Stachowiak, G., & Batchelor, A. W. (2016). Engineering Tribology (4th ed.). Butterworth-Heinemann.
- [3] Bhushan, B. (2013a). Introduction to Tribology (2nd ed.). Wiley.
- [4] Paul Michael, Nancy McGuire (2020, February). Efficient Hydraulic Systems Deliver The Power. Power & Motion. <https://www.powermotiontech.com/home/article/21122279/efficient-hydraulic-systems-deliver-the-power>.
- [5] Samuel Kärnell (2020). Division of Fluid and Mechatronic Systems, Linköping University. Fluid Power Pumps and the Electrification.
- [6] Toni Massoud, Rafael Pereira De Matos, Thierry Le Mogne, Michel Belin, Manuel Cobian, Benoît Thiébaud, Sophie Loehlé, Franck Dahlem and Clotilde Minfray (2019). Effect of ZDDP on lubrication mechanisms of linear fatty amines under boundary lubrication conditions. Université de Lyon, Ecole Centrale de Lyon, Laboratoire de Tribologie et Dynamique des Systèmes.
- [7] K. Oishi (2007). Development of Lead Free Copper Alloy "ECO BRASS®". Proceedings of the sixth International Copper-Cobre Conference, Toronto, Ontario, Canada, vol.1, 325-340.
- [8] Copper Development Association Inc (2022). BRASSES Copper-Zinc Alloys (Yellow Brasses). C27450 datasheet.
- [9] Lebronze Alloys (2017). EnviB ECO -CuZn21SiP - CW724R / C69300 datasheet.
- [10] Q8 Oils (2020). Q8 Holst 150: Advanced zinc-free hydraulic oil datasheet.
- [11] Phillips 66 Company (2021). Powerflow™ NZ HE Hydraulic Oil.
- [12] Shell (2019). Tellus S3 M 32: Premium Zinc-Free Industrial Hydraulic Fluid.
- [13] Kato, K., & Adachi, K. (2000). Wear mechanisms. In Modern Tribology Handbook: Volume One: Principles of Tribology (pp. 273-300). CRC Press.
- [14] Giovanni Straffelini (2001). A simplified approach to the adhesive theory of friction. Wear 249 (pp. 79-85). Elsevier.

- [15] Hamrock, B. J., Schmid, S. R., & Jacobson, B. O. (2004b). Fundamentals of Fluid Film Lubrication (Dekker Mechanical Engineering) (2nd ed.). CRC Press.
- [16] Brueninghaus Hydromatik (1983). Mannesmann Rexroth. A2V (Series 5).
- [17] Parker Hannifin Corporation (2012). Hydraulic Pump and Power Systems Division. P2060.
- [18] A. J. W. Moore and W. J. McG. Tegart (1952). Relation between friction and hardness. Proc. Roy. Soc. A, volume 212.
- [19] Chunyan Yang (2008) Role of Surface Roughness in Tribology: From Atomic to Macroscopic Scale. Quantum-Theorie der Materialien, Institut für Festkörperforschung (IFF), Forschungszentrum Jülich GmbH, Jülich, Deutschland.
- [20] Blau, P. J. (1991). Running-in: Art or engineering? Journal of Materials Engineering, 13(1), 47–53. <https://doi.org/10.1007/bf02834123>.
- [21] Otto, N.: „Experimentelle Untersuchung nachhaltiger Hydraulikfluide auf Ester- und Wasserbasis“, Ph.D. thesis RWTH Aachen, 2018, Aachen.
- [22] Steve Blank (2022, May). Artificial Intelligence and Machine Learning– Explained. <https://steveblank.com/2022/05/17/artificial-intelligence-and-machine-learning-explained/>.
- [23] Marian, M.; Tremmel, S (2021, September). Current Trends and Applications of Machine Learning in Tribology—A Review. <https://doi.org/10.3390/lubricants9090086>.
- [24] Ian Farrell (2021, December). Machine Learning & Artificial Intelligence in Tribology. <https://www.machinerylubrication.com/Read/32096/machine-learning-artificial-intelligence-in-tribology>.
- [25] Prölß, Maximilian & Schwarze, Hubert & Hagemann, Thomas & Zemella, Philipp & Winking, Philipp. (2018). Theoretical and Experimental Investigations on Transient Run-Up Procedures of Journal Bearings Including Mixed Friction Conditions.
- [26] Tribonet (2018, October). Fretting, fretting corrosion and fretting mechanisms. <https://www.tribonet.org/wiki/fretting-corrosion/>.
- [27] Spikes, Hugh & Zhang, Jie. (2014). History, Origins and Prediction of Elastohydrodynamic Friction. Tribology Letters. 56. 1-25. 10.1007/s11249-014-0396-y.
- [28] Andreas Velling (2020, June). Material Hardness – Types, Testing Methods & Units. <https://fractory.com/material-hardness/>.
- [29] John Anderson (2018). Hands On Machine Learning with Python.