Thèse n° 9607

Virtual metrology applied to milling process

Présentée le 25 mars 2022

Faculté des sciences et techniques de l'ingénieur
Groupe SCI STI DK
Programme doctoral en robotique, contrôle et systèmes intelligents

pour l'obtention du grade de Docteur ès Sciences

par

Paul Arthur Adrien Pierre DREYFUS

Acceptée sur proposition du jury

Prof. C. N. Jones, président du jury
Prof. D. Kyritsis, Prof. B. Faltings, directeurs de thèse
Prof. T. Wuest, rapporteur
Dr R. Perez, rapporteur
Prof. K. Alireza, rapporteur

École
polytechnique
fédérale
de Lausanne

2022

Acknowledgments

First, I would like to express my sincere thankfulness to my advisor Prof. Kiritsis that honored me by giving me the chance to be his last PhD. Thanks to him, those past 4 years were impregnated with confidence, understanding, motivation and good mood.
I would like also to thanks Prof. Faltings that, beside our few meetings, gave me insightful advice on technical subjects.

I would like to thank my mother, to have been continually believing in me since the beginning. I would never have gone half this journey without her. She taught me that nothing was impossible, and that failure was never an end, but only a step of our journey.
I would also like to thank my father for his support. I had the chance to make up for lost time during the end of my thesis and I am more than happy and glad to be able to share this success with him.

I also would like to warmly thank the people in the company that have trusted and supported me during those four years, including workers, technicians, engineers, managers and more. This journey was long and challenging, and we achieved it with success all together.

Throughout my academic career I had the chance to work with a lot of people, including amazing lab coworkers which taught me how a PhD life was going, and that for some of them, still support me regularly. Thank you so much.
I would also like to thank every single of the numerous student that did their semester or last year project with me. It was an honor for me to accompany them on their journey.

More personally, I would like to thank every single people that are or was part of CUBAliente association. I danced and shared with you during those 8 past years. I am glad to be able to still bring the Cuban sunshine in Lausanne with you. I would also want to thank some particular friends. Matteo thanks to have been that supportive and listening with me during those 7 past years. Patrick thank you to have known how to make me smile even during tough moments. Xavier and Son, I am really glad to have been able to do all my studies at your side. To all of them and a lot more, thank you so much.
Going a bit in the past, I would also like to thank Prof. Frédéric Kotecki which was my math professor in high-school and made me love math. I would certainly not be there without him.

Last but not least, thank you to my significant other which has been here on a daily basis and with which I shared, still share and will for sure share the best years of my life.

Lausanne, 22 mars 2022.

Abstract

Production quality and process efficiency are the two main drivers that lead any industrial strategy. To ensure product quality, a duality historically existed between two approaches, namely batch sampling and systematic sampling. In batch sampling, the batch homogeneity is assumed, and a subset is measured, whereas, in systematic sampling, every part of the batch is measured independently. The latter approach is too expensive to be considered. Therefore, batch sampling is implemented, which, however, involves sacrifices product quality, as the homogeneity assumption cannot always be true. Today, virtual metrology (VM) is on the brink of enabling a new standard for quality control by offering a third approach that combines the benefits of the others. VM estimates product quality while preserving process efficiency. It enables continuous systematic monitoring of production in hidden time. To do so, it leverages process data and relates them to product quality through machine learning algorithms. This thesis proposes a proof of concept and guidelines on how to design, tune, implement, and maintain a VM solution for one of the most common industrial operations, namely milling.

The literature lacked a comprehensive summary and a systematic review of the state of the art in VM. This motivated the realization of a systematic, complete, structured literature review of VM based on a content analysis of all research articles in the field published before 2021. The papers are classified thanks to a new framework highlighting shortcomings and future research directions. Finally, VM's use in various industrial fields is discussed, underlining its potential for every manufacturing industry. Two of the various findings of this literature review are the lack of development of VM for machine tool operations and the lack of research on the maintenance of data-based models to ensure their long-term performance. These two research gaps are addressed in the present thesis.

Thereafter, a significant step toward VM for milling operations is made, with a specific focus on the estimation of the dimensional quality of single pass milled parts. To do so, two experiments were conducted in an industrial environment. Various combinations of recorded process variables, data synchronization algorithms, dimensionality reduction algorithms, and regression algorithms were tested to maximize model accuracy. A new baseline for dimensional quality estimation in an industrial production environment is set with a mean absolute error of 14.4 µm. Besides the proof of concept, this part of the present thesis provides directions on how to design and tune a VM algorithm for milling operations.

Finally, a methodology is proposed for implementing a problem-oriented complete solution to ensure the maintenance of industrial data-based models. This solution is structured based on an operational framework, including a sampling decision system and an updating system. Each methodology step is described thanks to guidelines, while research gaps are highlighted. The methodology starts with a concept drift identification phase; thereafter, solutions are pre-selected based on the identified concept drifts. An optimization problem is then designed to select the solution that most respects costs and constraints. This section of the present thesis provides a methodology to implement a solution for the maintenance of industrial data-based models such as VM.

Keywords

Virtual metrology, quality estimation, soft sensor, machine tool, industrial process efficiency, machine learning, zero-defect manufacturing, industry 4.0, drift handling, data-based model maintenance.

Résumé

La qualité des produits et l'efficacité des processus sont les deux facteurs principaux qui guident la plupart des stratégies industrielles. Pour assurer la qualité de la production, il existe une dualité historique entre deux approches, l'échantillonnage par lot ou bien systématique. La première approche suppose une homogénéité des lots et seul un sous-ensemble est échantillonné afin d'être mesuré. En ce qui concerne l'échantillonnage systématique, tous les produits sont mesurés indépendamment. Cette dernière approche est généralement trop coûteuse pour être retenue. L'échantillonnage par lot est donc l'approche la plus utilisée, malgré le risque qu'elle présente pour la qualité de la production. En effet, la supposition d'homogénéité ne peut pas toujours être respectée. Aujourd'hui, la métrologie virtuelle (MV) est sur le point d'établir un nouveau standard pour le contrôle qualité au sein des industries manufacturières. Elle offre en effet, une troisième alternative qui combine les avantages des autres approches. La MV estime la qualité des produits, tout en préservant l'efficacité du processus de fabrication. Cela permet de surveiller la production systématiquement, de manière continue et en temps masqué. Pour cela, la MV utilise les données processus, et les corrèlent avec la qualité du produit grâce à des algorithmes d'apprentissage automatique. Cette thèse propose une démonstration de faisabilité et des directives pour la conception, le réglage, l'implémentation et la maintenance d'une solution de MV pour une des applications les plus communes dans l'industrie, l'opération de fraisage.

En l'absence d'un état de l'art exhaustif en matière de MV, il était nécessaire de proposer un résumé de la littérature systématique, complet et structuré de ce domaine, basé sur l'analyse du contenu de tous les articles scientifiques publiés sur le domaine avant 2021. Un nouveau « framework » permet la classification des manuscrits mettant en lumière des lacunes et des futures directions de recherche. Par la suite, les nombreux domaines d'application de la MV sont discutés, soulignant son potentiel pour toutes les industries manufacturières. Cet état de l'art permet premièrement de conclure au manque de développement de la MV sur les opérations basées sur les machines outils. Il met aussi en évidence les lacunes dans le domaine de la maintenance des algorithmes basés sur la donnée pour assurer la pérennité des performances des algorithmes de MV. Cette thèse traite de ces deux sujets.

Une étape importante vers la MV pour les opérations de fraisage est franchie. Un accent particulier est mis sur l'estimation de la qualité dimensionnelle des pièces fraisées en une seule passe, à l'aide d'outils d'apprentissage automatique. Pour ce faire, deux expériences sont conduites dans un environnement industriel. De nombreuses combinaisons de type de variables d'entrée, d'algorithmes de synchronisation des données, d'algorithmes de réduction de dimensionnalité et d'algorithmes de régressions, ont été testées dans le but de maximiser la précision de la solution de MV. Dans les conditions du cas d'étude, une nouvelle référence pour l'estimation de la qualité dimensionnelle dans un environnement industriel, est établi avec une erreur absolue moyenne de 14.4 µm. Outre la démonstration de faisabilité, ce chapitre de la thèse fournit des indications sur la manière de concevoir et de régler un algorithme de MV pour les opérations de fraisage.

Une méthodologie est proposée pour l'implémentation d'une solution complète de maintenance d'algorithme basé sur la donnée. Cette solution est structurée grâce à un framework opérationnel, qui inclut un algorithme de décisions d'échantillonnage et un système de mise à jour. Chaque étape de la méthodologie est décrite grâce à de nombreuses directives pour leur implémentation. De multiples opportunités de recherche sont également mises en lumière. La méthodologie débute avec une phase d'identification des dérives de concept ; par la suite, des solutions de maintenance sont présélectionnées, sur la base des caractéristiques des dérives précédemment identifiées. Un problème d'optimisation est ensuite défini afin de sélectionner la solution qui minimise les coûts, tout en respectant les contraintes. Ce chapitre fournit une méthodologie afin de mettre en œuvre une solution pour la maintenance de modèles industriels basés sur des données tels que la MV.

Mots-clés

Métrologie virtuelle, estimation qualité, capteur mou, machine outil, efficacité du processus industriel, machine learning, zero-defect manufacturing, industrie 4.0, gestion de dérive, maintenance d'algorithme basé sur la donnée.

Contents

List of Figures

List of Tables

List of Abbreviations

ART2	Adaptive-Resonance-Theory 2
BNN	Bayesian Neural Network
CD	Concept drift
CMP	Chemical-Mechanical Planarization
CNN	Convolutional Neural Network
CPS	Cyber physical system
CVD	Chemical Vapor Deposition
DD	Drift Detection
dEWMA	double Exponentially Weighted Moving Average
EWMA	Exponentially Weighted Moving Average
GAN	Generative Adversarial Network
GMM	Gaussian Mixture Model
GPR	Gaussian process regression
JIT	Just In Time
KMC	k-Means Clustering-based novelty detection
LASSO	Least Absolute Shrinkage and Selection Operator
MES	Manufacturing Execution System
ML	Machine Learning
MLR	Multiple Linear Regression
MoG	Mixture of Gaussian
MPC	Model Predictive Control
NN	Neural network
PCA	Principal Component Analysis
PECVD	Plasma-Enhanced Chemical Vapor Deposition
PI	Plasma Information
PLS	Partial least square
RNN	Recursive neural network
SDS	Sampling Decision System
SOM	Self-Organising Maps
SOTA	State Of The Art
SPD	Stability-Plasticity Dilemma
SVDD	Support Vector Data Description
SVR	Support vector regression
t-SNE	t-distributed Stochastic Neighbour Embedding
US	Updating system
VM	Virtual Metrology
ZDM	Zero Defect Manufacturing

Chapter 1 Introduction

1.1 Industry 4.0

Industries transform raw materials into higher-value goods. Today, most of the industrial companies are driven by profit. At the shopfloor level, this latter is represented by two main drivers: product quality, which affects sales, and process efficiency, which affects costs. These drivers have always existed in industry. While initially underemphasized, process efficiency began to flourish during the first industrial revolution. At that time, a transition from a society based on an agricultural-commercial economy to an industrial society began due to new technology enabling the extraction of energy from steam. Industry then, saw a second industrial revolution due to the development of electrical motors, which enabled critical innovations in production machinery. Thereafter, the development of new information and communication technology enabled the third industrial revolution, which modified the ways in which employees worked. Moreover, it modified production machinery by automatizing some production tasks. It could be claimed that every industrial revolution has been based on the development of disruptive technology that directly affects production machinery, enabling more efficient processes and higher-quality products. The industrial drivers have not changed – only the means of production have.

Today, due to the miniaturization of sensors and chips, an increasing number of systems are instrumented; these are called cyber-physical systems (CPSs). CPSs enable the production of a vast and continuous flows of data referred to as data streams. Data streams are characterized by a continuous flow of streaming data from various sources. All of these newly available data often carry important amount of unexploited information, which can be leveraged for process or product enhancement. However, data streams also produce certain constraints. For example, the flow of data must be analyzed with fast inference algorithms and requires a database to have a forgetting capability, as the flow is considered infinite, whereas the storage space cannot be. The impact of these constraints is discussed further later in this thesis. The sudden availability of data aided the development of novel, disruptive, nonlinear, data-based inference approaches known collectively as machine learning. Machine learning emerged due to the exponentially growing co-evolution of data availability and computational power. It soon became a critical research topic, as it enabled the modelization of systems which was still misunderstood. Moreover, data-based models can update themselves automatically can therefore adapt to unpredicted changes. This is impossible for model-based approaches, which limits their use on the shopfloor. Nearly all research fields have access to data and extract information using machine learning; industrial research is no exception. The potential product and process enhancement proved so important that the application of machine learning soon came to be referred to as the fourth industrial revolution, also known as Industry 4.0. In this case, machine learning enabled by CPS was the disruptive technology that permitted an evolution of production means. The goal of this ongoing industrial revolution is to enable the full modelization of a shopfloor, also called a digital twin. Through creating digital twins, numerous applications could disrupt the way in which a shopfloor is organized. In a previous publication, the author of this thesis proposed a framework (see Figure 1) that illustrates the impact of Industry 4.0 on shopfloor organization [1]. Thanks to virtual measurement, the quality of a product can be determined in real time, enabling a live automatic tuning of the production machine thanks to a “tuning assistant,” thus enhancing both product quality and production performance. This predictive maintenance will predict incoming failures and optimize the maintenance time to failure, thus enhancing production performance. Based on feedback from predictive maintenance and the tuning assistant, auto-scheduling will enable potential issues to be forecast at the level of the shopfloor as well as allow production to be scheduled in an optimal manner to improve process efficiency. Industry 4.0 is moving toward a fully autonomous robust production capable of handling very small lots to allow total product customization. Total product customization is a popular new paradigm that is not currently feasible due to the limited capability of industries.

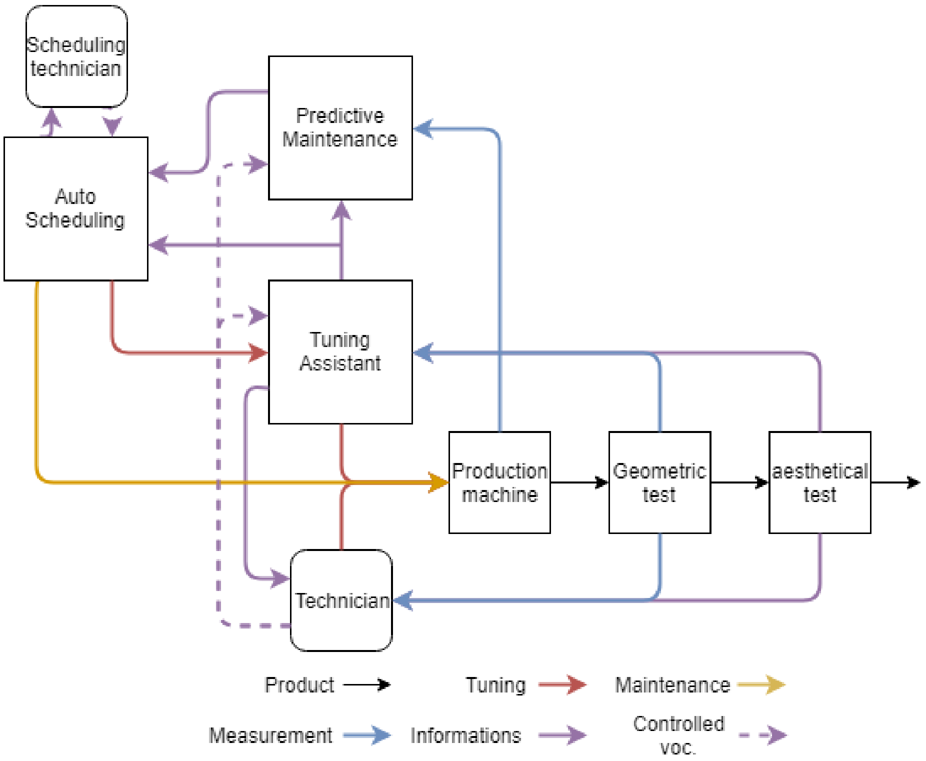


Figure 1: Framework of Industry 4.0's impact on the shopfloor.

The fourth industrial revolution promises important evolution in the future. However, in the short term, industries are still preparing for this transition. Indeed, to be able to monitor and control any machine or process, data must be available, and very few companies have begun to implement fully connected machinery. Production machines have a long lifetime, and their development is a lengthy process. Only recently have they started to include technology for extracting data. Previously, numerical command machines were not produced to extract data, which makes their retrofitting (i.e., updating old machines) highly complicated. This means that a very limited number of machines present on shopfloors today can obtain data from a process and generally can do so only in a limited manner. This industrial reality matches the state of the art of Industry 4.0 technology, as most current modelization solutions are in development, which highlights the work that still must be done to achieve the aforementioned digital twin objective. One of the most fundamental modelization technologies that is yet to be mastered is product quality estimation, which is also called virtual metrology (VM). Since VM is the privileged link between product and process, it is at the basis of the long-term vision for Industry 4.0.

1.2 Virtual metrology

The contemporary manufacturing domain is shifting to more sustainable manufacturing practices to more efficiently respond to complex market desires. Market demands have increased significantly, whereas product life cycles and manufacturing cycles have become considerably shorter, introducing new challenges for manufacturers [2], [3]. Furthermore, product customization has forced manufacturers to reduce batch sizes, which has increased defect rates [4], [5]. The key factor in achieving sustainable manufacturing is the assurance of product quality [6]. Therefore, manufacturers must measure and qualify the quality of their products at each stage of production [7], [8].The two main quality control strategies are systematic control, where every product is measured, and batch control, where only a few products from each batch are measured and assumed to be representative of the whole batch. Additionally, as stated by Psarommatis et al., the measuring process can be performed with two approaches, namely physical or virtual measurement [6]. As physically inspecting parts is time consuming and expensive, manufacturing companies currently prefer batch control to systematic control due to the cost and performance that can be achieved. However, batch control has a major disadvantage in that defective products may reach the consumer, which can result in devastating costs to a company [9], [10]. Moreover, even if defective parts are prevented from going to market, they usually cost, as they were transformed from defect occurrence to defect detection. Systematic control is the optimal practice for ensuring that all products conform to the desired quality characteristics [11], [12]. However, this approach is very expensive and inefficient, as measurement generally takes longer than production, which explains why batch control is the most widespread approach for ensuring product quality. Virtual inspection has become a reality due to modern technological advancements, and it has enabled manufacturers to apply systematic control without the cost and time disadvantages of physical inspection [6]. The concept of virtual inspection lies within a broader concept called VM, which enables the relaxation of the batch control homogeneity assumption while preserving production efficiency. VM was first developed for the semiconductor industry in 2005 [13]. In this industry, products are very expensive to produce, and physical inspections are often lengthy and destructive. These disadvantages induce industrial constraints, which may explain the motivation behind the development of VM.

In the present research, the following definition of VM is adopted: "VM involves estimating the characteristics that define a product's quality directly from production process data and mostly using data-driven algorithms." The approach is product-centered, meaning that it exploits data from the means of production to create information about a product [14]. VM can also incorporate a drift detector to add a safeguard against process evolution, and it may also be updatable and adaptable. Quality estimation can be performed using a sampling decision system (SDS) to optimize the real measurement frequency as well as a control algorithm to close the loop on the production mean.

VM lies within the zero-defect manufacturing (ZDM) approach. ZDM proposes several different methods with the shared goal of decreasing and mitigating failures in manufacturing processes—that is, it emphasizes doing things right the first time. These methods can be classified as detection, prediction and repair, and prevention methods [6]. VM is a part of the *detection* class of methods, as it enables the detection of defects that have already occurred. Furthermore, automatic machine control, which can repair or even prevent defects, is a production application enabled by VM [15]. The present study focused on ZDM because of its advanced capabilities compared with traditional quality improvement methods, such as Six Sigma, Lean Manufacturing, the Theory of Constraints, and Total Quality Management. The superiority of ZDM lies in the fact that after each defect is detected, production methods learn from it for the future, which does not occur in traditional quality improvement methods [16]. ZDM can be implemented through two approaches, namely the product-oriented and process-oriented approaches [6]. Both approaches lead to ZDM but have different starting points.

1.3 Industrial data-based model maintenance

With the emergence of Industry 4.0, an increasing number of processes are monitored, generating continuously tremendous quantities of data. This enables the implementation of impactful data-driven technologies, which can lead to higher levels of sustainability [17], [18]. VM and predictive maintenance are some potent data-driven concepts within ZDM that are heavily dependent on data, with their performance relying on the accuracy and adaptability of the corresponding models. Currently, most developed models are implemented on a data flow that is often considered stationary. Indeed, the models are built on data that can be seen as a short recording of reality; however, it is important to keep in mind the difference between reality and the reality represented by the data, as they can differ over time. This difference can arise from multiple sources; for instance, if the training dataset sampling time is too short or if some independent impactful variables are not measured. In any case, data-based models will become obsolete over time. The cause of this phenomenon is often referred to as concept drift (CD), which is described as unpredictable changes in the data stream distribution over time. A strategy for adapting data-based models to handle CD must be defined to develop any data-based industrial application. Indeed, all industrial technologies that interact with the shopfloor require a maintenance plan to ensure their long-term efficiency. Data-based models are no exception, and they are becoming a fundamental aspect of the well-known ZDM paradigm. Today, much effort is invested in the development of industrial data-based models, but very little is invested in their maintenance. In this study, efforts are made to compensate for this imbalance.

1.4 Research questions

The thesis first started with an industrial will to move toward Industry 4.0. As discussed, the future impact of such technologies is critical. In the future, it will be a crucial competitive advantage across all industries drivers regrouping for instance product quality or process efficiency. However, as with all new technology, the deployment of the infrastructure requires considerable time and investment. Today, pending questions are how to safely connect the production means to the servers and how to gather appropriate data from the correct source. To answer those questions, as for the snake that bites its own tail, prior knowledge on the technologies is required; however, data must be available to implement them. In other words, a company that could start its transition toward Industry 4.0 would have to invest in Industry 4.0 technology to increase its know-how and to refine its industry 4.0. requirement definition.

Therefore, this thesis is based on actual industrial problems, which are analyzed to extract research questions (see Figure 2). The first industrial question focused on product quality. As previously discussed, quality control technology is nonoptimal, highlighting the potential for significant improvements. The two drivers are product quality first and process efficiency second. Indeed, new quality controls that harm process efficiency can be counterproductive depending on a company's objectives. VM is the most fitting solution for addressing those requirements. As VM was only developed for semiconductors, the first research question is defined as follows: "Is the application of VM to estimate the dimensional quality of products milled with a machine tool feasible?" Applicability is defined by the utility of a technology for a manufacturer, which depends on the accuracy of the data-based model. The limit of applicability can be defined based on the production tolerance, which largely differs depending on the industries and products. The second industrial question dealt with the preservation of such technologies. Indeed, as with new production machines or processes, every element implemented on the shopfloor necessarily comes with a maintenance plan that ensures its long-term operation. Based on this question, the first driver was machine learning, as the system in need of maintenance is a data-based model. The second driver was "long term," as the first maintenance goal is to make the solution last as long as possible. The research question formulated based on those drivers is as follows: "How to maintain industrial data-based model?" Maintaining a model means detecting concept drift and updating the machine learning models. Industrial data-based model, correspond to model applied to industrial problems.

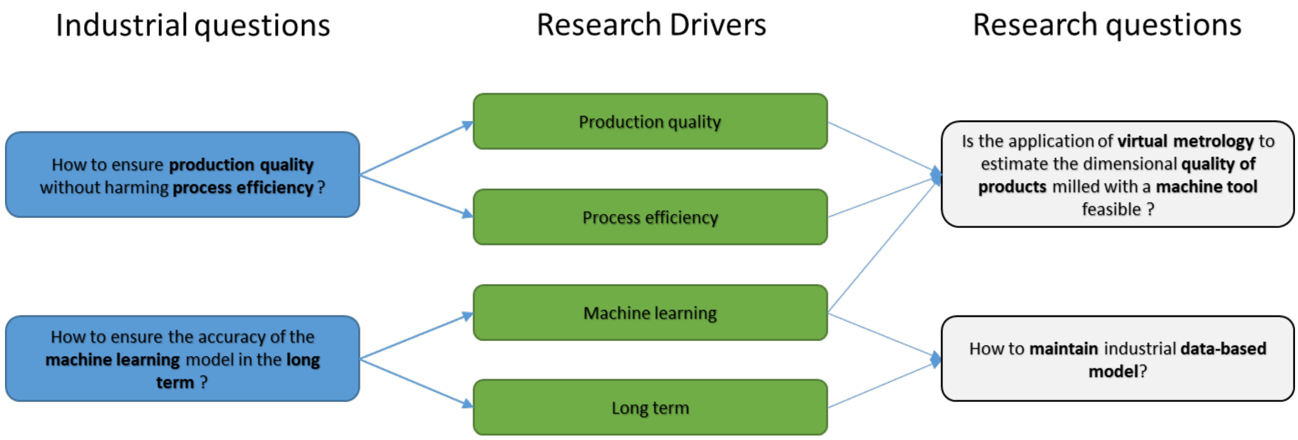


Figure 2: From industrial questions to the research questions.

The remained of this thesis is organized into three main chapters, followed by a chapter that presents the conclusion and directions for future research. In Chapter 2 a complete, ordered, and systematic state of the art of the industrial solution for VM is presented. The VM solution is structured thanks to a framework that includes all the elements studied in the state of the art related to VM. The aim of this chapter is to structure the overall solution and to provide an overview of the state of the art of VM. Chapter 3 presents a study of VM for machine tools based on two different industrial case studies where the VM models are developed, trained, and tested. This chapter is intended to be a proof of concept that VM can be applied to milling operations on machine tools. Chapter 4 presents an innovative methodology for optimizing the design and operation of industrial data-based model maintenance. Based on prior data acquisition and prior knowledge, CD is characterized. Thereafter, maintenance solutions composed of a sampling decision system (SDS) and an updating system are designed. Finally, based on industrial constraints, the optimal maintenance solution is selected. The proposed methodology is tested on a simulation.

Chapter 2 Virtual metrology framework and state of the art

Virtual metrology (VM) falls within a larger group of approaches described as soft metrology [19], which brings together concepts such as soft sensors that focus on chemical and continuous manufacturing applications. VM was initially developed for the field of semiconductor production. Today, however, VM is of interest to the entire manufacturing domain as it can significantly improve the sustainability of a manufacturing network [12], [20]–[22]. VM is a fairly new concept, which is verified by the fact that the first paper to define VM was published in 2005 [13]. Figure 3 illustrates the results of a bibliometric analysis on the number of publications per year, revealing a clear and steady upward trend. This trend is reasonable and expected as more manufacturers and researchers are understanding the need for alternative, more efficient methods of measuring the quality of their products.

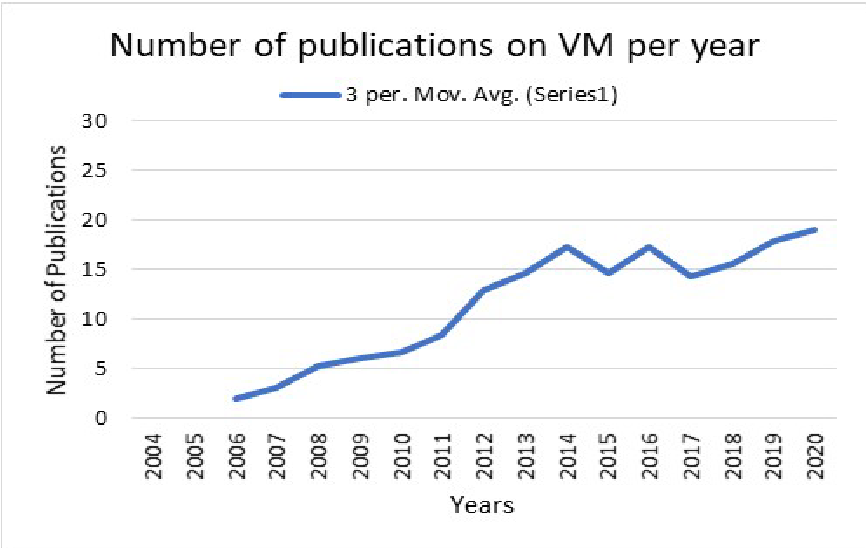


Figure 3: Publications on virtual metrology, from the first paper in 2005 to November 2020.

To respond at the research questions, a complete literature review needs to be done. It enables to highlight potential solutions as well as define potential research gap. To do so, a complete, structured, systematic review of the field of VM is done. It proposes a framework with which to structure and popularize all research activities linked with VM and reviews the technical shortcomings of all of its components, applications, and fields of use. State-of-the-art (SOTA) machine learning (ML) is also leveraged to highlight new opportunities for VM, as are other high-potential new fields of application.

In the following section, first the research method of the study is described. Then the proposed VM framework and its components is presented. After what they are analyzed and discussed. Followed by a discussion on the advancements and opportunities of VM industrial applications. Finally, the main findings and outcomes of the study are highlighted, mapping out directions for further research.

2.1 Research methodology

This systematic literature review used procedures [23] exemplified in an earlier review [24]. This section defines the methodology to make the process replicable and objective.

2.1.1 Data sources and search strategy

After the removal of duplicates, the literature on the SOTA of VM comprised 352 papers found through searching the Scopus, Web of Science, IEEE, and Engineer Village databases. A literature search querying phase usually forces keyword assumptions; however, for a complete review of the SOTA of VM, we used the unique query keyword of "Virtual Metrology." The date range of the systematic review was from 2004—when the first paper to propose the term VM was published—to November 2020.

2.1.2 Article selection

Once the 352 papers containing the keyword of 'Virtual Metrology' had been collected, four rejection criteria were defined to make the systematic review repeatable:

- **Accessibility**: Papers not accessible in English were not considered.
- **Crucial effects**: Articles ignored by the VM community would have negatively affected the variance of our statistical constructs. To avoid such effects, papers published more than 3 years previously and never cited were removed.
- **Repetition**: To ensure statistical quality and decrease potential bias, conference papers that had been republished as journal papers were removed.
- **Subject relevance**: For consistency, all the papers compared had to respect VM product-centered paradigms and be oriented toward an industrial topic. For instance, a paper dealing with the estimation of a machine's production efficiency and not with the quality of the final product would have been excluded from this review.

The effects of each of these selection–rejection criteria are summarized in Figure 4. In the bibliometric analysis, most of the papers were found to deal with semiconductor applications. However, the search was not limited to only this field because of the potential of VM in other manufacturing domains.

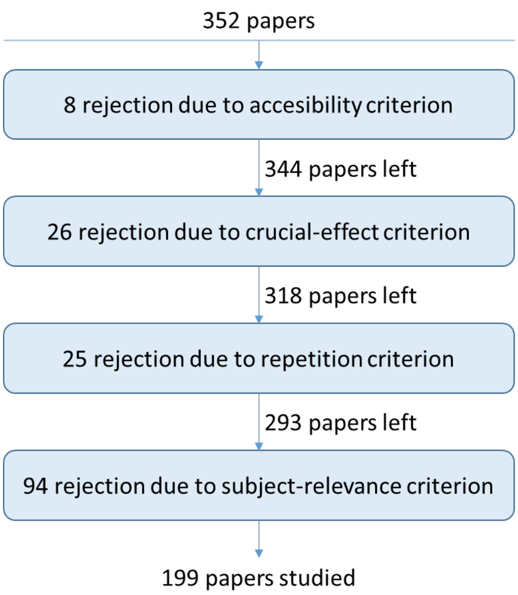


Figure 4: Selection of relevant articles as per the research methodology.

2.1.3 Quality assessment of the retained articles

A total of 199 papers were extracted according to the aforementioned criteria, and then they were classified depending on their topic and application. Every component and application of VM was studied, either exhaustively or in a quantitative manner. For the component and application counting too few papers to extract trends, an exhaustive review was performed; otherwise, a more quantitative approach to studying VM was employed. In this case, in addition to studying the statistics, some significant papers were also examined qualitatively. The criteria used to select them were as follows:

1. ML relevance: The proposed VM approach corresponded to commonly used, SOTA machine learning approaches.
2. Research impact: The paper was in the top 25% most cited papers in its domain of VM based on its Scopus citation record.
3. Recentness: The paper was published in 2018 or earlier.
4. Originality: The paper proposed an original, high-potential solution for resolving a known problem.

To be highlighted as significant research, a paper needed to be ML-relevant and fulfill at least one of the other criteria—either 2, 3, or 4. The fourth criterion enabled the consideration of under-rated solutions from another perspective. Every manufacturing application of VM was reviewed quantitatively, and they are presented later in the discussion section.

2.1.4 Data extraction and synthesis

To guide the answer of the research questions, 11 categories of VM were dichotomized into two groups. The technical group contained the components and applications of VM, as described in the framework presented in Table 1. The group of VM manufacturing applications classified industrial operations into the two categories in Table 2. Each paper was nonexclusively classified into one or more categories. To document the systematically reviewed papers, all of the papers are cited either in the text (for highlighted papers) or in the citation table in the Annex 1.

Table 1: Summary of technical group categories. Q: quantitative review. EXH: exhaustive review.

Category name	**Short explanation**	**Papers**	**Review**
Preprocessing	Data preprocessing before a quality estimation.	149	Q
Quality estimation	Quality estimation based on machine learning.	168	Q
Drift detection	Detection of concept drift.	22	EXH
Sampling decision system	Based on the detection of drift, this system adapts the measurement frequency, enabling the quality estimator to be updated and ensuring system reliability.	9	EXH
Updatability feature	Variation of the quality estimator, enabling it to be updated.	31	Q
Adaptability feature	Variation of the quality estimator, enhancing its data efficiency.	11	EXH
Multi-stage architecture	Architecture of the interconnections between different implementations of VM.	5	EXH
Machine control	Machine control based on quality estimation.	25	Q
Fab-wide architecture	VM implementation in industrial environments.	13	EXH

Table 2: Summary of manufacturing application categories. Q: quantitative review. EXH: exhaustive review.

Category name	**Short explanation**	**Papers**	**Review**
Semiconductor manufacturing	All studies of VM operations involving semiconductors.	158	Q
Other manufacturing domain	All studies of VM operations involving other manufacturing domains.	12	Q

2.2 Virtual metrology framework

Based on the analysis of the literature, the components and applications of VM were structured as a framework that depicted their interconnections. This framework is presented in Figure 5.

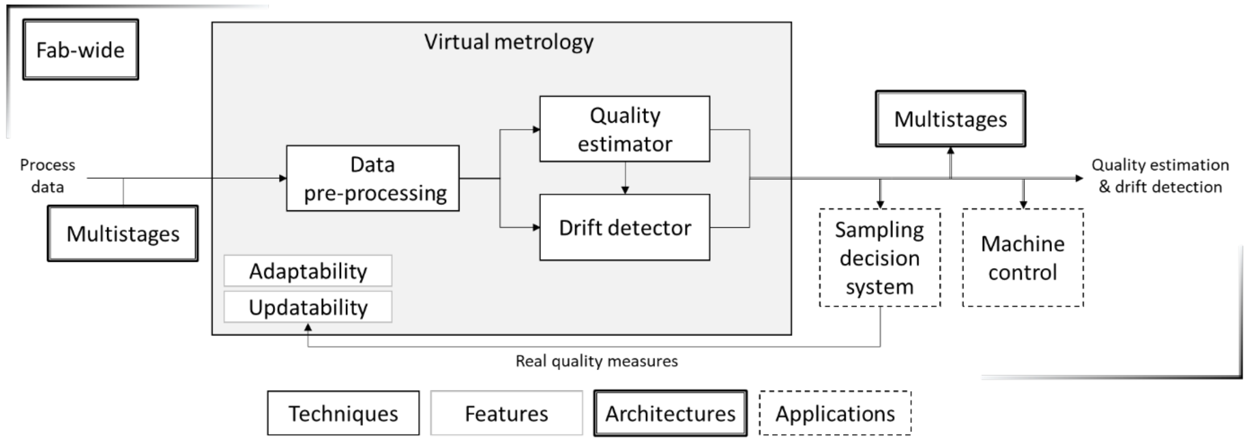


Figure 5: Virtual metrology framework.

The proposed framework was aimed at comprehensively connecting all of the components and applications of VM. The elements of the framework are summarized as follows: The VM block takes the raw process data as inputs. Those inputs then undergo preprocessing before being processed by the quality estimator and the drift detector. This *preprocessing step*, presented in Section 2.3.1, includes the nonexhaustive removal of outliers, reduction of dimensionality, and normalization of raw data. The quality estimator, presented in Section 2.3.2, extracts information from the preprocessed data to estimate the product quality. The drift detector, presented in Section 2.3.3, detects abnormal behavior in the preprocessing data and raises internal alarms when a drift is uncovered. All of these techniques can be enhanced to enable the VM to be updatable or more data-efficient. The updatability feature is discussed in Section 2.3.5, and the adaptability feature is discussed in Section 2.3.6. The data-efficiency enhancement provided by the adaptability feature enables faster prototyping and implementation of a new VM model. The implementation of VM in an industrial environment induces the need to define two connections: The first is the horizontal interconnection of the different VMs implemented along the production life cycle of the product, which is discussed in Section 2.3.7; the second is the vertical connection of the VM to the manufacturing execution system, known as a Fab-wide architecture, which is discussed in Section 2.3.9. Finally, the VM output is composed of a quality estimation and a drift detection alarm, which are used by two main applications: First, the sampling decision system (SDS), presented in Section 2.3.4, optimizes the product measurement rate; then, the production machine can be controlled based on the quality estimated, which is discussed in Section 2.3.8. Noteworthily, other Industry 4.0 applications can benefit from VM quality estimation [25]. For instance, predictive maintenance could benefit from VM since product quality is strongly linked to machines' health [26]–[28]. Vice-versa, VM can benefit significantly from predictive maintenance through the forecasting of upcoming concept drift [29]. Indeed, causes of machine failure are also causes of quality drift. However, because nearly no research exists on the link between those applications and VM, they were omitted from the proposed framework.

As depicted in Figure 6, each element of the framework can be classified by its role. This explains the place it holds in the implementation of VM.

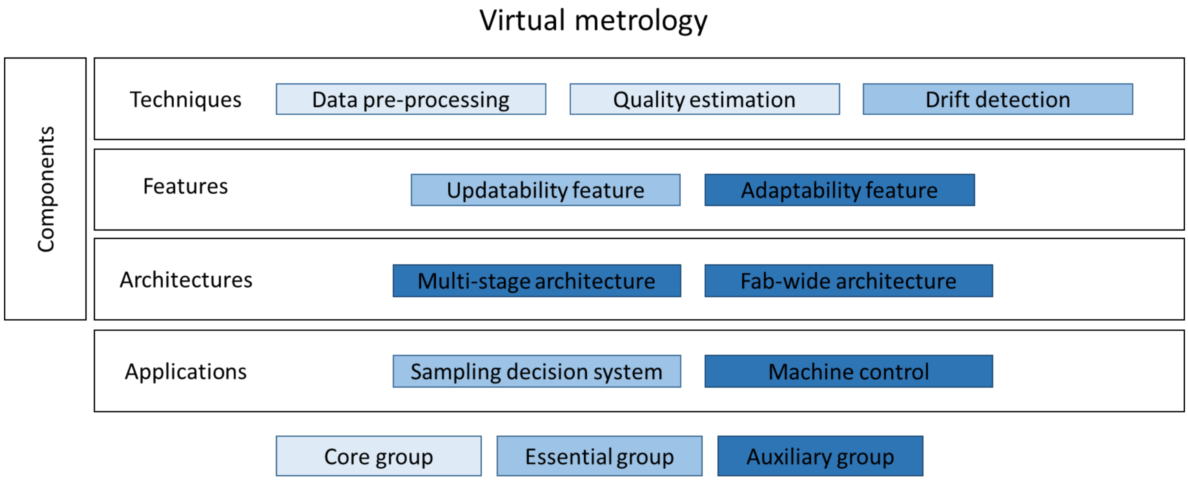


Figure 6: Virtual metrology terminology chart.

The most basic elements are those that extract information from the data. Most of the time, these are supervised or unsupervised ML algorithms, which obtain value from the process or from each other. They are referred to as *techniques*.

Features are technique modifications for achieving secondary objectives, such as reducing the minimum size of the training set or making the techniques updatable. They often come as technique add-ons, which still require technique modification to be implemented.

Architectures define the way VM is connected to the environment. They deal with the data flux horizontally—that is, all the operations required to manufacture the product—and vertically—that is, the enterprise software system.

Applications are external elements that use *techniques* to enhance or provide new benefits. As they are external from VM, they are not components. Every application can be described in the same way as VM, with techniques, features, and architectures.

It is worthwhile classifying the framework's elements by their necessity. Necessity is defined as the element's importance for the implementation of VM as presented figure 5.

The *core concept group* includes the elements that are strictly necessary for building VM and estimating product quality. The group is composed exclusively of techniques that are sufficient for implementing a basic VM solution. These techniques should be designed and implemented first.

The *essential group* includes the elements required for building a lasting VM solution. As with any Industry 4.0 solution, the VM environment evolves with time, and therefore, every technique needs to be updated. The essential group includes all of the elements that enable what could be called technique maintenance. These elements are not necessary for short-term VM implementation but are necessary for long-term VM implementation. For industrial deployment, essential elements must be taken into consideration.

The a*uxiliary group* contains the optional elements that create value but are not necessary for basic or long-term VM implementations. Such elements are still worth considering for enhancing the system in terms of industrial implementation or application.

2.3 Analysis and discussion of virtual metrology elements

2.3.1 Data preprocessing

Data preprocessing is a technique that transforms data to make it compatible with other techniques and also enhances the accuracy of quality estimation. Data preprocessing is a required technique for implementing VM, and thus, it is part of the core concept group. Preprocessing includes multiple operations that can all be performed offline and sometimes online, such as the preparation of the initial data, assessment of missing data (sometimes online), outlier removal (online), data normalization (online), and feature engineering (sometimes online).

Even though all of the retained papers described the use of preprocessing—as a fundamental step of ML—few explained which preprocessing phases were used, and even fewer explicitly mentioned which algorithms they employed. This explains why fewer papers were cited in our preprocessing statistics than in our quality estimation statistics and also why only outlier removal and feature selection were discussed. Moreover, this lack of transparency may also give rise to problems related to experiment repeatability.

Outlier removal

Outliers are aberrant data points lying outside of the expected distribution, which can lower the precision of the quality estimator if considered. Detecting and removing them from the dataset is a thoroughly discussed preprocessing operation in the VM field. The most used outlier removal approach is promptly described in the Table 3.

Manual outlier detection is the simplest and most used approach. Features can be plotted independently or not [30] for visualizing potential outliers. This method is purely offline, which is an important limitation for VM implementation. Statistics-based methods are also highly popular as they are the simplest online approach. A specific clustering method named ART2 was developed for VM outlier detection [31]. ART2 is widely used for VM as it is included in the automatic VM framework [32]. Lastly, a ML field called novelty detection proposes advanced solutions for detecting outliers. Some of those solutions have been tested successfully for VM [33]–[35]. The main advantage of novelty detection is that it can deal with data streams. Indeed, whereas the other solutions assume that the distribution they learned is fixed, a data stream accepts that it will evolve with time, adding a new challenge—whether it is "an aberrant point, an outlier, or just the evolution of the normality." Detecting aberrant points is not performed simply to increase the accuracy but also—and especially—to protect the quality estimator from the time-variation of its environment. This is a major challenge that should be prioritized more.

Table 3: outlier removal algorithms. All the citations are available in Table 19 of the Annex 1.

Methods name	**Algorithm's name**	**Description**	**Advantage**	**Limitation**
Manual outlier detection	Basic plot, t-SNE	Handmade based on visualization approaches.	• Simple	• Only offline
Statistics-based methods	Hotelling's T^2, Grubb's test	Statistical algorithm for testing the distribution distance between two sets of data.	• Simple	• Statistical assumptions
Principal component analysis (PCA)-based methods	PCA	Distance monitoring of each data point projected into another space.	• Simple	• Linear transformation
Specific method	ART2	Specific clustering algorithm.		
Novelty detection methods	SVDD, Gauss, MoG, KMC, (k-NN), GMM, Parzen window, SOM	Clustering or one-class classification algorithm that detects novelty.	• Nonlinear	• Computationally demanding

The field of novelty detection abounds with more advanced solutions that have yet to be studied for VM applications [36]. Outlier detection is highly connected to drift detection, which is discussed later in Section 2.3.3.

Dimensionality reduction

Dimensionality reduction consists of the modification of the feature's space, from a high dimensional space toward a low dimensional space, while maintaining the maximum amount of meaningful information. Working with a high dimensional feature space is detrimental to VM application due to the twin curses of dimensionality and computational intractability. Moreover, keeping features uncorrelated with the output can reduce the model's overall effectiveness [37], [38]. The VM feature's space size usually has a very high dimension due to the many process sensors sampling at a high frequency for several minutes at a time. This explains why dimensionality reduction is a fundamental preprocessing step in VM.

Dimensionality reduction can be achieved by removing less meaningful features, using feature selection, or transforming all of the features toward a lower dimensional space using feature extraction. Finally, dimensionality reduction can also be performed by reducing the size of each feature, as with data quantization.

Feature selection combines the dimensionality reduction approaches that preserve features by simply decreasing the feature space dimension. In other words, redundant or irrelevant features are removed. One advantage of feature selection is its transparency. Indeed, information about a feature's usefulness can, for example, explain the modification of the process itself. Feature-selection methods are usually divided into three main categories: filter, wrapper, and embedded methods. The most used feature selection algorithms are presented in the Table 4. Alternatively, feature selection could also simply be conducted by a process expert.

Table 4: Feature selection algorithms based on a similar table in [19], All of the citations are available in Table 20 of the Annex 1.

Methods name	**Algorithm's name**	**Description**	**Advantages**	**Limitations**
Filter method	Pearson, input covariance, ANOVA	Correlation-based approach	• Computationally efficient • Simple • Interpretable	• Mostly linear • Not associated with the inference model
Wrapper method	Forward, backward, or stepwise selection, genetic algorithm	Iterative approach	• Nonlinear • Interpretable • Associated with the inference model	• Computationally demanding • Difficult to generalize • Can overfit
Embedded method	Lasso, trees	Inference algorithm with in-built feature selection	• Can be nonlinear • Two in one • Interpretability	• Depends on the inference algorithm

Although computationally demanding, wrapper methods are highly popular due to their capability for high dimensionality reduction. In the reviewed papers, 59% of wrapper methods were used with linear approaches such as multiple linear regression (MLR) or partial least squares (PLS) as inference algorithms, whereas 41% were combined with nonlinear approaches such as neural networks (NNs), Gaussian process regression (GPR), or support vector regression (SVR). Taking into account the high number of iterations required to converge, weak learners such as linear learners were often used. The use of nonlinear learners highlights the nonlinearity of most VM problems. Indeed, applying linear algorithms to nonlinear problems will lead to underperformance. Similarly, filter methods are not used anymore. By contrast, the popularity of embedded methods has grown significantly, carried by multiple recent publications that have used ensemble methods as quality estimators.

Feature-extraction approaches transform an initial feature space into a smaller one. This transformation permits the removal of any hidden redundancy between features, which can be more efficient than feature selection in terms of reducing dimensionality. Moreover, the newly extracted feature space can increase the precision of the quality estimator and reduce the overall noise. The greatest limitation of the feature-extraction approach is the difficulty in extracting process knowledge from the new feature space once it has been transformed. Moreover, unlike the feature-selection approach, which transforms the feature space, the feature-extraction approach must be implemented both offline and online, thus increasing the calculation time. Feature selection and extraction are equally used in SOTA VM. The most used feature extraction algorithms are presented in the Table 5.

PCA and PLS are widely used methods for VM that are restrained by their linear assumption. To ease this limitation, a kernel can be associated with them to explore nonlinear relationships, a method usually called kernel PCA or kernel PLS [33], [39], [40]. Notably, all existing nonlinear feature-extraction approaches, including CNNs and autoencoders, have been described for the first time since 2017, reflecting the need to move toward nonlinear algorithms. Moreover, thanks to the fast evolution of CNN applications in the field of ML, the use of CNNs has also been increasing in VM applications, representing more than 70% of dimensionality reduction algorithms described in 2020.

Table 5: Feature-extraction algorithms. All of the citations are available in Table 20 of the Annex 1.

Algorithm name	**Description**	**Advantages**	**Limitations**
Principal component analysis (PCA)	Input features projected on the eigenspace, removing the smallest eigenvector	• Computationally efficient	• Linear • Unsupervised
Partial least squares (PLS)	Like PCA but matches both projection of the input and the output	• Computationally efficient	• Linear • Interpretability
Convolutional neural network (CNN)	Inference neural network algorithm with in-built feature extraction	• Nonlinear • Two in one	• Computationally demanding • Large learning set
Autoencoder	Neural network-based feature extraction	• Nonlinear	• Computationally demanding • Unsupervised • Large learning set

Data quantization, very commonly used for VM, reduces the dataset size by binning it by features and representing the bins using statistics. In addition to the most commonly generated statistics (such as mean, median, and maximum), excellent results have been reported for the segregation of steady-state and transient periods for building new statistics, such as steady-state level and duration, settling time, and rise time [41], [42]. Binning can also be performed at a fixed frequency or by focusing on specific zones of interest that correspond to production recipe steps [43]. Reducing the size of the dataset often comes at the price of a great loss of information, which may reduce the inference accuracy. However, some approaches require a qualitative feature and may have greater accuracy.

2.3.2 Quality estimation

Quality estimation is one of the core concepts of VM techniques. It is aimed at estimating product quality using measurable process variables. Defining a product's quality is application-dependent, but it usually has a single output, such as etching depth for plasma-etching applications. Moreover, quality estimations are often defined as continuous values, which encourages the use of regression algorithms for inference tasks. However, some papers have evaluated classification approaches that simply describe product quality as within tolerance or out of tolerance [33], [44]. Using classification for quality estimation is often called fault detection and classification. For instance, novel detection algorithms have been applied for discrete VM estimations of quality [45]. Most of the approaches used in VM are data-driven ML approaches. Discrete VM is restrictive and should only be used if data is too scarce to train a regression algorithm. Even if prior knowledge can, in some cases, be considered, algorithms learn thanks to data-driven representations of the system (i.e., a dataset). The approaches presented here are supervised as they always use the input–output couple to learn. Numerous algorithms are used to perform quality estimation, and they are presented in Table 6.

Table 6: Quality estimation algorithms based on a similar table in [19]. All of the citations are available in Table 22 of the Annex 1.

Method name	Algorithm name	Advantages	Limitations
Linear models	Multiple linear regression (MLR)	• Low complexity • Computationally and data-efficient • Interpretable	• Linear • Prone to instability • Sensitive to outliers
	Partial least square (PLS)	• Computationally and data-efficient • Feature extraction	• Linear • Interpretability • Risk of omitting real correlations
	Lasso	• Computationally and data-efficient • Feature selection • Interpretable	• Linear • Low performance with highly correlated variables • Low performance if the number of descriptors exceeds the number of observations
Neural networks	Multiple layer perceptron (MLP)	• Nonlinear • Ability to generate new features from a subset of features located in the training dataset	• Computationally demanding • Large dataset required • Black box
	Convolutional neural network (CNN)	• Nonlinear • Feature extraction • More robust	• Computationally demanding • Large dataset required • Black box
	Recursive neural network (RNN)	• Nonlinear • Time memory: effective for time series	• Computationally demanding • Large dataset required • Black box
	Bayesian neural network (BNN)	• Nonlinear • Causal • Prior-knowledge injection	• Computationally demanding • Large dataset required • Black box
Kernel methods	Gaussian process regression (GPR)	• Nonlinear • Data-efficient • In-built uncertainty quantification	• Computationally demanding • Poorly scalable • Hard to tune • Assumes Gaussianity
	Support vector regression (SVR)	• Non-linear • Data-efficient • Less prone to overfitting	• Poorly scalable • Difficult to tune
Ensemble methods	Bagging	• Nonlinear • Easily updatable • In-built uncertainty quantification • Reduces variance	• Can result in high bias
	Stacking	• Nonlinear • Enhances accuracy	• Depends on the learners that are used

Linear regressors are currently the most frequently used algorithms for quality estimation; however, although they provide fast inferences, they do not achieve high accuracy and can be unstable with nonlinear problems. Their popularity may be driven by their data efficiency in an industry where data is scarce considering continually changing production-machine health. Their popularity may also be driven by their noise rejection ability.

The second most used algorithm family is NNs. Even if MLP was for long the most used approach, today it is under-represented, replaced by more advanced neural network architectures. Recurrent NNs (RNNs) use time dependencies to increase their accuracy when dealing with time series. However, RNNs suffer from heavy data-efficiency issues, which explains their lack of popularity. Moreover, recent research demonstrated that CNNs can also capture time dependencies in a highly data-efficient manner and with greater accuracy [46], enabling greater accuracy and robustness against drifts, which could explain their recent popularity in the

VM field. Attention mechanism algorithms [47] have become a highly popular approach for replacing RNNs in SOTA ML. Future research on attention mechanisms for VM applications could be interesting. Papers featuring Bayesian NNs (BNNs) were particularly well represented in 2020 because BNNs have brought many new, valuable capabilities to the table, such as prior knowledge transfer from computer-aided manufacturing tools [48] and causality links. Causal inference is another popular topic in the field of VM as it ensures tremendous interpretability, enabling rapid correction after the detection of sources of process failure [49]. Hybrid architectures had excellent results in SOTA VM, such as a combination of an RNN and a CNN for enhancing robustness against concept drift [50]. Noteworthily, thanks to the evolution of calculation power, NN architectures and hyper parameters tend to now be optimized by genetic algorithms for VM applications [51], [52]. Such NN variations hold tremendous opportunities for enhancing the accuracy, precision, and interpretability of quality estimators; thus, they should be studied further.

The third most frequently used approach involves kernel methods. The use of SVR is declining while GPR is becoming ever more popular. GPR is a high-potential approach in the field of VM as it provides in-built uncertainty quantification, which is a critical advantage because it greatly simplifies the implementation of perennial VM solutions. GPR's greatest limitation is its lack of scalability, which can be limited if combined with the right dimensionality reduction algorithm such as lasso [53] or a CNN [54].

Despite a significant number of studies in 2020 adopting ensemble methods, they remain the least used approaches for quality estimation. The bagging algorithm has crucial advantages as it can not only estimate uncertainty like GPR but can also be updated simply through the addition or removal of an individual from the ensemble. Stacking is known to perform better than single learners for some problems [55], which tend to be validated for VM [30], [56], [57]. Most ensemble methods permit the easy implementation of essential group components, and although they are recent, they should be considered by practitioners for industrial applications.

2.3.3 Drift detection

The quality estimator learns a concept defined by the representation of the studied physical equations, namely the learning dataset. This concept is generally time dependent. Its evolution is called concept drift [58], and if not considered, it will constantly decrease the accuracy of quality estimations. When dealing with an issue as critical as product quality estimation, it is necessary to ensure the correctness of conjectures. This problem is known as the applicability or manufacturability problem [59]. The technique of drift detection (DD) is part of the essential group, and it is a necessary element for making VM perennial in time. DD assumes the role of the alarm by detecting concept drifts based on the inputs and feeding the information into the SDS. DD outputs a Boolean that indicates whether a drift is actually occurring. Implementing DD involves two steps, the first of which is detecting an abnormality, and the second is characterizing the concept drift. These two steps, including the algorithms used, are described in detail in the following two subsections.

Detecting an abnormality

The first group of algorithms used to detect drift in SOTA VM is called uncertainty quantification. Very often assuming Gaussianity, these algorithms identify a standard deviation of the quality estimation, capturing its "confidence." Based on a problem-dependent heuristic threshold, abnormality can be detected. One advantage of calculating the confidence is that besides DD, it can be leveraged to enhance both the quality estimators [60] and machine control algorithms [61], [62]. The automatic virtual metrology (AVM) solutions propose an add-on heuristic approach called the reliance index, which is based on the crossover of two different quality estimators [63]. This used to be the most popular method for detecting abnormality, but many other interesting alternatives are available today. For instance, regression approaches with in-built uncertainty quantification such as GPR and bagging are convincing for VM applications [64]. Moreover, in SOTA ML, multiple methods are yet to be studied, such as BNNs with uncertainty quantification, concrete dropout [65], [66], and deep ensembles [67].

As discussed in the outlier detection section, the second group of algorithms mostly originate from SOTA novelty detection. They are one-class classification or clustering algorithms. Generally, with the threshold indirectly included in their hyper parameters, they directly output a Boolean value for detecting an abnormality. They are known to be highly efficient at detecting drift [36].

Despite the central role of DD in the essential group, it has not been thoroughly studied for VM, and numerous algorithms remain to be explored.

Characterizing a concept drift

Detected abnormalities must be classified as either outliers or concept drift. These are different entities because outliers are not representative of the physical distributions [68]. Once a drift has been characterized, the product must be measured to update the quality estimator. However, measurements are expensive and measuring outliers is useless. Furthermore, including outliers in the training set is detrimental. To discriminate concept drift from outliers, the simplest solution used in VM is to measure output only if multiple successive abnormal samples have been detected [69]. Once the measurement has been performed, another VM approach suggests a second verification of whether the output is abnormal; if the output is normal and the input is abnormal, then the sample is considered an outlier [35]. Numerous methods exist for characterizing a concept drift that remain to be explored, particularly those originating from SOTA novelty detection.

2.3.4 Sampling decision system

The SDS is the main application of VM because it enables the production measurement rate to be adapted based on the virtual quality estimation. Its decisions are built around the quality estimator output, the DD output, and the time elapsed since the last measurement. SDS strategies simply based on time are called passive or time-based strategies, whereas SDS strategies only based on VM output are called active or event-based strategies. SDS strategies based on both time and VM output are called hybrid strategies. The Figure 7, which is inspired by a figure by [70] illustrates the different type of sampling strategy; They are also described in the Table 7.

Table 7: Sampling decision system strategies.

Method name	Algorithm name	Description	Advantages	Limitations
Time-based or Passive strategy	Statistical process control	Optimization of the fixed frequency	• Simple • No need for VM	• Under-optimized quality and efficiency
Event-based or Active strategy	Uncertainty sampling	Update when an event occurs		• Hidden-context sensitivity
Hybrid strategy	Simple or no interaction hybridization	Update at the fixed frequency or when an event occurs	• Partially optimized SDS	• Room for optimization
	Optimized interaction hybridization	Update at the optimized frequency or when an event occurs	• Optimized SDS	• Room for optimization • Complex

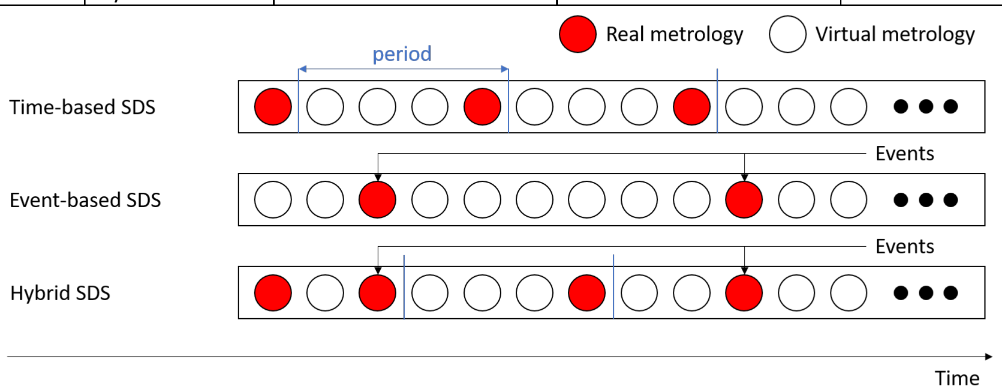


Figure 7: SDS strategies.

Event-based strategies compensate for the two aforementioned limitations of time-based strategies, as the measurement decision is based on the specific quality of each product. The main event-based SDS strategy is called uncertainty sampling [71]. It consists of triggering a measurement when the estimation is no longer trusted [72], [73], out of tolerance [74], or impossible, as with just-in-time (JIT) learning extrapolation problems [75]. The first approach, namely uncertainty sampling, has been integrated into all SDSs across the SOTA papers on VM. Event-based strategies should never be used alone because there are always sources of deviation that are hidden and then not visible for the DD, making the whole system fallible.

Today, time-based strategies are the main strategies used in production. The frequency at which measurements should be conducted are either optimized through a traditional statistical process control approach or simply defined heuristically. Most of the time, these strategies under-perform in terms of product quality because they assume a homogeneous batch, and also in terms of process efficiency because they assume the system to be time-independent.

Hybrid SDS strategies perform the best because they have the advantages of both time- and event-based strategies. A fixed-time-based frequency accounts for hidden drifts, while an event-based frequency takes care of the other types of drift [76]. Some approaches propose adapting the time-based frequency in function of the event occurrence [69]. By doing so, one can leverage some room for optimization from the fixed-frequency hybrid strategy; however, a theoretical background is nowadays missing to ensure that the model update frequency follows the hidden context evolution. In fact, this background is also missing when setting up the fixed-frequency hybrid strategy, which indicates great research opportunities.

2.3.5 Updatability feature

Once a labeled example is available, the updatability feature enables all techniques to be updated, including preprocessing [77]. The updatability feature enters the essential group as it enables VM to deal with concept drifts. To do so, it alters the way the learning set is stocked thanks to an external algorithm, thus avoiding the limitations of a huge learning set. The updatability features are described in the Table 8. The stability–plasticity dilemma (SPD) is fundamental for differentiating the various methods of updating [78]. A highly stable method will have high inertia and adapt slowly, whereas a method with high plasticity will adapt quicker to new concepts but may become unstable.

Table 8: Updatability algorithms. All of the citations are available in Table 23 of the Annex 1.

<table>
<tr><th>Algorithm name</th><th>Algorithm variation</th><th>Description</th><th>Advantages</th><th>Limitations</th></tr>
<tr><td rowspan="2">Moving window (MW)</td><td>Moving window</td><td>Keeps the youngest points in memory</td><td>• Simple
• SPD depending on window size</td><td rowspan="2">• Can forget meaningful points
• Bad against recurrent drift</td></tr>
<tr><td>Weighted moving window</td><td>The same but with an emphasis on the youngest point</td><td>• The same
• Increases plasticity</td></tr>
<tr><td colspan="2">Just-in-time (JIT) learning</td><td>Keeps the full set, trains on a subset on the fly</td><td>• Deals with recurrent drift
• Fast inference
• Good SPD</td><td>• No forgetting capability
• Only applicable to fast learners</td></tr>
</table>

Moving window (MW) is the most commonly used approach for dealing with data streams for VM thanks to its simplicity and its “forgetting capability”. Various MW ameliorations exist, such as adaptive MW or JIT learning MW [79], although only few studies have examined them for VM [80], [81].

JIT learning is popular for VM. It enables very large learning sets to be dealt with by retraining the inference algorithm at each inference on a small subset selected using different clustering algorithms [82], [83]. However, it is highly limited by its lack of ability to “forget,” making it impossible to use for applications that are not purely cyclical. Moreover, even if sometimes applied with slow learners [53], JIT learning needs to be applied online to be combined with fast learners, which explains why it is nearly exclusively used with linear algorithms. One-third of linear approaches use JIT learning, whereas all other approaches use MW.

One means of supplementing an updating algorithm—if one has prior knowledge about the physical system—is to model concept drifts separately to decompose the problem [84] with, for instance, NNs [85] or wavelet transform [86]. For example, maintenance—a common cause of concept drift—has been modeled multiple times [87]–[89]. This method is rarely sufficient as all of the causes of drift are generally unknown; it should be implemented as a complement to MW or JIT approaches.

Besides these methods found in SOTA VM papers, ensemble methods are often used for their high stability and plasticity. Indeed, ensemble methods can be updated easily by adding or removing new members; they would probably outperform the currently used methods for VM [90].

2.3.6 Adaptability feature

The adaptability feature aims to diminish the set-up time of a quality estimator, for new or different products, by reducing the number and size of the features required to train that quality estimator. In doing so, it increases VM's agility and responsiveness, which is increasingly useful in a world where personalization and product diversity have become the norm. Multiple approaches have been studied to reach those goals, and they are listed in Table 9.

Table 9: Adaptability approaches.

Approach name	**Description**	**Advantages**	**Limitations**
System identification	Analytically learns the difference between two products to compensate for it	• Simple • Keeps the same model	• Very low applicability
Transfer learning	Uses an existing VM of a similar product to pretrain a new product VM	• Faster learning time • Smaller learning set	• Needs similar products
Multi-task learning	Trains two similar tasks' VMs simultaneously	• Faster learning time • Smaller learning set	• Detrimental if the tasks are not similar
Semi-supervised learning	Generates training data from an unlabeled point	• Smaller learning set	• Sensitive to concept drift

Constant system identification has been applied for VM [91], [92]. However, it remains very limited as very few products differ from a simple constant.

Another method called transfer learning enables the learning time and the learning set size to be reduced by presetting optimized parameters with a learning algorithm that has already been trained on a similar task. Enhanced accuracy and computational efficiency been found for VM applications [93].

When multiple quality estimators must be jointly developed, multi-task learning can be implemented to reduce both the learning time and the learning set size. In SOTA VM, multi-task learning is applied to linear algorithms, ensemble methods, and kernel methods, and its data efficiency has been proven [94]. However, for different tasks to be jointly learned, they must be selected wisely as this approach could become counter-effective if the difference between them is too great.

Another approach is called semi-supervised learning, which automatically generates labeled data from unlabeled data to diminish the learning set size. For VM, all of the tested approaches label new points based on a consensus ensemble of homogeneous learners trained on different subsets of the dataset, as with the co-training algorithm [95], [96], or an ensemble of heterogeneous learners trained on the same dataset, as with the SAFER algorithm [97]. This approach has been noticed to be highly counter-effective if drift occurs, as this will lead to a biased consensus and the inclusion of false information in the training set.

Adaptability features are a crucial topic as labeled data is scarce and expensive in the industry, which is underdeveloped today. Thus, great research opportunities exist in this area.

2.3.7 Multi-stage architecture

VM's main goal is to estimate a product's quality based on production process data. Current quality usually depends on the quality of past operations. In other words, the quality at operation *n-1* affects operation *n*, and this is true along the production line. The multi-stage architecture in operation defines VM's interconnectedness; it deals with the integration and connection of multiple VM points along the product's production line to enhance the overall accuracy of quality estimation. Multiple multi-stage architectures exist; the present review proposes a three-group classification described in Table 10 and illustrated in Figure 8.

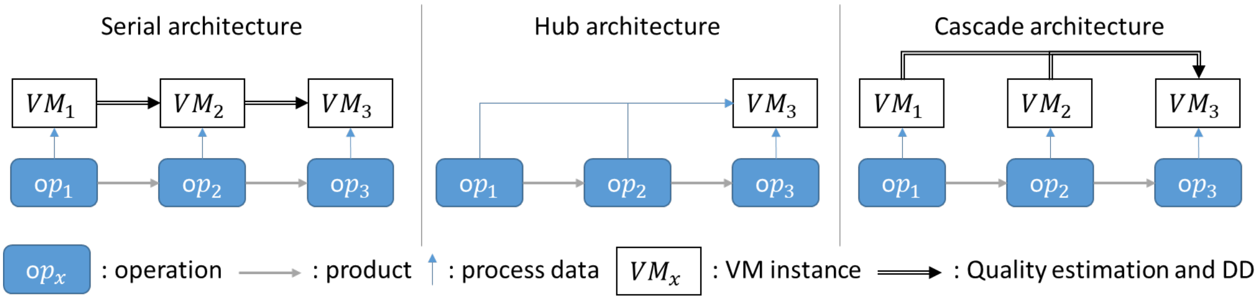


Figure 8: Multi-stage architecture diagrams.

Table 10: Multi-stage architectures.

Architecture name	**Description**	**Advantages**	**Limitations**
Serial architecture	Serially interconnected VM points.	• Efficient	• Information loss
Hub architecture	Use an existing VM of a similar product to pre-train a new product VM.	• Fewer VM points • Raw information	• Requires high computational power
Cascade architecture	Mix between serial and hub architecture.	• Depends on the exact architecture	• Depends on the exact architecture

A serial architecture improves VM accuracy without increasing the feature space of each VM point much [98]. It has been commonly used for VM and its applications, such as machine control, where it takes both current estimations and estimations from previous operations as inputs [99].

A hub architecture can increase VM's accuracy and is advantageous when the quality of past operations is not measurable or of low accuracy [30]. Moreover, developing and maintaining VM involve financial costs, which in some cases must be minimized. This architecture can be used either as a quality control point or as a final quality control verification. The hub architecture's main limitation comes from the curse of dimensionality as the feature space grows with each past operation considered, which is limiting in the real-time implementations.

A cascade architecture has pros and cons that are highly dependent on the final architecture. For example, one cascade architecture proposes independent VMs at each point of operation along the production line, but only feeds their estimations to the VM point for the last operation. This last VM point uses all preceding estimations plus the data from the last operation to estimate the product's final quality [100], [101].
As a guideline, the information flux should follow the operations depending on their correlation. Because most manufacturing operation dependence are not serial, most implementations should use cascade architectures. However, research in this field is scant. Therefore, more research must be conducted to compare the different multi-stage architectures and to determine how to develop them.

2.3.8 Machine control

Automatic machine control is an application of VM that leverages quality estimation and DD to control the means of production. By doing so, the production process can be stabilized and robustified. Most control approaches involve run-to-run control—they update process parameters at the end of each production run. This does not control the quality of each part but attempts to center the quality distribution around tolerance. VM can therefore deal with process drift [102] at the batch level, but it cannot ensure the quality of every piece produced. Multiple control algorithms are used with VM, and they are presented in Table 11.

Table 11: Control algorithms. All of the citations are available in Table 24 of the Annex 1.

Algorithm name	Description	Advantages	Limitations
Exponentially weighted moving average (EWMA)	Low pass filter with linear controller.	• Simple • Reduce noise	• Sensitive to trends (slow drift) • Linear controller
Double EWMA (dEWMA)	Same as EWMA, with a second EWMA for trends.	• Same as EWMA • Deals with trends	• Linear controller
Model predictive control (MPC)	Trajectory optimization based on the process model.	• Nonlinear controller • Optimal	• Requires a precise process model

The vast majority of control algorithms for VM use EWMA or dEWMA, which are exclusively implemented in run-to-run. Multiple variations have been studied for SOTA VM. The output quality measures can, for instance, be included in an EWMA filter when available [103]. Those newly included physical measures have a high weighting compared with virtual measures and they are used as temporary baselines. Some research using EWMA/dEWMA filters has taken information regarding uncertainty quantification into account. This can be used to fine-tune the weighting parameter in real time. High uncertainty will increase the weighting of actual measurements and vice-versa. It has also been shown to increase controller performance.

Another active research topic is the controller's robustness against concept drift. One approach that was proposed is to feedforward the measured quality to modify the controller gains. This allows a slow incremental drift to be detected and compensated for [104]. Another approach integrates knowledge about equipment condition into a dEWMA, which is estimated by a dynamic Bayesian network [61]. The main limitation of the EWMA/dEWMA filter is its assumption of linearity and its run-to-run implementation.

Model predictive control (MPC) is one of the few approaches that combines VM with online, real-time control [105]. Unfortunately, MPC is heavily dependent on the system simulation, which is mostly unavailable in VM applications. Another real-time approach merges the quality estimator and the controller by learning the inverse model, thus estimating the process parameters to achieve the desired quality of the output. Thus, quality is not estimated directly but the control task is performed. In this approach, a generative adversarial network is employed to train the CNN used for inference and to deal with concept drift caused by regular updates [62].

Real-time consideration may seem useless for the main VM application, namely SDS, which explains the popularity of run-to-run control. However, from a control perspective, real-time quality estimation is the holy grail as it would enable the real-time control of production apparatuses in terms of quality, which would be a huge step toward the zero defect manufacturing (ZDM) paradigm [106]. Although a dense body of knowledge on control theory exists, much research still needs to be done with a particular focus on real-time nonlinear approaches that incorporate information regarding uncertainty quantification.

2.3.9 Fab-wide architecture

The need for Fab-wide architectures is a fundamental aspect of any industrial big-data approach. A Fab-wide architecture is defined to connect the different organ of an industrial VM implementation. Its location is either on the shopfloor, on a computer on the production machine, either on the intranet, on the manufacturing execution system (MES), or on the cloud. The different architectures are presented in the Table 12.

Table 12: Fab-wide architectures.

Architecture Name:	Generic VM	MES VM	Partial cloud VM	Full cloud VM
Clouds			VM management server Model generation server Central database	VM management server Model generation server Central database VM unit
MES		VM management server Model generation server Central database		
Shop floor	VM unit	VM unit	VM unit	Data collection device

The generic VM architecture is the simplest that exists and should be the first step for new industrial VM implementations. The VM unit is as integrated as possible with the production apparatus [107]. It enables real-time capabilities while being highly flexible; it also diminishes the risk of data-packet asynchronicity problems or even losses.

MES VM architectures introduce three new components, which are shared for all of the different VM units. The model generation server consists of a multi-tenant server that can receive and run multiple VM creation models on different virtual machines simultaneously, thus ensuring user isolation [108]. The central database stocks all data and models. Using main memory database technology for VM applications has produced better results than disk-resident technology [109]. The VM management server deals with the control and maintenance of VM models [110], [111]. This approach enables better overall control of VM capability across a whole fab.

A *hot trend* in Industry 4.0. is products as services, where product providers rent services, maintain control of systems, and ensure their maintenance. The manufacturing industry or client rents VM solutions, outsourcing their implementation, operation, and maintenance problems. To implement products as services, VM solutions must be moved to the cloud. The partial cloud VM architecture pushes the model creation server, VM manager, and central database to the cloud [112]–[114]. In this case, the VM operation is kept close to the production apparatus, enabling real-time implementation. This final approach seems to be the optimal compromise because it successfully separates the online and offline parts.

The full cloud VM architecture places every possible component in the cloud, including the VM algorithm. The only part that remains on the machine is the data collection device [115]. Adding VM in a server dramatically increases the server's resistance to software failure thanks to the easy replacement of virtual machines; computational power can also easily be higher. On the other hand, real-time VM is compromised due to the lag between the cloud and shopfloor.

In terms of implementation, research has proposed multiple solutions for industrial VM. *Apache Hadoop* is a big-data-friendly, open-source ecosystem that can improve on the query processing speeds and data storage efficiency of handmade solutions [116]. Another solution is Microsoft's *Azure Machine Learning Studio*, which provides a drag-and-drop way to train, deploy, and manage

VM solutions [117]. Noteworthily, some Fab-wide architectures foresee not only the inclusion of VM but also other technologies such as predictive maintenance or ontology structuring.

2.4 Virtual metrology industrial application and discussion

VM is a generic ZDM tool that can theoretically be applied to every manufacturing application. Originally designed for semiconductor manufacturing, VM has recently been applied in a multitude of industrial fields to various process with different dynamics. It is impossible to find one algorithm for every component in the VM framework that will fit all the different dynamics. Accordingly, this section discusses the advancement and opportunities in those industrial applications. It is organized in 2 subsections, first discussing the semiconductor application and finally the other manufacturing domains.

2.4.1 Semiconductor manufacturing

SOTA VM for semiconductor applications can be grouped into the following five operations: photolithography, chemical vapor deposition (CVD), plasma-enhanced chemical vapor deposition (PECVD), plasma-etching, and chemical-mechanical planarization (CMP). All the citations for semiconductor manufacturing VM papers, classified by applications, are provided in Table 26 of the Annex 1.

The most studied semiconductor operation incorporating VM is plasma-etching, which usually involves removing a photoresist layer. As discussed in a 2010 review, this is a high-value-added operation usually run in open-loop, thus providing a strong motivation to control the quality of that loop using retroaction [106]. Following the review's recommendation, multiple control approaches have been tested based on VM, such as real-time MPC, which was highly successful and acted as a proof of concept [105]. Unfortunately, no other real-time implementation of VM has emerged from the SOTA. One might think that with the emergence of ZDM, real-time consideration will soon become a hot topic. Regarding plasma-etching, it is noteworthy that the use of plasma information (PI) as an input has enabled online control and opened new opportunities for VM; PI is still very commonly used [89]. Speaking of features, there is still a great deal of emphasis on dimensionality reduction with a preference for feature extractions. Plasma-etching is no different from other semiconductor applications in its strong need to reduce the number of processed inputs. As concluded by a 2009 review [118], due to nonlinearity, the best and most commonly used algorithm for the quality estimator is a NN. MLP was and still is the most popular choice. However, there is a notable increase in interest in CNN applications, which achieve increased accuracy and robustness compared with MLP [119]. Indeed, plasma-etching is not spared from concept drift such as residue deposition [106], justifying the potential of CNNs. On the other hand, the implementation of the essential group component is the optimal approach for dealing with concept drift, yet it remains under-researched [73].

The second most studied operation is CVD, along with its variant PECVD. Notably, most papers have not explicitly described the exact industrial operations involved, and thus, they remain vague. PECVD is an exception and is usually differentiated from CVD, although no clear differences exist between the algorithms used for both techniques. CVD is used to produce thin-films, either for adding a new layer or for doping an existing one. As with plasma-etching, great effort is expended to diminish the feature space size, with an emphasis on feature extraction algorithms. In terms of the quality estimator, a 2014 review on VM for CVD explained that the use of SVR is optimal for this application [120]. Since then, kernel methods have become the most used methods, among which GPR is the most popular. As GPR is not a scalable algorithm, its rise can be explained by the development of feature extraction algorithms. In terms of VM technical applications, very few papers have focused on machine control [87] for CVD. Because it enhances VM's impact on the shopfloor, it can only be beneficial for implementing CVD control, which motivates further research. Conversely, the implementation of essential group components is an active topic for CVD research. The most used approaches are MW for updatability, diverse DD methods, and SDS, demonstrating the clear presence of concept drift in this industrial application.

CMP is the third most studied application. It is used to planarize and clean the top layer, to finish the wafer, or to prepare it for its next operation. CMP can be viewed as a bridge between semiconductor manufacturing and other manufacturing domains. Indeed, planarization can also indicate "polishing," which is an operation found in many production industries. This application is quite recent as 50% of CMP VM papers have been published since 2018. Another point that makes CMP unique is that dimensionality reduction does not seem to be a problem. Very little research has been conducted on this subject for CMP, and most papers have not explained the use of one. Quality estimators are quasi-exclusively implemented with NNs but surprisingly not MLP. More advanced architectures are used such as RNNs, CNNs, or BNNs. BNN applications are mostly applied to CMP. CNNs and ensemble algorithm seem to be the approaches with the most potential for quality estimation across the industrial applications. Dealing with drift is another active topic, but remarkably the most used updatability feature is JIT learning, not MW. In terms of control, both classic controllers such as dEWMA and highly innovative controllers such as the one based on genetic adversarial network (GAN; [62] have been discussed. CMP is like the laboratory of VM for semiconductor manufacturing because many new technologies and algorithms have been tested with CMP, yet they still need to be tested for all the other industrial applications such as CNNs, BNNs, GANs, and more. It is the VM field to watch as it is the most active and innovative.

Finally, the least studied operation incorporating VM is photolithography, which shapes the final substrate. Very few papers have been published on its different components, making the resulting discussion less objective. Moreover, very little has been said in terms of pre-processing, concept drift handling, and applications. However, MLP is the most commonly used approach for quality estimation, indicating potentially high nonlinearity.

Lastly, it was found that since 2015, the scientific focus on CMP and more marginally represented semiconductor applications has grown significantly, such as copper-clad laminate [121], physical vapor deposition [30], laser chip manufacturing [122], roll-to-roll manufacturing [123], [124], and wafer sawing [56] among others. This clearly emphasizes the future opportunities for VM in new, less studied semiconductor applications.

2.4.2 Other manufacturing domains

Beyond the semiconductor domain, VM is also adopted in other domains, such as the following:

- Machining operations, including milling, turning, drilling, and spark machining
- Metal additive manufacturing
- Carbon fiber manufacturing

All of the citations for papers outside of the semiconductor-manufacturing domain are available in the Table 27 of the annex.

In the field of machining operations, the major challenge for VM is the drastic decrease in dataset signal-to-noise ratios compared with semiconductor manufacturing. The production environment is noisy with severe vibrations and temperature changes, which makes advanced preprocessing critical. Wavelet denoising, as described previously, is an interesting approach under these conditions [125]. In terms of preprocessing, as for semiconductor operations, machining operations must deal with numerous features, which necessitates efficient feature selection [42].

In terms of quality estimation, 100% of the papers describing VM in machining settings have used MLP as their inference algorithm. Indeed, MLP can deal with noise and leverage its presence to avoid overfitting [126]. Furthermore, the use of MLP can be explained by the high nonlinearity of machining operations. To reduce complexity, prior knowledge can be injected as post-processing. For instance, a deformation model of a hole can be used to enhance the accuracy of VM in a drilling operation [127]. Finally, one application, namely run-to-run control, has already been tested on machining operations successfully [128].

Research on the use of VM in other manufacturing domains besides semiconductors only began in 2017. Since then, 12 papers have been published on this topic, which is more than 10% of the published papers on VM in that time, demonstrating increasing interest. There are tremendous research opportunities on the use of VM in all manufacturing fields. The few results that have been presented can be viewed as proof of concept on the applicability of VM to the discussed manufacturing operations. The next step will be to implement the tested framework elements on different machines in order to validate the prior results and test untested algorithms, approaches and architectures, such as those in the essential and auxiliary groups.

2.5 Concluding remarks

This systematic review performed a comprehensive analysis of the literature on VM and then proposed a VM framework. This latter structure the state of the art. It is composed of the following components: preprocessing, quality estimation, DD, an SDS, the updatability feature, the adaptability feature, a multi-stage architecture, machine control, and a fab-wide architecture. Each element was defined in terms of its roles and importance. The detailed analysis performed revealed the following key findings:

- The use of nonlinear algorithms for dimensionality reduction and quality estimation is now the standard.
- Numerous high-potential approaches have been used for quality estimation. They bring new capabilities such as causality study for BNNs, automatic feature extraction for CNNs, uncertainty estimation, and simple updating for ensemble methods.
- The drift rejection capability of VM is enabled by three essential steps: DD, the SDS, and updating.
- An event-based SDS should not be used because of its sensitivity to hidden contexts. A hybrid SDS should always be used with VM.
- Three families exist for multi-stage architecture: serial, hub, and cascade architectures.
- Machine control based on VM is dominated by run-to-run approaches. Very few research studies have focused on real-time solutions.
- Driven by the product-as-services trend, the latest research on fabrication-wide VM architectures has focused on cloud implementation.
- CMP is the laboratory of VM, featuring all the newest approaches such as CNNs and BNNs. It is also the bridge between semiconductor manufacturing and other manufacturing domains.
- An increasing number of new VM industrial applications have been studied both for semiconductors and other manufacturing domains.

Beyond all of these considerations, VM-related research should be conducted on the different patterns illustrated by our study and by the proposed framework. Numerous proofs of concept of VM applications in new industrial operations such as machining and additive manufacturing, shows great research opportunity for new applications. The benefit of such a technology isn't to prove anymore. Considering the first research question, only one paper has applied VM to milling operation. Unfortunately, few information is provided in term of materials and methods used. It shows great research opportunities. For the second research question, the essential group proposed in the framework, which could also be called data-based maintenance, is shown to be a very important topic in the future decade. VM cannot be sustainable in an industrial environment without a proper drift rejection capability. The research on this topic is scarce, certainly because efforts are focused on the development of simple industry 4.0. solutions, illustrated as the core group in the framework, and not proper industrial solutions. As it is a fundamental aspect, maybe some companies hied this part of the solutions to protect themselves from concurrence. In conclusion, the state of the art showed great opportunity of research for VM application to milling machine. Those opportunities aligned perfectly with the research questions. In the next chapter the two research questions will be answered, thanks to case studies and simulations. In the chapter 3 the focus will be made on the core group implementation. In the chapter 4, the essential group will be discussed.

Chapter 3 Virtual metrology on milling machine

Product quality is crucial for every manufacturing company [16]. Existing solutions for ensuring quality are either expensive, such as systematic measurement, or fallible, such as batch measurement. Therefore, the ability to estimate the quality of a manufactured product has become one of the main objectives of Industry 4.0. Quality estimation has not only paved the way for zero-defect manufacturing [6] but also enabled significant gains in productivity and product quality for some industries, including the semiconductor industry [21]. Moreover, multiple European projects such as InterQ (which will end at the end of 2023) and Qu4lity (which will end in 2022) highlight the industrial interest in quality estimation. Although solutions for quality estimation have already been developed in precursor fields such as chemical engineering with soft sensors or semiconductors with virtual metrology (VM), traditional manufacturing fields such as machining lack such solutions.

In the last five years, VM studies have begun to target newer domains such as machining operations and additive manufacturing. In metal additive manufacturing, the product quality highly depends on the melt pool variation. One study measured such variations using a coaxial camera and other specific sensors [52]. Next, it performed feature selection with a triple-phase orthogonal greedy algorithm to finally feed a feed-forward NN. This enabled the estimation of the surface quality and density of products. Few studies have applied VM to machining operations. In 2016, a research group started to study quality estimation for milling operations [125]. The group noted that the major difficulties of applying such a technology to machine tools were the low signal-to-noise ratio, segmentation problems, and a high number of features. They proposed using wavelet preprocessing to reduce the noise in the dataset. To solve the segmentation problem, they used the G-code in the machining program to trigger the measurement. The researchers performed stepwise selection using a linear algorithm to select the most relevant feature. They measured the vibrations using an accelerometer, and the current in the spindle. They used a NN to precisely estimate the geometrical quality. Based on the presents author's knowledge, the aforementioned study is the only time that VM has been applied to milling operations.

Based on the aforementioned studies, it can be concluded that although quality estimation has been studied in detail for various applications, insufficient attention has been paid to other domains such as machining. To address this research gap which is linked to the 1st research question, two case studies were conducted in an industrial environment in which more than 1100 parts were milled. Process data such as spindle power consumption were measured. Finally, using machine learning tools, the product quality of each part related to the process data was gathered. The goal was to reach the best possible quality estimation accuracy on a milling machine. To do so, two technical aspects were prioritized:

- The data source, which data source to consider?
- The synchronization method, how to synchronize the time series?

Each one of those technical questions where the subject of an independent case study, always keeping in mind the main optimization criteria which was the model accuracy. Efforts were invested into the industrial usability of the case study, which motivates the use of industrial machines on the shopfloor to qualify the quality estimation in the same manner as a coordinate measuring machine also located in the shopfloor. In the first section, the first case study is presented focusing on the data source selection. In the second section, the second case study is presented. It is an enhanced version of the first case study focusing on the time series synchronization. Finally, a discussion concludes this chapter.

3.1 First case study: Data sources

The feature choice is a fundamental aspect of any data driven modelling projects. It is often constrained by economics or technological issues. In this section, the first case study is presented. The main goal being to find the features with the highest performance. A special focus is also made on the quantity of necessary data and the best preprocessing and modelling algorithm. To characterize the link between dimensional quality and process data, parts were milled on a computer numerical control (CNC) machine with different cutting depths while the process data were recorded. Subsequently, the parts were measured. Finally, the data were analyzed thanks to machine learning algorithms to extract existing correlations and being able to calculate the model accuracy. This latter enables, to analyzed the different studied algorithms. First the materials and methods present the case study design. Then the results are presented. Finally, they are discussed to extract conclusion from the case study.

3.1.1 Materials and methods

This sub-section first describes the materials used in the experiments, including the machines, the test parts, and the data sources. Then, the methods are described, including the experimental design and the data analysis procedure.

3.1.1.1 Materials

This sub-sub-section describes all of the materials used during the experiments, including the test parts, machines, and programs.

3.1.1.1.1 Machines

Test parts were milled on a twin-spindle three-axis CNC machine. The main spindle milled the part and the auxiliary spindle cut it with a saw. After sawing, the part was conveyored to a specific tray. The CNC machine was connected to a bar feeder for raw material supply. The main spindle was equipped with a 4-kW synchronous motor. The milling of parts was performed with an end mill 6 mm in diameter.

After being milled, all of the parts were measured in a coordinate measuring machine (CMM) using four touching probes connected in a diamond shape. Due to the evolving environment of the industrial shopfloor, the CMM was calibrated every 20 parts to correct for potential deviations due to temperature and humidity changes as well as other external deviations. Calibration repeatability errors were evaluated as under ±2 µm. Every part was manually deburred to avoid measuring chip stuck on the part.

3.1.1.1.2 Test part

For this study, a test part was developed. It was designed to provide a maximum number of test surfaces with different dimensional tolerances as can be seen in the Figure 9. Each surface was designed to be large enough (5 × 2 mm) to be precisely measured by a CMM once produced, but small enough to maximize the number of them. Moreover, as more than 500 parts were milled, it was more convenient to start from a bar. The face on each side of the test bar had three surfaces with varying dimensional tolerances and a reference surface, equaling a total of 12 studied surfaces and four reference surfaces. Only three studied surfaces were placed on each face to prevent bias caused by bar flexion. Every studied surface differed from the next by 0.005 mm in a decreasing manner. For instance, for 1 specific type of part, the face number twelve is milled 0.5 mm, the number 11 0.055 mm, … Like so, the range of the depth of cut studied goes from 1mm to 0.005 mm. The front face simply featured an engraved ID to ensure a link between each part and the process data. The engraved ID was also used for its foolproof capability, enabling the identification of every studied surface. The back surface was unused. The corner between each side face was not considered to conserve the same surface size for all of the different parts with varying depths of cut. The test parts were made of stainless steel.

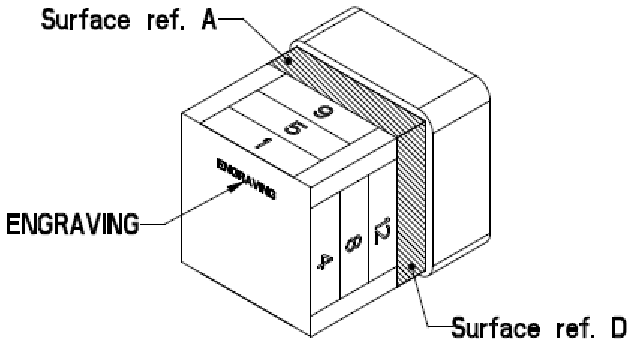


Figure 9: Perspective view of the test part. Note that the numbers are not engraved and are simply present in the image for the sake of comprehension. On the right of the part, the raw bar can be seen.

To machine the part, the CNC program of the main spindle started with a contouring operation where the squircle (a mix between a circle and a square) was transformed into a square. This first operation also defined all four reference surfaces, thereby leveling all of the faces. Then, in another example of the same tool, all of the 12 studied faces were milled in numerical order and in one pass. All of the studied faces were milled with the same movement and position as their reference faces. Then, the part was deburred. Finally, the part was engraved with an engraving tool.

3.1.1.1.3 Data selection

The physical equation linking the process variable to the depth of cut for milling operations is well known in the state of the art (Equation 1). The equation variables are presented in the Table 13.

$$P_c = \frac{a_p a_r f k_c V_c}{60000 \mu D}$$

Equation 1: Cutting power equation.

Table 13: Parameters of the power equation.

P_c	Spindle power	A directly measured feature
a_p	Cutting depth	The target
a_r	Cutting width	Can slightly vary; measured using the x-axis
k_c	Cutting force coefficient	Nonlinear in function of the cutting depth [129]
μ	Efficiency	Source of noise
f	Feed	Constant
V_c	Cutting speed	Constant
D	Tool diameter	Constant

Unfortunately, the equation presented above was not precise enough for the accuracy sought in this study. The cutting force coefficient is the main source of accuracy loss as its exact definition is not well defined in the state of the art. It is found heuristically most of the time. Although this equation may not be used as it is, it offers a hint about which variables should be studied: spindle and x-axis states.

3.1.1.1.4 Data sources

Data were gathered from two independent sources. The first was an external high-precision power sensor connected to the main spindle, which reached sub-watt sensibility. The second source was a sensor that read the numerical control (NC) bus, from which it extracted the power and intensity consumed, torque, measured position, and control tracking error of both the spindle and the x-axis cart. Both sources gathered data at a high frequency of nearly 200 Hz.

To define the dimensional quality, the CMM measured 10 points on all of the surfaces. From those points, the flatness of the reference surfaces was first controlled to ensure unbiased results due to burred parts or touch probe measures altered by dust. Once completed, all of the test surfaces' measured points were projected onto their reference plan, and then the mean of each test surface was calculated to determine the dimensional tolerances.

The data connection from the different data sources was fundamental. Both sources on the machine tool saved timestamps, enabling them to be merged time-wise. The G-code corresponding to the engraved ID was also extracted, enabling the time-series to be linked to the corresponding test part. During the measurement of dimensional quality, the ID of each part was manually inserted into the measurement data file, which ended the solid link between the data sources (Figure 10).

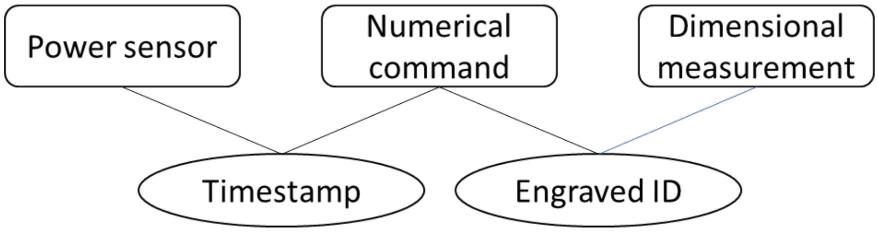


Figure 10: Link between all of the data sources. The blue line represents a link created manually.

3.1.1.2 Methods

This section describes and explains the design of the experiment, data analysis procedure, and metrics used.

3.1.1.2.1 Design of the experiments

To characterize the link between the process and dimensional quality, parts with different cutting depths from 0.005 mm to 1.1 mm were produced, using steps of 0.005 mm, to grid the depth of cut and the power consumption space. To do so, 16 different types of parts with 12 cut depths were manufactured.

Two similar experiments were conducted to create unbiased training, validation, and test sets. In the first experiment, parts of every type were produced 20 times each for a total of 340 parts. The second experiment was conducted the following day with new tools, and there were only 10 identical parts for each part type. In total, 510 parts were produced, creating 6120 quality-process couples. Next, 90% of the first experiment was used to train the algorithm and 10% was used for validation. To ensure generalizability of the results, the entire second experiment was used for the test set.

3.1.1.2.2 Preprocessing

All of the input data were standardized using the training set's mean and standard deviation. Outliers were detected manually through the planarity measurement of all surfaces. Indeed, the only outliers came from dust or burrs that the measurement sensor touched, and they were removed from the testing and training sets.

To synchronize the two data sources, the power consumption of the spindle – which was measured by both of them – was compared as can be seen in Figure 11. The beginning and end of each dataset (represented in the figure by the two crosses) were visually defined on both sources with the same constraints. First, the transient part of the signal was not considered to ensure that no bias was introduced into the study. Indeed, depending on the surfaces' positions, some took slightly longer to mill than others. Therefore, if this zone had been taken into account, the model could have learned this specificity, thus biasing the experiment. The second constraint was the systematic presence of some default on the extreme end of every surface due to material tearing. Therefore, this zone was avoided so that 110 points per surface were retained for the first data source and 300 were retained for the second. Finally, the two sources of inputs were synchronized, which enabled a slight reduction of the overall dimensionality.

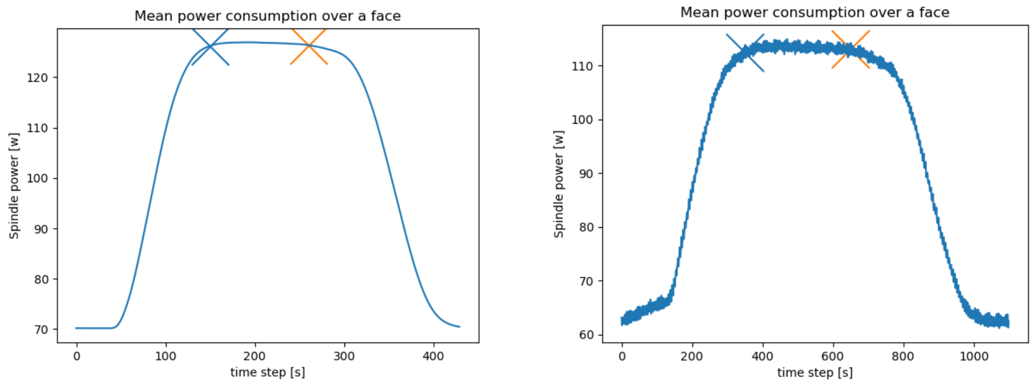


Figure 11: Mean power consumption over a face measured by the external power sensor (left) and the NC bus (right). The blue cross denotes the beginning of the dataset, and the red one cross denotes the end.

3.1.1.2.3 Feature engineering

With 11 inputs regrouping a total of more than 1000 features for estimating the dimensional quality, reducing the dimensionality was necessary. First, the dimensionality was reduced using expert knowledge. The power measured by the external sensor was obviously retained as a feature. However, the selection of the feature from the NC required more attention. All the features are classified in the Table 14.

Table 14: The different features of the second data source

	Axis x	**Axis spindle**
Power	Zero	Already measured
Torque	*Feature*	*Feature*
Current	Equal to Torque	Equal to Power
Control error	Zero	Zero
Position	Biased	Biased

The power in the x-axis and the control error in the x- and spindle axes were fixed and therefore removed. The current always had a linear correlation with the power or the torque over 99%, and thus, it was not retained. The position of the axis was biased for the x- and spindle axes because it carried information that could lead to enhanced performance using the experiment bias and not the studied process. Therefore, they were removed. The power measured by the internal NC sensor was not considered a feature for finding the highest estimation accuracy. However, once the optimal feature set was found, it was compared with the external spindle power sensor to quantify the performance without the external sensor investment. Finally, two features were retained from the second source, ensuring a total size of 300 each when the time-series size was taken into consideration. With a total feature size of 700, the dimensionality still required further reduction. Indeed, the estimation accuracy could have been harmed if there were too many features related to the number of examples. To deal with this problem, two feature extraction algorithms were tested, namely principal component analysis (PCA) and autoencoder (AE).

PCA extracts the feature space based on an orthogonal linear projection of the initial features along the eigenvector explaining the greatest amount of output variance. Dimensionality reduction is conducted by selecting the number of eigenvectors to retain. To select them, the amount of explained output variance is used as a metric and bounded by a heuristic threshold, which was defined as 99% for this study. As presented in the state of the art, PCA is very commonly used for feature extraction for quality estimation

[130]. Its popularity is due to its simplicity of use. AE is a NN based on an unsupervised algorithm, which aims to learn a dimensionally reduced representation of the feature space. To do so, it is composed of an encoder that reduces the dimensionality as well as a decoder that attempts to reconstruct the initial signal from the encoder's output. In doing so, it ensures the preservation of the most relevant aspects of the data. Its main advantage compared with PCA is its nonlinear capabilities. As discussed in Chapter 2, the use of AE started in 2019 with two studies that proved the efficiency of this approach [131]–[133].

To find the most effective combination of features, all seven possible source combinations were generated with and without the different feature extraction algorithms for a total of 28 preprocessed datasets. From those datasets, the lowest mean absolute error (MAE), the most important features, the highest-performing feature extraction algorithm, and the most accurate regression algorithm were determined.

3.1.1.2.4 Data analysis algorithms

Once all 28 datasets had been produced, they were all tested with three types of algorithms: linear algorithms, multi layer perceptrons (MLPs), and convolutional NNs (CNNs). In addition to machine learning good practices, the use of linear algorithms was motivated by the linear Equation 1.

Two linear algorithms were compared, namely multi-linear regression (MLR) and elastic net. MLR is the most basic linear regression algorithm. It is very simple and fast but may be susceptible to instability. It is a very common approach when estimating quality. Elastic net is a regularized version of MLR including both L1 and L2 regularization. The ratio between L1 and L2 is a hyperparameter and was grid searched in this study. For extreme ratios, elastic net becomes Lasso or ridge regression, which are both typical for regression in general. The main motivation for testing elastic net was its regularization capability and its greater stability compared with MLR.

To take the planned nonlinearity into consideration, MLPs were used. As, presented in Chapter 2, they are by far the most used for nonlinear quality estimation in state-of-the-art VM [130]. Finally, CNNs are a supervised family of NNs able to learn and optimally filter when applied to features. A CNN can extract latent variables and reduce the input noise. Often used for image recognition, CNNs have recently been used to perform time-series applications in VM [60], [130], [134]. Noteworthily, a CNN can perform feature extraction if there are a sufficient number of examples to tune it. However, when examples are scarce, such as in the present study, PCA can be applied upstream [135].

All of the parameter algorithms were optimized with the training dataset; the hyperparameters were tuned with the validation dataset; and the final metrics characterizing each algorithm were calculated using the test dataset. For MLPs, to deal with the initial weight variation, every model architecture was built hundreds of times with randomized weights, ultimately retaining the highest-performing model on the validation set.

An agnostic approach, which does not use inputs, was also employed to quantify the information extracted from the different features. To do this, the estimation was always set to the mean value of the output.

3.1.1.2.5 Metrics

The main metric used to represent the estimation error is the MAE. With the benefit of conserving units, it simplified the results analysis. It was used as a basis for tuning the models and selecting the best one.

$$MAE = \frac{1}{n}\sum_{i}^{n}|\widehat{y_i} - y_i\ |$$

Equation 2: Mean absolute error.

The other metric used was the minimum measurable tolerance interval (mTI). Commonly used by every manufacturing company, the ISO norm 22514-7 [136] provides a basis for qualifying new measurement systems. Through simple adaptation to make the norm compatible with virtual measurement systems, the mTI can be calculated using the following equation:

$$mTI = 4 * 6\sigma$$

Equation 3: Minimum measurable tolerance interval.

To do so, residual Gaussianity is assumed, and six standard deviations of the estimation error is considered times a constant that can evolve depending on industries. This constant has been set by industrial experts.

3.1.2 Results

Table 15 present the MAE of the training, validation, and test sets in functions of feature combination, feature extraction, and regression algorithms. As discussed previously, the test set was built to qualify a good extrapolation of the proposed results. This is why all the results are compared in terms of their accuracy on the test set. The agnostic algorithm results are also presented in Table 15.

Table 15: Mean absolute error results in mm.
P: Power, xT: x axis torque, sT: spindle axis torque. Numbers in bold denote the optimal result of the feature based on the test set performance.

		Raw data			PCA data			AE data		
		Train	**Val**	**Test**	**Train**	**Val**	**Test**	**Train**	**Val**	**Test**
P	**Linear**	0.0292	0.0325	0.0283	0.0309	0.0309	0.0295	0.0306	0.0323	0.0295
	MLP	0.0171	0.0171	0.0193	0.0171	0.0167	0.0190	0.0166	0.0171	0.0192
	CNN	**0.0166**	**0.0169**	**0.0189***	-	-	-	0.0196	0.0197	0.0209
xT	**Linear**	0.2267	0.2314	0.2394	0.2295	0.2201	0.2201	0.2137	0.2170	0.2459
	MLP	0.2374	0.2410	0.2651	0.0870	0.1396	0.2354	0.0457	0.1242	0.2188
	CNN	0.1781	0.1899	0.2701	**0.0672**	**0.1156**	**0.2166**	0.0501	0.1424	0.2465
sT	**Linear**	0.0283	0.0325	0.0315	0.0306	0.0325	0.0311	0.0239	0.0231	0.0254
	MLP	0.0104	0.0155	0.0236	0.0110	0.0166	0.0235	0.0132	0.0171	0.0242
	CNN	0.0145	0.0145	0.0325	0.0143	0.0168	0.0300	**0.0135**	**0.0180**	**0.0230**
P, xT	**Linear**	0.0273	0.0311	0.0298	0.0301	0.0304	0.0293	0.0274	0.0283	0.0269
	MLP	0.0106	0.0156	0.0222	0.0132	0.0163	0.0214	**0.0136**	**0.0167**	**0.0206**
	CNN	0.0133	0.0153	0.0207	0.0131	0.0180	0.0253	0.0148	0.0173	0.0230
P, sT	**Linear**	0.0271	0.0326	0.0295	0.0302	0.0320	0.0295	0.0179	0.0180	0.0207
	MLP	0.0119	0.0159	0.0220	**0.0145**	**0.0163**	**0.0206**	0.0121	0.0174	0.0237
	CNN	0.0138	0.0147	0.0269	0.0134	0.0165	0.0284	0.0143	0.0182	0.0228
xT, sT	**Linear**	0.0261	0.0325	0.0329	0.0311	0.0326	0.0318	**0.0213**	**0.0210**	**0.0250**
	MLP	0.0045	0.0153	0.0341	0.0100	0.0175	0.0375	0.0124	0.0170	0.0327
	CNN	0.0129	0.0139	0.0296	0.0103	0.0191	0.0529	0.0144	0.0203	0.0396
P, xT, sT	**Linear**	0.0297	0.0330	0.0348	0.0311	0.0326	0.0318	0.0212	0.0219	0.0250
	MLP	**0.0098**	**0.0158**	**0.0222**	0.0157	0.0167	0.0256	0.0128	0.0170	0.0264
	CNN	0.0108	0.0137	0.0334	0.0120	0.0189	0.0560	0.0145	0.0183	0.0241
Agnostic alg.		**0.2580**	**0.2376**	**0.2626**	-	-	-	-	-	-

For the regression algorithm, nonlinear approaches achieved better MAEs than linear ones. This reinforced the proposition of the researchers in [129], who stated that the specific force coefficient is known to be nonlinearly correlated with the radial cutting depth. For the nonlinear algorithms, both the NN and CNN performed better depending on the case.

In terms of features, the power estimated thanks to the external sensor measurement had the best performance with a MAE of 18 µm. One could remark that adding more features to the power did not enhance the MAE. Looking at the agnostic results, the torque of the x-axis was the only one that did not hold much useful information. With the optimal preprocessing algorithm, the other feature combinations had MAEs approximately 13 times better than the one of the agnostic approach, demonstrating effective information extraction. It was also interesting to compare the power measured by the external sensor with the power measured by the internal one. Unfortunately, the second one held less information and could not achieve an accuracy greater than 25 µm.

Figure 12 presents the CNN estimation of the dimensional quality based on the raw power measurement. Along the abscissa axis are all of the examples of the test set. One could remark that the performance was not homogeneous over the different cutting depths. For cutting depths over 0.78 mm, the MAE was slightly higher. Moreover, it was also higher between 0.43 and 0.47 mm. By contrast, when the error was smaller, the MAE dropped from 18 to 13 µm. As explained before, 16 different types of parts were produced, each with a different dimensional quality, to grid all of the research space. This explains the levels in the figure.

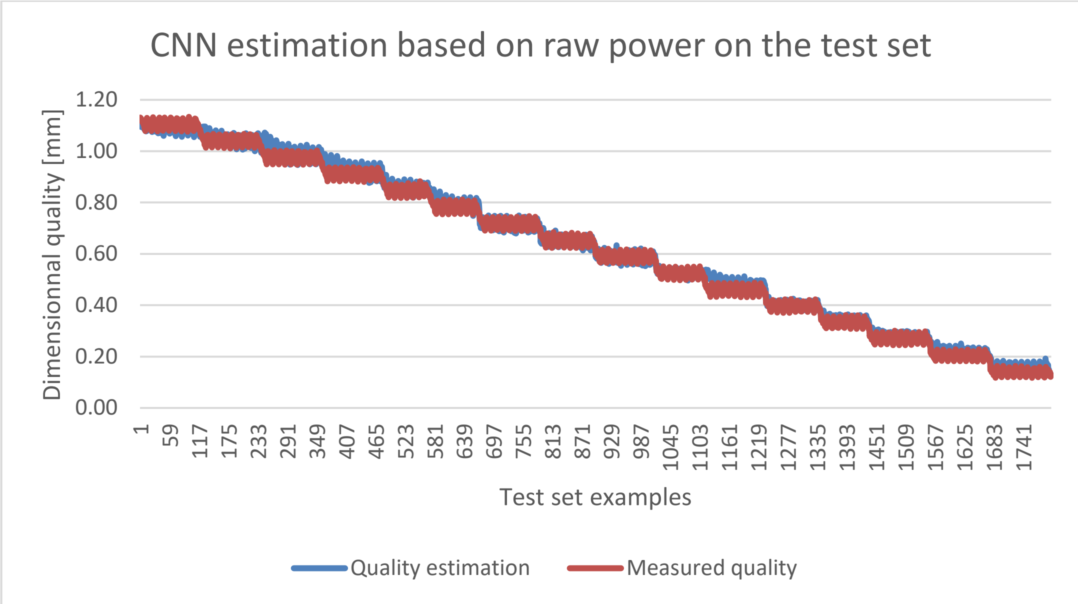


Figure 12: Test set of raw power estimated by the CNN model.

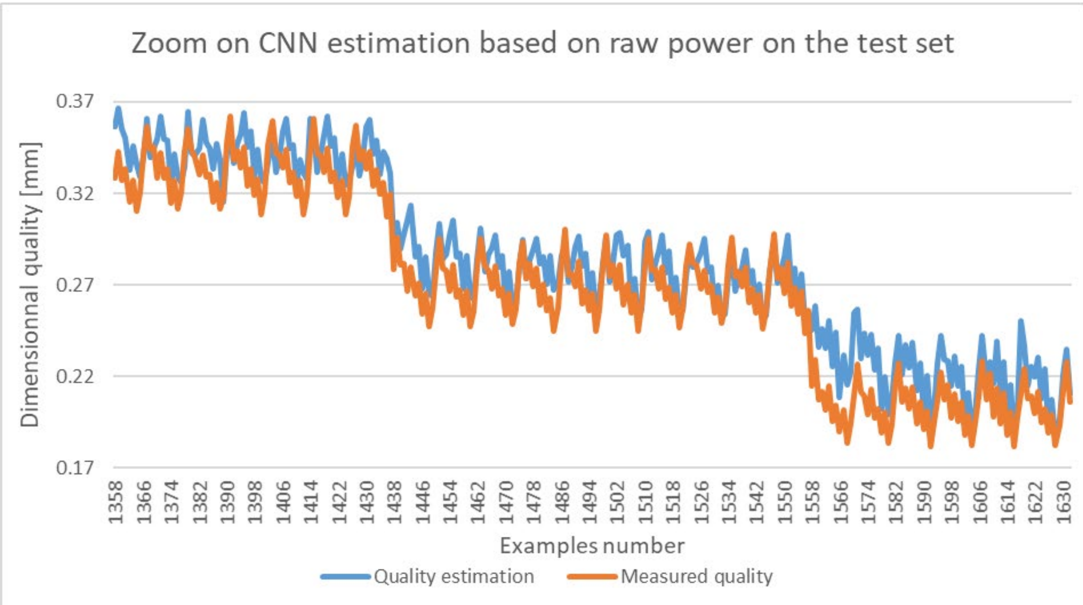


Figure 13: Zoom-in on the test set of raw power estimated by the CNN model.

Also, interesting to note is that good performance on the validation set is not always correlated with good performance on the test set. To characterize the impact of the training set size on model performance, it was reduced from 90% (3044 points) to 1% (34 points) while the test set's MAE was recorded. For every training set size, 20 datasets were sampled to build statistics. The optimal combination of feature, preprocessing, and regression algorithms, namely a CNN with raw power measurement, was used to obtain the test set's MAE. The results are presented in Figure 14. Surprisingly, the MAE increased slowly, providing suitable results with few points. After 34 points, the MAE started to increase exponentially. It was not added on the graph to maintain the scale. Notably, the max-min range increased by stepwise after 1691 and 676 points.

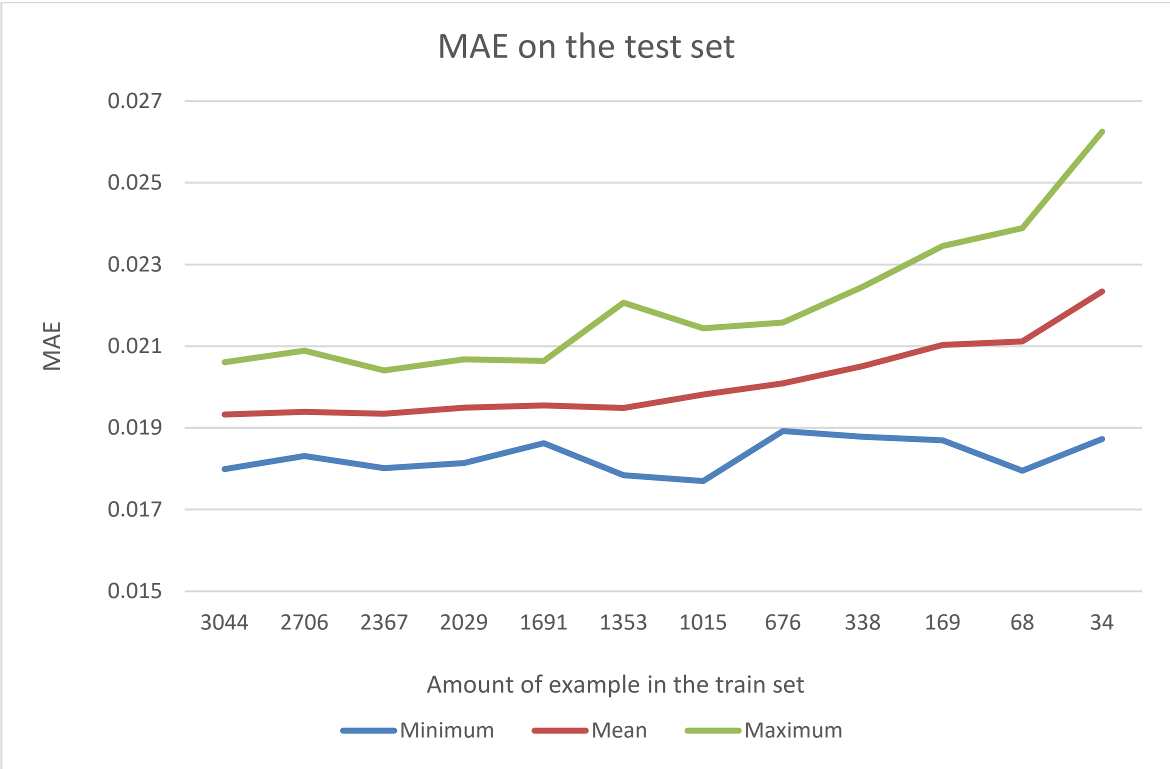


Figure 14: MAE of the test set for different size shuffled training dataset.

The mTI of the optimal model in terms of power consumption was 0.56 mm globally and 0.29 mm when cutting passes smaller than 0.43 mm. For comparison, the optimal accuracy without the external sensor was provided by the spindle torque preprocessed with an autoencoder, which had an overall mTI of 0.68 mm.

3.1.3 Discussion

The external sensor provided the optimal features. This could be explained by both the higher sensibility and precision of this sensor compared with the other ones, and also by its hidden inner filter layer, which could not be removed for practical reasons. Adding other features only increased the MAE, which could be explained by the overly small dataset size. One might extract the information of all of the features together by conducting the same experiment with a higher number of examples. In conclusion, under similar conditions, investment in a high-precision external sensor may decrease the MAE of 4 µm or the mTI of 0.39 mm, which may be required for high-precision industries, such as the automotive or aerospace industries.

The optimal performance was achieved with the CNN using raw power data. Depending on the studied depth of cut, the CNN achieved at best an accuracy of 13 µm with an overall mean of 19 µm. This variance is explained by two zones. First, at a high cutting depth, the dynamic seemed to change, which could be explained by higher vibrations. Second, between 0.43 and 0.47 mm, the performance dropped inexplicably. This drop occurred in the second experiment used for the test set, but it did not in the first experiment. This seemed to be an outlier as all of the points that were more difficult to estimate were produced together. However, as the source of this problem was not recorded during the experiment, the removal of those points could also mask a more serious underlying issue. Further investigations are required to ascertain the source of this complication in order to discard it. Accordingly, the optimal performance was determined to be 13 µm when the cut depth was under 0.43 mm, and 19 µm otherwise.

Following the adaptation of the ISO 22514 norm, the optimal mTI was 0.29 mm. It must be noted that the qualification of the control process is a strict process, which explains the importance of the gap between the MAE of the estimation and the final mTI that can be measured with the model. Using this industrial metric instead of using only machine learning ones ensures the industrial veracity and usability of the proposed results. It must also be highlighted that the CNN took longer to train than the basic NN, and that because the performance of PCA + NN was very close, it should be considered if calculation power limitations exist.

In addition, it was interesting to note that high performance on the validation set was not always correlated with good performance on the test set. As the test set was from the second experiment and the validation set was from the first experiment, this observation can be explained by the first experiment's overfitting. This definitely lowered the accuracy and could have been reduced further if the datasets were larger, and also if the training and validation sets were not sourced from the same experiment. This behavior at least highlights the generalizability of the proposed results.

Under the experiment conditions, Figure 14 highlights that good model performance was obtained with 1691 examples, which was still 6 µm higher than with the full dataset. With fewer data points, the min-max range started to increase. Taking this into consideration, if one wanted to implement the same system with a smaller dataset, this would risk a higher MAE. Good performance with few points highlights that the learned model may have a limited degree of freedom. It must be noted that the actual study was performed using a large range of cutting depths. As the model did not perform uniformly, some differences existed in the dynamics, which may require slightly more points to be dealt with. For instance, if the known variation around the mean of a given product is 40 µm, then the studied range may be defined as 60 µm, which is far smaller than this study's range and may increase the dynamic uniformity as well as decrease the number of points required. As discussed previously, these results do not mean that increasing the dataset size will not significantly enhance the model performance; it simply offers a view on the performance in this operating window of training set size. Finally, even if the results can be used as a baseline, the effective dataset size is often case-dependent and must be optimized on a case-by-case basis depending on the system and required tolerance.

Increasing the number of experiments would be interesting to capture more long-term potential drift as well as to characterize the generalizability and robustness of the proposed model. However, it must be taken into account that quality estimation cannot be entirely generalized with a system as complex as a machine tool. Indeed, the sources of deviation, namely the hidden variables impacting the quality, are numerous and time dependent. Knowing this, the quality estimator must be updated regularly to consider the system evolution [130]. This topic is discussed in the Chapter 4.

The transient parts at the beginning and end of the power signals certainly carry information that was ignored in the experiments because of the experiment design. It has motivated the organization of a second case study with a modified test part enabling the use of the transients. More generally, this second case study was the opportunity to focus into an important aspect of machine tool VM: time series synchronization.

3.2 Second case study: Synchronization methods

The second study focused on the synchronization methods. As it was following the first one, the test part design could have been enhanced. More than 600 parts were milled and measured in the same industrial environment. The machine tool's spindle power was measured. In contrast to the earlier experiment, this time, the full signal was considered, including its transient and steady parts. Even though it is sometimes ignored in VM applications, the transient part of the signal has been shown to carry important information. Some researchers have proposed a new feature extraction algorithm for semiconductors, tailored for 300-mm fabs, which has the particularity of extracting information from both the transient and steady parts of a time series [41]. This succeeded in improving the accuracy of the industrial VM model. Including the full signal, however, raised the new problem of how to synchronize the signals. Indeed, data coming from the shopfloor are prone to synchronization problems. One could note that using a wavelet transform could have helped those authors solve their synchronization problem. This remains a problem for VM applications in milling but also for more developed research domains such as soft sensors. Indeed, it is considered one of the main challenges to soft sensor solutions [137]. There are three state-of-the-art families of synchronization algorithms [137]. Firstly, there are indicator-variable techniques based on a measured or computed variable that is tracked and synchronized across the different time series. As explained above, an earlier study applying VM to a milling machine applied an indicator-variable technique based on the

G-code used in the machine tool's milling program [125]. Secondly, curve registration techniques synchronize some of the features of a times series, such as steady-state values. Some researchers have synchronized bell-shaped curves to forecast a building's electric load. They used a specific curve registration and got far better results with it [138]. Finally, dynamic time warping is a distance similarity metric that can be used to align time series by transforming them. However, it can simply be used for clustering or classification. One research team has developed a specific synchronization algorithm for soft sensors, known as KerDTW. Using the kernel trick and dynamic time warping, they managed to synchronize the batch trajectories of a penicillin fermentation process. They then used a relevant vector machine to do the classification [139]. In addition to these three methods, other approaches are considered outside the current framework of what constitutes the state of the art. The moving window algorithm, for instance, can mitigate synchronization errors while reducing the noise in the data. In soft sensor applications, one study compared different moving windows for synchronization and highlighted their good performance [140]. CNNs could also help to synchronize data. Indeed, although a simple low-pass filter brought by a moving window can prove good results, a CNN can train its own non-linear filters.

In the research described below, an indicator variable technique has been used, over which we compared curve registration, moving window, data quantization, wavelet transform, and CNN algorithms to find the best synchronization approach. This second case study had two objectives, quantifying optimal performance by considering the full-power signal; and comparing several ways to synchronize the spindle-power time series. As for with the first case study, great efforts have been made to ensure the industrial usability of this case study, which is why we used industrial machines on a shopfloor. First the case study is described in depth. Then the results are presented. Finally, they are discussed.

3.2.1 Materials and methods

To fulfil our goal of characterizing the link between the dimensional qualities of the parts milled and process data, we employed a computer numerical control (CNC) to cut different depths and we recorded the process data. The milled parts were subsequently measured. The data were then analyzed to extract any correlations. The subsections below describe the materials, machines, test parts, data sources, methods used, the experimental design, and the data analysis procedure.

3.2.1.1 Materials

The subsections below describe the experiments performed and the materials used in them, including the milling machines, their programs and the test part produced.

3.2.1.1.1 Machines

As in the first case study [141], the test parts were all milled on the same twin-spindle, three-axis CNC, with the same bar feeder. The main spindle was equipped with a 4-kW synchronous motor. The milling cutter tool used was an end mill with a 6-mm diameter.

After being milled, all the parts were also measured in a coordinate-measuring machine. Because of the many factors that can change the environment on a shopfloor (e.g., temperature and humidity deviations, vibrations), we recalibrated the coordinate-measuring machine each time 20 new parts had been milled. We estimated calibration repeatability errors to be under ±2 µm. Every milled part was deburred manually.

3.2.1.1.2 The test part

A stainless-steel test part was developed specifically for this study (Figure 15). It was designed to provide a maximum number of test surfaces, each with different dimensional tolerances, and to decorrelate the depth of cut on each face studied with that of its perpendicular neighbor. This was an issue in our previously published study, as it was possible to infer the dimensional quality simply the time length of each face. As can be seen in the Figure 15 on the left panel, cutting a deeper face A will reduce the time of cut of face D. To deal with this bias, only a small portion of the time series was considered. The present study explored the full signal. The face on each side of the test bar had five surfaces with varying dimensional tolerances and a reference surface, equaling a total of 10 studied surfaces and two reference surfaces. Five study surfaces were added on each side to prevent part flexion. The reference surfaces were enlarged to reduce any lever effects that might increase the measurement error on the more distant surfaces studied. Each subsequent surface studied was milled to differ from the next becoming 0.005 mm shallower in a decreasing manner. For instance, for one specific type of part, its first face was milled 0.5 mm, its second face was milled 0.055 mm, and so on. Thus, the cut depths studied ranged from 0.5 mm to 0.005 mm. An ID number was engraved into each part milled to ensure the link to its process data. This was a failsafe and foolproof method of ensuring that every studied surface could be identified.

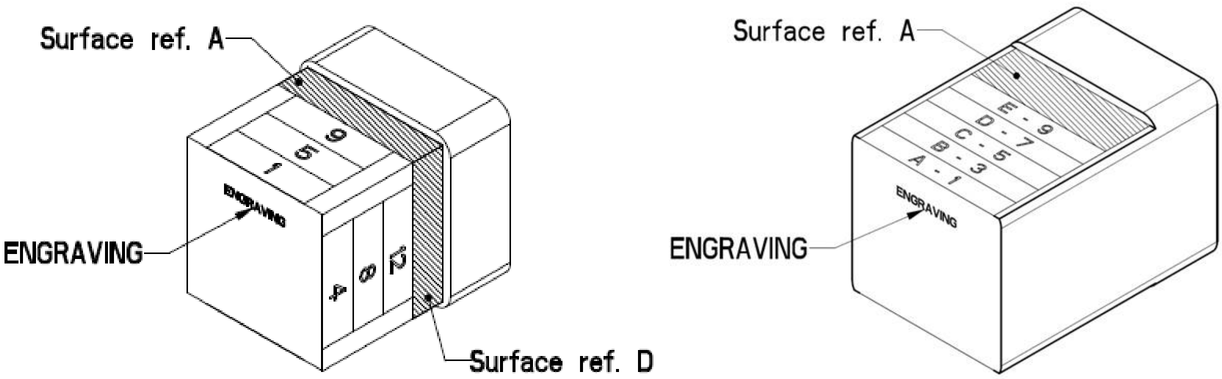


Figure 15: Perspective views of the old test part (left) and the new test part (right). Note that the numbers are not engraved and are simply present in the image for comprehension. Although hidden in this view, the left part has a second reference surface and five other studied surfaces under it.

The first milling operation levelled the studied surfaces and the reference face together. Next, all 10 studied surfaces were milled in one pass. All the studied surfaces were milled using the same machine movement and position as their reference surfaces. Once the part had been manually deburred, it was given an identification number using an engraving tool.

3.2.1.1.3 Data source

This study only examined the power consumed by the machine's spindle. The first case study had shown that power measurement was the most accurate means of estimating a parts dimensional quality [141]. These data were recovered from an external, high-precision power sensor with sub-watt sensitivity connected to the main spindle. Data were measured at a high frequency of nearly 200 Hz.

To define the dimensional quality, the CMM measured 10 points on each surface, studied and reference ones. Using those points, the reference surfaces' flatness was given an initial verification to ensure that burrs or dust would not bias the results or affect probe measurements. The test surfaces' measured points were then projected onto their reference plan; the mean measurements for each test surface were then calculated to establish dimensional tolerances.

3.2.1.2 Methods

This section describes and explains the experiment's design, the data analysis procedure, and the metrics used.

3.2.1.2.1 Experimental design

To characterize the link between the process and dimensional quality, the parts produced were cut to depths ranging from 0.005 mm to 0.5 mm, using steps of 0.005 mm, in order to grid the depth of cut and the power consumption space. In other words, all the research space is visited from 0.5mm to 0.005mm with a step of 0.005mm. To do so, we manufactured 10 different types of parts, each with 10 cut depths.

Contrastingly with the first case study, three similar experiments were conducted to create unbiased training, validation, and test sets. In the training experiment, 25 of each of the 10 types of parts were produced for a total of 250 parts. The validation experiment was conducted on the following day with new cutting tools but with only 20 versions of each part type. In the test experiment, each of the 10 part types were produced 15 times. The total of 600 parts produced created 6000 quality–process couples.

3.2.1.2.2 Preprocessing

As is commonly done for VM, all the input data were standardized using the training set's mean and standard deviation [63]. Outliers were first detected using the planarity measurements for all the surfaces, and these were then removed manually. Our research considered the full time-series generated by each measurement; from the moment the tool touched the part to the moment it was disengaged. Due to sampling and central processing unit priority issues, the time series were not all the same length. Moreover, the position of the curves on the time series varied, inducing features shift. First, the sensors listened to the bus of the numerical command, tracking specific milling program G-code to start or end the measurements. Many external sensors work this way with the numerical command. This method is part of the indicator-variable family of approaches, which certainly reduce the synchronization error but not totally in some cases, including ours [142]. A second synchronization step was needed, for which this study considered multiple solutions.

We began by studying a slope curve registration technique. As all the signals were bell-shaped, they were synchronized using heuristic stacking, either at the beginning or the center of the bell. The left-hand panel of Figure 16 shows the beginning synchronization, which tries to stack the signal from the moment the cutting tool touches the part. It synchronized high-frequency waves. The right-hand panel shows the centering synchronization, which synchronized bell curves themselves by centering them. Secondly, low-pass filters with different cutting frequencies were studied for their capability to average the shifts [140]. Figure 17 illustrates the two cutting frequencies studied. The filter with a higher cutting frequency will reduce the noise and the synchronization error but will reduce the overall information in the signal. They were implemented thanks to the mean of squared moving window. The moving window was applied on the centered, synchronized curves and not the left synchronized dataset as the high-frequency wave would have been damped by the low-pass filter. Wavelet preprocessing has already proven its efficiency in VM for milling operations [125]. Initially used for dealing with noise because it transformed the data in the frequency domain, wavelet preprocessing can also cancel potential data shift. Out of the multiple wave types tested, Gaussian-1 performed best. Also, due to its shift-invariant capability, we used the continuous wavelet transform tool. Finally, numerous papers have applied statistical data quantization to reduce dimensionality but with the risk of deleting useful information. As presented in chapter 2, the typical statistical metrics used are the minimum, maximum, and mean values of a time series [40], [130]. Some research has shown that they can be complemented by other metrics [118], such as rise time [41], [42] or the bell size. The present research used the mean, minimum, maximum, rise time, and bell size to transform the curves into five metrics.

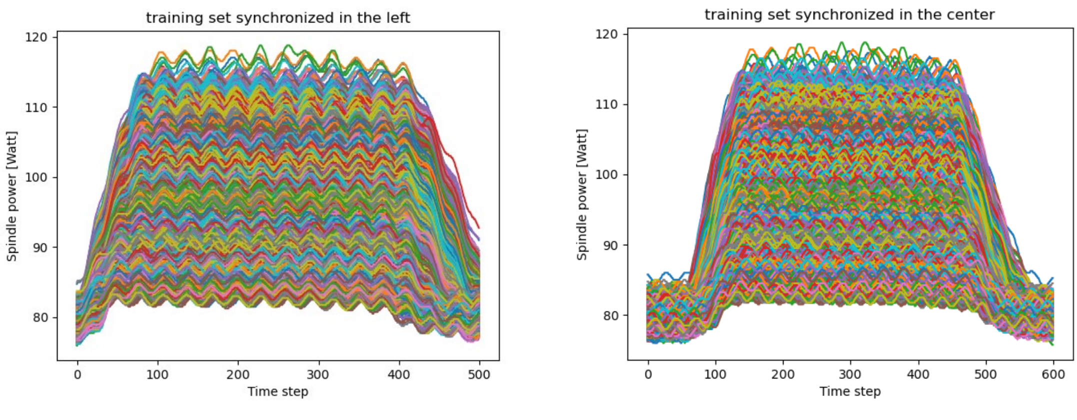


Figure 16: Training set synchronization examples. One is synchronized to the left (left panel) and the other is synchronized to the center (right panel).

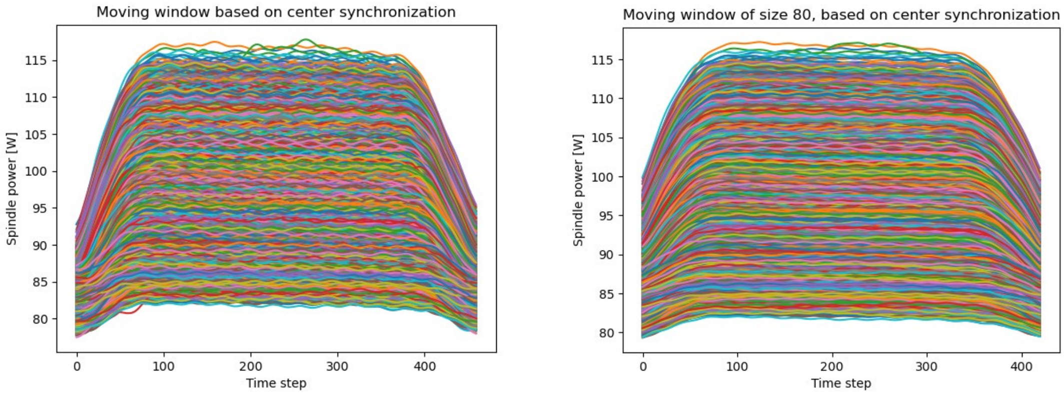


Figure 17: Different moving-window sizes.

To find the most effective preprocessing approach the datasets have been duplicate for each preprocessing options, then all the regression algorithms except the CNN algorithm were tuned independently on all the datasets. Because a CNN learns its own filters, it was only tested on non-synchronized data. Moreover, data augmentation is a common approach for helping the optimization algorithm understand that the signals could be shifted [143]. Every training data point was duplicated and randomly shifted. Finally, the test set's Mean Absolute Error (MAE) of all the judged solutions was then compared to assess the most performant preprocessing and regression algorithm pair on milling machine VM task.

3.2.1.2.3 Data analysis algorithms

Comparably with the first case study, four regression data-based algorithms were compared: multilinear regression (MLR), partial least square (PLS), multilayer perceptrons (MLPs), and convolutional neural networks (CNNs).

More stable than MLR, PLS enables a substantial reduction of dimensionality by transforming the features [144]. The most common linear approach is PLS, and MLPs are by far the most common algorithm used for non-linear quality estimation [130].

In the present research, a classic MLP was optimized thanks to Adam optimizer which is one of the most commonly used optimizers. Weight adjustment and early stop have commonly been used in VM to reduce overfitting [131], [135]. In this study, the MLP studied was designed considering architecture size from 1 hidden layer to 4 hidden layers.

In the present study, a CNN was used with a skip connection and a zero gate, which smoothed the optimization [145]. CNNs can begin by learning a linear representation of the model, which makes sense when we look at the performance of linear model of the first case study [141].

We optimized all the parameter algorithms using the training dataset. The hyperparameters were tuned using the validation dataset and a grid-search methodology for the rough part. Optimization continued manually to refine and extract the best set of the hyperparameters, as often implemented in VM applications [120], and the test dataset was used to calculate the final metrics characterizing each algorithm. Hundreds of architectures for each model were built, with randomized weights, to help the NNs deal with the initial variations in part weight. The best performing model was ultimately retained for the validation set.

An agnostic approach using no inputs was also used to quantify the information obtained from the different features. The necessary estimation for this was always set to the mean output value.

3.2.1.2.4 Metrics

The main metric used to represent the estimation error is the MAE, already presented in the first case study.

3.2.2 Results

Table 14 presents the MAEs of the test sets using different function synchronization methods and regression algorithms. As discussed above, the industrial test set was built to validate a good extrapolation of the proposed results. This is why all the results are compared in terms of their accuracy on the test set. The agnostic algorithm result is also presented.

Table 16: Mean absolute error results of test set in mm.

	Regression data-based algorithms	**No sync**	**Left sync**	**Center sync**	**Moving window 40**	**Moving window 80**	**Wavelet transform**	**Data quantization**
Linear	MLR	0.033275	0.03433	0.03427	0.029849	**0.029151**	0.030710	0.031714
	PLS	0.029098	0.033472	0.033671	0.025376	0.027433	**0.025324**	0.028318
NNs	MLP	0.026171	0.017976	0.02091	**0.017003**	0.018067	0.018780	0.020030
	CNN	**0.014443**						
	CNN + DA	0.018791						
Agnostic		0.114577						

All the algorithms tested outperformed the agnostic approach by a factor of at least three. Non-linear approaches to the regression algorithm resulted in better MAEs than the linear approach, supporting the findings of the first case study [141]. This had, in turn, bolstered support for earlier researchers [129], who had indicated that the specific force coefficient was known to be correlated non-linearly with the radial cutting depth. For linear algorithms, PLS always achieves better results than MLR. With a best MAE of 25.324 µm, PLS gave less accurate results than in the first case study [141]. For non-linear algorithms, CNNs do better than MLPs, giving the overall best MAE of 14.443 µm. The data augmentation did not improve the CNN filters to deal with the shifts. In terms of synchronization, linear approaches performed better when using heavier preprocessing, such as a low-pass filter using an 80-time-step window and wavelet transform. MLP does better with lighter preprocessing, like a low-pass filter using a 40-time-step window. CNN work better without pre-synchronization. As to curve registration methods, left synchronization seemed to be more effective than centered synchronization. However, on its own, the synchronization does not seem to be precise enough, as all of the trained models performed better when using other preprocessing steps. Using data quantization as input ensured high feature reduction but removed too much information.

Figure 18 presents the best CNN estimations of the parts' dimensional quality. The test-set parts' engraved numbers are visible along the horizontal axis. The results were more or less homogeneous at each different cutting depth. Sixteen different types of parts were produced, each with a different dimensional quality, to grid all of the research space, and be able to force the regression algorithm to generalize its model from 0.5mm to 0.005mm. It explains the levels in the figure. A drop can be seen in place of the 9th lot of produced part. Indeed, the acquisition software failed during this particular lot of the experiment. As the tests were done in an industrial environment, access to the milling machine was limited, explaining why this batch could not be produced again.

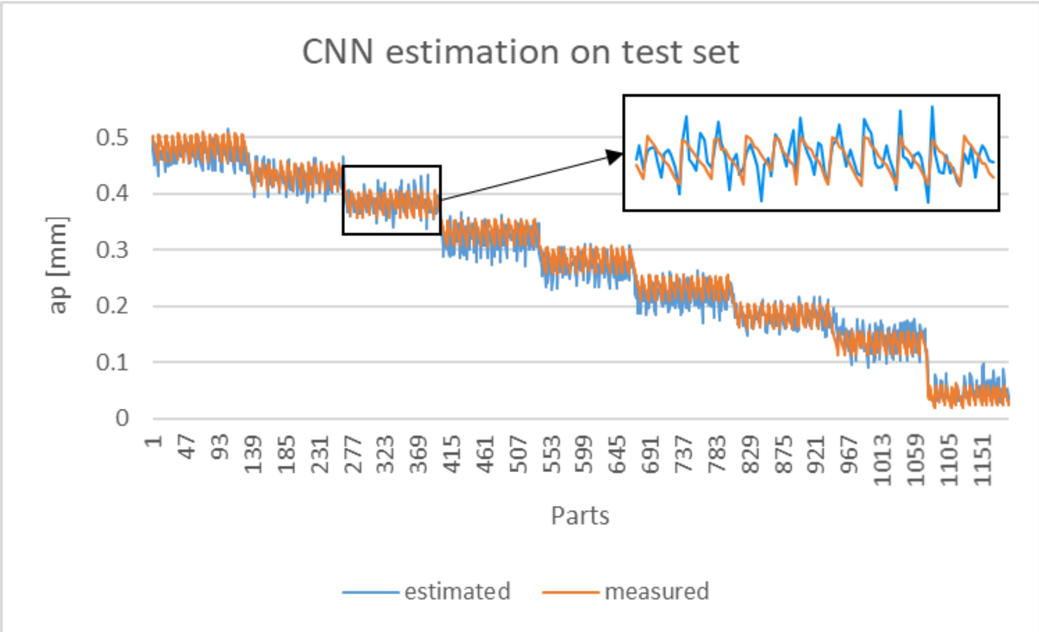


Figure 18: Test set estimated using the CNN model.

The training residual of the best CNN model on the training set is plotted in Figure 19. With a mean of 0.0044 mm, the residual is nearly perfectly centered, showing low bias. The standard deviation of the training set residual was 0.0110 mm, and the training set's MAE was 0.0094 mm. The Shapiro–Wilk test gave a *p*-value of 0.000004, showing that the residuals were not normally distributed.

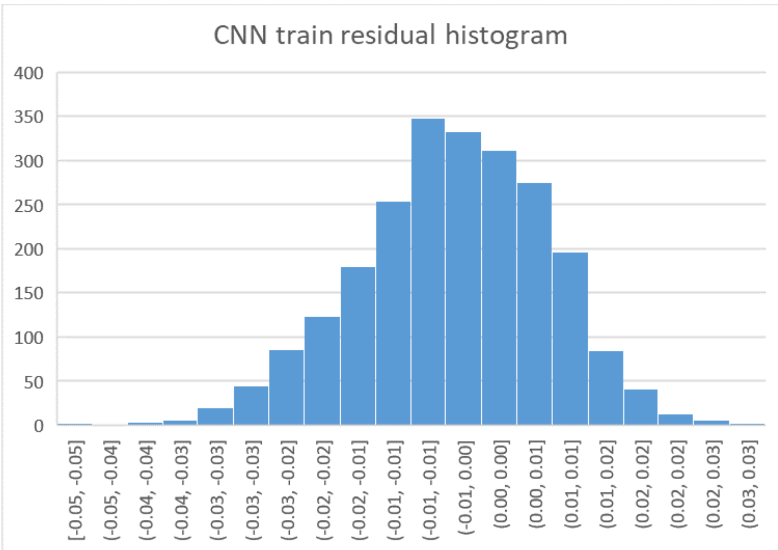


Figure 19: Histogram of training set residuals from the best CNN algorithm.

3.2.3 Discussion

The best dimensional quality estimation results from our experimental milling operations were achieved by a CNN, supporting the first case study conclusion [141]. Analysis of the residuals presented in Figure 19 revealed that the model seems unbiased. The

CNN's preprocessing ability enabled it to fit filters that assumed the roles of data synchronization and noise removal, two aspects which this study highlights. Indeed, most of the approaches did better with a low-pass filter than without. Noise removal and synchronization-error reduction are both advantages of a low-pass filter. The CNN design used a batch norm and a small network of five convolutional-maxpool layers. Even if theoretically applicable, the re-zero used did not actually improve the model's accuracy. This might be explained by the network's small size. Data augmentation was used to indicate the network's synchronization issue. However, the information generated seemed to be less than the noise introduced into the data as it reduced the CNN's accuracy. The MLP was the second-best performing approach tested, and it worked best with the 40-time-step moving window—the 80-time-step moving window seemed to start reducing non-linear information from the data. The MLP was very sensitive to the synchronization issue. Indeed, its MAE without synchronization preprocessing was its worst. In state-of-the-art VM, the MLP is the favorite non-linear estimation algorithm [130] and it should never be used without proper time-series synchronization. Although linear approaches performed better than the agnostic approach—which highlights just how much information they grasped—they underperformed compared to their non-linear relatives, from an important MAE margin of 10 µm. However, their fast prototyping-time and high data efficiency could motivate their use in some very particular cases. Otherwise, our general recommendation would be to use CNN for milling VM regression estimation. PLS and MLR are both remain very common as discussed in the chapter 2 [130]. The advantages of PLS over MLR, in terms of results, can be explained by its feature reduction capability. Moreover, the PLS capability of projecting the data into a new space could also attenuate the synchronization problem. This latter could explain why PLS is one of the approaches that performs better without synchronization than it does with left or center synchronization. In general, comparing left and center synchronization, left synchronization showed better results, highlighting the role of the high-frequency waves in the synchronization task. Our moving window results, however, showed that high frequency small waves did not contain much information in our specific case. Curve registration performed well for milling operation synchronization. The wavelet transform is also an interesting candidate as it is the only approach that has been applied to a similar application [125]. Wavelet transform provides the best linear performance with PLS; however, in this experiment, it does not perform as expected with non-linear algorithms. It must be underlined that wavelet transform, while enabling transient frequency transform, is very non-convex and complicated to tune. It remains an interesting approach to consider if enough computational power and time are available. Finally, our statistical quantization of the data was driven by this approach's use for other applications, as the one presented in chapter 2 [130]. The use of simple statistical means, maximums, and minimums as features did not perform well, which prompted our inclusion of more advanced statistics, inspired by control theory metrics. Although reducing 700 features to just 5 increased data and time efficiency, this significant feature reduction also removed some important information from the data, leading to poor accuracy when using both linear and non-linear approaches. Compared to our first study, the present one highlighted the importance of the information carried by the transient parts of the dataset. In future implementation of VM for milling operation, the full time series signal should always be considered, and the data synchronization and noise removal processes should also be planned.

It was interesting to note the significant divergences between the different datasets, highlighting the complex behavior of real-world industrial production systems. This does not only imply the means of production themselves—composed of a multitude of motors, compressors, and controllers—but also highlights the overall industrial environment, including numerous other production machines in the surroundings, vibrating slabs, and so on. In the training and tuning process, the environment often induced experimental overfitting, which has to be closely monitored. Highlighting the importance of producing dataset coming from separate experiments to avoid overfitting issues. This divergence between the different experiments, even if reducing the overall best accuracy presented, underlined the generalizability of our results. By demonstrating the presence of multiple concept drifts on a single shopfloor, in the era of the Industry 4.0, VM models need to be robust and flexible enough to remain accurate and undergo updating when necessary [130]. Even though this was not the main focus of the present research, its results also reflected this necessity.

3.3 Concluding remarks

The proposed case studies achieved to estimate the dimensional quality of a test part with a descent accuracy. Answering to the first research question, it can be seen as a first proof of concept that VM can be applied to industrial milling operation. Built during the second case study, the best performing model did achieve an MAE of 14.4 µm over the full studied range which is an improvement of 23.8% compared to the first case study. The solution could be applied like it is, if the acceptable measurement error corresponded to the MAE presented correspond to the company tolerances. The proposed solutions used a large, 0.5 mm window of dimensional quality, which gave greater importance on model generalizability than on model pure accuracy. Indeed, if this approach were industrialized, variations in quality would surely be 5–10 times lower, which foresees even better accuracy in the future. One of the main limitations to further decreasing MAE seems to be sensors themself. The error propagation is illustrated in Figure 20 .Using the causality modelization methodology defined in the Chapter 4, it shows the potential sources of error. S_q represents the measurement error of the quality Q. From the CMM requirement and our calibration, S_q has a repeatability of ±2 µm. S_0 regroups the effects of all the latent variables present in the equation linking quality and spindle power consumption. The first case study showed that the inbuilt sensors in machine tools are often not precise enough to estimate dimensional quality [141]. External data sources might have a positive impact on MAE and reduce S_0. Finally, S_p represents the power sensor measurement error. It is difficult to quantify the impact of S_p and S_0 but we know that if $S_Q = \pm 2\mu\text{m}$ then µm $S_0 + S_p \approx 10\ \mu\text{m}$, assuming that the model extracted all the possible information. One could also point out that S_0 had a major impact compared to S_p. The Equation 1 give a rough idea about the sensor resolution needed. In the proposed case study configuration, a depth of cut of 1 μm will induce a change in the spindle power of roughly 1 watt. The external power sensor, with a spindle of 32Kw, have a resolution of 0,5 watt which is way enough considering the usage of a spindle of 4Kw. Similarly, with a precision of 0.4watt, the error contribution of S_p can safely be capped at $1\mu\text{m}$. Moreover, the difference between the successive experiment, of both case studies, are important, highlighting the impact of S_0. Most of the effort should be driven on the reduction of S_0, by measuring, as it has been made with spindle power, other variables with high precision.

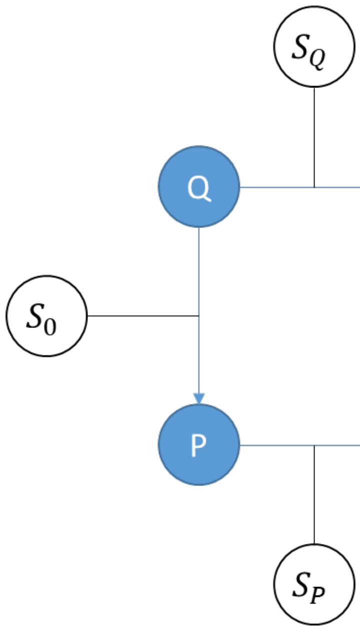


Figure 20: Directed graph of the VM problem, showing the different sources of error. Q is the dimensional quality. Q_m is the measured dimensional quality. P is the spindle power usage. P_m is the measured spindle power usage. S are source of error due to latent variables.

Finally, how NNs were initially weighted was an important factor in convergence, forcing all the hyperparameter groups to be trained multiple times before finding the best one. Highlighting the nonconvexity of the problem, it may come from a lack of data. More data may increase the final model's accuracy and tuning time. In any case, CNN was the most performant approach in both case study and should be considered if there is enough data available. The difference between the first case study, with a best MAE of 19 µm, and the second one with a MAE the 14.4 µm, is explained by the synchronization approach. It highlights the quantity of information contained in the transient part of the dataset for dimensional quality estimation. The transient should always be considered. The new metric derived from the ISO norm, proposed in the first case study, would need to be investigated further. It is clear, that such a norm will have to be specified in the future, the proposed solution is a first draft which will need to be refined.

The implementation of an industrial VM solution on machine tool, seems to be limited by two main challenges. Firstly, the model accuracy, which is industry dependent, and related to first research question. Multiple leads have been proposed to enhance it.

Secondly, the model long term reliability, which depends on the studied physical environment. Indeed, the data-based model learn a reality depicted by the data. If the data doesn't represent all the possible configurations, with enough time, the model will fail. The number of possible configurations depends from one case study to another. The best solution to this challenge is based on model maintenance. This latter is the subject of the second research question and is presented in the following chapter.

Chapter 4 Industrial data-based models' maintenance

In an industrial environment, models estimate quantities that are usually expensive or difficult to measure. Defining when to measure those quantities and how to use the obtained data to maintain the model is the topic of this chapter. Without these aspects of Industry 4.0, the implemented model is destined to fail, which could be very harmful considering the critical aspect of industry it supports. The estimation model must adapt to its environment. In this context, the use of model-based approaches is detrimental, as all sources of concept drift (CD) must be known and priorly modelled, which is not realistic in practice. Data-based models can use the data to update themselves, which is the focus of this thesis. A methodology for implementing a data-based model maintenance solution is required but currently missing from the literature.

This chapter answers the second research question by proposing a new methodology to allow for designing and implementing the most suitable solution for addressing CD in a defined environment. As it is easily generalizable, the methodology is designed for generic Industry 4.0 applications. It could be applied for virtual metrology (VM) without specific modifications. The characteristics of the involved CDs are used to discuss the different solutions through the various steps of the methodology. An optimization problem is designed to select the most suitable solution from among the various possible algorithms. Furthermore, a new causal approach for CD type acquisition is proposed. It originates from the adoption of causal representation, which results in a new type of CD called upstream CD. In the following sections, a literature review is first presented. While a general literature review was already presented in Chapter 2, this one is focused less on VM and more on general methodology for maintenance of data-based model implementation. Thereafter, the CD characteristics are defined to match the new CD vision proposed in this chapter. Subsequently, the methodology is presented. Next, practical implementations methodology are performed on a simulation to illustrate the previously discussed theory. Finally, concluding remarks regarding the findings and future perspectives are presented.

4.1 Literature review

The literature reveals proposed solutions that are designed to address CD issues. CD handling has been studied from diverse perspectives in numerous fields, and some examples are presented in this section. This will provide an overview of similar research and prepare for the proposed methodology to be introduced.

First, CD handling frameworks have been proposed in various fields. A framework is proposed in the novelty detection field [146] that tackles many questions related to the different mechanisms for handling CD. The framework divides the process into two steps: the offline phase when the model is designed and the online phase when the model is running. For the online phase, questions about the use of external feedback or forgetting mechanisms are addressed. Other topics, such as the treatment of outliers or recurring contexts, are also considered when proposing an adequate solution. In the virtual metrology field, a framework is proposed with a direct industrial approach [147]. Virtual metrology involves estimating the quality of a product using production process data to avoid costly physical measurements. The framework is composed of numerous elements including, among others, data pre-processing, the sampling decision system sampling decision system (SDS), model updating, as well as the model connection to the manufacturing execution system. Each step is tackled from a practical and industrial point of view. Updating the system and SDS are discussed in detail.

Second, numerous specific solutions for drift handling have been implemented in several fields. For instance, in the domain of active learning for data stream, one paper [148] proposes a solution that includes the measuring cost constraint in the solution. This is an attractive solution, given its industry-oriented approach. If the allocated budget is exceeded while a sampling is required, the measurement is not performed and there is no update. Different sampling strategies are presented such as random strategy or variable uncertainty strategy; however, the paper does not focus on the model update. Another paper proposes a unique, complete solution in the field of semi-supervised learning called SAND [149]; it is based on a semi-supervised adaptive novel class detection and classification over Data Stream. This solution uses an ensemble classifier composed of k-NN type models to classify the new incoming data. Outlier detection is applied to each new instance to identify the emergence of a novel class using novelty detection technique. A change detection technique is applied to the classifier confidence estimates to actively request samples for updating the classifiers. This solution makes it possible to reduce the measurements while their SDS is based on the classifier confidence estimates.

As described, frameworks do exist to structure the CD-handling solutions as well as solutions that have been implemented on specific applications. However, no methodology has been proposed to implement them. The CDs are never identified and characterized to support solution optimization. In the present research, a context-oriented methodology is proposed based on the link between the solution performance and the drift characteristics, thus enabling solutions to be more robust and generalizable. Many studies have stressed the need for generalized solutions that address the CD [150]–[152]; consequently, the proposed methodology considers the full CD handling framework from the SDS to the updating system (US) as well as the industrial cost and constraint to select the optimal solution.

In the following section, First a concept drift characteristics definition is proposed. Then, the methodology to implement the maintenance of a data-based model solution is described. Moreover, a simulation that illustrates the proposed methodology is presented. Following this, a general discussion on the importance of such a methodology for the industry is presented. Limitations of the research are also discussed, and some further directions for research are proposed.

4.2 Concept drift characteristics

CDs are the cause of drops in model performance. To handle them, they must be clearly defined. This research proposes a new vision of CD. A data-based model estimates $\hat{P}(Y|X_m)_t$ in an attempt to model the physical probability $P(Y|X_m)_t$, which is a local time-based probability of the physical function $Y = f(X_m, X_l)$, where X_m represents the measured features, X_l is the latent features, and Y is the output. The state of the art defines a CD as a change in local physical probability $P(Y|X_m)_t$ in time t such that $P(Y|X)_t \neq P(Y|X)_{t+1}$ [153]–[155]. This vision focuses on the physical local view and does not consider the data-based model. In this thesis, the focus is on model performance. Therefore, a CD is now defined as a significant drop in accuracy of the estimated model. This concept does not refer to the physical but rather the estimated one. The difference in definition enables a focus on the application of a CD-handling methodology, thus ensuring the preservation of model accuracy. This new vision has some differences in terms of CD characterization, which are presented in the following section. CDs can be classified thanks to three main independent degree of freedoms. Their types, recurrency and their geometry.

4.2.1 Concept drift types

The type of CD describes the visibility and relevance of a CD. In the literature, there are three recuring types of CD, with their names differing from paper to paper. They are presented as follows [153]–[155]:

- Hidden CD: The physical probability changes, but none of the features exhibit a change in their distribution. $P(Y|X)_t \neq P(Y|X)_{t+1}$ while $P(X)_t = P(X)_{t+1}$.
- Real CD: The physical probability changes, and a change in feature distribution is observed over time. $P(Y|X)_t \neq P(Y|X)_{t+1}$ and $P(X)_t \neq P(X)_{t+1}$.
- Virtual CD: The physical probability does not change, whereas a change occurs in the feature distribution. $P(Y|X)_t = P(Y|X)_{t+1}$ and $P(X)_t \neq P(X)_{t+1}$.

As the proposed CD definition focuses on model health, all of the CD types that decrease model accuracy are considered, including real and virtual ones. This motivated the development of a new CD definition considering two new degrees of freedom: CD visibility and the causal position of the CD. The first degree of freedom can be classified into two categories:

- Hidden CD (as defined above): The learned model becomes obsolete, but none of the features exhibit a change in their distribution.
- Visible CD: A change occurs in feature distribution, which is observed over time.

The second degree of freedom originates from the adoption of a deterministic point of view, which raises the question of causality between variables. It assumes a causal direction between two variables: X can cause Y, or Y can cause X. Therefore, a causality flow might be imagined in which the variables of interest composing the model are represented. Some variables are thus situated upstream and considered the main root causes of the target variables. Other variables are situated downstream and are considered the main effects of the target variable. Every CD that occurs in downstream variables will not have any effect on the model due to the direction of causality. However, CD in upstream variables could affect the model accuracy and thus be the source of a CD. Finally, variables situated inside the model in the causality flow naturally affect its accuracy when drifting. The causal point of view leads to the emergence of two new types of CD, namely upstream CD and inside CD, which are defined as follows:

- Upstream CD: This refers to CDs that are applied above both X and Y variables in the causality flow. The relationship between X and Y is not affected. The change in accuracy stems from the limitations of the trained model. Indeed, if an upstream CD affects model inputs and forces them out of the learned range, as defined by the training set, then the model will enter its extrapolation zone. The learned model becomes obsolete in terms of its extrapolation performance. Therefore, a difference from inside CD is that its amplitude defines its impact on not only model accuracy but also the model's extrapolation capability.
- Inside CD: This refers to CDs that are applied between Y and X variables in the causality flow. Such a CD will modify the relationship between X and Y, and the model accuracy will thus decrease.

By definition, an upstream CD is visible, and inside CDs could be visible or hidden depending on the causality direction of the variables of interest. Hidden inside CDs are simply called hidden CDs. Visible upstream CDs are called upstream CDs. The last CD type is visible inside CDs.

4.2.2 Concept drift recurrence

In addition to the visibility and relevance of a CD, its recurrence provides information about its criticality. This is an important aspect, as it strongly affects the update strategy and passive time frequency. Indeed, the spectrum of the CD occurrence frequency and the amplitude should form the bases for effectively designing a handling solution.

In the literature, the recurrence of a CD is often associated with a geometric qualitative description [153] that does not define the frequency at which the CD recurs. Therefore, further parameters are desirable. One paper discussed different aspects of CD recurrence [156]. It introduced cyclical CDs, where different concepts recur in a specific order. For instance, the four seasons could be defined as different concepts inside the cycle, with an incremental CD occurring between each season. The periodicity of the CD recurrence is another aspect composed of various dimensions, such as the frequency and duration of the concepts or the CDs in a cycle, which can be variable or fixed. The starting time of concepts and CDs could also be a way to characterize cyclical CDs.

4.2.3 Concept drift geometry

Geometric properties are characteristics that provide information about the shape of a CD when viewing a data distribution/time graph. Qualitative and quantitative descriptions exist, which differ in terms of their objectivity: The first is a subjective description that allows one to classify the different CDs under a limited number of categories, whereas the second offers an objective vision of the different CDs but allows an infinite number of different CDs. Qualitative and quantitative descriptions of CDs are detailed in the following two subsections.

4.2.3.1 Qualitative descriptions

In the literature, CDs are mainly described qualitatively. They are classified under four different categories: sudden/abrupt, gradual, incremental, and recurring CDs. The different qualitative descriptions are presented in Figure 21. A sudden CD describes an abrupt change in a concept where the transition between the two consecutive concepts is almost instantaneous and a clear distinction exists between both concepts. A gradual CD describes a smoother and more discrete transition where both concepts are observed. The current concept is gradually introduced, while the older one is gradually discarded. An incremental CD represents a continuous transition from one concept to a new one without returning to a previous state. Finally, a recurrent CD is more about the concept itself than the CD. Indeed, there is no description of the CD but only the idea of the recurrence of an old concept. Note that since recurrence was discussed in the previous section, it is not elaborated further here.

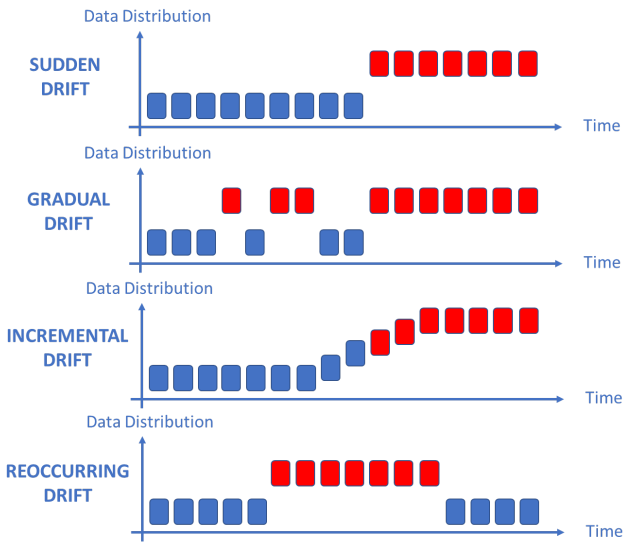


Figure 21: Qualitative descriptions of concept drift categories [27].

4.2.3.2 Quantitative descriptions

In the literature, some quantitative descriptions that can describe a CD more precisely. These descriptions are less common in the state of the art. Webb et al. presented four quantitative degrees of freedom for describing and differentiating a CD [156], which are presented below:

- Drift magnitude: This indicates the magnitude of difference in the concept before and after the CD. It is defined as follows:

$$Magnitude_{t,u} = D(t,u)$$

Equation 4: Drift magnitude definition.

where *D* is a distance function and *t* and *u* are two time steps.

One paper used the energy distance as a distance function to characterize the drift degree, which can be seen as magnitude [157]. The formula is provided as follows:

$$Energy\ distance = d(X, Y) = \frac{(2A - B - C)}{2A}$$

Equation 5: Energy distance used to characterize drift magnitude.

where $A = \frac{1}{mn}\sum_{i=1}^{n}\sum_{j=1}^{m}\|x_i - y_j\|$ is the average of the pairwise distance between two groups of samples, and $B = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\|x_i - x_j\|$ and $C = \frac{1}{m^2}\sum_{i=1}^{m}\sum_{j=1}^{m}\|y_i - y_j\|$ are the averages of the pairwise distance within X and Y, respectively.

- Drift duration: This indicates the duration of a CD starting at time *t* and ending at time *u*:

$$Duration_{t,u} = u - t$$

Equation 6: Drift duration definition.

- Path length: This is another way to differentiate CDs. It attributes importance to all steps that the concept has been through between times *t* and *u*. It could be defined as an indicator of the shape of the CD. The magnitude function is used to compute the path length of a CD between times *t* and *u*:

$$PathLen_{t,u} = \lim_{n \to +\infty} \sum_{k=0}^{n-1} D(t + \frac{k}{n}(u - t), \qquad t + \frac{k+1}{n}(u - t))$$

Equation 7: Path length definition.

- Drift rate: This quantifies how fast the concept is changing at time *t*:

$$Rate_t = \lim_{n \to \infty} nD\left(t - \frac{0.5}{n}, \qquad t + \frac{0.5}{n}\right)$$

Equation 8: Drift rate definition.

The quantitative descriptions are not as developed as the qualitative ones in the literature. Its development would be a relevant topic for further research as it would enable more accuracy drift characterization which may lead to better CD handling solution selection. Due to their extensive use in research, more solutions are based on methods using only qualitative descriptions of CDs. Thus, such descriptions were used for the solution design in the present work.

4.3 Methodology for maintenance of a data-based model

This paper proposes a solution to ensure the maintenance of an industrial data-based model. The operational framework to consider for the proposed solution is presented in Figure 22, which gives an overview of the different involved systems.

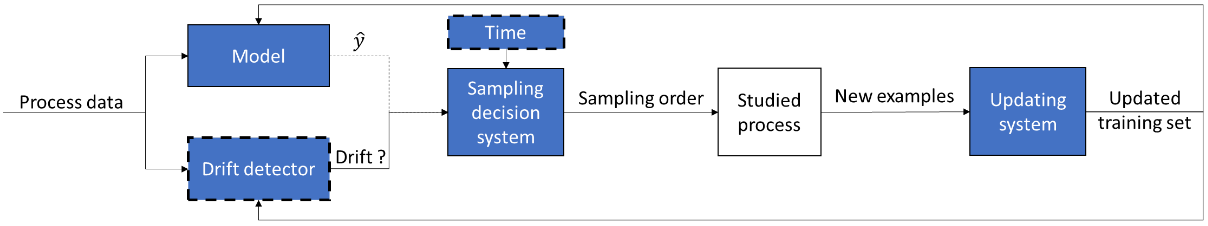


Figure 22: Operational framework

Process data, which can be machine parameters or sensor data, are fed as inputs of the framework. They are used as inputs in the estimation model as well as in the CD detector [158], [159]. The model returns the estimation of an unavailable physical variable and can be written as follows:

$$Y = f(X)$$

where X is the measured inputs variables, Y the targeted output variable, and f the estimated physical model.

The CD detector returns an alarm if there is a CD in the data stream. The outputs of the two previous blocks as well as time can be used in the SDS. The SDS defines when to measure new samples [160]. It triggers the sampling. Then the studied process which does the measuring. When new examples are available, they are used in the US to build an updated training set. This training set is given to the model as well as to the CD detector to keep them updated. The CD detector and time are dash-line blocks because they might not be used, depending on the chosen sampling strategy. Indeed, the SDS will need time if a passive strategy is chosen, and it could need a CD detector for an active strategy. These strategies will be discussed later in detail. There is a dash-line arrow between the model and SDS because this relationship is application dependent. For instance, the model estimate in virtual metrology is taken into account by the SDS.

The model maintenance solution regroups the SDS and US. There are many possible combinations and selecting the optimal one can be complex. To help with this task, a methodology is proposed for industrial implementation. The choice of the estimation model is independent of the methodology, which might or might not already be implemented. Its influence on the solution selection is not discussed in this paper. The proposed methodology is presented in Figure 23.

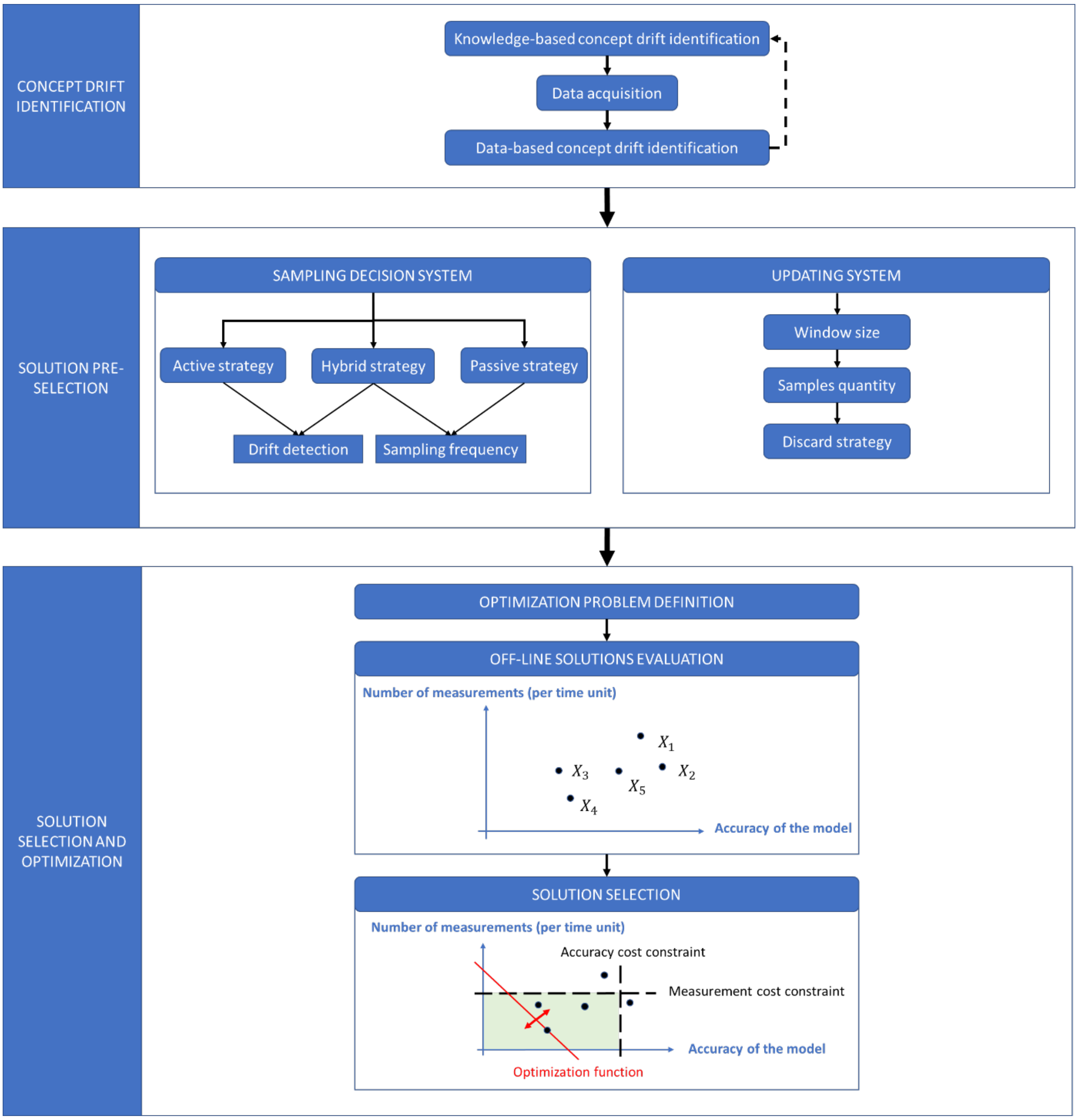


Figure 23: Methodology.

The different steps of the methodology are briefly presented are follows:

1. **Concept drift identification:** Identifies the characteristics of the involved CD. Those characteristics can afterwards be leveraged to guide the selection of a CD-handling solution. This is subdivided in three steps:
    - **Knowledge-based concept drift identification:** Uses the domain experts' prior knowledge to identify the CD characteristics and CD sources. This makes it possible to extract relevant information for the data acquisition, such as the required measurement sampling frequency or duration.
    - **Data acquisition:** Performs measurements used in the data-based CD identification and in the off-line solution evaluation.
    - **Data-based concept drift identification:** Extracts information from the measured data to identify characteristics of the involved CDs such as the recurrence or geometrical properties. This makes it possible to complete the CD description to enable the CD handling solution pre-selection.

These three steps can be repeated (dash line on Figure 2) until the obtained information is sufficient to move to the second stage of the methodology.

2. **Solution pre-selection:** Perform a reduction of the solutions space, by keeping only the solution that is known to be performant on the identified CD. The solutions are composed of a SDS as well as a US.
   - **Sampling decision system (SDS)**: Defines when to measure new samples. A good solution would make it possible to optimize the measurements cost by maintaining good model accuracy. **Active**, **hybrid** and **passive** strategies are the three different alternatives for SDS, as defined below:
     - **Passive strategy** is a triggering measurement based on a predefined sampling frequency.
     - **Active strategy** is based on CD detection to trigger a measurement when a CD is detected.
     - **Hybrid strategy** is the combination of active and passive strategies.
   - **Updating system (US):** Defines how to update the model when the new training data are available. This affects the model adaptation speed as well as the model performance described by the stability plasticity dilemma. Three criteria must be carefully tuned: the size of the moving window, the number of new examples to use, and the strategy to discard the less informative samples from the current update training set.

   The links relating the CD characteristics to the possible solutions indicate how those systems are pre-selected. Then the most performant solution needs to be selected.

3. **Solution selection and optimization:** Evaluates the pre-selected solutions to select the best one using an optimization problem.
   - **Optimization problem definition:** Defines the optimization problem and its parameters based on both the measurement and model accuracy constraints and cost.
   - **Off-line solutions evaluation:** Evaluates the pre-selected solutions. Each solution is implemented on a testing dataset to compute their relative accuracy and the number of measurements needed to reach it. This makes it possible to map the solution space on an optimization graph (black dots on the graph).
   - **Solution selection:** Selects the optimal solution that minimize the objective function (red slope on the graph) and satisfies the constraints (black dash line).

The different elements of the proposed methodology are discussed in more detail in the following sub-sections.

4.3.1 Concept drift identification

To be able to design the optimal model maintenance strategy, the CD characteristics must be known. CD identification is therefore a crucial and mandatory step of the proposed methodology. The importance of a thorough understanding of CD is at the core of the proposed methodology for implementing the adapted CD handling approach. CD identification can be separated into three different steps: knowledge-based CD identification, data acquisition, and data-based CD identification. For each of them, the different identifiable CD characteristics and related methods are presented in the following subsections, along with their relevance to the methodology.

4.3.1.1 Knowledge-based concept drift identification

Knowledge of the studied environment makes it possible to estimate intuitively the sources as well as the properties of potential CDs. Experts in the application field are required to identify as much information as possible regarding the involved CDs. The sources of the CD should first be identified; following this, the characteristics of the CD can be studied.

Trying to connect the different sources of CDs makes it possible to establish the causal relationships between variables. Accordingly, it is possible to identify the type of CD related to each source. A new tool has been designed with this purpose in mind. It is divided into different steps. First, a causal graph must be built based on generalizable rules. Then, the information about CD type must be extracted from the graph. To do so, instructions must be established to identify the type of CDs involved. Finally, the graph can also be used to select the most relevant features to measure.

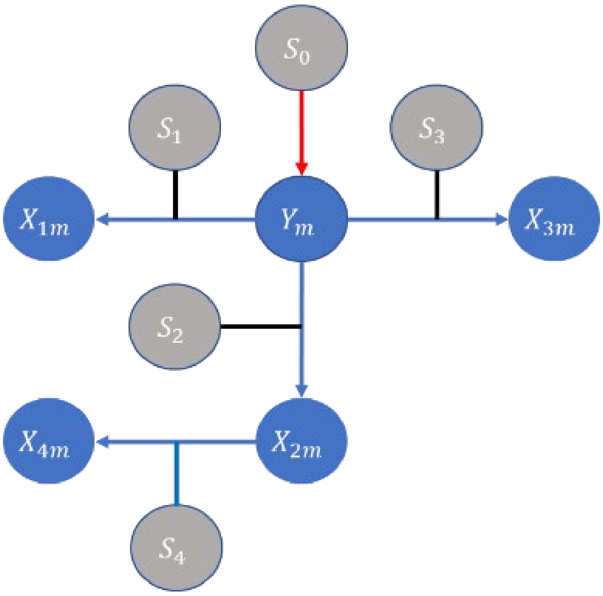


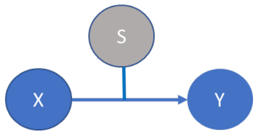
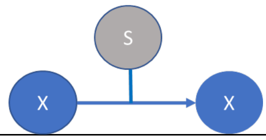
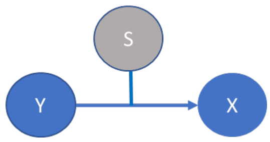
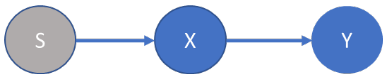
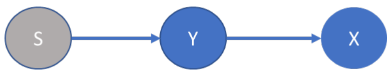
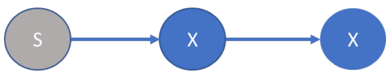
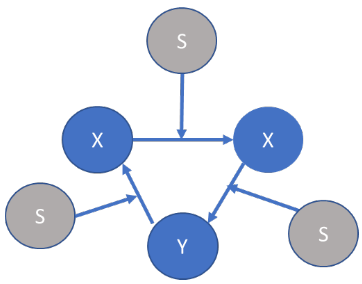
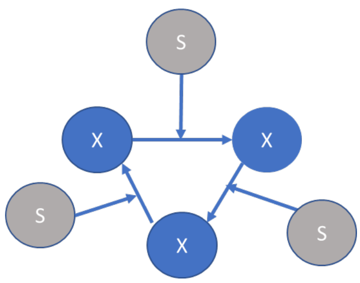
Figure 24: Example of causal representation.

The causal representation of the model is mapped in an oriented graph based on the system physics. Such a process requires knowledge of the involved variables and their causality. A general example where Y causes X is presented in Figure 24. The different components are described and the general rules for building such a graph are presented as follows:

- A node represents a variable (latent or observable). Blue circles are variables of interest (the ones involved in the learned model). X is the measured variables and Y is the estimated one. Grey circles represent sources of CD, which can come from multiple different latent variables.
- An arrow represents a function with the causal direction. Every variable (node) linked to an arrow is considered a variable of this function.
- Normal lines join latent variables/sources of CDs to an arrow, representing a possible CD. The colors of those lines correspond to the type of CD. In this example, red is upstream and black are VI CDs.
- A node is reached by only one arrow but can join different arrows. (A variable is defined by one function but can be part of different functions.)
- Arrows start from measure variables, except when a measure variable is at the top of the causality. In this special case, a source of CD must be added as the cause, and then the arrow starts from it.

Once the variables involved in the learned model (Xi, Y) are determined, using causal representation allows one to identify the type of CD involved. The basic patterns that can be found when building a causal graph are presented in the following table. X corresponds to a measured variable that is used in the estimated model; Y corresponds to a variable that is estimated using the model; and S represents latent variables and thus sources of CD. To quickly identify the type of CD involved in a problem, Table 17 is proposed.

Table 17: Directed graph dictionary.

N°	Combinations	Type of CD
1		Hidden
2		Visible inside
3		Visible inside
4		Upstream
5		Upstream
6		Visible (inside or upstream)
7		Visible
8		Visible (inside or upstream)

However, depending on the complexity of the physical model, some of the patterns in the table can be superposed or added. Depending on the case, the pattern definitions above cannot be followed. The following rules must be considered:

- Rule 1: If Y induces a measured variable, then hidden CDs no longer exist – they become visible CDs.
- Rule 2: If a source induces a measurable variable that further induces (below in the causality graph) the desired variable (Y), then it is an upstream CD. Otherwise, it is a VI CD.

Depending on the variable, it is possible to have different causal path suggesting either a VI or an upstream CD. Knowing how the information is distributed between the different causal paths would indicate which variable has the greatest impact on the model, and it would be possible to choose the proper type of CD. Hidden CDs, for instance, are always more difficult to treat than visible CDs. Moreover, as discussed previously, upstream CDs are slightly easier to deal with than VI CDs.

Another interesting use of causal representation is feature selection. Indeed, if the observable variables (Xi) are not yet fixed, it is possible to use the causal graph to choose them wisely. The type of CD is identified according to the measured variables. Hence, having the causal representation of the physical model allows one to more effectively select or to attribute more importance to the features used to build the learned model. Depending on the strategy, it could be interesting to focus on features that minimize the number of hidden CDs.

It is possible to classify the variables under the following three levels:

- First level: Relevant to use in the estimated model.
- Second level: Adds information but could be discarded.
- Third level: Useless; should be discarded from the model (variables that are not related to Y or that are situated downstream).

This approach becomes interesting when many features are used in a model. Reducing the number of features using this tool could increase the model's performance and reduce its computational cost.

In terms of geometry, it is difficult to obtain a precise description. However, based on a qualitative CD description magnitude, the duration or even the slope of the concept drift can be roughly estimated using the expert's knowledge.

In the same vein, recurrency can be roughly estimated in some cases based on the case study prior knowledge. For instance, in a manufacturing environment, if the tool wear is identified as a source of CD, then experts can roughly estimate the frequency of the necessary tool change. Therefore, concerning the geometry, no specific tool, but only prior knowledge, is required at this stage to estimate the recurrency of a CD.

In the next stage, data is acquired to enable the drift identification and the solution characterization. This stage makes it possible to design the data acquisition parameters:

- **The measured variables X** may induce a more hidden CD or CDs that are harder to detect.
- **The measurement sampling frequency** will be based on the Nyquist law make some CDs appear as noise.
- **The measurement sampling duration** may make the dataset not representative of the studied system as events could be missed or too slow to be impactful yet.
- **The number of examples** may limit the utilization of particular machine-learning algorithms.

4.3.1.2 Data acquisition

CDs cannot all be identified in off-line mode using the expert's knowledge. Measurements must be made to visualize and characterize CDs. The knowledge-based CD identification stage becomes fundamental to orient the selection of the data acquisition parameters previously defined. This makes it possible to reduce costs related to the measurement, which are difficult to estimate at this stage of the methodology. Data acquisition provides for building a dataset to later identify CD characteristics and evaluate the preselected maintenance solutions based on the proposed framework. The performance of the final solution will depend to a great extent on the quality of the dataset acquired, which itself depends on numerous parameters defined in the previous step. A user would in some cases want to leverage already available data so as to skip the data acquisition step. However, a close inspection of data acquisition parameters need to be done to validate the usability of the dataset.

4.3.1.3 Data-based concept drift identification

This step makes it possible to identify the characteristics of the involved CD by using the dataset built during the acquisition phase. However, the CD source cannot be directly identified, unless alarms are available in the dataset to warn of an incoming CD—for instance, a maintenance on a machine. The CD source can be identified by cross checking the characteristics identified in the data-based and knowledge-based identification phase. Knowing the CD source also makes it possible to reduce some CD effects by improving the studied process.

The type of drift is easily detected as X and Y are measured in the data acquisition step. If X and Y are changing, this indicates a visible upstream CD. If X is changing but not Y, this indicates a visible inside CD. Moreover, if Y is changing but not X, this indicates a hidden inside CD. Drift type can be easily used to correlate information from both knowledge-based and data-based phases.

The data-based model makes it possible to identify the occurrence frequency of a CD, providing for a clear definition of the potential source that had not been identified in the knowledge-based method. By contrast, if the sources are identified and known, the frequencies can be obtained by comparing the time between similar patterns.

CD geometry is easily identified using visual tools that allow CD magnitude to be represented over time [161]. This makes it possible to define quantitatively the CD geometry. This method could be used for different kinds of geometric characteristics, thus specifying or completing the information obtained in the knowledge-based method.

The overall identification process can be repeated if the CD identification seems incomplete after the data-based CD identification. This makes it possible to target new or more precise characterizations and adapt the dataset for better information extraction. Once the involved CDs are fully identified, it is possible to move to the next step—the solution pre-selection.

4.3.2 Solution pre-selection

At this stage of the methodology, the characteristics of the CD, involved in the working environment, are supposed to be known. Based on this knowledge, adequate solutions composed of a SDS, and an US must be pre-selected. They are discussed in this section.

4.3.2.1 Sampling decision system

To ensure the model's long-term sustainability, the training dataset must be regularly updated with fresh data. The SDS is the component that handles the measurement strategy, there are three types of strategies a SDS can manage.

Passive strategies, also called "time-based" strategies, sample measurement without any explicit detection. The measurements are made at a fixed frequency, which allows them to handle both hidden and visible CDs. A passive strategy does not require the implementation of a data-based algorithm; the only parameter to tune is the timer sampling frequency. Numerous timers can be defined, depending on the number of drifts. This is the simplest strategy, but it is not the most optimal one in terms of the number of measurement and CD handling.

The *active strategy* or "event-based strategy" decides to sample measurements based on a CD detector. Unlike the passive strategy, this approach can only deal with visible CDs. However, using real-time data can optimize CD rejection as well as the number of measurements needed. As it is always unlikely to have only visible CDs, a timer taken from the passive strategy can be added to the active strategy to act as a safeguard. In this case, its sampling frequency can be optimized, depending on the drift detection performance [147], [162], [163].

The *hybrid strategy* is the most complete approach as it combines the active and passive ones. When both visible and hidden CDs are present, the hybrid strategy should be used. Timers should be engineered for every hidden CD, and a drift detector should be implemented to deal with visible CDs.

4.3.2.1.1 Sampling frequency

The sampling frequency is defined by the geometry of the involved CD as well as the occurrence frequency. If the CD is gradual, the sampling frequency will be set based on the ratio of the CD slope and the acceptable drop in the accuracy of the model. If the CD is brutal, its slope will be infinite, and the sampling frequency should therefore be defined based on its occurrence frequency. The frequency can also be determined experimentally by performing tests with the training dataset [160]. It must be emphasized that too high a frequency would induce a high measuring cost. It could also affect the model's accuracy if the chosen samples are not providing relevant information to the update. Too low a frequency would miss too much CD, and the model would become obsolete over time.

4.3.2.1.2 Drift detection

Drift detection algorithms are data-based algorithms. In this paper they are classified into three categories: statistical tests, clustering, and "in-built"-based methods. Statistical test methods compare two data distributions to spot any significant change that would result in a CD. Distance functions are used to compare and quantify historical data distribution with the new data distribution [164]. Clustering methods, which are the most popular family of drift detection, examine the change in data density. Clusters are used to identify concepts. Several different clusters can coexist at the same time [146]. "In-built" methods are drift detectors integrated into the estimation algorithm. Most of the time, they will estimate the uncertainty of every inference and threshold it. Currently, there is no way to choose a category of algorithms that is dependent on the involved CD characteristics. Indeed, the literature does not contain research justifying the chosen drift detector based on the identified characteristics of a CD, and this will be a major gap to study in the future. Nevertheless, it is possible to discuss guidelines to tune the drift detector based on CD characteristics. Indeed, all the approaches have one or multiple hyper parameters to set the sensitivity of the drift detector. In the following paragraphs, this sensitivity is related to CD characteristics.

Concerning the CD geometry and particularly the CD magnitude, a higher sensitivity will enable the detection of smaller magnitude CDs; by contrast, a lower sensitivity will limit false positives. This is the same for the CD slope: the higher the slope, the less sensitive the detector must be. For instance, in statistical test methods, the robustness of the algorithms can be tuned by the choice of the hypothesis test. Too low a low threshold would lead to poor detector performance over small CDs. The process is the same with the clustering methods, where the sensitivity is tuned by changing the density threshold or the distance between clusters.

Tuning the sensitivity is related to the management of outliers, which can be a major source of false positives. As previously discussed, if the drifts have high amplitude and slopes, sensitivity will naturally reject outliers. However, other approaches could be used for outlier rejection. When the drift detector is triggered, the case could be added in a buffer. Depending on the amount of successive or similar elements in the buffer, the example could be classified as CD or as an outlier. The decision can be based either on heuristics or statistics [165], [166]. This adds another layer of protection against outliers and makes it possible to increase the maximum sensitivity that can be chosen.

The CD detection concerns only visible CDs, which are either upstream CDs or inside CDs. Upstream drifts start to have an impact when the model inputs X goes out of the training set in the extrapolation zone of the estimation model. By contrast, inside drifts will directly lower the model accuracy by changing the function between X and Y. Most of the time, upstream CDs will require sensitivity that is lower than inside CD. This effect will depend on the extent of the training set and on the model's extrapolation capability.

As previously explained, there is no existing rule to choose a suitable algorithm based on the CD degrees of freedom. Currently, the best solution is to choose different algorithms, tune them according to the CD geometry information, and test them on a testing dataset to be able to choose the most suitable one for the application.

4.3.2.2 Updating system

When the role of the SDS is to decide when to measure, the role of the US is to define how to update. Once the fresh measurements are acquired, the US can start the model adaptation. By modifying the learned concept, the US minimizes the effect of the CDs on the model's accuracy. As with the SDS, the longevity of the model will depend on which US is chosen and its tuning. Indeed, there are many factors that can influence the updating mechanism, which can even reduce the estimator MAE in the worst case. The selection and tuning choice depend on the CD's characteristic. In this section, the different degrees of freedom defining the US are discussed.

The slope of a CD is the first CD's characteristic which influences the updating strategy. From a qualitative point of view, the slope is defined as sudden or gradual. An important dilemma—one that highlights the difference between sudden and gradual CDs—is called the stability-plasticity dilemma. This dilemma results from the tradeoff between being stable and handling noise and outliers, on the one hand, or being plastic and adapting more quickly to CDs, on the other. In general, a US should invest more in the noise and outlier impact mitigation of the MAE. However, the more sudden the CDs are, the more plastic the US should be. Indeed, the stability becomes a flaw if the adaptation time is slower than the concept evolution. This phenomenon is represented in the most used approach for a US—the moving window. The window size is a parameter that illustrates the tradeoff between stability and plasticity. Small ones are suited for detecting abrupt CDs, whereas large ones are better at detecting gradual CDs [167]. The type of the sliding window is the first characteristic necessary to design a moving window. The simpler form is a fixed-size window—a method where the model is periodically updated using a window containing a fixed number of instances, where each new instance replaces another one in the window. (The strategy of deciding which point to discard relates to forgetting capabilities, and this is discussed later.) The size needs to be designed iteratively, as previously described, and will be adapted to one fixed ranged of CDs [168]. When dealing with different geometries of CDs, the optimal size of the moving window may depend on time. Accordingly, the use of a variable window size, where the size changes depending on the error of estimation [169], [170], can be considered. The higher the error, the smaller the window (and vice versa). The proposed framework assumes that the error of estimation is available only on measurements. Thus, it is possible to optimize the window size based on the error from the new measurements before performing the update. Important parameters to take into consideration when tuning the window size are outliers and noise. Indeed, both influence the performance. A small window will be more affected by noise and outliers, and the model accuracy will decrease, while large windows will dump their effects.

The second important CD characteristic for updating approach selection is CD magnitude, which describes the severity of the CD. During updates, the new examples will be added to the previous ones before retraining. However, if the training set is too large, the new information might be drowned. In the literature, this dilemma corresponds to the class-imbalanced problem, which also applies to regression. Therefore, the higher the CD magnitude, the greater the modification of the concept and the higher the number of required examples to restabilize the MAE. Each time the SDS triggers an alarm, the number of measurements performed can be higher than one. This is one important parameter that can mitigate the class-imbalance problem if it is well defined. However, with a moving window, a fresh data added is more of the time another data deleted. Too many measured points can be expensive and counterproductive. A solution to address this issue is called "instance weighting," a method whereby the model is updated by applying weights to each example to give more importance to some of them [167]. Different techniques exist for weighting the instances. Some assume that the most recent data is the most informative and thus give them more weight than the old ones; however, this assumption is not appropriate in every context. For instance, it does not hold in presence of recurrent concepts.

The third important CD characteristic to consider while tuning the US forgetting capability is its type. The moving window comes with a forgetting capability, which is mandatory as industrial implementation mostly comes in a data streams form. Data streams assume an infinite number of iterations; it would not be feasible to remove elements to assure a finite data storage and inference time. The forgetting capability should remove the less informative example. In the case of an inside CD, the relation between X and Y changes, making all the old examples less informative. The forgetting capability should therefore remove the oldest example. In the case of an upstream CD, the relation between X and Y does not change, which does not necessarily make the old examples less informative. Thus, the density can be representative of the example's informativeness. In cases where there is a diverse type of drift, the most conservative approach should be selected, which is the one based on the example's age.

The proposed approach makes it possible to select and tune the adequate algorithms to address the involved CD. Therefore, an ensemble of solutions, composed of an SDS as well as a US, can be extracted from this stage of the methodology. This is a first step in reducing the number of solutions. The next step will evaluate the pre-selected solution to eliminate the ones that do not respect the environmental constraints and to select the best solution that minimizes the costs.

4.3.3 Optimization and solution selection

The last step aims at choosing the most optimal solution among the pre-selected ones. First, an optimization problem is defined where the objective function and the different constraints are discussed. Then, an evaluation phase, where possible solutions are compared, is performed. Finally, as a result of the optimization problem, the best solution is selected.

4.3.3.1 Optimization problem definition

The optimal solution can now be selected because of the cost and constraints related to the number of measurements and the accuracy of the model. The optimization problem can be defined as such:

$$minimize\ C^T X$$

$$subject\ to$$

$$X \leq Constraints = \begin{pmatrix} Accuracy\ constraint \\ Measurement\ constraint \end{pmatrix}$$

$$X = \begin{pmatrix} Accuracy \\ Nbr\ of\ measures \end{pmatrix} \in Solutions, \quad C = \begin{pmatrix} Accuracy\ cost \\ Measurement\ cost \end{pmatrix}$$

Equation 9: optimization problem definition.

Where,

X is the vector of variables defined as the metrics of the tested solutions, which have been pre-selected in the previous stage of the methodology. It is composed of:

- **Accuracy** of the model when using the algorithms chosen for the solution X.
- **Number of measures required** by the relative solution X to maintain the model updated with the corresponding accuracy.

C is the cost vector characterized by two metrics:

- **Measurement cost** is the cost of measuring one unit for the company. This cost gathers resource costs such as the workforce used or the renting cost for the measure machine and potential shortfall. This cost varies from one industry to another and from one process to another.
- **Accuracy cost** defines the cost of risking a bad estimation. The range of acceptable error over the estimations must be defined. As for the measuring costs, this accuracy cost is dependent on the industry. For instance, in predictive maintenance applications, this cost could be represented by machine compensation cost and induced shortfall.

$C^T X$ is the objective function which must be minimized. It relates to the cost induced by the solution. The purpose is to find a solution that causes a minimal cost. $C^T X$ can be seen as a ratio between acceptable model performance variation and acceptable measurement budget. Acceptable model performance variation is defined by the product/process specification particular to the industry. Most of the time, this is regulated by customer expectations. The acceptable measurement budget is defined by the company, and it is relative to the attributed budget. Indeed, in the simplest case it would be a linear function.

The constraints are induced by the environments. There are two types of constraints: accuracy constraint and measurement constraint. The first type comes from the problem requirement definition. For instance, for VM, one could use the adaptation of the ISO norm to define this tolerance based on the industry [9]. The second one can come from the measurement material or human resource limitations.

Each parameter of the optimization problem must be defined. X is given by the solution pre-selection stage. The cost vector, the objective function, and constraints are given by the problem environment. Once the problem is defined, the evaluation of the pre-selected solution can be performed.

4.3.3.2 Offline solution evaluation

Pre-selected solutions need to be evaluated using the dataset built during the data acquisition step. The data acquisition step is subdivided into three subsets: (a) a training set for training the different algorithms of the solution, (b) a validation set to evaluate the performance of the solutions, and (c) a test set to evaluate the final performance of the selected solution. Therefore, for each pre-selected solution the following process is performed:

- Solution implementation in the framework.
- Training of the algorithms with the training dataset.
- Evaluation of the accuracy of the estimation model on the validation set.
- Evaluation of the number of measures required to reach such a degree of accuracy on the validation set.
- Mapping of the solution to be able to visually compare the different solutions. A graphic representation of the solution mapping is given in Figure 25.

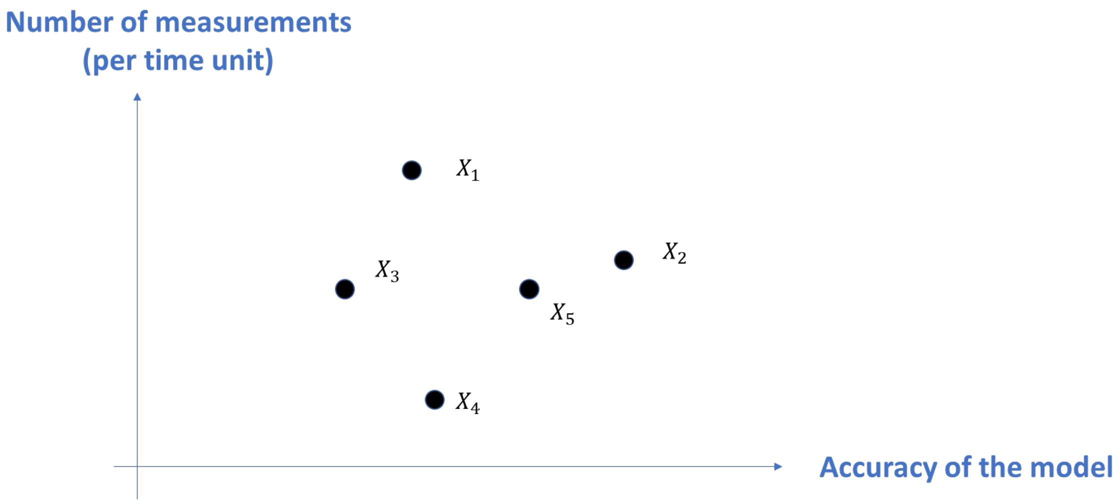


Figure 25 : Representation of the pre-selected solutions mapping. Black dots represent different solutions.

Once all pre-selected solutions are evaluated and mapped in a graph, the problem optimization can be implemented to select the best solution.

4.3.3.3 Solution selection

To select the most suitable solution, the optimization problem is implemented. By minimizing the optimization function, considering the constraint, the adequate solution is obtained. The following graph is produced where the optimal solution can be graphically identified.

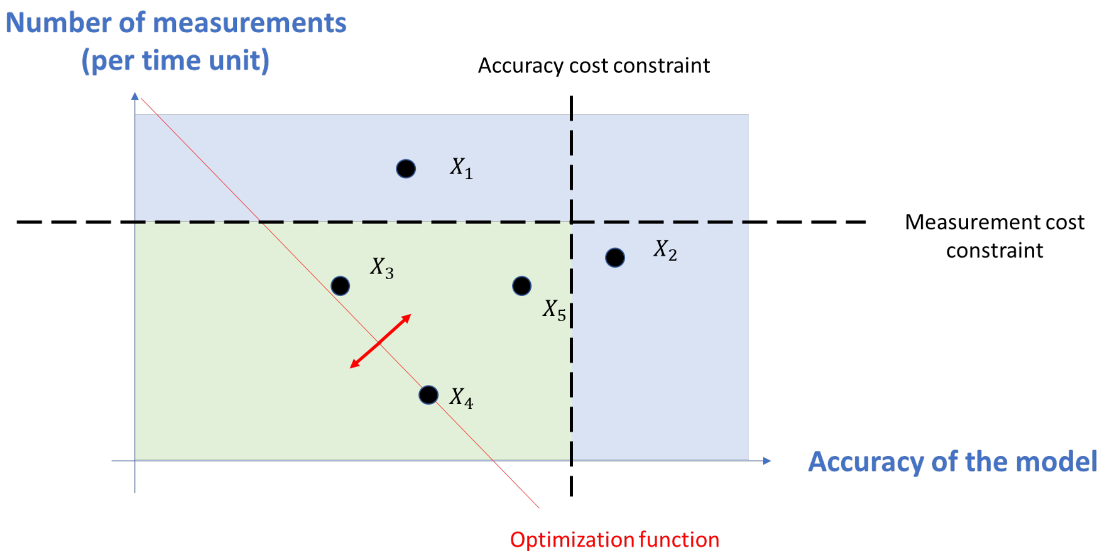


Figure 26: Optimization graph. Black dash lines represent the problem constraints. The green area is the valid solution space. The blue area is the invalid solution space. The red line represents the optimization function.

The two optimization constraints—the measurement constraint and the accuracy constraint—set the limits of the valid solution space (the green area on Figure 26). Solutions outside the valid space do not respect the problem constraints and therefore can be eliminated. The optimization function is represented by a ratio (the red slope on Figure 26)—in this case linear—between the acceptable model performance variation and acceptable measurement budget, which makes it possible to select the best solution. This ratio corresponds to a pareto efficiency situation where the pareto front is used as an optimization function to define the optimal solution called the pareto efficient. In this example, a linear function is considered for the pareto front, and it can be seen that the solution X_4 is the most suitable one. The selected solution can be tested on the test set to ensure its generalizability.

The proposed methodology makes it possible to select an adequate solution to address the identified CD for a given environment. The methodology is illustrated through a simulation in the next section.

4.4 Practical implementation

The following case study represents, on a simulated industrial example, the previously explained methodology.

This example pictures an enterprise wanting to use virtual metrology (VM) on the manufacture of a product. The VM algorithm should estimate the *cutting width* (a_e) realized by a machine tool during the milling of a certain region of a part on a known operation. Its inputs, provided by sensors placed on the machine, are the *cutting power* (P_c) and the *feed speed* (v_f). This simulation mainly aims at illustrating the methodology and not at comparing the CD detection algorithm or virtual metrology algorithm.

The company is interacting with a simulation developed with the equations presented in Table 18.

Table 18: Equations used for the simulation.

P_c (kW) : Actual Cutting Power	$P_c = \frac{a_p\, a_e\, v_f\, k_c}{60\ 10^6\ \eta}$
a_p (mm) : Depth of Cut	$a_p = 5$
a_e (mm): cutting width	$a_e = \mathcal{N}(2, 0.5)$
v_f (mm/min) : Feed speed	$v_f = \mathcal{N}$ (15000, 1000)
k_c (N/mm2) : Specific Cutting Force	$k_c = k_{c1}\, h_m^{-m_c} \left(1 - \frac{\gamma_0}{100}\right)$
k_{c1} (N/mm2) : variable	$k_{c1} = 2200$
h_m (mm) : chip thickness	$h_m = 1.75$
γ_0 (°) : rake angle	$\gamma_0 = 0$
η ϵ [0-1]: efficiency	$\eta = 0.8$

Three CDs occur in the simulation. The first CD (d_1) linearly changes the efficiency of the machine due to the wear of the tool, which is reset to zero when maintenance occurs. A second CD (d_2) suddenly changes the value of k_{c1}, due to the evolution of the raw material lot used for manufacturing the part. The third CD (d_3) linearly increases the mean of the Gaussian function that describes the cutting width (a_e), due to the wear of the mechanical stop, which is corrected at each maintenance.

The virtual metrology algorithm chosen for this experiment is a multilayer perceptron with two inputs, P_c and v_f, one output a_e, 2 hidden layers of 7 neurons each, and RELU activation function.

4.4.1 Concept drift identification

4.4.1.1 Knowledge-based concept drift identification

Experts identified tool wear as a possible source of gradual CD followed by a sudden CD when the tool is replaced. The tool lifetime is generally 30,000 manufactured parts long. For this reason, the dataset must be taken during at least this number of parts; moreover, it must be taken with a relatively high frequency to be reactive to the sudden CD happening after maintenance is performed and to other possible CDs with high frequency. Therefore, it has been decided to measure 3000 parts, one every 10.

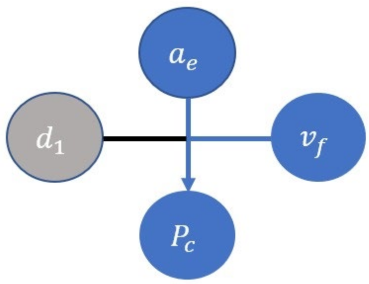


Figure 27: Preliminary directed graph.

Figure 27 shows the causal graph of this experiment. The supposed equation has been identified by experts, as recorded in the literature. The arrow represents equation 9; in other words, it represents the physical model linking the cutting width, the feed speed, the CD over the motor efficiency, and the cutting power. It is an anti-causal problem (Y is causing X), which means there is no hidden CD; moreover, d_1 is a source of a visible inside CD, which will affect the efficiency.

$$P_c = \frac{a_p \, a_e \, v_f \, k_c}{60 \;\; 10^6 \, \eta} \text{ kW}$$

Equation 10 : Cutting power equation

4.4.1.2 Data acquisition

After 7.000 measurements with different frequencies, the dataset was separated into a training and validation set. Figure 28 presents the plots of the validation set.

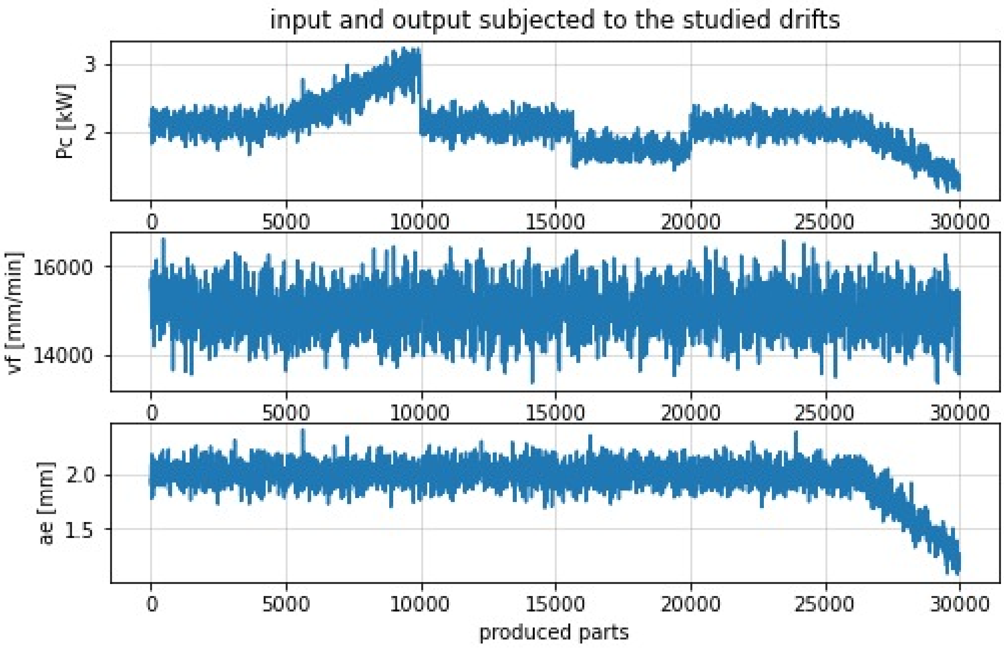


Figure 28: Concept drift on the dataset.

4.4.1.3 Data-based concept drift identification

The data presented in Figure 29 confirms the impact of the CD due to wearing of the tool (d_1), visible from the 5000th -produced part until the 10000th. Figure 29 also shows two unpredicted CDs: one, (d_2), that suddenly changes the value of the power (at points 16,000 and 20,000), and the second, (d_3), that is linear and affects simultaneously the cutting power and the cutting width (starting around the 27,000th manufactured part). This last one seems to be an upstream CD. The signal is somewhat noisy, which highlights the potential presence of outliers.

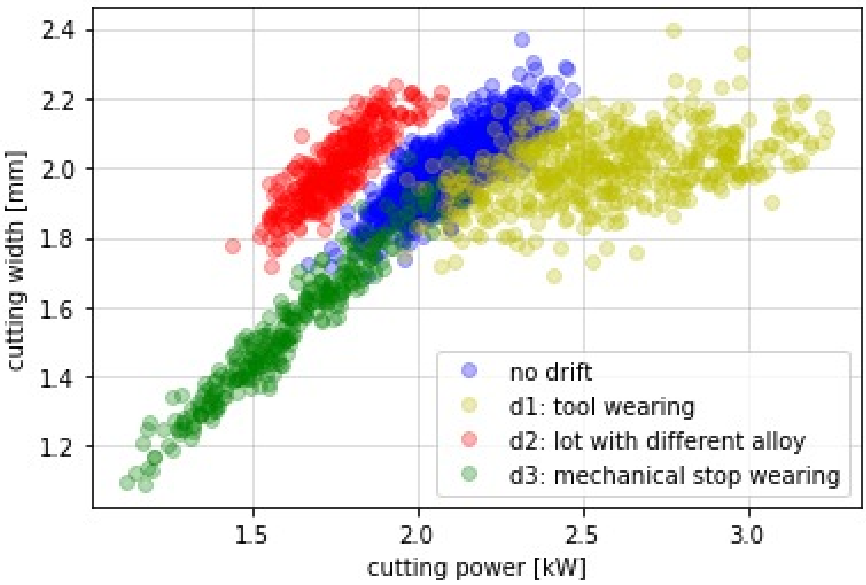


Figure 29: Cutting width in function of the cutting power for the different CD.

The different CDs of the evaluation set are presented in Figure 29 on a X-Y plot. First, d_3(in green) follows the same shape as without CD (in blue), but over an unexplored region. For this reason, it can be validated as an upstream CD. Both d_1 and d_2 CDs modify the concept between the cutting power and the cutting width. They are only identifiable on the cutting power signal. They are, therefore, classified as visible inside CDs.

4.4.1.4 *Knowledge-based concept drift identification (second iteration)*

After investigation, it has been noticed that the d_2 corresponds to one specific lot of raw parts. It corresponds to a sudden CD that is certainly related to a defective lot of raw material with a different inner constraint. The material properties affect the power consumption of the machine tool during the manufacturing of these parts. It was determined that it comes from the wearing of the mechanical stop that enables a precise loading of the raw material during the milling process. It has also been noticed that no CD occurs over v_f. It is still measured to enhance the accuracy of the VM model, as the sensor investment was already done. Therefore, one upstream CD and two visible inside CDs are identified, as shown in Figure 30.

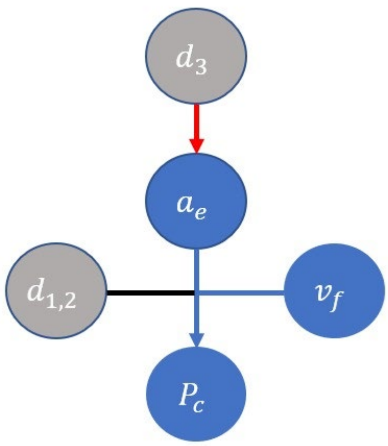


Figure 30: Updated directed graph.

4.4.2 Solution pre-selection

From the previous stage, the following information concerning the involved CD was identified:

- No hidden CD.
- Presence of an incremental upstream CD (d_3).
- Presence of two visible inside CDs, one gradual (d_1) and one sudden (d_2).

4.4.2.1 Sampling decision system

Since no hidden CD was detected, it has been decided to use an active strategy that includes a safeguard with a low passive sampling frequency, having as a period half of the acquired signal length (15000 parts), which is activated only when no CDs are detected by the active module over that period. Concerning active CD detection, it has been decided to compare two algorithms: a statistical method (CUSUM) [31] and a clustering method (OLINDDA) [25]. The clustering method is robust to outliers, which gives it a greater range for tuning its sensitivity. The presence of noise and outliers added to the quiet high amplitude of the different CDs motivate the use of a relatively low sensitivity for CD detection.

4.4.2.2 Updating system

For the updating of the multi-layer perceptron, a moving window is used. First, the slope of the CDs is relatively important, making them easier to detect but requiring higher flexibility and thus a smaller window size. Second, as the magnitude of the CDs can be important, the number of new points to insert in the window at each update should be larger than one. By contrast, as the window is relatively short, the number of new points cannot be too high. Lastly, as there is an inside CD, the older point is removed at each update. This entire process makes it possible to define a short range of interest for the moving window size and number of points needed for updates, all of which will be tested.

4.4.3 Solution selection and optimization

4.4.3.1 Optimization problem definition

The metrics defined by the enterprise depends on the cost of each measure (c_m) and the cost of a poor estimation (c_{MAE}). Each measure costs 31 USD, and each millimetre of MAE costs 6 USD. The optimization problem aims to minimize the sum of the two costs ($C^T X$); the variables of the optimization are the number of measures (n_m) and the mean absolute error of the produced part of a lot (MAE). These values are taken in the batch, including all the evaluated solutions. The enterprise cannot measure more than 1000 parts over a lot of 30,000 because of the measuring machine availability constraint. The optimization problem is therefore defined as follows:

$$\min_X (C^T X)\text{ , with } C = \begin{bmatrix} \frac{c_m}{n_p} \\ c_{MAE} \end{bmatrix},\ X = \begin{bmatrix} n_m \\ MAE \end{bmatrix}\text{, under } n_m < 1000$$

Equation 11: application of the optimization process

c_m **[USD/measure]:** cost per measure	c_m **= 31**
c_{mae} **[USD/mm/part]:** cost of the bad estimation	c_{mae} **= 6**
n_p **[parts]:** number of produced parts in the validation set	n_p **= 30000**
n_m **[measures]:** number of measures	value to optimize
MAE [mm]: mean square error of a lot of n_p parts	value to optimize

4.4.3.2 *Offline solution evaluation*

On the tuning phase, for each algorithm it was necessary to evaluate different sets of hyperparameters, including the parameters for updating the VM algorithm, which are the moving windows size and number of new points added to the moving window. More than 300 combinations have been tested offline for both drift detection algorithms. Figure 31 shows the MAE for each set of hyperparameters, over 30,000 produced parts in function of the number of measures based on the dataset acquired previously. In this case, it can be observed that extended CUSUM reaches better performances compared to OLINDA.

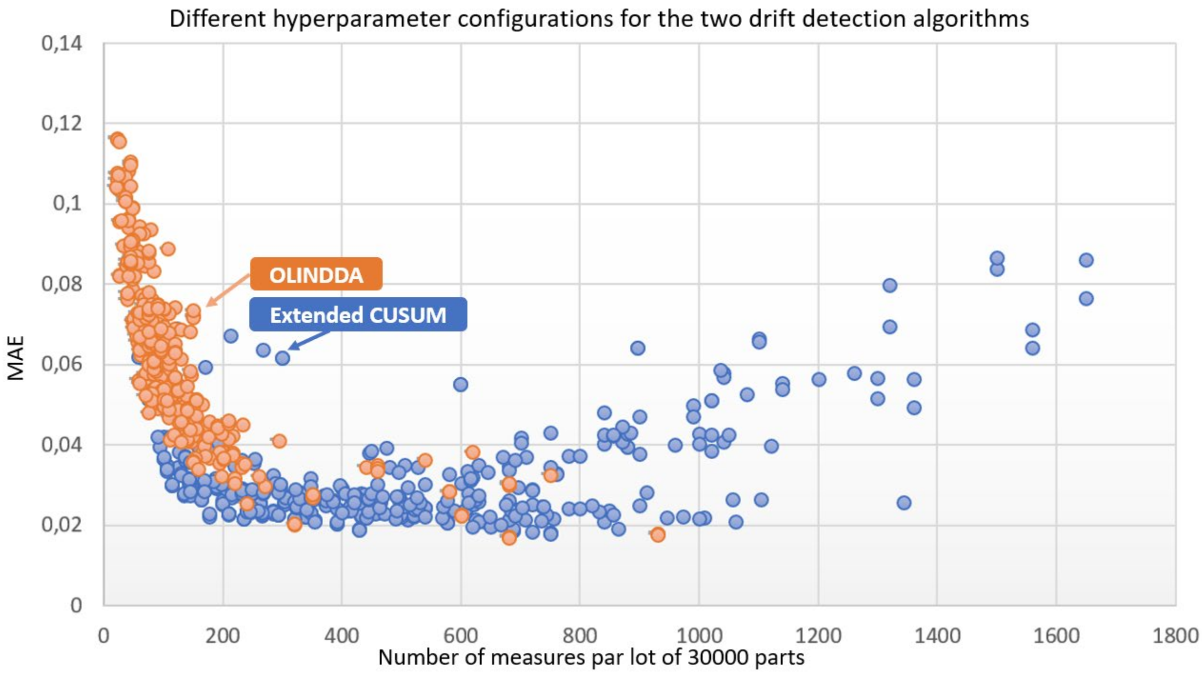


Figure 31: Offline solutions evaluation graph

4.4.3.3 *Solution selection*

As presented in Figure 32, the best approach for the enterprise, according to the cost and constraints previously defined, is extended CUSUM with an MAE of 0.03 mm and a number of measures equal to 114 over the test period. The total cost due to the poor estimation and the number of measures per produced part is equal to $C^T X$, which in this case is 0.30 USD/part. The size of the moving window used to update the estimator is 10 and the number of new points measured before each update is 3.

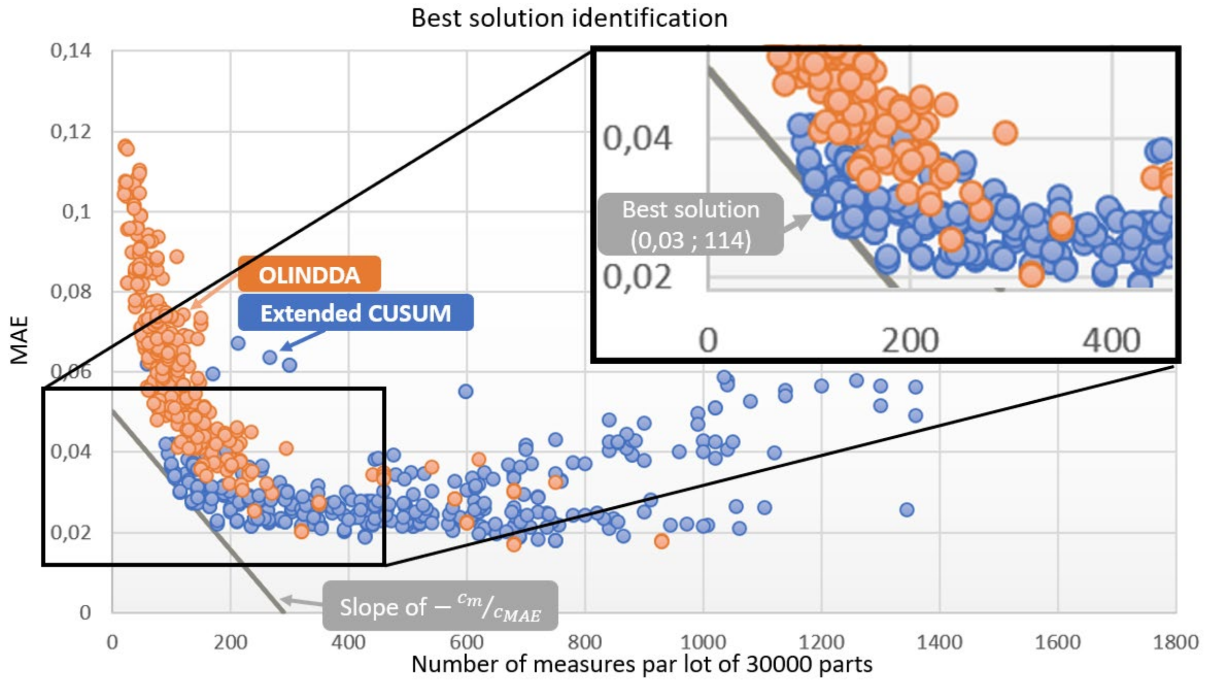


Figure 32: Identification of the optimal solutions

In order to test the performance of the selected algorithm, a test set of 1000 measures, one every 10 produced parts, is collected. Due to its small size, this test set only contains d_2 and d_3.

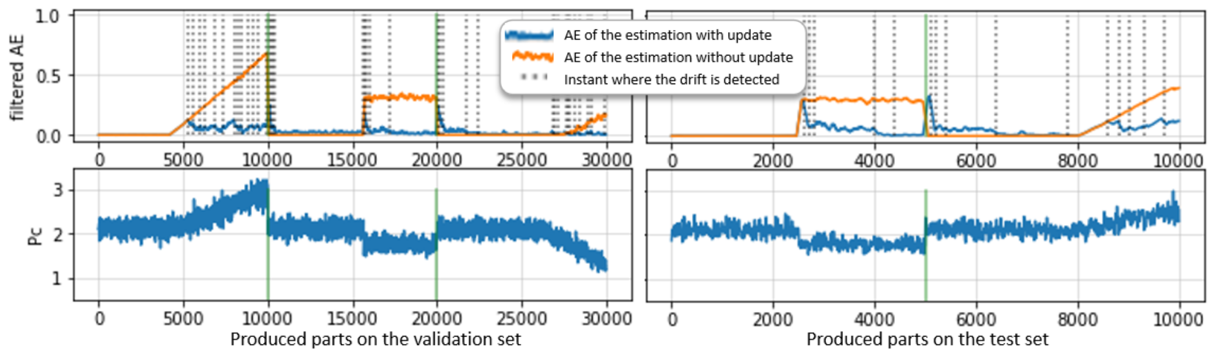


Figure 33: Validation and test set of the selected solution.

Figure 33 shows the absolute error of the VM algorithm estimation at each manufactured part for the validation and test set. The blue and orange lines of the upper graphs represent, respectively, the absolute error of the VM algorithm estimation with and without an update. The dotted line indicates the instant where a CD is detected, a measurement is triggered, and an update of the estimator is made.

On the test phase, 45 parts are measured, and the resulting MAE is 0.0455 mm, for a cost of 0.41 USD/part. The solution seems to generalize correctly and is validated.

4.5 Concluding remarks

This section presented some key discussion points regarding CD in general, and the following subsections focus on the different functionalities of the proposed framework. The scientific contribution of the current research work will have a significant impact on the performance of all data-driven technologies. Most Industry 4.0 technologies, such as machine learning, artificial intelligence, digital twins, VM, predictive maintenance, and zero-defect manufacturing (ZDM), are data driven. The creation of the models is one part of the solution, but there is another side that is currently not receiving the attention that it should be, namely the maintenance of the models and keeping them up to date so that they perform well and efficiently. Currently, most data-driven solutions are designed and developed to be static, meaning that after the training of the corresponding model, they are ready for deployment without considering the CDs that will occur over time. The proposed framework provides all of the directions required for turning the existing static models into a dynamic one. By considering CDs, the flexibility and adaptability of the models are increased significantly, which consequently increases accuracy, leading to a sustainable solution.

The phenomenon of CD has its origin in the dynamic nature of manufacturing systems and in general any system that changes over time. More specifically, in manufacturing systems, CDs are generally occurring because of the deterioration and wear of components, tools, and material, but also from the quality of the input material, human error, and environmental changes such as temperature or humidity in the shopfloor, and so on. Additionally, in the same spirit, feedback loops such as sensor data might change slightly over time because of the deterioration of the sensor itself, sending varying information. In an industrial environment, CDs occur systematically, which makes their consideration fundamental to ensure the maintenance of industrial data-based models.

Most of the Industry 4.0 technologies such as machine learning, artificial intelligence, digital twins, virtual metrology, predictive maintenance, and zero-defect manufacturing are data-driven technologies [32]. Currently, most of the data-driven solutions are designed and developed to be static, which means that it is ready for deployment after the training of the corresponding model, without considering the CDs that will occur as time passes. The proposed methodology enables the adaptation of existing static models to dynamic ones. Taking the CDs into consideration, the flexibility and adaptability of the models is increased significantly, which consequently increases long-term accuracy. In other words, the proposed methodology gives a general approach to ensuring the maintenance of data-based models for the industry.

The presented methodology proposes to take into consideration CD characteristics to design solutions that consider the practical issues of solution implementation. For the first stage of the methodology, CD identification, the participation of application experts appears to be crucial. The better the CD identification, the better the model maintenance and, thus, the smaller the measuring and accuracy costs. However, in some cases the CD identification can be difficult to perform, such as when several CDs impact the system simultaneously. In those cases, identification tools and methods could reach their limits. Therefore, further research could be done to develop new methods to enhance data-based CD identification capability. In terms of CD characterization, the literature is filled with diverse terminology that does not serve the interests of a uniform methodology. The new proposed CD types aim to facilitate the CD identification. Further research still needs to be undertaken on the normalization of CD geometric characteristics.

The second stage of the methodology, pre-selection solution, makes it possible to pre-select various solutions by choosing a suitable SDS and US to deal with the involved CD. Currently, the procedure to build solutions is to study the literature to find other works dealing with similar case studies. However, two similar applications can involve different CDs. There are not two identical industrial environments for the same process, so the CDs are obviously different. This reveals the limit of the previous methodology and the need to analyze the CD characteristics when selecting a solution. Indeed, if the algorithms are selected or defined by the involved CD, then they would not be application-dependent anymore. Therefore, one will be able to pre-select solutions of interest based only on the CD that must be addressed. SDS is the key component for handling potential CDs. The passive strategy is the simplest approach to implement. Most of the implementations include a single timer; however, having one timer per CD could be interesting to develop. In a general way, further research on developing techniques for choosing the sampling frequency could benefit not only the CD handling field, but also the industrial quality control field with the optimization of the batch measurement frequency. For active strategies, the implementation must integrate a timer safeguard in industrial environments. Indeed, CDs can be missed when doing the CD identification, or new CDs can appear over time. There are few methods to optimize the safeguard sampling frequency, so there is still space for improvement. Finally, the hybrid strategy appears to be the best choice, in general, for industry environments due to the high number of CDs.

The developed framework considers online updates; the update is made immediately once an example is available. However, in the literature incremental updates are described in which the model is updated once the entire window of new data instances is sampled [33]–[35]. This raises the question of when to update, as determined by the size of that window, which makes it possible to fully dissociate the SDS from the US. The main advantage of this technique is to minimize the computational cost at the price of a slower reaction to CDs. The computational cost could also be minimized by updating the model in cascade. Incremental updates could become interesting when training a specialized estimator as it is done in ensemble learning. Ensemble learning has not been discussed in this paper as it is a special case. Ensemble learning has the unique feature of being able to update by removing the specialized estimator and adding new estimators specialized with a newer batch of examples [36]. It is also possible to store old models in "sleeping mode" in case the concept comes back [37]. Such methods require memory to be able to stock the different concepts; the updating is then more about model management than real updating. The most efficient models are activated while inefficient ones are deactivated. This seems promising as it can deal with recurrence. However, such a method requires identifying concept signatures to be able to recognize when a sleeping model will be useful. Thus, this approach seems somewhat futuristic, and it is not a priority for future research. In a general way, this methodology has shown that updating mechanisms depend on different CDs' characteristics, which further reinforces the need to know the type and characteristics of the involved CDs in a problem.

The solutions developed with this methodology are suitable in most cases. However, when working in extreme constraints environments, this methodology might not be adequate as it is. Four different cases representing different scenarios of the optimization problem are presented in Figure 34 as a way to discuss alternative solutions. The two questions at the basis of the scenarios are as follows:

- Are all samples measurable?
- How much does it cost to measure a sample?

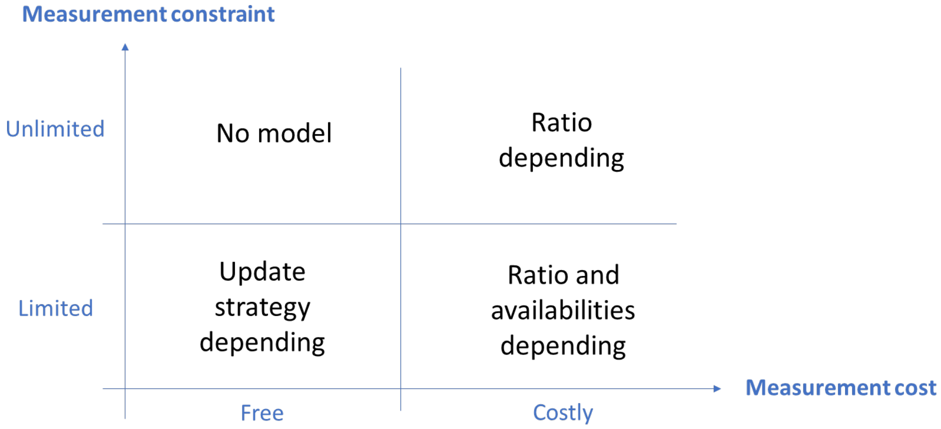


Figure 34: Measurement constraint graph.

If the samples to measure are unlimited and their measurements are costless, then there is no need for a model. Indeed, the measurements can be performed whenever they are required without caring about the relative cost. Hence, having an estimation model has no advantage over the physical measurement, and it does not make any sense to have one.

If the measurement frequency is limited but the measurement is free, the model enables measurement continuity while the real values are not available. As the measurement is free and more accurate than the model, measurements will be done every time possible. However, continuously updating the model with new points can be detrimental as the updating approaches; it is not possible to merely add the new data in the training set, but rather to exchange them. It can happen that the new examples hold less original information than the replaced ones. In this architecture, the SDS does not define when to measure but rather when to update. In this vein, active learning algorithms are suitable as they give as much relevant information as possible to the model. The US is thus connected in the SDS, which make sense in this case. Therefore, optimizing the forgetting mechanism of the US is fundamental in this case.

If the measurements are costly, the budget becomes a constraint and a ratio between acceptable model performance variation and acceptable measurement budget appears to be a good indicator for sampling decision strategy optimization. Most of the time, the first one is often hidden in the CD detector's hyperparameters for active strategies, whereas the second one is used to design the time-based measurement frequency of passive strategies. This ratio is similar to the one used by manufacturers for classical product quality control, with the key difference that it is not product quality that is out of tolerance but estimation accuracy. The measurement budget is, in general, an incentive, not a constraint [38]. If it is a constraint, the problem enters the class of costly and limited measurement problems [13]. The methodology proposes in this research is made to deal with this type of scenario.

When the measurement is costly and only partially available, a queue layer must be added to the strategy. Indeed, the time between the measurement order of the SDS and the actual measurement is unknown in this scenario. Next, the desired measurement must be queued. The queuing importance may depend on the measurement priority. Moreover, the ratio between the measurement frequency and the measurement availabilities can help greatly in defining the measurement strategy. If the measurement availability is low, it would be detrimental to miss the opportunity of measuring. To deal with this, one would want to complement a classical CD detection algorithm that does estimation with a forecasting CD prediction that is able to optimize the sampling strategy accounting for the measurement limitation. Further research on sampling strategies with limited and costly measurements would enhance the literature and be helpful for many applications.

With the last stage of the methodology, optimization, and solution selection, a clear vision and a good understanding on the problem are required. Indeed, it is mandatory to be able to define the right constraints and optimization function. Moreover, update process depends on the used estimation model. Therefore, it is recommended to consider the full operational framework when choosing a solution. This means considering the choice of the model with the SDS and US choices.

The proposed framework and methodology, as illustrated in Section 4.4, are directly applicable for VM. They answer the second research question by proposing an approach for maintaining the VM algorithm on the shopfloor. This chapter has provided an in-depth solution that is included in the essential group defined in Chapter 2. As previously discussed, the maintenance of industrial data-based models is a new topic that, out of necessity, will become a major part of any Industry 4.0 solution implementation. Having been based on an industrial need, this theme will soon become fundamental for both industry and research.

Chapter 5 Conclusion

Over the past decades, numerous aspects of industry have evolved, ranging from production technology with numerical commands to the organization of shopfloors with six sigma lean approaches. However, in terms of quality control strategies, the duality between batch sampling and systematic sampling has persisted. Today, virtual metrology (VM) is on the brink of enabling a new standard for quality control across all industries through a third strategy that combines the benefits of the others while minimizing their drawbacks. This thesis focused on implementing industrial VM for one of the most common types of production machine, namely machine tools. The contributions of this work serve as a proof of concept as well as guidelines on how to design, tune, implement, and maintain a VM solution on a shopfloor. The industrial applicability of the results was a main focus throughout the thesis, which explains the choice to conduct all of the case studies on an industrial undergoing shopfloor and not in a university laboratory. Moreover, the lack of knowledge on how to maintain such a solution seemed to be a crucial issue to address, as any industrial solution must have a maintenance plan. In the following paragraphs, the previous chapters are summarized, highlighting the most important findings and discussing further research. The present thesis is then concluded, by underlining some topics of interest which were developed during the thesis but not related with a research question.

First, a complete state of the art was outlined through searching all relevant literature using the keyword "virtual metrology," which, after screening, resulted in 199 papers. A clear lack of structure in the literature was highlighted, which motivated the design of a framework and terminology. Three groups were proposed to structure the state of the art of VM. First, the core group included quality estimation and preprocessing. It represented the necessary elements for implementing a VM prototype. Second, the essential group included the sampling decision system (SDS), drift detection, and updating features. It grouped the elements to build a VM maintenance plan that would allow the implementation of lasting industrial VM solutions. Finally, the auxiliary group included the automatic machine control, fab-wide architecture, multistage architecture, and adaptability features. It grouped the optional elements intended to enhance VM performance, adaptability, and connectivity with the surrounding environment. The proposed framework demonstrates how all those elements are connected and interact with one another. Thereafter, for all those elements, a systematic state of the art was established. Beside the technical aspect of the state of the art, a specific focus was on the industrial applications targeted by the researchers working on VM, based on the state of the art. Numerous findings were discussed in the concluding remarks section of Chapter 2. The most important findings, besides the framework itself, are summarized here. First, for dimensionality reduction and quality estimation, the use of nonlinear algorithms is now the standard. For quality estimation, various new capabilities have been proposed, such as a causality study for BNNs, automatic feature extraction for CNNs, uncertainty estimation for GPR, and simplified updating for ensemble methods. Some subframeworks to structure new VM subdomains were also proposed in this thesis, such as for multi-stage architectures classified as serial, hub, and cascade architectures. In terms of industrial model connection, namely fab-wide architecture, it is now driven by the product-as-services trend. The latest research focuses on cloud implementation.

In terms of research gaps, the core group, including quality estimation and data preprocessing, is well-developed and mastered for most semiconductor operations. However, it is undeveloped for other industries. As discussed, the benefits of VM no longer remain to be proven, but its applicability to newer domains does. Moreover, several research groups have started to work on these research gaps, but everything remains to be done. For the essential group elements, there is a notable lack of research. Due to the fundamental industrial necessity of making the shopfloor solution last, those elements will soon become a major research topic. As with the essential group elements, the auxiliary groups elements included underdeveloped subjects with high potential. For instance, run-to-run machine control using VM is a studied application, but only one paper focused on real-time control such as VM based model predictive control (MPC). As discussed in the introduction, real-time control is the future of the actual industry and will have a major impact on every shopfloor. However, it can only be enabled based on VM technology. The essential and auxiliary groups are underdeveloped because the core group is not mature yet. Before maintaining or enhancing a solution, the solution needs to be developed. However, solutions coming from the essential group will be required, as the core group will soon be mature for industrial implementation. This fact motivated and informed the two research questions of this thesis, which focused on the implementation of core group solutions on machine tools and the development of auxiliary group solutions. The first research question was as follows:

"Is the application of VM to estimate the dimensional quality of products milled with a machine tool feasible?"

To answer this question, a proof of concept of VM applications on milling machines was proposed. This allowed opportunities to be highlighted and will further help researchers and engineers to design and tune VM solutions. This study was based on two consecutive industrial case studies, which were conducted on a machine tool present on the shopfloor of a factory. The main goal was to estimate the dimensional quality of test parts with the greatest accuracy possible.
During the first case study, more than 500 parts were milled and measured in an industrial environment while sensors extracted spindle power consumption, spindle torque, and x-axis torque. More than 5,000 examples were produced and analyzed using machine learning tools. The main focus of this case study was on identifying the best group of variables to measure to reach the optimal model accuracy. Dimensional quality was successfully modeled based on spindle power measurement in an industrial environment. The most useful information was extracted exclusively from a high-accuracy external power sensor. The extracted NC data were not useful, as they did not improve the model accuracy. The optimal model achieved a 19-µm MAE, which was a highly encouraging first start for VM. The highest-performing models were nonlinear, which validated the nonlinear specific force coefficient hypothesis. Several aspects could be enhanced to further increase the accuracy, such as improving the test part model and increasing the number of examples or experiments. Such changes were applied in the second case study to improve the model accuracy. Due to the test part improvement, a major issue emerged that was previously hidden, namely the input time series synchronization. This case study demonstrated the critical importance of data synchronization in VM for machine tools. A multitude of synchronization approaches were compared to identify the optimal approach for synchronizing time series and extracting information from the transient parts. To compare those approaches, we milled more than 600 parts on an industrial shopfloor, which this time were separated into three datasets, while sensors extracted spindle power consumption data. Over 6,000 measurements were taken and analyzed using machine learning tools, including the CNN, partial least squares, multiple linear regression, and multilayer perceptron approaches. Multiple data synchronization methods were studied for VM on milling machines, including curve registration, moving windows, wavelet transform, and data quantization. Dimensional quality was successfully modeled, achieving a 14.4-µm MAE in an industrial environment, which is an improvement on the 23.8% achieved in the first case study. The best performing modeling and synchronization algorithm was a CNN without specific data augmentation and no external data synchronization. Several aspects of this approach could be enhanced to further decrease the MAE, including researching new external sensors and increasing the number of example parts manufactured. Furthermore, the applicability of this VM approach might be augmented by examining the effects of multiple measurement passes or by spreading experiments out over time to verify the model's robustness in the face of industrial concept drift (CD). The first research question can thus be answered as follows: Yes, it is possible to apply VM to milling machines for the estimation of dimensional quality with an accuracy of at least 14.4 µm.

"How to maintain industrial data-based model?"

To answer the second research question, a methodology to ensure the maintenance of industrial data-based models were developed. The main objective was to preserve the accuracy of industrial data-based models. This approach is one of the research findings, as it changes the definition of CDs themselves. A new drift type classification was proposed, which considers the impact of drift on data-based model accuracy. A new causal approach for detecting the type of CD was also proposed. The major finding of this research was the methodology, which starts with a drift identification phase. Unfortunately, very few papers have focused on the detection of CD characteristics. Once the CDs had been identified, a solution pre-selection began. Based on prior knowledge, the solution that seemed to best match the identified CD were considered. Even though many drift detection approaches exist, very few researchers have estimated the type of CD that their respective methods work best on. Then, once several solutions had been pre-selected, their performance was assessed on a dataset from the shopfloor. The two metrics that were used were model accuracy and the number of measurements or updates of the model required to obtain said accuracy. Based on those metrics, the optimal solution was selected through optimization based on predefined measurement cost and bad estimation cost. The framework describes all the elements of the solutions, such as the SDS, which determines when new examples should be measured, and an updating system that describes how to update the data-based model. For all those elements, the various existing mechanisms were detailed and discussed in Chapter 4. Their limitations as well as advantages were examined in terms of the function of CD characteristics. A simulation was proposed to illustrate the operation of the methodology.
Various literature gaps were found. For instance, the description of quantitative CD is a new and underdeveloped way of describing

drift geometry. Furthermore, CD identification tools are lacking in literature. Moreover, the sampling frequency setting for a passive strategy is currently set heuristically; an optimized approach would be a significant improvement. Finally, the links between CD detection algorithms and CD characteristics have not been sufficiently discussed in the state of the art. The methodology enables to have a clear overview of the practical implementation of such a solution. It answers the second research question.

As described, there are numerous ways to pursue this research. In these paragraphs the two main opportunities for further investigation are discussed in more details. To begin with, the limitations of accuracy of VM, as observed in the present thesis (14.4 µm) could be enhanced by investigating the use of one additional sensor. The actual accuracy of (14.4 µm) may be too big for precision milling applications. To generalize the industrialization of VM on machine tool, the accuracy needs to be increased. To do so, one would need to install a new sensor on the machine tool. The actual research indicates that for power-based VM, a part of the error of estimation comes from the motor efficiency variation. Therefore, the measurement needs to be done closer to the tool. This motivates the use of an additional force sensor. Depending on the configuration of the case study, it could be implemented on the part fixture or on the tool-holder. Anyhow, the case study should be done in an industrial environment to ensure an industrializable solution. Indeed, it is hardly possible to mimic the activity of a shopfloor in a laboratory. Multiple case study should be done, first starting with a single pass milling experiment, like the one performed in this research. It should characterize the performance of dimensional quality modelization based on the force sensor. Second, the process of VM should be adapted to the targeted product. To do so, the experiment could focus on multi passes milling, curve milling or different type of product quality such as geometrical tolerance or surface condition. It needs to be highlighted that, a VM model would be needed for every single quotation. Indeed, transfer learning would be important for industrial model development. Some solutions for VM model management have been discussed in Chapter 2. This second case study should be repeated, to train the initial models, every time VM would need to be implemented. Speaking of which, a normalized process would need to be defined in every industry needing to implement VM. In all the case studies, one would want to do three sub-experiments to build independent training, validation, and testing sets. The sub-experiments should not be spaced too much in time, to avoid CD. Indeed, CD would unnecessarily reduce the accuracy of the models, and CD handling should be managed in a later phase of development, by different algorithms. The second direction of further research would focus on CD handling in industrial environment. It is the logical follow-up for industrial implementation. However, as the case studies and the algorithms used are different, the first and second direction of research discussed in this paragraph should be considered as independent. This reflection should also be driven when defining the different positions needed in industry for the implementation of industrial data-based model. Indeed, developing a model and making it robust to CD are two interlinked but different processes which require different know-how. The present research proposes a methodology to develop CD handling solutions. This methodology is based on the characterization of CD and the link between the results of this characterization and the solutions to deal with CD. The characterization of CD could be enhanced theoretically. The link between CD characterization and the solutions to deal with CD can be done practically. To do so, first thanks to a simulation, then thanks to industrial case study, datasets with clear different type of CD needs to be produced. Then, the performance of numerous CD handling solutions needs to be assessed on the datasets. The different elements of CD handling could be compared independently. However, the different elements interact together, motivating a final full solution comparison. Doing so, guidelines could be produced to guide the choice of CD handling solutions. In a broader context, the proposed methodology would need industrial implantation to mature.

In addition to the research questions derived conclusions, some other topics were also highly promising. First, an adaptation of the ISO norm 22514-7 was proposed in Chapter 2. This would be a first step toward the inclusion of the VM system in the norm. This topic is fundamental, as the requirement definition for acquiring a new measurement system is based on this norm. Therefore, for the correct industrial implementation of virtual systems, the norm should be updated. Second, the link between VM and predictive maintenance is another under-researched topic. Applicative studies that demonstrate the benefit of this interaction would be of great interest. Finally, to conclude this thesis with forecasting, once VM achieves the correct estimation of product quality, as proposed in the zero-defect manufacturing (ZDM) paradigm, VM would be able to predict quality issues in advance. In this configuration, shopfloors may achieve the goal of ZDM.

This work is hoped to be of great use to you.

References

[1] P. Dreyfus and D. Kyritsis, “A framework based on predictive maintenance , zero-defect manufacturing and scheduling under uncertainty tools , to optimize production capacities of high-end quality products,” pp. 1–11.

[2] F. Psarommatis and D. Kiritsis, *A scheduling tool for achieving zero defect manufacturing (ZDM): A conceptual framework*, vol. 536. Springer International Publishing, 2018.

[3] E. Kiraci, P. Franciosa, J. A. Warnett, A. Attridge, and M. A. Williams, “Non-reproducible alignment and fitting algorithm effects on Laser Radar measurement,” *I2MTC 2017 - 2017 IEEE Int. Instrum. Meas. Technol. Conf. Proc.*, 2017, doi: 10.1109/I2MTC.2017.7969967.

[4] S. Ferretti, D. Caputo, M. Penza, and D. M. D’Addona, “Monitoring systems for zero defect manufacturing,” *Procedia CIRP*, vol. 12, pp. 258–263, 2013, doi: 10.1016/j.procir.2013.09.045.

[5] L. Zhou, S. M. Gupta, Y. Kinoshita, and T. Yamada, “Pricing Decision Models for Remanufactured Short-Life Cycle Technology Products with Generation Consideration,” *Procedia CIRP*, vol. 61, pp. 195–200, 2017, doi: 10.1016/j.procir.2016.11.208.

[6] F. Psarommatis, G. May, P. A. Dreyfus, and D. Kiritsis, “Zero defect manufacturing: state-of-the-art review, shortcomings and future directions in research,” *International Journal of Production Research*. 2020, doi: 10.1080/00207543.2019.1605228.

[7] Y. Chun and B. Bidanda, “Sustainable manufacturing and the role of the International Journal of Production Research,” *Int. J. Prod. Res.*, vol. 51, no. 23–24, pp. 7448–7455, 2013, doi: 10.1080/00207543.2012.762135.

[8] R. R. Inman, D. E. Blumenfeld, N. Huang, and J. Li, “Designing production systems for quality: Research opportunities from an automotive industry perspective,” *Int. J. Prod. Res.*, vol. 41, no. 9, pp. 1953–1971, 2003, doi: 10.1080/0020754031000077293.

[9] S. J. Cheah, A. S. Amirul, and F. Md. Taib, “Tracking hidden quality costs in a manufacturing company: An action research,” *Int. J. Qual. Reliab. Manag.*, vol. 28, no. 4, pp. 405–425, 2011, doi: 10.1108/02656711111121816.

[10] J. Tannock and S. Saelem, “Manufacturing disruption costs due to quality loss,” *Int. J. Qual. Reliab. Manag.*, vol. 24, no. 3, pp. 263–278, 2007, doi: 10.1108/02656710710730861.

[11] F. Psarommatis, A. Gharaei, and D. Kiritsis, “Identification of the critical reaction times for re-scheduling flexible job shops for different types of unexpected events,” *Procedia CIRP*, vol. 93, pp. 903–908, Jan. 2020, doi: 10.1016/j.procir.2020.03.038.

[12] F. Psarommatis and D. Kiritsis, “Identification of the Inspection Specifications for Achieving Zero Defect Manufacturing,” in *IFIP Advances in Information and Communication Technology*, 2019, vol. 566, pp. 267–273, doi: 10.1007/978-3-030-30000-5_34.

[13] P. Chen *et al.*, “Virtual metrology: a solution for wafer to wafer advanced process control,” *IEEE Int. Symp. Semicond. Manuf.*, pp. 155–157, 2005, doi: 10.1109/ISSM.2005.1513322.

[14] Y. H. Kuo and A. Kusiak, “From data to big data in production research: the past and future trends,” *Int. J. Prod. Res.*, vol. 57, no. 15–16, pp. 4828–4853, 2019, doi: 10.1080/00207543.2018.1443230.

[15] M. O. and J. P. M.A. Jebri, E.M. El Adel, G. Graton, “Virtual Metrology applied in Run-to-Run Control for a Chemical

Mechanical Planarization process," *13th Eur. Work. Adv. Control Diagnosis (ACD 2016)*, vol. 9, no. 11, pp. 63–70, 2016, doi: 10.1088/1742-6596/755/1/011001.

[16] F. Psarommatis, S. Prouvost, G. May, and D. Kiritsis, "Product Quality Improvement Policies in Industry 4.0: Characteristics, Enabling Factors, Barriers, and Evolution Toward Zero Defect Manufacturing," *Front. Comput. Sci.*, 2020, doi: 10.3389/fcomp.2020.00026.

[17] F. Psarommatis, J. Sousa, J. P. Mendonça, and D. Kiritsis, "Zero-defect manufacturing the approach for higher manufacturing sustainability in the era of industry 4.0: a position paper," *Int. J. Prod. Res.*, pp. 0–19, 2021, doi: 10.1080/00207543.2021.1987551.

[18] K. Grevenitis *et al.*, "A hybrid framework for industrial data storage and exploitation," in *Procedia CIRP*, Jan. 2019, vol. 81, pp. 892–897, doi: 10.1016/j.procir.2019.03.221.

[19] M. Vallejo, C. De La Espriella, J. Gómez-Santamaría, A. F. Ramírez-Barrera, and E. Delgado-Trejos, "Soft metrology based on machine learning: A review," *Meas. Sci. Technol.*, vol. 31, no. 3, 2019, doi: 10.1088/1361-6501/ab4b39.

[20] F. Psarommatis, "A generic methodology and a digital twin for zero defect manufacturing (ZDM) performance mapping towards design for ZDM," *J. Manuf. Syst.*, vol. 59, no. November 2020, pp. 507–521, 2021, doi: 10.1016/j.jmsy.2021.03.021.

[21] F. Cheng *et al.*, "Benefit Model of Virtual Metrology and Integrating AVM into MES," vol. 24, no. 2, pp. 261–272, 2011.

[22] M. Michl, "IMPLEMENTING VIRTUAL METROLOGY INTO SEMICONDUCTOR PRODUCTION," *Proc. 2011 Winter Simul. Conf.*, no. IDM, pp. 2017–2028, 2011, doi: 10.1109/WSC.2011.6147915.

[23] B. Kitchenham, "Procedures for performing systematic reviews," *Keele Univ. Natl. ICT Aust.*, 2004, doi: 10.1.1.122.3308.

[24] M. Rashid, M. W. Anwar, and A. M. Khan, "Toward the tools selection in model based system engineering for embedded systems - A systematic literature review," *J. Syst. Softw.*, 2015, doi: 10.1016/j.jss.2015.04.089.

[25] P. A. Dreyfus and D. Kyritsis, "A framework based on predictive maintenance, zero-defect manufacturing and scheduling under uncertainty tools, to optimize production capacities of high-end quality products," *IFIP Adv. Inf. Commun. Technol.*, vol. 536, pp. 296–303, 2018, doi: 10.1007/978-3-319-99707-0_37.

[26] G. Roeder *et al.*, "Framework for integration of virtual metrology and predictive maintenance," *ASMC (Advanced Semicond. Manuf. Conf. Proc.*, pp. 288–293, 2012, doi: 10.1109/ASMC.2012.6212913.

[27] Y. S. Hsieh, F. T. Cheng, H. C. Huang, C. R. Wang, S. C. Wang, and H. C. Yang, "VM-based baseline predictive maintenance scheme," *IEEE Trans. Semicond. Manuf.*, 2013, doi: 10.1109/TSM.2012.2218837.

[28] Y. S. Hsieh, F. T. Cheng, and H. C. Yang, "Virtual-metrology-based FDC scheme," 2012, doi: 10.1109/CoASE.2012.6386371.

[29] J. Moyne, M. Yedatore, J. Iskandar, P. Hawkins, and J. Scoville, "Chamber matching across multiple dimensions utilizing Predictive Maintenance, Equipment Health Monitoring, Virtual Metrology and Run-To-Run control," *ASMC (Advanced Semicond. Manuf. Conf. Proc.*, pp. 86–91, 2014, doi: 10.1109/ASMC.2014.6846983.

[30] C.-H. Chen, W.-D. Zhao, T. Pang, and Y.-Z. Lin, "Virtual metrology of semiconductor PVD process based on combination of tree-based ensemble model.," *ISA Trans.*, vol. 103, pp. 192–202, Aug. 2020, doi: 10.1016/j.isatra.2020.03.031.

[31] Y. Huang, F. Cheng, Y. Shih, and Y. Chen, "Advanced ART2 scheme for enhancing metrology- data-quality evaluation," *J. Chinese Inst. Eng.*, vol. 37, no. 8, pp. 1065–1080, 2014, doi: 10.1080/02533839.2014.912773.

[32] F. T. Cheng *et al.*, "Industry 4.1 for Wheel Machining Automation," *IEEE Robot. Autom. Lett.*, vol. 1, no. 1, pp. 332–339, 2016, doi: 10.1109/LRA.2016.2517208.

[33] P. H. Chou, M. J. Wu, and K. K. Chen, "Integrating support vector machine and genetic algorithm to implement dynamic wafer quality prediction system," *Expert Syst. Appl.*, vol. 37, no. 6, pp. 4413–4424, 2010, doi: 10.1016/j.eswa.2009.11.087.

[34] P. Kang, D. Kim, and S. Cho, "Evaluating the reliability level of virtual metrology results for flexible process control: a

novelty detection-based approach," *Pattern Anal. Appl.*, vol. 17, no. 4, pp. 863–881, 2014, doi: 10.1007/s10044-014-0386-6.

[35] D. Kim, P. Kang, S. kyung Lee, S. Kang, S. Doh, and S. Cho, "Improvement of virtual metrology performance by removing metrology noises in a training dataset," *Pattern Anal. Appl.*, 2015, doi: 10.1007/s10044-013-0363-5.

[36] E. R. Faria, I. J. C. R. Gonçalves, A. C. P. L. F. de Carvalho, and J. Gama, "Novelty detection in data streams," *Artif. Intell. Rev.*, vol. 45, no. 2, pp. 235–269, 2016, doi: 10.1007/s10462-015-9444-8.

[37] M. Kuhn and K. Johnson, *Applied predictive modeling*. 2013.

[38] D. Kim and S. Kang, "Effect of irrelevant variables on faulty wafer detection in semiconductor manufacturing," *Energies*, 2019, doi: 10.3390/en12132530.

[39] P. Kang *et al.*, "A virtual metrology system for semiconductor manufacturing," *Expert Syst. Appl.*, vol. 36, no. 10, pp. 12554–12561, 2009, doi: 10.1016/j.eswa.2009.05.053.

[40] P. Kang, D. Kim, H. J. Lee, S. Doh, and S. Cho, "Virtual metrology for run-to-run control in semiconductor manufacturing," *Expert Syst. Appl.*, vol. 38, no. 3, pp. 2508–2522, 2011, doi: 10.1016/j.eswa.2010.08.040.

[41] A. A. Ul Haq and D. Djurdjanovic, "Dynamics-inspired feature extraction in semiconductor manufacturing processes," *J. Ind. Inf. Integr.*, 2019, doi: 10.1016/j.jii.2018.12.001.

[42] C. C. Chen *et al.*, "A Novel Efficient Big Data Processing Scheme for Feature Extraction in Electrical Discharge Machining," *IEEE Robot. Autom. Lett.*, vol. 4, no. 2, pp. 910–917, 2019, doi: 10.1109/LRA.2019.2891498.

[43] T. Hirai and M. Kano, "Adaptive virtual metrology design for semiconductor dry etching process through locally weighted partial least squares," *IEEE Trans. Semicond. Manuf.*, vol. 28, no. 2, pp. 137–144, 2015, doi: 10.1109/TSM.2015.2409299.

[44] S. Tilouche, S. Bassetto, and V. P. Nia, "Classification algorithms for virtual metrology," 2014, doi: 10.1109/ICMIT.2014.6942477.

[45] D. Kim, P. Kang, S. Cho, H. J. Lee, and S. Doh, "Machine learning-based novelty detection for faulty wafer detection in semiconductor manufacturing," *Expert Syst. Appl.*, 2012, doi: 10.1016/j.eswa.2011.09.088.

[46] S. Bai, J. Z. Kolter, and V. Koltun, "An Empirical Evaluation of Generic Convolutional and Recurrent Networks for Sequence Modeling," 2018, [Online]. Available: http://arxiv.org/abs/1803.01271.

[47] A. Vaswani *et al.*, "Attention is all you need," 2017.

[48] C. H. Chen, P. Parashar, C. Akbar, S. M. Fu, M. Y. Syu, and A. Lin, "Physics-Prior Bayesian Neural Networks in Semiconductor Processing," *IEEE Access*, 2019, doi: 10.1109/ACCESS.2019.2940130.

[49] D. Kim, J. Koo, H. Kim, S. Kang, S. H. Lee, and J. T. Kang, "Rapid fault cause identification in surface mount technology processes based on factory-wide data analysis," *Int. J. Distrib. Sens. Networks*, 2019, doi: 10.1177/1550147719832802.

[50] K. Bum, L. Chang, and O. Kim, "Recurrent feature-incorporated convolutional neural network for virtual metrology of the chemical mechanical planarization process," *J. Intell. Manuf.*, vol. 31, no. 1, pp. 73–86, 2020, doi: 10.1007/s10845-018-1437-4.

[51] X. Jia, Y. Di, J. Feng, Q. Yang, H. Dai, and J. Lee, "Adaptive virtual metrology for semiconductor chemical mechanical planarization process using GMDH-type polynomial neural networks," *J. Process Control*, vol. 62, pp. 44–54, 2018, doi: 10.1016/j.jprocont.2017.12.004.

[52] H. Yang, M. Adnan, C. Huang, F. Cheng, Y. Lo, and C. Hsu, "An Intelligent Metrology Architecture With AVM for Metal Additive Manufacturing," *IEEE Robot. Autom. Lett.*, vol. 4, no. 3, pp. 2886–2893, 2019, doi: 10.1109/LRA.2019.2921927.

[53] L. Lik, T. Chan, X. Wu, J. Chen, L. Xie, and C. Chen, "Just-In-Time Modeling With Variable Shrinkage Based on Gaussian Processes for Semiconductor Manufacturing," vol. 31, no. 3, pp. 335–342, 2018.

[54] X. Wu, J. Chen, L. Xie, L. L. T. Chan, and C. I. Chen, "Development of convolutional neural network based Gaussian process regression to construct a novel probabilistic virtual metrology in multi-stage semiconductor processes," *Control Eng. Pract.*, 2020, doi: 10.1016/j.conengprac.2019.104262.

[55] S. Džeroski and B. Ženko, "Is combining classifiers with stacking better than selecting the best one?," *Mach. Learn.*, 2004, doi: 10.1023/B:MACH.0000015881.36452.6e.

[56] Y.-M. Hsieh, R. Lu, J.-W. Lu, F.-T. Cheng, and M. Adnan, "Automated Classification Scheme plus AVM for Wafer Sawing Processes," *IEEE Robot. Autom. Lett.*, 2020, doi: 10.1109/lra.2020.3000678.

[57] Z. Li, D. Wu, and T. Yu, "Prediction of Material Removal Rate for Chemical Mechanical Planarization Using Decision Tree-Based Ensemble Learning," *J. Manuf. Sci. Eng. Trans. ASME*, 2019, doi: 10.1115/1.4042051.

[58] G. Widmer, "Learning in the presence of concept drift and hidden contexts," *Mach. Learn.*, 1996, doi: 10.1007/bf00116900.

[59] C. Y. C. Jonathan and F. T. Cheng, "Manufacturability of multivariate applications in the semiconductor industry," *2006 IEEE Int. Conf. Autom. Sci. Eng. CASE*, no. 2, pp. 230–235, 2007, doi: 10.1109/COASE.2006.326885.

[60] M. Azamfar, X. Li, and J. Lee, "Deep Learning-Based Domain Adaptation Method for Fault Diagnosis in Semiconductor Manufacturing," *IEEE Trans. Semicond. Manuf.*, 2020, doi: 10.1109/TSM.2020.2995548.

[61] W. T. Yang, J. Blue, A. Roussy, J. Pinaton, and M. S. Reis, "A physics-informed Run-to-Run control framework for semiconductor manufacturing," *Expert Syst. Appl.*, 2020, doi: 10.1016/j.eswa.2020.113424.

[62] S. Kim, J. Jang, and C. O. Kim, "A run-to-run controller for a chemical mechanical planarization process using least squares generative adversarial networks," *J. Intell. Manuf.*, 2020, doi: 10.1007/s10845-020-01639-1.

[63] F. T. Cheng, Y. T. Chen, Y. C. Su, and D. L. Zeng, "Evaluating reliance level of a virtual metrology system," *IEEE Trans. Semicond. Manuf.*, vol. 21, no. 1, pp. 92–102, 2008, doi: 10.1109/TSM.2007.914373.

[64] H. Cai, J. Feng, Q. Yang, W. Li, X. Li, and J. Lee, "A virtual metrology method with prediction uncertainty based on Gaussian process for chemical mechanical planarization," *Comput. Ind.*, 2020, doi: 10.1016/j.compind.2020.103228.

[65] B. Lakshminarayanan, A. Pritzel, and C. Blundell, "Simple and scalable predictive uncertainty estimation using deep ensembles," 2017.

[66] J. Caldeira and B. Nord, "Deeply Uncertain: Comparing Methods of Uncertainty Quantification in Deep Learning Algorithms," pp. 1–11, 2020, [Online]. Available: http://arxiv.org/abs/2004.10710.

[67] H. M. D. Kabir, A. Khosravi, M. A. Hosen, and S. Nahavandi, "Neural Network-Based Uncertainty Quantification: A Survey of Methodologies and Applications," *IEEE Access*, vol. 6, pp. 36218–36234, 2018, doi: 10.1109/ACCESS.2018.2836917.

[68] K. Wadewale *et al.*, "Survey on Method of Drift Detection and Classification for time varying data set," *Comput. Biol. Med.*, 2016, doi: 10.1016/j.compeleceng.2013.11.024.

[69] F. T. Cheng, Y. S. Hsieh, C. F. Chen, and J. R. Lyu, "Automated sampling decision scheme for the AVM system," *Int. J. Prod. Res.*, vol. 54, no. 21, pp. 6351–6366, 2016, doi: 10.1080/00207543.2015.1072649.

[70] S. Kang and P. Kang, "An intelligent virtual metrology system with adaptive update for semiconductor manufacturing," *J. Process Control*, vol. 52, pp. 66–74, 2017, doi: 10.1016/j.jprocont.2017.02.002.

[71] D. D. Lewis and W. A. Gale, "A sequential algorithm for training text classifiers," *Proc. 17th Annu. Int. ACM SIGIR Conf. Res. Dev. Inf. Retrieval, SIGIR 1994*, pp. 3–12, 1994, doi: 10.1007/978-1-4471-2099-5_1.

[72] D. Kurz, C. De Luca, and J. Pilz, "A sampling decision system for virtual metrology in semiconductor manufacturing," *IEEE Trans. Autom. Sci. Eng.*, vol. 12, no. 1, pp. 75–83, 2015, doi: 10.1109/TASE.2014.2360214.

[73] G. A. Susto, "A dynamic sampling strategy based on confidence level of virtual metrology predictions," 2017, doi: 10.1109/asmc.2017.7969203.

[74] J. Wan, B. Honari, and S. McLoone, "A dynamic sampling methodology for plasma etch processes using Gaussian process regression," *2013 24th Int. Conf. Information, Commun. Autom. Technol. ICAT 2013*, no. Vm, pp. 1–6, 2013, doi: 10.1109/ICAT.2013.6684080.

[75] M. A. Jebri, E. M. El Adel, G. Graton, M. Ouladsine, and J. Pinaton, "The Impact of the Virtual Metrology on a Run-to-Run control for a Chemical Mechanical Planarization process," *IFAC-PapersOnLine*, vol. 50, no. 1, pp. 6154–6159, 2017, doi: 10.1016/j.ifacol.2017.08.980.

[76] F. T. Cheng, C. F. Chen, Y. S. Hsieh, H. H. Huang, and C. C. Wu, "Intelligent sampling decision scheme based on the AVM system," *Int. J. Prod. Res.*, vol. 53, no. 7, pp. 2073–2088, 2015, doi: 10.1080/00207543.2014.955924.

[77] K. Hyun Baek, K. Song, C. Han, G. Choi, H. Ku Cho, and T. F. Edgar, "Implementation of a robust virtual metrology for plasma etching through effective variable selection and recursive update technology," *J. Vac. Sci. Technol. B, Nanotechnol. Microelectron. Mater. Process. Meas. Phenom.*, vol. 32, no. 1, p. 012203, 2014, doi: 10.1116/1.4862254.

[78] G. Jaber, "An approach for online learning in the presence of concept change," p. 213, 2013.

[79] A. Urhan and B. Alakent, "Integrating adaptive moving window and just-in-time learning paradigms for soft-sensor design," *Neurocomputing*, vol. 392, pp. 23–37, 2020, doi: 10.1016/j.neucom.2020.01.083.

[80] J. V Ringwood and N. Macgearailt, "Weighted windowed PLS models for virtual metrology of an industrial plasma etch process," pp. 309–314, 2010.

[81] W. M. Wu, F. T. Cheng, and F. W. Kong, "Dynamic-moving-window scheme for virtual-metrology model refreshing," *IEEE Trans. Semicond. Manuf.*, 2012, doi: 10.1109/TSM.2012.2183398.

[82] S. kyung Lee, P. Kang, and S. Cho, "Probabilistic local reconstruction for k-NN regression and its application to virtual metrology in semiconductor manufacturing," *Neurocomputing*, 2014, doi: 10.1016/j.neucom.2013.10.001.

[83] C. Cheng and M. Sen Chiu, "A new data-based methodology for nonlinear process modeling," *Chem. Eng. Sci.*, 2004, doi: 10.1016/j.ces.2004.04.020.

[84] J. Iskandar and J. Moyne, "Maintenance of Virtual Metrology Models," *2016 27th Annu. SEMI Adv. Semicond. Manuf. Conf.*, pp. 393–398, 2016, doi: 10.1109/ASMC.2016.7491083.

[85] Yaw-Jen Chang, Yuan Kang, Chih-Liang Hsu, Chi-Tim Chang, and Tat Yan Chan, "Virtual Metrology Technique for Semiconductor Manufacturing," *2006 IEEE Int. Jt. Conf. Neural Netw. Proc.*, pp. 5289–5293, 2006, doi: 10.1109/IJCNN.2006.247284.

[86] Y. J. Chang, "Wavelet-based virtual metrology technique," *2010 IEEE Int. Conf. Mechatronics Autom. ICMA 2010*, pp. 367–371, 2010, doi: 10.1109/ICMA.2010.5588327.

[87] K. Y. Lin, C. Y. Hsu, and H. C. Yu, "A virtual metrology approach for maintenance compensation to improve yield in semiconductor manufacturing," *Int. J. Comput. Intell. Syst.*, 2014, doi: 10.1080/18756891.2014.947116.

[88] S. A. Lynn, J. Ringwood, S. Member, and N. Macgearailt, "Global and Local Virtual Metrology Models for a Plasma Etch Process," *IEEE Trans. Semicond. Manuf.*, vol. 25, no. 1, pp. 94–103, 2012, doi: 10.1109/TSM.2011.2176759.

[89] S. Park *et al.*, "Predictive control of the plasma processes in the OLED display mass production referring to the discontinuity qualifying PI-VM," *Phys. Plasmas*, 2020, doi: 10.1063/1.5135312.

[90] G. Ditzler, M. Roveri, C. Alippi, and R. Polikar, "Learning in Nonstationary Environments: A Survey," *IEEE Comput. Intell. Mag.*, vol. 10, no. 4, pp. 12–25, 2015, doi: 10.1109/MCI.2015.2471196.

[91] H. Tieng, C. F. Chen, F. T. Cheng, and H. C. Yang, "Automatic Virtual Metrology and Target Value Adjustment for Mass Customization," *IEEE Robot. Autom. Lett.*, vol. 2, no. 2, pp. 546–553, 2017, doi: 10.1109/LRA.2016.2645507.

[92] T. Pan, B. Sheng, D. S. Wong, and S. Jang, "A Virtual Metrology System for Predicting End-of-Line Electrical Properties Using a MANCOVA Model With Tools Clustering," vol. 7, no. 2, pp. 187–195, 2011.

[93] S. Kang, “On Effectiveness of Transfer Learning Approach for Neural Network-Based Virtual Metrology Modeling,” *IEEE Trans. Semicond. Manuf.*, 2018, doi: 10.1109/TSM.2017.2787550.

[94] C. Park, Y. Kim, Y. Park, and S. B. Kim, “Multitask learning for virtual metrology in semiconductor manufacturing systems,” *Comput. Ind. Eng.*, 2018, doi: 10.1016/j.cie.2018.06.024.

[95] C. Nguyen, X. Li, R. D. S. Blanton, and X. Li, “Partial Co-training for virtual metrology,” *IEEE Int. Conf. Emerg. Technol. Fact. Autom. ETFA*, pp. 1–8, 2018, doi: 10.1109/ETFA.2017.8247660.

[96] C. M. Nguyen, X. Li, R. D. S. Blanton, and X. Li, “Partial Bayesian Co-training for Virtual Metrology,” *IEEE Trans. Ind. Informatics*, 2020, doi: 10.1109/TII.2019.2903718.

[97] K. J. Kim, K. J. Kim, C. H. Jun, I. G. Chong, and G. Y. Song, “Variable Selection under Missing Values and Unlabeled Data in Semiconductor Processes,” *IEEE Trans. Semicond. Manuf.*, 2019, doi: 10.1109/TSM.2018.2881286.

[98] Y. C. Su, W. H. Tsai, F. T. Cheng, and W. M. Wu, “Development of a dual-stage virtual metrology architecture for TFT-LCD manufacturing,” 2008, doi: 10.1109/ROBOT.2008.4543767.

[99] A. A. Khan, J. R. Moyne, and D. M. Tilbury, “An approach for factory-wide control utilizing virtual metrology,” *IEEE Trans. Semicond. Manuf.*, vol. 20, no. 4, pp. 364–375, 2007, doi: 10.1109/TSM.2007.907609.

[100] G. A. Susto, S. Pampuri, A. Schirru, A. Beghi, and G. De Nicolao, “Multi-step virtual metrology for semiconductor manufacturing: A multilevel and regularization methods-based approach,” *Comput. Oper. Res.*, 2015, doi: 10.1016/j.cor.2014.05.008.

[101] A. Schirru, S. Pampuri, C. De Luca, and G. De Nicolao, “Multilevel kernel methods for Virtual Metrology in semiconductor manufacturing,” 2011, doi: 10.3182/20110828-6-IT-1002.01339.

[102] J. Moyne, E. Del Castillo, and A. M. Hurwitz, *Run-to-run control in semiconductor manufacturing*. 2000.

[103] C.-H. Lin, S.-T. Tseng, and H.-F. Wang, “Modified EWMA controller subject to metrology delay,” *IIE Trans.*, vol. 45, no. 4, pp. 409–421, 2013, doi: 10.1080/0740817X.2012.689242.

[104] M. Khakifirooz, M. Fathi, and C. F. Chien, “Modelling and decision support system for intelligent manufacturing: An empirical study for feedforward-feedback learning-based run-to-run controller for semiconductor dry-etching process,” *Int. J. Ind. Eng. Theory Appl. Pract.*, 2018.

[105] S. A. Lynn, N. Macgearailt, and J. V Ringwood, “Real-time virtual metrology and control for plasma etch,” *J. Process Control*, vol. 22, no. 4, pp. 666–676, 2012, doi: 10.1016/j.jprocont.2012.01.012.

[106] J. V Ringwood *et al.*, “Estimation and Control in Semiconductor Etch : Practice and Possibilities,” vol. 23, no. 1, pp. 87–98, 2010.

[107] H. C. Huang, Y. C. Su, F. T. Cheng, and J. M. Jian, “Development of a generic virtual metrology framework,” *Proc. 3rd IEEE Int. Conf. Autom. Sci. Eng. IEEE CASE 2007*, pp. 282–287, 2007, doi: 10.1109/COASE.2007.4341746.

[108] M. H. Hung, Y. Y. Li, Y. C. Lin, C. F. Wei, H. C. Yang, and F. T. Cheng, “Development of a novel cloud-based multi-tenant model creation service for automatic virtual metrology,” *Robot. Comput. Integr. Manuf.*, 2017, doi: 10.1016/j.rcim.2016.09.003.

[109] M. H. Hung, W. H. Tsai, H. C. Yang, Y. J. Kao, and F. T. Cheng, “A novel automatic virtual metrology system architecture for TFT-LCD industry based on main memory database,” *Robot. Comput. Integr. Manuf.*, 2012, doi: 10.1016/j.rcim.2012.01.002.

[110] F. T. Cheng, H. C. Huang, and C. A. Kao, “Developing an automatic virtual metrology system,” *IEEE Trans. Autom. Sci. Eng.*, vol. 9, no. 1, pp. 181–188, 2012, doi: 10.1109/TASE.2011.2169405.

[111] M. H. Hung, C. F. Chen, H. C. Huang, H. C. Yang, and F. T. Cheng, “Development of an AVM System Implementation Framework,” *IEEE Trans. Semicond. Manuf.*, vol. 25, no. 4, pp. 598–613, 2012, doi: 10.1109/TSM.2012.2206061.

[112] J. Y. Cheng, M. H. Hung, S. S. Lin, and F. T. Cheng, "New remote monitoring and control system architectures based on cloud computing," *Adv. Mater. Res.*, vol. 579, pp. 312–329, 2012, doi: 10.4028/www.scientific.net/AMR.579.312.

[113] Y. M. Hsieh, C. Y. Lin, Y. R. Yang, M. H. Hung, and F. T. Cheng, "Automatic Virtual Metrology for Carbon Fiber Manufacturing," *IEEE Robot. Autom. Lett.*, vol. 4, no. 3, pp. 2730–2737, 2019, doi: 10.1109/LRA.2019.2917384.

[114] M. H. Hung, Y. C. Lin, H. C. Huang, M. H. Hsieh, H. C. Yang, and F. T. Cheng, "Development of an advanced manufacturing cloud for machine tool industry based on AVM technology," 2013, doi: 10.1109/CoASE.2013.6654048.

[115] H. C. Huang, Y. C. Lin, M. H. Hung, C. C. Tu, and F. T. Cheng, "Development of cloud-based automatic virtual metrology system for semiconductor industry," *Robot. Comput. Integr. Manuf.*, 2015, doi: 10.1016/j.rcim.2015.01.005.

[116] J. Moyne, J. Samantaray, and M. Armacost, "Big Data Capabilities Applied to Semiconductor Manufacturing Advanced Process Control," *IEEE Trans. Semicond. Manuf.*, 2016, doi: 10.1109/TSM.2016.2574130.

[117] T. H. O, M. H. Hung, Y. C. Lin, and C. C. Chen, "Construction of Virtual Metrology Cloud Platform with Machine Learning Tools for Providing Factory-Wide Manufacturing Service," 2020, doi: 10.1007/978-981-15-3380-8_5.

[118] D. Zeng and C. J. Spanos, "Virtual metrology modeling for plasma etch operations," 2009, doi: 10.1109/TSM.2009.2031750.

[119] T. Tsutsui and T. Matsuzawa, "Virtual metrology model robustness against chamber condition variation using deep learning," *IEEE Trans. Semicond. Manuf.*, 2019, doi: 10.1109/TSM.2019.2931328.

[120] H. Purwins *et al.*, "Regression methods for virtual metrology of layer thickness in chemical vapor deposition," *IEEE/ASME Trans. Mechatronics*, vol. 19, no. 1, pp. 1–8, 2014, doi: 10.1109/TMECH.2013.2273435.

[121] M. Kim, S. Kang, J. Lee, H. Cho, S. Cho, and J. S. Park, "Virtual metrology for copper-clad laminate manufacturing," *Comput. Ind. Eng.*, vol. 109, pp. 280–287, 2017, doi: 10.1016/j.cie.2017.04.016.

[122] D. Hou, T. Liu, Y. T. Pan, and J. Hou, "AI on edge device for laser chip defect detection," 2019, doi: 10.1109/CCWC.2019.8666503.

[123] X. Jin, H. Shui, and M. Shpitalni, "Virtual sensing and virtual metrology for spatial error monitoring of roll-to-roll manufacturing systems," *CIRP Ann.*, 2019, doi: 10.1016/j.cirp.2019.04.108.

[124] H. Shui, X. Jin, and J. Ni, "Twofold Variation Propagation Modeling and Analysis for Roll-to-Roll Manufacturing Systems," *IEEE Trans. Autom. Sci. Eng.*, 2019, doi: 10.1109/TASE.2018.2844168.

[125] H. C. Yang, H. Tieng, and F. T. Cheng, "Total precision inspection of machine tools with virtual metrology," *J. Chinese Inst. Eng. Trans. Chinese Inst. Eng. A/Chung-kuo K. Ch'eng Hsuch K'an*, vol. 39, no. 2, pp. 221–235, 2016, doi: 10.1080/02533839.2015.1091279.

[126] D. Rolnick, A. Veit, S. Belongie, and N. Shavit, "Deep Learning is Robust to Massive Label Noise," 2017, [Online]. Available: http://arxiv.org/abs/1705.10694.

[127] H. Tieng, T. H. Tsai, C. F. Chen, H. C. Yang, J. W. Huang, and F. T. Cheng, "Automatic Virtual Metrology and Deformation Fusion Scheme for Engine-Case Manufacturing," *IEEE Robot. Autom. Lett.*, vol. 3, no. 2, pp. 934–941, 2018, doi: 10.1109/LRA.2018.2792690.

[128] D. Huang and J. Lv, "Run-to-run control of batch production process in manufacturing systems based on online measurement," *Comput. Ind. Eng.*, 2020, doi: 10.1016/j.cie.2020.106298.

[129] M. A. Rubeo and T. L. Schmitz, "Mechanistic force model coefficients: A comparison of linear regression and nonlinear optimization," *Precis. Eng.*, vol. 45, pp. 311–321, 2016, doi: 10.1016/j.precisioneng.2016.03.008.

[130] P. A. Dreyfus, F. Psarommatis, G. May, and D. Kiritsis, "Virtual metrology as an approach for product quality estimation in Industry 4.0: A systematic review and integrative conceptual framework," *Int. J. Prod. Res.*, 2021.

[131] J. Choi and M. K. Jeong, "Deep autoencoder with clipping fusion regularization on multistep process signals for virtual metrology," *IEEE Sensors Lett.*, 2019, doi: 10.1109/LSENS.2018.2884735.

[132] M. Maggipinto, A. Beghi, S. McLoone, and G. A. Susto, "DeepVM: A Deep Learning-based approach with automatic feature extraction for 2D input data Virtual Metrology," *J. Process Control*, 2019, doi: 10.1016/j.jprocont.2019.08.006.

[133] P. A. Dreyfus, F. Psarommatis, G. May, and D. Kiritsis, "Virtual metrology as an approach for product quality estimation: A systematic review," *Int. J. Prod. Econ.*, 2021.

[134] X. Yuan, S. Qi, Y. Shardt, Y. Wang, C. Yang, and W. Gui, "Soft sensor model for dynamic processes based on multichannel convolutional neural network," *Chemom. Intell. Lab. Syst.*, vol. 203, no. January, p. 104050, 2020, doi: 10.1016/j.chemolab.2020.104050.

[135] M. Maggipinto, M. Terzi, C. Masiero, A. Beghi, and G. A. Susto, "A Computer Vision-Inspired Deep Learning Architecture for Virtual Metrology Modeling with 2-Dimensional Data," *IEEE Trans. Semicond. Manuf.*, vol. 31, no. 3, pp. 376–384, 2018, doi: 10.1109/TSM.2018.2849206.

[136] International Organization for Standardization (ISO), "Statistical methods in process management — Capability and performance (ISO/DIS Standard No. 22514)," 2012, [Online]. Available: http://www.iso.org/iso/catalogue_detail?csnumber=63787.

[137] V. Brunner, M. Siegl, D. Geier, and T. Becker, "Challenges in the Development of Soft Sensors for Bioprocesses: A Critical Review," *Front. Bioeng. Biotechnol.*, vol. 9, no. August, pp. 1–21, 2021, doi: 10.3389/fbioe.2021.722202.

[138] D. Jeong, C. Park, and Y. M. Ko, "Short-term electric load forecasting for buildings using logistic mixture vector autoregressive model with curve registration," *Appl. Energy*, vol. 282, no. PB, p. 116249, 2021, doi: 10.1016/j.apenergy.2020.116249.

[139] K. Qiu, J. Wang, R. Wang, Y. Guo, and L. Zhao, "Soft sensor development based on kernel dynamic time warping and a relevant vector machine for unequal-length batch processes," *Expert Syst. Appl.*, vol. 182, no. May, p. 115223, 2021, doi: 10.1016/j.eswa.2021.115223.

[140] T. Offermans, E. Szymańska, L. M. C. Buydens, and J. J. Jansen, "Synchronizing process variables in time for industrial process monitoring and control," *Comput. Chem. Eng.*, vol. 140, 2020, doi: 10.1016/j.compchemeng.2020.106938.

[141] P. Dreyfus and D. Kiritsis, "Virtual metrology for milling operations: Single-pass dimensional quality estimation," *Expert Syst. Appl.*

[142] B. Lu, J. Stuber, and T. F. Edgar, "Integrated online virtual metrology and fault detection in plasma etch tools," *Ind. Eng. Chem. Res.*, vol. 53, no. 13, pp. 5172–5181, 2014, doi: 10.1021/ie401448d.

[143] J. P. Yun, W. C. Shin, G. Koo, M. S. Kim, C. Lee, and S. J. Lee, "Automated defect inspection system for metal surfaces based on deep learning and data augmentation," *J. Manuf. Syst.*, vol. 55, no. May, pp. 317–324, 2020, doi: 10.1016/j.jmsy.2020.03.009.

[144] D. Zeng, Y. Tan, and C. J. Spanos, "Dimensionality reduction methods in virtual metrology," *Proc. SPIE*, vol. 6922, no. April, pp. 692238-692238–11, 2008, doi: 10.1117/12.772739.

[145] T. Bachlechner, B. P. Majumder, H. H. Mao, G. W. Cottrell, and J. McAuley, "ReZero is All You Need: Fast Convergence at Large Depth," no. 1, pp. 1–14, 2020, [Online]. Available: http://arxiv.org/abs/2003.04887.

[146] E. R. Faria, I. J. C. R. Gonçalves, A. C. P. L. F. de Carvalho, and J. Gama, "Novelty detection in data streams," *Artif. Intell. Rev.*, vol. 45, no. 2, pp. 235–269, 2016, doi: 10.1007/s10462-015-9444-8.

[147] P. A. Dreyfus, F. Psarommatis, G. May, and D. Kiritsis, "Virtual metrology as an approach for product quality estimation in Industry 4.0: a systematic review and integrative conceptual framework," *Int. J. Prod. Res.*, 2021, doi: 10.1080/00207543.2021.1976433.

[148] I. Zliobaite, A. Bifet, B. Pfahringer, and G. Holmes, "Active learning with drifting streaming data," *IEEE Trans. Neural Networks Learn. Syst.*, vol. 25, no. 1, pp. 27–39, 2014, doi: 10.1109/TNNLS.2012.2236570.

[149] A. Haque, L. Khan, and M. Baron, "SAND: Semi-supervised adaptive novel class detection and classification over data stream," *30th AAAI Conf. Artif. Intell. AAAI 2016*, no. May 2015, pp. 1652–1658, 2016, doi: 10.1007/978-3-319-18032-8.

[150] L. Sun, Y. Ji, M. Zhu, F. Gu, F. Dai, and K. Li, “A new predictive method supporting streaming data with hybrid recurring concept drifts in process industry,” *Comput. Ind. Eng.*, vol. 161, Nov. 2021, doi: 10.1016/j.cie.2021.107625.

[151] H. Guo, S. Zhang, and W. Wang, “Selective ensemble-based online adaptive deep neural networks for streaming data with concept drift,” *Neural Networks*, vol. 142, pp. 437–456, Oct. 2021, doi: 10.1016/j.neunet.2021.06.027.

[152] Z. Liu, C. K. Loo, K. Pasupa, and M. Seera, “Meta-cognitive recurrent kernel online sequential extreme learning machine with kernel adaptive filter for concept drift handling,” *Eng. Appl. Artif. Intell.*, vol. 88, 2020, doi: 10.1016/j.engappai.2019.103327.

[153] J. Lu, A. Liu, F. Dong, F. Gu, J. Gama, and G. Zhang, “Learning under Concept Drift: A Review,” *IEEE Transactions on Knowledge and Data Engineering*, vol. 31, no. 12. IEEE Computer Society, pp. 2346–2363, 2019, doi: 10.1109/TKDE.2018.2876857.

[154] Y. Song, J. Lu, H. Lu, and G. Zhang, “Fuzzy Clustering-Based Adaptive Regression for Drifting Data Streams,” *IEEE Trans. Fuzzy Syst.*, vol. 28, no. 3, pp. 544–557, 2020, doi: 10.1109/TFUZZ.2019.2910714.

[155] J. Gama, I. Zliobaite, A. Bifet, M. Pechenizkiy, and A. Bouchachia, “A survey on concept drift adaptation,” *ACM Computing Surveys*. 2014, doi: 10.1145/2523813.

[156] G. I. Webb, R. Hyde, H. Cao, H. L. Nguyen, and F. Petitjean, “Characterizing concept drift,” *Data Min. Knowl. Discov.*, vol. 30, no. 4, pp. 964–994, 2016, doi: 10.1007/s10618-015-0448-4.

[157] W. Yang *et al.*, “Diagnosing Concept Drift with Visual Analytics,” in *Proceedings - 2020 IEEE Conference on Visual Analytics Science and Technology, VAST 2020*, 2020, pp. 12–23, doi: 10.1109/VAST50239.2020.00007.

[158] P. M. Gonçalves, S. G. T. De Carvalho Santos, R. S. M. Barros, and D. C. L. Vieira, “A comparative study on concept drift detectors,” *Expert Systems with Applications*, vol. 41, no. 18. pp. 8144–8156, 2014, doi: 10.1016/j.eswa.2014.07.019.

[159] S. G. T. C. Santos, R. S. M. Barros, and P. M. Gonçalves, “A differential evolution based method for tuning concept drift detectors in data streams,” *Inf. Sci. (Ny).*, vol. 485, pp. 376–393, 2019, doi: 10.1016/j.ins.2019.02.031.

[160] W. Liu, H. Zhang, Z. Ding, Q. Liu, and C. Zhu, “A comprehensive active learning method for multiclass imbalanced data streams with concept drift,” *Knowledge-Based Syst.*, vol. 215, Mar. 2021, doi: 10.1016/j.knosys.2021.106778.

[161] G. I. Webb, L. K. Lee, B. Goethals, and F. Petitjean, “Analyzing concept drift and shift from sample data,” *Data Min. Knowl. Discov.*, vol. 32, no. 5, pp. 1179–1199, Sep. 2018, doi: 10.1007/s10618-018-0554-1.

[162] S. Wassermann, T. Cuvelier, P. Mulinka, and P. Casas, “Adaptive and Reinforcement Learning Approaches for Online Network Monitoring and Analysis,” *IEEE Trans. Netw. Serv. Manag.*, vol. 18, no. 2, pp. 1832–1849, 2021, doi: 10.1109/TNSM.2020.3037486.

[163] F. T. Cheng, Y. S. Hsieh, C. F. Chen, and J. R. Lyu, “Automated sampling decision scheme for the AVM system,” *Int. J. Prod. Res.*, vol. 54, no. 21, pp. 6351–6366, 2016, doi: 10.1080/00207543.2015.1072649.

[164] J. Lu, A. Liu, F. Dong, F. Gu, J. Gama, and G. Zhang, “Learning under Concept Drift: A Review,” *IEEE Transactions on Knowledge and Data Engineering*, vol. 31, no. 12. IEEE Computer Society, pp. 2346–2363, Dec. 01, 2019, doi: 10.1109/TKDE.2018.2876857.

[165] E. J. Spinosaa, A. P. de L. F. de Carvalhoa, and J. Gamab, “Novelty detection with application to data streams,” in *Intelligent Data Analysis*, 2009, vol. 13, no. 3, pp. 405–422, doi: 10.3233/IDA-2009-0373.

[166] M. Masud, J. Gao, L. Khan, J. Han, and B. M. Thuraisingham, “Classification and novel class detection in concept-drifting data streams under time constraints,” *IEEE Trans. Knowl. Data Eng.*, vol. 23, no. 6, pp. 859–874, 2011, doi: 10.1109/TKDE.2010.61.

[167] I. Khamassi, M. Sayed-Mouchaweh, M. Hammami, and K. Ghédira, “Discussion and review on evolving data streams and concept drift adapting,” *Evol. Syst.*, vol. 9, no. 1, pp. 1–23, Mar. 2018, doi: 10.1007/s12530-016-9168-2.

[168] G. Hulten, L. Spencer, and P. Domingos, “Mining time-changing data streams,” *Proc. Seventh ACM SIGKDD Int. Conf. Knowl.*

Discov. Data Min., vol. 18, pp. 97–106, 2001, doi: 10.1145/502512.502529.

[169] L. I. Kuncheva and I. Žliobait, “On the window size for classification in changing environments,” *Intell. Data Anal.*, vol. 13, no. 6, pp. 861–872, 2009, doi: 10.3233/IDA-2009-0397.

[170] A. Akbar, A. Khan, F. Carrez, and K. Moessner, “Predictive analytics for complex IoT data streams,” *IEEE Internet Things J.*, vol. 4, no. 5, pp. 1571–1582, Oct. 2017, doi: 10.1109/JIOT.2017.2712672.

[171] C. Alippi and M. Roveri, “Just-in-time adaptive classifiers - Part I: Detecting nonstationary changes,” *IEEE Trans. Neural Networks*, vol. 19, no. 7, pp. 1145–1153, 2008, doi: 10.1109/TNN.2008.2000082.

[172] F. G. d. Costa, F. S. L. G. Duarte, R. M. M. Vallim, and R. F. d. Mello, “Multidimensional surrogate stability to detect data stream concept drift,” *Expert Syst. Appl.*, vol. 87, pp. 1339–1351, 2017, doi: 10.1016/j.eswa.2017.06.005.

[173] W. Liu, H. Zhang, Z. Ding, Q. Liu, and C. Zhu, “A comprehensive active learning method for multiclass imbalanced data streams with concept drift,” *Knowledge-Based Syst.*, vol. 215, 2021, doi: 10.1016/j.knosys.2021.106778.

[174] A. Bouchachia, “Incremental learning with multi-level adaptation,” *Neurocomputing*, vol. 74, no. 11, pp. 1785–1799, 2011, doi: 10.1016/j.neucom.2010.12.017.

[175] I. Katakis, G. Tsoumakas, and I. Vlahavas, “Tracking recurring contexts using ensemble classifiers: An application to email filtering,” *Knowl. Inf. Syst.*, vol. 22, no. 3, pp. 371–391, 2010, doi: 10.1007/s10115-009-0206-2.

[176] J. Shan, H. Zhang, W. Liu, and Q. Liu, “Online Active Learning Ensemble Framework for Drifted Data Streams,” *IEEE Trans. Neural Networks Learn. Syst.*, vol. 30, no. 2, pp. 486–498, 2019, doi: 10.1109/TNNLS.2018.2844332.

[177] I. Zliobaite, A. Bifet, B. Pfahringer, and G. Holmes, “Active learning with drifting streaming data,” *IEEE Trans. Neural Networks Learn. Syst.*, vol. 25, no. 1, pp. 27–39, 2014, doi: 10.1109/TNNLS.2012.2236570.

[178] H. Purwins *et al.*, “Regression methods for prediction of PECVD silicon nitride layer thickness,” 2011, doi: 10.1109/CASE.2011.6042426.

[179] A. Ferreira, A. Roussy, and L. Conde, “Virtual Metrology Models for predicting physical measurement in semiconductor manufacturing,” 2009, doi: 10.1109/ASMC.2009.5155973.

[180] M. H. Hung, T. H. Lin, F. T. Cheng, and R. C. Lin, “A novel virtual metrology scheme for predicting CVD thickness in semiconductor manufacturing,” *IEEE/ASME Trans. Mechatronics*, vol. 12, no. 3, pp. 308–316, 2007, doi: 10.1109/TMECH.2007.897275.

[181] S. K. S. Fan and Y. J. Chang, “An integrated advanced process control framework using run-to-run control, virtual metrology and fault detection,” *J. Process Control*, vol. 23, no. 7, pp. 933–942, 2013, doi: 10.1016/j.jprocont.2013.03.013.

[182] G. A. Susto and A. Beghi, “A virtual metrology system based on least angle regression and statistical clustering,” *Appl. Stoch. Model. Bus. Ind.*, vol. 29, no. 4, pp. 362–376, 2013, doi: 10.1002/asmb.1948.

[183] B. Kim and S. J. Hong, “In-situ virtual metrology for the silicon-dioxide etch rate by using optical emission spectroscopy data,” *J. Korean Phys. Soc.*, vol. 65, no. 2, pp. 168–175, 2014, doi: 10.3938/jkps.65.168.

[184] Y. Jang *et al.*, “Characteristics of a plasma information variable in phenomenology-based, statistically-tuned virtual metrology to predict silicon dioxide etching depth,” *Curr. Appl. Phys.*, 2019, doi: 10.1016/j.cap.2019.06.001.

[185] G. Roeder, S. Winzer, M. Schellenberger, S. Jank, and L. Pfitzner, “Feasibility Evaluation of Virtual Metrology for the Example of a Trench Etch Process,” vol. 27, no. 3, pp. 327–334, 2014.

[186] S. Pampuri, G. A. Susto, J. Wan, A. Johnston, P. O’Hara, and S. McLoone, “Insight extraction for semiconductor manufacturing processes,” *IEEE Int. Conf. Autom. Sci. Eng.*, vol. 2014-Janua, pp. 786–791, 2014, doi: 10.1109/CoASE.2014.6899415.

[187] A. U. Haq and D. Djurdjanovic, “Virtual metrology concept for predicting defect levels in semiconductor manufacturing,” *Procedia CIRP*, vol. 57, pp. 580–584, 2016, doi: 10.1016/j.procir.2016.11.100.

[188] A. Bleakie and D. Djurdjanovic, "Growing Structure Multiple Model System for Quality Estimation in Manufacturing Processes," *IEEE Trans. Semicond. Manuf.*, vol. 29, no. 2, pp. 79–97, 2016, doi: 10.1109/TSM.2016.2533159.

[189] J. C. H. Pan and D. H. E. Tai, "Implementing virtual metrology for in-line quality control in semiconductor manufacturing," *Int. J. Syst. Sci.*, 2009, doi: 10.1080/00207720802645204.

[190] P. Prakash, A. Schirru, P. Hung, and S. Mcloone, "MSC-Clustering and Forward Stepwise Regression for virtual metrology in highly correlated input spaces," *2012 SEMI Adv. Semicond. Manuf. Conf.*, pp. 45–50, 2012, doi: 10.1109/ASMC.2012.6212866.

[191] B. Lenz and C. Leicht, "Development of Smart Feature Selection for Advanced Virtual Metrology," pp. 145–150, 2014.

[192] T. H. Lin, M. H. Hung, R. C. Lin, and F. T. Cheng, "A virtual metrology scheme for predicting CVD thickness in semiconductor manufacturing," 2006, doi: 10.1109/ROBOT.2006.1641849.

[193] B. Q. Sheng, R. B. Zhang, T. H. Pan, and F. Tan, "Integrating MANCOVA and IMA (1, 1) approximation to enhance prediction accuracy for high-mix products manufacturing process," *Int. J. Adv. Comput. Technol.*, 2012, doi: 10.4156/ijact.vol4.issue5.1.

[194] T. H. Lin, F. T. Cheng, A. J. Ye, W. M. Wu, and M. H. Hung, "A novel key-variable sifting algorithm for virtual metrology," 2008, doi: 10.1109/ROBOT.2008.4543768.

[195] J. Feng, X. Jia, F. Zhu, J. Moyne, J. Iskandar, and J. Lee, "An online virtual metrology model with sample selection for the tracking of dynamic manufacturing processes with slow drift," *IEEE Trans. Semicond. Manuf.*, vol. 32, no. 4, pp. 574–582, 2019, doi: 10.1109/TSM.2019.2942768.

[196] S. Kang, D. Kim, and S. Cho, "Efficient Feature Selection-Based on Random Forward Search for Virtual Metrology Modeling," vol. 29, no. 4, pp. 391–398, 2016.

[197] T. H. Lin, F. T. Cheng, W. M. Wu, C. A. Kao, A. J. Ye, and F. C. Chang, "NN-based key-variable selection method for enhancing virtual metrology accuracy," *IEEE Trans. Semicond. Manuf.*, vol. 22, no. 1, pp. 204–211, 2009, doi: 10.1109/TSM.2008.2011185.

[198] S. Hwang, M. K. Jeong, and B. J. Yum, "Robust relevance vector machine with variational inference for improving virtual metrology accuracy," *IEEE Trans. Semicond. Manuf.*, 2014, doi: 10.1109/TSM.2013.2286498.

[199] S. Chen, T. Pan, Z. Li, and S. S. Jang, "Statistical key variable analysis and model-based control for improvement performance in a deep reactive ion etching process," *J. Semicond.*, 2012, doi: 10.1088/1674-4926/33/6/066002.

[200] S. Lynn, J. Ringwood, E. Ragnoli, S. McLoone, and N. MacGearailt, "Virtual metrology for plasma etch using tool variables," 2009, doi: 10.1109/ASMC.2009.5155972.

[201] S. Lynn, J. Ringwood, and N. Macgearailt, "Gaussian Process Regression for Virtual Metrology of Plasma Etch," 2010.

[202] P. Kang, D. Kim, and S. Cho, "Semi-supervised support vector regression based on self-training with label uncertainty : An application to virtual metrology in semiconductor manufacturing," *Expert Syst. Appl.*, vol. 51, pp. 85–106, 2016, doi: 10.1016/j.eswa.2015.12.027.

[203] S. Pampuri, A. Schirru, G. A. Susto, C. De Luca, A. Beghi, and G. De Nicolao, "Multistep Virtual Metrology Approaches for Semiconductor Manufacturing Processes," no. 1, 2012.

[204] S. ARIMA, H. BU, Y. ONUMA, K. HANDA, and A. NAGATA, "Feature Extractions from a High-dimension Low-samples Data for Multi-dimension Virtual Metrology," *IEEE Trans. Semicond. Manuf.*, 2018, doi: 10.1109/TSM.2017.2768241.

[205] C. Park and S. B. Kim, "Virtual metrology modeling of time-dependent spectroscopic signals by a fused lasso algorithm," *J. Process Control*, vol. 42, pp. 51–58, 2016, doi: 10.1016/j.jprocont.2016.04.002.

[206] M. Maggipinto, G. A. Susto, F. Zocco, and S. McLoone, "What are the most informative data for virtual metrology? a use case on multi-stage processes fault prediction," 2019, doi: 10.1109/COASE.2019.8842942.

[207] A. A. Popov, D. V. Bilevich, A. S. Salnikov, A. A. Kalentyev, and D. V. Garays, “Plasma-Chemical Etching Process Behavioral Models Based on Tree Ensembles and Neural Network,” 2018, doi: 10.1109/APEIE.2018.8546139.

[208] Y. Liu and X. Li, “Predictive Modeling for Advanced Virtual Metrology: A Tree-Based Approach,” 2018, doi: 10.1109/ETFA.2018.8502480.

[209] B. Lenz, B. Barak, J. Mührwald, and C. Leicht, “Virtual metrology in semiconductor manufacturing by means of predictive machine learning models,” *Proc. - 2013 12th Int. Conf. Mach. Learn. Appl. ICMLA 2013*, vol. 2, pp. 174–177, 2013, doi: 10.1109/ICMLA.2013.186.

[210] T. H. Pan, B. Q. Sheng, D. S. H. Wong, and S. S. Jang, “A virtual metrology model based on recursive canonical variate analysis with applications to sputtering process,” *J. Process Control*, 2011, doi: 10.1016/j.jprocont.2011.04.004.

[211] M. Terzi, C. Masiero, A. Beghi, M. Maggipinto, and G. A. Susto, “Deep learning for virtual metrology: Modeling with optical emission spectroscopy data,” *RTSI 2017 - IEEE 3rd Int. Forum Res. Technol. Soc. Ind. Conf. Proc.*, no. Ml, 2017, doi: 10.1109/RTSI.2017.8065905.

[212] S. Park, S. Jeong, Y. Jang, S. Ryu, H. J. Roh, and G. H. Kim, “Enhancement of the Virtual Metrology Performance for Plasma-Assisted Oxide Etching Processes by Using Plasma Information (PI) Parameters,” *IEEE Trans. Semicond. Manuf.*, 2015, doi: 10.1109/TSM.2015.2432576.

[213] G. A. Susto, A. Beghi, and C. De Luca, “A Virtual Metrology system for predicting CVD thickness with equipment variables and qualitative clustering,” 2011, doi: 10.1109/ETFA.2011.6059209.

[214] J. M. Gu and S. J. Hong, “Virtual metrology for TSV etch depth measurement using optical emission spectroscopy,” *2015 IEEE Electr. Des. Adv. Packag. Syst. Symp.*, pp. 27–30, 2015, doi: 10.1109/EDAPS.2015.7383700.

[215] G. A. Susto, M. Maggipinto, F. Zocco, and S. McLoone, “Induced Start Dynamic Sampling for Wafer Metrology Optimization,” *IEEE Trans. Autom. Sci. Eng.*, 2020, doi: 10.1109/TASE.2019.2929193.

[216] K. Hyun, T. F. Edgar, K. Song, G. Choi, H. Ku, and C. Han, “An effective procedure for sensor variable selection and utilization in plasma etching for semiconductor manufacturing,” *Comput. Chem. Eng.*, vol. 61, pp. 20–29, 2014, doi: 10.1016/j.compchemeng.2013.09.016.

[217] S. Park *et al.*, “Application of PI-VM for management of the metal target plasma etching processes in OLED display manufacturing,” *Plasma Phys. Control. Fusion*, 2019, doi: 10.1088/1361-6587/aae2db.

[218] D. Kurz, J. Pilz, A. Schirru, S. Pampuri, and C. De Luca, “A sampling decision system for semiconductor manufacturing - Relying on virtual metrology and actual measurements,” 2015, doi: 10.1109/WSC.2014.7020109.

[219] D. Sharma, H. Armer, and J. Moyne, “A Comparison of Data Mining Methods for Yield Modeling , Chamber Matching and Virtual Metrology Applications,” pp. 231–236, 2012.

[220] T. Tsuda *et al.*, “Advanced Semiconductor Manufacturing Using Big Data,” *IEEE Trans. Semicond. Manuf.*, 2015, doi: 10.1109/TSM.2015.2445320.

[221] H. C. Yang, H. Tieng, and F. T. Cheng, “Automatic virtual metrology for wheel machining automation,” *Int. J. Prod. Res.*, vol. 54, no. 21, pp. 6367–6377, 2016, doi: 10.1080/00207543.2015.1109724.

[222] W. Tsai, S. Member, F. Cheng, W. Wu, S. Member, and T. Lin, “Developing a Dual-Stage Indirect Virtual Metrology Architecture,” pp. 2107–2112, 2010.

[223] S. Pampuri, A. B. Johnston, Jian Wan, P. G. O’Hara, and S. McLoone, “On Regression Methods for Virtual Metrology in Semiconductor Manufacturing,” *25th IET Irish Signals Syst. Conf. 2014 2014 China-irel. Int. Conf. Inf. Communities Technol. (ISSC 2014/CIICT 2014)*, pp. 380–385, 2014, doi: 10.1049/cp.2014.0718.

[224] H. Roh *et al.*, “Development of the Virtual Metrology for the Nitride Thickness in Multi-Layer Plasma-Enhanced Chemical Vapor Deposition Using Plasma-Information Variables,” vol. 31, no. 2, pp. 232–241, 2018.

[225] K. Suthar, D. Shah, J. Wang, and Q. Peter He, “Feature-based Virtual Metrology for Semiconductor Manufacturing,” in *13*

International Symposium on Process Systems Engineering (PSE 2018), vol. 44, M. R. Eden, M. G. Ierapetritou, and G. P. B. T.-C. A. C. E. Towler, Eds. Elsevier, 2018, pp. 2083–2088.

[226] K. Suthar, D. Shah, J. Wang, and Q. P. He, “Next-generation virtual metrology for semiconductor manufacturing : A feature-based framework,” *Comput. Chem. Eng.*, vol. 127, pp. 140–149, 2019, doi: 10.1016/j.compchemeng.2019.05.016.

[227] T. Tanaka and S. Yasuda, “Prediction and control of transistor threshold voltage by virtual metrology (virtual PCM) using equipment data,” *IEEE Trans. Semicond. Manuf.*, 2013, doi: 10.1109/TSM.2013.2269147.

[228] H. K. Lee, K. H. Baek, and K. Shin, “Resolving critical dimension drift over time in plasma etching through virtual metrology based wafer-to-wafer control.”

[229] A. A. Khan, J. R. Moyne, and D. M. Tilbury, “Virtual metrology and feedback control for semiconductor manufacturing processes using recursive partial least squares,” *J. Process Control*, vol. 18, no. 10, pp. 961–974, 2008, doi: 10.1016/j.jprocont.2008.04.014.

[230] A. Ferreira *et al.*, “Virtual metrology models for predicting avera PECVD oxide film thickne,” *ASMC (Advanced Semicond. Manuf. Conf. Proc.*, no. December 2015, 2011, doi: 10.1109/ASMC.2011.5898187.

[231] K. Olson and J. Moyne, “Adaptive Virtual Metrology applied to a CVD process,” *ASMC (Advanced Semicond. Manuf. Conf. Proc.*, pp. 353–358, 2010, doi: 10.1109/ASMC.2010.5551478.

[232] W. M. Wu, F. T. Cheng, T. H. Lin, D. L. Zeng, and J. F. Chen, “Selection schemes of dual virtual-metrology outputs for enhancing prediction accuracy,” *IEEE Trans. Autom. Sci. Eng.*, 2011, doi: 10.1109/TASE.2010.2089451.

[233] S. Umeda, K. Nogi, D. Shiraishi, and A. Kagoshima, “Advanced process control using virtual metrology to cope with etcher condition change,” *IEEE Trans. Semicond. Manuf.*, 2019, doi: 10.1109/TSM.2019.2938546.

[234] M. A. Jebri, G. Graton, E. M. El Adel, M. Ouladsine, and J. Pinaton, “Virtual metrology on Chemical Mechanical Planarization process based on Just-In-Time Learning,” *2016 5th Int. Conf. Syst. Control. ICSC 2016*, pp. 169–174, 2016, doi: 10.1109/ICoSC.2016.7507082.

[235] B. Q. Sheng, R. B. Zhang, T. H. Pan, and F. Tan, “Dynamic canonical correlation analysis for online quality Estimation in an Industrial Sputtering Process,” *Int. J. Digit. Content Technol. its Appl.*, 2012, doi: 10.4156/jdcta.vol6.issue7.12.

[236] B. Q. Sheng and T. H. Pan, “Virtual metrology algorithm for TFT-LCD manufacutring process,” 2011, doi: 10.1109/FSKD.2011.6019961.

[237] J. Wan and S. McLoone, “Gaussian Process Regression for Virtual Metrology-Enabled Run-to-Run Control in Semiconductor Manufacturing,” *IEEE Trans. Semicond. Manuf.*, vol. 31, no. 1, pp. 12–21, 2018, doi: 10.1109/TSM.2017.2768241.

[238] T. Darwin, R. Garnett, and D. Djurdjanovic, “Gaussian process regression for virtual metrology of microchip quality and the resulting selective sampling scheme,” in *Lecture Notes in Mechanical Engineering*, 2018.

[239] B. Lenz and B. Barak, “Data Mining and Support Vector Regression Machine Learning in Semiconductor Manufacturing to improve virtual metrology,” 2013, doi: 10.1109/HICSS.2013.163.

[240] D. Kim, C. Lee, S. Hwang, and M. K. Jeong, “A robust support vector regression with a linear-log concave loss function,” *J. Oper. Res. Soc.*, 2016, doi: 10.1057/jors.2015.32.

[241] A. Chouichi, J. Blue, C. Yugma, and F. Pasqualini, “Chamber Mismatching Calibration with Adjusted Run-to-Run Regulation in Semiconductor Manufacturing,” 2019, doi: 10.1109/SMILE45626.2019.8965283.

[242] C. Y. C. Jonathan and F. T. Cheng, “Application development of virtual metrology in semiconductor industry,” 2005, doi: 10.1109/IECON.2005.1568891.

[243] H.-C. Yang, C.-H. Huang, M. Adnan, C.-H. Hsu, C.-H. Lin, and F.-T. Cheng, “An Online AM Quality Estimation Architecture From Pool to Layer,” *IEEE Trans. Autom. Sci. Eng.*, 2020, doi: 10.1109/tase.2020.3012622.

[244] F. Cheng, S. Member, H. Huang, and C. Kao, “Dual-Phase Virtual Metrology Scheme,” vol. 20, no. 4, pp. 566–571, 2007.

[245] Y. C. Su, F. T. Cheng, G. W. Huang, M. H. Hung, and T. Yang, “A quality prognostics scheme for semiconductor and TFT-LCD manufacturing processes,” 2004, doi: 10.1109/iecon.2004.1431887.

[246] Y. T. Huang and F. T. Cheng, “Automatic data quality evaluation for the AVM system,” *IEEE Trans. Semicond. Manuf.*, vol. 24, no. 3, pp. 445–454, 2011, doi: 10.1109/TSM.2011.2154910.

[247] S. Kang, “Joint modeling of classification and regression for improving faulty wafer detection in semiconductor manufacturing,” *J. Intell. Manuf.*, 2020, doi: 10.1007/s10845-018-1447-2.

[248] F. T. Cheng and Y. C. Chiu, “Applying the Automatic Virtual Metrology system to obtain tube-to-tube control in a PECVD tool,” *IIE Trans. (Institute Ind. Eng.*, 2013, doi: 10.1080/0740817X.2012.725507.

[249] H. Yang, C. Cheng, T. Su, C. Jan, W. Wu, and M. Wu, “Intelligent Sensing Unit for Estimation Roughness of Electrical Discharge Machining,” vol. 7, no. 3, pp. 125–131, 2017, doi: 10.5875/ausmt.v7i3.1431.

[250] C. A. Kao, F. T. Cheng, W. M. Wu, F. W. Kong, and H. H. Huang, “Run-to-run control utilizing virtual metrology with reliance index,” *IEEE Trans. Semicond. Manuf.*, 2013, doi: 10.1109/TSM.2012.2228243.

[251] Y. T. Huang, H. C. Huang, F. T. Cheng, T. S. Liao, and F. C. Chang, “Automatic virtual metrology system design and implementation,” 2008, doi: 10.1109/COASE.2008.4626524.

[252] Y. T. Chen, H. C. Yang, and F. T. Cheng, “Multivariate simulation assessment for virtual metrology,” *Proc. - IEEE Int. Conf. Robot. Autom.*, vol. 2006, no. May, pp. 1048–1053, 2006, doi: 10.1109/ROBOT.2006.1641848.

[253] L. van Dijk, F. Hasibi, M. Larrañaga, A. Pastol, A. Lam, and R. J. F. van Haren, “Smart implant-layer overlay metrology to enable fab cycle time reduction,” 2019, doi: 10.1117/12.2515185.

[254] M. F. Abdullah, M. K. Osman, N. M. Somari, A. I. C. Ani, S. P. R. S. Appanan, and L. K. Hooi, “In-situ particle monitor using virtual metrology system for measuring particle contamination during plasma etching process,” *Proc. - 6th IEEE Int. Conf. Control Syst. Comput. Eng. ICCSCE 2016*, no. November, pp. 507–511, 2017, doi: 10.1109/ICCSCE.2016.7893629.

[255] Y.-C. Su, M.-H. Hung, F.-T. Cheng, and Y.-T. Chen, “A Processing Quality Prognostics Scheme for Plasma Sputtering in TFT-LCD Manufacturing,” *IEEE Trans. Semicond. Manuf.*, vol. 19, no. 2, pp. 183–194, May 2006, doi: 10.1109/TSM.2006.873514.

[256] H.-G. Lee *et al.*, “Virtual overlay metrology for fault detection supported with integrated metrology and machine learning,” 2015, doi: 10.1117/12.2085475.

[257] W. T. Yang, J. Blue, A. Roussy, M. S. Reis, and J. Pinaton, “Virtual metrology modeling based on Gaussian Bayesian network,” 2019, doi: 10.1109/WSC.2018.8632485.

[258] W. T. Yang, J. Blue, A. Roussy, J. Pinaton, and M. S. Reis, “A Structure Data-Driven Framework for Virtual Metrology Modeling,” *IEEE Trans. Autom. Sci. Eng.*, 2020, doi: 10.1109/TASE.2019.2941047.

[259] Y. Su, T. Lin, and F. Cheng, “Accuracy and Real-Time Considerations for Implementing Various Virtual Metrology Algorithms,” vol. 21, no. 3, pp. 426–434, 2008.

[260] W. Tian, F. Zhao, Z. Sun, S. Shang, X. Mei, and G. Chen, “A Novel Performance Prediction Model for the Machining Process Based on the Interval Type-2 Fuzzy Neural Network,” *Math. Probl. Eng.*, 2020, doi: 10.1155/2020/5740362.

[261] L. Puggini and S. McLoone, “Extreme learning machines for virtual metrology and etch rate prediction,” 2015, doi: 10.1109/ISSC.2015.7163771.

[262] Y. Zhu, J. He, and R. D. Lawrence, “A general framework for predictive tensor modeling with domain knowledge,” *Data Min. Knowl. Discov.*, 2015, doi: 10.1007/s10618-014-0392-8.

[263] N. M. Somari *et al.*, “Particles contaminations detection during plasma etching process by using k-nearest neighbors and Fuzzy k-nearest neighbors,” *Proc. - 6th IEEE Int. Conf. Control Syst. Comput. Eng. ICCSCE 2016*, no. November, pp. 512–516, 2017, doi: 10.1109/ICCSCE.2016.7893630.

[264] M. C. Method, M. C. Mcmc, and A. C. Process, “Multiparametric Virtual Metrology Model Building by Job-Shop Data Fusion Using a Markov Chain,” vol. 26, no. 3, pp. 319–327, 2013.

[265] Q. P. He, J. Wang, H. E. Gilicia, J. D. Stuber, and B. S. Gill, “Statistics pattern analysis based virtual metrology for plasma etch processes,” 2012, doi: 10.1109/acc.2012.6315618.

[266] H. Jun, J. Shin, S. Kim, and H. Choi, “Development of a virtual metrology method using plasma harmonics analysis,” *AIP Adv.*, vol. 7, no. 7, 2017, doi: 10.1063/1.4993282.

[267] S. Do Kim, J. S. Kim, B. M. Mun, and S. J. Bae, “A Bayesian virtual metrology for quality inspection of mobile repeater systems,” *Manag. Prod. Eng. Rev.*, 2016, doi: 10.1515/mper-2016-0035.

[268] Q. Gong, G. Yang, C. Pan, and Y. Chen, “Performance analysis of single EWMA controller subject to metrology delay under dynamic models,” *IISE Trans.*, 2018, doi: 10.1080/24725854.2017.1386338.

[269] K. Hui and J. Mou, “Performance assessment of virtual metrology in APC applications for the viability of sampling reductions,” 2014, doi: 10.1109/CoASE.2014.6899410.

[270] G. A. Susto, A. Schirru, S. Pampuri, G. De Nicolao, and A. Beghi, “An information-theory and Virtual Metrology-based approach to Run-to-Run semiconductor manufacturing control,” *IEEE Int. Conf. Autom. Sci. Eng.*, pp. 358–363, 2012, doi: 10.1109/CoASE.2012.6386416.

[271] R. D. Molina-Arredondo, “Process control adjustment with feedback controller,” in *Lean Manufacturing in the Developing World: Methodology, Case Studies and Trends from Latin America*, 2014.

[272] Q. S. Gong, G. K. Yang, C. C. Pan, and M. S. Lee, “Stability and Control Performance Analysis of Double EWMA Controller with Metrology Delay,” *IEEE Trans. Semicond. Manuf.*, 2016, doi: 10.1109/TSM.2015.2504951.

[273] H. Kim, J. H. Park, and K. S. Lee, “A Kalman filter-based R2R control system with parallel stochastic disturbance models for semiconductor manufacturing processes,” *J. Process Control*, 2014, doi: 10.1016/j.jprocont.2014.09.001.

[274] K. Tamaki and S. Kaneko, “Multiparametric virtual metrology model building by job-shop data fusion using a markov chain monte carlo method,” *IEEE Trans. Semicond. Manuf.*, 2013, doi: 10.1109/TSM.2013.2257897.

[275] S. Park *et al.*, “Cause analysis of the faults in HARC etching processes by using the PI-VM model for OLED display manufacturing,” *Plasma Process. Polym.*, 2019, doi: 10.1002/ppap.201900030.

[276] C. Nguyen, X. Li, S. Blanton, and X. Li, “Efficient Classification via Partial Co-Training for Virtual Metrology,” 2020, doi: 10.1109/etfa46521.2020.9212012.

[277] S. McLoone, A. Johnston, and G. A. Susto, “A Methodology for Efficient Dynamic Spatial Sampling and Reconstruction of Wafer Profiles,” *IEEE Trans. Autom. Sci. Eng.*, 2018, doi: 10.1109/TASE.2017.2786213.

[278] F. T. Cheng, Y. S. Hsieh, J. W. Zheng, S. M. Chen, R. X. Xiao, and C. Y. Lin, “A scheme of high-dimensional key-variable search algorithms for yield improvement,” *IEEE Robot. Autom. Lett.*, 2017, doi: 10.1109/LRA.2016.2584143.

[279] H. Lee, C. O. Kim, H. H. Ko, and M. K. Kim, “Yield Prediction Through the Event Sequence Analysis of the Die Attach Process,” *IEEE Trans. Semicond. Manuf.*, 2015, doi: 10.1109/TSM.2015.2487540.

[280] F. Cheng, H. Tieng, S. Member, and H. Yang, “Industry 4 . 1 for Wheel Machining Automation,” *IEEE Robot. Autom. Lett.*, vol. 1, no. 1, pp. 332–339, 2016, doi: 10.1109/LRA.2016.2517208.

[281] Y. C. Su, F. T. Cheng, M. H. Hung, and H. C. Huang, “Intelligent prognostics system design and implementation,” *IEEE Trans. Semicond. Manuf.*, vol. 19, no. 2, pp. 195–206, 2006, doi: 10.1109/TSM.2006.873512.

[282] J. Moyne and B. Schulze, “Yield management enhanced advanced process control system (YMeAPC) part I: Description and case study of feedback for optimized multiprocess control,” 2010, doi: 10.1109/TSM.2010.2041294.

Annex 1: State of the art citation tables

Table 19: outlier detection methods citation table.

Manual outlier detection	[30], [142], [178]–[180]
Statistically-based methods	[43], [51], [80], [118], [181], [182]

Table 20: Dimensionality reduction citation table.

Feature selection: Filter methods	
Pearson covariance	[142], [144], [183], [184]
Input covariance	[182], [185]
ANOVA	[178], [120]
Feature selection: Wrapper methods	
Forward selection	[95], [96], [106], [178], [182], [186]–[190]
Backward elimination	[121], [178], [191]
Stepwise	[34], [39], [40], [45], [51], [88], [92], [103], [105], [118], [125], [144], [179], [180], [183], [184], [192]–[201]
Genetic algorithm	[33]–[35], [39]–[41], [118], [121], [202]
Others	[52], [92], [184], [193]
Feature selection: Embedded methods	
Lasso	[53], [100], [131], [182], [190], [203]–[206]
Tree approaches	[40], [57], [94], [185], [207]–[209]
Feature extraction	
PCA	[33], [39], [40], [45], [53], [88], [89], [92], [106], [118], [131], [135], [142], [179], [180], [182]–[184], [200], [201], [204], [206], [210]–[218]
PLS	[43], [52], [76], [77], [80], [99], [106], [113], [120], [125], [142], [144], [181], [183], [184], [189], [196], [210], [213], [214], [219]–[230]
CNN	[50], [54], [60], [62], [119], [132], [135], [211]
Autoencoder	[131], [132]

Table 21: Quality estimation citation table.

Linear methods	
Partial least squares	[43], [52], [76], [77], [80], [99], [113], [125], [132], [142], [144], [178], [181], [189], [196], [213], [214], [219]–[222], [224]–[226], [228]–[231]
Multiple linear regression	[15], [34], [35], [75], [77], [95], [111], [112], [124], [183], [184], [189], [190], [199], [232]–[234]
Lasso	[100], [182], [203], [205], [206]
Others	[87], [188], [210], [235], [236]

Kernel method	
Gaussian process regression	[53], [54], [64], [74], [88], [96], [201], [223], [237], [238]
Support vector regression	[35], [39], [120], [121], [191], [198], [202], [209], [239]–[241]
Others	[33], [101], [187], [204], [212]
Ensemble methods	
All	[30], [56], [57], [70], [94], [185], [207]–[209]
Neural networks	
Multi-layer perceptron	[21], [28], [52], [63], [70], [76], [91], [93], [105], [111]–[113], [118], [125], [144], [179], [180], [182], [192], [194], [197], [200], [207], [209], [213], [214], [221], [222], [232], [242]–[256]
Convolutional neural networks	[50], [60], [62], [119], [135], [211]
Bayesian neural networks	[48], [61], [257], [258]
Recurrent neural networks	[50], [98], [110], [259]
Others	[51], [85], [86], [260], [261]
Other methods	
All	[40], [82], [92], [103], [123], [188], [193], [195], [227], [262]–[267]

Table 22: Update approach citation table.

Moving window	[40], [64], [76], [77], [80], [81], [88], [110], [125], [142], [181], [185], [188], [195], [221], [229], [231], [233], [235], [236], [238], [246], [248], [250], [251], [259]
Just-in-time learning	[15], [43], [53], [75], [82], [234]

Table 23: Control approach citation table.

EWMA	[40], [75], [103], [106], [237], [248], [250], [268]–[271]
dEWMA	[15], [61], [99], [128], [181], [229], [271], [272]
MPC	[105]
Others	[62], [87], [104], [106], [273]

Table 24: Applications citation table.

Semiconductor manufacturing	[15], [22], [26]–[30], [33]–[35], [38]–[41], [43], [45], [48], [50], [51], [53], [54], [56], [57], [60]–[62], [64], [69], [70], [72]–[77], [80], [82], [85]–[89], [91]–[94], [97]–[101], [103]–[106], [109], [110], [115], [116], [118]–[120], [122]–[124], [131], [132], [135], [142], [144], [178]–[185], [187]–[196], [198]–[218], [220], [223]–[231], [233]–[239], [241], [242], [244]–[248], [251], [253]–[259], [261]–[263], [265], [266], [268]–[270], [273]–[279]
Traditional manufacturing	[42], [52], [108], [113], [125], [127], [128], [221], [243], [249], [260], [280]
Others	[44], [95], [112], [121], [232], [240], [267], [271]

Table 25: Semiconductor operations citation table.

CMP	[15], [50], [51], [57], [61], [62], [64], [75], [85], [195], [220], [234], [237], [250], [257], [258], [274]
CVD	[33], [53], [54], [72], [87], [98]–[101], [110], [120], [179], [180], [182], [191], [192], [198], [203], [204], [209], [213], [215], [218], [223], [230], [231], [236], [239], [242], [244], [246], [251], [259], [281]
PECVD	[27], [41], [69], [76], [178], [188], [224], [238], [248], [262]
Plasma etching	[39], [60], [73], [74], [77], [80], [86], [88], [89], [94], [103]–[106], [118], [119], [131], [135], [142], [183]–[185], [190], [193], [196], [200], [201], [205]–[207], [211], [212], [214], [216], [217], [225], [226], [228], [229], [233], [235], [254], [255], [261], [263], [265], [266], [282]
Photolithography	[31], [38], [70], [93], [99], [246], [247]
Others	[30], [56], [115], [123], [124], [199], [210], [275]

Table 26: Traditional manufacturing operations citation table.

Milling	[125], [128]
Turning	[32], [108], [221]
Drilling	[32], [127], [221]
Spark machining	[42], [249], [260]
Additive manufacturing	[52], [243]
Carbon fiber manufacturing	[113]

Curriculum Vitae

Paul-Arthur Dreyfus

Avenue du Tir Fédéral 29 – 1024 Ecublens – Suisse

+41 78 648 80 86 – Paul.dreyfus@epfl.ch

My LinkedIn profile

Swiss

French

28 yrs

Doctor ès sciences on Industry 4.0.

Creativity, innovation, management and efficiency

Education

DOCTORATE : "Virtual metrology applied on milling process" — Lausanne, 2018 – Feb. 2022

École Polytechnique Fédérale de Lausanne,
Doctoral school of robotics control and intelligent systems ICT for sustainable manufacturing group and laboratory of artificial intelligence

Virtual metrology is one of the **Industry 4.0.** and **zero-defect-manufacturing** pillar. It has proven its efficiency in multiple industries. It represents a major industrial challenge. My thesis aims at implementing the virtual metrology paradigm on **machine tools**. A focus is also made on the **long-term sustainability** of such technology.

MASTER and BACHELOR in mechanical engineering — Lausanne, 2012 - 2017

École Polytechnique Fédérale de Lausanne
Bachelor and Master in Mechanical engineering, automatic and mechatronic 5.35/6, Minor in Management of technologies and entrepreneurship.

Industrial experiences

WELL-KNOWN WATCHMAKING MANUFACTURE (The name is confidential) — Geneva, 2018-2022

Industrial doctorate: **project manager** with a serious technical, scientific and industrial experience for 4 years. **Development**, **marketing** and **management** of the project with teams of engineers and technicians. Regular meetings with the board members. Direct personal access to the industrial tools to run the case studies.

ROLEX — Geneva, February – August 2017

Master thesis as a project manager in the industrial robotics field 6/6,
Hardware, software and method enhancement into multiple projects.

CHOPARD — Geneva, Summer 2015

Design of watch's mechanisms in the technical office of the L.U.C.
Design of movement and test machine for watches with complication.

Management experiences

Student project management during doctorate — Lausanne, 2018 - 2022

Creation and management of EPFL students project : **recruitment**, **management**, **education** and **evaluation** of more than 20 master thesis and semester projects.
Very good post-employment evaluations and commentaries from my students.

Creation and presidency of the student's association CUBALIENTE (Latin dances). — Lausanne, 2013 - Now

Creation and management of the association. Hundred of dance events, regrouping 50 to 200 people, and around 50 staffs, weekly and quarterly organized.

Scientific journal publications

1st author publications :

Virtual metrology as an approach for product quality estimation in Industry 4.0: A systematic review and integrative conceptual framework — IJPR 2021
Complete state of the art of virtual metrology, structured thanks to a new framework. Quality estimation algorithms, algorithm long-term sustainability, industrial implementation and more.

Virtual metrology for milling operation: dimensional quality estimation — In review in ESwA 2021-2022
First case study estimating the dimensional quality of a milled part in an industrial environment. Results are encouraging, showing a model's mean absolute error of 13µm.

Virtual metrology for milling operation: dimensional quality estimation enhanced results — In review in ESwA 2021-2022
In this paper, a second case study is done to enhance the results of the first one.

A framework for the long-term sustainability of industry 4.0. data based models — In review in JMS 2021-2022
This paper proposes an innovative framework to deal with drift of industry 4.0. models.

Other publications :

Zero defect manufacturing: state-of-the-art review, shortcomings and future directions in research — IJPR 2019
Complete state of the art of the zero-defect manufacturing paradigm.

The Role of Big Data in the Context of Modeling Design and Operation of Manufacturing Systems — Book by Elsevier 2021
Book chapter : discussion on the impact of big data on modeling, design and operation of manufacturing system.

Awards

Best proposal award — Austin 2019
APMS 2019 International conference Advances in Production Management Systems

Best presentation award — Penang 2019
i3CDE 2019 Computational Design and Engineering

Languages

French : Mother tongue **English** : C1 Level **Spanish** : B1 Level

Computer skills

Machine learning : Model industrialization, Regression, Drift detection, Uncertainty estimation, Model update, Reinforcement learning.

Control theory : PID, MPC, Robust controller.

Programming : Python, Matlab, C/C++, Simulink, Labview, PhP, Mysql.

CAO : Catia, Solidwork, Inventor, Abaqus, Creo.

Office software : Office, Latex, Photoshop, After Effect.

Hobbies

Semi-professional : Latin dance professor and choreographer, First-aid worker.

Sports : Dance, Badminton.

Passions : Dance, Event organization and management, Watches, Automobile sport and mechanics.