POLITECNICO
MILANO 1863

POLITECNICO DI MILANO
DEPARTMENT OF MECHANICAL ENGINEERING
DOCTORAL PROGRAMME IN PROGNOSTICS AND HEALTH MANAGEMENT IN
MACHINE TOOLS AND MANUFACTURING INDUSTRY

ROBUST PROGNOSTICS APPROACHES FOR MACHINE TOOLS UNDER VARIABLE OPERATING CONDITIONS

Doctoral Dissertation of:
Luca Bernini
Matr. Nr. 947673

Supervisor:
Prof. Albertelli Paolo

Tutor:
Prof. Giglio Marco

The Chair of the Doctoral Program:
Prof. Bernasconi Andrea

2022 – XXXV

Un ringraziamento speciale va al professor Albertelli. Il primo ad aver creduto in me, accompagnandomi dalla laurea magistrale fino al termine di questo percorso di dottorato.

Grazie al team del MUSP, e in particolar modo al professor Monno, che mi hanno ospitato presso il loro centro di ricerca, permettendomi di svolgere tutte le attività in un clima accogliente e familiare.

Grazie alla mia famiglia che mi ha sostenuto e supportato in questi anni, anche passando momenti difficili, portandoci a questa soddisfazione immensa!

Grazie a tutti gli amici. A chi mi è stato accanto più da vicino, ma anche nonostante la distanza mi ha sempre saputo far sorridere e divertire.

Abstract

DIGITAL MANUFACTURING relies on the support of data analysis tools to increase reliability, safety and performances of manufacturing systems and machine tools. In this context, Prognostics and Health Management (PHM) proposes a set of digital tools with the final objective of predicting components remaining useful life. Variable operating regimes and the need for a high volume of experimental tests makes PHM for machine tools tough and challenging for an effective industrial implementation. In this thesis, an operational regime robust PHM architecture, based on low experimental effort, is proposed both for cutting tools and machine tool auxiliaries. The PHM strategy was conceived in order to describe tool wear by taking into account the phenomenological aspects of metal cutting. Thus, a solution based on process modelling was adopted. Physical mechanistic features were provided to an advanced real-time monitoring system, robust with respect to the operational regime. A further step ahead was taken through the development of an artificial intelligence tool capable of managing unmodeled process variability. A hybrid approach for tool wear was developed to adapt to unseen degradation curves. The approach was conceived having in mind the effect of cutting parameters on the tool degradation rate. Changing the problem perspective, process and tool monitoring was performed on the basis of machined surface images processing. Inference on the cutting parameters was performed based on milling technological signature. A PHM solution was also conceived for machine tool auxiliaries, with reference to machine tool hydraulic unit. Having in mind PHM challenges, the solution was developed with the support of physical models (digital twin) in order to reduce the experimental effort.

Contents

CHAPTER 1

Introduction

Manufacturing systems and, more specifically, machine tools are the means responsible for everyday goods production. In the last decades, manufacturing industries tried to face the growing demand of manufactured goods through the transition towards robotised and automated plants. Concurrently, reliability and safety standards are continuously improving, intrinsically leading to more complex production plants, composed by a higher number of parts [162]. As any other system or object of our world, machine tools and their components are affected by degradation processes which slowly or suddenly lead to failures. Component failures play crucial roles for the economical aspects of both manufacturing companies and their customers.

Prognostics and Health Management (PHM) of systems started to gain more attention from the industrial sector, since component failures play a crucial role for the satisfaction of increased production, reliability and safety demands. PHM represents the most advanced maintenance policy currently available; ideally, it autonomously identify when a repair is needed, based on signals acquired from the sensors installed on the machines.

Developing PHM approaches for the machine tool scenario is the general objective of this thesis work, with a special focus on tool condition monitoring (TCM) and prognostics. In this chapter an overview of the PHM framework is firstly proposed (section 1.1); further insights are given into TCM and hydraulic units prognosis (sections 1.1.2 and 1.1.1); the challenges related to the conceived approaches are summarised (section 1.2); concluding with the outline of the thesis structure (section 1.3).

1.1 PHM overview

PHM is addressed as a maintenance policy that allows for the prediction of the Remaining Useful Life (RUL) of a component. Actually, PHM consists of a set of consecutive modules, that perform different tasks [14]. These modules are reported by increased complexity and amount of output information:

- monitoring;
- diagnosis;
- prognosis.

Depending on the input data availability, the desired output level of information and the application scenario, some of these modules, may or may not be implemented. Preliminary analysis is thus mandatory, since it regards the collection of information about the component or system under investigation. It is needed to understand which of the components should be analysed through PHM (monitoring, diagnosis and/or prognosis methodology) or traditional policies like reactive or preventive maintenance. Indeed, several techniques may be used to perform this categorisation: failure-mode and effect analysis (FMEA), fault-tree analysis (FTA) and root cause analysis.

FMEA is a specific tool for identifying the most relevant, and thus critical, failure modes of a system. It studies the criticalities at different system levels, i.e. it considers the machine as an ensemble of main sub-units and increases the analysis depth until the component level is reached [161]. Furthermore, the effect of each failure is analysed with respect to system performances. The main limitation of FMEA is that no interaction between the failure modes is caught [14]. FTA is a complementary tool with respect to FMEA, by describing the interaction and causality between failure modes [94]. Root cause analysis tries to find what generated a particular failure. All these tools allow for a proper design of the maintenance policy, determining which failure modes of a system require a PHM strategy or not. PHM is implementable only on components that have a massive impact on the system repair costs, even including production losses.

The first step of a PHM policy consists in the implementation of a monitoring strategy. Monitoring is at the basis of the PHM structure. It allows for a fast identification of a fault initiation. Indeed, monitoring approaches may represent the trigger for successive steps of PHM, in the case of progressive degradation of a component [14], or they are responsible for a prompt detection of a catastrophic sudden failure [156]. Monitoring can be performed with different kind of approaches, being Statistical Process Control (SPC) (and its control charts [39, 156]) the most used tool [20]. Of course, other fields of data analysis have treated the monitoring problem. The exploding growth of artificial intelligence (AI) application is for sure the most relevant one. In this context, AI approaches tend to be addressed as anomaly detection algorithms or novelty detection ones. Monitoring and anomaly detection are based on the collection of a proper dataset, acquired in nominal condition for the machine/system. This dataset constitutes what is often referred to as "baseline" [14].

The second PHM step responds to the diagnosis problem. Diagnosis is the operation of recognising and discerning different failure modes (isolation). Meanwhile, diagnosis has the objective of assessing the criticality level of an initiated fault (quantification) [14, 20, 162]. Thus, diagnosis turns out to be either a pattern recognition

or classification problem [14]. Classification problems require a dataset in order to be applied and they are traditionally solved through machine learning approaches (also referred to as *data-driven* or *black-box* approaches). In general, two main categories of algorithms can be applied in the diagnostics phase: supervised and unsupervised ones. Within the supervised scenario, the dataset is constituted by organised samples, each of them assigned to a particular class, and thus, having a dedicated label. The input data are referred to as *predictors*. Based on the values assumed by the predictors, the objective of a classification algorithm is to infer the output class, i.e. *responses* [79] (in this case, the fault kind and its associated criticality level). Within this framework, each algorithm undergoes a *training* procedure, by which a degree-of-error is computed for each prediction, and the algorithm parameters are updated accordingly. In opposition to supervised learning approaches, unsupervised algorithms need to tackle a supplementary phase. In fact, the dataset is unlabelled and the algorithm attempts to find similarity regions in the distribution of data (clustering) [14, 79]. Once clustering is performed, a supervised classification approach can be applied to assign a new and unseen sample to the most similar cluster. Nevertheless, traditional machine learning algorithms are not directly applicable to raw data. In fact, they rely upon a human-based feature extraction and selection strategy. Feature extraction is the process of synthesising and reducing the dimensionality of raw data, trying to keep the highest amount of useful information. Feature selection is a further compression step [14], where the initial pool of extracted features is reduced either by transformation or ranking strategies.

Several machine learning approaches have been designed and adopted for supervised and unsupervised learning frameworks. K-Nearest Neighbours [57,58,180], Support Vector Machines [60,91,180], Linear Discriminant Analysis [82], Artificial Neural Networks [242] and Random Forest Classifiers [24] can be found among the most used algorithms for supervised problems, whereas K-means is just an example of clustering technique [79]. With analogous tasks, recently a new group of approaches started to be successfully implemented, namely deep learning algorithms. With respect to traditional machine learning approaches, deep learning embeds both feature extraction and selection procedures, avoiding the need for data-preprocessing performed by humans. At the same time, they require a huge amount of data samples, which are not always available especially in PHM applications [209]. Convolutional Neural Networks belong to this category, consisting of a deep set of hidden layers [86].

The last step of PHM maintenance policy consists in the prognosis phase. Prognosis has the objective of predicting the future evolution of a degradation characteristics, forecasting RUL of machine or components. RUL is computed as the difference between the prediction instant and the predicted failure time. Indeed, RUL is the PHM output containing the deepest level of information, since it allows not only to have an assessment of the component condition but also allows to anticipate the time of failure. Of course, being capable of predicting RUL requires high quantity and quality of input data. In opposition to diagnosis, prognosis falls within the scenario of time series prediction problems. Scientific literature agrees in categorising prognostics approaches in four main groups: knowledge-based, statistical-based, model-based and data-driven approaches [14, 162]. The first category deals with the implementation of models based on expert knowledge of faults. Approaches of this kind try to translate experts knowledge in a set of IF-THEN rules (the case of Expert Systems), or to model the system

in a linguistic fashion (the case of fuzzy logics) [14]. Statistical-based approaches (or grey box models) describe degradation in the form of a stochastic process, bringing to a relevant formal advantage, compared with respect to the other categories. In fact, statistical based approaches allow for the estimation of RUL probability density function. This provides support to the decision making phase, since it allows to consider the degradation process uncertainties. Furthermore, the prediction of RUL probability density function is suggested by the internation standards [96]. Filtering-based techniques belong to this category: Kalman Filter [64] and its extensions [65]; particle filters, capable of relaxing all their assumptions [137, 203]. Another example is constituted by Hidden Markov Models [220]. Model-based approaches guarantee the highest prediction performances, by modelling the degradation phenomenon starting from first principles [14]. Being the degradation formulation known a priori, fewer parameters must be estimated with respect to grey-box models, but they are available only in few applications. Some of them are reported as examples in the following: exponential lifetime prediction model for ball screw mechanisms under different feed modes [35], differential models for tool wear evolution in milling [235] and wear model for flank wear in turning [168]. At last, black-box models are represented by AI approaches and machine/deep learning techniques. They can learn complex non-linear relations within data, providing really flexible tools for prognostics. Their main shortcoming is associated to the amount of data necessary to build the model itself. On one hand, no physical law of the degradation process is needed to build the model, on the other hand, they lack of physical interpretation. ANN, Self-Organising-Maps (SOM) [37, 142, 193] and deep learning algorithms, SVM and Relevance Vector Machines (RVM) [88] are just a few examples of artificial intelligence techniques.

More recently, despite the rigorous grouping system, some of these main categories are combined in what are referred to as hybrid approaches, in order to enhance the strengths of the involved categories and reduce their shortcomings [20].

The above overview of PHM methodologies is valid for a general machine or system, independently from the field of application. Anyway, PHM can be applied at different levels of the system. Sub-units can be considered as a first subdivision, as shown in figure 1.1. Following the figure, in this thesis the focus is reserved to the two critical units/components of a machine tool: hydraulic units and cutting tools. In the next sections, some insights into PHM for these two areas are given.

1.1.1 Hydraulic units prognostics

The role of machine tools consists in the correct positioning of the end-effector which is typically a cutting tool. In order to reach this objective, machine tools are constituted by several units (or functional groups, fig. 1.1). Machine tool auxiliaries are represented for instance by spindle unit or by the feed-drive system. By the study of reliability analysis of CNC machine tools, the hydraulic unit resulted to be one of the most faulty ones and the cause of unexpected breakdowns and downtimes [29, 161]. Several CNC lathes were analysed in a five year period, from 2009 to 2014, and the hydraulic subsystem showed the highest failure rate (22.9% of the total failures) [29]. Twelve machining centres in another five year project, between 2005 and 2010, confirmed the result [217]. The necessity for the development of PHM on machine tool auxiliaries was highlighted by other two works [62, 63], addressing them as sources of unexpected downtimes,

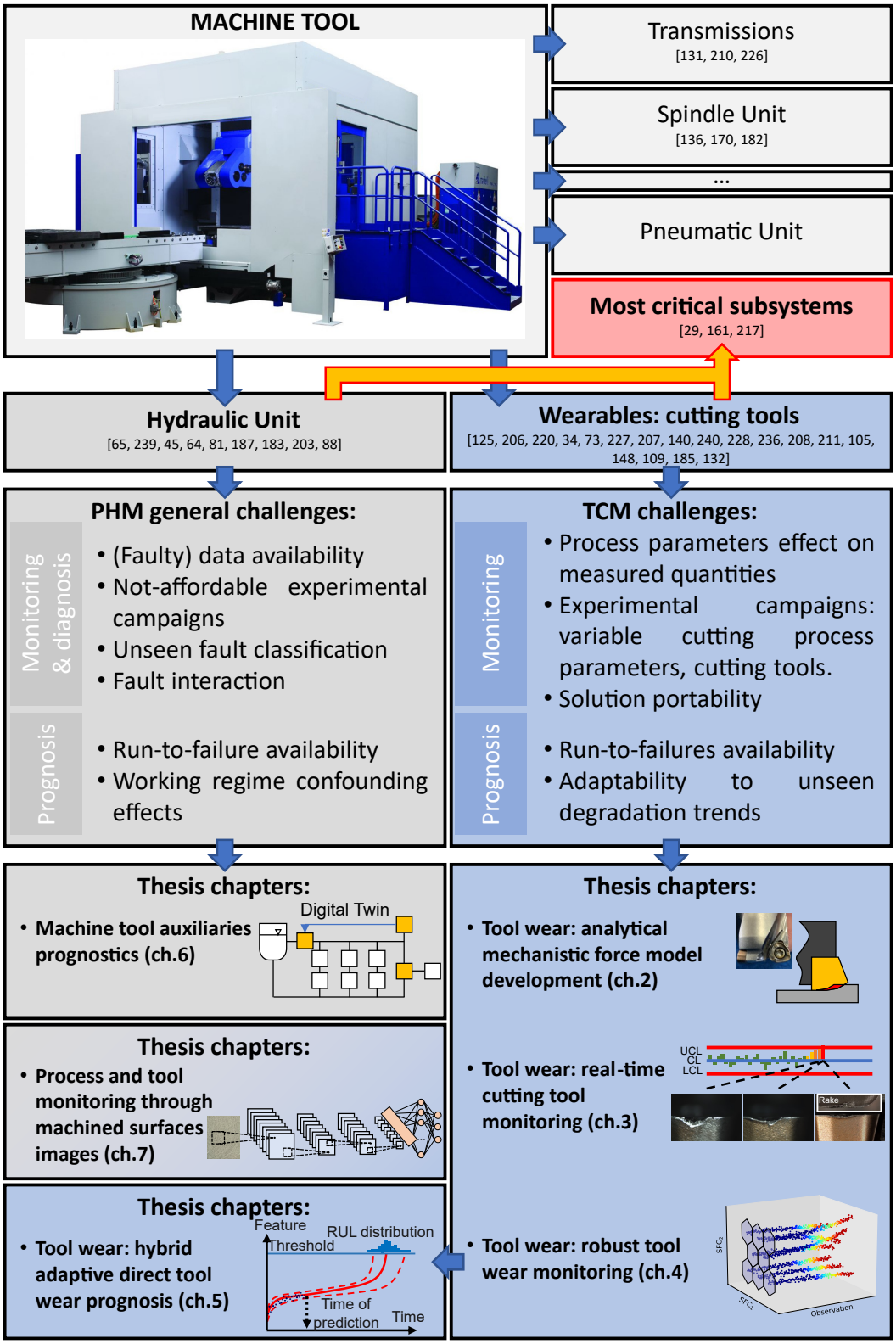


Figure 1.1: *Map of literature works and thesis chapter activities related to prognostics and health management on machine tool units. Challenges related to the analysed units are brought to the reader attention.*

with comparable losses to main machine tool components, thus the ability to foresee an incipient failure would certainly be a significant advantage for any customer [14]. In machine tools, several hydraulic circuits are present: high-pressure circuits are typically dedicated to the automatic tool changing unit; low pressure circuits are in charge of lubrication of main transmission and motors; other mid or low pressure circuits may be present for other auxiliary tasks, like hydraulic brakes. Finally, machine tools also feature a refrigerant hydraulic circuit and one for the addition of lubro-refrigerant cutting fluid [14]. Despite the above, research literature on PHM for hydraulic units and its components (e.g. pumps with their motor, pipes, valves, accumulators, and actuators) are scarce. In the aeronautic field, few works dealt with hydraulic circuit leakages. Research papers on oil monitoring can be found, but prognosis works are limited [198]. Despite some paper dealt with pump prognostics, not all the different kinds of pumps have been treated. For instance, gear pumps, commonly installed on machine tools, were not investigated yet [14].

1.1.2 Tool condition monitoring and prognosis

Tool condition monitoring (TCM) is a branch of PHM which makes reference to cutting tools. Indeed, the main objective of this policy regards the assessment of the cutting tool condition. Tool condition monitoring and prognosis impact on a manufacturing company is twofold: economical and environmental. In fact, the use of TCM solutions allow to optimally exploit tool life, avoiding dramatic failure and reducing early changes. This could be easily integrated with a resharpening policy, leading to dramatical advantages in terms of economical savings (for cutting tool users), while bringing increased profits (tool manufacturers), critical material reduction (e.g. Cobalt and Tungsten), CO_2 emissions and energetic footprint decreases. An example analysis of the impact of a combination of these two strategies is presented in the following[1]. Prognostics may allow an extension of 10% of cutting tool lives and the cost of resharpened cutting tools is assumed to be the 80% of a new one, resulting in a saving of 12% on the cutting tools cost (no production gain and scrap reduction is accounted for). On average a cutting tool can be resharpened about 4 times, with a 25% chance that a tool can not be resharpened at all. Resharpening doesn't need any additional material (thus, reducing of 80% Tungsten and Cobalt consumption) while recycling after 4 resharpenings allows saving the 95% of the employed Tungsten. Emitted CO_2 accounts for about 70% (when resharpening) with respect to a new one, while only 40% of the CO_2 (with respect to a new one) is emitted when recycling a cutting tool. The energetic footprint of a resharpened tool corresponds to about 30% out of the energetic footprint of a new cutting tool (75% when recycling a tool).

In literature, tool wear is addressed as a complex stochastic phenomenon [233], since it is the effect of several different causes. Thus, in order to make forecasts of its evolution, it is necessary to define a significant indicator that can well describe the status of the tool. Flank wear measurement is the most relevant and common degradation indicator for cutting tools [119, 156], so that almost any direct or indirect tool is benchmarked with its value.

[1] Assumptions are based on cutting tool market trends by the European Remanufacturing Network (https://www.remanufacturing.eu/), Sandvik Coromant (https://www.sandvik.coromant.com/en-gb/pages/default.aspx), MCM S.p.A. (https://www.mcmspa.it/?lSite=en), HTS Dynamics AS (https://www.htsdynamics.no/), ZUBIOLA S. Coop. (https://www.zubiola.es/en/inicio)

Literature treats TCM by means of two main algorithm categories: techniques either based on direct wear measurements or indirect wear measurements [60]. The first class of approaches relies on the analysis of quantities directly correlated to the degradation of the cutting edge, i.e. they are typically based on flank wear width measurements or on its extrapolation from cutting edge pictures. Direct inspection could be obtained through point wise scanning, profilometric acquisitions [151] or by 2D/3D calibrated pictures analysis [55, 139, 151]. Direct inspection methods have the advantage of providing a more robust tool for the assessment of tool condition, since the extrapolated flank wear measurement is slightly influenced by process parameters. Nevertheless, direct methods still have some shortcomings, like the accuracy of the inspection system and the post-processing of the measurements (e.g. image processing sensibility to light exposure, brightness, metal residuals and material reflectiveness). A second limitation in their applicability regards the fact that machining processes should be stopped every time the cutting tool have to be exposed to the inspection system. Direct measurements based approaches are not suitable for in-process monitoring. Nevertheless, the inspection task could be performed at periodic stops in masked time, for instance at tool changes, or triggered by user defined events according to optimised production strategies, providing useful tools to be integrated with indirect methods. Indirect methods are based on the observation of process related quantities, carrying indirect information of tool wear. The commonly adopted variables include axis torque or current [104], vibrations [194], acoustic emissions [211] and cutting forces [156]. Their advantage lays in a non-invasive measurement, thus being adequate tools for in-process TCM.

As said above, TCM generally performs only an assessment of tool condition, thus telling to the operator whether the tool is in acceptable conditions or not, at the assessment instant. Recently, following the PHM architecture, prognosis tools were applied in order to perform the prediction of the evolution of tool wear during the cutting process, i.e. to compute the remaining useful life (RUL) of cutting tools. The most used category of prognosis approaches is machine learning or data-driven methods. Flank wear evolution was described by the application of Support Vector Regression, starting from cutting forces, vibrational signals and machined surface pictures [34]. DenseNets, from the deep learning field, were used in order to correlate flank wear to cutting forces, vibrations and acoustic emission signals [73]. Convolutional neural networks were introduced to predict RUL of cutting tools starting from current, vibrations and acoustic emission signals [227]. Wang et al. conceived a hybrid methodology based on machine learning capable of integrating heterogeneous data (process parameters, power profiles and tool wear images). Wear severity was assessed by a convolutional neural network, while RUL prediction were performed by a recurrent neural network [207]. Even model-based solutions are emerging, for instance by including tool wear information in cutting force models. The model-based approach could be used for tool flank predictions through cutting force and temperature measurements [140]. Hidden Markov Models were applied for the prediction of tool flank wear, starting from indirect measurements (vibrations and acoustic emission signals) [220]. Particle filter was used to update the parameters of a linear degradation model by in-process vibrational measurements [224].

1.2 Challenges and objectives

PHM obviously received a great boost with the advent of the Industry 4.0 paradigm, through the availability of low cost sensors of any kind, the increased available computational resources, the interconnection between machine through the Internet of Things and the newly available Big Data analysis techniques. Nevertheless, the application of PHM strategies in the manufacturing field is still at its infancy. Several researches have proposed interesting solutions for the applicability of PHM in this field, but few of them were tested or found an application in a real industrial scenario [14, 162]. There are commonly agreed challenges that mine a robust application of PHM in the manufacturing industrial sector (also highlighted in fig. 1.1):

1. availability of data: despite the increasing effort towards a digitalisation of the manufacturing systems, it is worth noting that building prognosis-enabling datasets is far from being an easy task. The sensorisation of manufacturing systems and machine tools is just the first step of a bigger process. The creation of datasets for supervised learning approaches requires tons of data coming from the nominal and faulty machine. Furthermore, in order to be able to predict the RUL through prognosis approaches requires not only clustered datasets (i.e., the ones necessary for the application of diagnosis), but run-to-failure datasets. It must be pointed out that run-to-failures are experimental tests where a system, machine or component is operative from its infancy up to its failure. These experimental tests constitute the main limitation in the creation of prognosis datasets both for their economical and temporal impact on companies. Thus, the **minimisation of experimental run-to-failures** is the first objective for this thesis work.

2. operating conditions: most of the PHM works published within the scientific community regard applications with a single operating regime. Indeed, this context is the reference for high-volume production, where a large stock of identical products is produced. In this scenario, the acquired signals refer to the exact same operations through time, enabling the creation of effective prognosis datasets. However, the manufacturing field is full of low-medium volume production companies, where goods are produced in small batches or even as one-of-a-kind (e.g. aeronautical sector). This is the typical context for machine tools application. It must be highlighted that this challenge is strictly connected to the first one. A proper prognosis dataset should in fact contain run-to-failures performed in multiple working regimes. Thus, the **creation of an operational regime robust PHM framework** is the second main objective of this work.

3. static approaches: most of the literature methods rely upon static algorithms for the estimation of components health and their evolution forecast. This means that models are built (or learned, when referring to data-driven ones) and then applied as is. This has a twofold implication on the approaches effectiveness: the amount of data (even bigger than the training set) received during the implementation of the machine is not exploited; models for health predictions are not updated as long as data arrive from the system. For these reasons, **developing a hybrid and adaptive PHM framework for the machine tool field**, constitutes the third objective of this thesis.

1.3 Thesis structure

In order to present the thesis outline, it is useful to make reference to its graphical representation reported in figure 1.2. The thesis is composed of two main areas, highlighted with blue and grey backgrounds in figure 1.2.

The first area constitutes the main part of the thesis work and regards the development of a full PHM framework in the context of cutting tools wear, from process modelling up to tool wear prognostics. The choice of focusing on cutting tools is related to two central reasons. The first one is associated to the economical (and environmental) impact of tool consumption and substitution on manufacturing companies. Cutting tool provides large cost saving possibilities; up to 40% could be achieved through the monitoring of its health [185], while about 20% of the downtimes of machine tools was attributed to tool failures, resulting in reduced productivity and economic losses [121, 227]. The second reason is that the first two previously identified challenges are of paramount importance within the application of PHM to cutting tools.

Chapter 2 lays the foundation for tool condition monitoring (TCM). The idea is to develop a prognosis framework based catching the phenomenological aspects of metal cutting. Thus, process modelling becomes a useful mean to extract physical features, independent from the cutting conditions. Indeed, an analytical mechanistic force model for high-feed mills was developed, since they were employed for the experimental run-to-failures and a reference model is still not available in literature. These cutting tools featured a double-phased cutting edge with very low lead angles, that needed a proper formulation in order to take into account the variable engagement condition of the tool along the mill axial coordinate. The mechanistic model allowed to properly identify specific force coefficients (SFC) through traditional mean forces approaches. Chapter 3 makes the transition towards TCM. SFC are now identified in real time regressing instantaneous cutting forces acquisitions. SFC will represent a meaningful tool wear indicator, highlighting brittle degradation phenomena, while being theoretically independent from the cutting conditions. Chapter 4 regards the implementation of an unsupervised AI approach to deal with the residual variability included in the SFC estimation. This technique will allow for unsupervised description of gradual wear evolution of cutting tools. At last, in chapter 5 a hybrid and adaptive prognosis approach will be developed for cutting tool RUL prediction. The proposed algorithm will be capable of estimating the RUL probability density function, being updated online through direct wear measurements of turning cutting tools. The statistical and AI hybrid approach will estimate RUL in variable cutting conditions, learning just from one run-to-failure test (i.e., from a fixed set of cutting parameters and lubricating condition).

The second area is constituted by two parallel activities related to PHM in the machine tools context. In chapter 6, a full PHM framework was developed and presented for machine tool auxiliaries. Hydraulic unit was chosen as the subject for this activity, since it is recognised as one of the most critical parts and the cause of unexpected breakdowns and downtimes for machine tools [29, 161, 217]. Despite the above, research on PHM is rare for such subsystem. The conceived solution is based on the development of a digital twin of the system, capable of reproducing variable operating regimes and multiple concurring component faults. Datasets are built by digital twin simulations, while a diagnosis and prognosis solution is conceived and tested on them.

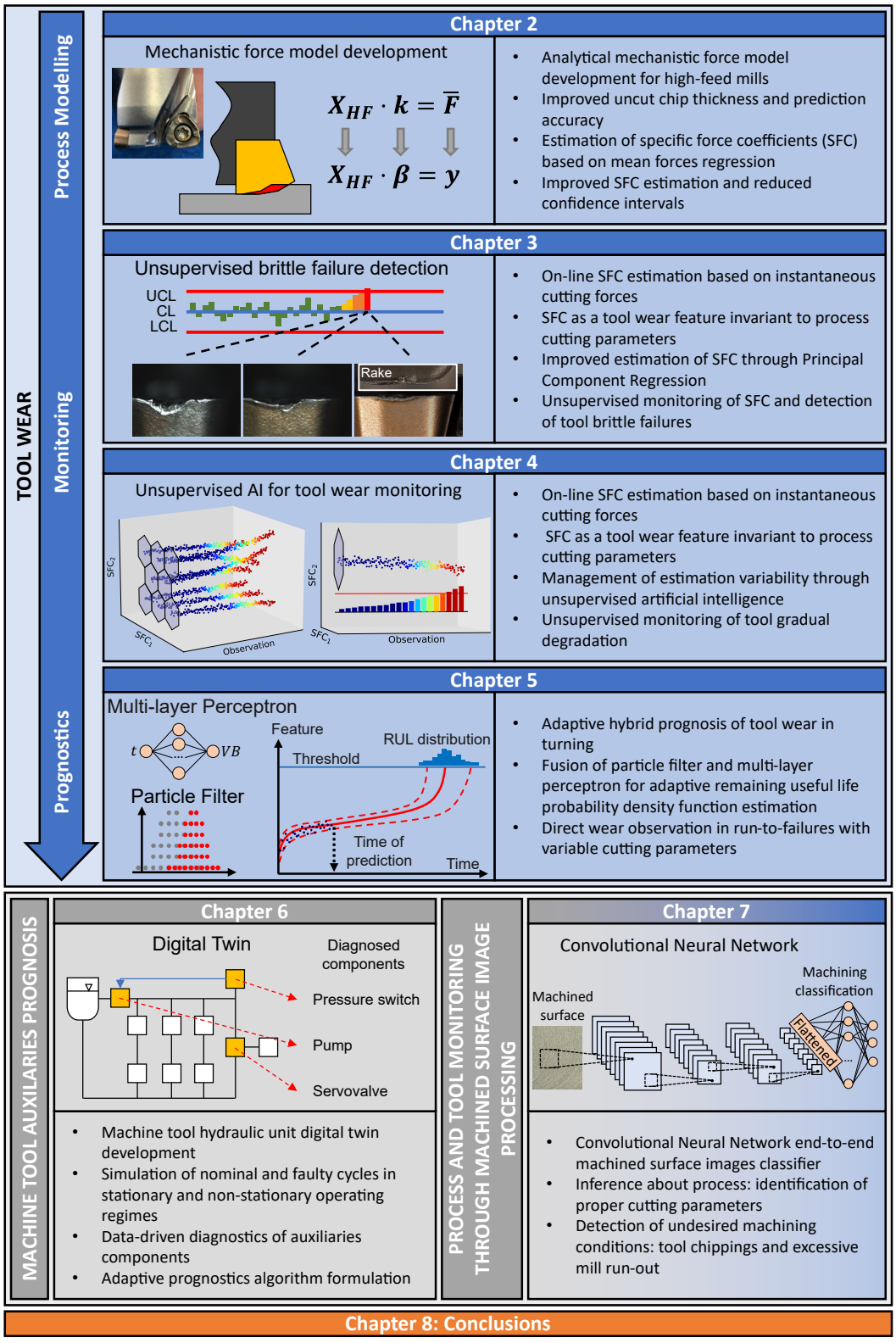


Figure 1.2: *Graphical representation of the thesis structure. Two main areas constitute the thesis structure: blue coloured areas refer to the tool wear PHM framework; grey areas include chapter 6 regarding a PHM strategy for machine tool auxiliaries, and chapter 7 for tool condition monitoring performed from the workpiece perspective.*

Table 1.1: *Fault states and associated HI for each component.*

Title	A novel prognostics solution for machine tool sub-units: the hydraulic case
Authors	Bernini L., Waltz D., Albertelli P., Monno M.
Journal	Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture
Status	Published [20]
Chapter	6
Title	Mechanistic force model for double-phased high-feed mills.
Authors	Bernini L., Albertelli P., Monno M.
Journal	International Journal of Mechanical Sciences
Status	Pre-proof published [18]
Chapter	2
Title	Mill condition monitoring based on instantaneous identification of specific force coefficients under variable cutting conditions
Authors	Bernini L., Albertelli P., Monno M.
Journal	Mechanical Systems and Signal Processes
Status	Published [19]
Chapter	3

In chapter 7, TCM is carried out within another perspective. The analysis of machined surfaces is performed through the construction of a deep learning convolutional neural network classifier. The proposed solution is capable of distinguishing between nominal and undesired effects, such as tool chippings and excessive run-out.

At last conclusion and possible future works are reported in the thesis conclusions (chapter 8).

1.4 Publications

Publications are mapped on thesis chapters through table 1.1:

CHAPTER 2

Tool wear: analytical mechanistic force model development

In this chapter, the framework of tool wear prognosis begins. The idea is to develop a solution where features for tool monitoring and prognosis catch physical and phenomenological aspects of the cutting process as well as tool wear. On this basis, mechanistic force models allow to describe through cutting condition independent coefficients the ongoing cutting process. Such coefficients will thus constitute the subject of the analysis of the next thesis chapters. Here, the analytical mechanistic model is conceived for high-feed mills in order to cover the associated research gap.

2.1 Introduction

Milling is a flexible chip removal process that allows for the production of several kind of goods such as aeronautical and biomedical ones [28]. The capability of predicting the forces acting on the tool is of fundamental importance. In fact, cutting forces influence on the milling operation is twofold: on one hand, they are representative of the performances of the operation; on the other hand, they allow for the estimation of related quantities, such as torque and power [231], that are needed to check the feasibility of the milling operation on the available machine [66], as well as to compare multiple operations in a sustainability perspective. Cutting forces are the cause for machine tool vibrations, that can be either forced [221], being responsible for reduced workpiece surface quality and geometrical accuracy, or regenerative [44], mining the process stability [2,201,202,204]. Indeed, cutting force models are the main sources of information and mathematical tools for stability analysis [9,178,218] and even chip formation mechanism determination [31]. Furthermore, within the Industry 4.0 paradigm, cutting forces and related quantities can be used to predict the remaining useful life of

the milling tool [2, 47, 140, 159]. Being able to predict cutting forces beforehand, allows to compare them with nominal ones. At last, cutting forces are the main source of information to analyse the energetic behaviour of a process in terms of specific cutting energies, which represent the main focus of this chapter.

In literature, force prediction models in milling applications are grouped in three main categories: physical models, numerical models and mechanistic models. The first category tries to link the cutting forces to the mechanical properties of the material and the geometrical configuration of the cut. Representative of this category is the famous work by Merchant [150]. In opposition to such methods, numerical approaches try to predict the cutting forces by studying the interaction between the tool and the workpiece [66]. In between the above categories, mechanistic models try to fuse the physical and empirical worlds. They relate cutting forces to geometrical quantities involved in the cutting process through empirical coefficients that need to be identified (specific force coefficients, SFC). In literature, there are two well established formulations: exponential models and linear models. Exponential models consider cutting forces as an exponential function of the average chip area. Models of this kind can be found in [25, 43, 216]. Linear models divide the cutting forces into two contributions instead. The first one is directly linked to the shearing action on the shear plane and friction on the rake face of the tool and it is proportional to the undeformed chip cross-sectional area by a specific cutting energy or cutting pressure. The second contribution is represented by the ploughing actions under the flank face of the cutter on the workpiece. This term is proportional to the cutting edge length through the edge SFC [26].

In literature, several works dealt with mechanistic modelling of different kind of mills: square shoulder mills, which feature the simplest geometry, and end mills. Square shoulder mills were the reference case for the Altintas model [26]. Research deeply analysed this kind of mills proposing several upgrades to the Altintas theory. More specifically, the main improvements on the Altintas model regarded the geometrical properties of the trajectory of the mill and its cutters. In fact, the original theory was based on some important assumptions: the cycloidal trajectory of each tooth does not allow for a closed solution for the chip thickness, which was approximated through a sinusoidal function of the feed per tooth, then. A second assumption regarded the cutter run-out: teeth were considered evenly spaced and featured by the same radius. These assumptions brought Wan *et al.* [199], firstly, Kumanchick and Schmitz [118], secondly, and Matsumura and Tumura [144], lastly, to propose extensions of the model, even considering different cutter geometries, e.g. end-mills. Cutting forces were decomposed in a nominal contribution and a run-out contribution (that vanishes in the coefficients identification); cycloidal motion of the mill teeth was introduced with higher order approximations, improving both the accuracy of the chip thickness formulation and the tooth path description. In [7] and [69], helical flute geometry was introduced, considering the helix angle in order to move from orthogonal cutting SFC to oblique cutting context; furthermore, more general tool geometries were considered, i.e. ball, ball nose, taper and ball taper mills. With [157], tool wear started to be included in mechanistic formulations of cutting forces. Actual chip-tool contact area and rubbing forces were included in the prediction of cutting forces.

Later, Zhu and Zhang proposed a tool wear model with adjustable coefficients. Once included in the mechanistic formulation, prediction of cutting forces reached a 98.5%

agreement with experimental instantaneous cutting forces [241]. Zhang *et al.* proposed a model that took into consideration teeth trochoidal trajectories, run-out and tool wear [233]. Even textured tools were analysed [33], where thin shear zone theory and unified cutting mechanics allowed to predict a reduction in the cutting thrust force. In order to compensate cutting forces measurements, inverse-based filtering with the inclusion of a mechanistic model considering workpiece deflection was introduced in [116]. With the objective of assessing cutting operation stability, structural dynamics of both tool and workpiece, and tool orientation were included in the analysis [129]. Sahoo *et al.* proposed a hybrid mechanistic approach, considering trochoidal tool trajectory, tool run-out, minimum chip thickness and all preceding teeth trajectories; SFC were extracted from FEM simulations [169]. Modification of mechanistic models were found even in exotic applications, like turn-milling [158], or like bone milling, where it was possible to include osteon orientation, tool geometry and edge effect with unified mechanics approach. Furthermore, heat flux was used to model surface temperature, too [138]. Some other improvements were proposed in the framework of micro-milling. Edge radius and material strengthening effects, together with overlapping tooth engagements, were firstly introduced [184]. Li *et al.* proposed a general undeformed chip thickness model accounting for trochoidal trajectory, tool run-out and the material removed by multiple preceding teeth [130]. Later, radius size-effect, tool deflection, variable entry and exit angles and minimum uncut chip thickness were added to the model [230]. Li and Wu presented cutting coefficients as functions of the instantaneous chip thickness, estimating them from FEM simulations [128], while Zhou *et al.* included the elastic recovery of the material in the uncut chip thickness formulation [238]. Run-out, tilt offset and tooth trajectory were integrated in the cutting force model by Zhang *et al.*, considering a distributed load on the tool to predict its deflection [232]. Jing *et al.* proposed a formulation including variable entry and exit angles, minimum chip thickness, tool run-out and elastic recovery. The mechanistic model assumed a hybrid fashion, estimating SFC through FEM simulations. They found that elastic recovery plays an important role only at low feed rates [106]. Wan *et al.* investigated the effects of uncut chip thickness on the shearing and ploughing force components. They proposed a formulation based on a combination of plastic formation theory and slip-line field theory, allowing to compute the shape of the dead metal zone [200]. More recently, the inclusion of cutters run-out, radius size-effect and variable entry and exit angles were also proposed for average uncut chip thickness formulations [234]. Zhou *et al.* introduced wear-varying cutting edge with non-linear associated behaviours, i.e. non-linear tool-chip friction and strain gradient plasticity effects, in order to estimate the shear flow stress [237]. Even the numerical computation of curved chip thickness models was introduced, capable of mimicking the chip flow on the rake face of the cutting insert [76]. At last, Kilic and Altintas proposed a generalisation of mechanics and dynamics of metal cutting operations, capable of simulating operations such as turning, milling, boring and drilling, while accounting for general cutter geometries, run-out and tool-workpiece relative vibrations [113, 114].

Altintas model allowed also for the identification of the proportionality coefficients. The coefficient estimations are typically performed through multivariate linear regression models on the average forces [156]. Different works recently dealt with the identification of the SFC of mechanistic models from various cutting tests based on the exten-

sions of the model [28, 54, 72]. Guo *et al.* provided a mathematical formulation for the mechanistic model of a five-axis ball-end mill together with the identification method for the model coefficients [74]. Gonzalo *et al.* proposed an instantaneous approach to fit the SFC for end-mills. They compared the identified coefficients with those obtained from orthogonal turning tests with the average identification procedure. Identified SFC resulted to be in accordance with those from orthogonal turning tests, providing a better physical description of cutting forces with variable chip thickness [66].

The design procedure of machining processes conducts to an optimisation problem. Typically, in order to obtain the lowest cutting times, the function to be optimised is the material removal rate. It depends on the axial depth of cut, the radial depth of cut and the feed rate. The analysis of machine tool vibrations identified the axial depth of cut as the main source of instability, namely chatter instability [6, 26, 222]. This discovery led tool manufacturers towards newer directions for the maximisation of the material removal rate. Indeed, they started to develop more complicated geometries for the cutters. Such tools are known as high-feed mills.

High-feed mills are a specific kind of High-Efficiency Milling (HEM) tool with a specialised profile, which rely on chip thinning in order to reach higher feed per tooth values. Indeed, high feed mills are devoted typically to roughing operations, with the objective of maximising the material removal rate and extend their life. They operate under really low axial depth of cuts, with the cutting action taking place mainly under the edge of the bottom profile. Thus, their geometrical characteristics allow for a reduced radial load on the spindle (lower tool deflection, too), while generating higher axial forces during the cut. Reduced radial loads make them a cutting tool suitable for thin wall milling and deep pocket contouring. High-feed mills are suited for ploughing operations, especially for light alloys machining (e.g. titanium alloys).They are also very practical for copy mill machining uneven surfaces and are highly effective in cavity milling and moulds and dies manufacturing. Researchers address cutters as "high-feed" when they can achieve up to ten times the feed rate and three times the material removal rate compared with conventional mills [49]. Such cutters may reach feed per tooth over mill radius ratios ranging from 0.03 up to 0.4 [49, 75, 89, 103, 141, 143, 235]. This capability was obtained through the use of really low lead angles and, more specifically, using double-phased cutters, i.e. cutters featured by two consecutive sections with different lead angles. Really low lead angles demand for such aggressive feed rates in order to avoid increased friction due to rubbing between the cutting edge and the workpiece. This would be the main root cause for increased cutting temperatures and lower tool lives. Furthermore, when adopted for face milling, they leave scallops on the machined surface, due to their geometrical characteristics. Scallops profiles are related to the radial overlap of subsequent passes. Low mill radial engagement (i.e., high radial overlap) leads to smaller (even null) scallops, while full radial engagements may cause the scallops to have a height equal to the axial depth of cut.

Literature analysis revealed absence, to the best of authors knowledge, of mechanistic modelling of high-feed mills. Furthermore, the improvements related to cycloidal trajectory of the tooth were relevant and appreciable for mills with a really high feed per tooth over mill diameter ratio, i.e. for micro-milling operations and not for high-feed milling, where this ratio is still considerably lower [118]. All the other improvements allowed for a better representation of the uncut chip thickness, and thus, to better cutting

force predictions. Anyway, they all introduced complex geometrical and mathematical formulations that needed numerical approaches based on discretisation (e.g. [76]) or time integration [113, 114], leading to slow computational operations. On the contrary, the conceived approach is completely analytical and in closed form, allowing for the exact solution of the equations with fast computational times, while keeping high prediction capabilities and outperforming literature analytical approaches. The developed model considers the variable entry and exit angles, while parametrically considering high-feed cutter geometries. At last, the conceived model allows for the identification of SFC based on average cutting forces. In this chapter, the analytical formulation of the mechanistic model for high-feed mills (that will be consistently indicated as HF) is introduced, while the differences with respect to the classical Altintas model (CL) will be analysed. In section 2.2 the HF model is presented, providing the main analytical formulas of the model. The mean fitting procedure is described, showing the construction procedure of the design matrix. The design of experiments is then explained together with the performed statistical analysis, including model fitting and uncertainty propagation on validation quantities. In section 2.3, the whole regression analysis output is furnished and its in-depth discussion is performed. In section 2.4, the novelties and contributions of the work are highlighted. At the end of the chapter, the theoretical background is included in appendix 2.A, with the equations related to the CL model. Two appendixes, 2.B and 2.C, describe the proof of the instantaneous and mean forces derivations, respectively.

2.2 Materials and methods

In this section, the whole work procedure was described following the order of figure 2.1. The HF model development is firstly presented, highlighting formulation differences with respect to the CL model (appendix 2.A). Then, the experimental set-up and the experimental campaign are introduced. The methods for the identification of the SFC and the validation analysis will conclude this section.

2.2.1 High-feed analytical model

The development of the HF model is presented in this section (fig. 2.1C-D). High-feed mills are featured by double-phased low lead angles cutters, as can be seen in figure 2.2a. The mill nominal radius is r_{nom}, while the number of cutters is N. The machining process is featured by the feed per tooth c and the rotational speed n. The formulation starts by defining the position in time of the cutters [54] through equation 2.1:

$$\varphi_j(t) = \varphi_0 + \frac{2\pi\,(j-1)}{N} + 2\pi n t \tag{2.1}$$

where t is the time; $j = 1, 2, ..., N$ denotes the j-th cutter; φ_0 is the angular position of cutter 1 at $t = 0$. The dependence on time of the cutter angles will be omitted in the following. The instantaneous chip thickness removed by cutter j is approximated by equation 2.2:

$$h(\varphi_j) = c \sin\varphi_j \tag{2.2}$$

In the proposed model, no run-out formulation is included. Thus, all the cutters are assumed to be placed at the same radial distance from the mill axis. This choice was

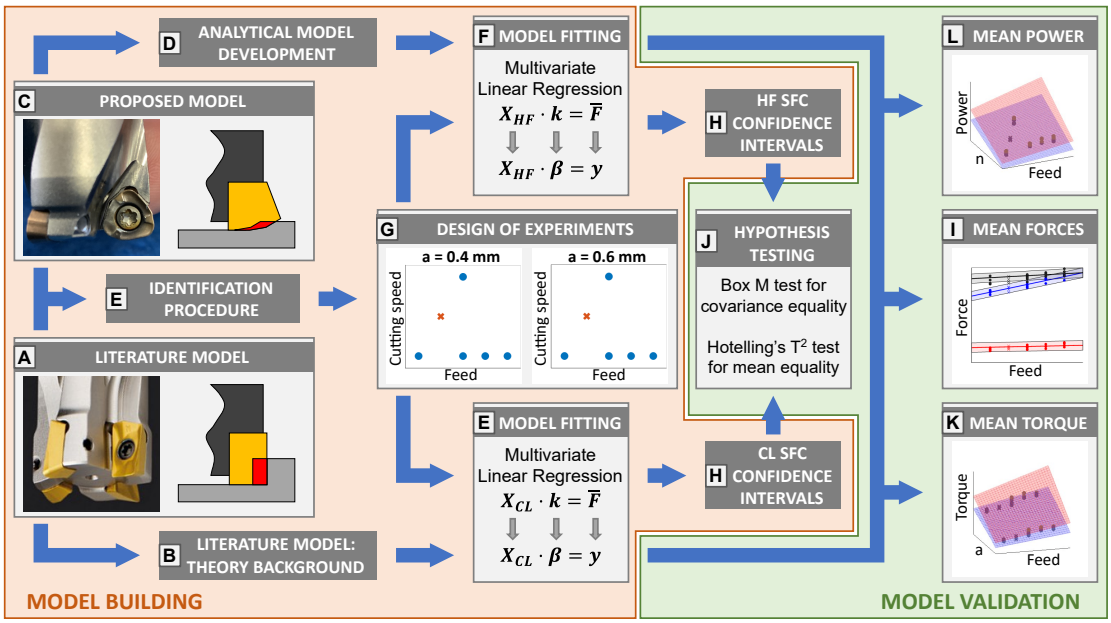
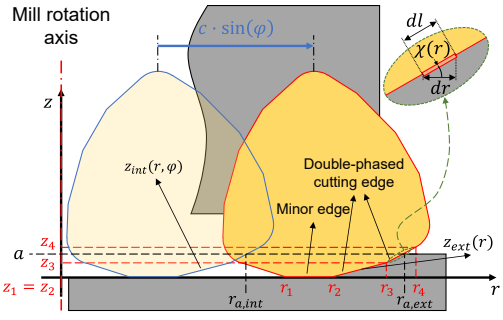
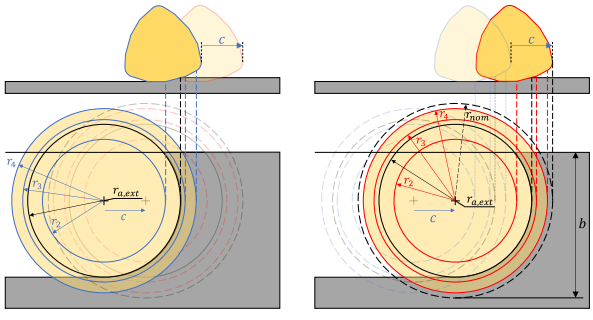


Figure 2.1: *Graphical representation of the developed work. From left to right: literature (A) and high-feed (C) mill cutters; reference (B) and proposed (D) analytical formulations allow to estimate SFC through a Multivariate Linear Regression Problem (E-F); the estimation is performed by cutting tests, designed to vary feed per tooth, cutting speed and axial depth of cut (G); the estimated SFC and their confidence intervals are compared in terms of mean and variance through multivariate hypothesis tests (H-J); the proposed model is validated and compared with the literature model through the prediction of mean forces (I), torque (K) and power (L).*



(a) *Representation of the double-phased cutter geometry and its relative nomenclature.*



(b) *Top view of the proposed model, with nomenclature and representation of the previous and current cutting tool reference circles.*

Figure 2.2: *High-feed model reference figure. The cutting insert profile is shown relatively to the previous cutter position for a generic cutter angle φ. Each vertex of the double-phased cutting edge can be projected on the top view, and can be seen as the associated circle of radius r_i.*

based on two main reasons. Firstly, when dealing with high-feed, the effect of run-out on the instantaneous chip thickness is negligible compared to the feed per tooth effect. Secondly, introducing a higher order term such as run-out, would lead to a not-trivial analytical closed formulation for cutting forces, and, as it will be explained in the next chapters, to a more complex identification of specific force coefficients (through instantaneous methods). For the next steps, it is necessary to make reference to figures 2.2a and 2.2b for the cutter geometry description. Looking at the plane $z-r$ that passes through the mill axis at the current position and the radius at given φ angle from the y-axis (fig. 2.2a), the cutter profile coordinates in actual and previous position (z_{ext} and z_{int}, respectively), are defined as functions of the radial coordinate r from the mill actual axis of rotation:

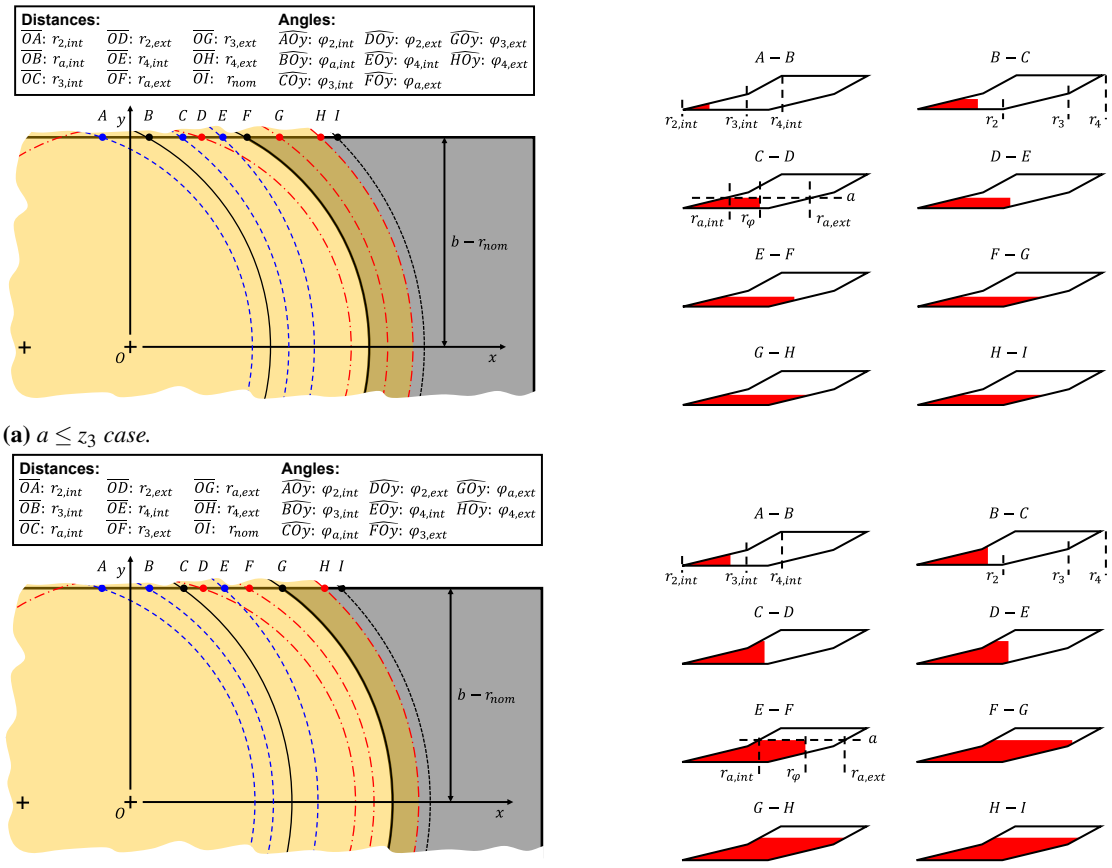
$$z_{ext}(r) = \begin{cases} 0, & \text{if } r_1 \leq r \leq r_2 \\ \frac{z_3}{r_3 - r_2}(r - r_2), & \text{if } r_2 < r \leq r_3 \\ z_3 + \frac{z_4 - z_3}{r_4 - r_3}(r - r_3), & \text{if } r_3 < r \leq r_4 \end{cases} \tag{2.3}$$

$$z_{int}(r,\varphi) = \begin{cases} \frac{z_3}{r_3 - r_2}\left[r - r_2 + h(\varphi)\right], \\ \qquad \text{if } r_2 - h(\varphi) < r \leq r_3 - h(\varphi) \\ z_3 + \frac{z_4 - z_3}{r_4 - r_3}\left[r - r_3 + h(\varphi)\right], \\ \qquad \text{if } r_3 - h(\varphi) < r \leq r_4 - h(\varphi) \\ z_4, \\ \qquad \text{if } r_4 - h(\varphi) < r \leq r_4 \end{cases} \tag{2.4}$$

where r_i and z_i, with $i = 2,3,4$ are shown in figure 2.2a and $z_1 = z_2 = 0$; the j index was omitted for simplicity. As can be seen from equations 2.3 and 2.4, high-feed mills present a radial extension of the cutter profile. This feature implies that φ_{st} is a function of the axial coordinate z (fig. 2.3a-2.3b). The developed model, in opposition to the CL model, takes into account this effect. First of all, the radial depth of cut b is defined with respect to the nominal radius of the mill r_{nom} (this is done to match what an operator does practically). A reference circle associated to each particular radius (i.e. r_i and $i = 1,2,3,4$) can be drawn in the reference model (figure 2.2b). It is possible to determine different critical angles, computed with respect to the y-axis, at which the radius of each circle ends on the edge of the workpiece (eq. 2.5):

$$\begin{aligned} \varphi_{i,ext} &= \arccos\frac{b - r_{nom}}{r_i}, \qquad i = 2,3,4 \\ \varphi_{i,int} &= \arctan\frac{r_i \sin\varphi_{i,ext} - c}{b - r_{nom}}, \qquad i = 2,3,4 \end{aligned} \tag{2.5}$$

Such angles determine the transition regions from an undeformed chip shape to a different shape (due to the engagement of a different section of the cutter profile, fig. 2.3a-2.3b). The possibility to introduce transition regions in the mathematical formulation of the model represents a difference with respect to the CL model. The axial depth of cut introduces the last two critical angles, at which the cutter profile is engaged at



(a) $a \leq z_3$ *case.*

(b) $a > z_3$ *case.*

Figure 2.3: *Previous (dashed) and current (dash-dotted) cutting tool reference circles. Circles with solid lines are generated by the axial engagement. The critical angles $\varphi_{i,int}$ and $\varphi_{i,ext}$ determine the regions where the shape of the undeformed chip changes. On the right the chip shapes vary according to the angular position of the tool. A different chip shape is obtained when the tool radius passes between two points (indicated as pairs over the chip shape itself).*

its maximum height. It is necessary to introduce the radii of intersection between the external and internal cutter profiles with the axial depth of cut line (eq. 2.6):

$$r_{a,ext} = \begin{cases} r_2 + \frac{r_3 - r_2}{z_3}\, a, & \text{if } a \leq z_3 \\ r_3 + \frac{r_4 - r_3}{z_4 - z_3}\,(a - z_3), & \text{if } a > z_3 \end{cases} \tag{2.6}$$
$$r_{a,int}(\varphi) = r_{a,ext} - h(\varphi)$$

It must be pointed out that $r_{a,int}(\varphi)$ is not a radius, but just the radial distance between the actual mill centre and the intersection of the previous cutter profile with the axial depth of cut line. It is then possible to compute the critical angles associated to $r_{a,ext}$ and $r_{a,int}(\varphi)$ (eq. 2.7):

$$\begin{aligned} \varphi_{a,ext} &= \arccos \frac{b - r_{nom}}{r_{a,ext}} \\ \varphi_{a,int} &= \arctan \frac{r_{a,ext} \sin \varphi_{a,ext} - c}{b - r_{nom}} \end{aligned} \tag{2.7}$$

At this point, the area and the edge contact length differentials can be defined by equation 2.8:

$$\begin{aligned} \mathrm{d}A(r) &= [z_{int}(r) - z_{ext}(r)]\,\mathrm{d}r = \\ &= \mathrm{d}A_{int}(r) - \mathrm{d}A_{ext}(r) \\ \mathrm{d}l(r) &= \frac{\mathrm{d}r}{\cos\chi(r)} \end{aligned} \tag{2.8}$$

where $\chi(r)$ is the lead angle at the current r radial coordinate.

The three components of the cutting forces (i.e. tangential, radial and axial) acting on cutter j are expressed in differential form (eq. 2.9):

$$\begin{aligned} \mathrm{d}F_{t,j}(\varphi_j) &= K_{t,c}\,\mathrm{d}A(\varphi_j) + K_{t,e}\,\mathrm{d}l(\varphi_j) \\ \mathrm{d}F_{r,j}(\varphi_j) &= K_{r,c}\,\mathrm{d}A(\varphi_j) + K_{r,e}\,\mathrm{d}l(\varphi_j) \\ \mathrm{d}F_{a,j}(\varphi_j) &= K_{a,c}\,\mathrm{d}A(\varphi_j) + K_{a,e}\,\mathrm{d}l(\varphi_j) \end{aligned} \tag{2.9}$$

where $K_{d,c}$ and $K_{d,e}$, with $d = t, r, a$, are the cutting and edge specific force coefficients (SFC), respectively. The first ones account for the shearing action on the shear plane and the friction effect on the rake face of the cutting edge, whereas the edge SFC are related to the ploughing action under the flank face of the cutting edge [54].

Substituting equation 2.8 in equation 2.9 and integrating it in the radial direction, leads to the expression of the instantaneous forces. The integration bounds are equal to $r_2 - h(\varphi)$ and $r_\varphi(\varphi)$, that is the length of the segment connecting the actual centre of the mill to the workpiece end (eq. 2.10):

$$r_\varphi(\varphi) = \begin{cases} \frac{b - r_{nom}}{\cos\varphi}, & \text{if } \varphi \le \varphi_{a,ext} \\ r_{a,ext}, & \text{if } \varphi > \varphi_{a,ext} \end{cases} \tag{2.10}$$

Since the engagement condition was already considered in the differential terms, the instantaneous cutting forces are:

$$\begin{aligned} F_{t,j}(\varphi_j) &= K_{t,c}\left[A_{int}(\varphi_j) - A_{ext}(\varphi_j)\right] + K_{t,e}\,l(\varphi_j) \\ F_{r,j}(\varphi_j) &= K_{r,c}\left[A_{int}(\varphi_j) - A_{ext}(\varphi_j)\right] + K_{r,e}\,l(\varphi_j) \\ F_{a,j}(\varphi_j) &= K_{a,c}\left[A_{int}(\varphi_j) - A_{ext}(\varphi_j)\right] + K_{a,e}\,l(\varphi_j) \end{aligned} \tag{2.11}$$

The force components are then projected in the feed, normal and axial direction of the milling operation, and the contribution of each cutter is summed up in equation 2.12:

$$\begin{aligned} F_x(t) &= -\sum_{j=1}^{N} F_{t,j}(\varphi_j)\cos\varphi_j - \sum_{j=1}^{N} F_{r,j}(\varphi_j)\sin\varphi_j \\ F_y(t) &= +\sum_{j=1}^{N} F_{t,j}(\varphi_j)\sin\varphi_j - \sum_{j=1}^{N} F_{r,j}(\varphi_j)\cos\varphi_j \\ F_z(t) &= +\sum_{j=1}^{N} F_{a,j}(\varphi_j) \end{aligned} \tag{2.12}$$

Forces are expressed as functions of time since $\varphi_j = \varphi_j(t)$.

In order to identify experimentally the six SFC, a procedure based on mean cutting forces was developed [54, 166]. Equation 2.11 is then substituted in equation 2.12 and integrated over one period of revolution T_P of the mill. The resulting expressions for mean forces are reported in equation 2.13:

$$\begin{aligned}
\overline{F}_x &= \frac{N}{2\pi}\left(-a_{11}\,K_{t,c} - a_{12}\,K_{t,e} - a_{13}\,K_{r,c} - a_{14}\,K_{r,e}\right) \\
\overline{F}_y &= \frac{N}{2\pi}\left(a_{21}\,K_{t,c} + a_{22}\,K_{t,e} - a_{23}\,K_{r,c} - a_{24}\,K_{r,e}\right) \\
\overline{F}_z &= \frac{N}{2\pi}\left(a_{35}\,K_{a,c} + a_{36}\,K_{a,e}\right)
\end{aligned} \tag{2.13}$$

where a_{do} are constant terms depending on the mill geometry, the feed per tooth (c) and the depths of cut (a and b).

The computation of the torque, requires more steps than for the CL model. This is due to the fact that the distance of the infinitesimal tangential force has an increasing arm with respect to its z coordinate of application. Starting from the infinitesimal tangential force, the associated infinitesimal torque is:

$$\mathrm{d}T_j(r) = K_{t,c}\,r\,\mathrm{d}A_j(r) + K_{t,e}\,r\,\mathrm{d}l_j(r) \tag{2.14}$$

By integrating equation 2.14 over the radial coordinate of the undeformed chip and summing up the contributions of the N cutters, the resulting torque is constituted by two terms:

$$T(t) = \sum_{j=1}^{N}\left\{K_{t,c}\,r^*_{A,j}\left[A_{int}(\varphi_j) - A_{ext}(\varphi_j)\right] + K_{t,e}\,r^*_{l,j}\,l(\varphi_j)\right\} \tag{2.15}$$

where $r^*_{A,j}$ and $r^*_{l,j}$ are the radial coordinates of the centre of mass of the undeformed chip area and edge contact length, respectively. The cutting power is computed by substituting equation 2.15 in equation 2.A.7. Complete expressions for A_{int}, A_{ext}, l, $r^*_{A,j}$ and $r^*_{l,j}$ were reported in appendix 2.B, while a_{do} terms in 2.C, due to their complexity. The different implementation of the CL and HF models was summarised in figure 2.4. The figure represents the step-by-step computations to be performed in order to calculate instantaneous cutting forces, torque and power.

2.2.2 Experimental set-up and Design of Experiments

The experimental set-up available was shown in figure 2.5. A Mandelli M5 machining centre was used, equipped with a Mitsubishi AJX06R203SA20S high-feed mill, with three JOMT06T216ZZER-JL MP9140 cutting inserts. The adopted cutters presented a nominal diameter equal to 20 *mm*. Their geometry featured two consecutive phases with lead angles equal to 12.5° and 24.5°. The cutter parameters after the geometrical simplification introduced by the model discussed in section 2.2.1 were reported in table 2.1. A Kistler 9255B dynamometric plate was adopted for the milling forces measurement, followed by a Kistler 5070 charge amplifier. A three-phase power meter was used to acquire the spindle power (with three LEM LF 205-S/SP3). These quantities were acquired through a NI cRIO 9039, with a NI 9215 acquisition card for cutting forces and NI 9205 acquisition card for spindle power. Siemens SinuCom NC was used to

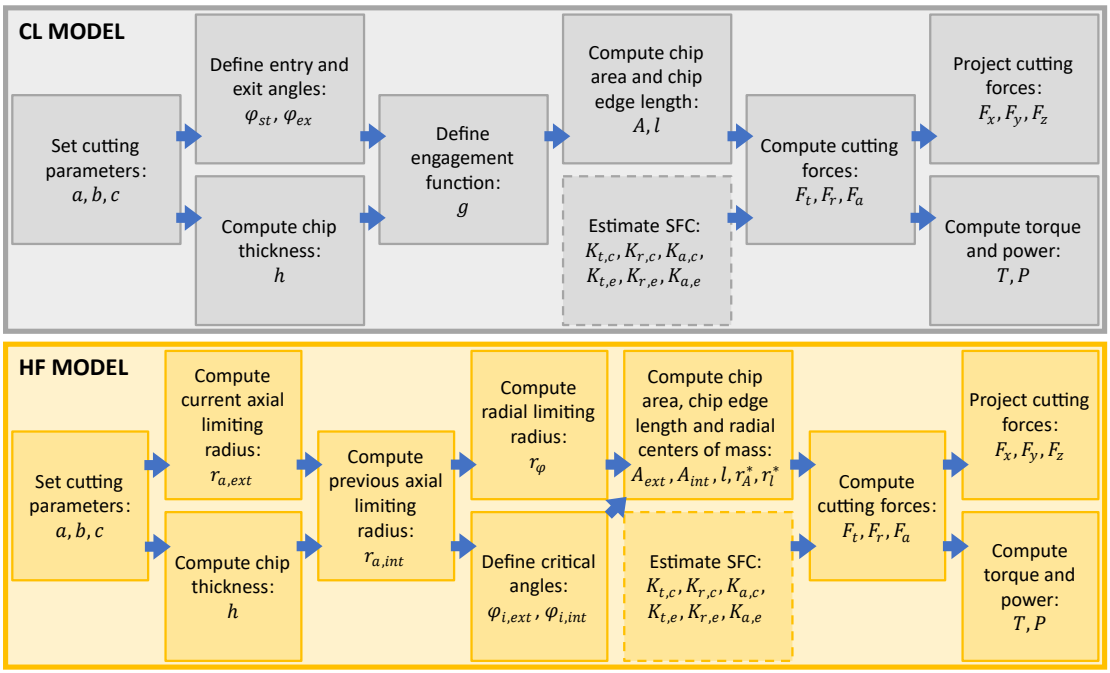


Figure 2.4: *Flow chart representing the needed steps to compute instantaneous cutting quantities (forces, torque and power) for the CL (appendix 2.A) and HF (section 2.2.1) models. Estimation of the SFC (dashed boxes) must be performed beforehand.*

Table 2.1: *Cutter geometrical parameters.*

r_1 $[mm]$	r_2 $[mm]$	r_3 $[mm]$	r_4 $[mm]$	z_1 $[mm]$	z_2 $[mm]$	z_3 $[mm]$	z_4 $[mm]$
5.35	6.57	8.53	9.03	0	0	0.40	0.62

acquire the spindle quadrature current and speed, axis positions and speeds. Cutting forces were acquired at high frequency (5 kHz), while SinuCom quantities were acquired at 250 Hz. In order to experimentally identify the SFC from the two analytical models (i.e. CL and HF), a set of 48 experiments was designed, constituted by 4 replicates of 12 different cutting conditions (table 2.2, fig. 2.1G). Tests were performed on a Ti_6Al_4V rectangular workpiece of dimensions $255x262$ mm. DoE consisted of two a levels: 0.4 and 0.6 mm. The a levels were chosen so that the first one was keeping only the first phase of the cutting edge engaged, while the highest level included also the cutting edge second phase. Five levels of c were adopted: 0.60, 0.65, 0.70, 0.75 and 0.80 $mm/tooth$. Three levels of cutting speed v_c were chosen 50, 60 and 70 m/min. All the tests were performed at $b = 13$ mm. The number of experiments was chosen in order to allow for a good estimation of SFC, following [51, 71, 92, 181, 188]. Five levels of c were chosen to explore the range of applicability of the tool, suggested by the tool manufacturer, and catch the trend of mean forces [71]. Cutting speed was varied in order to take into account the process variability. The parameter combinations were shown in table 2.2. The cuts with ID 4 and 10 were chosen as validation data. This was in accordance with reference literature in similar applications (e.g. SFC identification [166]) and linear regression problems [3]. Validation tests featured parameters that

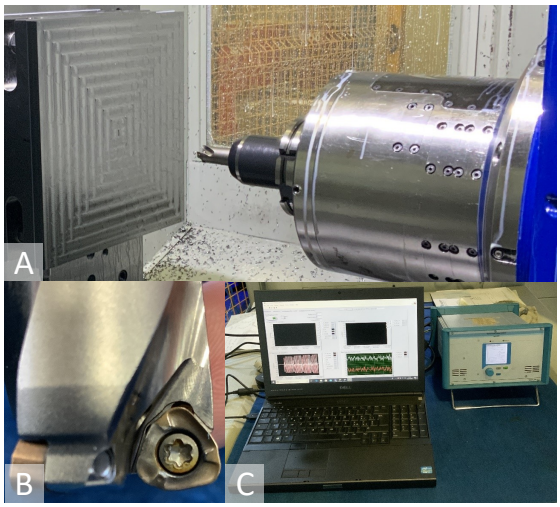


Figure 2.5: *Experimental set-up: A) Ti_6Al_4V workpiece and mill; B) Mitsubishi high-feed mill; C) DAQ.*

Table 2.2: *Design of Experiments parameter combinations.*

ID	c $[mm/tooth]$	a $[mm]$	v_c $[m/min]$	n $[rpm]$	Replicates	Model use
1	0.75	0.40	50	796	4	fitting
2	0.80	0.40	50	796	4	fitting
3	0.60	0.40	50	796	4	fitting
4	0.65	0.40	55	955	4	validation
5	0.70	0.40	70	1114	4	fitting
6	0.70	0.40	50	796	4	fitting
7	0.75	0.60	50	796	4	fitting
8	0.80	0.60	50	796	4	fitting
9	0.60	0.60	50	796	4	fitting
10	0.65	0.60	55	955	4	validation
11	0.70	0.60	70	1114	4	fitting
12	0.70	0.60	50	796	4	fitting

were unseen in model fitting experiments in terms of c and v_c. This choice was taken in order to verify linearity hypothesis and checking prediction accuracy. The regressor values for validation were determined in order to use the linear regression model for interpolation, and not for extrapolation (being this a general rule of applicability for linear regression models). In fact, by using the model for extrapolation, the assumptions can't be verified and the prediction intervals (PI) naturally increase [186]. The axial depth of cut a was considered as a block factor, while all the tests were casualised within each block.

Model Identification

Model identification procedure was firstly proposed by Altintas [26]. It consists in a multivariate linear regression problem (from which the notation is adopted [186]). This allows estimating the SFC (fig. 2.1E-F). The fitting procedure exploits the mean cutting forces measured during the experiments. The systems of linear equations 2.A.5

and 2.13 can be written in matrix fashion (eq. 2.16):

$$\begin{Bmatrix} \overline{F}_x \\ \overline{F}_y \\ \overline{F}_z \end{Bmatrix} = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{16} \\ x_{21} & x_{22} & \dots & x_{26} \\ x_{31} & x_{32} & \dots & x_{36} \end{bmatrix} \cdot \begin{Bmatrix} K_{t,c} \\ K_{t,e} \\ K_{r,c} \\ K_{r,e} \\ K_{a,c} \\ K_{a,e} \end{Bmatrix} \tag{2.16}$$

If M tests are experimented, the identification problem can be rewritten in a multivariate linear regression fashion (eq. 2.17):

$$\begin{Bmatrix} \overline{F}_{x,1} \\ \overline{F}_{y,1} \\ \overline{F}_{z,1} \\ \overline{F}_{x,2} \\ \overline{F}_{y,2} \\ \overline{F}_{z,2} \\ \dots \\ \overline{F}_{x,M} \\ \overline{F}_{y,M} \\ \overline{F}_{z,M} \end{Bmatrix} = \begin{bmatrix} x_{11,1} & x_{12,1} & \dots & x_{16,1} \\ x_{21,1} & x_{22,1} & \dots & x_{26,1} \\ x_{31,1} & x_{32,1} & \dots & x_{36,1} \\ x_{11,2} & x_{12,2} & \dots & x_{16,2} \\ x_{21,2} & x_{22,2} & \dots & x_{26,2} \\ x_{31,2} & x_{32,2} & \dots & x_{36,2} \\ \dots & \dots & \dots & \dots \\ x_{11,M} & x_{12,M} & \dots & x_{16,M} \\ x_{21,M} & x_{22,M} & \dots & x_{26,M} \\ x_{31,M} & x_{32,M} & \dots & x_{36,M} \end{bmatrix} \cdot \begin{Bmatrix} K_{t,c} \\ K_{t,e} \\ K_{r,c} \\ K_{r,e} \\ K_{a,c} \\ K_{a,e} \end{Bmatrix} + \begin{Bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \dots \\ \varepsilon_M \end{Bmatrix} \tag{2.17}$$

$$\mathbf{y} = \mathbf{X}_k \boldsymbol{\beta}_k + \boldsymbol{\varepsilon}_k$$

where $k = CL, HF$; $\mathbf{y}$ is the vector of responses, i.e. mean forces; $\mathbf{X}_k$ is the design matrix of either CL and HF model; $\boldsymbol{\beta}_k$ is the vector of regression coefficients, i.e. SFC; $\boldsymbol{\varepsilon}_k$ is the vector of residuals with $\varepsilon_{m,k} \sim \mathcal{N}(0, \sigma_k^2)$ uncorrelated random variables. The Ordinary Least Squares estimates for SFC are given by equation 2.18:

$$\hat{\boldsymbol{\beta}}_k = \left(\mathbf{X}_k^T \mathbf{X}_k\right)^{-1} \mathbf{X}_k \mathbf{y}, \qquad k = CL, HF \tag{2.18}$$

The reconstructed responses are (eq. 2.19):

$$\hat{\mathbf{y}}_k = \mathbf{X}_k \hat{\boldsymbol{\beta}}_k, \qquad k = CL, HF \tag{2.19}$$

The estimated residuals are (eq. 2.20):

$$\mathbf{e}_k = \mathbf{y} - \hat{\mathbf{y}}_k, \qquad k = CL, HF \tag{2.20}$$

It is then possible to estimate the variance of the process errors as (eq. 2.21):

$$\hat{\sigma}_k^2 = \frac{\mathbf{e}_k^T \mathbf{e}_k}{n_p - p}, \qquad k = CL, HF \tag{2.21}$$

where n_p is the number of data points (i.e. $3M$) and p the number of regression coefficients (i.e. 6 SFC). It is possible also to compute the covariance matrix of the regression coefficients $\boldsymbol{\Sigma}_{\boldsymbol{\beta}_k}$ (eq. 2.22):

$$\boldsymbol{\Sigma}_{\boldsymbol{\beta}_k} = \hat{\sigma}_k^2 \mathbf{C}_k, \qquad k = CL, HF \tag{2.22}$$

where $\mathbf{C}_k$ is the matrix $\left(\mathbf{X}_k^T \mathbf{X}_k\right)^{-1}$. Process variance allows for the computations of the regression coefficients $100\,(1-\alpha)\%$ confidence intervals (CI) through equation 2.23 (fig. 2.1H):

$$\hat{\beta}_{k,o} \;-\; t_{\alpha/2,n_p-p}\sqrt{\hat{\sigma}_k^2 C_{oo,k}} \;\leq\; \beta_{k,o} \;\leq\; \hat{\beta}_{k,o} \;+\; t_{\alpha/2,n_p-p}\sqrt{\hat{\sigma}_k^2 C_{oo,k}} \quad (2.23)$$

where $k = CL, HF$; α is the significance level, $C_{oo,k}$ is the o-th element of the diagonal of matrix $\mathbf{C}_k$; $t_{\alpha/2,n_p-p}$ is the upper $\alpha/2$ percentage point of a t distribution with $n_p - p$ degrees of freedom. Such limits are then the CI extrema for SFC.

2.2.3 Model Validation

Four main procedures were followed for the assessment and validation of the developed HF model (fig. 2.1I-J-K-L).

First validation step

This procedure consisted in the prediction of mean forces on model fitting and validation data (fig. 2.1I). $100 \cdot (1-\alpha)\%$ PI for single new response observations were computed through equation 2.24 [186]:

$$\hat{y}(\mathbf{x}_{0,k}) - t_{\alpha/2,n_p-p}\sqrt{\hat{\sigma}_k^2\left(1+\mathbf{x}_{0,k}^T C_k\,\mathbf{x}_{0,k}\right)} \leq y_0 \leq \hat{y}(\mathbf{x}_{0,k}) + t_{\alpha/2,n_p-p}\sqrt{\hat{\sigma}_k^2\left(1+\mathbf{x}_{0,k}^T C_k\,\mathbf{x}_{0,k}\right)} \quad (2.24)$$

where $\mathbf{x}_{0,k}$ is the new point for which an estimated response is required, with $k = CL, HF$; $\hat{y}(\mathbf{x}_{0,k})$ is the estimated response $\mathbf{x}_{0,k}^T \hat{\boldsymbol{\beta}}_k$.

Second validation step

This validation procedure consisted in the multivariate hypothesis testing of SFC estimates (fig. 2.1J). Since the CL and HF models are based on SFC carrying the same physical meaning, a statistical hypothesis testing procedure was applied to test whether or not the two models shown any statistically appreciable difference. In this case, two steps were applied: first of all, the Box M test for covariance equality [23] was applied to the two SFC groups. Since Box M test rejected the covariance equality in null hypothesis at 0.05 significance, an alternative to the Hotelling's Two-Sample T-squared test for difference in means was used. Nel and van der Merwe [155] procedures to solve Behrens-Fisher multivariate tests were adopted at 0.05 significance. Simultaneous CI for each difference between CL and HF SFC were computed at 0.05, highlighting the directions of meaningful differences.

Third and fourth validation steps

These steps were reported in the same section due to their similar methodology. The difference between the two, is that they are based on the prediction of two different

physical quantities: the mean spindle cutting torque (fig. 2.1K) and the mean spindle cutting power (fig. 2.1L). Such quantities were used for validation since none of them was introduced in the model fitting procedure. In this way, a more robust validation approach was introduced. The new variables were expressed as functions of the SFC, in order to make predictions. Starting from the CL model, mean torque and mean power follow equation 2.25 and 2.26:

$$\begin{aligned}\hat{\bar{T}}_{CL} &= r_{nom} \begin{Bmatrix} x_{35,CL} & x_{36,CL} & 0 & 0 & 0 & 0 \end{Bmatrix} \hat{\boldsymbol{\beta}}_{CL} = \\ &= \mathbf{a}_{T,CL}^T \hat{\boldsymbol{\beta}}_{CL}\end{aligned} \tag{2.25}$$

$$\begin{aligned}\hat{\bar{P}}_{CL} &= 2\pi n r_{nom} \begin{Bmatrix} x_{35,CL} & x_{36,CL} & 0 & 0 & 0 & 0 \end{Bmatrix} \hat{\boldsymbol{\beta}}_{CL} = \\ &= \mathbf{a}_{P,CL}^T \hat{\boldsymbol{\beta}}_{CL}\end{aligned} \tag{2.26}$$

where $x_{do,CL}$ is the predictor in row d and column o of the matrix in equation 2.16 for the CL model, computed at the nominal cutting parameters of a given experiment; $\mathbf{a}_{T,CL}^T$ is defined as $\begin{Bmatrix} x_{35,CL} & x_{36,CL} & 0 & 0 & 0 & 0 \end{Bmatrix}$ and $\mathbf{a}_{P,CL}^T$ is defined as $2\pi n r_{nom} \mathbf{a}_{T,CL}^T$.

For the HF model, it is necessary to take the mean over a tool revolution of the instantaneous torque (eq. 2.15) and power (substituting equation 2.15 in equation 2.A.7):

$$\hat{\bar{T}}_{HF} = \frac{1}{T_P} \int_0^{T_P} T(t)\, \mathrm{d}t = \mathbf{a}_{T,HF}^T \hat{\boldsymbol{\beta}}_{HF} \tag{2.27}$$

$$\hat{\bar{P}}_{HF} = \frac{1}{T_P} \int_0^{T_P} P(t)\, \mathrm{d}t = \mathbf{a}_{P,HF}^T \hat{\boldsymbol{\beta}}_{HF} \tag{2.28}$$

where the integrals can be expressed as two vectors $\mathbf{a}_{T,HF}^T$ and $\mathbf{a}_{P,HF}^T$ multiplied by the estimated SFC. It is then possible to propagate the uncertainty of the SFC estimations to the new variables of interest (eq. 2.25-2.28). Being the mean torque and mean power linear combinations of the SFC, the uncertainty is propagated through equation 2.29 [192]:

$$s_{q,k}^2 = \mathbf{a}_{q,k}^T \boldsymbol{\Sigma}_{\boldsymbol{\beta}_k} \mathbf{a}_{q,k} \tag{2.29}$$

where q stands for the quantity of interest, i.e. either T or P; k stands for the reference model, i.e. CL or HF; $s_{q,k}^2$ is the variance associated to the quantity q prediction using the k-th model; $\mathbf{a}_{q,k}$ is the vector that transforms SFC in the variable of interest (eq. 2.25-2.28); $\boldsymbol{\Sigma}_{\boldsymbol{\beta}_k}$ is the covariance matrix of the SFC coefficients (2.22).

2.3 Results and discussion

In this section, the results of the analysis of the experimental campaign data will be reported and discussed. The presentation of the results will start with the instantaneous forces prediction capabilities, followed by the three validation steps described in section 2.2.

2.3.1 Model identification

The model fitting procedure allowed to identify the SFC values and the associated 95% CI, which are reported in table 2.3. The fitting procedures computational times resulted

Table 2.3: *Identification of SFC for CL and HF models, together with their associated* 95% *CI.*

	CL			HF		
SFC	Lower	Mean	Upper	Lower	Mean	Upper
$K_{t,c}\,[MPa]$	969.7	1696.5	2423.3	1084.5	1255.2	1426.0
$K_{t,e}\,[N/mm]$	-114.8	262.1	639.1	70.0	86.5	103.1
$K_{r,c}\,[MPa]$	-523.5	203.3	930.0	135.5	306.3	477.0
$K_{r,e}\,[N/mm]$	-181.8	195.1	572.0	-0.9	15.7	32.2
$K_{a,c}\,[MPa]$	190.2	845.8	1501.4	-194.1	-39.2	115.8
$K_{a,e}\,[N/mm]$	422.8	746.9	1070.9	191.7	205.9	220.2

to be 12.3 *ms* and 46.7 *ms* for the CL and HF models, respectively. The computational times included the design matrix construction and the SFC identification. In order to provide a reliable estimate, they were reported as the average over 10000 cycles. Furthermore, such cycles were performed using Matlab® on a Dell XPS 15 7590 featuring an Intel® Core™ i7-9750H CPU @ 2.60GHz. It is clear how the developed HF model is capable of better catching the physical phenomenon underneath the cut than the CL one. In fact, identified SFC with HF model present narrower 95% CI. This means that the uncertainty in the estimation of the SFC is lower and it is confirmed even by the significance tests for the single regressors. Such tests can be checked directly from the CI of each regressor: a regressor results to be significant on the associated hypothesis test at 0.05 significance level (i.e. the regression coefficient is significantly different from 0), if its 95% CI doesn't include 0. Analysis of Variance (ANOVA) tables for the CL (table 2.4a) and HF (table 2.4b) regression models were computed through Minitab software.

The p-values of the single regression coefficient significance were explicitly shown in such tables. Using CL model, three SFC were not significant at 0.05 significance level (i.e. $K_{t,e}$, $K_{r,c}$ and $K_{r,e}$), whereas only two using the HF model (i.e. $K_{r,e}$ and $K_{a,c}$). It is possible also to note the amount of variance explained by the two regression models, i.e. R-squared (on model fitting data), looking at the *regression* row in *contribution* field. For both the models, R-squared is really high and comparable: 98.92% for the CL model versus 99.87% for the HF model. Usually, even adjusted R-squared, prediction R-squared and Test R-squared are computed, since they provide a more reliable information on the performances of the regression models. Adjusted R-square takes into account also the number of used predictors (in this case both the models use six regressors, so adjusted R-squared and R-squared provide the same information in the comparison of the two models). Prediction R-squared is computed by leaving out one sample per time from the fitting procedure, and tells how good a model is in explaining the variance for new predictions. Such coefficient is related to a leave-one-out cross-validation procedure for evaluating the model prediction capabilities and makes use of the PRESS statistic [5, 191]. Test R-squared has the same meaning of R-squared but it is computed only on model validation data, so it tells how much variability is explained by the model on a completely new set of data. The comparison between these coefficients for both the models is reported in table 2.5. Nevertheless, the two models are not comparable from the performances point of view. An anticipation of the Maximum Absolute Error, Root Mean Squared Prediction Error (RMSPE) and Mean Absolute Percentage Error (MAPE) is reported in figure 2.6. As can be seen, the con-

Table 2.4: *ANOVA tables comparisons for CL and HF regression models.*

Source	DF	Seq SS	Contribution	Adj SS	Adj MS	F-Value	P-Value
Regression	6	20203366	98.92%	20203366	3367228	1743.79	0.000
Ktc	1	8566401	41.94%	41291	41291	21.38	0.000
Kte	1	265945	1.30%	3665	3665	1.90	0.171
Krc	1	74264	0.36%	593	593	0.31	0.581
Kre	1	2030	0.01%	2030	2030	1.05	0.307
Kac	1	11254475	55.11%	12612	12612	6.53	0.012
Kae	1	40252	0.20%	40252	40252	20.85	0.000
Error	114	220132	1.08%	220132	1931		
Lack-of-Fit	18	201960	0.99%	201960	11220	59.27	0.000
Pure Error	96	18172	0.09%	18172	189		
Total	120	20423498	100.00%				

(a) *CL model.*

Source	DF	Seq SS	Contribution	Adj SS	Adj MS	F-Value	P-Value
Regression	6	20396977	99.87%	20396977	3399496	14612.50	0.000
Ktc	1	8738358	42.79%	49331	49331	212.05	0.000
Kte	1	60524	0.30%	24944	24944	107.22	0.000
Krc	1	132467	0.65%	2936	2936	12.62	0.001
Kre	1	821	0.00%	821	821	3.53	0.063
Kac	1	11274175	55.20%	58	58	0.25	0.617
Kae	1	190632	0.93%	190632	190632	819.42	0.000
Error	114	26521	0.13%	26521	233		
Lack-of-Fit	18	8349	0.04%	8349	464	2.45	0.003
Pure Error	96	18172	0.09%	18172	189		
Total	120	20423498	100.00%				

(b) *HF model.*

ceived model outperforms the CL model in the prediction of mean forces, mean torque and mean power. Further comments and discussions on this will be provided in the dedicated following sections.

Figures 2.7a and 2.7b represent instead the diagnostics of the regression analysis performed with CL and HF models, respectively. The top-left graph (A) is the quantile-quantile graph for normal distributions. It is used to check qualitatively the normality assumption of residuals together with the bottom-left graph (C), which represents the discrete probability density function of residuals. If data follow a normal distribution, they fall close to the red straight line. These graphs are typically accompanied by normality hypothesis tests. Here, Anderson-Darling (AD) normality test was performed on model fitting residuals. The results of the tests were reported in table 2.6. AD test assumes under the null hypothesis that the data follow a normal distribution. The p-value

Table 2.5: *Model performance summaries comparison.*

Model	S	R-sq	R-sq(adj)	PRESS	R-sq(pred)	AICc	BIC	Test S	Test R-sq
CL	43.9429	98.92%	98.87%	240795	98.82%	1257.28	1275.80	41.1673	98.89%
HF	15.2526	99.87%	99.86%	29221	99.86%	1003.33	1021.84	15.2180	99.85%

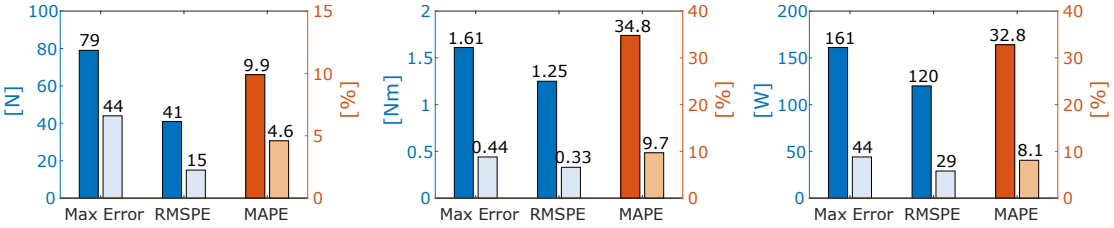


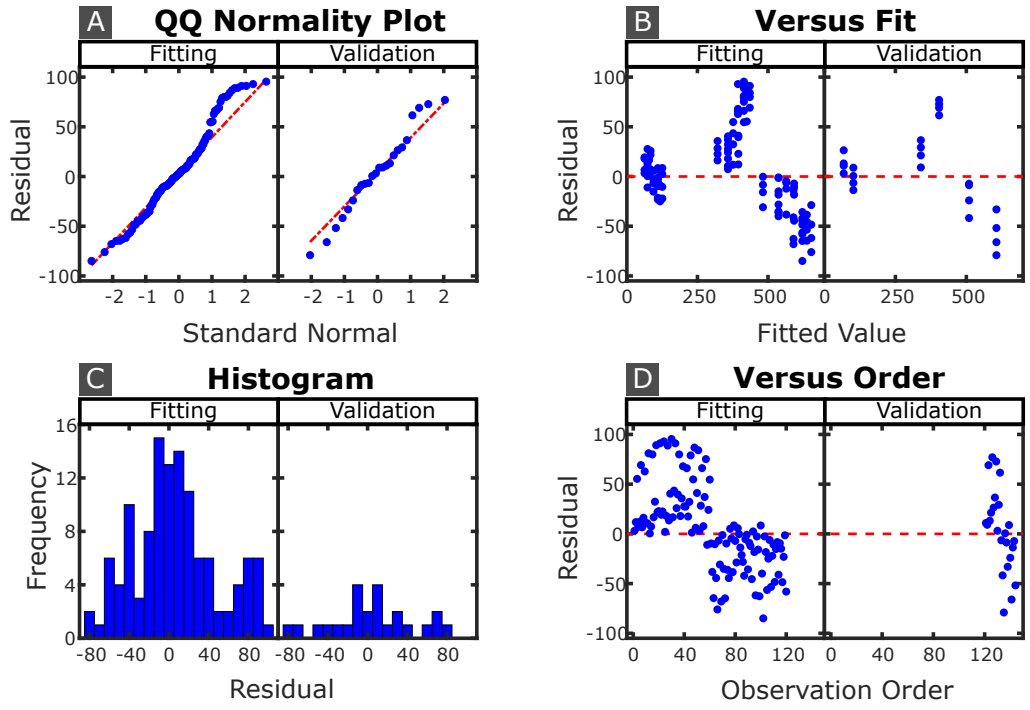
Figure 2.6: *Summary of error metrics for a comparison of HF and CL models. Maximum absolute error, root mean squared prediction error and Mean absolute percentage error metrics are reported for force (left), torque (centre) and power (right) predictions. Darker colours describe the CL model, while lighter bars refer to the HL model. Blue colours refer to left axes, whereas orange refers to right axes.*

Table 2.6: *Anderson-Darling normality tests.*

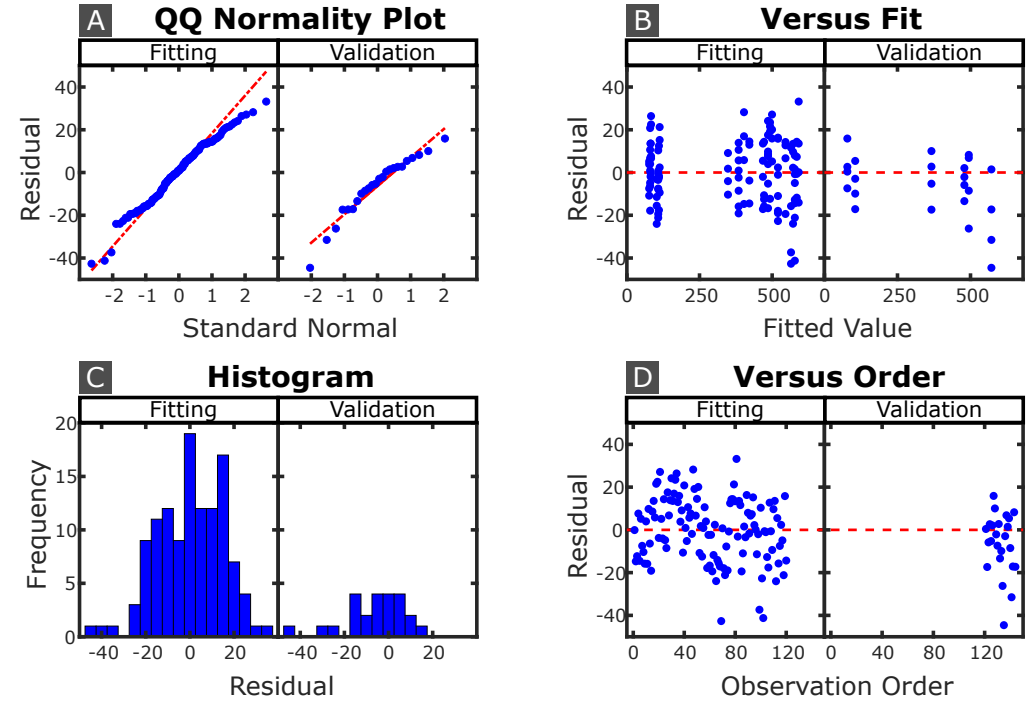
Model	Mean	StDev	N	AD	p-value
CL	6.470	42.52	120	0.995	0.012
HF	0.279	14.93	120	0.505	0.199

for CL residuals is 0.012 meaning that at 0.05 significance level, the null hypothesis is rejected. AD didn't reject the null hypothesis for the developed HF model, having a p-value of 0.199. This is another proof that the HF model is better representing the mean force data. The top-right graph (B) shows the residuals of regression versus the fitted values of mean forces. The interpretation of such graph is similar to the one in the bottom right corner (D), representing the residuals versus the experiment number. Looking at figure 2.7, it seems that homoscedasticity (i.e. model fitting residuals constant variance) is verified for both the models. Independence of model fitting residuals is not verified for the CL model, where residuals from observation number 1 to 60 are over the 0 line, while residuals from 61 to 120 are below the 0 line.

The different behaviour of the two models is also evident in the predicted instantaneous cutting forces. Figures 2.8a and 2.8b highlight the improved accuracy of the developed model in the description of the variable entry angles of the cutting inserts. In fact, the CL model features a portion of about 12° where no cutting inserts were engaged. On the contrary, the HF model correctly identifies a continuous cutting condition, where a cutting insert starts engaging the workpiece right in the moment where the previous one is exiting from it. This came from a peculiar choice in the nominal engagement b, determined to have exactly one cutter per time engaged in the workpiece. Nevertheless, it must be noted that the model catches a nominal behaviour of the instantaneous cutting forces, while experimental forces didn't show a null instantaneous force sample. This should be attributed to the dynamical components of the cutting process and to the run-out of the cutting tool. Radial vibrations of the tool tip cause small deviations in the radial engagement of the mill, causing a small delay between the entry of a cutter and the exit of the previous one. Radial cutters run-out causes small changes in their entry and exit angles, providing a similar effect. Furthermore, two other main considerations have to be brought to the reader attention: (i) the conceived model is, in general, better predicting the magnitude of the cutting forces; (ii) the HF model predictions for instantaneous cutting forces are featured by narrower 95% PI, due to the

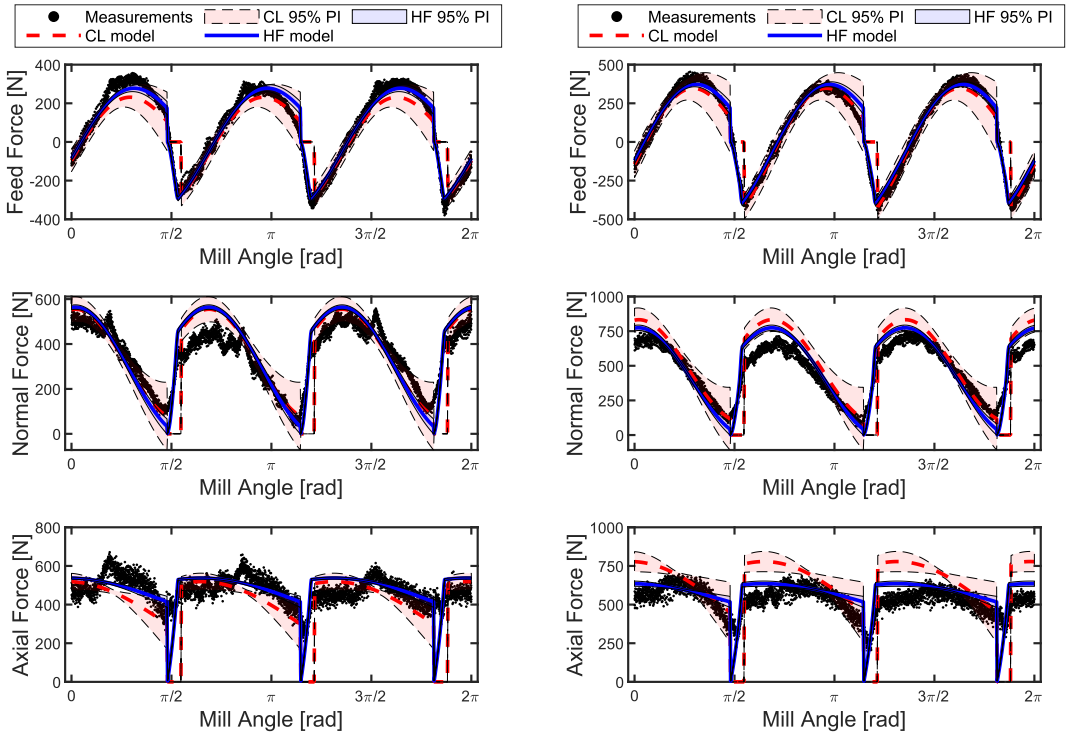


(a) *CL model.*



(b) *HF model.*

Figure 2.7: *Diagnostics of regression analysis. Quantiles-quantiles normality plot (A): the more residuals follow a normal distribution, the more the blue points lay on the red line. Residuals versus fitted responses (B): blue points should be randomly distributed around the red line to indicate their independence. Residuals histograms (C): should be as close as possible to a gaussian probability density function. Residuals versus observation (D): they should be randomly distributed and show constant variance.*



(a) *Test ID 4:* $a = 0.4\,mm$, $c = 0.65\,mm/tooth$ *and* $v_c = 55\,m/min$.

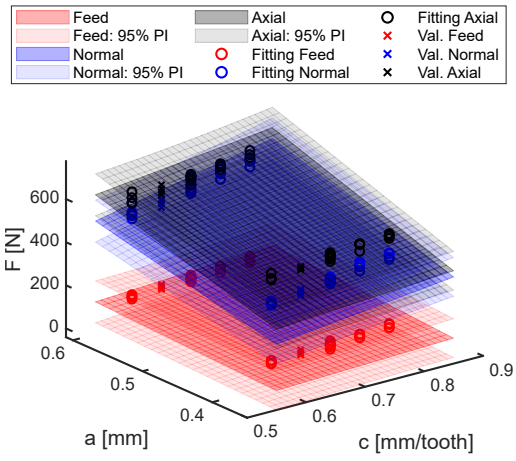
(b) *Test ID 10:* $a = 0.6\,mm$, $c = 0.65\,mm/tooth$ *and* $v_c = 55\,m/min$.

Figure 2.8: *Instantaneous force predictions of CL (dashed red line) and HF (solid blue line) models with respect to the mill angle. Coloured bands represent prediction intervals of the forces components and are drawn with analogous line style and colour.*

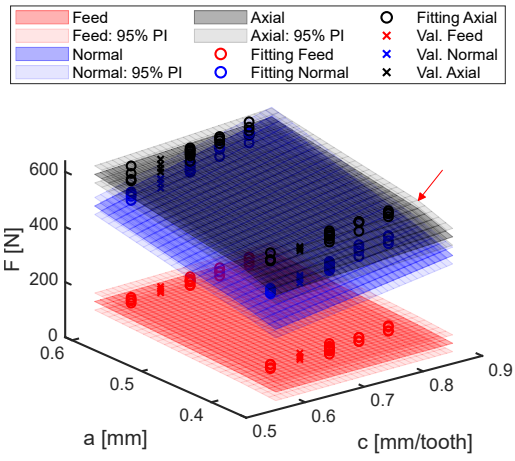
higher confidence in the regression coefficients estimation.

2.3.2 Model validation first step

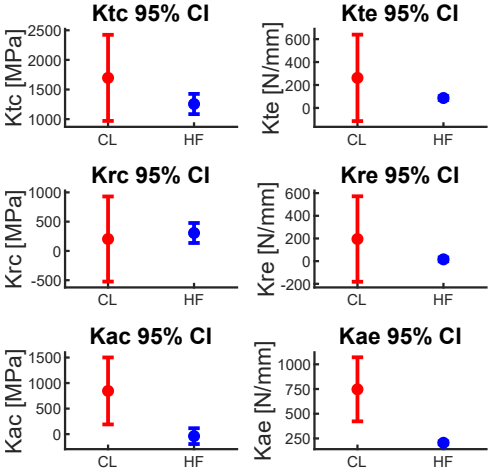
The behaviour pointed out in the last section is associated to the change in the axial depth of cut. For $a = 0.4\,mm$ the model is underestimating the mean cutting forces, while for $a = 0.6\,mm$ the model is overestimating them. The HF model is instead capable of well representing the mill cutting operation under both the axial depths of cut. This behaviour of the two models can be easily identified by looking at the mean force regression surfaces in the plane $a - c$. Such graphs are reported in figures 2.9a-2.9b, for the CL and HF models, respectively. In fact, the developed HF model includes two different behaviours, related to a lower then z_3 and a bigger than z_3 (i.e. the beginning of the engagement of the cutting edge second phase). Such distinction is visible in the response surface of the HF model (2.9b, red arrow). In correspondence of $a = z_3 = 0.4\,mm$, the response surface is featured by a sudden transition, due to the second cutting edge phase and to the modified chip geometry included in the model. The same wasn't observed in the CL model. This novelty allows the HF model to catch the different behaviours of the HF mills at different a values and not to under/overestimate the mean forces in the tested conditions. This phenomenon is even clearer while looking at the response curves of the two models at the tested a values, in a two-dimensional



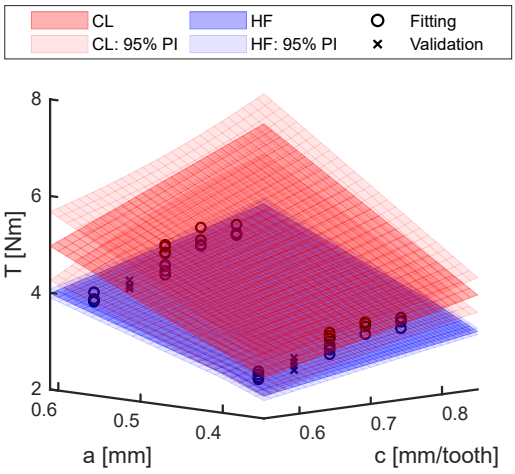
(a) *Prediction surfaces of CL model for mean forces.*



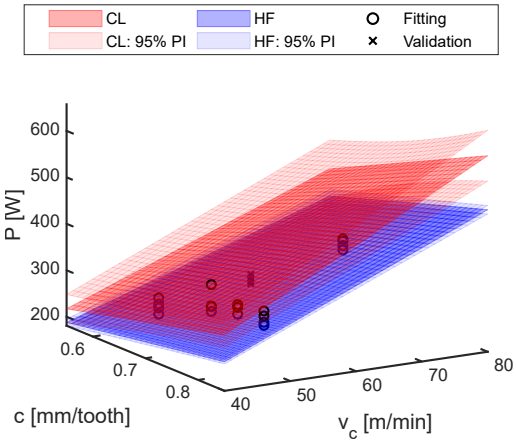
(b) *Prediction surfaces of HF model for mean forces.*



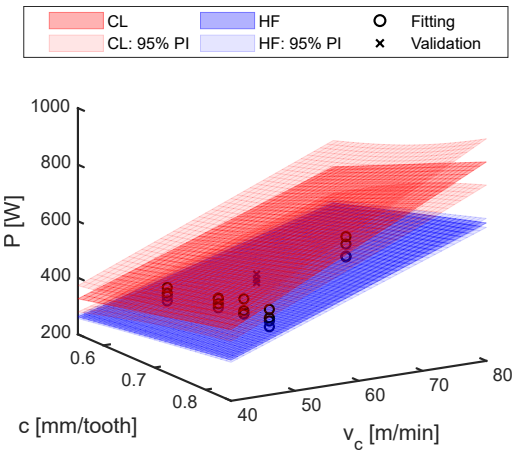
(c) *CL and HF SFC 95% CI comparison.*



(d) *Spindle torque: CL and HF models.*

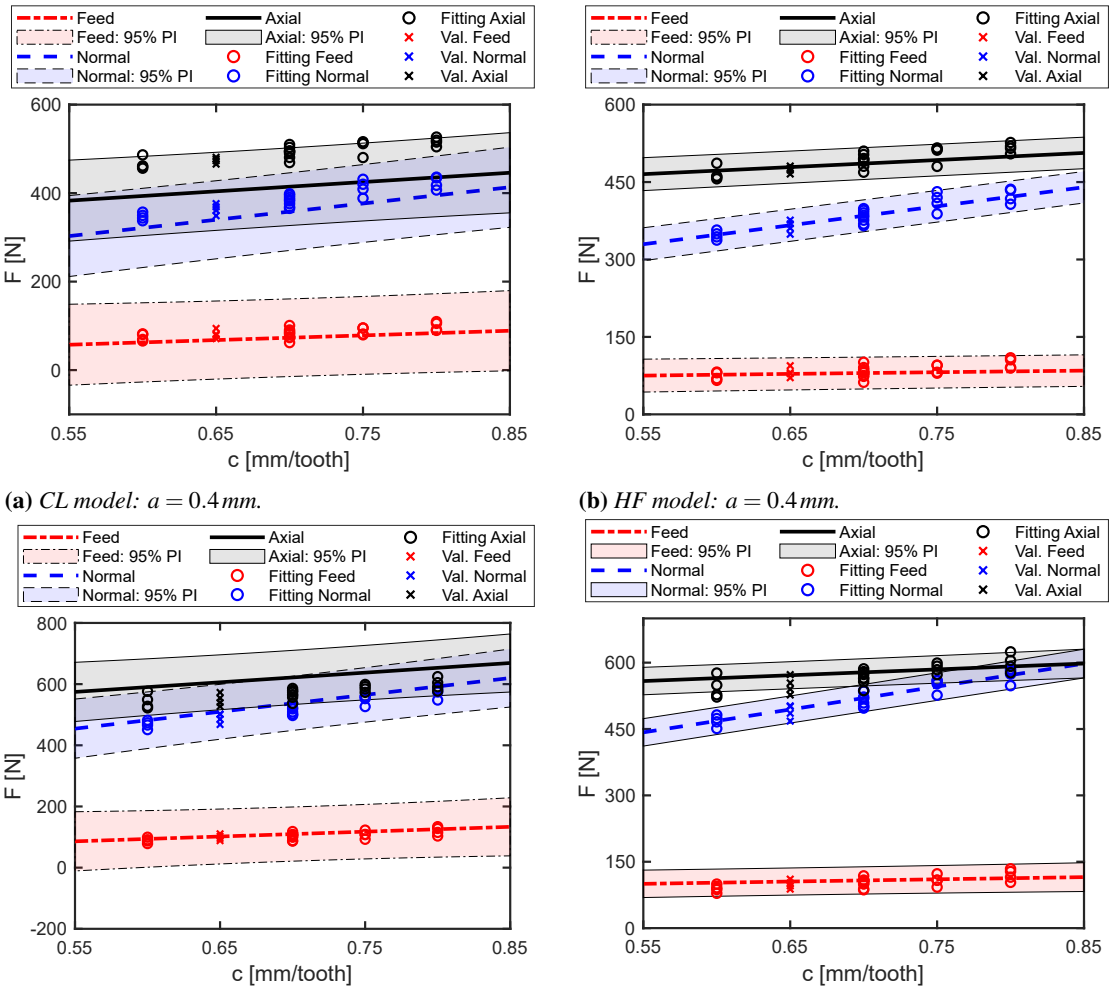


(e) *Spindle power: CL and HF models, $a = 0.4\,mm$.*



(f) *Spindle power: CL and HF models, $a = 0.6\,mm$.*

Figure 2.9: *Comparison summary between the CL and HF model for all the validation steps. (a) and (b) represent the first validation step, comparing the prediction capability of the two models with respect to mean forces. (c) summarises the second validation step, comparing the confidence intervals for the SFC estimation. (d) compares the prediction capabilities with respect to the mean torque, while (e) and (f) with respect to the mean power (thus, constituting the summary of the third validation step).*



(a) *CL model:* $a = 0.4\,mm$.

(b) *HF model:* $a = 0.4\,mm$.

(c) *CL model:* $a = 0.6\,mm$.

(d) *HF model:* $a = 0.6\,mm$.

Figure 2.10: *Mean force responses of CL and HF models versus c at the tested a values. Black solid lines represent the predicted mean axial forces, blue dashed lines describe normal forces, while red dashed-dotted lines represent feed forces. Circles represent data used for model fitting, whereas crosses represent validation data. Coloured bands represent prediction intervals of the forces components, being drawn with analogous line style and colour.*

plot (figures 2.10a-2.10d). From these figures, it is evident the difference between the two models. To quantitatively compare the two models, the Root Mean Squared Prediction Error (RMSPE - root mean squared error on the prediction of model validation data) was computed [22, 172]. The proposed model resulted to be remarkably more precise than the CL model, with an RMSPE of $15\,N$ against $41\,N$, respectively. The associated model validation maximum errors (in module) were $44\,N$ for HF and $79\,N$ for CL. Narrower SFC confidence intervals presented in table 2.3 translate in more accurate PI for the mean forces. Then, HF model is able to more reliably predict the mean cutting forces than the CL model. These results were confirmed also by the Mean Absolute Percentage Error (MAPE) metric [67, 111]. In fact, the MAPE obtained with the CL model was 9.9%, whereas the HL model allowed to reach a smaller value of 4.6%. Thus, the HF model dramatically decreased the MAPE on the validation tests,

Table 2.7: *Box M test for covariances equality.*

DF1	DF2	F	p-value
21	187857	142.00	0.000

Table 2.8: *Nel and van der Merwe procedures for Behrens-Fisher multivariate test for mean equality.*

DF1	DF2	F	p-value
6	169	17090.49	0.000

providing reliable predictions of mean forces.

2.3.3 Model validation second step

In order to statistically evaluate the difference between the two models, multivariate hypothesis testing was applied to the estimated regression coefficients. Literature proposes two approaches in order to compare the regression coefficients of two models [4]. The first one consists in the application of univariate 2t-tests on each pair of regression coefficients mean [40]. The second approach uses a joint ANOVA on the data, adding a categorical variable that assigns the data to the correspondent model. Interaction terms must be included in such regression problem [38]. The first approach is less conservative and tends to refuse the null hypothesis more frequently. This comes from the fact that the SFC are treated in a univariate fashion, while they are actually part of the same covariate set. Here, an extension of the method was proposed applying a multivariate approach. Box M test results were reported in table 2.7. The test refused the null hypothesis and, then, covariance equality of the CL and HF SFC. Because of such result, it was not possible to apply the classical Hotelling's T-squared test. The comparison between the means of the SFC was carried out through Nel and van der Merwe procedures to Behrens-Fisher multivariate tests [155]. The results of the test were reported in table 2.8. The p-value of the test was null, underlying statistical evidence about the difference in the HF and CL SFC means. It was then possible to compute the 95% simultaneous confidence intervals for the differences between each SFC pair. The intervals were reported in table 2.9. The intervals for the difference in means didn't include the zero, except for $K_{r,c}$. Then, the means of the SFC between CL and HF models were statistically different except for $K_{r,c}$. Such results confirmed that the CL model was not able to catch the physical phenomenon underneath high-feed milling

Table 2.9: *95% confidence intervals for mean differences between CL and HF SFC.*

SFC Difference	Lower	Upper
$\Delta K_{t,c}\,[MPa]$	-566.5	-316.0
$\Delta K_{t,e}\,[N/mm]$	-238.9	-112.3
$\Delta K_{r,c}\,[MPa]$	-22.3	228.2
$\Delta K_{r,e}\,[N/mm]$	-242.7	-116.1
$\Delta K_{a,c}\,[MPa]$	-998.0	-772.0
$\Delta K_{a,e}\,[N/mm]$	-595.4	-486.5

and further supported the need for the developed HF model. The difference between the estimated SFC using CL and HF models was highlighted by figure 2.9c. It is clear also from this figure that $K_{r,c}$ was the only equal SFC. The identified HF $K_{t,c}$ resulted to be significantly lower than the CL one. This difference is probably due to the fact that the HF model takes into account a much longer edge contact length, reflecting in a wider contribution of the ploughing effect under the flank face of the cutting edge. Of course, being the edge contact length longer, even the $K_{t,e}$ assumes a lower value. The axial force is dominated by the ploughing effects. The opposite occurs for the radial components which are basically constituted by the shearing action on the shearing plane.

2.3.4 Model validation third and fourth steps

To further validate the model, the prediction of variables of technological interest not included in the fitting procedure was carried out. Spindle torque and power are in fact useful to assess the feasibility of an operation on the available machine tool. The prediction surfaces of the spindle torque for the CL and HF model were reported in figure 2.9d. It is evident that the HF model outperformed the CL one both in the mean torque prediction as well as in the confidence of the prediction. The better performance of the HF model in the prediction of this quantities lies in two main contributions: on one hand, it is related to the better estimation of the SFC; on the other hand, on the fact that the contribution to the torque of an infinitesimal piece of cutting edge is proportional to its distance from the mill axis. The second factor was in fact considered in the HF model, while it was not present in the original CL model. Furthermore, the higher confidence in the prediction of spindle torque was due just to the first of the two contributions. In fact the prediction uncertainties were obtained from the SFC confidence intervals propagation. Being more confident on the SFC turned of course in a higher confidence in the prediction. The RMSPE for the torques were equal to $0.33\,Nm$ and $1.25\,Nm$, for the HF and CL respectively. The associated validation maximum errors (in module) were $0.44\,Nm$ for HF and $1.61\,Nm$ for CL. The MAPE metric relative to spindle torque underlined the improvements introduced with the HF model, assuming values of 9.7% and 34.8% for the HF and CL models, respectively. The same reasonings were valid for the spindle power. Figures 2.9e and 2.9f represented the spindle power versus the cutting speed and the feed per tooth under $a = 0.4\,mm$ and $a = 0.6\,mm$, respectively. For the cutting power, RMSPE resulted to be $29\,W$ and $120\,W$, for the HF and CL respectively. The associated validation maximum errors (in module) were $44\,W$ for HF and $161\,W$ for CL. The associated MAPE values were 8.1% for HF and 32.8% for CL.

2.4 Conclusions

In this chapter, a novel analytical mechanistic model for double-phased high-feed mills was proposed. Double-phased cutters allow to reach high feed per tooth values, while keeping a low axial depth of cut; this permits to increase the material removal rate of face-milling operations and, at the same time, to avoid cutting instability. The new formulation extended the reference literature model to include complex cutter geometry and variable engagement conditions along the axial coordinate of the tool. A longer

edge contact length led to lower specific force coefficients and to the predominance of ploughing under the flank face of the cutting insert.

The developed high-feed model introduced the above improvements while keeping the formulation analytical and in closed form, both for instantaneous and mean cutting forces expressions. The model computational time remains low, relying upon neither discretisation nor time integration. The model fitting procedure on mean forces takes 46.7*ms* for the proposed model, whereas 12.3*ms* for the literature one.

A reliable estimate of the specific force coefficients was achieved, with narrower confidence intervals with respect to the literature model, underlining a better representation of the phenomenon. Based on this, the developed model outperformed the literature one in mean forces fitting and prediction, while respecting all the regression assumptions. The same high-feed model could be used for multiple cutting conditions: low (0.4*mm*) and high (0.6*mm*) axial depth of cuts. The Root Mean Squared Prediction Errors were 15 N and 41 N, with associated validation maximum errors (in module) of 44 N and 79 N for the proposed and literature model, respectively.

The literature model was outperformed also in the prediction of spindle torque and power. The developed model predictions were considerably more accurate in terms of mean, reaching Root Mean Squared Prediction Errors of (0.33 Nm, 29 W), compared with (1.25 Nm, 120 W) of the literature one, respectively. The associated validation maximum errors (in module) were (0.44 Nm, 44 W) for proposed model and (1.61 Nm, 161 W) for the literature model. Narrower specific force coefficients confidence intervals translated into reduced uncertainties in the prediction of the above quantities. This allows for reliable comparisons between operations within a sustainable perspective.

Future works will regard the application of the developed model for energetic comparison between conventional and high-feed milling, as well as its application for the monitoring and prognostics of high-feed tools.

2.A Theoretical background

In this appendix, the CL model developed by Altintas [26] is presented (fig. 2.1A-B). The tool geometry is described by the diameter of the mill D and the number of cutters

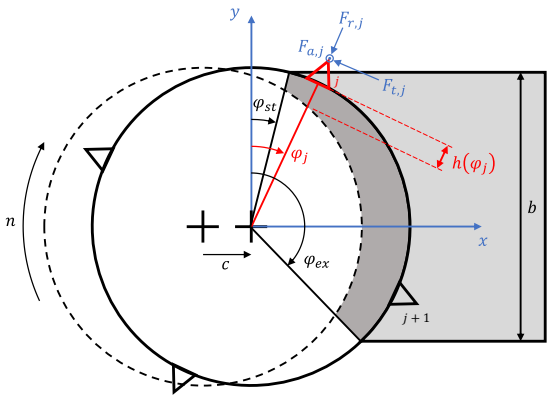


Figure 2.A.1: *Reference figure for the literature mechanistic model by Altintas. Figure represents the nomenclature of a simple milling operation: the angular position of cutters in the feed-normal reference frame, the radial engagement of the mill and the cutting forces in the cutter reference frame.*

N (fig. 2.A.1). The machining process is featured by the feed per tooth c and the rotational speed n. The formulation starts by defining the position in time of the cutters as for the HF model (eq. 2.1). The instantaneous chip thickness removed by cutter j is approximated by equation 2.2, too. Two angles are defined to determine whether the j-th cutter is processing material or not: φ_{st} determines the angular position at which the cutter starts to work, φ_{ex} determines the angular position at which the cutter exits the workpiece. The engagement function is then defined (eq. 2.A.1):

$$g(\varphi_j) = \begin{cases} 1, & \text{if } \varphi_{st} < \varphi_j < \varphi_{ex} \\ 0, & \text{otherwise.} \end{cases} \tag{2.A.1}$$

The three components of the cutting forces (i.e. tangential, radial and axial) acting on cutter j are expressed in differential form as for the HF model (eq. 2.9). dA and dl are instead the infinitesimal chip area and edge contact length and are defined through (eq. 2.A.2):

$$\begin{aligned} \mathrm{d}A(\varphi_j) &= h(\varphi_j)\,\mathrm{d}z \\ \mathrm{d}l(\varphi_j) &= \mathrm{d}z \end{aligned} \tag{2.A.2}$$

Substituting equation 2.A.2 in equation 2.9, and integrating in dz from 0 to the axial depth of cut a, the expressions for the three forces components are obtained:

$$\begin{aligned} F_{t,j}(\varphi_j) &= g(\varphi_j)\,\left[K_{t,c}\,a\,h(\varphi_j) + K_{t,e}\,a\right] \\ F_{r,j}(\varphi_j) &= g(\varphi_j)\,\left[K_{r,c}\,a\,h(\varphi_j) + K_{r,e}\,a\right] \\ F_{a,j}(\varphi_j) &= g(\varphi_j)\,\left[K_{a,c}\,a\,h(\varphi_j) + K_{a,e}\,a\right] \end{aligned} \tag{2.A.3}$$

Each force term must be multiplied by $g(\varphi_j)$ to include the engagement condition in the workpiece. The force components are then projected in the feed, normal and axial direction of the milling operation, and the contribution of each cutter is summed up through equation 2.12.

The identification of the six SFC is based on mean cutting forces as for the HF model [54, 166]. Indeed, the three force components are integrated over one period of revolution T_P of the mill (eq. 2.A.4):

$$\overline{F}_d = \frac{1}{T_P}\int_0^{T_P} F_d(t)\,\mathrm{d}t \tag{2.A.4}$$

where $d = x, y, z$. The resulting mean forces expressions are reported in eq. (2.A.5):

$$\begin{aligned}
\overline{F}_x =& \frac{N a c}{8\pi} \left[K_{t,c} \cos 2\varphi - K_{r,c} \left(2\varphi - \sin 2\varphi\right)\right]\Big|_{\varphi_{st}}^{\varphi_{ex}} + \\
&+ \frac{N a}{2\pi} \left(-K_{t,e} \sin\varphi + K_{r,e} \cos\varphi\right)\Big|_{\varphi_{st}}^{\varphi_{ex}} \\
\\
\overline{F}_y =& \frac{N a c}{8\pi} \left[K_{t,c} \left(2\varphi - \sin 2\varphi\right) + K_{r,c} \cos 2\varphi\right]\Big|_{\varphi_{st}}^{\varphi_{ex}} + \\
&- \frac{N a}{2\pi} \left(K_{t,e} \cos\varphi + K_{r,e} \sin\varphi\right)\Big|_{\varphi_{st}}^{\varphi_{ex}} \\
\\
\overline{F}_z =& \frac{N a}{2\pi} \left(-K_{a,c}\, c \cos\varphi + K_{a,c}\, \varphi\right)\Big|_{\varphi_{st}}^{\varphi_{ex}}
\end{aligned} \tag{2.A.5}$$

Equation 2.A.5 represents the mean cutting forces according to the CL model. Cutting torque T is defined in an instantaneous fashion through equation 2.A.6:

$$T(t) = \frac{D}{2} \sum_{j=1}^{N} F_{t,j}(t) \tag{2.A.6}$$

Cutting power P is instead (eq. 2.A.7):

$$P(t) = 2\pi n T(t) \tag{2.A.7}$$

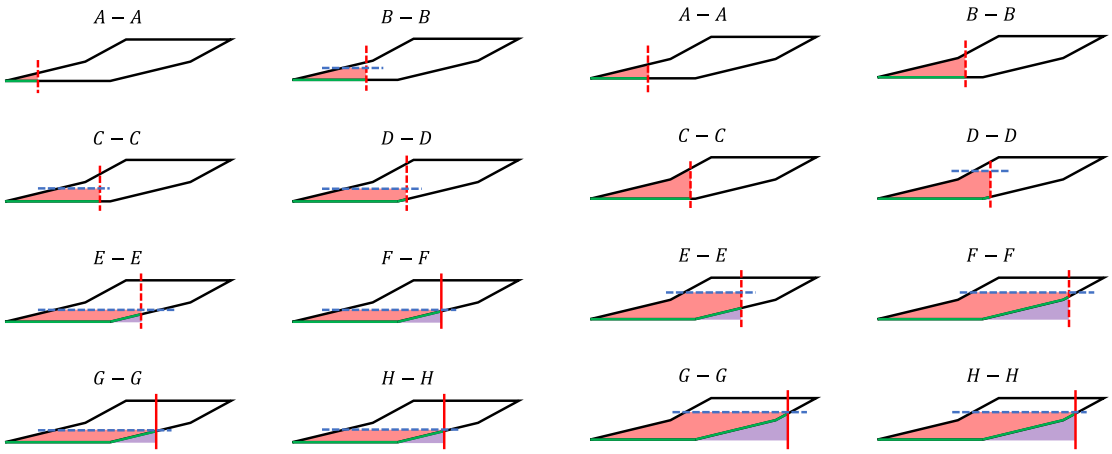
2.B Mathematical computations: instantaneous forces terms

Equation 2.11 included three terms that were not expanded in the body of the chapter, for clarity reasons. The scope of this appendix is to provide the mathematical expressions and derivations of such terms. The terms that are needed for the computation of the instantaneous forces in equation 2.11 are $A_{int}(\varphi_j)$, $A_{ext}(\varphi_j)$ and $l(\varphi_j)$. Since the formulation does not depend on the cutter index j, such index will be omitted in the next formulas.

2.B.1 Undeformed internal chip area $A_{int}(\varphi)$

This term corresponds to the area underneath the previous cutter position graph in radial-axial chart. The reference graph is shown in figure 2.B.1. The two cases represent the condition whether $a \leq z_3$ (fig. 2.B.1a) or $a > z_3$ (fig. 2.B.1b). This distinction is needed since it determines whether the second-phase of the cutter is engaged or not. In case $a \leq z_3$, $A_{int}(\varphi)$ can be expressed as a sum of areas of triangles and trapezoids:

$$A_{int}(\varphi) = \begin{cases}
0 & \text{if } 0 \leq \varphi \leq \varphi_{2,int} \\
\frac{1}{2} \frac{\left[r_\varphi(\varphi) - r_2 + h(\varphi)\right]^2}{r_3 - r_2} z_3 & \text{if } \varphi_{2,int} < \varphi \leq \varphi_{a,int} \\
\frac{1}{2} a \left[r_{a,int}(\varphi) - r_2 + h(\varphi)\right] + a \left[r_\varphi(\varphi) - r_{a,int}(\varphi)\right] & \text{if } \varphi_{a,int} < \varphi \leq \varphi_{a,ext} \\
\frac{1}{2} a \left[r_{a,int}(\varphi) - r_2 + h(\varphi)\right] + a \left[r_{a,ext} - r_{a,int}(\varphi)\right] & \text{if } \varphi_{a,ext} < \varphi \leq \pi \\
0 & \text{if } \pi < \varphi < 2\pi
\end{cases} \tag{2.B.1}$$



(a) *Case* $a \leq z_3$. **(b)** *Case* $a > z_3$.

Figure 2.B.1: *Representation of the main quantities defining the chip shape. The red area is* $A_{int}(\varphi)$*; violet area is the superimposition of* $A_{int}(\varphi)$ *and* $A_{ext}(\varphi)$*; green solid line is* $l(\varphi)$*; red dashed line is* $r_\varphi(\varphi)$ *and red solid line is* $r_\varphi(\varphi) = r_{a,ext}$*; blue dashed line represents the axial depth of cut limit.*

Instead, if $a > z_3$, the formulation becomes:

$$A_{int}(\varphi) = \begin{cases} 0, & \text{if } 0 \leq \varphi \leq \varphi_{2,int} \\ \frac{1}{2}\frac{\left[r_\varphi(\varphi)-r_2+h(\varphi)\right]^2}{r_3-r_2} z_3, & \text{if } \varphi_{2,int} < \varphi \leq \varphi_{3,int} \\ \frac{1}{2}(r_3-r_2)z_3 + \left[r_\varphi(\varphi)-r_3+h(\varphi)\right]z_3 + \frac{1}{2}\left[r_\varphi(\varphi)-r_3+h(\varphi)\right]^2 \frac{z_4-z_3}{r_4-r_3}, & \text{if } \varphi_{3,int} < \varphi \leq \varphi_{a,int} \\ \frac{1}{2}(r_3-r_2)z_3 + \frac{1}{2}(z_3+a)\left[r_{a,int}(\varphi)-r_3+h(\varphi)\right] + a\left[r_\varphi(\varphi)-r_{a,int}(\varphi)\right], & \text{if } \varphi_{a,int} < \varphi \leq \varphi_{a,ext} \\ \frac{1}{2}(r_3-r_2)z_3 + \frac{1}{2}(z_3+a)\left[r_{a,int}(\varphi)-r_3+h(\varphi)\right] + a\left[r_{a,ext}-r_{a,int}(\varphi)\right], & \text{if } \varphi_{a,ext} < \varphi \leq \pi \\ 0, & \text{if } \pi < \varphi < 2\pi \end{cases} \tag{2.B.2}$$

2.B.2 Undeformed external chip area $A_{ext}(\varphi)$

The same cases are also valid for $A_{ext}(\varphi)$, which represents the area under the graph of the current insert cutting edge in the radial-axial chart. Figures 2.B.1a and 2.B.1b are again the references. In case $a \leq z_3$, $A_{ext}(\varphi)$ is:

$$A_{ext}(\varphi) = \begin{cases} 0, & \text{if } 0 \leq \varphi \leq \varphi_{2,ext} \\ \frac{1}{2}\frac{\left[r_\varphi(\varphi)-r_2\right]^2}{r_3-r_2} z_3, & \text{if } \varphi_{2,ext} < \varphi \leq \varphi_{a,ext} \\ \frac{1}{2}a(r_{a,ext}-r_2), & \text{if } \varphi_{a,ext} < \varphi \leq \pi \\ 0, & \text{if } \pi < \varphi < 2\pi \end{cases} \tag{2.B.3}$$

Instead, if $a > z_3$, the formulation becomes:

$$A_{ext}(\varphi) = \begin{cases} 0, & \text{if } 0 \leq \varphi \leq \varphi_{2,ext} \\ \frac{1}{2}\frac{\left[r_\varphi(\varphi)-r_2\right]^2}{r_3-r_2} z_3, & \text{if } \varphi_{2,ext} < \varphi \leq \varphi_{3,ext} \\ \frac{1}{2}z_3(r_3-r_2) + z_3\left[r_\varphi(\varphi)-r_3\right] + \frac{1}{2}(z_4-z_3)\frac{\left[r_\varphi(\varphi)-r_3\right]^2}{r_4-r_3}, & \text{if } \varphi_{3,ext} < \varphi \leq \varphi_{a,ext} \\ \frac{1}{2}z_3(r_3-r_2) + \frac{1}{2}(z_3+a)(r_{a,ext}-r_3), & \text{if } \varphi_{a,ext} < \varphi \leq \pi \\ 0, & \text{if } \pi < \varphi < 2\pi \end{cases} \tag{2.B.4}$$

2.B.3 Edge contact length $l(\varphi)$

This term represents the length of the cutting edge which is involved in the cut of the chip. It is the length of the engaged part of the actual cutter. The same computations as before are done also for $l(\varphi)$ (fig. 2.B.1a and 2.B.1b). Here, a supplementary term is added to the integral, which is the minor cutting edge contribution. The resultant formulas for $a \leq z_3$ are:

$$l(\varphi) = \begin{cases} 0, & \text{if } 0 \leq \varphi \leq \varphi_{2,int} \\ r_\varphi(\varphi) - r_2 + h(\varphi), & \text{if } \varphi_{2,int} \leq \varphi \leq \varphi_{2,ext} \\ h(\varphi) + \left[r_\varphi(\varphi) - r_2\right] \sqrt{1 + \left(\frac{z_3}{r_3 - r_2}\right)^2}, & \text{if } \varphi_{2,ext} \leq \varphi \leq \varphi_{a,ext} \\ h(\varphi) + \sqrt{(r_{a,ext} - r_2)^2 + a^2}, & \text{if } \varphi_{a,ext} \leq \varphi \leq \pi \\ 0, & \text{if } \pi < \varphi < 2\pi \end{cases} \tag{2.B.5}$$

In case $a > z_3$:

$$l(\varphi) = \begin{cases} 0, & \text{if } 0 \leq \varphi \leq \varphi_{2,int} \\ r_\varphi(\varphi) - r_2 + h(\varphi), & \text{if } \varphi_{2,int} \leq \varphi \leq \varphi_{2,ext} \\ h(\varphi) + \left[r_\varphi(\varphi) - r_2\right] \sqrt{1 + \left(\frac{z_3}{r_3 - r_2}\right)^2}, & \text{if } \varphi_{2,ext} \leq \varphi \leq \varphi_{3,ext} \\ h(\varphi) + \sqrt{(r_3 - r_2)^2 + z_3^2} + \left[r_\varphi(\varphi) - r_3\right] \sqrt{1 + \left(\frac{z_4 - z_3}{r_4 - r_3}\right)^2}, & \text{if } \varphi_{3,ext} \leq \varphi \leq \varphi_{a,ext} \\ h(\varphi) + \sqrt{(r_3 - r_2)^2 + z_3^2} + \sqrt{(r_{a,ext} - r_3)^2 + (a - z_3)^2}, & \text{if } \varphi_{a,ext} \leq \varphi \leq \pi \\ 0, & \text{if } \pi < \varphi < 2\pi \end{cases} \tag{2.B.6}$$

2.B.4 Torque area equivalent radius $r_A^*(\varphi)$

Furthermore, two more quantities need to be expressed for equation 2.15. The first one is $r_A^*(\varphi)$. Starting with case $a \leq z_3$:

$$r^*_{A,int}(\varphi) = \begin{cases} 0, & \text{if } 0 \leq \varphi \leq \varphi_{2,int} \\ r_2 - h(\varphi) + \frac{2[r_\varphi(\varphi) - r_2 + h(\varphi)]}{3}, & \text{if } \varphi_{2,int} < \varphi \leq \varphi_{a,int} \\ \frac{1}{A_{int}(\varphi)} \left\{ \left[r_2 - h(\varphi) + \frac{2(r_{a,int}(\varphi) - r_2 + h(\varphi))}{3} \right] \frac{a[r_{a,int}(\varphi) - r_2 + h(\varphi)]}{2} + \right. \\ \left. + \frac{[r_\varphi(\varphi) + r_{a,int}(\varphi)] a[r_\varphi(\varphi) - r_{a,int}(\varphi)]}{2} \right\}, & \text{if } \varphi_{a,int} < \varphi \leq \varphi_{a,ext} \\ \frac{1}{A_{int}(\varphi)} \left\{ \left[r_2 - h(\varphi) + \frac{2(r_{a,int}(\varphi) - r_2 + h(\varphi))}{3} \right] \frac{a[r_{a,int}(\varphi) - r_2 + h(\varphi)]}{2} + \right. \\ \left. + \frac{[r_{a,ext} + r_{a,int}(\varphi)] a[r_{a,ext} - r_{a,int}(\varphi)]}{2} \right\}, & \text{if } \varphi_{a,ext} < \varphi \leq \pi \\ 0, & \text{if } \pi < \varphi < 2\pi \end{cases} \tag{2.B.7}$$

$$r^*_{A,ext}(\varphi) = \begin{cases} 0, & \text{if } 0 \leq \varphi \leq \varphi_{2,ext} \\ r_2 + \frac{2[r_\varphi(\varphi) - r_2]}{3}, & \text{if } \varphi_{2,ext} < \varphi \leq \varphi_{a,ext} \\ r_2 + \frac{2 \cdot (r_{a,ext} - r_2)}{3}, & \text{if } \varphi_{a,ext} < \varphi \leq \pi \\ 0, & \text{if } pi < \varphi < 2\pi \end{cases}$$

$$r^*_A(\varphi) = \begin{cases} \frac{r^*_{A,int}(\varphi) \cdot A_{int}(\varphi) - r^*_{A,ext}(\varphi) \cdot A_{ext}(\varphi)}{A_{int}(\varphi) - A_{ext}(\varphi)}, & \text{if } A_{int}(\varphi) - A_{ext}(\varphi) \neq 0 \\ 0, & \text{if } A_{int}(\varphi) - A_{ext}(\varphi) = 0 \end{cases}$$

While in case $a > z_3$, it is defined as follows:

$$r^*_{A,int}(\varphi) = \begin{cases} 0, & \text{if } 0 \le \varphi \le \varphi_{2,int} \\ r_2 - h(\varphi) + \frac{2[r_\varphi(\varphi) - r_2 + h(\varphi)]}{3}, & \text{if } \varphi_{2,int} < \varphi \le \varphi_{3,int} \\ \frac{1}{A_{int}(\varphi)} \left\{ \left[r_2 - h(\varphi) + \frac{2(r_3 - r_2)}{3} \right] \frac{(r_3 - r_2) z_3}{2} + \frac{[r_3 - h(\varphi) + r_\varphi(\varphi)][r_\varphi(\varphi) - r_3 + h(\varphi)] z_3}{2} + \right. \\ \left. + \left[r_3 - h(\varphi) + \frac{2(r_\varphi(\varphi) - r_3 + h(\varphi))}{3} \right] \frac{(z_4 - z_3)[r_\varphi(\varphi) - r_3 + h(\varphi)]^2}{2(r_4 - r_3)} \right\}, & \text{if } \varphi_{3,int} < \varphi \le \varphi_{a,int} \\ \frac{1}{A_{int}(\varphi)} \left\{ \left[r_2 - h(\varphi) + \frac{2(r_3 - r_2)}{3} \right] \frac{(r_3 - r_2) z_3}{2} + \frac{[r_3 - h(\varphi) + r_{a,int}(\varphi)][r_{a,int}(\varphi) - r_3 + h(\varphi)] z_3}{2} + \right. \\ + \left[r_3 - h(\varphi) + \frac{2(r_{a,int}(\varphi) - r_3 + h(\varphi))}{3} \right] \frac{(a - z_3)[r_{a,int}(\varphi) - r_3 + h(\varphi)]}{2} + \\ \left. + \frac{[r_{a,int}(\varphi) + r_\varphi(\varphi)][r_\varphi(\varphi) - r_{a,int}(\varphi)] a}{2} \right\}, & \text{if } \varphi_{a,int} < \varphi \le \varphi_{a,ext} \\ \frac{1}{A_{int}(\varphi)} \left\{ \left[r_2 - h(\varphi) + \frac{2(r_3 - r_2)}{3} \right] \frac{(r_3 - r_2) z_3}{2} + \frac{[r_3 - h(\varphi) + r_{a,int}(\varphi)][r_{a,int}(\varphi) - r_3 + h(\varphi)] z_3}{2} + \right. \\ + \left[r_3 - h(\varphi) + \frac{2(r_{a,int}(\varphi) - r_3 + h(\varphi))}{3} \right] \frac{(a - z_3)[r_{a,int}(\varphi) - r_3 + h(\varphi)]}{2} + \\ \left. + \frac{[r_{a,int}(\varphi) + r_{a,ext}][r_{a,ext} - r_{a,int}(\varphi)] a}{2} \right\}, & \text{if } \varphi_{a,ext} < \varphi \le \pi \\ 0, & \text{if } \pi < \varphi < 2\pi \end{cases}$$

$$r^*_{A,ext}(\varphi) = \begin{cases} 0, & \text{if } 0 \le \varphi \le \varphi_{2,ext} \\ r_2 + \frac{2[r_\varphi(\varphi) - r_2]}{3}, & \text{if } \varphi_{2,ext} < \varphi \le \varphi_{3,ext} \\ \frac{1}{A_{ext}(\varphi)} \left\{ \left[r_2 + \frac{2(r_3 - r_2)}{3} \right] \left[\frac{z_3(r_3 - r_2)}{2} \right] + \frac{z_3[r_\varphi(\varphi) + r_3][r_\varphi(\varphi) - r_3]}{2} + \right. \\ \left. + \left[r_3 + \frac{2(r_\varphi(\varphi) - r_3)}{3} \right] \frac{(z_4 - z_3)[r_\varphi(\varphi) - r_3]^2}{2(r_4 - r_3)} \right\}, & \text{if } \varphi_{3,ext} < \varphi \le \varphi_{a,ext} \\ \frac{1}{A_{ext}(\varphi)} \left\{ \left[r_2 + \frac{2(r_3 - r_2)}{3} \right] \left[\frac{z_3(r_3 - r_2)}{2} \right] + \frac{z_3(r_{a,ext} + r_3)(r_{a,ext} - r_3)}{2} + \right. \\ \left. + \left[r_3 + \frac{2(r_{a,ext} - r_3)}{3} \right] \frac{(z_4 - z_3)(r_{a,ext} - r_3)^2}{2(r_4 - r_3)} \right\}, & \text{if } \varphi_{a,ext} < \varphi \le \pi \\ 0, & \text{if } \pi < \varphi < 2\pi \end{cases}$$

$$r^*_A(\varphi) = \begin{cases} \frac{r^*_{A,int}(\varphi) \cdot A_{int}(\varphi) - r^*_{A,ext}(\varphi) \cdot A_{ext}(\varphi)}{A_{int}(\varphi) - A_{ext}(\varphi)}, & \text{if } A_{int}(\varphi) - A_{ext}(\varphi) \neq 0 \\ 0, & \text{if } A_{int}(\varphi) - A_{ext}(\varphi) = 0 \end{cases} \tag{2.B.8}$$

2.B.5 Torque edge contact length equivalent radius $r_l^*(\varphi)$

The second term that needs to be expressed for equation 2.15 is $r_l^*(\varphi)$. In case $a \leq z_3$, this term is equal to:

$$r_l^*(\varphi) = \begin{cases} 0, & \text{if } 0 \leq \varphi \leq \varphi_{2,int} \\ \frac{r_\varphi(\varphi)+r_2-h(\varphi)}{2}, & \text{if } \varphi_{2,int} \leq \varphi \leq \varphi_{2,ext} \\ \frac{1}{l(\varphi)} \left\{ \left[r_2 - \frac{h(\varphi)}{2} \right] h(\varphi) + \frac{[r_\varphi(\varphi)+r_2][r_\varphi(\varphi)-r_2]}{2} \sqrt{1+\left(\frac{z_3}{r_3-r_2}\right)^2} \right\}, & \text{if } \varphi_{2,ext} \leq \varphi \leq \varphi_{a,ext} \\ \frac{1}{l(\varphi)} \left\{ \left[r_2 - \frac{h(\varphi)}{2} \right] h(\varphi) + \frac{r_{a,ext}+r_2}{2} \sqrt{a^2+(r_{a,ext}-r_2)^2} \right\}, & \text{if } \varphi_{a,ext} \leq \varphi \leq \pi \\ 0, & \text{if } \pi < \varphi < 2\pi \end{cases} \tag{2.B.9}$$

In case $a > z_3$:

$$r_l^*(\varphi) = \begin{cases} 0, & \text{if } 0 \leq \varphi \leq \varphi_{2,int} \\ \frac{r_\varphi(\varphi)+r_2-h(\varphi)}{2}, & \text{if } \varphi_{2,int} \leq \varphi \leq \varphi_{2,ext} \\ \frac{1}{l(\varphi)} \left\{ \left[r_2 - \frac{h(\varphi)}{2} \right] h(\varphi) + \frac{[r_\varphi(\varphi)+r_2][r_\varphi(\varphi)-r_2]}{2} \sqrt{1+\left(\frac{z_3}{r_3-r_2}\right)^2} \right\}, & \text{if } \varphi_{2,ext} \leq \varphi \leq \varphi_{3,ext} \\ \frac{1}{l(\varphi)} \left\{ \left[r_2 - \frac{h(\varphi)}{2} \right] h(\varphi) + \frac{(r_3+r_2)(r_3-r_2)}{2} \sqrt{1+\left(\frac{z_3}{r_3-r_2}\right)^2} + \right. & \\ \left. + \frac{[r_\varphi(\varphi)+r_3][r_\varphi(\varphi)-r_3]}{2} \sqrt{1+\left(\frac{z_4-z_3}{r_4-r_3}\right)^2} \right\}, & \text{if } \varphi_{3,ext} \leq \varphi \leq \varphi_{a,ext} \\ \frac{1}{l(\varphi)} \left\{ \left[r_2 - \frac{h(\varphi)}{2} \right] h(\varphi) + \frac{(r_3+r_2)(r_3-r_2)}{2} \sqrt{1+\left(\frac{z_3}{r_3-r_2}\right)^2} + \right. & \\ \left. + \frac{(r_{a,ext}+r_3)(r_{a,ext}-r_3)}{2} \sqrt{1+\left(\frac{z_4-z_3}{r_4-r_3}\right)^2} \right\}, & \text{if } \varphi_{a,ext} \leq \varphi \leq \pi \\ 0, & \text{if } \pi < \varphi < 2\pi \end{cases} \tag{2.B.10}$$

2.C Mathematical computations: mean forces terms

Equation 2.13 presented ten terms which were not presented due to space issues. The goal of this appendix is to present the formulation of each term. Actually, only a subset of these terms must be defined, in fact some relationships can be found:

$$\begin{aligned} a_{11} &= \int_0^{2\pi} [A_{int}(\varphi) - A_{ext}(\varphi)] \cos\varphi \, \mathrm{d}\varphi & a_{12} &= \int_0^{2\pi} l(\varphi) \cos\varphi \, \mathrm{d}\varphi \\ a_{13} &= \int_0^{2\pi} [A_{int}(\varphi) - A_{ext}(\varphi)] \sin\varphi \, \mathrm{d}\varphi & a_{14} &= \int_0^{2\pi} l(\varphi) \sin\varphi \, \mathrm{d}\varphi \\ a_{21} &= \int_0^{2\pi} [A_{int}(\varphi) - A_{ext}(\varphi)] \sin\varphi \, \mathrm{d}\varphi = a_{13} & a_{22} &= \int_0^{2\pi} l(\varphi) \sin\varphi \, \mathrm{d}\varphi = a_{14} \\ a_{23} &= \int_0^{2\pi} [A_{int}(\varphi) - A_{ext}(\varphi)] \cos\varphi \, \mathrm{d}\varphi = a_{11} & a_{24} &= \int_0^{2\pi} l(\varphi) \cos\varphi \, \mathrm{d}\varphi = a_{12} \\ a_{35} &= \int_0^{2\pi} [A_{int}(\varphi) - A_{ext}(\varphi)] \, \mathrm{d}\varphi & a_{36} &= \int_0^{2\pi} l(\varphi) \, \mathrm{d}\varphi \end{aligned} \tag{2.C.1}$$

This means that only six terms are independent and need to be computed. Each term was computed for the two cases $a \leq z_3$ and $a > z_3$.

2.C.1 The case of $a \leq z_3$

For the sake of clarity, the derivation was presented just for the first terms. For the area terms, the integral was split into two parts, one related to the internal profile and one to the external profile. These terms were referred with $a_{do,A}$ and $a_{do,B}$ respectively. Starting from a_{11}, substituting equations 2.B.1 and 2.B.3 in the integral and splitting it, it follows:

$$
\begin{aligned}
a_{11,A} = & \int_{\varphi_{2,int}}^{\varphi_{a,int}} \frac{\left[r_\varphi(\varphi) - r_2 + c\sin\varphi\right]^2 z_3 \cos\varphi}{2(r_3 - r_2)} \,\mathrm{d}\varphi + \int_{\varphi_{a,int}}^{\varphi_{a,ext}} \left\{ \frac{a\left[r_{a,int}(\varphi) - r_2 + c\sin\varphi\right]}{2} + \right. \\
& \left. + a\left[r_\varphi(\varphi) - r_{a,int}(\varphi)\right] \right\} \cos\varphi \,\mathrm{d}\varphi + \int_{\varphi_{a,ext}}^{\pi} \left\{ \frac{a\left[r_{a,int}(\varphi) - r_2 + c\sin\varphi\right]}{2} + a\left[r_{a,ext} - r_{a,int}(\varphi)\right] \right\} \cos\varphi \,\mathrm{d}\varphi = \\
= & \left\{ \frac{z_3 (b - r_{nom})^2}{2(r_3 - r_2)} \ln\frac{1+\sin\varphi}{\cos\varphi} + \frac{z_3 r_2^2 \sin\varphi}{2(r_3 - r_2)} + \frac{z_3 c^2 \sin^3\varphi}{6(r_3 - r_2)} - \frac{z_3 r_2 (b - r_{nom})\varphi}{r_3 - r_2} + \right. \\
& \left. - \frac{z_3 c (b - r_{nom})\cos\varphi}{r_3 - r_2} - \frac{z_3 r_2 c \sin^2\varphi}{2(r_3 - r_2)} \right\} \Bigg|_{\varphi_{2,int}}^{\varphi_{a,int}} + \left\{ \frac{a(r_{a,ext} - r_2)\sin\varphi}{2} + (b - r_{nom})\, a\varphi + \right. \\
& \left. - a\, r_{a,ext} \sin\varphi + \frac{c\, a \sin^2\varphi}{2} \right\} \Bigg|_{\varphi_{a,int}}^{\varphi_{a,ext}} + \left\{ \frac{a(r_{a,ext} - r_2)\sin\varphi}{2} + \frac{a\, c \sin^2\varphi}{2} \right\} \Bigg|_{\varphi_{a,ext}}^{\pi}
\end{aligned}
$$

$$
\begin{aligned}
a_{11,B} = & \int_{\varphi_{2,ext}}^{\varphi_{a,ext}} \frac{\left[r_\varphi(\varphi) - r_2\right]^2 z_3 \cos\varphi}{2(r_3 - r_2)} \,\mathrm{d}\varphi + \int_{\varphi_{a,ext}}^{\pi} \frac{r_{a,ext}\, a \cos\varphi}{2} \,\mathrm{d}\varphi = \\
= & \left\{ \frac{z_3 (b - r_{nom})^2}{2(r_3 - r_2)} \ln\frac{1+\sin\varphi}{\cos\varphi} + \frac{r_2^2 z_3 \sin\varphi}{2(r_3 - r_2)} - \frac{r_2 z_3 (b - r_{nom})\varphi}{r_3 - r_2} \right\} \Bigg|_{\varphi_{2,ext}}^{\varphi_{a,ext}} + \\
& + \left\{ \frac{(r_{a,ext} - r_2)\, a \sin\varphi}{2} \right\} \Bigg|_{\varphi_{a,ext}}^{\pi}
\end{aligned}
$$

$$
a_{11} = a_{23} = a_{11,A} - a_{11,B} \tag{2.C.2}
$$

The second term is instead solved in the following:

$$
\begin{aligned}
a_{12} = & \int_{\varphi_{2,int}}^{\varphi_{2,ext}} \left[r_\varphi(\varphi) - r_2 + c\sin\varphi\right] \cos\varphi \,\mathrm{d}\varphi + \int_{\varphi_{2,ext}}^{\varphi_{a,ext}} \left\{ c\sin\varphi + \left[r_\varphi(\varphi) - r_2\right] \sqrt{1 + \left(\frac{z_3}{r_3 - r_2}\right)^2} \right\} \cos\varphi \,\mathrm{d}\varphi + \\
& + \int_{\varphi_{a,ext}}^{\pi} \left[c\sin\varphi + \sqrt{(r_{a,ext} - r_2)^2 + a^2} \right] \cos\varphi \,\mathrm{d}\varphi = \\
= & \left\{ (b - r_{nom})\varphi - r_2 \sin\varphi + \frac{c\sin^2\varphi}{2} \right\} \Bigg|_{\varphi_{2,int}}^{\varphi_{2,ext}} + \left\{ \frac{c\sin^2\varphi}{2} + (b - r_{nom}) \sqrt{1 + \left(\frac{z_3}{r_3 - r_2}\right)^2} \varphi + \right. \\
& \left. - r_2 \sqrt{1 + \left(\frac{z_3}{r_3 - r_2}\right)^2} \sin\varphi \right\} \Bigg|_{\varphi_{2,ext}}^{\varphi_{a,ext}} + \left\{ \frac{c\sin^2\varphi}{2} + \sqrt{(r_{a,ext} - r_2)^2 + a^2} \sin\varphi \right\} \Bigg|_{\varphi_{a,ext}}^{\pi}
\end{aligned}
$$

$$
a_{12} = a_{24} \tag{2.C.3}
$$

The third term is expressed as follows:

$$
\begin{aligned}
a_{13,A} = &\left\{ \frac{z_3\,(b-r_{nom})^2}{2\,(r_3-r_2)\cos\varphi} - \frac{z_3\,r_2^2\cos\varphi}{2\,(r_3-r_2)} + \frac{c^2\,z_3\cos\varphi\left(\cos^2\varphi-3\right)}{6\,(r_3-r_2)} + \frac{z_3\,r_2\,(b-r_{nom})\ln\cos\varphi}{r_3-r_2} + \right. \\
&- \frac{z_3\,c\,r_2}{r_3-r_2}\left(\frac{\varphi}{2}-\frac{\sin 2\varphi}{4}\right) + \frac{z_3\,c\,(b-r_{nom})}{r_3-r_2}\left[\ln\tan\left(\frac{\varphi}{2}+\frac{\pi}{4}\right)-\sin\varphi\right] \Bigg\}\Bigg|_{\varphi_{2,int}}^{\varphi_{a,int}} + \left\{ -\frac{a\,(r_{a,ext}-r_2)\cos\varphi}{2} + \right. \\
&- a\,(b-r_{nom})\ln\cos\varphi + a\,r_{a,ext}\cos\varphi + a\,c\left(\frac{\varphi}{2}-\frac{\sin 2\varphi}{4}\right) \Bigg\}\Bigg|_{\varphi_{a,int}}^{\varphi_{a,ext}} + \left\{ -\frac{a\,(r_{a,ext}-r_2)\cos\varphi}{2} + \right. \\
&\left. + a\,c\left(\frac{\varphi}{2}-\frac{\sin 2\varphi}{4}\right) \right\}\Bigg|_{\varphi_{a,ext}}^{\pi}
\end{aligned}
$$

$$
a_{13,B} = \left\{ \frac{z_3\,(b-r_{nom})^2}{2\,(r_3-r_2)\cos\varphi} - \frac{r_2^2\,z_3\cos\varphi}{2\,(r_3-r_2)} + \frac{r_2\,z_3\,(b-r_{nom})\ln\cos\varphi}{r_3-r_2} \right\}\Bigg|_{\varphi_{2,ext}}^{\varphi_{a,ext}} + \left\{ -\frac{(r_{a,ext}-r_2)\,a\cos\varphi}{2} \right\}\Bigg|_{\varphi_{a,ext}}^{\pi}
$$

$$
a_{13} = a_{21} = a_{13,A} - a_{13,B}
\tag{2.C.4}
$$

The fourth term is then presented:

$$
\begin{aligned}
a_{14} = &\left\{ -(b-r_{nom})\ln\cos\varphi + r_2\cos\varphi + c\left(\frac{\varphi}{2}-\frac{\sin 2\varphi}{4}\right) \right\}\Bigg|_{\varphi_{2,int}}^{\varphi_{2,ext}} + \left\{ c\left(\frac{\varphi}{2}-\frac{\sin 2\varphi}{4}\right) + \right. \\
&- (b-r_{nom})\sqrt{1+\left(\frac{z_3}{r_3-r_2}\right)^2}\ln\cos\varphi + r_2\sqrt{1+\left(\frac{z_3}{r_3-r_2}\right)^2}\cos\varphi \Bigg\}\Bigg|_{\varphi_{2,ext}}^{\varphi_{a,ext}} + \left\{ c\left(\frac{\varphi}{2}-\frac{\sin 2\varphi}{4}\right) + \right. \\
&\left. - \sqrt{(r_{a,ext}-r_2)^2+a^2}\cos\varphi \right\}\Bigg|_{\varphi_{a,ext}}^{\pi}
\end{aligned}
\tag{2.C.5}
$$

$$
a_{14} = a_{22}
$$

The fifth term is reported in the following:

$$
\begin{aligned}
a_{35,A} = &\left\{ \frac{z_3}{2\,(r_3-r_2)} \left[(b-r_{nom})^2\tan\varphi + r_2^2\varphi + \frac{c^2\varphi}{2} - \frac{c^2\sin 2\varphi}{4} - 2\,r_2\,(b-r_{nom})\ln\left(\tan\varphi+\sec\varphi\right) + \right. \right. \\
&\left. \left. - 2\,c\,(b-r_{nom})\ln\cos\varphi + 2\,r_2\,c\cos\varphi \right] \right\}\Bigg|_{\varphi_{2,int}}^{\varphi_{a,int}} + \left\{ \frac{a\,(r_{a,ext}-r_2)\,\varphi}{2} + a\,(b-r_{nom})\ln\left(\tan\varphi+\sec\varphi\right) + \right. \\
&\left. - a\,r_{a,ext}\varphi - a\,c\cos\varphi \right\}\Bigg|_{\varphi_{a,int}}^{\varphi_{a,ext}} + \left\{ \frac{a\,(r_{a,ext}-r_2)\,\varphi}{2} - a\,c\cos\varphi \right\}\Bigg|_{\varphi_{a,ext}}^{\pi}
\end{aligned}
\tag{2.C.6}
$$

$$
\begin{aligned}
a_{35,B} = &\left\{ \frac{z_3}{2\,(r_3-r_2)} \left[(b-r_{nom})^2\tan\varphi + r_2^2\varphi - 2\,r_2\,(b-r_{rnom})\ln\left(\tan\varphi+\sec\varphi\right) \right] \right\}\Bigg|_{\varphi_{2,ext}}^{\varphi_{a,ext}} + \\
&+ \left\{ \frac{(r_{a,ext}-r_2)\,a\varphi}{2} \right\}\Bigg|_{\varphi_{a,ext}}^{\pi}
\end{aligned}
$$

$$
a_{35} = a_{35,A} - a_{35,B}
$$

The last term is expressed as in the following:

$$
\begin{aligned}
a_{36} = & \left\{ (b-r_{nom}) \ln(\tan\varphi + \sec\varphi) - r_2\varphi - c\cos\varphi \right\}\Big|_{\varphi_{2,int}}^{\varphi_{2,ext}} + \left\{ -c\cos\varphi + \right. \\
& + \sqrt{1+\left(\frac{z_3}{r_3-r_2}\right)^2} \left[(b-r_{nom}) \ln(\tan\varphi + \sec\varphi) - r_2\varphi \right] \Big\}\Big|_{\varphi_{2,ext}}^{\varphi_{a,ext}} + \left\{ -c\cos\varphi + \right. \\
& \left. + \sqrt{(r_{a,ext}-r_2)^2 + a^2}\,\varphi \right\}\Big|_{\varphi_{a,ext}}^{\pi}
\end{aligned}
\tag{2.C.7}
$$

2.C.2 The case of $a > z_3$

All the computations performed for the $a \leq z_3$ case were carried out using the definitions of $A_{int}(\varphi)$, $A_{ext}(\varphi)$ and $l(\varphi)$ in the $a > z_3$ case, producing different formulations for the ten terms of equation 2.13. Starting from a_{11}, substituting equations 2.B.2 and 2.B.4 in the integral and splitting it, it follows:

$$
\begin{aligned}
a_{11,A} = & \left\{ \frac{z_3(b-r_{nom})^2}{2(r_3-r_2)} \ln\frac{1+\sin\varphi}{\cos\varphi} + \frac{r_2^2 z_3 \sin\varphi}{2(r_3-r_2)} + \frac{z_3 c^2 \sin^3\varphi}{6(r_3-r_2)} - \frac{r_2(b-r_{nom})\, z_3\varphi}{r_3-r_2} + \right. \\
& \left. - \frac{z_3 c (b-r_{nom})\cos\varphi}{r_3-r_2} - \frac{z_3 r_2 c \sin^2\varphi}{2(r_3-r_2)} \right\}\Big|_{\varphi_{2,int}}^{\varphi_{3,int}} + \left\{ \frac{(r_3-r_2) z_3 \sin\varphi}{2} + z_3(b-r_{nom})\varphi + \right. \\
& - z_3 r_3 \sin\varphi + \frac{c z_3 \sin^2\varphi}{2} + \frac{(z_4-z_3)(b-r_{nom})^2}{2(r_4-r_3)} \ln\frac{1+\sin\varphi}{\cos\varphi} + \frac{r_3^2(z_4-z_3)\sin\varphi}{2(r_4-r_3)} + \\
& + \frac{(z_4-z_3) c^2 \sin^3\varphi}{6(r_4-r_3)} - \frac{(z_4-z_3)(b-r_{nom}) r_3\varphi}{r_4-r_3} - \frac{r_3 c (z_4-z_3)\sin^2\varphi}{2(r_4-r_3)} + \\
& \left. - \frac{c(b-r_{nom})(z_4-z_3)\cos\varphi}{r_4-r_3} \right\}\Big|_{\varphi_{3,int}}^{\varphi_{a,int}} + \left\{ \frac{(r_3-r_2) z_3 \sin\varphi}{2} + \frac{(z_3+a)(r_{a,ext}-r_3)\sin\varphi}{2} + \right. \\
& + a(b-r_{nom})\varphi - a\, r_{a,ext}\sin\varphi + \frac{a c \sin^2\varphi}{2} \Big\}\Big|_{\varphi_{a,int}}^{\varphi_{a,ext}} + \left\{ \frac{(r_3-r_2) z_3 \sin\varphi}{2} + \right. \\
& \left. + \frac{(z_3+a)(r_{a,ext}-r_3)\sin\varphi}{2} + \frac{a c \sin^2\varphi}{2} \right\}\Big|_{\varphi_{a,ext}}^{\pi}
\end{aligned}
\tag{2.C.8}
$$

$$
\begin{aligned}
a_{11,B} = & \left\{ \frac{z_3(b-r_{nom})^2}{2(r_3-r_2)} \ln\frac{1+\sin\varphi}{\cos\varphi} + \frac{z_3 r_2^2 \sin\varphi}{2(r_3-r_2)} - \frac{z_3(b-r_{nom}) r_2\varphi}{r_3-r_2} \right\}\Big|_{\varphi_{2,ext}}^{\varphi_{3,ext}} + \left\{ \frac{z_3(r_3-r_2)\sin\varphi}{2} + \right. \\
& + z_3(b-r_{nom})\varphi - z_3 r_3 \sin\varphi + \frac{(z_4-z_3)(b-r_{nom})^2}{2(r_4-r_3)} \ln\frac{1+\sin\varphi}{\cos\varphi} + \frac{(z_4-z_3) r_3^2 \sin\varphi}{2(r_4-r_3)} + \\
& \left. - \frac{(z_4-z_3)(b-r_{nom}) r_3\varphi}{r_4-r_3} \right\}\Big|_{\varphi_{3,ext}}^{\varphi_{a,ext}} + \left\{ \left[\frac{(r_3-r_2) z_3}{2} + \frac{(z_3+a)(r_{a,ext}-r_3)}{2} \right] \sin\varphi \right\}\Big|_{\varphi_{a,ext}}^{\pi}
\end{aligned}
$$

$$
a_{11} = a_{23} = a_{11,A} - a_{11,B}
$$

The term a_{12} is then expressed as follows:

$$\begin{aligned}
a_{12} = &\left\{ (b-r_{nom})\,\varphi - r_2 \sin\varphi + \frac{c\sin^2\varphi}{2} \right\}\Bigg|_{\varphi_{2,int}}^{\varphi_{2,ext}} + \left\{ \frac{c\sin^2\varphi}{2} + (b-r_{nom})\sqrt{1+\left(\frac{z_3}{r_3-r_2}\right)^2}\,\varphi + \right. \\
&\left. - r_2\sqrt{1+\left(\frac{z_3}{r_3-r_2}\right)^2}\sin\varphi \right\}\Bigg|_{\varphi_{2,ext}}^{\varphi_{3,ext}} + \left\{ \frac{c\sin^2\varphi}{2} + \sqrt{z_3^2+(r_3-r_2)^2}\sin\varphi + \right. \\
&\left. + (b-r_{nom})\sqrt{1+\left(\frac{z_4-z_3}{r_4-r_3}\right)^2}\,\varphi - r_3\sqrt{1+\left(\frac{z_4-z_3}{r_4-r_3}\right)^2}\sin\varphi \right\}\Bigg|_{\varphi_{3,ext}}^{\varphi_{a,ext}} + \left\{ \frac{c\sin^2\varphi}{2} + \right. \\
&\left. + \sqrt{z_3^2+(r_3-r_2)^2}\sin\varphi + \sqrt{(r_{a,ext}-r_3)^2+(a-z_3)^2}\sin\varphi \right\}\Bigg|_{\varphi_{a,ext}}^{\pi}
\end{aligned} \tag{2.C.9}$$

$$a_{12} = a_{24}$$

The third term is presented in the following:

$$\begin{aligned}
a_{13,A} = &\left\{ \frac{z_3\,(b-r_{nom})^2}{2\,(r_3-r_2)\cos\varphi} - \frac{z_3\,r_2^2\cos\varphi}{2\,(r_3-r_2)} + \frac{z_3\,c^2\cos\varphi\left(\cos^2\varphi-3\right)}{6\,(r_3-r_2)} + \frac{z_3\,r_2\,(b-r_{nom})\ln\cos\varphi}{r_3-r_2} + \right. \\
&\left. - \frac{z_3\,r_2\,c}{r_3-r_2}\left(\frac{\varphi}{2}-\frac{\sin 2\varphi}{4}\right) + \frac{z_3\,(b-r_{nom})\,c}{r_3-r_2}\left[\ln\tan\left(\frac{\varphi}{2}+\frac{\pi}{4}\right)-\sin\varphi\right] \right\}\Bigg|_{\varphi_{2,int}}^{\varphi_{3,int}} + \left\{ -\frac{(r_3-r_2)\,z_3\cos\varphi}{2} + \right. \\
&\left. - z_3\,(b-r_{nom})\ln\cos\varphi + z_3\,r_3\cos\varphi + c\,z_3\left(\frac{\varphi}{2}-\frac{\sin 2\varphi}{4}\right) + \frac{(z_4-z_3)\,(b-r_{nom})^2}{2\,(r_4-r_3)\cos\varphi} + \right. \\
&\left. - \frac{(z_4-z_3)\,r_3^2\cos\varphi}{2\,(r_4-r_3)} + \frac{(z_4-z_3)\,c^2\cos\varphi\left(\cos^2\varphi-3\right)}{6\,(r_4-r_3)} + \frac{(z_4-z_3)\,(b-r_{nom})\,r_3\ln\cos\varphi}{r_4-r_3} + \right. \\
&\left. + \frac{(z_4-z_3)\,(b-r_{nom})\,c}{r_4-r_3}\left[\ln\tan\left(\frac{\varphi}{2}+\frac{\pi}{4}\right)-\sin\varphi\right] - \frac{r_3\,c\,(z_4-z_3)}{r_4-r_3}\left(\frac{\varphi}{2}-\frac{\sin 2\varphi}{4}\right) \right\}\Bigg|_{\varphi_{3,int}}^{\varphi_{a,int}} + \\
&+ \left\{ -\frac{(r_3-r_2)\,z_3\cos\varphi}{2} - \frac{(z_3+a)\,(r_{a,ext}-r_3)\cos\varphi}{2} - a\,(b-r_{nom})\ln\cos\varphi + \right. \\
&\left. + a\,r_{a,ext}\cos\varphi + a\,c\left(\frac{\varphi}{2}-\frac{\sin 2\varphi}{4}\right) \right\}\Bigg|_{\varphi_{a,int}}^{\varphi_{a,ext}} + \left\{ -\frac{(r_3-r_2)\,z_3\cos\varphi}{2} + \right. \\
&\left. - \frac{(z_3+a)\,(r_{a,ext}-r_3)\cos\varphi}{2} + a\,c\left(\frac{\varphi}{2}-\frac{\sin 2\varphi}{4}\right) \right\}\Bigg|_{\varphi_{a,ext}}^{\pi}
\end{aligned}$$

$$\begin{aligned}
a_{13,B} = &\left\{ \frac{z_3\,(b-r_{nom})^2}{2\,(r_3-r_2)\cos\varphi} - \frac{z_3\,r_2^2\cos\varphi}{2\,(r_3-r_2)} + \frac{z_3\,(b-r_{nom})\,r_2\ln\cos\varphi}{r_3-r_2} \right\}\Bigg|_{\varphi_{2,ext}}^{\varphi_{3,ext}} + \left\{ -\frac{z_3\,(r_3-r_2)\cos\varphi}{2} + \right. \\
&\left. - z_3\,(b-r_{nom})\ln\cos\varphi + z_3\,r_3\cos\varphi + \frac{(z_4-z_3)\,(b-r_{nom})^2}{2\,(r_4-r_3)\cos\varphi} - \frac{(z_4-z_3)\,r_3^2\cos\varphi}{2\,(r_4-r_3)} + \right. \\
&\left. + \frac{(z_4-z_3)\,(b-r_{nom})\,r_3\ln\cos\varphi}{r_4-r_3} \right\}\Bigg|_{\varphi_{3,ext}}^{\varphi_{a,ext}} + \left\{ -\left[\frac{(r_3-r_2)\,z_3}{2} + \frac{(z_3+a)\,(r_{a,ext}-r_3)}{2}\right]\cos\varphi \right\}\Bigg|_{\varphi_{a,ext}}^{\pi}
\end{aligned}$$

$$a_{13} = a_{21} = a_{13,A} - a_{13,B} \tag{2.C.10}$$

The fourth term is then shown:

$$
\begin{aligned}
a_{14} = & \left\{ -(b-r_{nom})\ln\cos\varphi + r_2\cos\varphi + c\left(\frac{\varphi}{2} - \frac{\sin 2\varphi}{4}\right)\right\}\Bigg|_{\varphi_{2,int}}^{\varphi_{2,ext}} + \left\{ c\left(\frac{\varphi}{2} - \frac{\sin 2\varphi}{4}\right) + \right. \\
& - (b-r_{nom})\sqrt{1+\left(\frac{z_3}{r_3-r_2}\right)^2}\ln\cos\varphi + r_2\sqrt{1+\left(\frac{z_3}{r_3-r_2}\right)^2}\cos\varphi\Bigg\}\Bigg|_{\varphi_{2,ext}}^{\varphi_{3,ext}} + \left\{ c\left(\frac{\varphi}{2} - \frac{\sin 2\varphi}{4}\right) + \right. \\
& - \sqrt{z_3^2+(r_3-r_2)^2}\cos\varphi - (b-r_{nom})\sqrt{1+\left(\frac{z_4-z_3}{r_4-r_3}\right)^2}\ln\cos\varphi + r_3\sqrt{1+\left(\frac{z_4-z_3}{r_4-r_3}\right)^2}\cos\varphi\Bigg\}\Bigg|_{\varphi_{3,ext}}^{\varphi_{a,ext}} + \\
& + \left\{ c\left(\frac{\varphi}{2} - \frac{\sin 2\varphi}{4}\right) - \sqrt{z_3^2+(r_3-r_2)^2}\cos\varphi - \sqrt{(r_{a,ext}-r_3)^2+(a-z_3)^2}\cos\varphi\right\}\Bigg|_{\varphi_{a,ext}}^{\pi}
\end{aligned}
$$

$$
a_{14} = a_{22}
\tag{2.C.11}
$$

The fifth term is reported in the following:

$$
\begin{aligned}
a_{35,A} = & \left\{ \frac{z_3}{2(r_3-r_2)}\left[(b-r_{nom})^2\tan\varphi + r_2^2\varphi + \frac{c^2\varphi}{2} - \frac{c^2\sin 2\varphi}{4} - 2r_2(b-r_{nom})\ln(\tan\varphi+\sec\varphi) + \right.\right. \\
& \left.\left. - 2(b-r_{nom})c\ln\cos\varphi + 2r_2 c\cos\varphi\right]\right\}\Bigg|_{\varphi_{2,int}}^{\varphi_{3,int}} + \left\{ \frac{z_3(r_3-r_2)\varphi}{2} + (b-r_{nom})z_3\ln(\tan\varphi+\sec\varphi) + \right. \\
& - r_3 z_3\varphi - c z_3\cos\varphi + \frac{z_4-z_3}{2(r_4-r_3)}\left[(b-r_{nom})^2\tan\varphi + r_3^2\varphi + \frac{c^2}{2} - \frac{c^2\sin 2\varphi}{4} + \right. \\
& \left.\left. - 2(b-r_{nom})r_3\ln(\tan\varphi+\sec\varphi) + 2r_3 c\cos\varphi - 2(b-r_{nom})c\ln\cos\varphi\right]\right\}\Bigg|_{\varphi_{3,int}}^{\varphi_{a,int}} + \\
& + \left\{ \frac{(r_3-r_2)z_3\varphi}{2} + \frac{(z_3+a)(r_{a,ext}-r_3)\varphi}{2} + a(b-r_{nom})\ln(\tan\varphi+\sec\varphi) - a r_{a,ext}\varphi + \right. \\
& \left. - a c\cos\varphi\right\}\Bigg|_{\varphi_{a,int}}^{\varphi_{a,ext}} + \left\{ \frac{(r_3-r_2)z_3\varphi}{2} + \frac{(z_3+a)(r_{a,ext}-r_3)\varphi}{2} - a c\cos\varphi\right\}\Bigg|_{\varphi_{a,ext}}^{\pi}
\end{aligned}
$$

$$
\begin{aligned}
a_{35,B} = & \left\{ \frac{z_3}{2(r_3-r_2)}\left[(b-r_{nom})^2\tan\varphi + r_2^2\varphi - 2r_2(b-r_{nom})\ln(\tan\varphi+\sec\varphi)\right]\right\}\Bigg|_{\varphi_{2,ext}}^{\varphi_{3,ext}} + \\
& + \left\{ \frac{z_3(r_3-r_2)\varphi}{2} + z_3(b-r_{nom})\ln(\tan\varphi+\sec\varphi) - r_3 z_3\varphi + \frac{z_4-z3}{2(r_4-r_3)}\left[(b-r_{nom})^2\tan\varphi + \right.\right. \\
& \left.\left. + r_3^2\varphi - 2(b-r_{nom})r_3\ln(\tan\varphi+\sec\varphi)\right]\right\}\Bigg|_{\varphi_{3,ext}}^{\varphi_{a,ext}} + \left\{ \frac{(r_3-r_2)z_3\varphi}{2} + \frac{(z_3+a)(r_{a,ext}-r_3)\varphi}{2}\right\}\Bigg|_{\varphi_{a,ext}}^{\pi}
\end{aligned}
$$

$$
a_{35} = a_{35,A} - a_{35,B}
\tag{2.C.12}
$$

The last term is instead equal to:

$$
\begin{aligned}
a_{36} = & \left\{ (b-r_{nom})\ln(\tan\varphi+\sec\varphi) - r_2\varphi - c\cos\varphi\right\}\Bigg|_{\varphi_{2,int}}^{\varphi_{2,ext}} + \left\{ -c\cos\varphi + \right. \\
& + \sqrt{1+\left(\frac{z_3}{r_3-r_2}\right)^2}\left[(b-r_{nom})\ln(\tan\varphi+\sec\varphi) - r_2\varphi\right]\Bigg\}\Bigg|_{\varphi_{2,ext}}^{\varphi_{3,ext}} + \left\{ -c\cos\varphi + \sqrt{z_3^2+(r_3-r_2)^2}\varphi + \right. \\
& + \sqrt{1+\left(\frac{z_4-z_3}{r_4-r_3}\right)^2}\left[(b-r_{nom})\ln(\tan\varphi+\sec\varphi) - r_3\right]\Bigg\}\Bigg|_{\varphi_{3,ext}}^{\varphi_{a,ext}} + \left\{ -c\cos\varphi + \sqrt{z_3^2+(r_3-r_2)^2}\varphi + \right. \\
& \left. + \sqrt{(a-z_3)^2+(r_{a,ext}-r_3)^2}\varphi\right\}\Bigg|_{\varphi_{a,ext}}^{\pi}
\end{aligned}
$$

(2.C.13)

CHAPTER 3

Tool wear: real-time cutting tool monitoring

In the context of prognostics and health management, this chapter describes an advanced real-time monitoring strategy for cutting tools. Based on the model developed in the previous chapter, a strategy for the real-time estimation of specific force coefficients is conceived. Their evolution in time allows for a phenomenological detection of out-of-control cutting conditions, related to tool wear. Identified specific force coefficients may also be integrated as features for cutting tool prognosis.

3.1 Introduction

The complexity of manufacturing systems has been increasing in order to accomplish the high reliability and safety that industry demands [162]. Furthermore, plants require high-quality machining and high-quality tools for automation. The development of condition monitoring solutions for machine tool components and processes became means to reach these goals [2]. Thus, research has been focusing on the study of cutting tool condition monitoring (TCM) and prognostics approaches [14]. This was related to the fact that the cutting tool provides large cost saving possibilities; up to 40% could be achieved through the monitoring of its health [185]. Furthermore, about 20% of the downtimes of machine tools was attributed to tool failures, resulting in reduced productivity and economic losses [121, 227]. TCM strategies represent also transition means towards sustainable manufacturing. Tool wear assessment and prediction can be an important resource to feed process parameter optimisation algorithms [12, 119, 120, 229]. In fact, a proper optimisation of cutting parameters during the cut can extend their remaining useful life, reducing the amount of waste materials [233, 236].

Basically, two strategies can be adopted for TCM, either based on direct measurements or indirect ones: the first one, where the assessment of the tool is carried out

Table 3.1: *Commercial TCM tools by sensor type [102].*

Physical quantity (sensor type)	Suppliers						
	Artis	Brankamp	Kistler	Montronix	Nordmann	Prometec	TMAC
Power	X	X		X	X	X	X
Torque	X		X	X			
Strain		X	X	X		X	X
Distance/displacement					X	X	
1-3 axis force sensor	X	X	X	X		X	
Measuring plate			X	X		X	
Acoustic emission	X	X	X	X	X	X	
AE fluid sensor					X	X	
Rotating AE sensor	X				X	X	
Vibration and ultrasound			X	X		X	X
Camera	X	X					
Laser	X				X		

by measurements of a wear characteristics (e.g. flank wear), whereas the second one, where quantities of potential interest are measured from the machine tool, such as cutting forces, spindle torque, etc. [121]. Despite the first methods provide accurate assessments of the tool condition (based on vision systems or proximity measurements, [240]), they are limited in application since it is necessary to quit the machining operations and in most of the cases the measurements are time consuming [121, 125]. Thus, in the recent years, the main contributions have regarded indirect TCM methods, due to their in-process applicability. Most of the indirect TCM applications exploit cutting forces (and related quantities, [228]), vibration measurements [34] or acoustic emission signals [208, 211, 236] in order to track the condition of the tool [119]. Many different machining applications have been investigated: drilling [105], turning [34, 125] and milling [109, 148, 185, 227]. Commercial tools are available for TCM, exploiting indirect measurements. Table 3.1 shows some important commercial TCM solutions and their mapping with respect to the used sensors [102]. The most common solutions are those relying upon cutting forces related quantities, acoustic emissions and vibrations. In table 3.2, commercial solutions (Artis, Brankamp, Kistler, Montronix, Nordmann, Prometec and TMAC[1]) are mapped upon the monitoring strategies [102]. Six strategies are available for commercial tools. Simple fixed limits are constant thresholds applied to the monitored variable; when overcome they trigger an alarm. Time defined limits are similar to fixed limits, but they assume multiple values depending on specific regions of a cutting operation. Part signature imposes thresholds which are functions of time, representative of specific components cutting. Pattern recognition stores particular signals shapes, when tool breakage has occurred; it continuously compares current signal profile with stored breakage profiles. Wear estimator estimates the flank wear in turning based on the relation of the three cutting forces components. Dynamic limits apply local thresholds following current signal trends, acting as a fixed thresholding system when abrupt signal changes occur (tool breakage). Cutting forces are typically measured through dynamometric plates, which are not affordable for an industrial implementation [14]. To avoid the use of dynamometers, indirect methods for cutting forces estimation [2] and novel integrated sensors [225] were developed

[1] Caron Engineering

Table 3.2: *Commercial TCM tools by monitoring strategy [102].*

Strategy	Suppliers						
	Artis	Brankamp	Kistler	Montronix	Nordmann	Prometec	TMAC
Simple fixed limits				X	X	X	X
Time defined limits			X		X	X	
Part signature	X	X	X	X	X		
Pattern recognition				X			
Wear estimator				X			
Dynamic limits						X	

and proposed. In order to extract wear related information, several feature extraction strategies can be applied to the measurements from time (e.g. mean, skewness and root mean square, [132]), frequency (e.g. maximum amplitude and average amplitude) or time-frequency domains (e.g. energy of each decomposed wavelet) [148, 227]. Such features are then fed into anomaly detection algorithms (monitoring), classification algorithms (diagnostics) or regression models (prognostics) [20]. Nevertheless, the above commercial TCM methods and traditional tool wear features are strongly influenced by the machining parameters. Thus, they can be applied only under fixed milling parameters and repetitive cutting operations [159]. In recent studies, most of the approaches makes use of machine learning [185] or deep learning algorithms, such as Convolutional Neural Networks [227]. Since a large set of high-quality data taken in a wide range of cutting conditions is requested, their application to real industrial scenarios is challenging [73, 135]. Scientific works were mapped in table 3.3, according to several classes, highlighting some useful aspects of the proposed methodologies like the algorithm domain.

Consequently, the necessity to develop solutions based on process modelling, either physical or mechanistic is evident. From this perspective, different models with various grades of complexity have been used in literature to predict the cutting forces for a given machining operation. Physical-based approaches are based on slip-line field analysis to model the chip formation mechanism (e.g. [53]); an extended review of state-of-art approaches of this kind can be found in [11]. Conversely, mechanistic models exploit simplified description of the process, but model coefficients (specific force coefficients, SFC) need to be experimentally identified [8]. The main feature of the SFC is that they are theoretically independent from some process parameters. At the same time, by representing the material shearing, the friction and ploughing effects involved in the cutting process, they carry the information about tool wear, becoming a relevant resource for TCM. Anyway, their identification is still based on multivariate linear regression on mean cutting forces measured during variable feed milling tests [8]. The Altintas model was widely studied and even upgraded with some additional features, such as improved chip thickness formulation including cutter run-out and cycloidal path [118, 144] or the inclusion of tool wear modelling in cutting forces expressions [140, 241], while several researches analysed SFC dependence with respect to the cutting parameters in multiple applications: square shoulder mills [28, 156] and ball-end milling [74]. Nouri *et al.* estimated such coefficients (exploiting mean forces) online during run-to-failure experiments. This was possible since cutting was executed continuously varying the feed per

Table 3.3: *Map of State-of-The-Art TCM approaches. Algorithms are described by application, employed sensors, measured features, algorithm and relative category, number of experimental run-to-failures.*

Ref.	Machining process	Measurements	Features	Algorithm	Class	Varying cutting parameters	Run-to-failures
[125]	Turning	Power consumption, vibrations	Power mean, vibration moving average (separately)	Thresholding	Grey-box	No	30 (some of them discarded)
[206]	Milling	Cutting forces	Energy from time-frequency domain features	Continuous Hidden Markov Models for state classification	Grey-box	No	7
[220]	Milling	Cutting forces, vibrations, acoustic emissions	rms (accelerometers)	Weighted Hidden-Markov Models for wear accumulation stage classification	Grey-box	No	3
[34]	Turning	Vibrations, cutting forces, surface pictures	Time, frequency, time-frequency domain features (signals); GLCM (pictures)	SVM	Black-box (for VB assessment)	Yes	9
[73]	Milling	Flank wear, cutting forces, vibrations, acoustic emission	Learned features (deep learning)	DenseNets	Black-box	No	2
[227]	milling	Machine tool current, vibrations and acoustic emissions	Time and frequency domains	Convolutional Neural Network	Black-box	Yes	15
[207]	Milling	Spindle power, milled surface images, tool wear images	Learned features (deep learning)	Convolutional Neural Network	Black-box	Yes	6300 images
[140]	Milling	Cutting forces and temperature	-	Tool wear model estimation from cutting forces and temperature	Model-based	Yes	9

Ref.	Machining process	Measurements	Features	Algorithm	Class	Varying cutting parameters	Run-to-failures
[240]	Micro-milling	Full tool wear images	Image decomposition through Morphological Component Analysis	Area estimation	-	Yes	-
[228]	Milling	Cutting forces, vibrations, spindle current	Specific force coefficients (fusion)	Upper-bound threshold	Model-based & Grey-box	Yes	13
[236]	Milling	Acoustic emission	Mean, std, minima of estimated Holder Exponents and quantities of singular points	Support Vector Machine	Black-box	Yes	18
[208]	Milling	Sound, acceleration and cutting bending moment	Upgraded principal component analysis features and energy proportion	Support Vector Machine	Black-box	Yes	18
[211]	Turning	Acoustic emission	Frequency domain features via spectrograms	Dirichlet process Gaussian mixture models	Grey-box	No	10
[105]	Drilling	Cutting forces	Mean, high-frequency components std, wavelet decomposition	-	-	No	3
[148]	Milling	Cutting force, vibration and acoustic emission	Time, frequency and time-frequency domain features	Gravitational search algorithm - back propagation Neural Network	Black-box	No	1
[109]	Milling	Cutting parameters and force signals	Functional data analysis	Linear regression	Grey-box	Yes	20
[185]	Milling	Vibrations, spindle current	Mean signal energy (vibrations)	Perceptron (wear level classification)	Black-box	Yes	3
[132]	Micro-milling	Flank wear, cutting forces	Flank wear area; rms, variance, skewness, kurtosis (forces)	Polynomial fitting	Grey-box	Yes	9

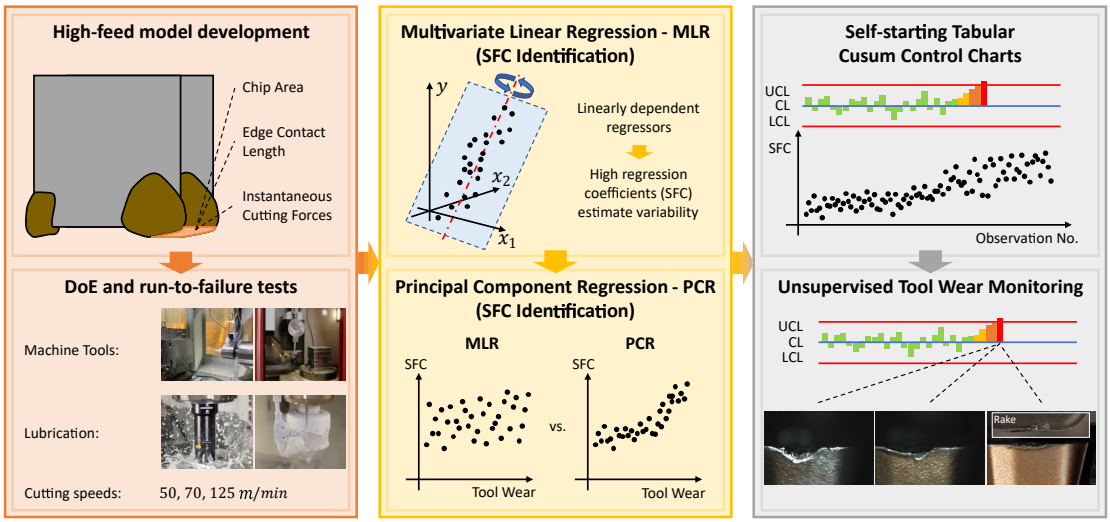


Figure 3.1: *Overall framework of the work. Left branch represents the high-feed cutters used and the experimental campaign parameters. Middle branch shows the comparison between the methods used for Specific Force Coefficients identification. Right branch shows the application of control charts for the unsupervised tool wear monitoring of critical tool chipping and notch wear.*

tooth, that cannot be guaranteed in real production scenarios [156]. To overcome this limitation, new approaches based on the identification of SFC from instantaneous cutting forces were developed and tested [54]. Nevertheless, up to the authors' knowledge, no-one still demonstrated the efficacy of such approaches for the real-time identification of SFC from instantaneous cutting forces in run-to-failure experiments and considering multiple cutting conditions.

The proposed work is intended to verify the applicability of SFC identification from instantaneous cutting forces in a TCM context. The chapter structure is organised as follows. In section 3.2, the whole methodology is presented. A background of the mechanistic model for square shoulder mills and an innovative mechanistic formulation for high-feed mills are reported in sections 3.2.1 and 3.2.2. The SFC identification through multivariate linear regression and principal component regression are discussed in sections 3.2.3 and 3.2.4, respectively. Self-starting tabular cusum control charts for the detection of out of control cutting processes are reported in section 3.2.5. In section 3.3 the experimental set-up and campaign are described. In section 3.4, the application of the conceived approach to the experimental data is presented and discussed with respect to the specific literature. In section 3.5, the main introduced novelties are highlighted and future works are outlined.

3.2 Methods

For sake of clarity, the approach adopted in this work is summarised in figure 3.1. Materials are summarised in the left branch of the figure, including a representation of the cutting tool and the experimental campaign (with two machine tools, conventional and cryogenic lubrication, and variable cutting speeds). The identification methods for the SFC are presented in the middle branch, while the goal of the proposed work is shown in the right branch of figure 3.1. In the following sections, all these points are

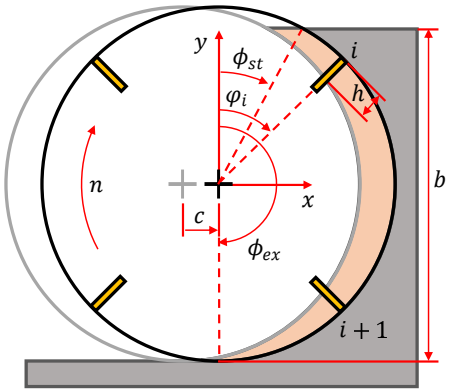


Figure 3.2: *Reference model for the geometry of milling operations.*

Table 3.4: *Cutter geometrical parameters.*

$r_{1c}\,[mm]$	$r_{2c}\,[mm]$	$r_{3c}\,[mm]$	$z_1\,[mm]$	$z_2\,[mm]$	$z_3\,[mm]$
6.57	8.53	9.03	0	0.40	0.62

thoroughly explained.

3.2.1 Theoretical background

The proposed solution is based on the identification of SFC during high-feed mill tests. Since a reliable mechanistic model for such mills is still not available, a specific formulation starting from [8] is developed and here presented. The mill cutter positions are described in time through equation (3.1):

$$\varphi_i(t) = \varphi_0 + \frac{2\pi i}{N} + 2\pi n t, \qquad i = 1,...,N \tag{3.1}$$

where φ_i is the angular position of the i-th cutter in the x-y reference plane (fig. 3.2); i is the cutter identifier; t is time; φ_0 is the initial phase of the first cutter; N is the number of teeth of the mill and n is its rotational speed. From now on, the time dependence is neglected. Altintas defines the instantaneous thickness h of the chip removed by a cutter as a function of the feed per tooth c, eq. (3.2):

$$h(\varphi) = c \sin\varphi \tag{3.2}$$

The expression for the tangential and radial cutting forces on a cutter, keeping in consideration the engagement conditions, are given by equation (3.3):

$$\begin{aligned} F_t &= g(\varphi)\,(K_{tc}\,a\,h + K_{te}\,a) \\ F_r &= g(\varphi)\,(K_{rc}\,a\,h + K_{re}\,a) \end{aligned} \tag{3.3}$$

where F_t and F_r are the tangential and radial forces acting on a single cutter; K_{tc}, K_{te}, K_{rc} and K_{re} are the SFC: the first and second are the tangential ones, the third and fourth are the edge ones; a is the axial depth of cut; $g(\varphi)$ is the engagement function defined by eq. (3.4):

$$g(\varphi) = \begin{cases} 1, & \text{if } \phi_{st} < \varphi < \phi_{ex} \\ 0, & \text{otherwise.} \end{cases} \tag{3.4}$$

and ϕ_{st}, ϕ_{ex} are the entry and exit angles determined by the engagement conditions of the mill (e.g. for downmilling, they are exressed by eq. (3.5)):

$$\begin{aligned} \phi_{st} &= \pi - \arccos\left(1 - \frac{2b}{D}\right) \\ \phi_{ex} &= \pi \end{aligned} \tag{3.5}$$

where b is the radial depth of cut; D is the diameter of the mill.

The single cutter contributions must then be projected on the x-y reference frame and summed up (eq. (3.6)):

$$\begin{aligned} F_x &= \sum_{i=1}^{N} \left(-F_{t,i}\cos\varphi_i - F_{r,i}\sin\varphi_i\right) \\ F_y &= \sum_{i=1}^{N} \left(+F_{t,i}\sin\varphi_i - F_{r,i}\cos\varphi_i\right) \end{aligned} \tag{3.6}$$

3.2.2 High-feed model

In this work the experimental campaign was conducted using high-feed mills, featured by a double-phased profile. Their shape is schematically shown in figure 3.3a. The proposed high-feed mechanistic model makes reference to figure 3.3a, from which the maximum allowable chip shape is derived. This shape corresponds to the region comprised between the current and previous tooth profiles, shifted by the chip thickness h. In order to describe the chip shape, it is necessary to define the cutter geometry. The geometry of the profile is uniquely defined in the radial-axial plane (r-z), figure 3.3b. The quantities r_{1c}, r_{2c}, r_{3c}, z_1, z_2 and z_3 are known for a given high-feed cutter, and they represent the radial and axial coordinates of the current (c) cutter fundamental points, defining the two phases of the cutting edge; χ_1 and χ_2 are the lead angles of the two cutter phases. The values associated to the cutters used for the experimentation are reported in table 3.4. The intersection of the axial depth of cut (a) line with the cutter geometry originates another important point, having coordinates $(r_{ac};a)$, given by eq. (3.7):

$$r_{ac} = \begin{cases} r_{1c} + \frac{a}{\tan\chi_1}, & \text{if } a \le z_2 \\ r_{2c} + \frac{a-z_2}{\tan\chi_2}, & \text{if } z_2 < a \le z_3 \end{cases} \tag{3.7}$$

r_{1p}, r_{2p}, r_{3p} and r_{ap} are the radial coordinates of the analogous points on the previous (p) cutter profile (i.e. related to the previous pass), and thus, they are determined through eq. (3.8):

$$r_{jp}(\varphi) = r_{jc} - h(\varphi), \qquad j = 1,2,3,a \tag{3.8}$$

where r_{jp} is a function of the position of the correspondent cutter angle φ; for the sake of conciseness this dependence will be omitted in the following. The radial engagement of the cutter is given by the radial depth of cut b, here expressed with reference to the cutter nominal radius R (figure 3.3c). Thus, each of the current and previous fundamental points is engaged when the cutter position reaches the critical corresponding angles

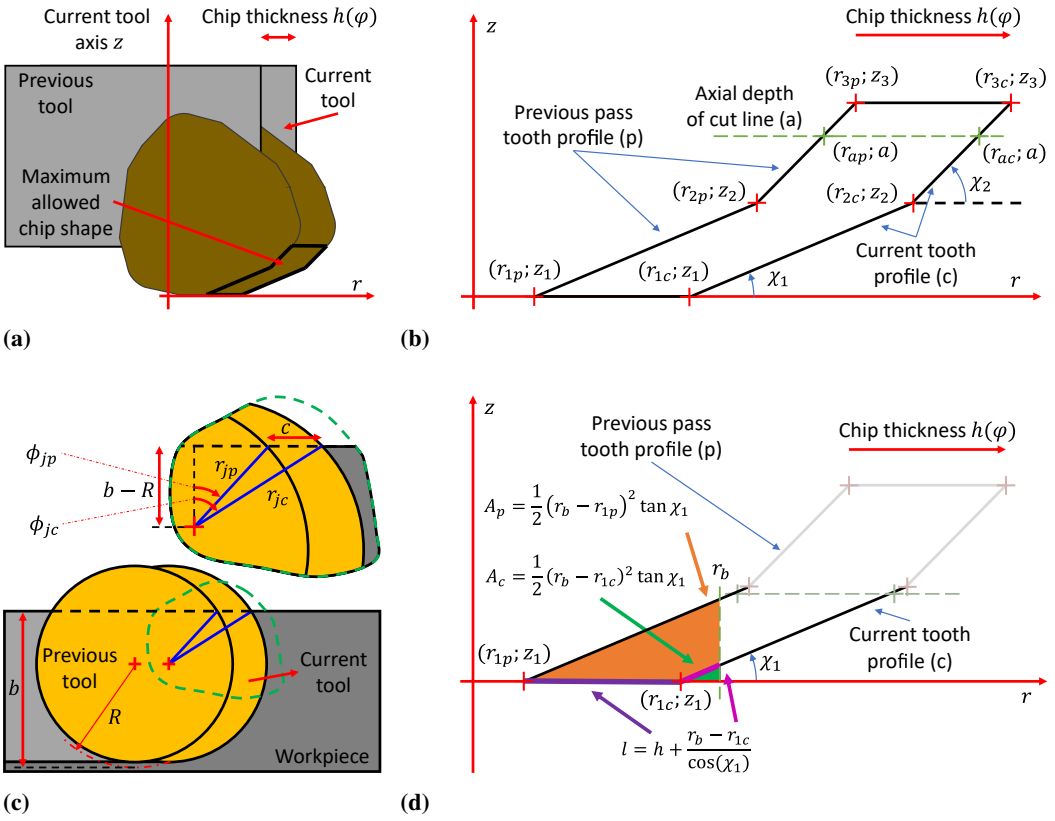


Figure 3.3: *Reference figure for the high-feed model: (a) derivation of the maximum allowable chip shape; (b) maximum allowable chip shape and high-feed cutter reference points; (c) derivation scheme for current and previous critical angles; (d) derivation scheme for current and previous chip area and chip edge length for an example case ($a <= z_2$, $\phi_{1p} \leq \varphi \leq \phi_{ap}$).*

ϕ_{jc} and ϕ_{jp} (eq. 3.9):

$$\begin{aligned} \phi_{jc} &= \arccos \frac{b-R}{r_{jc}} \\ \phi_{jp} &= \arctan \frac{r_{jc} \sin \phi_{jc} - c}{b-R} \end{aligned} \tag{3.9}$$

with $j = 1, 2, 3, a$. As a consequence, when a cutter position overcomes one of these angles, the chip shape changes. These relations are easily determined from the scheme reported in figure 3.3c through trigonometry.

It is important to define the actual radial engagement of a tooth $r_b(\varphi)$, given its current position (eq. 3.10):

$$r_b(\varphi) = \min\left(\frac{b-R}{\cos\varphi}, r_{ac}\right) \tag{3.10}$$

$r_b(\varphi)$ is initially limited by the workpiece (radial engagement) until it reaches the maximum engaged radial distance (determined by the axial depth of cut. The switch happens in correspondence of $\varphi = \phi_{ac}$. From here on, the φ dependences will be omitted. All the above-mentioned variables are needed to compute the actual chip area A and edge contact length l, taking in consideration even the progressive radial engagement of the

cutter. A and l are the results of the integration of the infinitesimal chip area dA and chip edge length dl (eq. (3.11)) over the radial coordinate from r_{1p} to r_b.

$$\begin{aligned} \mathrm{d}A(r) &= [z_p(r) - z_c(r)]\,\mathrm{d}r = \mathrm{d}A_p(r) - \mathrm{d}A_c(r) \\ \mathrm{d}l(r) &= \frac{\mathrm{d}r}{\cos\chi(r)} \end{aligned} \tag{3.11}$$

where $z_p(r)$ is the axial coordinate of the point on the previous cutter profile at a generic radial distance r from the mill axis (fig. 3.3d); $z_c(r)$ is the same as $z_p(r)$, but computed on the current cutter profile; dr is an infinitesimal variation in the radial direction; d$A_p(r)$ and d$A_c(r)$ are the infinitesimal areas below the previous and current tooth profile at radial coordinate r, respectively; $\chi(r)$ is the lead angle at the r radial coordinate. Thus, eq. (3.3) is modified in equation (3.12) and then, it can be introduced in eq. (3.6):

$$\begin{aligned} F_t &= K_{tc}\,(A_p - A_c) + K_{te}\,l \\ F_r &= K_{rc}\,(A_p - A_c) + K_{re}\,l \end{aligned} \tag{3.12}$$

where F_t, F_r, A_p, A_c and l dependencies on φ were omitted; A_p, A_c and l are the integrals of the above-mentioned differentials dA_p, dA_c and dl, respectively. A_p, A_c and l must be defined for the two cases $a \leq z_2$ and $z_2 < a \leq z_3$.

Case 1: $a \leq z_2$

The value of A_p is given by eq. (3.13):

$$A_p = \begin{cases} \frac{1}{2}\left(r_b - r_{1p}\right)^2 \tan\chi_1, & \text{if } \phi_{1p} < \varphi \leq \phi_{ap} \\ \frac{1}{2}\left(r_{ap} - r_{1p}\right) a + \left(r_b - r_{ap}\right) a, & \text{if } \phi_{ap} < \varphi \leq \phi_{ac} \\ \frac{1}{2}\left(r_{ap} - r_{1p}\right) a + \left(r_{ac} - r_{ap}\right) a, & \text{if } \phi_{ac} < \varphi \leq \pi \\ 0, & \text{otherwise.} \end{cases} \tag{3.13}$$

While A_c is given by eq. (3.14):

$$A_c = \begin{cases} \frac{1}{2}\left(r_b - r_{1c}\right)^2 \tan\chi_1, & \text{if } \phi_{1c} < \varphi \leq \phi_{ac} \\ \frac{1}{2}\left(r_{ac} - r_{1c}\right) a, & \text{if } \phi_{ac} < \varphi \leq \pi \\ 0, & \text{otherwise.} \end{cases} \tag{3.14}$$

l is given by eq. (3.15):

$$l = \begin{cases} r_b - r_{1p}, & \text{if } \phi_{1p} < \varphi \leq \phi_{1c} \\ h + \frac{r_b - r_{1c}}{\cos(\chi_1)}, & \text{if } \phi_{1c} < \varphi \leq \phi_{ac} \\ h + \frac{r_{ac} - r_{1c}}{\cos(\chi_1)}, & \text{if } \phi_{ac} < \varphi \leq \pi \\ 0, & \text{otherwise.} \end{cases} \tag{3.15}$$

The example scheme for retrieving the first case of each of these terms is shown in figure 3.3d. Other cases can be obtained by moving r_b progressively towards higher values.

Case 2: $z_2 < a \leq z_3$

The value of A_p is given by eq. (3.16):

$$A_p = \begin{cases} \frac{1}{2}\left(r_b - r_{1p}\right)^2 \tan\chi_1, & \text{if } \phi_{1p} < \varphi \leq \phi_{2p} \\ \frac{1}{2}\left(r_{2c} - r_{1c}\right) z_2 + \left(r_b - r_{2p}\right) z_2 + \frac{1}{2}\left(r_b - r_{2p}\right)^2 \tan\chi_2, & \text{if } \phi_{2p} < \varphi \leq \phi_{ap} \\ \frac{1}{2}\left(r_{2c} - r_{1c}\right) z_2 + \frac{1}{2}\left(r_{ap} - r_{2p}\right)\left(z_2 + a\right) + \left(r_b - r_{ap}\right) a, & \text{if } \phi_{ap} < \varphi \leq \phi_{ac} \\ \frac{1}{2}\left(r_{2c} - r_{1c}\right) z_2 + \frac{1}{2}\left(r_{ap} - r_{2p}\right)\left(z_2 + a\right) + \left(r_{ac} - r_{ap}\right) a, & \text{if } \phi_{ac} < \varphi \leq \pi \\ 0, & \text{otherwise.} \end{cases} \tag{3.16}$$

While A_c is given by eq. (3.17):

$$A_c = \begin{cases} \frac{1}{2}\left(r_b - r_{1c}\right)^2 \tan\chi_1, & \text{if } \phi_{1c} < \varphi \leq \phi_{2c} \\ \frac{1}{2}\left(r_{2c} - r_{1c}\right) z_2 + \left(r_b - r_{2c}\right) z_2 + \frac{1}{2}\left(r_b - r_{2c}\right)^2 \tan\chi_2, & \text{if } \phi_{2c} < \varphi \leq \phi_{ac} \\ \frac{1}{2}\left(r_{2c} - r_{1c}\right) z_2 + \frac{1}{2}\left(r_{ac} - r_{2c}\right)\left(z_2 + a\right), & \text{if } \phi_{ac} < \varphi \leq \pi \\ 0, & \text{otherwise.} \end{cases} \tag{3.17}$$

l is given by eq. (3.18):

$$l = \begin{cases} r_b - r_{1p}, & \text{if } \phi_{1p} < \varphi \leq \phi_{1c} \\ h + \frac{r_b - r_{1c}}{\cos(\chi_1)}, & \text{if } \phi_{1c} < \varphi \leq \phi_{2c} \\ h + \frac{r_{2c} - r_{1c}}{\cos(\chi_1)} + \frac{r_b - r_{2c}}{\cos(\chi_2)}, & \text{if } \phi_{2c} < \varphi \leq \phi_{ac} \\ h + \frac{r_{2c} - r_{1c}}{\cos(\chi_1)} + \frac{r_{ac} - r_{2c}}{\cos(\chi_2)}, & \text{if } \phi_{ac} < \varphi \leq \pi \\ 0, & \text{otherwise.} \end{cases} \tag{3.18}$$

It must be noted that the proposed model uses a circular approximation of the tooth trajectory. This choice was based on the fact that high-feed mills may reach up to ten times conventional milling *feed over mill radius ratios* [49], but they present a relatively low *non dimensional cutting parameter* [118], not comparable with micro milling ones. In case of really aggressive cutting processes, actual tool tip trajectory can be introduced following [118] and [54]. Nevertheless, critical angles in equation (3.9) must be adapted accordingly.

3.2.3 SFC identification

The identification of SFC is typically carried out by a multivariate linear regression based on average cutting forces in milling tests with variable feed per tooth. This procedure is well established in literature and references for performing such identification can be found in [8] as well as in [54].

In the proposed work, the fitting procedure was based on instantaneous cutting forces, following a procedure proposed by [54]. SFC estimation is performed every 3 tool revolutions (where, 3 was chosen to keep the SFC estimation reliable), and the estimation number is indicated with the o index. For each estimation, N_s measurement samples are used and index s will indicate the sample number for the o-th estimate.

Thus, for a given estimate number o, s will range from 1 to N_s. o ranges from estimation number 1 to the total number of estimates N_o. The model for the o-th SFC estimate is represented by eq. (3.19):

$$\mathbf{y}^{(o)} = \mathbf{X}^{(o)}\boldsymbol{\beta}^{(o)} + \boldsymbol{\varepsilon}^{(o)} \tag{3.19}$$

where $\mathbf{y}^{(o)}$ is the response vector; $\mathbf{X}^{(o)}$ is the design matrix; $\boldsymbol{\beta}^{(o)}$ is the vector of regression coefficients and $\boldsymbol{\varepsilon}^{(o)}$ is the vector of residuals [153] that considers the modelling errors. The elements of $\boldsymbol{\varepsilon}^{(o)}$ are assumed to be drawn from an uncorrelated random variable and $\boldsymbol{\varepsilon}^{(o)}$ is assumed to have null expected value and variance σ^2. The response vector $\mathbf{y}^{(o)}$ is a $2N_s$x1 vector containing instantaneous cutting forces as in eq. (3.20):

$$\mathbf{y}^{(o)} = \begin{Bmatrix} F_{x,1} & F_{x,2} & \dots & F_{x,N_s} & F_{y,1} & F_{y,2} & \dots & F_{y,N_s} \end{Bmatrix}^T \tag{3.20}$$

The column vector $\boldsymbol{\beta}^{(o)}$ of regression coefficients contains the SFC to be identified at the o-th estimation instant (eq. (3.21)):

$$\boldsymbol{\beta}^{(o)} = \begin{Bmatrix} K_{tc}^{(o)} & K_{te}^{(o)} & K_{rc}^{(o)} & K_{re}^{(o)} \end{Bmatrix}^T \tag{3.21}$$

The design matrix at o estimation number is derived by stacking column-wise the matrix in eq. (3.22) for each of the N_s instants, resulting in a $2N_s$x4 design matrix:

$$\mathbf{x}_s = \begin{bmatrix} x_{11s} & x_{12s} & x_{13s} & x_{14s} \\ x_{21s} & x_{22s} & x_{23s} & x_{24s} \end{bmatrix} \tag{3.22}$$

where the single elements are reported in eq. (3.23):

$$\begin{aligned} x_{11s} &= -\sum_{i=1}^{N} A(\varphi_{is}) \cos\varphi_{is} & x_{21s} &= \sum_{i=1}^{N} A(\varphi_{is}) \sin\varphi_{is} \\ x_{12s} &= -\sum_{i=1}^{N} l(\varphi_{is}) \cos\varphi_{is} & x_{22s} &= \sum_{i=1}^{N} l(\varphi_{is}) \sin\varphi_{is} \\ x_{13s} &= -\sum_{i=1}^{N} A(\varphi_{is}) \sin\varphi_{is} & x_{23s} &= -\sum_{i=1}^{N} A(\varphi_{is}) \cos\varphi_{is} \\ x_{14s} &= -\sum_{i=1}^{N} l(\varphi_{is}) \sin\varphi_{is} & x_{24s} &= -\sum_{i=1}^{N} l(\varphi_{is}) \cos\varphi_{is} \end{aligned} \tag{3.23}$$

where $A = A_p - A_c$ and l are the high-feed chip area and edge contact length. The least square solution for this problem gives the o-th estimation of the SFC, eq. (3.24):

$$\hat{\boldsymbol{\beta}}^{(o)} = (\mathbf{X}^{(o)T}\mathbf{X}^{(o)})^{-1}\mathbf{X}^{(o)T}\mathbf{y}^{(o)} \tag{3.24}$$

thus, the SFC ($\hat{\boldsymbol{\beta}}^{(o)}$) estimated through this formula will be indicated as $\hat{K}_{tc}^{(o)}$, $\hat{K}_{te}^{(o)}$, $\hat{K}_{rc}^{(o)}$ and $\hat{K}_{re}^{(o)}$. Actually, at the beginning of each cut, the initial tool phase φ_0 is unknown. The estimation of φ_0 can be obtained solving N_q regressions with a different tool phase φ_{0q} (eq. (3.25)):

$$\varphi_{0q} = \frac{2\pi}{N}\frac{q}{R}, \qquad q = 1, \dots, N_q \tag{3.25}$$

The regression with minimum least square error gives the best estimates for φ_0 and SFC [54].

3.2.4 Principal Components Regression

As it will be explained in section 3.4, the $\hat{\boldsymbol{\beta}}$ estimation (for this particular problem) tends to suffer from the multicollinearity phenomenon, which turns into a high variability of the estimated regression coefficients. In order to limit this effect, Principal Components Regression (PCR) was introduced [85]. PCR performs a multivariate linear regression on a subset of principal components of the design matrix. At first, each column of the design matrix and the response vector are standardised to have null mean and unitary standard deviation [79], bringing to the $\tilde{\mathbf{X}}$ matrix and $\tilde{\mathbf{y}}$ vector (omitting o dependency, eq. (3.26)):

$$\begin{aligned} \tilde{\mathbf{X}} &= \begin{bmatrix} \frac{\mathbf{x}_1-\bar{x}_1}{\sigma_{x_1}} & \frac{\mathbf{x}_2-\bar{x}_2}{\sigma_{x_2}} & \frac{\mathbf{x}_3-\bar{x}_3}{\sigma_{x_3}} & \frac{\mathbf{x}_4-\bar{x}_4}{\sigma_{x_4}} \end{bmatrix} \\ \tilde{\mathbf{y}} &= \frac{\mathbf{y}-\bar{y}}{\sigma_y} \end{aligned} \tag{3.26}$$

Then, PCA is performed on the matrix $\tilde{\mathbf{X}}$. This is achieved through the computation of the eigenvalues $\boldsymbol{\lambda}$ and eigenvector matrix $\mathbf{V}$ of the covariance matrix $\boldsymbol{\Sigma}$ of $\tilde{\mathbf{X}}$. The standardised design matrix $\tilde{\mathbf{X}}$ is projected on the principal components (PCs) directions by eq. (3.27):

$$\mathbf{Z} = \mathbf{X}\mathbf{V} \tag{3.27}$$

The principal components matrix $\mathbf{Z}$ is then used to perform a multivariate linear regression on the response vector $\tilde{\mathbf{y}}$. The o-th regression coefficients estimate performed by PCR (eq. (3.28)) is $\tilde{\boldsymbol{\alpha}}^{(o)}_{full}$:

$$\tilde{\boldsymbol{\alpha}}^{(o)}_{full} = (\mathbf{Z}^{(o)T}\mathbf{Z}^{(o)})^{-1}\mathbf{Z}^{(o)T}\tilde{\mathbf{y}}^{(o)} \tag{3.28}$$

At this point, it is necessary to remove the PCs with the lowest eigenvalues (i.e. the ones describing the lowest part of variability in the data). This is achieved by setting the last elements of $\tilde{\boldsymbol{\alpha}}^{(o)}_{full}$ to zero and obtaining $\tilde{\boldsymbol{\alpha}}^{(o)}_{PCR}$ [85]. In the studied case, two coefficients are set to zero, eq. (3.29):

$$\tilde{\boldsymbol{\alpha}}^{(o)}_{PCR} = \begin{Bmatrix} \tilde{\alpha}^{(o)}_{1,full} & \tilde{\alpha}^{(o)}_{2,full} & 0 & 0 \end{Bmatrix}^T \tag{3.29}$$

Finally, it is possible to transform $\tilde{\boldsymbol{\alpha}}^{(o)}_{PCR}$ back to the original space through eq. (3.30):

$$\tilde{\boldsymbol{\beta}}^{(o)}_{std} = \mathbf{V}^{(o)}\tilde{\boldsymbol{\alpha}}^{(o)}_{PCR} \tag{3.30}$$

Due to the standardisation of the data, the estimated coefficients must be scaled to match the original dimensions. The transformation gives the final set of regression coefficients $\tilde{\boldsymbol{\beta}}^{(o)}$ and it is given by eq. (3.31):

$$\tilde{\boldsymbol{\beta}}^{(o)} = \begin{Bmatrix} \frac{\tilde{\beta}^{(o)}_{1,std}}{\sigma^{(o)}_{x_1}} & \frac{\tilde{\beta}^{(o)}_{2,std}}{\sigma^{(o)}_{x_2}} & \frac{\tilde{\beta}^{(o)}_{3,std}}{\sigma^{(o)}_{x_3}} & \frac{\tilde{\beta}^{(o)}_{4,std}}{\sigma^{(o)}_{x_4}} \end{Bmatrix}^T \sigma^{(o)}_y \tag{3.31}$$

thus, the SFC ($\tilde{\boldsymbol{\beta}}^{(o)}$) estimated through this method will be called as $\tilde{K}^{(o)}_{tc}$, $\tilde{K}^{(o)}_{te}$, $\tilde{K}^{(o)}_{rc}$ and $\tilde{K}^{(o)}_{re}$.

3.2.5 Self-starting Tabular Cusum control chart

The regression coefficients become here the object of the monitoring strategy. As it will be explained in section 3.4, the $\tilde{K}_{re}$ coefficient is the most stable and correlated to the tool wear. Here, it will be used as a tool wear indicator to be monitored. The monitoring strategy is based on a statistical process control tool, that is a Self-starting Tabular Cusum control chart. This choice was based following Nouri *et al.*, who applied Cusum control chart to monitor SFC identified through a mean identification procedure. Cusum chart directly incorporates all the information in the sequence of sample values by plotting the cumulative sums of the deviations of the sample values from a target value [156]. When the process is in control, the mean of the monitoring variable corresponds to the target value. The Tabular Cusum works by accumulating deviations from the mean which are above and below the target. Secondly, with respect to a traditional Shewhart control chart, Cusum control charts are very sensible to mean drifts. This peculiarity is a consequence of the fact that a Shewhart control chart takes in consideration just the last drawn sample to check if the process is in control, while Cusum control chart works on the whole process history. The responsiveness and sensitivity of the Cusum is of fundamental importance when applying them in a self-starting fashion. In fact, the design phase is substituted with an online estimation of process mean and variance. If the control chart is not fast enough to detect an out of control process, the online estimation of mean and variance will include out of control conditions, not allowing for a successive detection. Thus, following the work by Farhadmanesh *et al.* the moving range related to the $\tilde{K}_{re}^{(o)}$ estimation is computed through eq. (3.32) [54]:

$$MR_o = |\tilde{K}_{re,o} - \tilde{K}_{re,o-1}| \tag{3.32}$$

The moving range is then averaged on 15 samples batches (eq.(3.33)), leading to the variable v_n:

$$v_n = \sum_{o=1}^{15} \frac{MR_o}{15} \tag{3.33}$$

v_n is then fed to a self-starting tabular Cusum control chart [152]. First of all, the running average is updated as soon as the n-th observation v_n becomes available (eq.(3.34)):

$$\overline{v}_n = \overline{v}_{n-1} + \frac{v_n - \overline{v}_{n-1}}{n} \tag{3.34}$$

The sum of squared deviations is updated, too (eq.(3.35)):

$$w_n = w_{n-1} + \frac{(n-1)(v_n - \overline{v}_{n-1})^2}{n} \tag{3.35}$$

The running standard deviation becomes (eq.(3.36)):

$$s_n = \sqrt{\frac{w_n}{n-1}} \tag{3.36}$$

The n-th observation is then standardised through eq. (3.37):

$$T_n = \frac{v_n - \overline{v}_{n-1}}{s_{n-1}} \tag{3.37}$$

It turns out that the quantity U_n is distributed as a standard normal random variable (if the monitored variable v is normally distributed), where U_n is defined by eq. (3.38):

$$U_n = \Phi^{-1}\left[CDF_{t,n-2}\left(a_n T_n\right)\right] \tag{3.38}$$

where $\Phi^{-1}(\cdot)$ is the inverse normal cumulative distribution; $CDF_{t,n-2}(\cdot)$ is the cumulative t distribution with $n-2$ degrees of freedom; $a_n = \sqrt{\frac{n-1}{n}}$ [152]. Thus, it is possible to apply the tabular Cusum control chart on U_n.

The Cusum control chart is based on the computation of two cumulative derivation with above and below the target value, C_n^+ and C_n^-, respectively. They are initialised to be null: $C_0^+ = 0$ and $C_0^- = 0$. Every time a U_n becomes available, the following expressions are computed (eq.(3.39)):

$$\begin{aligned} C_n^+ &= \max\left[0, U_n - (\mu_0 + K) + C_{n-1}^+\right] \\ C_n^- &= \max\left[0, (\mu_0 - K) - U_n + C_{n-1}^-\right] \end{aligned} \tag{3.39}$$

where K is typically set to $\frac{1}{2}\sigma_0$; μ_0 and σ_0 are the mean and standard deviation of U random variable, thus $\mu_0 = 0$ and $\sigma_0 = 1$. The process is considered out of control when either C_n^+ or C_n^- overcome the threshold H, commonly set to $5\sigma_0$ [152].

3.3 Materials

The experimental session was carried out on two different machine tools: a Mandelli M5 machine tool and a Sigma Flexi FFG group machine tool; both of them, featuring a Mitsubishi AJX06R203SA20S high-feed mill, with three JOMT06T216ZZER-JL MP9140 cutting inserts. The available set-ups were presented in figure 3.4. During the milling test, Kistler 9255B dynamometric plate was adopted to measure the cutting forces during the milling operations. The dynamometer was connected to Kistler 5070A charge amplifier (fig. 3.4c). The cutting forces were acquired through a NI cDAQ-9174, with a NI 9215 acquisition card. The sampling frequency of the acquisition was set to $5kHz$. For the inspection of wear on cutters, a Keyence VHX-7000 microscope was used (fig. 3.4d).

The experimentation consisted of 5 run-to-failure tests (i.e. machining the workpiece until the cutters were completely worn out). The run-to-failure experiments were performed until either mean flank wear reached $300\mu m$ or maximum flank wear reached $600\mu m$. The acquisition of the experimental tests was subdivided in subsequent stops in order to perform the flank wear assessment (fig. 3.4). The workpiece consisted of a 255x262mm block made of Ti_6Al_4V, grade 5. The experimentation was carried out in different cutting conditions, i.e. changing machine tool (M5 and Flexi), cutting speed ($50m/min$, $70m/min$ and $125m/min$) and lubrication set-ups (lubricant and cryogenic). The decision to use two different machine tools was based on the information about the portability of the conceived method. In fact, testing the methodology upon two different machine tools, with the same algorithm settings allows to understand how the algorithm is robust within a scenario closer to the industrial one. Manufacturing companies or production systems generally include more than one machine tool and having a monitoring system that works properly on all of them is strategic. The axial depth of cut a was set to $0.4mm$, the radial depth of cut b was set to $13mm$, while the

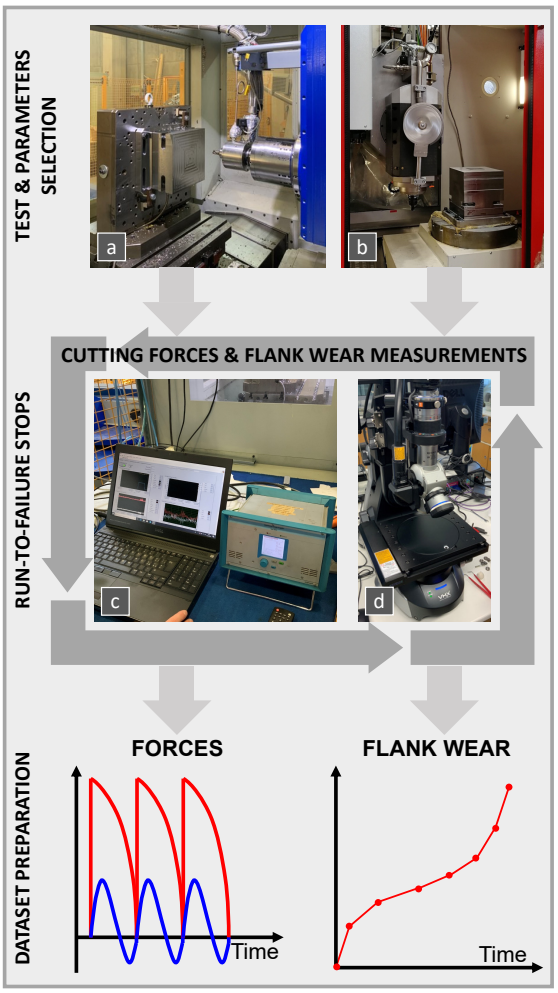


Figure 3.4: *Experimental set-up: a) Mandelli M5, Mitsubishi mill, Kistler dynamometer and cryogenic set-up; b) Sigma Flexi FFG group, Mitsubishi mill, Kistler dynamometer and conventional cooling set-up; c) DAQ system; d) Keyence microscope. Run-to-failures are performed through stops, in order to assess the flank wear.*

feed per tooth c was fixed to $0.7mm$. The cutting test descriptive data are reported in table 3.5. As can be seen in the table, a cutting test at $50m/min$ with traditional lubricant was not performed. This was associated to the experimental cost of such test. $50m/min$ is the optimal cutting speed for titanium alloys machining with conventional lubrication techniques. Lubricants main advantage is to reduce the friction between tool and workpiece and to remove heat from the cutting tool, thus dramatically increasing tool life. Additionally, high-feed mills work at really high material removal rates, thus leading to a not affordable economical cost of the test (in terms of material cost and machine tool cost). The same is not valid for cryogenic, which is not stable like conventional lubrication and has the strongest effects at high cutting speeds.

Table 3.5: *Design of Experiments*

Test	$v_c\,[m/min]$	$n\,[rpm]$	$v_f\,[mm/min]$	Lubrication	Machine
1	50	796	1671	Cryogenic	Flexi
2	70	1114	2340	Cryogenic	Flexi
3	70	1114	2340	Lubricant	Flexi
4	125	1989	4178	Lubricant	M5
5	125	1989	4178	Cryogenic	M5

3.4 Results and discussion

3.4.1 SFC estimation and cutting forces prediction

The novel high-feed mill formulation was fitted on the whole dataset of experimental data. Since the fitting procedure worked on instantaneous cutting forces, the accuracy of the model in the time interval used to retrieve the SFC is really high. An example of fitted cutting forces, compared with experimentally measured forces is reported in figure 3.5. For the example set of forces (taken from test 1), the R-squared was equal to 0.979 (following R-squared definition for regression through the , [50]). Obviously, the high fitting capability of the instantaneous cutting forces-based procedure, is constrained to the time interval in which it is computed. The prediction capabilities may be comparable or slightly worse than an average cutting forces-based procedure in terms of forecasting new cutting forces, especially when different cutting parameters are used. Nonetheless, for the conceived unsupervised approach, only a local/instantaneous estimation of SFC is of interest (i.e. regression coefficients) and not the prediction of new cutting forces.

3.4.2 MLR and PCR comparison

In figure 3.6, the evolution of the considered four SFC along the execution of all the tests using MLR and PCR is represented. Figures 3.6a, 3.6c, 3.6e, 3.6g and 3.6i showed the evolution of the SFC estimated through ordinary least squares solution of multivariate linear regression. It can be noted that the identification of SFC through this method suffers of high variability. This effect is linked to a multicollinearity phenomenon. Such phenomenon occurs when the regressors (i.e. the columns of the design matrix) tend to be linearly dependent one with respect to the other. This translates in a large variance

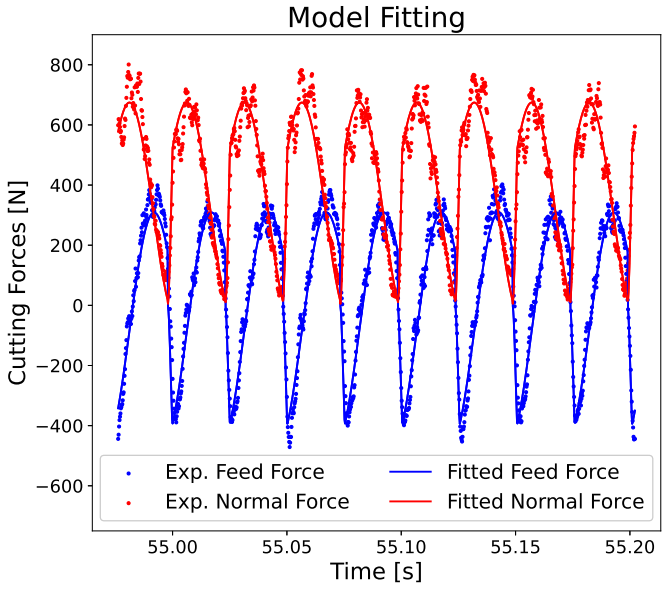
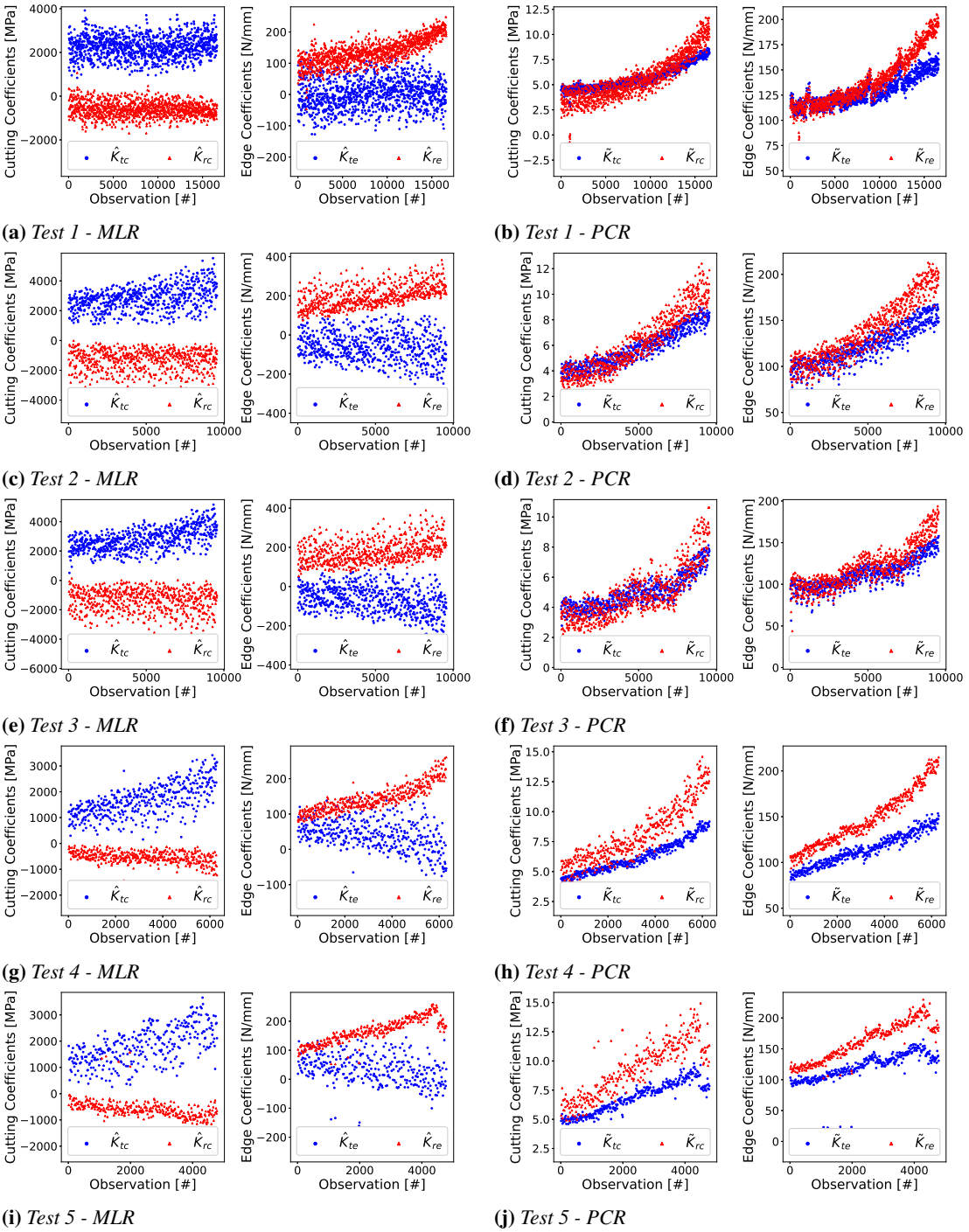


Figure 3.5: *Comparison between the o-th fitted instantaneous cutting forces and the o-th experimentally measured ones (thus including 3 mill revolutions). The example set of forces was taken from test 1: Flexi machine tool,* $v_c = 50\,m/min$*,* $c = 0.7\,mm/tooth$*, cryogenic,* $n = 796\,rpm$*,* $v_f = 1671\,mm/min$*.*

in the estimation of regression coefficients. The multicollinearity problem is typically solved by implementing biased regression methods. In general, these methods dramatically reduce the estimation variance of the regression coefficients, hopefully leading to a better bias-variance trade-off on the estimation; thus, to a more reliable estimation of the SFC. Here, as described in section 3.2, the Principal Component Regression algorithm is proposed to reduce the variance associated to the estimated SFC. The output of PCR is presented for all the tests in figures 3.6b, 3.6d, 3.6f, 3.6h and 3.6j. It must be noted, how PCR enhanced the reliability of the SFC estimation (i.e. by reducing the estimation variability), providing an intrinsic highlight of the wear phenomenon. On the other hand, PCR introduced a high bias in the cutting SFC (i.e., $\tilde{K}_{tc}$ and $\tilde{K}_{rc}$) which were reduced of almost three orders of magnitude. This is a result of the dimensionality reduction implemented by PCR; forgetting about the last two principal components (as explained in section 3.2), caused the cutting coefficient contributions to be almost removed. Furthermore, it is evident that $\tilde{K}_{re}$ is the coefficient presenting early signs and higher correlation with respect to tool degradation. This was expected and in agreement with literature analysis developed on mean forces-based estimation approaches, such as in [156].

3.4.3 Tool wear detection

Based on the results presented in section 3.4.2, it was decided to apply the self-starting control charts directly on $\tilde{K}_{re}$, as explained in section 3.2. The application of this kind of control charts allowed to build a completely unsupervised strategy to monitor the tool wear status along run-to-failures. In fact, they don't need a design phase to determine the variability and target of the monitored variable, but they are developed such that they estimate them as long as the data become available. The results of the application



(a) *Test 1 - MLR*

(b) *Test 1 - PCR*

(c) *Test 2 - MLR*

(d) *Test 2 - PCR*

(e) *Test 3 - MLR*

(f) *Test 3 - PCR*

(g) *Test 4 - MLR*

(h) *Test 4 - PCR*

(i) *Test 5 - MLR*

(j) *Test 5 - PCR*

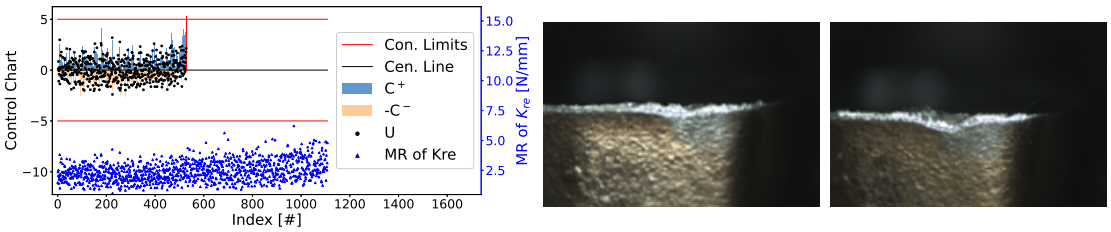
Figure 3.6: *Evolution of the specific force coefficients during the execution of tests. SFC on the left are obtained through multivariate linear regression (MLR) and ordinary least squares solution. SFC on the right are obtained through Principal Component Regression (PCR).*

of self-starting control charts on the $\tilde{K}_{re}$ moving range on the full set of tests are shown in figure 3.7. The figure for each experimental test the computed self-starting control chart main quantities. The red bar represents the first out of control observation of the process. Control charts graphs are accompanied by the flank wear and, if relevant, rake crater pictures at the stop before (pictures in the middle) and after (pictures on the right) the detection, in order to highlight the meaning of an out of control observation.

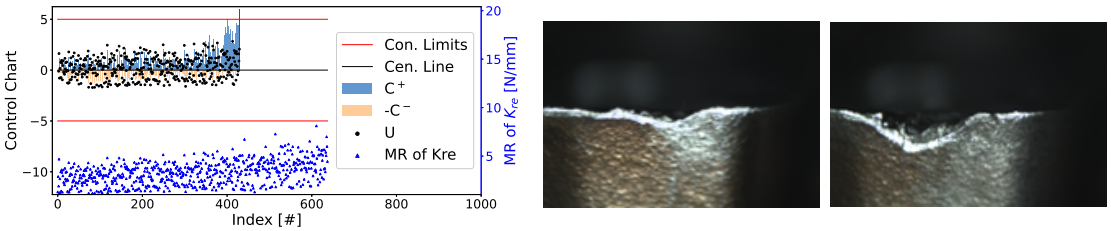
The evolution of the mean moving range (blue triangles) in figure 3.7 presents two different regions: a first stage, related to the initial tool life, where this coefficient is more or less stable, similar to the aspect of a random variable; a second stage with an increasing trend, higher variance and outliers. This behaviour is in agreement with the results found by Nouri *et al.*, who observed that the mean moving range of the K_{re} (estimated through a mean forces-based approach) faced a dramatic change in variance after a given number of performed cuts. They reported that this change in the behaviour of the moving range was correlated to either the occurrence of chipping on the cutting edge of tool inserts or a 3-dimensional evolution of the wear zone on the tool insert [156]. This was due to the sum of variations of infinitesimal cutting force directions along the worn cutting edges. The pictures of the cutting tool edges in figure 3.7 seem to be in agreement with the same hypothesis. The detection of the out of control observation falls in between the two stops corresponding to the proposed middle and right images. For what concerns tests 1, 2 and 3, the detection occurs when the cutting edge starts to feature an enhanced V wear shape, which is a 3D geometrical feature connecting the rake and flank faces (figures 3.7a, 3.7b and 3.7c). This phenomenon is associated to the occurrence of notch wear. After an initial period of regular flank land increase, localised accumulation of damages due to adhesion modify the cutting edge geometry leading to the formation of a notch. The notch is a 3D geometrical feature connecting the crater and the flank land and can be identified from a V shape on the flank surface. The proposed monitoring algorithm detects an out of control point when the notch wear starts to become critical. Regarding tests 4 and 5, the cutting inserts feature small chippings on the flank face, accompanied by bigger chippings on the rake face (as can be seen in figures 3.7d and 3.7e).

3.4.4 The influence of cutting conditions

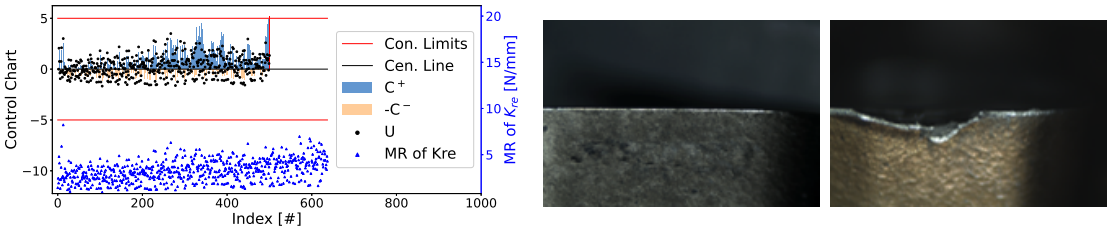
The algorithm detects an out of control cutting process when the chipping phenomenon takes place on the rake face. The difference between the first three tests and the last two seems to be correlated to the adopted cutting speeds. Low speed machining brings to more intensive adhesion phenomena [4], favouring notch wear occurrence. Furthermore, a difference between the two lubrication conditions was found at these cutting speeds. Cryogenic lubrication brought workpiece material to attach to the cutting edge (built-up-edge) on the cutting insert. Conversely, the cutting edge of conventional cooling at low speeds seemed cleaner, despite presenting notch wear, too. At high cutting speeds ($125 m/min$) the differences in the behaviour of the two lubrication conditions was less visible, bringing to small distributed chippings along the cutting edges. Despite these different phenomenological behaviours, the conceived algorithm behaved properly and was able to detect critical tool failures.



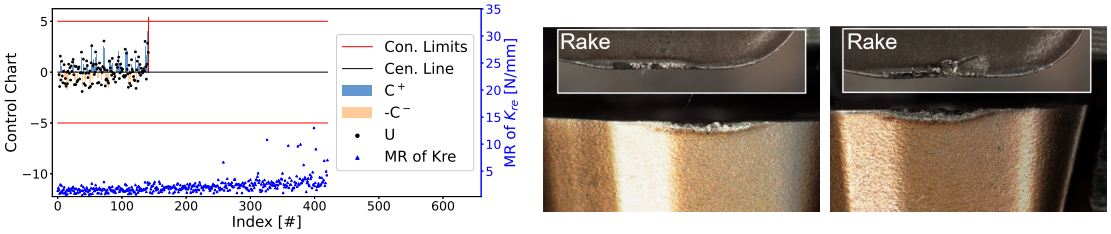
(a) *Test 1: Control Chart -* $VB_{m,d} = 305\mu m$ *(detection);* $VB_{m,3} = 221\mu m$ *(before);* $VB_{m,4} = 327\mu m$ *(after).*



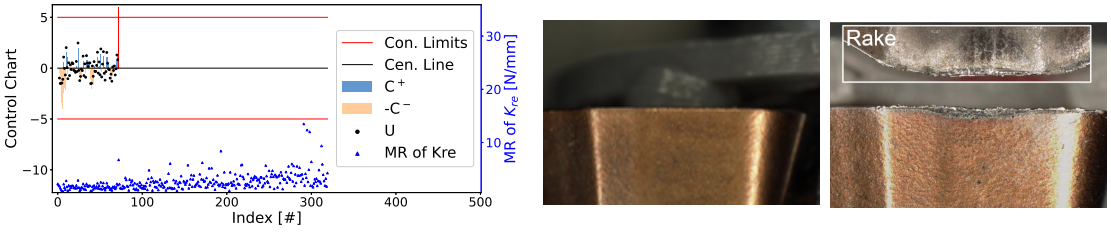
(b) *Test 2: Control Chart -* $VB_{m,d} = 499\mu m$ *(detection);* $VB_{m,1} = 386\mu m$ *(before);* $VB_{m,2} = 675\mu m$ *(after).*



(c) *Test 3: Control Chart -* $VB_{m,d} = 373\mu m$ *(detection);* $VB_{m,0} = 0\mu m$ *(before);* $VB_{m,1} = 475\mu m$ *(after).*



(d) *Test 4: Control Chart -* $VB_{m,d} = 293\mu m$ *(detection);* $VB_{m,1} = 293\mu m$ *(before);* $VB_{m,2} = 295\mu m$ *(after).*



(e) *Test 5: Control Chart -* $VB_{m,d} = 166\mu m$ *(detection);* $VB_{m,0} = 0\mu m$ *(before);* $VB_{m,1} = 180\mu m$ *(after).*

Figure 3.7: *Control chart detections among the tests. Photos represent the tool wear at the stop before (middle) and after (right) the control chart detection.*

3.4.5 Maximum flank wear measurements and sensitivity analysis

In figure 3.7, control charts were accompanied by the measured flank wears at the detection instants reported in the captions. It must be highlighted that flank wear measurements do not participate at all in the tool monitoring strategy. By the way, flank wear measurements were reported in order to contextualise the results of the proposed monitoring tool and to analyse from all the perspectives the out of control detections. The flank wear was computed making reference to the approaches explained in [3, 4] and standard ISO 3685 [93]. Thus, for the cutter i at the k-th stop, a set of N_f local measurements of the flank width, indicated as $VB_{ik,f}$, was available. The maximum flank width of the i-th cutter at the k-th stop is indicated as $VB_{m,ik}$ and is computed through eq. (3.40):

$$VB_{m,ik} = \max_{f \in [1,N_f]} VB_{ik,f} \tag{3.40}$$

Then, the maximum flank wear between the cutting inserts at the k-th stop is computed and reported as $VB_{m,k}$ (eq. (3.41)):

$$VB_{m,k} = \max_{i \in [1,N]} VB_{m,ik} \tag{3.41}$$

The maximum flank wear values at the stop before and after the detection point (in figure 3.7 reported as $VB_{m,k}$ and $VB_{m,k+1}$, with k assuming the corresponding stop number) are linearly interpolated to obtain maximum flank wear at the detection point (reported as $VB_{m,d}$). Thus, the detection point d falls between stops k and $k+1$. Let's call the observation numbers of these three points n_d, n_k and n_{k+1}, respectively. Maximum flank wear in correspondence of the detection point $VB_{m,d}$ is computed as follows (eq. (3.42)):

$$VB_{m,d} = VB_{m,k} + \frac{VB_{m,k+1} - VB_{m,k}}{n_{k+1} - n_k} (n_d - n_k) \tag{3.42}$$

In order to properly discuss the out of control detections, a sensitivity analysis of the algorithm was performed with respect to the control charts threshold. The results of such analysis were summarised in table 3.6. The first column of the table represents the case were $H = 5\sigma_0$. Such value is the standard one for a control chart of this kind [152]. Thus, these results are obtained without any kind of training procedure or supervision. The reference VB_m is $600\mu m$, which is the flank wear threshold at which the run-to-failures were physically stopped (in accordance with ISO 3685, [93], and [156]). The results are compared with the ones obtained by Nouri et al. in [156]. Five steel cutting tests were considered by the authors of the work. In order to compare results between the two condition monitoring systems, the reference mean flank wear was set to $300\mu m$, which was the tool wear criterion set by the researchers in accordance with the standard. Their results are reported for completeness in the last column of table 3.6. As can be seen, the mean of the relative percentage error (RPE) of the proposed method is -45% with respect to the reference threshold of $600\mu m$. This means that the conceived tool monitoring strategy is conservative with respect to the tool wear criterion suggested by the flank wear. Anyway, it is of primary importance to underline that, as presented in section 3.4.3, an out of control detection is corresponding to severe damages on the cutting tools (such as micro chippings, chippings and notches), which lead to more severe implications for workpiece quality and operator safety than progressive

tool wear [135]. This is a symptom of the higher order of information carried by the cutting force signals with respect to a single information provided by flank wear [219]. Furthermore, particular attention must be given to test 5 (the test facing the maximum relative error of -72%). In fact, the flank wear measurement is really low ($166\mu m$) compared to the reference threshold ($600\mu m$). This behaviour must be contextualised in the challenging cutting conditions of the test: cryogenic lubrication (which is not uniform as a conventional lubrication) and really high cutting speed ($125m/min$). These parameters led to the formation of micro chippings in the cutting edge. Moreover, it was found that cryogenic lubrication lowers the frictional effect on the cutting tool, justifying a low flank wear. Compared to Nouri et al. results, the proposed method is less conservative (-45% with respect to -57% mean RPE). The uniformity in the detection outputs is similar, by looking at the RPE standard deviations (20% and 13% in favor of Nouri et al. method). The higher variance can be explained on the basis of two reasons: Nouri et al. method is based on mean cutting forces (from which more stable stable SFC can be inferred) and a mean flank wear indicator (in opposition to the maximum flank wear of the proposed study). Anyway, their method can't be used in any context, since continuously variable feed per tooth must be adopted. Furthermore, it has to be noticed that the method is providing a relatively uniform indication of tool wear in an unsupervised way, while changing machine tools, cutting conditions, lubrication media and degradation phenomena. Nouri et al. lower variability reflects also in a lower maximum RPE of -66%.

Relaxing the hypothesis of a fully unsupervised monitoring tool, it is possible to change the threshold of the control chart. This assumption can be relaxed either in presence of a training dataset or as long as data from the operational field are collected. In these cases it would be possible to tune the control chart threshold. Thus, in table 3.6, a sensitivity analysis of the proposed method with respect to the threshold of the control chart is proposed. It is possible to notice that test 5 (the one that presented the maximum RPE) is always detected to be out of control at $166\mu m$. This is a symptom of the fact that a severe modification of the signals was recorded. By analysing figure 3.7e, it is possible to notice that when the detection occurred, an evident degradation of the mean moving range was triggered, justifying the results of the sensitivity analysis. Cryogenic lubrication introduced a more complex degradation phenomenon in the cutting tool, which was the formation and propagation of micro cracks and edge chippings (visible on the rake face of fig. 3.7e). This result is important since highlights the higher content of information carried by the cutting force signals. Furthermore, moving towards higher control chart thresholds allows to reduce how much the method is conservative (leading to lower mean RPE). This could be helpful in case of light machining operations, instead of hard-to-cut materials machining, such as Ti_6Al_4V.

3.5 Conclusion

In this chapter, an unsupervised tool wear monitoring strategy was conceived and validated. The proposition of a high-feed mechanistic model, an instantaneous forces-based specific force coefficients (SFC) fitting procedure (following [54]) and a self-starting tabular cusum control chart to detect an out of control cutting process were the key aspects. The main results of this work included:

Table 3.6: *Control chart threshold sensitivity analysis results. Mean, standard deviation and maximum of relative percentage errors (RPE) are computed with respect to a threshold of $VB_m = 600\mu m$ (maximum is computed using the absolute value of RPE, but is reported with the sign in order to keep the underestimation/overestimation case). For Nouri et al. [156] cases, RPE is reported with reference to $VB_{mean} = 300\mu m$.*

	Test	$H = 5\sigma_0$ $VB_{m,d}\,[\mu m]$	$H = 4\sigma_0$ (-20%) $VB_{m,d}\,[\mu m]$	$H = 6.5\sigma_0$ $(+30\%)$ $VB_{m,d}\,[\mu m]$	$H = 7.5\sigma_0$ $(+50\%)$ $VB_{m,d}\,[\mu m]$	$H = 10\sigma_0$ $(+100\%)$ $VB_{m,d}\,[\mu m]$	Nouri $VB_{mean,d}$ $[\mu m]$
	1	305	121	308	308	331	130
	2	499	472	501	544	552	198
	3	373	244	412	414	426	103
	4	293	293	293	293	294	116
	5	166	16	166	166	166	104
RPE	Mean [%]	-45	-62	-44	-43	-41	-57
RPE	St. Dev. [%]	20	29	21	24	24	13
RPE	Max [%]	-72	-97	-72	-72	-72	-66

- a new analytical mechanistic model for double-phased high-feed mills. The model includes the description of variable and gradual engagement of these cutting tools inside the material, while providing a parametric description of their geometry. The new formulation allows to identify SFC, where generic milling models can't.

- an improvement of the SFC fitting procedure proposed by [156] and [54]. It includes an instantaneous identification of SFC, removing the need for continuously variable feed per tooth during the cutting operation [156]. A principal component regression (PCR) approach was used in order to reduce the variability in the estimated SFC, due to the multicollinearity phenomenon. Multicollinearity doesn't allow to distinguish between the effects of the regressors. Thus, a small change in the experimental data may cause the coefficients to change according to what is referred to as see-saw effect. PCR implementation was proven to be efficient in limiting the multicollinearity phenomenon and naturally highlighting the tool wear information carried by SFC.

- an improvement of the solution proposed by [156] in terms of potentialities. Indeed, the conceived approach can be used to monitor the SFC with any occurring engagement condition, not relying on continuously variable feed during the workpiece machining in order to fit SFC. SFC can in fact be estimated directly from a small package of instantaneous cutting forces.

- the robustness with respect to working conditions. The conceived approach demonstrated to be valid among different cutting speeds (and thus degradation rates), lubrication conditions (i.e. conventional cooling and cryogenic, and thus chip formation mechanisms and wear phenomena) and machine tools, representing solid bases for solution portability.

- the correlation of the detection of out of control cutting process with 3D wear zones (notch wear phenomena) and chippings of the cutting edges. Providing a reliable and consistent detection of critical wear phenomena. Thus, the proposed

solution resulted in a prompt detection of these brittle degradation phenomena which were critical for the continuation of Ti_6Al4_V cutting [228].

- a complete unsupervised solution. The solution does not present any tuning or training procedure. The algorithm does not require any training datum in order to be fully operative. Control charts were implemented in a self-starting fashion and thresholds were chosen to be standard.
- relatively uniform prediction of gradual wear, indicated through maximum flank wear. The fully unsupervised monitoring strategy resulted in a flank wear mean relative prediction error of -45% with respect to a threshold of $600\mu m$ (run-to-failure end). The peak relative error reached -72%. Reference literature approaches reached a mean and peak relative error of -57% and -66% [156], respectively. The proposed method resulted to be less conservative than literature approach, while showing a higher variance in the detection. This was justified by the challenging variable conditions (different machine tools, lubrication media and cutting speeds) and the use of instantaneous cutting forces instead of mean ones.
- a sensitivity analysis of the control chart threshold with respect to the maximum flank wear at the detection point was performed. This resulted in a mean relative prediction error of -62% (when the threshold was reduced by the 20%, and of -41% (when the threshold was increased by 100%). These results give indications about how the predictions vary by tuning the control chart threshold in a supervised environment (if data from the field are collected and used to train the algorithm). Thus, the algorithm can be chosen to be either less conservative or more.

Future works of the proposed method include the introduction of prognostics solutions to estimate the Remaining Useful Life of the cutting tool and the SFC estimation based on more affordable measures like spindle currents (less invasive for industrial scenarios). Furthermore, research should focus on the correlation between the SFC evolution and the quality of cut left on the material surface.

CHAPTER 4

Tool wear: robust tool wear monitoring

The previous chapters of this thesis laid the basis for a cutting condition independent prognostics and health management tool. Nevertheless, the insurgence of unmodeled phenomena during the cutting process, such as different axes dynamics or tool run-out, may reduce the effectiveness of the previously presented real-time monitoring tool. Thus, a prognostics and health management layer capable of managing the process variability is here proposed by means of a robust artificial intelligence-based monitoring tool.

4.1 Introduction

Manufacturing systems require more and more accuracy, flexibility and reliability to face market demand. Thus, being capable of assessing the tool condition in real-time, is receiving increasing attention from the production system manufacturers [20]. Tool condition monitoring (TCM) techniques allow for the detection of worn cutting tools either from direct wear measurements (i.e., quantities directly associated to the tool wear like cutting edge pictures) or from indirect wear measurements (i.e. from quantities underlying information about wear). The development of an indirect TCM solution is the main focus of this chapter. In general, tool condition monitoring and prognostics strategies belong to four categories [14, 162]: knowledge-based, model-based, statistical-based and data-driven. Knowledge-based approaches include fuzzy logics and expert systems, trying to translate experts' knowledge in rules. Model-based approaches exploit dynamical models of wear evolution: they typically outperform other methods but generally they are not available for complex degradation phenomena. Statistical-based methods allow to identify model parameters from the data, introducing the concept of confidence of the tool condition estimation. Data-driven approaches build the model

and estimate the model coefficients directly from data, learning complex correlations between signals and degradation but, in general, preventing model interpretation.

Dealing with milling operations, the main limitation associated to TCM techniques based on indirect measurements regard the management of variable regimes or process conditions. The most used signals in TCM are vibrations (e.g. [34, 145]), cutting forces or torque (like in [34, 39, 105, 145]) and acoustic emissions [145, 211]. These quantities are dependent on process parameters, such as depth of cuts, feed per tooth as well as lubricating conditions and materials; this represents the main limitation for a consistent application of TCM approaches in industrial scenarios. Indeed, it would be necessary to perform run-to-failure tests with all the parameter combinations.

In this chapter, an unsupervised TCM strategy for gradual wear assessment is proposed, based on the online identification of Specific Force Coefficients (SFC) and artificial intelligence. SFC are estimated from the mechanistic milling model introduced in chapter 2 and 3, making reference to the Altintas' model [8]. Mechanistic milling models include a geometric description of the cutting process, thus, relating undeformed chip thickness formulation to the cutting forces based on the SFC. Many improvements have been introduced by researchers to the Altintas model: Kumanchik and Schmitz, together with Matsumura and Tumura included run-out and teeth trochoidal trajectory in the model [118, 144]; Wan *et al.* decomposed the cutting forces in nominal and run-out effects [199]; Kilic and Altintas developed a general model for chip removal operations including machine dynamics and run-out [113, 114]; Li *et al.* included the contribution of more than one previous teeth in the computation of undeformed chip thickness [130]; Zhang *et al.* included minimum chip thickness [230] while Zhou *et al.* introduced elastic recovery and variable entry/exit angles [238]; Zhang *et al.* proposed the associated average uncut chip thickness formulation [234]; at last, Hajdu *et al.* proposed a curved uncut chip thickness formulation [76]. In the conceived approach, the used mechanistic milling model considers variable engagement along the mill axis and cutter double-phased geometry.

SFC were selected as monitoring features for the tool condition by Nouri et al. [156], using a method based on mean forces, following the classical approach from [8]. Anyway, this method required experimental tests with continuously varying feed (not the typical case of parts production). Recently, methods based on instantaneous forces were introduced, relaxing such need [54, 74]. Nevertheless, such method brings to a higher uncertainty of estimated cutting coefficients, both due to machine dynamics, non-homogeneous materials and imperfect cutting models. Thus, it is necessary to properly analyse instantaneous SFC data.

In this chapter, a TCM method based on instantaneous SFC mapping is introduced, not requiring a predefined database of cutting operations for all the cutting combinations. The uncertainty in the estimation of SFC with instantaneous forces is shown to prevent a correct detection of an out-of-control cutting process. Thus, an unsupervised clustering technique (growing self-organising maps - GSOM) is introduced to deal with the estimation variability, adapting the solutions from [37, 142]. The chapter structure follows: in section 4.2, the developed procedure is explained in details, starting from the experimental campaign, followed by the algorithm development and the validation procedures; in section 4.3, the results are presented for two scenarios: optimisation on the full set of experiments; an industrial portability context. In section 4.4, final

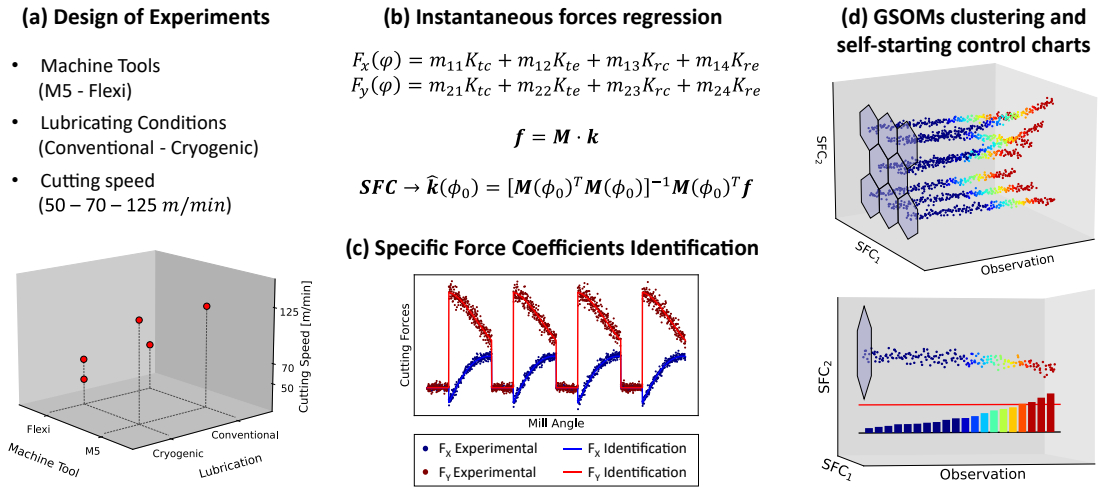


Figure 4.1: *Chapter framework: design of experiments in variable conditions (a); the multivariate regression approach based on instantaneous cutting forces (b-c); the unsupervised clustering of SFC based on GSOM (Growing Self-Organising Map), where a self-starting control chart is run inside each GSOM's region (d).*

considerations about the developed approach are drawn to the reader attention.

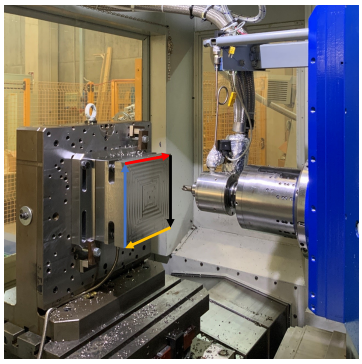
4.2 Materials and methods

The experiments and the developed methodology explanations are presented following figure 4.1.

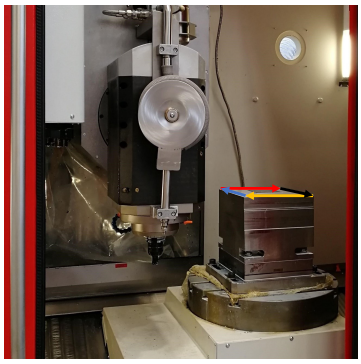
4.2.1 Experimental set-up and campaign

The experimental campaign was conducted on two different machine tools for milling applications. The first one is a Mandelli M5 machine tool (M5, from now on), featured by a horizontal spindle system. The other machine tool is a Sigma Flexi FFG group (Flexi, from now on), featuring a vertical spindle unit. The tooling system was composed of a Mitsubishi AJX06R203SA20S milling tool, mounting three high-feed cutting inserts (JOMT06T216ZZER-JL MP9140). The machine tools, M5 and Flexi, were shown in figure 4.2, together with the mill.

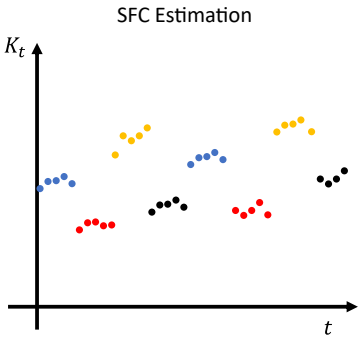
The high-feed cutting inserts are featured by a double-phased cutting edge, thus made of two consecutive sections with different lead angles (12.5° and 24.5°). Both the machines were equipped with two lubrication systems: a conventional lubrication system and a cryogenic system. Measurement set-up consisted of two hardware parts. From the machine side, the SinuCom NC acquisition system from Siemens was used to access and store quantities (axis positions, speeds and torques) from the Numeric Control of the machine tools at a frequency of $250Hz$. On the workpiece side, a dynamometric plate was installed in order to measure the cutting forces generated during the milling operations. The dynamometer is a Kistler 9255B, connected to a Kistler 5070A charge amplifier. The cutting forces were acquired through a NI cDAQ-9174, with a NI 9215 acquisition card with a sampling frequency of $5kHz$. For the inspection of wear on cutters, a Keyence VHX-7000 microscope was used.



(a) *Mandelli M5.*



(b) *Sigma Flexi FFG group.*



(c) *Qualitative example of axes dynamics effect on SFC estimates.*

Figure 4.2: *Adopted machine tools (a-b) and effect of unmodeled terms in specific force coefficients estimation.*

A set of experimental run-to-failure tests was performed, bringing any cutter from a "new" condition to a fully degraded condition. Run-to-failures were considered as concluded either when an average flank wear of $300\mu m$ or a maximum flank wear of $600\mu m$ was reached. The set of experimental tests consisted of five run-to-failures, for which the variable cutting conditions were reported in table 4.1. The workpiece was a 255x262mm block made of Ti_6Al_4V, grade 5. The experimentation was performed with variable cutting conditions, i.e. changing machine tools (M5 and Flexi), cutting speeds ($50m/min$, $70m/min$ and $125m/min$) and lubrication set-ups (lubricant and cryogenic). The axial depth of cut a was set to $0.4mm$, the radial depth of cut b was set to $13mm$, while the feed per tooth c was fixed to $0.7mm$. A summary of the design of experiments is reported in figure 4.1a.

Table 4.1: *Design of Experiments*

Test	$v_c\,[m/min]$	$n\,[rpm]$	$v_f\,[mm/min]$	Lubrication	Machine
1	50	796	1671	Cryogenic	Flexi
2	70	1114	2340	Cryogenic	Flexi
3	70	1114	2340	Lubricant	Flexi
4	125	1989	4178	Lubricant	M5
5	125	1989	4178	Cryogenic	M5

4.2.2 SFC estimation

The conceived approach is based on the estimation of SFC during the milling test by means of multivariate linear regression on instantaneous forces (fig. 4.2b-c). Since the geometrical features of the mill cutter are not traditional, it was necessary to use an adequate formulation for the uncut chip thickness. In this case, the reference model is the same proposed in chapter 3, within section 3.2.2. Thus, the mathematical formulations for the high-feed tangential (F_t) and radial (F_r) instantaneous cutting forces are here

reproposed (from eq. 3.12):

$$\begin{aligned} F_t &= K_{tc}\,(A_p - A_c) + K_{te}\,l \\ F_r &= K_{rc}\,(A_p - A_c) + K_{re}\,l \end{aligned} \tag{4.1}$$

where K_{tc}, K_{te}, K_{rc} and K_{re} are the SFC; A_p and A_c are the areas underneath the previous and current tooth passes; l is the chip edge length. Cutting forces in the feed (F_x) and normal (F_y) directions are obtained by projecting and summing up each cutter contribution, following eq. (3.12). At this point it is possible to formulate the instantaneous SFC identification procedure.

The SFC estimation $\hat{\boldsymbol{\beta}}^{(o)}$ for the generic o-th package of cutting forces consisting of 3 mill rotation is (from eq. (3.24)):

$$\hat{\boldsymbol{\beta}}^{(o)} = (\mathbf{X}^{(o)T}\mathbf{X}^{(o)})^{-1}\mathbf{X}^{(o)T}\mathbf{y}^{(o)} \tag{4.2}$$

where the SFC estimated through this formula will be indicated as $\hat{K}_{tc}^{(o)}$, $\hat{K}_{te}^{(o)}$, $\hat{K}_{rc}^{(o)}$ and $\hat{K}_{re}^{(o)}$; $\mathbf{X}^{(o)}$ is the design matrix presented in eq. (3.22)-(3.23); $\mathbf{y}^{(o)}$ is the vector containing the instantaneous cutting forces samples.

4.2.3 Growing Self-Organising Maps adaptation

Beside the approach conceived in the previous chapter, where principal component regression was used to reduce the effect of multi-collinearity, an AI approach capable of mapping similar behaviours of SFC and managing the SFC variability may be appropriate. Furthermore, this may also apply to different sources of variability for the SFC: the variability in machine dynamics along the cutting axis, tool variable run-outs, material heterogeneity are just some examples of possible causes for SFC variability issues. Here, an adaptation of the GSOM is proposed (fig. 4.2d).

The GSOM is an unsupervised neural network that divides the input space in a variable number of regions (i.e. creating a map), trough the Voronoi tessellation [142]. Each region V_m is defined by a centroid $\boldsymbol{\xi}_m$ with $m = 1,2,...,M$ (being M the total number of regions), eq. (4.3):

$$V_m = \{\mathbf{x} : \min_{m\in[1,M]} \|\mathbf{x} - \boldsymbol{\xi}_m\|\} \tag{4.3}$$

where $\mathbf{x}$ is a point belonging to the input space. Thus, a point of the input space belongs to the region with the nearest centroid. The centroids have the possibility to move over time, giving the self-organisation property to the map. Differently form Self-Organising Maps, GSOM are allowed to grow, i.e. to increase the number of centroids (and consequently regions) as time passes. This makes GSOM a particularly suitable tool for unsupervised learning scenarios. For the purposes of this thesis, the GSOM process begins with an initialisation phase. In this phase, the input space for the GSOM deployment is firstly selected. The input vectors are assumed to be constituted by p elements (here $p = 6$), as reported in eq. (4.4):

$$\mathbf{x}^T = \begin{Bmatrix} \hat{K}_{tc} & \hat{K}_{te} & \hat{K}_{rc} & \hat{K}_{re} & c & \phi_c \end{Bmatrix} \tag{4.4}$$

where c is the feed per tooth and ϕ_c is the angle between the feed direction and the x-axis in the x-y reference frame (i.e. the cutting direction). It has to be notice that

any other cutting parameter or measured quantity may be included here. Obviously the more the input space dimension is increased, the slower the algorithm is. The feed per tooth was included since it may influence the estimation of the SFC (e.g. as assumed by exponential mechanistic models [25,43,216]), whereas the cutting direction influences cutting forces due to the different dynamical compliance of the machine tool in the x and y directions. The initialisation phase includes also the choice relative to the initial number of regions M (here, $M = 1$ is selected) and the definition of the associated centroid (here, the first measured input vector is proposed as the centroid initial position). From here on, the deployment of the GSOM takes place.

A new SFC estimation is available every 3 mill revolutions, as described in chapter 3. Indeed, an input vector $\mathbf{x}^{(o)}$ is ready, too. First of all, the GSOM computes the *Best Matching Unit* (BMU, eq. (4.5)):

$$bmu = \arg \min_{m \in [1,M]} \|\mathbf{x}^{(o)} - \boldsymbol{\xi}_m^{(o-1)}\| \tag{4.5}$$

bmu represents the index of the region with the nearest centroid to the input vector (according to the Euclidean distance). Once the BMU is selected, the input vector is assigned and stored only within the BMU. A first hyper-parameter n of the network is here defined and referred to as *memory factor*. n is the number of past input vectors retained within each region. Once a region collects n input vectors, their mean value $\overline{\mathbf{x}}_m$ and their estimated covariance matrix $\mathbf{S}_m$ are computed. In order to decide whether the GSOM should enter in a growing phase or a learning phase, the squared Mahalanobi's distance $d_M^2\left(\mathbf{x}^{(o)}, \overline{\mathbf{x}}_{bmu}^{(o-1)}\right)$ between the input vector $\mathbf{x}^{(o)}$ and $\overline{\mathbf{x}}_{bmu}^{(o)}$ is computed (eq. 4.6)):

$$d_M^2\left(\mathbf{x}^{(o)}, \overline{\mathbf{x}}_{bmu}^{(o-1)}\right) = \left(\mathbf{x}^{(o)} - \overline{\mathbf{x}}_{bmu}^{(o-1)}\right)^T \left(\mathbf{S}_{bmu}^{(o-1)}\right)^{-1} \left(\mathbf{x}^{(o)} - \overline{\mathbf{x}}_{bmu}^{(o-1)}\right) \tag{4.6}$$

where $\left(\mathbf{S}_{bmu}^{(o-1)}\right)^{-1}$ is the inverse of the previously estimated covariance matrix $\mathbf{S}_m$ for the BMU at sample o. Eq. (4.6) represents a measure of how far the new input vector is from the distribution of the previous n samples. Since matrix $\mathbf{S}_m$ may result to be singular, the generalised inverse is computed. Thus, the eigenvalues ($\lambda_{m,i}$) and the eigenvectors ($\boldsymbol{\nu}_{m,i}$) of the matrix $\mathbf{S}_m$ are firstly derived such that $\mathbf{S}_m \boldsymbol{\nu}_{m,i} = \lambda_{m,i} \boldsymbol{\nu}_{m,i}$. Each eigenvalue is then normalised in order to represent the fraction of described variance by $\rho_{m,i} = \frac{\lambda_{m,i}}{\sum_i \lambda_{m,i}}$. The generalised inverse is then computed trough eq. (4.7):

$$\left(\mathbf{S}_{bmu}^{(o-1)}\right)^{-1} = \sum_{i, \rho_{m,i} \neq 0} \frac{1}{\lambda_{m,i}} \boldsymbol{\nu}_{m,i} \boldsymbol{\nu}_{m,i}^T \tag{4.7}$$

Based on Mahalanobi's distance it is possible to generate a prediction region for a new input vector:

$$\frac{(n-p)n}{(n-1)p} d_M^2(\mathbf{x}^{(o)}, \boldsymbol{\mu}_{bmu}^{(o-1)}) \sim \mathscr{F}(p, n-p) \tag{4.8}$$

where $\mathscr{F}(p, n-p)$ is a Fisher distribution with a numerator having p degrees of freedom and a denominator having $n-p$ degrees of freedom. This relation is valid for normally distributed past input vectors and creates an elliptical region in the input space:

thus, fixing a confidence level α, the associated ellipse squared radius is given by the $F_{1-alpha,n,n-p}$ quantile. If this assumption is not verified, the ellipse radius can be tuned, for example through Monte-Carlo approaches. Since in this context, no assumptions can be made on the distribution of the past input vectors, the ellipse radius will be referred to as R_t (growth threshold). The left hand side term in eq. (4.8) will be referred to as R. Thus, the GSOM is allowed to grow only when $R > R_t$. This condition, implies that the input vector is dramatically changed with respect to the previous n, thus representing a different cutting condition. It is assumed that gradual wear produces a more progressive evolution of the input vector.

It has to be noticed that eq. (4.6) can be computed only once n input vectors are stored in the BMU. Thus, the GSOM is allowed to grow only when a significant sample size is collected within the BMU. If the n samples are not yet collected, or if $R \leq R_t$, the BMU is only allowed to learn. The learning phase consists of a smoothing process where the centroid $\boldsymbol{\xi}_{bmu}$ is shifted towards the current input vector by eq. (4.9):

$$\boldsymbol{\xi}_{bmu}^{(o)} = \boldsymbol{\xi}_{bmu}^{(o-1)} + \eta \left(\mathbf{x}^{(o)} - \boldsymbol{\xi}_{bmu}^{(o-1)} \right) \tag{4.9}$$

where η is a hyper-parameter called learning rate [142] and governs the nervousness of the centroids.

The described procedure, allows to automatically cluster the input vectors and consequently the SFC as soon as they arrive to the GSOM. From here, a post-process of the SFC is carried out online, inside each region. When $\mathbf{x}^{(o)}$ is assigned to the BMU, and the growth or learning phases are performed, the computation of two important synthetic coefficients is carried out following the approach presented in [156]. These two coefficients ($K_t^{(o)}$ and $K_r^{(o)}$) are computed through eq. (4.10):

$$\begin{aligned} K_t^{(o)} &= \frac{K_{tc}^{(o)} K_{te}^{(o)}}{\overline{K}_{tc} \overline{K}_{te}} \\ K_r^{(o)} &= \frac{K_{rc}^{(o)} K_{re}^{(o)}}{\overline{K}_{rc} \overline{K}_{re}} \end{aligned} \tag{4.10}$$

where $\overline{K}_{tc}$, $\overline{K}_{te}$, $\overline{K}_{rc}$ and $\overline{K}_{re}$ are the SFC means computed on the first N input vectors collected in the BMU. It is then possible to compute a summary indicator $K^{(o)}$, starting from $K_t^{(o)}$ and $K_r^{(o)}$:

$$K^{(o)} = \sqrt{\left(K_t^{(o)}\right)^2 + \left(K_r^{(o)}\right)^2} \tag{4.11}$$

while $K_t^{(o)}$ and $K_r^{(o)}$ are representative of the effect of tool wear on the tangential and radial forces, respectively, $K^{(o)}$ carries global wear information [156]. These coefficients are computed in order to try to reduce the multicollinearity effect previously introduced in chapter 3.

From now on, the focus is redirected inside the *bmu* region. Indeed, the index $o-1$ will refer to the last element of the *bmu* collected before the o-th one. The moving range of $K^{(o)}$ is computed through eq. (4.12):

$$MR_{bmu}^{(o)} = |K_{bmu}^{(o)} - K_{bmu}^{(o-1)}| \tag{4.12}$$

Following the processing method proposed in [156], every N samples collected by the BMU, it is possible to compute a mean moving range which will be the monitored variable $v_{k,bmu}$ (eq. (4.13)):

$$v_{k,bmu} = \sum_{o=1}^{N} \frac{MR_{bmu}^{(o)}}{N} \tag{4.13}$$

Thus, each time N samples are collected within a region of the GSOM, a sample for the monitored variable $v_{k,bmu}$ becomes available. The monitoring process is performed through a self-starting control chart within each region of the GSOM. It has to be noticed that in the GSOM, only one region per time is activated and monitored. The control chart is thus reported following chapter 3 and [152]. The running average is updated once $v_{k,bmu}$ becomes available (eq. (4.14)):

$$\overline{v}_{k,bmu} = \overline{v}_{k-1,bmu} + \frac{v_{k,bmu} - \overline{v}_{k-1,bmu}}{k} \tag{4.14}$$

The sum of squared deviations becomes (eq.(4.15)):

$$w_{k,bmu} = w_{k-1,bmu} + \frac{(k-1)(v_{k,bmu} - \overline{v}_{k-1,bmu})^2}{k} \tag{4.15}$$

The running standard deviation is updated through eq.(4.16):

$$s_{k,bmu} = \sqrt{\frac{w_{k,bmu}}{k-1}} \tag{4.16}$$

The k-th observation undergoes a standardisation performed through eq. (4.17):

$$T_{k,bmu} = \frac{v_{k,bmu} - \overline{v}_{k-1,bmu}}{s_{k-1,bmu}} \tag{4.17}$$

where $T_{k,bmu}$ is the standardised monitored variable. Following chapter 3 and [152], the $U_{k,bmu}$ variable computed through eq. (4.18) is distributed as a standard normal:

$$U_{k,bmu} = \Phi^{-1}\left[CDF_{t,k-2}\left(a_{k,bmu}\, T_{k,bmu}\right)\right] \tag{4.18}$$

where $\Phi^{-1}(\cdot)$ is the inverse normal cumulative distribution; $CDF_{t,k-2}(\cdot)$ is a $k-2$ degrees of freedom cumulative t distribution; $a_{k,bmu} = \sqrt{\frac{k-1}{k}}$. A tabular Cusum control chart is then applied on $U_{k,bmu}$.

The accumulation of the two deviations above ($C^{+}_{k,bmu}$) and below ($C^{-}_{k,bmu}$) the target value are initialised to be null ($C^{+}_{0,bmu} = 0$ and $C^{-}_{0,bmu} = 0$) and updated as follows (eq.(4.19)):

$$\begin{aligned} C^{+}_{k,bmu} &= \max\left[0, U_{k,bmu} - (\mu_0 + K_{cc}) + C^{+}_{k-1,bmu}\right] \\ C^{-}_{k,bmu} &= \max\left[0, (\mu_0 - K_{cc}) - U_{k,bmu} + C^{-}_{k-1,bmu}\right] \end{aligned} \tag{4.19}$$

where K_{cc} is set to $\frac{1}{2}\sigma_0$; μ_0 and σ_0 are the mean and standard deviation of U ($\mu_0 = 0$ and $\sigma_0 = 1$). Then, the BMU is considered out of control when either $C^{+}_{k,bmu}$ or $C^{-}_{k,bmu}$

overcome the threshold H, which becomes here a GSOM hyper-parameter (typically set to 5, [152, 156]). If a region goes out-of-control, it is not allowed to return an in-control output for future samples. Having generated a clustered growing map of SFC values, it is necessary to define a method for determining the out-of-control state at the GSOM level (not only at a regional level). Thus, a combination of the out of control outputs of each region is proposed through eq. (4.20):

$$\Delta^{(o)} = \sum_{m=1}^{M^{(o)}} \delta_m^{(o)} \frac{n_{m,bmu}^{(o)}}{o} \tag{4.20}$$

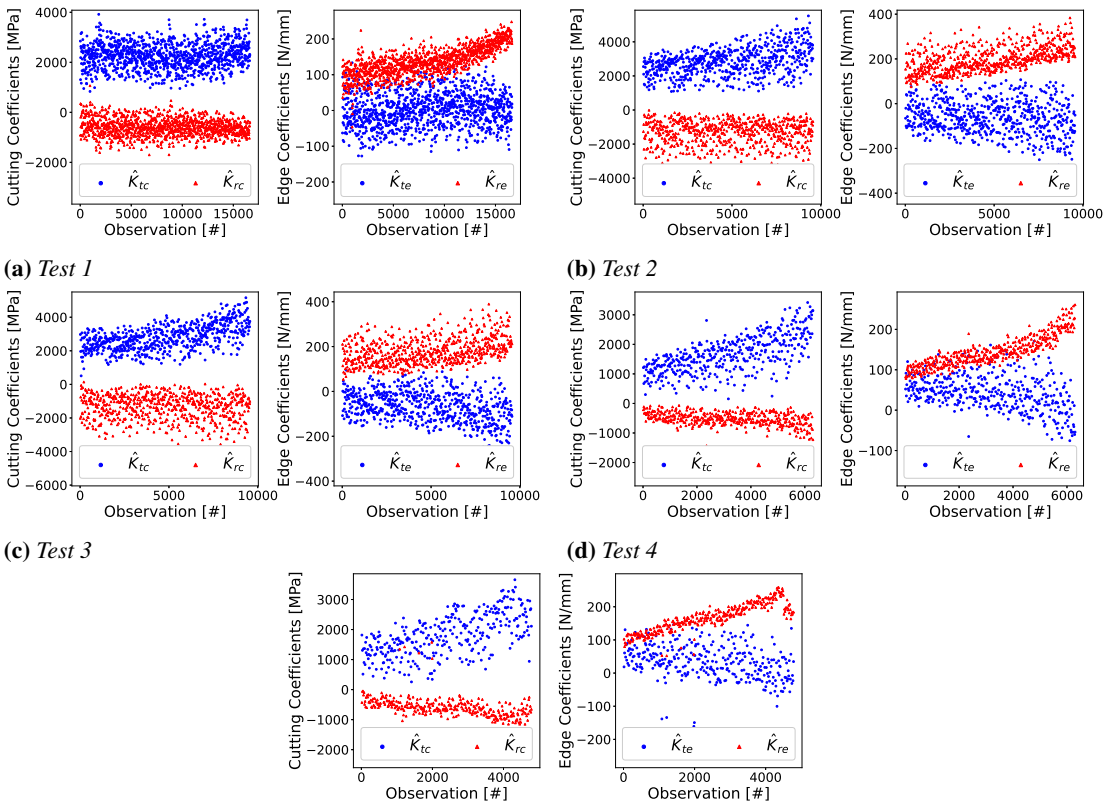
where $\Delta^{(o)}$ is the combined GSOM output indicator at the current o-th sample, varying between 0 and 1; m is the region index, while $M^{(o)}$ is the current number of regions in the GSOM; $\delta_m^{(o)}$ is a quantity equal to 0 or 1, if the m-th region is, at the current sample, in control or out of control, respectively; $n_{m,bmu}^{(o)}$ is the current number of times that region m was selected as the BMU; o is the current sample number. This indicator takes into account the control condition of each region through a weighted mean operation. The weight is heavier if the region is older and more frequented. The GSOM is considered out-of-control when $\Delta^{(o)}$ overcomes the threshold Δ_t, which constitutes the last hyper-parameter of the GSOM.

4.3 Results

The whole algorithm, starting from the SFC estimation up to the GSOM and control charts, was implemented in Python from scratch, using standard libraries (Numpy, Pandas, Scipy and Matplotlib). The analysis were performed on a Dell XPS 15 7590 featuring an Intel® Core™ i7-9750H CPU @ 2.60GHz.

As presented in section 4.2, a set of five run-to-failure tests was run. The instantaneous cutting forces of the five tests were used to fit the mechanistic force model presented in chapter 3. In figure 4.3, the estimations of SFC were presented for the whole experimental set. The evolution of the SFC based on instantaneous cutting forces shows some peculiarities. The SFC in fact present the multicollinearity effect[1]. As explained in chapter 3, the instantaneous identification process, tends to confuse between the effects of the four regressors on the predicted output (cutting forces). This implies a phenomenon in the SFC which is referred to as see-saw effect. This phenomenon consists in a correlation between the oscillations of the SFC. For instance, when the estimated K_{tc} increases, the K_{te} decreases. Such phenomenon is particularly evident in tests 2 (fig. 4.3b) and 3 (4.3c). This phenomenon is responsible for a high variability and instability of the SFC estimation over time, which hides tool wear influence on their values. Nevertheless, the oscillations are governed by unmodelled phenomena that cause cutting forces measurements to vary during workpiece machining. These phenomena may be related to run-out, cutting temperature, or by the different dynamic compliance of the machine tool with respect to the cutting directions.

[1]From a geometrical perspective, multicollinearity is easily explained in a bivariate regression problem. When two regressors are correlated, during the acquisition of a sample the points tend to be aligned in the 2D input space. Indeed, the regression curve (a plane) should be fitted to a cloud of points distributed nearly as a line. This means that random errors (points farther from the line) have a strong influence on the plane orientation, which is defined by the regression coefficients.



(a) *Test 1*

(b) *Test 2*

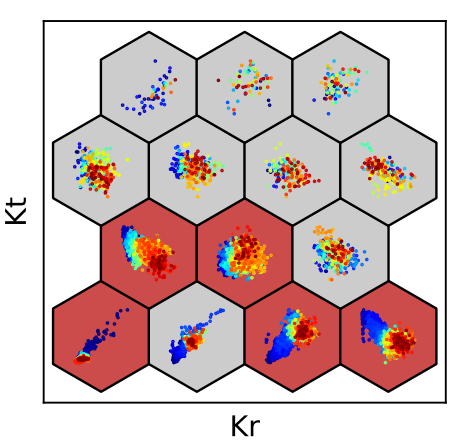
(c) *Test 3*

(d) *Test 4*

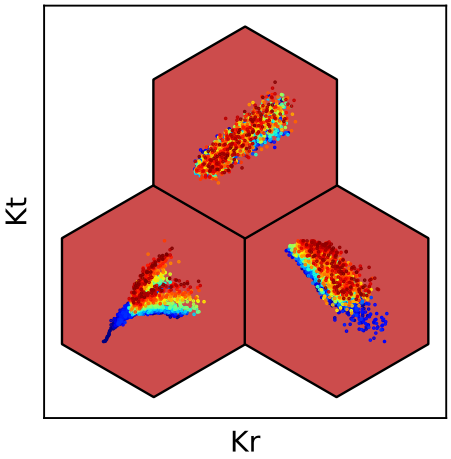
(e) *Test 5*

Figure 4.3: *Specific force coefficients estimation based on instantaneous cutting forces fitting. For each test, a pair of graph is showed: on the left, the cutting SFC; on the right, edge SFC.*

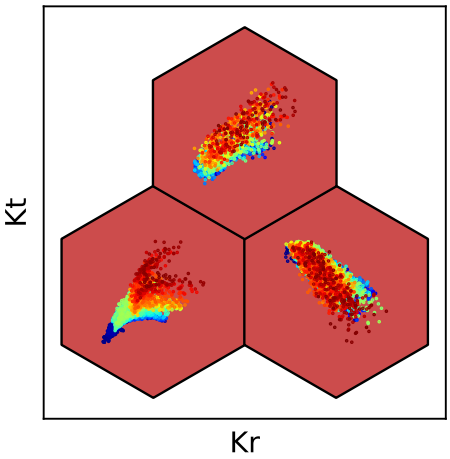
In order to face the undesired variability of the SFC, the GSOM was introduced. A visual representation of the GSOM application to each experimental test was shown in fig. 4.4. Figure 4.4 is a qualitative figure, summarising most of the information included in the GSOM clustering. The hexagonal cells are the representation of the regions created by the GSOM. In general, the number of generated cells is correlated to the overall variability of the estimated coefficients and to the number of samples of the test. In fact, when a sudden difference is found in two consecutive input vectors, it is more probable that the Mahalanobi's distance overcomes the growth threshold. Furthermore, the longer is the test, the higher is the probability of finding high variations in data. A second level of information is provided by the background colour of the cells. The figure shows the state of the GSOM map when a GSOM level out of control detection is observed. Thus, a grey cell background colour stands for an in-control cell at the end of the detection process; on the contrary, a red background colour represents an out-of control cell. Test cases number 2 and 3 present GSOM maps, where all the cells are out-of-control at the end of the process; whereas the remaining tests show also in-control cells. This behaviour is associated to the fact that the GSOM level out-of-control indicator $\Delta^{(o)}$ returns a detection when a Δ_t threshold is overcame. $\Delta^{(o)}$ varies between 0 and 1. Being Δ_t less than 1 (in this case, it is set to 0.7), it is not necessary that all the cells are out-of-control. Another useful point of view is represented by the scatter plots drawn inside any cell. These plots represent the pairs of K_r and K_t summary



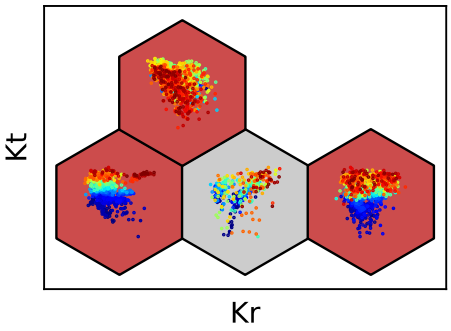
(a) *Test 1*



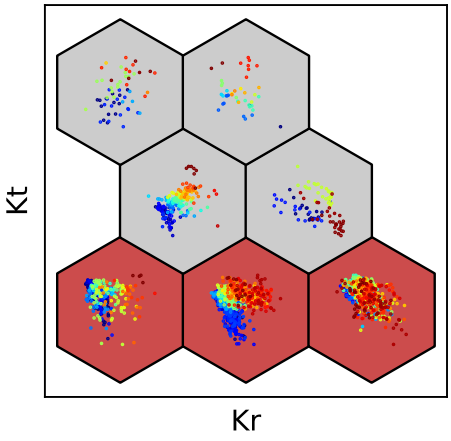
(b) *Test 2*



(c) *Test 3*



(d) *Test 4*



(e) *Test 5*

Figure 4.4: *Representation of the GSOM maps at the GSOM level out of control condition. Regions are represented by hexagonal cells. Grey background stands for in-control cells, red for out-of-control cells. 2D scatter plots represent the evolutions of the K_r and K_t coefficients in a normalised fashion. The colour represents the sample number from the first (blue) to the last (red).*

coefficients collected in the associated cell. The two coefficients are normalised and centred in the cell, thus no absolute values can be retrieved from the graph. At the same time, the scatter plots are drawn with progressive colours: blue points represent samples from the early stages of the cutting tool life, while red points refer to samples near the GSOM out-of-control detection. The scatter plots provide two levels of information. The first level is related to the efficacy of the region in describing the wear degradation. This can be noticed by a shift of the distribution of the two coefficients from the initial phases (blue), towards the cutting tool end-of-life (red). A strong separation within the data points represents a clear correlation with tool wear. This is also supported by the fact that, by construction, a cell can represent only a gradual progression of the coefficients. In general, diverging distribution of coefficients correlate to out-of-control cells, as expected. The frequency at which a cell is chosen as BMU represents a second level of information. The amount of times in which the cell was chosen as BMU is representative of the *experience* of the cell. In fact, frequently visited cells may be older than others, or more representative than other cells. It is important to notice that the $\Delta^{(o)}$ coefficient was thought to give more attention to cells with more experience. In this way, false positive and false negative cells have a lower effect on the overall performance of the GSOM. Some cells that present a meaningful trend of coefficients may be in-control because a significant number of expert cells already detected an out-of-control condition. The opposite case occurs in figure 4.4b, where the cell on top seems not to highlight a relevant shift in the coefficients, but it is still supported by expert out-of-control cells.

The results presented in the previous paragraph where obtained after an optimisation process of the GSOM hyper-parameters. More specifically, the results were optimised with respect to an average flank wear of $200\mu m$. This choice was related to the chosen cutting conditions following ISO standards indications [93] and scientific literature [4]. First of all, Ti_6Al_4V is a hard-to-cut titanium alloy. Cryogenic lubrication is a relatively new cooling technology, still more unstable than conventional media. Furthermore, some tests were performed at very high cutting speed. All these aspects support a more conservative choice of the flank wear threshold. The optimisation results are reported according to two different industrial scenarios: the first one, used for the above reported figures represents the optimisation of the approach over the full set of experiments; the second one represents the case where the solution is optimised for a machine tool and ported to another one.

4.3.1 Full set optimisation

As previously explained, the proposed GSOM algorithm is a completely unsupervised solution for automatically clustering high variability SFC estimations and detecting the correct end-of-life of the cutting tool. Here, the hyper-parameters of the GSOM are optimised in order to detect a tool with $200\mu m$ mean flank wear. An optimisation procedure over the full set of experimental run-to-failures is performed and the prediction error is measured on the full dataset. The optimisation method was performed through a two phases grid-search procedure [34]. A grid of parameters was selected and fully explored in order to reduce the searching space. A second grid was determined to find the combination leading to the minimum root mean squared percentage error (RMSPE, [22]). The hyper-parameter combinations of the grids are reported in

Table 4.2: *Grid-search hyper-parameters combinations for the two steps of optimisation. The optimisation algorithm tests all the combinations of hyper-parameter values.*

	η	R_t	n	N	H	Δ_t
1)	0.050, 0.075, 0.100	$\mathscr{F}(6, n-6)$, 6, 7, 8	10, 20, 50	10, 15, 30	3, 5, 10	0.3, 0.5, 0.7
2)	0.035, 0.045, 0.050	7, 8	10, 20, 50	30, 35	5, 7, 10	0.5, 0.6, 0.7

Table 4.3: *Relative percentage errors on the run-to-failure. Both scenarios are compared between the optimised GSOM and the optimised reference control chart (CC).*

Optimisation kind	Test 1	Test 2	Test 3	Test 4	Test 5
Full set (GSOM)	-23.0%	-5.2%	-2.6%	-9.1%	14.9%
Full set (CC)	-84.7%	-35.5%	-49.6%	-51.1%	-47.0%
Portability (GSOM)	-3.9% (TR)	1.77% (TR)	-2.8% (TR)	-40.6% (TE)	-27.1% (TE)
Portability (CC)	-86.3% (TR)	12.0% (TR)	-52.6% (TR)	-51.6% (TE)	-75.2% (TE)

table 4.2. The grid-search optimisation procedure consists in the evaluation of the full set of hyper-parameter combinations. The combination with the minimum RMSPE is chosen as the best. For this scenario, the best combination was: $\eta = 0.045$, $R_t = 7$, $n = 50$, $N = 35$, $H = 7$ and $\Delta_t = 0.7$. The algorithm was capable of predicting the end-of-life time with a RMSPE of 13.2%, with a tendency to underestimate the end-of-life of cutting tools. The relative percentage errors for the single tests are instead reported in the first row of table 4.3. In this scenario, the GSOM was capable to predict well almost all the tests. Only the first and last test presented a relative percentage error higher than 10%. The last cutting test remains the worst one, being overestimated with a relatively high error. Nevertheless, it must be underlined that the algorithm performed well despite the different cutting conditions of the tests. Furthermore, the solution is completely unsupervised, and the map is self generated along the evolution of each test. The results of the conceived approach were compared to the control chart presented in chapter 3, applied to the SFC estimated with multivariate linear regression, in order to make them comparable. The control chart was not applied with standard threshold and 15 samples as averaging dimension, as presented in chapter 3, but an optimisation of the two parameters was performed. Here, the control chart was optimised with respect to the full set of tests. The prediction results for the reference control chart showed an underestimation behaviour with a RMSPE of 56.1%. The predictions improvement of the GSOM were not negligible. Furthermore, in table 4.3 it is possible to compare the two algorithms on each test, by looking at the first two rows of the table. The GSOM improved the prediction performances on every test, even if on Test 5 the GSOM leads to an overestimation of the cutting tool life.

4.3.2 Portability scenario optimisation

This second scenario faces the case where the conceived solution is firstly optimised on the tests performed on a single machine tool (Flexi) and then, the algorithm needs to be used also for another machine tool. The proposed scenario allows to evaluate the portability of the conceived solution. The phases for the algorithm optimisation are the same and the associated hyper-parameter combinations can be found in table 4.2.

In this case, the algorithm is optimised in order to minimise the RMSPE on the first three run-to-failures, run on the Flexi machine tool. The performances of the algorithm are then tested on the fourth and fifth run-to-failures, performed on the M5 machine tool. The optimal combination of parameters was: $\eta = 0.035$, $R_t = 7$, $n = 50$, $N = 30$, $H = 10$ and $\Delta_t = 0.7$. The RMSPE on the optimised set of tests was found to be 2.9%, with a tendency to underestimate the cutting tool end-of-life. When testing the algorithm on the M5 set of run-to-failures, the predictions were underestimating the end-of-life with a RMSPE of 34.5%. The relative percentage errors for this scenario were reported in the third row of table 4.3. In this context, a really high performance was reached on the first three experiments. The algorithm was accurate with negligible errors. Of course, when testing the algorithm on unseen run-to-failures, the prediction errors rose. The algorithm underestimated both the tests. Nevertheless, the last two tests consisted of a double change in the cutting parameters. First of all, the machine tool was changed; secondly the cutting speed was almost doubled. These two changes increased the complexity of the estimations. Furthermore the optimisation set was constituted by a low number of tests. As for the previous scenario, the results of the conceived approach were compared to the control chart presented in chapter 3. The control chart was optimised with respect to the Flexi tests and tested on the M5 run-to-failures. An underestimation with RMSPE of 58.8% and 64.5 were obtained, respectively. Thus, GSOM led to dramatical improvements of the predictions. In table 4.3 the algorithm performances were compared on each test, by looking at the third and fourth rows. The GSOM improved the prediction performances on every test, both on Flexi and M5 machine tool.

4.4 Conclusions

In this chapter, a Growing-self Organising Map (GSOM) algorithm was introduced in order to perform tool condition monitoring in an unsupervised learning scenario. The algorithm was capable of managing the variability within the specific force coefficients (SFC) estimation, generated from the multicollinearity phenomenon and induced by unpredicted sources of variability, like machine tools dynamics dependent on the cutting direction or run-out. The conceived approach allowed to:

- cluster the SFC, by the automatic creation of regions with a similar behaviour of the coefficients. Thus, each region tends to monitor a gradual evolution of the SFC, while separating fast and sudden variations in their estimations.
- monitor the tool wear according to a voting system. Control charts are run inside any region and each of them contributes to a combined out-of-control indicator. The weight associated to each region is based on the region *experience*, i.e. based on its rate of being chosen as the best matching unit and the time of its creation. The voting system gave stronger weights to expert regions.
- outperform the prediction results of an optimised version of the chapter 3 control chart, in two different scenarios. The first scenario analysed the performances of the GSOM and the reference approach when optimised on the full set of available run-to-failures. The GSOM reached a RMSPE of 13.2%, generally underestimating the tool life, whereas the optimised control chart was capable to reach only

a 56.1% RMSPE, highlighting the improvements introduced by the algorithm. A second scenario tested the portability of the algorithm. Both the approaches were optimised on a machine tool and tested on run-to-failures performed on a different one. The GSOM RMSPE were of 2.9% and 34.5%, respectively, whereas the control chart application resulted in 58.8% and 64.5.

CHAPTER 5

Tool wear: hybrid adaptive direct tool wear prognosis

In previous chapters a robust tool monitoring approach based on process modelling was conceived. In this chapter, a hybrid prognosis tool is formulated in order to accomplish the estimation of cutting tool remaining useful life. Since cutting speed, and more generally process parameters, influences the rate of tool degradation, an adaptive prognosis strategy is presented on the basis of flank wear measurements. The proposed algorithm could also be adapted to integrate physical features like specific force coefficients with direct wear measurements.

5.1 Introduction

Digital manufacturing represents the synergy of production processes and data acquisition systems [147, 205]. Acquired data can be employed to support different phases of the production process. Prognostics and Health Management (PHM) is one of these. PHM consists in the monitoring, diagnosis and prognosis of machine faults [14, 20, 162]. Monitoring consists in the detection of abnormal states, i.e. deviations from normal operating conditions; diagnosis reflects in the classification of such deviations, performing the isolation and quantification of the abnormal states; at last, prognosis performs the prediction of abnormal states evolution up to a total failure [14].

Indeed, the main goal of prognosis is to predict the Remaining Useful Life (RUL) of faulty components, as well as the estimation of its probability density function (PDF) and thus, its uncertainty bounds [96]. The prediction of faults is beneficial both in terms of cost savings on materials and tools, but most importantly it allows to decrease maintenance time. When dealing with cutting tools, the estimation of failure before their occurrences may end up in saving up to 40% of maintenance costs [185] and about

20% of downtimes of machine tools [121, 227], responsible for production losses.

The focus of this chapter is on the prognostics of cutting tools for turning operations. Turning is a widespread technology that allow to produce axial symmetric solids of rotation. The turning process is a subtractive manufacturing technique which has been widely studied and modelled in the last decades. However, the possibility of developing adaptive strategies to estimate the remaining useful life of tools is a challenging task that has not been investigated deeply enough yet. Tool wear is a complex stochastic phenomenon [233]. In order to make forecasts of its evolution, it is necessary to define a significant indicator that represents it. Flank wear measurement is the most relevant and common degradation indicator for cutting tools [119, 156].

Two main branches of research deal with the prognosis of cutting tool wear: namely, indirect and direct tool wear approaches [60]. Indirect methods consist in the evaluation of tool wear degradation through the use of cutting process quantities. The commonly adopted variables include axis torque or current [104], vibrations [194], acoustic emissions [211] and cutting forces [156]. Main advantage of most indirect estimation techniques is that they could be developed to work in real-time, such that while machine is running, an updated wear indicator is available. However, all indirect methods are limited by the fact that there are lot of other effects greater in magnitude than tool wear. This makes indirect observation method still immature for general cutting conditions [14], even if it is possible to find some research works that try to deal with the normalisation of indirect quantities with respect to cutting parameters [47, 156]. On the other hand, direct measurement of flank wear is more robust against the cutting process variability. Direct inspection could be obtained through point wise scanning, profilometric acquisitions [151] or by 2D/3D calibrated pictures analysis [55, 139, 151]. Direct methods still have some shortcomings. Firstly, uncertainties are related to the accuracy of the inspection system and the post-processing of the measurements. Secondly, in general machining processes should be stopped and the cutting tool exposed to the inspection system in order to measure the tool-tip status. Nevertheless, the inspection task could be performed at periodic stops in masked time, for instance at tool changes, or triggered by user defined events according to optimised production strategies.

When dealing with prognosis approaches, several researches tried to apply machine learning or data-driven methods for the prediction of tool wear. Cheng *et al.* applied Support Vector Regression in order to predict flank wear evolution, starting from cutting forces, vibrational signals and machined surface pictures [34]. Guo *et al.* proposed a DenseNet algorithm in order to correlate flank wear to cutting forces, vibrations and acoustic emission signals. A multi-step encoder-decoder system predicted the short-term and long-term evolution of the flank wear [73]. A CNN was proposed by Zhang *et al.* to predict the RUL of cutting tools starting from current, vibrations and acoustic emission signals [227]. Wang *et al.* conceived a hybrid methodology based on machine learning capable of integrating heterogeneous data (process parameters, power profiles and tool wear images). Wear severity was assessed by a convolutional neural network, while RUL prediction were performed by a recurrent neural network [207]. Liu *et al.* proposed a cutting force model including tool wear information. The model-based approach could be used for tool flank predictions through cutting force and temperature measurements [140]. From statistical-based approaches, Yu *et al.* applied a weighted HMM framework for the prediction of tool flank wear, based on vibrations and acous-

tic emission signals [220]. Zhang *et al.* proposed a particle filter scenario including a linear degradation model. Its parameters are updated online from in-process vibrational measurements [224]. Scientific works were mapped in table 5.1, according to several classes, highlighting some useful aspects of the proposed methodologies like the availability of RUL PDF, the number of needed run-to-failures, and algorithm domain.

Despite the amount of research works on tool prognostics, several challenges are common between them:

- need of several run-to-failure experiments to train machine learning or data-driven models
- robustness of the solution with respect to cutting parameters
- prediction of the RUL PDF is missing
- algorithms are static, not updated through on-line measurements

Having in mind these challenges, a direct wear prognostic approach is here conceived. The solution is based on a hybrid adaptive algorithm, fusing the statistical framework with the data-driven world. In fact, a particle filter state observer is used to adapt on-line the weights and biases of a multi-layer perceptron, which maps the flank wear degradation curve with respect to time. The algorithm is assumed to take as input flank wear measurements, typical of direct tool condition monitoring systems.

The structure of this chapter is organised as follows: in section 5.2, the experimental set-up and campaign are firstly described, then the conceived adaptive and hybrid prognosis methodology is formulated, together with the description of the used performance metrics. In section 5.3 the analysis of the experimental data is reported together with its discussion. At last, conclusions are drawn in section 5.4.

5.2 Materials and methods

In this section, the experimental set-up and campaign are firstly explained; then all the elements constituting the conceived approach and the computed prognostics metrics are described, following the workflow of figure 5.1.

5.2.1 Experimental set-up and campaign

The experimental campaign consisted of a set of five run-to-failures in turning applications. The tests were performed on a SOMAB UNIMAB 400 lathe, equipped with an analogical numerical control. A carbide tool with a lead angle of 95° was adopted (ISO standard code: TNMG220404-M5 5625, tool radius equal to $r_\varepsilon = 0.4mm$, rake angle of 13° and a relief angle equal to 0°, with $Al_2O_3 - TiCN$ coatings) and fixed on tool holder, ISO code MTJNL2525M22. Hardened and tempered steel bars of UNI $39NiCrMo3$ were used to perform the wear tests. Run-to-failures were performed with conventional lubrication, using cutting fluid (oil-water emulsion with 5% of HOCUT 795 SC), adduced through nozzle on the cutting zone, in order to reproduce realistic industrial production scenarios.

Experimental tests were performed according to a full factorial design with one central point and no replicates (thus one test per corner and a central point test), following [153]. Two factors were chosen, namely cutting speed and feed per revolution.

Table 5.1: *Map of State-of-The-Art cutting tool prognosis approaches. Algorithms are described by application, employed sensors, measured features, algorithm and relative category, number of experimental run-to-failures, RUL pdf availability.*

Ref.	Machining process	Measurements	Features	Algorithm	Class	Varying cutting parameters	Run-to-failures	RUL pdf
[177]	Broaching	Cutting forces, strain, vibration	PCA on force covariance matrix	Support Vector Machine	Black-box	No	-	No
[10]	Drilling	Current	Energy of wavelet packet signal node	ARMA	Grey-box	Yes	2	No
[125]	Turning	Flank Wear, power consumption, vibrations	VB, power mean, vibration moving average (separately)	A sort of similarity method	Grey-box	No	30 (some of them discarded)	Yes
[223]	Milling	Vibrations	A set of time domain features	Neuro-fuzzy Neural Networks	Black-box	No	7800 (points)	No
[194]	Milling	Cutting forces, vibrations, acoustic emissions	Root mean square (rms), standard deviation (std), peak (forces); rms, kurtosis (accelerometers); mean, std (acoustic emission)	Dynamic Bayesian Networks	Grey-box & Black-box & experience-based	No	3	No
[133]	Milling	Cutting forces	Peak, amplitude, average, std	Fuzzy inference	Experience-based	-	-	-
[134]	Milling	Cutting forces	Time-domain features	Multiple regression models	Grey-box	No	2 (109600 points)	Yes
[101]	Milling	Cutting forces, vibration, acoustic emission	-	Artificial Neural Network (Extreme Learning Machine)	Black-box	No	1 (945 points)	No
[17]	Milling	Cutting forces, vibration, acoustic emission	Time-frequency domain features	Support Vector Regression	Black-box	No	2	Yes

Ref.	Machining process	Measurements	Features	Algorithm	Class	Varying cutting parameters	Run-to-failures	RUL pdf
[206]	Milling	Cutting forces	Energy from time-frequency domain features	Continuous Hidden Markov Models and Gaussian Process Regression	Grey-box	No	7	No
[100]	Milling	Cutting forces, vibrations	Four time domain features	Extreme learning machine	Black-box	No	1 (945 points)	Yes
[48]	Milling	Spindle power	Rms	Artificial Neural Network	Black-box	Yes	4	No
[164]	Milling	Cutting force, vibration, acoustic emission (workpiece side)	Energy features from wavelet transform and blind source separation	Non-linear regression with different model for each cutting tool tested	Grey-box	-	-	-
[214]	Milling	Cutting forces	Time and frequency domain features	Bayesian multi-layer perceptron	Grey-box & black-box	Yes	12	Yes
[220]	Milling	Cutting forces, vibrations, acoustic emissions	rms (accelerometers)	Weighted Hidden-Markov Models	Grey-box	No	3	No
[73]	Milling	Flank wear, cutting forces, vibrations, acoustic emission	Learned features (deep learning)	DenseNets and auto-encoder	Black-box	No	2	No
[207]	Milling	Spindle power, milled surface images, tool wear images	Learned features (deep learning)	Convolutional Neural Network and Recurrent Neural Network	Black-box	Yes	6300 images	No
[140]	Milling	Cutting forces and temperature	-	Tool wear model estimation from cutting forces and temperature	Model-based	Yes	9	No
[233]	Micro-milling	Flank wear measurements, cutting forces	-	Particle Filter and Long Short Term Memory network	Grey-box & black-box	Yes	9	Yes

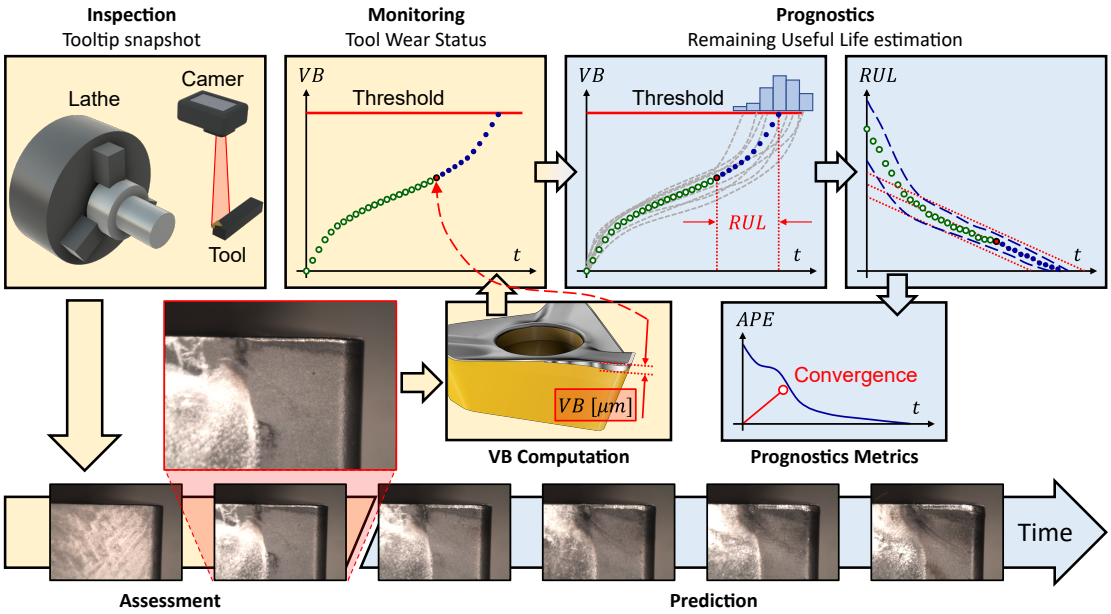


Figure 5.1: *Workflow for the conceived methodology. From tool pictures up to the estimation of the cutting tool Remaining Useful Life.*

Table 5.2: *Run-to-failure turning tests process parameters: cutting speed (v_c) and feed rate (f) are the two factors of the full-factorial design. Tool life (TL) is also reported.*

ID	$v_c\,[m/min]$	$f\,[mm/turn]$	TL $[s]$
1	220	0.1	765.6
2	220	0.3	345.6
3	190	0.2	798.2
4	160	0.1	2736.5
5	160	0.3	1062.1

These two parameters were chosen since they have strong influences on the cutting tool lives [107]. The factor levels for the run-to-failures were reported in table 5.2. The radial depth of cut was set to $2mm$. Direct measurement of tool flank wear were performed through recurrent visual inspection. Starting from calibrated macro pictures of cutting edges, the VB was computed with a manual procedure. Anyway, this step doesn't limit the application of the conceived prognostic approach, since similar results can be obtained through machine vision algorithms, as in [55, 139, 151]. It is important to have high quality pictures of clean cutting edges and properly define a region of interest. Then, thresholding, contouring, aligning and measuring should be performed. A Stereomicroscope Optika SZN-T with Motic SMZ-168T support was used to take flank wear width (VB) images during the cutting tests. The end-of-life of a cutting edge was set to a $VB = 0.15mm$. The corresponding end-of-life-time of the tool (TL) was reported in table 5.2 (fig. 5.2). The threshold was lower than what is suggested by ISO 3685 [93]. Anyway, the threshold was chosen in order to avoid too high dispersion of the results in terms of insert duration, seen in preliminary tests. This choice was taken to limit the experimental effort, too. In fact, carrying out many experiments up to

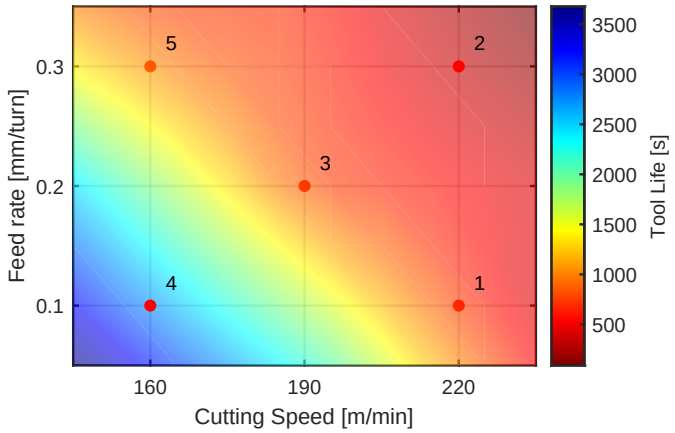


Figure 5.2: *Graphical representation of the design of experiments and tool life as a function of the selected cutting parameters.*

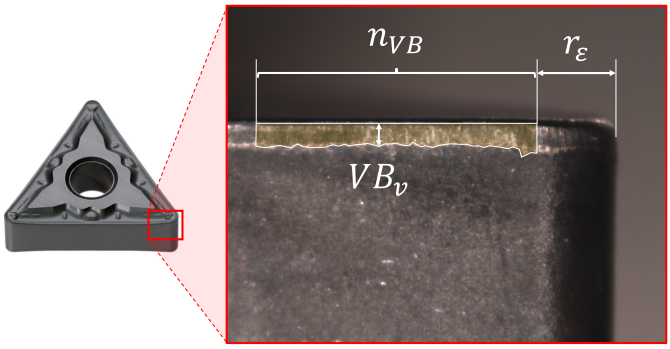


Figure 5.3: *Main quantities involved in the flank wear computation.*

a VB of $0.3mm$ would have required too many machining tool hours, considering the fact that flank wear measurements take an amount of time comparable or higher than the effective machining time [211]. Optimal cutting speeds, feed rates were adopted under conventional lubrication of steel cutting, which would have lead to unsustainable experimental costs. The computation of the VB is performed following international standards [93] (fig. 5.3). [t] As can be seen from figure 5.3, the flank wear width determination starts from a set of n_{VB} local measurements VB_v. The region of interest starts at a distance from the edge equal to the tool radius. Each measurements accounts for a differential width Δl_v. Thus, the average flank wear VB can be computed over the region of interest by eq. (5.1):

$$VB = \frac{1}{n_{VB}} \sum_{v=1}^{n_{VB}} VB_v \Delta l_v \tag{5.1}$$

5.2.2 PF-MLP definition

In this section, the hybrid (statistical-based and data-driven) adaptive approach for tool wear prognosis is presented. The solution is based on a combination of particle filter (PF) state observer, to set-up a bayesian update framework for a set of multi-layer perceptrons (MLP). The approach (in similar fashions) was investigated in other applica-

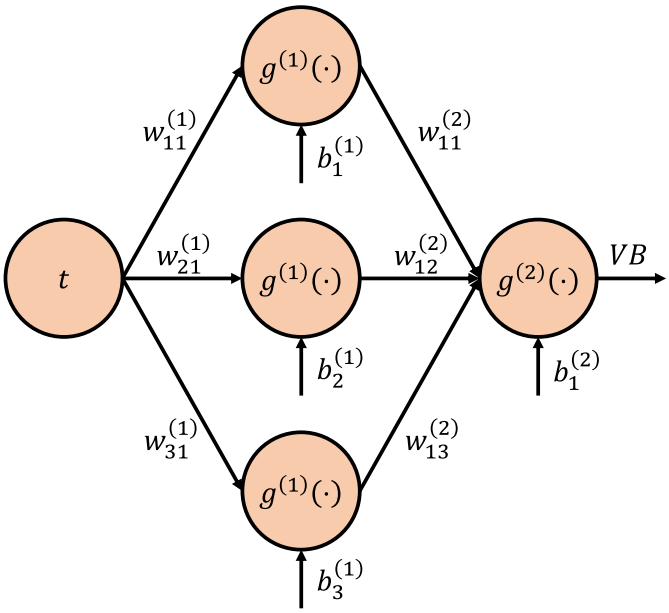


Figure 5.4: *Proposed MLP architecture. The input neuron propagates time values. The hidden layer features three neurons, while VB is the result of the output layer.*

tions (such as crack growth prediction and lithium-ion batteries prognostics), highlighting its interesting adaptive capabilities [27,175]. Furthermore, this algorithm allows for the prediction of the cutting tool RUL PDF, required by the international standards [96] but rarely available on state-of-art approaches. In order to present the approach it is necessary to start with the description of the MLP.

Multi-layer Perceptron

In the conceived scheme, the MLP represents a non-linear mapping of the VB evolution with respect to time. MLP was chosen due to its demonstrated universal fitting capabilities [41] but other functions (such as polynomials) may be used. The MLP features a brain-inspired architecture, with an input layer, at least one hidden layer and an output layer. In this case, the complexity of the network is reduced at minimum: the input and output layers are composed of one single neuron; only one hidden layer is present composed of three neurons. The structure and nomenclature of the MLP is shown in figure 5.4. MLP has a feed-forward structure, i.e. the input of the network (time t) is propagated through the layers, undergoing several processing operations. t may be also seen as the output of the input layer. The net input $n_i^{(l)}$ of a neuron i in layer l is computed through eq. (5.2):

$$n_i^{(l)} = \sum_{j=1}^{n_j} \left(w_{ij}^{(l)} o_j^{(l-1)} + b_i^{(l)} \right) \tag{5.2}$$

where j indicates the neuron of the preceding layer $l-1$; $w_{ij}^{(l)}$ is the weight of layer l, that multiplies the output $o_j^{(l-1)}$ of the previous layer $l-1$; b_i is the i-th neuron bias. Then, the net input is passed through the activation function $g^{(i)}(\cdot)$ obtaining the layer

output $o_i^{(l)}$ (eq. (5.3)):

$$o_i^{(l)} = g^{(l)}(n_i^{(l)}) \tag{5.3}$$

In the conceived scheme two different activation functions are used, depending on the layer. In the hidden layer, Elliot-sigmoid function (eq. (5.4)) was used [52]:

$$g^{(1)}(u') = \frac{u'}{1+|u'|} \tag{5.4}$$

where u' is a generic variable. In the output layer a pure linear (i.e., identity function) was used instead. Thus, considering the fact that first layer input (and output) is time t and the output of the last layer is the flank wear VB, the MLP non-linear function can be written as eq. (5.5):

$$VB(t) = \sum_{j=1}^{3} w_{1j}^{(2)} \cdot g^{(1)}(w_{j1}^{(1)} t + b_j^{(1)}) + b_1^{(2)} \tag{5.5}$$

Eq. (5.5) represents a non-linear parametric function of time, whose parameters are the weights and biases of the MLP. Weights and biases are randomly initialised. During the application of the algorithm, these parameters will be trained upon a single tool wear degradation curve.

Particle Filter

PF is a state observer able to relax all the Kalman Filter main assumptions. In Kalman Filter the distributional properties of the state of a dynamical system is assumed to follow a Gaussian distribution; the process disturbances and measurement noise are assumed to follow a Gaussian distribution, too; the system dynamics and observation processes are assumed to be linear [14, 54, 64]. PF removes these assumptions allowing for non-linear dynamical systems and non-gaussian states, disturbances and noises [14, 203]. In the proposed prognostics scheme, the MLP describes a possible degradation curve of the tool (describing the flank wear evolution in time). Although the degradation curve is initialised upon a historical run-to-failure, it may not represent the actual degradation of the current cutting tool. Thus, the MLP architecture should be updated with online measurements of the cutting tool flank wear. In order to perform such operation, the weights and biases are thus assumed to be states of a dynamical system, capable of evolving during time (eq. (5.6)):

$$\mathbf{x}_{k+1} = \mathbf{x}_k + \boldsymbol{\omega}_k \tag{5.6}$$

where the index k represents the discrete time step; $\mathbf{x}_k$ is the state vector at the current discrete time step, containing the whole set of weight and biases of the MLP; $\boldsymbol{\omega}_k$ is a vector of random variables assumed to have variance proportional to the corresponding weight or bias $\boldsymbol{\omega}_k \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_k)$; where the s-th element of $\boldsymbol{\Sigma}_k$ diagonal is $Q \cdot x_{k,s}$; Q represents disturbance intensity. Q is a hyper-parameter in PF framework: it should be large enough to explore all the possible degradation curves, but limited in order not to generate unrealistic degradation curves. Eq. (5.6) is referred to as *process equation*.

The observation process should associate the state vector to the system measurements, i.e. in this case, flank wear. Thus, the observation equation is represented by

the MLP function (eq. (5.5)). The MLP function will be now referred to as h in the observation equation (eq. (5.7)):

$$y_k = h(t_k, \mathbf{x}_k) + \eta_k \tag{5.7}$$

where y_k is the predicted flank wear through MLP; t_k is the current cutting time; η_k is the measurement noise, following a normal distribution with null mean and standard deviation equal to R.

Objective of the PF is the description of the posterior PDF of the MLP parameters, conditioned on the observed data $p(\mathbf{x}_k|y_{0:k})$ [27].

The PDF of the MLP parameters is approximated by a set of N_p samples, referred as *particles*. In practice, a particle, represents a realisation of the stochastic evolution of a MLP. Thus, in the PF framework, N_p MLP are generated, each of them creating a different map between time and VB. A PF iteration consists in two subsequent phases: a prediction phase and an update phase. The prediction phase consists in the estimation of the new particle values according to the process equation. For each particle, a realisation of the process disturbances is computed to explore the MLP feasible maps. Indeed, at the end of this step a full set of new MLP is obtained. This step doesn't make any use of information or measurement from the field, thus the new set of particles represents the "a priori" distribution of the MLP parameters.

The update phase is implemented to improve the PDF of the particles. A likelihood equation $L(y_{0:k}|\mathbf{x}_k)$ is computed for each particle by keeping in consideration all the measurements up to the discrete time step k (eq. (5.8)):

$$L(y_{0:k}|\mathbf{x}_k) = \exp\left(-\frac{1}{2R^2}\sum_{z=0}^{k}\left(y_z - h(t_z, \mathbf{x}_z)^2\right)\right) \tag{5.8}$$

where z is the index used to represent each discrete time step up to k. This equation expresses the likelihood of observing the measured VB curve, depending on the state $\mathbf{x}_k$ described by a particle. Thus, the likelihood help us to determine how much should we trust that particular particle (i.e, that particular MLP). Indeed, a weight $u_{k,s}$ is given to a particle s, equal to the associated measurement likelihood. Then, each particle weight is normalised such that they sum up to 1 (eq. (5.9)):

$$u_{k,s} = \frac{L(y_{0:k}|\mathbf{x}_{k,s})}{\sum_{s=1}^{N_p} L(y_{0:k}|\mathbf{x}_{k,s})} \tag{5.9}$$

where $\mathbf{x}_{k,s}$ is the state represented by particle s. Particle weights represent a probability density function for particles, since they are now summing up to 1. The resampling stage is then performed: the cumulative density function (CDF) of particle weights $U_{k,s}$ is generated through eq. (5.10):

$$U_{k,s} = \sum_{z'=1}^{s} u_{k,s} \tag{5.10}$$

where index z' is used here to represent each particle up to the s-th one. A new set of particles is obtained from the CDF $U_{k,s}$, through Monte-Carlo sampling. Monte-Carlo resampling stage works as follows: a random number r is drawn from a random

uniform distribution $\mathscr{U}(0;1]$; the s-th particle for which $U_{k,s}$ overcomes r is chosen; the extraction of r and the identification of the particle to be redrawn are repeated until a full set of new particles is extracted. The new set of particles represents the discrete approximation of the posterior PDF $p(\mathbf{x}_k|y_{0:k})$ of the MLP parameters. At this point the PF iteration is finished. The prediction and update stages are repeated every time a new flank wear measurement is available.

Cutting tool RUL prediction

At each time instant t_k, the diagnosis phase can be performed after the update phase. Each particle provides the estimate for the cutting tool flank wear $VB_{k,s}$, by evaluating the associated MLP curve at t_k. Thus, the full set of particles provides the approximated posterior PDF of the flank wear $p(VB_k|y_{0:k})$. In order to realise the final objective of PHM, prognosis must be integrated. With this aim a threshold value VB_t for the limit flank wear was set to $120\mu m$. This choice was taken in order to guarantee MLPs to describe increasing fitting curves above the threshold. In fact, MLP fitting curves are summations of sigmoid-like functions which have a horizontal asymptote. If MLP curves do not overcome the threshold, infinite RUL would be predicted, causing convergence problems. In an industrial application scenario, the training test should be run to reach a VB slightly higher with respect to the desired limit threshold for the tool. At every discrete time step k, after the update phase, it is possible to compute the estimated time of end-of-life $t_{EoL,k,s}$ of each particle s. This is performed by evaluating the associated MLP function for future values of time, until the predicted VB intersects the threshold line. The time at which the MLP non-linear map crosses the threshold gives the $t_{EoL,k,s}$ estimate. The set of estimated end-of-life times, constitutes the approximated probability density function $p(t_{EoL,k}|y_{0:k})$. RUL is computed as the difference between the time of end-of-life estimate $t_{EoL,k,s}$ and the prediction instant t_k, for every particle. As for the time of end-of-life, the set of RUL estimations gives the approximated posterior probability $p(RUL_k|y_{0:k})$.

5.2.3 Prognostic metrics

Prognostics metrics are a valuable tool for measuring the performances of a prognosis approach. In addition to quantitative assessment they also offer a comprehensive visual perspective that can be used in designing the prognostic system. Furthermore, they allow to compare the performances of an algorithm with respect to others. Here, three main prognostics metrics [173] are used to analyse the behaviour of the conceived approach on the experimental tests and to compare the performances across the cases.

The first metric is the absolute prediction error (APE). APE is computed as the absolute value of the difference between true RUL and the expected value of the approximated posterior RUL PDF. APE is here computed as a function of the normalised cutting tool life (λ). APE formula is reported in equation (5.11):

$$APE(\lambda) = |RUL_{true} - E[RUL_k]| \tag{5.11}$$

where $\lambda = t_k/t_{EoL,true}$. APE furnishes a direct measure of the RUL prediction error during the cutting tests. Here, it is used in place of relative accuracy, since RUL_{true} tends to zero for λ close to 1, leading by definition to a null accuracy. The second

Table 5.3: *Chosen hyper-parameters for the conceived approach: number of MLP corresponds to the number of particles; number of neurons per layer; R is the measurement noise standard deviation; Q is the process disturbances intensity; VB_t is the prognosis threshold; α is the admissible prediction error; β is the minimum acceptable probability mass.*

No. MLP [−]	MLP neurons [−]	R [−]	Q [−]	VB_t [μm]	α [−]	β [−]
250	1-3-1	3	10^{-2}	120	0.2	0.2

metric of interest is the prognostics horizon (PH) [173]. PH has been proposed in two fashions. Here, the probabilistic one is computed[1]. PH is defined by equation (5.12):

$$PH = t_{EoL,true} - t_{k_{\alpha\beta}} \tag{5.12}$$

where $t_{k_{\alpha\beta}}$ is defined as:

$$t_{k_{\alpha\beta}} = \min_k \int_{\alpha^-}^{\alpha^+} p(RUL_k, y_{0:k})dt \geq \beta \tag{5.13}$$

where $\alpha^- = t_{EoL,true}(1-\alpha)$ and $\alpha^+ = t_{EoL,true}(1+\alpha)$, with α representing the allowable prediction error; β is the minimum admissible probability mass. Here, PH is normalised with respect to the $t_{EoL,true}$. Indeed, PH represents the percentage of tool life at which the required prognostics accuracy is met. The higher PH is, the more time is available to intervene before the tool failure.

The last prognosis metric is convergence (C). Convergence is a meta-metric quantifying the speed of decay of another prognosis metric [173]. Here, the convergence of the APE is proposed. It is defined by equation (5.14):

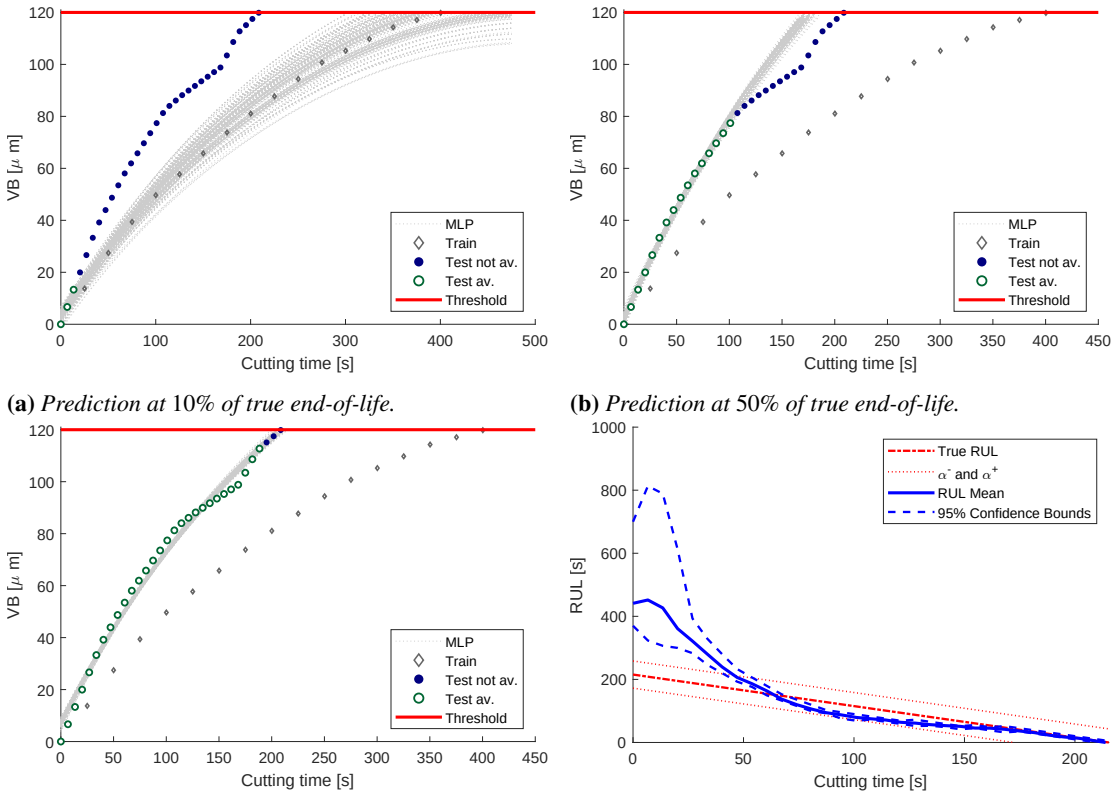
$$C_{APE} = \sqrt{APE_x^2 + APE_y^2} \tag{5.14}$$

where APE_x and APE_y are the x and y coordinates of the centre of mass of the area under the $APE(\lambda)$ curve. C_{APE} thus represents the euclidean distance from the origin of such centre of mass. The closer is this point with respect to the origin, the faster is the convergence of its metric (i.e., APE). Convergence is effective only for decreasing metrics [173].

5.3 Results and discussion

Results are presented in two phases. In a first section, the adaptability of the conceived approach, RUL estimation and metrics will be discussed for a single train-test run-to-failures combination. The reference case is when the MLP is initialised on test 5 and the algorithm is applied for cutting tool prognosis on test 2. A second section will regard the discussion of the whole set of results, thus all the combination of train and test run-to-failures. In both the cases, the hyper-parameters selected for the algorithm are reported in table 5.3.

[1]This version of PH can be adopted only for prognosis algorithms providing PDF estimates of the RUL. A deterministic version of PH is described in [173] for algorithms providing only RUL point estimates, too.



(a) *Prediction at* 10% *of true end-of-life.*

(b) *Prediction at* 50% *of true end-of-life.*

(c) *Prediction at* 90% *of true end-of-life.*

(d) *RUL estimate evolution.*

Figure 5.5: *RUL predictions when training on test 5, and testing on test 2. (a), (b) and (c) figures represent the adaptability of the prognosis approach. Grey diamond points represent the training data for MLP initialisation. Green circles are the already available flank measurements. Blue points are the upcoming experimental points. Red line is the end-of-life threshold. Grey dotted lines are the MLP non-linear functions generated by each particle. (d) represents the evolution of the estimated RUL (blue solid line) and associated confidence bounds (dashed blue line), compared to the true RUL (solid red line). Red dotted lines describe the allowable prediction bounds.*

5.3.1 Single train-test combination

In this section, the performances of the algorithm and the prognostics metrics are analysed for a reference combination of train-test run-to-failures. More specifically, the train test is 5 and the test run-to-failure is 2. This particular combination, allows to clearly identify the adaptivity characteristic of the conceived approach and to clearly discuss the adopted metrics. The discussion of this section makes reference to figure 5.5. The MLP configuration and evolution along the run-to-failure, as a consequence of the PF update, are presented at three different instants, i.e. 10% (fig. 5.5a), 50% (fig. 5.5b) and 90% (fig. 5.5c) of the testing cutting tool life. This choice was taken for conciseness reasons, while keeping enough information content. First of all, in fig. 5.5a, it is possible to see the difference between the training and test flank wear evolutions (diamonds and circular data points, respectively). Such difference is associated to the cutting parameters adopted for the cutting tests: training test 5 was performed at a high feed rate and low cutting speed, whereas test 2 was performed at high cutting speed and high feed rate. As expected run-to-failure 2 is faster than 5. At the top of the

graph, a solid red line represents the selected prognosis threshold, thus the flank wear forecast objective for prognosis. Diamond grey points were used to initially train the MLP. Thus, in a practical scenario, diamond data points constitute the only historical run-to-failure needed to train and run the approach. Green circles represent the data points received by PF from the field up to the current prediction time. As long as the run-to-failure goes on, more experimental data (blue points) will become available. The grey dotted lines show the MLP curves associated to the particle distributions. 10% of cutting tool life, represents a condition in which the algorithm is still very confident about the initialisation phase. Thus, the set of MLP faithfully pursuit the training run-to-failure data. Nevertheless, the process equation disturbances are already trying to further explore the MLP fitting space. Being the prediction time so early in the tool life, the spread of the MLP curves at the threshold flank wear is quite big. By sectioning the MLP curves with a vertical line at the prediction time (corresponding to the last green circle), and representing their intersections through a histogram, would provide the graphical representation of the approximated posterior PDF of flank wear $p(VB_k|y_{0:k})$. The same can be performed with the intersection of MLP curves with the threshold line, providing a visual representation of estimated RUL PDF at the current prediction time $p(RUL_k|y_{0:k})$. As long as time passes, more experimental data are exploited by the algorithm to adapt to the current tool degradation rate. In fact, at 50% of tool life (5.5b), the MLP curves are already laying on the experimental data, better representing the new degradation rate of the tool. Nevertheless, the cutting tool will face a sudden increase in the speed of degradation, as described by future experimental data (blue points). The algorithm is thus underestimating cutting tool RUL. It must be noted that the algorithm started growing more confident about its predictions, allowing for more similar MLP curves. Near the end-of-life of the cutting tool (5.5c), the algorithm is fully adapted to the tool wear trend and correctly predicts its RUL. This evolving behaviour of the algorithm is summarised in figure 5.5d. Here, main quantities related to $p(RUL_k|y_{0:k})$ are drawn as a function of k: namely, RUL estimate expected value (blue solid line), 2.5% and 97.5% percentiles, representing the 95% confidence bounds of the predictions (dashed blue lines). Their evolution is compared to the ground truth (true RUL red dash-dot line). Of course, the same behaviour explained above can be found here. An initial low confidence overestimating phase can be found up to 50s of cutting tool life; an underestimation phase between 50s and 180s; the final accurate RUL prediction for a cutting time bigger than 180s.

As previously explained, prognostic metrics allow to quantify the performances of the algortihm, compare different algorithms and different scenarios. Here, the metrics are computed with respect to the reference train-test combination. The first metric ($APE(\lambda)$) is reported in figure 5.6. APE gives an immediate visual idea on how the algorithm is converging to the ground truth RUL. The same phases described above can be identified here. A first part of the run-to-failure, where the algorithm was still confident about the training data can be found in the first 25% of cutting life ($\lambda = 0.25$). A second phase where the algorithm underestimated the RUL can be found in the APE hill between $\lambda = 0.25$ and $\lambda = 0.8$, keeping in mind that APE is the absolute value of prediction errors. In the last part of the tool life, where $\lambda > 0.8$, the prediction error almost became null. The convergence of APE is also graphically represented. Its value is equal to 71.9. This value provides a benchmark for the convergence time of

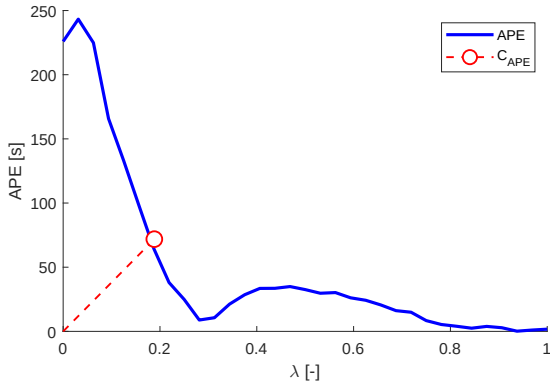


Figure 5.6: *Absolute prediction error APE (blue solid line) as a function of the normalised cutting time λ. Its convergence is represented with the red circle and red dashed line. This is the case in which training was performed on test 5 and validation on test 2.*

the algorithm in case other algorithms are tested on the same test. In the next section, convergence will be discussed much further in order to compare different train-test scenarios. The prognosis horizon is partially represented in figure 5.5d. In fact, the two red dotted lines represent the allowable prediction bounds, determined by α. PH is defined as the percentage of cutting tool life in which the beta criterion in eq. 5.13 is respected. In this case, the criterion is satisfied from 50s up to the cutting tool end-of-life. PH is thus equal to 81% of cutting tool life. This means that when the 19% of cutting tool life is expired, the algorithm starts to correctly predict the RUL. If PH is high, more time for preparing a maintenance action is available.

5.3.2 Full set performances

In this section, a cross-validation inspired approach is presented, with some modifications. Since the objective of the conceived algorithm is to reduce as much as possible the experimental effort to perform prognostics, the results presentation is carried out on a grid of cases. The grid is composed by different combinations of train and test run-to-failures. Thus, at each iteration, one of the test in table 5.2 is used for MLP initialisation, while the others are used to discuss the adaptivity performances of the hybrid methodology. In this sense, the proposed validation approach differs from Leave-One-Out Cross-Validation (LOO-CV), since LOO-CV consists of iteratively training the algorithm on all the tests but one (used for validation). Anyway, here, the idea is to analyse how the algorithm works with different training conditions and also suggest for the training run-to-failure best cutting parameters choice.

Two figures are reported representing this scenario: the first one reports the RUL predictions, the bounds for PH computation, true RUL and flank wear data for the training and test run-to-failures (fig. 5.7); the second one is the summary of prognostics metrics for the whole set of combinations (fig. 5.8). In both graphs the training run-to-failure number corresponds to the row index, while the test one to the column index.

The first results to be observed, are the ones on the main diagonal of figure 5.7. These represent the cases where the train test is the same as the validation one. It can be seen that, as expected, the RUL predictions are accurate and the confidence

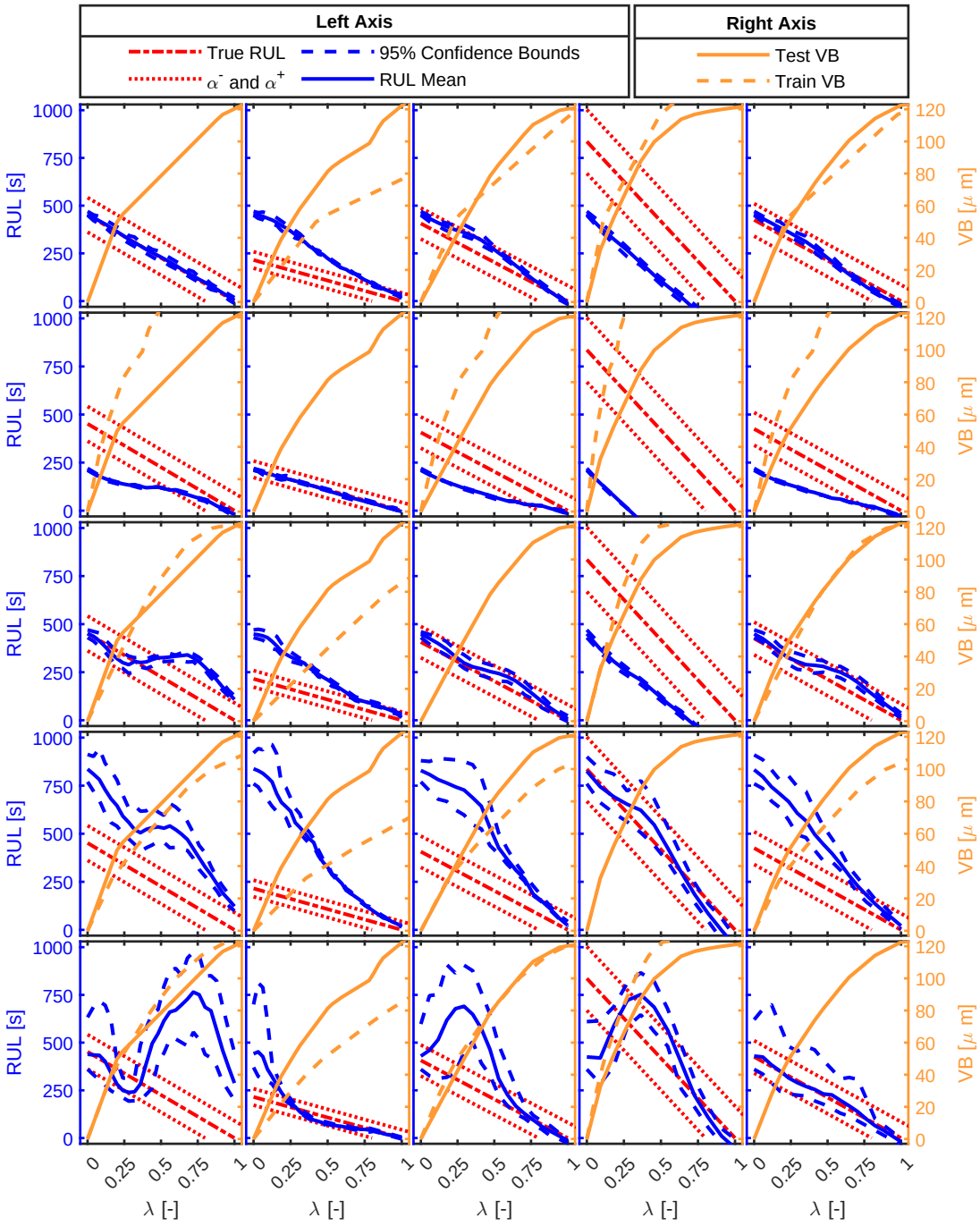


Figure 5.7: *Row index corresponds to the train test number, whereas column index to the test run-to-failure number. RUL predictions (blue solid lines) and their confidence bounds (blue dashed lines) are compared to the true RUL (red dash-dot line) and the allowed prediction bounds (dotted lines). Training (orange dashed line) and test (orange solid line) flank wear measurements are also compared.*

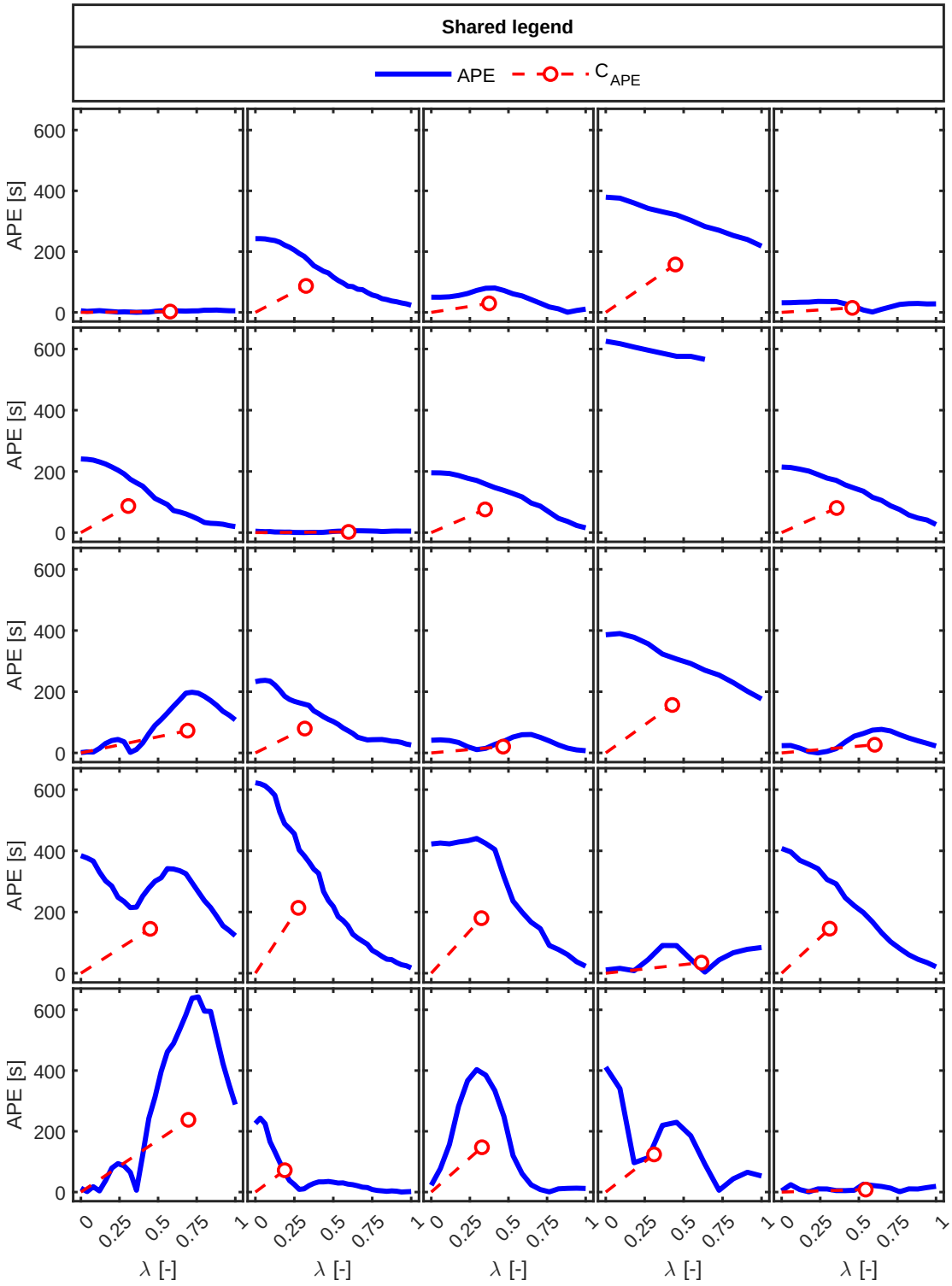


Figure 5.8: *Row index corresponds to the train test number, whereas column index to the test run-to-failure number. Prognostic metrics ($APE(\lambda)$ and C_{APE}) are represented for each train-test combination.*

Table 5.4: *Run-to-failure turning tests process parameters: cutting speed (v_c) and feed rate (f) are the two factors of the full-factorial design. Tool life (TL) is also reported.*

(a) *Prognostics horizon.*

		Test				
		1	2	3	4	5
Train	1	100%	19%	100%	0%	100%
	2	44%	100%	29%	0%	29%
	3	0%	19%	100%	0%	100%
	4	0%	13%	24%	100%	29%
	5	0%	81%	47%	45%	100%

(b) *Convergence.*

		Test				
		1	2	3	4	5
Train	1	2.7	87.1	29.5	157.5	14.8
	2	86.9	2.3	75.8	∞	80.5
	3	72.8	79.6	20.6	156.7	26.7
	4	145.0	213.6	179.9	34.7	145.6
	5	237.9	71.9	147.5	124.0	8.2

bounds are almost completely enclosed in the allowable prediction bounds. This is confirmed by table 5.4a, where the prognosis horizons of all the diagonal terms are 100%, meaning that from the beginning of the test, RUL is computed correctly. Also from the convergence of APE, it is possible to see that diagonal elements are minimum ones (tab. 5.4b).

The second possible analysis regards the identification of critical tests and the associated cutting parameters. In order to perform such analysis it is necessary to look at a column per time. In fact, a column represents a validation test in all the training scenarios. By looking at figure 5.7, two columns only present more difficulties in the adaptivity of the prognosis approach: namely, column 1 and column 4. The first column is well estimated when training is performed with the first two tests. Predictions become less accurate with training on test 3 and 4, while creating instability when training on test 5. This behaviour is highlighted even from the metrics perspectives 5.8, where APE tends to remain almost constant (training on 3 and 4), or diverges (training on 5). Column 1 of table 5.4a provides the same information, telling that for these three training cases, test 1 represents a critical condition from the prognosis horizon perspective. Convergence (table 5.4b) faces an increasing value, reaching the global maximum exactly when training is performed on test 5. Thus, high cutting speed tests and low feed rates tend to create adaptability problems.

Test 4 represents instead the worst case. In three training cases (1, 2 and 3) the algorithm is not capable of following the true RUL trajectory. Convergence of the APE provides the same information (table 5.4b), resulting in really high values (close to the global maximum) and even a non-convergent result (training on test 2). Furthermore, differently from test 1, the algorithm predictions seem to be rigid, with a null or small adaptation rate. This phenomenon is explained by the flank wear evolution visible

in fig. 5.7, where the test VB features constant degradation rate for the main part of tool life (up to $100\mu m$), when a sudden change in the rate is found. For the main part of the tool life, the training and test curves almost overlap, telling the algorithm that the current MLP curves are really good predictors and no adaptation is required. Despite this, the last parts of MLP curves (the ones responsible for RUL predictions) are dramatically underestimating tool life. Nevertheless, these data points are not available to the algorithm until the end-of-life is near, not giving enough time to the PF-MLP algorithm to adapt to the new rate. Test 4 was performed at low feed rate and low cutting speed. The results seem to converge to the fact that longer run-to-failures (low feed rates and low cutting speeds) are more critical to adapt to. No other critical tests were found.

The last analysis to be performed is the selection of the best training experiment case. From a practical perspective, this analysis helps a company to decide how to select the parameters for the training run-to-failure(s) to be performed. To this objective, it is necessary to have a look at figures 5.7, 5.8 and table 5.4 row-wise. In fact, once the row is fixed, the corresponding training test is set and the performances on all the other tests are verified. By looking at table 5.4a, rows 3 and 4 are excluded from being the best options. Training on test 3 is the only choice that leads to two null PH. Thus, it it is the only scenario in which two cutting tool RUL cannot be predicted correctly in advance. Furthermore, it presents also $PH = 19\%$ for test 2, which is one of the worst PH found in the whole set of combinations. Training on test 4 features only one critical test (test 1), but all the PH assume low values. On the contrary, within the scenario with test 1 as training sample, the algorithm performs really well on three run-to-failures. In these cases the RUL predictions are accurate right from the beginning. The same is confirmed by the RUL histories in fig. 5.7, and APE in fig. 5.8. The second best scenario is represented by row 5. Here, the algorithm performs fairly well on all the run-to-failures but test 1. PH assume quite high values (table 5.4a), all greater then 45% (which still means that RUL prediction is reliable at half of the tool life, leaving enough room for maintenance actions). Test 2 and test 5 presented really high PH. Row 5 seems to represent the context in which the prognosis algorithm adaptivity is enhanced (higher differences between training and test flank wear evolutions), even if PH are in general smaller with respect to case 1. Nevertheless, case 1 represents a safer option since the algorithm, at worst, underestimates the tool life. Test 1 and 5 represent intermediate conditions in which high feed rate or high cutting speeds are adopted, providing some indications about which cutting parameters should be adopted in order to train the PF-MLP algorithm.

5.4 Conclusions

The conceived approach allowed to develop a Prognostics and Health Management framework for tool wear monitoring and prognosis. A hybrid solution was conceived as a combination of the statistical world and the artificial intelligence world. From this perspective, the resulting methodology provided advantages with respect to the original separated components. Particle filter led to two benefits with respect to a standard multi-layer perceptron implementation: the capability to predict the remaining useful life in a statistical fashion, providing its probability density function estimation; the

adaptability of the multi-layer perceptron to the ongoing flank wear degradation, in opposition to its typical static implementation [14]. On the other hand, multi-layer perceptron increased the flexibility of particle filter for remaining useful life prediction. In fact, in order to implement particle filter for online prognosis, a dynamical degradation model must be known for the particular application. A combined implementation of particle filter and multi-layer perceptron doesn't require any dynamical formulation of the degradation process.

The main results obtained through the implementation of the proposed approach regards:

- a reduced experimental effort for the training dataset. The conceived approach was trained on a single run-to-failure experiment in order to test its adaptability. The algorithm was capable to follow online new degradation rates of cutting tools. In an industrial application scenario, the performances of the algorithm can be improved by adding new MLP initialisations to the initial particles, as long as cutting tools are worn out. This can be done also for a hybrid (direct and indirect) scenario, in which signal features (like specific force coefficients) are passed as inputs of the MLP. In fact, if a direct measurement tool is available to estimate flank wear, it is possible to automatically label indirect features. The same is not possible within a fully indirect approach.
- generalisation with respect to different cutting conditions. The adaptability feature of the conceived approach allowed to perform prognosis with fairly well performances on all the run-to-failures carried out with unseen cutting parameter combinations, limiting the needed training set dimension.
- estimation of the remaining useful life probability density function: specification required by international standards [96], but rarely satisfied in practical applications. This allowed to compute more robust prognosis metrics (prognostics horizon, [173]), as well as to support maintenance decision making, since remaining useful life prediction bounds were available, too [14].
- optimal remaining useful life prediction for already seen cutting parameters. The prognostics horizon for equal train and test data was always equal to 100%. The remaining useful life prediction was thus reliable for the whole cutting tool life.
- great adaptability to other cutting parameters. A combination of many train-test set highlighted the good performances on unseen tests. Only 6 combinations out of 25 train-test run-to-failure pairs didn't allow to correctly estimate tool remaining useful life. Three of them resulted in tool life underestimations (i.e. false positives), which represented a safer error condition with respect to overestimation. Only three PH fell under 20%. It is reasonable to expect the algorithm to perform well within the cutting parameters suggested by the tool manufacturer. Thus, the algorithm is expected to work properly when no abrupt wear phenomena occur (i.e., when the tool is used within the suggested technological region of applicability).

Regarding future works, an interesting challenge could be to join an adaptive real-time in process indirect estimation algorithm with a pit-stop direct measurement technique such that Prognostics and Health Management could be responsive and more

robust. Furthermore, the inclusion of process parameter data in the multi-layer perceptron architecture could increase the reliability of the predictions.

CHAPTER 6

Machine tool auxiliaries prognostics

Within this chapter, the focus is directed to machine tool auxiliaries prognostics and health management. Here, the objective is the development of a prognosis framework, not requiring any experimental run-to-failure for algorithm training. With this goal, a digital twin of the system is built. A hybrid adaptive algorithm is proposed, fed by a diagnosis layer trained upon digital twin simulations.

6.1 Introduction

In manufacturing systems, productivity targets are demanding to ensure maximum reliability and availability of machine tools [29], while breakdowns and failures need to be fully avoided [56, 161]. Due to these requirements and increased system complexity, preventive maintenance represents a conspicuous fraction of the total costs in an industrial scenario [99]. Consequently, the attractiveness of Prognostics and Health Management (PHM) solutions rises.

International standards describe the PHM framework [94–98] that consists of four modules: preliminary analysis, profile monitoring, diagnosis and prognosis [14]. The first regards the collection and investigation of all possible faults of the analysed system. The second performs the fault detection, i.e. the observation of a statistical discrepancy between the ongoing and a pre-recorded healthy working condition. The third deals with fault isolation and quantification, i.e. the localisation of the faulty component and its wear level assessment, respectively. The last step deals with the prediction of the Remaining Useful Life (RUL) of the component, i.e. the remaining time for which the component can perform the assigned task. RUL estimation should also contain information about its uncertainty through the estimation of its probability density function (pdf) [14].

Two steps are needed to create a proper dataset for PHM strategies: feature extraction and feature selection. The first consists of the computation of statistical quantities from sensor data. It enables dimensionality reduction of collected data, trying to condense the information as much as possible. Features can be computed in many domains: time, frequency and time-frequency domain [82, 83, 99]. Feature selection allows for a further reduction of the features pool. Spearman's and Pearson's correlation coefficients can be used to rank features and keep the most correlated ones [82, 83]. ANOVA could be applied when data are collected in classes [61].

Once the dataset is created, profile monitoring techniques can be adopted. Statistical Process Control (SPC) [152] is the typical tool, as proposed by Liu *et al.* [142] and Colosimo *et al.* [39].

Diagnosis is a classification problem. Several types of algorithms could be used to localise and identify the nature of the fault: Linear Discriminant Analysis (LDA), Support Vector Machines (SVM) [112], Mahalanobis-Taguchi Systems [183], filtering techniques, e.g. Unscented Kalman Filter (UKF) [165] and Artificial Neural Networks (ANN) are just a subset of possible solutions [79, 83, 242]. An innovation in this field could be progressive learning, introducing the capability of increasing the number of clusters during online learning [196]. Additionally, utility theory can be applied to introduce probabilities of critical conditions. It was successfully implemented to support decision making in maintenance actions on bearing faults in a sewage treatment plant [112].

Dealing with prognostics, four main algorithm categories can be distinguished depending on the data availability and the approach to the problem [14, 162]:

- Knowledge-based models: expert knowledge is translated in simple rules that the system can interpret. Such methods can be used only if robust knowledge of the degradation phenomenon and the machine is available. Expert systems, in which rules assume the form of *IF-THEN*, and fuzzy logic, giving a linguistic description of the system, are part of this category [14, 162];
- White-box models (model-based): they rely on a physical model of the degradation phenomenon [162]. Although the model structure is known *a priori*, experimental data are necessary to identify the model parameters [14]. Exponential lifetime prediction model for ball screw mechanisms under different feed modes [35], differential models for tool wear evolution in milling [235] and wear model for flank wear in turning [168] are just examples of this;
- Grey-box models (statistical-based methods): they rely on a dynamic stochastic description of the degradation phenomenon. The model is selected by the user and its coefficients are estimated through experimental data [14]. Their advantage is related to a statistical description of the RUL and, indeed, a strong support for decision making on maintenance actions. Filtering-based approaches such as Kalman filters [64] and its variations [65], or particle filters [137, 203] are all examples of grey-box approaches, as well as Hidden Markov Models (HMM) and Hidden semi-Markov Models [14]. Linear regression models can even be applied for RUL prediction [45].
- Black-box models (data-driven): they "learn" and describe the problem directly from the collected data. Data quantity and quality of both faulty and fault-free

data are of fundamental importance for successful implementation [94]. ANN, Self-Organising-Maps (SOM) [37, 142, 193] and deep learning algorithms, SVM and Relevance Vector Machines (RVM) [88] are just a few examples of artificial intelligence techniques.

Hybrid approaches have recently emerged, fusing multiple areas and exploiting their synergies. For instance, Sbarufatti *et al.* developed a prognostic solution for Li-ion batteries using particle filters to update Radial Basis Function Neural Networks. This solution could predict the RUL pdf, providing adaptiveness to new data [175]. Scientific works were mapped in table 6.1, according to several classes, highlighting some useful aspects of the proposed methodologies like the algorithm domain [14].

Actually, different challenges prevent PHM techniques to find a robust implementation in manufacturing:

- the system under analysis needs to be sensorised to provide meaningful signals regarding the fault conditions. No rules have been designed to choose the right ones;
- experimentation is needed: all the techniques available nowadays require training data (data from all the fault combinations or run-to-failures). Typically, only fault-free data are largely available and experimentation can be extremely expensive and time-consuming;
- developed solutions are often working-cycle dependent. This is a critical problem in many applications, e.g. with machining centres;
- how to estimate and deal with the uncertainty of the RUL prediction, i.e. the determination of its pdf.

Furthermore, scientific research is still lacking in the manufacturing field. The hydraulic unit is one of the most critical parts and the cause of unexpected breakdowns and downtimes [29, 161]. From a reliability analysis of ten CNC lathes from 2009 to 2014, the hydraulic subsystem showed the highest failure rate (22.9% of the total failures) [29]. A similar project on twelve machining centres between 2005 and 2010 confirmed the result [217]. Two other contributions highlighted the necessities of performing PHM on machine tool auxiliaries, being sources of unexpected downtimes and of comparable loss of production costs with respect to machine tool main components [62, 63]. Despite the above, research on PHM of CNC machines is rare in such subsystems, especially on hydraulic units. A case study on monitoring the filter's health state in an oil mist separator was recently presented in 2019 [62]. Authors trained machine learning on healthy data to model the environmental effects on the measured fan power. The reconstruction error was used as the health index (HI) of the filter. Instead, PHM in the machine tools field mainly focuses on tool monitoring and prognostics. Specific force coefficients were addressed as cutting-condition independent and tool-wear sensitive features by Nouri *et al.* [156]. Cheng *et al.* proposed a machine learning methodology based on cutting forces, vibration signals and machined surfaces' image features, typically used separately, to monitor the wear of different turning tools over various materials and cutting conditions [34]. McLeay *et al.* applied a Mahalanobis distance-based unsupervised algorithm to detect anomalies in the cutting process to assess the tool life in milling applications under fixed working conditions. The advantage

Table 6.1: *Map of State-of-The-Art PHM approaches [14]. Algorithms are described by machine tool units of interest, failure modes, employed sensors, features selection strategy, selected features, algorithm and relative category. Items belonging to the hydraulics class, are not belonging to MT hydraulic units works, but hydraulic systems in general.*

Ref.	MT unit	Failure modes	Sensors	Feature selection strategy	Selected features	Algorithm	Class
[131]	Feed-axis	Ball screw preload loss	Motor torque and current sensor, accelerometer mounted at ball screw nut	Maximization of signal to noise ratio and display of a monotonic trend	Features set extracted from vibration signals	Gaussian Process Regression	Grey-box
[210]	Feed-axis	Pitting on screw surface (as a result of accelerated life tests)	Accelerometers, one mounted on ball nut and one on bearing housing	Ranking based on trendability index; then correlation clustering to group features with the same information content and extract a representative feature from each class	Weighted Mahalanobis distance	Particle filter, based on an exponential Wiener process model, whose parameters are iteratively updated drawing on Bayesian theory	Grey-box
[226]	Feed-axis	Ball screw degradation	3 accelerometers, of which 2 mounted on bearings housing and one on screw nut	Principal Component Analysis	First five principal components	Dynamic fuzzy neural network, whose parameters have been tuned with a quantum genetic algorithm	Experience-based & black-box
[136]	Spindle	Balls, cage and races failures as a result of insufficient lubrication and salt water put on races to accelerate wear	Bearings and motor vibration and temperature	Fisher criterion	Minimum Quantization Error	Particle filter, operating on an exponential model identified with Bayesian statistics	Grey-box
[170]	Spindle	-	Bearings vibration and temperature, motor temperature and current	Correlation	-	Polynomial regression	Grey-box

Ref.	MT unit	Failure modes	Sensors	Feature selection strategy	Selected features	Algorithm	Class
[182]	Spindle	Failure as a consequence of either cage, inner race or outer race defect	Spindle vibration and temperature	Highest signal to noise ratio	Mahalanobis Distance (MD)	Linear approximation of MD evolution over time for RUL prediction	Grey-box
[65]	Hydraulics	Circuit leakages	Liquid level (quantized degradation messages as output)	-	Liquid level	Particle filter, based on a discrete time state space exponential model, whose slope is online estimated	Grey-box
[239]	Hydraulics	Oil contamination	Oil temperature, viscosity and dielectric constant	-	Particle contamination level	Particle filter	Grey-box
[45]	Hydraulics	Servo-valve clogging	Actuator (motor) current and valve position	Sensitiveness and correlation with respect to clogging	Integral of motor current	ARMA model	Grey-box
[64]	Hydraulics	Piston pump: looseness of regulator valve spring	Pressure	-	Average system pressure	Kalman filter; Montecarlo method for RUL confidence interval computation	Grey-box
[81]	Hydraulics	Piston pump: wear between valve plate and cylinder barrel	Oil return flow	Empirical mode decomposition	Oil return flow	Support Vector Machine	Black-box
[187]	Hydraulics	Piston pump: loose slipper	Vibration	-	Spectrum entropy, after discrete cosine transform and composite spectrum computation	Recurrent Neural Network	Black-box
[183]	Hydraulics	Centrifugal pump: seal and impeller failure and filter clogging	Vibration, temperature, flow, pressure	Highest signal to noise ratio	Mahalanobis distance	Linear dynamic model	Grey-box

Ref.	MT unit	Failure modes	Sensors	Feature selection strategy	Selected features	Algorithm	Class
[203]	Hydraulics	Slurry centrifugal pump: impeller failure	Vibration	Expertise	Sum of the amplitude of the harmonics in the neighbourhood of the vane pass frequency. Then, a moving-average wear indicator is used to extract the central tendency from the time profile of the previously defined feature	Particle filter, with an exponential model	Grey-box
[88]	Hydraulics	Slurry centrifugal pump: impeller failure	Vibration	Expertise	Standard deviation	Relevance Vector Machine, which estimates the parameters of two exponential functions	Black-box

of being an unsupervised method relies on the fact that only the normality condition is experimented for the training phase [145]. DaSilva *et al.* investigated tool wear evolution in the drilling of high-strength compacted graphite cast irons and individuated the spindle current signal as the best cost-benefit monitoring variable for tool wear [42]. HMM were also applied in different fashions for tool wear applications [194, 220]. Other subsystems are investigated by research literature to a lesser extent, such as those of the spindle and feed-drive. Chen *et al.* developed an overall machine tool monitoring method based on the frequency analysis of energy, intended as the collection of power, thermal, current and vibration measurements. Analytical modelling of screw, guide rail and bearing frequencies allowed the health status of the components to be estimated [32]. Moore *et al.* presented a test scenario in which machine learning and deep learning classifiers were applied for machining defects and machine tool failure-mode classification. Besides this, unsupervised algorithms for clustering were tested for novel failure-mode recognition [154]. Xia *et al.* developed a diagnosis solution for multiple units of flexible production line machining centres, including feed axes, spindles and converters. A neural network scheme was adapted to each machining centre to avoid the combination explosion of learning rules and miss-classification [215]. Multiple polynomial regression [170], the Mahalanobis-Taguchi System [182] and SOM [137] were applied to predict the RUL of rolling element bearing failures in the spindle system. Feed-drive system health and its influence on tool wear was investigated through a long-term operational modal analysis of vibration signals [104].

The focus of this chapter is mainly concentrated on the challenges mentioned above. A novel approach to deal with PHM was presented, trying to avoid run-to-failures through the use of a Dymola® model of the hydraulic unit of a machine tool. In section 6.2, the system and the model were described, together with its validation. The proposed solution was explained in section 6.3, supported by a graphical representation of the whole approach. Starting from the simulation of different working regimes (6.3.1), the creation of the datasets was described in subsection 6.3.2. The novel structure of the diagnostics phase, exploiting different algorithms for any component, was presented in subsection 6.3.5. In subsection 6.3.6, the innovative developed prognosis algorithm was presented, being set free from run-to-failures. Lastly, in section 6.4, a critical analysis of the entire process was carried out, while future developments and conclusions were reported in section 6.5.

6.2 Materials

6.2.1 System and model description

The system is the hydraulic unit of Mandelli's Spark machining centres. It is constituted by a high pressure (HP) sub-unit (that drives the tool clamping system and the braking system); a medium pressure sub-unit (that cools down and lubricates the biggest organs in the machine); a low pressure sub-unit (that cools down the oil by pumping it to the chiller).

One of the novelties at the basis of this research was to reproduce the faulty behaviour of the system through simulations. All the described sub-units were modelled together with different fault states of the system. A schematic representation of the

Table 6.2: *Fault states and associated HI for each component.*

Pump Leakage ($\frac{\dot{m}_l}{\dot{m}_{nom}} 100$ [%])	HI_P [%]	Sensor offset [bar]	HI_S [%]	Valve natural frequency [Hz]	HI_V [%]
0.0	100	0.0	100	29.8	100
0.6	80	0.5	80	25.0	50
1.2	60	1.0	60	21.2	0
1.8	40	1.5	40		
2.4	20	2.0	20		
3.0	0	2.5	0		

digital twin developed in Dymola®[1] was shown in figure 6.1. As explained above, the system digital twin consisted of three main parts as the hydraulic unit, in addition to each corresponding user. Each sub-unit works separately from the others a part from the fact that they share the oil coming from the same tank. In order to model the system under investigation different standard libraries of Dymola called Modelica and Cooling have been used, modifying all the components in order to make them suitable for the system under exam:

- Cooling is a library providing access to thermodynamical components, suitable for application requiring a fluid medium. It can be used for compressible and incompressible fluid applications; in this case, hydraulics. Standard components have a dedicated model, like pumps, compressors, basic valves, tanks, etc. The used oil was modelled following a standard fluid template with the corresponding physical characteristics;
- Modelica is a Dymola library containing all the basic blocks for signal generation and processing: constants, step sources, noise sources, addition blocks, integrators, etc.;
- at last, many components, e.g. those for governing the control logics of the actuated elements of the system have been generated from scratch.

The tuning procedure was based on the implementation of general characteristics of the components installed on the real hydraulic unit. Fine tuning was performed manually, optimising important system level behaviour characteristics of the hydraulic unit, as it will be later explained in section 6.2.2.

In order to focus only on relevant faults, the history of 15 years of maintenance reports was analysed. They contained data from several similar machines and different faults. The evolution of the fault occurrences over the 15 years was shown in figure 6.2. Although filter faults were the ones that occurred the most, they were typically subject to ordinary maintenance. Then, pump, sensor and valve faults were the most frequent and relevant ones. A HP pump leakage (fig. 6.1a), a pressure-switch offset (fig. 6.1b) and an increased opening time for the servo-valve of the tool clamping actuators (fig. 6.1c) were introduced in the model of the system [83, 84]. The pump was a

[1] Dymola is a commercial product of Dynasim AB (Dassault System) based on the open source modelling language Modelica. Dymola is a multi-domain modelling environment for simulation of complex dynamical systems. The advantage of this modelling environment is that the language is acausal. Acausal language means that blocks are not described by their inputs and outputs, but, instead a symbolic formula represents a component, providing dramatic advantages in the model formulation. Inversion and solution of complex formulae is managed directly by the integrated solver of the program.

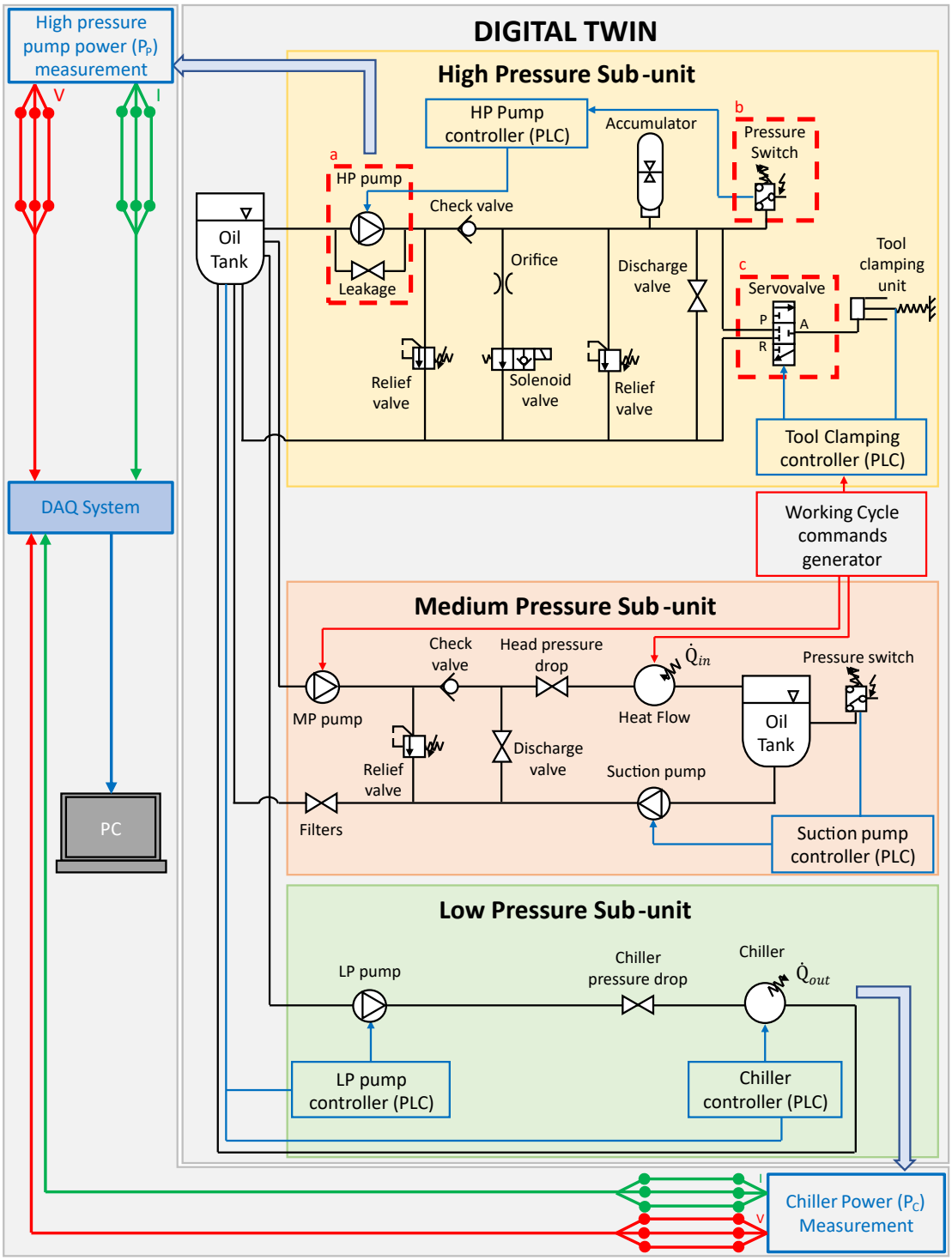


Figure 6.1: *Digital twin layout. Top-left and bottom right arrows represent the link from the digital twin to the measurement system. In the three dashed boxes, the components under analysis: a) HP pump; b) Pressure-switch; c) Servovalve*

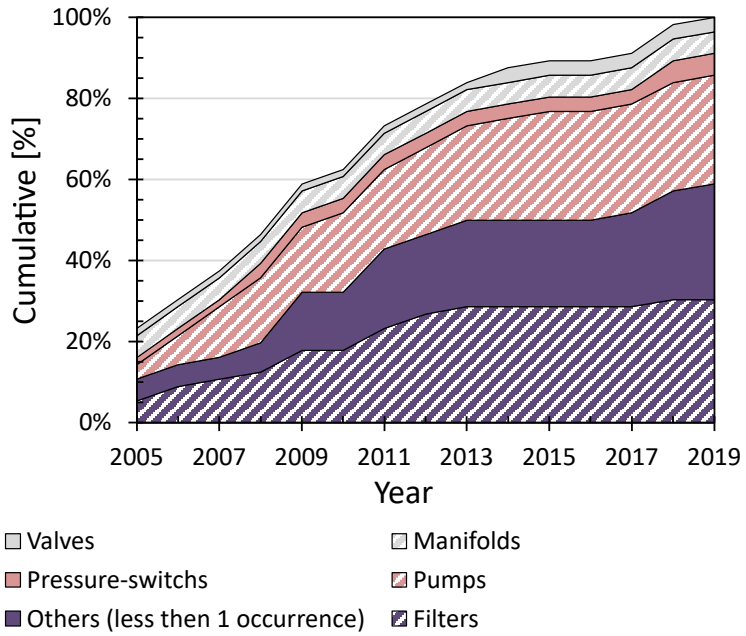


Figure 6.2: *Evolution of fault occurrences over the 15 years maintenance reports*

positive displacement VIVOIL XV1/P-4.9D. The sensor, the subject of the analysis, was a pressure-switch used to keep the HP system between 85 and 95 bar. The HP pump was controlled in a closed-loop by this sensor (IFM PN7071 025-MPa).

The HP pump was modelled through the following equations:

$$\omega = \frac{d\phi}{dt} \tag{6.1}$$

$$\tau = \frac{V_{disp} \cdot dp \cdot \eta_v}{2\pi \cdot \eta_g} \tag{6.2}$$

where ω and ϕ are the angular speed and position of the pump shaft, respectively; τ is the torque applied to the shaft; V_{disp} is the displacement of the pump; dp is the pressure difference between the pump ports; η_v and η_g are the volumetric and global efficiencies, respectively. The mass balance equation is:

$$\dot{m}_{in} = -\dot{m}_{out} = \dot{m} \tag{6.3}$$

where $\dot{m}_{in}$ and $\dot{m}_{out}$ represent the mass flows at the inlet and outlet pump ports respectively; $\dot{m}$ is the module and the minus sign represents a flow exiting the port. Specific enthalpies at the two ports are related to pressures by:

$$h_{out} = h_{in} + dh \tag{6.4}$$

$$p_{out} = p_{in} + dp \tag{6.5}$$

$$\rho_{in} = \rho \tag{6.6}$$

$$dh = \frac{dp}{\rho \cdot \eta_g} \tag{6.7}$$

where ρ is the working fluid density, defined as the inlet port oil density ρ_{in}. The mass flow and the pump power are:

$$\dot{m} = \rho \cdot \dot{V} = \frac{\rho \cdot \eta_v \cdot V_{disp} \cdot \omega}{2\pi} \tag{6.8}$$

$$P = \dot{m} \cdot dh \tag{6.9}$$

The pump leakage was introduced in the model as a valve described by the following equations:

$$\dot{m}_{leakage} = pos_{leakage} \cdot A_v \cdot \sqrt{dp \cdot \rho} \tag{6.10}$$

$$A_v = \frac{\dot{m}_{nom}}{\sqrt{dp_{nom} \cdot \rho_{nom}}} \tag{6.11}$$

where $\dot{m}_{leakage}$ is the leakage flow; A_v is the valve coefficient; dp_{nom}, ρ_{nom} and $\dot{m}_{nom}$ are nominal values for the fully opened valve; $pos_{leakage}$ represents the opening degree of the valve and regulates the leakage flow.

Since the pressure switch is a sensor, its output is just the pressure of the oil at the inlet port. It commands the HP pump through an on/off switch. It turns on the pump when the pressure decreases below the lower threshold th_l, whereas it turns off the pump when the pressure increases over the upper threshold th_u. Being aux an auxiliary variable:

$$aux = \begin{cases} th_l, & \text{when } p <= th_l \\ th_u, & \text{when } p >= th_l \end{cases} \tag{6.12}$$

where p is the pressure measured by the pressure switch. If p is outside the two control thresholds th_l and th_u, the command y is:

$$y = \begin{cases} 1, & \text{if } p <= th_l \\ 0, & \text{if } p >= th_u \end{cases} \tag{6.13}$$

If p is in between the two control thresholds, then:

$$y = \begin{cases} 0, & \text{if } aux > \frac{th_l + th_u}{2} \\ 1, & \text{otherwise.} \end{cases} \tag{6.14}$$

The pressure switch offset was introduced by adding a bias to the real pressure of the system, so that:

$$p = p_{real} + b \tag{6.15}$$

where p is the actual reading of the sensor, p_{real} is the real pressure at the sensor inlet port and b is the bias term.

The servovalve was modelled as two separated valves with complementary opening positions. The first one linked the HP port P to the actuator port A, the second one linked the actuator port A to the return line R (fig. 6.1c). The equations governing the servovalve behaviour are:

$$\dot{m}_{PA} = pos_{PA} \cdot A_v \cdot \sqrt{dp \cdot \rho} \tag{6.16}$$

$$\dot{m}_{AR} = pos_{AR} \cdot A_v \cdot \sqrt{dp \cdot \rho} \tag{6.17}$$

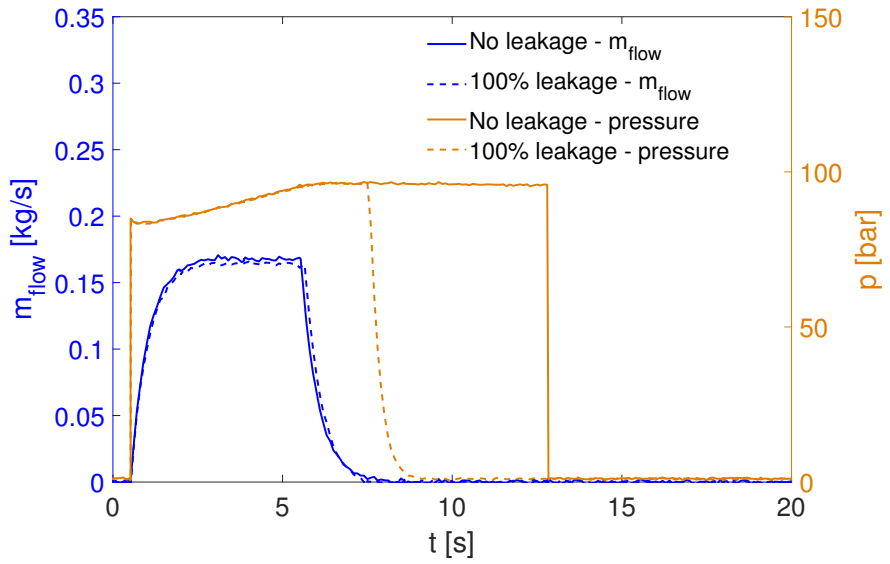


Figure 6.3: *Comparison between pressures and mass flows of the HP pump with and without leakage fault*

$$A_v = \frac{\dot{m}_{nom}}{\sqrt{dp_{nom} \cdot \rho_{nom}}} \tag{6.18}$$

where $\dot{m}_{PA}$ and $\dot{m}_{AR}$ are the oil flows through the valves; A_v is the valve coefficient; dp_{nom}, ρ_{nom} and $\dot{m}_{nom}$ are nominal operating points for fully opened valves. pos_{PA} and pos_{AR} are second order dynamic responses [167] to the positioning error of the actuator normalised on the actuator stroke. They are constrained to be in the range of 0 and 1. The servovalve fault was introduced by changing the natural frequency of the second order transfer functions. The simulated fault states, with the associated *HI* were reported in table 6.2. An example of the difference in the physical quantities of the system caused by the faults was shown in figure 6.3. The pressure at the outlet of the HP pump and the mass flow of the pump were compared in a full-health behaviour and in the 100% pump leakage case. The leakage implied a longer time to reach the required pressure and the pressure at the outlet port was quickly discharged.

6.2.2 Digital twin experimental validation

The validation of the digital twin was conducted by means of power acquisitions performed on a Mandelli's Spark 1600 machining centre:

- power absorbed by the HP pump electric motor;
- power absorbed by the chiller;

The experimental setup (fig. 6.4) was composed of a three-phase acquisition system for phase voltages and currents for each of the above units. Six LEM LF 205-S/SP3 and three NI9205 acquisition cards from National Instruments™ were adopted.

A brief comparison between the experimental and the simulated physical quantities during a healthy cycle was shown in figure 6.5a. Both the duty cycle and power consumption of the HP pump were respected during the idle time. The validation of the

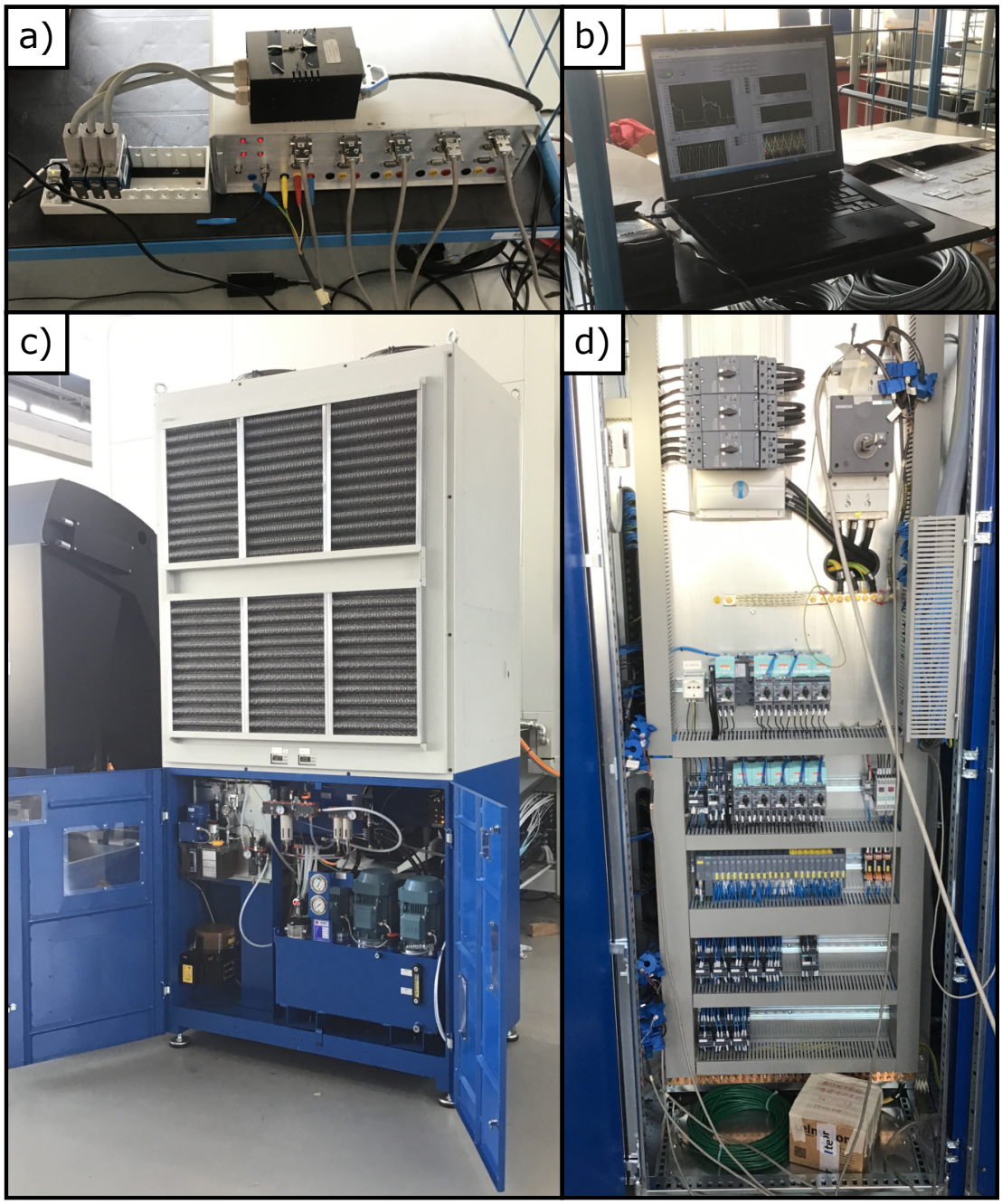


Figure 6.4: *The entire experimental setup: a) DAQ system and power meter; b) LabView® acquisition software; c) Hydraulic unit; d) electrical cabinet and LEM installation*

duty cycle and power consumption of the chiller was reported both for the machine warm-up phase (the spindle is activated to reach a steady-state thermal condition, fig. 6.5b) and idle state (fig. 6.5c).

6.3 Methods

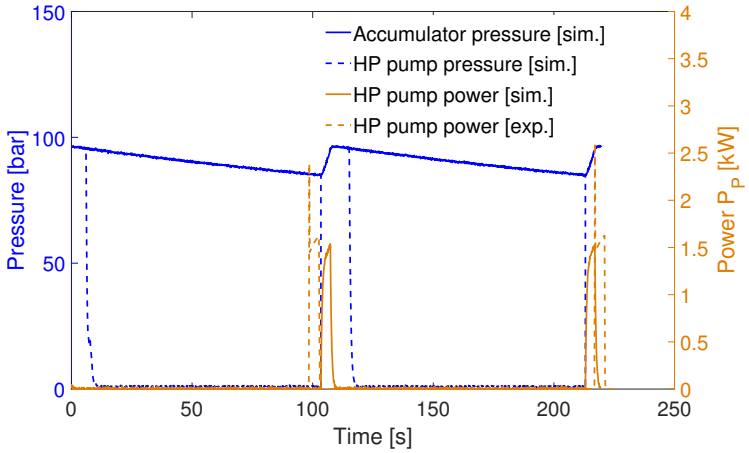
6.3.1 Synthetic data generation

In the machine tools scenario, collecting fault data could be a difficult task, perhaps infeasible. A digital twin of the system was used to recreate all possible combinations of fault states (fig. 6.6A) for the components under analysis while operating under different working regimes (fig. 6.6B), following scientific literature [1, 122, 126, 146, 174, 197] and analogously to what Helwig *et al.* did experimentally [82–84]. Indeed, in industrial scenarios, machine tools present high flexibility in working conditions. Hydraulic unit working cycles depend on various parameters such as the occurrence of tool changes, the duration of the machining operations and the loading condition (i.e. the heat transferred to the oil from the machine head/spindle during the operation). It is assumed that a single working cycle for PHM could be acceptable when the machine is dedicated to a single task (e.g. in mass-production industries), while this is not the case for most manufacturing companies. Different working regimes can cause dramatic changes in sensor outputs and, as a consequence, in the classification accuracy [83].

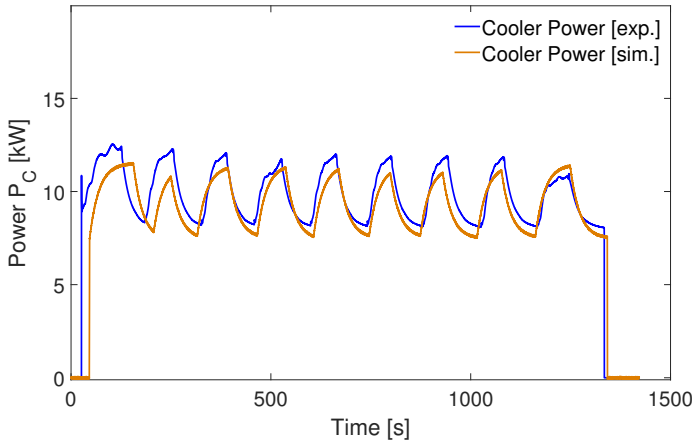
Then, as a novelty with respect to typical literature approaches, two kinds of working cycles were simulated to test the robustness of the proposed solution: stationary and non-stationary ones, which represented a machine conceived for a specific and a more flexible task, respectively (fig. 6.6B). Non-stationary cycles were created by mixing stationary cycles in fixed proportions, representing more realistic scenarios for a machine tool. Two stationary (SC1 and SC2) and two non-stationary working cycles (NSC1 and NSC2) were simulated:

- SC1 was composed of 300 seconds of machining (tool clamped) and 15 seconds for tool change (tool released), fig. 6.7a. The loading condition was represented by the average heat flow of 5kW removed from the head/spindle. This cycle comprised stationary subsequent phases of machining and tool changes;
- SC2 was composed of 150 second and 10 second phases respectively (figure 6.7b). The average heat flow absorbed by the oil was 3kW.
- NSC1 was composed of 70 percent of SC1 and 30 percent of SC2.
- NSC2 was composed of 30 percent and 70 percent respectively.

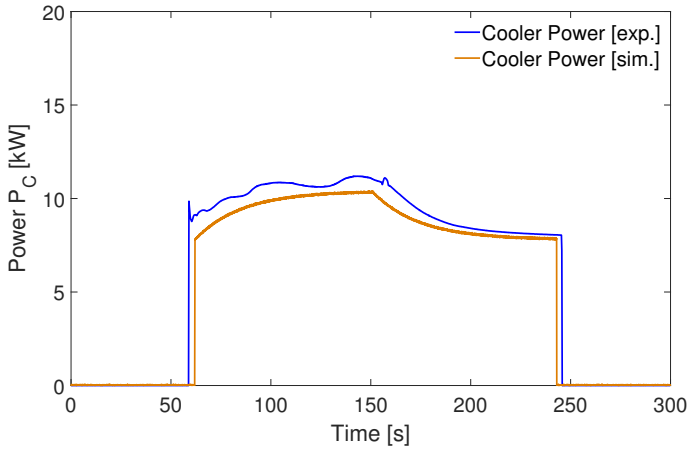
Figure 6.7c represented a qualitative structure of non-stationary cycles. Two datasets were created, one including SC1 and SC2, the other including NSC1 and NSC2. The output of the digital twin consisted of 41 physical quantities, theoretically measurable through sensors to be mounted on the machine. A total of 108 simulations for each dataset were generated combining all the *HI* listed in table 6.2. Indeed, the datasets covered all the possible scenarios: all components at full health, single faults and multiple occurring faults.



(a) *Acquisitions from HP pump of the unit during a healthy behaviour*



(b) *Acquisitions from chiller healthy behaviour in warm-up*



(c) *Acquisitions from chiller healthy behaviour in idle state*

Figure 6.5: *Comparison between experimental and simulated power*

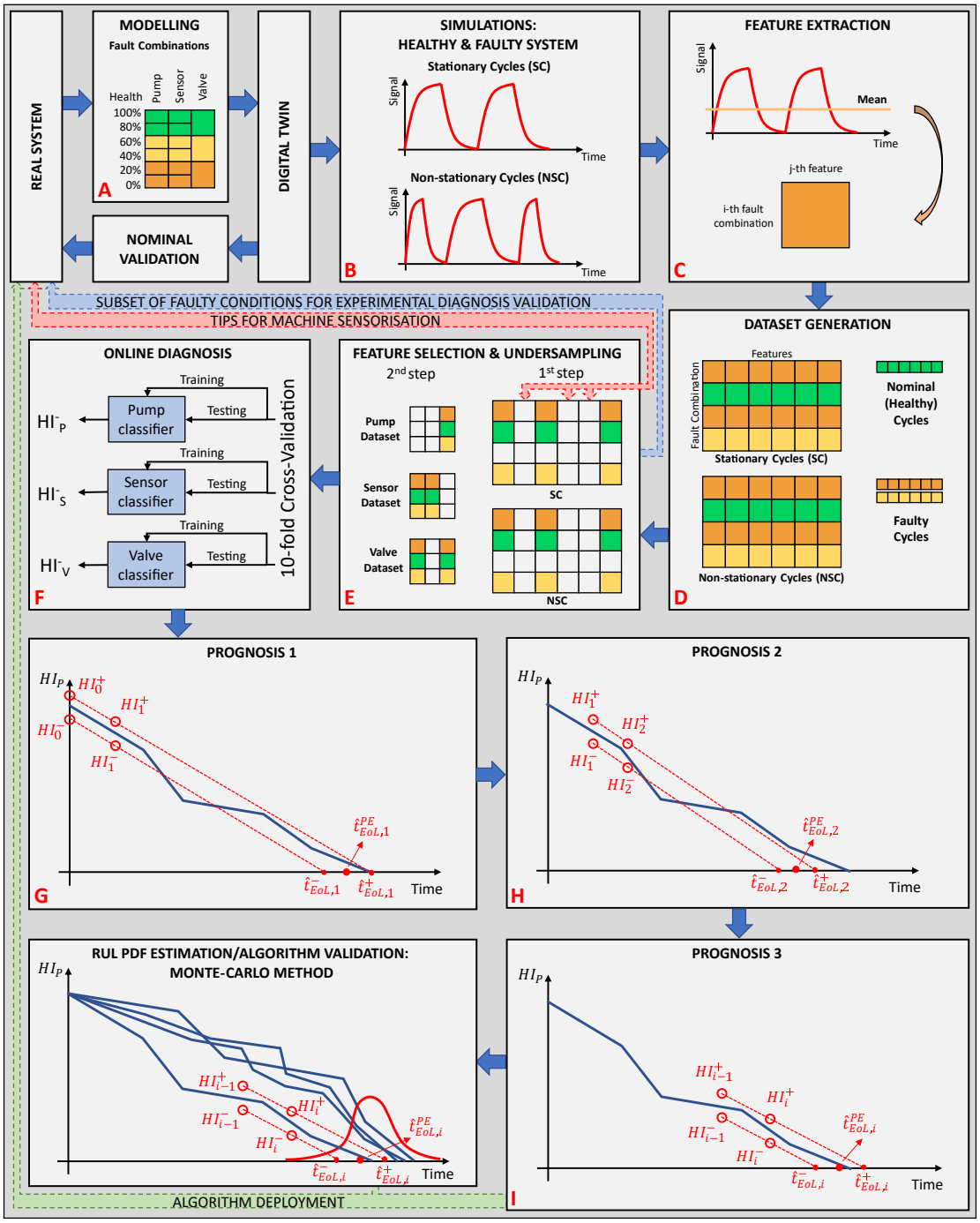
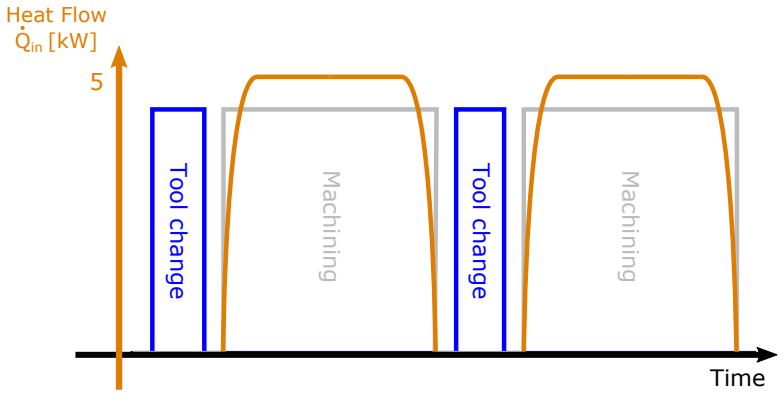
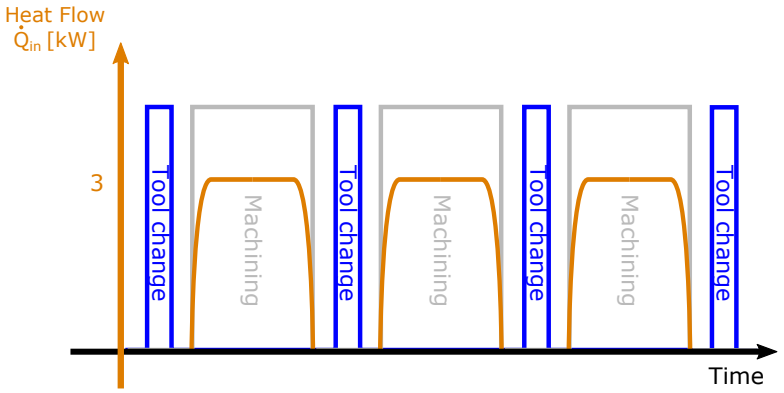


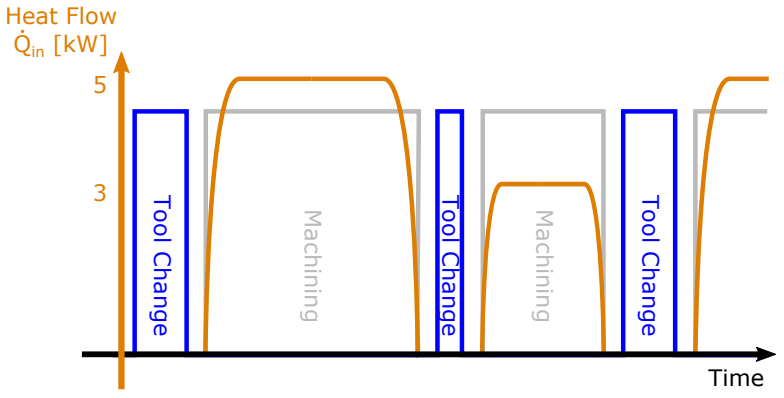
Figure 6.6: *Framework of the whole approach*



(a) *SC1*



(b) *SC2*



(c) *NSC*

Figure 6.7: *Qualitative structure of the simulated working cycles in terms of machining duty cycle, tool change and heat flow* $\dot{Q}_{in}$

6.3.2 Feature extraction

Three global features (from the whole cycle) were extracted for each signal: mean, Skewness and Kurtosis coefficients (fig. 6.6C). Global features provide a significant reduction of the dimensionality, but in some applications, they are not enough to obtain good results and local ones (from parts of the cycle) should be introduced. The features were then normalised to have null mean and unitary standard deviation [183]. For each fault combination, 20 repetitions were obtained adding random Gaussian noise proportional to signals' RMS. Resulting datasets consisted of 2160 rows and 123 columns (fig. 6.6D).

6.3.3 Feature selection/Machine sensorisation

An innovative aspect presented in this chapter regards the use of feature selection to obtain useful tips for machine sensorisation. Although applying dimensionality reduction techniques such as Principal Component Analysis or LDA leads to a smaller space to work on, the entire set of features is used and no sensor selection can be done. The proposed feature selection strategy was divided into two steps (fig. 6.6E). The first was dedicated to sensor selection. The score of each feature with respect to the components was computed separately through the One-Way-ANOVA F-statistic [153]:

$$F_k = \frac{\sum_{j=1}^{r} n_j(\bar{x}_{j,k} - \bar{x}_k)^2/(r-1)}{\sum_{j=1}^{r}\sum_{i=1}^{n_j}(x_{ij,k} - \bar{x}_{j,k})^2/(N-r)} \tag{6.19}$$

where r is the number of HI of a component; n_j is the number of samples in the $j-th$ group; $\bar{x}_{j,k}$ is the sample mean of the $k-th$ feature of the $j-th$ group; $\bar{x}_k$ is the mean of the $k-th$ feature; $x_{ij,k}$ is the value of the $k-th$ feature in the $i-th$ repetition of the $j-th$ group; N is the total number of rows of the dataset (fig. 6.8).

At the same time, the correlation matrix between features was computed, using Pearson's correlation coefficient:

$$\rho_{k,m} = \frac{\mathrm{Cov}(F_k, F_m)}{\sigma_{F_k} \cdot \sigma_{F_m}} \tag{6.20}$$

where $\rho_{k,m}$ is the correlation coefficient between the features $k-th$ and $m-th$ one; Cov is the sample covariance between features F_k and F_m; σ_{F_k} and σ_{F_m} are the sample standard deviations of features F_k and F_m. Starting from the feature with the highest score, up to ten features for each component were selected if each correlation with respect to those already selected was below a threshold of 0.4[2]. A final features pool was created by concatenating selected features for each component and excluding duplicates.

The second step consisted of reiterating the computation of the ANOVA scores only on the new features pool and selecting the best features for each component. The novelty of this approach was not only the use of non-correlated features but also the selection of which feature, and consequently which sensor, provided the most valuable information. This step could have consequences on the design of the machine sensorisation. By identifying the best features for the diagnosis purpose, only useful sensors

[2]There is not an absolute value for correlation cut-off. However, in the work proposed by Vatcheva *et al.*, the authors indicated that usually correlation cut-offs for regression can range from 0.5 - 0.8 [195]. Here, the idea is to be more conservative, being in a classification scenario. Furthermore, the aim is to also reduce the number of features as much as possible.

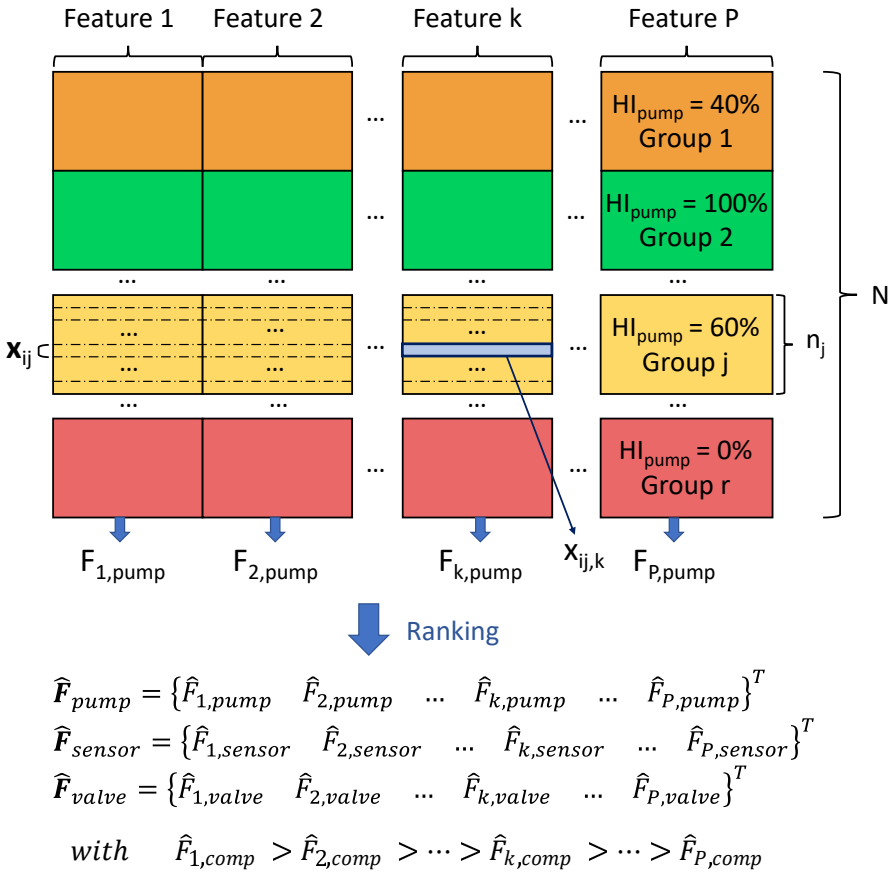


Figure 6.8: *Graphical representation of ANOVA dataset subdivision and scoring*

were traced back: seven sensors were used to compute the selected features pool representing 17% of the initial pool. Only the second step of feature selection was needed for the final implementation of the algorithm.

6.3.4 Undersampling/Design of experiments

The undersampling technique was applied to reduce the number of experiments for a future validation of the model under faulty conditions. Experimental campaigns can be time-consuming and expensive (e.g. Helwig *et al.* conducted a total of 2,205 tests on a test bench [83]). With regard to the objective of creating an experimental campaign based on the real system, selecting which scenarios to experiment first was needed.

Undersampling, typically applied to imbalanced datasets [15], was extended to identify the most valuable fault combinations for the classification purpose and consequently for the future experimentation on the real machine (fig. 6.6E). This technique could be applied since a classification model was developed for each component. The dataset was reduced three times (one for each component) and all the shared combinations were stored. In case a particular label is missing after this process, a fault-combination must be reintroduced manually to preserve all the classes of each component. Based on the nearest neighbours rule, *NearMiss* [15], the used undersampling technique selected a given number of samples from cluster *A* which were closest to each instance in cluster *B*.

Table 6.3: *Time required by different algorithms to perform a 10-fold cross-validation, training and testing of a single sample for each component analysed. The algorithms' complexity is also reported following computational performance metrics [173]: p, number of features; r, number of classes; t, number of trees; N, number of samples; e, number of training epochs; i, j, k, l, number of neurons in each layer.*

LDA	10F-CV [s]	Training [s]	Testing [s]	Complexity
Pump	0.190	0.015	<0.001	$\mathcal{O}(rp)$
Sensor	0.189	0.022	<0.001	
Valve	0.160	0.015	<0.001	
RFC	10F-CV [s]	Training [s]	Testing [s]	Complexity
Pump	9.935	0.300	<0.001	$\mathcal{O}(N^2pt)$
Sensor	6.559	0.416	<0.001	
Valve	4.710	0.263	<0.001	
ADB	10F-CV [s]	Training [s]	Testing [s]	Complexity
Pump	8.972	0.579	<0.001	$\mathcal{O}(pNt)$
Sensor	178.669	17.289	0.005	
Valve	362.898	33.593	0.009	
CNN	10F-CV [s]	Training [s]	Testing [s]	Complexity
Pump	68.730	0.858	<0.001	$\mathcal{O}(eN(ij+jk+kl))$
Sensor	97.438	0.900	<0.001	
Valve	125.235	1.094	<0.001	
QDA	10F-CV [s]	Training [s]	Testing [s]	Complexity
Pump	0.213	0.015	<0.001	$\mathcal{O}(p^2r)$
Sensor	0.232	0.015	<0.001	
Valve	0.080	0.010	<0.001	

As a result, 41 fault combinations were deleted, representing almost 38% of those that were started with. Undersampling was not needed for the final implementation of the algorithm.

6.3.5 Diagnostics

In this work, a tailored diagnosis solution was developed for each component fault (figure 6.6F). In literature, only one algorithm is selected and tuned for all the components under study. Here, a pool of algorithms was selected: LDA, Random Forest Classifier (RFC), Convolutional Neural Network (CNN), Quadratic Discriminant Analysis (QDA) and AdaBoost Classifier (ADB). The application scenario allowed for supervised learning techniques. In fact, the built datasets consisted of features values for known fault combinations (class labels). Two phases were needed: a training procedure, to update classifier parameters on training data, and a testing procedure to evaluate their performances. In order to compare the algorithms and select the most appropriate one for each component, classification accuracy (or classification rate) was introduced due to its easy interpretation. Performances were investigated both on stationary and non-stationary cycles. For the first ones, a 10-fold cross-validation (CV) was selected, according to the literature [79]. The mean of the accuracy was computed for each classification algorithm. For non-stationary cycles, algorithms were trained upon the

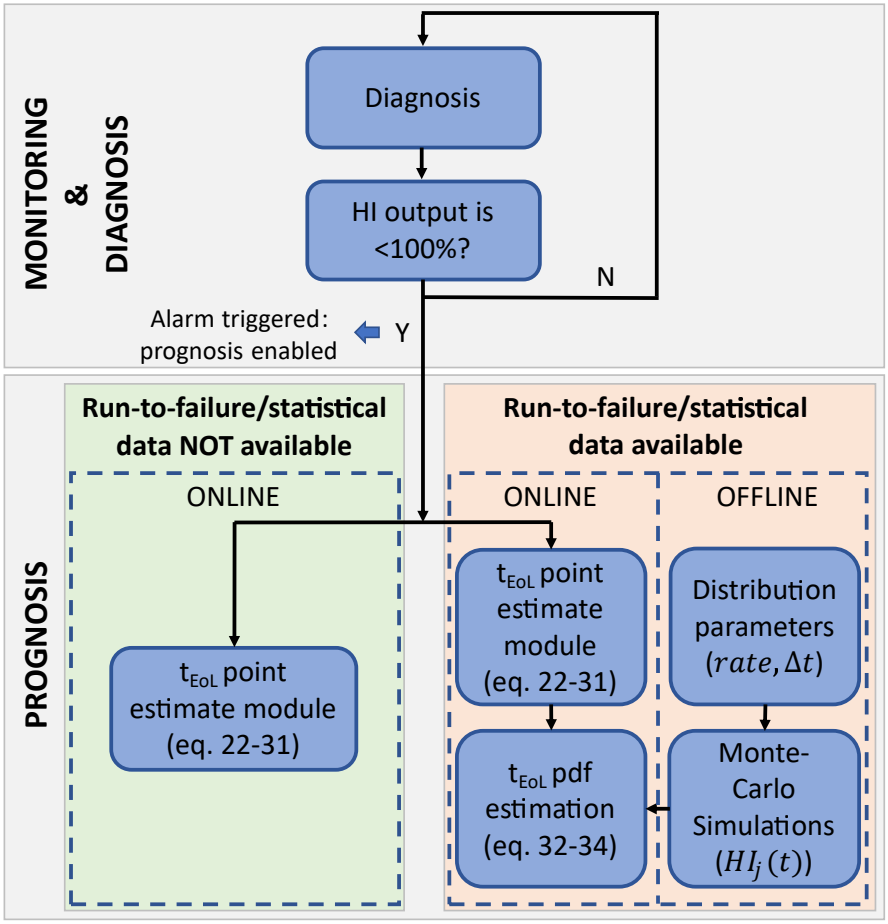


Figure 6.9: *Different strategies for prognosis in the presence, or not, of run-to failure data*

stationary dataset and tested on the non-stationary one. In this way, the robustness of the solution was validated on unseen and different working regimes. The time required by the algorithms to perform a 10-fold cross-validation, training and testing was reported in table 6.3, together with their complexity.

The algorithm with the highest final classification rate with respect to a given component was selected as its classifier (in case of a draw, the fastest one was chosen). Diagnosis was then run online at fixed intervals, providing HI as output for the three components.

6.3.6 Prognostics

The proposed prognosis solution can be applied in case run-to-failures or statistical data are available or not. The two associated procedures and algorithm structures were represented in figure 6.9.

A novelty of this prognosis approach is that it is based on the output of the diagnosis module, i.e. a set of HI representing the separated health status of the components. Indeed, the approach is able to take into consideration single fault scenarios and multiple occurring faults. By the way, it's worth noting that in this way, diagnosis performances have an impact on the prognostics solution. Diagnosis returns, at regular intervals of

time Δt_d, a discrete estimate HI^- of the real HI degradation pattern (i.e. ideally the nearest severity used for diagnosis training, figure 6.10a in red). Discretisation depends on how many fault states the classification algorithms were trained on. The gap between two levels of classified HI is:

$$\delta HI = \frac{100\%}{r-1} \tag{6.21}$$

where r corresponds to the number of fault states considered. For instance, for the pump leakage, the HI can be classified to be only 0%, 20%, 40%, 60%, 80% or 100% ($\delta HI = 20\%$ and $r = 6$). A prognosis module is generated for each component and acts independently of the others. Furthermore, the prognosis is triggered when the HI goes below 100% (fig. 6.9). For the sake of clarity, just one model is considered and explained.

End of Life (EoL) time point estimate

The procedure in the left branch of fig. 6.9 could be completely liberated from experimental tests and gives a point estimate of the RUL of a component.

At the end of the degradation process, a monotonic decreasing stair-like HI history is constructed ranging from 100% to 0%. When a new value of the HI is available, it is compared to the previous one. A set of three variables t_i, $HI^-(t_i)$ and $HI^+(t_i)$ (eq. 6.24) was stored whenever:

$$HI^-(t_i) < HI^-(t_i - \Delta t_d) \tag{6.22}$$

and holds:

$$t_i = k_i \Delta t_d \tag{6.23}$$

$$HI^+(t_i) = HI^-(t_i) + \delta HI \tag{6.24}$$

with $i = 0, \ldots, r-1$ (fig 6.10a). At the end of the component life, a set of r instants and r pairs of HI were collected (starting from the instant $t_0 = 0$, $HI_0^- = 100\%$ and $HI_0^+ = 100\% + \delta HI$).

Except for the starting time instant t_0, at any t_i, an updated estimation of the EoL time is performed as follows (fig. 6.6G-H-I):

1. Compute the line between the last two HI^+ and HI^- as a function of time t:

$$HI^-(t) = \frac{HI^-(t_i) - HI^-(t_{i-1})}{t_i - t_{i-1}}(t - t_i) + HI^-(t_i) \tag{6.25}$$

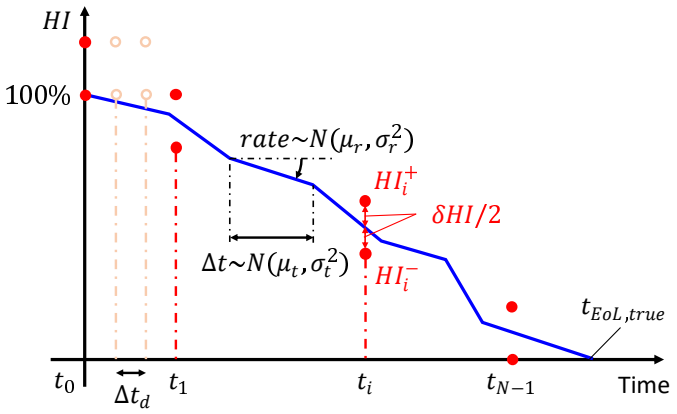
$$HI^+(t) = \frac{HI^+(t_i) - HI^+(t_{i-1})}{t_i - t_{i-1}}(t - t_i) + HI^+(t_i) \tag{6.26}$$

   with $i = 1, \ldots, r-1$.

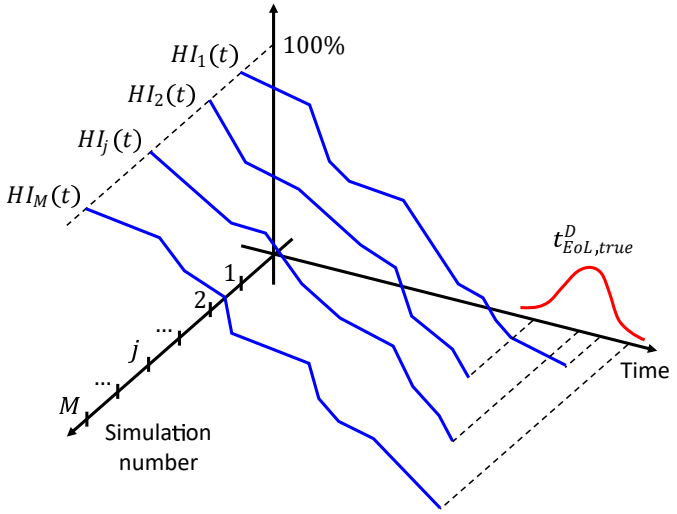
2. Compute the EoL estimates imposing:

$$HI^-(\hat{t}^-_{EoL,i}) = 0 \tag{6.27}$$

$$HI^+(\hat{t}^+_{EoL,i}) = 0 \tag{6.28}$$



(a) *The actual value of HI is represented by the solid line. Dots represent the HI_i at the time instant t_i. Parameters Δt and rate sampled from their distributions in any Monte Carlo simulation are shown*



(b) *Monte Carlo simulations of M $HI(t)$*

Figure 6.10: *Monte Carlo graphical representation*

from which:

$$\hat{t}^{-}_{EoL,i} = t_i - HI^{-}(t_i)\frac{t_i - t_{i-1}}{HI^{-}(t_i) - HI^{-}(t_{i-1})} \tag{6.29}$$

$$\hat{t}^{+}_{EoL,i} = t_i - HI^{+}(t_i)\frac{t_i - t_{i-1}}{HI^{+}(t_i) - HI^{+}(t_{i-1})} \tag{6.30}$$

with $i = 1, \ldots, r-1$.

3. Compute the unbiased $i-th$ point estimate for the EoL time:

$$\hat{t}^{PE}_{EoL,i} = \frac{\hat{t}^{-}_{EoL,i} + \hat{t}^{+}_{EoL,i}}{2} \tag{6.31}$$

with $i = 1, \ldots, r-1$.

The last step is necessary since $\hat{t}^{-}_{EoL,i}$ tends to underestimate the real EoL time, while $\hat{t}^{+}_{EoL,i}$ tends to overestimate it.

Monte Carlo RUL pdf estimation

If run-to-failure tests or statistical data were available, a statistical method for RUL pdf estimation is obtained through the Monte Carlo approach in the right branch of fig. 6.9. This method is based on the definition of a monotonic decreasing piecewise function which represents the "real" degradation trend of the HI of a component fault. This is true when a linear degradation pattern with different working loads is operated on the machine. The Monte Carlo approach was designed to extract a set of M completely new degradation patterns (Monte Carlo simulations) $HI_j(t)$ with $j = 1,2,...,M$ (see figure 6.10b). The variability was introduced by sampling the degradation *rate* of each single piece of the function from a Gaussian distribution and sampling the duration of each single piece Δt from another Gaussian distribution (see figure 6.10a). Applying equations 6.22-6.31 to all the M Monte Carlo samples, it was possible to extract the distribution relative to the $i-th$ EoL point estimate as the collection of the $\hat{t}^{PE}_{EoL,i,j}$. It was also possible to define all the last time instants of the real patterns $HI_j(t)$:

$$t_{EoL,true,j} \Rightarrow HI_j(t = t_{EoL,true,j}) = 0 \tag{6.32}$$

with $j = 1,2,...,M$ and their collection $t^D_{EoL,true}$. The prediction error is defined as:

$$e_{i,j} = \frac{\hat{t}^{PE}_{EoL,i,j} - t_{EoL,true,j}}{\hat{t}^{PE}_{EoL,i,j}} \tag{6.33}$$

while the error distribution associated with the $i-th$ estimate e^D_i is the collection of the $e_{i,j}$ with $j = 1,2,\ldots,M$. t_{EoL} pdf can instead be computed as:

$$\hat{t}^D_{EoL,i} = \hat{t}^{PE}_{EoL,i}(1 - e^D_i) \tag{6.34}$$

The described Monte Carlo approach (up to eq. 6.33) was used to test the performances of the point estimate module, either in terms of estimation error mean or variance. Here, distributions shown in figure 6.10a were hypothesised.

6.4 Results

6.4.1 Diagnosis

Diagnosis results for all types of working cycle were shown in table 6.4. The column labels are explained in the following:

- SC12: classification rates for each component based on 10-fold CV for stationary working cycles 1 and 2;
- NSC1 and NSC2: classification rates for non-stationary working cycles with training done using stationary working cycles 1 and 2;
- suffix FS indicates that the classification rates were obtained with the application of the feature selection.

Table 6.4: *Classification rates of the five best algorithms.*

LDA	SC12	NSC1	NSC2	SC12 FS	NSC1 FS	NSC2 FS
Pump	100.00%	99.79%	99.42%	100.00%	100.00%	99.98%
Sensor	64.14%	78.03%	76.23%	65.79%	75.21%	77.50%
Valve	72.18%	76.85%	69.91%	72.31%	79.00%	75.97%
RFC	SC12	NSC1	NSC2	SC12 FS	NSC1 FS	NSC2 FS
Pump	95.88%	97.62%	96.50%	100.00%	100.00%	100.00%
Sensor	79.10%	56.85%	56.60%	79.10%	56.85%	56.60%
Valve	88.10%	67.94%	60.74%	88.10%	67.94%	60.74%
ADB	SC12	NSC1	NSC2	SC12 FS	NSC1 FS	NSC2 FS
Pump	100.00%	99.91%	99.79%	100.00%	100.00%	100.00%
Sensor	69.84%	70.84%	79.84%	71.32%	74.28%	76.62%
Valve	73.73%	77.34%	74.05%	73.73%	77.34%	74.05%
CNN	SC12	NSC1	NSC2	SC12 FS	NSC1 FS	NSC2 FS
Pump	99.49%	100.00%	99.76%	99.98%	100.00%	99.97%
Sensor	81.23%	79.67%	85.16%	79.75%	90.94%	82.61%
Valve	76.62%	77.61%	77.82%	74.93%	81.62%	81.57%
QDA	SC12	NSC1	NSC2	SC12 FS	NSC1 FS	NSC2 FS
Pump	96.76%	99.86%	99.00%	100.00%	100.00%	100.00%
Sensor	63.70%	49.75%	54.98%	70.12%	66.18%	73.87%
Valve	71.13%	81.27%	75.51%	71.41%	79.00%	76.41%

6.4.2 Prognosis

In order to evaluate the performances of the algorithm, a Monte Carlo based analysis was carried out. In figure 6.11, the evolution of the EoL time distributions with respect to the estimation number were shown. Parameters of the distributions shown in figure 6.10a were hypothesised. As expected, as long as the time of the estimate is approaching the real End-of-Life time of the component, the distributions of the estimated EoL time are getting better, both in terms of expected value and variance. It is worth noting that the first estimate is underestimating the real EoL time. This is due to the fact that, at time instant $t_0 = 0$, the algorithm is considering $HI_0^- = 100\%$ and $HI_0^+ = 100\% + \delta HI$. This leads to an overestimation of the real $HI(t)$ pattern (fig. 6.12). At $t_0 = 0$, HI^- is coincident with the real pattern, while in other t_i, the real pattern is exactly the mean value of HI_i^- and HI_i^+. This operation is lifting up the first point of the pair over which the linear interpolation is computed at t_1. The overestimated pair of HI then leads to an excessively negative predicted rate of degradation which turns into an underestimation of the RUL. This issue can be easily addressed by adding a correction term in HI_0^- and HI_0^+:

$$HI_0^- = 100\% - \frac{\delta HI}{2} \tag{6.35}$$

$$HI_0^+ = 100\% + \frac{\delta HI}{2} \tag{6.36}$$

By adopting this correction, the lift effect due to the first HI, overestimation could be softened. However, the prognostic solution was, estimate by estimate, converging

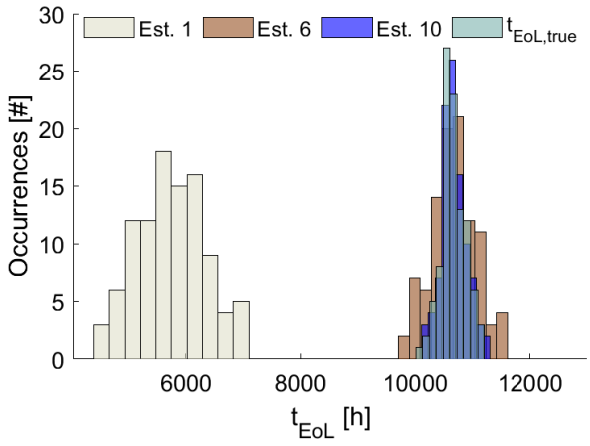


Figure 6.11: *Estimates distribution of t_{EoL} evolution with respect to the estimate number. Note that the estimates are ten but just three are shown in order not to make graphical confusion due to superimposition. Also, the distribution of the real EoL time of Monte Carlo samples is shown*

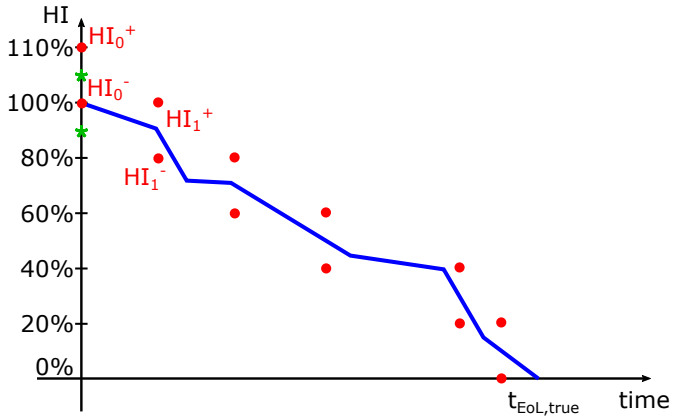


Figure 6.12: *The underestimation problem of the first estimate is due to the different selection criteria of the HI_0^+ and HI_0^-. The corrected HI are shown by asterisks*

on the true EoL time. Such a phenomenon was also evident in the distribution of the prediction errors. In figure 6.13, the bias of the first estimate is evident, while the progression of the estimate is reducing the errors. The error mean and the 95% interval of error trends during the life of the component were plotted and showed the tendency to become null and decrease respectively over the time. Furthermore, the behaviour of the algorithm with respect to the different number of levels adopted in the classification procedure was analysed. In this case, 7, 9, 10, 11 and 13 levels of classifiers were tested.

In figure 6.13, an insignificant difference between the different classifiers is shown: the classifier with the lowest number of fault states shows slightly smaller error intervals, but at the same time, the last estimate is much earlier during the life of the component. In this way, the last estimate of the classifiers with a higher number of fault states are further towards the component EoL allowing a narrower error interval (note that the first estimate is performed earlier but with a bigger error interval). On the other hand, classifiers with a higher number of fault states, require a higher number

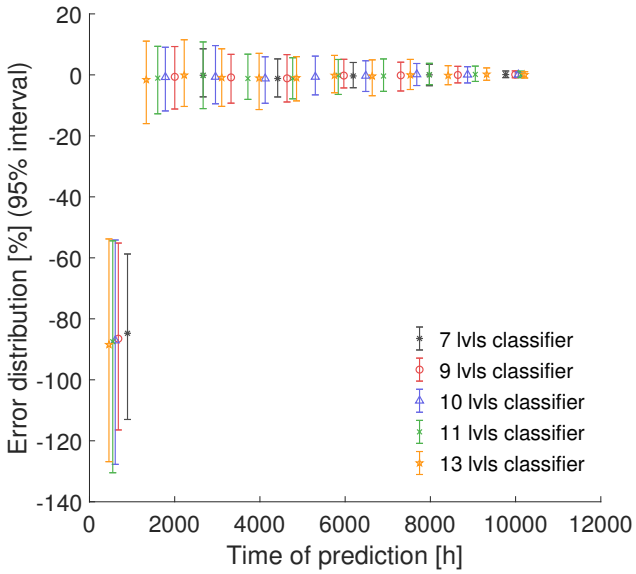


Figure 6.13: *95% error intervals for different classifiers during a component's life. Five different classifiers are tested with 7, 9, 10, 11, 13 fault states*

of tests for training and tend to be more critical in the diagnosis phase. Analysis of performances of the algorithm with respect to other distributions is out of the scope of this chapter and will be a matter for future work by the authors.

6.5 Conclusions

In this chapter, a PHM solution for a machine tool hydraulic unit was presented. Despite the hydraulic unit being one of the most critical part of machine tools [29, 161, 217], scientific literature was still lacking in this research field. The unavailability of a large amount of faulty data in the life of a machine tool brought about the decision to implement a digital twin of the hydraulic unit. The model was used to generate simulations of the healthy and faulty machine during multiple working cycles:

- such a solution was demonstrated to be efficient in addressing the working regime variability, i.e. the main limitation for the applicability of prognosis approaches in industry;
- the use of a digital twin allowed the support of the sensorisation and the design of experiments for a future validation of the model under fault conditions;
- a tailored multi-classifier solution was developed for any component, whereas typical literature solutions are based on a single classifier approach. QDA performed an excellent pump fault diagnosis, while CNN was the best classifier for sensor and valve faults.
- the proposed prognosis solution took into account the interaction between different faults, exploiting diagnosis outputs trained on all the fault combinations.

- the developed algorithm was able to estimate the RUL probability density function through a Monte Carlo approach.

Proposals for future works include the deployment of the algorithm on a test rig of the system, experimental validation of the digital twin in the presence of faults based on feature selection and undersampling support, and robustness tests on new and unseen working cycles.

CHAPTER 7

Process and tool monitoring through machined surfaces images

This chapter proposes a high-level monitoring and diagnosis tool oriented to production management, both for cutting process and tool. Starting from the analysis of machined surface images, the technological signature of the process is analysed to diagnose the status of the cutting process or the cutting tool. The proposed approach allows to make inference about the process, detecting suitable and not-suitable cutting parameters. Undesired effects such as tool chipping or excessive run-out are detected, too.

7.1 Introduction

Industry 4.0 introduced the paradigm of smart manufacturing plants and machining centres. The term *smart* underlies the capability of manufacturing technology to use advanced data analytics to improve the machining performances and supporting the operator [147, 205]. In the last decades, many techniques were developed to deal with this goal, most of them being part of the world of artificial intelligence. Research in artificial intelligence received a boost especially due to advancements in information processing technology and in machine learning techniques, capable of learning models and decision rules on training sensorial or pictorial data [115]. Machine learning algorithms will constitute the benchmark for the approach conceived in this chapter. With the advent of cheaper sensors, increased storage capacities and more powerful computers, deep learning started to gain attraction with the implementation of neural networks and the discovery of more effective training algorithms for them. Deep learning represented the natural evolution of machine learning, gaining the capacity to automatically identifying discriminatory characteristics of signal and pictures from training data [160, 209]. Machine learning firstly, and deep learning, secondly, were involved

in many fields related to image processing and classification ranging from land-use categories classification [78], object recognition and classification in road scans [124], traffic signs recognition [149], texture analysis in turning processes [59] and surface defects recognition [58, 180, 190], prediction of car prices from images [30], robot assembly operations [30] and manufacturing trajectory smoothing [127].

In this chapter, a deep learning approach is proposed with the aim of improving the choice of machining parameters and to detect undesirable machining conditions. The conceived approach is based on a shallow convolutional neural network (CNN) capable of detecting wrong cutting speeds, feed per tooth, machining conditions as well as technological parameters from machined workpiece surface images. Thus, from the CNN, the operator can retrieve useful information on how to take corrective actions on process parameters to reach higher production quality levels for the following parts to be machined.

The structure of this chapter is set as follows: in section 7.2, a state-of-the-art of traditional texture descriptors, machine learning and deep learning classifiers is presented; in section 7.3 the setup, the experimentation and the dataset presentation are thoroughly explained, together with the definition of the f1-score performance metric; in section 7.4.1, state-of-the-art texture descriptors extraction and machine learning classifiers with application similar to those of this chapter (with benchmark purposes) are described; in section 7.4.2, the conceived CNN architecture and hyper-parameters selection is set out; in section 7.5, the f1-score performance metric (mean and standard deviation) is compared among the state-of-the-art approaches and the developed CNN, together with a proper discussion of the results; at last, conclusions are drawn at the end of the chapter, in section 7.6.

7.2 State-of-the-art

7.2.1 Traditional texture descriptors

In order to prepare images for machine learning classification, it is necessary to manually extract features from them. Traditional texture features may belong to the spectral, structural and statistical world [59] or coming from fractal analysis [110]. When dealing with milled surface images, statistical features are the most used and relevant for the analysis. Three of them are the most used: Gray-Level Co-Occurrence Matrix (GLCM), Histogram of Oriented Gradients (HOG) and Local Binary Pattern (LBP).

GLCM is a second-order statistical texture descriptor, i.e. accounting for the relative position of two pixels. In fact, GLCM computes the number of occurrences of pairs of pixels with a given intensity and a given displacement [59]. GLCM was introduced by Haralick *et al.* [78] and was widely used in many image classification and analysis problems, while being selected as a benchmark for deep learning algorithms performances evaluation [58, 59, 190]. HOG computes the frequency of occurrence of the orientation of the gradient in a localised portion of an image. HOG was created for object detection in images and was adopted in image classification and analysis problems [30, 60, 160] (even as a benchmark). LBP computes the histogram of a transformation to decimal number of 8 digit binary numbers obtained on 3x3 pixel cells of the image. LBP was identified as one of the most used and effective statistical descriptors for classification tasks on textures. Its success was related to its invariance with

respect to grayscale and rotation. Several modifications to its definition were also developed to overcome some related challenges, e.g. noise sensitivity [180]. LBP found applications in several contexts, not only related to machining: Song *et al.* applied it for defect recognition [180], Garcia-Ordas *et al.* employed it for tool wear monitoring based on computer vision [60], Fu *et al.* used it as a benchmark for steel surface defect classification [58] and Hou *et al.* used it as a benchmark for cancer sub-types classification [86].

7.2.2 Machine learning classifiers for texture descriptors

Machine learning classifiers are based on the manual extraction of features/descriptors. Thus, GLCM, HOG and LBP are given as input to machine learning approaches to assign classes to the original data. Machine learning approaches can perform such task without being explicitly programmed, but by proper *learning* on training data [115]. Some of the most common machine learning approaches for classification purposes are k-Nearest Neighbours (KNN), Support Vector Machine (SVM) and Random Forest (RF) classifiers [13, 68].

KNN is a simple classifier based on the maximum occurring class in the k nearest training data according to a specific distance measure [57]. It was widely used in literature for classification tasks, such as steel surface defects detection [58, 180]. SVM finds the optimal hyperplane to separate features belonging to two or more groups. It found applications in the medical sector through image classification [91], steel surface defects detection [180], object detection in road environments [124], tool wear classification [60] and benchmarking [30, 58, 176, 212]. RF combines a set of trivial classifiers (called decision trees) and select the output class based on a voting system [24]. RF were used in classification tasks such as object detection of RGB-D images (RGB images with depth information), unordered point set context prediction [70] and as a benchmark [160].

7.2.3 Convolutional Neural Networks are end-to-end classifiers

Deep learning methods typically assume the shape of deep neural networks. Deep learning has the advantage, over machine learning, to automatically extract discriminatory features and descriptors from the training data. In this way, they directly operate on the raw input data/images. Nevertheless, they typically require a huge amount of training data to learn how to classify the provided data [209], thus generally preventing its use for limited datasets. The most used architecture for the classification of images is the CNN [86]. CNNs are based on stacked convolutional layers that perform automatic extraction of low-level and high-level features across the layers, while simultaneously classifying images with the last one. During the recent years, many *standard* architectures have succeeded each other, trying to solve commonly faced problems and increase CNNs performances. The deeper the network is, the more accurate and powerful it is. Based on this evidence, He *et al.* [80] introduced the Residual Neural Networks (ResNets). ResNets allowed to overcome the problem related to the fact that a deeper and more sophisticated network architecture may collapse on its shallower counterpart simply with a layer learning the identity function [80]. Facing the problem of gradient vanishing effect during training of deeper and deeper CNNs, Huang *et al.* proposed Dense Networks (DenseNets) by connecting with short paths multiple layers

of CNNs [90]. Such architecture reached ResNets performances with fewer parameters, less computational costs and less optimisation difficulties. Tan *et al.* introduced a novel compound method to uniformly scale networks architectures, developing the Efficient Networks (EfficientNets) [189]. Howard *et al.* proposed a new generation of Neural Networks specifically designed for their efficiency, i.e. computational costs. They were thought to work on mobile devices and were thus called Mobile Networks (MobileNets V3) [87]. At last, several researches tried to deal with large images and high resolution images, that naturally lead to computationally expensive CNN networks. Patch-based CNNs were introduced by analysing patches of large images [149, 176].

CNNs, with all their architectures, were applied on several classification tasks based on pictorial input data: object classification, street view house numbers recognition [90], industrial inspection and surface defect detection [160, 209], face detection and image recoloring [213], traffic signs detection and recognition [149] are just few examples.

Differently from state-of-the-art CNN architectures, typically featured by wide and deep structures, a shallow architecture made of 3 stacked convolutional blocks is introduced in this chapter. The choice was taken to respond to common challenges of deep learning approaches, i.e. to rely on large and exhaustive training datasets (typically featuring more than 50000 images). The CNN is thought to learn from patches of images in order to contain the number of parameters and reduce the architecture complexity when facing high-resolution images. Moreover, a CNN was never used as a mean to infer the goodness of cutting conditions, in terms of process parameters, from images of the machined surfaces. This will open, in the future, the possibility to retrieve process parameter corrective actions from the CNN for adaptive control strategy purposes.

7.3 Materials

This section presents the milling parameters of interest for this work and how the ability of producing many specimen influenced the overall classifier development process in terms of data preprocessing and evaluation metric.

7.3.1 Experiments

The research activity was focused on developing models that are able to recognise and classify different process and technological parameters for milling operations by leveraging only images of machined surfaces. The parameters of interest were:

- process parameters: machining conditions, feed rate and cutting speed;
- technological parameters: tool diameter and insert nose radius.

An experimental campaign was designed and carried out in order to produce the specimens necessary to collect the images to be analysed with the classification models. The specific values of the parameters of interest to the purpose of this work were reported in Table 7.1. To reduce the required number of specimens, machining costs and times, it was decided to mill up to 2 sides of the prismatic specimens, and to machine different portions of the same surface with different combinations of cutting parameters (fig. 7.1).

The starting dataset consisted of 100 very high resolution (on average about 8 megapixels) RGB images with a portrait aspect ratio, similar to the one reported in Fig. 7.3.

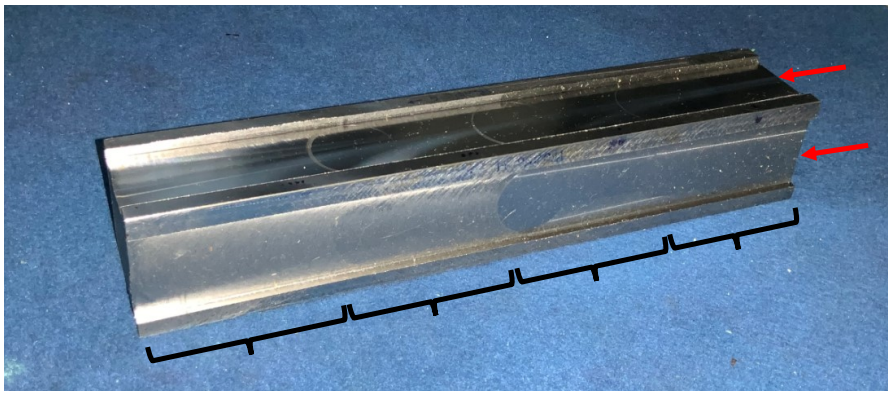


Figure 7.1: *Example of a specimen used for the milling tests. Red arrows show machined sides. The surfaces are machined in portions with different combinations of cutting parameters indicated by black braces.*

Table 7.1: *Parameters of interest for which the experimental campaign was carried out.*

Parameter type	Parameter	Values
process	machining conditions	nominal, run-out, chipped tooth
	feed rate [mm/tooth]	0.10, 0.15, 0.20, 0.60, 0.70, 0.80
	cutting speed [m/min]	80, 150, 220
technological	tool diameter [mm]	27.9, 32.0
	nose radius [mm]	0.4, 0.8, 1.0, 1.2, 1.5

The images were collected with a Keyence VHX-7000 digital microscope focused on a machined area of the specimen (using 20x magnification). Different combinations of milling parameters lead to different patterns on the surfaces, and some of these patterns can be recognised with ease, some can not. The objective of the classifiers was to correctly recognise the aforementioned parameters by analysing the input images. The use of a digital microscope allowed to have a uniform light exposure of taken images. A similar solution could be achieved for an industrial online system. In this case, a semi-autonomous robotic arm was prepared for a future application and validation of the solution. An ABB robot was set-up to work on the side of the machine tool pallet. It features an AVT Prosilica GT2300 cam with optics, assisted by a proper lighting system (four LED Bar CCS in addition to a LED Ring CCS). This set-up (fig. 7.2) has the objective to reproduce a favourable industrial environment for the conceived approach.

7.3.2 Data preprocessing and dataset preparation

The collected images featured a very high resolution (about 8 megapixels). Independently of the classification approach, dealing with a high number of pixels would be cumbersome and, more importantly, would imply excessive computational costs. For context, state-of-the-art classification models were trained with image sizes ranging from 32x32 [30, 80, 90, 160, 209] to 600x600 pixels [189], with one of the most commonly adopted being 224x224 pixels [13, 30, 58, 80, 87, 90, 189]. Furthermore, due to

Figure 7.2: *Robotic arm set-up for a future validation of the conceived approach in an industrial environment.*

experimental costs related to machining, the resulting dataset was quite small, and consisted of 100 total samples. For reference, to train state-of-the-art classification neural networks, researchers leveraged datasets such as Imagenet [46], CIFAR-10 [117] and MNIST [123] consisting of 14 million, 60000 and 70000 images, respectively.

To overcome these limitations, it was decided to split each raw image into smaller patches [60,86,149,176] with a dimension of 224x224 pixels. Thanks to this operation, it was possible to:

- reduce the dimensions of the input images, dramatically decreasing the computational costs required to develop and evaluate the models;
- increase the number of samples in the dataset, since each raw image got split in about 150 patches.

With that being said, it must be highlighted that patches from the same image should not be treated as completely distinct samples: as visible from Fig. 7.4, two neighbouring patches were likely characterised by similar patterns. If one of those patches would be used to train the classifier while the other one to test it, their similarity could be a source of information leakage, i.e. leading to an overestimation of the model prediction capabilities and hiding the model actual prediction performances. Therefore, to avoid any information leakage between train and test partitions, it was decided to proceed as follows:

1. for each parameter of interest, separate the 100 available raw images into train and test partitions;
2. split each raw image into as many 224x224 pixels patches as possible.

Consequently, the image preprocessing pipeline is completed by the following steps, which are commonly adopted in the literature [190]:

1. convert the images' color space from RGB to grayscale;

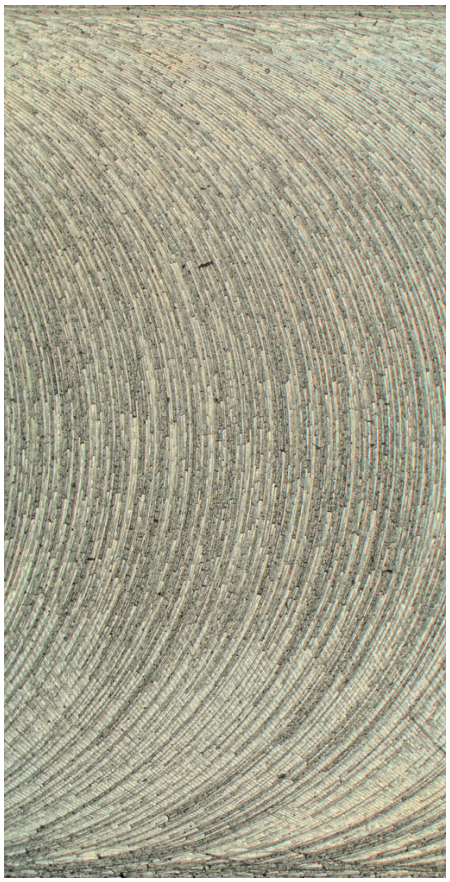


Figure 7.3: *Raw sample image of milled surface.*

2. scale the pixels values of each sample from the range [0,255] to the range [0,1];

3. centre each sample by subtracting the mean pixel value (this mean value is computed from the training partition only and then applied to both training and testing partitions).

A factor that can negatively influence the performance of the classifiers is the different percentage of samples within each class, which leads to unbalanced datasets. For instance, in the case of the machining condition parameter, collecting data from undesired machining conditions is not trivial and can be expensive as well as risky. Thus, among the 100 available samples, a large portion belonged to *nominal conditions*, whereas only a limited number of samples to *chipped tooth* and *run-out*. Having less samples for a given class means that the classifier will struggle to learn relevant features for that class, hindering its overall performance. To compensate this limitation the classifiers were developed leveraging a 5-fold cross-validation process. This process consisted in evaluating a classifier by looking at a metric averaged over 5 different train-test splits. At each split, a different set of 20% of the available samples was reserved for testing. Furthermore, since some target parameters (i.e. process and technological ones) presented classes populated by a very limited number of samples (less than 10), a stratified split approach was used to preserve the percentage of samples

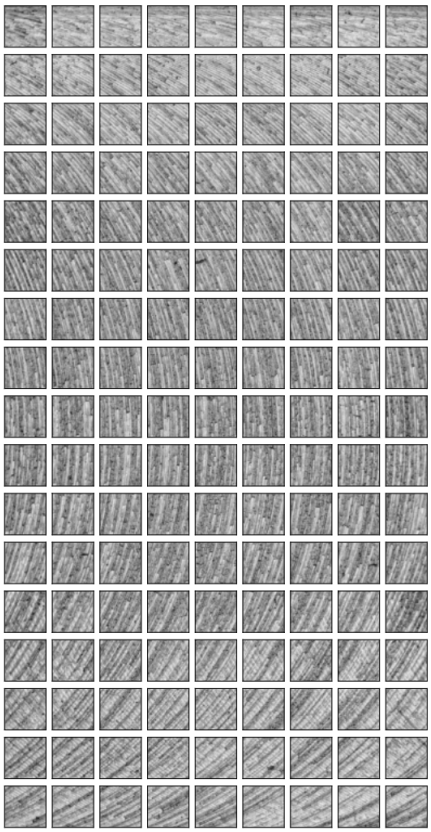


Figure 7.4: *Preprocessed sample image of milled surface: patches of 224x224 pixels are extracted from the image, then converted to grayscale. If this sample image is selected for training purposes, then all of its patches are assigned to the training set.*

for each class in each train-test split.

7.3.3 Evaluation metric

Because of its simplicity and interpretability, the *accuracy* is the most widely used classification metric, and it is defined as:

$$accuracy = \frac{correct\,predictions}{total\,predictions} \tag{7.1}$$

However, this metric is not suitable for unbalanced datasets. For instance, if a model was to be trained with a dataset consisting of 99 images collected from *nominal machining conditions* and 1 image collected from *run-out*, it could simply learn to assign the 100 samples to the first class and still score 99% accuracy. Once deployed to production, it would still assign any new sample to the first class, since it did not learn how to distinguish between different machining conditions.

For this reason, a different classification metric was chosen to better quantify the performance of the classifiers when dealing with unbalanced datasets: the f-score. This

metric is defined in [36] as the harmonic mean of precision and recall:

$$F_\beta = \frac{(\beta^2+1)(P \cdot R)}{\beta^2(P+R)} \tag{7.2}$$

where:

- β is a parameter that controls a balance between precision and recall;
- P is precision, defined as $\frac{true\ positive}{true\ positive\ +\ false\ positive}$;
- R is recall, defined as $\frac{true\ positive}{true\ positive\ +\ false\ negative}$.

When the same weight is assigned to precision and recall, meaning $\beta = 1$, the f-score becomes what is commonly referred to as f1-score [171]:

$$F_1 = \frac{2PR}{P+R} \tag{7.3}$$

Considering the previous example, the classifier returns an f1-score equal to 0, highlighting the fact that it actually did not learn anything useful and can't be deployed to production.

In Section 7.5, the classifiers performance will be evaluated, for the 5-folds, in terms of mean value and standard deviation of testing f1-score.

7.4 Methods

This section presents the traditional texture descriptors, namely LBP, HOG and GLCM, that were used to train the machine learning models (SVM, KNN, RF) to classify the milled surfaces images according to the parameters of interest (explained in Section 7.3.1). The results obtained through these approaches will constitute the benchmarks for the conceived deep learning approach performances. As a final remark, it is necessary to remind that the data preprocessing steps implemented before computing the texture descriptors are the same presented in Section 7.3.2, and the evaluation metric is the one presented in Section 7.3.3.

7.4.1 Theoretical background

Traditional texture descriptors: LBP, HOG, GLCM

From a general point of view, the common starting point for each traditional texture descriptor are the preprocessed images. Once the descriptors are computed, it is possible to proceed in two ways:

1. use the resulting images as inputs to the ML models;
2. perform a further step and compute features from the resulting images, and use those features as input to the ML models.

Because the ML models implemented in this chapter are not compatible with images (images in the form of 2D arrays), it was decided to implement option 2. It could be argued that the images can be converted into a stack of 1D arrays, but this implies working

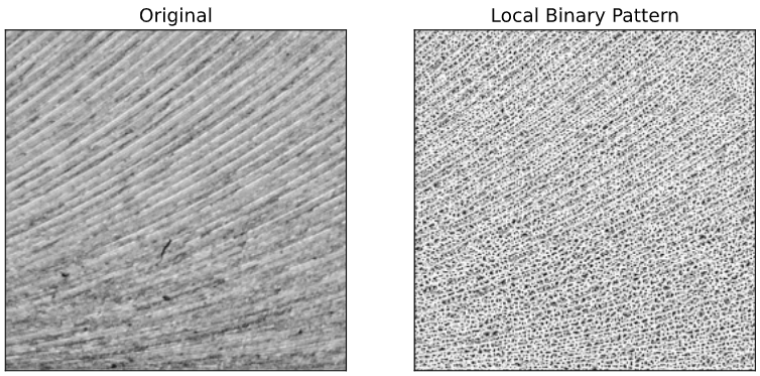


Figure 7.5: *Left: 500x500 pixels bottom-left corner of a sample image. Right: 500x500 pixels of the same portion plotted in terms of LBP values.*

with very large inputs, that increase dramatically the computational costs. Furthermore, it would lead to the state where the number of features vastly exceeds the number of samples, commonly referred to as *curse of dimensionality* [16]. Consequently, once the LBP, HOG and GLCM descriptors were computed for each starting sample, a further step was added to obtain a set of feature vectors.

LBP computes a histogram with the distribution of the binary configurations of the pixels of the image, based on thresholding the surrounding window of each pixel with the intensity of the centre pixel. Generally, the LBP descriptor works on 3x3 pixels cells, and the centre pixel's value sets the threshold. Each neighbouring pixel is converted into a binary value according to this criteria:

- if the neighbouring pixel's value is larger than the threshold (i.e., 3x3 window central pixel value), the pixel is converted to 1;
- if the neighbouring pixel's value is smaller than the threshold (i.e., 3x3 window central pixel value), the pixel is converted to 0.

The resulting sequence of zeros and ones is converted to a binary number and thus to an integer number (according to the binary conversion system). This value is assigned to the centre pixel of the starting 3x3 cell; the same procedure is repeated for all the other pixels in the image to obtain an output image. Figure 7.5 clarifies how a sample 500x500 pixels window changes when LBP is computed, while Figure 7.6 is relative to a 100x100 pixels window from the same original sample. A histogram of the output image was computed and its histogram bin values were selected as the feature vector. The idea is that surfaces that show different patterns should present different LBP histograms, and thus, the bins should have different values. Indeed, it should be possible to train ML classifiers to recognise those differences.

HOG descriptors are mainly used to describe the structural shape and appearance of an object. However, since HOG is able to capture the local intensity gradients and edge directions, it is even a good texture descriptor. By focusing on the gradient and orientation of the edges (magnitude and direction), it is possible to understand if the pixels belong to an edge and find its orientation. The idea is to leverage the edges and their orientations to distinguish between different cutting parameters that generate different

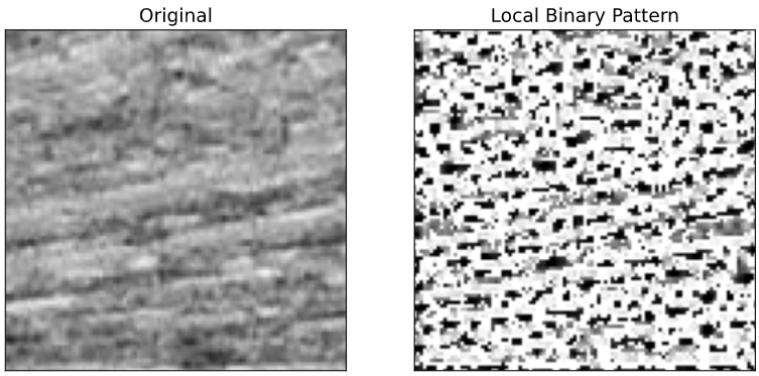


Figure 7.6: *Left: 100x100 pixels bottom-left corner of a sample image. Right: 100x100 pixels of the same portion plotted in terms of LBP values.*

patterns on the metallic surfaces. The orientations were calculated in localised portions, meaning that the starting image was decomposed into smaller regions (or *cells*) and, for each region, the gradients and orientations were calculated. Consequently, a histogram for each region was generated from the gradients and orientations of the pixel values. An example output from the HOG descriptor is reported in Fig. 7.7 for a window of size 500x500 pixels and Fig. 7.8 for a window of size 100x100 pixels. Moreover, from Fig. 7.8 it is possible to see how the starting image was decomposed into smaller regions (with dimensions 32x32 pixels) and a total of 9 possible orientations was set. To limit the issues caused by the very high dimensionality implied by the adoption of this texture descriptors, it was decided to pass the HOG values through a principal component analysis (PCA) [108] stage to obtain a feature vector with length 8 for each sample.

GLCM consists in computing statistical experiments on the matrix (or matrices) containing the co-occurrences of the pixel intensities at given angles and distances. Such statistics experiments intuitively provide measures of properties such as smoothness, coarseness and regularity on the distribution of pixels on the texture. By definition a GLCM is the probability of the joint occurrence of gray-levels i and j, where $i \leq G$

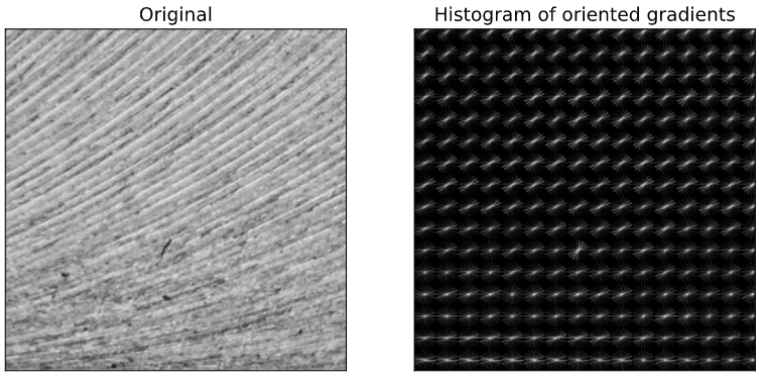


Figure 7.7: *Left: 500x500 pixels bottom-left corner of a sample image. Right: 500x500 pixels of the same portion plotted in terms of HOG.*

and $j \leq G$ and G identifies the gray-level depth of the image, within a defined spatial relation in an image. This spatial relation is defined in terms of a distance D and an angle θ. From the GLCM, it is possible to compute statistical features that can then be stacked together to build the feature vector. In particular, the features evaluated according to [77] were:

- contrast;
- dissimilarity;
- homogeneity;
- energy;
- correlation.

In this work, to compute the 5 features it was decided to set D equal to 100 and θ equal to 0°. Those values might not be the best ones, and evaluating the best possible combination is actually very time consuming since a sweep of all possible combinations would be required. This is one of the disadvantages of trying to classify images when leveraging highly engineered features such as GLCMs. An example of the evaluated GLCM was reported in Fig. 7.9 for two different values of cutting speed. It is possible to notice that the resulting GLCMs were slightly different, and these matrices were leveraged to compute the statistical parameters mentioned above. In this example, the difference between the two raw samples was highlighted by the *contrast* feature, while the other features didn't highlight large differences.

Machine learning classifiers: SVM, KNN, RF

To establish a performance baseline for the classification problem presented in this chapter, it was decided to utilise three of the most widely known supervised machine learning classifiers, namely SVM, KNN and RF.

Considering a binary classification problem, meaning classifying between just two classes, the objective of the SVM was to find the hyperplane that maximises the distance to the nearest input data point of any class, as per Fig. 7.10 (the distance is called

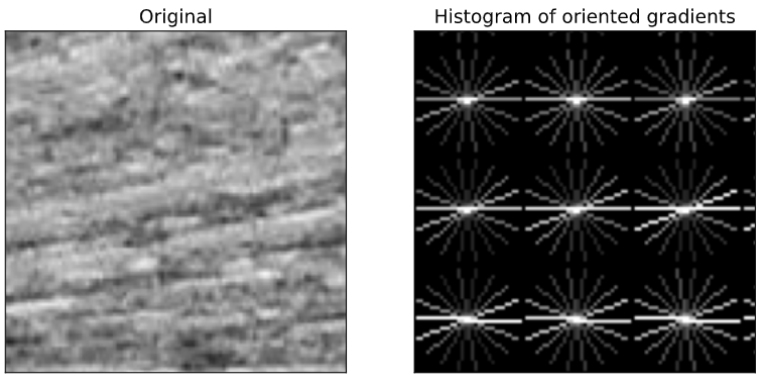


Figure 7.8: *Left: 100x100 pixels bottom-left corner of a sample image. Right: 100x100 pixels of the same portion plotted in terms of HOG.*

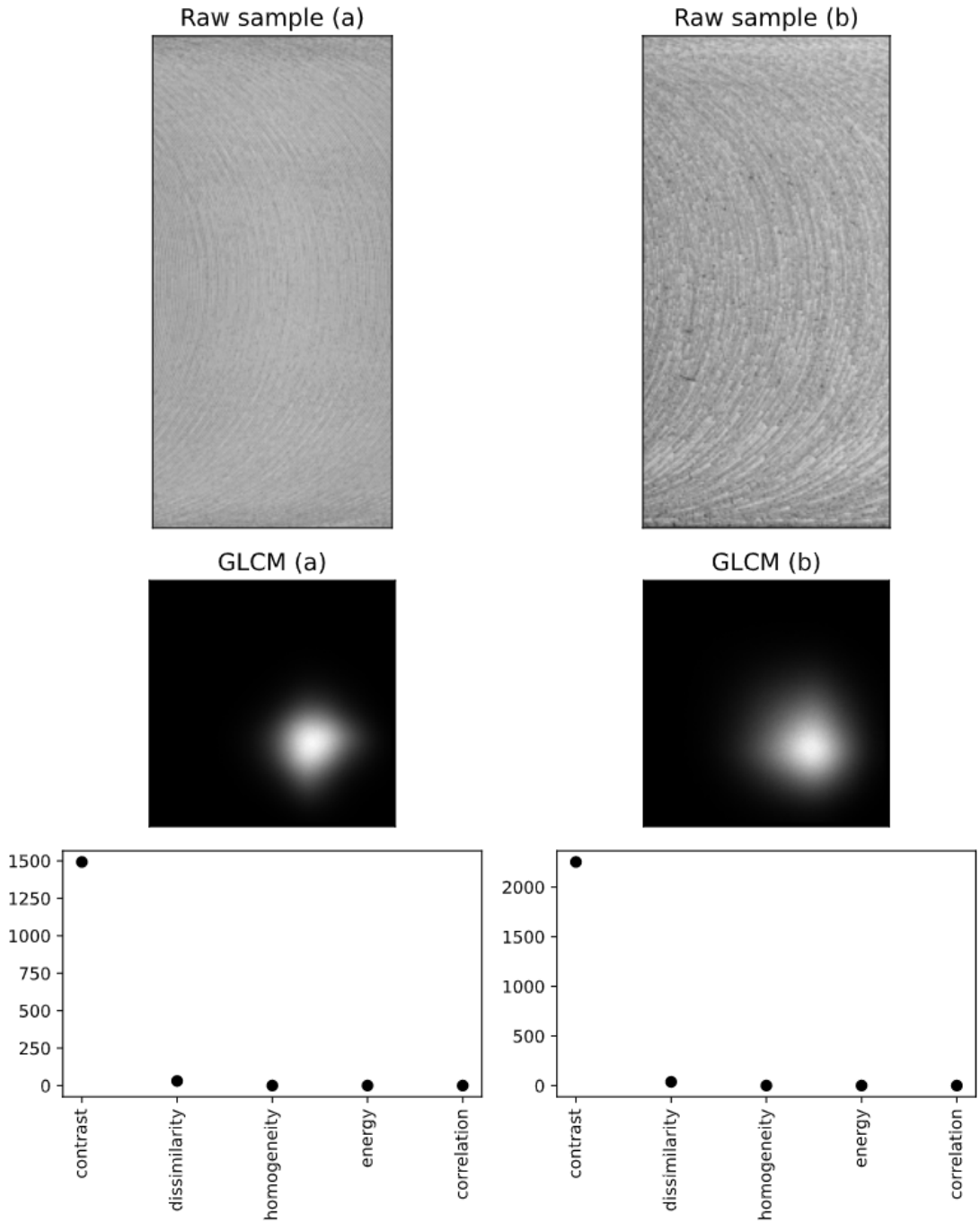


Figure 7.9: *Top: two raw samples representing different cutting speed values (a: high cutting speed, b: low cutting speed). Middle: GLCM from raw samples (a) and (b). Bottom: features computed from GLCM (a) and (b) respectively.*

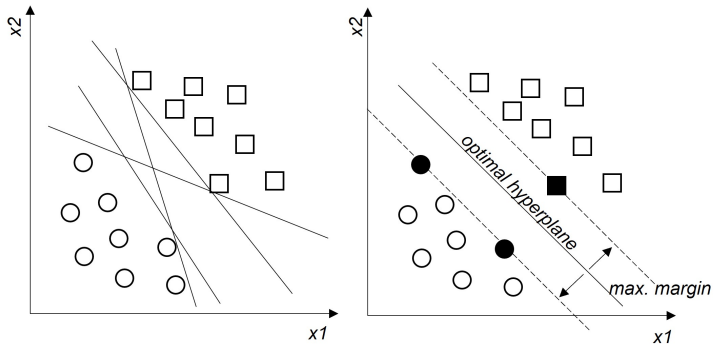


Figure 7.10: *SVM defines the hyperplane that maximises margin between the samples of the two classes.*

margin, the nearest sample is called support). SVMs showed to be able to perform both linear and non-linear classification; for the latter, a *kernel trick* [21] is exploited to map the input data onto a higher dimensional space. The idea behind the *kernel trick* is that if the input data are not linearly separable in the current space, they may be separable in a higher dimensional one, as shown in Fig. 7.11, computed via combinations of simple functions and the starting input data. Two of the most known kernel functions are polynomial and gaussian radial basis functions. The disadvantage of the *kernel trick* is that by increasing the dimensionality of the problem, the testing error increases, too; thus, to compensate this, more data are required to train the SVM.

The KNN classification algorithm [57] is quite intuitive: for any new sample, a distance metric is computed to find the k-nearest samples and assigns it to a class according to a voting mechanism based on the most recurring class among the k-neighbours. The main parameter that requires to be tuned is k, which sets the number of neighbours considered to classify a new sample (i.e. participate in the voting mechanism):

- if k is too high the testing error increases because the boundaries between classes are less distinct. This means that samples of different classes can be mixed together and assigned to a single region (underfitting). When a new sample falls

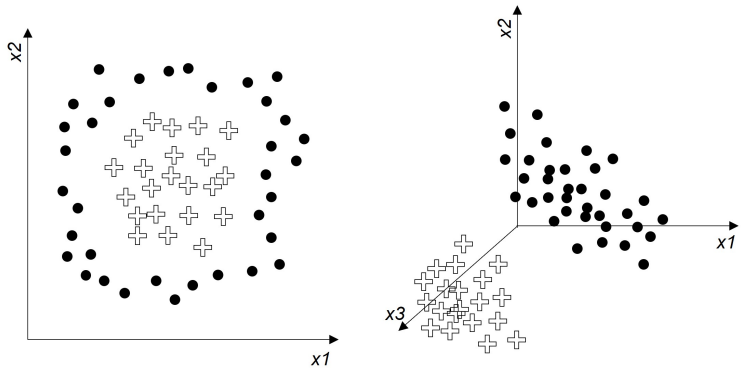


Figure 7.11: *SVM: example of kernel trick application and how it allows to define a plane that separates the two classes in a higher dimensional space.*

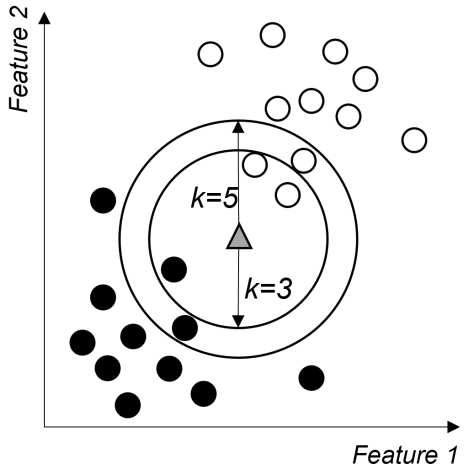


Figure 7.12: *KNN: example of classification with different neighbours number.*

within that boundary, it may be incorrectly classified because the most recurring class among its k-neighbours is actually different from the one of the region (which is the correct one).

- if k is too low the computational cost and time increase dramatically, especially if the dataset is large and the samples have high dimensionality. Furthermore, KNN may end up learning the noise (overfitting).

So, k is usually tuned by applying some optimisation methods that test different possible values and evaluate the performance of the KNN classifier accordingly.

Random forest classifiers are built by combining multiple decision trees that operate as an ensemble [24]. The main features that characterise the random forest classifiers are:

- bootstrap aggregating algorithm (bagging) is leveraged to assign the input data to each decision tree;
- each decision tree operates on a random subset of features to reduce the correlation between different trees;
- the classification output of the random forest is selected based on the most voted class by the decision trees.

To better adapt the random forest classifier to the specific task it is possible to tune a variety of parameters, such as: the number of estimators (decision trees), the number of features for each estimator and the maximum depth of each estimator. The number of estimators sets the number of decision trees that will be used to build the random forest classifier. It is not always clear if this parameter should be tuned or set to the maximum value possible given the available computing resources [163]. The number of features for each estimator sets the number of features that will be leveraged by each

decision tree to classify any given sample, and it is usually set equal to the square root of the number of total features available. The depth of the estimators can be limited to a maximum value to avoid overfitting the training data and to improve the bias-variance trade-off (very deep decision trees lead to very good performance at the expenses of becoming very sensitive to small variations in the input data, which leads to large variations in the predicted class).

From an operative perspective, the three algorithms (SVM, KNN, RF) were tuned and trained with a common strategy that leveraged a combination of cross-validation and grid-search. As reported in the pseudo-code in Algorithm 1, the first step was loading the data (meaning the raw images) and computing a set of traditional feature descriptors such as LBP, HOG or GLCM. The different sets of tunable hyper-parameters characterising each model in this work were the following:

- SVM: C, kernel, polymonial kernel degree;
- KNN: number of neighbours, weights, power parameter for Minkowski distance;
- RF: number of estimators, maximum number of features,

For each hyper-parameter multiple possible values were set. The second *for* loop split the dataset in 80% -20% training-testing proportion. In the third *for* loop the model was tuned and trained. Here, two approaches were deployed at the same time: grid-search was used to generate all possible hyper-parameters' values combinations, 5-folds cross-validation was used to evaluate which combination led to the best testing results. The third *for* loop was actually repeated 20 times (see line 7 in Algorithm 1): the objective was to find the most recurring best architecture from the grid-search cross-validation process to limit as much as possible the detrimental effects of imbalanced datasets.

Algorithm 1 SVM, KNN, RF train-test loop

```
1:  set parameters
2:  set models
3:  for parameter in parameters do
4:      load data
5:      compute feature (LBP, HOG, GLCM)
6:      set hyper − parameters grid boundaries
7:      for i in range(20) do
8:          stratified train-test split of data (80-20)
9:          initialise Grid-Search as gs
10:         initialise StratifiedShuffleSplit as sss
11:         initialise Cross-Validation as cv
12:         for model in models do
13:             initialise model
14:             train and tune model with gs, sss and cv
15:             compute F1score
16:             save tuned model architecture
17:         end for
18:     end for
19:     find most recurring model architecture
20: end for
```

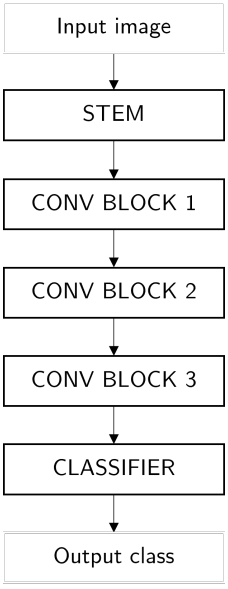


Figure 7.13: *Compact CNN architecture overview.*

7.4.2 Custom CNN for surface classification

This section explains the proposed CNN model architecture, highlighting some of the design choices that were to compensate the limitations introduced in Section 7.3.2.

It is possible to notice from the graph of Fig. 7.13 that the proposed CNN was very shallow and consisted in a series of stacked convolutional blocks. This went against the design trends of deeper (and sometimes wider) models that characterised recent state-of-the-art convolutional neural networks for image classification such as ResNets [80], DenseNets [90] and EfficientNets [189], just to name a few.

Keeping the network shallow and limiting the number of trainable parameters was necessary due to the dataset shortcomings highlighted in Section 7.3.2:

- training deep models, with many trainable parameters, leveraging only a limited number of samples would inevitably lead to overfitting. This overfit tendency would hinder the model performances when deployed to production;
- having a model with few traininable parameters facilitates the training and inference processes when dealing with limited computing resources.

Nonetheless, if necessary, it is possible to increase the depth of the proposed model by stacking more convolutional blocks. Similarly, the model can be widened by increasing the number of filters generated by each convolution operation.

From an operative perspective, the input images were fed to a first convolutional block (*stem block*), with the purpose of increasing the number of channels while decreasing the spatial dimensions. The stem was very similar to implementations found in the literature, consisting of a convolutional layer with a large filter size followed by batch-normalisation, rectified linear unit (relu) activation function and a max pool operation for further downsampling. The resulting outputs were passed through a stack of three inverted residual linear blocks (called *conv block*, inspired from Mo-

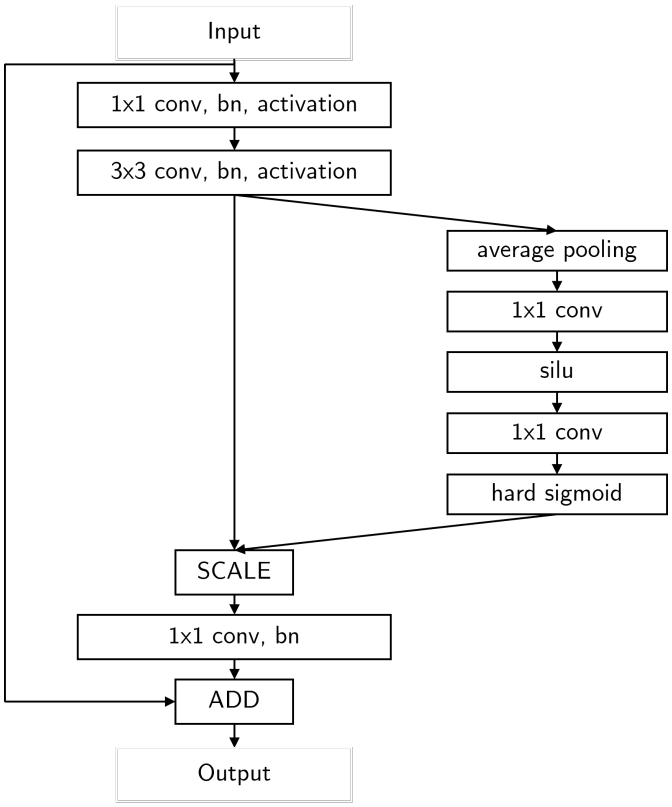


Figure 7.14: *Convolutional block detailed view.*

bileNetV3 [87] and EfficientNets) detailed in Fig. 7.14: the *conv blocks* were characterised by state-of-the-art features such as depth-wise convolutions, squeeze-excite blocks, residual connections and hard-sigmoid activation function. The inverted residual linear block was integrated in the proposed CNN and custom developed to:

- be lightweight;
- optimise the accuracy-latency trade off on limited resources (mobile devices), which are desirable features for the conceived research purposes;
- efficiently generate feature maps that synthesise meaningful information from each sample, making it possible to recognise the different machining parameters of interest.

The last stage of the proposed CNN was the *classifier block*, where the output filters were averaged and passed through a combination of convolutional layer and softmax layer in order to obtain the predicted classifications.

A detailed summary of all the parameters that defined the architecture of the proposed CNN (dimensions, kernel sizes, filters, strides, activation functions) was reported in Table 7.2. Some further details: the squeeze-excite block adopted convolutional layers as weights layers and was characterised by the absence of batch-normalisation layers, the reduction factor was set to 8; in the *classifier block* the dropout layer had a survival probability of 90%. The proposed CNN model was trained with stochastic

Table 7.2: *Custom CNN architecture.*

Output size	Layers		Activation
54x54	conv, 11x11, 32, stride=4		relu
26x26	max pool, 3x3, stride=2		
13x13	$conv, 1x1, 64$ $dwconv, 3x3, 64, stride = 2$ $conv, 1x1, 48$	x1	relu
7x7	$conv, 1x1, 96$ $dwconv, 3x3, 96, stride = 2$ $conv, 1x1, 64$	x1	hard-sigmoid
7x7	$conv, 1x1, 128$ $dwconv, 3x3, 128, stride = 2$ $conv, 1x1, 64$	x1	hard-sigmoid
1x1	global average pool dropout		
1x1	conv, 1x1, classes		softmax

gradient descent algorithm (SGD) with initial learning rate 0.9 and momentum 0.9; the learning rate was adjusted after every batch according to the *OneCycle* policy [179]. Number of epochs was set to 20 and batch size to 128.

7.5 Results

In this section the classification results obtained with two approaches were reported, being them:

- traditional texture descriptors (LBP, HOG, GLCM) and machine learning classifiers (SVM, KNN, RF);
- end-to-end classifier model (CNN).

For each process (machining condition, feed rate, cutting speed) and technological (tool diameter, tool nose radius) parameter, the classifiers performance was evaluated through the f1-scores over the 5-fold cross-validation process. The results in the following tables described the classification performances in terms of testing f1-score mean value and its standard deviation.

7.5.1 Machining condition classification results

Table 7.3 reported the testing performances of each model when classifying the samples according to the machining conditions. It is reminded that the available samples in the dataset were collected for 3 machining conditions:

- nominal;

- run-out;
- chipped tooth (insert).

When classifying the three machining conditions, the best performing model was the proposed end-to-end CNN, which was marginally outscoring the combinations KNN+GLCM and RF+LBP (0.86838 compared to 0.86667 and 0.86364). It is interesting to notice how the CNN showed slightly more reliable testing performance, highlighted by a lower standard deviation than KNN+GLCM ($\pm$0.02786 vs $\pm$0.03275). This could be an indicator of the better CNN's generalisation capabilities compared to KNN+GLCM, and of its effectiveness in learning synthetic features that capture meaningful information within each sample. With that being said, the classifier and feature pairs were able to perform very well on average, with f1-scores comfortably above the 80% threshold. An unexpected behavior was shown by SVM when paired with LBP and GLCM: SVM was completely unable to properly separate the samples when leveraging features obtained from these two texture descriptors, and this was true across the different parameters analysed (see tables in the next pages). Considering that the SVM's hyper-parameters were tuned via grid-search approach and that the samples were cycled through cross-validation, it is fair to assume that this phenomenon is due to a poor choice in the features selection process (namely in terms of type of features and quantity of features). It is also true that LBP and GLCM features did perform well when leveraged to train other ML models, so this problem appears to be limited to the SVM.

Table 7.3: *Machining condition classification results.*

ML algorithm	Texture descriptor	F1-score (5-fold)
SVM	LBP	0.10000 ± 0.00000
	HOG	0.84375 ± 0.03204
	GLCM	0.10000 ± 0.00000
KNN	LBP	0.80714 ± 0.07319
	HOG	0.85385 ± 0.04770
	GLCM	0.86667 ± 0.03275
RF	LBP	0.86364 ± 0.02335
	HOG	0.85455 ± 0.01508
	GLCM	0.82083 ± 0.04502
CNN		$\mathbf{0.86838 \pm 0.02786}$

7.5.2 Feed rate classification results

Table 7.4 reported the testing performances of each model when classifying the samples according to the feed rate. It is reminded that the available samples in the dataset were collected for a total of 6 different feed rates:

- 0.10 mm/tooth;
- 0.15 mm/tooth;

- 0.20 mm/tooth;
- 0.60 mm/tooth;
- 0.70 mm/tooth;
- 0.80 mm/tooth.

For this classification task, the best performing model was the proposed CNN, clearly outscoring all the tested models-texture descriptors combinations by quite some margin (the second best model-texture descriptor combination is
SVM+HOG, trailing by more than 10 basis points). As noticed for the machining conditions classification, the proposed CNN was very stable, registering the second-lowest f1-score spread ($\pm$0.04601). With that being said, it is possible to notice how each model was struggling to correctly classify the samples according to the feed rate. The proposed CNN was the only model that achieved meaningful performances, with the remaining models registering very low f1-scores (below 50%). This is likely due to the models assigning most of the testing samples to just one or two classes (out of the six available), therefore the training process didn't have the expected impact. One other reason could be that recognising this cutting parameter simply by looking at an image is hard even for human experts. This may be amplified for models trained on a limited number of samples.

Table 7.4: *Feed rate classification results.*

ML algorithm	Texture descriptor	F1-score (5-fold)
SVM	LBP	0.28182 ± 0.05600
	HOG	0.40000 ± 0.12649
	GLCM	0.15000 ± 0.03780
KNN	LBP	0.32500 ± 0.05244
	HOG	0.45625 ± 0.12082
	GLCM	0.39167 ± 0.09432
RF	LBP	0.37143 ± 0.09063
	HOG	0.35000 ± 0.06124
	GLCM	0.28889 ± 0.06009
CNN		$\mathbf{0.58205 \pm 0.04601}$

7.5.3 Cutting speed classification results

Table 7.5 reported the testing performances of each model when classifying the samples according to the cutting speed. It is reminded that the available samples in the dataset were collected for a total of 3 different cutting speeds:

- 80 m/min;
- 150 m/min;
- 220 m/min.

Similarly to the previous classification tasks, the best performing approach for the cutting speed classification was the proposed CNN, registering a mean f1-score of 0.85294 over the 5-folds. The second best performing approach was represented by the combination KNN+LBP with a mean f1-score of 0.81429. Once again, the proposed CNN showed very stable classification performance, highlighted by a testing f1-score standard deviation of ± 0.04400 which is the tightest across all other models-feature combinations. For reference, the KNN+LBP combination shows a f1-score spread that is almost 50% wider (± 0.06268) than the one of the proposed CNN. The overall testing performances of all models were acceptable, with most model-feature combinations registering f1-scores above the 70% threshold, signaling that it is feasible to recognise different values of the cutting speed parameter.

Table 7.5: *Cutting speed classification results.*

ML algorithm	Texture descriptor	F1-score (5-fold)
SVM	LBP	0.66667 ± 0.05590
	HOG	0.74412 ± 0.07682
	GLCM	0.63636 ± 0.08090
KNN	LBP	0.81429 ± 0.06268
	HOG	0.78750 ± 0.11087
	GLCM	0.59167 ± 0.11583
RF	LBP	0.74167 ± 0.07360
	HOG	0.73750 ± 0.09910
	GLCM	0.64375 ± 0.08210
CNN		$\mathbf{0.85294 \pm 0.04400}$

7.5.4 Tool diameter classification results

Table 7.6 reported the testing performances of each model when classifying the samples according to the tool diameter. It is reminded that the available samples in the dataset were collected for a total of 2 different tool diameters:

- 27.9 mm;
- 32.0 mm.

For this classification task it is possible to see that all the approaches performed very well, registering testing f1-scores above the 90% mark. With that being said, the benchmark was set by the SVM+HOG combination, showing a mean test f1-score of 0.98636 which slightly outperformed the proposed CNN model (0.97464). At the same time, the proposed CNN showed excellent performance reliability
across the 5-folds: it registered a test f1-score deviation of just ± 0.00785, which was a third of the deviation registered by SVM+HOG combination. This confirmed both the effectiveness of the features generation and learning processes and the generalisation capabilities of the proposed CNN model, even when classifying a technological parameter such as the tool diameter. It has to be reminded that this classification task was somewhat simpler compared to the other 4 tasks, since it required to distinguish

between just 2 classes. Consequently, since the total number of samples was fixed and equal to 100, more samples were available to describe each class, which is something desirable to effectively train machine learning models.

Table 7.6: *Milling tool diameter classification results.*

ML algorithm	Texture descriptor	F1-score (5-fold)
SVM	LBP	0.89333 ± 0.07528
	HOG	$\mathbf{0.98636 \pm 0.02335}$
	GLCM	0.80000 ± 0.04472
KNN	LBP	0.91000 ± 0.02236
	HOG	0.98462 ± 0.02402
	GLCM	0.82188 ± 0.05468
RF	LBP	0.93333 ± 0.04330
	HOG	0.96429 ± 0.04756
	GLCM	0.79286 ± 0.08381
CNN		0.97464 ± 0.00785

7.5.5 Nose radius classification results

Table 7.7 reported the testing performances of each model when classifying the samples according to the insert nose radius. It is reminded that the available samples in the dataset were collected for a total of 5 different nose radii:

- 0.4 mm;
- 0.8 mm;
- 1.0 mm;
- 1.2 mm;
- 1.5 mm.

The proposed CNN model was the best performing approach when classifying the samples according to the nose radius of the milling tool's inserts. It managed to outperform the SVM+HOG combination, registering a mean test f1-score of 0.79475 compared to 0.76250. As previously shown across the other classification tasks, the proposed CNN was a very reliable approach, with a test f1-score deviation of ± 0.04724, which is almost half the one scored by the SVM+HOG combination (± 0.08561). Overall, the results were mixed: HOG features were those leading the machine learning models to the best performances, while LBP and GLCM were unable to generate features that correctly identify the different nose radii from the surface images. It should be noted that, similarly to the feed rate, recognising different values of nose radii through vision only is a complex task even for experts.

7.5.6 Average classification results

From a global perspective, it was possible to notice that the proposed CNN was performing very well across the spectrum of classification tasks: it was the best performing

Table 7.7: *Milling tool insert's nose radius results.*

ML algorithm	Texture descriptor	F1-score (5-fold)
SVM	LBP	0.50000 ± 0.08165
	HOG	0.76250 ± 0.08561
	GLCM	0.59000 ± 0.09947
KNN	LBP	0.67727 ± 0.10574
	HOG	0.71429 ± 0.08522
	GLCM	0.48571 ± 0.08187
RF	LBP	0.57000 ± 0.13038
	HOG	0.76000 ± 0.08216
	GLCM	0.47143 ± 0.07559
CNN		$\mathbf{0.79475 \pm 0.04724}$

model in the case of machining conditions, feed rate, cutting speed, nose radius classification, and it is marginally outperformed in the case of tool diameter (-1.2% w.r.t. SVM+HOG combination). This can be seen from Table 7.8, where the average performances of the different model-feature combinations and proposed model across the 5 classification tasks were reported:

- the proposed CNN was, on average, the best performing model across all tasks, with an average f1-score of 0.81455. It was the most consistent too, highlighted by a f1-score spread of ± 0.12996;
- the second best approach was, on average, the KNN paired with HOG features. This approach showed an average f1-score of 0.75930 (-5.5% w.r.t. the proposed CNN) and a standard deviation of ± 0.17571 (+4.6% w.r.t. the proposed CNN);
- the worst approach were, on average, the SVM paired with GLCM features or LBP features (as anticipated when discussing the machining conditions results);
- from the texture descriptor point of view the most effective one appears to be the HOG (it is reminded that this feature was paired with PCA), while the least effective one was the GLCM.

7.6 Conclusions

In this chapter, a deep learning approach for machined surface classification tasks is presented. A shallow end-to-end Convolutional Neural Network (CNN) classifier was built and trained upon 100 raw surface very high resolution images, split in 224x224 pixels batches. The CNN learned to classify images for 5 classification tasks: machining conditions, feed rates, cutting speeds, tool diameters and nose radii. The proposed approach was compared to state-of-the-art machine learning techniques (Support Vector Machines, k-Nearest Neighbours and Random Forests), fed with traditional surface feature descriptors (Local Binary Patterns, Histogram of Oriented Gradients and Gray-Level Co-Occurrence Matrix). The developed approach resulted to outperform state-of-the-art machine learning techniques in all the classification tasks, except for

Table 7.8: *Average classification performances across parameters of interest (machining condition, feed rate, cutting speed, tool diameter, nose radius).*

ML algorithm	Texture descriptor	Global F1-score
SVM	LBP	0.48836 ± 0.27908
	HOG	0.74735 ± 0.19355
	GLCM	0.45527 ± 0.27900
KNN	LBP	0.70674 ± 0.20472
	HOG	0.75930 ± 0.17571
	GLCM	0.63152 ± 0.18542
RF	LBP	0.69601 ± 0.20383
	HOG	0.73327 ± 0.20770
	GLCM	0.60355 ± 0.20057
CNN		$\mathbf{0.81455 \pm 0.12996}$

the milling tool diameter showing a mean f1-score of 97.5% (just 1.2% less than the best state-of-the-art algorithm). Furthermore, the CNN results were 5.5% better then the best machine learning approach when considering the average classification performances. Thus, the proposed CNN structure demonstrated to be robust, reliable, flexible and accurate enough for a possible industrial deployment. Nevertheless, there are still margins of improvement especially regarding the feed rate classification, where the CNN still struggles to reach optimal performances. This will be the subject of authors' future works, together with the deployment of the conceived approach in an adaptive process control framework (i.e. the CNN will suggest milling parameter changes and tool failure detection).

CHAPTER 8

Conclusions

In this thesis, the main objective was to develop prognostics and health management solutions for the manufacturing scenario, with particular attention to the machine tool system. The main branches were three and focused: on the cutting tool, from the analysis of machine signals; on machine tool auxiliaries, with particular attention to the hydraulic unit; on workpiece quality and cutting tool diagnosis with the analysis of machined surfaces. The cutting tool was chosen due its temporal and economical impact on the manufacturing systems: researchers affirmed that 40% savings could be recovered through tool condition monitoring [185]; about 20% of machine tools downtimes was attributed to tool failures, resulting in reduced productivity and economic losses [121, 227]. Tool condition monitoring and prognosis have an economical and environmental impact on manufacturing companies. In fact, cutting tool life can be optimised through TCM solutions, avoiding dramatic failure and reducing early tool changes. This allows to couple them with a cutting tool resharpening policy, leading to economical savings (for cutting tool users), while bringing increased profits (tool manufacturers), critical material reduction (e.g. Cobalt and Tungsten), CO_2 and energetic footprint decrease. The hydraulic unit was chosen in representation of machine tool auxiliaries. This was based on several research findings, which highlighted the crucial role of hydraulic units in machine tool reliability [29, 161, 217].

Three main challenges were identified, preventing a diffused and effective implementation of PHM strategies in the manufacturing sector. Typical literature solutions were based on the creation of huge supervised learning datasets, requiring several run-to-failures to be performed. These experimental tests resulted in strong economical and temporal impacts for companies. The second main challenge regarded the variability of operating conditions, especially for machine and cutting tools and typical of low-medium volume production companies. That is the case where goods are manufactured

in small batches or even as one-of-a-kind products (e.g. aeronautical sector). Furthermore, modifications in signals related to varying cutting conditions or parameters tend to be stronger than degradation effect. Last challenge consisted in the fact that literature approaches are, in general, static. This translated in models built on training data and not capable to adapt to the data coming from the field.

Derived from these challenges, the objective of the thesis were set to **minimising the need for experimental tests**, and **conceiving an operational regime robust PHM architecture** in a **hybrid and adaptive** fashion.

Regarding to the cutting tool thesis section, in chapter 2 the foundations for an operational regime robust PHM solutions were laid. In fact, a novel analytical mechanistic model for double-phased high-feed mills was formulated and proposed. A mechanistic model allows to predict useful cutting quantities (e.g. cutting forces, torque and power) based on specific cutting force coefficients, which represent specific energies or cutting pressures. In chapter 2, a specific formulation for double-phased cutters was developed, in order to be capable of predicting quantities of interest in high feed milling operations. The new formulation extended the reference literature model to include complex cutter geometry and variable engagement conditions along the axial coordinate of the tool. The predominance of ploughing action under the flank face of the cutting insert, led to lower specific force coefficients. The new formulation allowed to estimate specific force coefficients in a more reliable manner, underlying the better representation of the cutting process with respect to the application of the reference one. The literature model was outperformed also in the prediction of spindle torque and power. This was true both in terms of mean value predictions, and in terms of their prediction bounds. This effect was strictly associated to the narrower confidence intervals related to the specific force coefficients estimation.

Chapter 3, represented the first link between the developed mechanistic model to a tool condition monitoring system. In fact, in this context, specific force coefficients were adopted as tool wear correlated features, but being normalised from the operational regime. Indeed, this chapter strictly dealt with the first two objectives of the thesis. A fully unsupervised tool wear monitoring strategy was conceived based on the instantaneous forces-based specific force coefficients (SFC) fitting procedure. In fact, the identification method presented in the second chapter of the thesis would have required continuously changing feed rate run-to-failures to be applied in a tool condition monitoring or prognosis framework. Thus, the PHM solution would have resulted in no advantages on the experimental effort and limited application scenarios. On the contrary, the identification procedure of chapter 3, relying upon the instantaneous cutting forces, allowed to estimate specific force coefficients without the need of a training dataset. A principal component regression approach was needed in order to reduce the variability in the estimated specific force coefficients brought by the multicollinearity phenomenon. Multicollinearity would prohibit to separate univocally the effects of the regressors. Thus, small changes in the experimental data would have caused the see-saw effect on the specific force coefficients, generating a correlation among their oscillations. Principal component regression implementation proved to be efficient in limiting this phenomenon and strengthened their correlation to tool wear. Self-starting tabular cumulative control charts provided an effective fully unsupervised tool to detect out of control cutting samples. The detections provided by the developed tool condition

monitoring system were connected to the development of tool chippings and notches. Another important result was related to the robustness of the conceived approach with respect to multiple cutting speeds, lubrication conditions and machine tools, responding to the first two main objectives of the thesis.

In chapter 4, another tool condition monitoring strategy was proposed. In this case, the objective was related to the indirect assessment of flank wear starting from the specific force coefficients of mechanistic force models. Instantaneous identification procedures resulted to be subjected to the multicollinearity phenomenon, as previously reported. The multicollinearity phenomenon is strengthened by not modelled physical phenomena such as tool run-out or machine tool variable dynamic compliance, leading to different tool tip vibrations when cutting along different directions. An artificial intelligence-based unsupervised approach was implemented for online clustering of specific cutting force coefficients. Growing self-organising maps allowed to group data as soon as they became available. Different cells of the map contained progressive evolutions of specific force coefficients, separating sudden changes in their value. Self-starting control charts in each region constituted a voting system to detect cutting tool flank wear critical values. This chapter was complementary with chapter 3; together, they represented brittle and gradual tool wear monitoring strategies, respectively. Even in this chapter, multiple machine tools, cutting speeds and lubrication conditions were successfully analysed.

The last phase of prognostics and health management of cutting tools was presented in chapter 5. A prognosis approach was developed and applied to perform predictions of cutting tool flank wear degradation curves in turning applications. It was based on direct measurements of flank wear, obtainable through camera images and their post-processing. The conceived approach combined particle filter, coming from the statistical world, and multi-layer perceptron, from artificial intelligence. The hybridisation of the algorithm permitted to leverage on the advantages of the two fields. On one side, the solution provided adaptability to online flank wear measurements and unseen cutting tool degradation curves. On the other side, remaining useful life was estimated together with its confidence intervals and its probability density function. The approach didn't need for tool wear dynamical models, but leveraged on only one run-to-failure. Nevertheless, it performed well on validation sets presenting unseen tool degradation rates, caused by variable cutting parameters. In this chapter, the construction of a prognostics and health management framework for cutting tools was brought to a conclusion, contributing to all the main challenges and objectives of the thesis work. The developed prognostics and health management framework can be applied in most of the metal cutting operations such as milling, turning and drilling.

In chapter 6, the subject of the analysis switched to machine tool auxiliaries. A machine tool hydraulic unit digital twin was developed in order to respond to the necessity of training data. No experimental run-to-failures were needed since a set of simulations was performed for stationary and non-stationary cycles. The digital twin included also the possibility to simulate faulty components. Three different elements of the high pressure circuit of the hydraulic unit were modelled in nominal conditions and with multiple levels of fault severity. A full prognostics and health management architecture was developed for this case. A supervised learning framework was set up

for diagnosis starting from digital twin simulations. This phase allowed to detect, isolate and quantify the failures, while the prognosis phase was able to account for the interaction between different faults. In fact, it was based on the output of diagnosis, which was trained on all the fault combinations. Remaining useful life and its probability density function were estimated through a Monte Carlo approach. The solution was thus conceived to achieve all the main thesis objectives, with special regard to the first two. The robustness of the solution was verified by training the data-driven algorithms on stationary cycles and testing them on non-stationary ones.

Chapter 7 changed the perspective of the thesis, looking at prognostics and health management from the workpiece side. Very high resolution images of machined surfaces where collected through a digital microscope. The images were taken for machining processes performed with different technological and cutting parameters. Different depths of cut, cutting speeds, feed rates, tool nose radii and diameters were investigated. Furthermore, some samples were performed considering undesired machining conditions such as excessive tool run-out and tool chipping. A convolutional neural network, from the deep learning pool of algorithms, was conceived and applied to the analysis of image patches. The proposed algorithm structure demonstrated to be robust, reliable, flexible and accurate enough in the recognition of all the above cases, allowing for a possible industrial implementation. This last chapter contributed to the robustness with respect to variable working regimes challenge, providing a tool both for cutting tool and process diagnosis.

8.1 Future works

The conceived prognostics and health management framework for cutting tools was applied to milling operations, since they represent the worst case scenario in terms of cutting condition variability from a practical perspective. Despite prognosis was performed on turning run-to-failures, direct tool prognosis algorithm would be unchanged for milling. Furthermore, the developed tools provide useful methods also for other metal cutting operations. Possible future works are thus associated to the adaptation and validation of the single modules to other cutting processes.

The vision for the proposed prognosis method regards the use of specific force coefficients in place of the flank wear measurements, or in combination with them. This would allow for a continuous indirect prognosis of the cutting tool life. Furthermore, other cutting process quantities could be included. Specific force coefficients were estimated starting from cutting forces measurements. They are typically measured through dynamometers which may not be affordable in industrial applications. The solution should be integrated in indirect cutting forces estimation systems, otherwise other process measurements should be used for the specific force coefficients estimation, e.g. cutting tool vibrations or cutting torque measurements.

The developed adaptive approach is currently estimating remaining useful life of tools based on direct wear measurements. Nevertheless, the integration of indirect tool wear indicators like specific force coefficients could be included in the formulation. The hybrid architecture didn't exploited any supplementary information from the cutting process context, like machining parameters. Furthermore, the developed approach

could be improved in such a way that new run-to-failures coming from the field possibly increase its adaptability and fitting capability. This may be done by including new data in the starting multi-layer perceptron particles. When direct measurements are available from the field this can be done automatically.

Research should also focus on the correlation between the SFC evolution and the quality of cut left on the material surface. This links also to the workpiece side monitoring and diagnosis. The performances of the conceived prognostics and health management approach (and literature approaches) could be enhanced by the fusion of manufactured product features and finishing. The tool condition predictions as well as the process monitoring and diagnosis could be implemented in an adaptive process control scenario, defining production objectives to be met, subjected to machine tool component lives, cutting tool life and surface quality constraints.

Bibliography

[1] P. Aivaliotis, K. Georgoulias, and G. Chryssolouris. The use of Digital Twin for predictive maintenance in manufacturing. *International Journal of Computer Integrated Manufacturing*, 32(11):1067–1080, 2019. Publisher: Taylor & Francis.

[2] P. Albertelli, M. Goletti, M. Torta, M. Salehi, and M. Monno. Model-based broadband estimation of cutting forces and tool vibration in milling through in-process indirect multiple-sensors measurements. *International Journal of Advanced Manufacturing Technology*, 82(5):779–796, February 2016. Publisher: Springer-Verlag London Ltd.

[3] Paolo Albertelli, Valerio Mussi, and Michele Monno. Development of generalized tool life model for constant and variable speed turning. *The International Journal of Advanced Manufacturing Technology*, 118(5):1885–1901, January 2022. Publisher: Springer Science and Business Media Deutschland GmbH.

[4] Paolo Albertelli, Valerio Mussi, Matteo Strano, and Michele Monno. Experimental investigation of the effects of cryogenic cooling on tool life in Ti6Al4V milling. *The International Journal of Advanced Manufacturing Technology*, 117(7):2149–2161, December 2021. Publisher: Springer Science and Business Media Deutschland GmbH.

[5] David M. Allen. The Relationship Between Variable Selection and Data Agumentation and a Method for Prediction. *Technometrics*, 16(1):125–127, February 1974.

[6] Y. Altintas and E. Budak. Analytical Prediction of Stability Lobes in Milling. *CIRP Annals*, 44(1):357–362, 1995.

[7] Y. Altintas and P. Lee. A General Mechanics and Dynamics Model for Helical End Mills. *CIRP Annals*, 45(1):59–64, 1996.

[8] Yusuf Altintas. *Manufacturing Automation: Metal Cutting Mechanics, Machine Tool Vibrations, and CNC Design*. Cambridge University Press, Cambridge, 2012.

[9] Yusuf Altintas, Gabor Stepan, Erhan Budak, Tony Schmitz, and Zekai Murat Kilic. Chatter Stability of Machining Operations. *Journal of Manufacturing Science and Engineering*, 142(11):110801, November 2020.

[10] Yinhui Ao and George Qiao. Prognostics for drilling process with wavelet packet decomposition. *The International Journal of Advanced Manufacturing Technology*, 50(1-4):47–52, September 2010.

[11] P. J. Arrazola, T. Ozel, D. Umbrello, M. Davies, and I. S. Jawahir. Recent advances in modelling of metal machining processes. *CIRP Annals*, 62(2):695–718, January 2013.

[12] Abdullah Aslan. Optimization and analysis of process parameters for flank wear, cutting forces and vibration in turning of AISI 5140: A comprehensive study. *Measurement*, 163:107959, October 2020.

[13] Mohsen Azimi, Armin Dadras Eslamlou, and Gokhan Pekcan. Data-Driven Structural Health Monitoring and Damage Detection through Deep Learning: State-of-the-Art Review. *Sensors*, 20(10), May 2020. Publisher: MDPI AG.

[14] Marco Baur, Paolo Albertelli, and Michele Monno. A review of prognostics and health management of machine tools. *The International Journal of Advanced Manufacturing Technology*, 107(5):2843–2863, March 2020. Publisher: The International Journal of Advanced Manufacturing Technology.

[15] Marcelo Beckmann, Nelson F. F. Ebecken, and Beatriz S. L. Pires de Lima. A KNN Undersampling Approach for Data Balancing. *Journal of Intelligent Learning Systems and Applications*, 07(04):104–116, 2015.

[16] Richard Bellman. Dynamic Programming. *Science*, 153(3731):34–37, July 1966. Publisher: American Association for the Advancement of Science.

[17] T. Benkedjouh, K. Medjaher, N. Zerhouni, and S. Rechak. Health assessment and life prediction of cutting tools based on support vector regression. *Journal of Intelligent Manufacturing*, 26(2):213–223, April 2015.

[18] L. Bernini, P. Albertelli, and M. Monno. Mechanistic force model for double-phased high-feed mills. *International Journal of Mechanical Sciences*, page 107801, October 2022.

[19] Luca Bernini, Paolo Albertelli, and Michele Monno. Mill condition monitoring based on instantaneous identification of specific force coefficients under variable cutting conditions. *Mechanical Systems and Signal Processing*, 185:109820, February 2023.

[20] Luca Bernini, David Waltz, Paolo Albertelli, and Michele Monno. A novel prognostics solution for machine tool sub-units: The hydraulic case:. *Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture*, December 2021. Publisher: SAGE PublicationsSage UK: London, England.

[21] Bernhard E Boser, Isabelle M Guyon, and Vladimir N Vapnik. A Training Algorithm for Optimal Margin Classifiers. In *Proceedings of the Fifth Annual Workshop on Computational Learning Theory*, pages 144–152. Association for Computing Machinery, 1992.

[22] Alexei Botchkarev. Evaluating Performance of Regression Machine Learning Models Using Multiple Error Metrics in Azure Machine Learning Studio. *SSRN Electronic Journal*, 2018.

[23] G. E. P. Box. A GENERAL DISTRIBUTION THEORY FOR A CLASS OF LIKELIHOOD CRITERIA. *Biometrika*, 36(3-4):317–346, 1949.

[24] Leo Breiman. Random Forests. *Machine Learning*, 45(1):5–32, October 2001.

[25] E. Budak. Analytical models for high performance milling. Part I: Cutting forces, structural deformations and tolerance integrity. *International Journal of Machine Tools and Manufacture*, 46(12-13):1478–1488, October 2006.

[26] E. Budak, Y. Altintas, and E. J. A. Armarego. Prediction of Milling Force Coefficients From Orthogonal Cutting Data. *Journal of Manufacturing Science and Engineering*, 118(2):216–224, May 1996.

[27] Francesco Cadini, Claudio Sbarufatti, Matteo Corbetta, Francesco Cancelliere, and Marco Giglio. Particle filtering-based adaptive training of neural networks for real-time structural damage diagnosis and prognosis. *Structural Control and Health Monitoring*, 26(12):1–19, 2019.

[28] G. Campatelli and A. Scippa. Prediction of milling cutting force coefficients for Aluminum 6082-T4. *Procedia CIRP*, 1(1):563–568, 2012.

[29] Fei Chen, Xiaojuan Chen, Qunya Xie, and Binbin Xu. Reliability Analysis of Numerical Control Lathe Based On The Field Data. In *Proceedings of the 2015 6th International Conference on Manufacturing Science and Engineering*, pages 1018–1024. Atlantis Press, 2015.

[30] Steven Chen, Edward Chou, and Richard Yang. The Price is Right: Predicting Prices with Product Images. *arXiv*, 2018.

[31] Xuelin Chen, Jinyuan Tang, Han Ding, and Anmin Liu. A new geometric model of serrated chip formation in high-speed machining. *Journal of Manufacturing Processes*, 62:632–645, February 2021.

[32] Xuezhen Chen, Zhiyong Song, Hai Li, Qinghong Gong, Ying Li, and Feng Wang. Research on fault early warning and the diagnosis of machine tools based on energy fault tree analysis. *Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture*, 233(11):2147–2159, 2019.

[33] Yuhan Chen, Jun Wang, and Qinglong An. Mechanisms and predictive force models for machining with rake face textured cutting tools under orthogonal cutting conditions. *International Journal of Mechanical Sciences*, 195:106246, April 2021.

[34] Minghui Cheng, Li Jiao, Xuechun Shi, Xibin Wang, Pei Yan, and Yongping Li. An intelligent prediction model of the tool wear based on machine learning in turning high strength steel. *Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture*, 234(13):1580–1597, 2020.

[35] Qiang Cheng, Baobao Qi, Zhifeng Liu, Congbin Yang, and Jigui Zheng. Positioning accuracy degradation and lifetime prediction of the ball screw considering time-varying working conditions and feed modes. *Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture*, 235(6-7):943–957, 2021.

[36] Nancy Chinchor. MUC-4 Evaluation Metrics. In *Proceedings of the 4th Conference on Message Understanding*, pages 22–29. Association for Computational Linguistics, 1992.

[37] Michael E. Cholette, Jianbo Liu, Dragan Djurdjanovic, and Kenneth A. Marko. Monitoring of complex systems of interacting dynamic systems. *Applied Intelligence*, 37(1):60–79, 2012.

[38] Clifford C Clogg, Eva Petkova, and Adamantios Haritou. Statistical Methods for Comparing Regression Coefficients between Models. *American Journal of Sociology*, 100(5):1261–1293, 1995.

[39] Bianca Colosimo, Giovanni Moroni, and Marco Grasso. Real-time tool condition monitoring in milling by means of control charts for auto-correlated data. *Journal of Machine Engineering*, 10:5–17, 2010.

[40] N. A. C. Cressie and H. J. Whitford. How to Use the Two Samplet-Test. *Biometrical Journal*, 28(2):131–148, 1986.

[41] G. Cybenko. Approximation by superpositions of a sigmoidal function. *Mathematics of Control, Signals, and Systems*, 2(4):303–314, December 1989.

[42] L.R.R. da Silva, V.T.S. Del Claro, C.L.F. Andrade, W.L. Guesser, M.J. Jackson, and A.R. Machado. Tool wear monitoring in drilling of high-strength compacted graphite cast irons. *Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture*, 235(1-2):207–218, 2021.

[43] Jian-Wei Dang, Wei-Hong Zhang, Yun Yang, and Min Wan. Cutting force modeling for flat end milling including bottom edge cutting effect. *International Journal of Machine Tools and Manufacture*, 50(11):986–997, November 2010.

[44] Xue-Bin Dang, Min Wan, Wei-Hong Zhang, and Yun Yang. Chatter analysis and mitigation of milling of the pocket-shaped thin-walled workpieces with viscous fluid. *International Journal of Mechanical Sciences*, 194:106214, March 2021.

[45] Cintia de Oliveira Bizarria, Takashi Yoneyama, Cintia De Oliveira Bizarria, and Takashi Yoneyama. Prognostics and health monitoring for an electro-hydraulic flight control actuator. *IEEE Aerospace Conference Proceedings*, pages 1–9, 2009. Publisher: IEEE ISBN: 9781424426225.

[46] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. ImageNet: A large-scale hierarchical image database. In *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pages 248–255. Institute of Electrical and Electronics Engineers (IEEE), 2009.

[47] Jianming Dou, Shengjie Jiao, Chuangwen Xu, Foshu Luo, Linhu Tang, and Xinxin Xu. Unsupervised online prediction of tool wear values using force model coefficients in milling. *The International Journal of Advanced Manufacturing Technology*, 109(3-4):1153–1166, July 2020.

[48] Cyril Drouillet, Jaydeep Karandikar, Chandra Nath, Anne-Claire Journeaux, Mohamed El Mansori, and Thomas Kurfess. Tool life predictions in milling using spindle power with the neural network technique. *Journal of Manufacturing Processes*, 22:161–168, April 2016.

[49] J. Duplak, M. Hatala, D. Duplakova, and J. Steranka. Comprehensive analysis and study of the machinability of a high strength aluminum alloy (EN AW-AlZn5.5MgCu) in the high-feed milling. *Advances in Production Engineering & Management*, 13(4):455–465, December 2018.

[50] Joseph G Eisenhauer. Regression through the Origin. *Teaching Statistics*, 25(3), 2003.

[51] Alessandro Elefanti, Paolo Albertelli, Matteo Strano, and Michele Monno. Estimation of cutting and friction coefficients in dry and cryogenic milling through experiments and simulations. In *PROCEEDINGS OF THE 22ND INTERNATIONAL ESAFORM CONFERENCE ON MATERIAL FORMING: ESAFORM 2019*, page 080006, Vitoria-Gasteiz, Spain, 2019.

[52] David L. Elliott. A better activation function for artificial neural networks. Technical report, 1993.

[53] N. Fang, I. S. Jawahir, and P. L.B. Oxley. Universal slip-line model with non-unique solutions for machining with curled chip formation and a restricted contact tool. *International Journal of Mechanical Sciences*, 43(2):557–580, February 2001. Publisher: Elsevier Science Ltd.

[54] M. Farhadmanesh and K. Ahmadi. Online identification of mechanistic milling force models. *Mechanical Systems and Signal Processing*, 149:1–18, 2021. Publisher: Elsevier Ltd.

[55] Laura Fernandez-Robles, Lidia Sanchez-Gonzalez, Javier Diez-Gonzalez, Manuel Castejon-Limas, and Hilde Perez. Use of image processing to monitor tool wear in micro milling. *Neurocomputing*, 452:333–340, September 2021.

[56] Susana Ferreiro, Egoitz Konde, Santiago Fernandez, and Agustin Prado. INDUSTRY 4.0: Predictive Intelligent Maintenance for Production Equipment. In *European Conference of the Prognostics and Health Management Society*, pages 1–8, 2016. Issue: June 2016.

[57] E Fix and J L Hodges. *Discriminatory Analysis: Nonparametric Discrimination: Consistency Properties*. USAF School of Aviation Medicine, 1951.

[58] Guizhong Fu, Peize Sun, Wenbin Zhu, Jiangxin Yang, Yanlong Cao, Michael Ying Yang, and Yanpeng Cao. A deep-learning-based approach for fast and robust steel surface defects classification. *Optics and Lasers in Engineering*, 121:397–405, October 2019. Publisher: Elsevier Ltd.

[59] E. S. Gadelmawla. A vision system for surface roughness characterization using the gray level co-occurrence matrix. *NDT \& E International*, 37(7):577–588, October 2004.

[60] Maria Teresa Garcia-Ordas, Enrique Alegre-Gutierrez, Rocio Alaiz-Rodriguez, and Victor Gonzalez-Castro. Tool wear monitoring using an online, automatic and low cost system based on local texture. *Mechanical Systems and Signal Processing*, 112:98–112, November 2018. Publisher: Academic Press.

[61] A. Gilioli, C. Sbarufatti, A. Manes, and M. Giglio. Compression after impact test (CAI) on NOMEX^{TM} honeycomb sandwich panels with thin aluminum skins. *Composites Part B: Engineering*, 67(November 2013):313–325, 2014. Publisher: Elsevier Ltd.

[62] Thomas Gittler, Adam Gontarz, Lukas Weiss, and Konrad Wegener. A fundamental approach for data acquisition on machine tools as enabler for analytical Industrie 4.0 applications. *Procedia CIRP*, 79:586–591, 2019.

[63] Thomas Gittler, Fabian Stoop, David Kryscio, Lukas Weiss, and Konrad Wegener. Condition monitoring system for machine tool auxiliaries. *Procedia CIRP*, 88:358–363, 2020.

[64] Joao Paulo P. Gomes, Bruno P. Leao, Wlamir O.L. Vianna, Roberto K.H. Galvao, and Takashi Yoneyama. Failure prognostics of a hydraulic pump using Kalman Filter. *Proceedings of the Annual Conference of the Prognostics and Health Management Society 2012, PHM 2012*, 4(1):464–468, 2012. ISBN: 9781936263059.

[65] Joao Paulo Pordeus Gomes, Leonardo Ramos Rodrigues, Bruno P. Leao, Roberto Kawakami Harrop Galvao, Takashi Yoneyama, Bruno P. Leao, Roberto Kawakami Harrop Galvao, and Takashi Yoneyama. Using Degradation Messages to Predict Hydraulic System Failures in a Commercial Aircraft. *IEEE Transactions on Automation Science and Engineering*, 15(1):214–224, 2018.

[66] Oscar Gonzalo, Jokin Beristain, Haritz Jauregi, and Carmen Sanz. A method for the identification of the specific force coefficients for mechanistic milling simulation. *International Journal of Machine Tools and Manufacture*, 50(9):765–774, 2010. Publisher: Elsevier.

[67] Paul Goodwin and Richard Lawton. On the asymmetry of the symmetric MAPE. *International Journal of Forecasting*, 15(4):405–408, October 1999.

[68] Bhushan Gopaluni, Aditya Tulsyan, Benoit Chachuat, Biao Huang, Jong Min Lee, Faraz Amjad, Seshu Kumar Damarla, Jong Woo Kim, and Nathan P. Lawrence. Modern Machine Learning Tools for Monitoring and Control of Industrial Processes: A Survey. *IFAC-PapersOnLine*, 53(2):218–229, 2020.

[69] Janez Gradisek, Martin Kalveram, and Klaus Weinert. Mechanistic identification of specific force coefficients for a general end mill. *International Journal of Machine Tools and Manufacture*, 44(4):401–414, March 2004.

[70] David Griffiths and Jan Boehm. A Review on Deep Learning Techniques for 3D Sensed Data Classification. *Remote Sensing*, 11(12), June 2019. Publisher: MDPI AG.

[71] N. Grossi, L. Sallese, A. Scippa, and G. Campatelli. Chatter Stability Prediction in Milling Using Speed-varying Cutting Force Coefficients. *Procedia CIRP*, 14:170–175, 2014.

[72] N. Grossi, L. Sallese, A. Scippa, and G. Campatelli. Speed-varying cutting force coefficient identification in milling. *Precision Engineering*, 42:321–334, 2015. Publisher: Elsevier Inc. ISBN: 0552758728.

[73] Baosu Guo, Qin Zhang, Qinjing Peng, Jichao Zhuang, Fenghe Wu, and Quan Zhang. Tool health monitoring and prediction via attention-based encoder-decoder with a multi-step mechanism. *The International Journal of Advanced Manufacturing Technology*, August 2022.

[74] Minglong Guo, Zhaocheng Wei, Minjie Wang, Shiquan Li, and Shengxian Liu. An identification model of cutting force coefficients for five-axis ball-end milling. *International Journal of Advanced Manufacturing Technology*, 99(1-4):937–949, 2018. Publisher: The International Journal of Advanced Manufacturing Technology.

[75] Virginija Gyliene and Valdas Eidukynas. The Numerical Analysis of Cutting Forces in High Feed Face Milling, Assuming the Milling Tool Geometry. *Procedia CIRP*, 46:436–439, 2016.

[76] David Hajdu, Asier Astarloa, and Zoltan Dombovari. Cutting force prediction based on a curved uncut chip thickness model, November 2021. Number: arXiv:2111.00795 arXiv:2111.00795 [math].

[77] Mryka Hall-Beyer. GLCM Texture: A Tutorial v. 3.0 March 2017, 2017.

[78] Robert M. Haralick, K. Shanmugam, and I. Dinstein. Textural features for Image Classification. *IEEE Transactions on Systems, Man, and Cybernetics*, SMC-3(6):610–621, 1973.

[79] Trevor Hastie, Robert Tibshirani, and Jerome Friedman. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer, New York, 2009.

[80] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep Residual Learning for Image Recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 770–778, December 2016.

[81] Zhaomin He, Shaoping Wang, Kang Wang, and Kai Li. Prognostic analysis based on hybrid prediction method for axial piston pump. In *IEEE 10th International Conference on Industrial Informatics*, pages 688–692, Beijing, China, July 2012. IEEE.

[82] Nikolai Helwig, Steffen Klein, and Andreas Schutze. Identification and Quantification of Hydraulic System Faults Based on Multivariate Statistics Using Spectral Vibration Features. *Procedia Engineering*, 120:1225–1228, 2015. Publisher: Elsevier B.V.

[83] Nikolai Helwig, Eliseo Pignanelli, and Andreas Schutze. Condition monitoring of a complex hydraulic system using multivariate statistics. *Conference Record - IEEE Instrumentation and Measurement Technology Conference*, 2015-July:210–215, 2015. ISBN: 9781479961139.

[84] Nikolai Helwig, Eliseo Pignanelli, and Andreas Schutze. Detecting and Compensating Sensor Faults in a Hydraulic Condition Monitoring System. *SENSOR 2015*, 0:641–646, 2015.

[85] Jerry L.; Hintze. NCSS: User's Guide III. Regression and Curve Fitting, 2007. Publication Title: NCSS Statistical System.

[86] Le Hou, Dimitris Samaras, Tahsin M Kurc, Yi Gao, James E Davis, and Joel H Saltz. Patch-Based Convolutional Neural Network for Whole Slide Tissue Image Classification. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2424–2433, 2016.

[87] Andrew Howard, Mark Sandler, Grace Chu, Liang-Chieh Chen, Bo Chen, Mingxing Tan, Weijun Wang, Yukun Zhu, Ruoming Pang, Vijay Vasudevan, Quoc V. Le, and Hartwig Adam. Searching for MobileNetV3. *arXiv*, May 2019. arXiv: 1905.02244.

[88] Jinfei Hu and Peter W. Tse. A relevance vector machine-based approach with application to oil sand pump prognostics. *Sensors (Switzerland)*, 13(9):12663–12686, 2013.

[89] Fengqing Huang and Xiaoliang Jin. Surface texture generation using high-feed milling with spindle speed modulation. *Precision Engineering*, 72:13–24, November 2021.

[90] Gao Huang, Zhuang Liu, Laurens van der Maaten, and Kilian Q. Weinberger. Densely Connected Convolutional Networks. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2261–2269, July 2017.

[91] Guang Bin Huang, Qin Yu Zhu, and Chee Kheong Siew. Extreme learning machine: Theory and applications. *Neurocomputing*, 70(1-3):489–501, December 2006.

[92] Xinda Huang, Xiaoming Zhang, Haikuo Mou, Xiaojian Zhang, and Han Ding. The influence of cryogenic cooling on milling stability. *Journal of Materials Processing Technology*, 214(12):3169–3178, December 2014.

[93] International Standards. *ISO 3685: tool-life testing with single-point turning tools.* ISO, November 1993.

[94] ISO Central Secretary. Condition monitoring and diagnostics of machines - Data interpretation and diagnostics techniques - Part 1: General guidelines. Standard ISO 13379-1:2012. International Organization for Standardization, Geneva, CH. 2012.

[95] ISO Central Secretary. Condition monitoring and diagnostics of machines - Data interpretation and diagnostics techniques - Part 2: Data-driven applications. Standard ISO 13379-2:2015. International Organization for Standardization, Geneva, CH. 2015.

[96] ISO Central Secretary. Condition Monitoring and Diagnostics of Machines - Prognostics - Part 1: General Guidelines. Standard ISO 13381-1:2015. International Organization for Standardization, Geneva, CH. 2015.

[97] ISO Central Secretary. Quality management systems - Requirements. Standard ISO 9001:2015. International Organization for Standardization, Geneva, CH. 2015.

[98] ISO Central Secretary. Condition monitoring and diagnostics of machines - General guidelines. Standard ISO 17359:2018. International Organization for Standardization, Geneva, CH. 2018.

[99] Andrew K.S. Jardine, Daming Lin, and Dragan Banjevic. A review on machinery diagnostics and prognostics implementing condition-based maintenance. *Mechanical Systems and Signal Processing*, 20(7):1483–1510, 2006.

[100] Kamran Javed, Rafael Gouriveau, Xiang Li, and Noureddine Zerhouni. Tool wear monitoring and prognostics challenges: a comparison of connectionist methods toward an adaptive ensemble model. *Journal of Intelligent Manufacturing*, 29(8):1873–1890, December 2018.

[101] Kamran Javed, Rafael Gouriveau, Noureddine Zerhouni, Ryad Zemouri, and Xiang Li. Robust, reliable and applicable tool wear monitoring and prognostic: Approach based on an improved-extreme learning machine. In *2012 IEEE Conference on Prognostics and Health Management*, pages 1–9, Denver, CO, USA, June 2012. IEEE.

[102] K. Jemielniak. Commercial Tool Condition Monitoring Systems. *The International Journal of Advanced Manufacturing Technology*, 15(10):711–721, September 1999.

[103] Wei Ji, Xianli Liu, Lihui Wang, and Shilong Sun. Experimental evaluation of polycrystalline diamond (PCD) tool geometries at high feed rate in milling of titanium alloy TC11. *The International Journal of Advanced Manufacturing Technology*, 77(9-12):1549–1555, April 2015.

[104] Pingjia Jia, Youmin Rong, and Yu Huang. Condition monitoring of the feed drive system of a machine tool based on long-term operational modal analysis. *International Journal of Machine Tools and Manufacture*, 146(March):103454, 2019. Publisher: Elsevier Ltd.

[105] Amaia Jimenez, Miguel Arizmendi, and Jose Manuel Sanchez. Extraction of tool wear indicators in peck-drilling of Inconel 718. *International Journal of Advanced Manufacturing Technology*, 114(9-10):2711–2720, June 2021. Publisher: Springer Science and Business Media Deutschland GmbH.

[106] Xiubing Jing, Rongyu Lv, Yun Chen, Yanling Tian, and Huaizhong Li. Modelling and experimental analysis of the effects of run out, minimum chip thickness and elastic recovery on the cutting force in micro-end-milling. *International Journal of Mechanical Sciences*, 176:105540, June 2020.

[107] Daniel Johansson, Soren Hagglund, Volodymyr Bushlya, and Jan-Eric Stahl. Assessment of Commonly used Tool Life Models in Metal Cutting. *Procedia Manufacturing*, 11:602–609, 2017.

[108] Ian Jolliffe. Principal Component Analysis. In Miodrag Lovric, editor, *International Encyclopedia of Statistical Science*, pages 1094–1096. Springer Berlin Heidelberg, Berlin, Heidelberg, 2011.

[109] Sonja Jozic, Branimir Lela, and Drazen Bajic. A new mathematical model for flank wear prediction using functional data analysis methodology. *Advances in Materials Science and Engineering*, 2014, 2014. Publisher: Hindawi Publishing Corporation.

[110] A. A. Kassim, Zhu Mian, and M. A. Mannan. Tool condition classification using Hidden Markov model based on fractal analysis of machined surface textures. *Machine Vision and Applications*, 17(5):327–336, October 2006.

[111] Ummul Khair, Hasanul Fahmi, Sarudin Al Hakim, and Robbi Rahim. Forecasting Error Calculation with Mean Absolute Deviation and Mean Absolute Percentage Error. *Journal of Physics: Conference Series*, 930:012002, December 2017.

[112] Raghad M. Khorsheed and Omer Faruk Beyca. An integrated machine learning: Utility theory framework for real-time predictive maintenance in pumping systems. *Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture*, 235(5):887–901, 2021.

[113] Z.M. Kilic and Y. Altintas. Generalized mechanics and dynamics of metal cutting operations for unified simulations. *International Journal of Machine Tools and Manufacture*, 104:1–13, May 2016.

[114] Z.M. Kilic and Y. Altintas. Generalized modelling of cutting tool geometries for unified process simulation. *International Journal of Machine Tools and Manufacture*, 104:14–25, May 2016.

[115] Dong-Hyeon Kim, Thomas J.Y. Kim, Xinlin Wang, Mincheol Kim, Ying-Jun Quan, Jin Woo Oh, Soo-Hong Min, Hyungjung Kim, Binayak Bhandari, Insoon Yang, and Sung-Hoon Ahn. Smart Machining Process Using Machine Learning: A Review and Perspective on Machining Industry. *International Journal of Precision Engineering and Manufacturing-Green Technology*, 5(4):555–568, August 2018. Publisher: Korean Society for Precision Engineering.

[116] Kadir Kiran and Mehmet Cengiz Kayacan. Cutting force modeling and accurate measurement in milling of flexible workpieces. *Mechanical Systems and Signal Processing*, 133:106284, November 2019.

[117] Alex Krizhevsky. Learning Multiple Layers of Features from Tiny Images. Technical report, April 2009.

[118] Lee M. Kumanchik and Tony L. Schmitz. Improved analytical chip thickness model for milling. *Precision Engineering*, 31(3):317–324, 2007.

[119] Mustafa Kuntoglu, Abdullah Aslan, Danil Yurievich Pimenov, Usame Ali Usca, Emin Salur, Munish Kumar Gupta, Tadeusz Mikolajczyk, Khaled Giasin, Wojciech Kaplonek, and Shubham Sharma. A Review of Indirect Tool Condition Monitoring Systems and Decision-Making Methods in Turning: Critical Analysis and Trends. *Sensors*, 21(1):108, December 2020.

[120] Mustafa Kuntoglu, Abdullah Aslan, Haci Saglam, Danil Yurievich Pimenov, Khaled Giasin, and Tadeusz Mikolajczyk. Optimization and Analysis of Surface Roughness, Flank Wear and 5 Different Sensorial Data via Tool Condition Monitoring System in Turning of AISI 5140. *Sensors*, 20(16):4377, August 2020.

[121] Satya Kurada and Colin Bradley. A review of machine vision sensors for tool condition monitoring. *Computers in Industry*, 34(1):55–72, 1997.

[122] Benjamin Lamoureux, Jean Remi Masse, Nazih Mechbal, Jean-Remi Masse, and Nazih Mechbal. Diagnostics of an aircraft engine pumping unit using a hybrid approach based-on surrogate modeling. *2013 IEEE Conference on Prognostics and Health Management (PHM)*, pages 1–8, 2013. ISBN: 9781467357227.

[123] Yann LeCun and Corinna Cortes. MNIST handwritten digit database. 2010. Medium: http://yann.lecun.com/exdb/mnist/.

[124] Matti Lehtomaki, Anttoni Jaakkola, Juha Hyyppa, Jouko Lampinen, Harri Kaartinen, Antero Kukko, Eetu Puttonen, and Hannu Hyyppa. Object Classification and Recognition From Mobile Laser Scanning Point Clouds in a Road Environment. *IEEE Transactions on Geoscience and Remote Sensing*, 54(2):1226–1239, February 2016. Publisher: Institute of Electrical and Electronics Engineers Inc.

[125] Christophe Letot, Roger Serra, Maela Dossevi, and Pierre Dehombreux. Cutting tools reliability and residual life prediction from degradation indicators in turning process. *The International Journal of Advanced Manufacturing Technology*, 86(1):495–506, December 2015. Publisher: Springer.

[126] Urko Leturiondo, Oscar Salgado, Lorenzo Ciani, Diego Galar, and Marcantonio Catelani. Architecture for hybrid modelling and its application to diagnosis and prognosis with missing data. *Measurement: Journal of the International Measurement Confederation*, 108:152–162, 2017. Publisher: Elsevier Ltd.

[127] Bingran Li, Hui Zhang, Peiqing Ye, and Jinsong Wang. Trajectory smoothing method using reinforcement learning for computer numerical control machine tools. *Robotics and Computer-Integrated Manufacturing*, 61, February 2020. Publisher: Elsevier Ltd.

[128] Huaizhong Li and Bing Wu. Development of a hybrid cutting force model for micromilling of brass. *International Journal of Mechanical Sciences*, 115-116:586–595, September 2016.

[129] Jianhui Li, Z. Murat Kilic, and Yusuf Altintas. General Cutting Dynamics Model for Five-Axis Ball-End Milling Operations. *Journal of Manufacturing Science and Engineering*, 142(12):121003, December 2020.

[130] Kexuan Li, Kunpeng Zhu, and Tao Mei. A generic instantaneous undeformed chip thickness model for the cutting force modeling in micromilling. *International Journal of Machine Tools and Manufacture*, 105:23–31, June 2016.

[131] Pin Li, Xiaodong Jia, Jianshe Feng, Hossein Davari, Guan Qiao, Yihchyun Hwang, and Jay Lee. Prognosability study of ball screw degradation using systematic methodology. *Mechanical Systems and Signal Processing*, 109:45–57, September 2018.

[132] Si Li and Kunpeng Zhu. In-situ tool wear area evaluation in micro milling with considering the influence of cutting force. *Mechanical Systems and Signal Processing*, 161:107971, December 2021.

[133] X. Li, M. J. Er, B. S. Lim, J. H. Zhou, O. P. Gan, and L. Rutkowski. FUZZY REGRESSION MODELING FOR TOOL PERFORMANCE PREDICTION AND DEGRADATION DETECTION. *International Journal of Neural Systems*, 20(05):405–419, October 2010.

[134] Xiang Li, Junhong Zhou, Hao Zeng, Yoke Wong, and Geok Hong. An Intelligent Predictive Engine for Milling Machine Prognostic Monitoring. In *2006 IEEE International Conference on Industrial Informatics*, pages 1075–1080, Singapore, August 2006. IEEE.

[135] Xuebing Li, Xianli Liu, Caixu Yue, Steven Y. Liang, and Lihui Wang. Systematic review on tool breakage monitoring techniques in machining operations. *International Journal of Machine Tools and Manufacture*, 176:103882, May 2022.

[136] Linxia Liao and Radu Pavel. Machinery time to failure prediction - Case study and lesson learned for a spindle bearing application. In *2013 IEEE Conference on Prognostics and Health Management (PHM)*, pages 1–11, Gaithersburg, MD, USA, June 2013. IEEE.

[137] Linxia Liao, Haixia Wang, and Jay Lee. Bearing Health Assessment and Fault Diagnosis Using the Method of Self-Organizing Map. In *61st Meeting of the Society for Machinery Failure Prevention Technology*, 2007.

[138] Zhirong Liao, Dragos Axinte, and Dong Gao. On modelling of cutting force and temperature in bone milling. *Journal of Materials Processing Technology*, 266:627–638, April 2019.

[139] Wan-Ju Lin, Jian-Wen Chen, Jian-Ping Jhuang, Meng-Shiun Tsai, Che-Lun Hung, Kuan-Ming Li, and Hong-Tsu Young. Integrating object detection and image segmentation for detecting the tool wear area on stitched image. *Scientific Reports*, 11(1):19938, December 2021.

[140] Delin Liu, Zhanqiang Liu, Jinfu Zhao, Qinghua Song, Xiaoping Ren, and Haifeng Ma. Tool wear monitoring through online measured cutting force and cutting temperature during face milling Inconel 718. *The International Journal of Advanced Manufacturing Technology*, August 2022.

[141] Hui Liu, Daniel Schraknepper, and Thomas Bergs. Investigation of residual stresses and workpiece distortion during high-feed milling of slender stainless steel components. *Procedia CIRP*, 108:495–500, 2022.

[142] Jianbo Liu, Dragan Djurdjanovic, Kenneth Marko, and Jun Ni. Growing structure multiple model systems for anomaly detection and fault diagnosis. *Journal of Dynamic Systems, Measurement and Control, Transactions of the ASME*, 131(5):1–13, 2009.

[143] Hiroshi Matsuda, Hiroyuki Sasahara, and Masaomi Tsutsumi. Generation of a regularly aligned surface pattern and control of cutter marks array by patch division milling. *International Journal of Machine Tools and Manufacture*, 48(1):84–94, January 2008.

[144] Takashi Matsumura and Shoichi Tamura. Cutting force model in milling with cutter runout. *Procedia CIRP*, 58:566–571, 2017.

[145] Thomas McLeay, Michael S. Turner, and Keith Worden. A novel approach to machining process fault detection using unsupervised learning. *Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture*, 235(10):1533–1542, 2021.

[146] Kamal Medjaher and Noureddine Zerhouni. Framework for a Hybrid Prognostics. *Chemical Engineering Transactions*, 33:91–96, 2013.

[147] Parikshit Mehta, Prahalada Rao, Zhenhua Wu, Vukica Jovanovic, Olga Wodo, and Mathew Kuttolamadom. Smart Manufacturing: State-of-the-Art Review in Context of Conventional and Modern Manufacturing Process Modeling, Monitoring and Control. In *International Manufacturing Science and Engineering Conference*, 2018.

[148] Xiangfei Meng, Jingjie Zhang, Guangchun Xiao, Zhaoqiang Chen, Mingdong Yi, and Chonghai Xu. Tool wear prediction in milling based on a GSA-BP model with a multisensor fusion method. *International Journal of Advanced Manufacturing Technology*, 114(11-12):3793–3802, June 2021. Publisher: Springer Science and Business Media Deutschland GmbH.

[149] Zibo Meng, Xiaochuan Fan, Xin Chen, Min Chen, and Yan Tong. Detecting small signs from large images. *Proceedings - 2017 IEEE International Conference on Information Reuse and Integration, IRI 2017*, 2017-January:217–224, November 2017. arXiv: 1706.08574 Publisher: Institute of Electrical and Electronics Engineers Inc. ISBN: 9781538615621.

[150] M. Eugene Merchant. Basic Mechanics of the Metal-Cutting Process. *Journal of Applied Mechanics*, 11(3):A168–A175, 1944.

[151] Ovidiu Moldovan, Simona Dzitac, Ioan Moga, Tiberiu Vesselenyi, and Ioan Dzitac. Tool-Wear Analysis Using Image Processing of the Tool Flank. *Symmetry*, 9(12):296, November 2017.

[152] Douglas C Montgomery. *Introduction to Statistical Quality Control*. John Wiley & Sons, Inc., 2008.

[153] Douglas C Montgomery. *Design and Analysis of Experiments*, volume 2. John Wiley & Sons, Inc., 2012. Publication Title: Design ISSN: 00401706.

[154] James Moore, Jon Stammers, and Javier Dominguez-Caballero. The application of machine learning to sensor signals for machine tool and process health assessment. *Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture*, 235(10):1543–1557, 2021.

[155] D. G. Nel and C. A. Van Der Merwe. A solution to the multivariate behrens-fisher problem. *Communications in Statistics - Theory and Methods*, 15(12):3719–3735, 1986.

[156] Mehdi Nouri, Barry K. Fussell, Beth L. Ziniti, and Ernst Linder. Real-time tool wear monitoring in milling using a cutting condition independent method. *International Journal of Machine Tools and Manufacture*, 89:1–13, 2015. Publisher: Elsevier.

[157] Kashfull Orra and Sounak K. Choudhury. Mechanistic modelling for predicting cutting forces in machining considering effect of tool nose radius on chip formation and tool wear land. *International Journal of Mechanical Sciences*, 142-143:255–268, July 2018.

[158] Harry Otalora-Ortega, Patxi Aristimuno Osoro, and Pedro Arrazola Arriola. Uncut chip geometry determination for cutting forces prediction in orthogonal turn-milling operations considering the tool profile and eccentricity. *International Journal of Mechanical Sciences*, 198:106351, May 2021.

[159] Tianhang Pan, Jun Zhang, Xing Zhang, Wanhua Zhao, Huijie Zhang, and Bingheng Lu. Milling force coefficients-based tool wear monitoring for variable parameter milling. *The International Journal of Advanced Manufacturing Technology*, 120(7-8):4565–4580, June 2022.

[160] Je-Kang Park, Bae-Keun Kwon, Jun-Hyub Park, and Dong-Joong Kang. Machine learning-based imaging system for surface defect inspection. *International Journal of Precision Engineering and Manufacturing-Green Technology*, 3(3):303–310, July 2016. Publisher: Korean Society for Precision Engineering.

[161] Rajkumar B. Patil and Basavraj S. Kothavale. Failure Modes and Effects Analysis (FMEA) of Computerized Numerical Control (CNC) Turning Center. *International Review of Mechanical Engineering (IREME)*, 12(1):78–87, 2018.

[162] Ying Peng, Ming Dong, and Ming Jian Zuo. Current status of machine prognostics in condition-based maintenance: A review. *International Journal of Advanced Manufacturing Technology*, 50(1-4):297–313, 2010.

[163] Philipp Probst and Anne-Laure Boulesteix. To Tune or Not to Tune the Number of Trees in Random Forest. *Journal of Machine Learning Research*, 18(181):1–18, 2018.

[164] Bazi Rabah, Tarak Benkedjouh, and Rechak Said. Tool Wear Condition Monitoring Based on Blind Source Separation and Wavelet Transform. In Mohammed Chadli, Sofiane Bououden, Salim Ziani, and Ivan Zelinka, editors, *Advanced Control Engineering Methods in Electrical Engineering Systems*, volume 522, pages 377–389. Springer International Publishing, Cham, 2019. Series Title: Lecture Notes in Electrical Engineering.

[165] Benham Razavi and Clarence W. De Silva. Condition Monitoring in a Hydraulic System of an Industrial Machine Using Unscented Kalman Filter. *International Journal of Information Acquisition*, 07(03):177–192, 2010.

[166] Mark A. Rubeo and Tony L. Schmitz. Mechanistic force model coefficients: A comparison of linear regression and nonlinear optimization. *Precision Engineering*, 45:311–321, July 2016.

[167] Karl-Erik Rydberg. *Hydraulic Servo Systems: Dynamic Properties and Control*. Linkoping University Electronic Press, Linköping, 2016.

[168] Chithajalu Kiran Sagar, Amrita Priyadarshini, Amit Kumar Gupta, and Devanshi Mathur. Experimental investigation of tool wear characteristics and analytical prediction of tool life using a modified tool wear rate model while machining 90 tungsten heavy alloys. *Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture*, 235(1-2):242–254, 2021.

[169] Priyabrata Sahoo, Tej Pratap, and Karali Patra. A hybrid modelling approach towards prediction of cutting forces in micro end milling of Ti-6Al-4V titanium alloy. *International Journal of Mechanical Sciences*, 150:495–509, January 2019.

[170] Divya Sardana, Raj Bhatnagar, Radu Pavel, and Jon Iverson. Data driven predictive analytics for a spindle's health. In *2015 IEEE International Conference on Big Data (Big Data)*, pages 1378–1387. IEEE, 2015.

[171] Yutaka Sasaki. The truth of the F-measure, 2007. Publication Title: Teach Tutor Mater.

[172] Antans Sauhats, Roman Petrichenko, Zane Broka, Karlis Baltputnis, and Dmitrijs Sobolevskis. ANN-based forecasting of hydropower reservoir inflow. In *2016 57th International Scientific Conference on Power and Electrical Engineering of Riga Technical University (RTUCON)*, pages 1–6, Riga, Latvia, October 2016. IEEE.

[173] Abhinav Saxena, Jose Celaya, Bhaskar Saha, Sankalita Saha, and Kai Goebel. Metrics for offline evaluation of prognostic performance. *International Journal of Prognostics and Health Management*, 1(1):2153–2648, 2010.

[174] C. Sbarufatti, G. Manson, and K. Worden. A numerically-enhanced machine learning approach to damage diagnosis using a Lamb wave sensing network. *Journal of Sound and Vibration*, 333(19):4499–4525, 2014. Publisher: Elsevier.

[175] Claudio Sbarufatti, Matteo Corbetta, Marco Giglio, and Francesco Cadini. Adaptive prognosis of lithium-ion batteries based on the combination of particle filters and radial basis function neural networks. *Journal of Power Sources*, 344:128–140, 2017.

[176] Atharva Sharma, Xiuwen Liu, Xiaojun Yang, and Di Shi. A patch-based convolutional neural network for remote sensing image classification. *Neural Networks*, 95:19–28, November 2017. Publisher: Elsevier Ltd.

[177] Dongfeng Shi and Nabil N. Gindy. Tool wear predictive model based on least squares support vector machines. *Mechanical Systems and Signal Processing*, 21(4):1799–1814, May 2007.

[178] Kundan K. Singh, V. Kartik, and Ramesh Singh. Modeling dynamic stability in high-speed micromilling of Ti-6Al-4V via velocity and chip load dependent cutting coefficients. *International Journal of Machine Tools and Manufacture*, 96:56–66, September 2015.

[179] Leslie N. Smith and Nicholay Topin. Super-convergence: very fast training of neural networks using large learning rates. *https://doi.org/10.1117/12.2520589*, 11006:369–386, May 2019. arXiv: 1708.07120 Publisher: SPIE ISBN: 9781510626775.

[180] Kechen Song and Yunhui Yan. A noise robust method based on completed local binary patterns for hot-rolled steel strip surface defects. *Applied Surface Science*, 285:858–864, November 2013. Publisher: Elsevier B.V.

[181] Thanassis Souflas, Harry Bikas, Mani Ghassempouri, Alessandro Salmi, Eleonora Atzeni, Abdollah Saboori, Ivan Brugnetti, Anna Valente, Federico Mazzucato, and Panagiotis Stavropoulos. A comparative study of dry and cryogenic milling for Directed Energy Deposited IN718 components: effect on process and part quality. *The International Journal of Advanced Manufacturing Technology*, 119(1-2):745–758, March 2022.

[182] Ahmet Soylemezoglu, S. Jagannathan, and Can Saygin. Mahalanobis Taguchi System (MTS) as a Prognostics Tool for Rolling Element Bearing Failures. *Journal of Manufacturing Science and Engineering*, 132(5), 2010.

[183] Ahmet Soylemezoglu, Sarangapani Jagannathan, and Can Saygin. Mahalanobis-Taguchi system as a multi-sensor based decision making prognostics tool for centrifugal pump failures. *IEEE Transactions on Reliability*, 60(4):864–878, 2011. Publisher: IEEE.

[184] Y.V. Srinivasa and M.S. Shunmugam. Mechanistic model for prediction of cutting forces in micro end-milling and experimental comparison. *International Journal of Machine Tools and Manufacture*, 67:18–27, April 2013.

[185] P. Stavropoulos, A. Papacharalampopoulos, E. Vasiliadis, and G. Chryssolouris. Tool wear predictability estimation in milling based on multi-sensorial data. *International Journal of Advanced Manufacturing Technology*, 82(1-4):509–521, 2016.

[186] Xiaogang Su, Xin Yan, and Chih-Ling Tsai. Linear regression. *Wiley Interdisciplinary Reviews: Computational Statistics*, 4(3):275–294, May 2012.

[187] Jian Sun, Hongru Li, and Baohua Xu. Prognostic for hydraulic pump based upon DCT-composite spectrum and the modified echo state network. *SpringerPlus*, 5(1):1293, December 2016.

[188] Ehsan Tahmasebi, Paolo Albertelli, Tommaso Lucchini, Michele Monno, and Valerio Mussi. CFD and experimental analysis of the coolant flow in cryogenic milling. *International Journal of Machine Tools and Manufacture*, 140:20–33, May 2019.

[189] Mingxing Tan and Quoc V. Le. EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. *ArXiv*, May 2019. arXiv: 1905.11946.

[190] Xian Tao, Dapeng Zhang, Wenzhi Ma, Xilong Liu, and De Xu. Automatic metallic surface defect detection and recognition with convolutional neural networks. *Applied Sciences*, 8(9), September 2018. Publisher: MDPI AG.

[191] Thaddeus Tarpey. A Note on the Prediction Sum of Squares Statistic for Restricted Least Squares. *The American Statistician*, 54(2):116, May 2000.

[192] Joel Tellinghuisen. Statistical error propagation. *Journal of Physical Chemistry A*, 105(15):3917–3921, 2001.

[193] Markus Timusk, Mike Lipsett, and Chris K. Mechefske. Fault detection using transient machine signals. *Mechanical Systems and Signal Processing*, 22(7):1724–1749, 2008.

[194] D. A. Tobon-Mejia, K. Medjaher, and N. Zerhouni. CNC machine tool's wear diagnostic and prognostic by using dynamic Bayesian networks. *Mechanical Systems and Signal Processing*, 28:167–182, 2012.

[195] Kristina P. Vatcheva, MinJae Lee, Joseph B. McCormick, Mohammad H. Rahbar, Kristina P. Vatcheva, and MinJae Lee. Multicollinearity in Regression Analyses Conducted in Epidemiologic Studies. *Epidemiology: Open Access*, 06(2):1–20, 2016.

[196] Rajasekar Venkatesan and Meng Joo Er. A novel progressive learning technique for multi-class classification. *Neurocomputing*, 207:310–321, 2016.

[197] Oscar Villalobos, Colin Burvill, and Jacek Stecki. Fault Diagnosis of Electrohydraulic Systems. *Proceedings of the JFPS International Symposium on Fluid Power*, 2005(6):658–663, 2005. ISBN: 493107006X.

[198] James M. Wakiru, Liliane Pintelon, Peter N. Muchiri, and Peter K. Chemweno. A review on lubricant condition monitoring information analysis for maintenance decision support. *Mechanical Systems and Signal Processing*, 118:108–132, March 2019.

[199] M Wan, W H Zhang, G Tan, and G H Qin. New algorithm for calibration of instantaneous cutting-force coefficients and radial run-out parameters in flat end milling. *Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture*, 221(6):1007–1019, June 2007. Publisher: IMECHE.

[200] Min Wan, Dan-Yang Wen, Ying-Chao Ma, and Wei-Hong Zhang. On material separation and cutting force prediction in micro milling through involving the effect of dead metal zone. *International Journal of Machine Tools and Manufacture*, 146:103452, November 2019.

[201] Min Wan, Wei Yin, Wei-Hong Zhang, and Hui Liu. Improved inverse filter for the correction of distorted measured cutting forces. *International Journal of Mechanical Sciences*, 120:276–285, January 2017.

[202] Min Wan, Heng Yuan, Jia Feng, Wei-Hong Zhang, and Wei Yin. Industry-oriented method for measuring the cutting forces based on the deflections of tool shank. *International Journal of Mechanical Sciences*, 130:315–323, September 2017.

[203] Dong Wang and Peter W. Tse. Prognostics of slurry pumps based on a moving-average wear degradation index and a general sequential Monte Carlo method. *Mechanical Systems and Signal Processing*, 56:213–229, 2015. Publisher: Elsevier.

[204] Dongqian Wang, Lars Penter, Albrecht Hänel, Steffen Ihlenfeldt, and Marian Wiercigroch. Stability enhancement and chatter suppression in continuous radial immersion milling. *International Journal of Mechanical Sciences*, page 107711, September 2022.

[205] Jinjiang Wang, Yulin Ma, Laibin Zhang, Robert X. Gao, and Dazhong Wu. Deep learning for smart manufacturing: Methods and applications. *Journal of Manufacturing Systems*, 48:144–156, July 2018. Publisher: Elsevier B.V.

[206] Mei Wang and Jie Wang. CHMM for tool condition monitoring and remaining useful life prediction. *The International Journal of Advanced Manufacturing Technology*, 59(5-8):463–471, March 2012.

[207] Peng Wang, Ziye Liu, Robert X. Gao, and Yuebin Guo. Heterogeneous data-driven hybrid machine learning for tool condition prognosis. *CIRP Annals*, 68(1):455–458, 2019.

[208] Runqiong Wang, Qinghua Song, Zhanqiang Liu, Haifeng Ma, and Zhaojun Liu. Multi-condition identification in milling Ti-6Al-4V thin-walled parts based on sensor fusion. *Mechanical Systems and Signal Processing*, 164:108264, February 2022.

[209] Daniel Weimer, Bernd Scholz-Reiter, and Moshe Shpitalni. Design of deep convolutional neural network architectures for automated feature extraction in industrial inspection. *CIRP Annals*, 65(1):417–420, January 2016. Publisher: Elsevier USA.

[210] Juan Wen and Hongli Gao. Remaining useful life prediction of the ball screw system based on weighted Mahalanobis distance and an exponential model. *Journal of Vibroengineering*, 20(4):1691–1707, June 2018.

[211] Chandula T. Wickramarachchi, Timothy J. Rogers, Thomas E. McLeay, Wayne Leahy, and Elizabeth J. Cross. Online damage detection of cutting tools using Dirichlet process mixture models. *Mechanical Systems and Signal Processing*, 180:109434, November 2022.

[212] Junhui Wu, Yun Ye, Yu Chen, and Zhi Weng. Spot the Difference by Object Detection. *arXiv e-prints*, January 2018. arXiv: 1801.01051.

[213] Sai Wu, Mengdan Zhang, Gang Chen, and Ke Chen. A New Approach to Compute CNNs for Extremely Large Images. In *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*, volume Part F131841, pages 39–48, Singapore, Singapore, November 2017. Association for Computing Machinery, New Yorl, NY, USA.

[214] Yue Wu, G. S. Hong, and W. S. Wong. Prognosis of the probability of failure in tool condition monitoring application-a time series based approach. *The International Journal of Advanced Manufacturing Technology*, 76(1-4):513–521, January 2015.

[215] Yuanmeng Xia, Wei Wang, Zhiyong Song, Zhaoling Xie, Xuezhen Chen, and Hai Li. Fault diagnosis of flexible production line machining center based on PCA and ABC-LVQ. *Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture*, 235(4):594–604, 2021.

[216] Yiqing Yang, Qiang Liu, and Bin Zhang. Three-dimensional chatter stability prediction of milling based on the linear and exponential cutting force model. *The International Journal of Advanced Manufacturing Technology*, 72(9-12):1175–1185, June 2014.

[217] Zhao-jun Yang, Chuan-hai Chen, Fei Chen, Qing-bo Hao, and Bin-bin Xu. Reliability analysis of machining center based on the field data. *Eksploatacja i Niezawodnosc*, 15(2):147–155, 2013.

[218] Hyung Suk Yoon and Kornel F. Ehmann. Dynamics and stability of micro-cutting operations. *International Journal of Mechanical Sciences*, 115-116:81–92, September 2016.

[219] Zhichao You, Hongli Gao, Liang Guo, Yuekai Liu, Jingbo Li, and Changgen Li. Machine vision based adaptive online condition monitoring for milling cutter under spindle rotation. *Mechanical Systems and Signal Processing*, 171:108904, May 2022.

[220] Jinsong Yu, Shuang Liang, Diyin Tang, and Hao Liu. A weighted hidden Markov model approach for continuous-state tool wear monitoring and tool life prediction. *The International Journal of Advanced Manufacturing Technology*, 91(1):201–211, 2017.

[221] Xing Yuan, Shuting Wang, Xinyong Mao, Hongqi Liu, Zhaoshun Liang, Qiushuang Guo, and Rong Yan. Forced vibration mechanism and suppression method for thin-walled workpiece milling. *International Journal of Mechanical Sciences*, 230:107553, September 2022.

[222] Danian Zhan, Shanglei Jiang, Jinbo Niu, and Yuwen Sun. Dynamics modeling and stability analysis of five-axis ball-end milling system with variable pitch tools. *International Journal of Mechanical Sciences*, 182:105774, September 2020.

[223] Cunji Zhang, Xifan Yao, Jianming Zhang, and Hong Jin. Tool Condition Monitoring and Remaining Useful Life Prognostic Based on a Wireless Sensor in Dry Milling Operations. *Sensors*, 16(6):795, May 2016.

[224] Jianlei Zhang, Binil Starly, Yi Cai, Paul H. Cohen, and Yuan-Shin Lee. Particle learning in online tool wear diagnosis and prognosis. *Journal of Manufacturing Processes*, 28:457–463, August 2017.

[225] Pengfei Zhang, Dong Gao, Yong Lu, Fenglei Wang, and Zhirong Liao. A novel smart toolholder with embedded force sensors for milling operations. *Mechanical Systems and Signal Processing*, 175:109130, August 2022.

[226] Xiaochen Zhang, Hongli Gao, and Haifeng Huang. Screw Performance Degradation Assessment Based on Quantum Genetic Algorithm and Dynamic Fuzzy Neural Network. *Shock and Vibration*, 2015:1–11, 2015.

[227] Xiaoyang Zhang, Sheng Wang, Weidong Li, and Xin Lu. Heterogeneous sensors-based feature optimisation and deep learning for tool wear prediction. *International Journal of Advanced Manufacturing Technology*, 114(9-10):2651–2675, June 2021. Publisher: Springer Science and Business Media Deutschland GmbH.

[228] Xing Zhang, Yang Gao, Zhuocheng Guo, Wei Zhang, Jia Yin, and Wanhua Zhao. Physical model-based tool wear and breakage monitoring in milling process. *Mechanical Systems and Signal Processing*, 184:109641, February 2023.

[229] Xing Zhang, Tianhang Pan, Angyang Ma, and Wanhua Zhao. High efficiency orientated milling parameter optimization with tool wear monitoring in roughing operation. *Mechanical Systems and Signal Processing*, 165:108394, February 2022.

[230] Xuewei Zhang, Kornel F. Ehmann, Tianbiao Yu, and Wanshan Wang. Cutting forces in micro-end-milling processes. *International Journal of Machine Tools and Manufacture*, 107:21–40, August 2016.

[231] Xuewei Zhang, Tianbiao Yu, Yuanxing Dai, Sheng Qu, and Ji Zhao. Energy consumption considering tool wear and optimization of cutting parameters in micro milling process. *International Journal of Mechanical Sciences*, 178:105628, July 2020.

[232] Xuewei Zhang, Tianbiao Yu, and Wanshan Wang. Prediction of cutting forces and instantaneous tool deflection in micro end milling by considering tool run-out. *International Journal of Mechanical Sciences*, 136:124–133, February 2018.

[233] Xuewei Zhang, Tianbiao Yu, Pengfei Xu, and Ji Zhao. In-process stochastic tool wear identification and its application to the improved cutting force modeling of micro milling. *Mechanical Systems and Signal Processing*, 164:108233, February 2022.

[234] Yu Zhang, Si Li, and Kunpeng Zhu. Generic instantaneous force modeling and comprehensive real engagement identification in micro-milling. *International Journal of Mechanical Sciences*, 176:105504, June 2020.

[235] PeiYi Zhao, Kai Cheng, Bin Jiang, and LinHan Zuo. Development of the innovative differential tool wear modeling for high-feed milling and its experimental verification. *Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture*, 235(1-2):85–97, January 2021.

[236] Chang'an Zhou, Kai Guo, and Jie Sun. Sound singularity analysis for milling tool condition monitoring towards sustainable manufacturing. *Mechanical Systems and Signal Processing*, 157:107738, August 2021.

[237] Lin Zhou, Ben Deng, Fangyu Peng, Minghui Yang, and Rong Yan. Semi-analytic modelling of cutting forces in micro ball-end milling of NAK80 steel with wear-varying cutting edge and associated nonlinear process characteristics. *International Journal of Mechanical Sciences*, 169:105343, March 2020.

[238] Yudong Zhou, Yanling Tian, Xiubing Jing, and Kornel F. Ehmann. A novel instantaneous uncut chip thickness model for mechanistic cutting force model in micro-end-milling. *The International Journal of Advanced Manufacturing Technology*, 93(5-8):2305–2319, November 2017.

[239] Junda Zhu, Jae M. Yoon, David He, Yongzhi Qu, and Eric Bechhoefer. Lubrication Oil Condition Monitoring and Remaining Useful Life Prediction with Particle Filtering. *International Journal of Prognostics and Health Management*, 4(3), November 2020.

[240] Kunpeng Zhu and Xiaolong Yu. The monitoring of micro milling tool wear conditions by wear area estimation. *Mechanical Systems and Signal Processing*, 93:80–91, September 2017.

[241] Kunpeng Zhu and Yu Zhang. A generic tool wear model and its application to force modeling and wear monitoring in high speed milling. *Mechanical Systems and Signal Processing*, 115:147–161, January 2019.

[242] Gui Lan Zuo, Shang Ding Lai, and Yue Cheng. Study on the Fault Diagnosis of Gear Pump Based on PNN Neural Network. *Advanced Materials Research*, 1044(1):873–876, 2014. ISBN: 9783038352778.