THESE de DOCTORAT DE L'UNIVERSITE DE LYON
opérée au sein de
INSA de Lyon

Ecole Doctorale N° 162
Mécanique, Énergétique, Génie Civil, Acoustique (MEGA)

Spécialité/ discipline de doctorat :
Mécanique

Soutenue publiquement le 27/03/2020, par :

Rabii JAZA

Prediction of the tribological behaviour of a contact with third body particles:

Relating the morphological descriptors of the third body particles with the rheological parameters of the contact

Devant le jury composé de :

RICHARD Caroline	Professeur des universités	Polytech Tours	Rapporteur
FABRE Agnès	Maître de conférence HDR	Arts et Métiers ParisTech	Rapporteur
DENAPE Jean	Professeur des universités	Université de Toulouse	Examinateur
DUCOTTET Christophe	Professeur des universités	Université Jean Monnet	Examinateur
MASSI Francesco	Professor	La Sapienza University	Examinateur
BERTHIER Yves	Directeur de recherche Emérites	INSA Lyon	Directeur de thèse
DESCARTES Sylvie	Ingénieur de recherche HDR	INSA Lyon	Encadrante
MOLLON Guilhem	Maître de conférence	INSA Lyon	Encadrant

Abstract

This thesis work is the first part of a long-term work where we try to answer the question whether it is possible to set a link between the morphological aspects of the third body particles and the rheological parameters of the contact where they were created. Both measurements, morphological and rheological, are almost impossible to obtain without opening the contact itself. Therefore, such a link could be a game changer especially in machine monitoring and failure prediction, which is the final goal of this project. This work consists in a proof of concept. In this effort, we evaluate the efficiency of supervised machine learning algorithms in linking back the third body particles with the test from which they originate. In addition, we assess the ability of the algorithms in predicting the rheological properties of the contact from the morphological descriptors of the wear debris it produced.

We choose to hold our own tribological tests using a classic pin-on-disk tribometer. We select the 819AW (35NCD16) steel for the first bodies in contact. To ensure the production of diverse third body particles, we conduct nine tests organized in three sets. One experimental condition (rotational speed of the disk, gaseous environment of the contact, covered distance by the pin during the contact) vary between the tests of a given set. During the tests, we record *in situ* signals such as the tangential force and recorded video to supervise the outlet of the contact. The wear debris are characterized using two different sets of data: Rheological and Morphological sets.

The rheological parameters in this project are not the measurements used commonly, as they cannot be gathered during the experiment itself without opening the contact as mentioned before. The rheological data refers to six chosen descriptors calculated from the *in situ* signals that describe the flow of the third body particles. Regarding the morphological dataset, we chose five different descriptors to characterize the particles *post mortem* after the tribological tests were terminated. Those descriptors are calculated through image analysis algorithms of SEM images.

The machine learning algorithms use first the morphological database to determine in which test each particle was created. We put in place two different algorithms: logistic regression and neural network. The algorithms have a general success rate of classification equal to almost 40 % in predicting the test using only the five morphological descriptors of the particles.

Investigating in depth this success rate showed the limits of this method. However, it also proves that the approach has an interesting potential if proper adjustments are amde to both the algorithms and the database.

The next task assigned to the machine learning algorithms is to predict the rheological measurements from the morphological database. For that purpose, we put in place two other algorithms: linear and polynomial regression based. We try each time to predict a single rheological value from the morphological database.This method is not as successful if we adopt the prediction error (the absolute difference between the real value of the rheological data and the predicted one) as a success indicator. However, the predicted values of all the rheological data evolved around the averages of the real value of each one. This 'failure' is mainly the result of the variability difference between the rheological data and the morphological descriptors within the same test. It allows detecting the current limitations in the methodology with a great accuracy and it paves the way to future more accurate rheology prediction by machine learning algorithms.

Résumé

Ce travail de thèse est une preuve de concept. C'est la première partie d'un travail beaucoup plus grand où nous essayons de savoir s'il est possible d'établir un lien entre les aspects morphologiques des particules du troisième corps et les paramètres rhéologiques du contact où elles ont été créés. Les mesures rhéologiques sont difficiles à obtenir *in situ* sans ouvrir le contact et séparer les deux premiers corps. Un tel lien pourrait aider énormément l'industrie, en particulier dans la surveillance des machines et la prévision des défaillances, ce qui est l'objectif à long terme du projet auquel appartient ce travail de thèse. Dans cet effort, nous évaluons l'efficacité des algorithmes d'apprentissage automatique supervisé pour relier les particules du troisième corps aux tests d'où elles proviennent. De plus, nous évaluons la capacité des algorithmes à prédire les propriétés rhéologiques du contact à partir des descripteurs morphologiques des débris d'usure qu'il a produit.

Nous avons choisi de mener nos propres tests tribologiques en utilisant un tribomètre classique pion-disque. Nous avons opté pour l'acier 819AW (35NCD16) pour les premiers corps en contact. Pour assurer la création de diverses particules du troisième corps avec des morphologies différentes, nous avons effectué neuf tests organisé en trois ensembles. Une condition expérimentale (Vitesse de rotation du disque, environnement gazeux du contact, distance parcourue par le pion lors du contact) variait entre les tests d'un ensemble. Lors des tests, nous avons enregistré des signaux *in situ* tels que la force tangentielle et enregistré des vidéos pour surveiller la sortie du contact. Les débris d'usure sont caractérisés à l'aide de deux base de données différents: une rhéologiques et une morphologiques.

Les paramètres rhéologiques de ce projet ne correspondent pas à la définition usuelle du terme, car des mesures rhéologiques directes ne peuvent recueillies au cours de l'expérience elle-même qu'avec ouvrir le contact comme mentionné précédemment. Les données rhéologiques se réfèrent donc ici à six descripteurs choisis et calculée à partir des signaux *in situ* qui décrivent les débits des particules du troisième corps. Pour les données morphologiques, nous avons choisi cinq descripteurs différents pour caractériser les particules *post mortem* après la fin des tests tribologiques en utilisant des images microscopiques issues du MEB. Ces descripteurs sont calculés grâce à des algorithmes d'analyse d'images.

Les algorithmes d'apprentissage automatique ont utilisé au premier lieu la base de données morphologiques pour déterminer dans quel test chaque particule a été créée. Nous avons mis

en place deux algorithmes différents: la régression logistique et le réseau de neurones. Les algorithmes ont obtenu un taux de réussite globale de classification égal à près de 40% pour prédire le test en utilisant uniquement les cinq descripteurs morphologiques des particules. Une étude approfondie de ce taux de réussite a montré les limites de cette méthode. Cependant, cela a également prouvé que l'approche choisie présentait un potentiel intéressant si des ajustements mineurs étaient appliqués à la fois sur les algorithmes et la base de données.

L'autre tâche assignée au algorithmes d'apprentissage automatique était de prédire les mesures rhéologiques à partir de la base de données morphologiques. Pour cela, nous avons mis en place deux autres algorithmes: basés sur la régression linéaire et polynomiale. Nous avons essayé à chaque fois à prédire une seule valeur rhéologique à partir de la base de donnée morphologique. Cette méthode n'a pas été aussi fructueuse que la précédente si nous adoptons l'erreur de prédiction (la différence absolue entre la valeur réelle des données rhéologiques et celle prédite) comme indicateur de succès. Cependant, les valeurs prédites de toutes les données rhéologiques ont évolué autour des moyennes de la valeur réelle de chacune. Cet 'échec' était principalement le résultat de la différence de variabilité entre les données rhéologiques et les descripteurs morphologiques au sein du même test. Il a permis de détecter les limites actuelles de notre approche et de déterminer avec précision les améliorations à apporter à la base de données afin d'atteindre des objectifs de prédiction suffisamment fiable de la rhéologie du troisième corps

Content table

I. State of the art

1. Tribology

Peter JOST (Jost, 1966) first introduced the appellation 'Tribology' in 1966 . It comes from the Greek root 'Tribo' that means 'I rub'. Tribology is the science of friction and wear of solids into contact. It regroups sciences and technologies that study the different interactions between contacting surfaces, their causes and effects. Oxford dictionary defines tribology as 'the science of interacting surfaces in relative motion'.

Despite the name being rather new, the recorded interest in this science goes back to ancient times. We can trace it back to the Palaeolithic period (3.3 million years ago) where the drills made for digging holes or producing fire were equipped with the early edition of bearings made mainly from bones (Davidson, 1957). The Egyptians, around 1880 BC, were also among the first to use the fundamentals of tribology also (Layard, 1853). As we can see in **Figure I-1** (Dowson, 1998), the man in the red circle is applying a lubricant extracted from animal fat to lubricate the way and facilitate the pulling of the statue weighing almost 720 tons by the slaves.
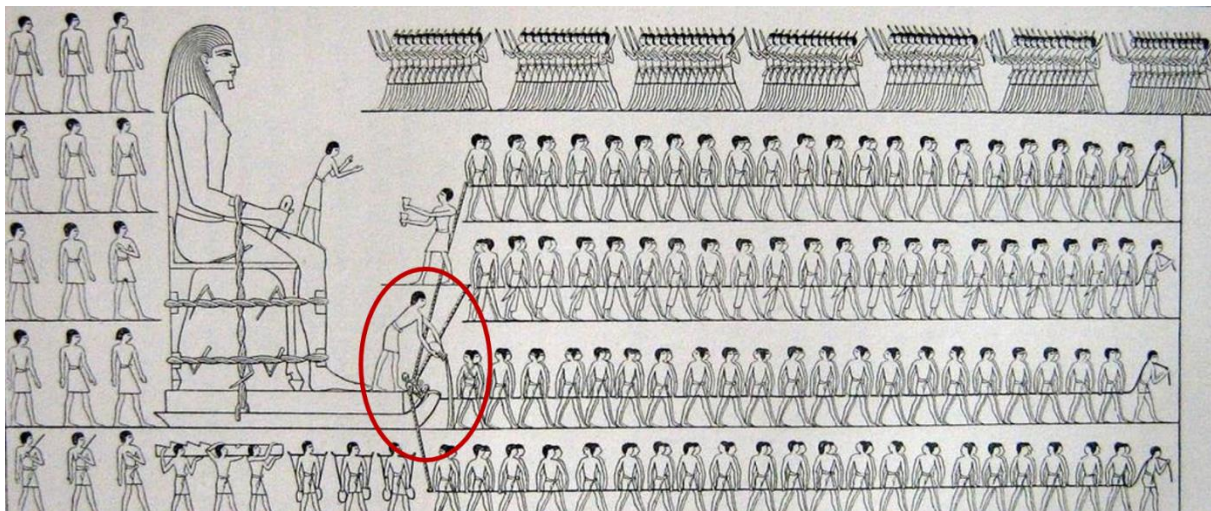

Figure I-1: Moving of an Egyptian colossus

The use of the wheel started from 3500 BC and therefore we can assume that people back then were interested in reducing the friction so the movement would be more fluid. The **Figure I-2** below shows the first attempts for the studded wheels at nearly 1338 AD.

Figure I-2: Drawing of two wheeled harvest cart with studded wheels.

The world needed to wait until the 15th century and for Leonardo da Vinci to give a scientific approach to friction and to attempt the first experimental study for tribology. Da Vinci was the first to conclude that the friction force is directly proportional to the load and that it is independent from both the area of contact and the sliding velocity using his famous sliding blocks over a flat surface. However, his work did not have any influence on the history of tribology because it was not published. His researches and work were published by DOWSON (Dowson, 1998) and presented in his study of the history of tribology more than 5 centuries later. In the 16th century came a scientific breakthrough with Robert HOOKE who put in place one of the first laws in friction and Leonard EULER who analytically defined the friction and gave the notation μ to the friction coefficient.

To highlight the importance of Tribology, a recent published study shows that almost 23% of the world's total energy consumption is the direct result of tribological contacts (Holmberg and Erdemir, 2017). Also according to JOST (Jost, 1966), the gross national product of a nation can rise by 1% if its industry adapts better tribological methods.

In the rest of this section, we will be explaining in more details the mechanical aspects of this project such as the surface and the third body approach used.

1.1. Surface

By definition, the surface of any volume is the set of points separating this volume from the environment surrounding it (Barrau, 2004). As a simpler definition, the surface is considered in most cases to be the exterior layer of a volume. However in reality, it is not that simple. In

metals, for instance, the surface is devided into 3 different layers, each chemically different from the other as seen in **Figure I-3** (Gatos and C., 1968; Buckley, 1978);

- Bulk layer which is the basic structure of the metal.
- Worked hardened layer which has a different crystaline structure from the bulk layer.
- Oxides (outer layer) created from the interactions of the metal with its environment and from the deposition of impurites.

In tribology, the surface we are often interested in studying is the near surface layer (a grouping of the oxides and the worked hardened layer) with a thickness between $10^{-6} \, mm$ and $10^{-2} \, mm$.
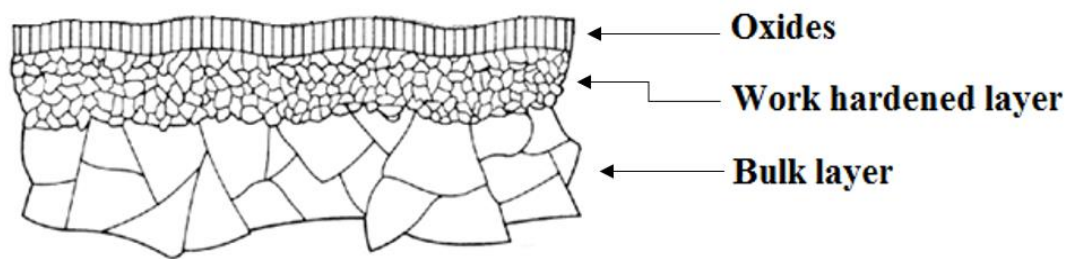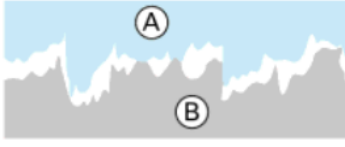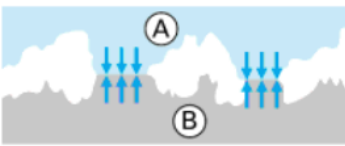
Figure I-3: Theoretical view of a theoretical surface representation.

The second big misconseption about the surface is its flatness. We generally consider a 'smooth' surface to be flat. However, when a solid is examined microscopically it is more likely to present several superficial irregularities . The surface may present two classes of geometrical errors:

- Macroscopic errors, generally related to the manufacturing machine and visible to the naked eye.
- Microscopic errors, related to operations such as grinding, cutting or blasting and usually only seen under the microscope.

The topography of the surface is characterized by its roughness. The surface roughness gives an idea also about the real area of the contact and the pressure's distribution during the friction period. In **Table I-1**, we present some cases of the two contacting surfaces with different states and we discuss the effect it has on the contact itself.

Table I-1: Two bodies in contact: causes and effects.

Phenomenon	Contact description	Effects
Imbrication		• Resistance to the movement • Deformation, rupture…
Plastic or elastic deformation		• Change in the topography of both surfaces • Hardening of the materials
Plastic deformation : Penetration (hardness of A > hardness of B)		• Resistance to the movement • Ploughing of B by A
Adhesion		• Resistance to the movement

When two solid bodies are in contact and relative motion under a certain normal force F, a tangential force T appears and opposes to the sense of the movement, as shown in **Figure I-4**. T is the friction force. On one hand, the friction is beneficial to the contact if our aim is to immobilize two solids with respect to each other. For example, without friction we would not be able to walk because it is the force keeping our feet from sliding. On the other hand, it can be a disadvantage if our aim is to keep the two bodies in contact moving. For instance, we would not want any undesirable effects on a motion ensured by a ball bearing from the contact of the balls with either the inner ring, the outer ring or the cage. In the example in **Figure I-4** (Brunetière, 2016), T is the friction force, F is the normal force that the body 2 applies on the body 1, V is the sliding velocity and L is the contact length.
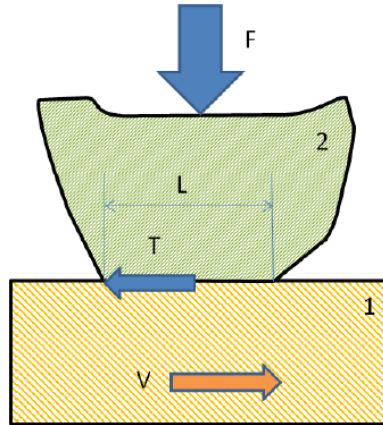
Figure I-4: Two bodies in contact.

A logical result of keeping two bodies in contact while each one is trying to move in the opposite direction is the wear of the contacting surfaces. In certain cases to limit this damage, a certain lubricant is introduced to the contact. For example, when using a ball bearing, it is common to use grease to limit the friction consequences.

Depending on the geometries of the bodies used, the contact can be:

- Point (sphere/sphere, sphere/plane, …)
- Linear (cylinder/plane)
- Surface (plane/plane)

When we put two objects in contact, their surface roughness causes that the contact takes place only at the asperities that the surface presents. The sum of those spots is the so-called real contact surface. In the case of sliding friction, the surface is constantly changing in time. It is then more reasonable to think in terms of applied force than pressure.

1.2.   Third body approach

Godet first introduced the third body approach in the 70s. This theory stresses that, apart from the first period in the contact's life, two-body contacts are unfound in the industry. In most cases third bodies are created immediately between the contacting surfaces transforming a two-body contact into a three-body one (Godet, 1984). It is a mechanical transposition of the lubrication concepts to dry friction (Denape, 2014). This concept was introduced to understand better the dynamics of the interfaces and to bring a better understanding of the mechanisms of friction and wear in general (Berthier, 1996).

1.2.1. Tribological triplet

The analysis of rubbing contacts is not complete without the understanding of the phenomena occurring at different scales, hence the suggestion of the tribological triplet, as represented in **Figure I-5**. Adopting this concept divides the tribological problem into three different parts; the mechanism, the first bodies and the third body (Vincent, Berthier and Maurice, 1992).
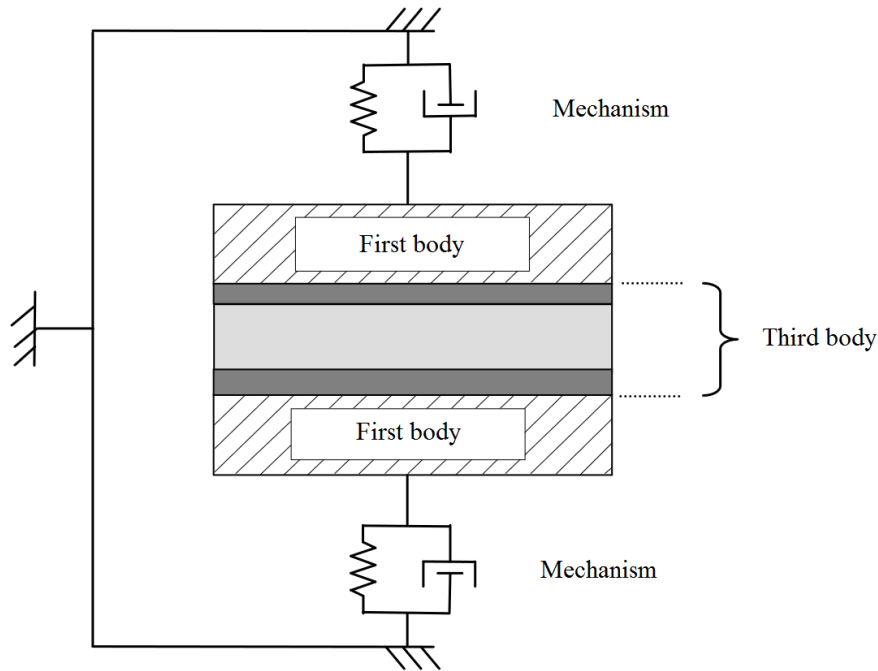
Figure I-5: Representation of the tribological triplet.

1.2.1.1. Mechanical system

The mechanical system is the structure holding the first bodies and the third body in place, transmitting the normal loads to the contact and setting the operating conditions. We can differentiate between two conditions associated to the mechanical system (Bill and C., 1982): the contact conditions (Normal load, vibration levels, …) and the environmental conditions (humidity, temperature, nature of the surrounding environment …).

1.2.1.2. First bodies

The first bodies are the two bodies that limit the contact. They intervene by their geometry, mechanical and chemical characteristics that defines how the deformation will occur that will create the elementary contact areas. The intervention of the materials on the contact is indirect because it happens via the geometry of the first bodies. However, their action is direct on the

elementary contact areas since their materials respond to the local tribological stress. The first bodies react to the applied mechanical stress with two different answers:

- Structure change that leads to the detachment of particles ranging in size from a nanometre to a few micrometres.
- Cracking.

Any kind of coating of the surfaces in considered as a part of the first bodies.

1.2.1.3. Third body

Holding for a period two first bodies in contact undergoing a movement of opposite translation with a normal force applied on one of them results into the progressive destruction of the their surfaces. By definition, the third body is the intercalary element between both surfaces. It is considered as a new body because the physico-chemical analysis proved that its properties are different from those of the first bodies.

In the case of a lubricated contact, the third body is the fluid (and all it contains) introduced between the surfaces. This third body is viewed as artificial. Nevertheless, in the case of dry contact, the observed third body turns out to be solid. It comes mainly from particles detached from the surfaces of the first bodies. It can have different forms such as separate particles with different sizes and shapes or a paste form depending on the mechanical loads applied, on the environment where they were created and on the nature of the first bodies. The rust and the particles coming from external pollution (dust, dirt …), present on the surface of the first bodies, are considered as parts of the third body.

The third body particles are generally related to mechanical aspects, such as separating the two first bodies (Godet, 1990), transmitting the load (Lofficial and Berthier, 1987) and accommodating the velocity (Berthier, Vincent and Godet, 1988). However, their role is not limited only to that. Third body particles guarentee the thermal continuity in the contact (Ling and Simkins, 1963) and the layer helps contain the temperature difference in its thickness (Kennedy, 1984).

Understanding the friction phenomena requires understanding the third body. The impact of this layer of particles in the contact has been proven at all scales. One of the most important discoveries in nanotribology was superlubricity in the contact of atomically arranged smooth surfaces (Dienwiebel *et al.*, 2004). However, the existence of third body particles trapped between the contacting surfaces cancel the superlubricity effect (Filippov *et al.*, 2008). The

same result can be due to impurities between the surfaces, even at the atomic level (Müser, Wenning and Robbins, 2001). Admitting that the contacting surfaces change is a key element in solving macroscopic tribological problems. The third body layer is the essential element to understand the tribological properties of the combustion engine for example (Scherge, Shakhvorostov and Pöhlmann, 2003).

1.2.2. Tribological circuit

To finish the tribological investigation, it is necessary to reconstitute and evaluate the flows of the third body going into, through and out of the contact. **Figure I-6** describes those flows present in the contact. The different third body flows are necessary to understand the relationship between the third body particles and the wear in the contact. We can find the following three different types of flows (Berthier, 1990):

- The source flow Q_s is what fuels the contact by the destruction of the surfaces of the two first bodies Q_s^{int} or by injecting artificial lubricant Q_s^{ext}.
- The external flow Q_e is the flow of the particles that escape the contact. We can distinguish between the reintroduced flow Q_r and the wear flow Q_w. The difference between those two flows is that the reintroduced flow does not leave the contact yet the wear flow is made of particle that will no longer be in the contact.
- The internal flow Q_i of the particles trapped between the surfaces in contact. .

Figure I-6: The tribological circuit.
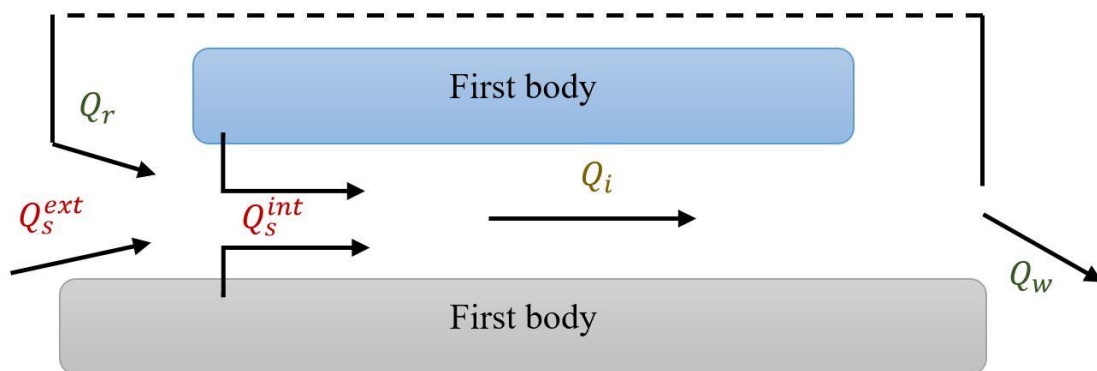
This approach is used with solid third body in analogy with the fluid third bodies. It helps understand more the changes in friction and correlate them to the rheology and the different flows of the third body in the contact. As an example, if we have a 'thick' enough layer of third body, then the wear flow Q_w and the internal third body source flow Q_s^{int} are directly correlated

because for every particle that gets ejected from the contact a new particle is detached from the first bodies to keep the balance (Fillot *et al.*, 2004).

1.2.3. Solid third body rheology

Previous studies proved the importance of introducing the notion of rheology to better analyse the solid third body in dry contacts (Dowson, University of Leeds. Institute of Tribology. and Institut national des sciences appliquées de Lyon., 1996; Souchon, Renaux and Berthier, 1998). The rheology is the science dealing with the deformation and flow of matter, according to Oxford dictionary.

As mentioned before, third body particles are found in the contact as separate particles, chunks of material or even in a paste form. This discontinuous state makes it impossible to consider the Hertzian contact stress because the load is supported by blocks of material.

The surface of the first bodies and the third body are very sensitive to the external environment like the temperature and the humidity. Therefore, the nature of the solid third body particles is generally a result of external factors that are difficult to control.

For those reasons and because of the lack of solid third body behaviour laws and experimental protocols, its rheology cannot be measured directly, but is instead evaluated. This assessment is based on the 'cohesion' and the 'ductility' of the third body (Descartes and Berthier, 2002). The term 'cohesion' is given by analogy with the cohesion of powder mix. When studying two third body samples, the one with the lower cohesion is the powdery sample containing different small particles. The more cohesive one tends to be a very compact sample. The term 'ductility' is chosen by analogy with the plastic flow characteristics of metals. A sample with low ductility means that the third body spreads little and covers a small area of the contact. While a sample with high ductility is a sample, where the third body spreads easily and covers a large part of the contact area.

The characterization process for those two properties is the result of coupling video recording of the test, measurements of the friction factor and post mortem analyses of the wear. This part will be further explained in the following sections.

1.3.   Wear mechanisms

Wear is the phenomenon of material removal from a surface due to its interaction with another one. This event leads to mass and form loss and changes in the mechanical and chemical properties of the surface. It is associated to the wear flow.

Velocity, the temperature and the time affect the evolution of wear. While some wear forms are progressive in time, other forms can evolve unexpectedly when certain critical values are reached. Usually, we try to minimize the wear due to its negative effects, but we tend to forget that it can also present positives aspects. As a simple example, if there was no wear, we would not be able to use a pencil to write. Generally, the wear of a mechanism is the result of a combination of events. It is rarely the result of one well-known cause.

At the beginning of the friction process, the screen of the surface is the part that undergoes the thermal, mechanical and chemical effects. Then those effects get through to the core of the surface. In general, a surface goes through three stages during its lifetime:

- Lapping: Rapid wear with a decreasing wear speed.
- Normal functioning: Low and constant wear.
- Aging: Rapid wear with an increasing wear speed.

There are generally four wear modes agreed on:

- *Adhesive wear*: When two surfaces are held in contact with each other, and because of the non-perfect surfaces state discussed earlier, they only touch in certain points. When a normal force is applied to the contact, friction and wear happen to those points. Adhesive wear refers to the undesired detachment and the transfer of wear particles from one surface in the contact to the other.
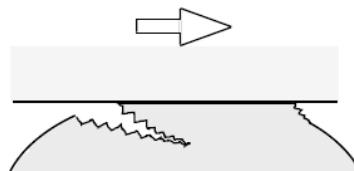

Figure I-7: Adhesive wear.

- *Abrasive wear*: When two surfaces with different roughness are put in contact and if the contact has interlocking of inclined and curved regions, ploughing can take place. That generates the loss of material from one surface, the one

with the lower roughness; this is what the ASTM (American Society for Testing and Materials) international defines as abrasive wear.
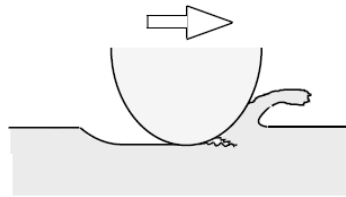
Figure I-8: Abrasive wear.

- *Fatigue wear*: This type of wear is caused after a certain number of repeated contacts. The progressive and localized damage takes place on one surface of the two in contact.
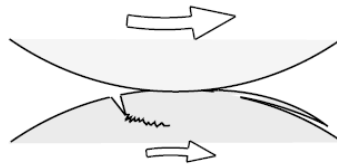
Figure I-9: Fatigue wear.

- *Corrosive wear*: When the contact is held in an environment containing corrosive liquids or gases, a certain number of reactions take place between the two surfaces. Corrosive wear is the material degradation from one surface or both in condition where both corrosion and wear mechanisms are present.
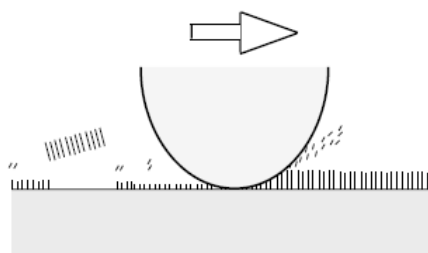
Figure I-10: Corrosive wear.

1.4. Conclusion

As we saw in this section, contacts between two solids include a large amount of complexity and practitioner solving friction problems may tend to oversimplify them by focusing on a particular scale instead of considering the tribological triplet entirely. In an effort to ease the

understanding of the contact, we focus in this thesis on a thorough examination of the third body. Learning from the morphology of the latter and linking it to the type of wear the surfaces are enduring will help shorten the time of the contact investigation and maintenance.

The linking part will be put in place with machine learning algorithms that will learn from past data to predict the future. This recent technology is presented in the next part of this report

2. Machine Learning

Machine learning is a big part of Artificial Intelligence, which is a new concept in technology but not necessarily to human kind. Actually, we can trace Artificial Intelligence back to Greek mythology with Talus (Stuart and Peter, 2005). Talus is the first 'Humanoid' allegedly created by Hephaistos, the Greek god of Technology, to protect Athens. It was described as a combination of living and non-living neurological interfaces. It could feel, think and above all learn just as humans do.

Human cognition depends on this remarkable ability, which we take for granted most of the time, to extract knowledge from few examples and generalize it. Definitely, animals can differentiate between objects, but only humans and maybe some other mammals are able to get the characteristics of an object and then generalize it to know that object under any circumstances. Let us study how a child is taught the meaning of words. For this example, we will use the word 'Horse'. Parents will provide a certain number of pictures of horses so the child will connect the word to those images. However, it will not stop there; the toddler will be able to get the characteristics of the horse and will be able to tell if a new picture is of a horse or not. Of course, there will be some mistakes like with pictures of Zebras, Ponies and Donkeys (Woodward and Needham, 2009). This ability to generalize the few information we get from a small number of examples is very critical, not just to understand the meaning of words, but also to conclude the properties of things, cause and effect relationships and many other domains from our daily lives.

With the enormous number of possibilities that learning offers to humans, it comes with no surprise that, since machines were invented, the question of whether they could one day have that ability was asked. '*The Perceptron*' was the first working machine learning algorithm discovered by Rosenblatt in 1958. *The Perceptron* learns a discriminant linear function from a sequence of examples described by their attributes (Rosenblatt, 1958). Certainly, the

discriminant analysis was discovered in 1930 but *The Perceptron* was the first algorithm able to induce a concept or a function from a sample examples. This was the turning point in machine learning history and the research in that field has since then made a significant progress.

It took humanity hundreds of centuries to create the first learning algorithm. However, in less than half a century, we went from *the Perceptron* to almost infinite possibilities. Machine learning algorithms cover almost all domains. Machine learning can recognize a human from its face (Rowley, Baluja and Kanade, 1998), fingerprints (Alonso-Fernandez *et al.*, 2009) and eye iris (Dobeš *et al.*, 2004). In the transport field, we now find autonomous vehicles (Bojarski *et al.*, 2016). Machine learning algorithms are capable now of finding tumours in MR images (Devos *et al.*, 2007), studying the biomedical signals and predicting the progression of a disease (Sejdić and Falk, 2018). Thanks to machine learning, machines are beating humans at chess (Feng-Hsiung Hsu, 1999), the strategy game Go (Silver *et al.*, 2017), and at solving Rubik cubes where an AI machine holds the world record.

There are two main reasons behind this recent exponential progress:

- **The technological breakthrough in hardware**: To get the best results of a machine-learning algorithm it is crucial to have the right hardware and infrastructure. The highest performing computers in the beginning of the Artificial intelligence era were the supercomputers. However, they were only accessible to a small group of people due to their high cost. For example, the IBM 7030 Stretch supercomputer cost nearly 8 million dollars when it was created in 1961. However, companies now are creating special units just for Machine Learning and Artificial Intelligence purposes for cheap prices due to the lower cost of transistors. To understand the magnitude of the evolution of processors, according the Shane Legg (cofounder of Google DeepMind), the phase of 'teaching' a machine learning algorithm that takes one day on the newest processors nowadays, would have taken a quarter of a million years on the 80486 intel processor created in the nineties.
- **Data availability**: As explained before, machine learning needs to learn from example samples. Researchers estimate that each day, we are producing two and a half quintillion (10^{18}) bytes of data (10^6 Terabytes of data) and all this information is stored in a way or another. In fact IBM reports (IBM, 2016) confirm that 90% of the data stored has been created in the last two years.

In the rest of this section, *the intelligent agents* will be described first, followed by a presentation of the different models of machine learning: *Supervised learning* and *Unsupervised learning*. The different applications of machine learning in image analysis will be the last part of this section.

2.1.   Intelligent agents

An agent is by definition anything that understands and identifies its environment using sensors and acts on this environment through actuators (Stuart and Peter, 2005). For example, a human agent has eyes, ears and other organs that act as sensors to its environment and it has muscles, vocal cords and other organs that are his actuators.
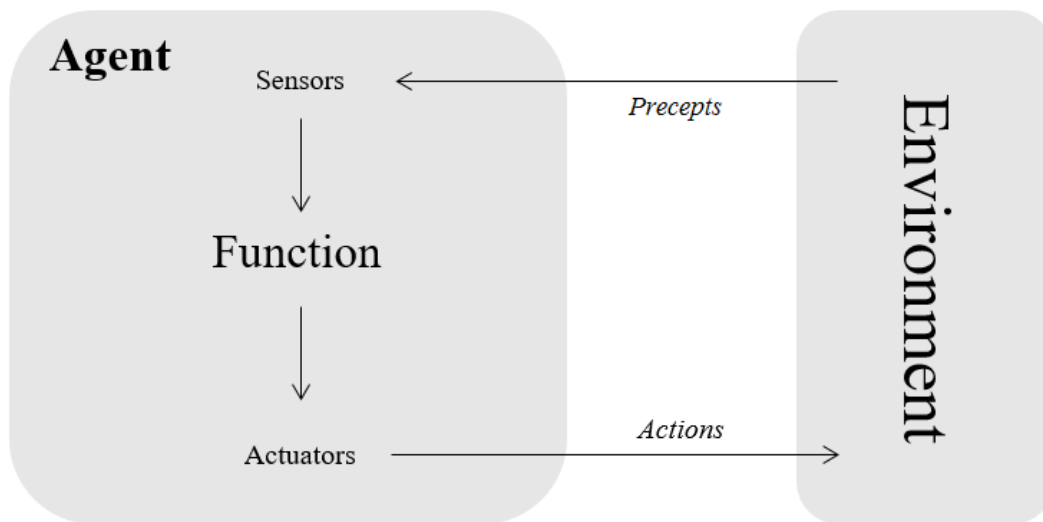

Figure I-11: Simplified architecture of an intelligent agent.

The agent, as described in **Figure I-11**, reacts with its environment following an approach described by a function or a strategy that associates every series of actions to a sequence of precepts. Once we created an agent and implemented it in its environment, we need to keep track of it and to judge if the agent is a 'wise' agent or not. In fact, when we implement an agent in an environment, it generates a sequence of actions depending on the sequence of stimulus it gets. Those reactions will make the environment experience a set of states. We consider an agent acceptable or wise if this set of states is preferable. Moreover, we quantify this acceptance by the performance measure.

The performance is a quantified quota that reflects the precision of the function or the strategy used by the agent. It can be associated to the number of times that the agent choses the good decision, in some simple tasks such as the classification. However and for much more complex jobs like the sequential decision problems, the performance must consider all the actions completed by the agent over the time because every one of them can have a different effect on the overall decision made.

2.2.   Machine Learning methods

The right agent for a certain environment needs to be wise but also it needs to learn as much as possible from what it experiences. Simply, a learning agent is an agent with a performance rate getting better over the time, using its own experience. The initial configuration of the agent's function may show some knowledge of the environment, but as the agent interacts more with it, its function may change and its action may be adjusted.

The first who wanted to solve the famous question 'Can machines think and learn?' was Turing (Turing, 1950). He initially thought about programming all the possible actions and relating them to every percept possible. Nevertheless, he abandoned this idea quickly for three major reasons. First of all, the agent's designer cannot predict the variety of situations that the agent might encounter. Secondly, the designer cannot know if there will be a change on the environment so he would not know how to anticipate the reaction to it. Finally, it is sometimes impossible for the designers to know how to program a task.

During the years, there have been many methods to teach machines how to learn from their experiences. One of the main differences between those techniques is the *feedback*. The type of the feedback determines the learning mode. In the *supervised learning*, the feedback matches the action the agent choses to apply depending on the set of percept it received. Therefore, the environment becomes the supervisor of the agent that corrects its errors (Kearns and Vazirani, 1994) . However, in the *unsupervised learning*, the agent does not receive any feedback. The agent starts with a collection of unidentified data and it builds its own strategy to describe this data by detecting similarities on its own without any interference from the outside world (Xu and Wunsch, 2008).

Between those two extreme sides, we have many other modes of learning such as *the Semi-supervised learning* and *the Reinforcement learning* that will not be implemented in this project. Therefore, we will skip their introduction.

2.2.1. Supervised learning

The basis of supervised learning is the idea of learning from examples. In this type of machine learning, we have two main sets of data to deal with; Z and Z^*. We call Z the training set and Z^* the test set of respectively N and N^* elements of input-output pairs. The supervised learning algorithms takes the training set $Z = \{(x_1; y_1), (x_2; y_2), \dots, (x_N; y_N)\}$, where $y_i = f(x_i)$, and discovers a resemblance function $\boldsymbol{h}$ to predict the unknown function $\boldsymbol{f}$. The learning process is a constant search for all the possible improvement we can implement to $\boldsymbol{h}$. To judge the efficiency of our resemblance function, we see how well it performs on our test set Z^*. The function $\boldsymbol{h}$, called also the hypothesis function, is said to generalize well if it predicts y with the smallest error possible for the x in the test set.

There are two different types of problems in the supervised learning; *Classification* and *Regression*. When the output set y is a number (such as the grade of a student), the problem is called regression. However, the problem is called classification if the output set is a restricted series of values (such as for the state of a student; absent, present or expelled).

Curve-fitting algorithms are one of the simplest examples for supervised learning algorithms. Suppose we have a number of outputs for an unknown mathematical function f as shown with the blue circles in **Figure I-12**. We want to search for the best function that will try to fit the largest number of points we already have from the set of possible candidates.
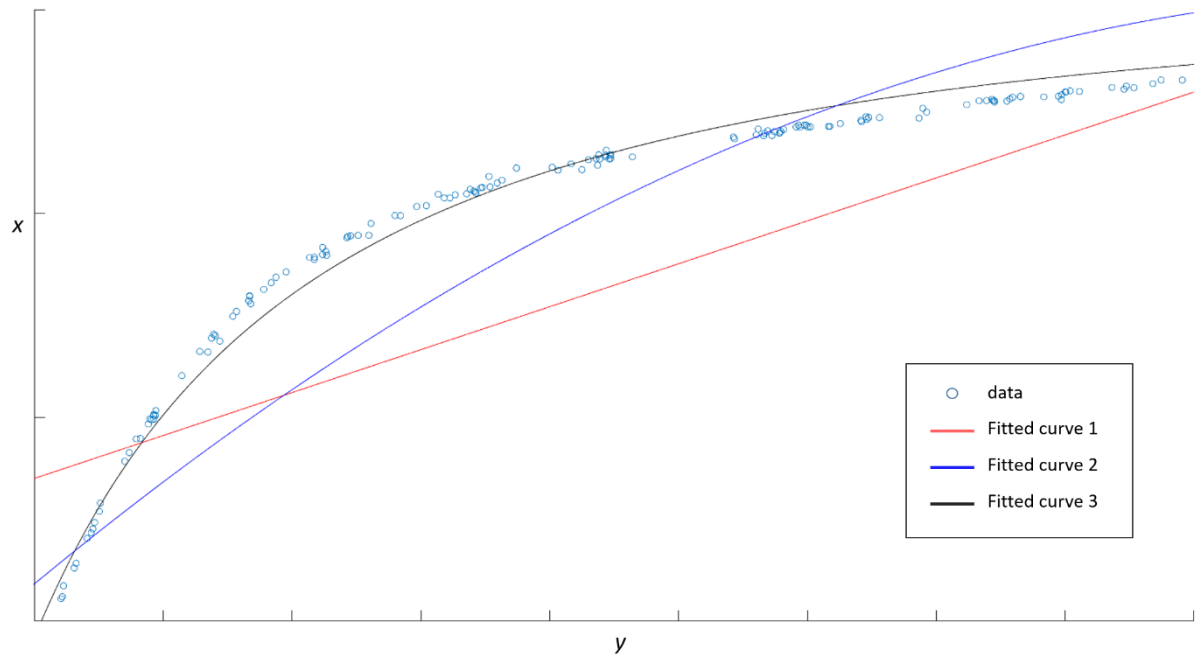
Figure I-12: Curve fitting examples.

These algorithms are categorized as supervised learning technics because they have a set of outputs in the training set that will send feedback about their exactitude to the agent working with them.

2.2.2. Unsupervised learning

The unsupervised learning does not require as much input data as its supervised sibling to start with. It uses a training set of inputs that is neither classified nor labelled and acts on it according to similarities and differences that the agent detects without supervision. No target outputs are chosen but the algorithm identifies them alone.

A popular example of unsupervised learning algorithms is Clustering, as shown in **Figure I-13**. In cluster analysis, the input set is separated into groups or 'clusters' based on some measure of similarity or shared characteristic. We find two categories of clustering; *Hard clustering*, where each input belongs to exactly one and only one cluster and *Soft clustering* where each input may belong to several clusters.
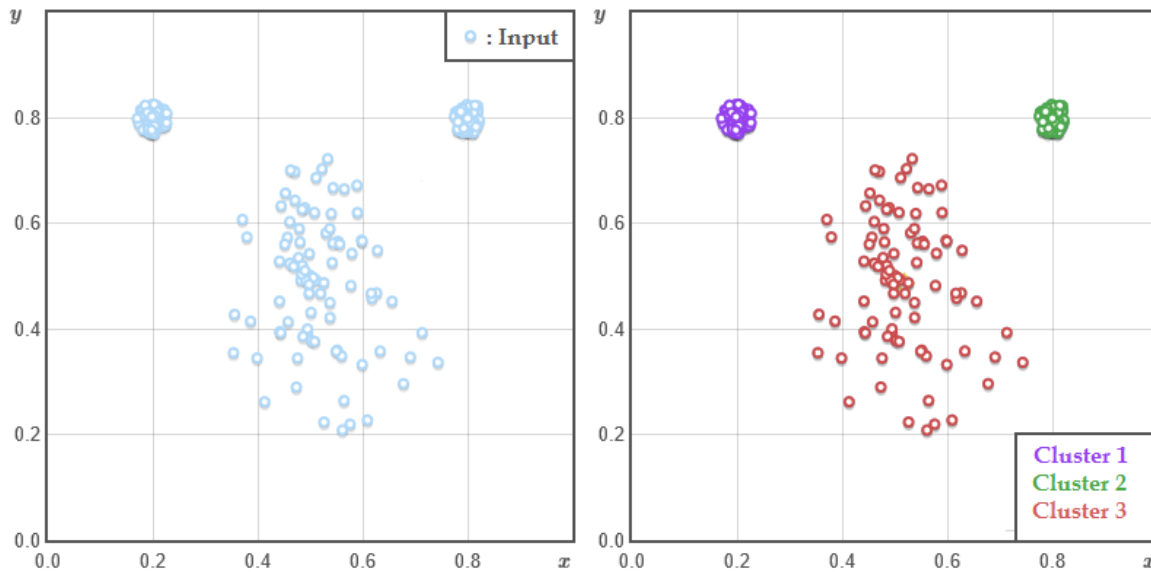
Figure I-13: Clustering result (left figure: raw data / right figure: division into three clusters).

Trying to evaluate the performance of unsupervised learning algorithms is not as easy as for the supervised ones. Actually, in some cases it is not even possible because judging how good the unsupervised agent was in performing a certain task will depend mostly on why it was used in the first place. Nevertheless, we can try to measure the quality of the output of an unsupervised system. There are two types of validity indices; *external indices* and *internal indices*. For example, in Clustering, an external index is a quota of agreement between two results where the first one is from a known cluster and the second is the one we are trying to fit in a cluster (Dudoit and Fridlyand, 2002).So, we are judging the result of clustering algorithm by a known cluster. The internal index is a measure of the 'goodness' of a clustering algorithm without using any external information (Thalamuthu *et al.*, 2006). Therefore, we are using quantities and features built in the input set, such as the Silhouette index, Distance between two clusters index (CD), Homogeneity index, *etc...*

2.3. Machine Learning for image classification

In this project, we are trying to explore the use of machine learning algorithms for image classification and computer vision purposes. Whether we use supervised or unsupervised machine learning technic and given a set of data describing the images, a trained algorithm should be able to learn the different classes of images we have and infer the class of the new ones.

To solve a classification problem we need two equally important elements: the databases and the learning algorithm. In our study, those databases are made from a combination of morphological and rheological data. Those terms will be defined in the next section of this manuscript.

2.3.1. Feature descriptors extraction technics

Collecting the data necessary to teach a machine-learning algorithm is one of the most critical phases (Due Trier, Jain and Taxt, 1996). We find low-level descriptors interested in pixel-level information and higher-level descriptors requiring an intermediate representation to get the information (such as a texture pattern description) (Srinivasan and Shobha, 2008). Not to overcomplicate this part of the workload, we chose to work with high-level descriptors.

Wear debris have been studied for a long time, mainly to correlate their morphology to the condition of the machinery (Davies, 1998; Davies, Prickett and Grosvenor, 1998; Price and Roylance, 1998; Myshkin, Markova and Andrei, 2005). According to those studies, third body features are divided into two main categories: qualitative and quantitative features.

To quantify an object we talk about its 'size'. The size of particle can be characterized by many figures such as the length, the width, the perimeter, the area … Those measurements are not easy to calculate in our case due to the complexity of the edges of the third body particles and the separation between the particle itself and its background as shown in the **Figure I-14**.

The qualitative features are the elements that specify the 'configuration' of an object. And for third body particles 2 characteristics are pointed; the form and the proportions (Starr *et al.*, 2001). The form describes the similarity of a particle to a regular geometrical structure such as a cube or a sphere. The proportion is the relative difference between that regular structure and the particle studied.

The exact features used in this work will be specified in the following sections.
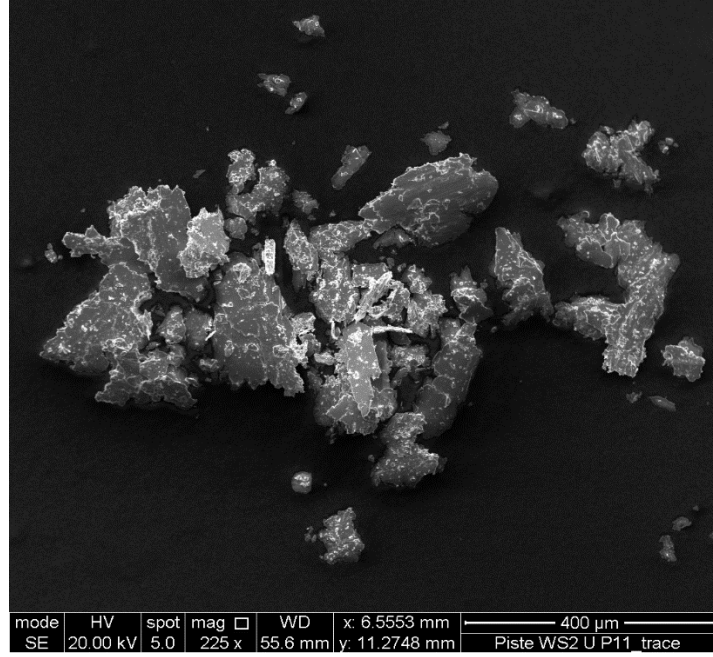
Figure I-14: A sample of third body particles.

2.3.2. Image classification algorithms

First, we need to specify the labels for some of the variables that will be used by the algorithms. To do so, we summarize in a general manner the logic that almost every machine learning algorithms follow:

1) From a set of data (Z) called learning database, where the input (X) and the output (Y) are both known, the algorithm forms a hypothesis function h_θ. In most cases, it is the weighted (θ) sum of the input (X). However, it might take other forms like a polynomial function. The output of the hypothesis function ($\widehat{Y}$) is supposedly the result the operator is searching for.

2) The algorithm calculates the 'cost' (J) of using that set of weights (θ). The cost function quantifies the performance of the algorithm when it is applied on learning database (Z) but measuring the difference between the expected result (Y) and the computed result ($\widehat{Y}$).

3) It modifies the value of the weights (θ) of the hypothesis function (h_θ) to minimize the cost function (J).

4) The classifier repeats step 2 and 3 for a certain number of time (NI: Number of Iterations). This is where we suppose the algorithm is learning.

5) The operator tests the accuracy of the final hypothesis function of a test database (Z^*).

We proceed in the next section to specify the two best-known algorithms used to solve classification problems. Those results of those algorithms are presented in chapter 5, section 2.

2.3.2.1. Logistic regression

The logistic regression was developed as a statistical model in the late of the 1960s to study binary data (its units take only two possible values) especially in the medical field(Cox, 1977). In spite of its label, logistic regression is used to solve classification problem(Hosmer and Lemeshow, 2000; El-Habil, 2012) such as the text classification(Nigam, Lafferty and McCallum, 1999) and co-reference resolution(Kehler, 1997) (Finding to which person or thing an expression refers to in a text). By calculating probabilities using logistic (Sigmoid) equation, it measures the relationship between an output class and one or more independent input variables. For example, it classifies a person as a female or male based on its height, weight and hair length.

We can distinguish between two types of classification problems; Binary and Multiclass classification.

- *Binary classification*

A binary classification problem is a classification problem that has only two classes to choose from. In this case, the probability for an input object to belong to one class is the complement probability for its belonging to the other class. If we have 0.7 as the probability of the input belonging to class 1 then automatically its probability for belonging to class 2 is 0.3. Therefore, we fix a threshold (Th) to determine to which class our input belongs to. The binary classification is as follows

$$\begin{cases} if\ h_\theta \geq Th \rightarrow y = 0|1 \\ \qquad\quad else \rightarrow y = 1|0 \end{cases}$$

As mentioned earlier, the algorithm classifies based on its own probability calculation. The algorithm associate a set of weights (θ) to the set of inputs (X) and calculates the classical weighted sum ($\sum \theta_i x_i$). Each weight θ_i is a real number ($\in \mathbb{R}$) that can indicate the importance of the feature x_i it is associated with. The result of the weighted sum is also a real number ($\in \mathbb{R}$) and therefore cannot be considered as a probability. The weighted sum passes through a sigmoid function that maps the real numbers to a $[0, 1]$ interval. The sigmoid function has the following equation and it is plotted in **Figure I-15.**
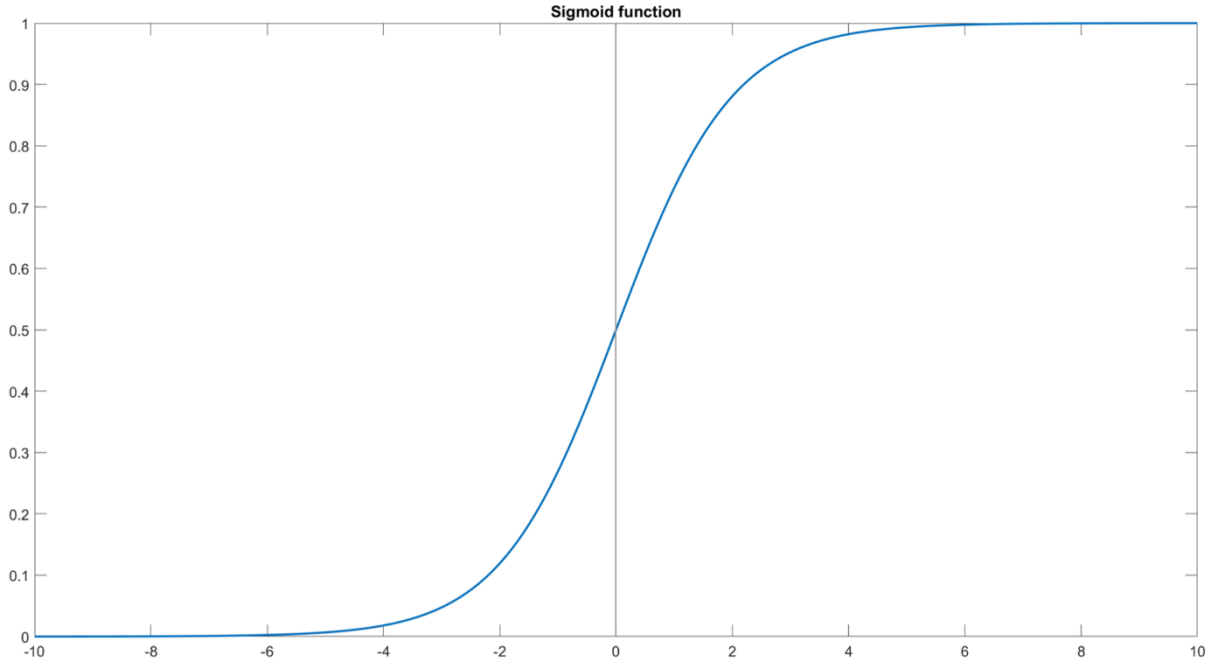
$$s(z) = \frac{1}{1 + e^{-z}}$$

Figure I-15: Sigmoid function representation.

Therefore, the hypothesis function for a logistic regression classifier is as follows:

$$h_\theta(X) = s(\theta \cdot X) = \frac{1}{1 + e^{-\theta \cdot X}}$$

$$with\ \theta \cdot X = [\theta_0 \quad \theta_1 \quad \cdots \quad \theta_n] \cdot \begin{bmatrix} x_0 \\ x_1 \\ \vdots \\ x_n \end{bmatrix}$$

Once we defined the hypothesis function, we now investigate the cost function. Many functions can be applied such as the hinge loss function (Tang, 2013), the Cauchy-Schwarz divergence function (Czarnecki, Józefowicz and Tabor, 2015), etc. However, each function has its own effect on the learning process (Janocha and Czarnecki, 2017). Therefore, in this work, we use the loss function, which is the most used with in classification problems: the cross entropy cost function. The equation for this loss function is as follows:

$$J(\theta) = -\frac{1}{m} \sum_{i=1}^{m} \left[y^{(i)} log(\hat{y}^{(i)}) + (1 - y^{(i)}) log\left((1 - \hat{y}^{(i)})\right) \right]$$

$$with:\ \hat{y}^{(i)} = h_\theta(x^{(i)}), m = size(Z)$$

If the correct answer $y^{(i)}$ is 0 (Class 1), the first term of the sum $\left(y^{(i)}log\left(\hat{y}^{(i)}\right)\right)$ will be null and the cost function will be 0 if our predicted result $\left(\hat{y}^{(i)}\right)$ is 0 too. The more the calculated result approaches 1 (Class 2, which is the wrong result), the bigger the cost function. The same applies if the correct answer $y^{(i)}$ is 1 (Class 2). This time the second term vanishes from the cost function, and the further from 1 the bigger the cost function.

At the start of the learning process, the weights are randomly chosen. However, the point of machine learning is to get better in time, and this purpose is reached by finding the best set of weights θ that minimizes the cost function $J(\theta)$. To do so, we need an optimization algorithm. Many are found in literature such as Conjugate gradient algorithm, Broyden–Fletcher–Goldfarb–Shanno (BFGS) algorithm, …Studying their effect on the performance of the machine learning algorithm shows that is depends on whether the data is highly correlated or not(Thomas P. Minka, 2003). We chose to work with Gradient Descent due to its good results when applied on logistic regression problems(Mayooran, 2018).

This algorithm searches for the first minimum it finds in a mathematical function. Our cost function $J(\theta)$ is a convex function since the exponential function $\left(e^{-\theta x}\right)$ is a convex function and the logarithmic function of a convex function is a convex function itself. Therefore, the gradient descent procedure will converge to the only minimum there is. It searches for the minimum by exploring in which direction the function is rising and moving in the opposite way.

In **Figure I-16**, we schematize the Gradient Descent algorithm. The function plotted can simulate the loss function $J(\theta)$. It is plotted as a function of the set of weights θ. The random initialisation of θ results into obtaining θ_1 and J_1. We calculate the gradient (the slope of the tangent at J_1) at θ_1. Since it is negative, the algorithm instructs us to move in the opposite direction, meaning that we need to increase θ and we obtain θ_2. The amount by which we move from θ_1 to θ_2 is controlled by the learning rate α.
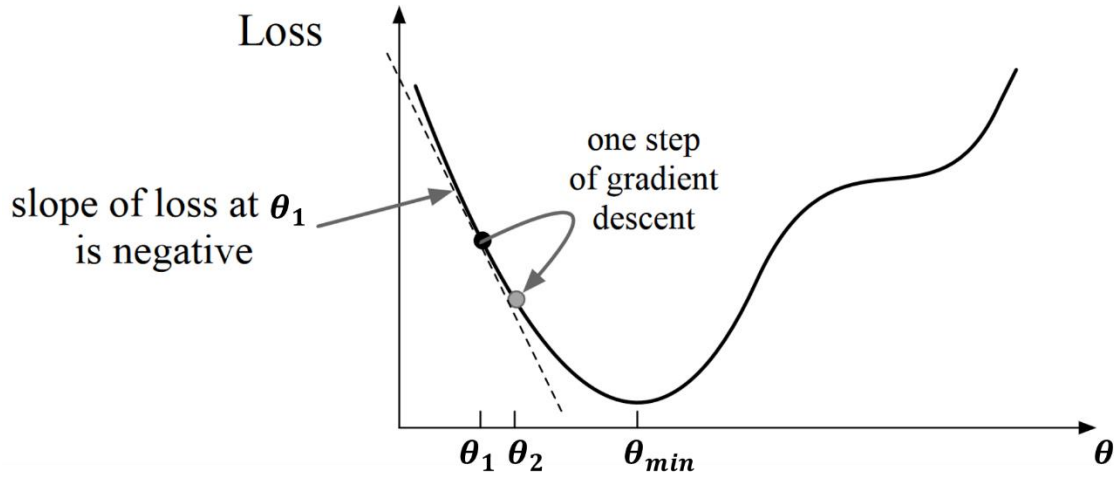
Figure I-16: Gradient Descent algorithm explained.

Mathematically, the Gradient Descent is translated as follows:

$$Repeat \left\{ \theta_j = \theta_j - \alpha \frac{\partial}{\partial \theta_j} J(\theta) \right\}$$

By developing the derivative, we find:

$$Repeat \left\{ \theta_j = \theta_j - \alpha \sum_{i=1}^{m} \left((\hat{y}^{(i)} - y^{(i)}) x_j^{(i)} \right) \right\}$$

This ensures that all the weights are updated at the same time. The number of times we repeat the optimisation process is defined during the implementation ($NI : Number\ of\ Iterations$).

- *Multiclass classification*

In classification problems, we do not always have just two classes to choose from. For example, if we need to determine the state of a car whether it is new, used or destroyed based on its fuel consumption and CO_2 emission, we have three different classes. This kind of problems is called Multiclass(Aly and Aly, 2005).

In order to solve the problem that has C classes, two approaches are accepted: **one-vs-all** and **one-vs-one**:

- **one-vs-all**: we divide our problem into C binary classification problems. In fact, the classifier is taking one class and then lumping all the others into a second class each time. We train therefore C binary classifiers. The input takes the class that its

classifier associated the highest probability $\hat{y} = \max\left(h_\theta^{(i)}(x)\right)$, with $i \in \{1, 2, \ldots, C\}$.

- **one-vs-one**: in this case, we train $\binom{C}{2} = \frac{C(C-1)}{2}$ separate binary classifiers. The input needs to pass through the $\binom{C}{2}$ classifiers before it gets a class. This later is associated to the class with the highest probability. One of the drawbacks of this technic is the high number of classifiers needed compared with the **one-vs-all** technic which can slow down the calculations.

For this thesis, we chose to work with **one-vs-all** approach to solve the multiclass classification.

- *Regularization:*

During the learning phase, we can have two different results:

- Under-fitting: A hypothesis function that cannot map the training dataset. Either this is caused by a too simple function or too few features used to characterize each example.
- Over-fitting: A hypothesis function that maps perfectly the training dataset but does not perform well on a test database. This problem is due to a too sophisticated function or the use of too many features.

In order to fix the under-fitting problem, the developer generally opts to increase the number of the features used or to change the model chosen for the hypothesis function. However, to fix the over-fitting, the developer either reduces the number of features (manually or using a selection algorithm) or regulates the hypothesis function.

In this work, we are studying the features we chose to describe the third body particles. Therefore, we need to keep their number unchanged, which made us opt for the regularization solution.

The regularization fixes the over-fitting problem by reducing the effect of the weights θ without changing their values. This penalisation is implemented in the cost function by adding a multiple of the regularization norm $N(\theta)$ by the regularization parameter λ. This transforms the cost function into:

$$J_{Reg}(\theta) = J(\theta) + \lambda N(\theta)$$

Many regularization models can be utilized such as the LASSO(Tibshirani, 1996), the RLAD(Wang, Gordon and Zhu, 2006), the Dantzig Selector(Candes and Tao, 2007) … To each model its effect on the learning procedure. For this study, we chose the model most common in classification problems: the Tikhonov regularization(Rong-En Fan *et al.*, 2008) also known as the ridge regression where $N(\theta) = \frac{1}{m}\sum \theta_i^2$. This transforms the cost function into:

$$J_{Reg}(\theta) = -\frac{1}{m}\sum_{i=1}^{m}\left[y^{(i)} log(\hat{y}^{(i)}) + (1 - y^{(i)}) log\left((1 - \hat{y}^{(i)})\right)\right] + \lambda\frac{1}{m}\sum_{i=1}^{n} \theta_i^2$$

The new regularized cost function is still a convex function since it is a sum two convex functions. Therefore, we keep the use of gradient descent technique in order to find the minima of the function.

2.3.2.2. Artificial neural networks

Artificial neural networks are limited simulations of the human brain. At a very basic stage, brain neurons are composed of three parts: the dendrites, the cell body and the axon. The dendrites receive the electrical stimulation that is treated by the cell body and the output is sent by the axon. In artificial neural networks, the feature attributes correspond to the signal that the dendrites receive the output corresponds to the signal the axon sends and the hypothesis function corresponds to the cell body. The similarities can be seen more clearly in **Figure I-17**.
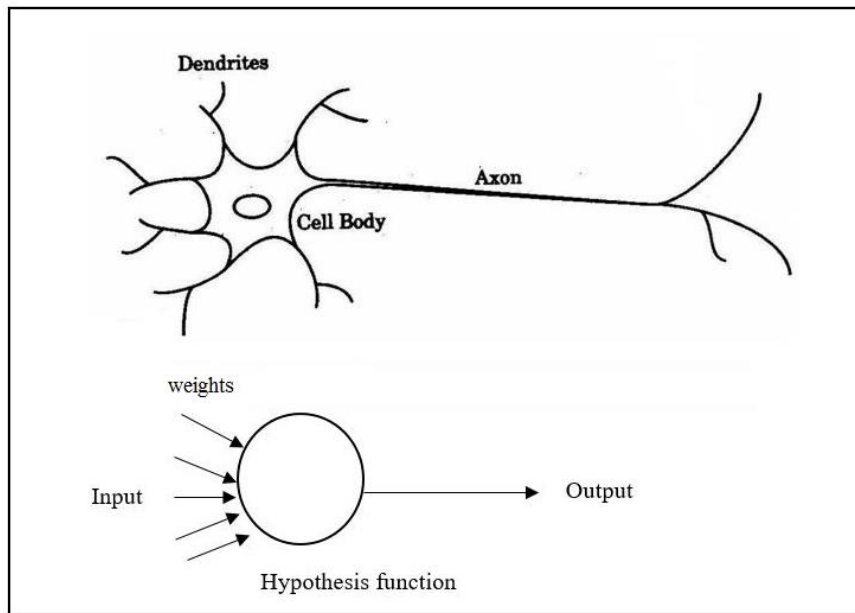

Figure I-17: Comparison between biological and artificial neurons

The artificial neural network (ANN) is a system of interconnected 'neurons'. Those connections have weights that are adjusted during the training phase of the learning. Usually the network has multiple layers of interconnected neurons. Each layer responds to a different combination of inputs that is the output of the layer before it. The artificial neural networks take a different approach from that used by conventional algorithms. They do not follow a set of instructions. ANN corrects itself using the training dataset by modifying the weights of each layer. This is why data used in the ANN should be selected carefully.

Artificial neural networks are implemented in different domains such as handwritten digits recognition, the diagnosis of disease conditions (Kalyan *et al.*, 2014) and face detection algorithms. From the benefits of the artificial neural networks, we can cite their ability to learn directly from data and the more data we have the more accuracy we obtain. Nevertheless, they have some disadvantages. Using ANN to solve a problem will definitely provide results but it will not explain how those results were obtained. Thus, it is usually the last resolution scientist resort to for answering a problem.

We keep the same logistic function (s) as in the logistic regression. A very simplistic representation of the neural network can be as follows:

$$Input \rightarrow [\ ] \rightarrow Output$$

The *Input* vector goes through *hidden layers* that results into calculating the result of the algorithm. The hidden layers are matrices (made of nodes) s calculated the same way as the hypothesis function in the logistic regression. To explain more we showcase an example in **Figure I-18** of an uncomplicated neurone network with one hidden layer with three nodes and an input that has four quantities:
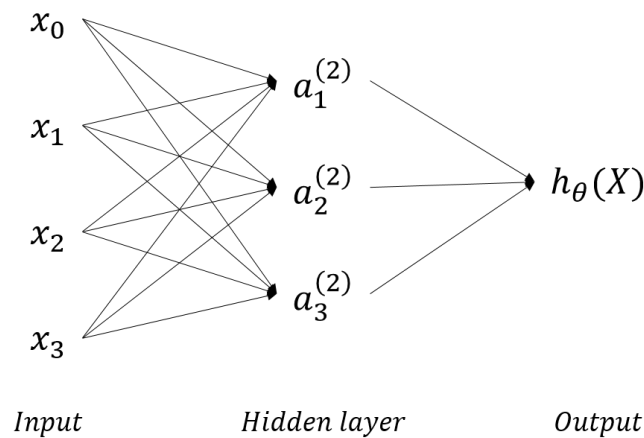

Figure I-18: Neural Network example.

The value of each node of this neural network is as follows:

$$a_1^{(2)} = s\left(\Theta_{10}^{(1)}x_0 + \Theta_{11}^{(1)}x_1 + \Theta_{12}^{(1)}x_2 + \Theta_{13}^{(1)}x_3\right)$$

$$a_2^{(2)} = s\left(\Theta_{20}^{(1)}x_0 + \Theta_{21}^{(1)}x_1 + \Theta_{22}^{(1)}x_2 + \Theta_{23}^{(1)}x_3\right)$$

$$a_3^{(2)} = s\left(\Theta_{30}^{(1)}x_0 + \Theta_{31}^{(1)}x_1 + \Theta_{32}^{(1)}x_2 + \Theta_{33}^{(1)}x_3\right)$$

$$h_\Theta(X) = s\left(\Theta_{10}^{(2)}a_0^{(2)} + \Theta_{11}^{(2)}a_1^{(2)} + \Theta_{12}^{(2)}a_2^{(2)} + \Theta_{13}^{(2)}a_3^{(2)}\right)$$

To calculate the value of the nodes a in a layer l, we can use the following equation:

$$a^{(l)} = s\left(\theta^{((l-1))} \times a^{(l-1)}\right)$$

knowing $\theta^{((l-1))}$ is a matrix with the dimensions $(p \times k)$ where p is the number of nodes at the layer (l) and k is the number of nodes at the layer $(l-1)$. Therefore at the end of the network, the hypothesis function is as follows:

$$h_\theta(X) = a^{(L)} = s\left(\theta^{((L-1))} \times a^{(L-1)}\right) = s\left(\theta^{((L-1))} \times s\left(\theta^{((L-2))} \times s\left(\ldots \times s\left(\theta^{(1)} \times X\right)\ldots\right)\right)\right)$$

where L is the number of layers in the network.

In the case of multiclass classification, the output $a^{(L)}$ of the neural network is a matrix with the dimensions $(1 \times n)$, with n being the number of classes we have.

The regularized cost function will keep the same form as in the logistic regression with few differences. It is given by:

$$J(\Theta) = -\frac{1}{m}\sum_{i=1}^{m}\sum_{k=1}^{K} y_k^{(i)} \log\left(\left(\hat{y}^{(i)}\right)_k\right) + \left(1 - y_k^{(i)}\right)\log\left(\left(1 - \left(\hat{y}^{(i)}\right)_k\right)\right) + \frac{\lambda}{2m}\sum_{l=1}^{L-1}\sum_{i=1}^{s_l}\sum_{j=1}^{s_l+1}\left(\Theta_{j,i}^{(l)}\right)^2$$

With : - m: number of examples used.

- K: number of classes (outputs).

- L: number of layers in the network.

- s_l: number of nodes in layer l

We can note in this cost function that the double sum adds up the logistic regression costs calculated for each cell in the output layer and the triple sum adds the squares of all the θ put in place in the network developed.

During logistic regression, we were able to use gradient descent directly as our optimization algorithm for the weights. However, it is not the case with neural networks. In neural networks, except for the last layer, none of the outcome of the intermediate nodes is pre-known. To solve this problem, the backpropagation(Rumelhart, Hinton and Williams, 1985) technique was proposed. Knowing the architecture of the network and given a cost function, this method

calculates the gradient of the cost function with respect to the different weights in place. This method starts with the gradient of the last layer (the output layer) and moves 'backward' into the network calculating the gradients until it reaches the input layer. Each gradient for a layer l depends on the gradient from the layer $l + 1$.

2.4.   Conclusion

A general state of the art review was established to know where we stand in this new technology. We went through the machine learning techniques that will be used in this work. A detailed explanation of the constituents of the learning database will be found in the next chapter along with a review of studies that dealt with similar subjects.

3. Machine Learning and Tribology

Despite being associated with one fifth of the world's total energy (Holmberg and Erdemir, 2017), the subject of tribology has a moderate understanding among the overall population. This problem comes from its poor availability. In other words, tribology seems like a science for the few enlightened. At the moment, only few experts around the world have the sufficient knowledge to understand what is happening in the contact and to estimate its lifetime.

The classical tribological analysis procedure did not include the quantification of wear debris, until it was proven that the study of those particles helps characterize and determine the wear mode (Stachowiak, Kirk and Stachowiak, 1991; Roylance and Raadnui, 1994a). Consequently, the microscopic examination of those particles was added to the tribological investigation.

3.1.   Third body morphological descriptors

The examination of the third body particles started as individual characterization of the debris. Scientists tried to relate every shape they found to a possible source. Seven shapes have been proposed, only based on visual characterization, with their possible origins (Hunt, 1993; Surapol and BJ, 1995; Podsiadlo and Stachowiak, 1997) as shown in **Table I-2**.

Table I-2: Link between particle shapes and their origins.

Shape observed	Possible origins
Sphere	Metal fatigue
Distorted smooth avoids	Quarry dusty
	Atmosphere dust
Chucks and slabs	Metal fatigue
	Bearing petting
	Rock debris
Platelets and flakes	Running in metal wear
	Paint
	Copper in grease
Curls, spirals and slivers	Machining debris produced in high temperature
Rolls	Similar to platelets but in rolled form
Strands and fibres	Polymers, cotton and wood fibres
	Occasionally metal

Many other researches have worked on giving metric dimensions to the feature describers when relating them to the wear type taking place in the contact (Roylance and Raadnui, 1994b; Xu, Luxmoore and Deravi, 1997; Roylance and Hunt, 1999; Roylance, Williams and Dwyer-Joyce, 2000). For example, we find that abrasive wear may lead to particles that have a thickness from 2μm to 5μm and a length from 25 μm to 100 µm. However, in fatigue wear, particles are found as spherical particles with a radius from 3 μm to 10 μm, as a group of small particles with a width ranging from 10μm to 100μm or as laminar particles with a length starting from 20 μm to 50 μm (Anderson, 1982).

This visual analysis of third body particles and the conclusions based on it are time consuming, subjective since they are based on an experienced opinion and not always consistent. To overcome these limitations and with the help of the accelerated technological progress we witness in computer science, tribologists resorted to automated image analysis algorithms to give objective feature description of the third body particles and introduced the numerical morphological descriptors.

The morphological descriptors characterize the body of the particle and they outline two major features; the shape and the surface texture. The shape is related to the contour and the outline of the particle while the texture relates the smaller details on the particle. In this project, we will only study the shape descriptors. The numerical descriptors vary from the simplest to the more complex. Common parameters are the length, the equivalent circle diameter, the

perimeter and the area. However, we also find the more complex descriptors such as the Spike parameter, Form factor, Elongation, Circularity … Each descriptor gives an information about the morphology of the particle. Hence, we must incorporate different descriptors (Blott and Pye, 2008) in order to get a full idea about the particle.

One of the works that standout in this field is the work of Kowandy et al (Kowandy *et al.*, 2007). In fact, it has some similarities with our project such as studying the morphological properties of third body particles created during a tribological experiment and then relating those descriptors to the wear regime taking place in the contact. In their work, the chosen descriptors were the length, the perimeter, the area and the Spike parameter. The Spike parameter is an indicator for the angularity of the particle (Kaye, 1993). In order to get the microscopic images of the third body sample, the scientists chose to move the third body particles from their original place to a glass plate. The displacement of the particles may have caused the modification of their morphology due to their delicacy. In addition, the number of particles analysed was not consistent from one test to another. However, this research proved the efficiency of using wear debris morphology to understand the wear that takes place in the contact, which encouraged us to explore the effect of using descriptors that are more complicated. Moreover, even though the friction coefficient and the wear rate were noted during this research, the link between those properties and the morphological state of the third body particles was not directly investigated.

3.2.   Machine Learning applications in Tribology

As more advanced technologies emerge, their application in different fields is always suggested. The same happened with machine learning and its capability in solving tribology problems. Many studies have tried to relate both fields for numerous purposes, like for example to optimize metal composites for a better tribological behaviour of their surfaces (Thankachan, Soorya Prakash and Kamarthin, 2018) and to diagnose mechanical defaults through the vibration signal of a machine (Meng Li, 2011).

However, the attempts to relate the morphological descriptors of the third body to its rheological parameters are not as numerous. Decost (Decost and Holm, 2015) worked on classifying materials from their microstructural features. In this work, the classification system created needs to work on a big database. In addition, it uses both simulation results and

microscopic images of concrete third body particles. Therefore, it requires computer clusters that are expensive and complicated to operate.

We can also find a research that aimed to diagnose the state of the mechanical contacts from the third body's morphological state (Laghari, Memon and Khuwaja, 2007). A hardware tool called Leica Quantimet Q500MC was used during this work. It is an easy-to-use image analyser system. It has a built in camera to shoot the third body sample and a software for the image processing procedure. This program dictates the morphological properties of the detected particles. In addition, the classification put in place did not take into consideration the rheological properties of the whereabouts of the creation of the third body.

4. Conclusion

The typical tribological analysis follows a two-phased approach: (1) a large-scale observation of the mechanical measurements of the system (*in situ*), and (2) close microscopic observations of the contact's surface and its boundaries after the test (*post mortem*). The *in situ* measurements may consist in recordings of the evolution of the coefficient of friction in time and/or the monitoring of the contact with high-speed camera to observe the behaviour of the third body particles (rheological data). However, the *post mortem* part typically consists in a qualitative judgement of the overall characteristics of the wear debris. Those 'measurements' may be highly biased and time consuming since their acquisition requires individual examination by a tribology expert.

With the aim of improving the analysis procedure and with the study of the third body particles created being a promising route, the improvements suggested in this field varied. The research was divided into either working on producing more morphological descriptors (more than 200 features (Wang and Wang, 2013) have been defined) and then choosing only the most efficient depending on the study case (Wang and Wang, 2013) or optimizing the image analysis algorithms themselves for a more precise results (F. Li, C. Xu, G.-Q. Ren, 2005; Liu *et al.*, 2016). However, some other measurements describing the tribological test were ignored, such as the rheological data.

The aim of this thesis is to evaluate the efficiency of linking the morphological data that describes the third body particles with its rheological measurement collected during the tribological experiment in order to be able to understand and predict the wear. The

morphological descriptors are quantitative measurements of the third body particles obtained by digitally analysing microscopic images of the debris found on the first bodies.

Instead of collecting data from previous experiments, we opted for a methodology where we conduct our own experiments. The main intention behind conducting our tribological tests is the production of an important quantity of third body particles in different controlled conditions. The amount of the particles is critical to obtain accurate morphological descriptors. Modifying the test conditions will result in a variety of morphological and rheological data. Possessing those two parameters for different tests under different conditions will create a rich database needed for the last part of this project, which consists in using machine-learning algorithms to link the two classes of properties. Classical steps of training and validation on separate subsets of the database are performed, and the success ratio in the validation phase is the indicator used to optimize machine-learning parameters. The chosen algorithm is expected to put in place a link between the rheological and morphological measurements and therefore to be able to predict one property from the other.

II. Experimental methodology

In this project, we do not seek to recreate a particular industrial contact but to investigate a concept. To do so, we carry out our own tribological tests to have control on the majority of the parameters. Thus, a tribological test bench has been set up. The bench consists of mainly two components: the tribometer and the measurement chain. The tribometer holds the different mechanical parts to ensure the ability to carry out the tribological tests. The measurement chain provides the necessary signals to monitor the tests and to study their different results.

We aim to study the morphological state of the third body particles and its rheological measurements and check the possibility of relating those two properties. To achieve this aim, we need a varied database of the two classes of properties. Therefore, tests under different condition are performed. It allows modifying the morphological and rheological states of the results. In addition, the material used for the first bodies and the contact conditions must generate an important amount of debris in order to get a sufficient number of morphological descriptors values.

1. Test bench

1.1. Tribometer

The tribological tests are carried out on a pin-on-disk tribometer previously assembled, shown in **Figure II-1**. This tribometer was improved during this thesis. We upgraded the tangential force measurement and installed new equipment to meet our needs (gas box, cameras, lightings …). Those changes are detailed later in this document. We study the sliding of the pin on a given track of the disk at a relatively low linear speed. The tribometer consists of an arm holding the pin in contact with the disk's surface, a motor to keep the disk in a continuous rotational movement and a force sensor to follow the tangential force sensor mounted on the supporting arm. The sensor sends its signal to an acquisition system (OROS35) to analyse it. OROS program presents the results as graphs of the coefficient of friction (COF) versus time and records those values.

The counterweights (8) are used to oppose to the weight (1) of the pin (6), its supporting arms and the different component on the arm itself (force sensor, fixing screws …). When applying $0 \, N$, the system is in a state of equilibrium and the normal force detected at the tip of the pin is very close to zero.
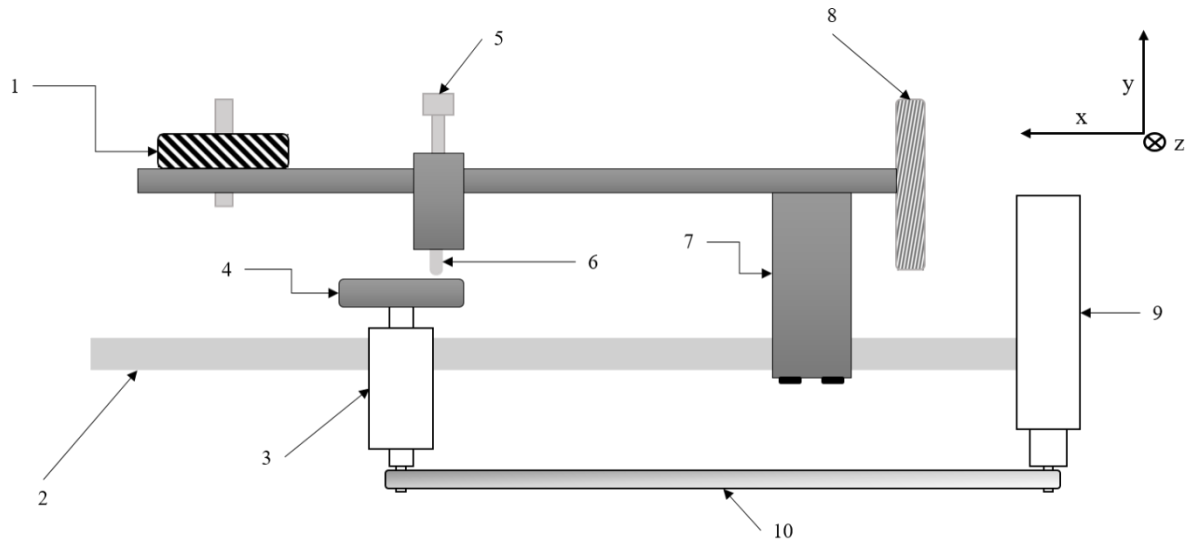
1: Normal load weights   2: Table

3: Disk supporting axis   4: Disk

5: Tangential force sensor   6: Pin

7: Support arm for the pin   8: Counterweights

9: Motor   10: Serpentine belt

Figure II-1: Presentation of the pin disk tribometer.

Once the pin is secured on its supporting arm, this arm keeps two rotational degree of freedom. We use as a reference the axis system in **Figure II-1** (top right corner). The pin's supporting arm can rotate around the *Z-axis* so that the pin can go up and down. It is only stopped by the disk's surface. In addition, the arm has approximately $\pm 2°$ of rotational freedom angle to turn around the *Y-axis*.

1.1.1. Materials

We chose the $35NCD16$ steel ($35NiCrMo16$ steel according to the DIN). It is a low-carbon alloy steel. This material is highly used in the manufacturing of highly stressed parts regardless of their dimensions across different industrial fields thanks to its high hardenability and important impact strength level. The chemical composition given by the supplier of this material and its mechanical properties are presented in **Table II-1** and **Table II-2** respectively.

Table II-1: Chemical composition of the 35NCD16 steel (percentage of the mass content).

C	Mn	Si	Ni	Cr	Mo	S	Fe
0,32 - 0,39	0,50 - 0,80	0,10 - 0,40	3,60 - 4,10	1,60 - 2,00	0,25 - 0,45	$\leq 0,025$	Balance

E	ν	Rp$_{0,2\%}$(MPa)	R$_m$ (MPa)	Hardness
205 000	0,3	1400	1870	≈50 HRC

1.1.2. First bodies

The dimensions of both the pin and the disk are imposed by the tribometer. The disk, presented in **Figure II-2,** has a 109 mm diameter and a 10 mm height. Whereas the pin displayed in **Figure II-3** is also a cylinder that has a 5 mm of radius and a 25 mm height. The pin has a hemisphere-like tip that measures 5 mm of radius. Both the pin and the disk have a surface roughness equal to $R_a = 0.2$.
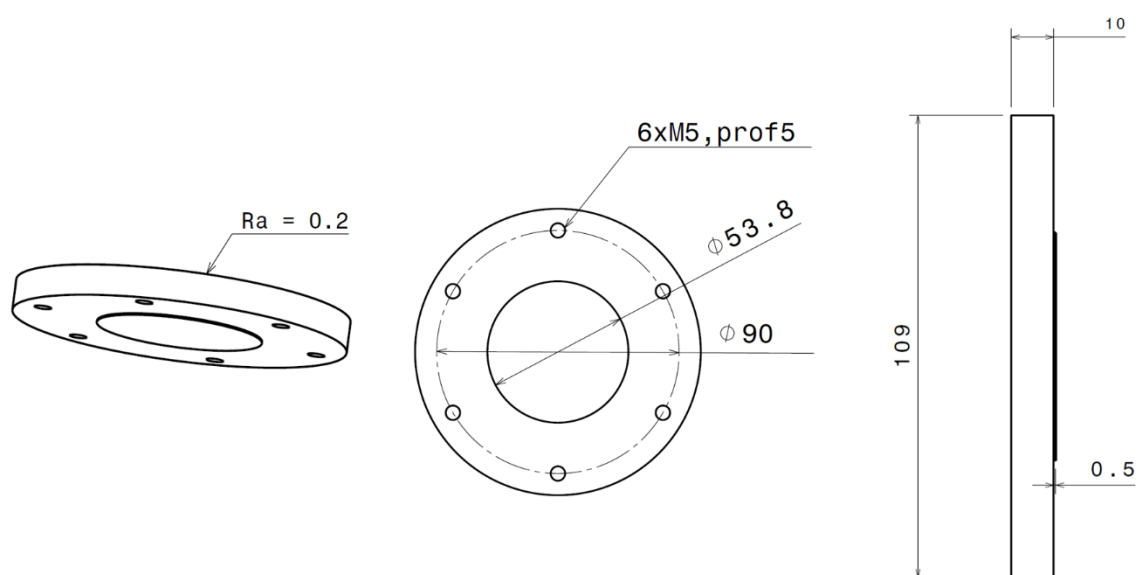
Figure II-2: Design drawing of the disk.
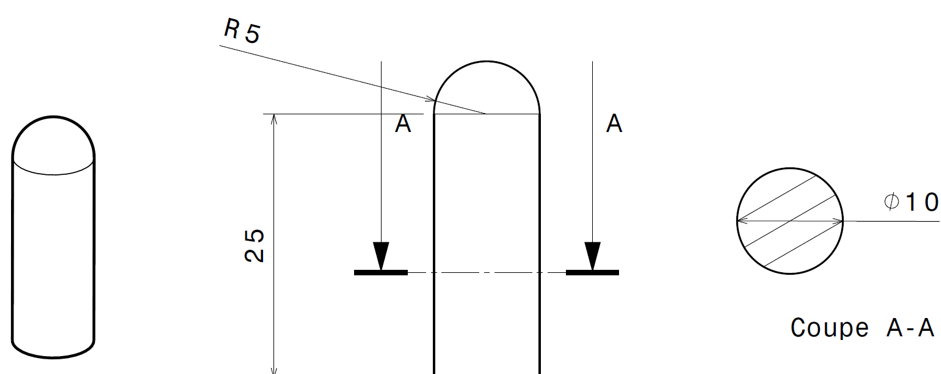
Figure II-3: Design drawing of the pin.

A private company covered the production of the pins and disks. The manufacturer offers two different states for the same roughness: a linear pattern (**Figure II-4.a**) and a circular pattern (**Figure II-4.b**). The linear pattern causes the reflected light to blind the camera every

half a lap. Therefore, we opt to use the circular pattern. In **Figure II-5**, we present a table that contains the results of analysing the surface using a profilometer. Those parameters are calculated following the red line that is visible in the figure.
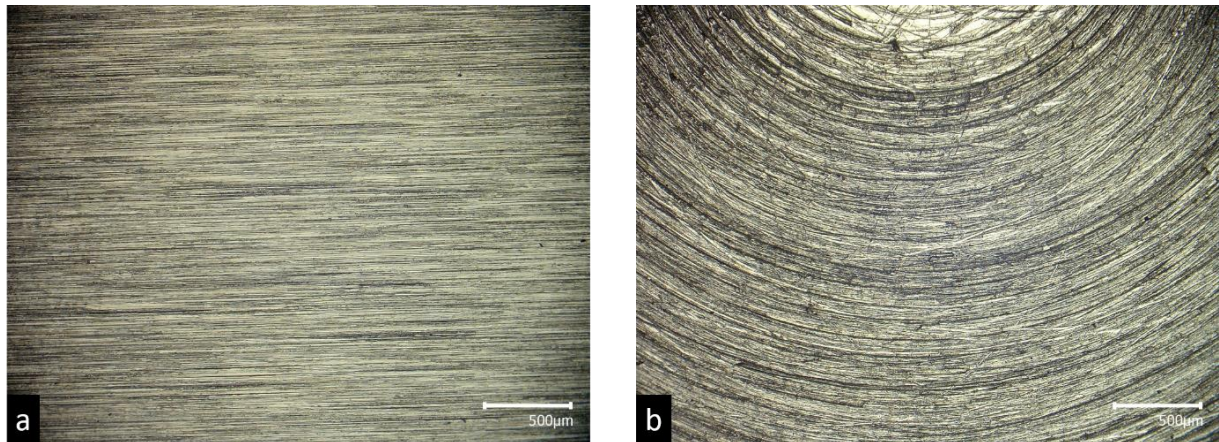
Figure II-4: the two surface states (taken near the centre of the disk to accentuate the surface) provided visualised using an optical microscope.

For each test, we use a new pin. Before the test, the surfaces of both the pin and disk are cleaned to remove any oil-like residue or particles resulting from the manufacturing process or the handling practices that may have an effect on the sliding test. The cleaning method consists of a set of steps:

- Remove the big apparent particles on the surface using a soft brush.
- Wash with water and dry immediately to avoid rust.
- Clean the surface wish Acetone.
- Put in an ultra-sonic cleaner for 10 minutes in a container submerging the piece in Acetone.
- Dry immediately.
- Clean the surface with Ethanol.
- Put in an ultra-sonic cleaner for 10 minutes in a container submerging the piece in Ethanol.
- Dry immediately.

ISO 4287			
Paramètres d'amplitude - Profil de rugosité			
Rp	0.449	µm	*Filtre gaussien, 0.08 mm*
Rv	0.589	µm	*Filtre gaussien, 0.08 mm*
Rz	1.04	µm	*Filtre gaussien, 0.08 mm*
Rc	0.732	µm	*Filtre gaussien, 0.08 mm*
Rt	1.14	µm	*Filtre gaussien, 0.08 mm*
Ra	0.213	µm	*Filtre gaussien, 0.08 mm*
Rq	0.255	µm	*Filtre gaussien, 0.08 mm*
Rsk	-0.25		*Filtre gaussien, 0.08 mm*
Rku	2.26		*Filtre gaussien, 0.08 mm*
Paramètres liés au taux de portance - Profil de rugosité			
Rmr	97.7	%	*c = 1 µm sous le pic le plus haut, Filtre gaussie...*
Rdc	0.475	µm	*p = 20%, q = 80%, Filtre gaussien, 0.08 mm*

Figure II-5: The disk's surface state.

1.2. Measurements

1.2.1. Forces

The recordings of the tangential force are critical for this study. They are essential to study the friction phenomena since we can deduce directly the friction factor values from it $\left(friction\ factor = \frac{tangential\ Force}{Normal\ Force} \right)$.. Consequently, it is important not only to choose wisely the parameters of the sensor but also to store the tangential force values during the test since it is not a constant during the test. The tangential force is transmitted to the force sensor via a mechanical joint. Since the sensor is mounted directly above the pin, there is no need to convert its signal. The sensor used is the SIKA FTCA50. It is a S-type sensor that can measure compression and tension. Its measuring ranges from 1 to 50 N. However, we cannot exceed 20 N in our case, which is the limit of the signal conditioner used.

There are two ways to analyse the signal. On one hand, we plot the entire signal, in 1D or 2D, so we can study its entire evolution is time that can give us an idea about micro (momentary) events as shown in **Figure II-6.A**.. On the other hand, we can plot the average of the signal per lap as shown in **Figure II-6.B**, which is the most common way found in literature. This highlights the general evolution of the friction coefficient.

Before using a sensor, it needs to be calibrated. This process consists in applying a set of known loads and analysing the response of the sensor. We systematically obtained a linear response. Still, the sensor's results did not have an excellent repeatability when repeating the calibration process in the same controllable conditions. The repeatability of the results was judged by calculating the standard deviation (SD) between five fitting equations of five different tests. The SD is equal to 2 % and the highest difference between the slopes of the

equations was equal to 13 %. This error comes mainly from how the sensor is installed in the tribometer. The sensor is fixed with two arms. The rod ends of those arms allows the sensor to keep only the ability of rotating around itself. This degree of freedom has an impact on the sensor's response. Since we can neither fix the tilt degree of the sensor nor perfectly reproduce it from one test to another, we chose to recalibrate it each time after fixing it before the test and we recheck the recalibration after each test was finished.
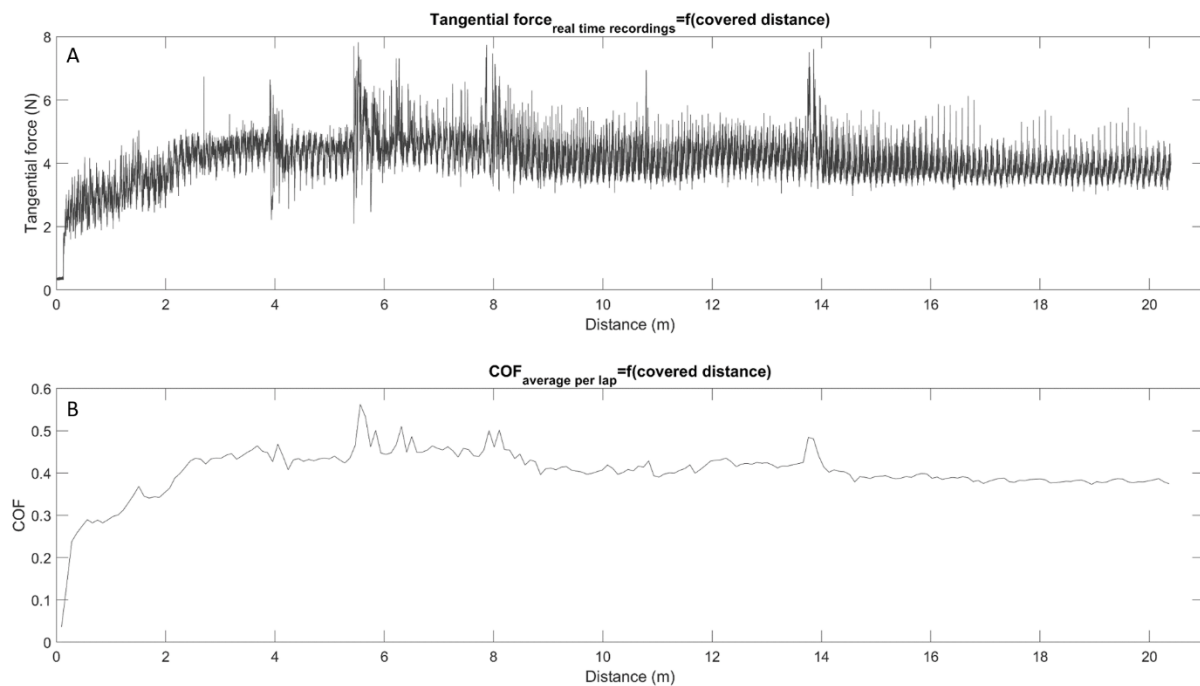
Figure II-6: A) The instantaneous evolution of the friction coefficient. B) The average of the friction coefficient per lap.

The calibration method is as follows:

- The sensor is fixed on the supporting arm of the pin. This arm is lowered as if the test was to be conducted. However, using a set of blocks, the pin is kept a few millimetres away from the disk.
- A set of calibrated masses is used. The calibration load is applied using a string that is attached to the pin at its other end (a wheel is used to change the direction of the force from vertical to horizontal). We start with $100\ g$ and adding weights until we reach $600\ g$. Then we remove the masses in the reverse order to check the reversibility of the measurements.
- The string and the wheel are then removed from the bench and the arm is fully lowered so the pin touches the surface of the disk. If the arm is risen again, the calibration method needs to be run again.

1.2.2. Accelerometer

The accelerometer is a sensor that sends three different signals surveying the vibration levels of three axis of the objects it is attached to. The three signals are read through OROS just like the tangential force signal. The sensor is placed directly on the pin in our tests as shown in **Figure II-7**, as well as the three axis that it follows.

| 1: Disk | 2: Accelerometer |
| 3: Pin | 4: Support arm for the pin |
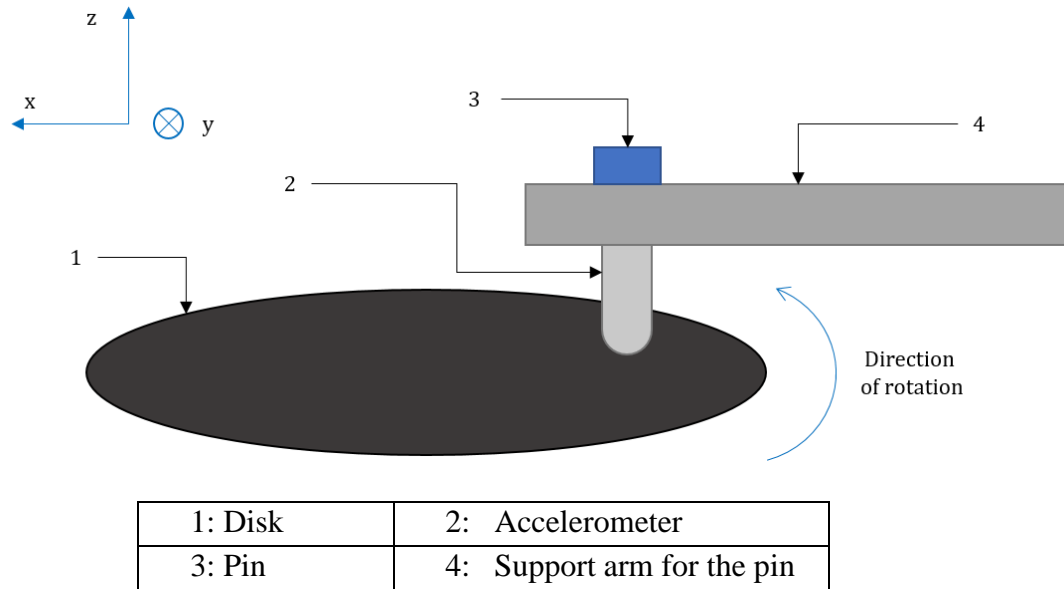
Figure II-7: Accelerometer placement and corresponding system of axes.

The accelerometer's signals will be used in our case to attempt to explain the different events we record in the friction factor graphs. For example, if we record a peak in the friction factor signal and a change in the signal from the z axis and a change in the signal from the y axis at the same time, we can assume that a large third body particle passed in the contact, which led to sudden pull (y axis) and rise (z axis) of the pin. This will be further detailed in the next chapter.

1.3. Gas box

It was proven that interactions take place between the materials in contact and the environment of the experiment. In fact, many studies confirm that many chemical reactions due to gas consumption can change the way materials behave when submitted to tribological loadings (Gouider *et al.*, 2004; Kasem *et al.*, 2007; Colas, 2013). Therefore, the environment is considered as an additional variable to be changed to vary the morphological measurements and the rheological data.

Due to the location of the pin and the disk, it is impossible to change the gaseous environment of the contact without covering the entire top part of the tribometer (the supporting arm for the pin, the weights for applying the normal load, force sensor and the counterweights as shown in **Figure II-1**). The designed solution is a cube to cover those parts. It is made of Plexiglas, to maintain the possibility of filming the contact with the cameras being out. The sketch of the cube with its different dimensions is shown in **Figure II-8**.
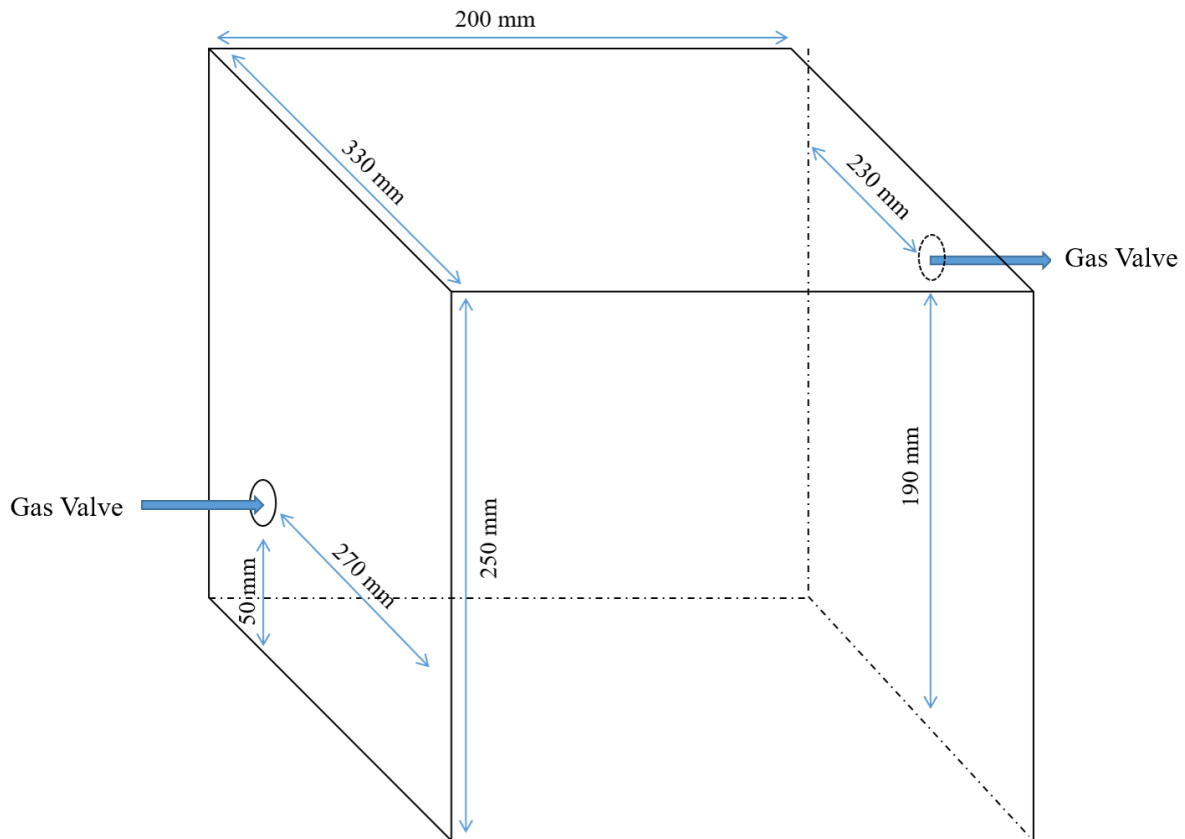
Figure II-8: Sketch of the gas cabin.

The original design of the tribometer did not take into consideration the possibility of adding the gas box as a feature, hence the lack of fixation mechanism. In order to avoid heavy changes in the tribometer, the box is fixed on the table using insulation tape. Therefore, certain leaks are bound to be present. To correct this phenomenon, a low gas flow is let in during the test to compensate for the loss.

The two valves are on opposite sides of the box and have different heights. Each one of them can be used as inlet or outlet depending on the gas used. For example, if the gas chosen is lighter than air, then we chose the higher valve as the inlet and the lower one as outlet. To fill the box at first, we close the outlet valve and let a flow of the gas get inside the chamber until the

pressure inside is higher than the pressure outside. When opening the outlet valve, the air (the lower gas) will go out first. For gases more dense than air, we switch the two valves.

2. Image acquisition

2.1. *In situ* monitoring

The utility of using *in situ* real time visualisation of the tribological test has been proven in the observation and the quantification of the wear debris (Sliney, 1978; Scharf and Singer, 2003; Wahl, Chromik and Lee, 2008), especially with cameras evolving and their specifications getting better. Primarily, those recordings help us deduce the rheological properties of the third body particles, which are the ductility and the cohesion of those particles (Descartes and Berthier, 2002). In addition, the *in situ* recordings can help explain the micro events we can find in the COF figures since they follow the contact in real time.

The test bench is equipped with two high-speed cameras (GO 5000). This model was chosen after a comparison between different cameras (GoPro 4, Phantom V710, RS pro x10) and it had the best results for our needs. The camera's specifications are presented in **Table II-3**.

Table II-3: Camera specifications.

Sensor	CMOS
Resolution	$2560\ (Horizontal) * 2048\ (Vertical)$
Pixel size	$5\ \mu m$
Frame rate	Up to $62\ fps$

The cameras are equipped with a 35 mm lens and a variety of extension tubes to enhance the focal distance to get the best recordings possible. Additional light sources are important in order to reduce the exposure time and to pick up sharp details of the object. The lightings mounted are two XL2150 from Spectrum Illumination. They are ultra-high power LEDs producing $559\ lm$ per LED of pure white light. The lights can be moved and their angle can be changed to ensure that the coverage area is the same as the contact area.

The cameras can be placed to follow different spots with the help of their supporting fixing arms. We can follow the inlet of the contact, the outlet, the sliding track (away from the contact) and we can follow the contact perpendicularly in the case of using a transparent first body.

During our tribological test, we chose to supervise only the outlet of the contact between the pin and the disk in order to follow the behaviour of the third body particles.

2.2. *Post-mortem* observation

At the end of each tribological test, the two first bodies are separated and dismounted from the tribometer. Both the sliding track on the disk and the contact zone on the tip of the pin are observed through a scanning electron microscope (FEI Quanta600), shown in **Figure II-9**. We chose not to move the third body particles from the surfaces where they laid during their creation in order not to affect their form in any way. We need to be aware that only the act of opening the contact between the pin and disk, even though necessary to go through studying of the results of the sliding test, adds a number of errors to the analysis.
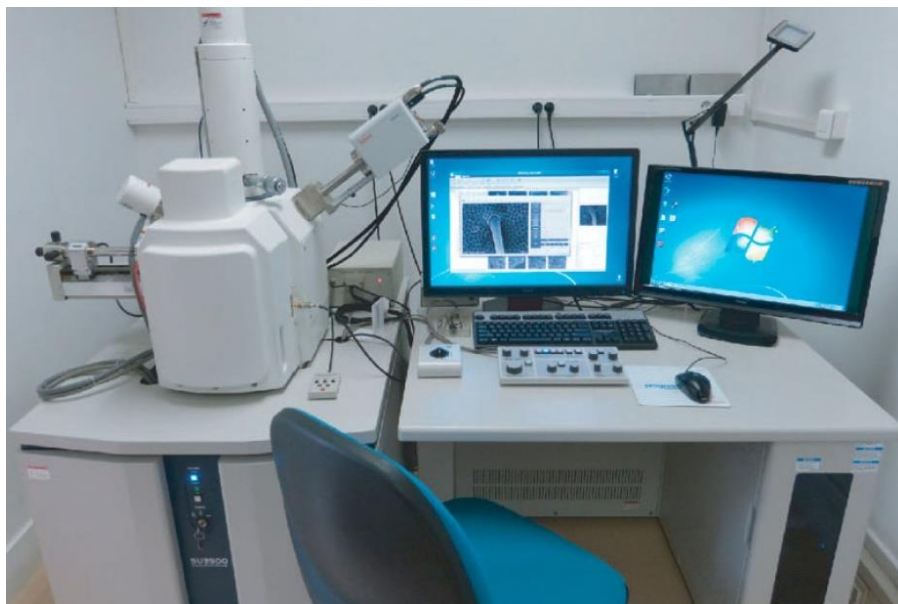
Figure II-9: Scanning electron microscope.

The Scanning electron microscope (SEM) applies a focused electron beam on the sample we want to investigate microscopically. Depending on the specimen's properties and the electron beam energy, a number of signals are emitted back from the sample, such as the Auger electrons, the Secondary electrons, the Backscattered electrons … Mainly, we only use the Secondary and the Backscattered electrons.

- *The Secondary electrons* (SE): When the atoms of the sample have enough energy from the electron beam, secondary electrons are created. However, due to the small energy of the SE, only the surface's SE stay significant enough to be detected.

Therefore, this method is used primarily to describe the morphology of the sample's surface. The SE are sensible also to the angle between the surface and the electron beam: the energy of the SE get larger when that angle deviates from 90°. This is called the edge effect. The brighter the spots on the image are the more width they have and the darker the parts are the flatter they actually are.

- *The Backscattered electrons* (BSE): The backscattered electrons are the electrons reflected back from the samples. They have higher energy levels. Thus, they describe deeper regions of the sample than its surface. The BSE are susceptible to the edge effect also. Therefore, they describe the topography of the sample in more details than the SE. However, they are very sensitive to the chemical composition of the sample used. Therefore, the resulting image can also describe its chemical composition (Z-contrast). The resulting image will not give in details the different chemical elements that make the sample or present on its surface but it can indicate that there are different element present depending on the brightness levels.

In **Figure 10**, we present three images of the same sample: SE (**Figure 10.a**), BSE (**Figure 10.b**) and Z-contrast image (**Figure 10.c**). We can see that the information from the SE and BSE image are different. The BSE image emphasis more on the rough edges of the surface when the SE is more of what is expected to see by the naked eye at that scale. The Z-contrast shows different brightness levels, which indicate the presence of different chemical elements. To investigate those elements more, we can use the EDX sensor attached to the SEM that sends back a report with detailed information about its chemical composition.
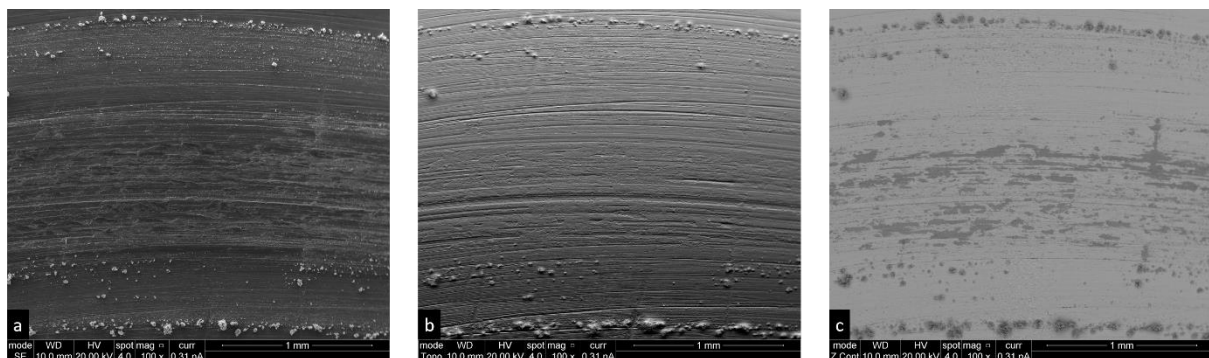
Figure II-10: The three imaging options from the SEM: a) SE image. b) BSE image. c) Z-contrast image.

In this project, we use the SEM primarily to investigate the morphology of the third body particles. Therefore, the microscopic images need to cover the totality of the contact area on the pin and the sliding track on the disk to gather the majority of the wear debris particles created during the tribological test.

The microscopic images will be analysed thoroughly to extract all the information they offer. Thus, we opt to collect images with the best resolution possible while still keeping a reasonable scanning time. Not only the SEM's configurable image properties are considered, but also the file formats when saving the images were appropriately selected.

3. Test conditions

As expressed earlier in this chapter, both the morphological and rheological data of the test need to vary. To ensure it, we chose to work under varying test conditions. The test bench allows us to change the following parameters: the linear speed of the disk, the normal load and the gaseous environment.

3.1. Applied normal force

The normal load is applied using the weights stacked on their placement on the pin's supporting arm (**Figure II-1**). The distance between the masses and the pin is a constant that we cannot control due to the tribometer design. Therefore, in order to have normal loads that are for example a set of multiples of five, we need a very specific set of masses. We chose to keep the applied force a constant at $10\ N$, which means applying a weight of $600\ g$, in order not to surpass the tangential force sensor limits. The shape of the first bodies and the normal load ensure a point contact in the initial state of the experiment with an initial Hertzian maximum pressure equal to 977 MPa at the contact point. The initial average Hertzian pressure in the contact point is equal to 651 MPa.

3.2. Sliding speed

The linear speed depends on both the rotational velocity and the radius of the track used on the disk: $linear\ speed = Rotational\ speed * Radius * 2\pi$. The command box of the motor controls the rotational speed. Theoretically, it can vary between 0 and 50 laps per minute, with a 0.1 lap per minute increments. However, the disk is not mounted directly on the motor's shaft but it is fixed on a separate rod connected through a belt to ensure the rotational movement. We notice while using the motor that the moment created by a command velocity under 2.5 laps

per minute will not suffice to move the disk. In addition, if the rotational velocity is superior to 15 laps per minute, significant vibrations are noticed. Those vibrations come from the motor itself and the movement of the transmission belt. We fix the rotational speed value to be in the range $[3, 10\ rpm]$. The linear speed can still have a wider range since it depends also on the radius of the sliding track itself along with the rotational speed.

3.3. Additional properties

Due to the difference between the radius of the disk and the radius of the pin (10 times bigger) and since the supporting arm of the pin can move forward and backward, we chose to partition the surface of the disk into 5 different sections so we can conduct 5 independent tests per disk. The disk theoretical division follows **Figure II-10**.
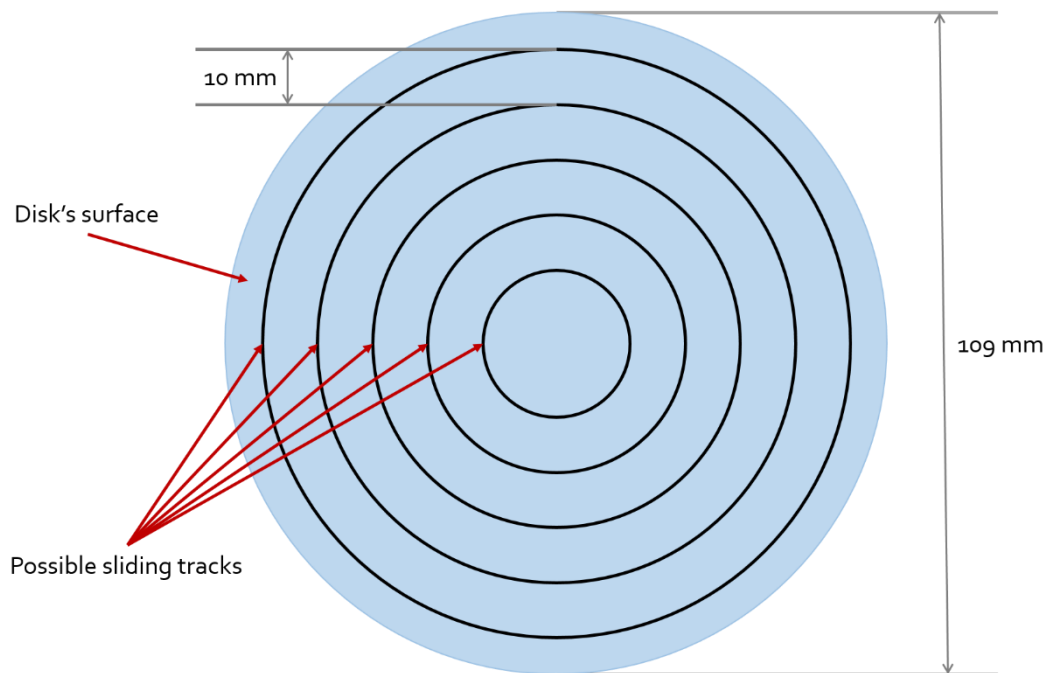
Figure II-11: The division of the disk's surface to accommodate five sliding traces.

The gaseous environment is changed with the help of the container designed for the tribometer. The gas chosen for this project are air and argon. The latter is wildly used with low-alloyed steel, especially in welding applications. As explained earlier, a low gas flow is maintained through the test to compensate for the leaks. The advantage of working with argon is that it is heavier than air so the contact is guaranteed to be immersed in the gas since it is at the bottom of the cabin. The inlet valve (lower valve) of the gas is a few millimetres under the contact between the pin and disk, so that we ensure the presence of the gas in the contact but

do not disturb the created particles by any gas flow. We compute the flow to be equal to $Q = 8.4 * 10^{(-3)}\ m^3.s^{(-1)}$ and that we need $10\ seconds$ to fill the gas tank.

Accordingly, we close the outlet valve for around $10\ s$. Then we open the outlet valve and so the air goes out. We keep the flow of the argon the same for 10 more seconds to be sure that all the air went out. Afterwards we close the outlet valve and lower the pressure of the Argon container to $0.02\ bars$ in order to correct the leaks that take place inside the box.

The temperature and humidity are not imposed in this project but they are recorded during each test. Keeping record of both those parameters may help to explain the differences between the results between the tests. The two properties are noted from a phone application that uses its sensors. The maximum error induced from the sensors is 10 %, according to the application description.

In this work, we conducted nine different tests to compare the effect of changing the different controllable conditions on the morphological and rheological measurements. **Table II-4** recaps the different test carried out:

- Set 1: {Test 1, Test 2, Test 3}: Different linear speeds.
- Set 2: {Test 4, Test 5, Test 6}: Different Gaseous environments.
- Set 3: {Test 7, Test 8, Test 9}: Different covered distances.

Table II-4: Experimental conditions' summary.

Set	Tag	Radius (m)	Normal force (N)	Linear speed (m/s)	Temperature (°C)	Humidity (%)	Gaseous environment	Covered distance (m)
Set1	Test 1	0,02		**0,7536**	22	51	Air	$\cong 25$
	Test 2	0,015		**0,5652**	20,3°	48,90		$\cong 24$
	Test 3	0,01		**0,3768**	23	50		$\cong 25$
Set 2	Test 4	0,02	$10\ N$	0,628	21,8	43,7	**Air**	$\cong 18$
	Test 5	0,02		0,628	20	50	**Argon**	$\cong 15$
	Test 6	0,02		0,628	23	50	**Argon**	$\cong 17$
Set 3	Test 7	0,01		0,3768	27°	42	Air	$\cong \mathbf{25}$
	Test 8	0,01		0,3768	23,4°	49		$\cong \mathbf{22}$
	Test 9	0,01		0,3768	21°	51		$\cong \mathbf{15}$

4. Conclusion

In this chapter, we detailed the project's strategy where we explained the aim and the contribution of this work in the tribological analysis approach. We introduced the material used for the first bodies with its different mechanical and chemical properties. We then described then the tribological bench with its components and their utilities. We explained the different visualisation methods used for observing the sliding track and the third body particles on the first bodies. Finally, we summarized the test conditions and listed the nine different tests, which results will be analysed in a following chapter

In the rest of this manuscript, we analyse the results of the tests conducted. We describe the approach used to form the rheological database (Chapter 3) and the morphological database (Chapter 4). Finally, we present the machine learning algorithms (Chapter 5) developed and investigate their predictive ability.

III. Rheological data

Two types of information, that are necessary to build the database needed for the machine learning algorithms, are extracted from the tests conducted in this project. In this chapter, we detail the first type of information, namely the rheological data. As explained in the first chapter, rheology is the science of the flow and the deformation of materials, the third body in our case. A direct measurement of its rheological properties would require the use of a rheometer. However, this is not possible since the third-body cannot be extracted from the contact without significant disturbance. Besides, its properties change depending on many factors such as the loading conditions, the environment, and the internal structure of the first body materials. We must thus rely on indirect measurements.

In this project, we chose a number of parameters that describe the third body created and the state of the first bodies after the test is conducted. Even though, they are not commonly agreed upon, they will suffice for our need especially since the aim of this work is to be a proof of concept for the machine learning and its ability to establish a link between the rheological data and the morphological descriptors for the third body particles created during each test.

This chapter can be divided into two sections. In the first section, we detail the analysis protocol to obtain the rheological measurements from a single test. We detail the computations for every rheological parameter for this test. The second part is where we present the rheological dataset for all the tests from each set.

We summarize the different tests and their experimental parameters in **Table III-1**. In set 1 we changed the rotational speed of the disk. We varied the gaseous environment where the contact takes place between ambient air and argon between the tests grouped in set 2. We changed the sliding distance of the tests in set 3.

Table III-1: Experimental conditions' summary.

Set	Tag	Normal force (N)	Linear speed (m/s)	Temperature (°C)	Humidity (%)	Gaseous environment	Distance covered (m)
Set1	Test 1		**0,75**	22	51	Air	$\cong 25$
	Test 2		**0,57**	20	49		$\cong 24$
	Test 3		**0,38**	23	50		$\cong 25$
Set 2	Test 4	10 N	0,63	22	44	**Air**	$\cong 18$
	Test 5		0,63	20	50	**Argon**	$\cong 15$
	Test 6		0,63	23	50	**Argon**	$\cong 17$
Set 3	Test 7		0,38	27	42	Air	$\cong 25$
	Test 8		0,38	23	49		$\cong 22$
	Test 9		0,38	21	51		$\cong 15$

1. Analysis protocol

The analysis protocol is a standardized procedure that allows us to obtain the rheological data from the information gathered during and after the tribological tests. The outputs of the tests are divided into two types; the dynamic data and the post-mortem visualisation. The dynamic data is the ensemble of signals (the tangential force and the three vibration levels) recorded during the sliding test along with the in-situ video recordings of the contact outlet. The post-mortem visualisations are the microscopic views of the contact area on the pin and the sliding track on the disk. From each type of data, we calculate a number of descriptors that form the rheological database.

1.1. Dynamic results

1.1.1. Friction factor curves

Figure III-1 displays the evolution in the total travelled distance of the average per lap of the Coefficient of Friction (COF) for test 5. The first observation we can state is that the COF keeps varying during the test and does not keep a constant value.

Upon closer study of the plot, we can distinguish three different phases:

- *Phase 1 (Before the black dot)*: An important increase in the COF is observed, where it goes from 0 to almost 0.24.
- *Phase 2 (Between the black dot and the red square)*: The COF changes in value but not in an extreme way. We can say it keeps a more or less constant value.
- *Phase 3 (After the red square)*: Sudden value changes (peaks) appear in the curve. Both the peaks value and their frequency of occurrence increase in time. The value to which the COF goes back to after passing each peak changes during the test.
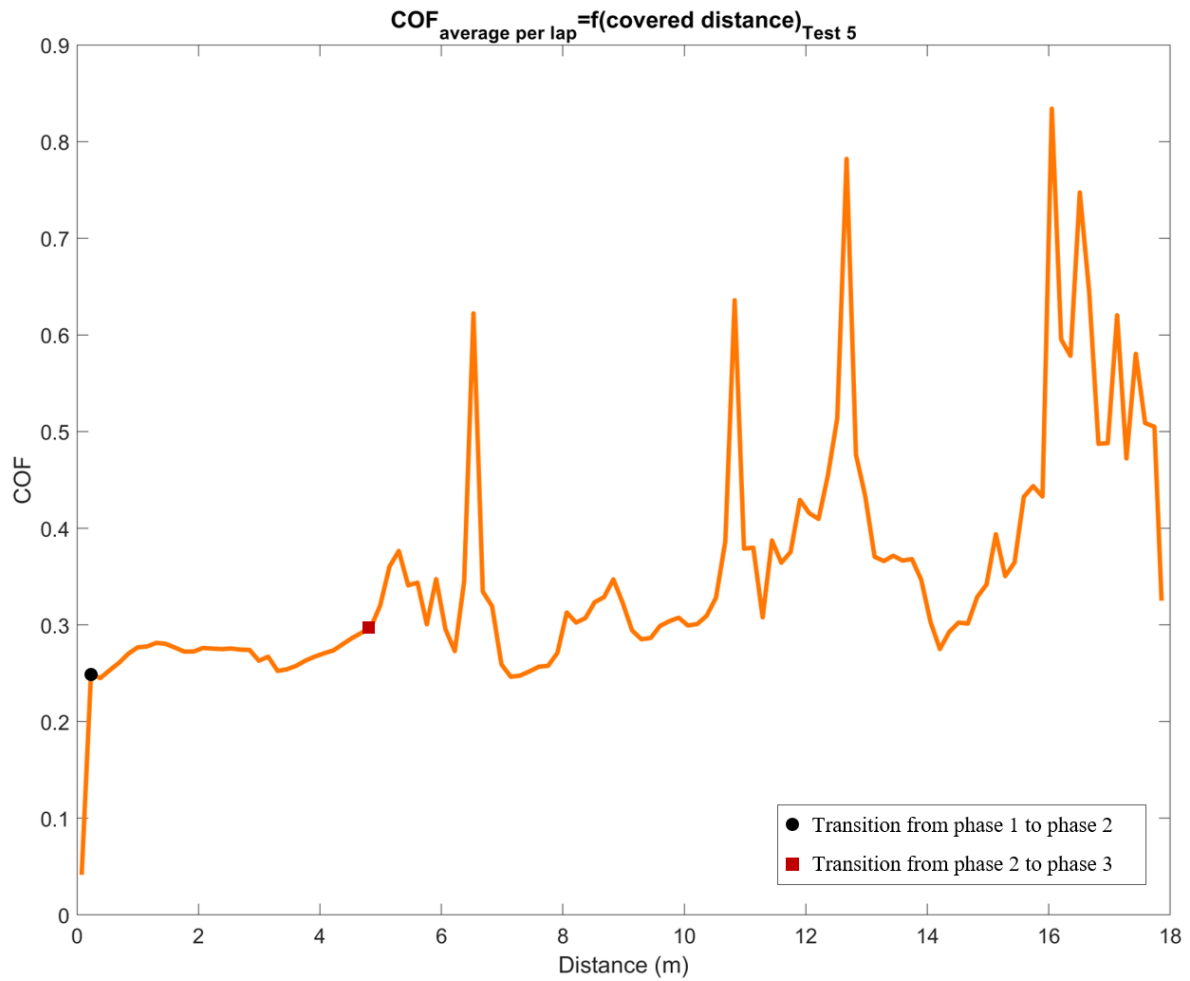
Figure III-1: Evolution of the average per lap of the coefficient of friction during test 5:

The peak phenomenon is interesting since it represents a sudden but temporary increase in friction. In order to understand the origin of these events more deeply, it is necessary to use instantaneous data since **Figure III-1** only presents the average per lap of the COF.

1.1.2.   Instantaneous signals

Figure III-2 shows the evolution of the average per lap of the coefficient of friction along with the instantaneous recordings of the tangential force and the vibration levels of the three Cartesian axes detected on the pin. All the signals are in function of the travelled distance.
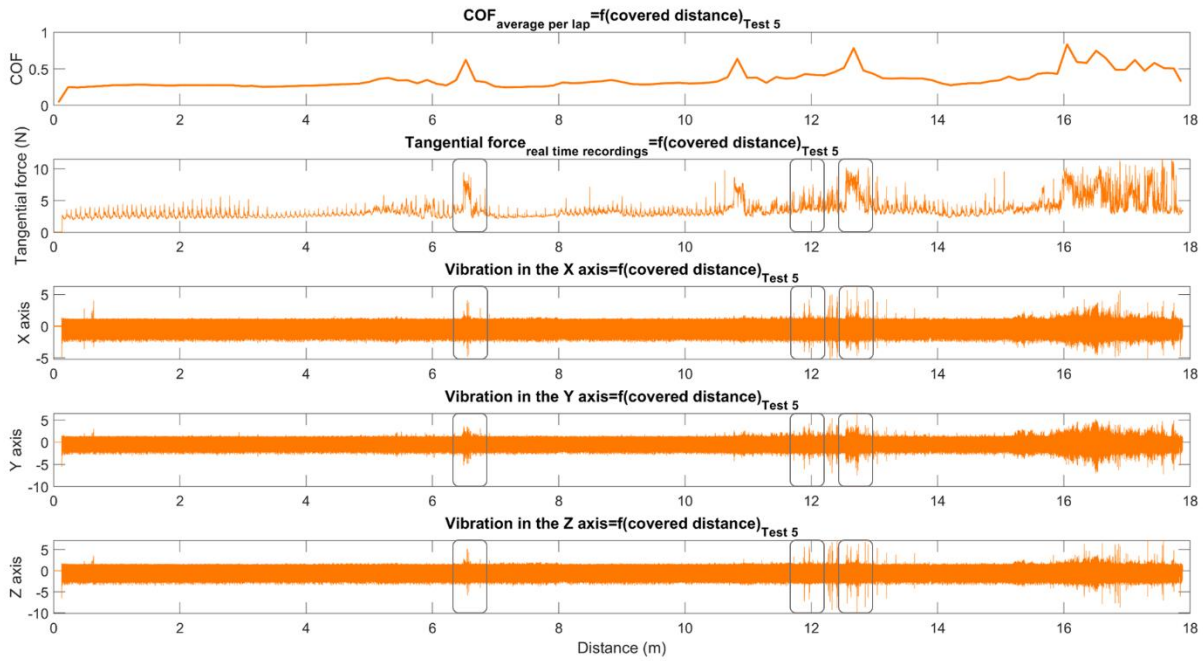
Figure III-2: The evolution of the average per lap of the COF, the tangential force and the vibration levels in the three Cartesian axes recorded during test 5.

Interestingly, the first fact we can identify from the tangential force curve is that there are two types of peaks: the repeatable spikes and the sudden peaks. The repeatable spikes are found during all the duration of the experiment. They may change value but their periodicity does not change. The sudden peaks however happen abruptly and interrupt the repeatable spikes. They are differentiated by their higher values when compared to the repeatable ones. First, we study the sudden peaks. Examination of the vibration levels in the three axes shows that whenever we record a sudden change in the tangential force there is a simultaneous change in the vibration levels. We highlighted these events using the grey rectangles in **Figure III-2**. Therefore, the movement of the pin while those peaks take place does not follow a line (one axis) or a plane (two axes) but it moves in space (three axes). In addition, the peak registered are not momentary but they last in time. Their duration is not a constant and it changes during the test.

To understand the repeatable spikes, we focus on the [3m 4m] interval, as shown in **Figure III-3**. A lap in this test is equal to $0.125\ m$. We can see that the peaks occur periodically (with different intensities). However when we consider the totality of the tangential force, the spikes keep their periodicity but their intensity changes. Different theories are elaborated in order to explain those spikes of which we cite:

- As we saw in the last chapter, the disk's surface initially had a certain texture. However, in **Figure III-4**, we can notice that the disk's surface still has a texture after

the sliding test. This texture is a mixture of the third body layer and the disk's surface itself since we did not clean the latter after the sliding test.

- The belt used to transfer the rotational movement from the motor to the disk is made of rubber and is called a timing belt. It has teeth that match the toothed pulleys (one on each axis of the motor and the disk). This match needs to be perfectly measured and timed in order to maintain the movement. However, since the belt is rubber, it tends to elongate when it starts to wear, which may create a stick-slip like movement.
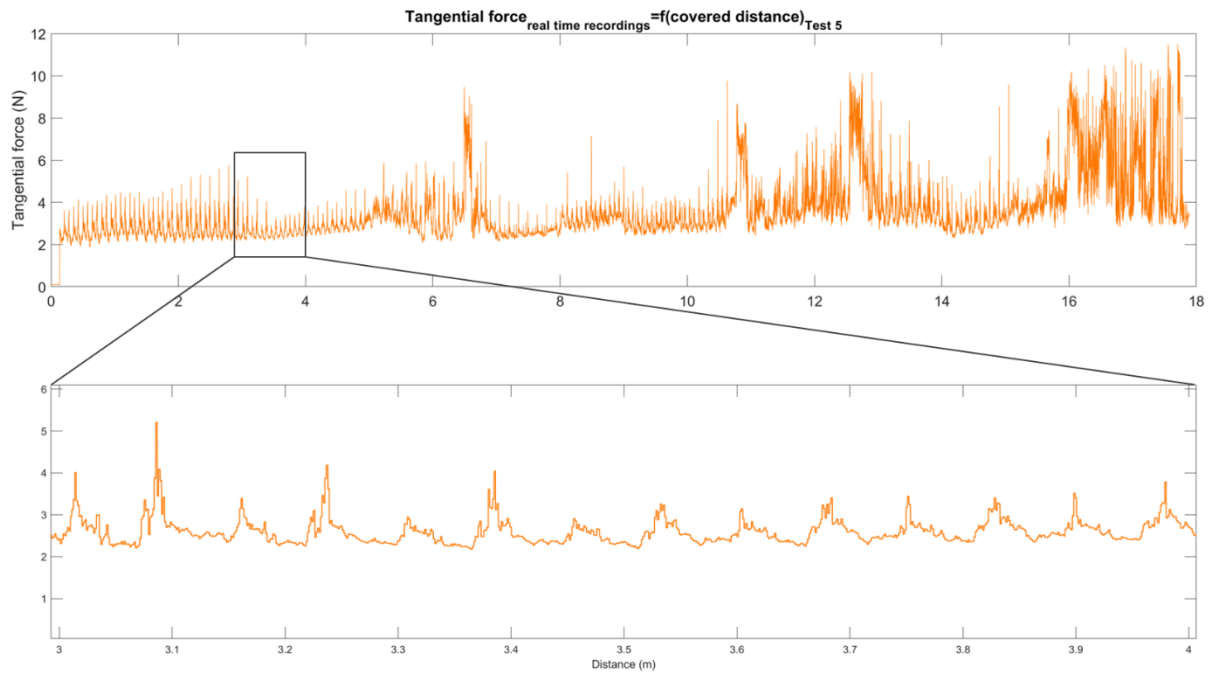
Figure III-3: Isolated section of the tangential force signal (interval [3m 4m]).
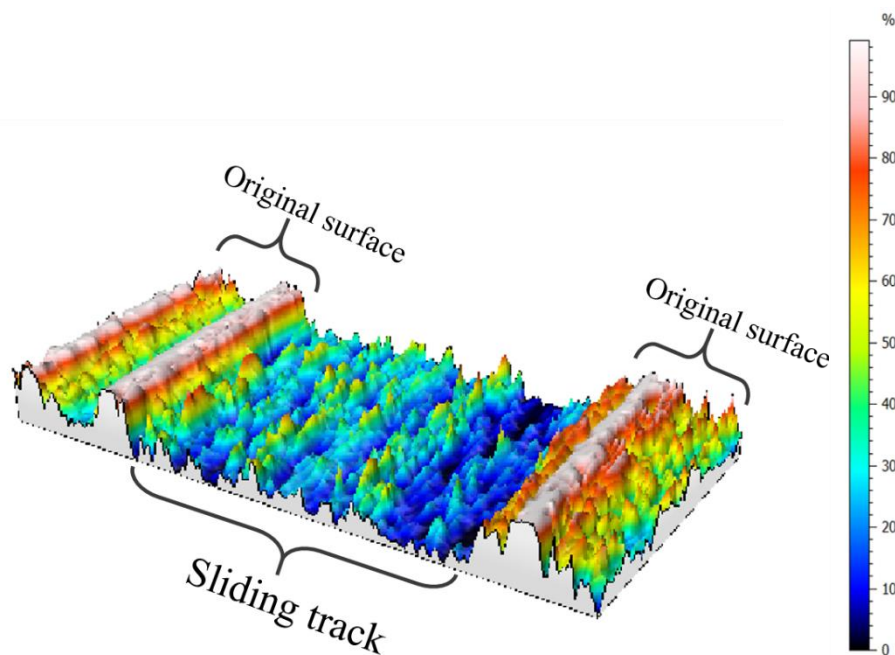
Figure III-4: Disk's state after the sliding test (without cleaning).

1.1.3. *In-situ* observations

During the test 5, we use a camera to record the outlet of the contact, in order to monitor the state of the sliding track and the behaviour of the third body particles. Studying the video recordings shows that the sliding track and the wear debris pass through three different phases:

- *Phase I* (From the start until lap 9): At this observation scale and this stage, no modification on the disk's surface is visible and no third body particle can be seen.
- *Phase II* (From lap 9 until lap 45): The reflective properties of the sliding track change. In this phase the sliding track becomes clearer and wider. Third body particles are now visible but in a small number, and they are only present on the sliding track itself.
- *Phase III* (From lap 45 until the end): The sliding track gets brighter in time and it has a more pigmented aspect, which indicate small third body particles. In addition, their location is no longer limited to the sliding track, but they become visible on its boundaries.

In **Table III-2**, we present a representative frame from each different phase to back up the descriptions we stated earlier. The frames show the same spot on the disk each time to have a correct judgement on the state of the contact. The black tape visible on the pin is necessary to reduce the reflective light on the disk's surface and to make the observation possible. Moreover, in **Figure III-5**, we divide the COF curve in three phases using the *in situ* video records. We can notice that the phases we distinguished when studying the COF (Section 1.1.1. of this Chapter) almost coincide with the *in situ* phases. The error degree is due to the different scales of examination: the COF curve is an intermediate scale (between the microscopic and the macroscopic scales) observation to the movement of the pin (and therefore the contact) while the *in situ* video is a macroscopic investigation of the contact.
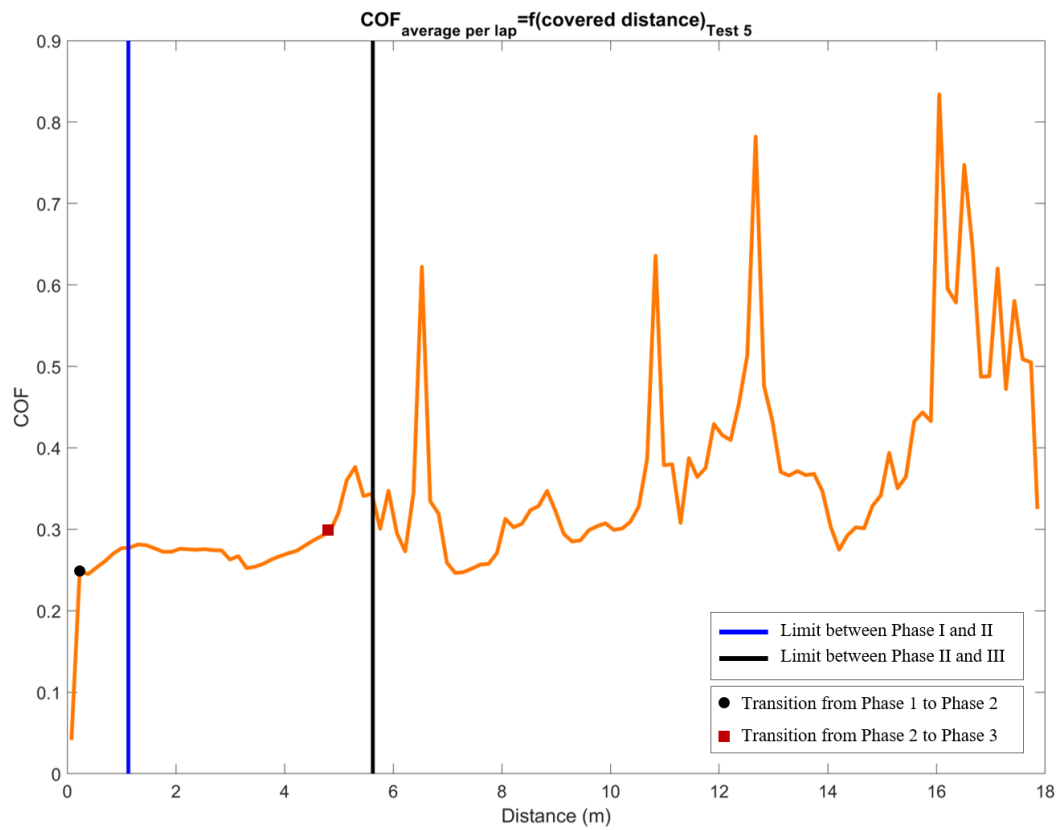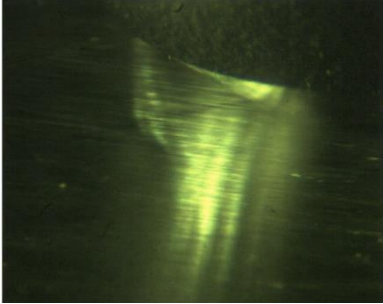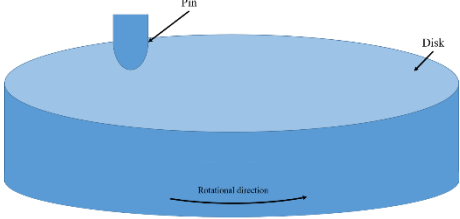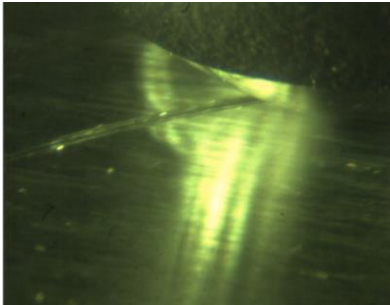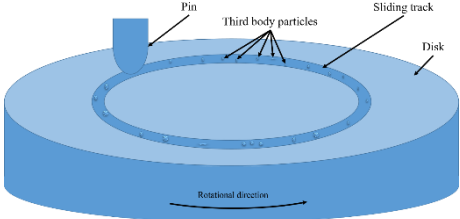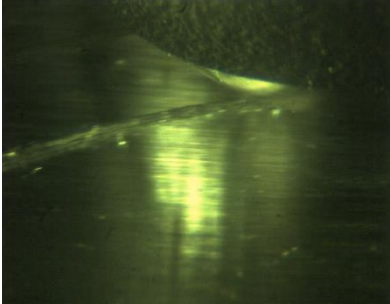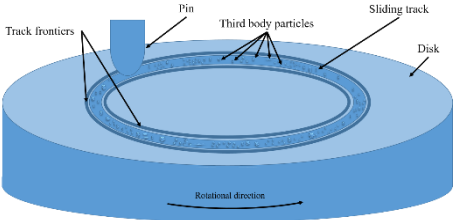
Figure III-5: The evolution of the average of the COF per lap during test 5 in the sliding distance and the different phases detected.

Table III-2: The evolution of the sliding track and the behaviour of the wear debris during test 5.

Test Phase	Test 5 Argon	Sliding track evolution
Phase I	 From the start until cycle 9	
Phase II	 From cycle 9 until cycle 45	
Phase III	 From cycle 45 until the end	

Starting at the lap 103 ($\cong 13\ m$ of sliding), some third body particles appear to become attached or glued to the surface of the pin from the outlet zone (the only side visible in the *in situ* records). The exact explanation for this phenomenon is unclear and we can only speculate for the reasons of this phenomenon. The steel used is a ferromagnetic material. Keeping the two first bodies in contact for a period of time ($\cong 30\ min$ for test 5) may create a small magnetic field. This latter does not have an effect on the contact itself but it could influence the third body particles. This phenomenon could also be related to an electrostatic effect or adhesion-related. However, it does pull some of the third body particles to the pin's surface, as it is visible

in **Figure III-6**. To simplify the discussion, we will refer to this phenomenon as 'the attraction phenomenon' when encountered again.

Figure III-6: The attraction phenomenon during Test 5.

1.1.4. Dynamic rheological measurements

In order to describe the dynamic results of the tests, we chose to use the average and the standard deviation of the COF at the established period. This period is when the contact between the pin and the disk reaches some kind of steady state, and we passed the running-in period. After studying all the conducted tests, we chose to fix the $10\ m$ point as our turning position. Before the $10\ m$, we judge that the contact is still in a running-in state, and after that sliding distance it passes to the steady state period.

Our first rheological descriptors have the following definitions. If we consider our raw data vector S, which has a size equal to N:

- Average per lap of COF signal: $\mu = \frac{1}{N}\sum_{i=1}^{N} S_i$

- Standard deviation of COF signal: $\sigma = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}|S_i - \mu|^2}$

Both descriptors are calculated using MATLAB.

1.2. *Post-mortem* inspection

In order to complete the tribological analysis, we investigate the contact zone on the pin and the sliding track on the disk. When the targeted sliding distance fixed is reached, the test is stopped and the contact is opened by separating the pin and the disk. Both first bodies are observed under the scanning electron microscope without being cleaned.

1.2.1.   The pin

In **Figure III-7**, we can see the SEM observation of the used zone on the pin after the separation of the first bodies. This zone can be divided into three sections: The inlet area, the contact area and the outlet area. When the first third body particles are created (start of the second *in situ* phase), and following the movement direction, the inlet zone is the area where the deposited particles first enter into contact with the pin. The contact area is the zone where the disk and the pin connect. Whereas the outlet area is the last point (for a given lap) where the third body particles exiting the contact area touch the disk. Inside the contact area itself, lighter and darker zones are clearly visible.

In **Figure III-8**, we display the state of the contact area, without cleaning it after the sliding test, on the pin using a profilometer. This observation shows a certain texture that occupies this zone.
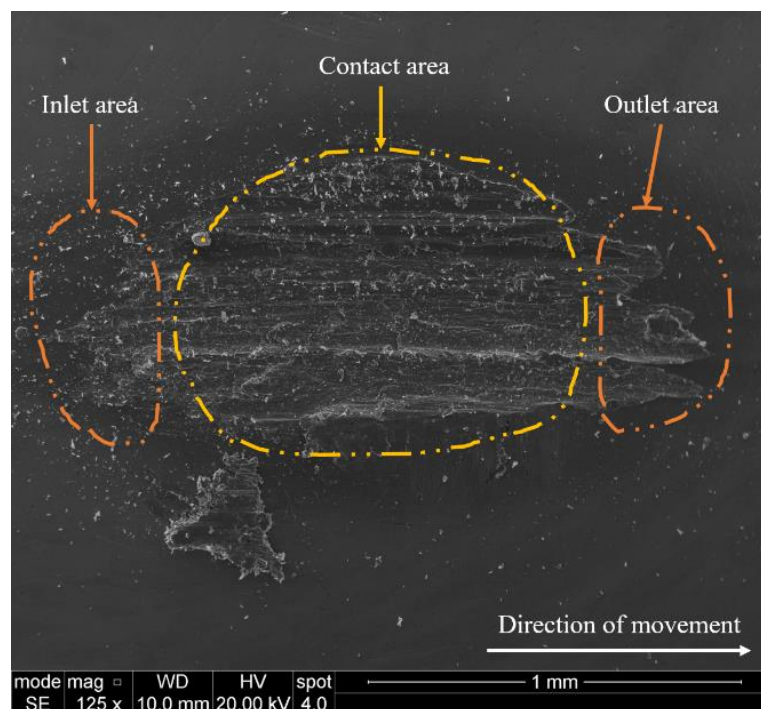
Figure III-7: General SEM view of the used area on the pin after test 5.
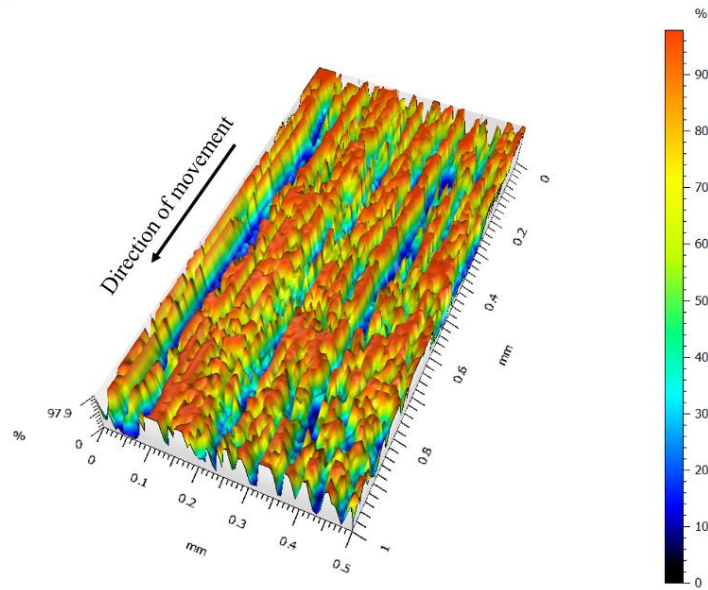
Figure III-8: Representation of the topography of the used area on the pin after test 5.

On the outlet area, we can see the effect of material removal that left the scratched-like marks. Those signatures may be be a result of material adhesion or the flow of third body particles in the contact. Therefore, they can be an indicator of the severity of those flows specifically before exiting the contact.

1.2.2. The disk

We present in **Figure III-9**, the state of the sliding track on the disk at the end of test 5. Just like the pin, we can distinguish three different regions that we can name appropriately when we link this observation with the *in situ* video: the ejection zone, the contact boundary and the contact area. The ejection zone is filled with third body particles that are, as the name suggests it, ejected from the contact and, unless intentionally moved by the user, do not have any contribution to the contact after their ejection. However, the boundaries are filled with particles that may still be used in the contact.
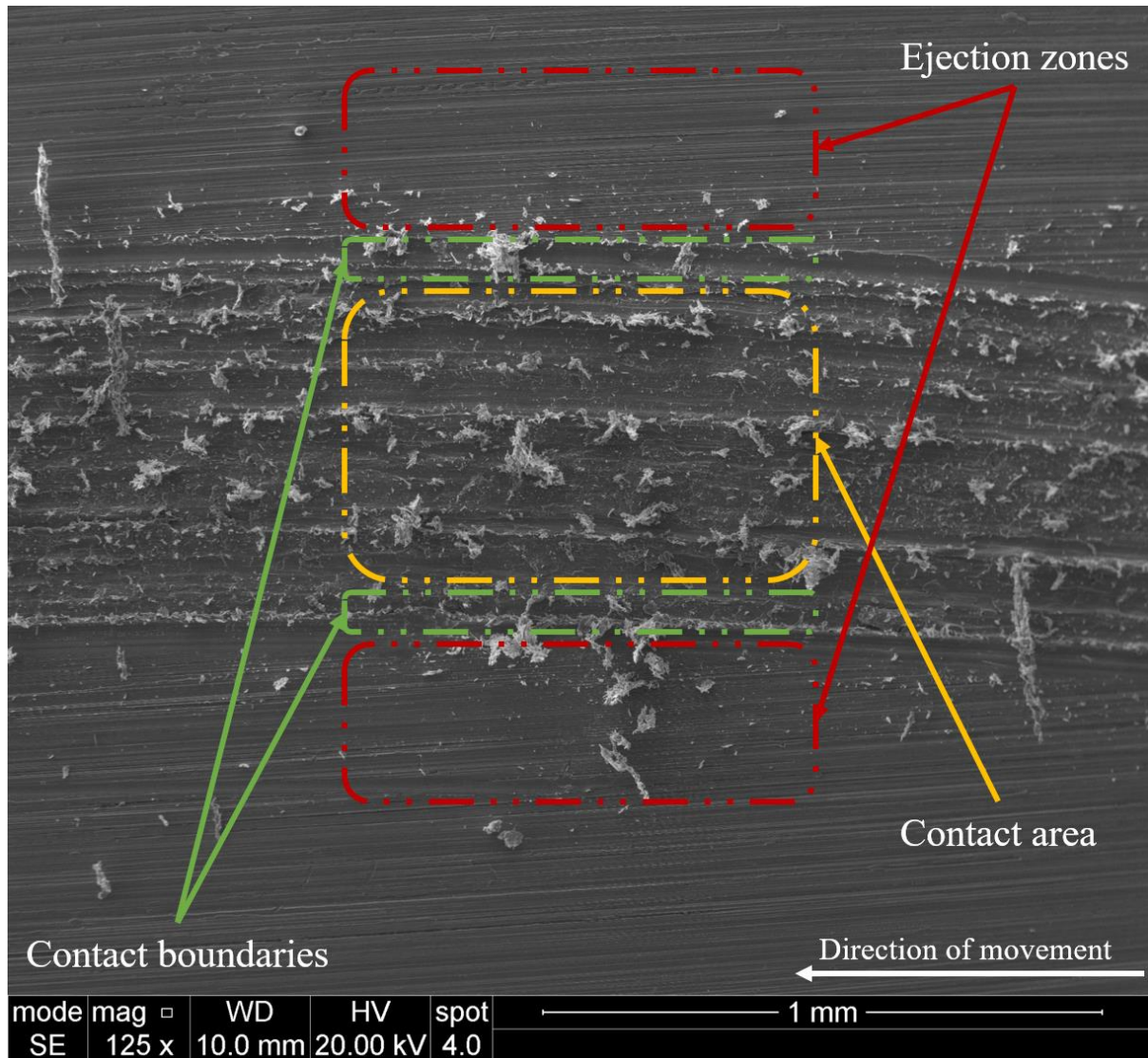
Figure III-9: General SEM of the sliding track on the disk after test 5.

1.2.3. Third body particles

During the sliding test, a layer of third body particles is created, commonly known as wear debris. To complete the study of the contact, this outcome must be investigated. The entire sliding track on the disk and the contact zone (with their different areas) are microscopically observed using the SEM. The analysis in this section is only qualitative.

The theoretical diameter of our point contact according to Hertz's calculations is $0.62 \ mm$ and during the test, it evolves. We chose to work with a maximum of 1500 as a magnification factor to have images containing the largest number of particles while covering the entirety of the sliding zone and the contact area. Choosing a much higher magnification factor $(2500 - 5000)$ would result in getting images with fewer and more defined particles. However, it would need much more images to cover the contact zones on the pin and the disk.

Studying the result from test5, we can see that the third body particles are found in two configurations on the pin: separated 'individual' particles or a cohesive layer. **Figure III-10** displays those two configurations found on the pin. Circled in blue are the cohesive layers found and in yellow some the separated particles. The number of particles found is more significant than the layers that can be detected visually.

Direction of movement
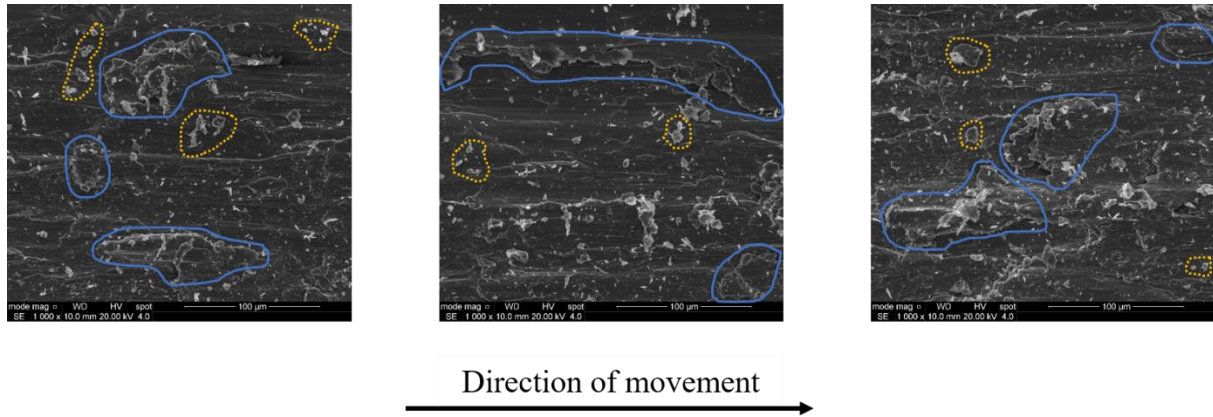
Figure III-10: Third body created during test 5 on the pin.
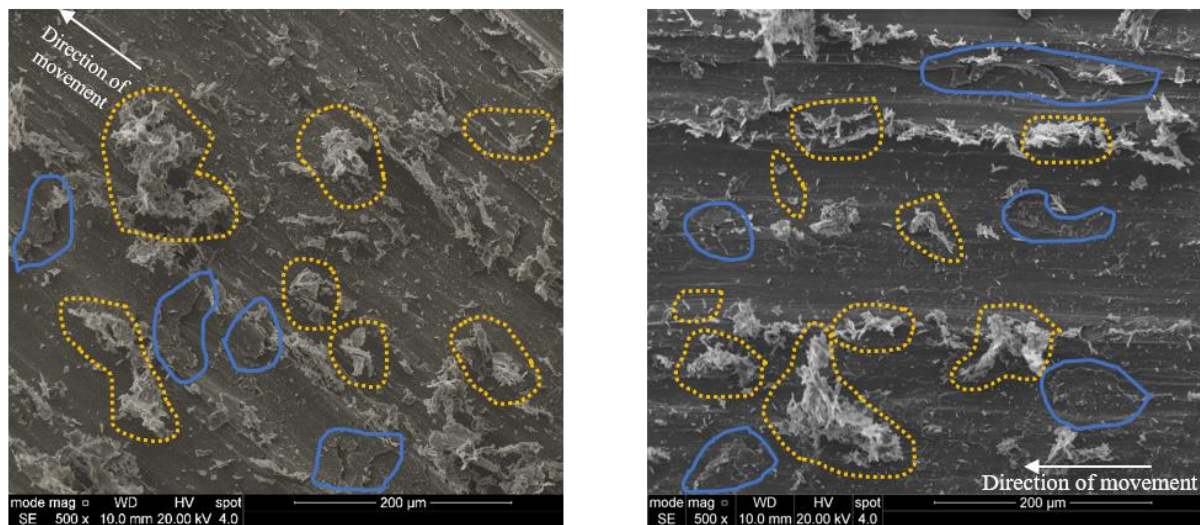
Figure III-11: Third body particles created during test 5 on the disk.

1.2.4. *Post-mortem* rheological measurements

From this section, we chose four descriptors that can help us identify the sliding tests conducted. The descriptors are the following:

- The percentage of the cohesive zone relative to the entire contact area on the pin: The cohesive zones are defined as the zones that appear darker on the contact area on the pin. That change in grey level is due to a pressure change. The darker the zone is, the greater the pressure applied in it is. Therefore, this indicator can inform

us on the spatial distribution of the pressure in the contact. As we can see in **Figure III-12**, we select the dark zone in the contact on the pin and calculate their cumulated surface. Our descriptor is the percentage of the total surface of the cohesive zones relatively to the surface of the contact area.
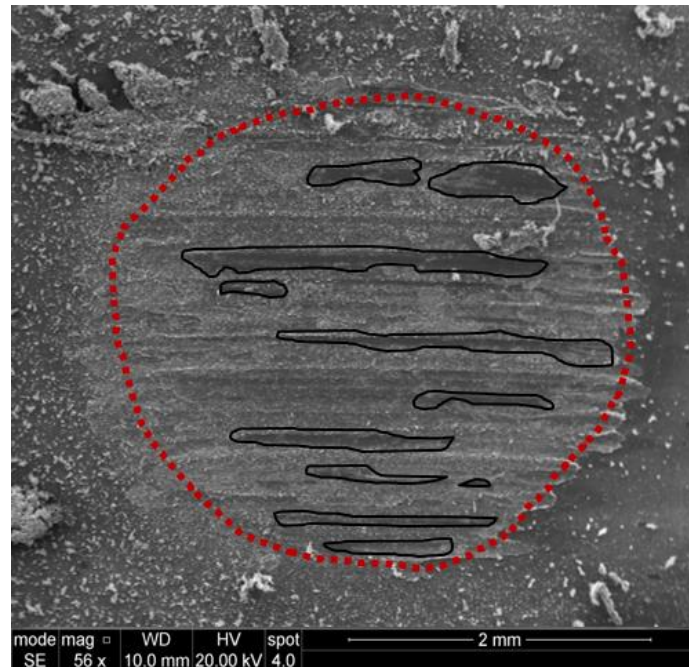
Figure III-12: Selection of the cohesive zones (black continuous outline) in relative to the total surface of the contact area (red dotted outline).

- The intensity of the flow: We judge it mainly using the output area on the pin. After studying the results of different tests, we conclude that there are mainly three flow levels, as visible in **Figure III-13**: Mild (1), Medium (2) and Extreme (3). The different flows are judged visually. We judge the flow on how the exit path of the particles on the contact zone on the pin looks: the more significant the exist marks are the more intense the flow is. Each flow is given the numerical value that is in parenthesis in front of it name.

| Mild flow | Medium flow | Extreme flow |
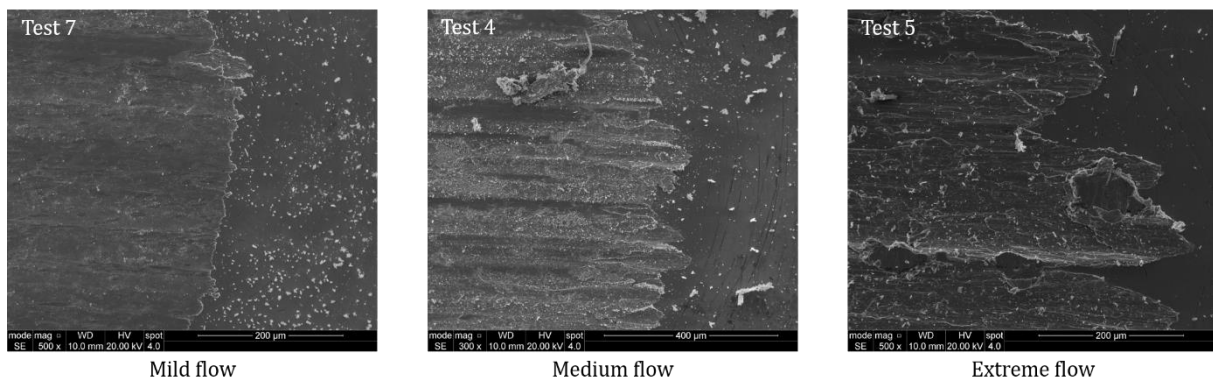
Figure III-13: Different intensity levels of the flow.

- The diameter of the contact area on the pin: Since the contact area resembles a circle, we try to estimate its diameter. We calculate it using a MATLAB algorithm where the user selects the length that he deems correct.
- The width of the contact on the sliding track: This parameter is calculated to check if the contact area on the pin covers all the contact area on the sliding track. It is calculated using the same MATLAB code as the first descriptor.

1.3. Summary

At the end of the examination protocol, each test should provide six rheological measurements that characterize the mechanical-rheological behaviour of that test. The descriptors (measurements) chosen are summarized **Table III-3**. In the next sections of this chapter, we use RD as short for Rheological Descriptor to avoid using the full definition of each descriptor.

Table III-3: Rheological data summary.

Label	Description	Value
RD_1	Average of the friction factor in established state (After the $10\ m$ mark).	$\mu = \dfrac{1}{N}\sum_{i=1}^{N} S_i$
RD_2	Standard deviation of the real time friction factor signal in established state (After the $10\ m$ mark).	$\sigma = \sqrt{\dfrac{1}{N-1}\sum_{i=1}^{N}|S_i - \mu|^2}$
RD_3	The percentage of the cohesive zone relatively to the entire contact area on the pin	$\in [0,100]$
RD_4	The intensity of the flow	$\in \{1; 2; 3\}$
RD_5	The diameter of the contact area on the pin	$\mathbb{R}^+$: positive real
RD_6	The width of the contact on the sliding track	

2. Set 1 analysis

Table III-4 summarizes the conditions of the tests in set 1, where the varying parameter is the linear speed. The modification of the linear speed is implemented by changing the radius of the sliding track and not by changing the rotational speed. In fact, this parameter is fixed

between the three tests at $6\ laps\ per\ minute$. Using this method, we do not change the vibration errors captured by the sensors originating from the motor and/or the transferring belt.

Table III-4: Set 1 conditions.

Tag	Radius (m)	Normal force (N)	Linear speed (m/s)	Temperature (°C)	Humidity (%)	Gaseous environment	Distance covered (m)
Test 1	0,02		**0,75**	22	51		≅ 25
Test 2	0,015	10	**0,56**	20	49	Air	≅ 24
Test 3	0,01		**0,38**	23	50		≅ 25

2.1. Dynamic results

Figure III-14 shows the evolution of the average per lap of the coefficient of friction (COF) during the tests in set 1. Each curve presents three different phases starting and finishing at different points:

- *Phase 1(before the block dot)*: The COF starts at a near zero value and increases swiftly until it reaches a maximum value before it decreases for the first time.
- *Phase 2 (between the black dot and the red square)*: The COF evolution is characterized by the occurrence of significant peaks.
- *Phase 3 (after the red square)*: The COF achieves a more or less constant value. Small peaks (in comparison with those of *Phase 2*) are detected.

Numerous differences between the coefficient of friction from the three tests are noticed from studying **Figure III-14**:

- The length of each phase detected varies between the tests. The phases in the curve from test 2 are the longest, followed by those from test 1 and finally those ones from test 3. The duration of the phases does not seem to be linearly correlated to the linear speed of sliding
- At the end of *Phase 1*, the coefficient of friction is almost equal between test 1 and test 3. However, the COF from test 2 is much larger.
- The peaks that the COF witnessed during *Phase 2* are different between the three tests in both their values and their frequency of occurrence.
- After establishing an almost constant COF in *Phase 3*, the test 2 and test 3 reach similar values. However, the value from test 1 is different. This difference can be explained by the duration difference between the tests. Since we use different speeds

to achieve almost the same covered distance, the duration changes between the tests (the fastest speeds corresponds to the shortest test timewise). However, we find no direct (linear) correlation between the difference of the final value of the COF and the linear speed chosen.
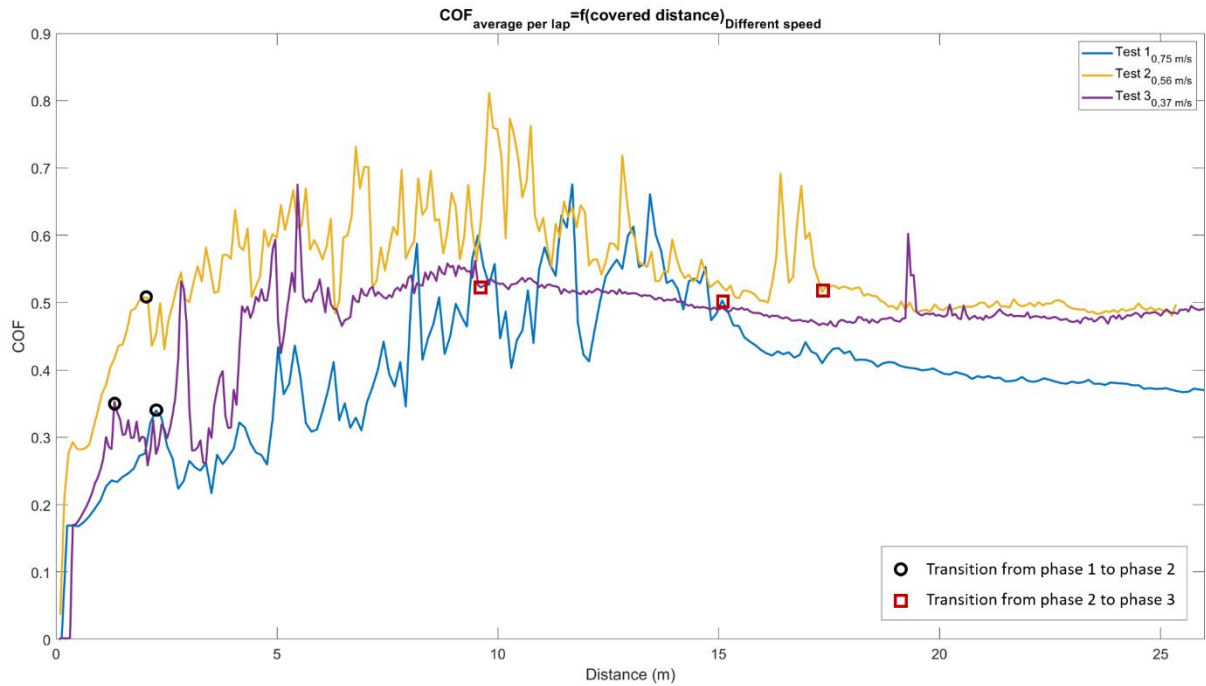
Figure III-14: The evolution of the friction coefficient average per lap for the three different tests of set 1.

The accelerometer was installed only in test 1 and test 2. The instantaneous evolution of the tangential force and the vibration levels following the three Cartesian axes are displayed in **Figure III-15**. We can see that at each instantaneous peak we register in the tangential force, we register a peak in the vibration levels as well. This means that the peaks in the tangential force are a direct result of a movement of the pin and not some external event. In addition, after the peak the tangential force does not go directly back to its level before the peak but it still varies. We also notice that before entering *Phase 3*, the vibration levels become more important than they were before, and then stay at that level.

Unfortunately, the *in situ* visualisations for the three tests were lost due to a technical fault with the registering devices. Therefore, we lack their analysis section.
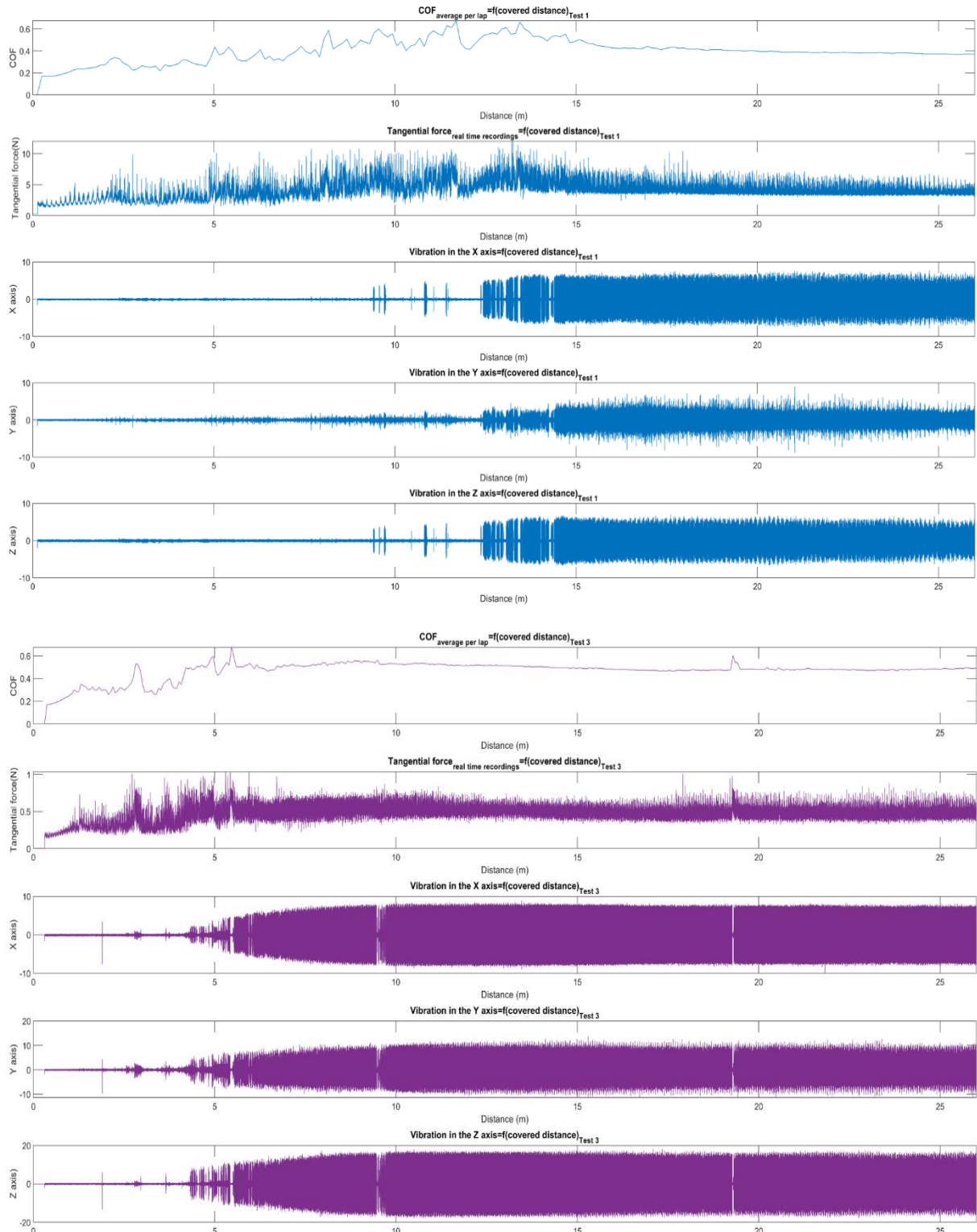
Figure III-15: The average per lap of the COF and the instantaneous recordings of the tangential force and the vibration levels in the three Cartesian axis recorded during test 1 (Blue on top) and test 3 (Purple on the bottom).

2.2. *Post mortem* investigations

In **Table III-5**, we present the images of the used areas on the pin and the sliding track on the disk at the end of each test and after opening the contact.

2.2.1. Pin

The used section of the pins from all the tests can be divided into 4 different zones: the inlet area, the contact area, the outlet area and the third body portal. The latter section seems to have the same visual properties (dense and compact) in the three tests with a slight difference in size and it is not present in all the tests, which explains its absence from the example treated in the analysis protocol section. During the sliding tests, third body particles are created from the contact between first bodies. However, those particles leave the contact at a certain point to move on the turning sliding track. Eventually, and since there is no obstacle that will make them leave the track, the particles meet the pin in the inlet area and try to re-enter the contact. During the contact, the third body particles start assembling creating a very large agglomerate, just before the inlet area, that we name third body portal. The size difference between the three portal sections is not necessarily a property of the contact. If the test conductor is not careful when removing the pin from its supporting arm, the shaking movement could result into parts of the portal breaking apart and parts of it falling from the pin's surface on the disk's surface.

In order to determine the third body flows as defined in the analysis protocol earlier (page 61), we study the outlet zone on the pin. In **Table III-6**, we represent this section for the three tests in set 1. In test 2, we judge that the flow that took place cannot be considered as mild nor extreme when compared to the examples of each type given earlier. Therefore, it is considered as a medium. However to determine the flow type in test 1 and test 3, we turn to the topography images, in **Figure III-16**. We can see that in the test 3, the marks (highlighted in **Figure III-16** with red arrows) look deeper than those from test 1. We judge that the flow in test 1 is medium and the flow in test 3 is extreme.

Table III-5: The state of the used area of the pin and the sliding track on the disk after the tests in set 1.

	Test 1	Test 2	Test 3
Outlet area			
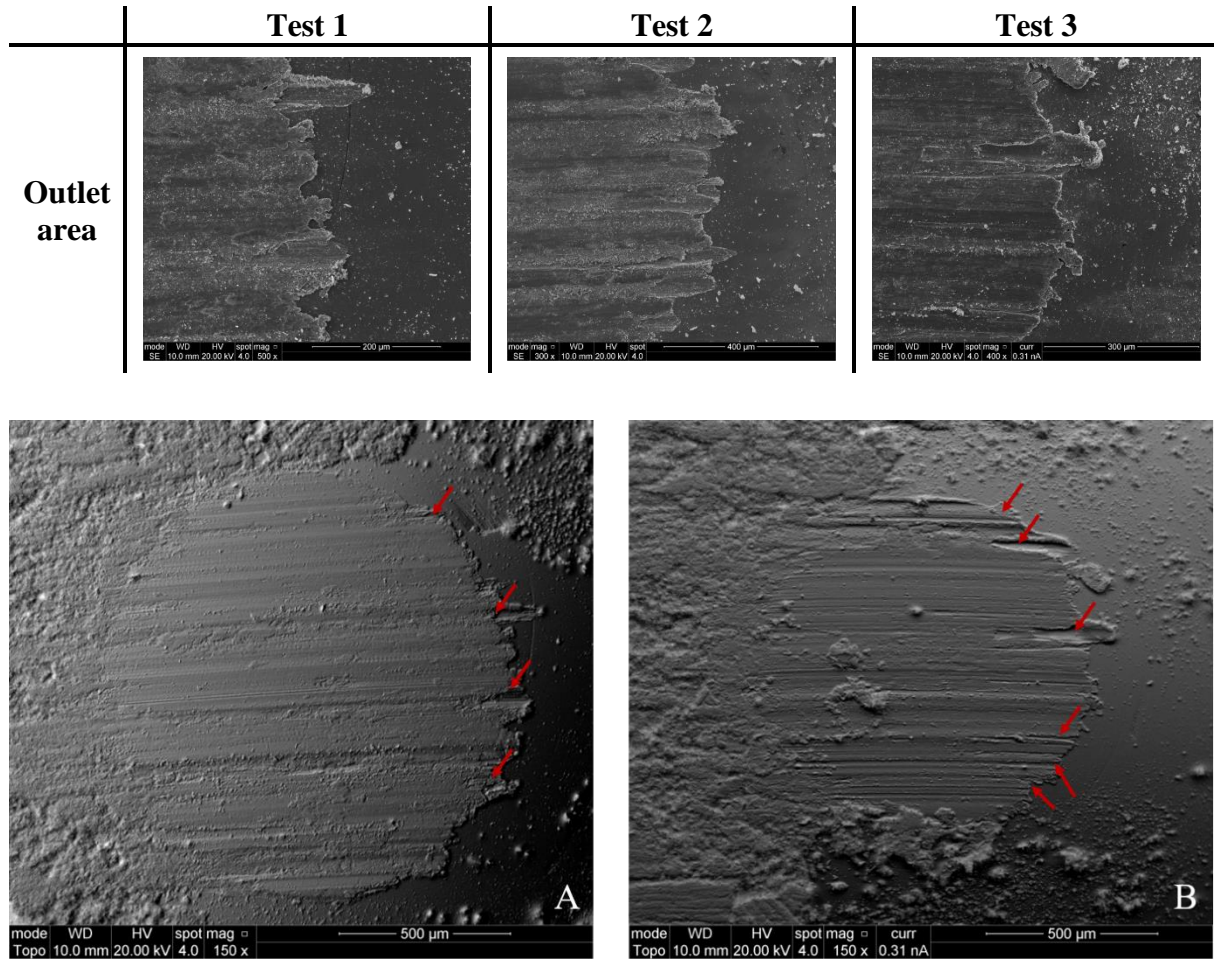
Figure III-16: Topography (mode BSE in the MEB) images of the used area from: A) Test 1. B) Test 3.

2.2.2. Disk

The three sliding tracks can be divided into 3 different areas: the ejection zones, the contact boundaries and the contact area. Between the three tests, the ejection area and the contact boundary are full of separated particles and much smaller particles fill the contact area. The blocks of particles seen in the upper ejection area of the sliding track resulting from test 2 have most likely fallen from the third body portal during the contact opening procedure or during the test itself.

In **Figure III-17**, we see the topography images of the contact boundaries and area of the sliding track from each test in set 1. Studying those images, we find a texture that is characterized by parallel curved lines (highlighted in yellow) all along the contact area and its boundaries. Those should not be mistaken with the lines that exist in the ejection zone that are highlighted in red.
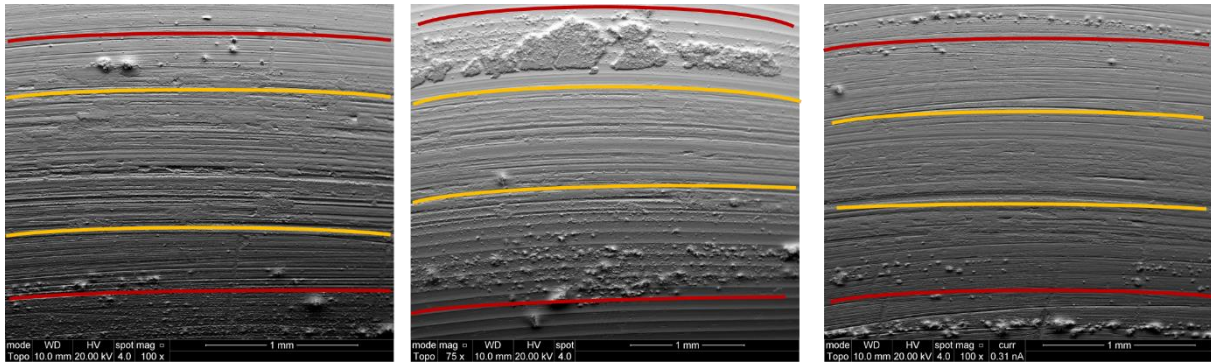
Figure III-17: Topography state of the sliding track after the tests in set 1.

When analysing the pin topography from test 3 in **Figure III-18**, it appears that it presents a particular texture (section highlighted in blue). The sliding track's topography shows a similar (complementary) texture in the spot where the pin and disk connect.
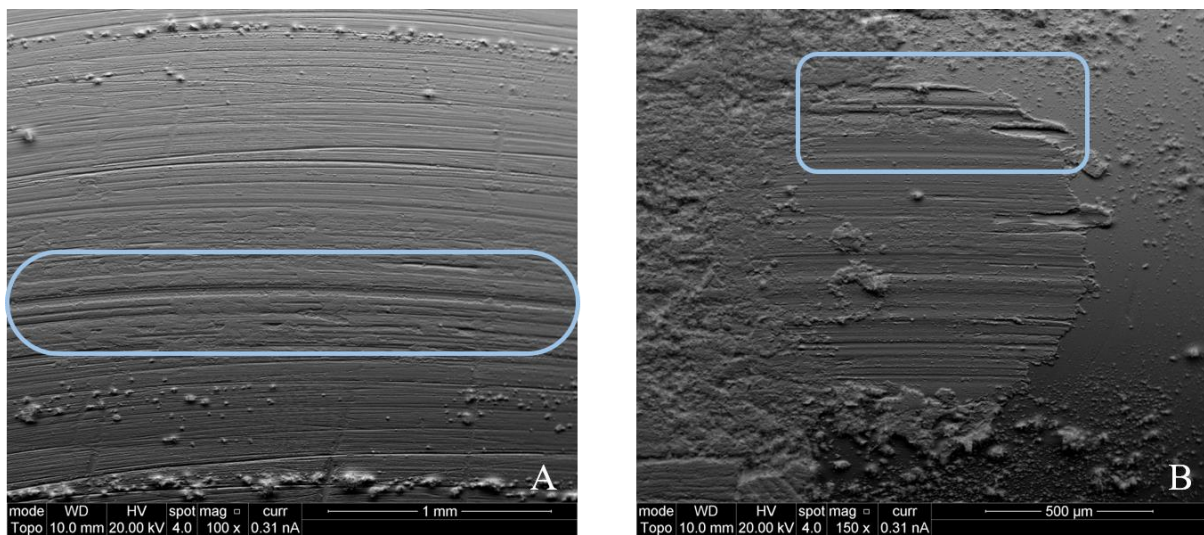
Figure III-18: Relating the pin's and the disk's contacting areas from test 3.

Examining the used area on the pin and the sliding track on the disk, we can propose the following scenario to explain the various peaks we observe in **Figure III-14**:

1) Third body particles are stuck in the inlet of the contact during the sliding test until a certain a threshold (in relation with the maximum supported tangential force before the pin's arm is moved) is reached, which explains the creation of the third body portal observed on the pins after each test.

2) Certain parts of the third body particles pass through the contact area. This event results in the peaks we observe in the tangential force recordings and the change in the vibration levels in all the three directions because this mass moves the pin.

3) The bulk of the particles that passed change the topography of the contact area shortly, which justifies the perturbations we see in the tangential force after the peak is recorded.

4) After a certain number of laps, the third body particles either are ejected or stuck again in the inlet of the contact. This explains how the tangential force goes back to the same pattern similar to that before the peak.

2.2.3. Third body particles

We cannot study the wear debris created during the sliding tests from the microscopic views covering the totality of the used areas on the pin and the disk. We need to magnify more but still respect the 1500 limit that allows us to stay in the macroscopic scale. At this point, the images are inspected visually.

In **Table III-7**, we display the variety of images of third body particles found on the pin and the disk from the different zones determined earlier. We distinguish two types of findings: separated particles/agglomerates (yellow dotted outline) and cohesive layers (blue continuous outline). It is the same observation in all the tests from set 1, which either means that the linear speed at which the disk turns has no effect on the creation of the third body or that the investigation method is not sophisticated enough.

The found cohesive layers found seem to have an elongated form following the direction of the slip, as we can see in **Figure III-19**. However, they also look crushed in different directions indicating complex accommodation regimes. The separated particles are created when parts of the cohesive layer reach they elasticity limit and break. **Figure III-19** illustrates this theory by highlighting the breaking phenomenon.

From this figure and knowing the direction of rotation, we can interpret the different observations. The yellow arrows show the direction that the spreading of this layer seems to be following. The green sections are where we notice that parts of the layer start to break. We can see that such events can be of two types: very small particles (with a powder-like aspect at this scale) and much larger fragments.
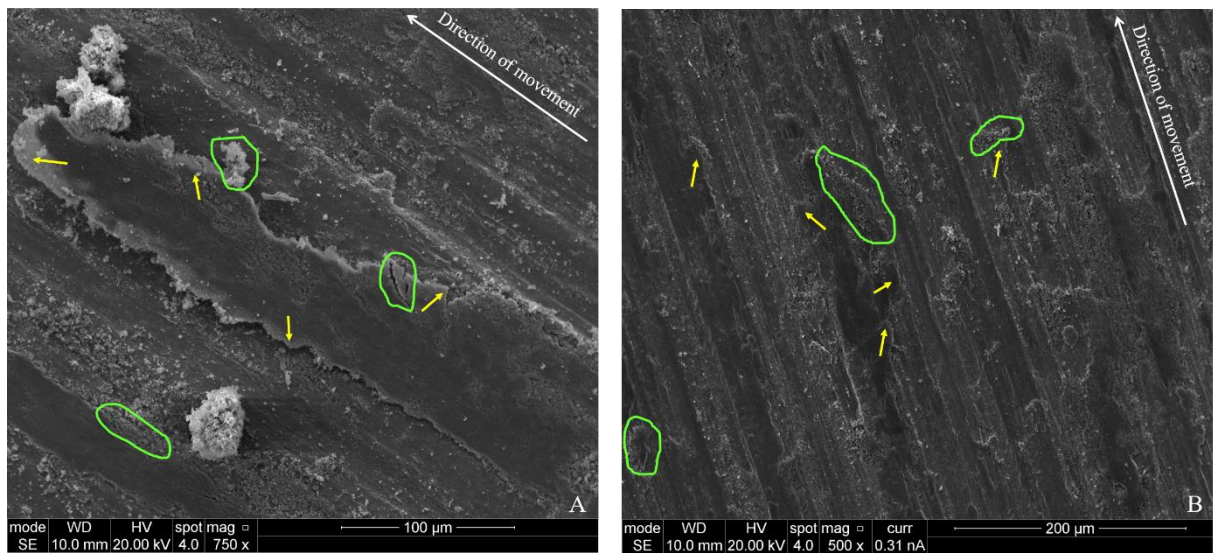
Figure III-19: The state of the cohesive layer in A) Test 1. B) Test 3.

Table III-7: The state of the third body particles produced during the tests from set 1 (separated particles in yellow and cohesive layer in bleu).

	Contact area on the pin	Sliding track on the disk
Test 1 0.75 m/s		
Test 2 0.56 m/s		
Test 3 0.38 m/s		

2.3. Set 1 rheological measurements

Table III-8 summarizes the different rheological measurements describing the test from set 1, where we changed the linear speed by decreasing it from test 1 to test 3.

Table III-8: Set 1 rheological data (defined page 60).

	Test 1 $_{0.75 \text{ m/s}}$	**Test 2 $_{0.56 \text{ m/s}}$**	**Test 3 $_{0.38 \text{ m/s}}$**
RD_1	0.44	0.54	0.49
RD_2	0.11	0.09	0.05
RD_3 (%)	16	18	36
RD_4	2	2	3
RD_5 (mm)	1.22	1.25	0.94
RD_6 (mm)	1	1.1	0.84

3. Set 2 analysis

The experimental parameters of the three tests conducted under set 2 are summarized in **Table III-9**. Studying the effect of the gaseous environment was the main interest in this set. Test 4 was conducted in air while Test 5 and Test 6 were conducted in argon with the help of the gas enclosure defined earlier. The other controllable parameters are fixed between the three tests in order to associate any resulting variation to the gas used.

Table III-9: Set 2 conditions.

Tag	Radius (m)	Normal force (N)	Linear speed (m/s)	Temperature (°C)	Humidity (%)	Gaseous environment	Distance covered (m)
Test 4				22	44	**Air**	
Test 5	0,02	10	0,63	20	50	**Argon**	$\cong 18$
Test 6				23	50	**Argon**	

3.1. Dynamic results

Following the analysis protocol, we study first the evolution of the COF curves shown in **Figure III-20**. Each curve is the evolution of the average per lap of the coefficient of friction in distance. We can distinguish 3 different phases in each curve:

- *Phase 1(before the black dots)*: The COF increases quickly from zero to a value at which the raise stops.
- *Phase 2(between the black dots and the red squares)*: The COF keeps increasing but at a very low rate.
- *Phase 3(after the red squares)*: Noticeable peaks appear in the evolution of the COF. The value of the COF keeps changing after each peak.

From this figure, we can detect many differences between the COF results from the tests conducted in Argon and the test conducted in Air:

- The lengths of the phases detected in tests 5 and 6 are almost equal. While the length of the phases detected in test 4 are different.
- The value of the COF at the end of *phase 1* is different between the three tests.
- During *phase 2*, test 4 demonstrates small peaks. However, the COF's value goes back to its previous value after passing each peak. The COF from test 5 and test 6 do not show any peaks during that phase, but keep a more or less constant value.
- In *phase 3*, the curves from test 5 and test 6 record very important peaks until the end of each test. In test 4, the evolution of the COF passes by peaks as well but they are less significant than those registered in the other tests. In addition, at the end of test 4 the interface seems to have reached a steady state.
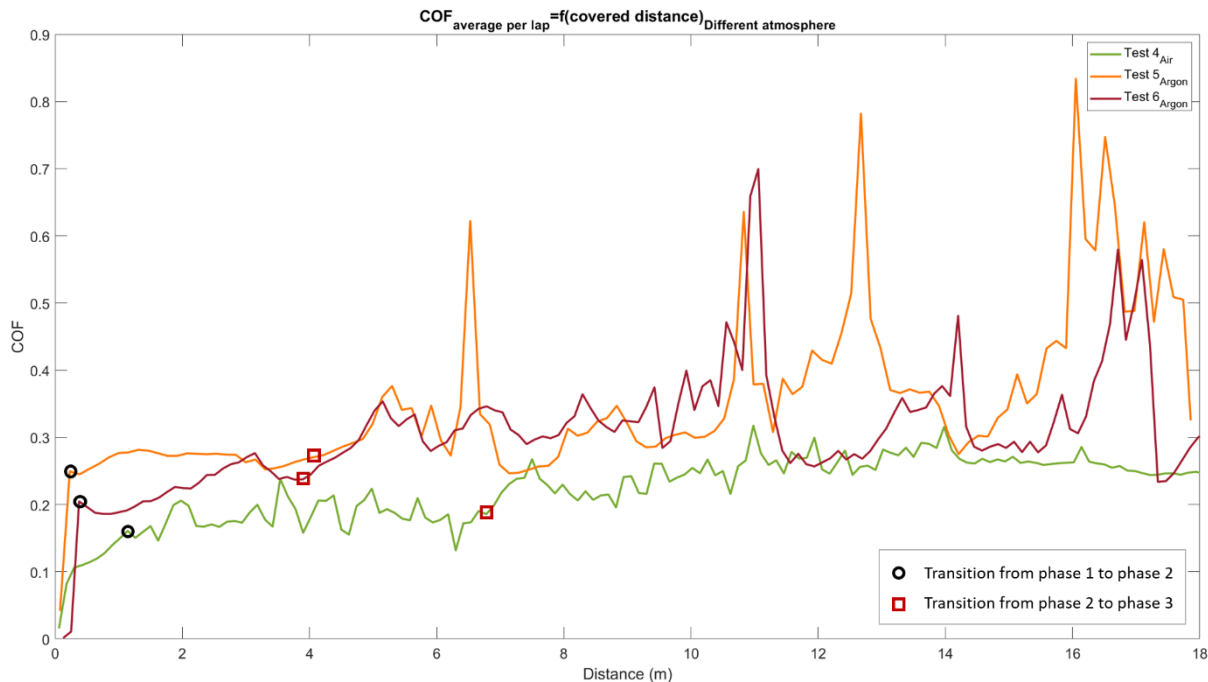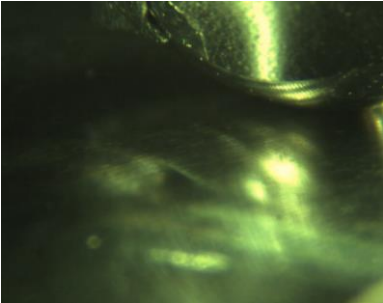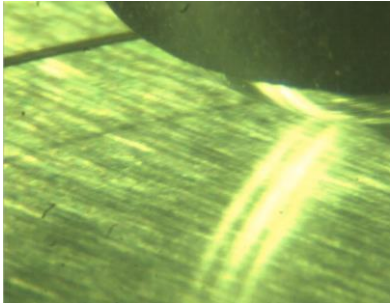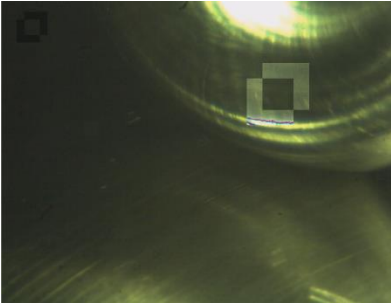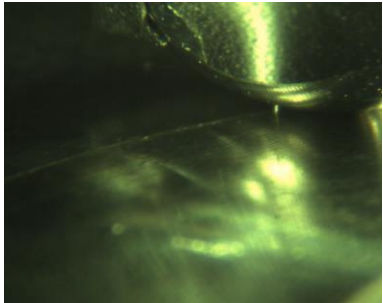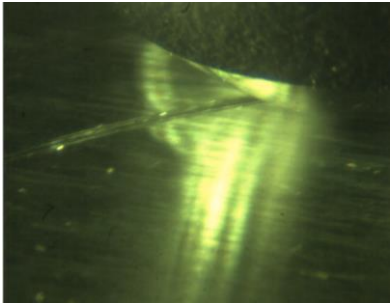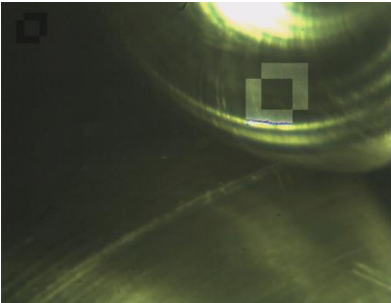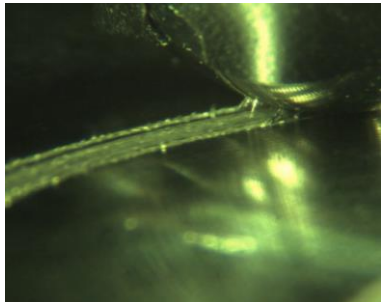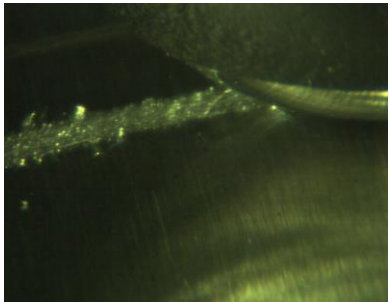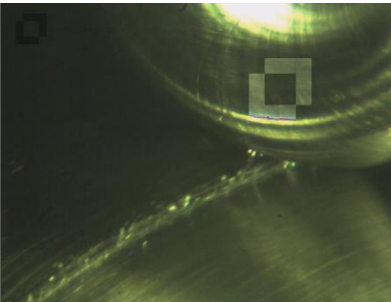
Figure III-20: The evolution of the friction coefficient average per lap for the three different tests of set 2.

At this point, we can conclude that, at least when studying the COF curves, our materials behave differently in argon than in air.

During the three tests, we record the outlet of the contact as described in the last chapter. Studying the resulting films proves that the three tests pass through the same 3 phases that were detected when studying the test 5 (used as an explanatory example for the analysis protocol). In **Table III-10**, we show different frames to highlight each phase. The frame is always of the same spot in order to be able to compare the changes between the three phases. Using just the *in situ* visualisation, we notice differences between the two atmospheres. During the test in air, we can detect that the third body particles are not as big as in argon. In addition, the ejection zone on the disk is more established (distinguishable from the sliding track) in air.

Table III-10: The evolution of the sliding track during the three tests in set 2.

Test / Phase	Test 4 Air	Test 5 Argon	Test 6 Argon
Phase I	From the start until cycle 9	From the start until cycle 9	From the start until cycle 39
Phase II	From cycle 9 until cycle 27	From cycle 9 until cycle 45	From cycle 39 until cycle 79
Phase III	From cycle 27 until the end	From cycle 45 until the end	From cycle 79 until the end

3.2. *Post mortem* investigations

After each test, we separate the first bodies and observe them under the SEM. **Table III-11** presents the used areas on the pins and the disks after the tests in set 2.

Table III-11: The state of the used area of the pin and the sliding track on the disk after the tests in set 2.

3.2.1. Pin

Focusing on the used areas on the pins, we can see that the three pins can be divided just like the pins from set 1. Essentially, we find three main shared zones: the inlet area, the contact area and the outlet area. However, we only find the third body portal in test 4 and test 6. In addition, in test 6, we can detect a new zone located after the outlet area that we call 'Third body escape'. This zone is the same zone we observe in the *in situ* visualisations after a certain number of laps due to the attraction phenomenon. It is seen in every tests of set 2, as proven in **Figure III-21**.

Third body portals are observed in tests 4 and 6 but they do not share the same visual characteristics, as appears in **Figure III-22**. The portal in test 4 (**Figure III-22.A**) looks denser and more compact than that from test 6 (**Figure III-22.B**). The particles that form the portal zone in test 6 have a fibre-like look. The absence of the portal zone in the pin resulting from test 5 is a result of accidentally shaking violently the pin when removing it from its fixing arm.

Starting from cycle 100

Starting from cycle 103

Starting from cycle 84
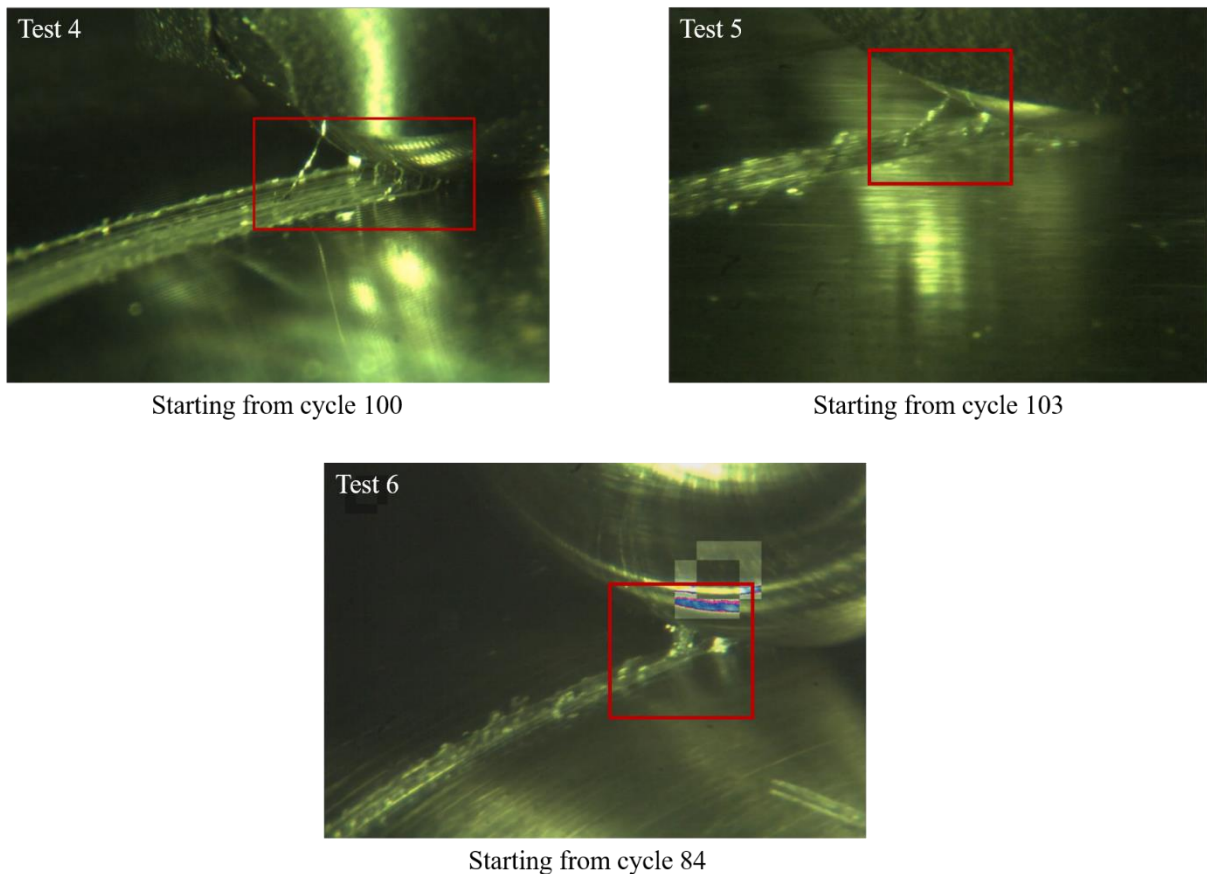
Figure III-21: The attraction phenomenon during the tests in set 2.
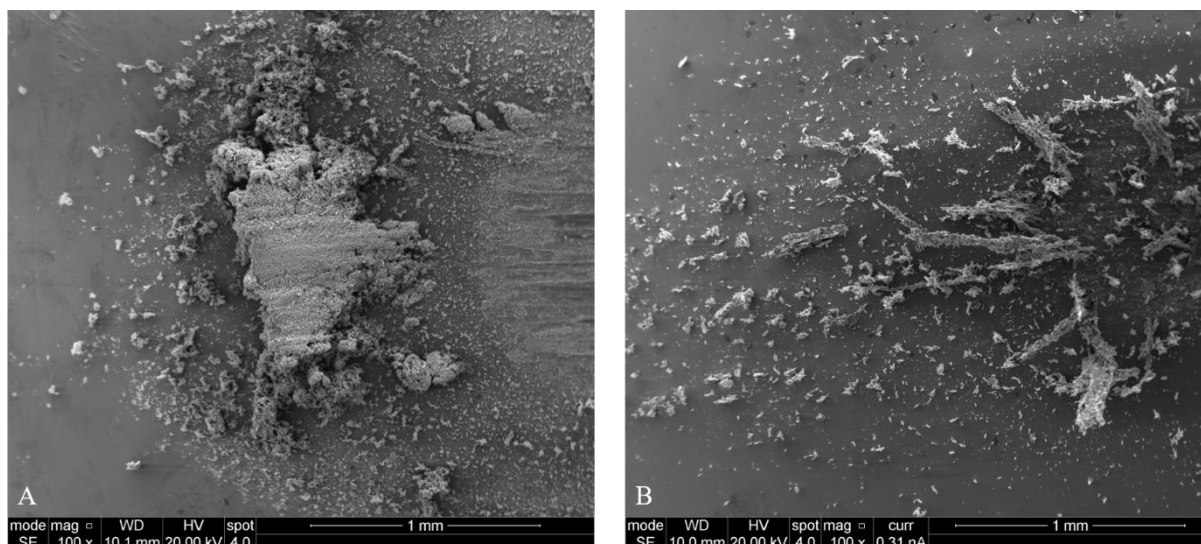
Figure III-22: The state of the third body portal from: A) Test 4. B) Test 6.

To evaluate the flow that took place in each test, we study the outlet area as defined in the analysis protocol. The three areas are shown in **Table III-12**. We can conclude that the flow in test 4 was medium while in test 5 it was extreme. However, it is necessary to use the topography image of the used area in the pin from test 6 in order to make a relevant assessment, as shown in **Figure III-23**. This latter reveals that the visible scares (pointed at by the red arrows) on the outlet area are deep and significant. Therefore, we assign the extreme attribute to the flow in test 6.

Table III-12: The outlet area state from the tests in set 2.

	Test 4 Air	Test 5 Argon	Test 6 Argon
Outlet area			

mode | WD | HV | spot | mag □ | curr | 1 mm
Topo | 10.0 mm | 20.00 kV | 4.0 | 100 x | 0.31 nA
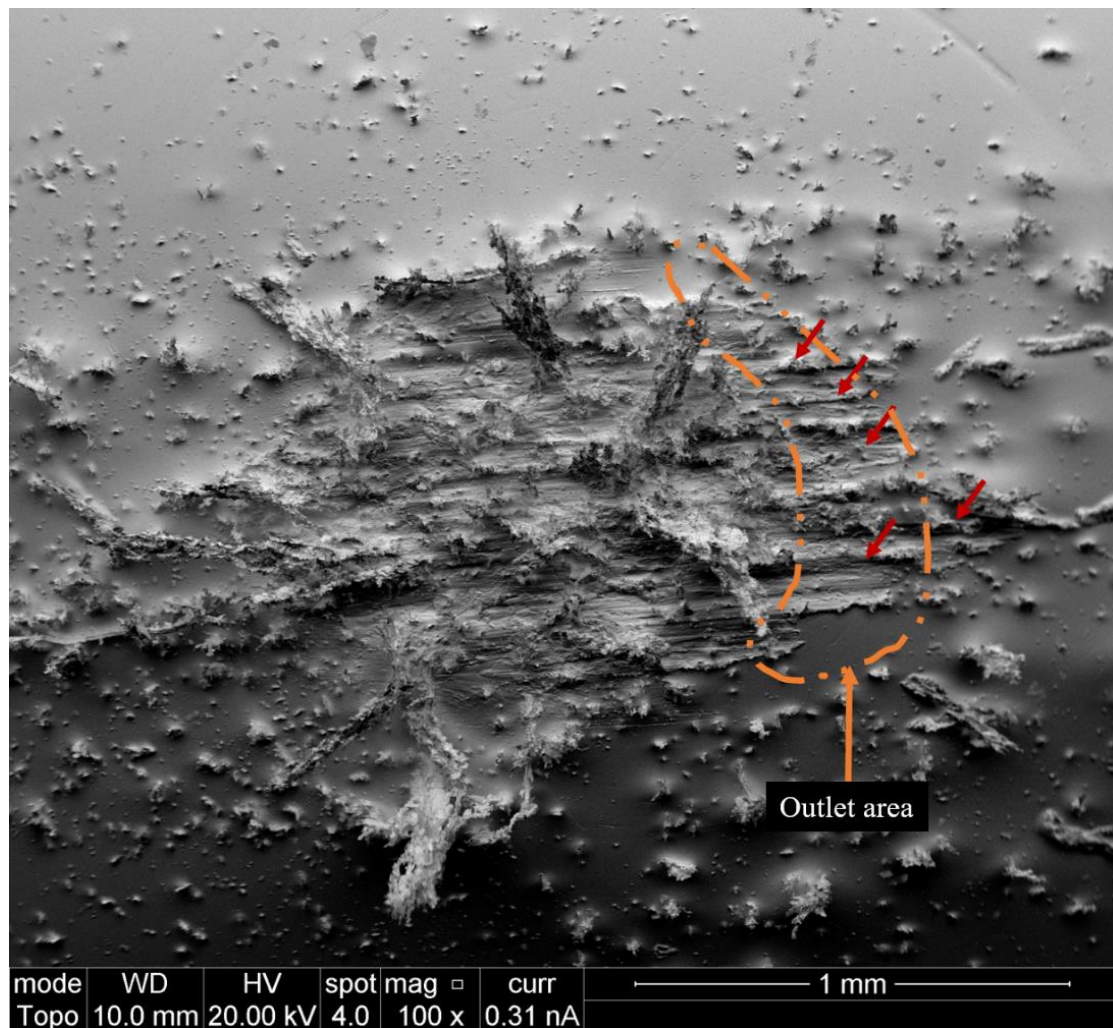
Figure III-23: Topography images of the used area from Test 6.

3.2.2. Disk

Studying the three different sliding tracks from the two gaseous environments shows that they share the same division into three zones: the ejection areas, the contact boundaries and the contact area.

In the air (Test 4), the components of each area are the same as in set 1: separated particles with different sizes in the ejection area and the contact boundary and much smaller particles in the contact area. However, in the tests conducted in argon (Test 5 and Test 6), the sliding tracks are full of large particles (compared to those from Test 4). A close analysis of **Figure III-24**, where we show the topography images of the sliding tracks from the tests in set 2, reveals how much different the third body distribution is when changing from air to argon.
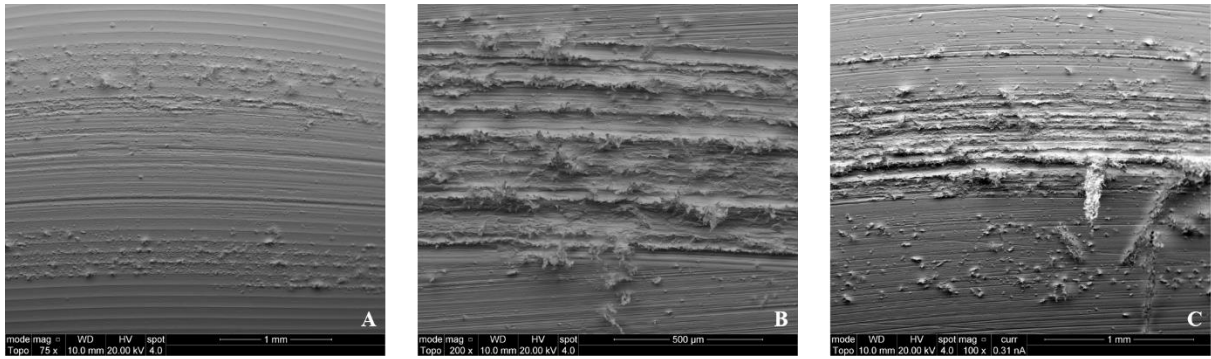
Figure III-24: Topography (BSE mode in the MEB) images of the sliding track from A) Test 4. B) Test 5. C) Test 6.

Even though the third body particles are not scattered alike between the two environments, the texture of the tracks is similar between the tests. In **Figure III-25**, we highlight the detected texture with yellow parallel lines just like in Set 1.
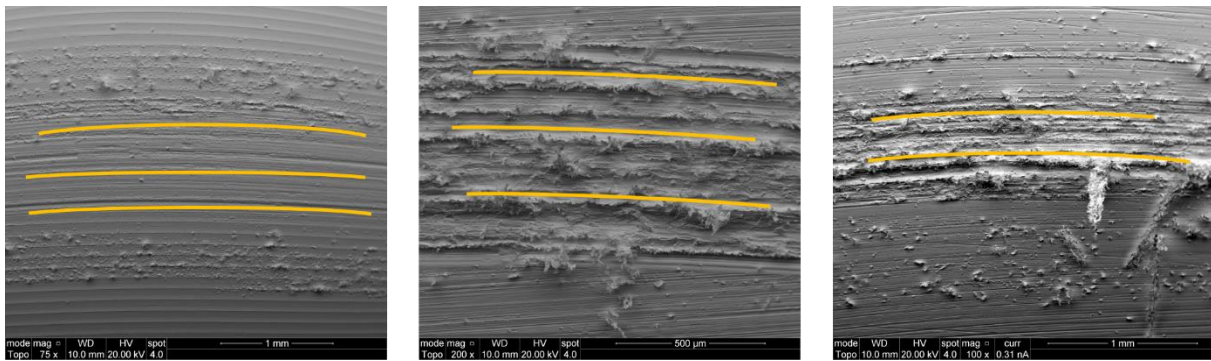
Figure III-25: Highlighted texture in the topography (BSE mode in the MEB) images of the sliding track from A) Test 4. B) Test 5. C) Test 6.

3.2.3.   Third body particles

We focus in this section on the third body particles found on both the pin and the disk after the tests in set 2. We respect the recommendations mentioned in the analysis protocol.

The third body is present under two types, which are similar to those described when analysing the results of set 1: individual separate particles and cohesive layers adhering to the surface of one of the first bodies. We display illustrations of both types in the three tests in **Table III-13.**

A study of the particles on the disk shows the effect of changing the gaseous environment from air to argon. In fact, in air, it appears that most of the created particles are transported to the ejections areas (**Figure III-26.A**) and only small particles are found on the contact area (**Figure III-26.B**). However, when using argon, the particles are not ejected as easily since the ejection areas in test 5 and 6 (**Figure III-27.A**) are not as full as in test 4, and they seem to

remain longer in the contact area (**Figure III-27.B**). The particles produced in argon seem to adhere to each other and to the surfaces of the first bodies more than those created in air.
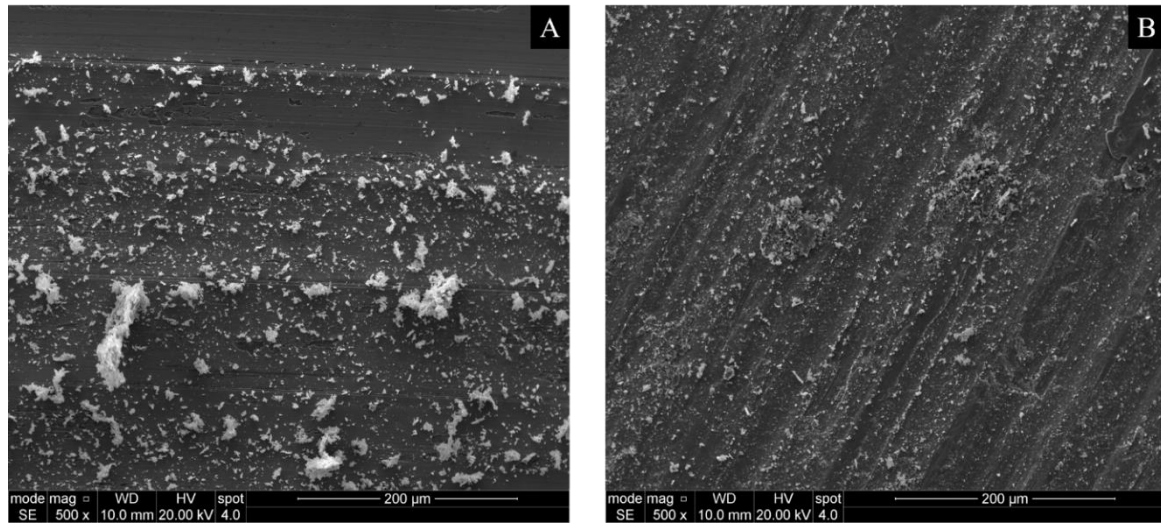
Figure III-26: Comparison between the third body particles (SE mode in the MEB) found in A) The ejection area. B) The contact area from test 4.
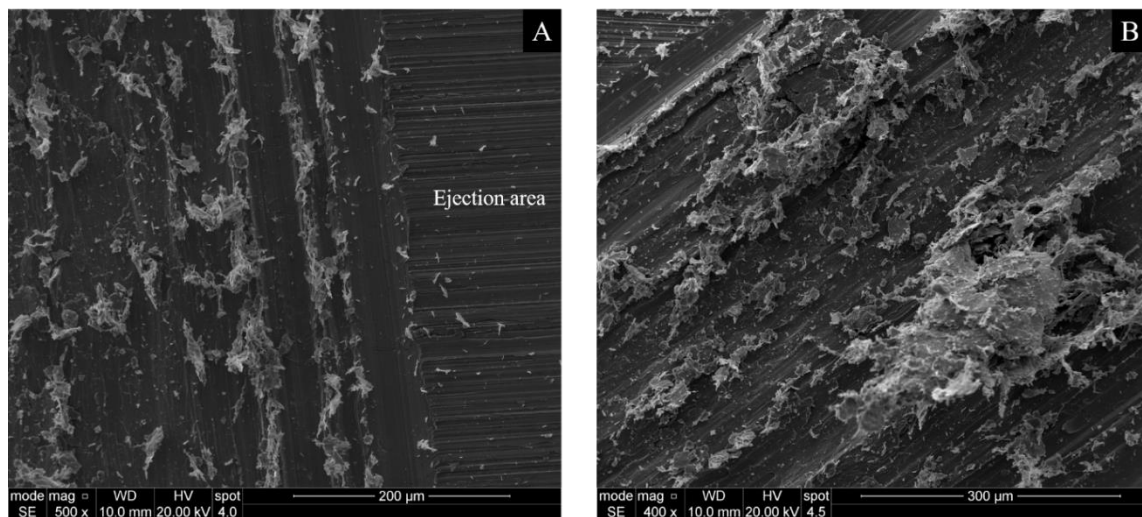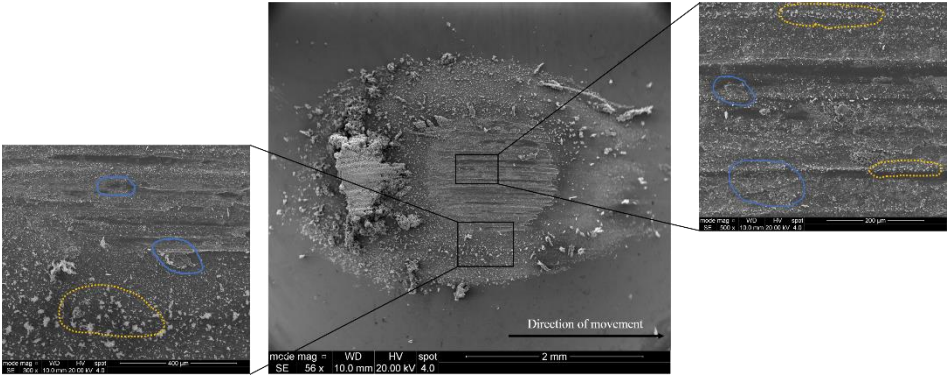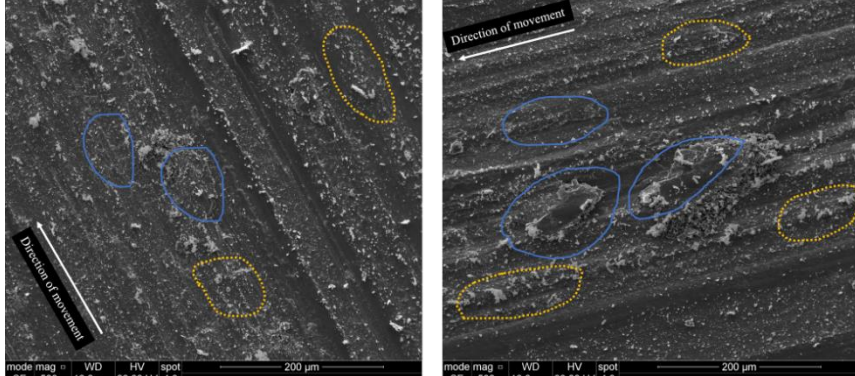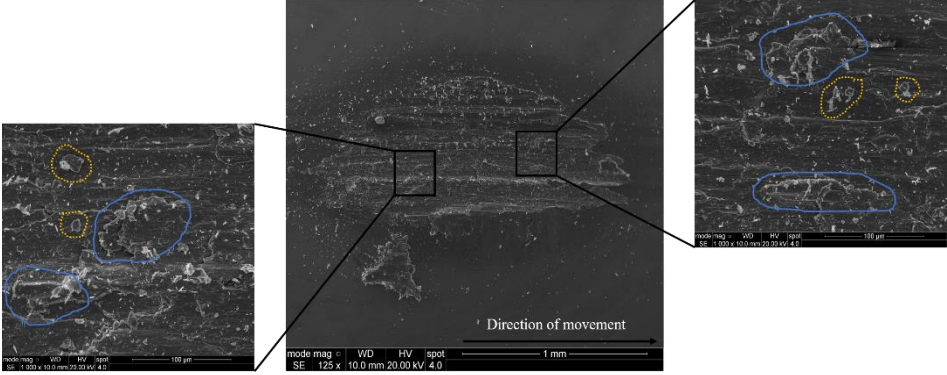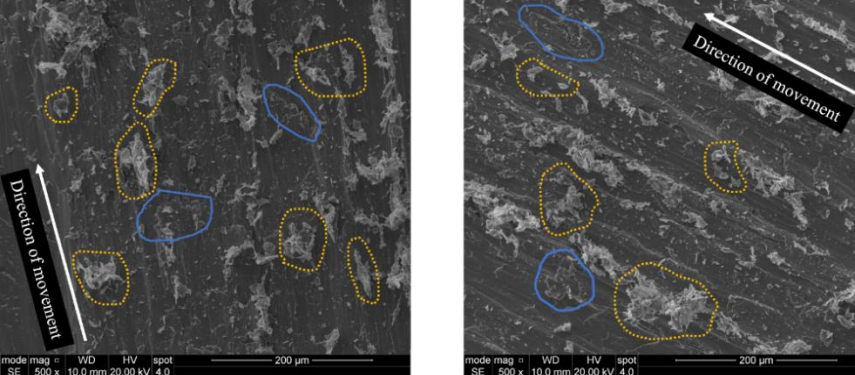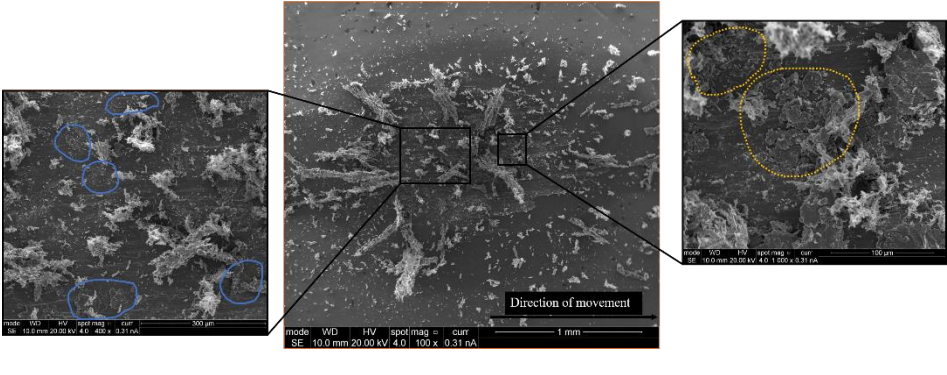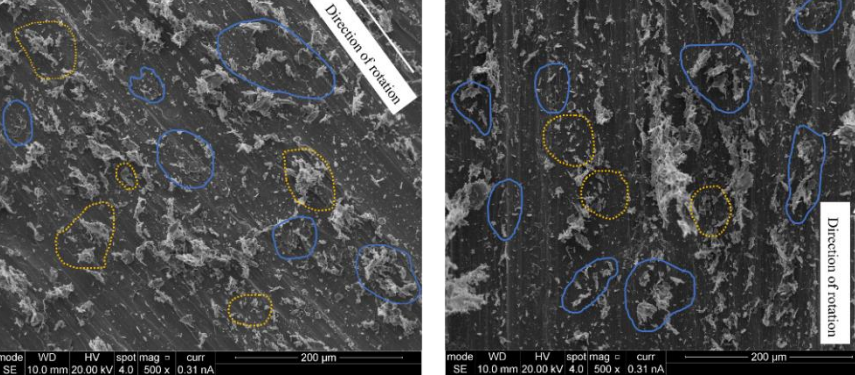
Figure III-27: Comparison between the third body particles (SE mode in the MEB)found in A) The ejection area. B) The contact area from test 5.

Table III-13: The state of the third body particles produced during the tests from set 2 (separated particles in yellow and cohesive layer in blue).

The peaks registered in the COF signals in **Figure III-20** are explained by the same scenario detailed in set 1. Nevertheless, since the particles in tests 5 and 6 stick together and to the surfaces of the first bodies more than those created in test 4, the threshold that the tangential force needs to reach before breaking the particles stuck in the portal zone is more important in tests 5 and 6. Hence, the peaks in the COF's curves for those tests are more important than the peaks in the COF's curve from test 4.

3.3. Set 2 rheological measurements

We assign the numerical values to each rheological measurement as defined in the analysis protocol in order to describe each test. **Table III-14** detail each measurement for the tests in set 2. The percentage of the dark zones in the contact area on the pin from test 6 could not be calculated due to the high concentration of third body particles that made that task impossible.

Table III-14: Set 2 rheological data.

	Test 4 Air	Test 5 Argon	Test 6 Argon
RD_1	0.26	0.43	0.35
RD_2	0.04	0.17	0.12
RD_3 (%)	12	21	-
RD_4	2	3	3
RD_5 (mm)	1.25	0.7	0.94
RD_6 (mm)	1.2	0.68	0.72

4. Set 3 analysis

During set 3, we study the effect of varying the total sliding distance covered during the tests. Similarly to the two previously analysed sets, we conduct three different tests and **Table III-15** summarizes the conditions for each of them. The change in the covered distance is a result of a change in the duration of each test. This ensures that the effects originating from the rotating motor or the transferring belt are the same between the tests.

Table III-15: Set 3 conditions

Tag	Radius (m)	Normal force (N)	Linear speed (m/s)	Temperature (°C)	Humidity (%)	Gaseous environment	Distance covered (m)
Test 7				27	42		$\cong$ 24
Test 8	0,01	10	0,37	23	49	Air	$\cong$ 22
Test 9				21	51		$\cong$ 15

4.1. Dynamic results

During the tests, we record the tangential force in order to calculate the coefficient of friction. In **Figure III-28**, we print the average per lap coefficient of friction and its evolution in distance. We distinguish three different phase when studying the three curves:

- *Phase 1(before the black dots)*: The COF rises in value almost instantly from zero to a maximum before which it decreases for the first time.
- *Phase 2(between the black dots and the red squares)*: The COF's value experiences large fluctuations, that are characterized by the peaks we record. The COF's average value is increasing during this phase.
- *Phase 3(after the red squares)*: The peaks are less apparent in the curves and the COF's average value reaches a more or less constant value.
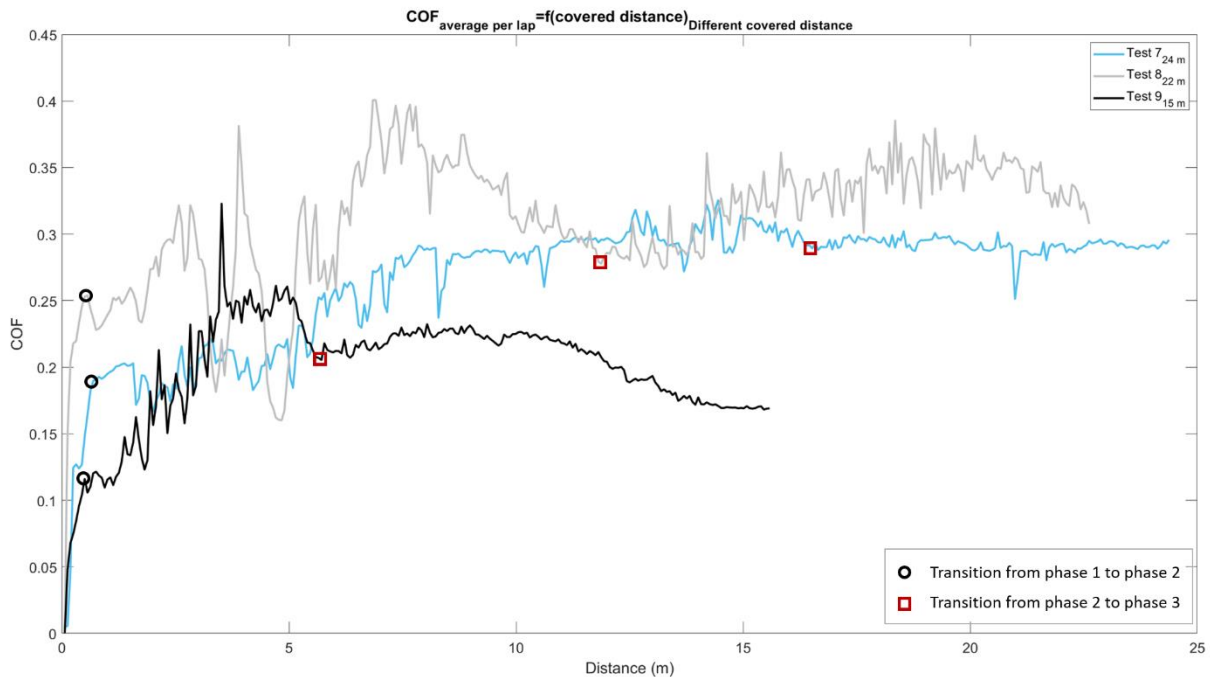
Figure III-28: The evolution of the friction coefficient average per lap for the three different tests of set 3.

With the tests having the same conditions, we supposed that the tests' results would be similar. However, studying the COF curves demonstrates multiple differences between the tests:

- Even thought, the length of *Phase 1* is similar between the tests, the length of the two other phases is not. For *Phase 2*, test 7 has the longest duration and test 9 has the shortest.
- The value each curve achieves at the end of *Phase 1* is not the same for the three tests. Test 8 has the highest COF at the end of that phase followed by test 7 while test 9 has the lowest.
- The peaks values differ between the tests in *Phase 2*. In tests 8 and 9, the peaks are very important in value when compared to the average value of the COF in those tests. In test 7 we observe smaller peaks.
- Test 7 and test 9 attain an established state in *Phase 3* recognisable by the small peaks and the constant COF value. In test 8, the peaks decrease in value in *Phase 3* when compared to those in *Phase 2*. However, they are more noticeable than in the other two tests.
- In test 9, the COF value decreases at the end of *Phase 3*, unlike the other two tests. This drop is mainly caused by a drop in the instant peaks values. In **Figure III-29** we can see this phenomena highlighted furthermore.
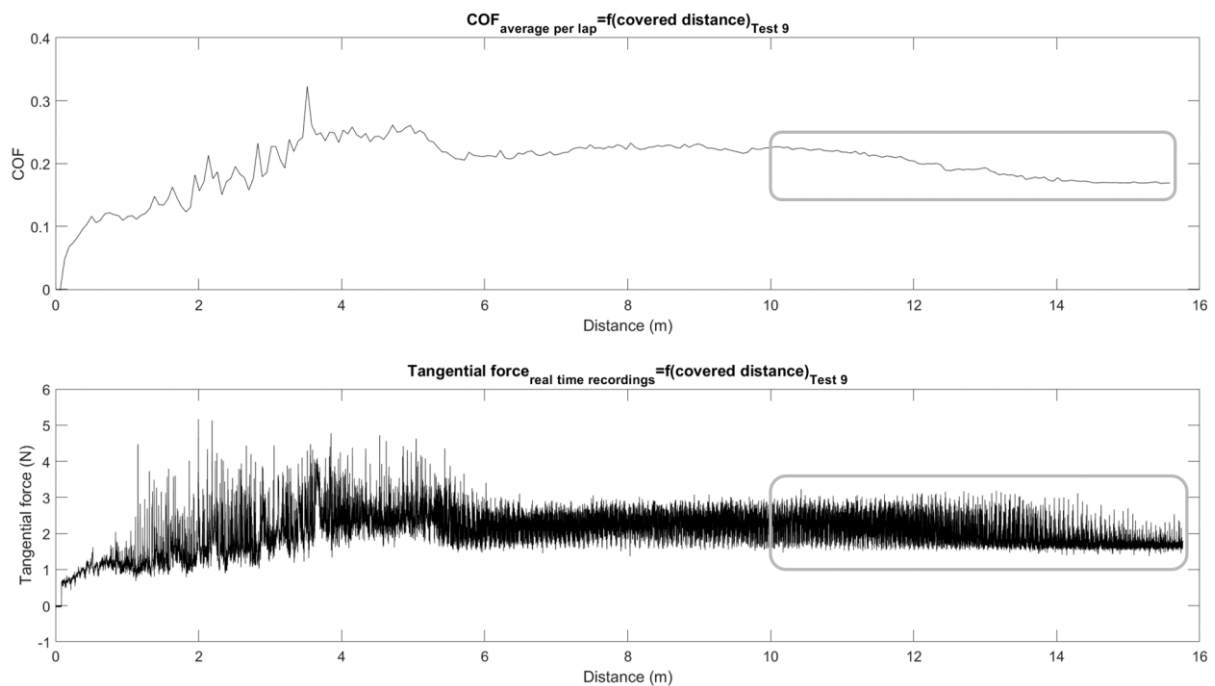
Figure III-29: The average per lap and the instantaneous evolution of the coefficient friction in distance during test 9.

The *in situ* films are only available for the test 7 and 9. They show that the contact goes through the same three phases as in set 2. The attraction phenomenon starts in test 9 from 49 laps as shown in **Figure III-30**.
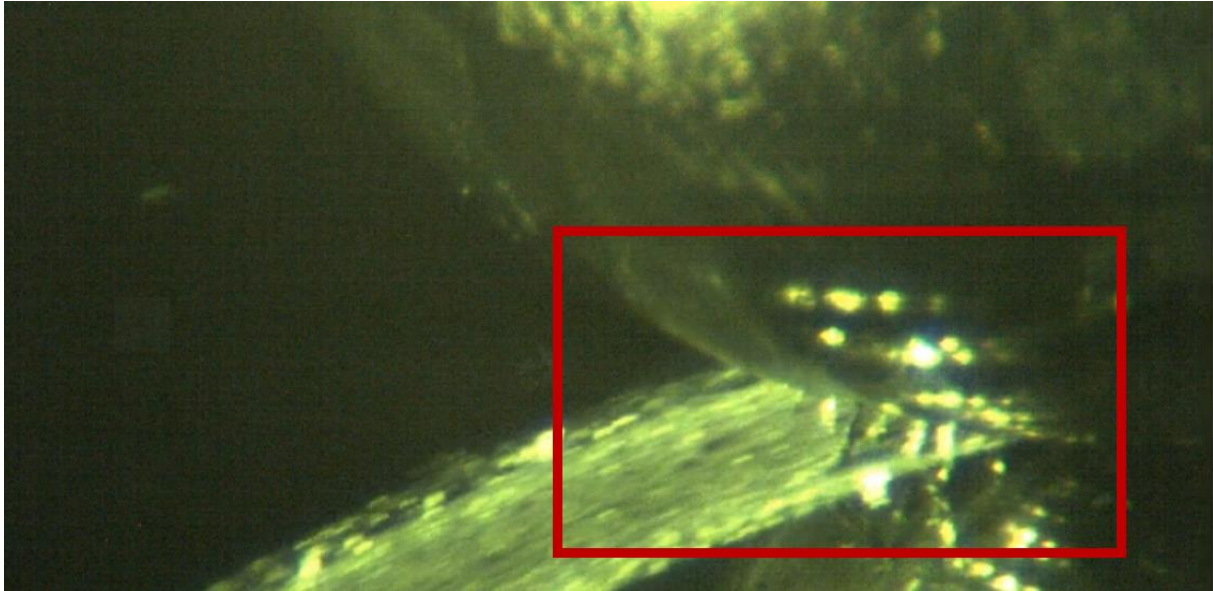
Figure III-30: The attraction phenomenon during the tests in test 9.

4.2. *Post mortem* investigations

Following the analysis protocol, we observe the pins and the sliding tracks after the tests of set 3 under the SEM. **Table III-16** shows the used areas from the first bodies at the end of each test.

4.2.1. Pin

When studying the used areas on the pin, we distinguish the same four zones as in set 1 and set 2: the Third body portal, the Inlet area, the Contact area and the Outlet area. The Third body portals, even though different is size, hasve the same structure between the three sets; compact and dense. The size difference can be a result of moving the pins when the test is stopped to the SEM or the progressive creation of this section during the test.

We can also see that the existence of the particles is not limited to the used areas but they are dispersed all over the observable area, especially in tests 8 and 9.

When we analyse the topographical images of the used areas on the pins, in **Figure III-31**, we can see that a certain texture is present on all the contact areas. It is highlighted using the dark blue lines in **Figure III-31**.

Table III-16: The state of the used area of the pin and the sliding track on the disk after the tests in set 3.

	Pin	Disk
Test 7 24 m		
Test 8 22 m		
Test 9 15 m		
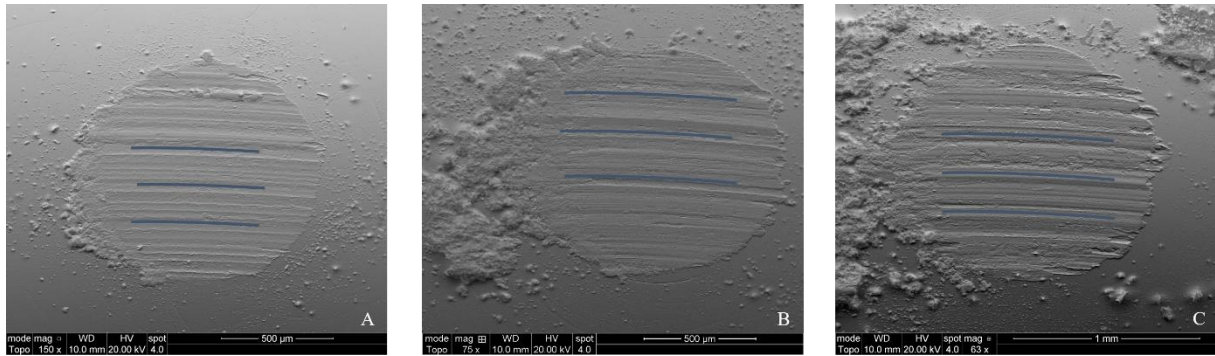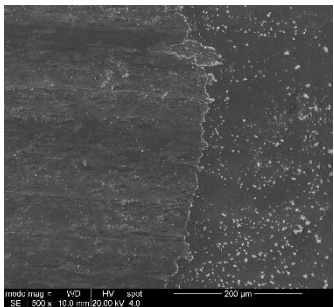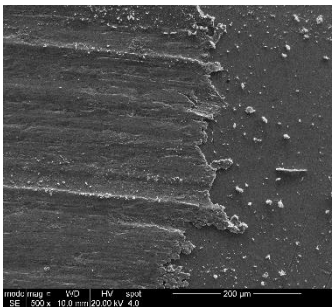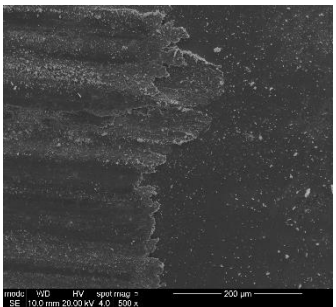
Figure III-31: Highlighted texture in the topography images of the used area on the pin from A) Test 7. B) Test 8. C) Test 9.

In **Table III-17**, we focus on the outlet area to judge the flow of the wear debris that took place during the tests. The flow in test 7 qualifies as mild since the borders where the used area ends is barely visible. In tests 8 and 9, we judge the flow as medium.

Table III-17: The outlet area state from the tests in set 3.

	Test 7 24m	Test 8 22m	Test 9 15m
Outlet area			

4.2.2. Disk

When inspecting the sliding tracks on the disk, we find the three usual areas, like in the other two sets; the ejection area, the contact boundary and the contact area. Similar to test 4 from set 2, the contact area at this scale of observation looks almost empty and the majority of the particles are in the ejection area. The particles we see in the image from test 9 have fallen from the pin during the removal process.

Studying the topography images of each sliding track shows a texture that reassembles to parallel lines. Those lines are highlighted in **Figure III-32**. This texture is similar to the one we find on the contact areas of the pins.
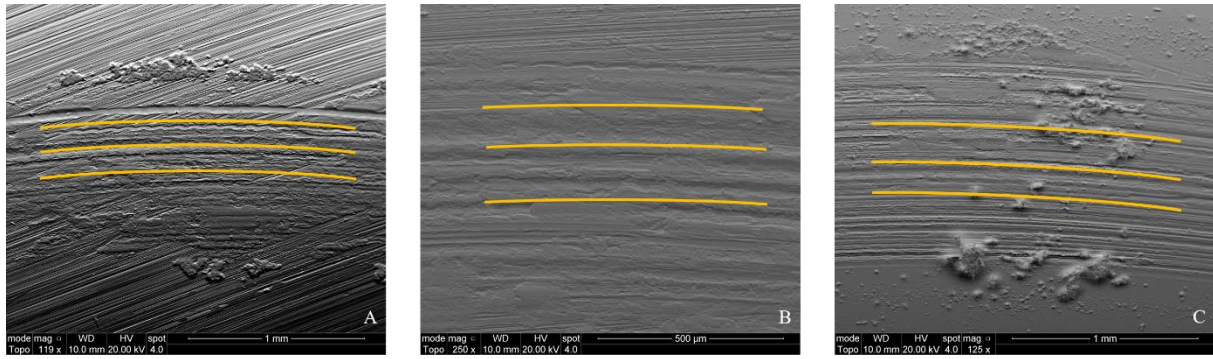
Figure III-32: Highlighted texture in the topography images of the sliding track from A) Test 7. B) Test 8. C) Test 9.

4.2.3.   Third body particles

Respecting the different instructions in the analysis protocol, we proceed to examine the third body created during set 3.

Just like the other two sets of tribological tests, we find the wear debris under two forms: separated particles and cohesive layer attached to the surface of one of the first bodies, as seen in **Table III-18**. The separated fragments are highlighted in yellow and the layers in blue. We can see in **Figure III-33** that the bigger fragments are always found in the ejection areas of the sliding tracks.
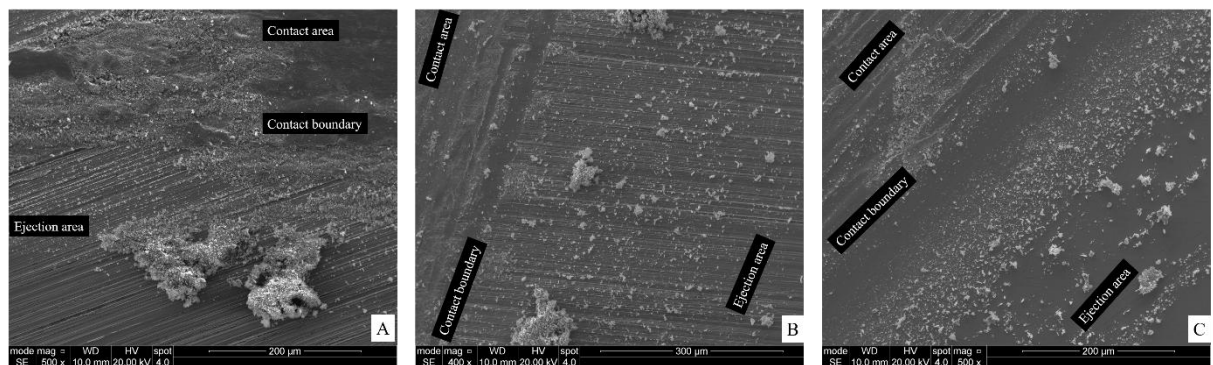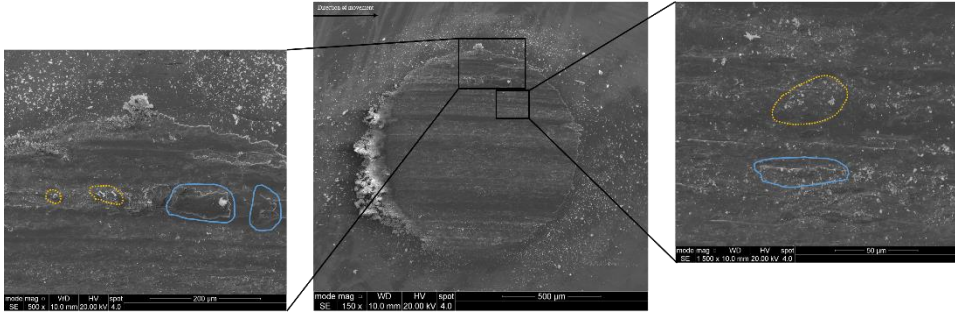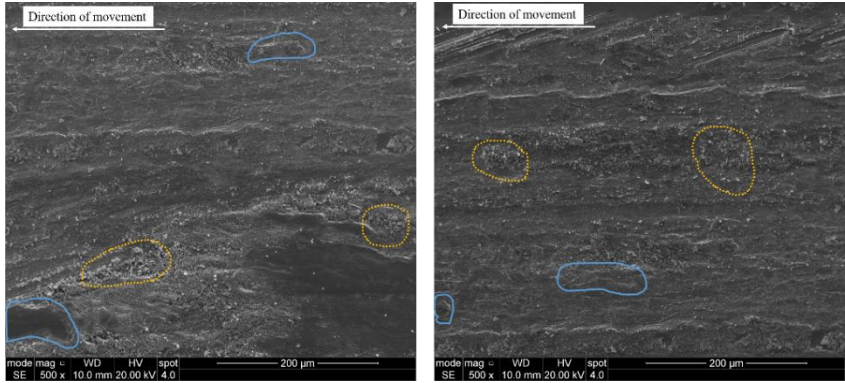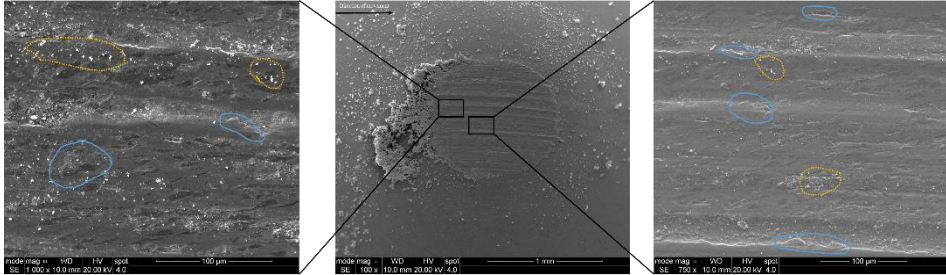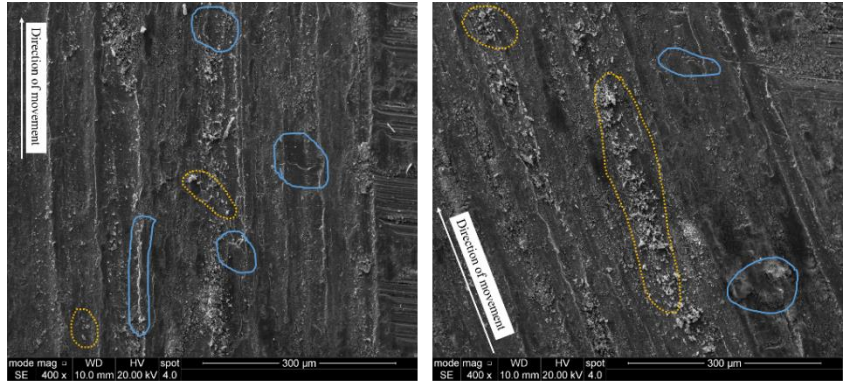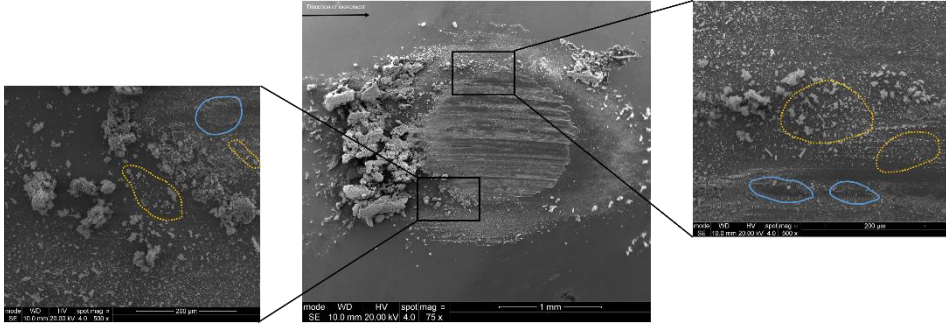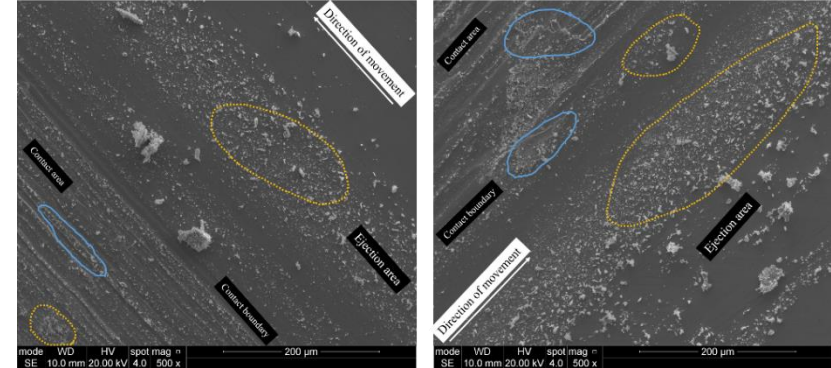
Figure III-33: Comparison between the three different areas found in the sliding track from A) Test 7. B) Test 8. C) Test 9.

Table III-18: The state of the third body particles produced during the tests from sets 3 (separated particles in yellow and cohesive layer in blue).

	Contact area on the pin	Sliding track on the disk
Test 7 **24 m**		
Test 8 **22 m**		
Test 9 **15 m**		

4.3. Set 3 rheological measurements

We calculate the different rheological measurements as defined in the analysis protocol. **Table III-19** lists the descriptors for each test of set 3 where we changed the total sliding distance.

Table III-19: Set 3 rheological data.

	Test 7 $_{24m}$	Test 8 $_{22m}$	Test 9 $_{15m}$
RD_1	0.29	0.38	0.2
RD_2	0.03	0.07	0.04
RD_3 (%)	20	21	35
RD_4	1	2	2
RD_5 (mm)	0.9	1	1.2
RD_6 (mm)	0.7	0.7	0.9

5. Conclusion

In this chapter, we analysed the experimental results of the tribological tests we conducted in order to be able to compare between the tests and study the effect of changing an experimental condition (Linear speed, Gaseous environment and Total covered distance). We choose six measurements that we calculate to characterize each test. **Table III-20** summarizes the rheological database describing all the nine tests concluded. Even though the rheological parameters vary from one test to another, we increase the probability of guessing correctly the test they presented by having more than one parameter. In addition, in order to calculate those descriptors, we need to totally end the tribological test and separate the first bodies. In the next chapter, we detail the methods we put in place to create the morphological database to describe the different particles created during each test.

Table III-20: The rheological database

	Set 1			Set 2			Set 3		
	Test 1 0.75 m/s	Test 2 0.56 m/s	Test 3 0.38 m/s	Test 4 Air	Test 5 Argon	Test 6 Argon	Test 7 24m	Test 8 22m	Test 9 15m
RD_1	0.44	0.54	0.49	0.26	0.43	0.35	0.29	0.38	0.2
RD_2	0.11	0.092	0.055	0.037	0.17	0.123	0.03	0.065	0.04
RD_3 (%)	16	18	36	12	21	-	20	21	35
RD_4	2	2	3	2	3	3	1	2	2
RD_5 (mm)	1.22	1.25	0.94	1.25	0.7	0.94	0.9	1	1.2
RD_6 (mm)	1	1.1	0.84	1.2	0.68	0.72	0.7	0.7	0.9

IV. Morphological data

Conventional analysis of the wear particles has traditionally been performed based on the judgment of tribology experts. However, even though their interpretations are precise and accurate, it is not an effective method when having a large quantitative data set. Besides, the knowledge extracted from such a subjective evaluation is difficult to transmit to non-experts, and therefore to use in other fields of the scientific or industrial community. Therefore, the use of modern technology could bring some improvements. We chose to develop an image-processing algorithm that describes the third body particles created during a tribological test, using their microscopic images. This is performed by calculating a number of selected morphological descriptors. Using this method will reduce the subjectivity issue and improve the repeatability of the tribological analysis.

This chapter starts by specifying the parameters used to capture the microscopic images of the wear debris particles. We then detail the image analysis algorithm that computes the morphological descriptors chosen in this project. Finally, we present the results of processing the different images from the three different sets of tests.

1. Microscopic images

1.1. Imaging parameters

The scanning electron microscope (SEM) offers many parameters that the user can adjust to have control on the characteristics (histogram, contrast type …) and the quality of the acquired images. Those parameters are fixed in our study in order to avoid adding variables that can affect the image processing results while upholding an acceptable image quality. **Table IV-1** presents the values of the parameters for every image used in our database. However, two SEM parameters change between the tests, which are the contrast and the brightness levels. The magnification chosen for the images is fixed as explained earlier for a maximum of 1500.

Table IV-1: SEM parameters for the acquired images.

SEM parameters	Value
Scanning velocity (µs)	30
Resolution (px)	2048*1768
Working Distance (mm)	10
High Voltage (kV)	20
Spot size	4 (current: 0.31 nA at 20 kV)
Contrast	[60 - 78]
Brightness	[38 - 43]
Polarization of the SE detector grid	250 V

1.2. Visualisation procedure

During our project, we choose to limit as much as possible the disturbance of the third body particles at the end of the test, in order not to influence their shape and spatial organization (distribution and localisation). During the visualisation procedure, we cover the entirety of the sliding track (zones defined in chapter 3) on the disk and the contact surface on the pin, in order to gather a maximum number of particles. This strategy does not take into consideration whether a considered particle was active (on the sliding track on the disk or on the contact zone on the pin) or passive (ejected) in the contact. In the case of an ejected particle, we do not have any information about the ejection time in relation with the contact's lifespan.

Even though the general observation of the third body particles proved its efficiency in the reconstitution of the contact's life(Colas *et al.*, 2013), this method did not help in differentiating between the different tests with different conditions, at the scale of the whole contact, as we saw in the previous chapter. It only resulted into identifying two types of wear debris: the separated particles (non-cohesive) and the cohesive layer. Therefore, a more detailed analysis method is required. Since we will be judging the morphology of the particles, the user must obtain images with particles that are easily identifiable. If the algorithm fails to identify the edges of each particle, the user should be able to correct it. We need to collect morphological data about as many particles as possible in order to have a collection that represents the totality of the particles created during the sliding test. However, the adequate quantity of particle morphologies to collect in order to obtain statistically meaningful results can only be estimated at the very end of the process, based on the success rates of the machine learning algorithms. This will be one of the sections of the next chapter.

2. Image processing

We developed a human-machine interface to facilitate the interactions between the user (who is not usually a developer) and the image analysis algorithms. We chose to develop our own algorithms in order to include the functionalities we need and to make the program compliant enough for it to adapt to likely future requirements. The processing algorithm calculates a number of morphological descriptors to characterize each particle.

2.1. Morphological descriptors

Normally the morphology of an object refers to a description of its appearance, which includes its shape and texture. However, in our project, the morphological descriptors will only focus on the shape of the third body particles. We do not take into consideration the texture aspect, although they represent a promising idea for future work. Many studies were conducted to determine the best descriptors to characterize the shape of sediment particles (Blott and Pye, 2008). However, we need to select a number of descriptors that does not overcomplicate the study and yet still gives an acceptable idea about the shape of the particles.

Four important aspects of the shape are considered crucial to judge the form of a particle, i.e., the elongation, the roundness, the regularity and the circularity (Mollon and Zhao, 2012). In **Figure IV-1**, we represent a geometrical explanation for the four morphological descriptors mentioned. A fifth measurement is added to quantify the 2D surface of the particles. We compute five descriptors for each third body particle:

- Elongation $= \frac{width}{Length}$: this descriptor defines the aspect ratio of the particle. The width of the particle in our case is calculated by minimization with respect to the angle of rotation θ. The closer the elongation measurement is to one, the less elongated the particle is.

- Roundness $= \frac{\sum R_{in}}{n * R_{insc}}$: It refers to the presence or the absence of sharpness in the particles edges. In the equation, the numbers R_{in} are the radii of a large number of circles that fit inside the particles and touch its outline at the same time (n is their number).

- Regularity $= \log_{10}\left(\frac{P}{P-P_{conv}}\right)$: This term indicates the presence of either projections or indentations on the particle's surface. In this equation, P is the perimeter of the particle and P_{conv} is its convex perimeter. The more regular the particle is (more convex), the bigger the value of the regularity descriptor is.

- Circularity $= \sqrt{\frac{R_{insc}}{R_{circ}}}$: It calculates how similar the shape of the particle is to a circle by dividing the radius of the largest inscribed circle R_{insc} by the radius of the circumscribed circle R_{circ}. The closer the particle is to a circle, the closer to one the circularity factor.

- Area: this is the value of the apparent 2D surface. Mathematically, this descriptor is calculated by multiplying the area of a single pixel by the number of pixels confined within the contour detected. Calculating the area with the metric system (opposite to pixel based) helps to reduce the magnification level effect.
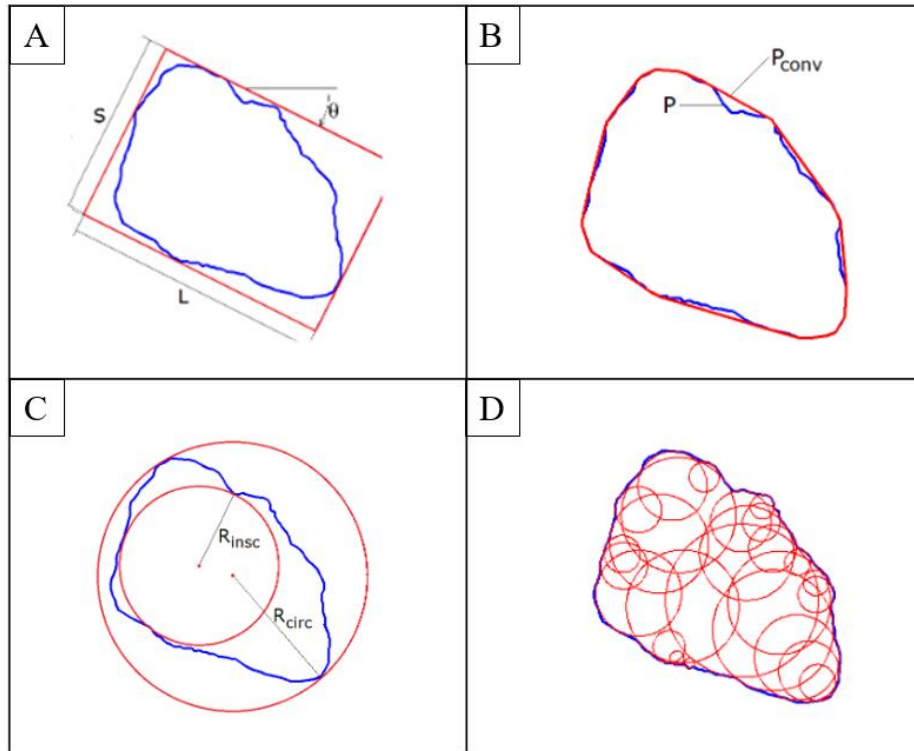
Figure IV-1: Geometrical representation of the morphological descriptors adopted in this study: A) Elongation. B) Regularity. C) Circularity. D) Roundness.(Mollon and Zhao, 2012)

The outline detected is only the contour of the particle as seen in the 2D image. It does not take into account the position of the particles relatively to the surface as shown in **Figure IV-2**. We observe the same particle but change the tilt degree of the disk's support. Even though the particle does not change, its aspect does due to its three dimensional shape and the two

dimension representation that we work with. To evaluate the difference between the two tilt degrees we present in **Table IV-2** a comparison of the descriptors obtained for two different tilt angles.
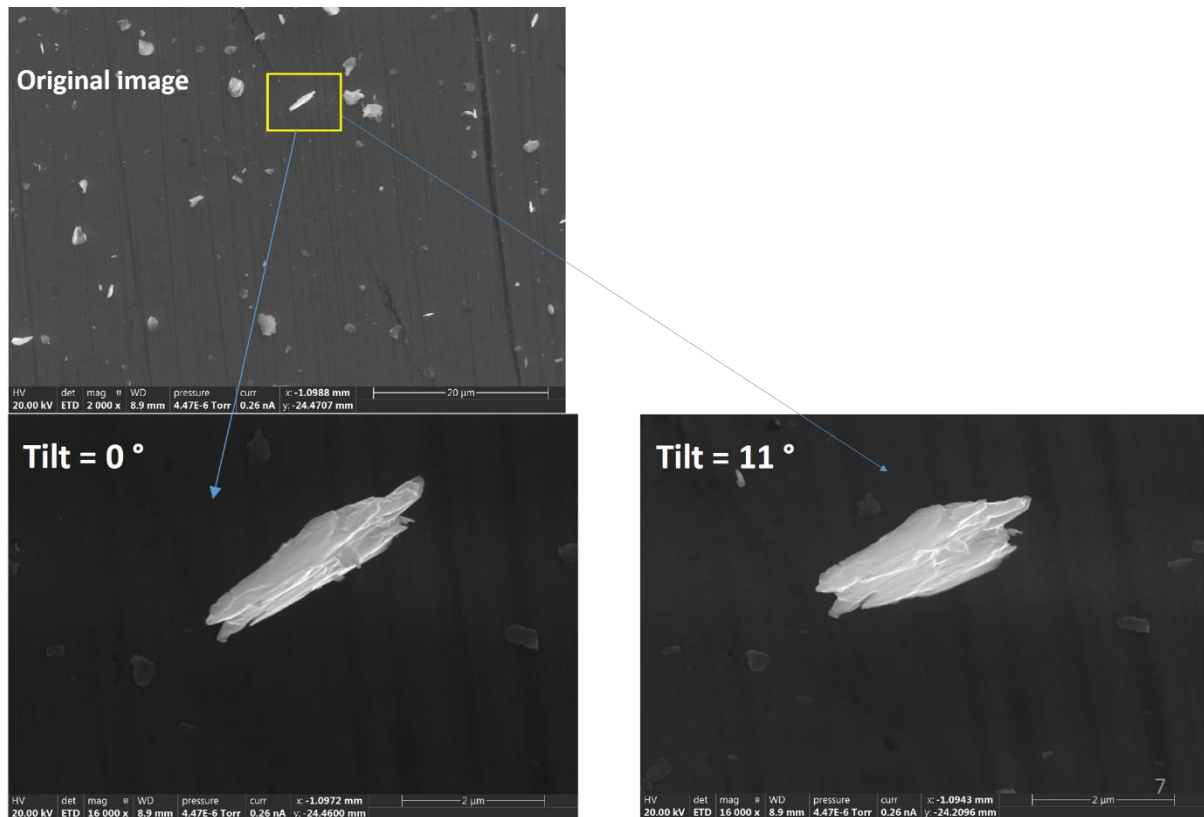
Figure IV-2: The tilt effect on the 2D SEM image.

Table IV-2: Results of image analysis of the particles in Figure VI-2

Tilt degree	Elongation	Regularity	Roundness	Circularity	Area (µm²)
0°	0,23	0,86	0,09	0,48	2,2
11°	0,36	0,87	0,06	0,59	2,6

In **Figure IV-3**, we display the morphological calculations for two particles from a sample of third body particles. We can see that the algorithm detects the edges of each particle and it sends back the five chosen morphological descriptors chosen. Particle 1 is almost a circle, which explains the elongation and the circularity factors that are almost equal to one. However, particle 2 is more elongated and less circular, therefore its elongation and circularity measurements are smaller than one. The roundness is small for both particles since they do not exhibit very sharp edges. The regularity factor results are also correct since particle 1 is more convex than particle 2.

	Elongation	Regularity	Roundness	Circularity	Area (µm)
Particle 1	0,9	1,47	0,07	0,92	662
Particle 2	0,52	0,8	0,07	0,61	929
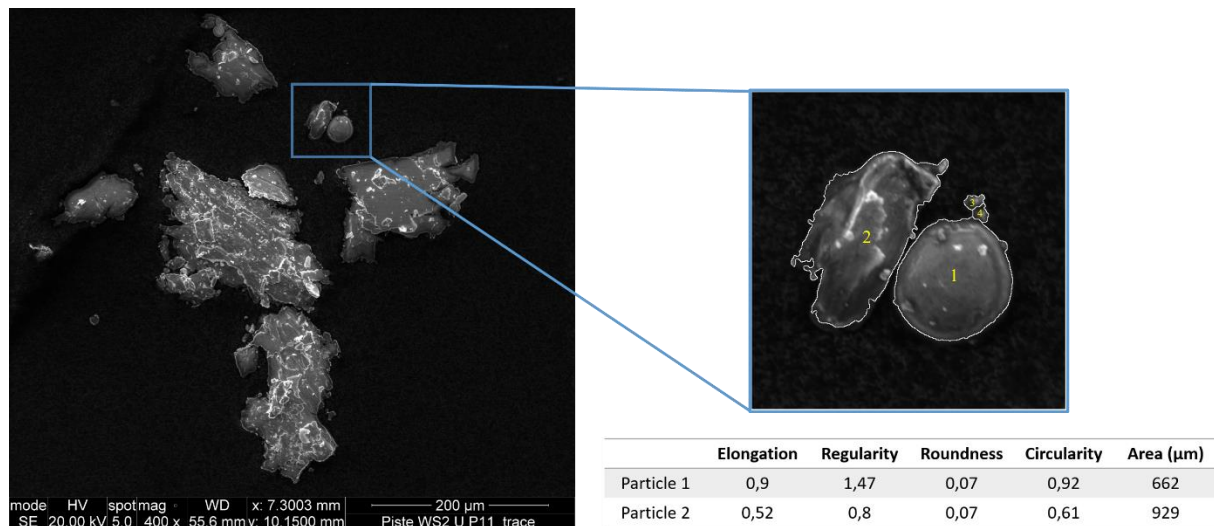
Figure IV-3: Example of morphological descriptor calculations.

2.2. Image analysis algorithms

At this point, we generate the third body particles through the tribotests and acquire their microscopic images. The general aim of the image processing algorithms is to calculate the morphological descriptors to characterize each particle visible in the image. The code is written in MATLAB. It is especially convenient for our purpose since it has several pre-programmed toolboxes, one of which being an image processing toolbox with different functions for image analysis. Besides, it is a standard programming language in the academic environment.

In order to be able to calculate the morphological descriptors of each particle, the algorithm has two intermediate steps that need to pass by. First, it must differentiate between the different particles present in the image and its background that is, in our case, either one of the surfaces of the pin or the disk or some other much larger particles. This separation is grey-level based. In all our microscopic images, the background is darker than the particles on it as can be seen in **Figure IV-4** that shows multiple images from different tests. Secondly, the physical edges of each particle (i.e. its geometrical contour) must be detected individually. Since we opt not to move the particles from where they lay at the end of the tests, the particles may set on each other or overlap, as we see in the different examples in **Figure IV-5**. In this case, those particles are not taken into account. The detection of the edges is a semi-automatic procedure so the user may fix them if he considers the edges detected are not the correct ones. To facilitate the different interactions with the code written, we developed a human-machine interface where all the developed functions are accessible through buttons, as seen in **Figure IV-6**.
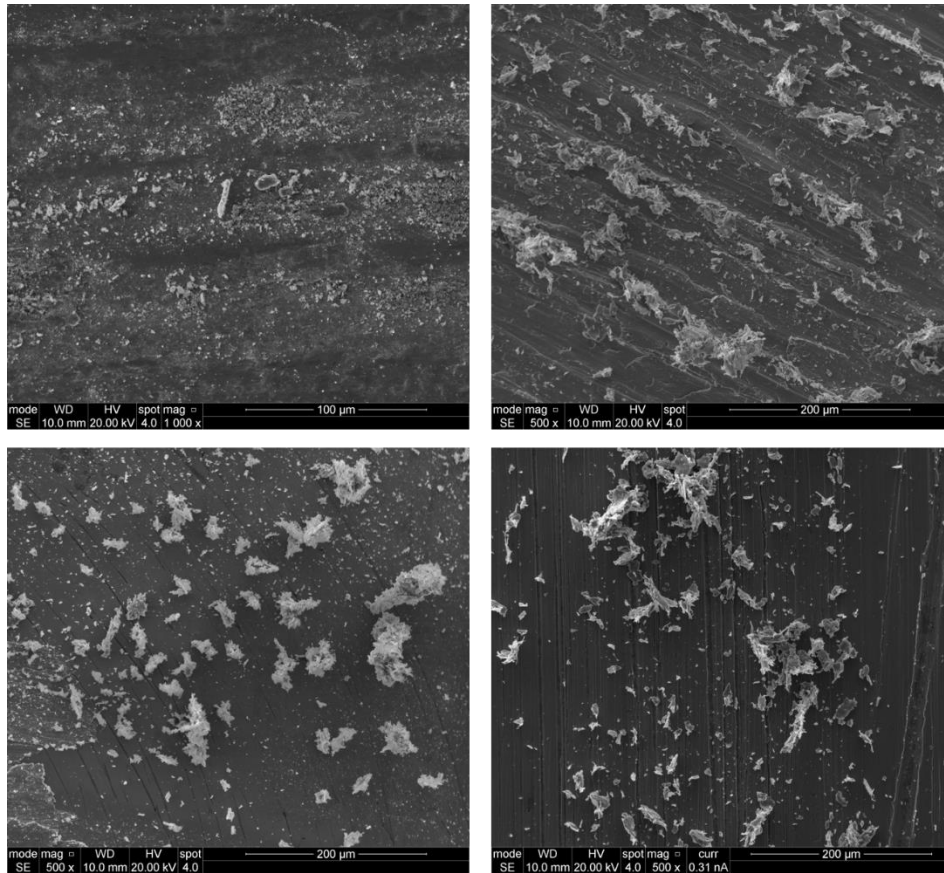
Figure IV-4: Different microscopic images of the first body surfaces resulting from the tribological tests.
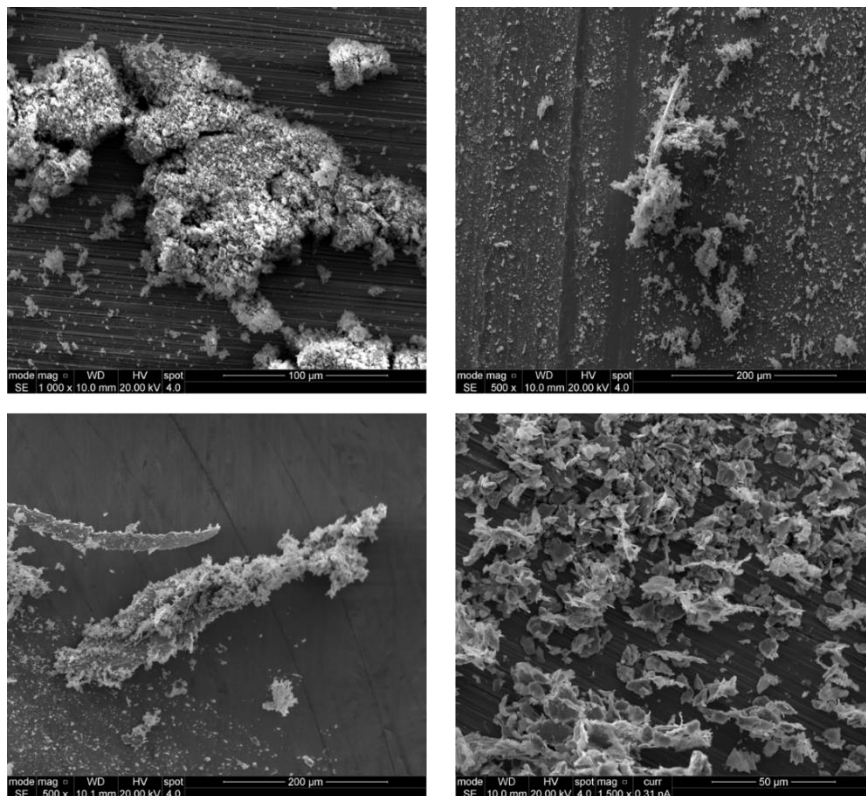
Figure IV-5: Illustrative unusable images of third body particles.

Figure IV-6: The main human-machine interface.

2.2.1. Human Machine interface

There are different buttons on the interface that we will explain further. We provide hereafter some brief descriptions of the general functions of the buttons.

The buttons developed in the interface are:

- 'Close': It closes all the open figures and clears the memory of the used variable. It is as if the software had just opened.
- 'Close figures': It closes the open figures only.
- 'Upload': This button opens the file navigator to upload the image we want to work with.
- 'Reimport image': It reloads the last original image used by the algorithm.
- 'Scale': It calculates the length of a pixel using the scale found on the image.
- 'Crop': It selects a region of interest of the image.
- 'Tilt': It gives the possibility of turning the image around its centre to a degree the user fixes.

- 'Delete surface': It deletes a selected surface from the image. The algorithm tries to fill the deleted space with a gradient of colours that simulates its surrounding. This function works best for deleting more or less small sections of the image.
- 'Adjust': This button allows changing the contrast. It opens a separate window with the histogram of the image. The contrast adjustment is done by changing the display range of the different histogram levels.
- 'Watershed': This button returns the separated particles present on the image depending on the grey level chosen using the slider.

2.2.2.   Algorithm function flow

While developing the algorithm, we kept the non-expert users that will be using it in mind. For example, the naming of the buttons is kept as simple as possible. In addition, the buttons that are not useful at a given state are kept disabled until the moment they are needed, in order to avoid errors. **Figure IV-7** contains a flowchart that summarizes the function flow of the image-processing algorithm.

- *Upload procedure*

Opening the interface gives only the user the possibility to load an image at first (in addition to the option of closing the interface). All the other buttons are kept disabled. After uploading the image, the scale button is enabled in order to urge the user to set the scale ratio. After this action is completed, all the other buttons are enabled.

- *Region of interest selection*

If the image has different particles, the user generally may want to divide it into different regions to process them separately. To do so, he may use the 'Crop' button. If there are some zones deemed erasable, the user can select such areas and delete them. The 'Adjust' button gives the possibility to regulate the contrast level of the selected image. If at any point, the user is not satisfied by the results of the choice made, he can always reload the original image using the 'Reimport image' button.
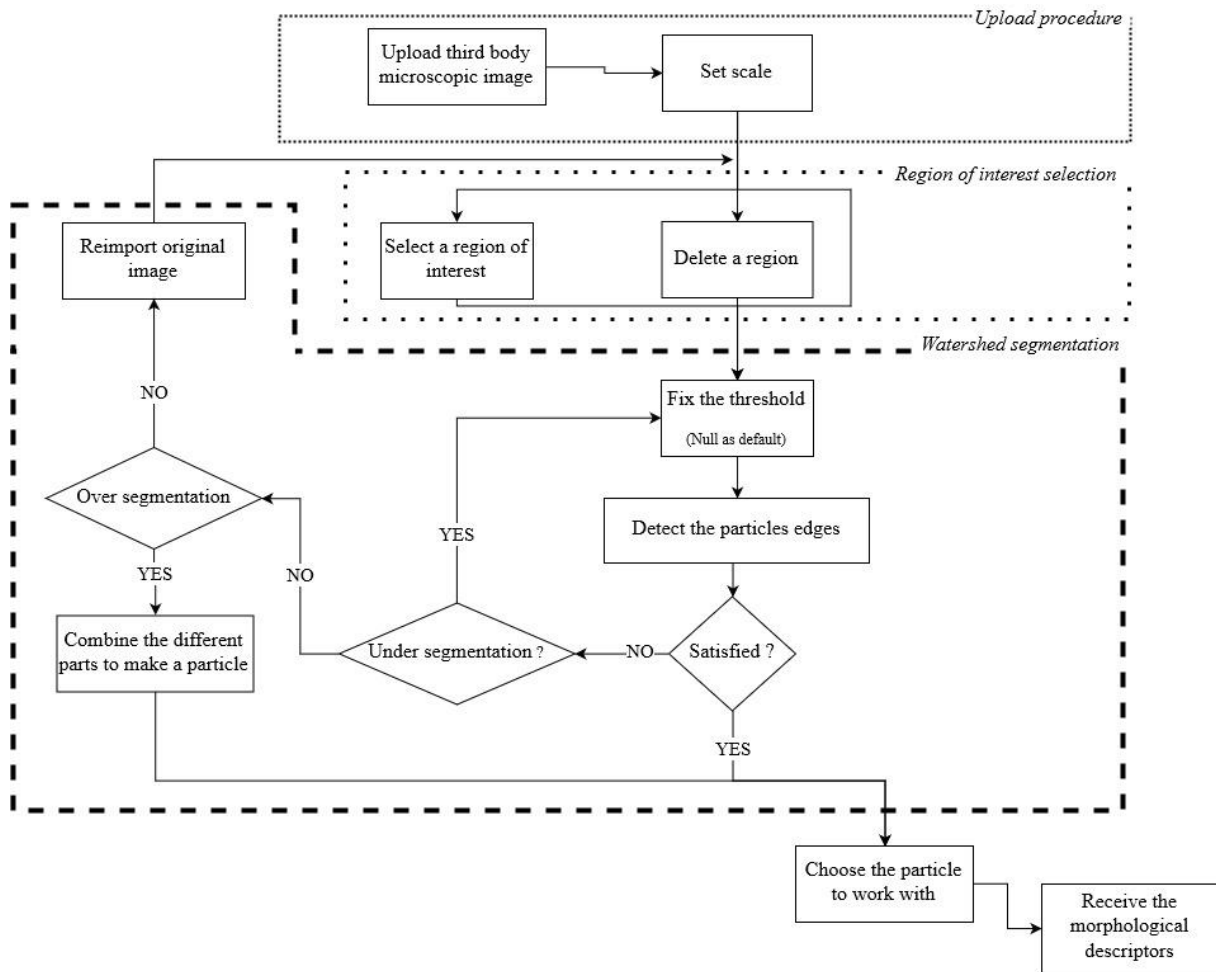
Figure IV-7: General flowchart of the image-processing algorithm.

- *Watershed segmentation*

When the region of interest is satisfactorily selected, the user needs to select a grey threshold level for the segmentation procedure using the slider on the human-machine interface (the threshold level is set to zero by default). Pressing on the 'Watershed' button starts the segmentation procedure that results into the detection of the edges of the particles. In order to reach this goal we use the watershed segmentation method, which is wildly used when studying medical (Pham, Xu and Prince, 2000; Atta-Fosu *et al.*, 2016) and material science (Waggoner *et al.*, 2013; Myasnikov, 2017) images. A region-based segmentation separates the entire area into disjoint portions. When reading an image it considers the values of the pixels as heights and transforms it into a topographical image. In a topographical image, we have three types of points:

- Points of minima: if a drop of water is placed on it, it will not move.
- Basins: If a drop of water is placed on it, it will fall to one single and known minima.

- Watershed lines: if a drop of water is placed on it, it can head to more than one minima. This line is the border between two regions that we are looking for.

This procedure is already programmed in Matlab. However, it is left to the user to prepare the proper matrix for its input. The raw SEM image needs to be treated before starting the watershed processing. The input image for the watershed function needs to be a grey-level image from where it can detect the appropriate minima for each particle.

Figure IV-8 illustrates the different transformations that the raw image goes through and the result of the watershed algorithm. First, we apply a filter that cleans the background noise and removes all the elements that are connected to the borders of the images. When the user selects the region of interest and if a particle touches the border of that selected region, we consider it incomplete and this filter helps to remove it and only focuses on the ones in the centre. In **Figure IV-9**, we present two cases to display the results of the filter when a particle touches the border and when it does not. We can see that the portion of the particle that touches the border is supressed after the filter is applied.

Then, the image's histogram is adjusted to increase its contrast. This filter saturates the bottom 1% and the top 1% of all the pixels values. At this point, we use the threshold value (Set earlier using the slider in the human-machine interface) and make our image binary. The resulting image may still contain some defaults referred to as 'holes'. A hole is a region of pixels that does not match its surroundings. A filter is used to correct those defaults and we can see the difference between first binary image and the second corrected one, especially when examining the central part of the particles.

At this point, we have a binary image where the particles are in white and the background is in black. Only now, we can proceed to apply the watershed segmentation function based on the Distance Transform Approach. This approach is regularly used along the watershed function (He, Fang and Liu, 2013; Win, Choomchuay and Hamamoto, 2017). It calculates simply the distance from every pixel to its nearest nonzero valued pixel. In a binary picture like the one we have, the distance transform calculates the distance of every black pixel to its nearest white pixel. A white pixel will be replaced in the distance transform result by a zero valued pixel since its nearest white pixel is itself. The distance transform will result into transforming the binary image into a grey-level image. Only now, we can proceed to apply the watershed segmentation function. At the end, we can have three outcomes: a correct segmentation, an over segmentation and an under segmentation. The three cases are seen in **Figure IV-8**.

The correct segmentation case gives the accurate borders of the particles in the image and the user can proceed to the next steps. Under-segmentation results into detected edges that contain several particles within them or contains parts of the background along with the particle itself. To correct this problem, the user should increase the threshold and retry the segmentation procedure. The over-segmentation has two possible results, either the edges do not cover the particle or the edges divide the particle into several portions. The solution for former case is to decrease the threshold level and repeat the segmentation process. In the latter case, the user can merge semi-automatically the areas until only the edges of the chosen particle correspond with its borders.

- *Morphological measurements*

Once the user judges that the detected edges are accurately describing the contour of the particle, the algorithm proceeds to opening a new window where, when choosing a particle, the user receives its five morphological descriptors, as we saw earlier in **Figure IV-3**. The algorithm also saves the contour of the chosen particle as an image and the measurements in an excel file to be used later.
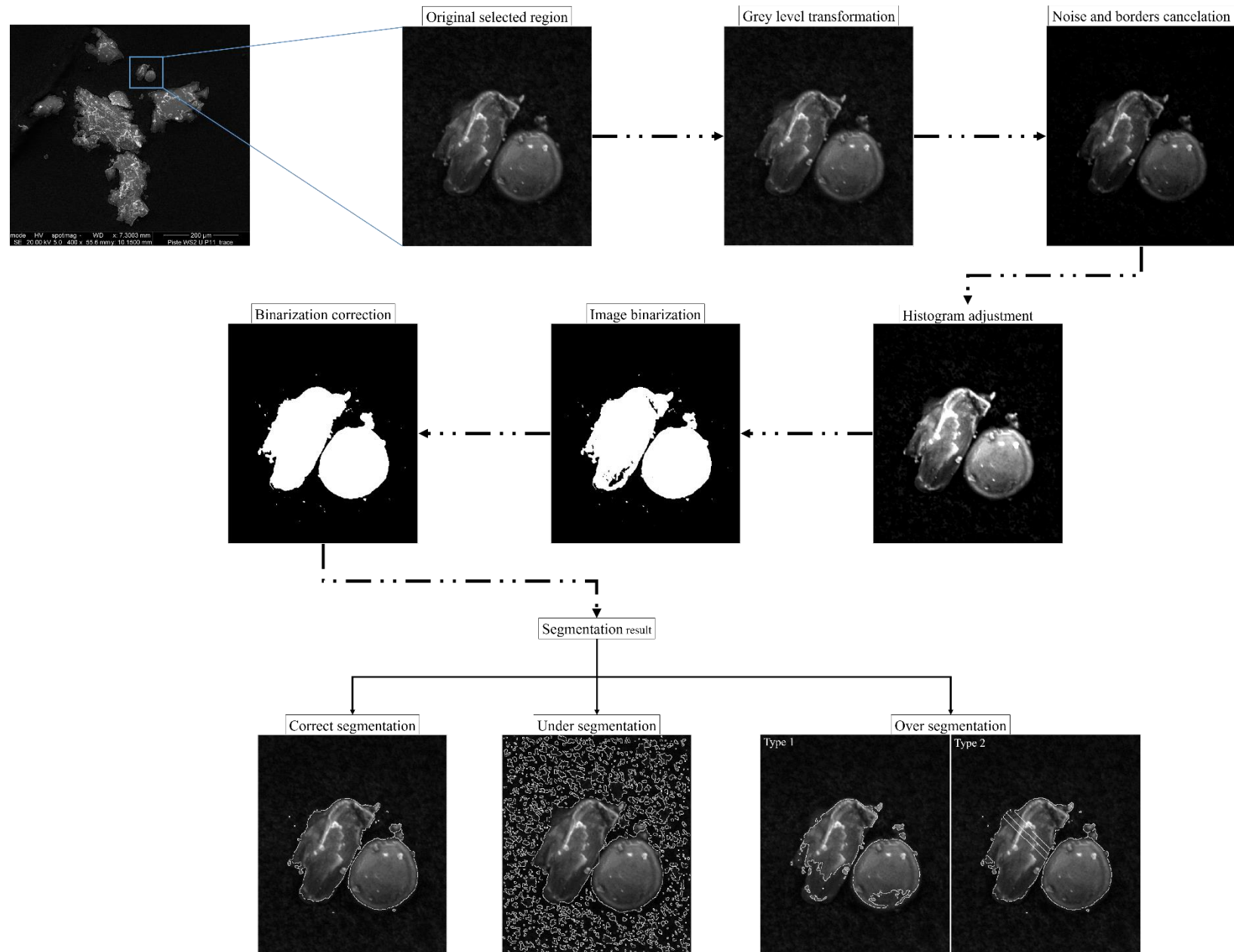
Figure IV-8: Edge detection procedure step by step
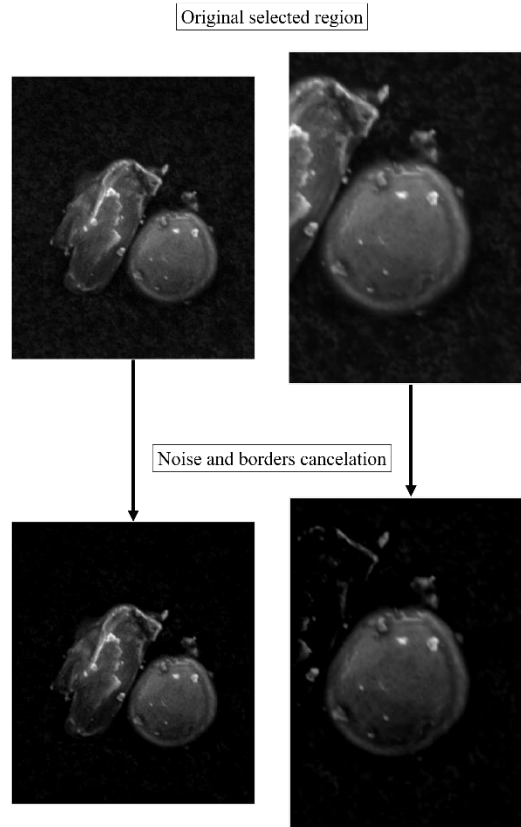
Figure IV-9: Noise and border cancelation effect.

3. Third body morphological descriptors calculations

To be able to compare the results between the different tests conducted, we calculate the average of each descriptor. To do so we need to fix a minimum number of particles that need to be treated from each test to ensure that the average we calculated is applicable on all the particles with a certain degree of error. This is slightly complicated to do, in our case, since we do not know the total number of particles n that is created during each test.

We choose for our study that the margin of error should not be more than 10% at the 95% confidence interval. This means that if we study 100 new particles from test 1, for example, 95 of those particles should have descriptors equal to the averages calculated with a margin of error less than or equal to 10%. The margin of error (MoE) can be estimated using the following equation:

$$\text{MoE} = Z * \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \leq 10\% = 0.1$$

The precise calculations are included in **Appendix 1**. The number of particles n that we need to analyse from each test is equal to 96 particles. It is the minimum amount of particles to guarantee that the margin of error is at 10 % maximum at the 95 % confidence interval. Our project aims to study the practicability of linking the rheological and morphological databases through machine learning algorithms. Consequently, having a 10 % margin of error is acceptable for this study but it is also quite high. Decreasing it is one of the future perspective of this study.

The particle selection method is handled by the user in order to eliminate as much as possible the particle/agglomerate confusion that may take place if the selection procedure is fully automatic. In fact, if we developed fully automatic algorithms without human intervention, it would consider any closed area that has a higher grey level than the fixed threshold as a particle. However, this assumption cannot be considered as correct in our case because not moving the third body particles after the tests created agglomerates of particles. The agglomerate is a number of third body particles that are hard to distinguish them from each other separately, because of their spatial distribution when the microscopic images was taken. However, we find two main types of agglomerates. In F**igure IV-10**, we present both types; one that can be used in this project and the other that is to be avoided. **Figure IV-10.A** (left image) is an agglomerate that consists of a number of small particles stacked on top of a much larger particle that serves as a base for them. This only changes the texture of the underlying particle, but this texture is not included in this work. Its shape is intact and can still be considered. However, the agglomerate in **Figure IV-10.B** (right image) is to be avoided because it consists of particles that are lying very close to each other (on top of each other in some cases) without leaving the possibility to understand where the edge of a particle does start and when its neighbours' one does. Even though such agglomerates are not accounted for when calculating the morphological descriptors of the third body particles, their presence can be one other descriptor that may be used in future work.

Even as a semi-automatic procedure, judging whether what we are observing is a particle or an agglomerate is a human based decision and remains very subjective. The scale of observation of the particles plays a very important role in judging if we are considering a particle or an

agglomerate. In **Figure IV-11**, we demonstrate the significance of the scale of observation. Both red squares on the top image and on the bottom image are highlighting the same region. In we only had access to **Figure IV-11.A** (top image), we would judge that this zone contains a particle. However, when we zoom more and get **Figure IV-11.B** (bottom image), we can clearly see that it is an agglomerate and not a single particle. It should therefore not be taken into account for the analysis process.

From each test, we gather the morphological information of at least 100 particles both from the contact surface on the pin and from the sliding track on the disk. We calculate the average of each along with the standard deviation between the values to evaluate their homogeneity.
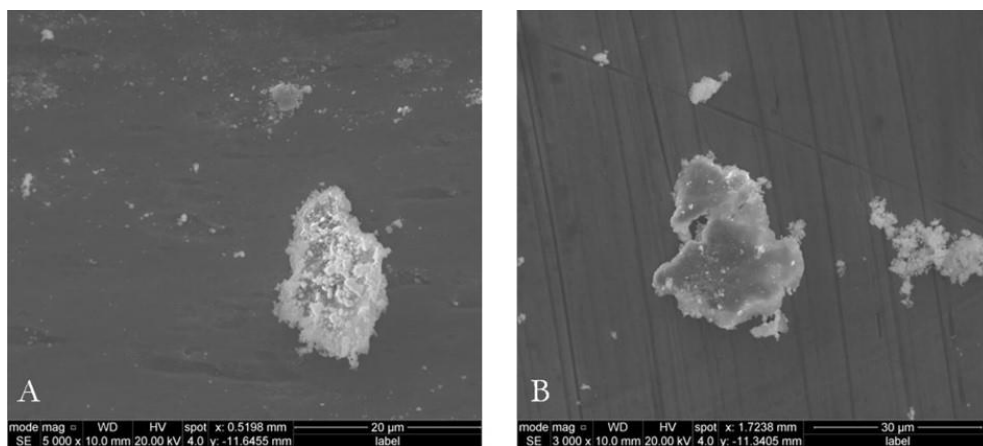
Figure IV-10: Two types of agglomerates found after the sliding test.
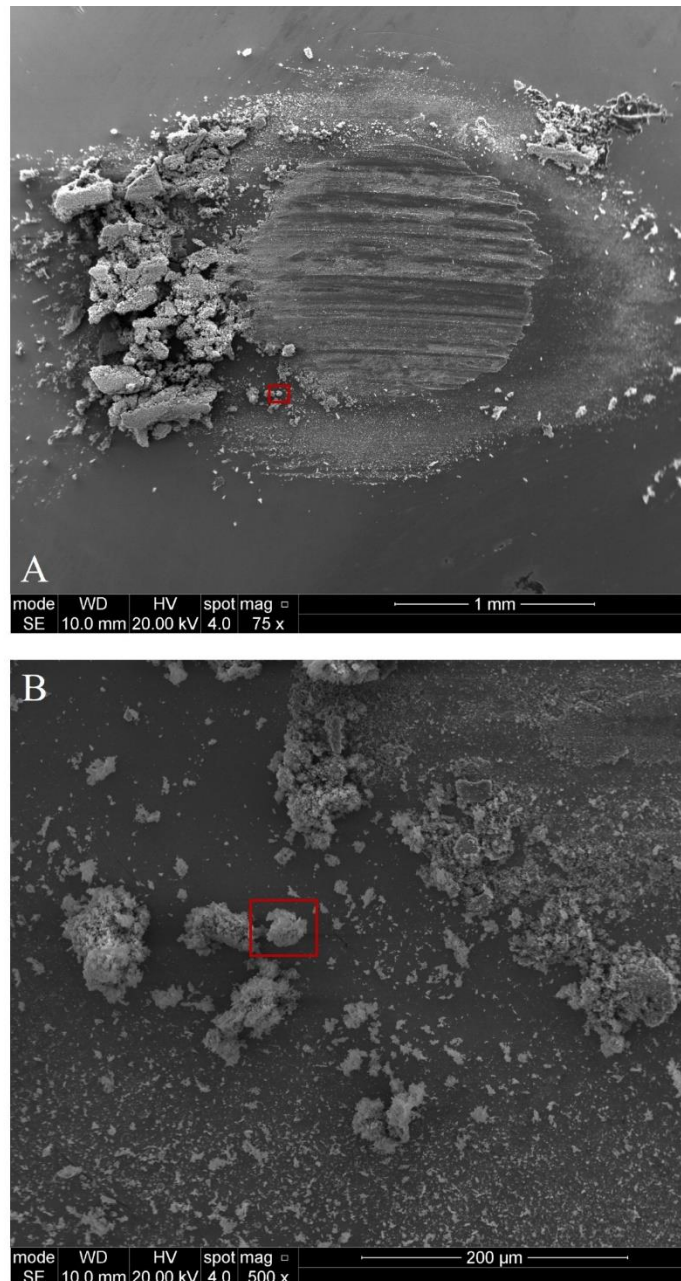
Figure IV-11: The scale effect in differentiating between the particles and the agglomerates.

3.1. Set 1 morphological results

The linear speed changed between the three tests in set 1. Test 1 has the fastest linear speed ($0,75\,m/s$) and it goes slower for test 2 ($0,56\,m/s$) and slower again for test 3 ($0,38\,m/s$). This speed modification procedure is explained in *Chapter 3.Section 3.Paragraph 1.* The results of the morphological analysis of the third body particles produced during the three tests are presented in **Figure IV-12**. In the first four descriptors, we display an example of a particle that has the average value of that descriptor.

Studying the morphological descriptors shows that, except for the elongation descriptor that is similar between the three tests, the other morphological descriptors of test 2 are noticeably different from the other two tests. The particles from test 2 are less regularly shaped, less round and slightly less circular than the particles from test 1 and test 3. Nevertheless, the particles from test 2 are much bigger, with respect to area calculations, than the particles from the other two experiments. However, the standard deviation shows that the individual measures vary considerably inside the same test.

3.2. Set 2 morphological results

In set 2, the varying experimental condition is the gaseous environment where the contact takes place, using the gas box. The specifications of the box and the procedure are described in *Chapter 2.Section 3.Paragraphe 3*. The two gases used are air in test 4, and argon in tests 5 and 6. **Figure IV-13** summarizes the morphological results as in to the previous section.

The regularity, the roundness and the area indicators can discriminate between particles formed in argon and those in air. Nevertheless, considering that the two tests conducted in argon (test 5 and test 6) are in theory identical, the results of the morphological analysis of the particles resulting from those two tests vary considerably. For example, the analysis show that particles from test 5 are on average twice as big (area wise) as the particles from test 6. However, no distinguishable difference between the results of the elongation and the circularity descriptors of the three tests in set 2.

3.3. Set 3 morphological results

For the tribological tests in set 3, we varied the sliding distance by changing the duration of the test itself. We avoided changing the radius of the sliding track or the linear velocity mainly in order not to change the vibration levels introduced by the motor. The pin travelled for $24\,m$, $22\,m$ and $15\,m$ in test 7, test 8 and test 9 respectively.

As seen in **Figure IV-14**, the particles produced during the tests in the third set have the same level of elongation and circularity. Yet, the bigger the covered distance, the bigger the regularity and the roundness descriptors and the smaller the particle area.
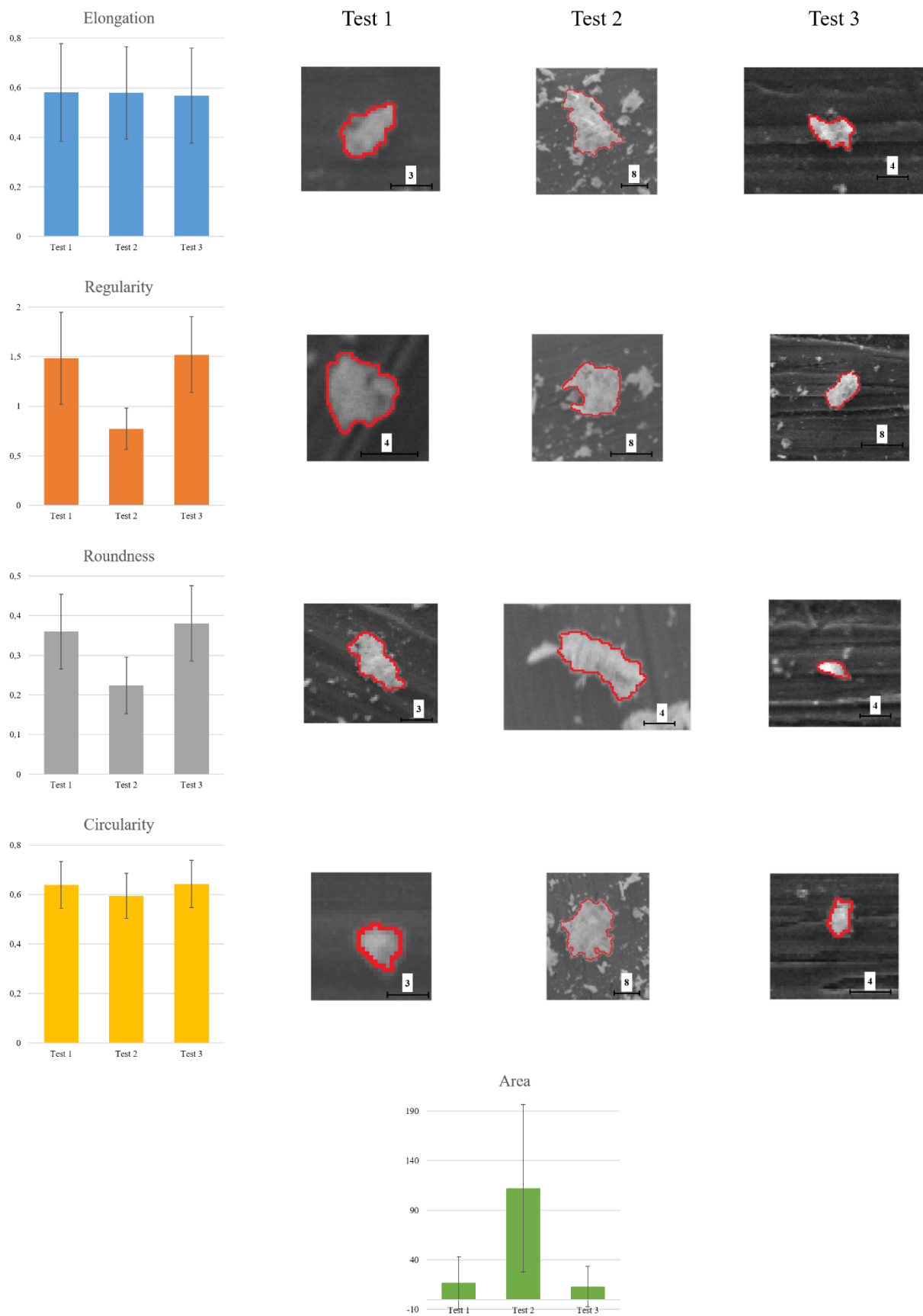
Figure IV-12: Morphological descriptors of the particles from Set 1 (the scale displayed is in μm*).*
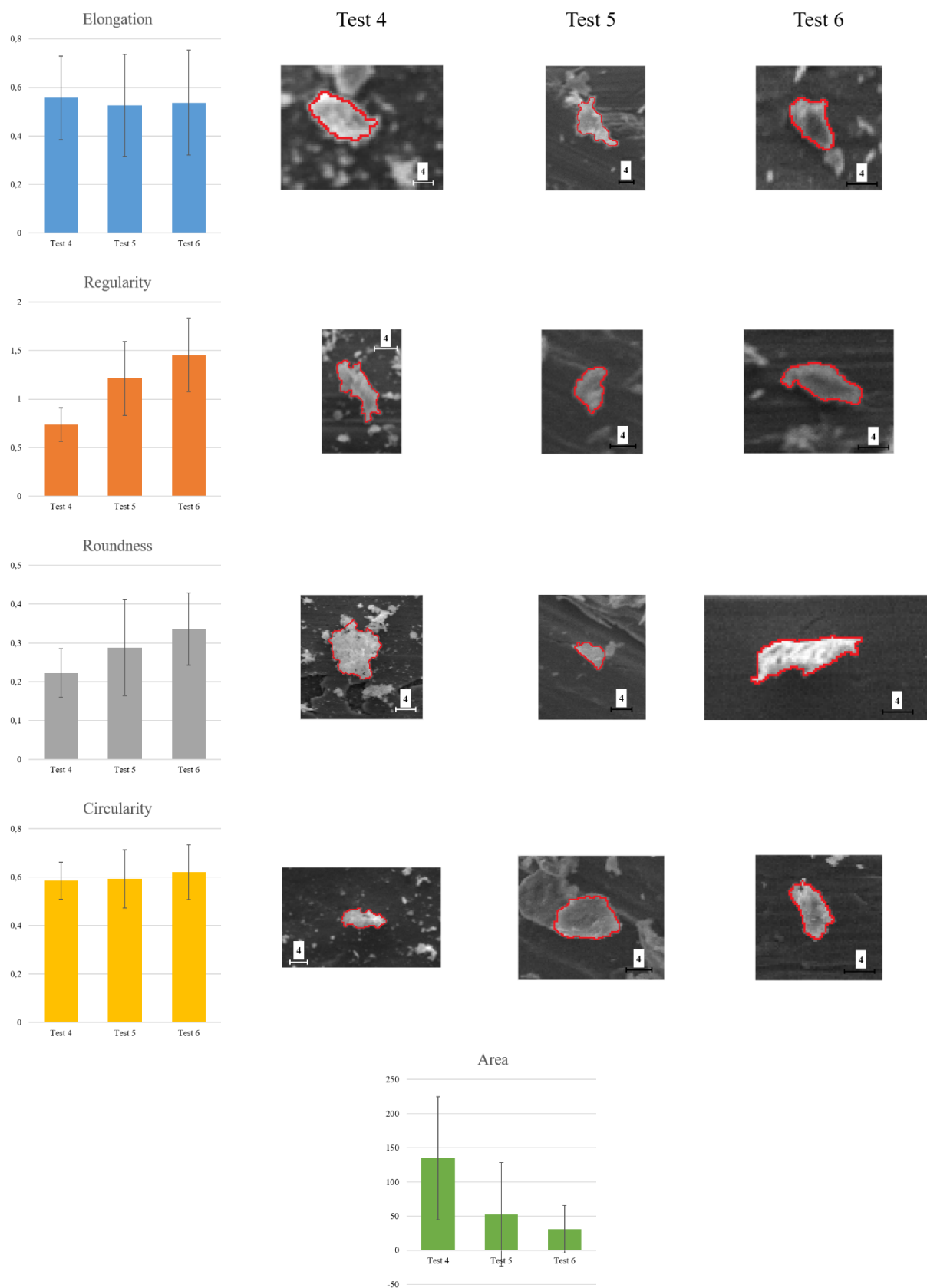
Figure IV-13: Morphological descriptors of the particles from Set 2 (the scale displayed is in µm).
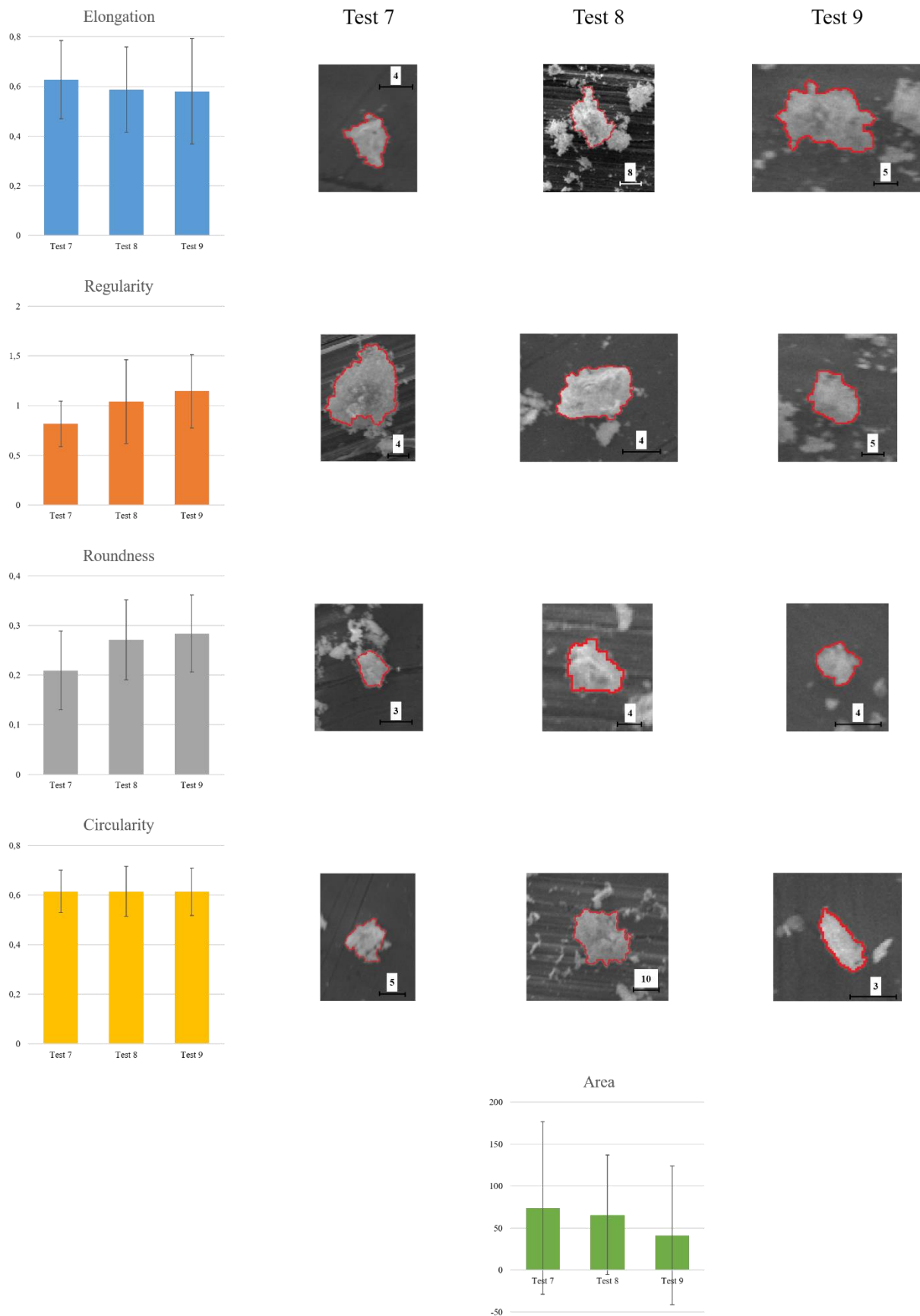
Figure IV-14: Morphological descriptors of the particles from Set 3 (the scale displayed is in µm).

4. Conclusion

In this chapter, we analyse the particles created during the different tribological tests in order to gather a number of morphological descriptors. We put in place a standardized process to choose which particles are to be considered and how many are needed in order for the analysis to be statistically significant.

We summarize the results in **Figure IV-15**. Certainly, differences between the tests are noted, especially in the regularity, the roundness and the area descriptors. However, there is no direct correlation between those variations and the experimental changes. Since no evident pattern seems to appear, a new approach based on Machine Learning algorithms will be described and used in the next chapter to evaluate its relevance in this context.
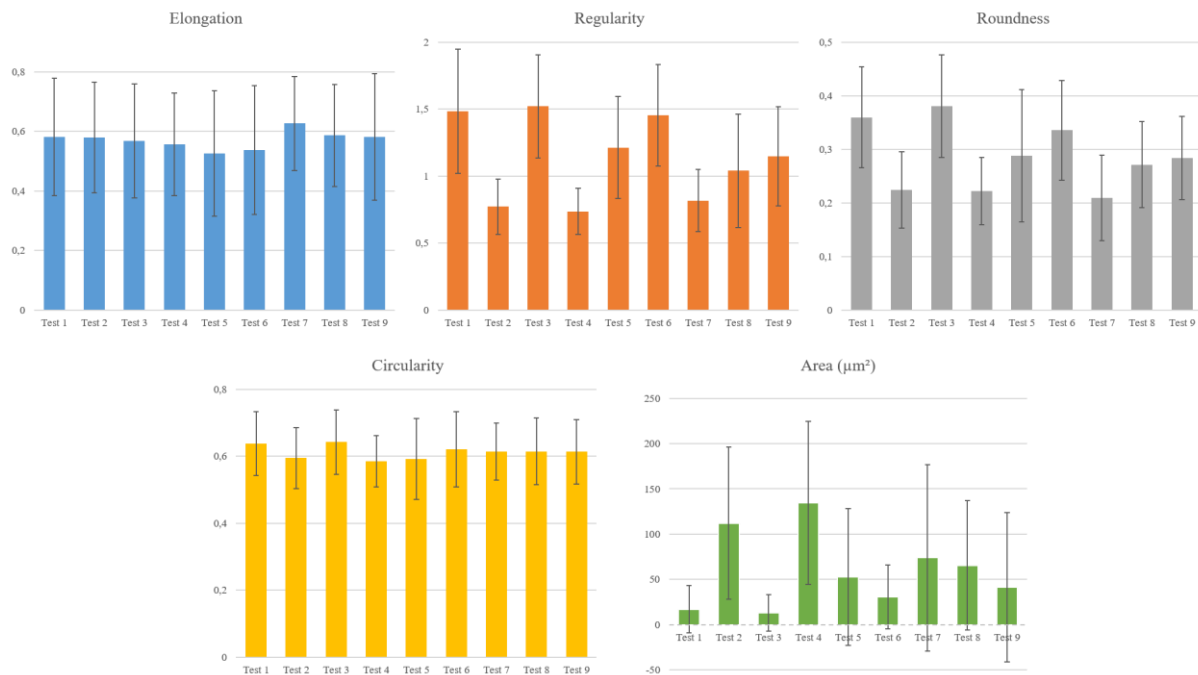
Figure IV-15: Morphological descriptors of the particles from all the tests conducted.

V. Learning from the morphological descriptors

Up to this point, we conducted nine different tests under three sets where experimental conditions changed each time. During each tribological test, we recorded different signals from which we calculated six rheological parameters to describe the test itself. It must be reminded that these parameters are not the conventional rheological quantities usually measured with a rheometer, but rather consist in a number of parameters that describe the flow of the third body particles. Along with the rheological parameters, we chose five morphological descriptors in order to characterize the external shape of the third body particles created during the tribological tests.

In this chapter, we evaluate the efficiency of machine learning algorithms, using the morphological information available, in:

- discriminating between the tests,
- predicting one the rheological parameters recorded. We start by describing the algorithms developed.

Then, we present the findings in each situation. Finally, we regroup all the results in the conclusion section.

1. Database structure

One of the crucial steps of setting up a machine-learning algorithm is the database and its elements. After finishing the experimental phase, we have six rheological measurements describing the nine tests and five morphological descriptors to characterize the third body particles created during the tribological tests. The database can make all the difference in the machine learning results. Therefore, we define the ones used during this project. **Table V-1** presents the exact number of particles studied from each test.

Table V-1: Number of analysed particles by test.

Test	1	2	3	4	5	6	7	8	9	Total
Number of particles	232	109	224	100	124	233	97	99	100	=1318

Since we have two different questions to answer using machine learning, it is necessary to have to different databases to use. A database that is created for machine learning purposes has two main type of data; input and output (in reference to the algorithms). The input data in this is the morphological database (1318 particles with 6 descriptors each). The output data will

however change depending on the objective we aim for. When we are trying to identify the test where each particle was created, the output database is the test's ID. For example, the 232 particles from test 1 will take the value 1. The 109 particles from test 2 will take the value 2, and so on. However, if the machine learning is used to predict the rheological parameter value, the output data will be the value of the rheological parameter for every particles.

For all the algorithms coded, the main database must be divided into two sub-bases: the training database and the test database. The training database is 75% of the main database, leaving the rest for the test database. From each test, we divide the particles in a random way between the two databases. We will remain in a supervised learning methodology, which means that each element of the database (particle) is defined by an input (the descriptors) and the corresponding output (e.g. the test ID where they were created).

2. Tribotest identification using the morphological data

In this thesis, we aim to study the relationship between the shape of the particles and the rheological conditions of the tribological tests. The algorithm needs to associate the particle (input) to a test (output) where it might have been created. Hence, we focus mainly on classification algorithms (known also as classifiers): logistic regression and neural networks. To judge the efficiency and the accuracy of an algorithm, we calculate the success rate: the percentage of the correctly predicted tests.

The machine learning algorithms were coded using Matlab, the same environment that was used for the image analysis work. They follow the flowchart presented in **Figure V-1**. Step 1 on the flowchart start by initializing the weights θ and then changing their values to minimize the cost function as much as possible. The algorithms adapt automatically to the size of both databases. However, certain parameters, which will be defined in the next sections, are left to be changed manually by the user to test their effect.
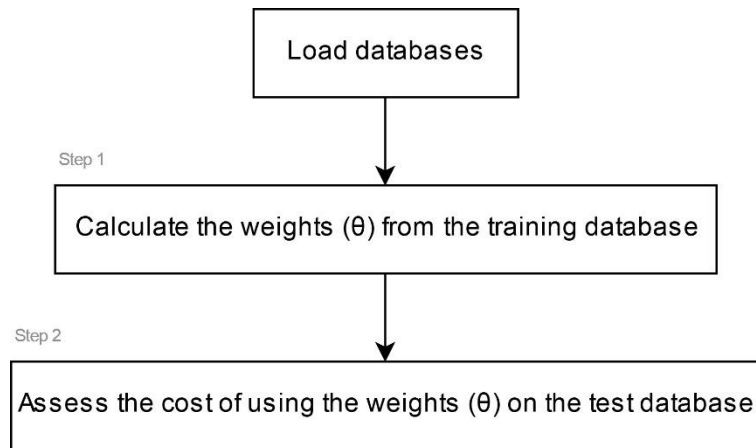
Figure V-1: General flowchart of the machine learning algorithms.

Both algorithms developed were tested using the MNIST database (LeCun and Cortes, 2010; Li Deng, 2012). This database is composed of pictures of handwritten digits and using the pixels forming the image of each digit the algorithm needs to identify it. Each image is a $20 * 20$ matrix (400 pixels) and the database had 5000 elements (500 image for each digit from 0 to 9). The accuracy of the algorithms is between 95% and 97.5% using both the logistic regression and the neural network.

2.1. Logistic regression

As explained in the first chapter, logistic regression algorithms are classifiers that predict to which class each input belongs, using their features. Since we have more than two tests (classes), then our classification is multiclass, and we put in place a **one-versus-all** methodology. The algorithms calculate the probability of the belonging of the particles to each test and we associate the one that has the highest score (probability).

In order to study the efficiency of the algorithms, two variables need to be studied, previously specified in Chapter I, section 2.3.2.1:

- The regularization parameter λ.
- The number of iterations in the gradient descent algorithm NI.

We use the morphological database described before in order to study if those descriptors alone are sufficient to predict the test where the particle was made. We start by inspecting the outcome of varying the algorithms' internal parameters λ and NI.

As we see in **Figure V-2**, we vary both parameters and compute the success rate of the algorithm when applied to the training database and the test database. Studying the different plots, we can see that increasing λ parameter does not always mean having better performing algorithms. Besides, the success rate functions hit a plateau when we keep increasing the NI. The combination $[\lambda = 0.01; NI = 500]$ guarantees the best success percentages during both the training phase (44%) and the test phase (35%). The success percentage means that the algorithm predicts correctly the test from which originates about on particles over three. This success rate is modest when compared to common applications of machine learning techniques, but remains way above a purely random assessment (it would be one successful prediction over nine). Since we have only 44% success rate on the training set, which is deemed low, it is interesting to investigate more in the results obtained.

In **Table V-2,** we present the success rate for each test from the two databases. Using this method showcases that the general success rate of 35% is not shared equally between the tests but it is actually an average. The algorithm identifies with good success particles from test 3 and test 6 but it does not perform as well on the other tests. Moreover, it is unable to identify any of the particles from test 8 and 9. In order to check if a certain test is dominating the classification procedure, we plot all the labels that the algorithms associate to every particle in the test database and the training database.

Table V-2: Success rate per class for morphological database for the logistic regression classifier..

	Training Database	Test Database
Test 1	34 %	19 %
Test 2	28 %	15 %
Test 3	76 %	73 %
Test 4	31 %	8 %
Test 5	10 %	6 %
Test 6	92 %	84 %
Test 7	39 %	32 %
Test 8	0	0
Test 9	0	0
	44 %	36 %
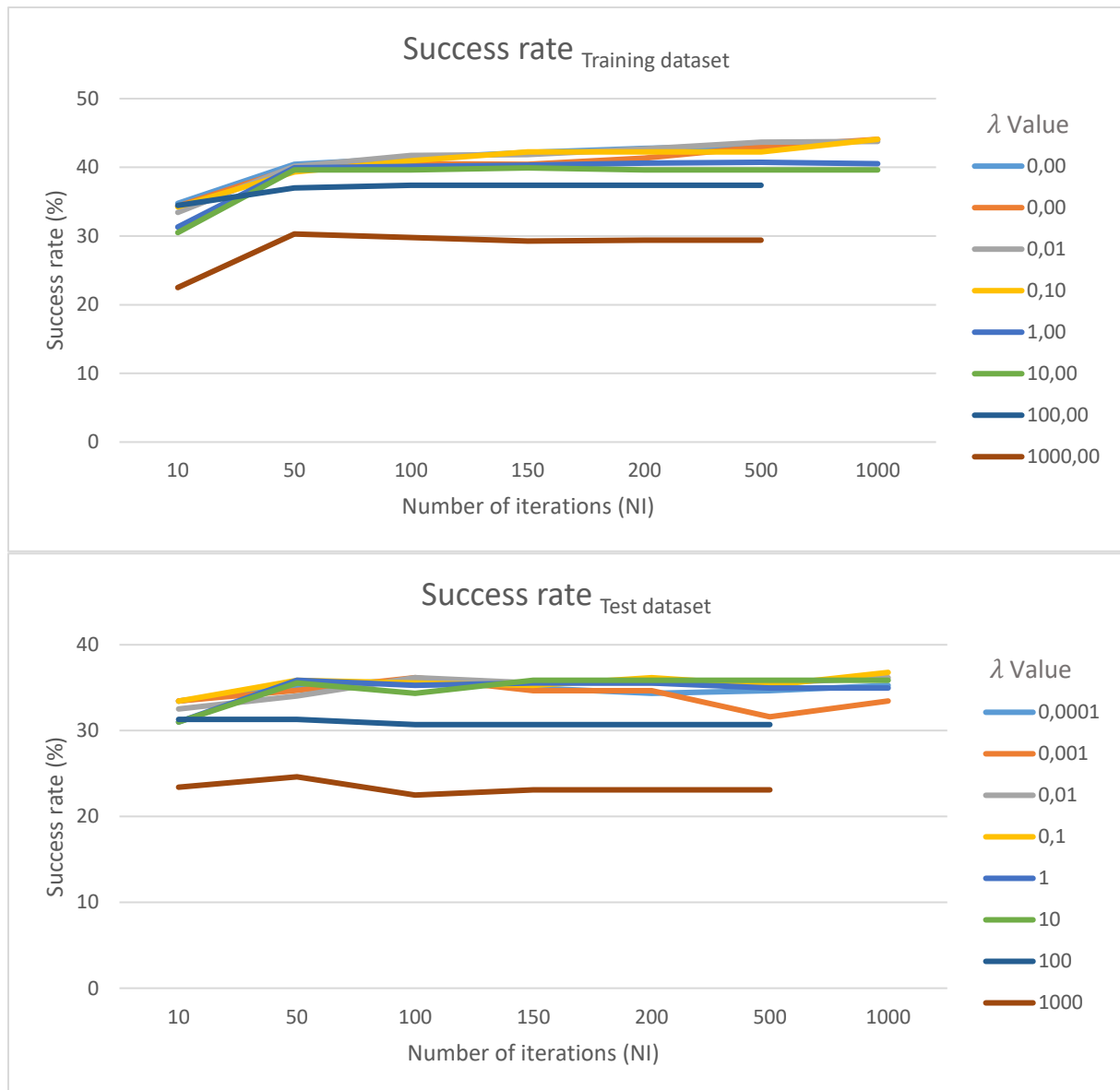
Figure V-2: Success rate of the logistic regression classifier when varying λ and NI.

In **Figure V-3**, we plot the tests attributed to the particles in both the training and the test databases. More specifically, each column corresponds to the particles originating from a given test, and the colours in this column indicate what test was attributed to these particles by the machine-learning algorithm. This is done for both the training and the test databases. Several remarks can be noted:

- Majority of particles from test 1 are mislabelled as particles from test 3, in both databases.
- In the test database, majority of particles from tests 4, 5, 7, 8 and 9 are mislabelled as particles from test 6.
- The percentages in **Table V-2** of tests 3 and 6 are confirmed since an important number of particles from both those test are correctly classified.
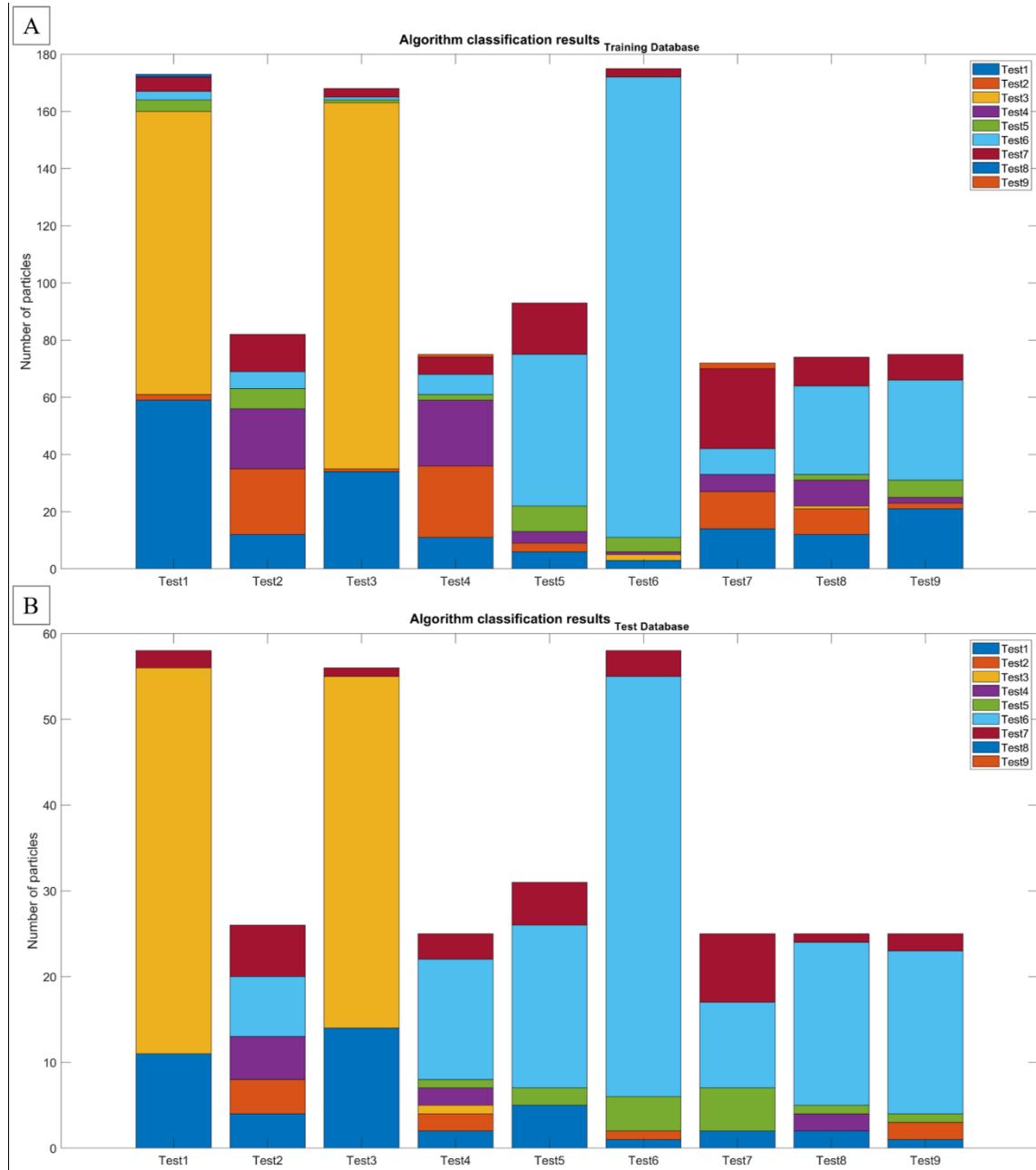
Figure V-3: The algorithm classification results per particle per test using: A) the morphological training dataset. B) from the morphological test dataset.

We are inclined to check if there is a correlation between those classification results and the five morphological descriptors used in this database. In **Figure V-4** and **Figure V-5**, we plot the probability scores calculated by the algorithms and the values of the descriptors of the particles respectively from the training and the test database. Those figures are plotted with the following protocol: the horizontal axis corresponds to all the particles in the database. They are first sorted by the number of the test which created them (leading to the nine sections dividing the x-axis). Then, they are sorted in each of these subsets by the score they received in their own test of origin (for example, the score that the algorithm gave to a test 4 particle for originating from test 4) in a decreasing order. This information is colour-coded, as well as the

score the particle got in the other eight tests. This makes the upper part of the figure. In the lower part follows the same horizontal sorting for all the particles, buy we provide the input values corresponding to the particles in terms of morphological descriptors.

Studying the results of applying the logistic regression classifier on the training dataset presented in **Figure V-4**, we can note some possible correlations between the morphological measurement and the test the algorithm associated to the particle:

- Particles labelled as particles generated during tests 1 and 3 have low regularity and high roundness measurements when compared to the rest of the tests.
- Particles identified from test 2 are particles with low regularity, low roundness and high area values.
- Particles classified as particles from test 6 have high regularity and high roundness levels.

It is interesting to notice that particles having a good score in test 1 also have a good one in test 3, and those scores are highly correlated. It is thus clear that the algorithm has difficulties to discriminate between these tests, based on the input database it is using. The same observation applies in a lesser extent to tests 2, 4 and 7, which also have similar profiles.

Examining the results of the classification on the test database shown in **Figure V-5**, we can notice that the algorithms tends to classify the particles with high regularity as particles generated in test 6. In addition, the particles that are labelled as particles from test 1 or test 3 are characterized by their low regularity and high roundness. The particles classified as particles from test 2 all share low regularity, high roundness and relatively high area measurements.

To estimate the effect of each of the five descriptors, we eliminate one at a time from the morphological database and evaluate the success rate of the algorithm. In **Figure V-6**, we display the results of this numerical experiment. We can see that the regularity parameter is the most important when using a logistic regression classifier since the success rate dropped significantly when removing it. It passed from 36 % when using all five descriptors to 27 % when removing the regularity descriptor from the morphological database. The 'WO' in the Y-axis on **Figure V-6** is an abbreviation for 'without'.
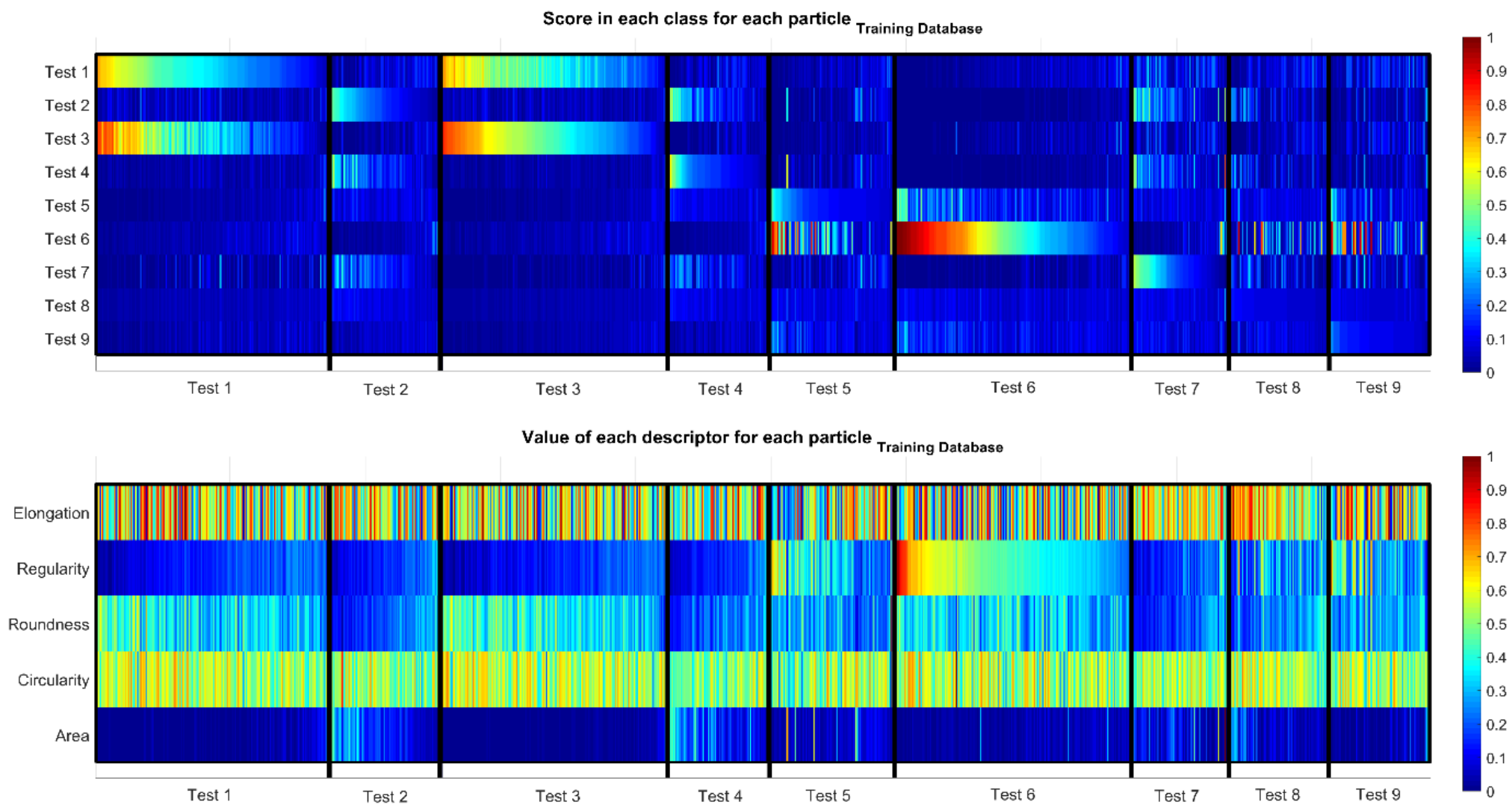
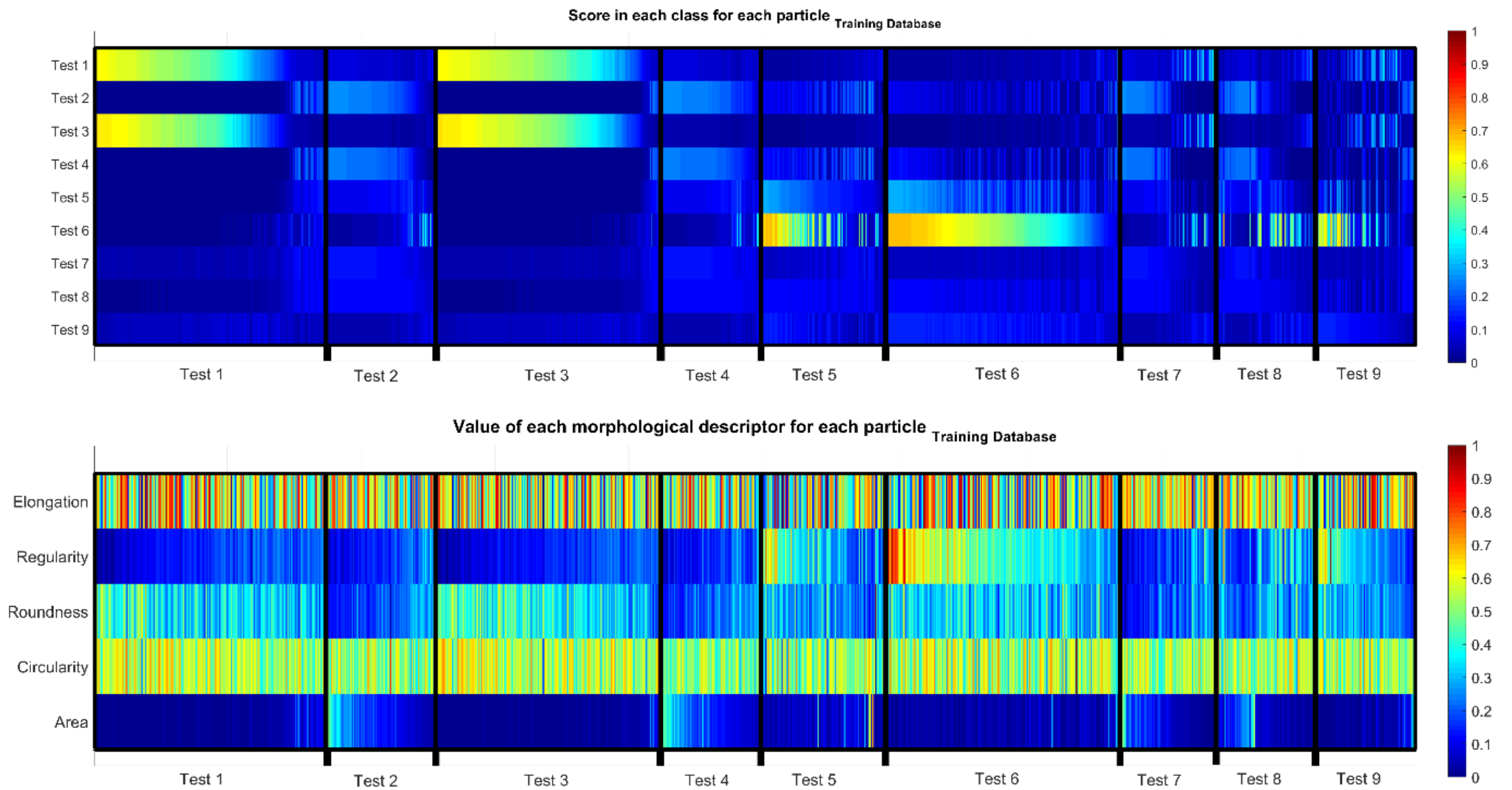*Figure V-4: Logistic regression classification score results and the morphological descriptors values for the particles from the training database.*
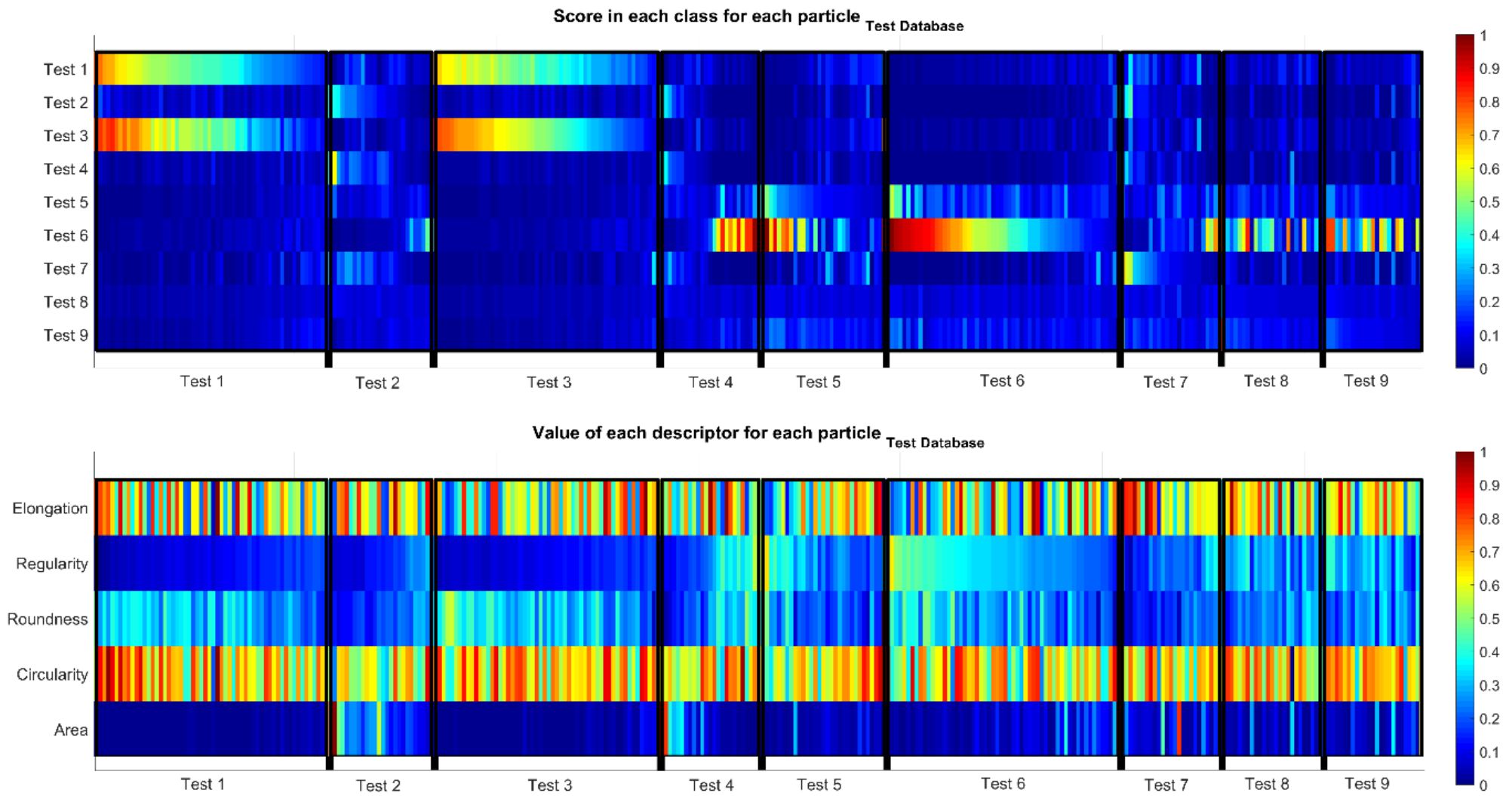
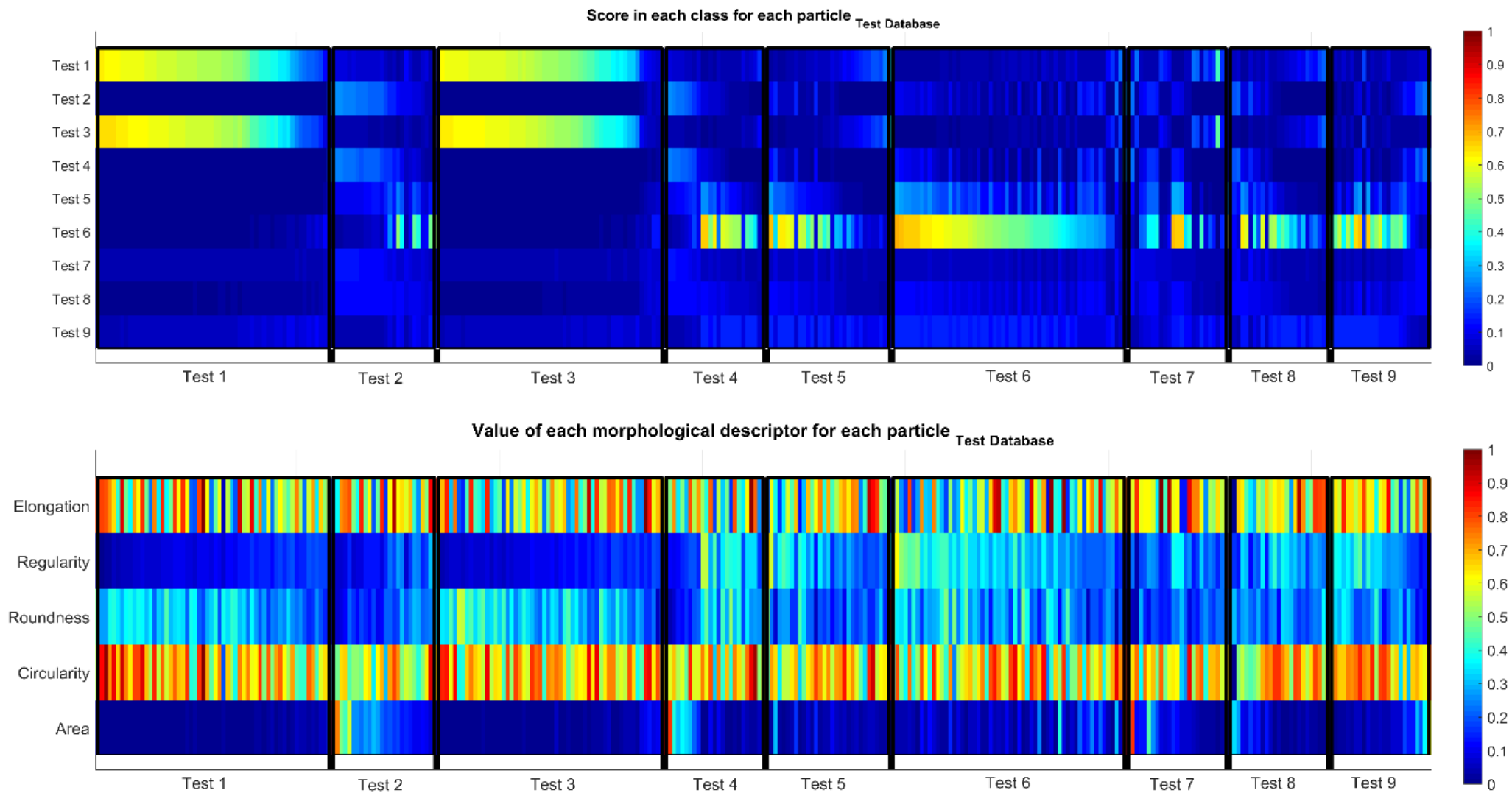*Figure V-5: Classification score results and the morphological descriptors values for the particles from the test database.*
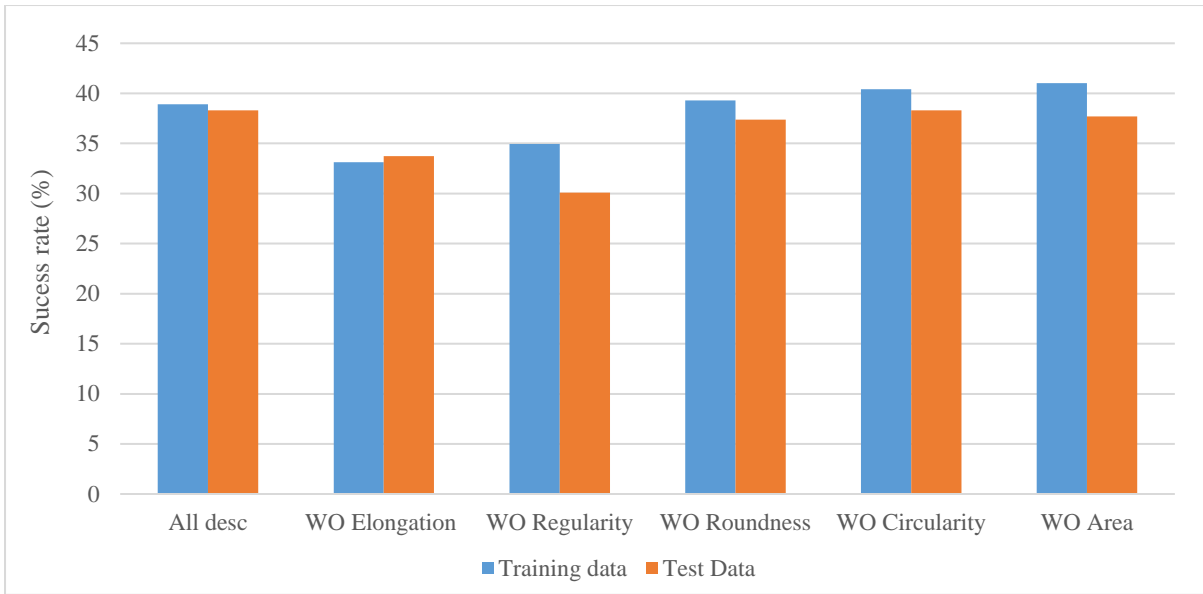
Figure V-6: Results of removing one descriptor from the morphological database using a logistic regression classifier.

2.2. Neural network

We test now the neural network classifier and its ability in predicting the test from where each particle came from, using only the morphological database. As mentioned in the first chapter, the neural network algorithm will calculate, like the logistic regression classifier, the probability for a particle to belong to each test. The test with the highest probability (score) will be assigned to the particle.

In a neural network algorithm, there are four parameters to be determined, described earlier in Chapter I, section 2.3.2.2:

- The regularization parameter λ.
- The number of iterations in the gradient descent algorithm NI.
- The Number of hidden layers NH.
- The size of each layer l(the number of nodes a_i explained more in the chapter I, section 2.3.2.2.) that is SH_l

The number of hidden layers has been a subject of many researches to enhance the efficiency of the neural network algorithms. It was proven that a single hidden layer network with a finite number of nodes is capable of approximating 'any function that maps one finite space to another'(Hornik, 1991; Alexeev, 2010). Therefore, we fix the number of hidden layers NH at 1. Adding more layers is kept as perspective for this work if deemed necessary. In addition,

since with the logistic regression fixing the NI to 500 gives the best results, we keep it at this level with the neural network. Therefore, we end up with two parameters to study.

We evaluate how efficient a neural network algorithm is when using the morphological database to classify the particles. We begin by focusing on the two parameters at hand for the algorithm λ and SH_l.

In **Figure V-7**, we display the results of varying the two studied parameters and the resulting success rates. Increasing the lambda λ, makes the algorithm better until a certain threshold is reached. As we see for $\lambda = 1$ or $\lambda = 100$ the success rate of the algorithm is much lower than when λ is smaller. However, when changing the size of the hidden layer SH_l, the success rate of the algorithm hits a plateau at a certain point (just after setting $SH_l = 5$).
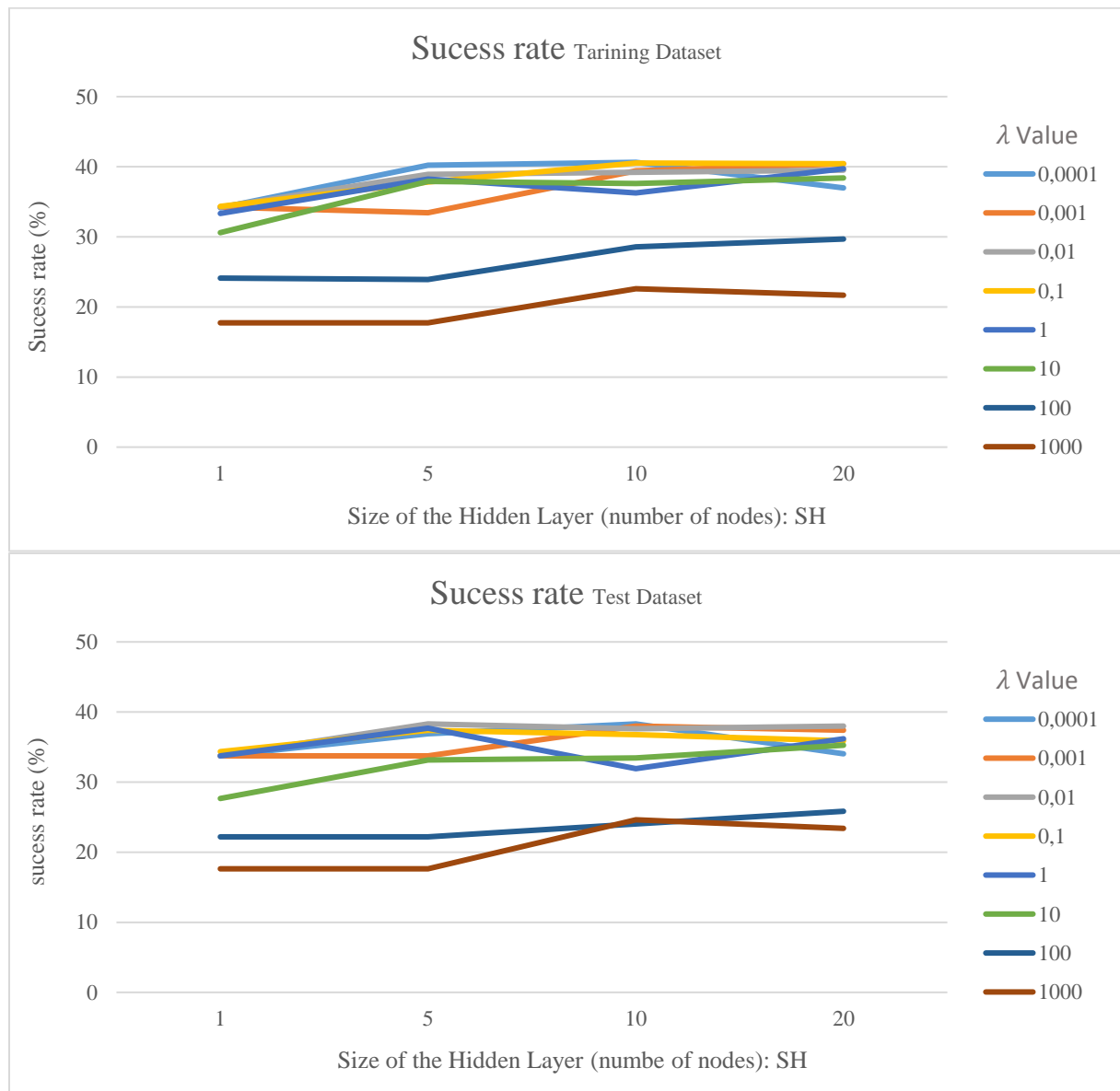
Figure V-7: Success rate of the neural network classifier when varying λ and SHl.

We choose the combination that ensures the highest success rate on the test database, which is [$\lambda = 0.01$; $SH_l = 5$]. This combination resulted into having 39 % and 38 % as success rates when applied on the training and the test databases respectively. At this point, we tried changing the number of iterations NI to verify if 500 was the value that gave us the best success rates on the test database. In **Figure V-8**, we plot the success rates when we only vary the NI and keep the lambda and the size of the hidden layer at the values that we chose earlier. Even though, the algorithm performs better on the training database when Ni is equal to 1000 and more, the success rate of the algorithm on the test database declines. The combination [$NI = 500$; $\lambda = 0.0$; $SH_l = 5$] ensures 39 % and 38 % as success rates when the algorithm is applied on the training and the test morphological databases respectively.
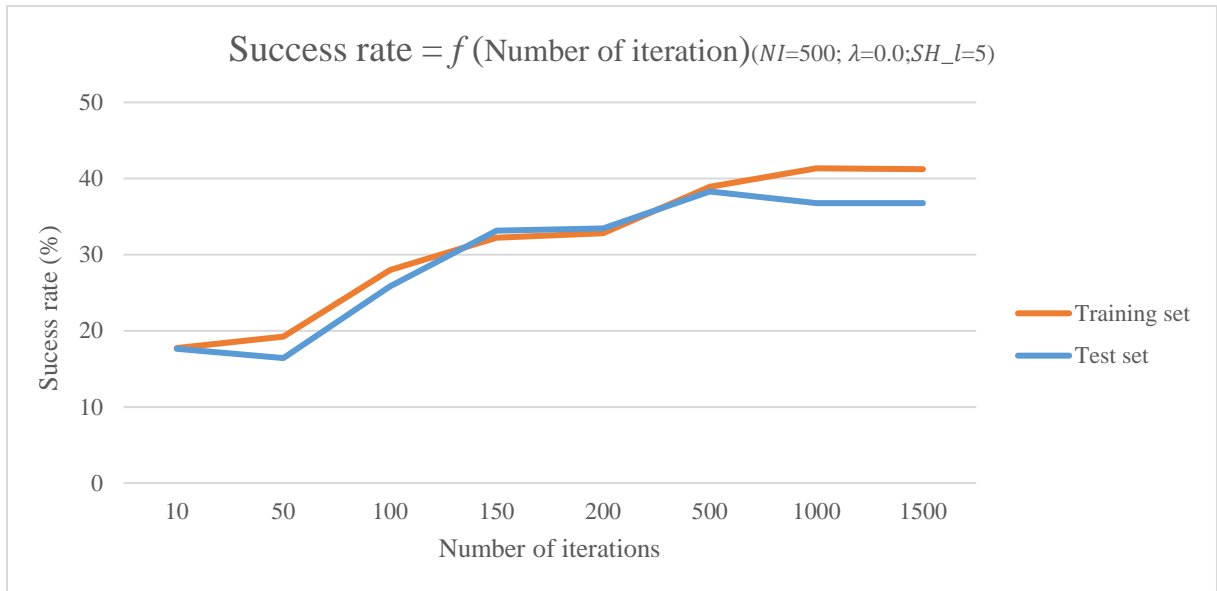
Figure V-8: Success rate of the neural network when changing the number of iteration NI.

Just like with the logistic regression classifier, we study in depth the results of classification using the neural network. **Table V-3** details the success rates for every class. The algorithm is able to classify correctly the majority of the particles from test 2, 3 and 6. However, it is unable to identify any particles from tests 5, 7, 8 and 9.

	Training Database	Test Database
Test 1	28 %	17 %
Test 2	73 %	58 %
Test 3	63 %	79 %
Test 4	4 %	8 %
Test 5	0	0
Test 6	95 %	95 %
Test 7	0	0
Test 8	0	0
Test 9	0	0
	39 %	38 %

In **Figure V-9**, we plot how the algorithm classifies the particles from each test. We can notice that:

- The algorithm mislabels particles from test 1 as if they were generated in test 3.
- The algorithm mislabels particles from tests 4, 5, 7, 8 and 9 as particles from test 6.

To investigate more about those results, we plot the scores of the algorithm associated to every particle from the training database (**Figure V-10**) and the test database (**Figure V-11**). Studying those two figures, we can assume that the algorithm associates some characteristics to identify the particles. We can see that the algorithm labels particles with low regularity, low roundness and low area measures as particles from test 2. Particles with low regularity and high roundness are labelled as particles from test 3. The classifier designates particles with high regularity and high roundness as particles from test 6.
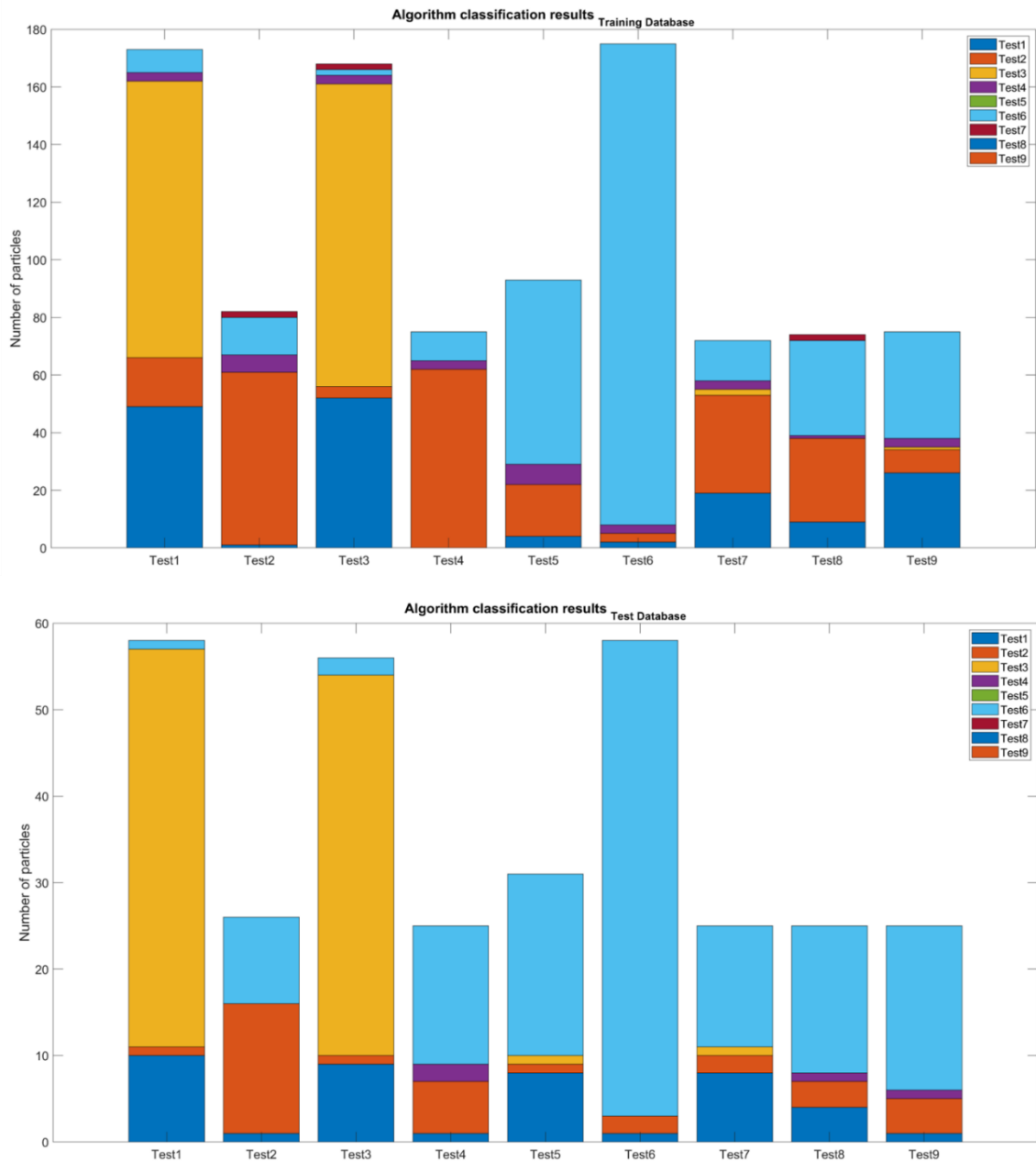
Figure V-9: The neural network algorithm classification results per particle per test using the morphological database.

The same test, as with the logistic regression classifier, is performed with the neural network algorithm to evaluate the effect of each descriptor in the morphological database. In **Figure V-12** we can see that the elongation and the regularity descriptors are more important than the others since their effect is more noticeable. The 'WO' in the Y-axis on **Figure V-12** is an abbreviation for 'without'.

Figure V-10: Neural neurone classification score results and the morphological descriptors values for the particles from the training database

Figure V-11: Neural neurone classification score results and the morphological descriptors values for the particles from the test database.

Figure V-12: Results of removing one descriptor from the morphological database using a neural network classifier.

3. Relating the morphological descriptors with the rheological data

Until this point, the output of the machine-learning algorithm was the number of the test from where the particle was generated. In this section, we test if the algorithm is capable of predicting the rheological measurements from the morphological descriptors of the particles. This changes the nature of our problem because we no longer have a classification objective but a regression one. More specifically, we are not asking the algorithms to predict a class to which the grains may belong to, but a numerical quantity. The database contains all the particles from the tests. It is divided into two sub-bases: the training (75 %) and the test database (25 %). The input part of the databases is the five morphological measurements of the database. For the output, we choose to use the RD_1 (The average of the friction coefficient after the established regime) but the same work can be done for the other rheological measurements (Appendix 2).

3.1. Linear regression

Unlike the logistic regression, where the algorithm computes the output from a pre-known set of individual values (classes), in linear regression the output is a continuous signal. In our work, the machine-learning algorithm using linear regression calculates an estimation of one of the rheological measurements (continuous signal) directly through the hypothesis function. Whereas the logistic regression algorithm that calculates a probability of belonging to each on of the nine tests (finite set of values) of the particle based on its morphological data and assigns the test that has the highest probability. The hypothesis function for the linear regression algorithm is the following:

$$h_\theta(X) = [\theta_0 \quad \theta_1 \quad \dots \quad \theta_n] \begin{bmatrix} x_0 \\ x_1 \\ \vdots \\ x_n \end{bmatrix} = \theta * X$$

n is the number of features (descriptors) used.

To evaluate how the weights (θ) are performing we put in place the following regularized cost function:

$$J_\theta = \frac{1}{2m} \sum_{i=1}^{m} (\hat{y}^{(i)} - y^{(i)})^2 + \lambda \sum_{j=1}^{n} \theta_j^2$$

m is the number of examples in the database

In addition, we use the gradient descent technique to minimise the cost function by correcting the weights of the hypothesis function.

The main difficulty we face with linear regression algorithms how to quantify its success at predicting the output wanted. In a classification problem, it consists in a binary comparison between the predicted output and the correct one. For the regression algorithm, we calculate the average of the absolute value of the difference between the predicted RD_1 and its correct value. This parameter is named 'average prediction error' (APE). The lower the value of the APE the better the algorithm is performing.

In **Table V-4,** we remind the values of the RD_1 for every test. We remind that the particles from one test all have the same values of the rheological measurements since the microscopic images were taken after the tests were stopped.

Table V-4: The RD_1 values for the different tests.

	RD_1
Test 1	0,44
Test 2	0,54
Test 3	0,49
Test 4	0,26
Test 5	0,43
Test 6	0,35
Test 7	0,29
Test 8	0,38
Test 9	0,2

In **Figure V-13**, we plot the results of the regression algorithm predicting the value of RD_1 from the five morphological descriptors of the particles. We varied the λ value and printed the average prediction error for both the training and the test databases. The peaks we notice in **Figure V-13** in the values of the average prediction error when $\lambda = 0.001$ or when $\lambda = 0.004$ are not due to a singular error that made the APE value rise, but the individual APE is fluctuating significantly higher for those two values of λ, as presented in **Figure V-15**. As we can see in **Figure V-14**, where we plot the real values of RD₁ and the values predicted when using $\lambda = 0.001$ (orange line) and $\lambda = 1$ (grey line), the values making the orange plot are more fluctuant

than those making the grey one, which explains the high value of their average in comparison to the other values of λ. This applies also to explain the peak when $\lambda = 0.004$.

Just as was the case with the classifier, this general indicator does not work equally well. **Table V-5** contains APE detailed for every test when varying λ. **Figure V-15** provide a visual representation of the contents of the training and the test database results respectively. We can note that the algorithm performs best on tests 1, 6 and 8. However, tests 2, 4 and 9 have the highest average prediction errors that means they are not well predicted by the algorithm. Those conclusions do not change when varying the λ value even though the average prediction error changes in value.

The linear regression algorithm kept predicting well the particles from test 6 like when we used neural network or logistic regression. However, even though it did not predict successfully the same tests as in the ones before, we always have a minimum of 3 three tests well predicted.
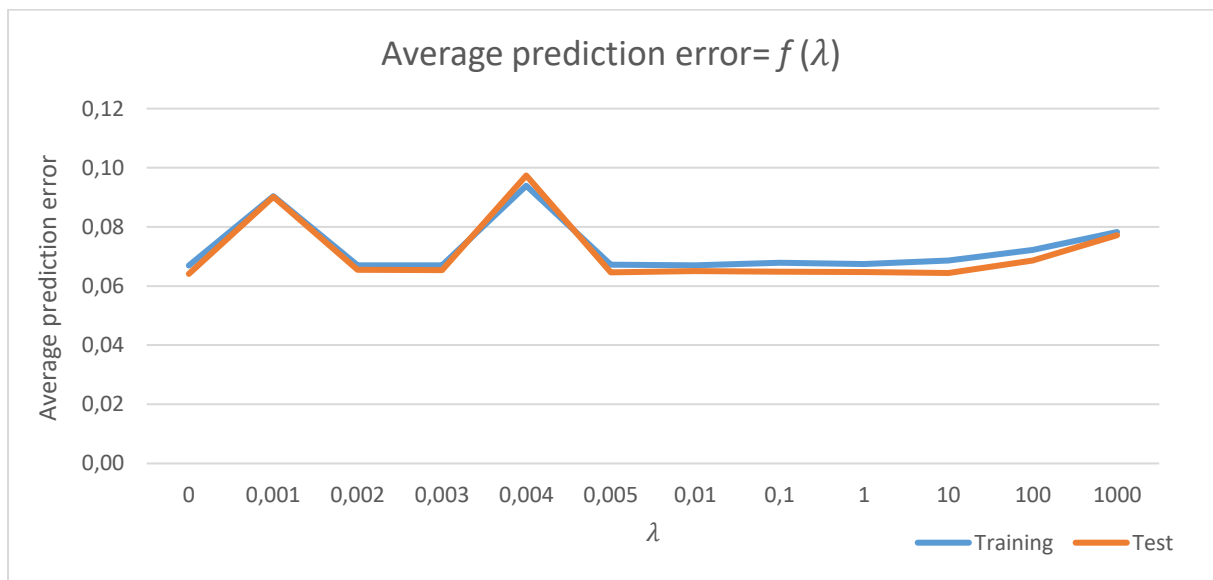
Figure V-13: The evolution of the APE when varying the values of λ when using the linear regression technic.
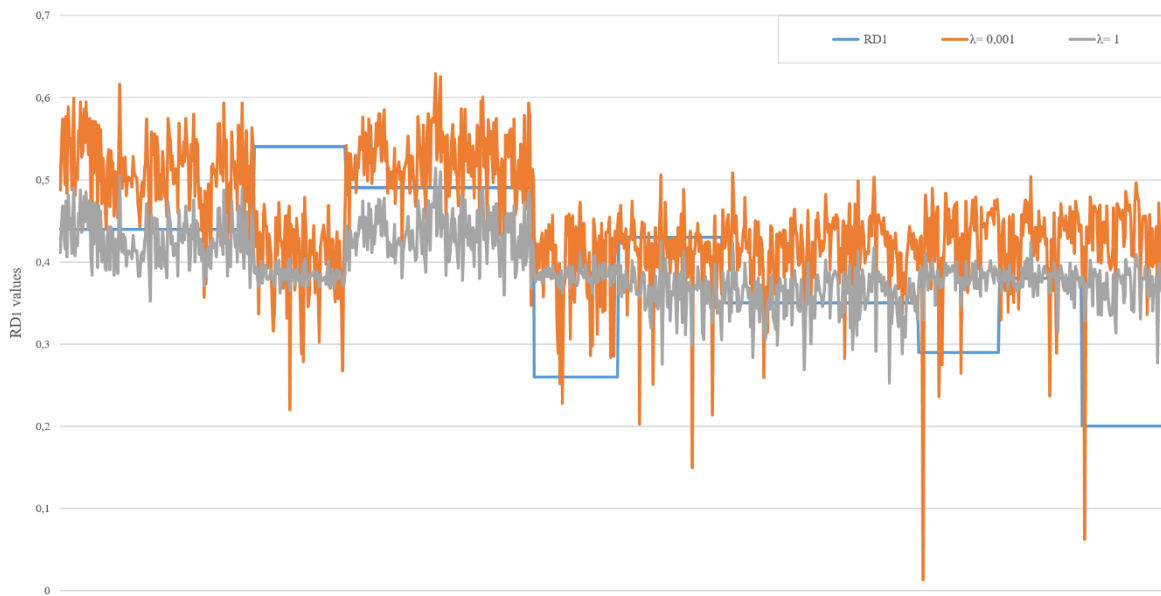
Figure V-14: the effect of changing the λ value for the linear regression algorithm.

Table V-5: Individual APE for each test when varying the values of λ when using the linear regression technic.

		λ											
		0	0,001	0,002	0,003	0,004	0,005	0,01	0,1	1	10	100	1000
Training Database	Test 1	0,03	0,07	0,03	0,03	0,09	0,03	0,03	0,03	0,03	0,02	0,03	0,04
	Test 2	0,16	0,14	0,16	0,16	0,18	0,16	0,16	0,16	0,16	0,15	0,15	0,16
	Test 3	0,05	0,04	0,05	0,05	0,14	0,05	0,05	0,05	0,06	0,06	0,08	0,09
	Test 4	0,13	0,14	0,12	0,12	0,1	0,13	0,12	0,13	0,13	0,13	0,13	0,12
	Test 5	0,07	0,03	0,07	0,07	0,06	0,07	0,07	0,07	0,07	0,06	0,05	0,04
	Test 6	0,03	0,07	0,02	0,02	0,04	0,02	0,02	0,03	0,02	0,02	0,03	0,04
	Test 7	0,09	0,14	0,09	0,09	0,07	0,09	0,09	0,09	0,09	0,1	0,1	0,1
	Test 8	0,01	0,06	0,01	0,01	0,02	0,01	0,01	0,02	0,01	0,01	0,01	0,01
	Test 9	0,17	0,23	0,17	0,17	0,17	0,17	0,17	0,17	0,17	0,18	0,19	0,19
Test database	Test 1	0,03	0,08	0,03	0,03	0,1	0,03	0,03	0,03	0,03	0,02	0,02	0,04
	Test 2	0,16	0,13	0,17	0,17	0,19	0,17	0,17	0,17	0,16	0,16	0,16	0,16
	Test 3	0,05	0,04	0,05	0,05	0,15	0,05	0,05	0,05	0,06	0,06	0,08	0,09
	Test 4	0,11	0,16	0,11	0,12	0,12	0,11	0,12	0,11	0,11	0,11	0,12	0,13
	Test 5	0,08	0,04	0,08	0,08	0,06	0,08	0,08	0,08	0,08	0,06	0,05	0,04
	Test 6	0,02	0,07	0,02	0,02	0,03	0,02	0,02	0,02	0,02	0,02	0,02	0,04
	Test 7	0,08	0,14	0,07	0,07	0,08	0,08	0,08	0,07	0,08	0,09	0,1	0,11
	Test 8	0,03	0,06	0,03	0,03	0,02	0,03	0,02	0,03	0,03	0,02	0,01	0,01
	Test 9	0,17	0,23	0,17	0,17	0,19	0,17	0,17	0,17	0,17	0,18	0,19	0,2

The colour of the cell is darker when the value it contains is higher than the minimum (0,01).
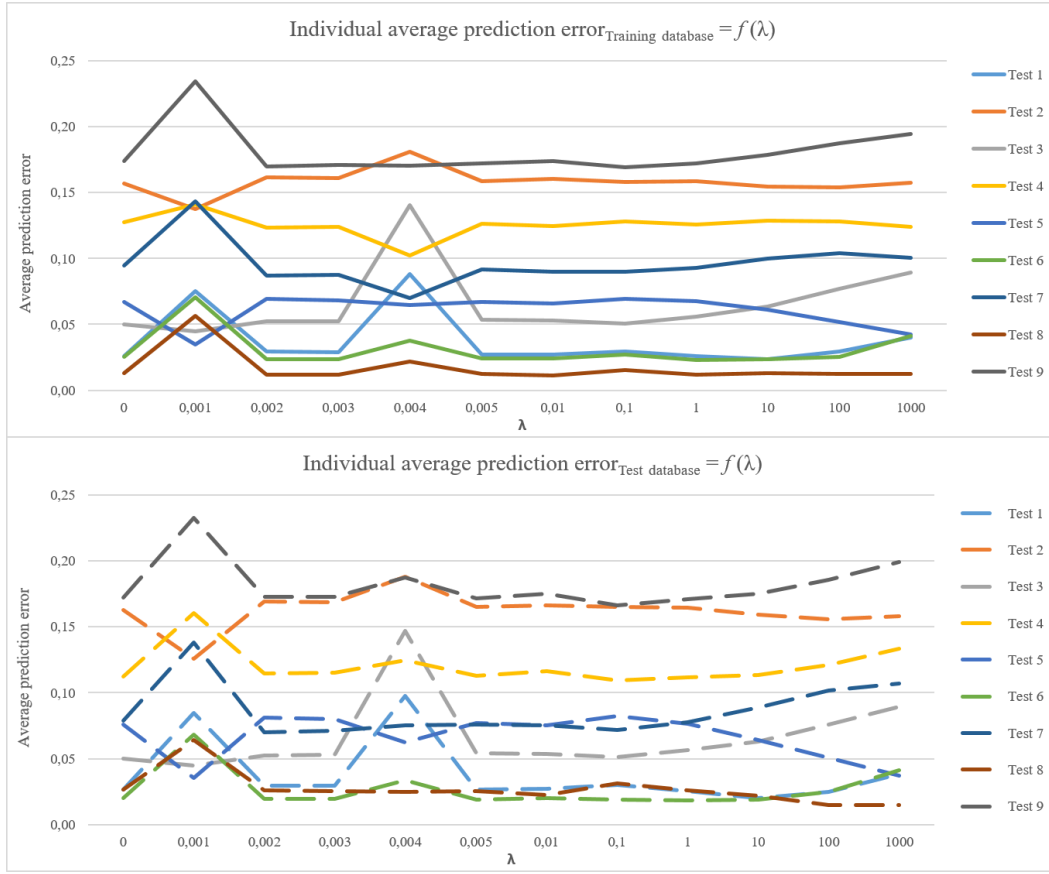
Figure V-15: Individual APE for each test from the training and the test databases when using the linear regression technic.

3.2. Polynomial regression

In order to improve the results of the regression algorithm, we change the hypothesis function from its linear form to a richer polynomial equation that may be written under the following form:

$$h_\theta = \sum_{i=1}^{p} \theta_i X^i$$

p is the polynomial order of the equation.

The same database is used using with this hypothesis function. The input is the five morphological descriptors that describe the shape of the particle. The values of the input vector are normalized when they pass the pth power. It is simply by averaging the values $\left(\frac{X - mean\,(X)}{std(X)}\right)$. It is a necessary step to keep all the input data in the same range when passing them to a bigger power in order for them not to dominate the hypothesis function.

In **Table V-6** and **Table V-7**, we present the average prediction error for the different values of λ and p. To study more deeply the results we present them in **Figure V-16** and **Figure V-17**. We can notice that there is no real improvement in transforming the hypothesis equation into a polynomial function. However, the only benefit of using the polynomial regression is correcting the results when $\lambda = 0.001$ and $\lambda = 0.004$. It is

noticeable that the two peaks that were present in the le linear regression ($p = 1$) are not visible anymore. It is also interesting to notice that the results does not improve (in some cases worsen) the higher the degree of polynomial is.

Table V-6: APE for the polynomial regression algorithm for each value of λ and p applied on the training database.

P \ λ	0	0,001	0,002	0,003	0,004	0,005	0,01	0,1	1	10	100	1000
1	0,067	0,090	0,067	0,067	0,094	0,067	0,067	0,068	0,067	0,069	0,072	0,078
2	0,066	0,066	0,066	0,066	0,066	0,066	0,066	0,066	0,066	0,067	0,068	0,070
3	0,066	0,066	0,066	0,066	0,066	0,066	0,066	0,066	0,066	0,066	0,067	0,070
4	0,066	0,066	0,066	0,066	0,066	0,066	0,066	0,066	0,066	0,066	0,067	0,070
5	0,066	0,066	0,066	0,066	0,066	0,066	0,066	0,066	0,066	0,066	0,067	0,070
10	0,066	0,066	0,066	0,066	0,066	0,066	0,066	0,066	0,066	0,066	0,067	0,070

Table V-7: APE for the polynomial regression algorithm for each value of λ and p applied on the test database.

P \ λ	0	0,001	0,002	0,003	0,004	0,005	0,01	0,1	1	10	100	1000
1	0,064	0,090	0,066	0,065	0,097	0,065	0,065	0,065	0,065	0,064	0,069	0,077
2	0,065	0,065	0,065	0,065	0,065	0,065	0,065	0,065	0,065	0,065	0,066	0,069
3	0,065	0,065	0,065	0,065	0,066	0,065	0,065	0,065	0,065	0,064	0,066	0,070
4	0,065	0,065	0,065	0,065	0,065	0,065	0,065	0,065	0,065	0,065	0,065	0,070
5	0,065	0,065	0,065	0,065	0,065	0,065	0,065	0,065	0,065	0,065	0,065	0,070
10	0,091	0,094	0,094	0,090	0,098	0,092	0,090	0,092	0,091	0,085	0,067	0,071
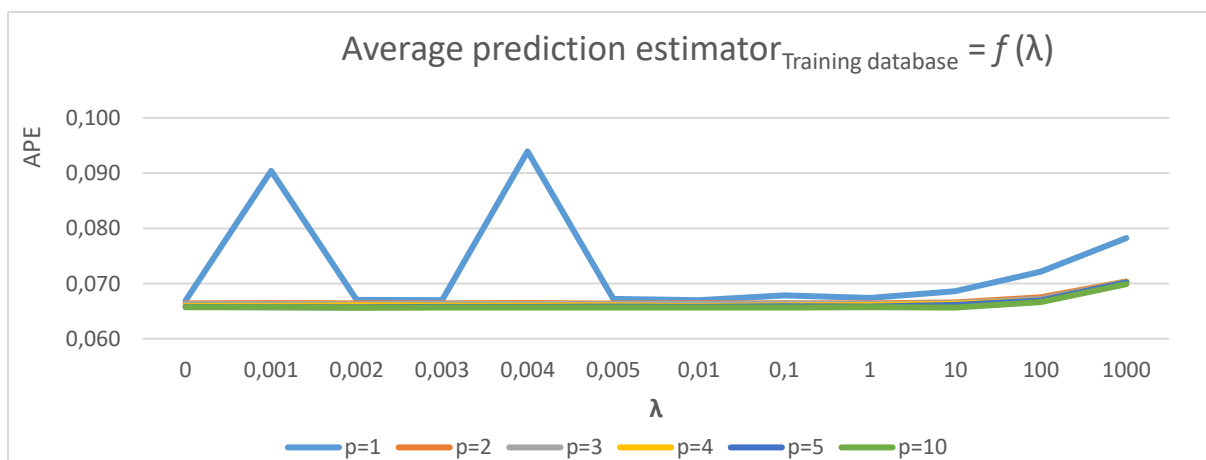
Figure V-16: Plot of the APE values when varying the λ and the p when the polynomial regression algorithm is applied on the training database.
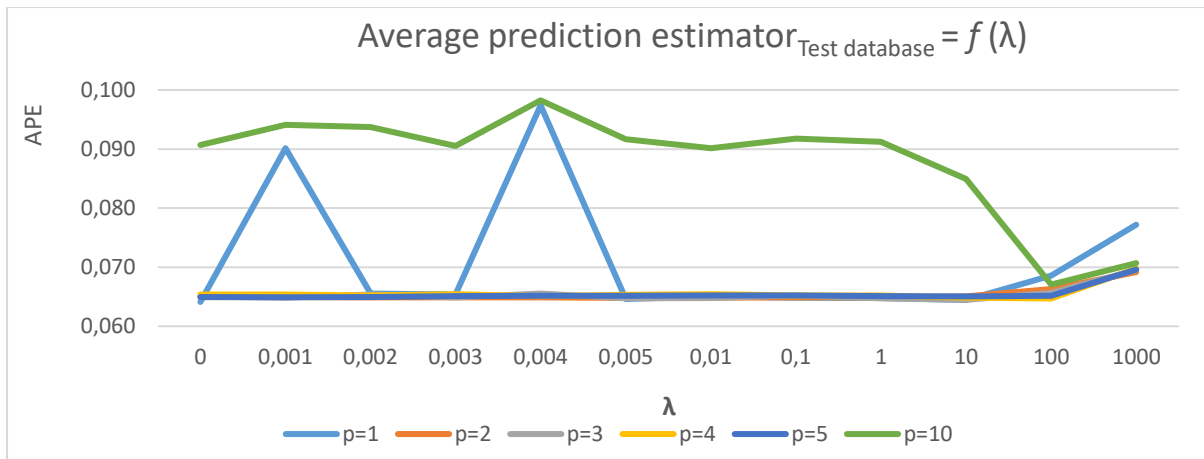
Figure V-17: Plot of the APE values when varying the λ and the p when the polynomial regression algorithm is applied on the test database.

4. Conclusion

We proposed in this chapter different machine learning algorithms to help identify the test where each third body particles was generated. This identification method was based on a database of morphological descriptors to characterize each particle and rheological measurements of the conducted tribotests.

In **Table V-8**, we summarize the different success rates for each of the machine learning algorithms when applied on both databases. If we ask a person who is not an experienced tribologist, he will have a success rate equal to 11 % (one chance in nine to guess randomly the test). We prove that using machine-learning classifier at least triples that success rate.

Table V-8: Summary of the Machine Learning algorithms used.

	logistic regression		Neural network	
	Training Database	**Test Database**	**Training Database**	**Test Database**
Morphological input database	44 %	35 %	39 %	38 %

Using a logistic regression algorithm, we obtained a 36 % success rate on a test database when using a morphological database. The generalised 36 % success rate is not shared between the tests. The logistic regression classifier distinguishes successfully particles from test 3 and 6 yet it mislabels the majority of the particles from the other tests. The algorithm used mostly the regularity and the roundness for its classification: Particles from test 3 have low regularity and high roundness and Particles from test 6 have high regularity and high roundness.

We put in place a second classifier based on artificial neural network to compare its performance with the logistic regression algorithm. The neural network classifier resulted in a 38 % success rate when a morphological test database is used. Similar to the logistic regression algorithm, this general percentage is not equally shared between the particles. The algorithm successfully labels particles from tests 2, 3 and 6 and

totally mislabels the particles from the rest of the tests. Accordingly, the neural network classifier gives slightly better results than the logistic regression counterpart did. It mostly uses three descriptors in order to obtain those results. The same relationship between the roundness and the regularity of the particles that was discovered when working with Logistic Regression was found. However, particles from test 2 were distinguished due to their low regularity, low roundness and low surface calculations.

The descriptors that seem successful at differentiating between particles are the regularity and the roundness. The regularity descriptor effect is evaluated immediately when removing it from the database and the considerable loss we had in the success rate when using both algorithms (**Figure V-6** and **Figure V-12**). The effect of roundness effect is seen when investigating the results of the classification. Those two descriptors will have to be investigated more deeply but they are essential for our work.

Both elongation and circularity were not exploited by the algorithms to differentiate between the tests. If we take a deep look at those descriptors, we find that they vary more than the other descriptors between the particles. No pattern could have been found in our work, however this is not a sufficient reason enough to eliminate those two descriptors from future databases. A proper investigation for the effect of each descriptor must be held in order to know those which could be ignored.

To assess the success of an algorithm we use, its results on the tests database (the unknown database). Based on that quantity, the neural network classifier performed better than the logistic regression classifier when applied on the Morphological database but only by 3%. Such a small difference does not justify eliminating one algorithm.

As a second part of this chapter, we present the results of coupling the morphological database with one of the rheological descriptors. This means that we trained an algorithm to predict the values of the rheological parameters from the morphological database. The results from this attempt were not as easy to interpret as the results from the first part. It is no longer a binary result (test correctly identified or not) but we have two continuous signals to compare between. The predicted value of RD_1 evolves around the average of the real value that we want predicted, as seen in the example shown in **Figure V-14**. We chose to evaluate how correct the algorithms results by using the average prediction error ($APE = |predicted\ value - real\ value|$). Using this evaluation parameter, the results of this part are not as promising as those obtained in the first part of this chapter.

. In **Appendix 2**, we present the results of using the other rheological measurements instead of just the RD_1. The results do not change much and the algorithm kept providing its predictions around the average value of that particular rheological parameter.

VI. General conclusion and discussion

In this document, we present the work accomplished during this thesis. It is a preliminary effort in order to build a system that might someday be able to predict tirobological failure in an industrial context based on the wear debris created. Only expert tribologists can deduct from a visual inspection of the third body particles the contact conditions where they were created. We suggest a proof of concept that machine-learning algorithms can succeed in predicting from which test a third body particle was created using a set of descriptors.

In the worst-case scenario, the machine learning algorithms tripled the success rates of a random guess when applied on the databases provided. Encouraged by the classification results, we tried predicting one of the rheological measurements RD using the morphological descriptors of the particles. We trained a linear regression algorithm for this purpose. The algorithm kept predicting a value around the average of the rheological data set we are predicting. This is explained by the lack of variability in the values of each rheological measurement between the particles of the same test.

Even though this work proves that using machine-learning algorithms to predict the test where a particle was created is possible, we cannot yet chose one algorithm over the other due to the very small difference in success rates between the two, and many areas should be inspected and explored in a thorough manner.

The neural networks used were fairly simple since they contained only one hidden layer. In future work, we can test the efficiency of increasing the number of hidden layers alongside changing the values of other parameters$(\lambda; NI; SH_l)$ of the algorithm. In addition, our database used for training the classifiers and testing their efficiency is not made of an equal number of particles between the tests. Fixing a constant number of particles to be analysed from every test should be a priority.

One other area to investigate more is the different morphological descriptors and the rheological measurements themselves. When the accelerometer was installed during the tribological experiments, we supervised the movements of the pin, which can be very interesting to study. Accelerometer signals are widely used nowadays especially to feed machine learning algorithms. They succeeded at predicting earthquakes in a simulated environment (Rouet-Leduc *et al.*, 2017) and at sprint assessment for athletes (Gurchiek *et al.*, 2019). In **Figure VI-1**, we can see the accelerometer recordings from a test in air and one in argon. There are many features that can be analysed, such as the change in the level of the vibrations between the two signals during the stabilisation phase (after the $10\ m$ milestone), the difference between the movements recorded for the three axis of the pin, etc. In addition, the real time video recordings should be examined and be explored more deeply. A second camera can be installed to record the entry of the contact or the evolution of the track itself away from contact. Additionally, other lenses may be tested in order to zoom in more on certain regions and to monitor how the particles behave during the experiment.

The real time recorded COF (tangential force) signal can be examined with more details for more accurate rheological measurements. Calculating the average of the COF may not be the best descriptor of the real time

COF recordings. We can see in **Figure VI-2** that the average of the average of the COF signal decreases at the end of the experiment but this is due to a decline in the instant peaks values found in the real time recordings, not a continuous change in the whole interface behaviour.

the accuracy of the rheological measurements that are based on visual inspection needs to be evaluated and examined like the case with the diameter of the contact area on the pin and the width of the sliding track on the disk. A more systematic and quantitative way to determine them would bring a better repeatability to the analysis.
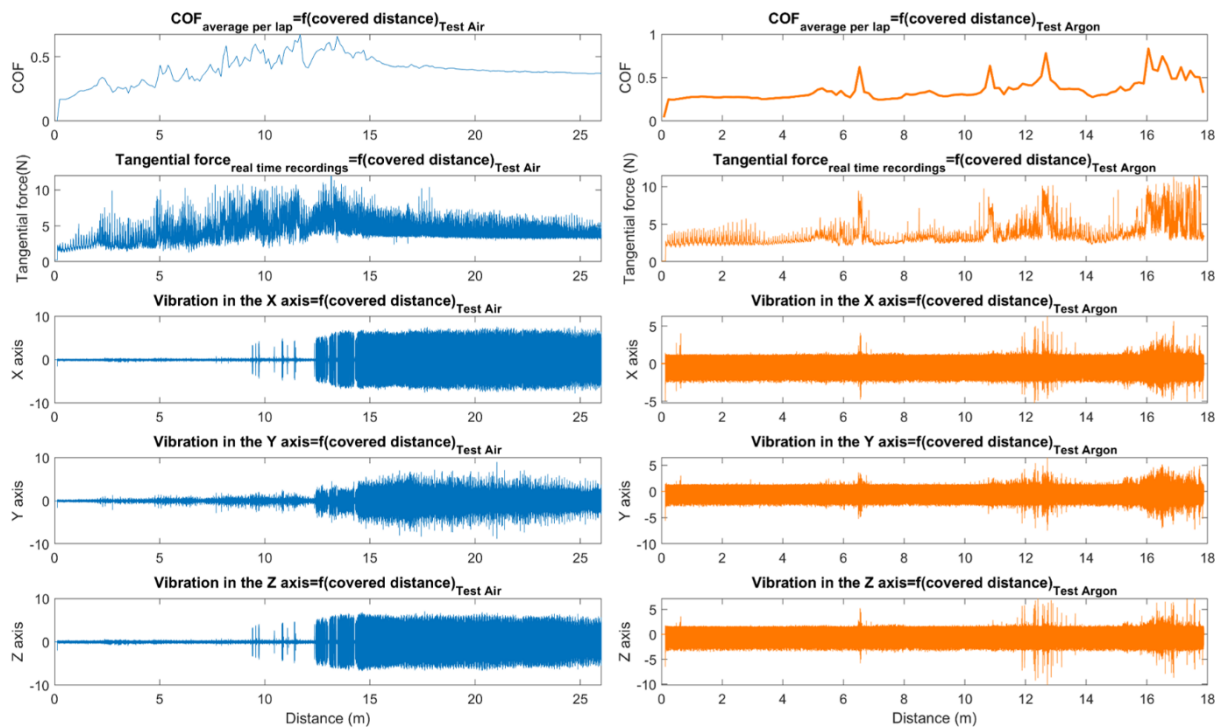
Figure VI-1: Difference between the real time recordings between a test in Air (Blue signal) and a test in Argon (Orange signal).
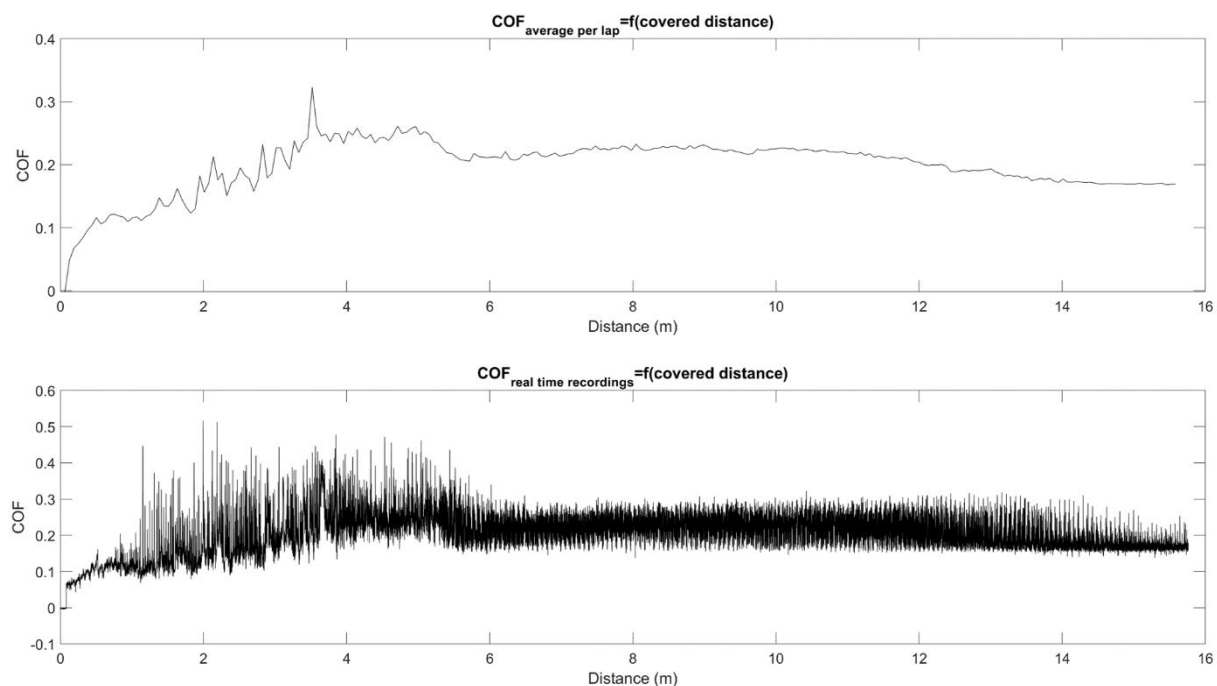
Figure VI-2: the average per lap VS the real time recordings of the COF signal.

The morphological descriptors we chose until now only describe the 2*D* projection of the particles from the microscopic images. However, those images are not always sufficient to describe the 3*D* nature of the third body particles. We can see in **Figure VI-2** the difference between the images of the same particle with a 11° tilt difference. Almost all the five morphological result are different. In **Table VI-1,** we summarize the set of morphological result for both microscopic images. Adding a descriptor to take into account the missing third dimension is mandatory.
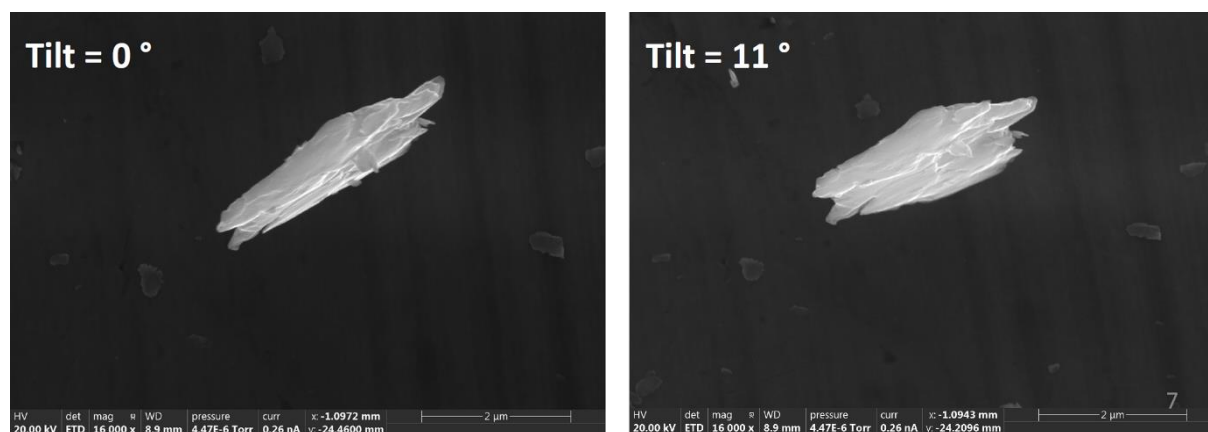
Figure VI-3: The tilt effect on the particle representation.

Table VI-1: Difference between the morphological descriptors when the tilt degree is changed.

Tilt degree	Elongation	Regularity	Roundness	Circularity
0°	0,23	0,86	0,09	0,48
11°	0,36	0,87	0,06	0,59

The visualisation scale of the third body particles is a major field that needs to be investigated more in future work. This scale is the basis of our judgement in making the distinction between an agglomerate and a particle. In **Figure VI-4**, we can clearly observe more the effect of the scale when we magnify what we considered a particle (Top left mage from **Figure VI-4**) and figure out that it was an agglomerate (Bottom right image from **Figure VI-4**). In our project, we only took into consideration what we judged as particles and analysed them to obtain the morphological descriptors. However, the existence of the agglomerates should be taken into account to describe the results of an experiment.
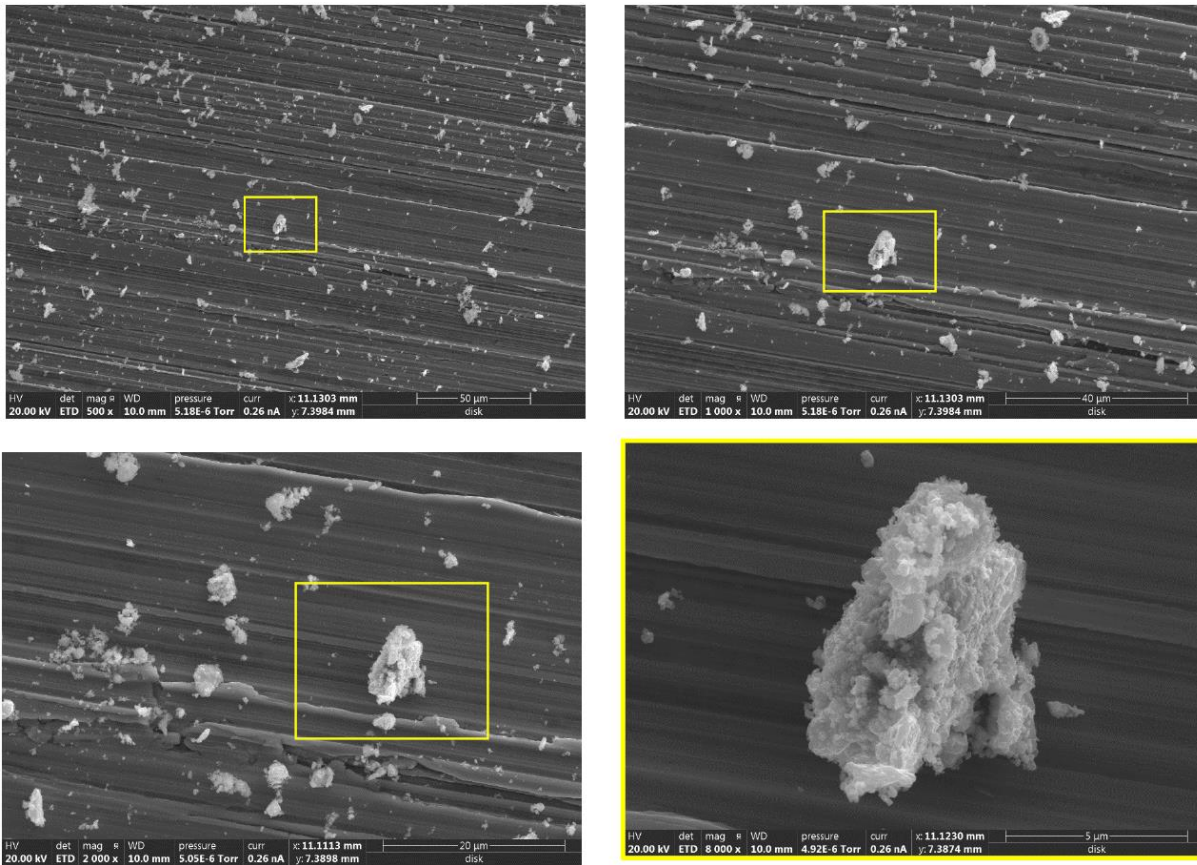
Figure VI-4: The scale of observation effect.

In our study, we only focused on the shape and the form of the particles from SEM images. One more area of improvement for this proposed work is using a second microscope such as the optical microscope. In **Figure VI-5**, we display the same third body particles observed using the SEM and the optical microscope. The information obtained are complementary and help have a more complete idea about the particles. The first descriptor that comes to mind to add to the morphological database is a texture descriptor.
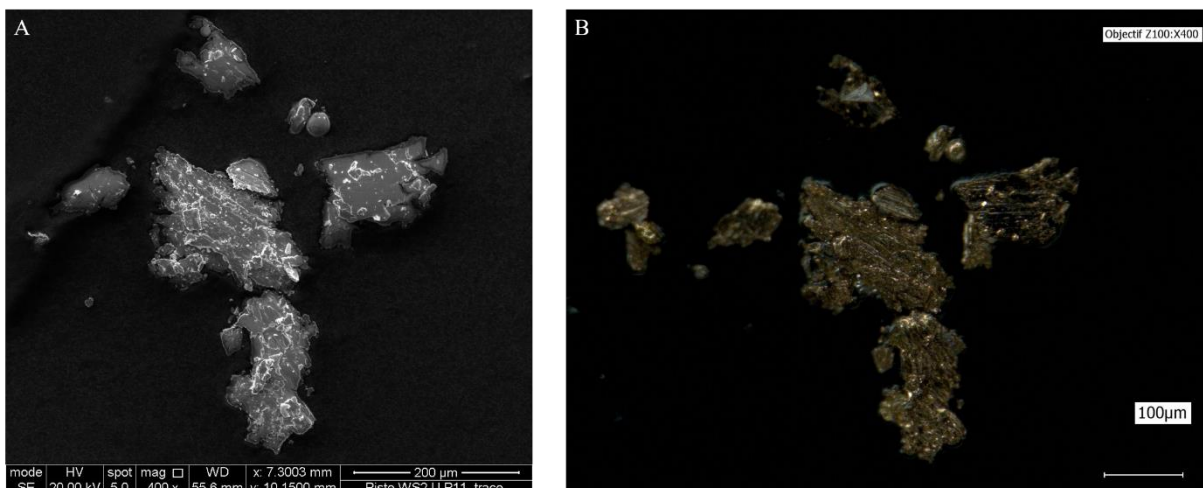
Figure VI-5: Microscopic image of a batch of third body particles A) from the SEM. B) from the optical microscope.

The rheological measurements are obtained from the real time signals that describe the totality of the tribological experiments. In **Figure VI-6**, we display the real-time evolution of the COF value during one of the tribological tests. It is evident that the coefficient of friction is varying both in time and in space, and that

some sections of the disc have characteristic values of friction which are different from those of other sections. However, since we do not have a way to estimate an exact time of creation for the third body particles, we associate the same COF value (the average of the real time recordings during the stable stage) to all of the particles found, which is not accurate. As a solution for this point, we suggest that the experimental procedure should be stopped during the tests. Particles that are already in the ejection zone should be analysed since they will no longer be reintroduced to the contact and take the value of the COF at that moment of time. This will ensure more variability in the COF value inside the particles from the same test. Besides, when analysing the final state of the track, the location of each processed particle should be noted in order to establish a connection between this particle and the local (not space and time averaged) value of the rheological parameters (e.g. local value of the coefficient of friction in this area of the disc). This would tremendously enrich the database and lead to a more accurate processing by the machine-learning algorithms.
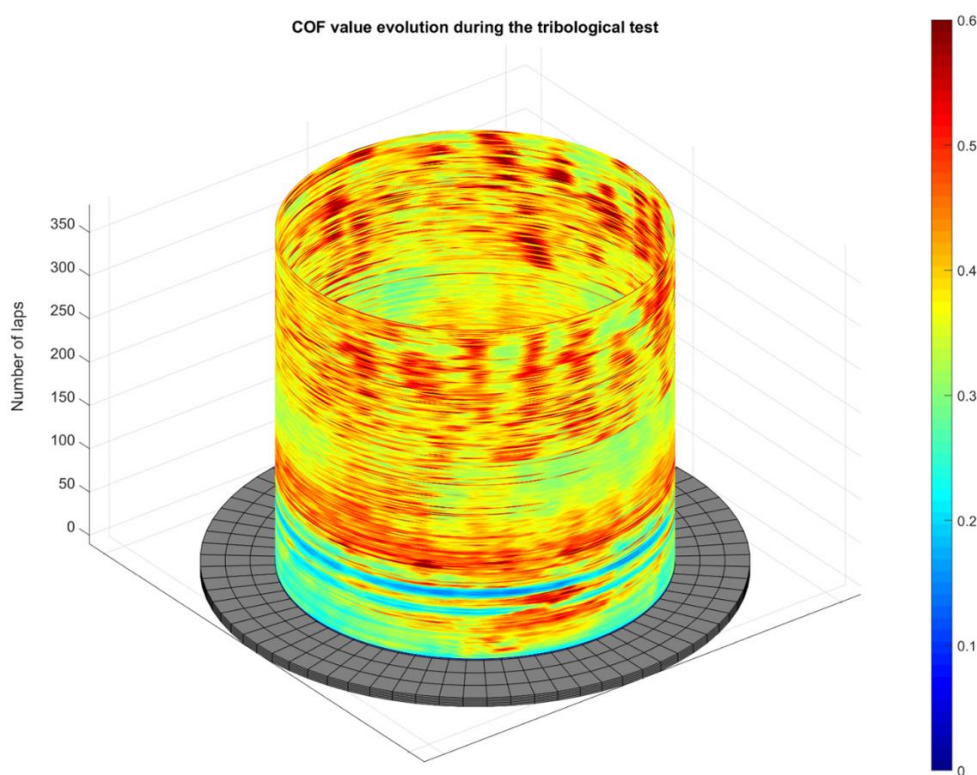
Figure VI-6: Evolution of the COF value during a sliding test.

During this thesis, we adapted the pin-disk tribometer to our needs. Some aspects of the experiment need to be reinforced, such as the accuracy of the execution of the given instruction for the rotational speed by the motor. In addition, the gas chamber was fixed using insulating tape during the experiments and the possible leakage was corrected by keeping a steady flow of the gaz. The fixation method should be improved and the chamber should be fixed permanently on the tribometer. Even when the experiment does not require a change in the gaseous environment, the chamber helps keep the experimental surroundings clean from any foreign particles (any tiny particles in the air that are not created during the tribological experiment). .

As future perspective for this work, we aim to expand the database of the wear debris used to include third body particles created through numerical simulation. Simulations to recreate the tribological experiments are

in their first steps (Mollon, 2016, 2018). **Figure VI-7** shows samples resulting from the present algorithms. The simulations are very promising since they offer direct accessibility to both morphological and rheological data and through them different flow regimes can be reproduced. However, the simulation scale is very small when compared to the contact surface and the numerical results are not experimentally verified yet.

| Powdery aspect | Agglomerated aspect | Plastic aspect |
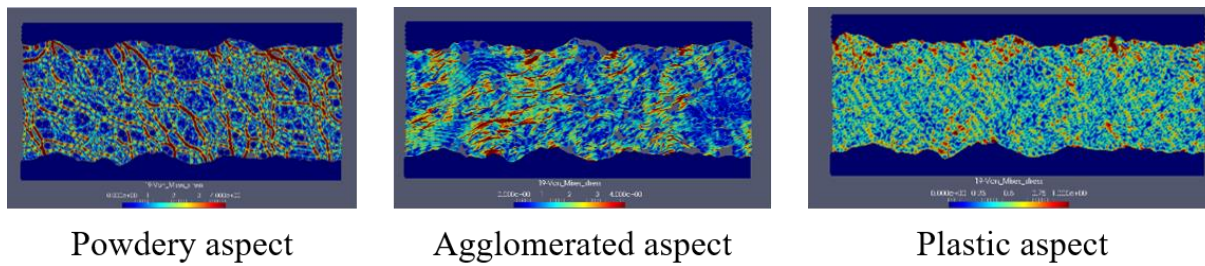
Figure VI-7: Different aspect created through numerical simulations

This thesis marks one of the building blocks of a much bigger project, where there is an industrial aim. The final product should be able to predict machine failure from a database created by combining the rheological signals registered using the different sensors installed and the morphological state of the wear debris photographed through the cameras that observe the third body particles. Such product will be useful in machine monitoring and in planning the preventive maintenance.

Appendix

Appendix 1: Standard Normal Distribution

Since we work with the averages of the morphological descriptors (μ_i), it is necessary for our study to include a statistical element to determine the minimum number n of particles that needs to be studied in order to have meaningful results. We cannot for example analyse one particles from test 1 and a thousand from the other tests. The comparison of the descriptors will not be useful.

We fix therefore a margin of error (MoE) equal to 10 % at the 95 % confidence interval. The confidence interval chosen in itself means that if we were to study a 100 particles from any test, approximately 95 particles will have descriptor value that is contained in the interval defined by $[\mu_i - MoE; \; \mu_i + MoE]$.

The margin of error has the following formula :

$$\text{MoE} = Z * \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \leq 10\%$$

with the variables defined as :

Z: The value that corresponds to having 95% as a confidence interval.

$\hat{p}$: The sample proportion. It is the percentage of the population that meets our criteria.

n: The sample size we are searching for.

To determine the Z value we follow the next steps:

- *Step 1*: Subtract the confidence interval value from 100 % to calculate the α.
$$\alpha = 100 - 95 = 5\%$$
- *Step 2*: Convert α to the decimal numeral system.
$$\alpha = 0.05$$
- *Step 3*: Divide α by 2
$$\alpha/2 = 0.025$$
- *Step 4*: Subtract $\alpha/2$ from 1 to have a.
$$a = 1 - \alpha/2 = 0.975$$
- *Step 5*: Look up a in the z-table (red square in the **Table below**)
$$Z = 1.96 \text{ for a 95 \% confidence interval.}$$

After determining the value of Z, only two unknown variables remain in the MoE equation:

$$1.96 * \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \leq 10\%$$

To ensure that we have the maximum n, we need that $\hat{p}$ that maximises $\hat{p}(1-\hat{p})$, which is $\hat{p} = 0.5$.

Therefore, we have :

$$n \geq \frac{\hat{p}(1-\hat{p})}{\left(\frac{0.1}{Z}\right)^2}$$

➔ $n \geq 96$ particles.

This Statistical calculation resulted in determining that we need to analyse at least 96 particles from each test in order to ensure that our data respect the 10 % margin of error at the 95 % confidence interval.

Table 1: Z-table values (in red square is the Z value for 95 % confidence interval).

z	0	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
+0	.50000	.50399	.50798	.51197	.51595	.51994	.52392	.52790	.53188	.53586
+0.1	.53983	.54380	.54776	.55172	.55567	.55966	.56360	.56749	.57142	.57535
+0.2	.57926	.58317	.58706	.59095	.59483	.59871	.60257	.60642	.61026	.61409
+0.3	.61791	.62172	.62552	.62930	.63307	.63683	.64058	.64431	.64803	.65173
+0.4	.65542	.65910	.66276	.66640	.67003	.67364	.67724	.68082	.68439	.68793
+0.5	.69146	.69497	.69847	.70194	.70540	.70884	.71226	.71566	.71904	.72240
+0.6	.72575	.72907	.73237	.73565	.73891	.74215	.74537	.74857	.75175	.75490
+0.7	.75804	.76115	.76424	.76730	.77035	.77337	.77637	.77935	.78230	.78524
+0.8	.78814	.79103	.79389	.79673	.79955	.80234	.80511	.80785	.81057	.81327
+0.9	.81594	.81859	.82121	.82381	.82639	.82894	.83147	.83398	.83646	.83891
+1	.84134	.84375	.84614	.84849	.85083	.85314	.85543	.85769	.85993	.86214
+1.1	.86433	.86650	.86864	.87076	.87286	.87493	.87698	.87900	.88100	.88298
+1.2	.88493	.88686	.88877	.89065	.89251	.89435	.89617	.89796	.89973	.90147
+1.3	.90320	.90490	.90658	.90824	.90988	.91149	.91308	.91466	.91621	.91774
+1.4	.91924	.92073	.92220	.92364	.92507	.92647	.92785	.92922	.93056	.93189
+1.5	.93319	.93448	.93574	.93699	.93822	.93943	.94062	.94179	.94295	.94408
+1.6	.94520	.94630	.94738	.94845	.94950	.95053	.95154	.95254	.95352	.95449
+1.7	.95543	.95637	.95728	.95818	.95907	.95994	.96080	.96164	.96246	.96327
+1.8	.96407	.96485	.96562	.96638	.96712	.96784	.96856	.96926	.96995	.97062
+1.9	.97128	.97193	.97257	.97320	.97381	.97441	.97500	.97558	.97615	.97670
+2	.97725	.97778	.97831	.97882	.97932	.97982	.98030	.98077	.98124	.98169
+2.1	.98214	.98257	.98300	.98341	.98382	.98422	.98461	.98500	.98537	.98574
+2.2	.98610	.98645	.98679	.98713	.98745	.98778	.98809	.98840	.98870	.98899
+2.3	.98928	.98956	.98983	.99010	.99036	.99061	.99086	.99111	.99134	.99158
+2.4	.99180	.99202	.99224	.99245	.99266	.99286	.99305	.99324	.99343	.99361
+2.5	.99379	.99396	.99413	.99430	.99446	.99461	.99477	.99492	.99506	.99520
+2.6	.99534	.99547	.99560	.99573	.99585	.99598	.99609	.99621	.99632	.99643
+2.7	.99653	.99664	.99674	.99683	.99693	.99702	.99711	.99720	.99728	.99736
+2.8	.99744	.99752	.99760	.99767	.99774	.99781	.99788	.99795	.99801	.99807
+2.9	.99813	.99819	.99825	.99831	.99836	.99841	.99846	.99851	.99856	.99861
+3	.99865	.99869	.99874	.99878	.99882	.99886	.99889	.99893	.99896	.99900
+3.1	.99903	.99906	.99910	.99913	.99916	.99918	.99921	.99924	.99926	.99929
+3.2	.99931	.99934	.99936	.99938	.99940	.99942	.99944	.99946	.99948	.99950
+3.3	.99952	.99953	.99955	.99957	.99958	.99960	.99961	.99962	.99964	.99965
+3.4	.99966	.99968	.99969	.99970	.99971	.99972	.99973	.99974	.99975	.99976
+3.5	.99977	.99978	.99978	.99979	.99980	.99981	.99981	.99982	.99983	.99983
+3.6	.99984	.99985	.99985	.99986	.99986	.99987	.99987	.99988	.99988	.99989
+3.7	.99989	.99990	.99990	.99990	.99991	.99991	.99992	.99992	.99992	.99992
+3.8	.99993	.99993	.99993	.99994	.99994	.99994	.99994	.99995	.99995	.99995
+3.9	.99995	.99995	.99996	.99996	.99996	.99996	.99996	.99996	.99997	.99997
+4	.99997	.99997	.99997	.99997	.99997	.99997	.99998	.99998	.99998	.99998

Appendix 2: Predicting the rheological descriptors from the morphological database.

The linear and polynomial regression were used to predict the rest of the rheological descriptors. In this appendix, we present the result of this study. The aim of the figures resulting from the linear regression is not for individual inspection. It is to show how the predictions, independently from the value of λ, evolve around the average f the real values of the rheological descriptor the algorithm is trying the study (learn). We remind in **Table 1**, the rest of the rheological descriptors. In **Table 2**, we find the value of all the rheological data from all the tests.

Table 0-2: Rheological data description.

Label	Description
RD_2	Standard deviation of the average per lap friction factor in established state (After the $10\ m$ mark).
RD_3	The percentage of the cohesive zone relative to the entire contact area on the pin
RD_4	The intensity of the flow
RD_5	The diameter of the contact area on the pin
RD_6	The width of the contact on the sliding track

Table 0-3: Rheological descriptors values per test.

Test	RD_2	RD_3	RD_4	RD_5	RD_6
Test 1	0,11	16	2	1,22	1
Test 2	0,092	18	2	1,25	1,1
Test 3	0,055	36	3	0,94	0,87
Test 4	0,037	12	2	1,25	1,2
Test 5	0,17	21	3	0,7	0,68
Test 6	0,123	–	3	0,94	0,72
Test 7	0,029	20	1	0,9	0,75
Test 8	0,065	21	2	1	0,7
Test 9	0,038	35	2	1,2	0,92

1) Predictions for RD_2

1.1) Linear regression results
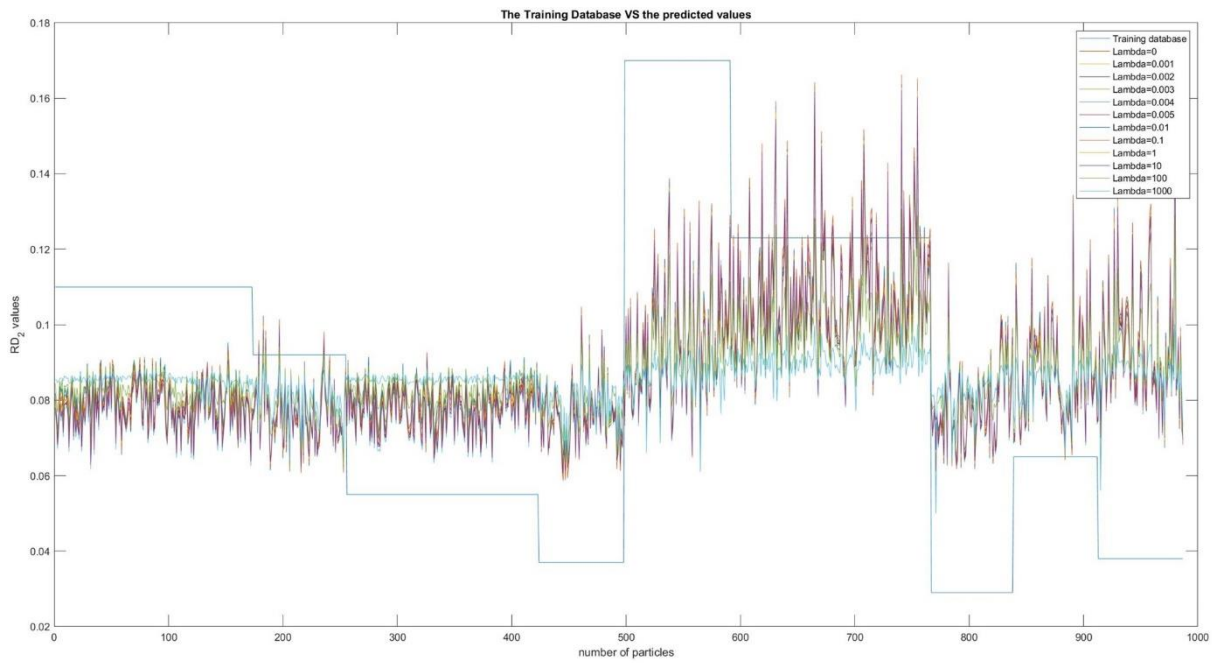
Figure 1: Results of applying the linear regression algorithm on the training database.
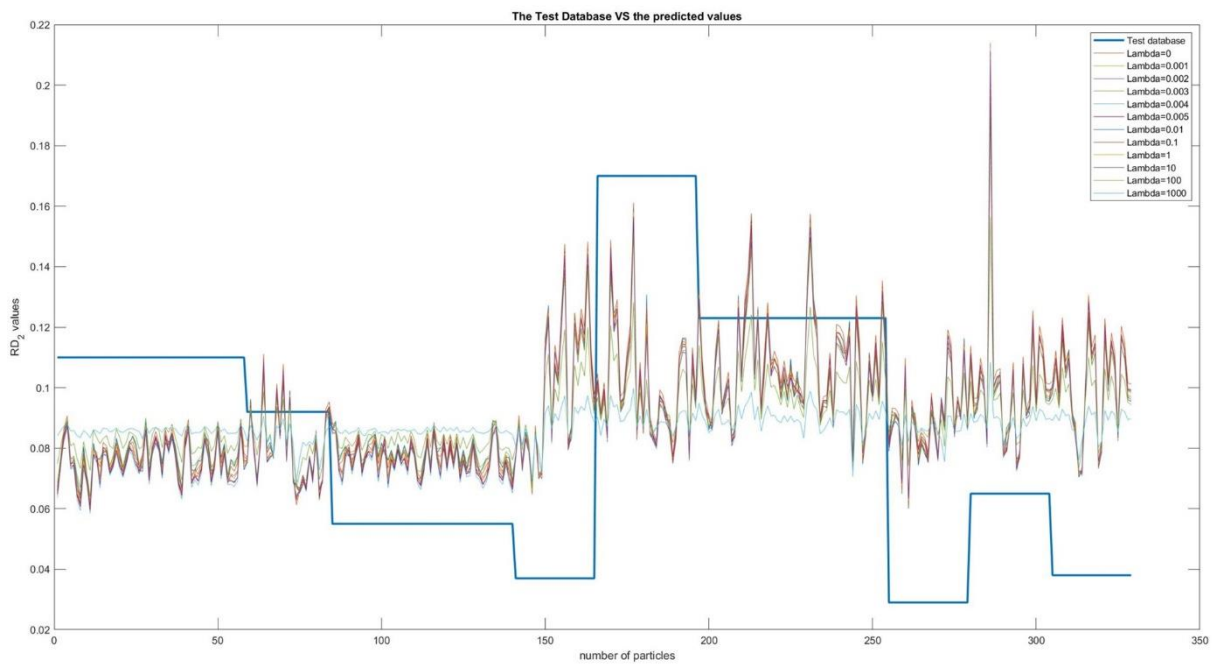
Figure 2: Results of applying the linear regression algorithm on the test database.

1.2)   Polynomial regression results
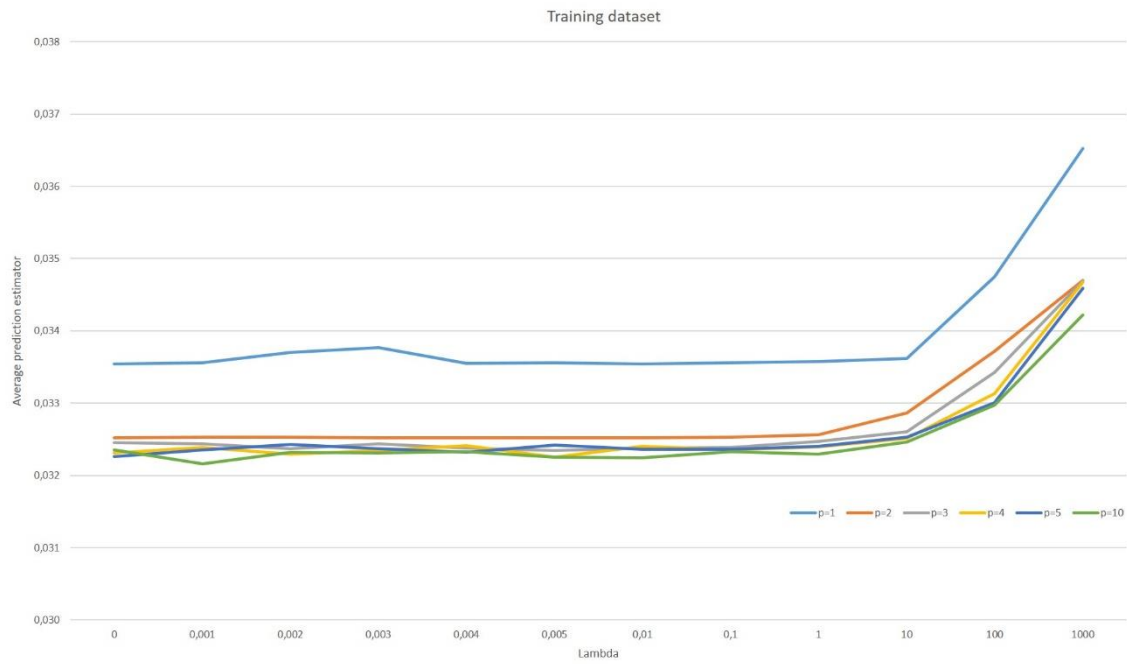
Figure 3: Average prediction estimator calculated on the training dataset when polynomial regression algorithms are applied.
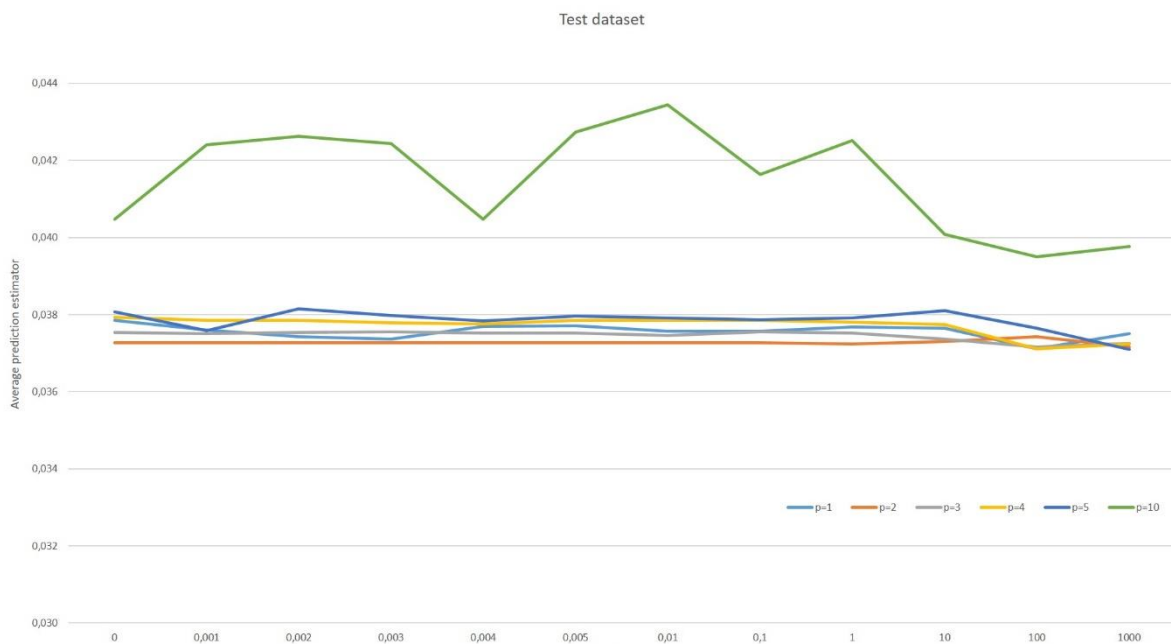
Figure 4: Average prediction estimator calculated on the test dataset when polynomial regression algorithms are applied.

2) Predictions for RD_3

2.1) Linear regression results
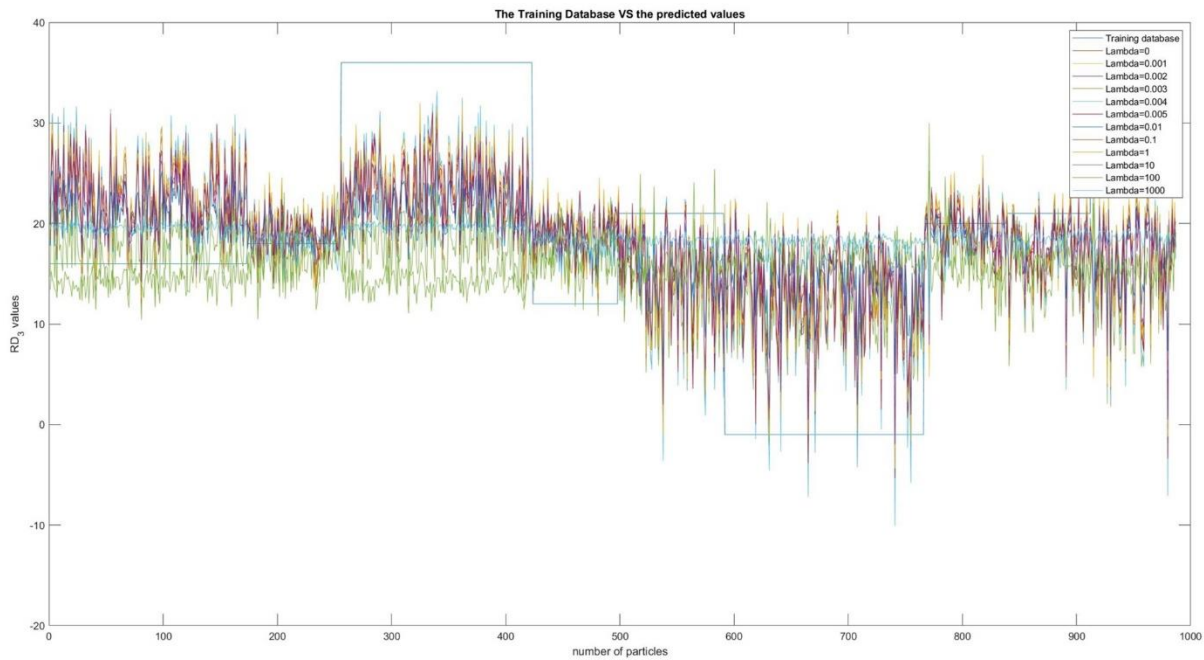
Figure 5: Results of applying the linear regression algorithm on the training database.
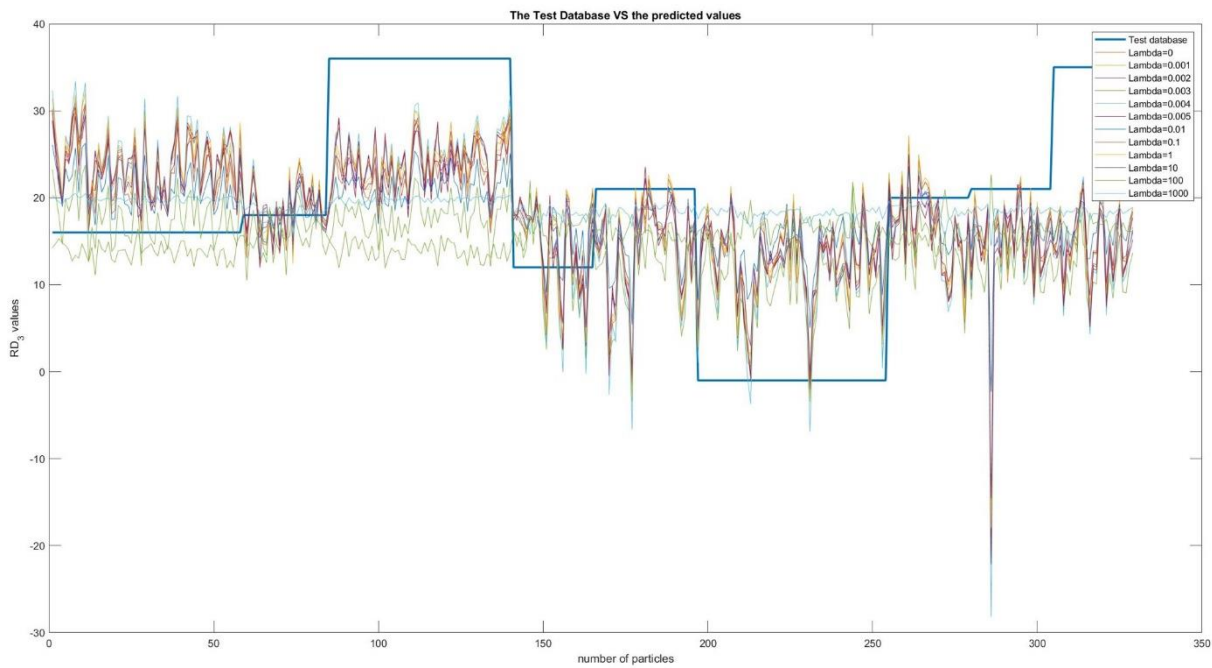
Figure 6: Results of applying the linear regression algorithm on the test database.

2.2) Polynomial regression results
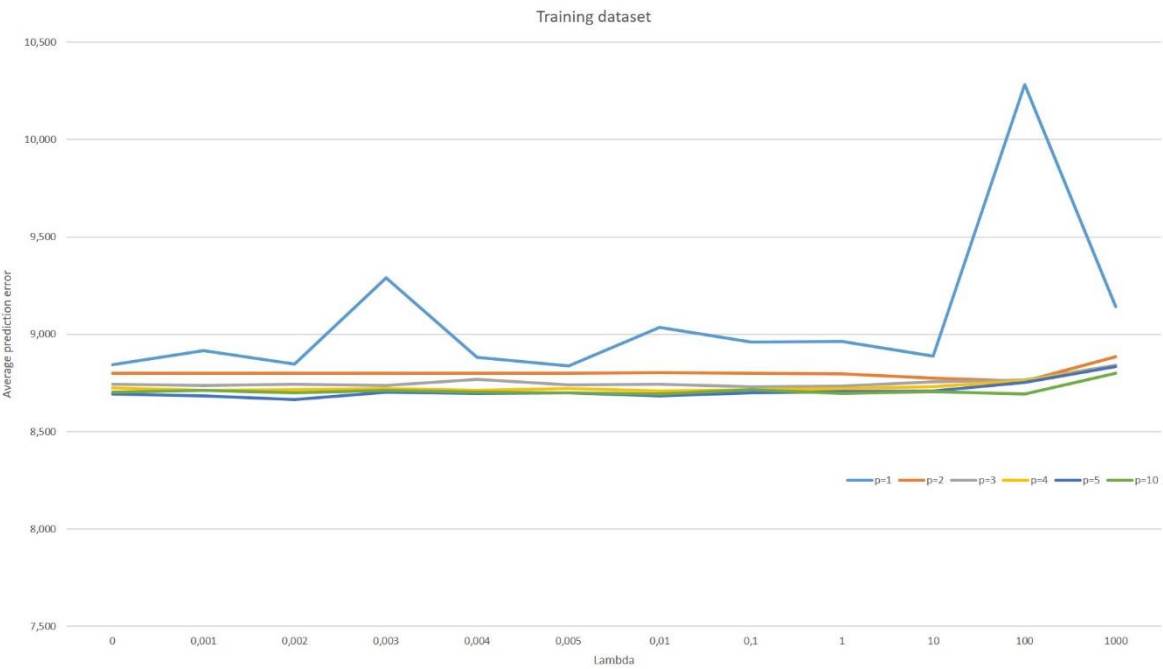
Figure 7: Average prediction estimator calculated on the training dataset when polynomial regression algorithms are applied.
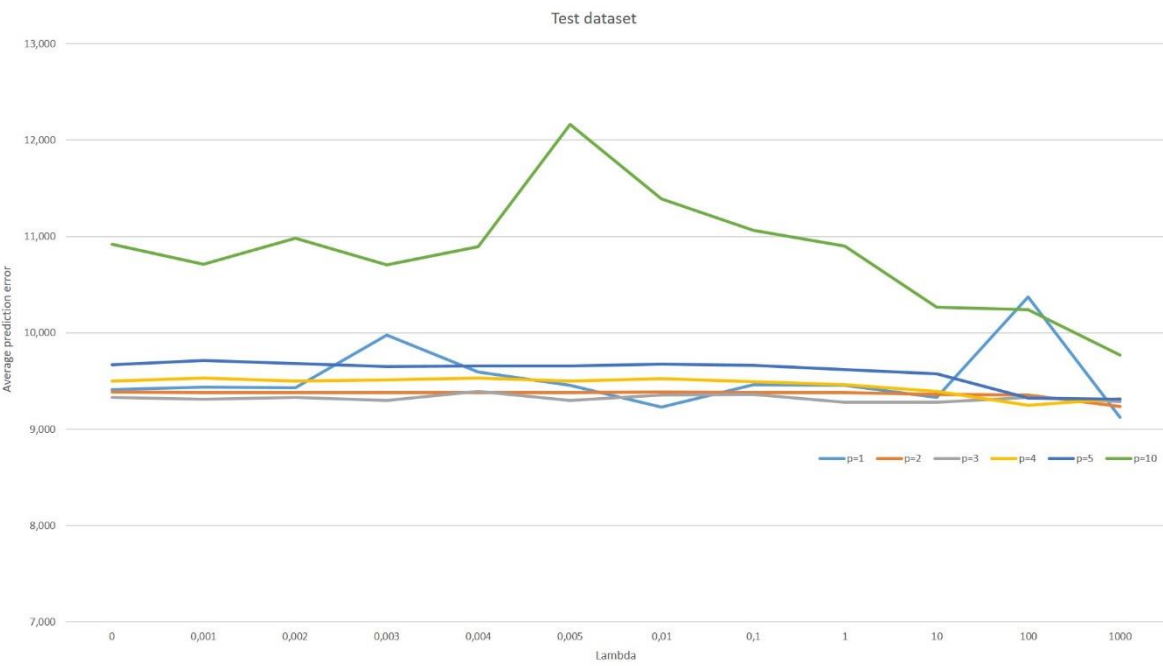
Figure 8: Average prediction estimator calculated on the test dataset when polynomial regression algorithms are applied.

3) Predictions for RD_4

3.1) Linear regression results
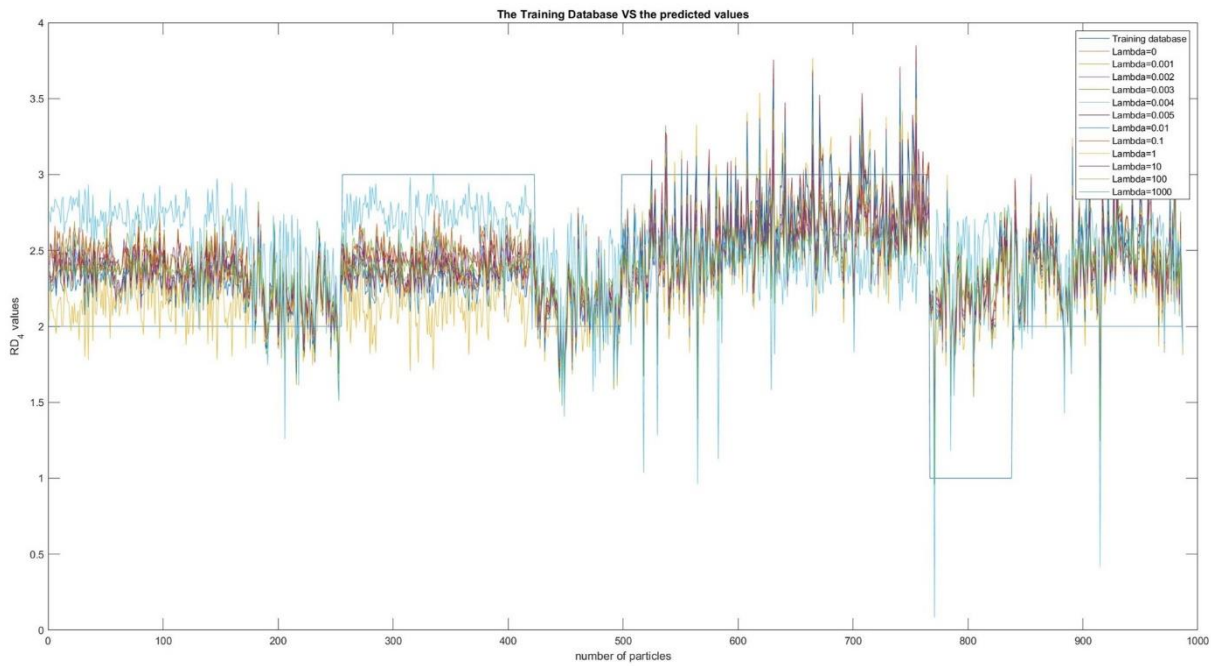
Figure 9: Results of applying the linear regression algorithm on the training database.
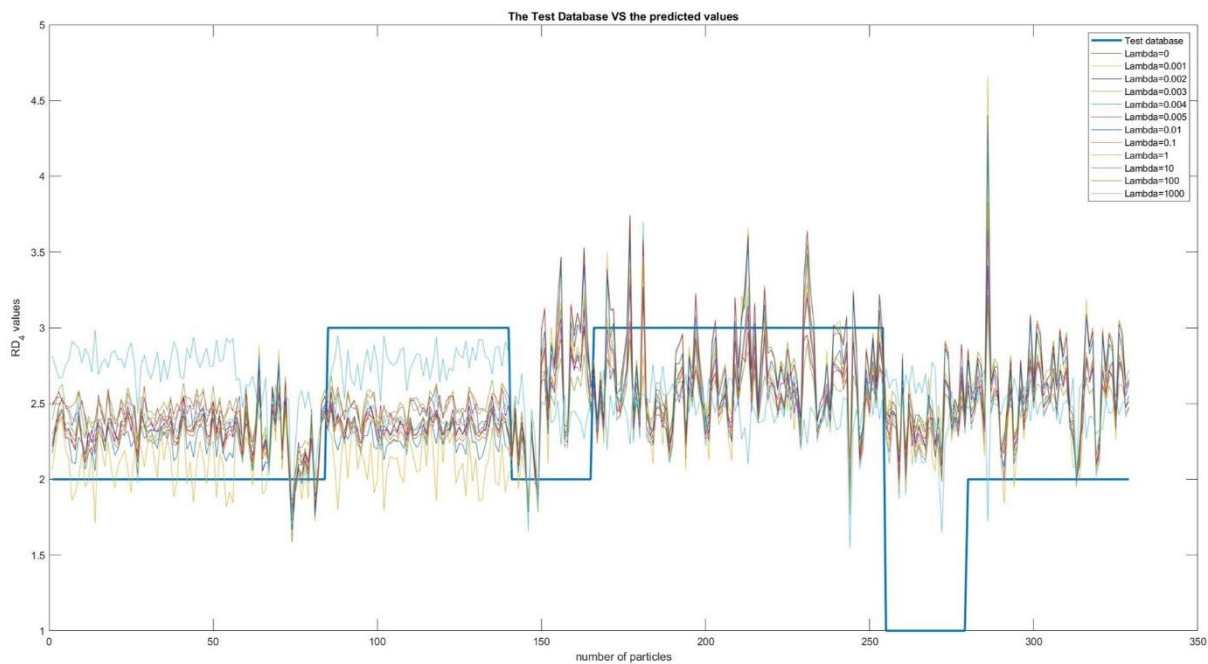
Figure 10: Results of applying the linear regression algorithm on the test database.

3.2) Polynomial regression results
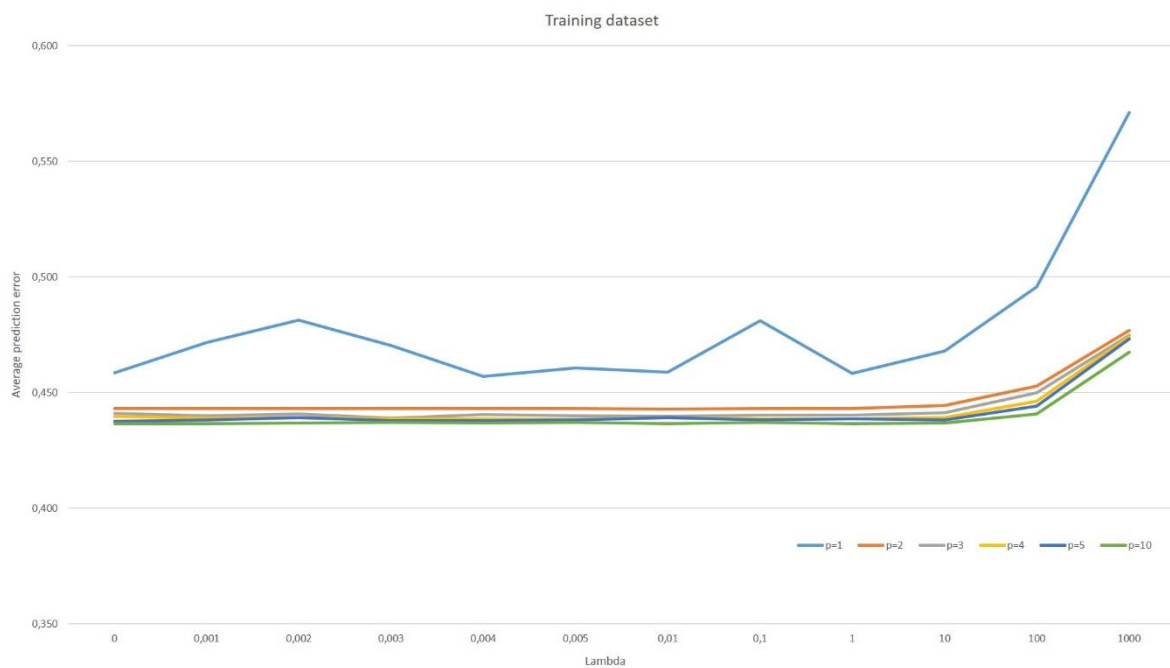
Figure 11: Average prediction estimator calculated on the training dataset when polynomial regression algorithms are applied.
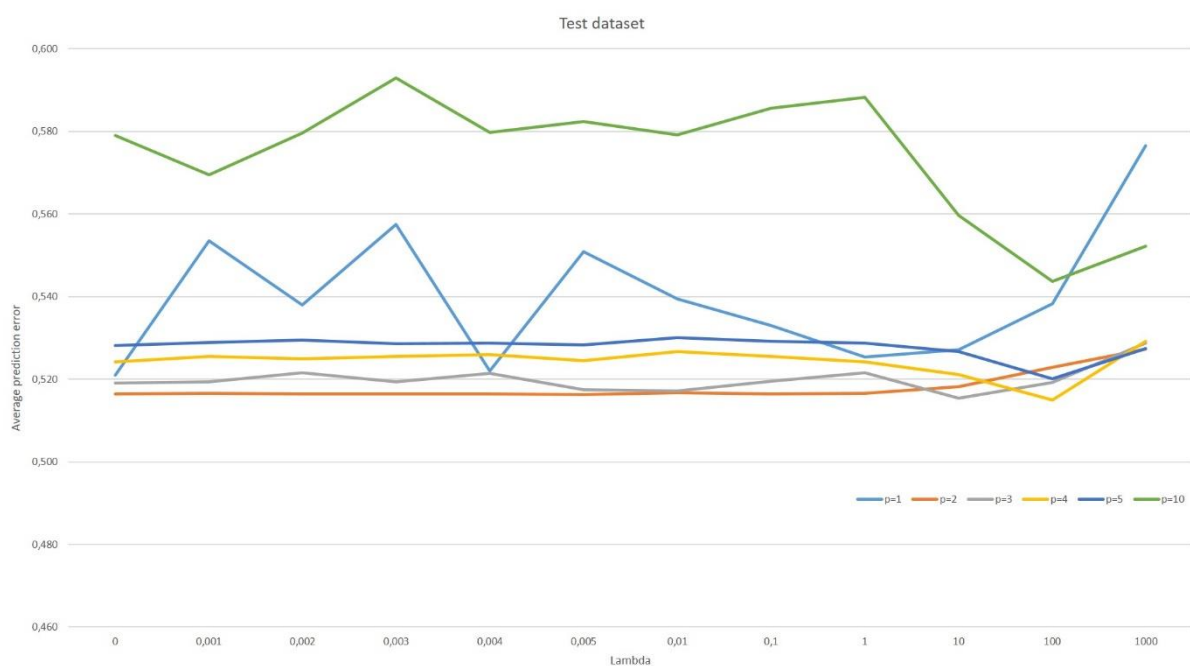
Figure 12: Average prediction estimator calculated on the test dataset when polynomial regression algorithms are applied.

4) Predictions for RD_5

4.1) Linear regression results
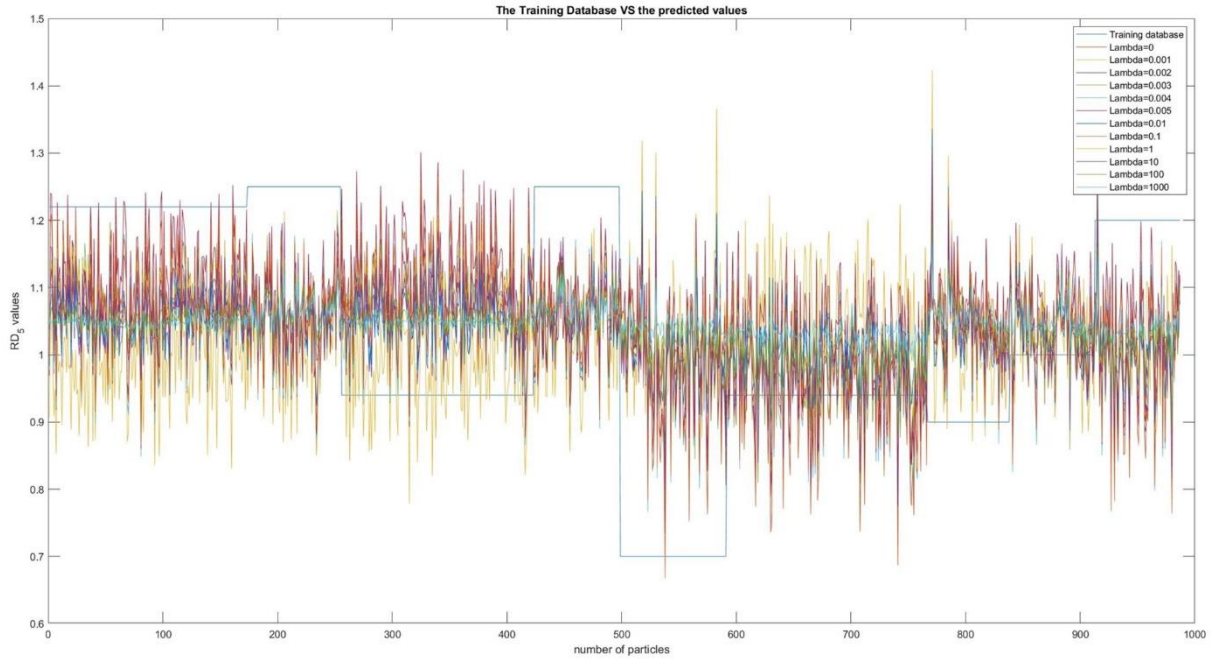
Figure 13: Results of applying the linear regression algorithm on the training database.
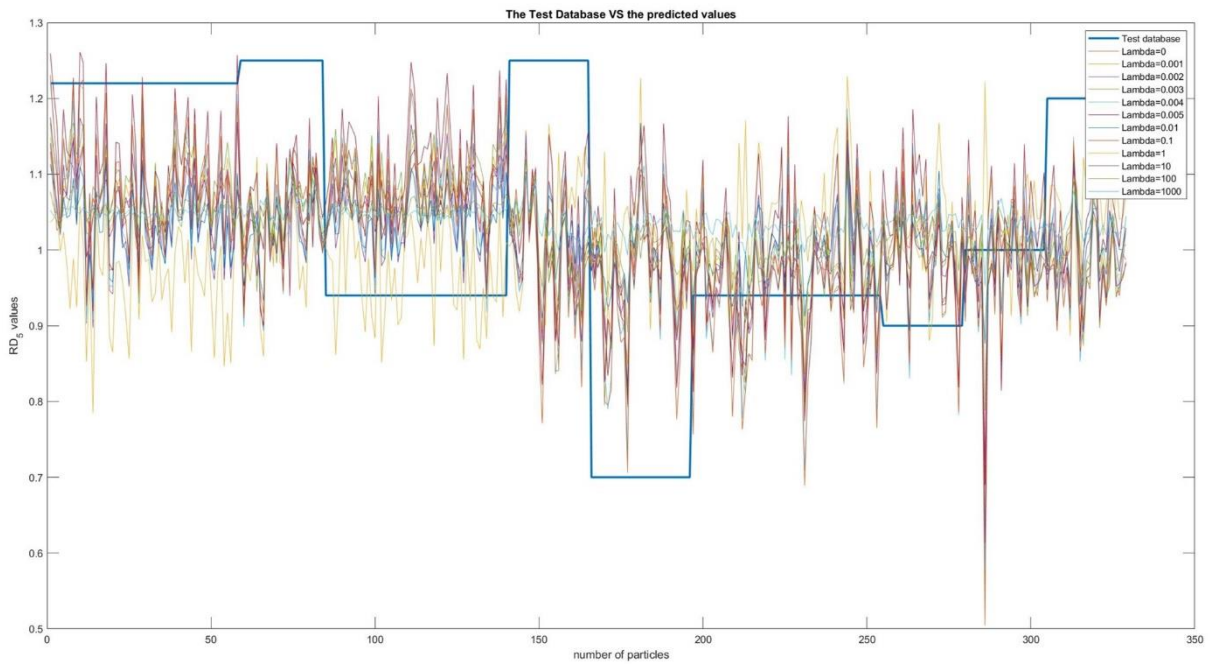
Figure 14: Results of applying the linear regression algorithm on the test database.

4.2) Polynomial regression results
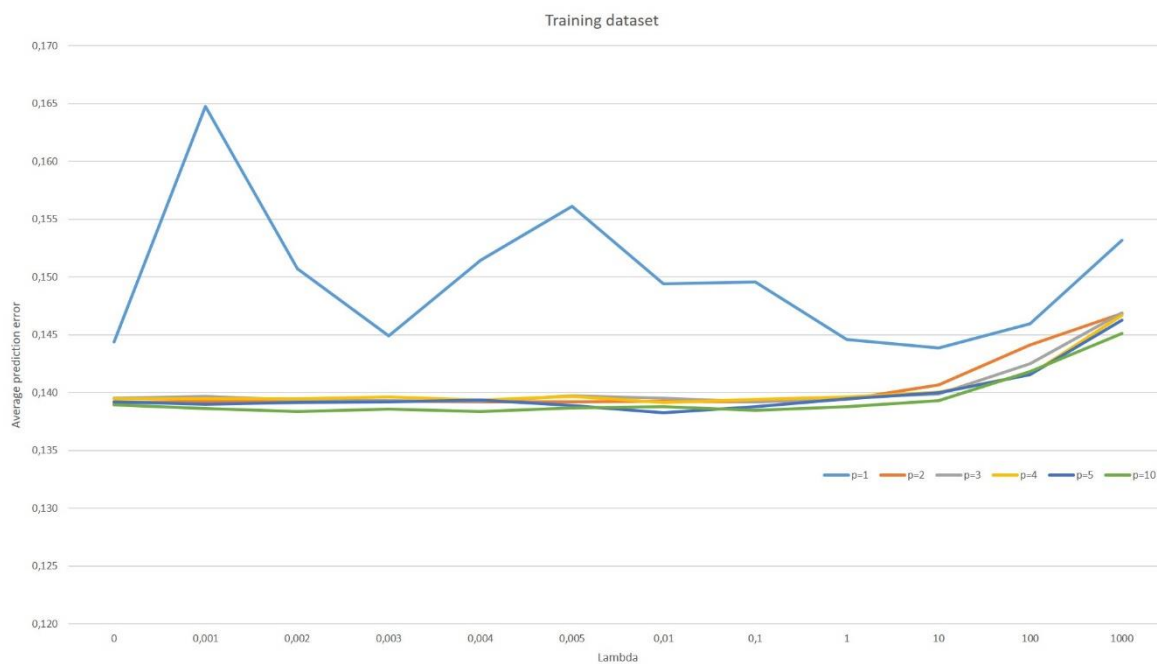
Figure 15: Average prediction estimator calculated on the training dataset when polynomial regression algorithms are applied.
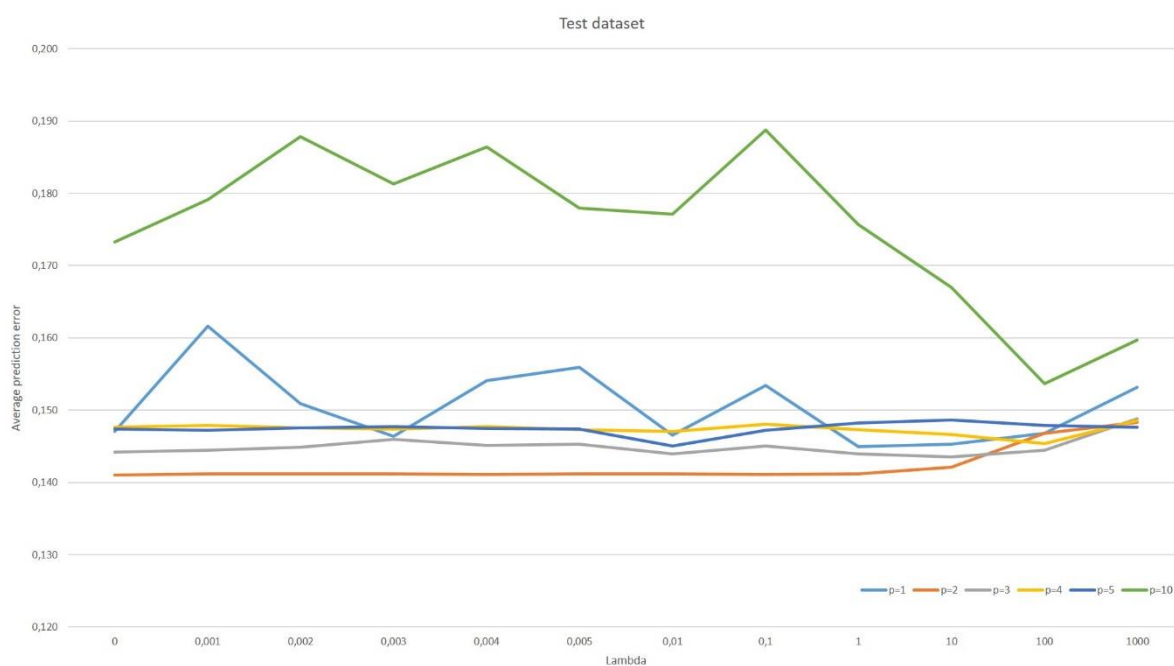
Figure 16: Average prediction estimator calculated on the test dataset when polynomial regression algorithms are applied.

5) Predictions for RD_6

5.1) Linear regression results
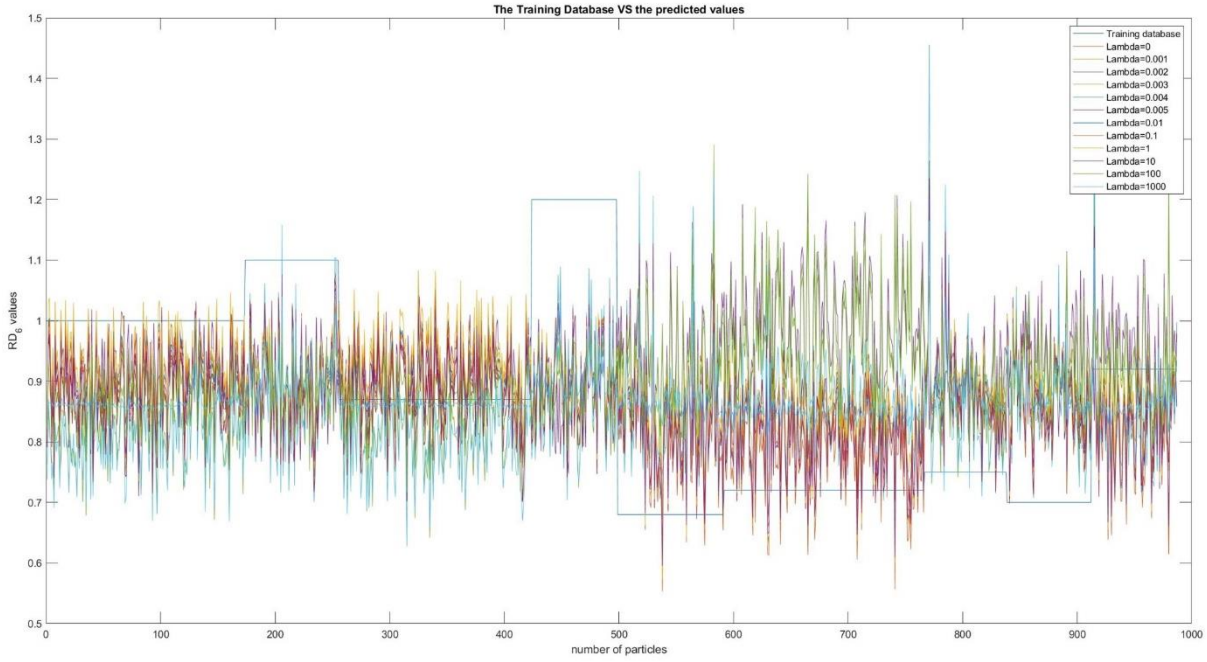
Figure 17: Results of applying the linear regression algorithm on the training database.
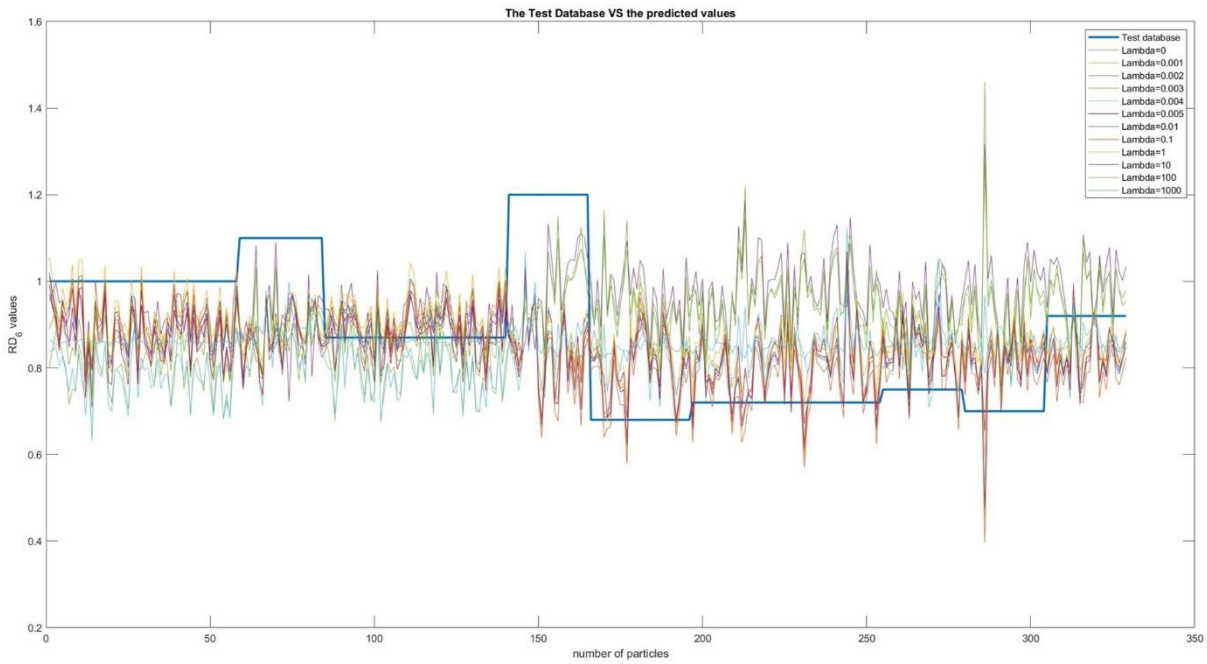
Figure 18: Results of applying the linear regression algorithm on the test database.

5.2) Polynomial regression results
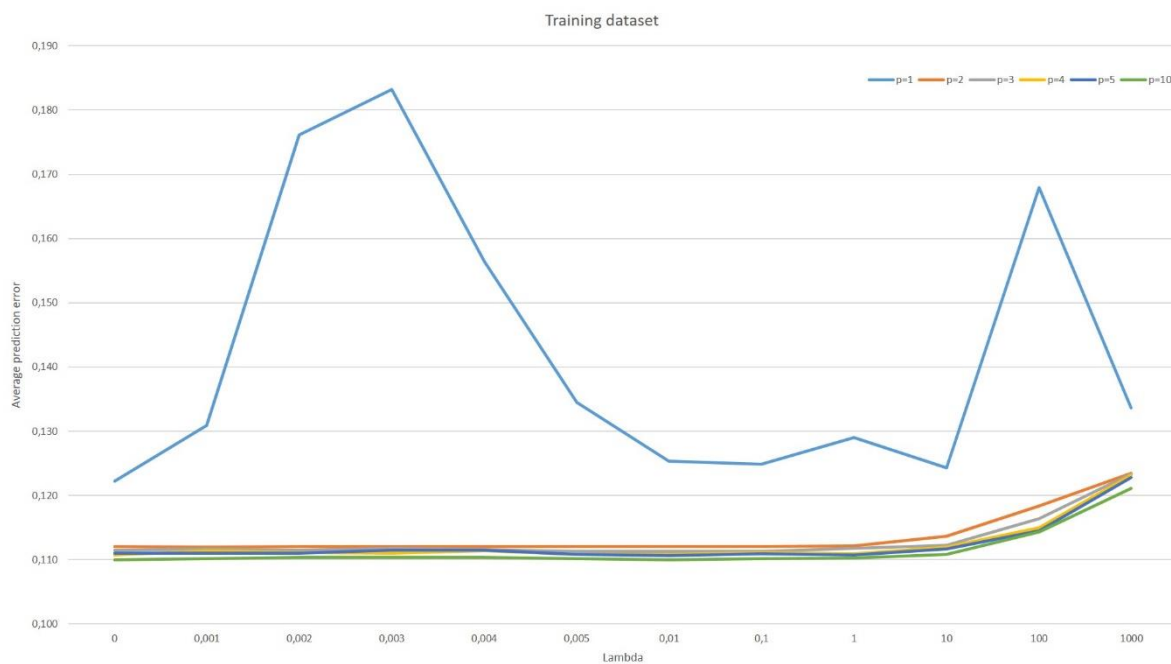
Figure 19: Average prediction estimator calculated on the training dataset when polynomial regression algorithms are applied.
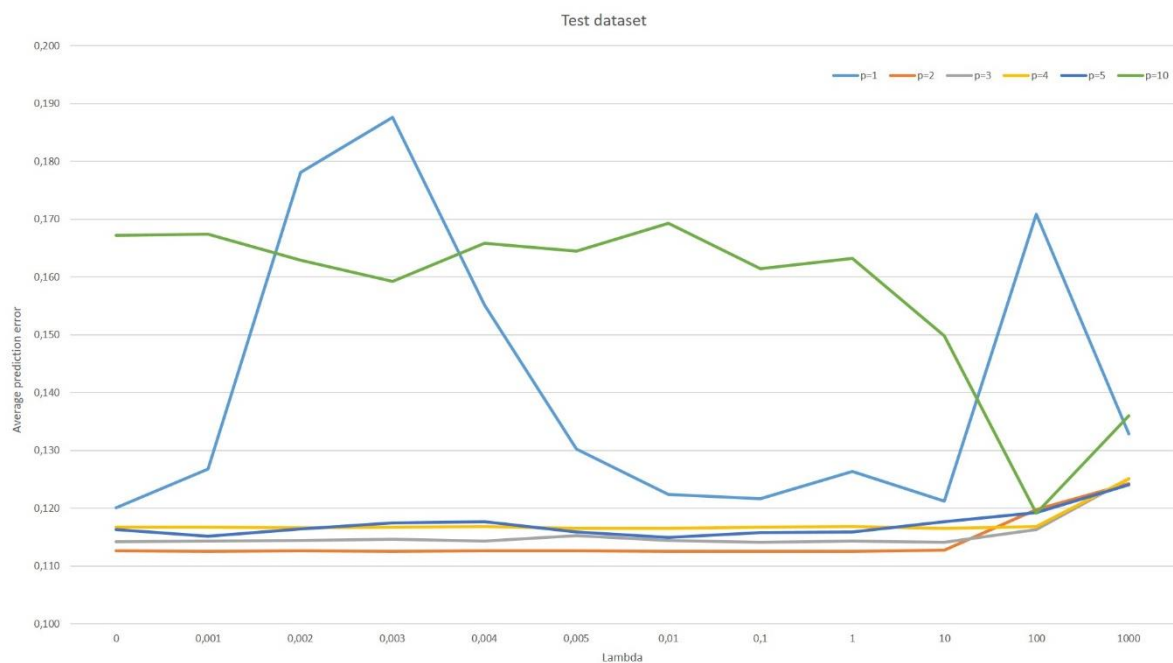
Figure 20: Average prediction estimator calculated on the test dataset when polynomial regression algorithms are applied.

Bibliography

Alexeev, D. V. (2010) 'Neural-network approximation of functions of several variables', *Journal of Mathematical Sciences*. Springer, 168(1), pp. 5–13. doi: 10.1007/s10958-010-9970-5.

Alonso-Fernandez, F. *et al.* (2009) 'Fingerprint Recognition', in *Guide to Biometric Reference Systems and Performance Evaluation*. London: Springer London, pp. 51–88. doi: 10.1007/978-1-84800-292-0_4.

Aly, M. and Aly, M. (2005) 'Survey on Multiclass Classification Methods'. Available at: http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.175.107 (Accessed: 4 November 2019).

Anderson, D. P. (1982) 'Wear Particle Atlas. Revised'. Available at: https://apps.dtic.mil/docs/citations/ADA125512 (Accessed: 30 January 2019).

Atta-Fosu, T. *et al.* (2016) '3D Clumped Cell Segmentation Using Curvature Based Seeded Watershed', *Journal of Imaging*. Multidisciplinary Digital Publishing Institute, 2(4), p. 31. doi: 10.3390/jimaging2040031.

Barrau, O. (2004) 'Étude du frottement et de l'usure d'acier à outils de travail à chaud'. Available at: http://oatao.univ-toulouse.fr/7382/ (Accessed: 11 November 2018).

Berthier, Y. (1990) 'Experimental evidence for friction and wear modelling', *Wear*. Elsevier, 139(1), pp. 77–92. doi: 10.1016/0043-1648(90)90210-2.

Berthier, Y. (1996) 'Maurice Godet's Third Body', *Tribology Series*. Elsevier, 31, pp. 21–30. doi: 10.1016/S0167-8922(08)70766-1.

Berthier, Y., Vincent, L. and Godet, M. (1988) 'Velocity accommodation in fretting', *Wear*. Elsevier, 125(1–2), pp. 25–38. doi: 10.1016/0043-1648(88)90191-3.

Bill and C., R. (1982) 'Review of factors that influence fretting wear'. Available at: https://ntrs.nasa.gov/search.jsp?R=19830036041 (Accessed: 11 November 2018).

Blott, S. J. and Pye, K. (2008) 'Particle shape: A review and new methods of characterization and classification', *Sedimentology*, 55(1), pp. 31–63. doi: 10.1111/j.1365-3091.2007.00892.x.

Bojarski, M. *et al.* (2016) 'End to End Learning for Self-Driving Cars'. Available at: http://arxiv.org/abs/1604.07316 (Accessed: 10 January 2019).

Brunetière, N. (2016) *Introduction to Tribology*. Available at: https://anarkia333data.center/en/node/2701 (Accessed: 3 February 2020).

Buckley, D. H. (1978) 'Tribological properties of surfaces', *Thin Solid Films*. Elsevier, 53(3), pp. 271–283. doi: 10.1016/0040-6090(78)90221-3.

Candes, E. and Tao, T. (2007) 'The Dantzig selector: Statistical estimation when p is much larger than n', *The Annals of Statistics*. Institute of Mathematical Statistics, 35(6), pp. 2313–2351. doi: 10.1214/009053606000001523.

Colas, G. *et al.* (2013) 'Decrypting third body flows to solve dry lubrication issue - MoS2 case study under ultrahigh vacuum', *Wear*. Elsevier, 305(1–2), pp. 192–204. doi: 10.1016/j.wear.2013.06.007.

Colas, G. (2013) 'Utilisation raisonnée de contaminants pour caractériser la rhéologie des premiers et troisième corps solides', p. 245.

Cox, D. R. (David R. (1977) *The analysis of binary data*. Chapman and Hall.

Czarnecki, W. M., Józefowicz, R. and Tabor, J. (2015) 'Maximum Entropy Linear Manifold for Learning Discriminative Low-dimensional Representation'. Available at: http://arxiv.org/abs/1504.02622 (Accessed: 1 November 2019).

Davidson (1957) 'Bearings since the Stone Age', *Engineering*, (183).

Davies, A. (ed.) (1998) *Handbook of Condition Monitoring*. Dordrecht: Springer Netherlands. doi: 10.1007/978-94-011-4924-2.

Davies, A., Prickett, P. W. and Grosvenor, R. I. (1998) 'Future developments in condition monitoring techniques and systems', in *Handbook of Condition Monitoring*. Dordrecht: Springer Netherlands, pp. 541–560. doi: 10.1007/978-94-011-4924-2_23.

Decost, B. L. and Holm, E. A. (2015) 'A computer vision approach for automated analysis and classification of microstructural image data', *Computational Materials Science*. Elsevier B.V., 110, pp. 126–133. doi: 10.1016/j.commatsci.2015.08.011.

Denape, J. (2014) 'Third body concept and wear particle behavior in dry sliding conditions', *Tribological aspects in modern aircraft industry*, 640, pp. 1–12. doi: 10.4028/www.scientific.net/KEM.0.1.

Descartes, S. and Berthier, Y. (2002) 'Rheology and flows of solid third bodies: Background and application to an MoS1.6coating', *Wear*, 252(7–8), pp. 546–556. doi: 10.1016/S0043-1648(02)00008-X.

Devos, A. . *et al.* (2007) 'Classification of Brain Tumours by Pattern Recognition of Magnetic Resonance Imaging and Spectroscopic Data', in *Outcome Prediction in Cancer*. Elsevier, pp. 285–318. doi: 10.1016/B978-044452855-1/50013-1.

Dienwiebel, M. *et al.* (2004) 'Superlubricity of Graphite', *Physical Review Letters*. American Physical Society, 92(12), p. 126101. doi: 10.1103/PhysRevLett.92.126101.

Dobeš, M. *et al.* (2004) 'Human eye iris recognition using the mutual information', *Optik*, 115(9), pp. 399–404. doi: 10.1078/0030-4026-00388.

Dowson, D. (1998) *History of tribology*. Professional Engineering Pub. Available at: https://books.google.fr/books/about/History_of_Tribology.html?id=5GgfAQAAIAAJ&redir_esc=y (Accessed: 11 November 2018).

Dowson, D., University of Leeds. Institute of Tribology. and Institut national des sciences appliquées de Lyon. (1996) *The third body concept : interpretation of tribological phenomena : proceedings of the 22nd Leeds-Lyon Symposium on Tribology, held in the Laboratoire de mécanique des contacts, Institut national des sciences appliquées de Lyon, France, 5th-8th September 1995*. Elsevier.

Dudoit, S. and Fridlyand, J. (2002) 'A prediction-based resampling method for estimating the number of clusters in a dataset', *Genome Biology*. BioMed Central, 3(7), p. research0036.1. doi: 10.1186/gb-2002-3-7-research0036.

Due Trier, Ø., Jain, A. K. and Taxt, T. (1996) 'Feature extraction methods for character recognition-A survey', *Pattern Recognition*. Pergamon, 29(4), pp. 641–662. doi: 10.1016/0031-3203(95)00118-2.

El-Habil, A. M. (2012) 'An Application on Multinomial Logistic Regression Model', *Pakistan Journal of Statistics and Operation Research*, 8(2), p. 271. doi: 10.18187/pjsor.v8i2.234.

F. Li, C. Xu, G.-Q. Ren, and J.-W. G. (2005) 'Image Segmentation of Ferrography Wear Particles Based On Mathematical Morphology', *Journal Of Nanjing University of Science and Technology*, 29(1), pp. 70–72.

Feng-Hsiung Hsu (1999) 'IBM's Deep Blue Chess grandmaster chips', *IEEE Micro*, 19(2), pp. 70–81. doi: 10.1109/40.755469.

Filippov, A. E. *et al.* (2008) 'Torque and Twist against Superlubricity', *Physical Review Letters*. American Physical Society, 100(4), p. 046102. doi: 10.1103/PhysRevLett.100.046102.

Fillot, N. *et al.* (2004) 'Le rôle des particules adhésives dans l'usure d'un contact sec.' Available at: https://hal.archives-ouvertes.fr/hal-00507340 (Accessed: 18 January 2019).

Gatos and C., H. (1968) 'Structure of surfaces and their interactions'. Available at: https://ntrs.nasa.gov/search.jsp?R=19690021024 (Accessed: 11 November 2018).

Godet, M. (1984) 'The third-body approach: A mechanical view of wear', *Wear*. Elsevier, 100(1–3), pp. 437–452. doi: 10.1016/0043-1648(84)90025-5.

Godet, M. (1990) 'Third-bodies in tribology', *Wear*. Elsevier, 136(1), pp. 29–45. doi: 10.1016/0043-1648(90)90070-Q.

Gouider, M. *et al.* (2004) 'Mass spectrometry during C/C composite friction: carbon oxidation associated with high friction coefficient and high wear rate', *Wear*. Elsevier, 256(11–12), pp. 1082–1087. doi: 10.1016/S0043-1648(03)00534-9.

Gurchiek, R. D. *et al.* (2019) 'Sprint assessment using machine learning and a wearable accelerometer', *Journal of Applied Biomechanics*. Human Kinetics Publishers Inc., 35(2), pp. 164–169. doi: 10.1123/jab.2018-0107.

He, P., Fang, K. L. and Liu, X. H. (2013) 'Improved watershed algorithm based on morphology and distance transform', in *Applied Mechanics and Materials*, pp. 1071–1075. doi: 10.4028/www.scientific.net/AMM.333-335.1071.

Holmberg, K. and Erdemir, A. (2017) 'Influence of tribology on global energy consumption, costs and emissions', *Friction*, 5(3), pp. 263–284. doi: 10.1007/s40544-017-0183-5.

Hornik, K. (1991) 'Approximation capabilities of multilayer feedforward networks', *Neural Networks*. Pergamon, 4(2), pp. 251–257. doi: 10.1016/0893-6080(91)90009-T.

Hosmer, D. W. and Lemeshow, S. (2000) *Applied logistic regression*. Wiley. Available at: https://www.researchgate.net/publication/320432727_Application_of_Binary_Logistic_Regression_in_Clinical_Research (Accessed: 1 November 2019).

Hunt, T. M. (1993) *Handbook of wear debris analysis and particle detection in liquids*. Elsevier Applied Science. Available at: https://www.springer.com/gp/book/9781851669622 (Accessed: 7 March 2019).

IBM (2016) '10 Key Marketing Trends for 2017', p. 18. Available at: https://www-01.ibm.com/common/ssi/cgi-bin/ssialias?htmlfid=WRL12345USEN (13.10.2017).

Janocha, K. and Czarnecki, W. M. (2017) 'On Loss Functions for Deep Neural Networks in Classification'. Available at: http://arxiv.org/abs/1702.05659 (Accessed: 31 October 2019).

Jost, H. P. (1966) *Lubrication (tribology) : education and research; a report on the present position and industry's needs.* London: H.M.S.O. Available at: https://copac.jisc.ac.uk/id/1171591?style=html (Accessed: 12 March 2019).

Kalyan, K. *et al.* (2014) 'Artificial neural network application in the diagnosis of disease conditions with liver ultrasound images.', *Advances in bioinformatics*. Hindawi, 2014, p. 708279. doi: 10.1155/2014/708279.

Kasem, H. *et al.* (2007) 'Interdependence between wear process, size of detached particles and CO_2 production during carbon/carbon composite friction', *Wear*. Elsevier, 263(7–12), pp. 1220–1229. doi: 10.1016/J.WEAR.2007.01.077.

Kaye, B. H. (1993) 'Applied Fractal Geometry and the Fineparticle Specialist. Part I: Rugged boundaries and rough surfaces', *Particle & Particle Systems Characterization*. John Wiley & Sons, Ltd, 10(3), pp. 99–110. doi: 10.1002/ppsc.19930100302.

Kearns, M. J. and Vazirani, U. V. (1994) *An introduction to computational learning theory*. MIT Press. Available at: https://mitpress.mit.edu/books/introduction-computational-learning-theory (Accessed: 11 November 2018).

Kehler, A. (1997) 'Probabilistic Coreference in Information Extraction'. Available at: https://www.aclweb.org/anthology/W97-0319/ (Accessed: 2 November 2019).

Kennedy, F. E. (1984) 'Thermal and thermomechanical effects in dry sliding', *Wear*. Elsevier, 100(1–3), pp. 453–476. doi: 10.1016/0043-1648(84)90026-7.

Kowandy, C. *et al.* (2007) 'Correlation between the tribological behaviour and wear particle morphology—case of grey cast iron 250 versus Graphite and PTFE', *Wear*. Elsevier, 262(7–8), pp. 996–1006. doi: 10.1016/J.WEAR.2006.10.015.

Laghari, M. S., Memon, Q. A. and Khuwaja, G. A. (2007) 'Knowledge Based Wear Particle Analysis', (January), pp. 998–1002.

Layard, A. . (1853) *Discoveries in the Ruins of Nineveh and Babylon , Iand II.*

LeCun, Y. and Cortes, C. (2010) 'MNIST handwritten digit database.' Available at: https://www.bibsonomy.org/bibtex/2935bad99fa1f65e03c25b315aa3c1032/mhwombat (Accessed: 11 November 2019).

Li Deng (2012) 'The MNIST Database of Handwritten Digit Images for Machine Learning Research [Best of the Web]', *IEEE Signal Processing Magazine*, 29(6), pp. 141–142. doi: 10.1109/MSP.2012.2211477.

Ling, F. F. and Simkins, T. E. (1963) 'Measurement of Pointwise Juncture Condition of Temperature at the Interface of Two Bodies in Sliding Contact', *Journal of Basic Engineering*, 85(3), p. 481. doi: 10.1115/1.3656655.

Liu, H. *et al.* (2016) 'The Segmentation of Wear Particles Images Using   J   -Segmentation Algorithm', *Advances in Tribology*, 2016, pp. 1–10. doi: 10.1155/2016/4931502.

Lofficial, G. and Berthier, Y. (1987) 'Load carrying in slow reciprocating mechanisms', *Tribology Series*. Elsevier, 12, pp. 281–290. doi: 10.1016/S0167-8922(08)71077-0.

Mayooran, T. (2018) 'A Gradient-Based Optimization Algorithm for Ridge Regression by Using R', *nternational Journal of Research and Scientific Innovation (IJRSI*, Volume V(Issue IV). Available at: https://www.researchgate.net/publication/323915259_A_Gradient-Based_Optimization_Algorithm_for_Ridge_Regression_by_Using_R (Accessed: 1 November 2019).

Meng Li (2011) 'An intelligent fault diagnosis system of rolling bearing', in *Proceedings 2011 International Conference on Transportation, Mechanical, and Electrical Engineering (TMEE)*. IEEE, pp. 544–547. doi: 10.1109/TMEE.2011.6199261.

Mollon, G. (2016) 'A multibody meshfree strategy for the simulation of highly deformable granular materials', *International Journal for Numerical Methods in Engineering*. John Wiley and Sons Ltd, 108(12), pp. 1477–1497. doi: 10.1002/nme.5258.

Mollon, G. (2018) 'Mixtures of hard and soft grains: micromechanical behavior at large strains', *Granular Matter*. Springer New York LLC, 20(3). doi: 10.1007/s10035-018-0812-3.

Mollon, G. and Zhao, J. (2012) 'Fourier-Voronoi-based generation of realistic samples for discrete modelling of granular materials', *Granular Matter*, 14(5), pp. 621–638. doi: 10.1007/s10035-012-0356-x.

Müser, M. H., Wenning, L. and Robbins, M. O. (2001) 'Simple Microscopic Theory of Amontons's Laws for Static Friction', *Physical Review Letters*. American Physical Society, 86(7), pp. 1295–1298. doi: 10.1103/PhysRevLett.86.1295.

Myasnikov, E. V. (2017) 'HYPERSPECTRAL IMAGE SEGMENTATION USING DIMENSIONALITY REDUCTION AND CLASSICAL SEGMENTATION APPROACHES', *Computer Optics*, 41(4), pp. 564–572. doi: 10.18287/2412-6179-2017-41-4-564-572.

Myshkin, N. K., Markova, L. V and Andrei, Y. G. (2005) 'Condition monitoring and predictive analysis of tribosystems by wear debris', *Practicing Oil Analysis* , 7, pp. 3–10. Available at: https://www.researchgate.net/publication/288656758_Condition_monitoring_and_predictive_analysis_of_tri bosystems_by_wear_debris (Accessed: 21 January 2019).

Nigam, K., Lafferty, J. D. and McCallum, A. (1999) 'Using Maximum Entropy for Text Classification'. Available at: https://www.semanticscholar.org/paper/Using-Maximum-Entropy-for-Text-Classification-Nigam-Lafferty/656859af2ed88cfa23f2bd063c1816a8fc04c47e (Accessed: 2 November 2019).

Pham, D. L., Xu, C. and Prince, J. L. (2000) 'Current Methods in Medical Image Segmentation', *Annual Review of Biomedical Engineering*. Annual Reviews 4139 El Camino Way, P.O. Box 10139, Palo Alto, CA 94303-0139, USA , 2(1), pp. 315–337. doi: 10.1146/annurev.bioeng.2.1.315.

Podsiadlo, P. and Stachowiak, G. . (1997) 'Characterization of surface topography of wear particles by SEM stereoscopy', *Wear*. Elsevier, 206(1–2), pp. 39–52. doi: 10.1016/S0043-1648(96)07324-3.

Price, A. L. and Roylance, B. J. (1998) 'Detection and diagnosis of wear through oil and wear debris analysis', in *Handbook of Condition Monitoring*. Dordrecht: Springer Netherlands, pp. 377–419. doi: 10.1007/978-94-011-4924-2_15.

Rong-En Fan *et al.* (2008) 'LIBLINEAR: A Library for Large Linear Classification', *The Journal of Machine Learning Research*, 9, pp. 1871–1874. Available at: https://dl.acm.org/citation.cfm?id=1442794 (Accessed: 4 November 2019).

Rosenblatt, F. (1958) 'The perceptron: A probabilistic model for information storage and organization in the brain.', *Psychological Review*, 65(6), pp. 386–408. doi: 10.1037/h0042519.

Rouet-Leduc, B. *et al.* (2017) 'Machine Learning Predicts Laboratory Earthquakes', *Geophysical Research Letters*. Blackwell Publishing Ltd, 44(18), pp. 9276–9282. doi: 10.1002/2017GL074677.

Rowley, H. A., Baluja, S. and Kanade, T. (1998) 'Neural network-based face detection', *IEEE Transactions on Pattern Analysis and Machine Intelligence*. IEEE Computer Society, 20(1), pp. 23–38. doi: 10.1109/34.655647.

Roylance, B. J. and Hunt, T. M. (1999) *The wear debris analysis handbook*. 1st ed. Oxford UK: Coxmoor Pub. Co. Available at: https://www.worldcat.org/title/wear-debris-analysis-handbook/oclc/40966389 (Accessed: 30 January 2019).

Roylance, B. J. and Raadnui, S. (1994a) 'The morphological attributes of wear particles — their role in identifying wear mechanisms', *Wear*, 175(1–2), pp. 115–121. doi: 10.1016/0043-1648(94)90174-0.

Roylance, B. J. and Raadnui, S. (1994b) 'The morphological attributes of wear particles — their role in identifying wear mechanisms', *Wear*, 175(1–2), pp. 115–121. doi: 10.1016/0043-1648(94)90174-0.

Roylance, B. J., Williams, J. A. and Dwyer-Joyce, R. (2000) 'Wear debris and associated wear phenomena—fundamental research and practice', *Proceedings of the Institution of Mechanical Engineers, Part J: Journal of Engineering Tribology*, 214(1), pp. 79–105. doi: 10.1243/1350650001543025.

Rumelhart, D. E., Hinton, G. E. and Williams, R. J. (1985) 'Learning Internal Representations by Error Propagation'. Available at: https://apps.dtic.mil/docs/citations/ADA164453 (Accessed: 4 November 2019).

Scharf, T. W. and Singer, I. L. (2003) 'Quantification of the thickness of carbon transfer films using Raman tribometry', *Tribology Letters*, 14(2), pp. 137–146. doi: 10.1023/A:1021942822261.

Scherge, M., Shakhvorostov, D. and Pöhlmann, K. (2003) 'Fundamental wear mechanism of metals', *Wear*. Elsevier, 255(1–6), pp. 395–400. doi: 10.1016/S0043-1648(03)00273-4.

Sejdić, E. and Falk, T. H. (2018) *Signal Processing and Machine Learning for Biomedical Big Data.* Available at: https://www.crcpress.com/Signal-Processing-and-Machine-Learning-for-Biomedical-Big-Data/Sejdic-Falk/p/book/9781498773454 (Accessed: 10 January 2019).

Silver, D. *et al.* (2017) 'Mastering the game of Go without human knowledge', *Nature*. Nature Publishing Group, 550(7676), pp. 354–359. doi: 10.1038/nature24270.

Sliney, H. E. (1978) 'Dynamics of Solid Lubrication as Observed by Optical Microscopy', *A S L E Transactions*, 21(2), pp. 109–117. doi: 10.1080/05698197808982866.

Souchon, F., Renaux, P. and Berthier, Y. (1998) 'Helical scan head and tape contact behavior: optimization of tribological and magnetic aspects', *Tribology International*, 31(8), pp. 479–484. doi: 10.1016/S0301-679X(98)00074-7.

Srinivasan, G. and Shobha, G. (2008) 'Statistical Texture Analysis'. Available at: https://www.semanticscholar.org/paper/Statistical-Texture-Analysis-Srinivasan-Shobha/6563afbb8c50452a70d0c002e9feb0fa5e1274d4 (Accessed: 31 May 2019).

Stachowiak, G. W., Kirk, T. B. and Stachowiak, G. B. (1991) 'Ferrography and fractal analysis of contamination particles in unused lubricating oils', *Tribology International*. Elsevier, 24(6), pp. 329–334. doi: 10.1016/0301-679X(91)90002-Q.

Starr, A. *et al.* (2001) 'Data fusion applications in intelligent condition monitoring'. Available at: https://www.semanticscholar.org/paper/Data-fusion-applications-in-intelligent-condition-Starr-Willetts/8de24748f6692b7dc7fefdcbbafa5ed98ef870ad (Accessed: 30 January 2019).

Stuart, R. and Peter, N. (2005) *Artificial Intelligence : A modern approach*, *Belgian Journal of Zoology*. doi: 10.1017/S0269888900007724.

Surapol, R. and BJ, R. (1995) 'The classification of wear particle shape', *Lubrication Engineering*, 51(5), pp. 432–437. Available at: https://www.researchgate.net/publication/279903465_The_classification_of_wear_particle_shape (Accessed: 22 May 2019).

Tang, Y. (2013) 'Deep Learning using Linear Support Vector Machines'. Available at: http://arxiv.org/abs/1306.0239 (Accessed: 1 November 2019).

Thalamuthu, A. *et al.* (2006) 'Evaluation and comparison of gene clustering methods in microarray analysis', *Bioinformatics*, 22(19), pp. 2405–2412. doi: 10.1093/bioinformatics/btl406.

Thankachan, T., Soorya Prakash, K. and Kamarthin, M. (2018) 'Optimizing the Tribological Behavior of Hybrid Copper Surface Composites Using Statistical and Machine Learning Techniques', *Journal of Tribology*. American Society of Mechanical Engineers, 140(3), p. 031610. doi: 10.1115/1.4038688.

Thomas P. Minka (2003) 'A comparison of numerical optimizers for logistic regression'. Available at: https://www.researchgate.net/publication/237254359_A_comparison_of_numerical_optimizers_for_logistic _regression (Accessed: 4 November 2019).

Tibshirani, R. (1996) 'Regression Shrinkage and Selection via the Lasso', *Journal of the Royal Statistical Society. Series B (Methodological)*. WileyRoyal Statistical Society, pp. 267–288. doi: 10.2307/2346178.

Turing, A. M. (1950) 'COMPUTING MACHINERY AND INTELLIGENCE', *Mind*, LIX(236), pp. 433–460. doi: 10.1093/mind/LIX.236.433.

Vincent, L. (Leo V., Berthier, Y. (Yves B. and Maurice, G. (Maurice G. (1992) 'Standardization of fretting fatigue test methods and equipment', *ASTM*. ASTM, pp. 33–48. Available at: https://books.google.fr/books?id=9elVJSHFwR0C&pg=PA33&lpg=PA33&dq=ASTM,+1992,+pp.+33-48&source=bl&ots=m6nQW690zc&sig=ACfU3U1EUfP60qg7eEHZpvWNDqEviNIOAQ&hl=en&sa=X&ved=2ahUKEwjP7qyg-fbfAhWXAmMBHYZjDZkQ6AEwBHoECAIQAQ#v=onepage&q=ASTM%2C 1992%2C pp. 33 (Accessed: 18 January 2019).

Waggoner, J. *et al.* (2013) '3D Materials Image Segmentation by 2D Propagation: A Graph-Cut Approach Considering Homomorphism', *IEEE Transactions on Image Processing*, 22(12), pp. 5282–5293. doi: 10.1109/TIP.2013.2284071.

Wahl, K. J., Chromik, R. R. and Lee, G. Y. (2008) 'Quantitative in situ measurement of transfer film thickness by a Newton's rings method', *Wear*, 264(7–8), pp. 731–736. doi: 10.1016/j.wear.2007.04.009.

Wang, J. and Wang, X. (2013) 'A wear particle identification method by combining principal component analysis and grey relational analysis', *Wear*, 304(1–2), pp. 96–102. doi: 10.1016/j.wear.2013.04.021.

Wang, L., Gordon, M. and Zhu, J. (2006) 'Regularized Least Absolute Deviations Regression and an Efficient Algorithm for Parameter Tuning', in *Sixth International Conference on Data Mining (ICDM'06)*. IEEE, pp. 690–700. doi: 10.1109/ICDM.2006.134.

Win, K. Y., Choomchuay, S. and Hamamoto, K. (2017) 'Automated segmentation and isolation of touching cell nuclei in cytopathology smear images of pleural effusion using distance transform watershed method', in *Second International Workshop on Pattern Recognition*. SPIE, p. 104430Q. doi: 10.1117/12.2280807.

Woodward, A. and Needham, A. (2009) *Learning and the infant mind*. Oxford University Press.

Xu, K., Luxmoore, A. R. and Deravi, F. (1997) 'Comparison of shape features for the classification of wear particles', *Engineering Applications of Artificial Intelligence*. Pergamon, 10(5), pp. 485–493. doi: 10.1016/S0952-1976(97)00017-1.

Xu, R. and Wunsch, D. C. (2008) *Clustering*.

List of tables

List of figures